Multiple sustained increases in atmospheric oxygen and marine productivity through the Neoproterozoic and Paleozoic Eras

Richard Stockey

r.g.stockey@soton.ac.uk

University of Southampton https://orcid.org/0000-0001-5541-7987

Devon Cole

Georgia Institute of Technology https://orcid.org/0000-0002-5669-3817

Una Farrell

Trinity College Dublin

Heda Agić

Durham University

Thomas Boag

Princeton University

Jochen Brocks

The Australian National University https://orcid.org/0000-0002-8430-8744

Donald Canfield

University of Southern Denmark

Meng Cheng

Chengdu University of Technology

Peter Crockford

Carleton University

Huan Cui

Mississippi State University https://orcid.org/0000-0003-0705-3423

Tais Dahl

University of Copenhagen https://orcid.org/0000-0003-4629-8036

Lucas Mouro

Harvard University https://orcid.org/0000-0001-7829-0683

Keith Dewing

Geological Survey of Canada

Stephen Dornbos

University of Wisconsin-Milwaukee

Joseph Emmings

CGG https://orcid.org/0000-0003-2084-0501
Robert Gaines
Pomona College https://orcid.org/0000-0002-3713-5764
Timothy Gibson
Yale University
Benjamin Gill
Virginia Polytechnic Institute and State University https://orcid.org/0000-0001-7402-0811
Geoffrey Gilleaudeau
George Mason
Karin Goldberg
Kansas State University https://orcid.org/0000-0002-0897-0533
Romain Guilbaud
CNRS https://orcid.org/0000-0001-9701-9522
Galen Halverson
McGill University https://orcid.org/0000-0003-3061-7928
Emma Hammarlund
Lund University
Kalev Hantsoo
Johns Hopkins University
Miles Henderson
The University of Texas Permian Basin
Charles Henderson
University of Calgary
Malcolm Hodgskiss
University of Cambridge
Amber Jarrett
Northern Territory Geological Survey
David Johnston
Harvard University
Pavel Kabanov
Geological Survey of Canada
Julien Kimmig
Staatliches Museum für Naturkunde Karlsruhe https://orcid.org/0000-0001-8032-4272
Andrew Knoll
Harvard University https://orcid.org/0000-0003-1308-8585
Marcus Kunzmann
Fortescue Metals Group Ltd
Matthew LeRoy
Virginia Polytechnic Institute and State University

Chao Li
Chengdu University of Technology https://orcid.org/0000-0001-9861-661X
David Loydell
University of Portsmouth https://orcid.org/0000-0003-3967-0047
Francis Macdonald
University of California, Santa Barbara https://orcid.org/0000-0001-8416-4894
Joseph Magnall
41GFZ German Research Centre for Geosciences
N. Tanner Mills
The University of Texas at Austin
Lawrence Och
Swiss Federal Institute of Aquatic Science and Technology
Brennan O'Connell
University of Cambridge
Anais Pagès
Department of Water and Environmental Regulation
Shanan Peters
University of Wisconsin-Madison https://orcid.org/0000-0002-3346-4317
Susannah Porter
University of California Santa Barbara https://orcid.org/0000-0002-4707-9428
Simon Poulton
University of Leeds https://orcid.org/0000-0001-7621-189X
Samantha Ritzer
Stanford University
Alan Rooney
Yale University https://orcid.org/0000-0002-5023-2606
Shane Schoepfer
Western Carolina University
Emily Smith
Johns Hopkins https://orcid.org/0000-0001-9260-9355
Justin Strauss
Dartmouth College https://orcid.org/0000-0003-3298-3227
Gabriel Uhlein
Federal University of Minas Gerais
Tristan White
Yale University
Rachel Wood
University of Edinburgh https://orcid.org/0000-0002-0165-5987
Christina Woltz

University of California Santa Barbara
Inessa Yurchenko
The University of Texas at Dallas
Noah Planavsky
Yale University https://orcid.org/0000-0001-5849-8508
Erik Sperling
Stanford University

Article

Keywords:

Posted Date: November 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3438740/v1

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Geoscience on July 2nd, 2024. See the published version at https://doi.org/10.1038/s41561-024-01479-1.

Multiple sustained increases in atmospheric oxygen and marine productivity through the Neoproterozoic and Paleozoic Eras

Richard G. Stockey[1, 2*], Devon B. Cole[3], Una C. Farrell[4], Heda Agić[5,6], Thomas H. Boag[7], Jochen J. Brocks[8], Don E. Canfield[9,10], Meng Cheng[11], Peter W. Crockford[12,13], Huan Cui[14], Tais W. Dahl[15], Lucas Del Mouro[16,17], Keith Dewing[18], Stephen Q. Dornbos[19], Joseph F Emmings[20], Robert R. Gaines[21], Timothy M. Gibson[22], Benjamin C Gill[23], Geoffrey J. Gilleaudeau[24], Karin Goldberg[25], Romain Guilbaud[26], Galen Halverson[27], Emma U. Hammarlund[28], Kalev Hantsoo[29], Miles A. Henderson[30], Charles M. Henderson[31], Malcolm S.W. Hodgskiss[32,33], Amber J.M. Jarrett[34], David T. Johnston[35], Pavel Kabanov[18,31], Julien Kimmig[36], Andrew H. Knoll[16], Marcus Kunzmann[37], Matthew A. LeRoy[23], Chao Li[38,39], David K. Loydell[40], Francis A. Macdonald[6], Joseph M. Magnall[41], N. Tanner Mills[42], Lawrence M. Och[43], Brennan O'Connell[33], Anais Pagès[44], Shanan E. Peters[45], Susannah M. Porter[6], Simon W. Poulton[46], Samantha R. Ritzer[2], Alan D. Rooney[22], Shane Schoepfer[47], Emily F. Smith[29], Justin V. Strauss[48], Gabriel Jubé Uhlein[49], Tristan White[22], Rachel A. Wood[50], Christina R. Woltz[2,6], Inessa Yurchenko[51], Noah J. Planavsky[22], Erik A. Sperling[2*]

[1]School of Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, Southampton, UK.

[2]Earth and Planetary Sciences, Stanford University, 450 Jane Stanford Way, Stanford, CA, 94305, USA.

[3]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA

[4]Geology, School of Natural Sciences, Trinity College Dublin, Dublin 2, Ireland

[5]Durham University, Department of Earth Sciences, DH1 3LE, UK.

[6]Department of Earth Science, University of California Santa Barbara, Santa Barbara, CA 93106, USA.

[7]Department of Geosciences, Princeton University, Princeton, NJ, 08544, USA.

[8]Research School of Earth Sciences, The Australian National University, Canberra, ACT, 2601, Australia.

[9]Department of Biology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark.

[10]Key Laboratory of Petroleum Geochemistry, Research Institute of Petroleum Exploration and Development, China National Petroleum Corporation (Petrochina), Beijing, 100093, China.

[11]Institute of Sedimentary Geology, Chengdu University of Technology, Chengdu 610059, China.

[12]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, 02543, USA.

[13]Department of Earth Sciences, Carleton University, Ottawa ON, K1S 5B6, Canada.

[14]Department of Geosciences, Mississippi State University, Mississippi State, MS 39762, USA

[15]University of Copenhagen, Globe Institute, Øster Voldgade 5-7, DK-1350 Copenhagen K, Denmark.

[16]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.

[17]University of São Paulo, Institute of Geosciences, R. do Lago, 562 – Butantã, São Paulo, Brazil.

[18]Geological Survey of Canada, 3303-33rd St NW, Calgary, Alberta, Canada.

[19]Department of Geosciences, University of Wisconsin-Milwaukee, Milwaukee, WI, 53211, USA.

[20]CGG, Tyn Y Coed, Llanrhos, Llandudno LL30 1SA, UK.

[21]Pomona College, Geology Department, Claremont, CA, 91711, USA.

[22]Department of Earth and Planetary Sciences, Yale University, New Haven, CT, 06437, USA.

[23]Department of Geosciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24061, USA.

[24]Department of Atmospheric, Oceanic, and Earth Sciences, George Mason University, Fairfax, VA, USA.

[25]Department of Geology, Kansas State University, 108 Thompson Hall, KS 66506, USA.

[26]Géosciences Environnement Toulouse, Université de Toulouse, CNRS, Toulouse, France.

[27]Department of Earth and Planetary Sciences/Geotop, McGill University, 3450 University St., Montreal, QC H3A0E8, Canada.

[28]Tissue Development and Evolution (TiDE), Department of Laboratory Medicine, Lund University, Sölvegatan 19, 223 84 Lund, Sweden.

[29]Department of Earth and Planetary Sciences, Johns Hopkins University, Olin Hall, Baltimore, MD 21218, USA.

[30]Geosciences, The University of Texas Permian Basin, 4901 E. Univeristy Blvd., Odessa, TX 79762, USA.

[31]Dept. of Geoscience, University of Calgary, 2500 University Drive NW, Calgary, AB, Canada.

[32]Geological Survey of Norway, Leiv Eirikssons vei 39, 7040 Trondheim, Norway.

[33]Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK.

[34] Northern Territory Geological Survey, Level 3 Paspalis Centrepoint, 48-50 Smith Street, Darwin, Australia.

[35]Deptartment of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, MA 02138, USA.

[36]Abteilung Geowissenschaften, Staatliches Museum für Naturkunde Karlsruhe, Karlsruhe, 76133, Germany.

[37]Fortescue Metals Group Ltd, L2/87 Adelaide Terrace, East Perth WA 6004, Australia.

[38]State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation & Institute of Sedimentary Geology, Chengdu University of Technology, Chengdu 610059, China.

[39]International Center for Sedimentary Geochemistry and Biogeochemistry Research, Chengdu University of Technology, Chengdu 610059, China.

[40]School of the Environment, Geography and Geosciences, University of Portsmouth, Burnaby Road, Portsmouth, PO1 3QL, UK

[41]GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany.

[42]Institute for Geophysics, The University of Texas at Austin, Austin, Texas, USA

[43]Zurich, Switzerland. (Unaffiliated)

[44]Department of Water and Environmental Regulation, 8 Davidson Terrace, Joondalup WA 6027, Australia.

[45]Deptartment of Geoscience, University of Wisconsin-Madison, Madison WI 53706, USA.

[46]School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK

[47]Geosciences and Natural Resources Department, College of Arts and Sciences, 331 Stillwell Building, Western Carolina University, Cullowhee, NC, 28789, USA.

[48]Department of Earth Sciences, Dartmouth College, HB6105 Fairchild Hall, Hanover NH 03755, USA.

[49]CPMTC, Geology Department, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil.

[50]School of GeoSciences, University of Edinburgh, James Hutton Road, Edinburgh EH9 3FE, UK.

[51]The University of Texas at Dallas, Department of Geosciences, ROC 21, 800 West Campbell Road, Richardson, TX 75080-3021, USA.

[*]Corresponding authors. Emails: r.g.stockey@soton.ac.uk, esper@stanford.edu

Abstract

A geologically rapid Neoproterozoic oxygenation event is commonly linked to the appearance of marine animal groups in the fossil record. However, there is still debate about what evidence from the sedimentary geochemical record – if any – provides strong support for a persistent shift in surface oxygen immediately preceding the rise of animals. We present statistical learning analyses of a large dataset of geochemical data and associated geological context from the Neoproterozoic and Paleozoic sedimentary record, and then use Earth system modeling to link trends in redox sensitive trace metal and organic carbon concentrations to the oxygenation of Earth's oceans and atmosphere. We do not find evidence for the wholesale oxygenation of Earth's oceans in the late Neoproterozoic. We do, however, reconstruct a moderate long-term increase in atmospheric oxygen and marine productivity. These changes to the Earth system would have increased dissolved oxygen and food supply in shallow-water habitats during the broad interval of geologic time that the major animal groups first radiated. This approach provides some of the most direct evidence for potential physiological drivers of the Cambrian radiation, while highlighting the importance of later Paleozoic oxygenation in the evolution of the modern Earth system.

Introduction

Earth's oceans and atmosphere are traditionally thought to have undergone two major episodes of oxygenation[1]. The first, the Great Oxidation Event, occurred between 2.2 and 2.4 billion years ago and resulted in an increase in atmospheric oxygen from trace concentrations to levels above 0.1% of the present atmospheric level (PAL)[2,3]. A second, Neoproterozoic Oxygenation Event (NOE) has long been inferred around the Ediacaran-Cambrian boundary (~538.8 million years ago, Ma), originally interpreted as the time when atmospheric oxygen concentrations increased from Proterozoic to near modern levels (canonically from 1-10% to ~100% PAL)[1,4]. This presumed Neoproterozoic Oxygenation Event has received extensive attention due to its broad temporal coincidence with the appearance of the first unambiguous macroscopic animal fossils in the later Ediacaran and Cambrian[5]. Ocean-atmosphere oxygenation has widely been invoked as a plausible environmental driver of early animal evolution, since all extant animal groups rely on environmental oxygen to complete their life cycle, and observations from modern oceans indicate that many Cambrian animal body plans and ecologies would not have been permitted at very low oxygen levels[6].

Over the past decade, however, the inferred trajectory of Neoproterozoic oxygenation has been challenged. Multiple lines of evidence support the general persistence of anoxic water masses in the deep ocean (and intermittently on the shelves) until the mid-Paleozoic Era[7–11], punctuated by a series of extreme 1-10 Myr-scale oscillations in the oxygenation of Earth's oceans and atmosphere (oceanic oxygenation events) through much of the Neoproterozoic Era and into the Cambrian[12–14]. These data call into question a late Neoproterozoic pO_2 increase to near-modern levels. They further highlight the importance of spatial and temporal scale in framing questions about Neoproterozoic environmental change and the evolution of animals. Many geochemical

proxies for paleoredox record fluctuations in bottom-water oxygenation along the outer shelf and slope, whereas the majority of modern marine animals and those recorded in the fossil record live in shallow shelf environments. It is therefore the oxygenation of these shallow environments that would control the degree of hypoxic stress experienced by marine animals. While Earth system boundary conditions such as continental configuration can play a major role in global deep ocean oxygenation at 10Myr timescales[15], dissolved [O_2] in shallow marine animal habitats is expected to be primarily controlled by ocean-atmosphere gas exchange. Furthermore, geochemical proxies commonly record fluctuations on 10kyr to 1Myr timescales, whereas Neoproterozoic-Paleozoic animal radiations were evolutionary singularities, with new body plans and ecologies remaining viable on 100Myr timescales. In order to test permissive environment hypotheses relating to environmental oxygen and the diversification of early animals, we must therefore specifically investigate the long-term oxygenation of Earth's continental shelf environments.

Dramatic changes in organic carbon burial have also been inferred across the Neoproterozoic-Paleozoic transition. On short-term, physiological timescales, increased benthic carbon flux is expected to provide increased food supply to benthic marine ecosystems, potentially acting as an alternative or synergistic physiological driver of ecological and evolutionary change[16]. If sustained over geologic timescales, increased organic carbon burial would also have driven increased oxygen flux to the atmosphere[17], in turn increasing the oxygenation of shallow-water environments where dissolved [O_2] is dominated by air-sea gas exchange. Better understanding organic carbon burial through the Neoproterozoic and Paleozoic is therefore critical for constraining the oxygenation of shallow shelf-slope environments where most early animals were living. Organic carbon flux is also a key environmental parameter impacting the reduction rates of trace metals such as U and Mo in shallow sediments[18], complicating inferences based on some of

157 the most commonly used geochemical proxies for late Neoproterozoic ocean oxygenation[9,14,19–21].
158 Any dramatic increase in benthic organic carbon flux around the Ediacaran-Cambrian transition is
159 therefore important to understand as a potential physiological driver, as an underexplored indicator
160 of shallow-water oxygenation, and as a complicating factor in reconstructing global ocean
161 oxygenation from commonly used geochemical proxies.

162 Here, we combine approaches from statistical learning, biogeochemical modeling, and
163 ecophysiology to better constrain changes in both global ocean biogeochemistry and
164 physiologically important environmental conditions in marine animal habitats through the
165 Neoproterozoic and Paleozoic. We conduct statistical learning analyses of a large dataset of
166 geochemical data and associated geological context assembled by the Sedimentary Geochemistry
167 and Paleoenvironments Project[22] (SGP) for shales from the Tonian through Carboniferous (~1000-
168 300 Ma). We investigate sedimentary records of the trace metals molybdenum and uranium,
169 proxies that are widely used to investigate global ocean oxygenation because of their high burial
170 rates in modern anoxic settings relative to modern oxygenated settings[14,19–21,23]. Based on these
171 modern observations, faithful archives of seawater Mo and U concentrations are expected to record
172 the extent of bottom-water anoxia globally. We also investigate trends in the local redox of anoxic
173 environments using the iron speciation proxy and trends in shale total organic carbon (TOC) as a
174 proxy for organic carbon burial flux. Reconstructing the global scale biogeochemical processes
175 that motivate this study from these sedimentary geochemical proxies is complicated by local
176 biogeochemical and physical factors during deposition as well as by post-depositional processes.
177 Uneven geological sampling in space and time also complicate reconstructions. The statistical
178 analyses we apply in this study are designed to address these issues, enabling us to reconstruct
179 meaningful biogeochemical trends from the aggregated sedimentary geochemical record.

First, we reconstruct temporal trends in the mean distributions of sedimentary geochemical data, accounting for sampling biases using a weighted bootstrap analysis that incorporates the spatial and temporal proximity of samples in the dataset[24] (*Materials and Methods*). Improving on previous data compilations[9,16,19,22,25], this analysis accommodates the impact of geographical sampling bias to generate global mean trends in sedimentary concentrations[24], as well as benefiting from the improved data density in the SGP dataset[22]. We then conduct a statistical learning analysis designed to isolate global long-term trends in the biogeochemical processes that motivate our use of these proxies, separated from the complex caveats and interactions that affect interpretations of raw sedimentary data. We use a Monte Carlo random forest framework (*Materials and Methods*) to generate partial dependence analyses that isolate the marginal effect of geologic time on the mean value of each geochemical proxy with all identified confounding geochemical and geologic context variables held constant. For Mo and U, these analyses are designed to identify trends in the seawater inventories of these metals, faithfully tracking the oxygenation of global bottom-waters. For iron speciation, these analyses are designed to reconstruct how sulfide levels have varied in sampled low-oxygen shelf-slope settings independent of other oceanographic factors. For total organic carbon, these analyses are designed to reconstruct trends in organic carbon flux to sampled environments independent of other oceanographic factors, with supplementary analyses further indicating how organic carbon delivery to anoxic settings changed through this time interval. Although the absolute magnitude of these results are a reflection of the preserved sedimentary record (for instance, with organic carbon concentrations being higher than expected global averages because of the high concentrations found in shelf-slope settings where most of our sedimentary record comes from), the directional trends produced in these statistical learning analyses enable us to reconstruct changes in global biogeochemical cycles.

Finally, we combine our statistical reconstructions with Earth system modeling to investigate trends in atmospheric pO_2, global ocean biogeochemistry, and the oxygenation of shallow marine habitats over the Neoproterozoic and Paleozoic eras. Specifically, we apply a combined biogeochemical modeling framework to investigate the impacts of atmospheric oxygen and marine productivity on seawater Mo and U concentrations, organic carbon burial rates, and dissolved [O_2] in shallow shelf environments. First, we conduct an ensemble modeling experiment using an Earth system model of intermediate complexity (cGENIE[26]) to investigate the impact of different stable atmospheric oxygen and marine productivity scenarios[27] on three-dimensional global ocean simulations of dissolved [O_2]. The benthic dissolved oxygen profiles from these ocean simulations are then used to estimate the simulated extents of redox-sensitive sinks, which are used to force coupled Mo and U mass balance models and thus estimate dissolved trace metal inventories for each oxygen-productivity scenario. Estimates of global marine organic carbon burial rates are then further calculated for each scenario based upon binned mean rates from the CANOPS biogeochemical model[27,28]. The simulated seawater Mo and U concentrations and organic carbon burial rates for each stable pO_2-productivity scenario are ultimately combined with our statistical reconstructions of ocean biogeochemistry to provide new long-term (~10Myr timescale) estimates of atmospheric oxygen, marine productivity, seafloor redox and shallow shelf dissolved [O_2] through the Neoproterozoic and Paleozoic eras.

Geochemical proxy records

Trace metal concentrations in anoxic black shales do increase in the late Neoproterozoic in our temporal-spatial weighted analyses of bootstrapped means, but they subsequently decrease in the early Paleozoic before increasing again in the Devonian (Fig. 1A and C). Distributions of both Mo

in euxinic shale and U in anoxic shale show similar trends when temporal and spatial sampling biases are accounted for. When trace metals are standardized to TOC, there is considerable noise in bootstrap means and no clear trend through the Neoproterozoic and Early Paleozoic, although both Mo/TOC and U/TOC increase in the later Devonian (Extended Fig. 3). These analyses alone therefore suggest that there was no major sustained increase in marine Mo or U concentrations until the Devonian, contrasting sharply with previous interpretations of trace metal data.

A major (~130%) Ediacaran-Cambrian increase in TOC is also captured in our spatial-temporal bootstrap analyses, followed by further increases in the mid Paleozoic (Fig. 1D). There is a low proportion of euxinic (anoxic, sulfidic) samples for most of the Tonian through Carboniferous based upon iron speciation data, indicating that the majority of anoxic samples in the Neoproterozoic and Paleozoic were deposited under ferruginous (anoxic, non-sulfidic) bottom water conditions (Fig. 1B) according to standard iron speciation thresholds (see further discussion of the iron speciation proxy in *Materials and Methods*). The late Ediacaran and Devonian are notable exceptions, with relatively high proportions of anoxic shale samples classified as euxinic. Some transient trends, such as high late Tonian (~800Ma) Mo and TOC values are difficult to interpret due to the limited number of euxinic shale units sampled through those intervals (Fig. 1), although this interval of the late Tonian merits further investigation given correlated shifts in other proxy records[5].

Deconvolved biogeochemical trends

Temporal reconstructions of mean global Mo and U, isolated from the impacts of local redox, organic carbon, and depositional and post-depositional processes (*Materials and Methods*) show minor, partially transient increases in trace metal concentrations in the late Neoproterozoic

followed by major increases in the mid-Paleozoic (Fig. 2A and C). The trajectories of the Mo and U partial dependence plots generated in our Monte Carlo random forest analyses differ slightly in the timings and rates of change. This remains the case when iron speciation data are used as predictor data, rather than thresholds (Extended Figure 6), indicating the robustness of these trends to specific proxy interpretations (c.f. Pasquier et al.[29]). This potentially suggests minor differences in the response of these trace metals to Earth system boundary conditions, although we are hesitant to overinterpret these differences given both the different sample distributions between analyses and the broad temporal resolution of this study. Notably, these analyses are designed to investigate changes in the mean state of the Earth system on relatively long geologic timescales, and are therefore not expected to capture <~10 Myr redox instabilities, such as those observed in the Ediacaran-Cambrian U cycle[13] and produced by biogeochemical models at similar timescales[12]. Variable importance plots (Extended Fig. 4) illustrate that TOC (followed by geologic age and [Al]) is the most important predictor of Mo and U concentrations in our random forest models.

In contrast to our trace metal reconstructions, deconvolved TOC reconstructions do exhibit a major sustained increase in the late Neoproterozoic and early Cambrian (Fig. 2). This ~70% increase in deconvolved mean values is in broad agreement with compilations of raw data[16], but of lower magnitude than previously estimated after accounting for confounding variables and sampling biases. This Ediacaran-Cambrian TOC increase is followed by a steady increase from the Middle Ordovician onward in analyses including all shale samples. TOC analyses that are restricted to anoxic shales (reflecting deconvolved trends in organic carbon burial in anoxic settings alone) show a more muted Ordovician-Devonian increase, while TOC analyses of anoxic shales with Fe_{Py}/Fe_{HR} as a predictor variable (reflecting deconvolved trends in organic carbon burial in anoxic settings independent of temporal changes in the Fe-S biogeochemistry of anoxic

environments) show a major stepwise Ediacaran-Cambrian increase followed by relatively stable values until a more muted Devonian increase (Extended Fig. 5). Analyses of variable importance (Extended Fig. 4) indicate that Al concentrations, lithology, and geographic coordinates (latitude and longitude) are the most important predictors of TOC in our primary analyses. The proportion of euxinic samples moderately increases in the late Ediacaran before decreasing in the early Paleozoic, then increases dramatically in the Silurian-Devonian in agreement with previous statistical analyses[8].

Implications for oxygen and productivity

In our ensemble Earth system modeling experiment, we estimate the impact of atmospheric oxygen concentrations and marine productivity on the biogeochemical processes that we aim to isolate in our statistical learning analyses, as well as on the oxygenation of both the global ocean and shallow marine habitats. Our cGENIE simulations of 3D ocean biogeochemistry show that the oxygenation of shallow shelf and global bottom water environments respond differently to changes in atmospheric pO_2 and marine PO_4 combinations that are predicted to be stable on geologic timescales[27] (Fig. 3; *Materials and Methods*). Most of the global seafloor is overlain by reducing bottom waters at relatively low atmospheric O_2 levels (below ~25-100% PAL depending on marine PO_4), with the proportions of anoxic and suboxic seafloor varying as a function of atmospheric oxygen and marine productivity (Figs. 3A, D; Extended Figs. 8-9). At relatively high atmospheric oxygen levels (~25-100% PAL – again depending on marine PO_4), the majority of the global seafloor – including the deep ocean – is overlain by oxygenated bottom waters. In contrast, dissolved [O_2] in shallow shelf environments scales essentially linearly with atmospheric oxygenation (Fig. 3F, with minor decreases in shelf oxygenation with increasing productivity.

Changing the continental configuration, global climate state, or biological pump strength in cGENIE does not substantially impact these results (Extended Fig. 10). Although continental configuration may have played a role in structuring deep ocean ventilation during the Paleozoic[15], this does not challenge our key observation that shallow shelf oxygenation scales essentially linearly with atmospheric oxygen while deep ocean oxygenation exhibits a contrastingly non-linear relationship with atmospheric oxygen.

We link these oceanographic trends to the biogeochemical dynamics isolated in our statistical learning analyses by coupling cGENIE outputs to biogeochemical models of molybdenum, uranium and organic carbon. The predominance of reducing seafloor conditions at low atmospheric pO_2 levels results in a very limited sensitivity of seawater Mo and U concentrations to changing atmospheric oxygen below ~25% PAL atmospheric pO_2 (Fig. 3B, E). Our simulations demonstrate that both anoxic and suboxic conditions can act as major controls on the trace metal inventory of seawater, with the same marine Mo and U concentrations simulated by a range of biogeochemical landscapes defined by the balance of the two reducing sinks[30] (Extended Fig. 7). Seawater trace metal concentrations therefore broadly track the extent of oxic seafloor (or, equivalently, $f_{anox+subox}$, the combined extent of reducing seafloor), and the seawater inventories of Mo and U are much more sensitive to both atmospheric oxygen concentrations and marine productivity above ~25% PAL pO_2. In contrast to modeled trace metal concentrations, organic carbon burial rates are sensitive to changing atmospheric oxygen and marine productivity across the whole logarithmic range of pO_2-productivity scenarios investigated here (Fig. 3C). Global average organic carbon burial rates (and therefore average TOC in shales, unless separately impacted by other factors) are predicted to increase relatively continuously with logarithmic

317 increases in atmospheric oxygen and marine productivity, similar to shallow shelf dissolved
318 oxygen concentrations (Fig. 3F, Extended Figs. 8-9).
319
320 **Discussion**
321 By integrating statistically deconvolved proxy records with a novel combination of
322 biogeochemical models, we are able to provide new long-term reconstructions of Neoproterozoic-
323 Paleozoic ocean-atmosphere oxygenation and marine productivity. Coupling increased sampling
324 with improved statistical treatment of confounding geologic context and geochemical variables
325 enables us to both reevaluate classical trace metal evidence for a stepwise Neoproterozoic
326 oxygenation and to realize the potential of global TOC records for constraining changes in the
327 oxygenation of the atmosphere and shallow marine environments, as well as changes in marine
328 productivity. Although previous studies have inferred a Paleozoic shift in marine redox[8–11], our
329 analyses provide the first direct reconciliation of canonical evidence for a Neoproterozoic
330 oxygenation event[19,20] with mid-Paleozoic ocean-atmosphere oxygenation. Our analyses further
331 establish the existence of sustained changes in ocean-atmosphere oxygenation and productivity at
332 longer timescales than Ediacaran-Cambrian oscillations in ocean-atmosphere oxygenation[12,13].
333 The contrasting sensitivities of modeled seawater trace metal concentrations and organic
334 carbon burial rates to atmospheric oxygenation and marine productivity provide a plausible
335 environmental mechanism for the two major transitions observed in our deconvolved
336 reconstructions of Mo, U and TOC (Fig. 2). In the late Neoproterozoic, we reconstruct a minor
337 increase in marine trace metal concentrations but a major stepwise increase in organic carbon
338 burial. These contrasting responses likely indicate a late Neoproterozoic increase in atmospheric
339 oxygen, marine productivity, and organic carbon burial without a major change in global deep-

ocean oxygenation (Fig. 4). Relatively low seawater trace metal concentrations were maintained because the majority of bottom waters remained reducing, even though raw shale metal concentrations may have increased on average[19,20] due to the impact of increased organic carbon loading and accompanying metal sequestration. In the mid-Paleozoic, we reconstruct major increases in both marine trace metal concentrations and organic carbon burial, indicating that during this interval atmospheric oxygen and marine productivity increased to levels at which both proxies are expected to be sensitive. The seemingly contradictory implications of deconvolved TOC and Mo-U records can therefore be reconciled if the Ediacaran-Cambrian increase in atmospheric oxygen and marine productivity was insufficient to oxygenate deep ocean water masses and the global ocean did not become persistently oxygenated to near-modern levels until the Silurian-Devonian.

Although trace metal proxies are commonly used to study ocean oxygenation through Earth history because of their sensitivity to the areal extent of reducing bottom waters, our results highlight the oblique view that trace metals provide of shallow marine animal habitats. While we refute suggestions that most of the global ocean became fully oxygenated in the late Neoproterozoic[1], we do reconstruct a late Neoproterozoic increase in dissolved [O_2] in shallow shelf environments. We therefore argue that whether there was a Neoproterozoic oxygenation event depends on spatial and temporal perspective. From a global oceanographic viewpoint encompassing the deep ocean, there was essentially no NOE. From an ecophysiological perspective, considering that shallow shelf environments are where most animals live and most of our fossil record is preserved, there was a NOE. Specifically, the analyses indicate that shallow marine habitats were likely suboxic or severely hypoxic (0 to 22 μmol/kg O_2) on average for most of the Neoproterozoic, hypoxic (22 to 63 μmol/kg) in the early Paleozoic, and generally would not

363 be considered oxic from a modern ecophysiological perspective (≥63 μmol/kg)[6] until the Devonian
364 (Fig. 4e). Across spatial oxygen gradients in the modern ocean, the differences in shallow marine
365 oxygen concentrations that we reconstruct between the Neoproterozoic and Paleozoic eras
366 correspond to substantial differences in the functional and taxonomic diversity of marine animal
367 communities[6].

368 Establishing drivers of the sustained Ediacaran-Cambrian and Silurian-Devonian shifts we
369 reconstruct in oxygenation and productivity is beyond the scope of this study, although we suggest
370 that biogeochemical feedbacks related to volcanic reductant flux[31] and the ecological expansion of
371 land plants[9,11] warrant particular research attention. Changes in proxy records may also have been
372 modulated by changes in temperature and plate tectonic configuration[15], and these invite further
373 study, however, they are unlikely to explain the scale of changes reconstructed in our statistical
374 analyses (Extended Fig. 10). Moreover, changes in continental configuration primarily impact
375 deep water oxygenation[15], highlighting the power of our multiproxy trace metal and organic carbon
376 approach for interrogating the oxygenation of shallow water marine animal habitats.

377 The increase in marine productivity that we reconstruct in the late Neoproterozoic also has
378 physiological implications for early marine animals. Our reconstructions support hypotheses
379 linking food supply to the Cambrian radiation[16,32] and open doors for further analyses evaluating
380 the joint roles of oxygen and food supply in driving the trends we observe in the fossil record[33].
381 The second major increase in atmospheric pO_2, marine primary production, and shallow marine
382 dissolved [O_2] that we reconstruct for the Silurian-Devonian has similar implications for the
383 potential roles of oxygen and food supply in the Devonian radiation of fishes[34]. We expect
384 mechanistic ecophysiological modeling approaches – similar to those designed to investigate the
385 capability of other inferred environmental changes to drive specific reconstructed biodiversity

386 dynamics (e.g. refs[35,36]) – to be critical in establishing the specific roles that oxygen and food
387 supply may have played as drivers of the early animal evolution. Nonetheless, these analyses
388 establish sustained directional changes in both dissolved oxygen and export-driven food supply in
389 late Neoproterozoic shallow marine habitats, with similar magnitudes to environmental gradients
390 that play key roles in structuring the biodiversity and composition of modern marine ecosystems[6].
391 We thus demonstrate reductions in key ecophysiological stressors at an appropriate time to have
392 been bottom-up drivers of the polyphyletic radiation of marine animals during the Ediacaran-
393 Cambrian transition.
394

Materials and Methods

Data processing

The complete Phase 1 dataset was downloaded from the Sedimentary Geochemistry and Paleoenvironments Project website (sgp-search.io; data API – {“type”:“nhhxrf”,“filters”:{},“show”:[“fe”,“fehr_fe_t”,“fe_py_fe_hr”,“toc”,“alu”,“mo”,“u”,“fe_t_al”,“coord_lat”,“coord_long”,“basin_type”,“meta_bin”,“environment_bin”,“lithology_name”,“max_age”,“min_age”,“interpreted_age”,“site_type”,“strat_name”,“strat_name_long”]}). This dataset was filtered to include only sedimentary samples with interpreted geological ages between 300 and 1000 Ma (approximately Tonian through Carboniferous). The dataset was then further filtered to include only samples described as having shale-like or fine-grained lithologies (argillite, clay, claystone, dolomudstone, lime mudstone, meta-argillite, metapelite, metasiltstone, mud, mudstone, oil shale, pelite, phosphorite, shale, silt, siltite, siltstone, slate, plus samples assigned no lithology) and samples assigned marine depositional environments (basinal, outer shelf, inner shelf, plus samples assigned no depositional environment). We further removed samples with exceptionally high Mo or U concentrations that would be considered ore-grade metalliferous rocks (using a cut-off of 1000 ppm). The distribution of categorical geological context variables in this primary dataset through geologic time is shown in Extended Figure 1. The full filtering process and number of samples in the dataset after each filtering step is shown in Supplementary Data File 1.

Spatial-temporal weighted bootstrap analysis

For our spatial-temporal weighted analyses of bootstrapped means, we remove samples without geographic coordinate data, and subset the primary SGP dataset using proxy-specific filters

418 (Extended Data Table 1). Next, we bin the shale samples for each proxy into 25 Myr time bins. In
419 each 25 Myr time bin, we assign samples weights based upon their temporal and spatial proximity
420 to other samples in the bin using the inverse weighting algorithm of Mehra et al.[24]. We then use
421 these weights to generate distributions of 1000 weighted bootstrapped means for each time bin and
422 plot those distributions as box and whisker plots. We present histograms illustrating the number
423 of lithostratigraphic units sampled per time bin for each weighted bootstrap analysis. We conduct
424 the same analysis for shale Mo/TOC and U/TOC values (Extended Figure 3). We also present box
425 and whisker plots of the raw data used in these analyses (Extended Figure 2). Full spatial-temporal
426 weighted bootstrap methods are shown in Supplementary Data File 2.
427
428 *Random forest analyses*
429 We conduct Monte Carlo random forest analyses to generate statistical reconstructions of changes
430 in four key marine biogeochemical variables from the Tonian through Carboniferous – seawater
431 [Mo], seawater [U], proportion of sampled anoxic depositional environments that were euxinic,
432 and organic carbon burial. Each of these variables are semi-quantitatively reconstructed as a
433 statistical model of changes in the mean value of a sedimentary geochemical proxy (shale [Mo],
434 shale [U], proportion euxinic based on iron speciation, TOC) with geological time when all
435 geologic context and geochemical variables known to impact the incorporation and preservation
436 of the desired biogeochemical signal in sedimentary rocks are held constant. For each
437 biogeochemical variable, a different combination of predictor variables is used, and additional
438 filtering steps may be applied based upon the geological and biogeochemical processes associated
439 with the specific proxy (Extended Data Table 2). Some filters in these analyses (also in the
440 bootstrap analyses above) involve iron speciation parameters that are interpreted according to

standard thresholds[37,38]. These thresholds have recently been questioned[29], although that compilation involved a large number of samples that would be deemed inappropriate for iron speciation based on criteria in Raiswell et al.[38] and thus continued re-analyses of these thresholds are necessary. To test whether these iron speciation thresholds were influencing our results, we ran our molybdenum analyses (the analyses most dependent on iron speciation) with iron speciation values both as explicit filters on the dataset to isolate anoxic samples (Fig. 2) and as predictor variables in the analysis (Extended Fig. 6). The results are qualitatively very similar, indicating that ongoing debate about the use of specific iron speciation thresholds does not impact our results.

To maximize the number of samples that can be used in these models, samples that have data entered for the variable of interest, estimated age, and any other necessary geochemical data for redox classification (Fe_{HR}/Fe_T and Fe_{Py}/Fe_{HR} for Mo; Fe_{HR}/Fe_T or Fe_T/Al for U) but are missing other variables incorporated in the random forest analyses are included. For samples with partial data, missing data for categorical geologic context (lithology, site type, basin type, environmental bin, metamorphic bin) and geographic coordinates are randomly assigned in each Monte Carlo simulation from the full range of feasible values for each variable (as random forest analyses require a complete data matrix). For samples missing [Al] (in analyses that include [Al]), we impute [Al] by generating a random value between the 25th and 75th percentile of [Al] values within the sample's assigned lithology (based upon the entire primary dataset described above and in Supplementary Data File 1). For samples without an assigned lithology that are missing [Al], a random [Al] value between the 25th and 75th percentile of [Al] values for all lithologies is imputed and then lithology is randomly assigned. In our Monte Carlo approach, these random value assignments are conducted 100 times for each analysis, corresponding to 100 separate random

forest models per proxy/biogeochemical process. The results of these 100 random forest models are then evaluated to estimate the uncertainty associated with geologic age models (see description below) and samples with partial data.

All samples in our Monte Carlo random forest analyses are assigned ages using on section-based age models and assigned geological age uncertainties. Samples that were not assigned maximum and minimum ages by SGP contributors are assigned age uncertainties of 25 Myrs (□ 12.5 Myrs relative to assigned interpreted age) for Neoproterozoic samples and 10 Myrs (□ 5 Myrs relative to assigned interpreted age) for Paleozoic samples. In each Monte Carlo random forest analysis, we randomly assign each sample a geologic age based upon its estimated maximum and minimum ages and its relationship with other samples in the sampled section in order to accommodate both uncertainties in geologic age models and the principle of stratigraphic superposition. These ages are assigned separately for each of 100 random forest models independently to incorporate uncertainty in assigned geologic ages. Age models are assigned by categorizing sections into 8 categories under 3 broad classifications:

1. Sections with height/depth data
    a. Samples are assigned continuously ascending interpreted age estimates.
    b. Samples are all assigned the same interpreted age estimate.
    c. Samples are assigned interpreted age estimates in clusters or steps.
2. Sections without height/depth data
    a. Samples are assigned continuously ascending interpreted age estimates.
    b. Samples are all assigned the same interpreted age estimate.
    c. Samples are assigned interpreted age estimates in clusters or steps.
3. Sections with only one sample

a. Sample has stratigraphic height/depth

b. Sample does not have stratigraphic height/depth

In each scenario, ages are randomly assigned within younging upward sequences to incorporate both age uncertainty and the individual sample's stratigraphic relationship to other samples in the section. Full age model methods are shown in Supplementary Data File 3.

We present the results of our Monte Carlo random forest analyses as partial dependence plots. Partial dependence plots show the isolated marginal effect of a variable of interest (geologic time in the analyses presented in Fig. 2) on the predicted outcome of the random forest model[39]. We use partial dependence plots rather than other feature effect methods (e.g. accumulated local effects plots) for both statistical and biogeochemical reasons: 1) correlations between geologic time and model variables are consistently low in all model treatments, and 2) while the mean tendency of proxy variables may shift through geologic time, observations that are less likely in a certain geological time interval (e.g. rift basins or low TOC shale in the late Paleozoic) are still represented in our dataset (Extended Fig. 1; Appendices 1-2). This means that we never ask our models to make predictions in unrealistic feature space where the random forest model has not been trained. 3) Most importantly, we use global feature effects models because they exclude the impacts of other confounding variables linked to authigenic enrichments in sedimentary rocks. Partial dependence plots allow us (at least to a first approximation) to isolate changes in marine biogeochemistry, by asking questions such as "How do molybdenum concentrations in fine-grained siliciclastic rocks change independent of other secular changes in sampled sedimentary environments, including changes in average organic carbon loading, local redox and depositional environment?". Accumulated local effects plots and other local feature effects models, on the other hand, would be expected to better represent average sedimentary enrichments for each geologic

time interval (without, for example, standardizing for long-term variations in average local redox) because they do factor in changes in other predictor variables (in our case characteristics of sedimentary rocks). Consequently, they are less likely to independently track the oceanographic and Earth system changes that we are most interested in investigating. We present partial dependence plots as envelopes, summarizing the 100 random forest models in each Monte Carlo analysis (Fig. 2). The plotted envelopes are generated by linear interpolation of the 100 individual partial dependence plots generated per analysis at 0.1 Myr time intervals and computing the 5^{th}, 25^{th}, 75^{th} and 95^{th} percentiles of the interpolated partial dependence plot populations at each time step. Full Monte Carlo random forest methods are shown in Supplementary Data File 3.

Earth system modeling

We generated three-dimensional realizations of feasible ancient ocean biogeochemistry using the cGENIE Earth system model of intermediate complexity[26]. We conducted an ensemble modeling experiment, varying atmospheric pO_2 and marine PO_4 in 11 logarithmic increments between 1% and 100% of present atmospheric/oceanic levels (1%, 1.6%, 2.5%, 4%, 6.3%, 10%, 16%, 24%, 40%, 63%, 100%). Our main text results (ensemble #1) use a Cryogenian-Ediacaran (635Ma) continental configuration, a shallow (50% of modern, 294.9745m) e-folding depth to parameterize organic remineralization, 3336 ppm (12 PAL) atmospheric pCO_2, and appropriate Cryogenian-Ediacaran estimate of the solar constant (1295.9701 Wm^{-2}, a 5.2653% reduction from modern), following Reinhard et al.[40]. All ensembles use a single limiting nutrient (PO_4) scheme to parameterize biological export, following Meyer et al.[41]. We conduct additional ensemble experiments as sensitivity analyses to investigate the impacts of the marine biological carbon pump (#2), continental configuration (#3), global climate (#4, #5), and combined continental

configuration and global climate (#6). Ensemble #2 uses a modern e-folding depth (589.9451m) and is otherwise identical to ensemble #1. Ensemble #3 uses an Ordovician continental configuration and solar constant[42] and is otherwise identical to ensemble #1. Ensemble #4 and ensemble #5 use atmospheric pCO_2 concentrations of 834 ppm (3 PAL) and 5560 ppm (20 PAL) respectively and are otherwise identical to ensemble #1. Ensemble #6 uses an Ordovician continental configuration and solar constant[42], an atmospheric pCO_2 concentration of 834 ppm (3 PAL) and is otherwise identical to ensemble #1. All cGENIE simulations are run for 10,000 years, at which point benthic oxygen concentrations have reached steady state.

For each experiment, we extract the mean annual dissolved oxygen concentrations for the final model year. We then categorize the bottom water cells of each ocean realization into three categories based upon dissolved oxygen concentrations: anoxic (dissolved $[O_2] \leq 0$ μmol/kg), suboxic (dissolved $[O_2] > 0$ μmol/kg; dissolved $[O_2] \leq 4.8$ μmol/kg), oxic (dissolved $[O_2] \geq 4.8$ μmol/kg)[6]. As all cGENIE configurations in this study use equal area grid cells, the proportional extents of anoxic, suboxic and oxic seafloor are then computed as a direct proportion of all (non-land) model grid cells. For each experiment we also calculate mean shelf $[O_2]$ by computing the mean dissolved $[O_2]$ of all ocean cells adjacent to land in the top 3 layers of the cGENIE ocean (<283.8 m water depth).

Stable atmospheric pO_2 and marine productivity states

The comparatively coarse spatial resolution and associated computational efficiency of CANOPS means that it is viable to run large model ensembles with an open oxygen cycle on geologic timescales, whereas computational constraints mean that it would be unrealistic to run large

ensemble cGENIE experiments at the scale conducted here for the timescales required for an open oxygen cycle to reach equilibrium. We therefore use results from an ensemble of model experiments using the CANOPS Earth system model[27] to establish a subset of our cGENIE Earth system model experiments that are expected to be stable on geologic timescales.

Trace metal mass balance

We use a three-sink Mo-U mass balance model[30] to simulate seawater [Mo] and [U] for each cGENIE Earth system simulation. We use the extents of anoxic, suboxic and oxic seafloor calculated from our cGENIE ocean realizations (Extended Fig. 8) as forcings for the extents of redox-sensitive sinks in the mass balance model. The naming conventions f_{anox} (fractional extent of anoxic seafloor), f_{subox} (fractional extent of suboxic seafloor) and f_{oxic} (fractional extent of oxic seafloor), are used equivalently to f_{eux}, f_{red} and f_{oxic} in Stockey et al.[30] to directly match the parameters extracted from our cGENIE ocean simulations. The modern biogeochemical data underlying the parameterization of these sinks and associated fluxes remains the same. All other flux subscripts are correspondingly updated such that:

$$\frac{\mathrm{d}N_{\mathrm{sw}}}{\mathrm{dt}} = \mathrm{F_{riv}} - \mathrm{F_{oxic}} - \mathrm{F_{subox}} - \mathrm{F_{anoxic}}$$

Thus, the change in the seawater inventory of Mo or U (N_{sw}) with respect to time is a function of the metal flux into the global ocean via rivers (F_{riv}), minus the metal fluxes into redox-sensitive depositional environments (F_{oxic}, F_{subox}, F_{anox}). The pseudospatial scaling algorithm of Stockey et al.[30] (based on Reinhard et al.[43]) is also used here to incorporate the estimated attenuation of trace metal burial rates with water depth for scenarios with expansive reducing conditions in deep marine environments. For all parameters in the model analyses presented in this study, we use the

mean value (midpoint) of the uniform distributions used to define model parameters in Stockey et al.[30].

We further present a global sensitivity analysis of this Mo-U mass balance model, illustrating contours of equal seawater [Mo] or [U] for varying balances of f_{anox} and f_{subox} (Extended Fig. 7). This analysis uses the same parameters as the model applied to the cGENIE simulations but f_{anox} and f_{subox} are varied in 31 logarithmic steps between 0.1% and 100%.

Organic carbon burial

We use results from the ensemble CANOPS Earth system model experiment used to establish stable Earth system states[27] in order to estimate organic carbon burial rates for each atmospheric pO_2 and marine productivity scenario. Global marine organic carbon burial rates (as a function of present oceanic levels) are binned into the logarithmically-spaced atmospheric pO_2 and marine productivity scenarios and the mean simulated organic carbon burial flux is calculated for each scenario.

Acknowledgments

We thank Anne-Sofie Ahm, Wing Chan, Matthew Clarkson, Katherine Doyle, Julie Dumoulin, Tiffani Fraser, Jon Husson, Benjamin Johnson, Florian Kurzweil, Alfred Lenz, Xinze Lu, Yu Liu, Austin Miller, Peter Petrov, Sylvain Richoz, Patrick Sack, Clinton Scott, Sarah Slotznick, Samuel Spinks, Sabrina Tecklenburg, Danielle Thompson, Huajian Wang, Lei Xang and Jianghai Wang for their contribution to the Sedimentary Geochemistry and Paleoevironments Project dataset used in this study. We thank Andy Ridgwell and Alexandre Pohl for helpful discussions. AJMJ publishes with permission of the Senior Executive Director, Northern Territory Geological Survey.

602 We thank Stanford University, the Stanford Research Computing Center, the IRIDIS High
603 Performance Computing Facility and associated High Performance Computing support services at
604 the University of Southampton for providing computational resources and support during this
605 research. This research and the SGP are funded by NSF grants EAR-1922966 and EAR-2143164
606 to EAS. We further thank the donors of The American Chemical Society Petroleum Research Fund
607 for partial support of this research (61017-ND2).
608

609 **Code Availability**

610 R code used to run the statistical analyses is available in R Markdown format in Supplementary
611 Data Files 1-3. The code for the version of the 'muffin' release of the cGENIE Earth system model
612 used in this paper, is tagged as v0.9.45, and is assigned a DOI: 10.5281/zenodo.xxxxxxx.
613 Configuration files for the specific experiments presented in the paper can be found in the
614 directory: genie-userconfigs/PUBS/submitted/Stockey_et_al.NatGeo.2023. Details of the
615 experiments, plus the command line needed to run each one, are given in the readme.txt file in that
616 directory. All other configuration files and boundary conditions are provided as part of the code
617 release.A manual detailing code installation, basic model configuration, tutorials covering various
618 aspects of model configuration, experimental design, and output, plus the processing of results, is
619 assigned a DOI: 10.5281/zenodo.7545814 (v0.9.35).
620
621

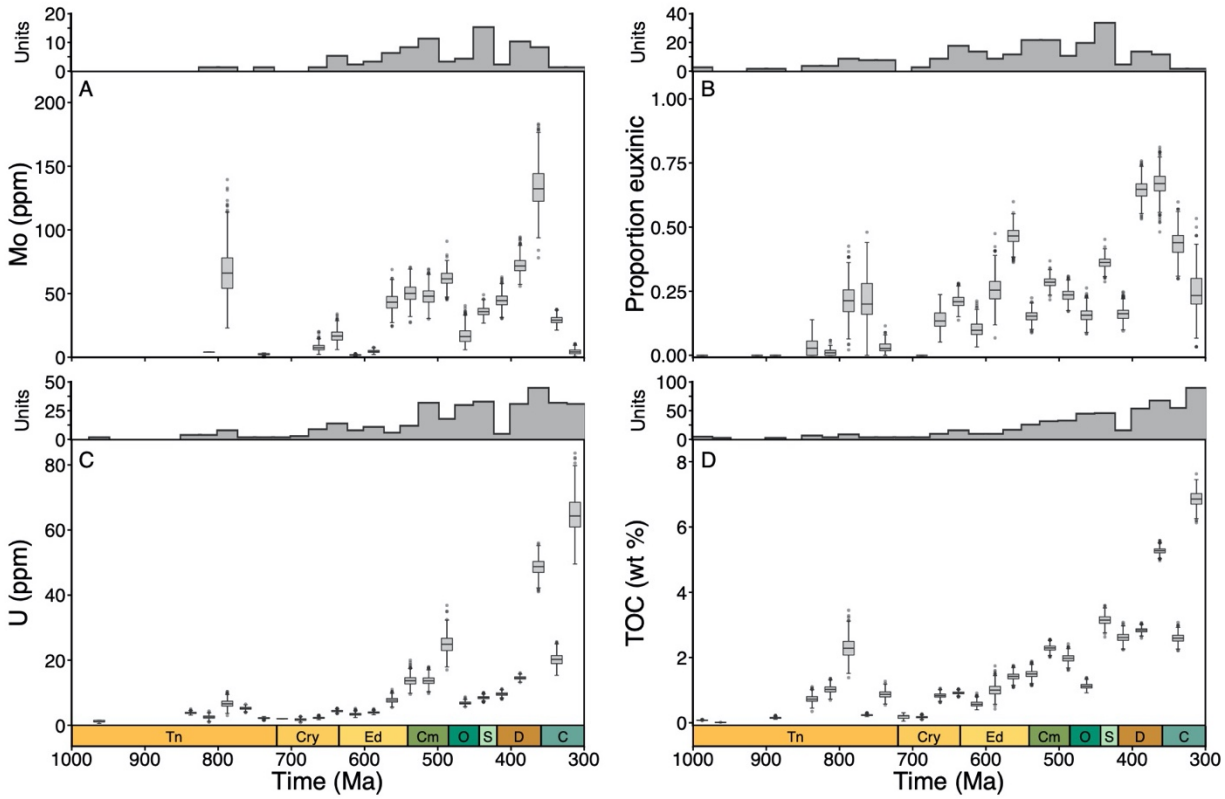


Figure 1: Spatial-temporal weighted bootstrapped means of key geochemical proxies from sampled shales. A) molybdenum concentrations in euxinic shales, B) proportion of anoxic shales that are euxinic based on iron speciation, C) uranium concentrations in anoxic shales, D) total organic carbon (TOC) in all shales. Box and whisker plots illustrate the distribution of 1000 weighted bootstrapped means per 25-million-year time bin. The weighting algorithm inverse weights samples based upon their spatial and temporal proximity to other samples in the time bin[24]. Histograms show the number of lithostratigraphic units used in the bootstrap analyses for each time bin.

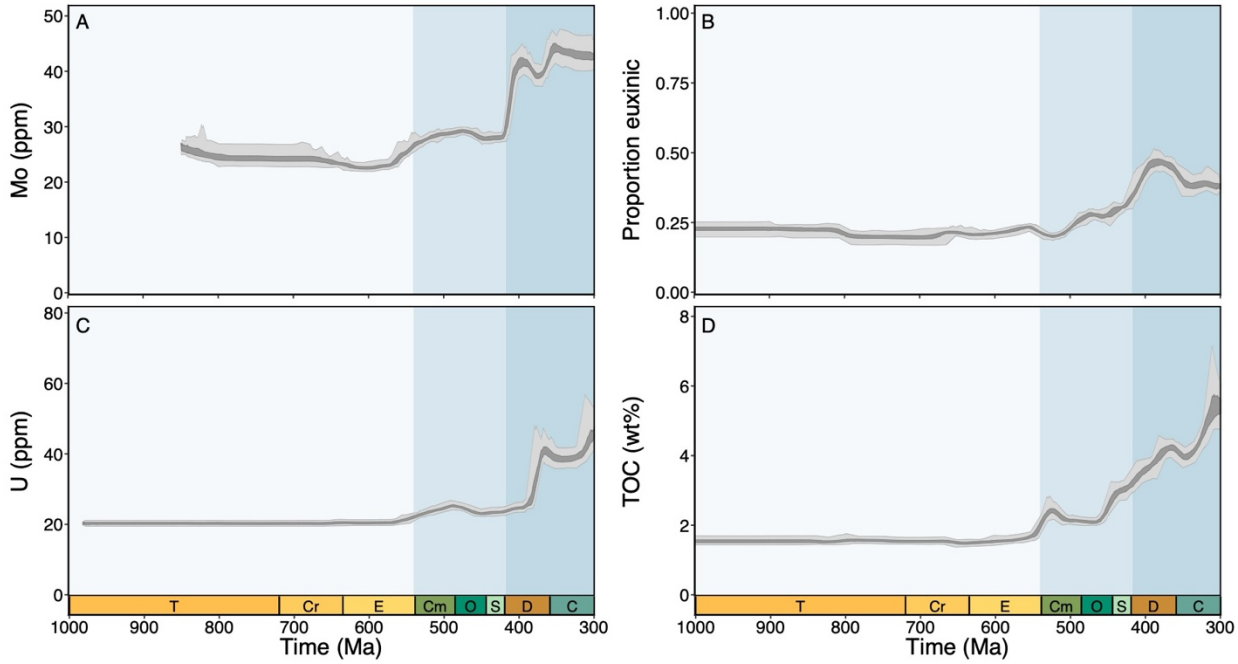


Figure 2: Statistical reconstructions of deconvolved marine biogeochemical signals for key geochemical proxies in sampled shales. Partial dependence plots illustrate the marginal effect of geologic time on A) molybdenum, B) proportion of euxinic depositional environments, C) uranium, D) total organic carbon, when all other variables expected to influence the incorporation of these proxies into fine-grained sedimentary archives are held constant. Dark grey envelopes represent the 25th to 75th percentiles of the distribution of interpolated partial dependence plot values from 100 Monte Carlo random forest analyses for each timestep, light grey envelopes represent the 5th to 95th percentiles of the same distributions. See Extended Data Table 2 for full model predictor variables.

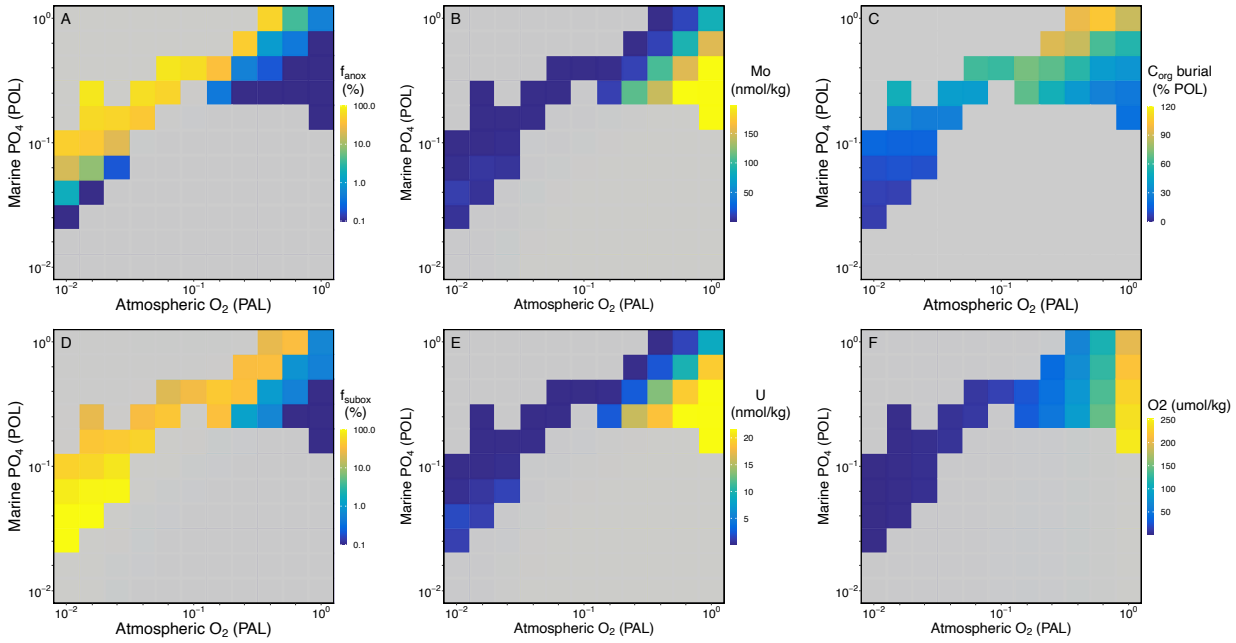


Figure 3: Heatmaps of key biogeochemical variables from combined modeling approach for each logarithmically scaled atmospheric pO_2 and marine PO_4 scenario. A) f_{anox} – the fractional extent of anoxic bottom waters in cGENIE global ocean models, B) global seawater Mo concentrations, C) global marine organic carbon burial rates as a function of present oceanic levels, D) f_{subox} – the fractional extent of suboxic bottom waters in cGENIE global ocean models, E) global seawater U concentrations, F) mean shelf dissolved [O_2]. Only atmospheric pO_2 and marine PO_4 scenarios that are expected to be stable on geologic timescales based on Cole et al.[27] are included.

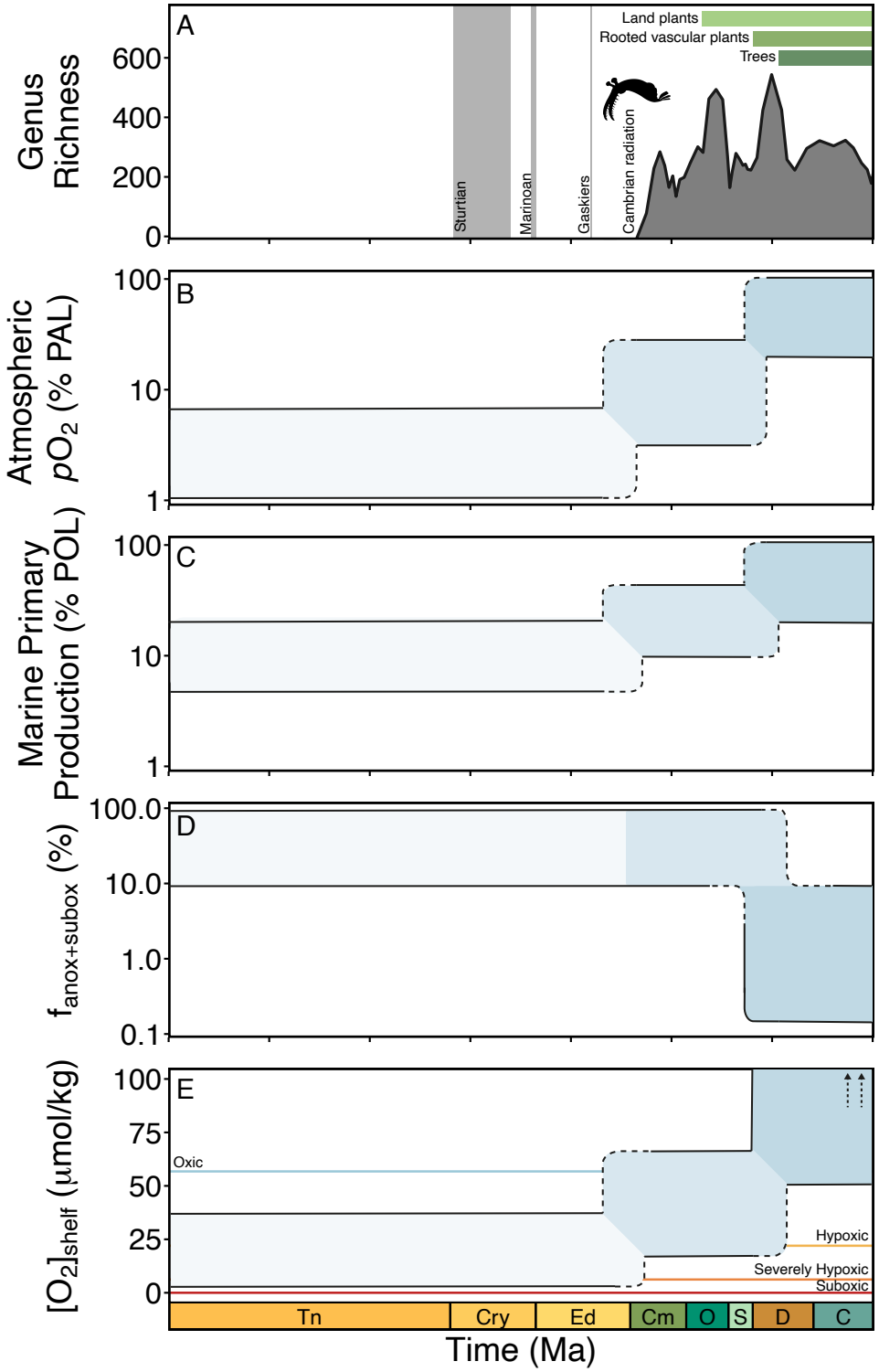
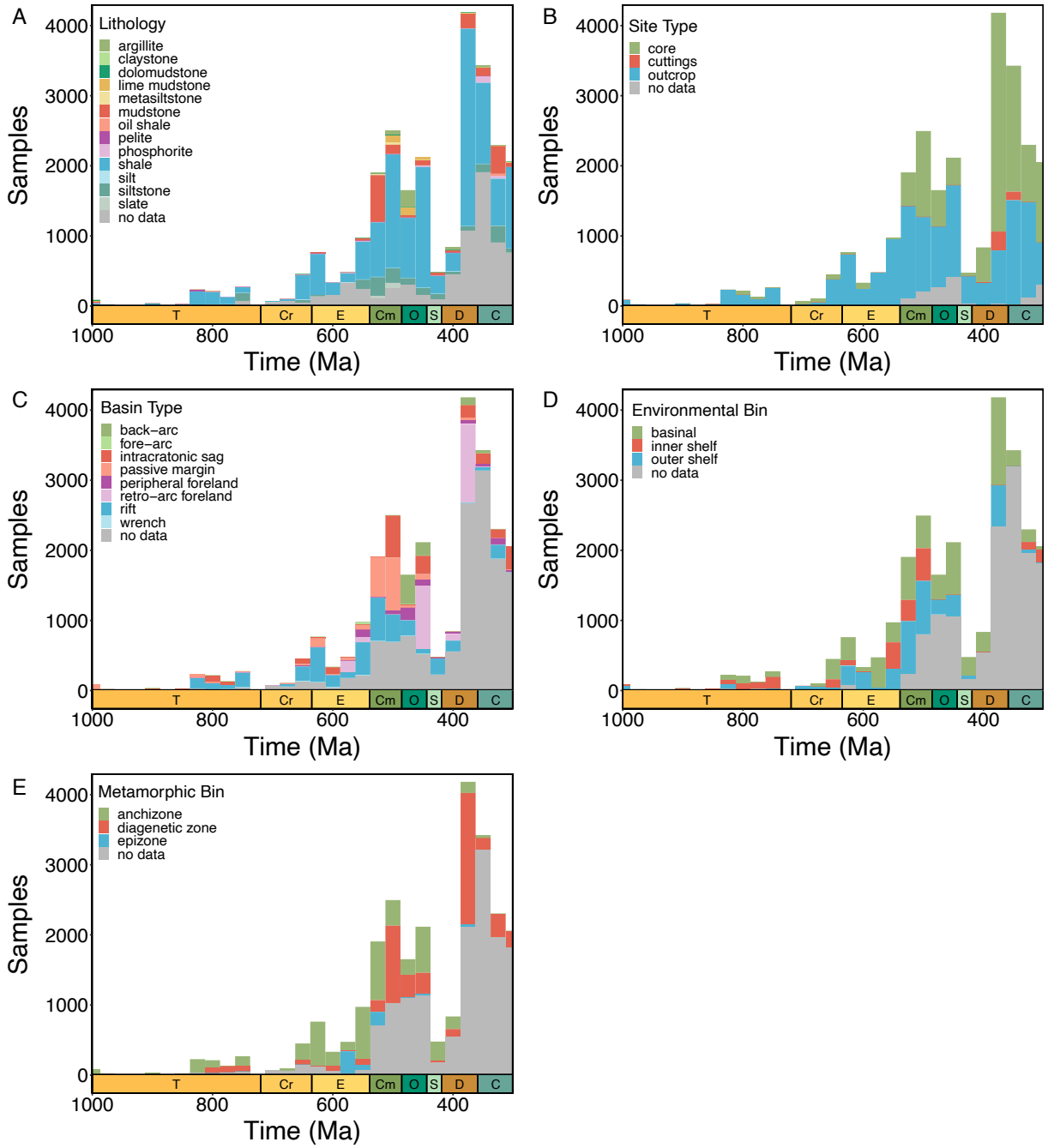
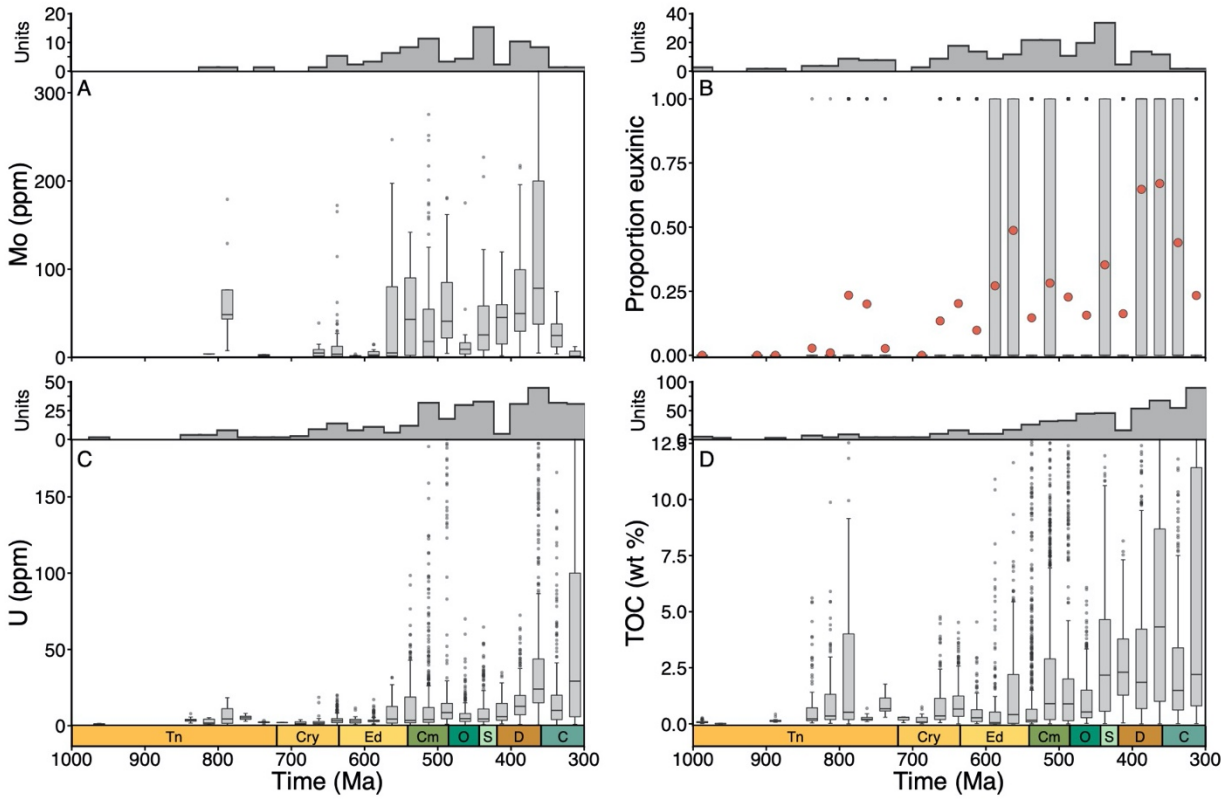


Figure 4: Summary of reconstructed Neoproterozoic-Paleozoic Earth system evolution. A) Marine animal genus richness[44], dates of Snowball Earth events[16], evolution of land plants[45], B) estimated ranges for atmospheric pO_2, C) estimated ranges for marine primary production, D) estimated ranges for the extent of reducing seafloor ($f_{anox+subox}$. or 1- f_{ox}), E) estimated ranges for

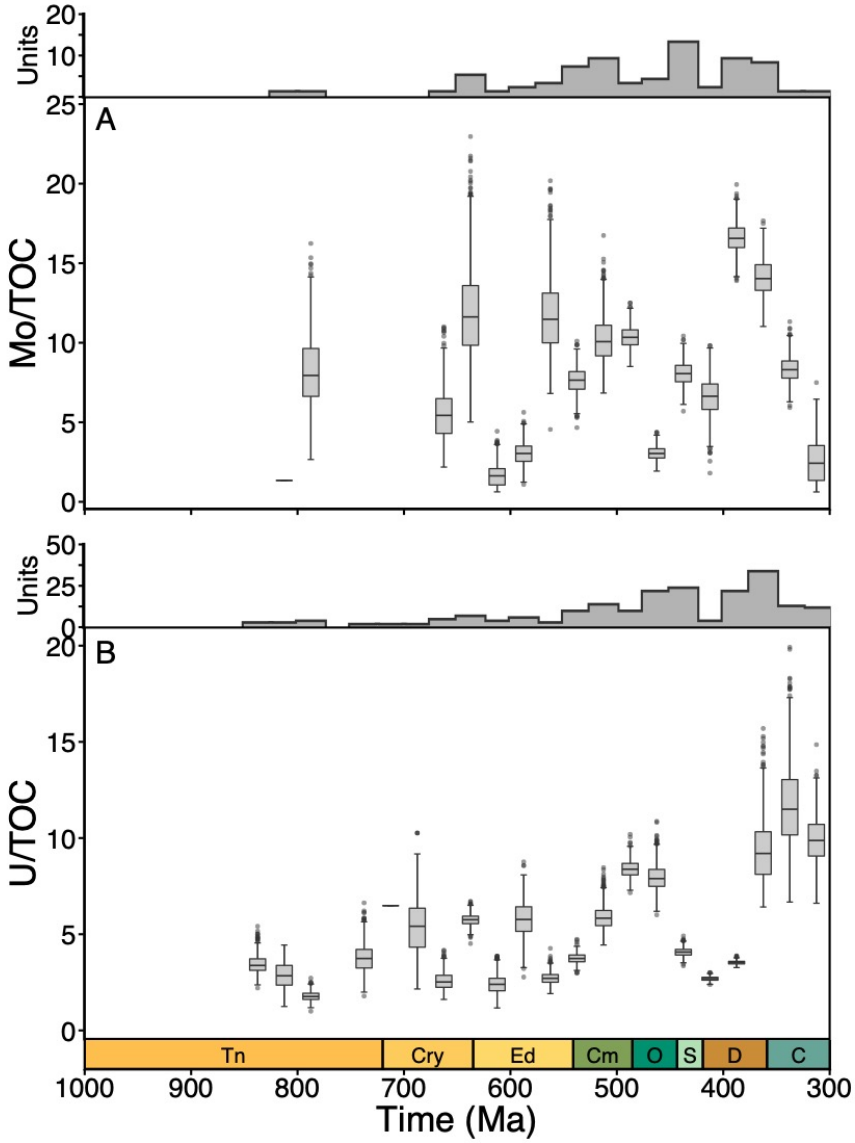
mean dissolved oxygen concentrations in shelf environments, with hypoxic thresholds[6] based on modern ecophysiology. *Anomalocaris canadensis* silhouette credit: Caleb M. Brown.



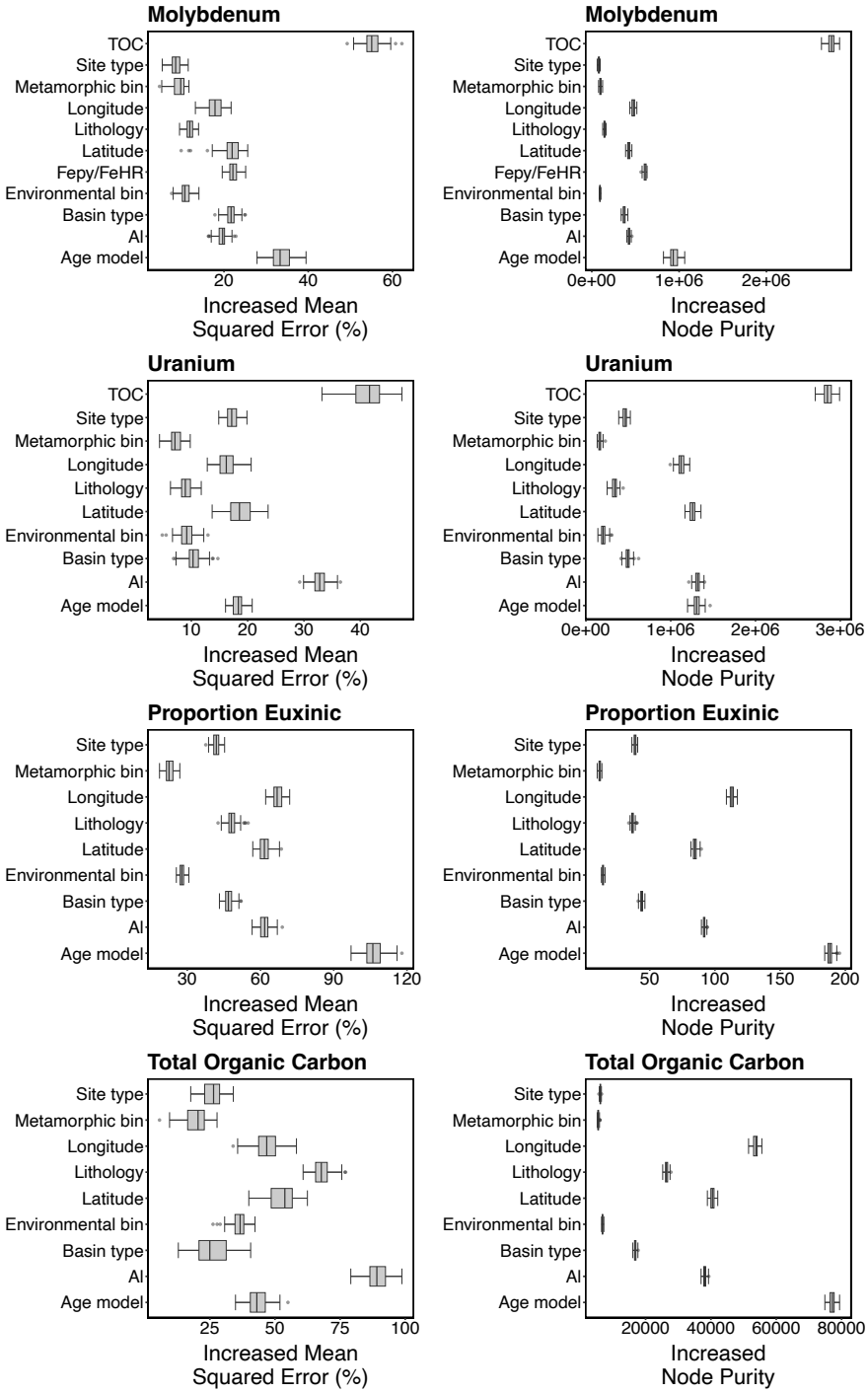
Extended Figure 1: Histograms showing the distribution of geologic context variables for the primary SGP dataset used in this study. A) Lithology, B) Site Type, C) Basin Type, D) Environmental Bin, E) Metamorphic Bin. Details regarding these geological and geographic context variables can be found on the SGP wiki (https://github.com/ufarrell/sgp_phase1/wiki/Database-description#geological-context).



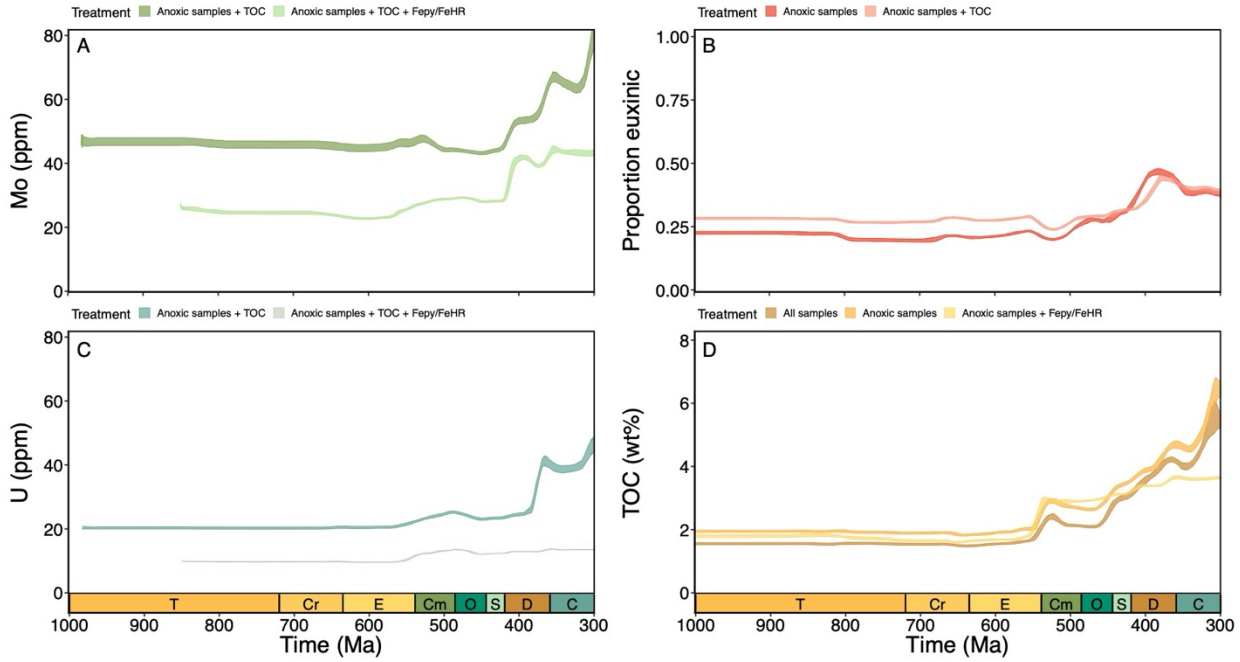
Extended Figure 2: Distributions of raw geochemical data in sub-datasets used in spatial-temporal weighted bootstrap analyses (Fig. 1 of main text). A) molybdenum concentrations in euxinic shales, B) proportion of anoxic shales that are euxinic based on iron speciation, C) uranium concentrations in anoxic shales, D) total organic carbon (TOC) in all shales. Box and whisker plots illustrate the distribution of all data. As box and whisker plots of proportion euxinic data are not very informative about the structure of the data, we overlay red data points illustrating the mean of the data for each time bin for panel B. Histograms show the number of lithostratigraphic units used in the bootstrap analyses for each time bin.



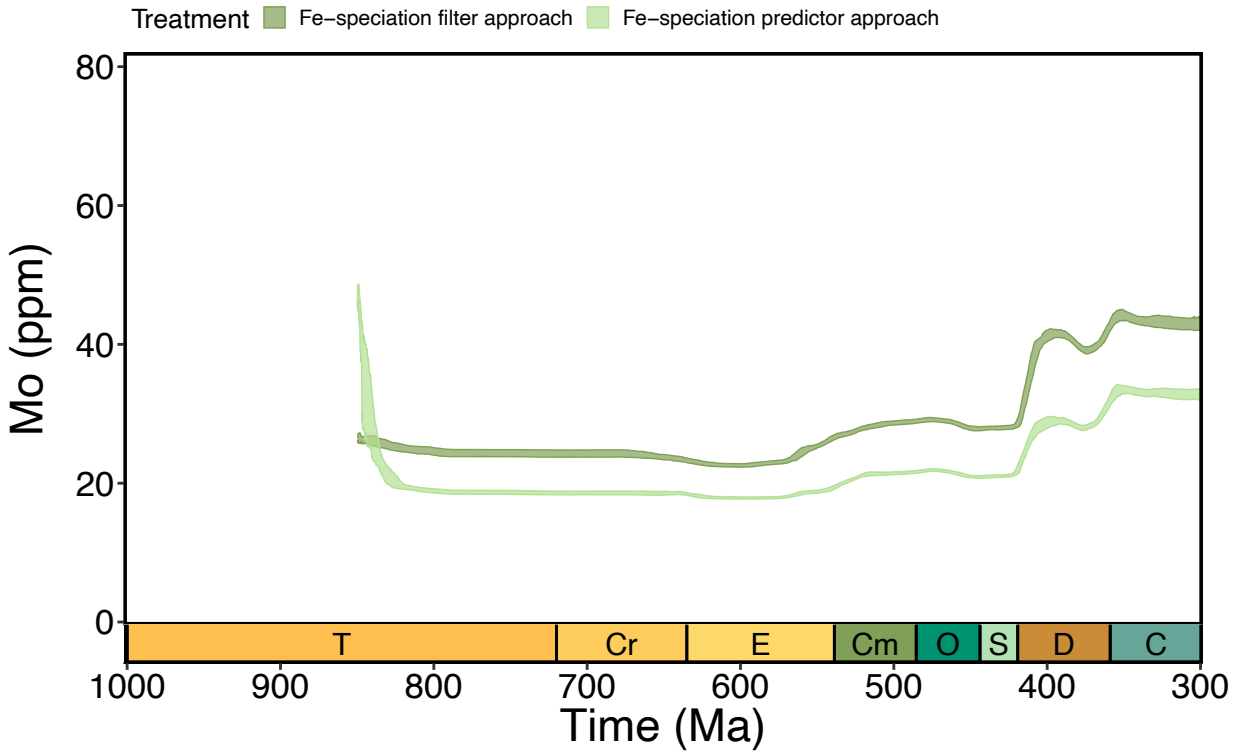
Extended Figure 3: Spatial-temporal weighted bootstrapped means of key geochemical proxies from sampled shales. A) Mo/TOC ratios in euxinic shales, B) U/TOC ratios in anoxic shales. Box and whisker plots illustrate the distribution of 1000 weighted bootstrapped means per 25 million year time bin. Spatial-temporal weighting algorithm inverse weights samples based upon their spatial and temporal proximity to other samples in the time bin. Histograms show the number of lithostratigraphic units used in the bootstrap analyses for each time bin.



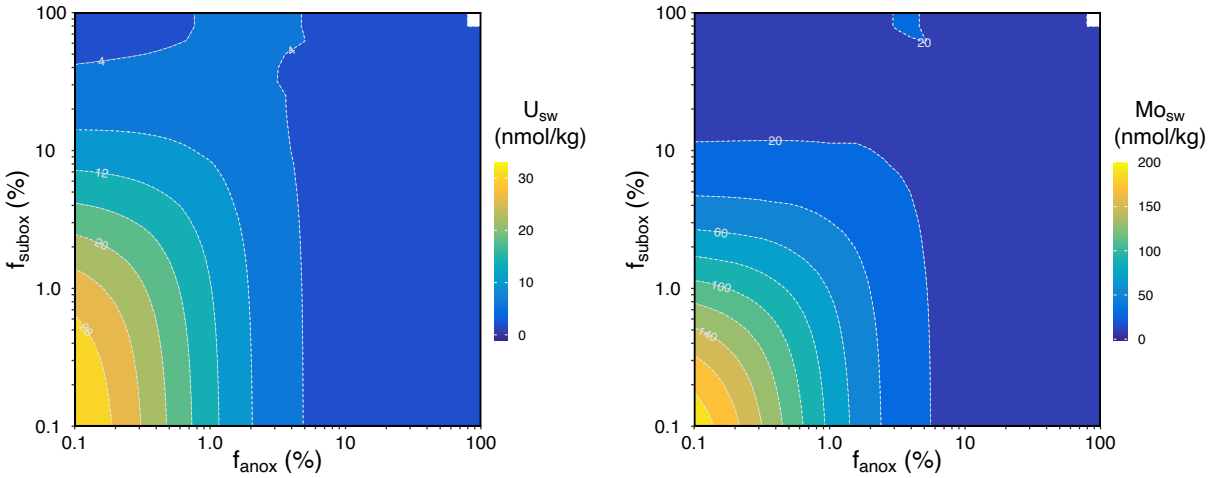
Extended Figure 4: Random forest variable importance plots. A, C, E, and G show increased mean squared error estimates for molybdenum, uranium, proportion euxinic, and total organic carbon analyses, respectively. B, D, F, and H show increased node purity estimates for molybdenum, uranium, proportion euxinic, and total organic carbon analyses, respectively. Box and whisker plots summarize the results from the 100 Monte Carlo random forest analyses presented in the main text for each proxy. For both variable importance metrics, higher values indicate that a variable is more important in determining the predictions of the model.



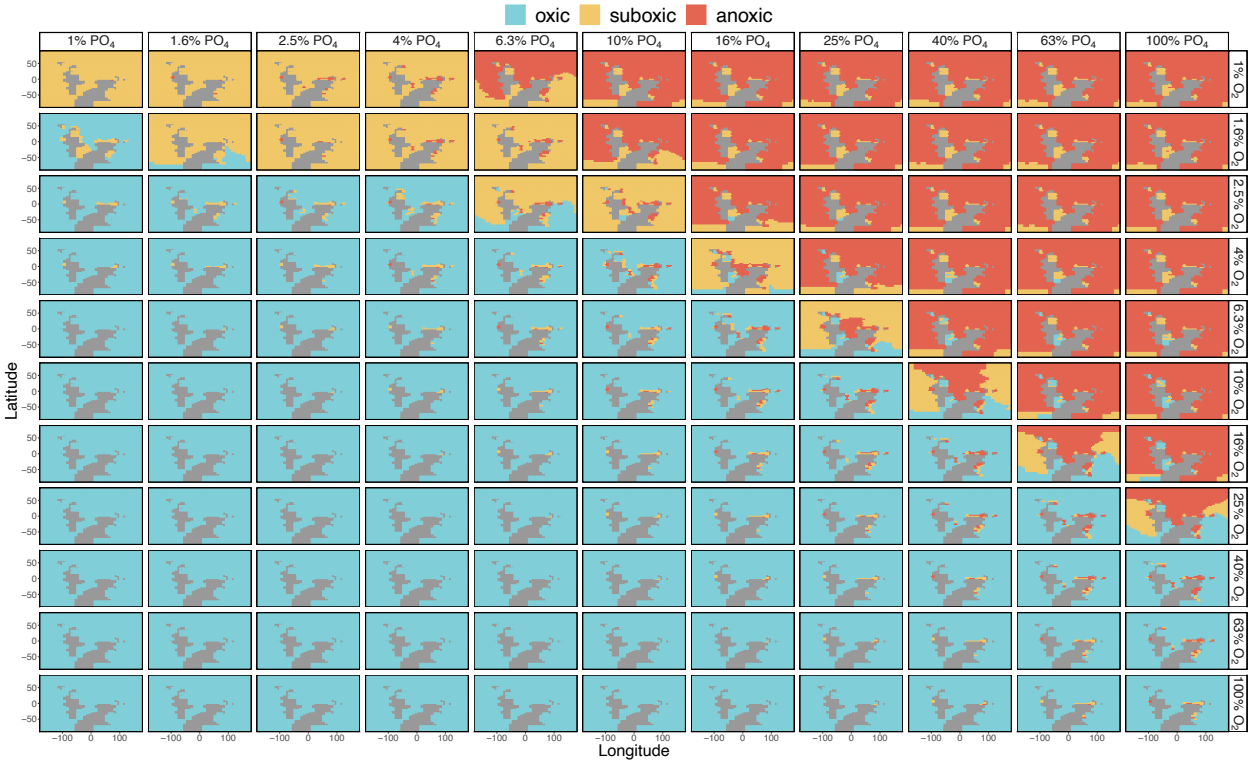
Extended Figure 5: Impact of alternative redox treatments on random forest analyses. Partial dependence plots illustrate the marginal effect of geologic time on A) molybdenum, B) proportion of euxinic depositional environments, C) uranium, D) total organic carbon. All analyses included here include the full suite of geologic context variables used in all random forest analyses. Redox filters are color-coded for each proxy. Envelopes represent the 25th to 75th percentiles of the distribution of interpolated partial dependence plot values from 100 Monte Carlo random forest analyses for each timestep. See Extended Data Tables 2-3 for full model predictor variables.



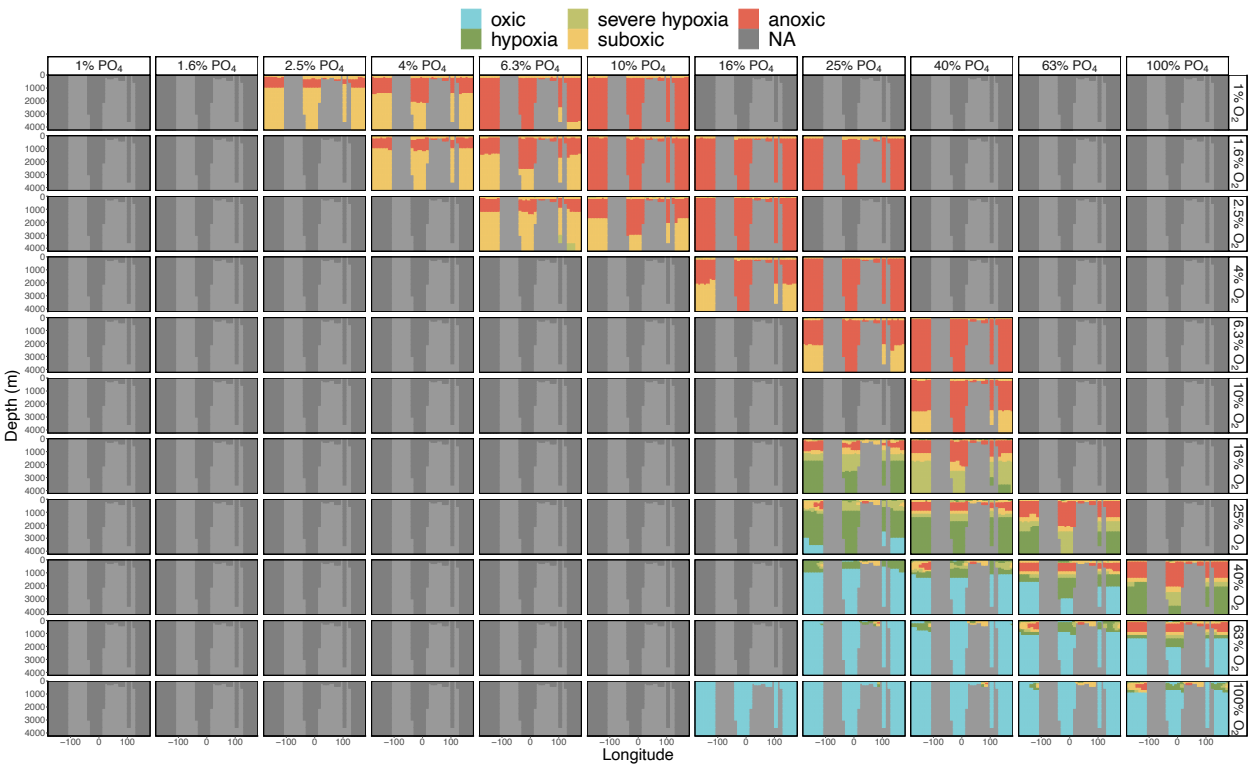
Extended Figure 6: Impact of using iron speciation values as predictors rather than filters (as in main text) in molybdenum random forest analyses. These analyses eliminate any assumptions linked to specific iron speciation thresholds.



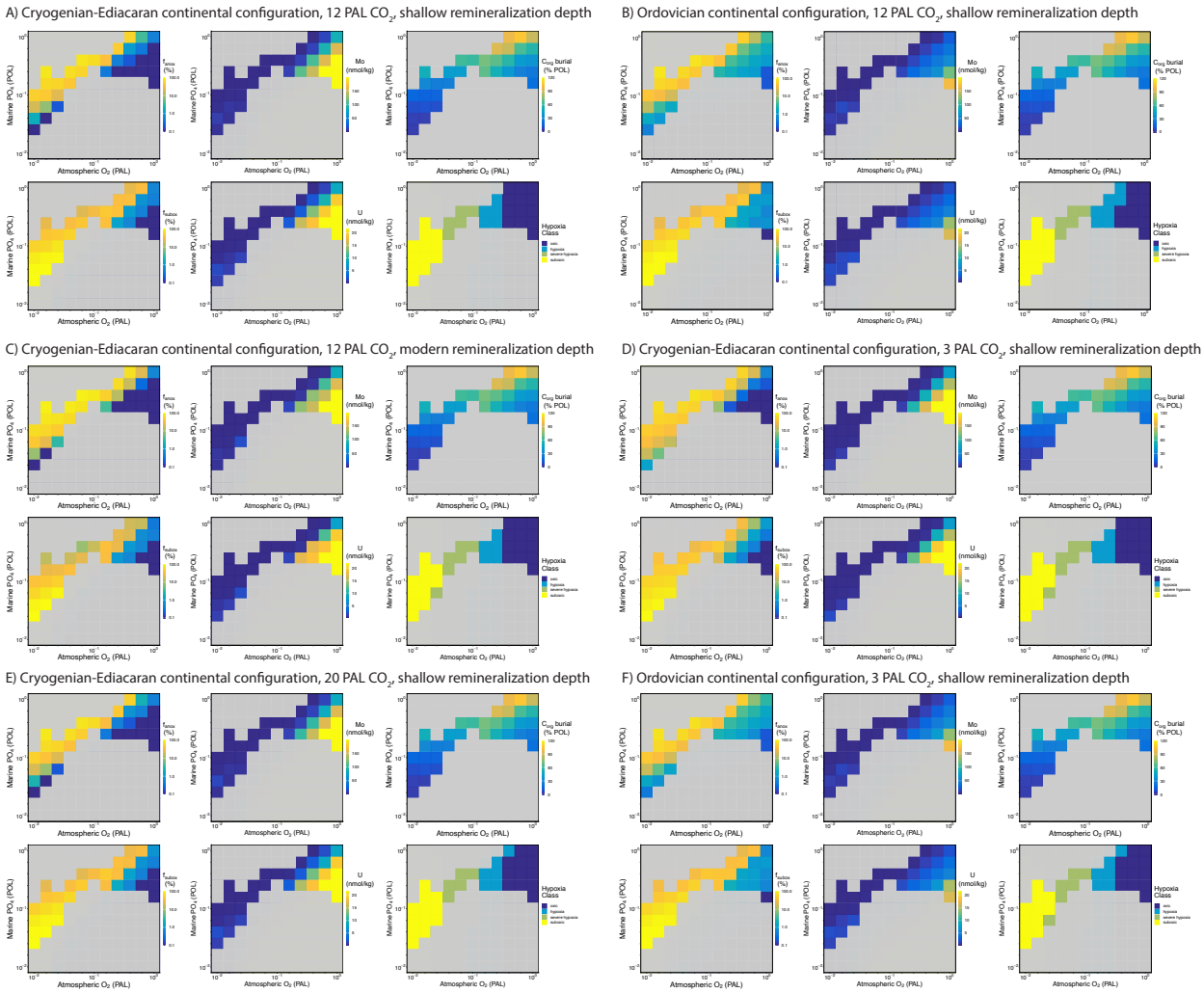
Extended Figure 7: Mo-U mass balance global sensitivity analysis. Impact of f_{anox} and f_{subox} on the estimated concentrations of molybdenum and uranium in seawater using the Mo-U mass balance model of Stockey et al.[30].



Extended Figure 8: Maps of redox classifications for cGENIE bottom waters. Maps illustrate bottom water dissolved oxygen concentrations binned into oxic (dissolved [O2] ≥ 4.8 μmol/kg), suboxic (dissolved [O2] > 0 μmol/kg; dissolved [O2] ≤ 4.8 μmol/kg) and anoxic (dissolved [O2] ≤ 0 μmol/kg) categories for the cGENIE ensemble experiment presented in Figure 3.



Extended Figure 9: Hypoxia classification of cGENIE equatorial transects. Ocean transects show equatorial cross sections of the three-dimensional ocean models presented in Figure 3 binned into anoxic, suboxic, severe hypoxia, hypoxia and oxic categories (following Sperling et al.[6]) to illustrate the impact of oxygen-productivity scenarios on broad physiological thresholds for marine animals.



Extended Figure 10: cGENIE configuration sensitivity. In this sensitivity analysis, we replicate Figure 3 for A) Cryogenian-Ediacaran continental configuration at 12 PAL CO_2 with shallow remineralization depth (Fig. 3), B) Ordovician continental configuration at 12 PAL CO_2 with shallow remineralization depth, C) Cryogenian-Ediacaran continental configuration at 12 PAL CO_2 with modern remineralization depth, D) Cryogenian-Ediacaran continental configuration at 3 PAL CO_2 with shallow remineralization depth, E) Cryogenian-Ediacaran continental configuration at 20 PAL CO_2 with shallow remineralization depth, F) Ordovician continental configuration at 3 PAL CO_2 with shallow remineralization depth.

Extended Data Table 1: Data filtering and biogeochemical justifications for spatial-weighted bootstrap mean analyses.

Proxy	Required geochemical data	Filters	Description	Figure
Mo	Mo, Fe_{HR}/Fe_T, Fe_{Py}/Fe_{HR}	$Fe_{HR}/Fe_T \geq 0.38$ AND $Fe_{Py}/Fe_{HR} \geq 0.7$	Mo in euxinic shale only	1A
U	U, [Fe_{HR}/Fe_T OR Fe/Al]	$Fe_{HR}/Fe_T \geq 0.38$ OR Fe/Al ≥ 0.53	U in anoxic shale only	1C
Proportion euxinic (binary coding based on Fe_{Py}/Fe_{HR})	Fe_{HR}/Fe_T, Fe_{Py}/Fe_{HR}	$Fe_{HR}/Fe_T \geq 0.38$	Proportion euxinic in anoxic shale only	1B
TOC	TOC	None	TOC in all shales	1D
Mo/TOC	Mo, Fe_{HR}/Fe_T, Fe_{Py}/Fe_{HR}, TOC	$Fe_{HR}/Fe_T \geq 0.38$ AND $Fe_{Py}/Fe_{HR} \geq 0.7$ AND TOC ≥ 0.3 (wt %)	Mo/TOC in euxinic shale only	Extended 3A
U/TOC	U, [Fe_{HR}/Fe_T OR Fe/Al], TOC	[$Fe_{HR}/Fe_T \geq 0.38$ OR Fe/Al ≥ 0.53] AND TOC ≥ 0.3 (wt %)	U/TOC in anoxic shale only	Extended 3B

765 **Extended Data Table 2:** Data filtering, predictor variables and biogeochemical justifications for
766 primary Monte Carlo random forest analyses. In each row, the identified geochemical proxy is
767 the response variable of the random forest analysis.
768

Proxy	Required geochemical data	Filters	Predictor variables	Description	Figure
Mo	Mo, Fe_{HR}/Fe_T, Fe_{Py}/Fe_{HR}, TOC	$Fe_{HR}/Fe_T \geq 0.38$	Age model, site type, metamorphic bin, basin type, site latitude, site longitude, lithology name, environmental bin, TOC, Fe_{Py}/Fe_{HR}, Al	Samples from anoxic environments only. Control for depositional environment, post-depositional alteration, organic carbon loading, detrital input and sulfide levels.	2A
U	U, [Fe_{HR}/Fe_T OR Fe/Al], TOC	$Fe_{HR}/Fe_T \geq 0.38$ OR $Fe/Al \geq 0.53$	Age model, site type, metamorphic bin, basin type, site latitude, site longitude, lithology name, environmental bin, TOC, Al	Samples from anoxic environments only. Control for depositional environment, post-depositional alteration, organic carbon loading and detrital input.	2C
Proportion euxinic (binary coding based on Fe_{Py}/Fe_{HR})	Fe_{HR}/Fe_T, Fe_{Py}/Fe_{HR}	$Fe_{HR}/Fe_T \geq 0.38$	Age model, site type, metamorphic bin, basin type, site latitude, site longitude, lithology name, environmental bin, Al	Samples from anoxic environments only. Control for depositional environment, post-depositional alteration	2B

				and detrital input.	
TOC	TOC	None	Age model, site type, metamorphic bin, basin type, site latitude, site longitude, lithology name, environmental bin, Al	Control for depositional environment, post-depositional alteration and detrital input.	2D

769
770

Extended Data Table 3: Data filtering, predictor variables and biogeochemical justifications for supplemental Monte Carlo random forest analyses. In each row, the identified geochemical proxy is the response variable of the random forest analysis.

Proxy	Required geochemical data	Filters	Predictor variables	Description	Figure
Mo	Mo, [Fe_{HR}/Fe_T OR Fe/Al], TOC	$Fe_{HR}/Fe_T \geq 0.38$ OR $Fe/Al \geq 0.53$	Age model, site type, metamorphic bin, basin type, site latitude, site longitude, lithology name, environmental bin, TOC, Al	Samples from anoxic environments only. Control for depositional environment, post-depositional alteration, organic carbon loading and detrital input.	Extended 5A
U	U, Fe_{HR}/Fe_T, Fe_{Py}/Fe_{HR}, TOC	$Fe_{HR}/Fe_T \geq 0.38$	Age model, site type, metamorphic bin, basin type, site latitude, site longitude, lithology name, environmental bin, TOC, Fe_{Py}/Fe_{HR}, Al	Samples from anoxic environments only. Control for depositional environment, post-depositional alteration, organic carbon loading, detrital input and sulfide levels.	Extended 5C
Proportion euxinic (binary coding based on Fe_{Py}/Fe_{HR})	Fe_{HR}/Fe_T, Fe_{Py}/Fe_{HR}, TOC	$Fe_{HR}/Fe_T \geq 0.38$	Age model, site type, metamorphic bin, basin type, site latitude, site longitude, lithology name, environmental bin, TOC, Al	Samples from anoxic environments only. Control for depositional environment, post-depositional alteration, organic carbon	Extended 5B

				loading and detrital input.	
TOC	TOC, [Fe_{HR}/Fe_T OR Fe/Al]	$Fe_{HR}/Fe_T \geq 0.38$ OR $Fe/Al \geq 0.53$	Age model, site type, metamorphic bin, basin type, site latitude, site longitude, lithology name, environmental bin, Al	Samples from anoxic environments only. Control for depositional environment, post-depositional alteration and detrital input.	Extended 5D
TOC	TOC, Fe_{HR}/Fe_T, Fe_{Py}/Fe_{HR}	$Fe_{HR}/Fe_T \geq 0.38$	Age model, site type, metamorphic bin, basin type, site latitude, site longitude, lithology name, environmental bin, Fe_{Py}/Fe_{HR}, Al	Samples from anoxic environments only. Control for depositional environment, post-depositional alteration, detrital input and sulfide levels.	Extended 5D

775
776

777

References

1. Lyons, T. W., Reinhard, C. T. & Planavsky, N. J. The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307–315 (2014).
2. Farquhar, J., Bao, H. & Thiemens, M. Atmospheric influence of Earth's earliest sulfur cycle. *Science* **289**, 756–758 (2000).
3. Crockford, P. W. *et al.* Triple oxygen isotope evidence for limited mid-Proterozoic primary productivity. *Nature 2018 559:7715* **559**, 613–616 (2018).
4. Kump, L. R. The rise of atmospheric oxygen. *Nature* **451**, 277–278 (2008).
5. Cole, D. B. *et al.* On the co-evolution of surface oxygen levels and animals. *Geobiology* vol. 18 260–281 (2020).
6. Sperling, E. A., Knoll, A. H. & Girguis, P. R. The ecological physiology of Earth's second oxygen revolution. *Annu Rev Ecol Evol Syst* **46**, 215–235 (2015).
7. Stolper, D. A. & Keller, C. B. A record of deep-ocean dissolved O_2 from the oxidation state of iron in submarine basalts. *Nature* **553**, 323–327 (2018).
8. Sperling, E. A. *et al.* A long-term record of early to mid-Paleozoic marine redox change. *Sci Adv* **7**, eabf4382 (2021).
9. Dahl, T. W. *et al.* Devonian rise in atmospheric oxygen correlated to the radiations of terrestrial plants and large predatory fish. *Proc Natl Acad Sci U S A* **107**, 17911–17915 (2010).
10. Lu, W. *et al.* Late inception of a resiliently oxygenated upper ocean. *Science* **361**, 174–177 (2018).
11. Wallace, M. W. *et al.* Oxygenation history of the Neoproterozoic to early Phanerozoic and the rise of land plants. *Earth Planet Sci Lett* **466**, 12–19 (2017).
12. Krause, A. J., Mills, B. J. W., Merdith, A. S., Lenton, T. M. & Poulton, S. W. Extreme variability in atmospheric oxygen levels in the late Precambrian. *Sci Adv* **8**, eabm8191 (2022).
13. Wei, G. Y. *et al.* Global marine redox evolution from the late Neoproterozoic to the early Paleozoic constrained by the integration of Mo and U isotope records. *Earth Sci Rev* **214**, 103506 (2021).
14. Sahoo, S. K. *et al.* Oceanic oxygenation events in the anoxic Ediacaran ocean. *Geobiology* **14**, 457–468 (2016).
15. Pohl, A. *et al.* Continental configuration controls ocean oxygenation during the Phanerozoic. *Nature* **608**, 523–527 (2022).
16. Sperling, E. A. & Stockey, R. G. The temporal and environmental context of early animal evolution: Considering all the ingredients of an "explosion". *Integr Comp Biol* **58**, 605–622 (2018).
17. Berner, R. A. *The Phanerozoic Carbon Cycle*. (Oxford University Press, 2004). doi:10.1093/oso/9780195173338.001.0001.
18. Tribovillard, N., Algeo, T. J., Lyons, T. & Riboulleau, A. Trace metals as paleoredox and paleoproductivity proxies: An update. *Chem Geol* **232**, 12–32 (2006).
19. Scott, C. *et al.* Tracing the stepwise oxygenation of the Proterozoic ocean. *Nature* **452**, 456–459 (2008).
20. Partin, C. A. *et al.* Large-scale fluctuations in Precambrian atmospheric and oceanic oxygen levels from the record of U in shales. (2013) doi:10.1016/j.epsl.2013.03.031.
21. Sahoo, S. K. *et al.* Ocean oxygenation in the wake of the Marinoan glaciation. *Nature* **489**, 546–549 (2012).

22. Farrell, Ú. C. *et al.* The Sedimentary Geochemistry and Paleoenvironments Project. *Geobiology* **19**, 545–556 (2021).
23. Morford, J. L. & Emerson, S. The geochemistry of redox sensitive trace metals in sediments. *Geochim Cosmochim Acta* **63**, 1735–1750 (1999).
24. Mehra, A. *et al.* Curation and analysis of global sedimentary geochemical data to inform Earth history. *GSA Today* **31**, 4–9 (2021).
25. Sperling, E. A. *et al.* Statistical analysis of iron geochemical data suggests limited late Proterozoic oxygenation. *Nature* **523**, 451–454 (2015).
26. Ridgwell, A. *et al.* Marine geochemical data assimilation in an efficient Earth system model of global biogeochemical cycling. *Biogeosciences* **4**, 87–104 (2007).
27. Cole, D. B., Ozaki, K. & Reinhard, C. T. Atmospheric oxygen abundance, marine nutrient availability, and organic carbon fluxes to the seafloor. *Global Biogeochem Cycles* **36**, e2021GB007052 (2022).
28. Ozaki, K. & Tajika, E. Biogeochemical effects of atmospheric oxygen concentration, phosphorus weathering, and sea-level stand on oceanic redox chemistry: Implications for greenhouse climates. *Earth Planet Sci Lett* **373**, 129–139 (2013).
29. Pasquier, V., Fike, D. A., Révillon, S. & Halevy, I. A global reassessment of the controls on iron speciation in modern sediments and sedimentary rocks: A dominant role for diagenesis. *Geochim Cosmochim Acta* **335**, 211–230 (2022).
30. Stockey, R. G. *et al.* Persistent global marine euxinia in the early Silurian. *Nat Commun* **11**, 1–10 (2020).
31. Alcott, L. J., Mills, B. J. W. & Poulton, S. W. Stepwise Earth oxygenation is an inherent property of global biogeochemical cycling. *Science* **366**, 1333–1337 (2019).
32. Brocks, J. J. The transition from a cyanobacterial to algal world and the emergence of animals. *Emerging Topics in Life Sciences* vol. 2 181–190 Preprint at https://doi.org/10.1042/ETLS20180039 (2018).
33. Laakso, T. A., Sperling, E. A., Johnston, D. T. & Knoll, A. H. Ediacaran reorganization of the marine phosphorus cycle. *Proc Natl Acad Sci U S A* **117**, 11961–11967 (2020).
34. Dahl, T. W. & Hammarlund, E. U. Do large predatory fish track ocean oxygenation? *Commun Integr Biol* **4**, 92–4 (2011).
35. Boag, T. H., Gearty, W. & Stockey, R. G. Metabolic tradeoffs control biodiversity gradients through geological time. *Current Biology* **31**, 2906-2913.e3 (2021).
36. Penn, J. L., Deutsch, C., Payne, J. L. & Sperling, E. A. Temperature-dependent hypoxia explains biogeography and severity of end-Permian marine mass extinction. *Science* **362**, eaat1327 (2018).
37. Poulton, S. W. & Canfield, D. E. Ferruginous conditions: A dominant feature of the ocean through Earth's history. *Elements* **7**, 107–112 (2011).
38. Raiswell, R. *et al.* The iron paleoredox proxies: A guide to the pitfalls, problems and proper practice. *Am J Sci* **318**, 491–526 (2018).
39. Molnar, C. *Interpretable machine learning - A guide for making black box models explainable*. *Lulu. com* (2020).
40. Reinhard, C. T. *et al.* The impact of marine nutrient abundance on early eukaryotic ecosystems. *Geobiology* **18**, 139–151 (2020).
41. Meyer, K. M., Ridgwell, A. & Payne, J. L. The influence of the biological pump on ocean chemistry: implications for long-term trends in marine redox chemistry, the global carbon cycle, and marine animal ecosystems. *Geobiology* **14**, 207–219 (2016).

42. Stockey, R. G., Pohl, A., Ridgwell, A., Finnegan, S. & Sperling, E. A. Decreasing Phanerozoic extinction intensity as a consequence of Earth surface oxygenation and metazoan ecophysiology. *Proc Natl Acad Sci U S A* **118**, 2021 (2021).
43. Reinhard, C. T. *et al.* Proterozoic ocean redox and biogeochemical stasis. *Proc Natl Acad Sci U S A* **110**, 5357–62 (2013).
44. Kocsis, Á. T., Reddin, C. J., Alroy, J. & Kiessling, W. The r package divDyn for quantifying diversity dynamics using fossil sampling data. *Methods Ecol Evol* **10**, 735–743 (2019).
45. D'Antonio, M. P., Ibarra, D. E. & Boyce, C. K. Land plant evolution decreased, rather than increased, weathering rates. *Geology* **48**, 29–33 (2020).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SGPAppendix1.pdf
- SGPAppendix2.pdf
- SGPAppendix3.pdf