Geochemical investigation of the co-evolution of life and environment in the Neoproterozoic Era

Junyao Kang

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Geosciences

Shuhai Xiao, Chair

Benjamin Gill, Member

Madeline Schreiber. Member

Rachel Reid. Member

February 2nd, 2024

Blacksburg, Virginia

Keywords: Neoproterozoic, Nitrogen Isotope, Iron Speciation, Redox, Tonian, Nitrate, Eukaryotes, Carbon Isotope, North China Craton, Chemostratigraphy, Sulfur Isotope, Pyrite, Trace Element, LA-ICP-MS, Machine Learning, Random Forest, XGBoost

Geochemical investigation of the co-evolution of life and environment in the Neoproterozoic Era

Junyao Kang

ABSTRACT

The co-evolution of life and the environment stands as a cornerstone in Earth's 4.5-billion-year history. Environmental fluctuations have wielded substantial influence over biological evolution, while life forms have, in turn, reshaped Earth's surface and climate. This dissertation centers on a critical period in Earth's history—the Neoproterozoic Era—when profound environmental shifts potentially catalyzed pivotal eukaryotic evolutionary events. By delving deeper into Neoproterozoic paleoenvironments, I aim at a clearer understanding of life-environment co-evolution in this crucial era.

The first chapter focuses on an important juncture—the transition from prokaryote to eukaryote dominance in marine ecosystems during the Tonian Period (1000 Ma to 720 Ma). To assess whether the availability of nitrate, an important macro-nutrient, played a critical role in this evolutionary event, nitrogen isotope compositions ($\delta^{15}N$) of marine carbonates from the early Tonian (ca. 1000 Ma to ca. 800 Ma) Huaibei Group in North China were measured. The data indicate nitrate limitation in early Neoproterozoic oceans. Further, a compilation of Proterozoic sedimentary $\delta^{15}N$ data, together with box model simulations, suggest a ~50% increase in marine nitrate availability at ~800 Ma. Limited nitrate availability in early Neoproterozoic oceans may have delayed the ecological rise of eukaryotes until ~800 Ma when increased nitrate supply, together with other environmental and ecological factors, may have contributed to the transition from prokaryote-dominant to eukaryote-dominant marine ecosystems.

Recognizing the spatial and temporal variations in Neoproterozoic oceanic environments, the second chapter lays the groundwork for a robust stratigraphic framework for the early Tonian Period. Employing the dynamic time warping algorithm, I constructed a global stratigraphic framework for the early Tonian Period using $\delta^{13}C_{carb}$ data from the North China, São Francisco, and Congo cratons. This exercise confirms the generally narrow range of $\delta^{13}C_{carb}$ fluctuations in the early Tonian, but also confirms the presence of a negative $\delta^{13}C_{carb}$ excursion of notable magnitude (~9 ‰) at ca. 920 Ma in multiple records, suggesting that it was global in scope. This negative excursion, known as the Majiatun excursion, is likely the oldest negative excursion in the Neoproterozoic Era and marks the onset of the dynamic Neoproterozoic carbon cycle.

Shifting focus to the late Neoproterozoic, the third chapter delves into the origins of Neoproterozoic superheavy pyrite, whose bulk-sample $\delta^{34}S$ values are greater than those of contemporaneous seawater sulfate and whose origins remain controversial. Two supervised machine learning algorithms were trained on a large LA-ICP-MS pyrite trace element database to distinguish pyrite of different origins. The analysis validates that two models built on the co-behavior of 12 trace elements (Co, Ni, Cu, Zn, As, Mo, Ag, Sb, Te, Au, Tl, and Pb) can be used to accurately predict pyrite origins. This novel approach was then used to identify the origins of pyrite from two Neoproterozoic sedimentary successions in South China. The first set of samples contains isotopically superheavy pyrite from the Cryogenian Tiesi'ao and Datangpo formations. The second set of samples contains pyritic rims from the Ediacaran Doushantuo Formation; these pyrite rims are associated with fossiliferous chert nodules and do not have superheavy sulfur isotopes. For the superheavy pyrite, the models consistently show high confidence levels in

identifying its genesis type, and three out of four samples were inferred to be of sedimentary origins. For the pyritic nodule rims, the models suggest that early diagenetic pyrite was subsequently altered by hydrothermal fluids and therefore shows mixed signals. The third chapter highlights the importance of pyrite trace elements in deciphering and distinguishing the origins of pyrite in sedimentary strata.

Geochemical investigation of the co-evolution of life and environment in the Neoproterozoic Era

Junyao Kang

GENERAL AUDIENCE ABSTRACT

Understanding how life and the environment have shaped our planet's story over 4.5 billion years is like piecing together an intricate puzzle. On the one hand, changes in the environment kickstarted big shifts in how life evolved. On the other hand, living creatures have also left their mark on Earth's landscapes and climate. This dissertation focuses on unraveling the mysterious Neoproterozoic Era (1 billion to 538 million years ago), a time when Earth saw some of its most dramatic changes.

A significant aspect of my investigation delves into the evolutionary dynamics within ancient marine ecosystems. Specifically, I'm exploring a critical juncture when organisms with more complex cellular structures, known as eukaryotes, became ecologically more important than prokaryotic life forms in many aspects of Earth systems. By examining ancient rock formations from China, I have found evidence suggesting that nitrate, a vital nutrient, was scarce in the Neoproterozoic oceans. However, around 800 million years ago, there appears to have been a significant surge in nitrate availability. This surge potentially catalyzed a pivotal phase in evolution, possibly driving the shift from prokaryote to eukaryote dominance in these ancient waters.

Second, there is a challenge to delineate a robust timeline for the early Neoproterozoic Era. Imagine trying to piece together a story from a time when there were no calendars or clear dates. Employing advanced statistical methods and comparing chemical signals preserved in carbonate rocks from disparate global locations, I endeavor

to craft a coherent timeline for this crucial period. Within this timeline, a noteworthy anomaly in the carbon cycle emerged around 920 million years ago known as the Majiatun excursion. This anomaly represents a significant shift in the Neoproterozoic carbon cycle.

Furthermore, my investigation plunges into the geochemistry of sulfur, an important element in shaping ancient marine environments. Certain sedimentary rocks harbor anomalous sulfur isotope signatures in the mineral pyrite (also known as fool's gold), hinting at dramatic environmental transformations during the late Neoproterozoic. Employing advanced analytical techniques and machine learning methodologies, I seek to discern the origins and implications of these anomalous sulfur isotope signals found in pyrite, unraveling their significance in reconstructing the environmental dynamics of ancient oceans.

Dedicated to my partner Changxin and my parents, Guojing and Yuansong, without whose love and support I would not be here today.

ACKNOWLEDGEMENTS

I wish to extend my deepest gratitude to my advisor, Shuhai, who not only served as a mentor but also stood as a remarkable role model throughout this thesis journey. Your guidance, unwavering support, and invaluable insights were pivotal in shaping this work. Your dedication to academic excellence has been a constant source of inspiration.

A special thank you to my undergraduate advisor at Peking University, Lin, whose guidance initially introduced me to the captivating world of Neoproterozoic geology. Your early mentorship laid the foundation for my academic pursuits and greatly influenced the direction of this thesis.

My heartfelt appreciation goes to the members of my thesis committee, Ben, Rachel, and Maddy, for their constructive critiques and invaluable suggestions that greatly enhanced the depth and quality of this work.

I am humbled by the journey that led me here. As a child, I never dreamed of becoming a scientist, especially a geoscientist, yet through the encouragement and support of mentors like Shuhai, Lin, Ben, Rachel, and Maddy, I discovered a passion for scientific inquiry that has shaped this thesis.

To my family, Changxin, my mom, and my dad, whose enduring patience, love, and understanding sustained me through this demanding journey, I owe an immeasurable debt of gratitude.

I extend sincere appreciation to my friends from high school in Chengdu, my undergraduate years at Peking University, and those here at Virginia Tech, for their steadfast belief in this endeavor and their unwavering support during challenging phases.

Lastly, my heartfelt thanks to all the participants and individuals who generously shared their time and insights for this study. Their contributions have been invaluable to the fruition of this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ATTRIBUTION

CHAPTER 1

Junyao Kang: Conceptualization, Methodology, Formal Analysis, Investigation, Data Curation, Writing – Original Draft. **Benjamin Gill:** Conceptualization, Investigation, Writing – Review & Editing. **Rachel Reid:** Investigation, Writing – Review & Editing. **Feifei Zhang:** Investigation, Writing – Review & Editing. **Shuhai Xiao:** Conceptualization, Supervision, Project Administration, Funding Acquisition, Writing – Original Draft.

CHAPTER 2

Junyao Kang: Conceptualization, Methodology, Formal Analysis, Investigation, Data Curation, Writing – Original Draft. **Qing Tang:** Investigation, Writing – Review & Editing. **Benjamin Gill:** Conceptualization, Methodology, Investigation, Writing – Review & Editing. **Rachel Reid:** Investigation, Writing – Review & Editing. **Wentao Zheng:** Investigation, Writing – Review & Editing. **Shuhai Xiao:** Conceptualization, Supervision, Project Administration, Funding Acquisition, Writing – Original Draft.

CHAPTER 3

Junyao Kang: Conceptualization, Methodology, Formal Analysis, Investigation, Data Curation, Writing – Original Draft. **Daniel D. Gregory:** Conceptualization, Methodology, Data Curation, Investigation, Writing – Review & Editing. **Benjamin Gill:** Conceptualization, Investigation, Writing – Review & Editing. **Shiqiang Huang:**

Investigation, Writing – Review & Editing. **Changxin Lai:** Methodology, Formal Analysis, Writing – Review & Editing. **Zhaoshan Chang:** Investigation, Writing – Review & Editing. **Huan Cui:** Investigation, Writing – Review & Editing. **Ivan Belousov:** Data Curation, Writing – Review & Editing. **Shuhai Xiao:** Conceptualization, Supervision, Project Administration, Funding Acquisition, Writing – Original Draft.

INTRODUCTION

Understanding the co-evolution of life and environments is an important component of studying the 4.5-billion-year history of the Earth. Environmental changes have played a critical role in shaping the course of biological evolution, and reciprocally, life has significantly influenced and transformed Earth's surface and climate. This interaction between life and the environment is increasingly becoming an important research topic due to the ongoing rapid and profound environmental changes threatening diverse organisms on the Earth. Specifically, delving into the underlying mechanism and long-term impact of these changes is essential in our quest to predict how situations will evolve in the near future and to develop sustainable strategies for mitigating environmental threats.

In this dissertation, I focused on a critical interval in Earth history — the Neoproterozoic Era, when dramatic environmental changes may have driven important eukaryotic evolutionary events, including the rise of eukaryotes to ecological dominance (*1*) and the origin of animals (*2*). Previous research suggested that changes in atmospheric oxygen levels (*3, 4*), oceanic redox structures (*5-7*), and nutrient cycles (*8, 9*) may have played significant roles in driving the Neoproterozoic ecosystem revolution, but the details remain unclear. By further investigating and reconstructing Neoproterozoic paleoenvironments, we can better understand the co-evolution of life and environment during this pivotal era. In addition, it may help us understand the boundary conditions and the resilience of ecosystems under current climate changes.

In my first chapter, I focus on the potential driver of an important event in the evolutionary history of the biosphere — the transition from a prokaryote-dominant to a eukaryote-dominant marine ecosystem. Existing fossil and biomarker evidence indicates that this transition likely occurred sometime in the Tonian Period (i.e., early Neoproterozoic, 1000 to 720 million years ago) (*1, 10-12*). As phosphorus has long been regarded as the ultimate limiting nutrient on productivity in marine ecosystems on geological time scales (*13*), its increasing supply during the Tonian Period has been proposed to facilitate the rise of eukaryotic primary producers (*1, 8, 9*). In principle, the bioavailability of nitrogen, another macronutrient in marine ecosystems, could have played a similar role, depending on the relative abundance of bioavailable phosphate vs nitrate. To understand the Neoproterozoic nitrogen cycle and its potential role in this evolutionary event, I investigated the ca. 1000–800 Ma Huaibei Group from the North China Craton, using a multi-geochemical proxy approach (Fe speciation and N isotopes). The new data from the Huaibei Group, together with existing Mesoproterozoic-Neoproterozoic nitrogen isotope records, were used to reconstruct the evolution of the Nitrogen cycle during this interval.

Given the spatial variations in oceanic environments during the Neoproterozoic, it is essential to gather geochemical data from diverse locations across the globe to accurately evaluate global environmental shifts. Consequently, before embarking on a compilation of global data, a robust stratigraphic framework needs to be established. In the second chapter, my primary focus is on the preliminary construction of a $\delta^{13}C$ chemostratigraphic framework for the early Tonian Period. Due to the lack of radiometric dates and index fossils within Tonian strata, carbon isotope chemostratigraphic data from paleogeographically distant cratons, including the North China, São Francisco, and Congo (*14-17*), were used to build this framework. In order to achieve stratigraphic correlations with statistical significance and reproducibility, a novel application of the dynamic time warping algorithm was used to correlate the chemostratigraphic data (*18*).

In the third chapter, I shift my focus to the sulfur cycle in the late Neoproterozoic Era, recognizing its critical role in regulating oceanic redox conditions. Sulfate reduction, a key process within this cycle, consumes organic matter, leading to the localized accumulation of H_2S and the buildup of toxic euxinic or sulfidic conditions (*19*). The sulfur isotope composition of sedimentary pyrite and seawater sulfate is a powerful tool for studying the ancient sulfur cycle. Typically, biogenic pyrite has lower $\delta^{34}S$ values than seawater sulfate because microbial sulfate reduction prefers the lighter sulfur isotope (*20*). However, anomalously high $\delta^{34}S_{pyrite}$ values (> $\delta^{34}S$ of contemporaneous seawater sulfate) have been reported from a number of Cryogenian and Ediacaran sedimentary basins. While several hypotheses have been proposed to link its origin with the Neoproterozoic biogeochemical sulfur cycle (e.g., *21, 22, 23*), a recent hypothesis proposed by Cui et al. (*24*) has attributed the isotopically superheavy pyrite to thermogenic sulfate reduction related to hydrothermal alteration. Hence, prior to drawing inferences for Neoproterozoic paleoenvironments, it is imperative to assess the origins of superheavy pyrite. To achieve this goal, my research explored the application of machine learning algorithms in distinguishing pyrite of different origins based on in-situ trace element measurements obtained through LA-ICP-MS.

REFERENCES

1. J. J. Brocks *et al.*, The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* **548**, 578-581 (2017).
2. D. B. Mills, W. R. Francis, D. E. Canfield, Animal origins and the Tonian Earth system. *Emerging Topics in Life Sciences* **2**, 289-298 (2018).
3. N. J. Planavsky *et al.*, Low mid-Proterozoic atmospheric oxygen levels and the delayed rise of animals. *Science* **346**, 635-638 (2014).
4. D. B. Cole *et al.*, A shale-hosted Cr isotope record of low atmospheric oxygen during the Proterozoic. *Geology* **44**, 555-558 (2016).
5. R. Guilbaud, S. W. Poulton, N. J. Butterfield, M. Zhu, G. A. Shields-Zhou, A global transition to ferruginous conditions in the early Neoproterozoic oceans. *Nature Geoscience* **8**, 466-470 (2015).
6. G.-Y. Wei *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-Science Reviews* **214**, 103506 (2021).
7. F. Zhang *et al.*, Uranium isotope evidence for extensive shallow water anoxia in the early Tonian oceans. *Earth and Planetary Science Letters* **583**, 117437 (2022).
8. C. T. Reinhard *et al.*, Evolution of the global phosphorus cycle. *Nature* **541**, 386–389 (2017).
9. N. J. Planavsky *et al.*, A sedimentary record of the evolution of the global marine phosphorus cycle. *Geobiology* **00**, 1-7 (2022).
10. P. A. Cohen, R. B. Kodner, The earliest history of eukaryotic life: uncovering an evolutionary story through the integration of biological and geological data. *Trends in Ecology & Evolution* **37**, 246-256 (2022).
11. S. M. Porter, Insights into eukaryogenesis from the fossil record. *Interface Focus* **10**, 20190105 (2020).
12. D. B. Mills *et al.*, Eukaryogenesis and oxygen in Earth history. *Nature Ecology & Evolution* **6**, 520-532 (2022).
13. T. Tyrrell, The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* **400**, 525-531 (1999).
14. H. Zhao *et al.*, New geochronologic and paleomagnetic results from early Neoproterozoic mafic sills and late Mesoproterozoic to early Neoproterozoic successions in the eastern North China Craton, and implications for the reconstruction of Rodinia. *GSA Bulletin* **132**, 739-766 (2020).
15. D. A. D. Evans *et al.*, Return to Rodinia? Moderate to high palaeolatitude of the São Francisco/Congo craton at 920 Ma. *Geological Society, London, Special Publications* **424**, 167-190 (2016).
16. R. I. Trindade, M. S. D'Agrella-Filho, P. Y. Antonio, W. Teixeira, in *Ancient supercontinents and the paleogeography of Earth*. (Elsevier, 2021), pp. 445-464.
17. D. A. Evans, in *Ancient Supercontinents and the Paleogeography of Earth*. (Elsevier, 2021), pp. 549-576.
18. C. C. Hay, J. R. Creveling, C. J. Hagen, A. C. Maloof, P. Huybers, A library of early Cambrian chemostratigraphic correlations from a reproducible algorithm. *Geology* **47**, 457-460 (2019).

19. R. A. Berner, Burial of organic carbon and pyrite sulfur in the modern ocean: its geochemical and environmental significance. *Am. J. Sci.;(United States)* **282**, (1982).
20. D. Canfield, Biogeochemistry of sulfur isotopes. *Reviews in mineralogy and geochemistry* **43**, 607-636 (2001).
21. B. Shen *et al.*, Stratification and mixing of a post-glacial Neoproterozoic ocean: Evidence from carbon and sulfur isotopes in a cap dolostone from northwest China. *Earth and Planetary Science Letters* **265**, 209-228 (2008).
22. J. B. Ries, D. A. Fike, L. M. Pratt, T. W. Lyons, J. P. Grotzinger, Superheavy pyrite ($\delta^{34}S_{pyr} > \delta^{34}S_{CAS}$) in the terminal Proterozoic Nama Group, southern Namibia: A consequence of low seawater sulfate at the dawn of animal life. *Geology* **37**, 743–746 (2009).
23. C. Cai *et al.*, Enigmatic super-heavy pyrite formation: Novel mechanistic insights from the aftermath of the Sturtian Snowball Earth. *Geochimica et Cosmochimica Acta* **334**, 65-82 (2022).
24. H. Cui *et al.*, Questioning the biogenicity of Neoproterozoic superheavy pyrite by SIMS. *American Mineralogist: Journal of Earth and Planetary Materials* **103**, 1362-1400 (2018).

CHAPTER 1 Nitrate limitation in early Neoproterozoic oceans delayed the ecological rise of eukaryotes

Published in Science Advances: J. Kang, B. Gill, R. Reid, F. Zhang, S. Xiao, Nitrate limitation in early Neoproterozoic oceans delayed the ecological rise of eukaryotes. Science Advances 9, eade9647 (2023).

Authors

Junyao Kang,[1,2] Benjamin Gill,[1,2] Rachel Reid,[1,2] Feifei Zhang,[3] Shuhai Xiao[1,2]

Affiliations

1. Department of Geosciences, Virginia Tech, Blacksburg, VA, USA
2. Global Change Center, Virginia Tech, Blacksburg, VA, USA
3. State Key Laboratory for Mineral Deposits Research, School of Earth Sciences and Engineering, and Frontiers Science Center for Critical Earth Material Cycling, Nanjing University, Nanjing 210023, China

ABSTRACT

The early Neoproterozoic Era witnessed the initial ecological rise of eukaryotes at ca. 800 Ma. To assess whether nitrate availability played an important role in this evolutionary event, we measured nitrogen isotope compositions ($\delta^{15}N$) of marine carbonates from the early Tonian (ca. 1000–ca. 800 Ma) Huaibei Group in North China. The data reported here fill a critical gap in the $\delta^{15}N$ record and indicate nitrate limitation in early Neoproterozoic oceans. A compilation of Proterozoic sedimentary $\delta^{15}N$ data reveals a stepwise increase in $\delta^{15}N$ values at ~800 Ma. Box model simulations indicate that this stepwise increase likely represents a ~50% increase in marine

nitrate availability. Limited nitrate availability in early Neoproterozoic oceans may have delayed the ecological rise of eukaryotes until ~800 Ma when increased nitrate supply, together with other environmental and ecological factors, may have contributed to the transition from prokaryote-dominant to eukaryote-dominant marine ecosystems.

INTRODUCTION

Although eukaryotes as a clade likely diverged from their living sister clade (i.e., share a most recent common ancestor) in the Paleoproterozoic, they did not become phylogenetically and ecologically diverse until sometime in the Tonian Period (1000–720 Ma) (*1-3*). The delayed transition from a prokaryote-dominant to a eukaryote-dominant marine ecosystem marks an important milestone in the evolutionary history of the biosphere, and is documented by multiple lines of evidence. First, phylogenetically resolved crown-group eukaryotes, including the red alga *Bangiomorpha pubescens* and the green alga *Proterocladus antiquus*, emerged in the Tonian Period ca. 1.0 Ga (*1*). Various micro-/macrofossils also indicate that important eukaryotic evolutionary innovations occurred during this time, including multicellularity, cell differentiation, biomineralization, and eukaryovory (i.e., a feeding strategy in which one eukaryotic organism consumes another eukaryotic organism) (*4*). Further, there is an increase in global species richness in the late Tonian Period (ca. 800 Ma) (*5*), although an increase of within-assemblage fossil diversity is not apparent (*4*). The morphological disparity of eukaryotic fossils also expanded in the Tonian Period relative to the preceding Mesoproterozoic Era (*6*). More importantly, lipid biomarkers (see discussion in Supplementary Material) suggest an ecological rise of eukaryotic primary producers during the Tonian Period [(*7*), but see (*8*)].

What factors may have driven the transition from a prokaryote-dominant to a eukaryote-dominant marine ecosystem? Given the coupling of predation and animal diversity (*9, 10*), it is possible that the rise of eukaryovory may have driven the phylogenetic diversification of eukaryotes in the Tonian (*5, 11*). However, this hypothesis does not account for the ecological dominance of eukaryotes, which is likely also linked to nutrient supplies and the rise of eukaryotic primary producers. Phosphorus has long been regarded as the ultimate limiting nutrient on geological timescales (*12*). It has been proposed that an enlarged marine phosphorus reservoir after the Sturtian glaciation in the Cryogenian (*13*) was responsible for the ecological rise of eukaryotic algae as the dominant primary producers (*7*). Widespread ferruginous conditions in pre-Cryogenian oceans may have caused marine phosphorus limitation due to enhanced phosphorus scavenging into the sediment by iron minerals (*13*) and weakened phosphorus remineralization in anoxic environments (*14*). Under such conditions, cyanobacteria outcompete phototrophic eukaryotes because of their smaller cell size and greater efficiency of nutrient uptake (*7, 15*). During the Cryogenian, enhanced post-glacial weathering likely introduced a substantial amount of phosphorus into the ocean (*13, 16*). With increased phosphate availability, eukaryotic phototrophs became more competitive than cyanobacteria, and the ecological rise of eukaryotic algae with a larger cell size led to a more efficient biological pump and fundamental changes in the marine carbon cycle and redox structure (*13, 16*).

In principle, the bioavailability of nitrogen, another macronutrient in marine ecosystems, could have also played a similar role in driving the ecological rise of eukaryotic primary producers. It has been suggested that fixed nitrogen, instead of phosphorus, limits primary productivity on glacial-to-interglacial timescales (*17*). Earth system models show that bioavailable nitrogen may have been a limiting nutrient during parts of the Proterozoic when pO_2 was at intermediate levels of 0.1–0.4× present atmospheric level (*13*). The marine biogeochemical cycle of nitrogen, which has multiple redox states, is closely coupled with oceanic redox structures. Nitrogen fixation reduces atmospheric N_2 to biologically accessible ammonium. In the presence of O_2, even at nM levels (*18*), ammonium is oxidized to nitrate through nitrification. In suboxic-to-anoxic seawater and sedimentary porewaters, however, nitrate is reduced to gaseous species and removed from the ocean mainly through denitrification and anammox (*19, 20*). The availability of nitrate has long been regarded as a key factor dictating the competition between eukaryotic and bacterial primary producers (*21*). Eukaryotes do not fix nitrogen and thus their growth is dependent on the assimilation of bioavailable nitrogen (*22*). In contrast, diazotrophic cyanobacteria fix nitrogen whenever needed, although they can also release enzymes to efficiently scavenge bioavailable nitrogen from the seawater (*21*). Thus, cyanobacteria can outcompete photosynthetic eukaryotes and maintain ecological dominance when nitrate is limited, but eukaryotes have an upper hand when nitrate is readily available, due to their ability to rapidly transport nitrate into the cell for further assimilation (*23*) and to store nitrate in intracellular vacuoles (*21, 24*). This nitrate-controlled dichotomy is readily seen in modern oceans, where diatoms dominate in nitrate-rich areas and cyanobacteria in nitrate-limited areas (Fig. 1. 1). With low atmospheric oxygen levels and widespread anoxic conditions in the pre-800 Ma oceans (*25-27*), it is expected that nitrogen loss may be extensive due to denitrification/anammox and nitrate was not widely available (*28*). Therefore, it is possible that nitrate availability may be as important as phosphorus supply in limiting or fueling the ecological rise of photosynthetic eukaryotes. However, the current understanding of the early Neoproterozoic nitrogen cycle is extremely limited due to a lack of data, particularly for the early Tonian Period, prompting us to generate data from the Huaibei Group and to integrate a modeling approach in order to elucidate the potential role of nitrate in this critical evolutionary event.

To fill the early Tonian gap of $\delta^{15}N$ data, we investigated the ca. 1,000–800 Ma Huaibei Group from the North China Craton, using a multi-geochemical proxy approach (Fe speciation and N isotopes). The Huaibei Group offers a window onto the shallow marine redox structure and nitrogen cycle before the proliferation of eukaryotic algae. Noting that $\delta^{15}N$ can be influenced by local environmental conditions and can show appreciable spatial variations (*29, 30*), we compiled a Proterozoic $\delta^{15}N$ database, which was then analyzed using time-series techniques in order to identify secular trends from local signals. Finally, we used a simple box model to provide a broader view of nitrate availability and its potential relationship with the ecological rise of eukaryotes. This integrative approach helps us to quantitatively understand the early Neoproterozoic nitrogen cycle and to evaluate the possibility of a nitrate driver for the ecological rise of eukaryotes.

Geological background

The North China Craton (NCC) is a major tectonic unit in China and one of the best studied cratons in the world. Tonian sedimentary rocks, including the Huaibei Group, occur in the Xuhuai rift system along the southeastern margin of the NCC. The Huaibei Group consists of 13 formations and can be divided into three parts (fig. 1. S1), with eukaryotic fossils reported from the middle and upper parts of the group. The lower part of the succession consists of the Lanling, Xinxing, and Jushan formations that are predominantly siliciclastic rocks, including quartz sandstone and shale with carbonate concretions. The middle-upper Huaibei Group is dominated by carbonate rocks and consists of the Jiayuan, Zhaowei, Niyuan, Jiudingshan, Zhangqu, Weiji, and Shijia formations. Stromatolites and molar-tooth structures are common in this part of the succession, and putative vase-like microfossils, leiospheric acritarchs, *Chuaria*, and *Tawuia* have also been reported from this portion of the Huaibei Group (*31*). The upper part of the succession consists of the Wangshan Formation (stromatolitic dolostone and limestone), the Jinshanzhai Formation (quartz sandstone, shale, limestone, and stromatolitic dolostone), and the uppermost Gouhou Formation (shale and siltstone) (*32*). The carbonaceous compression fossils *Chuaria*, *Tawuia*, and *Protoarenicola* have been reported from the Jinshanzhai Formation, and *Chuaria*, *Tawuia*, *Dictyosphaera* sp., *Trachyhystrichosphaera aimika*, and *Valeria lophostriata* have been recovered from the Gouhou Formation (*31, 33*). The exceptional fossil preservation suggests low degrees of thermal maturity (*33*), which is consistent with relatively low vitrinite reflectance (R_0: 2.2–2.5) of the Huaibei Group (*34*) and Raman spectroscopic data of organic-walled microfossils from age equivalent strata in the same rift system (*35*).

The depositional age of the Huaibei Group is relatively well-constrained by available radiometric, biostratigraphic, and chemostratigraphic data (fig. 1. S1). Authigenic monazite from the Xinxing Formation constrains the initial deposition of the Huaibei Group to be 1086 ± 17 Ma (*36*). Dolerite sills intruding the Wangshan Formation are dated at 906 ± 10 Ma (*37*), providing minimum age constraints on the intruded strata. The authigenic xenotime from the Shijia Formation suggest a minimum depositional age of 919 ± 23 Ma (*36*). These ages constrain most of the Huaibei Group between ~1086 and ~906 Ma. Based on an age model using strontium isotope chemostratigraphy, the deposition of most of the carbonate strata occurred between ca. 980 Ma and ca. 920 Ma (*38*). However, the youngest detrital zircons from the Jinshanzhai Formation yield a maximum depositional age of 820 ± 11 Ma (*39*), and the existence of *Protoarenicola* in the same formation suggests a Tonian age (>720 Ma). Hence, a ~100 Myr unconformity may be present between the Wangshan and the Jinshanzhai formations and therefore in this study, we assigned a

<820 Ma age to the Jinshanzhai Formation. The presence of *Trachyhystrichosphaera aimika*, *Valeria lophostriata*, and *Dictyosphaera* sp. in the Gouhou Formation, *Protoarenicola* in the Jinshanzhai Formation, *Pararenicola* in the Shijia Formation, and *Tawuia* at multiple horizons in the Huaibei Group supports a Tonian age for this group. The $\delta^{13}C_{carb}$ chemostratigraphy is also consistent with the Tonian global $\delta^{13}C$ record (*31*).

In this study, we focus on the middle and upper parts of the Huaibei Group, which were mostly deposited on a shallow-water carbonate platform (*40*). Development of storm deposits, ripple marks, and cross stratification suggest extensive mixing of the water column (*41*). Additionally, paleontological data from the Jinshanzhai and Gouhou formations also indicate an open marine environment (*33*). A total of 141 carbonate samples were collected from four sections in the Huaibei region of North China, including the Zhaowei section (Jiayuan, Zhaowei, Niyuan, and Jiudingshan formations), Liangtang section (Jiudingshan Formation), Lushan section (Zhangqu, Weiji, and Shijia formations), and Langan section (upper Wangshan, Jinshanzhai, Gouhou formations). These samples were analyzed for iron speciation and nitrogen isotope composition.

Iron speciation

Iron speciation is a widely used approach to evaluate local redox conditions in ancient sedimentary basins. This proxy focuses on iron species that are highly reactive (Fe_{HR}) towards dissolved sulfide, including iron in carbonate (Fe_{carb}), iron oxides (Fe_{ox}), magnetite (Fe_{mag}), and iron that has reacted with dissolved sulfide to form pyrite (Fe_{py}) (*42*). In sediments deposited under an anoxic water column, enrichment of highly reactive iron occurs via shelf-to-basin iron transport, resulting in elevated ratios of Fe_{HR} to the total iron (Fe_T) in the sediment (*43*). It has been shown, through analyses of modern and ancient sediments, that $Fe_{HR}/Fe_T < 0.22$ is indicative of sediments deposited under oxic water columns, and $Fe_{HR}/Fe_T > 0.38$ is indicative of sediments deposited under anoxic water columns. Furthermore, ferruginous (anoxic, iron-replete) and euxinic (anoxic, sulfide-replete) water columns can be distinguished using Fe_{py}/Fe_{HR} ratios, because proportionally more Fe_{HR} would react with hydrogen sulfide and precipitate as pyrite in the euxinic water columns than in ferruginous conditions. A threshold Fe_{py}/Fe_{HR} of 0.7–0.8 is widely used to separate ferruginous (< 0.7) from euxinic conditions (> 0.8) (*42*). More recently, with Holocene data from the eastern Mediterranean Sea (*44*), a lower threshold (0.6–0.8) has been suggested (*45*). While the iron speciation proxy was initially applied in the analysis of fine-grained siliciclastic rocks, recent work demonstrates that Fe_{HR}/Fe_T is also a reliable redox proxy for carbonate rocks with $Fe_T > 0.5$ wt% (*46*). Here we compiled a Fe speciation dataset of carbonate samples ($Fe_T > 0.5$ wt%) from modern and ancient anoxic basins to further evaluate the effectiveness of Fe_{py}/Fe_{HR} in carbonate rocks (see Supplementary Material).

The nitrogen isotope proxy

The nitrogen isotope composition ($\delta^{15}N$) of sedimentary rocks is a powerful tool for understanding the ancient nitrogen cycle. Sedimentary nitrogen has a biological origin and ultimately comes from primary producers (*19, 20*). Primary producers assimilate nitrogen from seawater, and hence reflect the $\delta^{15}N$ of the ocean reservoir. In the ocean system, bioavailable nitrogen is ultimately supplied by N_2 fixation and removed by a combination of pathways, including heterotrophic denitrification and anammox in the water column and sediments, as well as the burial of nitrogen in sediments. On geologic timescales, the $\delta^{15}N$ of the ocean reservoir is controlled by the isotopic fractionation associated with these input and output fluxes. For the input flux, nitrogen fixation is associated with a small fractionation ($\varepsilon_{NH_4^+\text{-}N_2}$) of –2‰ to +1‰ when Mo-based nitrogenase enzyme is used (*19*). Alternative pathways using V-/Fe-based nitrogenases can cause fractionations of –6‰ to –8‰; however, based on the nitrogen isotope record and phylogenetic data, these enzymes are thought to be negligible in Precambrian oceans (*19, 47*). Given the $\delta^{15}N$ of atmospheric N_2 ($\delta^{15}N$ = 0‰) (*48*), in a fully anoxic ocean (Fig. 1. 2A) where ammonium is the dominant species and assimilation is quantitative (i.e., nitrate is completely removed) (*19*), a near-zero (–2‰ to +1‰) $\delta^{15}N$ for marine sediments would be expected. On the other hand, when nitrate is abundant (Fig. 1. 2B and C), its partial removal from the water column through denitrification and anammox can impart substantial isotopic fractionations of –10‰ to –30‰ ($\varepsilon_{N_2-NO_3^-}$) (*49*), leading to ^{15}N enrichment in the residual nitrate pool. When denitrification/anammox occurs in the sediments, however, nitrate loss is quantitative and the net isotopic fractionation approaches 0‰ (*49*). The nitrate $\delta^{15}N$ signals are then transferred to sediments by organisms assimilating nitrate as the nitrogen source. As shown in Fig. 1. 2B and C, nitrogen cycles in a Proterozoic ocean with oxygenated surface water and in the modern ocean with an oxygen minimum zone are similar. The major difference is the extent of nitrogen removal and therefore the size of nitrate reservoir. In modern oceans, the average $\delta^{15}N$ value of sediments is about +5‰ due to partial nitrate loss in oxygen minimum zones (*29*), whereas average $\delta^{15}N$ of Proterozoic sediment is expected to be lower (but still above 0‰) due to stronger nitrate loss and proportionally greater ammonium assimilation. In summary, positive $\delta^{15}N$ values of sediments can be regarded as evidence for nitrate availability in the contemporaneous oceans, and near-zero $\delta^{15}N$ values indicate nitrate scarcity and/or extensive nitrogen loss.

Based on previous nitrogen cycle models (*50, 51*), we applied a steady-state box model to understand how nitrate availability in the ocean affects nitrogen isotope signals preserved in sediments (fig. 1. S2). Since our focus is the evolution of the nitrate reservoir on geological timescales (>10^6 years), steady state models are appropriate (*50, 51*). In addition, because the box model does not capture spatial heterogeneity and temporal dynamics, the simulated outputs only represent global averages. In this model, two nitrogen reservoirs were considered (*51*) (fig. 1. S2). The first reservoir (N_{fixer}/ammonium) includes organic nitrogen of diazotrophic organisms (N_{fixer}, mostly cyanobacteria) and ammonium derived from the remineralization of this organic matter. The second reservoir ($N_{assimilator}$/nitrate) includes nitrate and organic nitrogen of nitrate-assimilating organisms ($N_{assimilator}$, mostly eukaryotes). We only considered net input and output fluxes related to these two reservoirs. For example, although dissimilatory nitrate reduction to ammonium (DNRA) was argued to be important in Proterozoic ferruginous oceans (*52*), this nitrogen transformation pathway is not included in our model because it is a recycling flux that is accounted for when the net nitrification is parameterized. For the N_{fixer}/ammonium reservoir, the only input flux is N_2 fixation (F_{fix}). A large portion of this input flux is balanced by net nitrification (F_{remin}) coupled with remineralization, which supplies nitrate to the second reservoir ($N_{assimilator}$/nitrate). Only a minor portion escapes from remineralization and is buried in marine sediments (F_{fixer_burial}). For the $N_{assimilator}$/nitrate reservoir, F_{remin} is the major input, with two minor input fluxes (atmospheric deposition F_{depo} and riverine input F_{river}). In this study, F_{depo} and F_{river} were considered when the model was validated by reproducing modern ocean $\delta^{15}N$ values (see Supplementary Materials). However, because these two fluxes are orders of magnitude smaller than F_{fix} and their impact on model results is negligible (*51*), they were not considered when modeling the Precambrian nitrogen cycle. The major output fluxes from the $N_{assimilator}$/nitrate reservoir are denitrification/anammox in the water column (F_{wcd}) and sediments (F_{sd}), with a minor part of $N_{assimilator}$ buried and preserved as sedimentary N ($F_{assimilator\_burial}$). In our model, F_{sd} was held as a constant equivalent to its modern value, because even surface sediments under oxic water columns can reach the denitrification zone (*53*). Ultimately, the ratio of F_{sd} to F_{wcd} controls $\delta^{15}N$ of the nitrate reservoir ($N_{assimilator}$/nitrate). We note that in extreme cases where marine anoxia and water-column denitrification/anammox are prevalent, nitrate is exhausted in the water column and denitrification/anammox in the sediment is diminished. In such cases, it is inappropriate to hold F_{sd} at its modern value. Although not implemented in this model, in such extreme cases isotopic fractionation associated with both water-column and sedimentary denitrification/anammox approaches to zero, and $\delta^{15}N$ of the nitrate reservoir ($N_{assimilator}$/nitrate) would be near-zero and insensitive to F_{sd}/F_{wcd}.

When in steady state, the fluxes into and out of the ammonium and nitrate reservoirs can be described as follow (Eq. 2 and Eq. 3).

$$F_{fix} = F_{remin} + F_{fixer\_burial} \tag{2}$$

$$F_{remin} = F_{wcd} + F_{sd} + F_{assimilator\_burial} \tag{3}$$

The isotope mass balance can be described in Eq. 4 and Eq. 5.

$$F_{fix} \times (\delta^{15}N_{atmo} + \varepsilon_{fix}) = F_{remin} \times \delta^{15}N_{fixer/ammonium} + F_{fixer\_burial} \times \delta^{15}N_{fixer/ammonium} \tag{4}$$

$$\begin{aligned} F_{remin} \times \delta^{15}N_{fixer/ammonium} &= F_{wcd} \times (\delta^{15}N_{assimilator/nitrate} + \varepsilon_{wcd}) \\ &+ F_{sd} \times (\delta^{15}N_{assimilator/nitrate} + \varepsilon_{sd}) + F_{assimilator\_burial} \times \delta^{15}N_{assimilator/nitrate} \end{aligned} \tag{5}$$

where $\delta^{15}N_{atmo}$, $\delta^{15}N_{fixer/ammonium}$, $\delta^{15}N_{assimilator/nitrate}$ are nitrogen isotope compositions of atmospheric N_2, N_{fixer}/ammonium reservoir, and $N_{assimilator}$/nitrate reservoir; and ε_{fix}, ε_{wcd}, and ε_{sd} are fractionation factors associated with F_{fix}, F_{wcd}, and F_{sd} (see Supplementary Materials for the parameterization of ε_{fix}, ε_{wcd}, and ε_{sd}, which are defined as $\varepsilon = \delta^{15}N_{product} - \delta^{15}N_{reactant}$). Fractionations associated with denitrification and anammox are not differentiated and are represented by a single fractionation factor that integrates both denitrification and anammox, as is commonly done in field site measurements (*49*). This choice is also supported by culture experiments showing similar isotopic effect of denitrification and anammox (*54*). Furthermore, we assumed no isotopic fractionation associated with F_{remin}, $F_{assimilator\_burial}$, and F_{fixer_burial} (*19, 20, 51*). Solving equations (2) to (5), we can obtain the nitrogen isotope compositions of the two oceanic reservoirs (Eq. 6 and Eq. 7).

$$\delta^{15}N_{fixer/ammonium} = \delta^{15}N_{atmo} + \varepsilon_{fix} \tag{6}$$

$$\delta^{15}N_{assimilator/nitrate} = \delta^{15}N_{fixer/ammonium} - (F_{wcd} \times \varepsilon_{wcd} + F_{sd} \times \varepsilon_{sd}) / F_{remin} \tag{7}$$

Since nitrate is the focus of this study, we formulated $f_{assimilator}$ (Eq. 8) as the proportion of nitrogen that is buried from the $N_{assimilator}$/nitrate reservoir. Thus, $f_{assimilator}$ can be viewed as an indicator of nitrate availability in the ocean; when nitrate level is high, we expect a greater $f_{assimilator}$ value or a larger contribution of nitrate burial to the total nitrogen burial.

$$f_{assimilator}=F_{assimilator\_burial}/F_{total\_burial} \tag{8}$$

As the total burial flux (F_{total_burial}) is a mixture of F_{fixer_burial} and $F_{assimilator\_burial}$, the nitrogen isotope composition of sediments ($\delta^{15}N_{sed}$) can be expressed as follows (Eq. 9).

$$\delta^{15}N_{sed} = (1 - f_{assimilator}) \times \delta^{15}N_{fixer/ammonium} + f_{assimilator} \times \delta^{15}N_{assimilator/nitrate} \tag{9}$$

Because both $F_{assimilator\_burial}$ and nitrate loss through denitrification/anammox are controlled by the extent of anoxia in the ocean, $F_{assimilator\_burial}$ is expected to decrease when nitrate loss increases. We thus added an extra constraint on the $F_{assimilator\_burial}$ (Eq. 10) (*51*).

$$F_{assimilator\_burial} \propto \frac{1}{F_{total\_denit}} \tag{10}$$

where $F_{total\_denit} = F_{wcd} + F_{sd}$. Since our focus is on $f_{assimilator}$, or $F_{assimilator\_burial}$ as a proportion of F_{total_burial} (Eq. 8), we used a constant F_{total_burial} (= 25 Tg/yr) in our model calculation. We varied

$f_{assimilator}$ from 0% to 100% to understand how variation in nitrate availability would affect $\delta^{15}N_{sed}$. Finally, the uncertainty of ε_{fix} and ε_{wcd} was evaluated using a Monte Carlo simulation, in which ε_{fix} and ε_{wcd} were sampled from uniform distributions (ε_{fix}, –2 to 1‰; ε_{wcd}, –30 to –22‰). Although higher ε_{wcd} values (–15 to –10‰) have been reported from cellular-level culture experiments (*55*), we used community fractionation factors from field studies to better capture global average conditions.

RESULTS

Twenty-five samples with Fe_T greater than 0.5% and eleven samples with Fe_T between 0.4% and 0.5% were analyzed for iron speciation (Fig. 1. 3A). All of these samples have Fe_{HR}/Fe_T greater than 0.38 (average 0.68, range 0.38–0.94) and Fe_{py}/Fe_{HR} less than 0.6 (average 0.05, range 0–0.40) (Fig. 1. 3B and C).

The compilation of Fe speciation data from carbonate deposits in modern and ancient anoxic basins (Fig. 1. 4), where inference of redox condition was based on independent proxies (e.g., redox-sensitive trace metal concentrations) or Fe speciation data from adjacent beds of fine-grained siliciclastic sediments (Table 1. S1), shows that carbonate samples with $Fe_T > 0.5$ wt% can be used in iron speciation analysis for paleo-redox investigation (see Supplementary Material). The compilation also validates the recently revised threshold of $Fe_{py}/Fe_{HR} = 0.6–0.8$ (*45*) for the differentiation of ferruginous and euxinic conditions based on carbonate samples with $Fe_T > 0.5$ wt%. In light of this, carbonate samples from the Huaibei Group with $Fe_T > 0.5$ wt% provide reliable Fe speciation data, whereas those with Fe_T between 0.4 wt% and 0.5 wt% are considered, but with caution.

A total of 123 carbonate samples from the Huaibei Group were decarbonated and the residues were analyzed for total organic carbon content (TOC), total nitrogen content (TN), organic carbon isotope composition ($\delta^{13}C_{TOC}$), and total nitrogen isotope composition ($\delta^{15}N_{bulk}$) (Fig. 1. 3D and fig. 1. S3). Average TOC is 0.05 wt.% (ranging from 0.01 to 0.21 wt.%, n = 123), average TN is 0.006 wt.% (ranging from 0.001 to 0.032 wt.%, n = 112), average $\delta^{13}C_{TOC}$ is –27.0‰ (ranging from –29.8 to –19.8‰, n = 123), and average $\delta^{15}N_{bulk}$ is 2.0‰ (ranging from –2.0 to 4.2‰, n = 112). About 74% of the samples have $\delta^{15}N_{bulk}$ less than 2.5‰ (n = 83) and ~54% samples have $\delta^{15}N_{bulk}$ between 1.5‰ and 2.5‰ (n = 61). A positive excursion was observed in the Jiudingshan Formation, where $\delta^{15}N_{bulk}$ rises from 1.4‰ to 4.2‰ and then declines to 1.7‰. Short-term fluctuations of $\delta^{15}N_{bulk}$ values were also observed in this section.

Numerical modeling results show that $\delta^{15}N_{sed}$ has a non-linear relationship with the proportional contribution of nitrate-assimilating eukaryotes ($f_{assimilator}$) to sedimentary nitrogen. $\delta^{15}N_{sed}$ values are highest when $f_{assimilator}$ is ~0.5, but decrease to near-zero values when $f_{assimilator}$ approaches zero or unity (Fig. 1. 5E). When $f_{assimilator}$ approaches zero, the ocean is mostly anoxic and nitrate is not available. Sedimentary nitrogen is largely derived from nitrogen-fixing cyanobacteria, which have near-zero $\delta^{15}N$ values. When $f_{assimilator}$ approaches unity, the ocean is mostly oxic and water-column denitrification/anammox approaches zero. As a result, $\delta^{15}N_{assimilator/nitrate}$ and sedimentary $\delta^{15}N$ approach 0‰.

The modeling results also show that modeled $\delta^{15}N_{sed}$ is affected by the fractionation factor associated with nitrogen fixation [ε_{fix}, allowed to vary between –2‰ and +1‰ (*51*)] and the fractionation factor associated with water-column denitrification/anammox [ε_{wcd}, allowed to vary between –30 to –22‰ (*49*)] (Fig. 1. 5A–D). Briefly, as ε_{fix} increases, heavier nitrogen is introduced to the ocean through nitrogen fixation and $\delta^{15}N_{sed}$ reaches higher values. In contrast, as ε_{wcd}

decreases, lighter nitrogen is removed from the ocean through water-column denitrification/anammox and $\delta^{15}N_{sed}$ reaches higher values.

The $\delta^{15}N$ data from the Huaibei Group was incorporated in a compilation of sedimentary nitrogen isotope data (n = 1824) from ~1.6 Ga to ~0.539 Ga, which is updated from a database compiled by Stüeken et al. (*19*). The compiled dataset captures a wide range of marginal marine depositional environments in order to reflect the global nitrogen cycle. Data are grouped into 100 Myr time bins and plotted in Fig. 1. 6A. The dataset is dominated by $\delta^{15}N$ data derived from fine-grained siliciclastic lithologies (1194 samples) as opposed to carbonate rocks (630 samples). This is especially true for the Mesoproterozoic data, with only 85 out of 574 samples from carbonate rocks.

DISCUSSION

Evidence for primary oceanic signals

Previous petrographic, paleontological, and geochemical work on the same suite of samples analyzed here indicates excellent preservation of the sedimentary fabrics, minimal diagenetic alteration, and low degrees of thermal maturity (*31, 56*). Here, we further assess possible post-depositional alteration of the Fe speciation and $\delta^{15}N$ signals.

Fe speciation is subjected to diagenetic and metamorphic alterations since iron and sulfur can be removed from or added to the sediments after deposition. Given that most carbonate samples are micritic with no notable recrystallization or late-stage cements (*31*), post-depositional alteration is likely limited. One specific concern for the Fe speciation proxy is the reaction between highly reactive iron minerals and dissolved sulfide from diagenetic fluids, which could potentially elevate the Fe_{py}/Fe_{HR} of samples with $Fe_{HR}/Fe_{T} > 0.38$ and thus lead to false signals of euxinia (*57*).

Since all analyzed samples have low pyrite contents and no euxinic signal is recorded in the Huaibei samples, this indicates this bias is not a concern for our samples. Another form of post-depositional alteration comes from deep burial dolomitization, which introduces reactive iron from external sources (*42*) and elevates the Fe_{HR}/Fe_T ratio. However, previous studies have shown that dolomite in the Huaibei Group and many other Neoproterozoic successions are likely syndepositional or early diagenetic (*58, 59*), suggesting a lack of highly reactive iron input later during diagenesis. Additionally, Fe_{carb} is not correlated with Mg/Ca (fig. 1. S4), indicating that dolomitization is not a major controlling factor on Fe_{carb} and Fe_{HR}.

$\delta^{15}N$ signals in sedimentary organic matter can also be altered by diagenesis, metamorphism, and detrital contributions. Isotope modifications during biodegradation in the water column and in surface sediments are controlled by oxygen exposure time, which is related to dissolved oxygen content, water depth, and sedimentation rate. When oxygen exposure is limited, isotopic alteration during this process is negligible (< 1‰) (*20*). Considering that the Huaibei carbonates were deposited on a shallow-water carbonate platform (where sedimentation rate tends to be high) and under ferruginous conditions (see discussion below), isotope alteration during early diagenesis is likely minimal. In late diagenesis, nitrogen can be removed from organic matter as NH_4^+ through either microbial remineralization or thermal maturation. This ammonium can be recaptured by clay minerals (as substitutes for monovalent cations) in the sediment or rock under closed system conditions (*60*). Instead of a net loss of sedimentary nitrogen, this process merely redistributes N between organic matter to clay minerals. Therefore, $\delta^{15}N$ of bulk samples ($\delta^{15}N_{bulk}$, including both organic-N and clay-NH_4^+) remains relatively robust against diagenetic alteration. Regarding low-grade metamorphism (from anthracite to semi-graphite stages), which is suggested by low vitrinite reflectance (*34*) and organic-walled microfossil preservation (*33*) in the Huaibei Group, previous studies have demonstrated that metamorphic denitrogenation can cause nitrogen loss but induces little isotopic fractionation (*20*). However, these studies focused on coal deposits with high organic matter contents (*20*). Since progressive metamorphism causes preferential loss of N relative to C, and preferential loss of light isotopes (^{14}N and ^{12}C), correlations between $\delta^{15}N_{bulk}$ and TN, $\delta^{15}N_{bulk}$ and C/N, $\delta^{15}N_{bulk}$ and $\delta^{13}C_{org}$ are expected (*51*). However, none of these correlations are observed in our data (fig. 1. S5A–C). Nitrogen from detrital clay minerals can also potentially mask the primary δ^{15}N signals. The strong correlation between TN and TOC of decarbonated samples (fig. 1. S5D), however, indicates that TN is dominated by organic nitrogen, although the non-zero intercept does suggest that a small fraction of TN is derived from clay minerals. But even this small amount of clay-bound nitrogen may have been, at least partially, recaptured nitrogen from the decomposition of organic matter.

Ferruginous shallow waters and nitrate limitation in the early Tonian ocean

High Fe_{HR}/Fe_T ratios (>0.38) but low Fe_{py}/Fe_{HR} ratios (<0.6) of all Huaibei samples with $Fe_T > 0.5\%$ indicate that the Huaibei carbonates were deposited under a persistently anoxic and ferruginous water column (Fig. 1. 3). The development of ferruginous conditions in the Xuhuai rift basin is consistent with previously published Fe speciation data from Tonian siliciclastic rocks in the adjacent Huainan region, which also indicate consistently ferruginous conditions elsewhere in the basin (*27*). Indeed, Guilbaud et al. (*27*) argued for a transition from euxinic to ferruginous mid-depth waters in the earliest Tonian that was driven by a greater Fe_{HR} flux into the oceans relative to that of sulfate, ultimately related to change in weathering regime during the amalgamation of Rodinia. Considering that the ferruginous conditions are recorded in carbonate rocks of the Huaibei Group deposited on a shallow-water carbonate platform, the chemocline must have been relatively shallow, at least in the Xuhuai rift basin. The expansion of ferruginous seawaters onto carbonate platforms is also in line with recently published uranium isotope data, which indicate extensive shallow water anoxia on a global scale in the early Tonian Period (*56*).

As discussed above, ferruginous shallow waters would be expected to cause extensive nitrogen loss through denitrification and anammox, leading to a marine nitrogen cycle dominated by N-fixation and sedimentary N characterized by near zero $\delta^{15}N$ values. The Huaibei Group has an average $\delta^{15}N_{bulk}$ of 2.0‰ (Fig. 1. 3), indicating extensive but not quantitative removal of nitrate through denitrification and anammox, which would likely result in nitrate limitation. Variations in $\delta^{15}N_{bulk}$ values throughout the Huaibei Group may be related to small-scale changes in the oxygen level of surface waters while bottom waters maintained ferruginous conditions. The positive excursion in the Jiudingshan Formation probably represents a transient expansion of oxic shallow waters. These small-scale and transient variations are consistent with a small nitrate reservoir, which was sensitive to local changes in input or output fluxes.

Based on numerical modeling results (Fig. 1. 5), a mean $\delta^{15}N_{bulk}$ of ~2.0‰ from the Huaibei Group would require a relatively low contribution from nitrate assimilators (mostly eukaryotes) to sedimentary nitrogen. These $\delta^{15}N_{bulk}$ values suggest a $f_{assimilator}$ of only 0.11 (when $\varepsilon_{fix} = -0.5‰$, $\varepsilon_{wcd} = -26‰$) on average; although a $f_{assimilator}$ of 0.87 (when $\varepsilon_{fix} = -0.5‰$, $\varepsilon_{wcd} = -26‰$) can mathematically satisfy the mass balance equations, this is inconsistent with the geochemical data indicative of shallow-water ferruginous anoxia and a relatively small marine nitrate reservoir. Under a scenario with high ε_{fix} (1‰) and low ε_{wcd} (−30‰), $f_{assimilator}$ would decrease further to 0.03. Even at relatively low ε_{fix} (−2‰) and high ε_{wcd} (−22‰), $f_{assimilator}$ is still relatively low (0.24). In comparison, $f_{assimilator}$ in modern oceans is about 0.7 (see Supplementary Materials). Thus, our integrated Fe speciation and nitrogen isotope data, coupled with numerical modeling results, suggest that the Xuhuai basin was probably characterized by ferruginous anoxia, a relatively small marine nitrate reservoir, and a relatively low contribution of nitrate assimilators to sedimentary nitrogen. Considering the global-scale shallow-water anoxia evidenced by uranium isotope data (*56*), the patterns we observed in the Xuhuai basin may be representative of the global ocean. If so, an inescapable implication is that nitrate-assimilating eukaryotes likely played a limited ecological role in early Tonian oceans.

Long-term δ^{15}N trend, nitrate availability, and ecological rise of eukaryotes

The compilation of sedimentary nitrogen isotope data (n = 1824) span the beginning of the Mesoproterozoic at 1.6 Ga, when acritarchs started to appear abundantly in the fossil record (*1*), to the end of the Neoproterozoic at ~0.539 Ga, when multicellular eukaryotes, including possible stem-group animals, became ecologically dominant (*61*).

We argue that the long-term trend in δ^{15}N values over this time interval can inform us about changes in the marine nitrogen cycle and nitrate availability. In particular, the lowest quartile of δ^{15}N data in each time bin is informative. We focus on the lowest quartile for two reasons: (1) it is likely more informative of the extent of nitrate-limited regions, and (2) it helps us to better tackle the potential impact of diagenetic alteration. With regard to the first reason, there was likely greater spatial heterogeneity in nitrate concentrations in Proterozoic oceans relative to modern oceans, given that the average nitrate concentration was lower and the residence time was shorter. The residence time of fixed nitrogen in modern oceans is less than 2000 years (*62*). With stronger nitrogen loss through denitrification/anammox in Proterozoic oceans, the residence time may be shorter than the ocean mixing time, resulting in spatial heterogeneity in both nitrate availability and δ^{15}N values (e.g., figs. 1. S6 and S7). Considering that oxic oases may have been present in

Mesoproterozoic oceans and nitrate supplies may have been locally abundant (*63*), such isotopic heterogeneity was likely accentuated. Since the focus of this study was on the ecological rise of eukaryotes on a global scale, we are more interested in time intervals and regions where nitrate availability could have limited eukaryote success, rather than local environments where eukaryotes may have managed to survive. As such, the lowest quartile of the $\delta^{15}N$ compilation is more informative of the extent of nitrate-limited regions. For the second reason, the lowest quartile helps us to better address the potential issue of diagenetic alteration. Some of the variations in the compiled $\delta^{15}N$ data may be related to diagenesis and metamorphism, which tend to elevate $\delta^{15}N$ values. Therefore, instead of tracking the temporal pattern of the entire dataset, we subjected the lowest quartile in each 100-Myr time bin to a LOWESS regression analysis to understand the long-term evolution of $\delta^{15}N$ (Fig. 1. 6A). Nonetheless, as sensitivity tests and a validation of the trend captured by the lowest quartile data, we also conducted the same LOWESS regression on the lowest 50% data and the entire dataset (fig. 1. S6). We would also like to note that $\delta^{15}N$ data from the terminal Ediacaran Period (551 to 539 Ma) were not included in the analysis because this interval is marked by a brief return to extensive oceanic anoxia (*64*) and thus a likely return to a more broadly nitrate-limited ocean. Indeed, the terminal Ediacaran is characterized by rather low $\delta^{15}N$ values (*19, 20*) and witnessed the extinction of both macro- and microscopic eukaryotes (*65, 66*), perhaps representing a transient return to a prokaryote-dominant world ecologically akin to pre-Cryogenian oceans.

LOWESS analysis of the lowest quartile data revealed a stepwise increase in $\delta^{15}N$ values around 800 Ma, from less than 2‰ to ~3‰. The deviation in the 1.1–1.2 Ga time bin is probably related to sampling bias: data in this time bin come from only one section that was deposited in a well-oxygenated basin (*67*). Due to the scarcity of data in the 800–900 Ma time bin, the bootstrapped LOWESS result indicates that the rise in $\delta^{15}N$ could have started as early as 900 Ma, but a changepoint detection analysis revealed that the most significant change in the mean and variance of the lowest 25% $\delta^{15}N$ data occurred at ca. 820 Ma (Fig. 1. 6A). We note that more data from the 800–900 Ma time bin are needed to accurately define the rise of $\delta^{15}N$. Sensitivity tests of LOWESS regression and changepoint detection analysis on the lowest 50% data and the entire dataset (fig. 1. S6) also revealed a stepwise rise at around 800 Ma, although these analyses show more temporal variation. These analyses are further supported by a two-group analysis of the entire dataset: the mean $\delta^{15}N$ values of pre-800 Ma (3.4 ± 2.2‰) and post-800 Ma (4.6 ± 1.7‰) sub-datasets are significantly different ($P < 0.0001$, Z-test), and the difference remains statistically significant for the bootstrapped sub-datasets ($P < 0.0001$, Z-test) (Fig. 1. 6B, C). The post-800 Ma subgroup has slightly higher total nitrogen (TN) and organic carbon (TOC) contents, but this difference is small and therefore the difference in $\delta^{15}N$ values is unlikely related to differences in either TOC or TN (fig. 1. S8). Finally, the variability of the bootstrapped $\delta^{15}N$ values is higher in the pre-800 Ma sub-dataset (Fig. 1. 6B–C), consistent with a smaller nitrate reservoir with greater spatial heterogeneity and temporal dynamics.

We can use the same box model described above to better assess the evolution of nitrate availability during the Mesoproterozoic and Neoproterozoic eras. For the lowest quartile of the compiled data, a rise in $\delta^{15}N$ from 2‰ to 3‰ at ~800 Ma indicates an increase in $f_{assimilator}$ from 0.11 (95% CI: 0.04−0.21, Monte Carlo simulation to account for uncertainty in ε_{fix} and ε_{wcd}) to 0.16 (95% CI: 0.08−0.31), suggesting a ~50% increase in $f_{assimilator}$ (Fig. 1. 5). For the entire dataset, the mean $\delta^{15}N$ values of the pre- and post-800 Ma subgroups cannot be achieved with low ε_{fix} and high ε_{wcd} values. If the median values are used ($\varepsilon_{fix} = -0.5‰$, $\varepsilon_{wcd} = -26‰$), an increase in average $\delta^{15}N$ from 3.4‰ to 4.6‰ implies a change in $f_{assimilator}$ from 0.18 to 0.27, again representing a

~50% increase. It should be emphasized that, because this box model does not capture spatial heterogeneity and temporal dynamics, the calculated 50% increase in nitrate burial flux represents a global average. If we assume a first-order relationship between the reservoir size and input/output fluxes, this change in nitrate burial flux represents a 50% increase in nitrate availability in the global ocean. Although the nitrate levels of post-800 Ma oceans may have remained low relative to that of modern oceans, this significant increase may have provided an important resource to support the proliferation of eukaryotic primary producers and other eukaryotes further downstream in the food chain.

We note that an increase in δ^{15}N values of sedimentary nitrogen can also be achieved, in principle, by a decrease in ε_{wcd} (i.e., greater fractionation during water column denitrification/anammox). However, considering that nitrate concentration inserts a primary control on the scale of fractionation (*55*), lower ε_{wcd} values imply higher nitrate levels in the ocean. Thus, the main conclusion still holds that higher $\delta^{15}N_{sed}$ values imply greater nitrate availability.

The ultimate driver for a stepwise increase in marine nitrate availability at ~800 Ma is unclear. It is possible that this increase was caused by the rise of atmospheric oxygen levels at ~800 Ma, as indicated by the shale chromium isotope record [(*25, 26*); but see (*68, 69*)]. The breakup of the supercontinent Rodinia (*70*) and the rise of mountains (*71*) may have intensified continental weathering (*38*) and enhanced nutrient supply (e.g., phosphorus) into the ocean, which would incentivize nitrogen fixation to match the availability of other macronutrients, promoting the increase in primary productivity and the subsequent rise of oxygen levels in the atmosphere and shallow oceans. Oxic shallow waters would then weaken water-column denitrification and anammox, increasing nitrate availability and supporting eukaryotic primary producers. The ecological rise of eukaryotes, in turn, would lead to a more efficient biological pump, push the oxygen demand to greater depths, and deepen the chemocline (*72*), which would further increase the oceanic nitrate reservoir size. This positive feedback highlights the complex interactions among biological, atmospheric, oceanic, and tectonic processes that would have influenced the growth of the marine nitrate reservoir and the ecological rise of eukaryotes (Fig. 1. 7).

To sum up, the Fe speciation and nitrogen isotope data reported here from the early Tonian Huaibei Group in the North China Craton fill an important gap in the oceanic redox proxy record of the Proterozoic. Together, the data indicate ferruginous conditions in shallow marine environments, where the contribution of nitrate-assimilating eukaryotes to sedimentary nitrogen was limited. A simple box model calculation indicates that, in the Huaibei basin, nitrate assimilators contributed ~10% to sedimentary nitrogen, compared with ~70% in modern oceans. With data from the Huaibei Group added to the compilation of Mesoproterozoic-Neoproterozoic $\delta^{15}N$ data, our analysis indicates a stepwise increase in sedimentary $\delta^{15}N$ values at ~800 Ma. This increase implies a ~50% increase in marine nitrate availability and eukaryotic contribution to sedimentary nitrogen. It is possible that tectonic, oceanic, atmospheric, and biological processes together shaped the history of marine redox conditions and nitrate supply which, together with the increased availability of other nutrients such as phosphorus, supported the ecological rise of eukaryotes.

MATERIALS AND METHODS

Iron speciation

Total iron (Fe_T) was measured by heating ~0.3 g powered sample at 950 ºC for 6–8 hours, and then dissolving the powder in concentrated hydrochloric acid at 140 ºC for 48 hours. Iron in carbonate (Fe_{carb}), iron oxides (Fe_{ox}), and magnetite (Fe_{mag}) were extracted from ~0.15 g powered sample following the sequential extraction method of Poulton and Canfield (*73*). The iron concentration of these extracted species was measured using the Ferrozine method (*74*) and were analyzed on a Thermo Scientific GENESYS 10S UV-VIS spectrophotometer at Virginia Tech after they were left to react with the Ferrozine reagent for ~12 hours to ensure the complexation reaction reached completion. Pyrite sulfur was extracted from ~5 g powered sample using the chromium reduction method (*75*) and its concentration was determined by gravimetry of the produced Ag_2S precipitate. Iron found in pyrite (Fe_{py}) was calculated using the stoichiometry of pyrite and the amount of pyrite sulfur extracted. The precision (one standard deviation, here and throughout) of the sequential iron extraction based on sample replicates and internal lab standards was ~8%. The precision of the gravimetric pyrite content analysis was ~6% and total iron was ~5%.

Nitrogen and carbon isotopes

Samples were decarbonated using overnight digestion in 4 M hydrochloric acid, and then centrifuged to separate undissolved residue. The undissolved residue was then rinsed three times with 18.2 mΩ-cm deionized water to remove any residual hydrochloric acid. The undissolved residue was dried at 50 ºC and then homogenized. An aliquot of 2–50 mg decarbonated residue was weighed and packaged in a tin capsule for organic carbon isotope analysis; similarly, 50–200 mg decarbonated residue was weighed for nitrogen isotope analysis. Carbon and nitrogen isotopic compositions were determined separately on an Isoprime 100 isotope ratio mass spectrometer (IRMS) coupled with an Elementar vario ISOTOPE cube elemental analyzer at Virginia Tech. For carbon isotope analysis, the temperature of the combustion column and the reduction column was set at 1050 ºC and 650 ºC, respectively. For nitrogen isotope analysis, the combustion column was heated up to 1150 ºC while the reduction column was kept at 650 ºC. Samples were normalized to the VPDB and Air scales using two-point linear normalization. CH-6 ($\delta^{13}C_{VPDB}$ = –10.45 ± 0.03‰), CH-7 ($\delta^{13}C_{VPDB}$ = –32.15 ± 0.05‰), Elemental Microanalysis Low Organic Content Soil ($\delta^{13}C_{VPDB}$ = –22.88 ± 0.40‰), and Elemental Microanalysis Wheat Flour ($\delta^{13}C_{VPDB}$ = –27.21 ± 0.13‰) were used as carbon isotope standards. USGS25 ($\delta^{15}N_{Air}$ = –30.41 ± 0.16‰), USGS26 ($\delta^{15}N_{Air}$ = 53.75 ± 0.15‰), and Elemental Microanalysis Urea ($\delta^{15}N_{Air}$ = –0.30 ± 0.2‰) were used as nitrogen isotope standards. The precision of organic carbon isotope analysis was 0.4‰ (one standard deviation, here and throughout), based on the repeated measurement of 11 Elemental Microanalysis Low Organic Content Soil standards and 5 Elemental Microanalysis Wheat Flour standards. $\delta^{13}C$ measurements of Elemental Microanalysis Low Organic Content Soil (–23.0 ± 0.5‰) and Elemental Microanalysis Wheat Flour (–27.5 ± 0.1‰) agree well with the expected values. The precision of nitrogen isotope analysis was 0.1‰, based on the repeated measurement of 16 Elemental Microanalysis Urea standards. $\delta^{15}N$ Measurements of Elemental Microanalysis Urea standards (–0.5 ± 0.1‰) agree well with the expected value. The reproducibility for $\delta^{15}N$ on duplicated samples (n = 20) was 0.4‰.

Nitrogen isotope data compilation

Published nitrogen isotope data have been screened and only samples below or near greenschist metamorphic grade that likely better reflect primary oceanic signals were included (n = 1824). As bulk nitrogen isotope values may be more accurate (*19*), $\delta^{15}N_{bulk}$ data were prioritized when available (n = 1759); if $\delta^{15}N_{bulk}$ data are not available, then kerogen nitrogen isotope data were used (n = 65).

Statistical analysis

Locally-Weighted Scatterplot Smoothing (LOWESS): LOWESS is a non-parametric, locally weighted regression method for smoothing scatterplots. It is ideal for analyzing time series plots with no theoretical functions. At each data point, a local polynomial with a fixed degree is fitted to the nearest *f* percentage of the dataset using weighted least squares. Data points within this *f* percentage of the dataset are weighted using the tri-cube weight function. More weight is given to points closer to the target point. We used a polynomial degree of 2 to minimize bias. For the Huaibei Group data (Fig. 1. 3D), we used an *f* value of 30%, and the confidence interval was calculated from the standard errors. For the compiled dataset (Fig. 1. 6A), we used an *f* value of 50% to minimize local fluctuations and capture large-scale trends. To further minimize sampling bias, the data were bootstrapped 10,000 times and LOWESS regression was performed each time. The 95% confidence interval of the LOWESS regression was calculated based on the 10,000 LOWESS regression curves. All LOWESS regression analyses were performed in the R Statistical Software using the spatialEco and loess packages (*76*).

Monte Carlo simulation: In this study, we assume a uniform distribution for ε_{fix} (from –2‰ to 1‰) and ε_{wcd} (from –30‰ to –22‰). Following the law of large numbers, values for ε_{fix} and ε_{wcd} were randomly drawn from these two uniform distributions to calculate the $\delta^{15}N_{sed}$ value using the box model described in the text (fig. 1. S2). This process was repeated 10,000 times to estimate the 95% confidence interval of $\delta^{15}N_{sed}$ (Fig. 1. 5E).

Changepoint detection: For time series data, changepoint detection aims to estimate the point at which the statistical properties of this dataset change based on the likelihood test statistics (*77*). In this study, we focused on change in both the mean and variance of the $\delta^{15}N$ dataset. The analysis was performed in R Statistical Software using the changepoint package (*77*). We assumed a normal distribution of the data. The default MBIC penalty was used to find the cut-off point, and the AMOC (at most one change) method was used to detect the most significant change point.

REFERENCES

1. P. A. Cohen, R. B. Kodner, The earliest history of eukaryotic life: uncovering an evolutionary story through the integration of biological and geological data. *Trends in Ecology & Evolution* **37**, 246-256 (2022).
2. S. M. Porter, Insights into eukaryogenesis from the fossil record. *Interface Focus* **10**, 20190105 (2020).
3. D. B. Mills *et al.*, Eukaryogenesis and oxygen in Earth history. *Nature Ecology & Evolution* **6**, 520-532 (2022).
4. S. Xiao, Q. Tang, After the boring billion and before the freezing millions: evolutionary patterns and innovations in the Tonian Period. *Emerging Topics in Life Sciences* **2**, 161-171 (2018).
5. P. A. Cohen, L. A. Riedman, It's a protist-eat-protist world: recalcitrance, predation, and evolution in the Tonian–Cryogenian ocean. *Emerging Topics in Life Sciences* **2**, 173-180 (2018).
6. J. W. Huntley, S. Xiao, M. Kowalewski, 1.3 billion years of acritarch history: An empirical morphospace approach. *Precambrian Research* **144**, 52-68 (2006).
7. J. J. Brocks *et al.*, The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* **548**, 578-581 (2017).
8. L. K. Eckford-Soper, K. H. Andersen, T. F. Hansen, D. E. Canfield, A case for an active eukaryotic marine biosphere during the Proterozoic era. *Proceedings of the National Academy of Sciences* **119**, e2122042119 (2022).
9. J. W. Huntley, M. Kowalewski, Strong coupling of predation intensity and diversity in the Phanerozoic fossil record. *Proceeding of the National Academy of Sciences of the United State of America* **104**, 15006-15010 (2007).
10. J. R. Meyer, R. Kassen, The effects of competition and predation on diversification in a model adaptive radiation. *Nature* **446**, 432-435 (2007).
11. S. Porter, The rise of predators. *Geology* **39**, 607-608 (2011).
12. T. Tyrrell, The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* **400**, 525-531 (1999).
13. C. T. Reinhard *et al.*, Evolution of the global phosphorus cycle. *Nature* **541**, 386–389 (2017).
14. M. A. Kipp, E. E. Stüeken, Biomass recycling and Earth’s early phosphorus cycle. *Science Advances* **3**, eaao4795 (2017).
15. C. T. Reinhard *et al.*, The impact of marine nutrient abundance on early eukaryotic ecosystems. *Geobiology* **18**, 139-151 (2020).
16. N. J. Planavsky *et al.*, The evolution of the marine phosphate reservoir. *Nature* **467**, 1088-1090 (2010).

17. P. G. Falkowski, Evolution of the nitrogen cycle and its influence on the biological sequestration of CO2 in the ocean. *Nature* **387**, 272-275 (1997).
18. J. Füssel *et al.*, Nitrite oxidation in the Namibian oxygen minimum zone. *The ISME journal* **6**, 1200-1209 (2012).
19. E. E. Stüeken, M. A. Kipp, M. C. Koehler, R. Buick, The evolution of Earth's biogeochemical nitrogen cycle. *Earth-Science Reviews* **160**, 220-239 (2016).
20. M. Ader *et al.*, Interpretation of the nitrogen isotopic composition of Precambrian sedimentary rocks: Assumptions and perspectives. *Chemical Geology* **429**, 93-110 (2016).
21. A. D. Anbar, A. H. Knoll, Proterozoic ocean chemistry and evolution: A bioinorganic bridge? *Science* **297**, 1137-1142 (2002).
22. J. A. Sohm, E. A. Webb, D. G. Capone, Emerging patterns of marine nitrogen fixation. *Nature Reviews Microbiology* **9**, 499-508 (2011).
23. S. E. Fawcett, M. W. Lomas, J. R. Casey, B. B. Ward, D. M. Sigman, Assimilation of upwelled nitrate by small eukaryotes in the Sargasso Sea. *Nature Geoscience* **4**, 717-722 (2011).
24. Q. Dortch, J. Clayton, S. Thoresen, S. Ahmed, Species differences in accumulation of nitrogen pools in phytoplankton. *Marine Biology* **81**, 237-250 (1984).
25. N. J. Planavsky *et al.*, Low mid-Proterozoic atmospheric oxygen levels and the delayed rise of animals. *Science* **346**, 635-638 (2014).
26. D. B. Cole *et al.*, A shale-hosted Cr isotope record of low atmospheric oxygen during the Proterozoic. *Geology* **7**, 555–558 (2016).
27. R. Guilbaud, S. W. Poulton, N. J. Butterfield, M. Zhu, G. A. Shields-Zhou, A global transition to ferruginous conditions in the early Neoproterozoic oceans. *Nature Geoscience* **8**, 466-470 (2015).
28. K. Fennel, M. Follows, P. G. Falkowski, The co-evolution of the nitrogen, carbon and oxygen cycles in the Proterozoic ocean. *American Journal of Science* **305**, 526-545 (2005).
29. J. E. Tesdal, E. D. Galbraith, M. Kienast, Nitrogen isotopes in bulk marine sediment: linking seafloor observations with subseafloor records. *Biogeosciences* **10**, 101-118 (2013).
30. E. E. Stüeken, A test of the nitrogen-limitation hypothesis for retarded eukaryote radiation: Nitrogen isotopes across a Mesoproterozoic basinal profile. *Geochimica et Cosmochimica Acta* **120**, 121-139 (2013).
31. S. Xiao *et al.*, Biostratigraphic and chemostratigraphic constraints on the age of early Neoproterozoic carbonate successions in North China. *Precambrian Research* **246**, 208–225 (2014).
32. B. Wan *et al.*, Repositioning the Great Unconformity at the southeastern margin of the North China Craton. *Precambrian Research* **324**, 1-17 (2019).
33. Q. Tang, K. Pang, X. Yuan, B. Wan, S. Xiao, Organic-walled microfossils from the Tonian Gouhou Formation, Huaibei region, North China Craton, and their biostratigraphic implications. *Precambrian Research* **266**, 296-318 (2015).
34. Z. Li, D. Liu, X. Wu, G. Wang, T. Wang, Hydrocarbon-generating potentials of the Neoproterozoic in Xuhuai area, the east part of southern North China. *Chinese Journal of Geology* **47**, 154-168 (2012).

35. K. Pang *et al.*, Raman spectroscopy and structural heterogeneity of carbonaceous material in Proterozoic organic-walled microfossils in the North China Craton. *Precambrian Research* **346**, 105818 (2020).
36. S. Zhang *et al.*, SIMS Pb-Pb dating of phosphates in the Proterozoic strata of SE North China Craton: Constraints on eukaryote evolution. *Precambrian Research* **371**, 106562 (2022).
37. X. Su, P. Peng, S. Foley, W. Teixeira, M.-G. Zhai, Initiation of continental breakup documented in evolution of the magma plumbing system of the ca. 925 Ma Dashigou large igneous province, North China. *Lithos* **384**, 105984 (2021).
38. Y. Zhou *et al.*, Reconstructing Tonian seawater 87Sr/86Sr using calcite microspar. *Geology* **48**, 462-467 (2020).
39. D.-B. Yang *et al.*, U–Pb ages and Hf isotope data from detrital zircons in the Neoproterozoic sandstones of northern Jiangsu and southern Liaoning Provinces, China: implications for the Late Precambrian evolution of the southeastern North China Craton. *Precambrian Research* **216**, 162-176 (2012).
40. F. Sun *et al.*, Provenance analysis of the late Mesoproterozoic to Neoproterozoic Xuhuai Basin in the southeast North China Craton: Implications for paleogeographic reconstruction. *Precambrian Research* **337**, 105554 (2020).
41. L. Wang, Shandong University of Science and Technology, Shandong, China (2009).
42. R. Raiswell *et al.*, The iron paleoredox proxies: A guide to the pitfalls, problems and proper practice. *American Journal of Science* **318**, 491-526 (2018).
43. T. W. Lyons, S. Severmann, A critical look at iron paleoredox proxies: New insights from modern euxinic marine basins. *Geochimica et Cosmochimica Acta* **70**, 5698-5722 (2006).
44. A. Benkovitz *et al.*, Tracing water column euxinia in Eastern Mediterranean Sapropels S5 and S7. *Chemical Geology* **545**, 119627 (2020).
45. S. W. Poulton, in *Elements in Geochemical Tracers in Earth System Science,* T. Lyons, A. Turchyn, C. Reinhard, Eds. (Cambridge University Press, 2021), pp. 1-28.
46. M. O. Clarkson, S. W. Poulton, R. Guilbaud, R. A. Wood, Assessing the utility of Fe/Al and Fe-speciation to record water column redox conditions in carbonate-rich sediments. *Chemical Geology* **382**, 111-122 (2014).
47. A. K. Garcia, H. McShea, B. Kolaczkowski, B. Kaçar, Reconstructing the evolutionary history of nitrogenases: Evidence for ancestral molybdenum ‐ cofactor utilization. *Geobiology* **18**, 394-411 (2020).
48. B. Marty, L. Zimmermann, M. Pujol, R. Burgess, P. Philippot, Nitrogen isotopic composition and density of the Archean atmosphere. *Science* **342**, 101-104 (2013).
49. A. H. Devol, Denitrification, anammox, and N2 production in marine sediments. *Annual review of marine science* **7**, 403-423 (2015).
50. T. Algeo, P. Meyers, R. S. Robinson, H. Rowe, G. Jiang, Icehouse–greenhouse variations in marine denitrification. *Biogeosciences* **11**, 1273-1295 (2014).
51. M. A. Kipp, E. E. Stüeken, M. Yun, A. Bekker, R. Buick, Pervasive aerobic nitrogen cycling in the surface ocean across the Paleoproterozoic Era. *Earth and Planetary Science Letters* **500**, 117-126 (2018).
52. C. C. Michiels *et al.*, Iron-dependent nitrogen cycling in a ferruginous lake and the nutrient status of Proterozoic oceans. *Nature Geoscience* **10**, 217-221 (2017).

53. D. E. Canfield *et al.*, Pathways of organic carbon oxidation in three continental margin sediments. *Marine geology* **113**, 27-40 (1993).
54. B. Brunner *et al.*, Nitrogen isotope effects induced by anammox bacteria. *Proceedings of the National Academy of Sciences* **110**, 18994-18999 (2013).
55. K. Kritee *et al.*, Reduced isotope fractionation by denitrification under conditions relevant to the ocean. *Geochimica et Cosmochimica Acta* **92**, 243-259 (2012).
56. F. Zhang *et al.*, Uranium isotope evidence for extensive shallow water anoxia in the early Tonian oceans. *Earth and Planetary Science Letters* **583**, 117437 (2022).
57. A. M. Hutchings, A. V. Turchyn, A quantification of the effect of diagenesis on the paleoredox record in mid-Proterozoic sedimentary rocks. *Geology* **49**, 1143-1147 (2021).
58. A. v. S. Hood, M. W. Wallace, Synsedimentary diagenesis in a Cryogenian reef complex: Ubiquitous marine dolomite precipitation. *Sedimentary Geology* **255**, 56-71 (2012).
59. J. Kang, C. Li, L. Dong, S. Xiao, Early Diagenetic Dolomite as a Potential Archive of Paleo-redox Fluctuations in an Early Tonian Marine Basin? *Acta Geologica Sinica (English edition)* **96**, 607-620 (2022).
60. P. Müller, CN ratios in Pacific deep-sea sediments: Effect of inorganic ammonium and organic nitrogen compounds sorbed by clays. *Geochimica et Cosmochimica Acta* **41**, 765-776 (1977).
61. S. Xiao, M. Laflamme, On the eve of animal radiation: Phylogeny, ecology and evolution of the Ediacara biota. *Trends in Ecology & Evolution* **24**, 31-40 (2009).
62. J. A. Brandes, A. H. Devol, A global marine‐fixed nitrogen isotopic budget: Implications for Holocene nitrogen cycling. *Global biogeochemical cycles* **16**, 67-61-67-14 (2002).
63. Z. Wang *et al.*, Coupled nitrate and phosphate availability facilitated the expansion of eukaryotic life at circa 1.56 Ga. *Journal of Geophysical Research: Biogeosciences* **125**, e2019JG005487 (2020).
64. F. Zhang *et al.*, Extensive marine anoxia during the terminal Ediacaran Period. *Science Advances* **4**, eaan8983 (2018).
65. A. V. Kolesnikov, V. V. Marusin, K. E. Nagovitsin, A. V. Maslov, D. V. Grazhdankin, Ediacaran biota in the aftermath of the Kotlinian Crisis: Asha Group of the South Urals. *Precambrian Research* **263**, 59-78 (2015).
66. S. D. Evans *et al.*, Environmental drivers of the first major animal extinction across the Ediacaran White Sea-Nama transition. *Proceedings of the National Academy of Sciences* **119**, e2207475119 (2022).
67. E. E. Stüeken, S. Viehmann, S. V. Hohl, Contrasting nutrient availability between marine and brackish waters in the late Mesoproterozoic: Evidence from the Paranoá Group, Brazil. *Geobiology* **20**, 159-174 (2022).
68. G. J. Gilleaudeau *et al.*, Oxygenation of the mid-Proterozoic atmosphere: clues from chromium isotopes in carbonates. *Geochemistry Perspectives Letters* **2**, 178-187 (2016).
69. D. E. Canfield *et al.*, Highly fractionated chromium isotopes in Mesoproterozoic-aged shales and atmospheric oxygen. *Nature Communications* **9**, 1-11 (2018).
70. Z. X. Li *et al.*, Assembly, configuration, and break-up history of Rodinia: A synthesis. *Precambrian Research* **160**, 179–210 (2008).
71. M. Tang, X. Chu, J. Hao, B. Shen, Orogenic quiescence in Earth's middle age. *Science* **371**, 728-731 (2021).

72. T. M. Lenton, R. A. Boyle, S. W. Poulton, G. A. Shields-Zhou, N. J. Butterfield, Co-evolution of eukaryotes and ocean oxygenation in the Neoproterozoic era. *Nature Geoscience* **7**, 257-265 (2014).
73. S. W. Poulton, D. E. Canfield, Development of a sequential extraction procedure for iron: implications for iron partitioning on contitentally derived particulates. *Chemical Geology* **214**, 209-221 (2005).
74. L. L. Stookey, Ferrozine---a new spectrophotometric reagent for iron. *Analytical chemistry* **42**, 779-781 (1970).
75. D. E. Canfield, R. Raiswell, J. T. Westrich, C. M. Reaves, R. A. Berner, The use of chromium reduction in the analysis of reduced inorganic sulfur in sediments and shale. *Chemical Geology* **54**, 149-155 (1986).
76. J. S. Evans, spatialEco. (2021).
77. R. Killick, I. Eckley, changepoint: An R package for changepoint analysis. *Journal of statistical software* **58**, 1-19 (2014).
78. W. Gregg, C. Rousseaux. (Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, USA, 2017).
79. N. J. Planavsky *et al.*, A sedimentary record of the evolution of the global marine phosphorus cycle. *Geobiology* **00**, 1-7 (2022).
80. J. A. Zumberge, D. Rocher, G. D. Love, Free and kerogen ‐ bound biomarkers from late Tonian sedimentary rocks record abundant eukaryotes in mid ‐ Neoproterozoic marine communities. *Geobiology* **18**, 326-347 (2020).

FIGURES

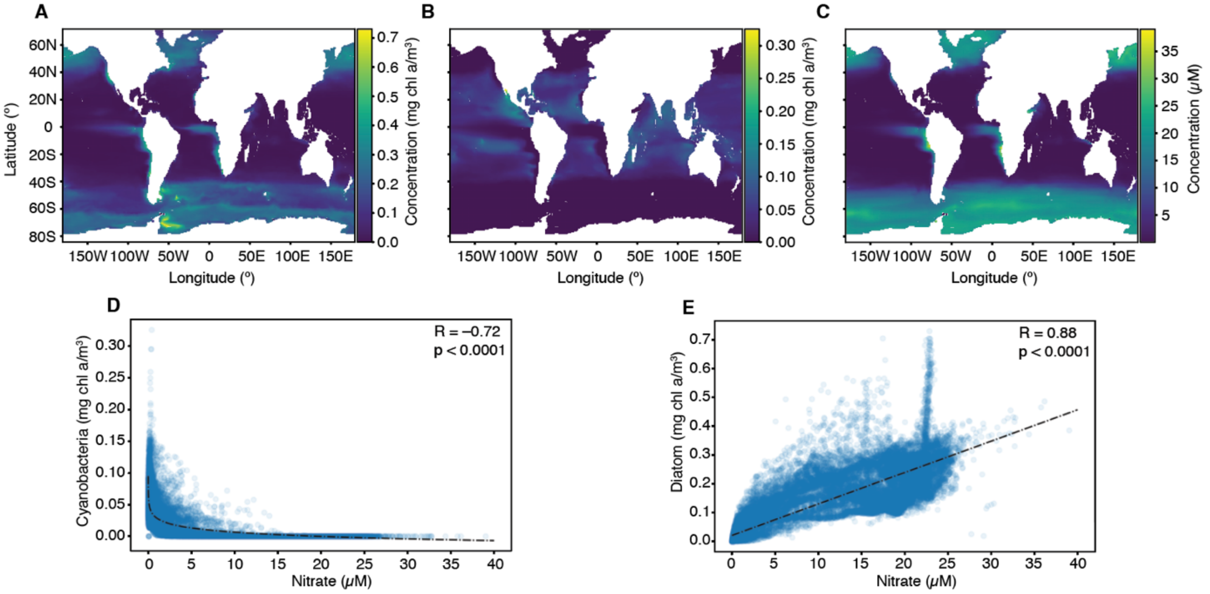
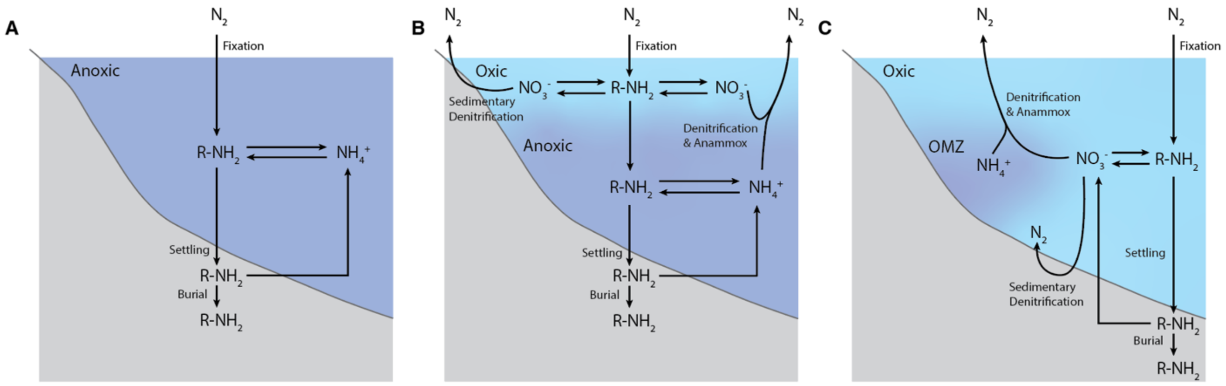


Fig. 1. 1. Distribution of diatoms, cyanobacteria, and nitrate in modern oceans.

(A–C) Time-averaged (01/2011–12/2015) concentration maps of first major feature (the most prominent distribution pattern) of diatom (A), cyanobacteria (B), and nitrate (C). Major spatial features were extracted using singular value decomposition (SVD) (see Supplementary Materials). Note that maps show concentrations in first major features rather than actual concentrations in oceans. Concentration data are from NASA Ocean Biogeochemical Model (NOBM) (*78*). (D–E) Cross-plots of cyanobacteria (D) and diatom (E) vs. nitrate concentrations in modern oceans, showing that diatoms are abundant in nitrate-rich areas whereas cyanobacteria are dominant in nitrate-limited areas. Time-averaged data of the first major component were used in the plots. For diatom and nitrate, simple linear regression model (y = a + b*x) was applied. For cyanobacteria, log function (y = a + b*log(x)) was applied. R and p values were calculated using Pearson correlation coefficient.



Fully anoxic ocean

Anoxic ocean with oxic surface water

Modern ocean with oxygen minimum zone (OMZ)

Fig. 1. 2. Simplified nitrogen cycles in Precambrian and modern oceans.

R-NH$_2$ represents organic nitrogen. (A) Nitrogen cycle in a fully anoxic ocean (e.g., in much of the Archean). Arrows between organic nitrogen and ammonium represent ammonification and nitrogen assimilation processes. (B) Nitrogen cycle in an anoxic ocean with oxic surface water (e.g., in much of the Proterozoic). Arrows from organic nitrogen to nitrate represent a combination of ammonification and nitrification processes. Arrows from nitrate to organic nitrogen represent the nitrogen assimilation process. (C) Nitrogen cycle in the modern ocean with an oxygen minimum zone (OMZ).

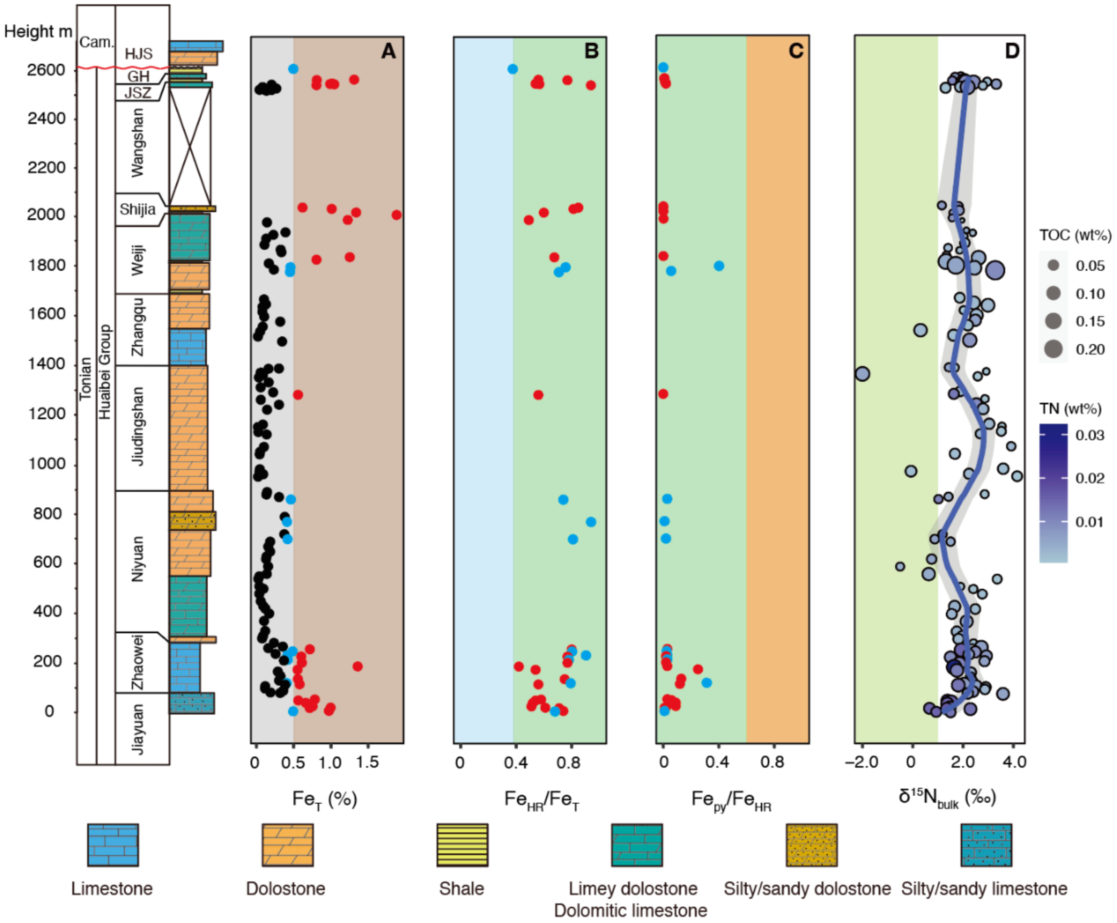


Fig. 1. 3. Geochemical profiles of Fe$_T$ (A), Fe$_{HR}$/Fe$_T$ (B), Fe$_{py}$/Fe$_{HR}$ (C), and δ^{15}N$_{bulk}$ (D) of the Huaibei Group samples.

Red points in (A–C) denote samples with Fe$_T$ > 0.5%; blue points, Fe$_T$ 0.4–0.5%; black points, Fe$_T$ < 0.4%. Blue band in (B) marks sediments deposited under oxic conditions and green band, anoxic conditions. In (C), green band represents ferruginous conditions, and orange band, euxinic conditions. In (D), green band marks near-zero δ^{15}N$_{bulk}$ values indicative of dominance of nitrogen fixation. Color and size of data points are keyed to reflect TN and TOC contents. Blue line is LOWESS regression fit with 95% confidence interval (gray shadow) calculated from standard errors.

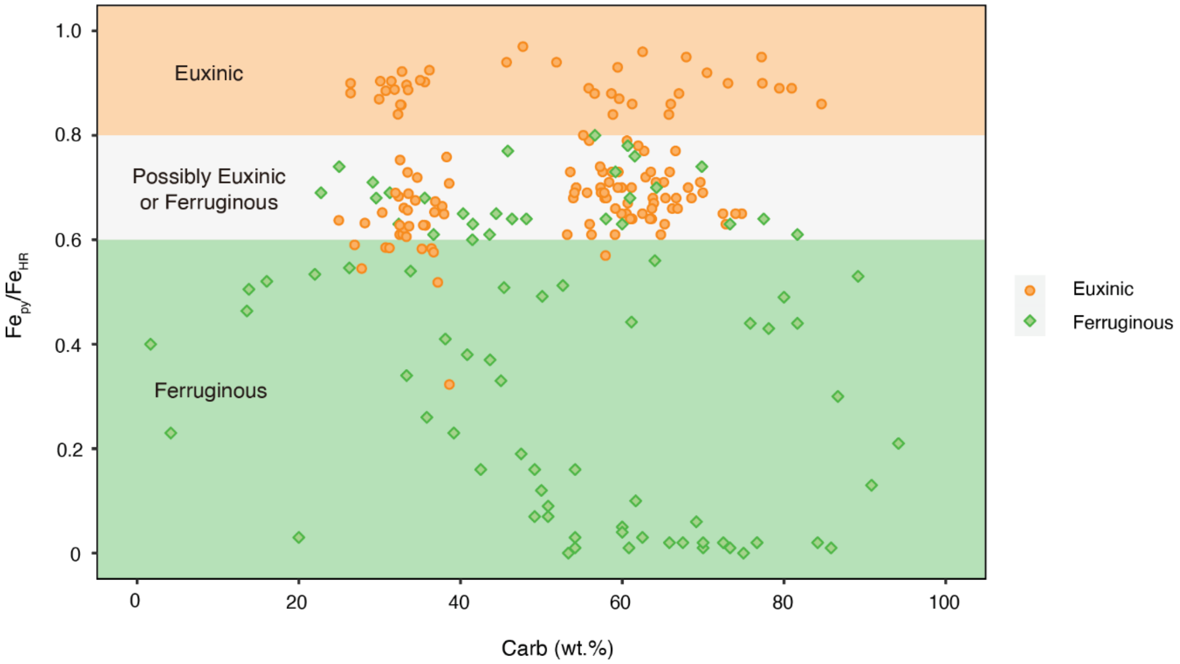


Fig. 1. 4. Fe$_{py}$/Fe$_{HR}$ ratios plotted against carbonate percentage (C$_{arb}$) for independently assessed euxinic and ferruginous samples.

Modern and ancient samples with Fe$_T$ > 0.5% were plotted. The data compilation confirms that the recently proposed euxinic/ferruginous threshold (0.6−0.8) is valid for distinguishing euxinic versus ferruginous conditions in carbonate samples.

See Table 1. S1 for data sources and independent assessment of redox conditions.

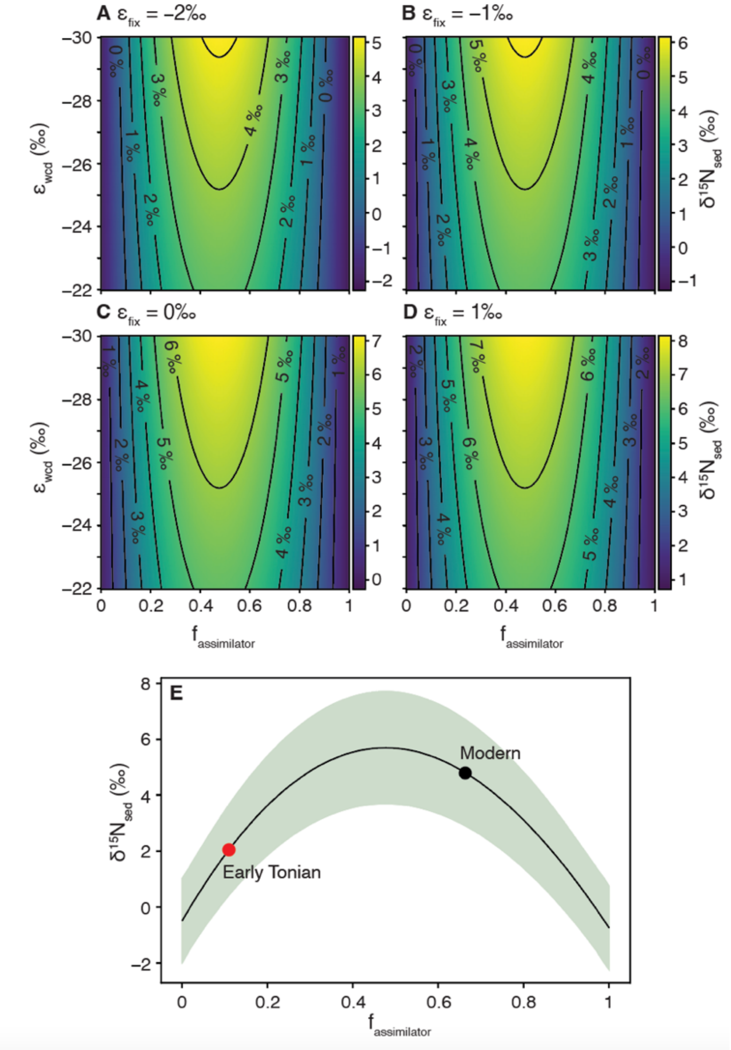


Fig. 1. 5. Results of numerical modeling, showing effect of $f_{assimilator}$, ε_{fix}, ε_{wcd} on $\delta^{15}N_{sed}$. (A–D) Model results showing effect of ε_{fix} and ε_{wcd} on sedimentary nitrogen isotope composition ($\delta^{15}N_{sed}$). Each plot shows modeled $\delta^{15}N_{sed}$ values with a fixed ε_{fix} value but variable ε_{wcd} values. When ε_{fix} is higher or ε_{wcd} is lower, $\delta^{15}N_{sed}$ becomes higher at a fixed $f_{assimilator}$; note that the ε_{wcd} axis is inverted. Relationship between $\delta^{15}N_{sed}$ and $f_{assimilator}$ is not linear: $\delta^{15}N_{sed}$ is highest at intermediate $f_{assimilator}$ values, but decreases when $f_{assimilator}$ approaches to zero (due to greater contribution from nitrogen-fixing organisms) or unity (due to weaker denitrification/anammox). (E) Modeled $\delta^{15}N_{sed}$ values with Monte Carlo sampling of ε_{fix} (–2 to 1‰, uniform distribution)

and ε_{wcd} (–30 to –22‰, uniform distribution). Black line represents median values and green shade shows 95% confidence interval. In modern oceans, $f_{assimilator}$ is ~0.7 and $\delta^{15}N_{sed}$ is ~5‰ (black dot). $\delta^{15}N_{sed}$ of Huaibei Group samples is ~2.0‰ (red dot), and $f_{assimilator}$ is estimated to be < 0.5 in more anoxic Tonian oceans.

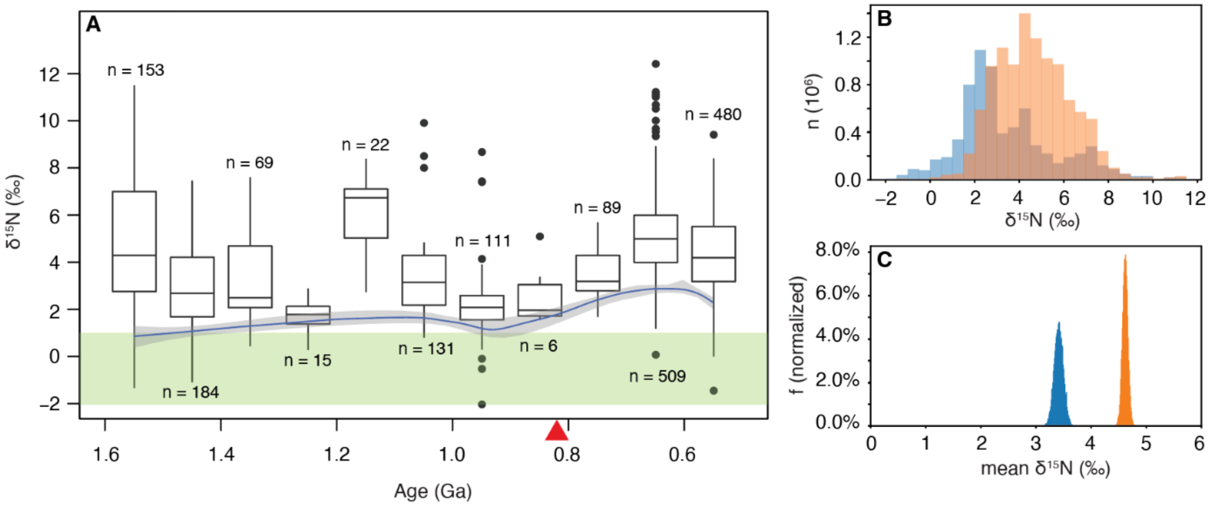


Fig. 1. 6. Nitrogen isotope composition ($\delta^{15}N$) of sedimentary rocks through time.

(A) Compilation of published nitrogen isotope data from Mesoproterozoic to Neoproterozoic. Boxplots show distribution of $\delta^{15}N$ values in each 100 Myr time bin. Blue line is LOWESS regression fit of lowest 25% $\delta^{15}N$ data in every 100 Myr time bin; see fig. 1. S6 for LOWESS regression fit of lowest 50% data and the entire dataset. Gray shade is 95% confidence interval calculated from 10,000 bootstrapping experiments. Red triangle marks stepwise increase from a changepoint analysis of lowest 25% $\delta^{15}N$ data. Most significant change in both mean and variance occurs at ca. 820 Ma. Green shade marks $\delta^{15}N$ values characteristic of a nitrogen cycle dominated by nitrogen fixation (e.g., $f_{assimilator} < 0.16$, see Fig. 1. 5). (B) Frequency distributions of $\delta^{15}N$ values from bootstrapping experiments (n = 10,000) of pre-800 Ma (blue) and post-800 Ma (orange) samples. (C) Frequency distributions of mean $\delta^{15}N$ values from bootstrapping experiments (n = 10,000) of pre-800 Ma (blue) and post-800 Ma (orange) samples.

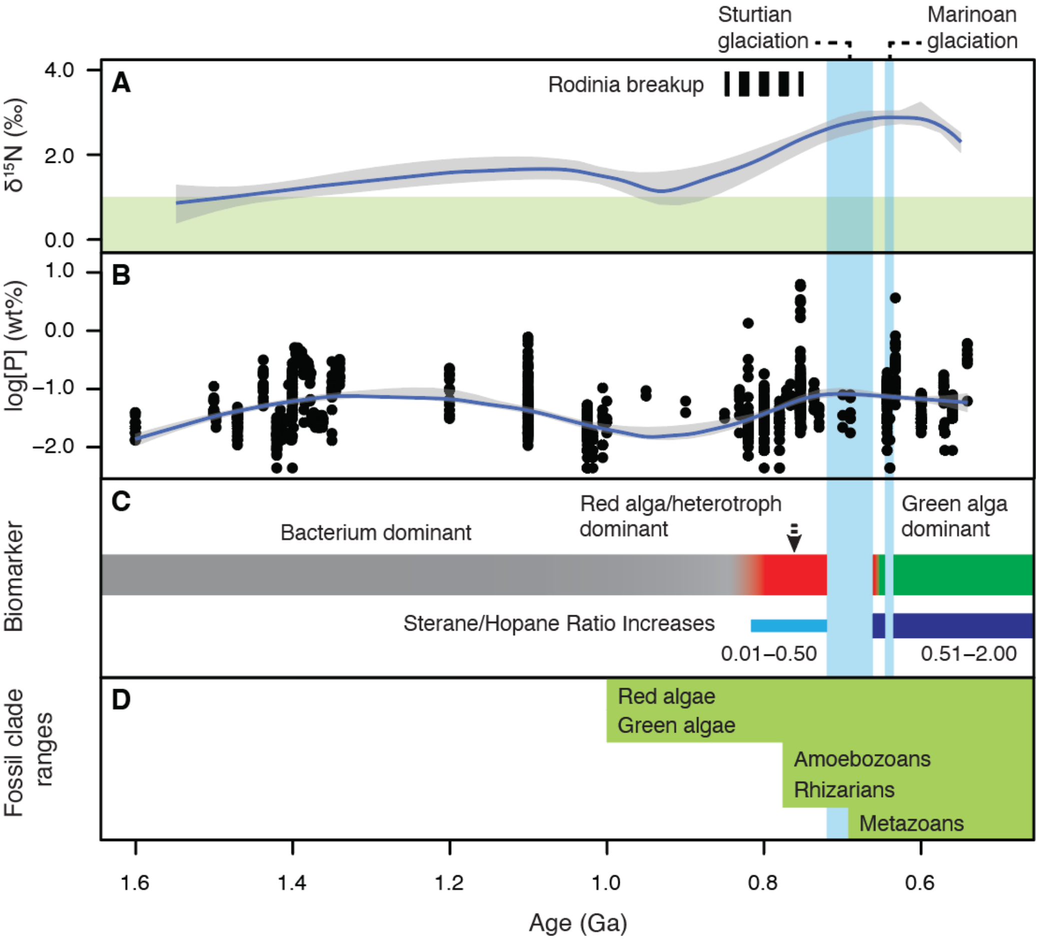


Fig. 1. 7. Secular variations in δ^{15}N and phosphorus content, in comparison with biomarker and eukaryotic fossil record of the Mesoproterozoic and Neoproterozoic eras.

(A) LOWESS regression curve of lowest 25% δ^{15}N data in every 100 Myr time bin. A stepwise rise in δ^{15}N at ca. 800 Ma coincides with breakup of supercontinent Rodinia and implies increasing marine nitrate availability. (B) Phosphorus concentrations of marine siliciclastic sediments through time (*13, 79*). LOWESS regression is based on the entire dataset. An increase in late Tonian implies greater phosphorus availability in the surface ocean. (C) Biomarker data (see Supplementary Material) (*80*). Note oldest steranes at ca. 820 Ma, increasing S/H ratios, and inferred dominance of various primary producers. (D) Approximate stratigraphic ranges of major eukaryotic clades (*5*).

CHAPTER 2 Carbonate $\delta^{13}C$ chemostratigraphy of the Neoproterozoic Hunjiang Group in North China and a tentative stratigraphic framework for the early Tonian Period

Authors

Junyao Kang[1,2], Qing Tang[3], Benjamin Gill[1,2], Rachel Reid[1,2], Wentao Zheng[4], Shuhai Xiao[1,2]

Affiliations

1. Department of Geosciences, Virginia Tech, Blacksburg, VA, USA
2. Global Change Center, Virginia Tech, Blacksburg, VA, USA
3. State Key Laboratory for Mineral Deposits Research, School of Earth Sciences and Engineering, and Frontiers Science Center for Critical Earth Material Cycling, Nanjing University, Nanjing, China
4. Department of Earth Science, University of California, Santa Barbara, CA, USA

ABSTRACT

A robust stratigraphic framework for the Tonian Period is necessary for us to better understand evolutionary and environmental changes during this critical geological time interval. Given the limited availability of radiometric and biostratigraphic data from Tonian strata, chemostratigraphic correlations have been used to build this framework. Here we report new carbonate carbon isotope chemostratigraphic data from the Hunjiang Group in the Baishan area of Jilin Province in North China. The carbon isotope data (average 3.00‰ ± 1.15‰, 1SD), together with existing detrital zircon and microfossil evidence, confirm an early Tonian age of the Hunjiang Group. Using the dynamic time warping algorithm, we then constructed an early Tonian global stratigraphic framework with $\delta^{13}C_{carb}$ data from the North China, São Francisco, and Congo cratons. This exercise confirms the generally narrow range of $\delta^{13}C_{carb}$ fluctuations in the early Tonian, but also substantiates a negative $\delta^{13}C_{carb}$ excursion of notable magnitude (~9 ‰) and of potential global extent at ca. 920 Ma. This negative excursion, known as the Majiatun excursion, is likely the oldest negative excursion in the Neoproterozoic Era and marks the onset of the dynamic Neoproterozoic carbon cycle.

INTRODUCTION

The Tonian Period witnessed critical biological and environmental changes in Earth history, including important evolutionary innovations in early eukaryotes (*1*), the rise of eukaryotes to ecological dominance (*2, 3*), changes in macronutrient supply (*4, 5*), and the return to ferruginous-dominant deeper oceans (*6*). To understand these global changes and their underlying mechanisms, a robust temporal framework for the Tonian Period is imperative. However, the lack of radiometric dates and index fossils from Tonian strata has hindered the application of chrono- and biostratigraphic correlations. Although several taxa have been proposed as Tonian index fossils, including *Cerebrosphaera* (*7*) and *Trachyhystrichosphaera aimika* (*8*), their applicability is limited because they are rarely preserved in carbonate rocks, which constitute a significant portion of the Tonian sedimentary record. Conversely, carbonate carbon isotope ($\delta^{13}C_{carb}$) chemostratigraphy has long been validated as a useful tool for stratigraphic correlation of Precambrian carbonate rocks (*9, 10*). However, as the carbon isotope composition of carbonate can be influenced by local sedimentary and diagenetic processes (*11-13*), a comprehensive assessment of global carbon isotope signals requires chemostratigraphic data from multiple localities worldwide to provide a reliable basis for building a solid correlative framework. In addition, conventional chemostratigraphic frameworks rely on interpretive correlations (*9*), which may introduce subjectivity and potential bias. To mitigate these issues, the application of a dynamic time warping algorithm on carbon isotope chemostratigraphic correlations can be used to achieve stratigraphic correlations with statistical significance and reproducibility (*14*).

Early Tonian strata exposed along the southeastern margin of the North China Craton (NCC) have been crucial for illuminating the global picture of Tonian Earth history, thanks to their minimal degree of metamorphic alteration, extensive stratigraphic coverage of the early Neoproterozoic, and exceptional preservation of both macro- and microfossils (*15-20*). Carbonate carbon isotope data have been previously reported from early Tonian sequences in several localities of NCC, including the Huaibei Group on the southern margin (*17*), and the Wuxingshan and Jinxian groups and Sangwon Supergroup on the eastern margin (*21, 22*). These data are important in building up a comprehensive stratigraphic framework for the early Tonian. Conversely, while the early Tonian sequences (e.g., the Hunjiang Group) in the northeastern part of North China are impressive due to their significant thickness of carbonates and well-preserved microfossils in siliciclastics (*23-26*), there is a scarcity of chemostratigraphic data in this region, resulting in incomplete stratigraphic correlations for the NCC.

In this study, we obtained new $\delta^{13}C_{carb}$ data from the Hunjiang Group in the Baishan area of southern Jilin Province in China. After integrating the Baishan data with existing $\delta^{13}C_{carb}$ data and radiometric dates from multiple areas in the NCC, we used the dynamic time warping algorithm (*14*) to build a relatively complete $\delta^{13}C$ record for the early Tonian Period in the NCC. Finally, using the same algorithm, we correlated the NCC record with previously published $\delta^{13}C$

data and geochronometric constraints from the São Francisco and Congo cratons to establish a global chemostratigraphic correlative framework for the early Tonian Period (Fig. 2. 1A).

STRATIGRAPHIC SETTINGS AND SAMPLE LOCALITIES

The NCC, which was stabilized around 1.9 Ga (*27-29*), is one of the most extensively examined cratons globally. Multiple Proterozoic sedimentary basins developed on the NCC from ca. 1.9 Ga to ca. 0.9 Ga (*28*). At ca. 1.1 Ga, siliciclastic and carbonate sediment started to accumulate in the Xuhuai rift system along the southeastern margin of the NCC (*30-32*). The Xuhuai rift system was offset by the Triassic-Jurassic strike-slip Tan-Lu fault (*33, 34*), and the Tonian sedimentary strata are now distributed in the Baishan, Pyongnam, Dalian, Huaibei, Huainan, and Jiaolai areas (Fig. 2. 1A–B).

The sampled Baishan section is located in southern Jilin Province. Here, a sedimentary sequence of late Mesoproterozoic to early Neoproterozoic age sits unconformably above the metamorphic basement of the Mesoproterozoic Laoling Group and below the phosphatic conglomerate of the early Cambrian Shuidong Formation (*35*) (Figs. 2. 1C and 2). The sequence consists of, in ascending order, the Baifangzi Formation, the Xihe Group, and the Hunjiang Group. The Xihe Group, including the Diaoyutai, Nanfen, and Qiaotou formations, is dominated by quartz sandstone, silty shale, and shale. The Hunjiang Group, which conformably overlies the Xihe Group and is divided into the Wanlong, Badaojiang, and Qinggouzi formations in ascending order, is dominated by limestone, silty limestone, dolomitic limestone, and black shale. Microfossils characteristic of late Mesoproterozoic to early Neoproterozoic age, including *Trachyhystrichosphaera aimika* and *Proterocladus antiquus*, have been reported from the Qinggouzi Formation (*26*). In this study, we focused on carbonate strata of the upper Wanlong

Formation and lower Badaojiang Formation, which outcrop in the Baishan section. For the Wanlong Formation, the sampled interval transitions from silty to dolomitic carbonate mudstones (Figs. 2. 2 and 3A–E). Molar tooth structures, which are composed of homogenous microsparitic calcite and show evidence of pre-compaction cementation, are pervasive in the silty carbonate mudstones (Fig. 2. 3C–E). For the Badaojiang Formation, the sampled interval consists of carbonate mudstone and stromatolitic boundstones (Figs. 2. 2 and 3F–I). Stromatolites in the Badaojiang Formation consist of dark- and light-colored laminae, with the former typically composed of rhombic dolomite crystals (Fig. 2. 3G–H).

The depositional ages of the Xihe and the Hunjiang groups are poorly constrained (Fig. 2. 2). Detrital zircon data place a maximum depositional age of 1147.3 ± 9.08 Ma on the Qiaotou Formation, and 1137.4 ± 9.84 Ma on the Wanlong Formation (*36*). Additionally, the youngest detrital zircons from the Diaoyutai Formation in the Dalian area to the south yield an age of 1056 ± 22 Ma (*37*), and a diabase sill intruding the Qiaotou Formation in the Dalian area has been dated at 947.8 ± 7.4 Ma (*38*). Thus, the sampled interval of this study can be broadly constrained between late Mesoproterozoic to early Neoproterozoic based on available biostratigraphic and radiometric data.

METHODS

In this study, 85 carbonate samples were collected from the Wanlong and Badaojiang formations of the Hunjinag Group at the Taipinghu Outcrop, Baishan, Jilin, China (41° 54′ 24.24″ N, 126° 18′ 09.26″ E) (Fig. 2. 1C). Standard petrographic thin sections were made and examined on a Zeiss AxioscopeA1 microscope with an Axiocam 512 digital camera, and an Olympus SZ16 microscope with a DP27 digital camera at Virginia Tech. Guided by petrographic observations, powders for geochemical analyses were taken from freshly cut carbonate samples using a Dremel microdrill.

Geochemical Analysis

Carbonate $\delta^{13}C$ and $\delta^{18}O$ measurements were conducted at Virginia Tech on an Isoprime 100 Isotope Ratio Mass Spectrometer coupled to a MultiFlowGEO headspace analyzer. For each sample, 0.4−0.6 mg of powder was treated overnight with glacial phosphoric acid at 70 °C. The evolved CO_2 was then analyzed on the mass spectrometer. Sample carbon and oxygen isotope compositions were normalized to the Vienna Pee Dee belemnite (VPDB) scale. IAEA-CO-1 ($\delta^{13}C = 2.492 \pm 0.030$ ‰, $\delta^{18}O = -2.4 \pm 0.1$ ‰), IAEA-CO-9 ($\delta^{13}C = -47.321 \pm 0.030$ ‰, $\delta^{18}O = -15.6 \pm 0.2$ ‰), and NIST NBS-18 ($\delta^{13}C = -5.014 \pm 0.035$ ‰, $\delta^{18}O = -23.2 \pm 0.1$ ‰) were used as carbon and oxygen isotope standards. The precision of carbonate carbon isotope analysis was 0.14 ‰ (1 SD), based on repeated measurements of IAEA-CO-1 ($n = 8$), IAEA-CO-9 ($n = 8$), and NIST NBS-18 ($n = 8$). The precision of carbonate oxygen isotope analysis was 0.49 ‰ (1 SD), based on repeated measurements of IAEA-CO-1 ($n = 8$), IAEA-CO-9 ($n = 8$), and NIST NBS-18 ($n = 8$).

Elemental concentrations of Mg, Ca, Mn, and Sr were analyzed at Virginia Tech on an inductively coupled plasma mass spectrometer (ICP-MS). For each sample, approximately 0.1 g powder was dissolved in glacial acetic acid. After complete digestion, samples were centrifuged and the supernatant was separated. The supernatant was dried down and redissolved in 1% acetic acid, which was then analyzed for major and trace element concentrations on a Thermo Electron iCAP-RQ ICP-MS per Standard Method 3125-B (*39*). The precision for Mg, Ca, Mn, and Sr concentrations is better than 5%.

Dynamic Time Warping

Dynamic time warping (DTW) is a dynamic programming technique widely used to compare and align time-series datasets. In geosciences, DTW has already been applied to the alignments of paleoclimate records (*40*), secular paleomagnetic variations (*41*), and carbon isotope chemostratigraphy (*14*). In this study, the algorithm by Hay et al. (*14*) was used to align chemostratigraphic data from the North China, São Francisco, and Congo cratons.

To align a candidate sequence (y_m) to a target sequence (x_n), an n × m cost matrix containing all feasible pairings between data points from these two sequences is constructed. The value of each element in the cost matrix is determined by calculating the squared difference between a data point in the target sequence and its corresponding data point in the candidate sequence. The goal is to find an optimal path that matches each index (m) from the candidate sequence to an index (n) in the target sequence while minimizing the sum of the cost matrix across all m (*14*). This algorithm employs two stratigraphically significant parameters to determine the optimal alignment. The edge parameter quantifies the extent of overlap between the two sequences, with a higher edge value indicating a stronger degree of overlap (*14, 42*). The g parameter characterizes the similarity in sediment accumulation rates between the two sections during the shared depositional interval, where a value of g greater than 1 penalizes stretching or squeezing (*14, 42*). Once the two sequences are aligned, the cross-correlation at zero lag of the overlapped interval, which equals the Pearson correlation coefficient, is calculated when the overlap is >10%. The statistical significance (p-value) of the correlation is then evaluated by the test statistic. In this study, alignments were calculated for all combinations of edge values ranging from 0.01 to 0.25 with increments of 0.01 and g values ranging from 0.98 to 1.01 with increments of 0.01. When comparing two sections, we explored two distinct scenarios. In each scenario, one section was designated as the target sequence while the other served as the candidate sequence. The alignment with the highest cross-correlation value, after considering geologic constraints, was adapted to build the stratigraphic framework.

In this study, early Tonian $\delta^{13}C_{carb}$ data from Dalian of Liaoning Province (*22*), Pyongnam of North Korea (*21*), Baishan of Jilin Province (this study), and Huaibei of Anhui and Jiangsu provinces (*17*) were chosen for DTW analysis to construct a composite chemostratigraphic sequence for the NCC, which is then aligned with early Tonian $\delta^{13}C_{carb}$ data from the São Francisco Craton (*43*) and the Congo Craton (*44*) using DTW to obtain a global chemostratigraphic sequence. To improve the accuracy of alignments, the uppermost parts of the Dalian, Huaibei, and Congo sequences, which are separated from the rest of the data within the sequence by major depositional hiatuses (>100 meters), were removed before DTW analysis. Throughout the analysis, a sequential correlation approach was employed. Each new sequence is correlated with the aggregated sequence resulting from the previous correlation step.

RESULTS

The $\delta^{13}C_{carb}$ values of the Wanlong and Badaojiang formations at the Baishan section are consistently positive and fall between 0 and 5 ‰ (average 3.0 ± 1.2 ‰, 1SD; Figs. 2. 2 and 4). From the Wanlong to the Badaojiang Formation, the $\delta^{13}C_{carb}$ values show a small-scale increasing and then decreasing trend. The variation in $\delta^{13}C_{carb}$ is decoupled from that of $\delta^{18}O_{carb}$, which is mostly between –4 ‰ and –8 ‰ in the Wanlong Formation and between –8 ‰ and –21 ‰ in the Badaojiang Formation. For the Wanlong and Badaojiang formations, carbonate contents and Mg/Ca ratios are consistent with the carbonate mineralogy of the units. Mn/Sr ratios (average 1.1 ± 1.0, 1SD) are relatively low throughout the measured section.

Early Tonian $\delta^{13}C_{carb}$ chemostratigraphic records from North China, São Francisco, and Congo cratons were correlated using DTW (Fig. 2. 5). The alignment between the Pyongnam and Dalian sequences yielded a maximum cross-correlation value of 0.93 ($p < 1e{-}10$, $g = 0.98$, $edge = 0.08$; Fig. 2. 6A). Subsequently, the alignment between the Huaibei sequence and the newly merged dataset containing the Pyongnam and Dalian data produced a maximum cross-correlation value of 0.74 ($p < 1e{-}10$, $g = 0.98$, $edge = 0.01$; Fig. 2. 6A). Next, the addition of the Baishan data yielded a maximum cross-correlation value of 0.86 ($p < 1e{-}6$, $g = 0.99$, $edge = 0.03$; Fig. 2. 6A). Finally, for the correlation between the North China Craton and São Francisco Craton (the Vazante Group), an alignment with a cross-correlation value of 0.75 ($p < 1e{-}10$, $g = 1.0$, $edge = 0.21$, Fig. 2. 6B) was selected (see discussion in section 5.4). For the alignment between the North China Craton and Congo Craton (the Mbuji-Mayi Supergroup), the cross-correlation value was not calculated because the overlap is less than 10 % (Fig. 2. 6B).

DISCUSSION

Evidence for primary oceanic signals

Based on the available petrographic, paleontological, and geochemical data, we argue that $\delta^{13}C_{carb}$ values from the Baishan sequence reflect primary oceanic signals. First, the sampled carbonates are mostly micritic limestones, although dolomitization is evident in some samples (Fig. 2. 3B, F, and G). Additionally, molar tooth structures consist of homogenous calcite microspar, which is thought to have precipitated during early diagenesis from porewater in direct connection with seawater (*45*). The exceptional fossil preservation in the Qinggouzi Formation

(*26*), which overlies the Badaojiang Formation, also suggests low degrees of thermal maturity and diagenetic alteration. Second, although there is a weak correlation between $\delta^{13}C_{carb}$ values and Mg/Ca ratios (Fig. 2. 4A–B), it is noteworthy that $\delta^{13}C_{carb}$ fluctuations, which can be driven by dolomitization (*46*), do not coincide with the lithological shift and the Mg/Ca change observed at the middle of the Wanlong Formation (Fig. 2. 2). Hence, $\delta^{13}C_{carb}$ values are likely not controlled by carbonate mineralogy. Instead, the higher $\delta^{13}C_{carb}$ values of dolomitic limestone in the upper Wanlong Formation and lower $\delta^{13}C_{carb}$ values of limestone in the Badaojiang Formation may reflect secular changes of oceanic signals. Third, as diagenesis tends to enrich Mn but deplete Sr (*47, 48*), the low Mn/Sr ratios (<2) of the Baishan samples suggest minimal diagenetic alterations (*49*) (Fig. 2. 2). This assessment is consistent with the weak correlation between $\delta^{13}C_{carb}$ values and Mn/Sr ratios (Fig. 2. 4C). Fourth, although $\delta^{18}O_{carb}$ exhibits large variations, $\delta^{13}C_{carb}$ remains in a relatively narrow range, resulting in a rather weak correlation between $\delta^{13}C_{carb}$ and $\delta^{18}O_{carb}$ (Fig. 2. 4D, $R = 0.44$ for Wanlong samples, $R = -0.0027$ for Badaojiang samples), and indicating that $\delta^{13}C_{carb}$ is better buffered against diagenetic alteration than is $\delta^{18}O_{carb}$ (*50, 51*).

Age constraints of the Baishan succession

The new chemostratigraphic data from the Baishan sequence, combined with existing geochronological and paleontological data, suggest a Tonian age of the Hunjiang Group. The late Mesoproterozoic to early Neoproterozoic $\delta^{13}C_{carb}$ record is characterized by consistently positive values with moderate variation (*17*). In addition, the early Neoproterozoic $\delta^{13}C_{carb}$ values (mostly 2 – 5 ‰) are slightly greater than those of the late Mesoproterozoic (0 – 4 ‰) (*17*). Therefore, $\delta^{13}C_{carb}$ values (average 3.00 ± 1.15 ‰, 1SD) recorded in the Wanlong and Badaojiang formations are consistent with an early Neoproterozoic age (Fig. 2. 2). This Tonian/early Neoproterozoic age is also supported by detrital zircon data (1147.3 ± 9.08 Ma) from the underlying Qiaotou Formation (*36*). The occurrence of *Trachyhystrichosphaera aimika* and *Proterocladus antiquus* from the Qinggouzi Formation further suggests a Tonian age of the Baishan strata (*26*). Finally, a bipartite network analysis of microfossil assemblages indicates that the Qinggouzi Formation is best correlated with the Tonian Gouhou Formation from the Huaibei Group (*26*).

The carbon isotope record of the NCC

The new carbon isotope data from the Baishan strata were integrated with existing data from Pyongnam (*21*), Dalian (*22*), and Huaibei (*17*) areas in the NCC for stratigraphic correlation using DTW (Fig. 2. 5 and 6). To initiate this process, we aligned and then merged the Pyongnam and Dalian data because both of them capture a distinct mid-amplitude (~9 ‰) negative excursion at ca. 920 Ma. The alignment with the highest cross-correlation value (0.93) was selected. This choice is supported by the alignment of radiometric dates within this framework and the absence of unrealistically large depositional gaps in the merged chemostratigraphic sequence (i.e., data points remained evenly distributed), given there is no evidence for large unconformities in the sections. In the next step, the Huaibei sequence was correlated with and incorporated in the newly merged dataset containing the Pyongnam and Dalian data. Here again, we opted for the alignment with the highest cross-correlation value (0.74), following the same rationale as in the initial integration. As a result, the ~5 ‰ decreasing trend at the uppermost portion of the Huaibei sequence was aligned with the descending limb of the ca. 920 Ma negative excursion identified in the Pyongnam and Dalian sequences. Finally, the

Baishan sequence was integrated into the dataset. It is important to acknowledge that while we chose the alignment with the highest cross-correlation value (0.86) for this integration, uncertainties persist due to the dearth of radiometric dates and distinct isotopic features in the Baishan sequence. This newly established, statistically significant integration unveils two distinct phases of the early Tonian $\delta^{13}C_{carb}$ record in the NCC: an interval marked by relatively moderate $\delta^{13}C$ variations with consistently positive values extending over approximately 100 million years, followed by a mid-amplitude negative excursion at ca. 920 Ma.

A tentative global stratigraphic framework for the early Tonian Period

To construct a global stratigraphic framework for the early Tonian Period, we compared the newly integrated $\delta^{13}C_{carb}$ record of the NCC with Tonian $\delta^{13}C_{carb}$ data from the São Francisco and Congo cratons (Fig. 2. 5 and 6). Specifically, data from the Vazante Group (*43*) deposited near the western border of the São Francisco Craton and the Mbuji-Mayi Supergroup (*44*) deposited on the Congo Craton were used for this exercise. Both units have been assigned an early Tonian age based on available radiometric dates, paleontological evidence, and carbon and strontium isotope data (*43, 44, 52-54*). Previous research suggested a Neoproterozoic paleogeographic connection between the NCC and the São Francisco and Congo cratons on the basis of similar sediment provenance (*32, 55*) and igneous rocks of similar lithology and ages (*56-58*). However, paleomagnetic data present a contrasting perspective, showing that the NCC and the São Francisco and Congo cratons were situated at different latitudes and were unlikely to have been paleogeographically connected in the early Tonian Period (*38, 59*). Given that the paleomagnetic data provide more direct evidence of paleogeographic location, and that sediment provenance and magmatic data do not necessarily demand unique paleogeographic reconstructions, we argue that the compilation of $\delta^{13}C_{carb}$ data from these three cratons would likely give a global rather than local picture of early Tonian chemostratigraphy.

In contrast to the relatively straightforward alignments within the NCC, the cross-basin comparison of data from the three cratons presents greater challenges due to the potential influence of local processes on the $\delta^{13}C_{carb}$ record. When conducting chemostratigraphic alignment between the NCC and the São Francisco Craton, data were first normalized because the Vazante Group has consistently lower $\delta^{13}C$ values compared with that of the NCC, which can be driven by local processes during or after deposition. For instance, cross-shelf seawater $\delta^{13}C$ gradients, due to the influx of remineralized organic carbon from intertidal creeks, have been observed in the modern Great Bahama Bank (*11*). Additionally, the formation of authigenic carbonate during early diagenesis can decrease the $\delta^{13}C$ values of bulk samples (*60*). Given the available radiometric dates, it is highly likely that the negative carbon isotope excursions recorded in both cratons are synchronous. To facilitate their correlation, we selected an alignment with a cross-correlation value of 0.75 (Fig. 2. 6B), and while alignments with higher cross-correlation values were available, they either resulted in large depositional hiatuses that are not supported by sedimentary evidence (Fig. 2. 7A) or failed to accurately match the excursions (Fig. 2. 7B).

For the chemostratigraphic correlation between the NCC and the Congo Craton, we propose that the most negative $\delta^{13}C$ values within an increasing up section trend in the lower Mbuji-Mayi Supergroup (specifically, the BIIa − BIIc subgroups) capture the rising limb of the ca. 920 Ma negative excursion. Previous research suggested that the negative values within the BIIa − BIIc subgroups corresponded to the Bitter Springs Anomaly (*44*). However, available radiometric dates indicate that the BIIa – BIIc subgroups are older than 948 ± 20 Ma (*61, 62*), thus predating the ca. 810 Ma Bitter Springs event (*63-65*). To accurately align the negative excursions, we used an edge value of 0.005, a selection driven by the limited overlap between the two $\delta^{13}C_{carb}$ profiles. Additionally, the cross-correlation value was not calculated for this alignment, because that the overlap between the records was less than 10% (Fig. 2. 6B).

By consolidating data from these three cratons, we have established a tentative chemostratigraphic framework for the early Tonian Period (Fig. 2. 5, 6, and 8). This framework reaffirms the moderate variability of the early Tonian $\delta^{13}C$ record, but also underscores the possibility that the ca. 920 Ma negative $\delta^{13}C_{carb}$ excursion could potentially be a global signal, given the presence of this excursion across multiple cratons. Accordingly, this excursion likely represents the oldest negative excursion in the Neoproterozoic Era and marks the onset of Neoproterozoic carbon cycle volatility (Fig. 2. 8).

The ultimate cause of the ca. 920 Ma negative $\delta^{13}C_{carb}$ excursions is unclear. Zhang et al. (*22*), who first documented this excursion in the NCC and termed it the Majiatun anomaly, argued that it was caused by a rapid movement of the NCC from tropical to subtropical latitudes, leading to a reduction in organic carbon burial. A similar mechanism has been invoked to explain the later Bitter Springs anomaly. Briefly, tropical land masses may have moved toward the poles during a true polar wander event (*66*), thereby decreasing organic carbon burial in tropical deltas and reducing surface ocean $\delta^{13}C$ values (*66, 67*). For the Majiatun excursion, there is indeed paleomagnetic data suggesting a mid- to high-latitude migration for the São Francisco and Congo cratons at ca. 920 Ma *(59, but see 68 for positing São Francisco and Congo cratons near the equator at ca. 920 Ma)*. However, to further assess this hypothesis, additional high-resolution geochronological and geomagnetic data from both the NCC and São Francisco/Congo Craton are needed to chart continental movements throughout the excursion. On the other hand, given the prevalence of ca. 920 Ma mafic sills in the NCC and São Francisco/Congo Craton (*56-59, 69*), an alternative driver for the Majiatun excursion may be extensive volcanic activity (Fig. 2. 8). Similar to the Siberian Traps and the Central Atlantic Magmatic Province, which may have been the ultimate cause of negative $\delta^{13}C_{carb}$ excursions related to the end-Permian and end-Triassic mass extinctions, respectively (*70, 71*), the ca. 920 Ma mafic sills on the NCC and São Francisco/Congo Craton may represent a large igneous province (*59*) and may have contributed to the Majiatun negative $\delta^{13}C_{carb}$ excursion. It is also necessary to highlight the disparate temporal scales associated with the two proposed mechanisms. While LIPs are characterized by relatively brief durations (< 2 million years) (*72*), the process of continental movement unfolds over much longer timescales (> 10 million years) (*73, 74*). Consequently, additional geochronological data may help delineate the temporal extent of this excursion, thereby deepening our understanding of the underlying driving mechanism.

CONCLUSIONS

Available radiometric, paleontological, and chemostratigraphic data suggest that the Hunjiang Group at the Baishan area in North China is Tonian in age. Integrating new and existing $\delta^{13}C_{carb}$ chemostratigraphic data from the North China, São Francisco, and Congo cratons, we constructed a stratigraphic framework for the early Tonian Period, using the dynamic time warping algorithm. The framework confirmed the moderate variability of the $\delta^{13}C_{carb}$ record in the early Tonian and suggested the ca. 920 Ma Majiatun excursion as a potential global $\delta^{13}C_{carb}$ chemostratigraphic feature, which is a different event from the Bitter Springs Anomaly. Although the driver of this excursion remains unclear, it likely marks the onset of Neoproterozoic carbon cycle volatility.

REFERENCES

1. S. Xiao, Q. Tang, After the boring billion and before the freezing millions: evolutionary patterns and innovations in the Tonian Period. *Emerging Topics in Life Sciences* **2**, 161-171 (2018).
2. J. J. Brocks *et al.*, The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* **548**, 578-581 (2017).
3. J. A. Zumberge, D. Rocher, G. D. Love, Free and kerogen‐bound biomarkers from late Tonian sedimentary rocks record abundant eukaryotes in mid‐Neoproterozoic marine communities. *Geobiology* **18**, 326-347 (2020).
4. J. Kang, B. Gill, R. Reid, F. Zhang, S. Xiao, Nitrate limitation in early Neoproterozoic oceans delayed the ecological rise of eukaryotes. *Science Advances* **9**, eade9647 (2023).
5. C. T. Reinhard *et al.*, Evolution of the global phosphorus cycle. *Nature* **541**, 386–389 (2017).
6. R. Guilbaud, S. W. Poulton, N. J. Butterfield, M. Zhu, G. A. Shields-Zhou, A global transition to ferruginous conditions in the early Neoproterozoic oceans. *Nature Geoscience* **8**, 466-470 (2015).
7. Y. Cornet *et al.*, New insights on the paleobiology, biostratigraphy and paleogeography of the pre-Sturtian microfossil index taxon Cerebrosphaera. *Precambrian Research* **332**, 105410 (2019).
8. K. Pang, Q. Tang, B. Wan, X. Yuan, New insights on the palaeobiology and biostratigraphy of the acritarch Trachyhystrichosphaera aimika: A potential late Mesoproterozoic to Tonian index fossil. *Palaeoworld* **29**, 476-489 (2020).
9. G. P. Halverson, P. F. Hoffman, D. P. Schrag, A. C. Maloof, A. H. N. Rice, Toward a Neoproterozoic composite carbon-isotope record. *GSA Bulletin* **117**, 1181-1207; doi: 1110.1130/B25630.25631 (2005).
10. A. H. Knoll, M. R. Walter, Latest Proterozoic stratigraphy and Earth history. *Nature* **356**, 673-678 (1992).
11. E. C. Geyman, A. C. Maloof, Facies control on carbonate δ13C on the Great Bahama Bank. *Geology* **49**, 1049-1054 (2021).
12. J. A. Higgins *et al.*, Mineralogy, early marine diagenesis, and the chemistry of shallow-water carbonate sediments. *Geochimica et Cosmochimica Acta* **220**, 512-534 (2018).
13. S. Xiao *et al.*, Using SIMS to decode noisy stratigraphic δ13C variations in Ediacaran carbonates. *Precambrian Research* **343**, 105686 (2020).
14. C. C. Hay, J. R. Creveling, C. J. Hagen, A. C. Maloof, P. Huybers, A library of early Cambrian chemostratigraphic correlations from a reproducible algorithm. *Geology* **47**, 457-460 (2019).
15. L. Dong *et al.*, Restudy of the worm-like carbonaceous compression fossils *Protoarenicola*, *Pararenicola*, and *Sinosabellidites* from early Neoproterozoic successions in North China. *Palaeogeography Palaeoclimatology Palaeoecology* **258**, 138-161 (2008).

16. Q. Tang, K. Pang, X. Yuan, B. Wan, S. Xiao, Organic-walled microfossils from the Tonian Gouhou Formation, Huaibei region, North China Craton, and their biostratigraphic implications. *Precambrian Research* **266**, 296-318 (2015).
17. S. Xiao *et al.*, Biostratigraphic and chemostratigraphic constraints on the age of early Neoproterozoic carbonate successions in North China. *Precambrian Research* **246**, 208–225 (2014).
18. C. Yin, Micropaleoflora from the late Precambrian in Huainan region of Anhui Province and its stratigraphic significance. *Professional Papers of Stratigraphy and Palaeontology, Chinese Academy of Geological Sciences* **12**, 97-119 (1985).
19. Q. Tang, K. Pang, X. Yuan, S. Xiao, A one-billion-year-old multicellular chlorophyte. *Nature Ecology & Evolution* **4**, 543-549 (2020).
20. D. Li, G. Luo, Q. Tang, Z. She, S. Xiao, New record of the green algal fossil Proterocladus and coexisting microfossils from the Meso-Neoproterozoic Diaoyutai Formation in southern Liaoning, North China. *Precambrian Research* **393**, 107104 (2023).
21. H. Park *et al.*, Deposition age of the Sangwon Supergroup in the Pyongnam basin (Korea) and the Early Tonian negative carbon isotope interval. *Acta Petrologica Sinica* **32**, 2181-2195 (2016).
22. Z. Zhang *et al.*, Oldest-known Neoproterozoic carbon isotope excursion: Earlier onset of Neoproterozoic carbon cycle volatility. *Gondwana Research* **94**, 1-11 (2021).
23. L. Gao, Y. Xing, G. Liu, Neoproterozoic micropalaeoflora from Hunjiang area, Jilin Province and its sedimentary environment. *Professional Papers of Stratigraphy and Palaeontology* **26**, 1-23 (1995).
24. X. Qiao *et al.*, Outcrop sequence stratigraphy of the Neoproterozoic in the Hunjiang Area, Jilin Province. *Reginal Geology of China* **2**, 99-111 (1995).
25. L. Yin, in *Stratigraphy and Palaeontology of Systemic Boundaries in China: Precambrian–Cambrian Boundary (1),* Nanjing Institute of Geology and Palaeontology Academica Sinica, Ed. (Nanjing University Press, Nanjing, 1987), pp. 415-494.
26. W. Zheng *et al.*, The Proterozoic Qinggouzi microfossil assemblage and its biostratigraphic constraints on the Great Unconformity in northeastern North China Craton. *Precambrian Research* **395**, 107130 (2023).
27. T. M. Kusky, J. Li, Paleoproterozoic tectonic evolution of the North China Craton. *Journal of Asian Earth Sciences* **22**, 383-397 (2003).
28. M. Zhai, B. Hu, T. Zhao, P. Peng, Q. Meng, Late Paleoproterozoic–Neoproterozoic multi-rifting events in the North China Craton and their geological significance: a study advance and review. *Tectonophysics* **662**, 153-166 (2015).
29. G. Zhao, M. Sun, S. A. Wilde, S. Li, Assembly, accretion and breakup of the Paleo-Mesoproterozoic Columbia Supercontinent: records in the North China Craton. *Gondwana Research* **6**, 417-434 (2003).
30. Q. Meng, H. Wei, Y. Qu, S. Ma, Stratigraphic and sedimentary records of the rift to drift evolution of the northern North China craton at the Paleo-to Mesoproterozoic transition. *Gondwana Research* **20**, 205-218 (2011).
31. X. Su *et al.*, Multi-stage evolution of the Xuhuai rift: Insights from the occurrence and compositional profiles of doleritic sills in the southeastern margin of the North China Craton. *Gondwana Research* **82**, 221-240 (2020).

32. F. Sun *et al.*, Provenance analysis of the late Mesoproterozoic to Neoproterozoic Xuhuai Basin in the southeast North China Craton: Implications for paleogeographic reconstruction. *Precambrian Research* **337**, 105554 (2020).
33. J. Xu, G. Zhu, W. Tong, K. Cui, Q. Liu, Formation and evolution of the Tancheng-Lujiang wrench fault system: a major shear system to the northwest of the Pacific Ocean. *Tectonophysics* **134**, 273-310 (1987).
34. G. Zhu *et al.*, 40Ar/39Ar dating of strike-slip motion on the Tan–Lu fault zone, East China. *Journal of Structural Geology* **27**, 1379-1398 (2005).
35. Bureau of Geology and Mineral Resources of Jilin Province, *Regional Geology of Jilin Province. Chinese Ministry of Geology and Mineral Resources, Geological Memoirs, series 1, number 10.* (Geological Publishing House, Beijing, 1988).
36. Z. Wen, L. Fulai, L. Chaohui, Detrital zircon U-Pb ages of the late Mesoproterozoic–Neoproterozoic Qiaotou formation in the Liao-Ji area of the North China Craton: implications for Rodinia reconstruction. *International Geology Review*, 1-20 (2020).
37. D. Yang *et al.*, U–Pb ages and Hf isotope data from detrital zircons in the Neoproterozoic sandstones of northern Jiangsu and southern Liaoning Provinces, China: implications for the Late Precambrian evolution of the southeastern North China Craton. *Precambrian Research* **216**, 162-176 (2012).
38. H. Zhao *et al.*, New geochronologic and paleomagnetic results from early Neoproterozoic mafic sills and late Mesoproterozoic to early Neoproterozoic successions in the eastern North China Craton, and implications for the reconstruction of Rodinia. *GSA Bulletin* **132**, 739-766 (2020).
39. A. American Public Health, A. American Water Works, F. Water Environment, *Standard methods for the examination of water and wastewater.* (APHA-AWWA-WEF, Washington, D.C., ed. 20th, 1998).
40. L. E. Lisiecki, P. A. Lisiecki, Application of dynamic programming to the correlation of paleoclimate records. *Paleoceanography* **17**, 1-1-1-12 (2002).
41. C. J. Hagen, B. T. Reilly, J. S. Stoner, J. R. Creveling, Dynamic time warping of palaeomagnetic secular variation data. *Geophysical Journal International* **221**, 706-721 (2020).
42. H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing* **26**, 43-49 (1978).
43. C. J. Alvarenga *et al.*, Carbonate chemostratigraphy of the Vazante Group, Brazil: A probable Tonian age. *Precambrian Research* **331**, 105378 (2019).
44. F. Delpomdor, A. Préat, Early and late Neoproterozoic C, O and Sr isotope chemostratigraphy in the carbonates of West Congo and Mbuji-Mayi supergroups: a preserved marine signature? *Palaeogeography, Palaeoclimatology, Palaeoecology* **389**, 35-47 (2013).
45. A. Kriscautzky, L. C. Kah, J. K. Bartley, Molar-Tooth Structure as a Window into the Deposition and Diagenesis of Precambrian Carbonate. *Annual Review of Earth and Planetary Sciences* **50**, 205-230 (2022).
46. U. Bold *et al.*, Effect of dolomitization on isotopic records from Neoproterozoic carbonates in southwestern Mongolia. *Precambrian Research* **350**, 105902 (2020).
47. U. Brand, J. Veizer, Chemical diagenesis of a multicomponent carbonate system; 1, Trace elements. *Journal of Sedimentary Research* **50**, 1219-1236 (1980).

48. P. K. Swart, The geochemistry of carbonate diagenesis: The past, present and future. *Sedimentology* **62**, 1233–1304 (2015).
49. A. J. Kaufman, A. H. Knoll, Neoproterozoic variations in the C-isotope composition of sea water: Stratigraphic and biogeochemical implications. *Precambrian Research* **73**, 27-49 (1995).
50. U. Brand, J. Veizer, Chemical diagenesis of a multicomponent carbonate system; 2, Stable isotopes. *Journal of Sedimentary Research* **51**, 987-997 (1981).
51. K. V. Lau, D. S. Hardisty, Modeling the impacts of diagenesis on carbonate paleoredox proxies. *Geochimica et Cosmochimica Acta* **337**, 123-139 (2022).
52. B. K. Baludikay, J.-Y. Storme, C. François, D. Baudet, E. J. Javaux, A diverse and exquisitely preserved organic-walled microfossil assemblage from the Meso-Neoproterozoic Mbuji-Mayi Supergroup (Democratic Republic of Congo) and implications for Proterozoic biostratigraphy. *Precambrian Research* **281**, 166-184 (2016).
53. F. d. A. Caxito *et al.*, in *Stratigraphy & Timescales*. (Elsevier, 2019), vol. 4, pp. 73-132.
54. F. Delpomdor, C. Blanpied, A. Virgone, A. Préat, Paleoenvironments in Meso–Neoproterozoic carbonates of the Mbuji-Mayi Supergroup (Democratic Republic of Congo)–Microfacies analysis combined with C–O–Sr isotopes, major-trace elements and REE+ Y distributions. *Journal of African Earth Sciences* **88**, 72-100 (2013).
55. F. Sun, P. Peng, D. Zheng, P. Zuo, Reappraising the provenance of early Neoproterozoic strata in the southern–southeastern North China Craton and its implication for paleogeographic reconstruction. *Minerals* **12**, 510 (2022).
56. J. Cederberg, U. Söderlund, E. P. Oliveira, R. E. Ernst, S. A. Pisarevsky, U-Pb baddeleyite dating of the Proterozoic Pará de Minas dyke swarm in the São Francisco craton (Brazil)–implications for tectonic correlation with the Siberian, Congo and North China cratons. *Gff* **138**, 219-240 (2016).
57. A. de Oliveira Chaves, R. E. Ernst, U. Söderlund, X. Wang, T. Naeraa, The 920–900 Ma Bahia-Gangila LIP of the São Francisco and Congo cratons and link with Dashigou-Chulan LIP of North China craton: New insights from U-Pb geochronology and geochemistry. *Precambrian Research* **329**, 124-137 (2019).
58. P. Peng *et al.*, Neoproterozoic (~ 900 Ma) Sariwon sills in North Korea: Geochronology, geochemistry and implications for the evolution of the south-eastern margin of the North China Craton. *Gondwana Research* **20**, 243-254 (2011).
59. D. A. D. Evans *et al.*, Return to Rodinia? Moderate to high palaeolatitude of the São Francisco/Congo craton at 920 Ma. *Geological Society, London, Special Publications* **424**, 167-190 (2016).
60. D. P. Schrag, J. A. Higgins, F. A. Macdonald, D. T. Johnston, Authigenic carbonate and the history of the global carbon cycle. *Science* **339**, 540-543 (2013).
61. L. Cahen, D. Ledent, N. J. Snelling, Données géochronologiques dans le Katangien inférieur du Kasai oriental et du Shaba nord-oriental (République du Zaïre). *Musée royal de l'Afrique Centrale, Tervuren, Rapport annuel du Département de Géologie et de Minéralogie*, 59-70 (1974).
62. F. Delpomdor *et al.*, Depositional age, provenance, and tectonic and paleoclimatic settings of the late Mesoproterozoic–middle Neoproterozoic Mbuji-Mayi Supergroup, Democratic Republic of Congo. *Palaeogeography, palaeoclimatology, palaeoecology* **389**, 4-34 (2013).

63. P. A. Cohen, J. V. Strauss, A. D. Rooney, M. Sharma, N. Tosca, Controlled hydroxyapatite biomineralization in an ~810 million-year-old unicellular eukaryote. *Science Advances* **3**, e1700095 (2017).
64. G. P. Halverson, S. M. Porter, T. M. Gibson, Dating the late Proterozoic stratigraphic record. *Emerging Topics in Life Sciences* **2**, 137-147 (2018).
65. F. A. Macdonald *et al.*, Calibrating the Cryogenian. *Science* **327**, 1241-1243 (2010).
66. A. C. Maloof *et al.*, Combined paleomagnetic, isotopic, and stratigraphic evidence for true polar wander from the Neoproterozoic Akademikerbreen Group, Svalbard, Norway. *Geological Society of America Bulletin* **118**, 1099–1124 (2006).
67. D. P. Schrag, R. A. Berner, P. F. Hoffman, G. P. Halverson, On the initiation of a snowball Earth. *Geochemistry, Geophysics, Geochemistry (G^3)* **3(4)**, DOI number 10.1029/2001GC000219 (2002).
68. R. I. Trindade, M. S. D'Agrella-Filho, P. Y. Antonio, W. Teixeira, in *Ancient supercontinents and the paleogeography of Earth*. (Elsevier, 2021), pp. 445-464.
69. R. Ernst, M. Wingate, K. Buchan, Z.-X. Li, Global record of 1600–700 Ma Large Igneous Provinces (LIPs): implications for the reconstruction of the proposed Nuna (Columbia) and Rodinia supercontinents. *Precambrian Research* **160**, 159-178 (2008).
70. A. Marzoli *et al.*, Extensive 200-million-year-old continental flood basalts of the Central Atlantic Magmatic Province. *Science* **284**, 616-618 (1999).
71. P. R. Renne, M. T. Black, Z. Zichao, M. A. Richards, A. R. Basu, Synchrony and causal relations between Permian-Triassic boundary crises and Siberian flood volcanism. *Science* **269**, 1413-1416 (1995).
72. R. E. Ernst *et al.*, Large igneous province record through time and implications for secular environmental changes and geological time - scale boundaries. *Large igneous provinces: A driver of global environmental and biotic changes*, 1-26 (2021).
73. D. Evans *et al.*, Return to Rodinia? Moderate to high palaeolatitude of the São Francisco/Congo craton at 920 Ma. *Geological Society, London, Special Publications* **424**, 167-190 (2016).
74. D. A. Evans, in *Ancient Supercontinents and the Paleogeography of Earth*. (Elsevier, 2021), pp. 549-576.
75. T. He *et al.*, Measuring the 'Great Unconformity' on the North China Craton using new detrital zircon age data. *Geological Society, London, Special Publications* **448**, 145-159 (2017).
76. S. Zhang, Y. Zhao, H. Ye, G. Hu, Early Neoproterozoic emplacement of the diabase sill swarms in the Liaodong Peninsula and pre-magmatic uplift of the southeastern North China Craton. *Precambrian Research* **272**, 203-225 (2016).
77. B. Hu *et al.*, Mesoproterozoic magmatic events in the eastern North China Craton and their tectonic implications: Geochronological evidence from detrital zircons in the Shandong Peninsula and North Korea. *Gondwana Research* **22**, 828-842 (2012).
78. X. Su, P. Peng, S. Foley, W. Teixeira, M. Zhai, Initiation of continental breakup documented in evolution of the magma plumbing system of the ca. 925 Ma Dashigou large igneous province, North China. *Lithos* **384**, 105984 (2021).
79. R. Zhu *et al.*, Geochronology, geochemistry and petrogenesis of the Laozhaishan dolerite sills in the southeastern margin of the North China Craton and their geological implication. *Gondwana Research* **67**, 131-146 (2019).

80. S. Zhang *et al.*, SIMS Pb-Pb dating of phosphates in the Proterozoic strata of SE North China Craton: Constraints on eukaryote evolution. *Precambrian Research* **371**, 106562 (2022).
81. X. Fu *et al.*, New paleomagnetic results from the Huaibei Group and Neoproterozoic mafic sills in the North China Craton and their paleogeographic implications. *Precambrian Research* **269**, 90-106 (2015).
82. K. Azmy, B. Kendall, R. A. Creaser, L. Heaman, T. F. de Oliveira, Global correlation of the Vazante Group, São Francisco Basin, Brazil: Re–Os and U–Pb radiometric age constraints. *Precambrian Research* **164**, 160-172 (2008).
83. J. Rodrigues *et al.*, Provenance of the Vazante Group: New U–Pb, Sm–Nd, Lu–Hf isotopic data and implications for the tectonic evolution of the Neoproterozoic Brasília Belt. *Gondwana Research* **21**, 439-450 (2012).
84. G. M. Cox *et al.*, Continental flood basalt weathering as a trigger for Neoproterozoic Snowball Earth. *Earth and Planetary Science Letters* **446**, 89-99 (2016).

FIGURES

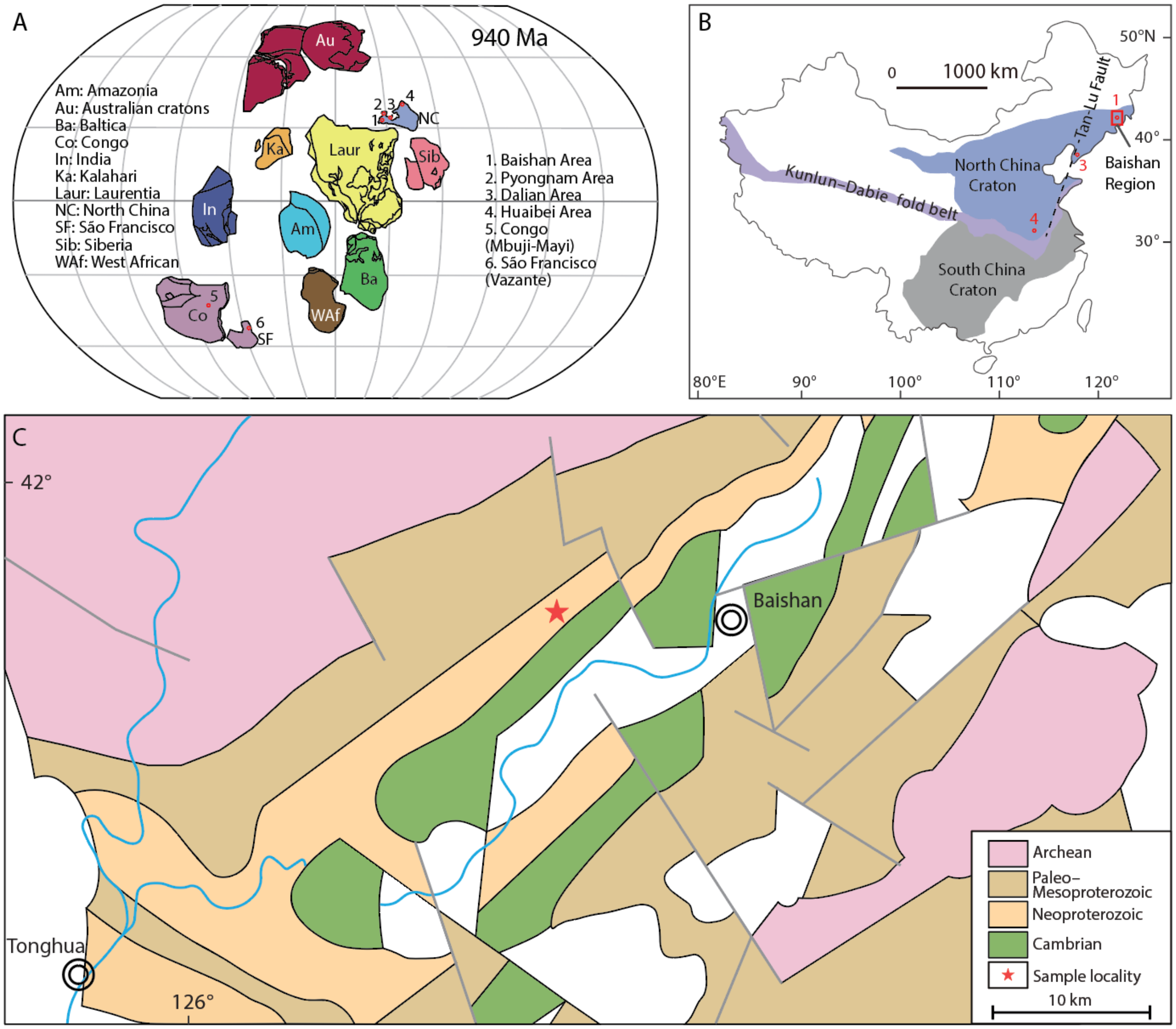


Fig. 2. 1. Geological maps.

(A) Paleogeographic reconstruction of the Rodinia supercontinent at ~940 Ma. Modified from Evans (*74*). Locations of Tonian sections are marked as 1–6. (B) Simplified geological map of China showing the location of the Baishan region (red square) within the North China Craton. Locations of Tonian sections are marked as 1, 3, and 4. (C) Geological map of the Baishan region showing the sample location of the Hunjiang Group (red star). Modified from Zheng et al. (*26*).

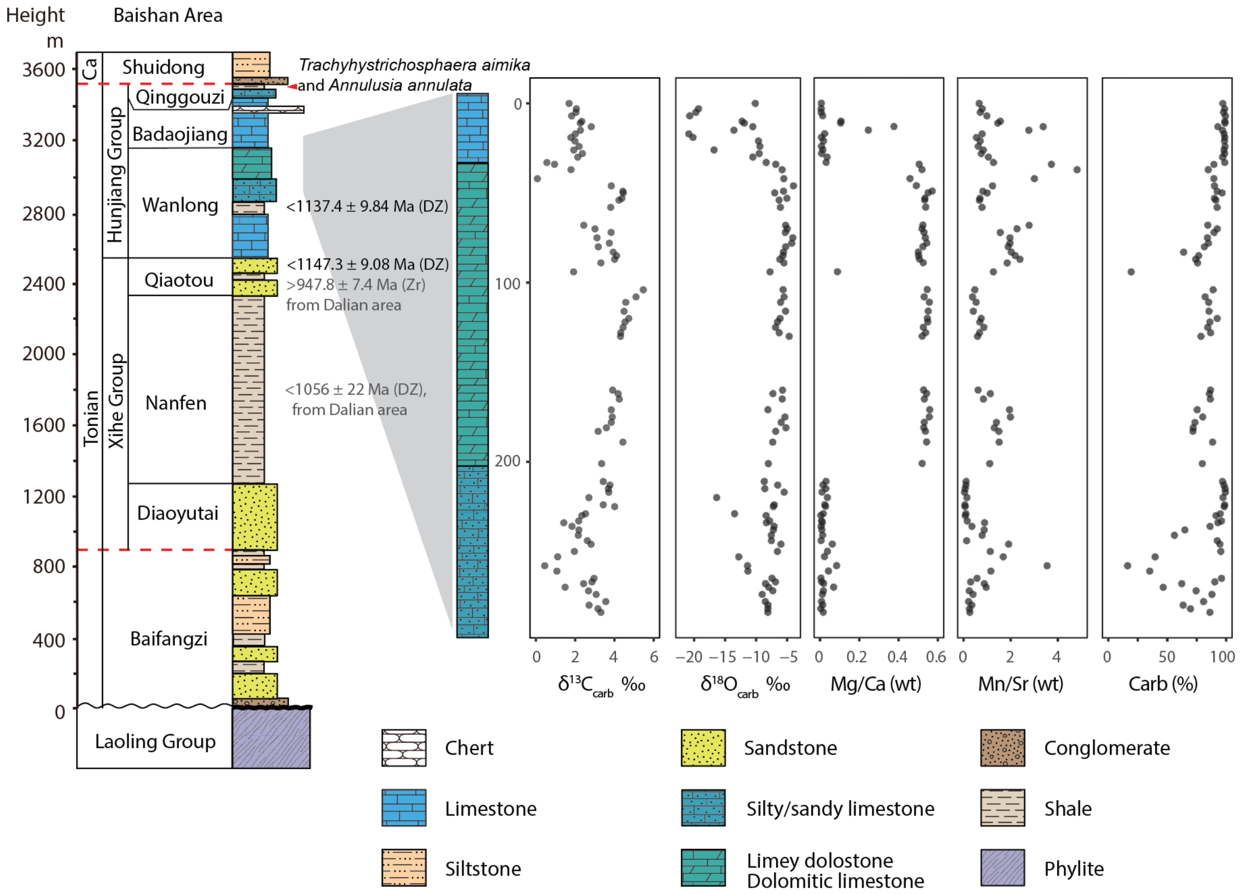


Fig. 2. 2. Lithostratigraphic columns, chronostratigraphic constraints, and geochemical profiles ($\delta^{13}C_{carb}$, $\delta^{18}O_{carb}$, Mg/Ca, Mn/Sr, carbonate content) of the Hunjiang Group in the Baishan region.

Radiometric dates (see text for references) and fossil occurrences are marked at their approximate stratigraphic horizons. Red dashed lines within the lithostratigraphic columns represent unconformities. Ca = Cambrian.

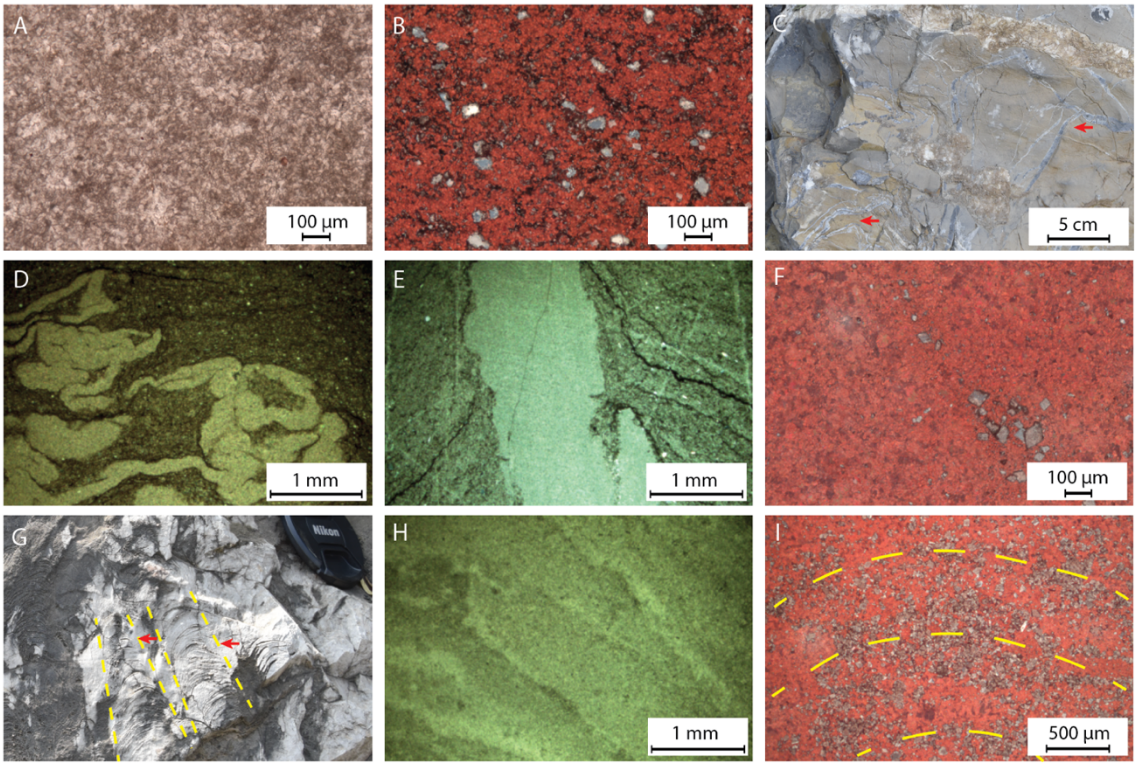


Fig. 2. 3. Field and petrographic photographs of the Wanlong and Badaojiang formations in the Baishan region.

(A) Transmitted photomicrograph of ARS-stained dolomitic limestone of the uppermost Wanlong Formation. (B) Transmitted photomicrograph of ARS-stained silty limestone of the upper Wanlong Formation, under cross-polarized light. (C) Molar tooth structures (red arrows, outlined by yellow dashed lines) from the Wanlong Formation in outcrop view. (D–E) Transmitted photomicrograph of molar tooth structures under cross-polarized light. (F) Transmitted photomicrograph of ARS-stained limestone of the Badaojiang Formation. Unstained crystals are dolomite rhombs. (G–I) Stromatolites in the Badaojiang Formation. (G) Stromatolite (red arrows) on the outcrop. (H) Transmitted photomicrograph at lower magnification to better illustrate laminae in stromatolitic limestone. (I) Transmitted photomicrograph of ARS-stained stromatolitic limestone. Yellow dashed lines denote stromatolitic laminae that are better discernable on hand samples.

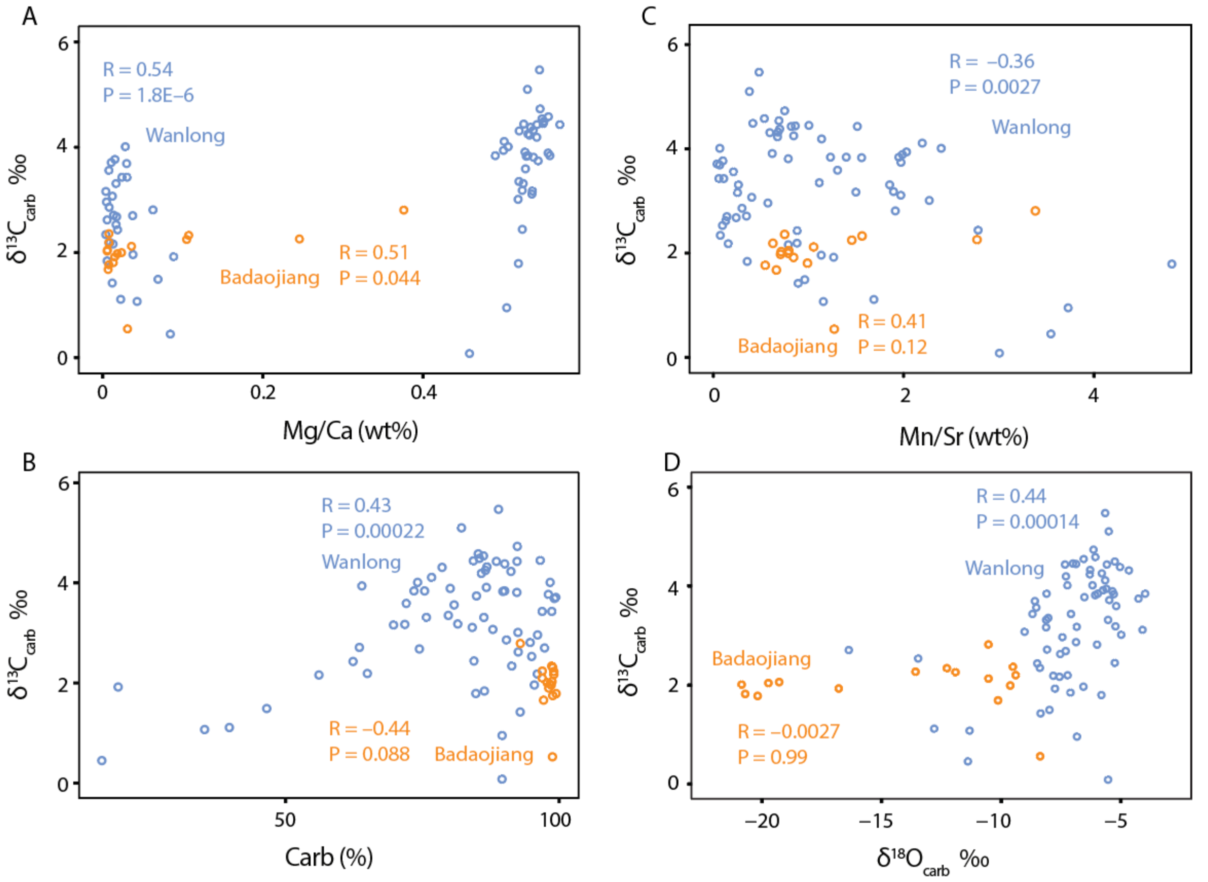


Fig. 2. 4. Cross-plots showing the relationship between $\delta^{13}C_{carb}$ and Mg/Ca, Mn/Sr, carbonate content, and $\delta^{18}O_{carb}$.

R and p values were calculated using Pearson correlation coefficient. No significant correlation is observed in $\delta^{13}C_{carb}$ vs. Mn/Sr, carbonate content, and $\delta^{18}O_{carb}$. A weak correlation is observed between $\delta^{13}C_{carb}$ and Mg/Ca. See text for discussion.

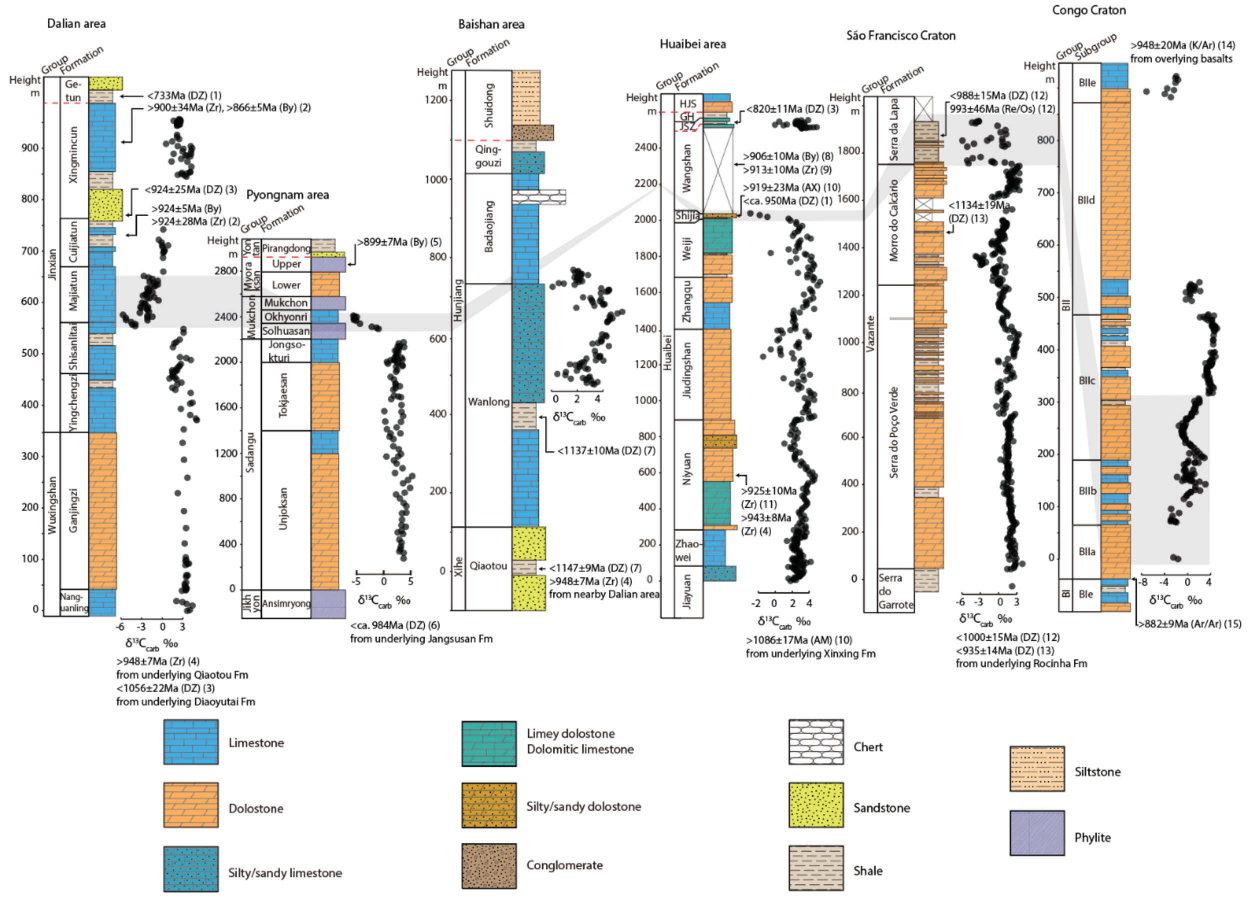


Fig. 2. 5. $\delta^{13}C_{carb}$ profiles of Tonian carbonate sequences from the North China, São Francisco, and Congo cratons.

Modified from Zhang et al. (*22*), Park et al. (*21*), Xiao et al. (*17*), Alvarenga et al. (*43*), and Delpomdor and Préat (*44*). The gray band marks the ca. 920 Ma Majiatun negative $\delta^{13}C_{carb}$ excursion and likely correlative excursions on the São Francisco and Congo cratons. AM–authigenic monazite; AX–authigenic xenotime; By–baddeleyite; DZ–detrital zircon; Zr–magmatic zircon. References for radiometric ages: (1) He et al. (*75*); (2) Zhang et al. (*76*); (3) Yang et al. (*37*); (4) Zhao et al. (*38*); (5) Peng et al. (*58*); (6) Hu et al. (*77*); (7) Wen et al. (*36*); (8) Su et al. (*78*); (9)Zhu et al. (*79*); (10) Zhang et al. (*80*); (11) Fu et al. (*81*); (12) Azmy et al. (*82*); (13) Rodrigues et al. (*83*); (14) Cehen et al. (*61*); (15) Delpomdor et al. (*62*).

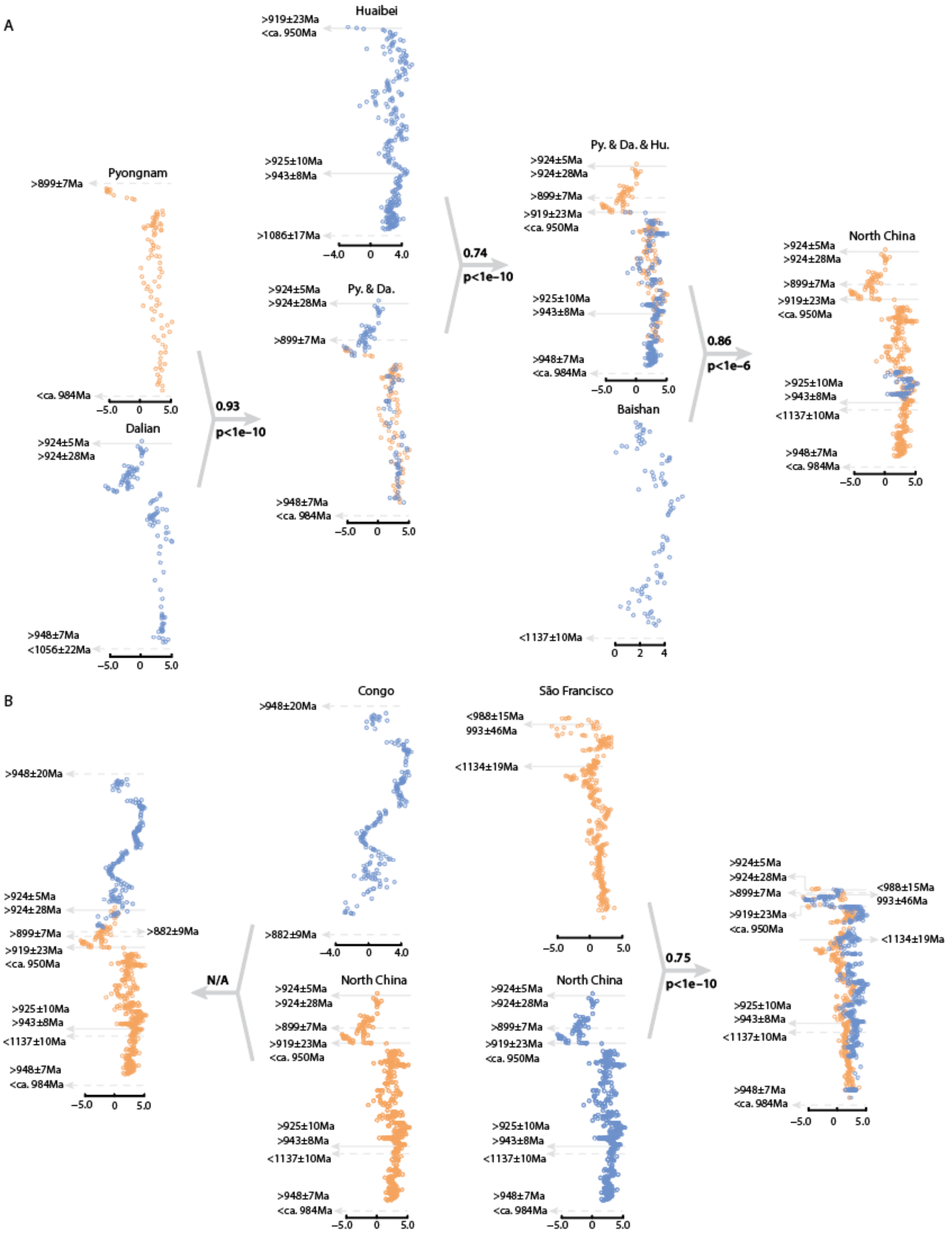


Fig. 2. 6. Chemostratigraphic correlations for sections from NCC (A) and sections from São Francisco and Congo cratons (B) using dynamic time warping (DTW).

When comparing two sequences, the blue-colored sequence is the candidate sequence. The numbers on the arrows are cross-correlation values and p values of the alignments. See Fig. 2. 5. caption for sources of radiometric ages. Solid lines represent ages within the target strata. Dashed lines represent ages above or below the target strata.

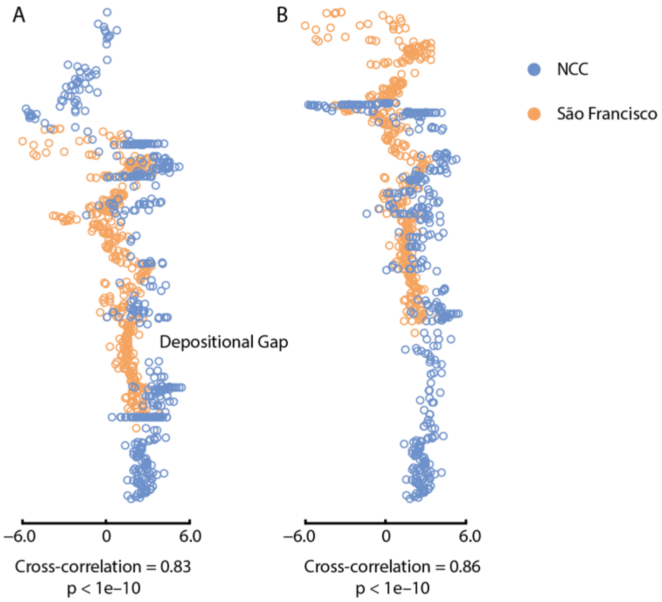


Fig. 2. 7. Alternative alignments between the North China sequence and the São Francisco sequence.

(A) This alignment implies a large depositional gap in the North China sequence. (B) This alignment fails to match the negative excursion in the two sequences.

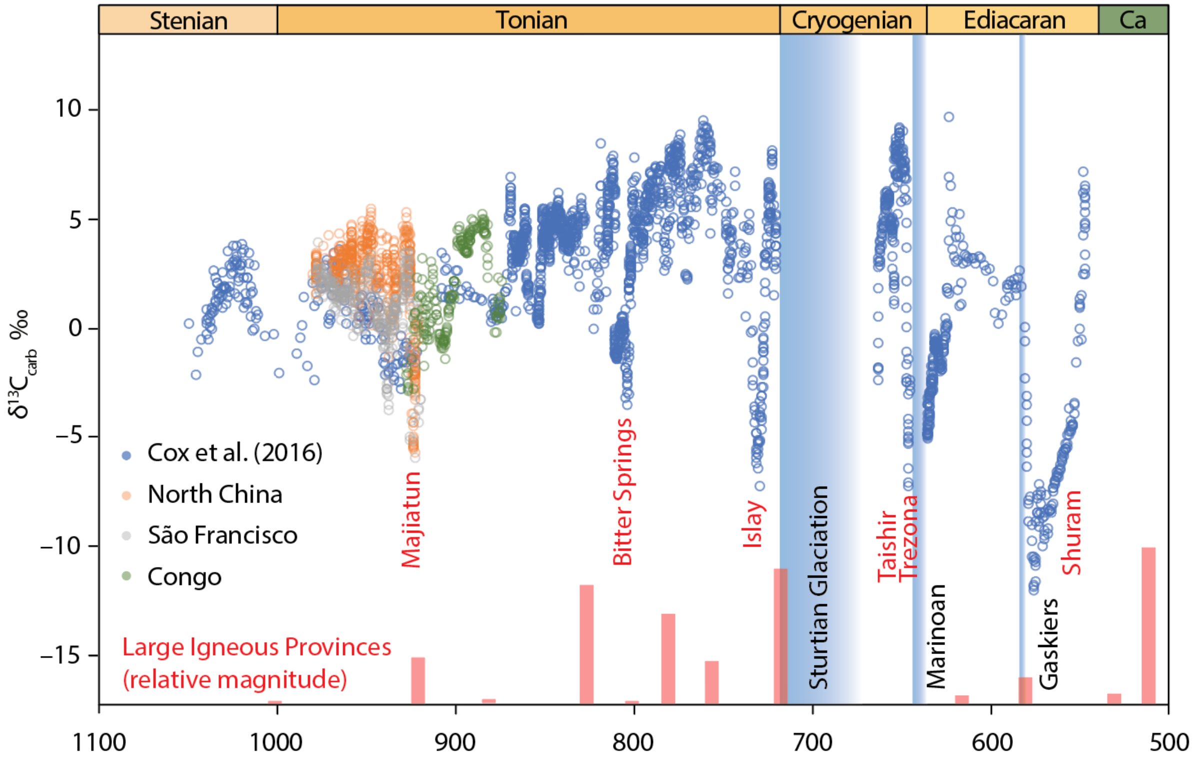


Fig. 2. 8. Compilation of carbonate carbon isotope data for the Neoproterozoic Era. LIP record was marked as vertical bars. Modified after Cox et al. (*84*) and Ernst et al. (*72*). The ca. 920 Ma negative excursion (Majiatun excursion) identified in this study is likely the oldest negative excursion in the Neoproterozoic Era.

CHAPTER 3 Trace element evidence for diverse origins of superheavy pyrite in Neoproterozoic sedimentary strata

Published in Geochimica et Cosmochimica Acta: J. Kang et al., Trace element evidence for diverse origins of superheavy pyrite in Neoproterozoic sedimentary strata. Geochimica et Cosmochimica Acta 364, 1-9 (2024).

Authors

Junyao Kang[1,2], Daniel D. Gregory[3], Benjamin Gill[1,2], Shiqiang Huang[4], Changxin Lai[5], Zhaoshan Chang[4], Huan Cui[6], Ivan Belousov[7], Shuhai Xiao[1,2]

Affiliations

1. Department of Geosciences, Virginia Tech, Blacksburg, VA, USA
2. Global Change Center, Virginia Tech, Blacksburg, VA, USA
3. Department of Earth Sciences, University of Toronto, Toronto, Ontario, Canada
4. Department of Geology and Geological Engineering, Colorado School of Mines, Golden, CO, USA
5. Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
6. Department of Geosciences, Mississippi State University, Mississippi State, MS, USA
7. CODES (ARC Centre of Excellence in Ore Deposits), School of Physical Sciences, University of Tasmania, Hobart, Tasmania, Australia

ABSTRACT

Sedimentary pyrite has long been used as an archive of marine environments in Earth history. To capture reliable paleoenvironmental signals, however, we need to first evaluate pyrite

in sedimentary strata as it can be altered and masked by later diagenetic and/or hydrothermal processes. Here, we trained two supervised machine learning algorithms on a large LA-ICP-MS pyrite trace element database to distinguish pyrite of different origins. The analysis validates that two models built on the co-behavior of 12 trace elements (Co, Ni, Cu, Zn, As, Mo, Ag, Sb, Te, Au, Tl, and Pb) can be used to accurately predict pyrite origins. Further statistical analysis suggests four trace element clusters behaving differently among sedimentary (syngenetic and early diagenetic), synsedimentary hydrothermal (syngenetic hydrothermal), and post-sedimentary hydrothermal (epigenetic hydrothermal) pyrite, which is probably driven by chemical and physical properties of source fluids, interactions between elements, competition among coprecipitating minerals, and pyrite growth rate. Armed with this initial success and aided by new LA-ICP-MS trace element data from 9 samples, we then demonstrated the efficacy of this approach in identifying the origins of pyrite from two Neoproterozoic sedimentary successions in South China. The first set of samples contain isotopically superheavy pyrite (i.e., whose bulk-sample δ^{34}S values greater than those of contemporaneous seawater sulfate and whose origins remain controversial) from the Cryogenian Tiesi'ao and Datangpo formations. The second set of samples contain pyritic rims (associated with fossiliferous chert nodules and thought to be critical in exceptional fossil preservation) from the Ediacaran Doushantuo Formation. For the superheavy pyrite, the models consistently show high confidence levels (mostly >80% probability) in identifying its genesis type, and three out of four samples were given sedimentary origins. For the pyritic nodule rims, the models suggest that early diagenetic pyrite was subsequently altered by hydrothermal fluids and therefore shows mixed signals. The study highlights the importance of pyrite trace elements in deciphering and distinguishing the origins of pyrite in sedimentary strata.

INTRODUCTION

Pyrite is a ubiquitous sulfide mineral in organic-rich marine sediments and sedimentary rocks. Its morphology, abundance, trace element content, and sulfur isotope composition have been widely used to reconstruct oceanic redox conditions and the sulfur cycle (*1-4*), which have played important roles in the evolution of life (*5*). However, the origins of pyrite can be diverse and controversial. Specifically, the presence of hydrothermal pyrite and hydrothermally overprinted sedimentary pyrite can compromise the reliability of the pyrite as a paleoenvironmental proxy, because the chemical and isotopic compositions of such pyrite may reflect in part or in whole the hydrothermal fluids rather than ocean waters.

One prominent case is the origin of isotopically superheavy pyrite — pyrite with bulk-sample $\delta^{34}S$ values greater than those of contemporaneous seawater sulfate — that is widely present in Neoproterozoic and particularly Cryogenian sedimentary rocks (*4, 6-8*). Sedimentary pyrite formation is driven by microbial sulfate reduction where microorganisms use dissolved sulfate as a terminal electron acceptor to oxidize organic matter, producing sulfide (*9*). This sulfide then goes on to react with iron to eventually form pyrite. Sedimentary pyrite typically has bulk-sample $\delta^{34}S$ values lower than seawater sulfate because microbial sulfate reduction prefers the lighter sulfur isotope (*10*). Thus, although rare occurrences of superheavy pyrite have been reported from Phanerozoic rocks and modern environments (*11-13*), its common occurrence in the Neoproterozoic has been puzzling and several hypotheses have been proposed to explain its origins (*6, 14-18*). For example, superheavy pyrite in the cap dolostone atop late Cryogenian (Marinoan) glacial diamictite was interpreted as a result of oceanic stratification where the bottom-water sulfate inherited high $\delta^{34}S$ values derived from syn-glacial isotopic distillation, but the surface-water sulfate had lower $\delta^{34}S$ values derived from post-glacial continental weathering (*14*). In this scenario, carbonate-associated sulfate and pyrite in the cap dolostone were derived from the surface- and bottom-water, respectively, and are thus isotopically decoupled (*14*). Other researchers attributed superheavy pyrite in Ediacaran carbonate rocks to low seawater sulfate concentrations and intense aerobic reoxidation of sulfide (*6*). Still others proposed that superheavy pyrite in the Cryogenian Datangpo Formation overlying Sturtian (early Cryogenian) glacial diamictite was related to the oxidative loss of early ^{34}S-depleted pyrite and the formation of diagenetic pyrite from ^{34}S-enriched sulfate diffusing downward from the deglaciation-facilitated euxinic waters (*15*). On the other hand, the sedimentary origin of Neoproterozoic superheavy pyrite has been questioned by some researchers. New in-situ sulfur isotope measurements suggest that some Cryogenian superheavy pyrite was derived from thermogenic sulfate reduction, hence it is likely late diagenetic or hydrothermal in origin and irrelevant to the oceanic environment (*7*). Therefore, it is important to independently test the origin of Neoproterozoic superheavy pyrite; if it is of hydrothermal origin, then it does not reflect the Neoproterozoic sulfur cycle and marine environments.

Trace element concentrations in pyrite can be used to constrain its origin since trace element enrichment or depletion reflects the formation mechanism and the composition of the source fluid (*3, 19, 20*). Furthermore, these trace elements are relatively stable in unaltered pyrite and can be preserved even in mid-greenschist facies (*21*). The challenge, however, is that the behavior of trace elements in pyrite is complex and no simple criteria can be used to distinguish the different origins of pyrite. In other words, different types of pyrite often have overlapping trace element concentrations. For example, gold can have high concentrations in both early diagenetic pyrite and late metamorphic or hydrothermal pyrite (*21*). To overcome this challenge, multiple trace elements should be compared simultaneously, which can be implemented using rigorous statistical analysis (*20*).

In this study, we used supervised machine learning, in combination with detailed petrographic analysis, to decipher the complex relationships between trace element patterns and the origin of pyrite. First, as pyrite grains from the same rock sample can have different origins, the study of pyrite origins should focus on individual pyrite grains. Hence, in-situ trace element analysis techniques, such as laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS) and secondary ion mass spectrometry (SIMS), were used for geochemical analysis of individual pyrite grains. Second, we used 12 trace elements (Co, Ni, Cu, Zn, As, Mo, Ag, Sb, Te, Au, Tl, and Pb) as input features to develop an optimal classification model to identify the genesis type of each pyrite. These 12 elements were selected based on data availability and previous machine learning applications (*20*); incorporating a greater number of elements would reduce the database size (i.e., fewer samples are qualified) and degrade model performance (*20*). Two widely used supervised classification algorithms, Random Forests (*22*) and XGBoost (*23*), were trained/optimized on a large LA-ICP-MS pyrite trace element database (*20*). The present study followed the standard machine learning practices, similar to those utilized in other investigations in the field of geochemistry (*24-27*).

The labeled training dataset for supervised machine learning is adapted from LA-ICP-MS pyrite trace element data compiled by Gregory et al. (*20*). This compilation includes sedimentary pyrite through geological history and hydrothermal pyrite from different ore deposits. For the purpose of this study, all data are grouped and labeled into three classes: sedimentary, synsedimentary hydrothermal, and post-sedimentary hydrothermal pyrite. First, sedimentary pyrite in this database consists of syngenetic and early diagenetic pyrite. Samples have been rigorously screened using independent petrographic and geochemical criteria to exclude large inclusions, later hydrothermal overprint, recrystallization, or metamorphism (*20*). Second, synsedimentary hydrothermal pyrite, also known as syngenetic hydrothermal pyrite, is defined as pyrite precipitating on or near the seafloor but partially influenced by hydrothermal fluids. Here hydrothermal fluids refer to either fluids from volcanic-dominated systems (volcanogenic massive sulfide, VMS) (*28*) or sedimentary basinal brines (sedimentary exhalative deposits, SEDEX) (*29*). Lastly, post-sedimentary hydrothermal pyrite, also known as epigenetic hydrothermal pyrite, refers to pyrite found in igneous or metamorphosed rocks without contemporaneous seawater inputs. Hydrothermal pyrite in iron oxide copper gold (IOCG), orogenic Au, porphyry Cu, and hydrothermal breccia deposits are included in this category. We note that this is not an exhaustive labeling of pyrite precipitating in different conditions, but it is sufficiently diverse to help us understand the differences among these three classes of pyrite. Meanwhile, both in-sample-locality and out-of-sample-locality test datasets were constructed from the same trace element database. The in-sample-locality test dataset includes data from geologic units/deposits that were part of the training dataset. The out-of-sample-locality test dataset includes pyrite analyses from geologic units/deposits that were not part of the training dataset, which can help us better evaluate the generalizability of the models on unseen instances, including unencountered geologic time intervals and ore deposits.

After training, classification models (Random Forests and XGBoost) were used to determine the origins of two sets of ambiguous Neoproterozoic pyrite samples with existing petrographic and SIMS δ^{34}S data, and to understand their Neoproterozoic marine environments. The first set came from the Cryogenian Tiesi'ao and Datangpo formations in South China. The SIMS $\delta^{34}S_{pyrite}$ data were previously reported in Cui et al. (*7*), who proposed a hydrothermal origin of the superheavy pyrite in these units. The second sample set came from the Ediacaran Doushantuo Formation in South China. Its SIMS $\delta^{34}S_{pyrite}$ data were previously reported in Xiao et al. (*30*). These authors argued that rapid microbial sulfate reduction during early diagenesis facilitated the precipitation of pyrite and silica, forming the pyritic rims and fossiliferous chert nodules. The proposed mechanism may improve the understanding of the taphonomic role of silicification in the preservation of Ediacaran fossils. In this study, we used LA-ICP-MS to analyze these two sample sets for trace element concentrations in pyrite, and then used machine learning models to assess the origins of pyrite grains in the two sets of Neoproterozoic samples.

MATERIALS AND METHODS

Ambiguous pyrite from Neoproterozoic strata

A total of nine best-preserved and most-representative samples were analyzed in this study for pyrite trace element concentrations using LA-ICP-MS (see Supplemental Material for geological settings). Four of these samples (Hy1, Hy31, Hy55a, and Hy59) came from a drill core (4.5 cm in diameter) of the Cryogenian Tiesi'ao and Datangpo formations in South China with SIMS $\delta^{34}S_{pyrite}$ data reported in Cui et al. (*7*). Hy1 is a shale sample from the Datangpo Formation that overlies the Sturtian-age Tiesi'ao glacial diamictite. Two types of pyrite are present in this sample. The first was named pyrite flowers (*7*), with a framboidal pyrite core surrounded by zoned pyrite overgrowth (Fig. 3. 1 and Fig. 3. S1). The second is characterized by a framboidal pyrite core, a thin pyrite rim, and an iron oxide layer in-between (Fig. 3. 1 and Fig. 3. S1). Pyrite flowers show abnormally high $\delta^{34}S$ values (mean: +61.6‰) and were interpreted to be hydrothermal in origin (*7*), whereas pyrite cores within iron oxide show much lower $\delta^{34}S$ values (mean: +22.2‰) and were thought to have formed through syngenetic or early diagenetic processes (*7*). Hy31 is also a shale sample from the Datangpo Formation. There is no framboidal pyrite in this sample. Instead, most pyrite grains are subhedral and some grains can reach millimeters in size, with micrometer-sized sub-grains cemented by later-stage pyrite (Fig. 3. 1 and Fig. 3. S2). High $\delta^{34}S$ values of pyrite grains (mean: +66.3‰) and cement (mean: +62.0‰) were suggested to reflect hydrothermal processes (*7*). Hy55a is a Mn-rich carbonate sample from the basal Datangpo Formation. In this sample, nodule- and sausage-shaped pyrite replaced (and thus postdates) pre-existing rhodochrosite layers and lath-shaped illite crystals, and inherited the texture of the hosting rhodochrosite (Fig. 3. S3). The pyrite is composed of individual framboids and lacy overgrowth (Fig. 3. 1 and Fig. 3. S3G). On the basis of petrographic observations (e.g., pyrite postdating rhodochrosite and illite) and high $\delta^{34}S$ values of both the framboids (mean: +56.9‰) and the overgrowth (mean: +57.8‰), Cui et al. (*7*) argued for a hydrothermal origin of pyrite in Hy55a. Hy59 came from the uppermost Tiesi'ao Formation, which consists of glacial diamictite deposited during the Sturtian snowball Earth glaciation. Pyrite framboids and nodules with low $\delta^{34}S$ values (mean, framboids: +16.4‰; nodules: +26.5‰) are common in this specimen (Fig. 3. 1 and Fig. 3. S4). Unlike pyrite in Hy31 or Hy55a, these framboids and nodules were thought to have formed through syngenetic and early diagenetic processes (*7*).

The remaining five samples, JLW15.1, JLW23.5, JLW34.8, JLW76, and SX88.15, are outcrop samples of argillaceous dolostones and calcareous mudstones from members II and III of the Ediacaran Doushantuo Formation exposed at the Jiulongwan (JLW) and Sixi (SX) sections in the Yangtze Gorges area of South China. These samples were selected for comparison with the superheavy pyrite described above. They had previously been described in Xiao et al. (*30*). Pyrite in these samples occurs as pyritic rims around chert nodules or as disseminated pyrite in the carbonate matrix (Fig. 3. 1 and Fig. 3. S5–9). Chert nodules are typically nucleated on a microbial mat fragment, which is surrounded by a silica cortex and then a pyritic rim. Rim pyrite is immersed in a groundmass of silica and occurs as subhedral pyrite grains. A centrifugal decrease in pyrite crystal size is observed in the pyritic nodule rim. The disseminated pyrite is subhedral to anhedral in shape and randomly distributed in the matrix. Compared with the superheavy pyrite in the Cryogenian Tiesi'ao and Datangpo Formations[7], both rim pyrite (mean, JLW15.1: +24.6‰; JLW23.5: +18.1‰; JLW34.8: +26.6‰; SX88.15: +25.9‰) and disseminated pyrite (mean, JLW23.5: +30.6‰; JLW76: +29.1‰) in the Doushantuo samples have much lower $\delta^{34}S$ values. Based on petrographic and isotopic evidence, Xiao et al. (*30*) suggested that both the rim pyrite and chert nodules formed during early diagenesis. They argued that bacterial sulfate reduction of an organic-rich nucleus not only provided a H_2S source for pyrite formation but also decreased porewater pH to facilitate silicification (*30*).

LA-ICP-MS Trace Element Measurement

LA-ICP-MS pyrite trace element analyses were performed at the LA-ICP-MS Laboratory of the Department of Geology and Geological Engineering, Colorado School of Mines of the USA, using a RESOlution SE 193 nm ArF excimer laser ablation system equipped with an S155 sample chamber, and an Agilent 8900 ICP-MS. Laser spot analysis was carried out with an on-sample fluence of 2.7 J/cm^2, a repetition rate of 5 pulses/sec (5 Hz), and a laser spot size of 14 to 30 µm chosen according to pyrite grain sizes. The ablated material was carried by He gas, then mixed with Ar gas in a funnel sitting right above the sample cell, and subsequently introduced to the ICP-MS. The analyzed isotopes included ^{57}Fe, ^{59}Co, ^{60}Ni, ^{65}Cu, ^{66}Zn, ^{75}As, ^{95}Mo, ^{107}Ag, ^{121}Sb, ^{125}Te, ^{197}Au, ^{205}Tl, and ^{208}Pb. A dwell time of 10 milliseconds was used for all the above isotopes. In addition, ^{29}Si, ^{31}P, ^{43}Ca, ^{49}Ti were also analyzed with a dwell time of 5 milliseconds to monitor the potential intersection of mineral inclusions. Each laser spot was pre-ablated with one laser shot to eliminate possible surface contamination. That was followed by 15 seconds of washing time for the aerosol to pass through the ICP-MS. A gas blank was collected for 20 seconds, followed by 30 seconds of sample signal collection. The primary external standard was STDGL3 glass made at CODES, University of Tasmania (*31*). The internal standard was Fe assuming a stoichiometric concentration of 46.55 weight percent. Data reduction was performed using the software Iolite v. 4.0 (*32*). When selecting an interval on the time-resolved signal (counts per second) plot of each laser spot for data reduction, mineral inclusions were avoided to minimize contaminations. For the same purpose, the data were further screened to ensure that no analyses had higher than 1% Zn, 2% As, 1% Cu, 1% Ni, and 2% Co (*20*). When an element was below the detection limit, half of the detection limit was taken as the concentration.

SEM and EDS observation

SEM and EDS observations were performed at the Virginia Tech Institute of Critical Technology and Applied Science, on a JEOL IT-500HR SEM via a backscattered electron detector at an accelerating voltage of 15 or 20 keV and a working distance of ~10.0 mm. Elemental maps were acquired using an AZtecLive Automated Microanalysis System with an UltimMax100 silicon drift detector at an accelerating voltage of 20 keV.

Random Forests and XGBoost

The LA-ICP-MS pyrite trace element database compiled by Gregory et al. (*20*) was used in this study. To train and evaluate models, the labeled trace element dataset was divided into a training dataset, an in-sample-locality test dataset, and an out-of-sample-locality test dataset. A total of 240 pyrite analyses from each of the three labels (sedimentary, synsedimentary hydrothermal, and post-sedimentary hydrothermal) were randomly selected to form a balanced training dataset. For sedimentary pyrite, 16 different geological units each contributed 15 analyses. For synsedimentary hydrothermal pyrite, 120 analyses were from SEDEX deposits and the remaining 120 analyses were from VMS deposits. For post-sedimentary hydrothermal pyrite, hydrothermal breccia, IOCG, orogenic Au, and porphyry Cu deposits each contributed 60 analyses. The rest of analyses within these geologic units/deposits were grouped together to form an in-sample-locality test dataset. Meanwhile, an out-of-sample-locality test dataset was constructed using pyrite analyses from geologic units/deposits that were not part of the training dataset.

Random Forests and XGBoost, two powerful decision-tree ensemble methods (see Appendix A for more details), were performed in Python using the sklearn (*33*) and XGBoost (*23*) packages, respectively. All data were standardized by subtracting the mean and scaling to unit variance prior to inputting into the models. A five-fold cross-validation method (GridSearchCV in sklearn), which randomly and evenly splits the training dataset into five parts for training and validation, was used to determine optimal hyperparameters. For Random Forests, the feature importance is decided by the mean decrease in impurity when a feature is used to build a split. For XGBoost, the feature importance is decided by the gain, which is the improvement in accuracy brought by a feature. Another widely-used supervised machine learning method, support vector machine (SVM), was also used to analyze the data, although its performance was not as good as Random Forests and XGBoost.

Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA)

Two unsupervised machine learning algorithms, HCA and PCA, were used to investigate trace element patterns of sedimentary and hydrothermal pyrite in the database, similar to analyses of Emmings et al. (*34*) for origins of sedimentary pyrite through geologic time and Hu et al. (*25*) for origins of high titanium magnetite. HCA is a method of cluster analysis that aims to group similar observations together into several clusters. In order to identify elements sharing similar behaviors, data of each trace element were treated as one observation. To be specific, HCA in this study focused on partitioning 12 observations (elements) into different clusters. PCA is a dimensionality reduction method that aims to increase data interpretability while preserving the maximum amount of information. In PCA, each pyrite analysis was treated as one observation with 12 features (trace elements). The high-dimensional data were projected onto several principal components capturing most of the variance of trace element concentrations and likely recording trace element patterns, which can be used to validate the HCA results. Since LA-ICP-MS trace element data vary over several orders of magnitude and show high skewness, all data were power transformed using the Yeo-Johnson method (*35*) before HCA and PCA. Analyses were performed in Python using the sklearn (*33*) package.

RESULTS

New LA-ICP-MS data for Hy, JLW, and SX samples

A total of 177 in-situ pyrite trace element analyses of 9 samples (Hy, JLW, and SX specimens) passed the screening (see Section 2.2.). For each analyzed element, concentrations vary up to 4–5 orders of magnitude. For samples Hy31, JLW15.1, JLW23.5, JLW76, and SX88.15, most analyses of Zn and Te were below detection limits. For JLW34.8, only Te analyses were below detection limits. For Hy31, in addition to Zn and Te, the majority of Mo, Ag, Sb, Au, and

Tl analyses were also close to or below detection limits. Analyses below detection limits were replaced by half the detection limit as machine learning inputs (*20*).

The LA-ICP-MS database for machine learning analysis

A total of 3632 analyses from 71 different deposits and sedimentary units were used for supervised machine learning model training and testing (Table S1) (*20*). Of these, 1123 analyses of sedimentary pyrite are from 28 sedimentary units covering the geological interval of 2572−125 Ma and therefore recording the variability of trace element concentrations over geologic time, 1345 analyses of synsedimentary hydrothermal pyrite are from 18 ore deposits, and 1164 analyses of post-sedimentary hydrothermal pyrite are from 25 ore deposits. A total of 720 analyses (240 analyses from each label), from 16 sedimentary units ranging from 2572 to 125 Ma and 33 ore deposits covering 6 different ore types, were randomly selected as the training dataset. The remaining 2487 analyses from these deposits/units were used for initial in-sample-locality testing. Furthermore, 425 analyses from 12 sedimentary units ranging from 2170 to 180 Ma in age and 10 ore deposits covering 3 different ore types, which were not seen in the training and initial testing datasets, were used for out-of-sample-locality testing. By creating such diverse training and testing datasets, we believe that a representative variation in pyrite forming environments has been taken into account. Again, for each analyzed element, concentrations vary up to 5−6 orders of magnitude. When analyses were below detection limits, either half the detection limit or values from the literature were used as an estimation. Considering a significant number of Te and Au analyses are affected by this estimation, we also examined machine learning models on a refined dataset that excluded the Te and Au data.

Machine learning model performance and predictions on Hy, JLW, and SX samples

Random Forests and XGBoost models both yielded robust performance on in-sample-locality and out-of-sample-locality test datasets (Table 3. 1, Fig. 3. 2A−D), which is significantly better compared with the SVM model (Fig. 3. S10). Average precision, recall, and F1-score are all >90%. For sedimentary pyrite, the precision even reaches 96% (Random Forests) and 98% (XGBoost) for the out-of-sample-locality test dataset. Mo is the most important feature in both models (Fig. 3. 2E−F). Meanwhile, Ni, Cu, Zn, Ag, Sb, Te, Au, Tl, and Pb all have significant contributions. Additionally, models trained by the dataset without Au and Te, whose concentrations are below the detection limit in many of the samples in the dataset, show lower but still promising precision, recall, and F1-score (Table 3. S2, Fig. 3. S11).

Random Forests and XGboost models, trained with and without Au and Te, show consistent predictions on pyrite analyses of Hy, JLW, and SX samples (Fig. 3. 3 and Fig. 3. S12). The predictions are provided in the ternary plots as the probability of being identified as one of the three types of pyrite (Fig. 3. 3). Pyrite grains in Hy1 (9 grains analyzed), Hy55 (3 nodules analyzed), and Hy59 (2 nodules and 5 aggregates of framboids analyzed) were identified as showing sedimentary signals (mostly >80% probability for both models) whereas pyrite grains in Hy31 were classified as post-sedimentary hydrothermal pyrite (mostly >80% probability for both models). For JLW and SX samples, the majority of the pyritic rims in JLW15.1, JLW23.5, and SX88.15 and disseminated pyrite in JLW23.5 (4 grains analyzed) and JLW76 (20 grains analyzed) were identified as post-sedimentary hydrothermal pyrite with a few exceptions (pyrites identified as sedimentary or synsedimentary hydrothermal pyrite). A significant portion of these analyses (53% for the Random Forest model, 13% for the XGBoost model) falls into the gray triangle in ternary plots because they were given a <50% probability of being identified as a leading label.

Conversely, all analyses on JLW34.8 show synsedimentary hydrothermal origins (>50% probability for the Random Forest model, >80% probability for the XGBoost model).

DISCUSSION

Trace element patterns of sedimentary and hydrothermal pyrite

Despite the considerable variability in trace element signals within each pyrite type (e.g., variation in trace element compositions in sedimentary pyrite due to changes in ocean chemistry; (*34*), both models exhibited robust predictive capabilities for determining the provenance of pyrite. The consistent precision, recall, and F1-score values obtained from both the in-sample-locality and out-of-sample-locality tests indicate a high level of generalizability in the models. Based on feature importance, ten out of twelve elements show significant contributions to both models, suggesting a complex relationship between trace element concentrations and the origin of pyrite. In other words, there is no single element that can be used to differentiate sedimentary and hydrothermal pyrite, which further supports the necessity of using machine learning modeling of a suite of elements to identify the origins of pyrite.

Decision Tree ensembles (i.e., Random Forest and XGBoost), however, cannot offer further insights into trace element behavior. To investigate any existing trace element patterns of sedimentary and hydrothermal pyrite, we applied hierarchical cluster analysis (HCA) on the power transformed (*35*) trace element database. HCA reveals four major clusters of elements (Fig. 3. 4A and Fig. 3. S13). Firstly, Te and Au are enriched in post-sedimentary and VMS-type synsedimentary hydrothermal pyrite. The enrichment is possibly related to changes in the chemical and physical properties of fluids, zone refinement (i.e., enrichment through dissolution and reprecipitation), and direct input from magmatic and metamorphic hydrothermal fluids (*21, 36*).

A coupled Au-Te transfer in fluids and enhancement of Au partitioning with increasing Tellurium concentrations (*37*) may also play a role in the enrichment. We note, however, that average Au and Te concentrations from the literature (*20*) were used for most SEDEX pyrite. These data were assumed to be reasonable, as Au and Te are usually below the detection limit in SEDEX deposits (*20*). Secondly, Co and Ni are relatively depleted in the majority of VMS-type synsedimentary hydrothermal pyrite. This is probably due to coprecipitating minerals outcompeting pyrite for metal incorporation in their respective structure (*38*). For example, Co and Ni, with similar ionic radii and charges to Zn, can substitute into sphalerite forming from the hydrothermal fluid (*38*). Other researchers have also proposed that the depletion in Co and Ni can be attributed to the temperature of pyrite formation. Pyrite formed at high temperatures tends to concentrate Co and Ni. As these local fluids undergo cooling and reach the low-temperature seafloor, both the fluids themselves and the precipitated pyrite are depleted in these elements (*39*). Meanwhile, the solubility of Co and Ni can be affected by the formation of Cl^- complexes in hydrothermal fluids (*39*). Thirdly, As distribution is relatively ubiquitous. No significant difference in As concentration is observed among the three types of pyrite. Lastly, Mo, Tl, Ag, Sb, Pb, Cu, and Zn are depleted in post-sedimentary hydrothermal pyrite. High concentrations of these elements in pyrite forming near the seafloor may be related to the decomposition of particulate and dissolved organic matter and the reduction of Mn-Fe (oxyhydr)oxide (*34*). On the other hand, for post-sedimentary hydrothermal pyrite, higher temperature and slower pyrite growth rate enable trace elements to be partitioned into other distinct sulfide phases (e.g., sphalerite and galena) rather than existing as solid solutions or micro-inclusions in pyrite (*21*). Among the 12 elements investigated, Mo is the only element showing a unidirectional change from sedimentary to post-sedimentary hydrothermal pyrite — Mo concentration decreases from left to right in Fig. 3. 4A. This pattern can be attributed to the higher concentrations of Mo in seawater compared to most hydrothermal fluids (*40, 41*). As we transition from sedimentary to post-sedimentary hydrothermal pyrite, the contribution of seawater to pyrite formation progressively diminishes, accompanied with an escalating prominence of hydrothermal fluids in the mineralization process. This unidirectional change likely accounts for the selection of Mo as the most important feature by both machine learning models.

HCA results are further supported by principal component analysis (PCA) (Fig. 3. 4B). The first two principal components (PC1 and PC2) capture 59% of the total variance (Fig. 3. S14) and accurately record the concentration changes of the four clusters of elements. PC1 records changes in the third and fourth clusters of elements (As, Mo, Tl, Ag, Sb, Pb, Cu, Zn), whereas PC2 records changes in the first and second clusters (Au, Te, Co, Ni) plus Mo and Tl from the fourth cluster. Along the PC1 direction, post-sedimentary hydrothermal pyrite is separated from the sedimentary and synsedimentary hydrothermal pyrite, which is consistent with the insight from HCA that the fourth cluster of elements is enriched in pyrite forming near the seafloor. Additionally, the separation of post-sedimentary hydrothermal and VMS-type synsedimentary hydrothermal pyrite from sedimentary and SEDEX-type synsedimentary hydrothermal pyrite along the PC2 direction confirms the observation that the first cluster of elements is enriched in post-sedimentary hydrothermal and VMS-type synsedimentary hydrothermal pyrite. The differentiation between SEDEX-type pyrite and VMS-type pyrite along the PC2 is further enhanced due to the fact that SEDEX-type pyrite is enriched in Co and Ni while VMS-type pyrite is depleted in these two elements. Meanwhile, due to the observed overlap of pyrite analyses from different origins within the PC1 and PC2 biplot, the viability of employing the first two principal components for discriminating different pyrite is compromised, which further reinforces the complex nature of this target.

Origins of Neoproterozoic pyrite

Our machine-learning models partially confirmed the findings of Cui et al. (*7*). Specifically, pyrite grains in Hy31 were identified as post-sedimentary hydrothermal pyrite, and grains in Hy59 were identified as sedimentary pyrite. A post-sedimentary hydrothermal origin of pyrite in Hy31 is consistent with low trace element concentrations of Mo, Ag, Tl, Sb, and Zn. However, both models posit a sedimentary origin for pyrite grains in Hy1 and Hy55a, whereas Cui et al. (*7*) argued for a hydrothermal origin. One of their major arguments was based on the paragenesis of Hy55a, where pyrite replaced and thus postdate rhodochrosite and illite. Key to this argument is the time when rhodochrosite and illite formed. Rhodochrosite can form during early diagenesis as porewaters become anoxic and manganese (oxyhydr)oxides are reduced (*42*). For example, calcian rhodochrosite has been reported from organic-rich marine muds in the Baltic Sea (*43*). Recent studies even suggested that manganese-rich carbonate can form in the water column of redox-stratified lakes (*44*). Furthermore, the preservation of rhodochrosite laminae that warp around pyrite nodules in Hy55a (Fig. 3. S3A−F) suggests a sedimentary or early diagenetic origin. As for the formation of illite, laboratory experiments have demonstrated microbially-induced illitization at room temperature and atmospheric pressure within two weeks (*45*). Field observation also recorded microbially induced smectite-to-illite transformation during diagenesis (*46, 47*). Thus, our model predictions are not necessarily inconsistent with petrographic observations, and the formation of superheavy pyrite could be linked to the contemporaneous oceanic environment. If so, regional and global models of the oceanic sulfur cycle should be considered to seek an explanation for unusual sulfur isotope signals from the Neoproterozoic superheavy pyrite.

For the JLW and SX samples from the Ediacaran Doushantuo Formation, our models reveal a more complex diagenetic history than originally thought (*30*). Instead of a sedimentary origin, rim pyrite and disseminated pyrite in JLW15.1, JLW23.5, JLW76, and SX88.15 were classified as post-sedimentary hydrothermal pyrite. We note, however, that a large portion of analyses shows ambiguous prediction results (i.e., they are in the gray triangles in ternary diagrams in Fig. 3. 3C–D). They were given a <50% probability of being identified to the leading label (Fig. 3. 3 and Fig. 3. S12), which in this case is the post-sedimentary hydrothermal pyrite. We propose two explanations for the low probability. First, since our hydrothermal training data come from a limited number of ore deposits, it is impossible for us to cover every pyrite-forming hydrothermal environment. Accordingly, these subhedral pyrite grains may have formed in a hydrothermal environment not represented in the training dataset. Second, as some Doushantuo pyrite analyses have been identified to sedimentary or synsedimentary hydrothermal origins with some confidence, it is possible that some of these ambiguous pyrite grains were originally sedimentary pyrite but later altered by hydrothermal fluids and therefore show mixed signals. In this scenario, JLW34.8 may have been altered by synsedimentary hydrothermal fluids and the four other Doushantuo samples may have been further altered by post-sedimentary processes. The possibility of hydrothermal alteration is also consistent with petrographic observations: no pyrite framboids have been observed in any of the five Doushantuo samples analyzed in this study. Considering that not all analyses identified as post-sedimentary hydrothermal pyrite (i.e., the label with the highest possibility predicted by the model) fall into the ambiguous zone (probability <50%) and more than half (89% for XGBoost, 53% for Random Forests) of them were given a high probability (>50%), we argue the hydrothermal alteration hypothesis may be more plausible. Given the petrographic evidence for an early diagenetic origin of the pyritic rim around Doushantuo chert nodules, the silicification model proposed by Xiao et al. (*30*) may still be valid, although the trace element data do show that the pyrite may have been hydrothermally altered at a later time.

CONCLUSIONS

To sum up, this is the first study applying supervised machine learning to understand the origins of Neoproterozoic pyrite. The two models tested here, Random Forests and XGBoost, show strong performance in distinguishing sedimentary and hydrothermal pyrite using in-situ trace element data. Our analysis shows that superheavy pyrite in the Cryogenian Datangpo Formation can be sedimentary or hydrothermal in origin. Future studies are needed to investigate what the ^{34}S-enriched pyrite can tell us about the sulfur biogeochemical cycle. Our analysis of pyritic rims surrounding fossiliferous chert nodules from the Ediacaran Doushantuo Formation revealed a more complex history than previously thought. Early diagenetic pyrite precipitated around fossiliferous chert nodules and may have played a key role in chert nodule formation and fossil preservation (*30*), but was subsequently altered and overprinted by hydrothermal fluids. This study highlights machine learning models as a powerful tool to study the origins of pyrite and to effectively screen pyrite samples for paleoenvironmental research.

REFERENCES

1. R. T. Wilkin, H. L. Barnes, S. L. Brantley, The size distribution of framboidal pyrite in modern sediments: An indicator of redox conditions. *Geochimica et Cosmochimica Acta* **60**, 3897-3912 (1996).
2. T. W. Lyons, J. P. Werne, D. J. Hollander, R. Murray, Contrasting sulfur geochemistry and Fe/Al and Mo/Al ratios across the last oxic-to-anoxic transition in the Cariaco Basin, Venezuela. *Chemical Geology* **195**, 131-157 (2003).

3. D. D. Gregory, The Pyrite Trace Element Paleo-Ocean Chemistry Proxy. *Elements in Geochemical Tracers in Earth System Science*, (2020).
4. D. A. Fike, A. S. Bradley, C. V. Rose, Rethinking the ancient sulfur cycle. *Annual Review of Earth and Planetary Sciences* **43**, 593-622 (2015).
5. A. D. Anbar, A. H. Knoll, Proterozoic ocean chemistry and evolution: A bioinorganic bridge? *Science* **297**, 1137-1142 (2002).
6. J. B. Ries, D. A. Fike, L. M. Pratt, T. W. Lyons, J. P. Grotzinger, Superheavy pyrite ($\delta^{34}S_{pyr} > \delta^{34}S_{CAS}$) in the terminal Proterozoic Nama Group, southern Namibia: A consequence of low seawater sulfate at the dawn of animal life. *Geology* **37**, 743–746 (2009).
7. H. Cui *et al.*, Questioning the biogenicity of Neoproterozoic superheavy pyrite by SIMS. *American Mineralogist: Journal of Earth and Planetary Materials* **103**, 1362-1400 (2018).
8. X. Lang *et al.*, Cracking the superheavy pyrite enigma: possible roles of volatile organosulfur compound emission. *National Science Review* **8**, (2021).
9. R. A. Berner, Sedimentary pyrite formation. *American journal of science* **268**, 1-23 (1970).
10. K. S. Habicht, D. E. Canfield, Sulfur isotope fractionation during bacterial sulfate reduction in organic-rich sediments. *Geochimica et Cosmochimica Acta* **61**, 5351-5361 (1997).
11. V. Ferrini, M. Fayek, C. D. Vito, S. Mignardi, J. Pignatti, Extreme sulphur isotope fractionation in the deep Cretaceous biosphere. *Journal of the Geological Society* **167**, 1009-1018 (2010).
12. S.-Y. Hu *et al.*, Variability of sulfur isotopes and trace metals in pyrites from the upper oceanic crust of the South China Sea basin, implications for sulfur and trace metal cycling in subsurface. *Chemical Geology* **606**, 120982 (2022).
13. P. Cadeau *et al.*, The Dziani Dzaha Lake: A long-awaited modern analogue for superheavy pyrites. *Geobiology* **20**, 444-461 (2022).
14. B. Shen *et al.*, Stratification and mixing of a post-glacial Neoproterozoic ocean: Evidence from carbon and sulfur isotopes in a cap dolostone from northwest China. *Earth and Planetary Science Letters* **265**, 209-228 (2008).
15. C. Cai *et al.*, Enigmatic super-heavy pyrite formation: Novel mechanistic insights from the aftermath of the Sturtian Snowball Earth. *Geochimica et Cosmochimica Acta* **334**, 65-82 (2022).
16. T.-B. Liu, J. B. Maynard, J. Alten, in *Evolution of Early Earth's Atmosphere, Hydrosphere, and Biosphere—Constraints from Ore Deposits: Geological Society of America Memoir 198,* S. E. Kesler, H. Ohmoto, Eds. (Geological Society of America, Boulder, Colorado, 2006), pp. 205-222.
17. Y. Peng *et al.*, Widespread contamination of carbonate-associated sulfate by present-day secondary atmospheric sulfate: Evidence from triple oxygen isotopes. *Geology* **42**, 815-818 (2014).
18. R. Tostevin *et al.*, Constraints on the late Ediacaran sulfur cycle from carbonate associated sulfate. *Precambrian Research* **290**, 113-125 (2017).
19. D. D. Gregory *et al.*, Trace element content of sedimentary pyrite in black shales. *Economic Geology* **110**, 1389-1410 (2015).

20. D. D. Gregory *et al.*, Distinguishing ore deposit type and barren sedimentary pyrite using laser ablation-inductively coupled plasma-mass spectrometry trace element data and statistical analysis of large data sets. *Economic Geology* **114**, 771-786 (2019).
21. R. R. Large *et al.*, Gold and trace element zonation in pyrite using a laser imaging technique: Implications for the timing of gold in orogenic and Carlin-style sediment-hosted deposits. *Economic Geology* **104**, 635-668 (2009).
22. L. Breiman, Random forests. *Machine learning* **45**, 5-32 (2001).
23. T. Chen, C. Guestrin, paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016.
24. X.-M. Li *et al.*, Discrimination of Pb-Zn deposit types using sphalerite geochemistry: New insights from machine learning algorithm. *Geoscience Frontiers* **14**, 101580 (2023).
25. B. Hu *et al.*, The origin and discrimination of High-Ti magnetite in magmatic-hydrothermal systems: Insight from machine learning analysis. *Economic Geology* **117**, 1613-1627 (2022).
26. G. Chen *et al.*, Reconstructing Earth's atmospheric oxygenation history using machine learning. *Nature Communications* **13**, 5862 (2022).
27. I. Mukherjee, R. Corkrey, R. Large, L. Danyushevsky, Abiotic and biotic constraints on Earth's ancient colonisers in the Proterozoic. *Precambrian Research* **393**, 107089 (2023).
28. M. D. Hannington, in *Treatise on Geochemistry (Second Edition),* H. D. Holland, K. K. Turekian, Eds. (Elsevier, 2014), pp. 463-488.
29. P. Emsbo, R. R. Seal, G. N. Breit, S. F. Diehl, A. K. Shah, "Sedimentary exhalative (sedex) zinc-lead-silver deposit model," (US Geological Survey, 2016).
30. S. Xiao, J. Schiffbauer, K. McFadden, J. Hunter, Petrographic and SIMS pyrite sulfur isotope analyses of Ediacaran chert nodules: Implications for microbial processes in pyrite rim formation, silicification, and exceptional fossil preservation. *Earth and Planetary Science Letters* **297**, 481-495 (2010).
31. I. Belousov *et al.*, STDGL3, a Reference Material for Analysis of Sulfide Minerals by Laser Ablation ICP-MS: An Assessment of Matrix Effects and the Impact of Laser Wavelengths and Pulse Widths. *Geostandards and Geoanalytical Research* **47**, 493-508 (2023).
32. C. Paton, J. Hellstrom, B. Paul, J. Woodhead, J. Hergt, Iolite: Freeware for the visualisation and processing of mass spectrometric data. *Journal of Analytical Atomic Spectrometry* **26**, 2508-2518 (2011).
33. F. Pedregosa *et al.*, Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825-2830 (2011).
34. J. F. Emmings *et al.*, Pyrite mega-analysis reveals modes of anoxia through geological time. *Science advances* **8**, eabj5687 (2022).
35. I. K. Yeo, R. A. Johnson, A new family of power transformations to improve normality or symmetry. *Biometrika* **87**, 954-959 (2000).
36. S. Fuchs, M. D. Hannington, S. Petersen, Divining gold in seafloor polymetallic massive sulfide systems. *Mineralium Deposita* **54**, 789-820 (2019).
37. I. Belousov *et al.*, Pyrite compositions from VHMS and orogenic Au deposits in the Yilgarn Craton, Western Australia: Implications for gold and copper exploration. *Ore Geology Reviews* **79**, 474-499 (2016).
38. D. Genna, D. Gaboury, Deciphering the hydrothermal evolution of a VMS system by LA-ICP-MS using trace elements in pyrite: an example from the Bracemac-McLeod

deposits, Abitibi, Canada, and implications for exploration. *Economic Geology* **110**, 2087-2108 (2015).

39. M. Keith, K. M. Haase, R. Klemd, S. Krumm, H. Strauss, Systematic variations of trace element and sulfur isotope compositions in pyrite with stratigraphic depth in the Skouriotissa volcanic-hosted massive sulfide deposit, Troodos ophiolite, Cyprus. *Chemical Geology* **423**, 7-18 (2016).
40. J. H. Trefry *et al.*, Trace metals in hydrothermal solutions from Cleft segment on the southern Juan de Fuca Ridge. *Journal of Geophysical Research: Solid Earth* **99**, 4925-4935 (1994).
41. S. Metz, J. H. Trefry, Chemical and mineralogical influences on concentrations of trace metals in hydrothermal fluids. *Geochimica et Cosmochimica Acta* **64**, 2267-2279 (2000).
42. J. E. Johnson, S. M. Webb, C. Ma, W. W. Fischer, Manganese mineralogy and diagenesis in the sedimentary rock record. *Geochimica et Cosmochimica Acta* **173**, 210-231 (2016).
43. T. Neumann, U. Heiser, M. A. Leosson, M. Kersten, Early diagenetic processes during Mn-carbonate formation: evidence from the isotopic composition of authigenic Ca-rhodochrosites of the Baltic Sea. *Geochimica et Cosmochimica Acta* **66**, 867-879 (2002).
44. E. M. Herndon, J. R. Havig, D. M. Singer, M. L. McCormick, L. R. Kump, Manganese and iron geochemistry in sediments underlying the redox-stratified Fayetteville Green Lake. *Geochimica et Cosmochimica Acta* **231**, 50-63 (2018).
45. H. Dong *et al.*, A critical review of mineral–microbe interaction and co-evolution: mechanisms and applications. *National science review* **9**, nwac128 (2022).
46. J. Kim *et al.*, Naturally occurring, microbially induced smectite-to-illite reaction. *Geology* **47**, 535-539 (2019).
47. J. Aubineau *et al.*, Microbially induced potassium enrichment in Paleoproterozoic shales and implications for reverse weathering on early Earth. *Nature Communications* **10**, 2670 (2019).

FIGURES

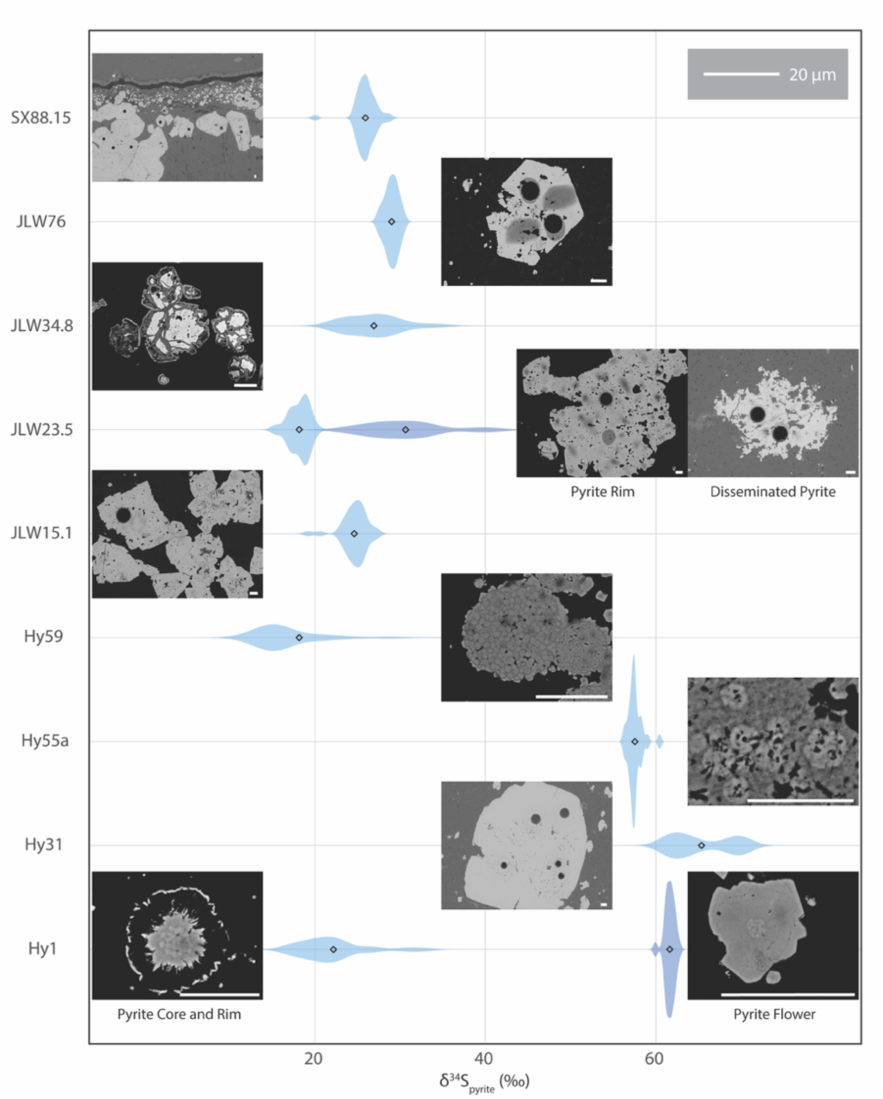


Fig. 3. 1. SIMS $\delta^{34}S_{pyrite}$ data and representative BSE-SEM images of samples analyzed in this study.

$\delta^{34}S_{pyrite}$ data were summarized as kernel density plots with mean $\delta^{34}S_{pyrite}$ marked as diamonds. See Section 2.1. for detailed petrographic descriptions. Rim pyrite in JLW15.1, JLW23.5, JLW34.8, and SX88.15 was analyzed. Disseminated subhedral pyrite in JLW23.5 and JLW76 was analyzed. Black dots and dark gray shades in SEM images are, respectively, LA-ICP-MS and SIMS analytical spots. See Supplementary Material for detailed SEM images and elemental maps (Figs. S1−S9).

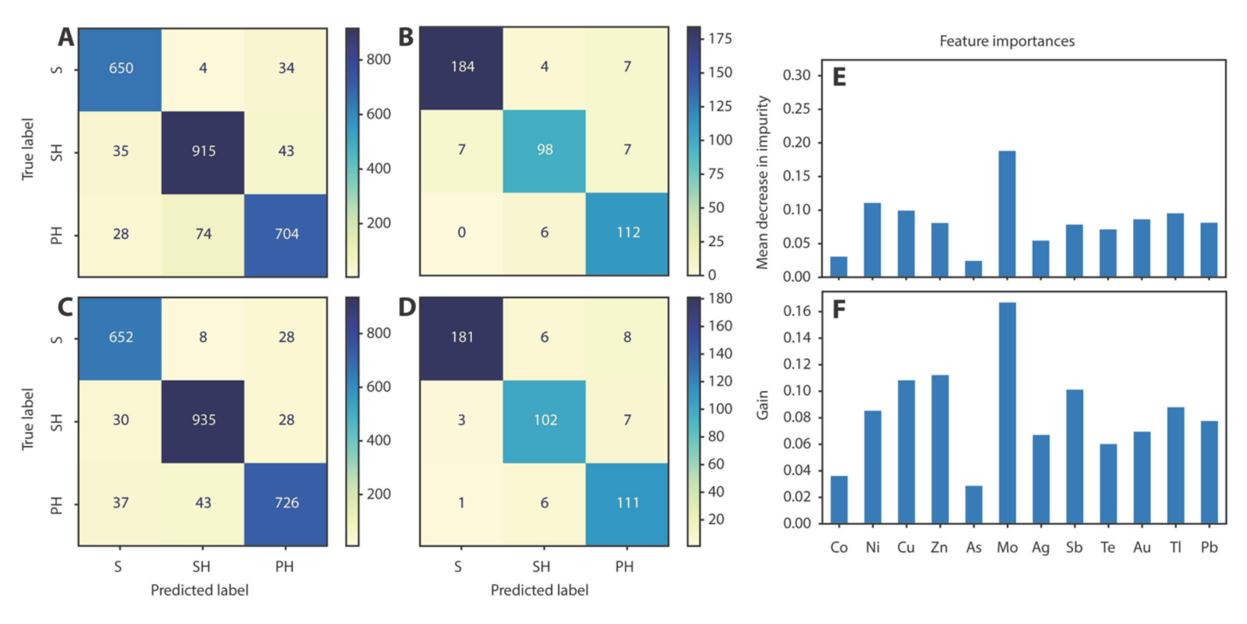


Fig. 3. 2. Supervised machine learning model confusion matrix and feature importance. (A) and (B) are confusion matrixes of the Random Forests model on in-sample-locality test and out-of-sample-locality test datasets, respectively. (C) and (D) are confusion matrixes of the XGBoost model on in-sample-locality test and out-of-sample-locality test datasets, respectively. Confusion matrixes visualize model performance on identifying sedimentary (S), synsedimentary hydrothermal (SH), and post-sedimentary hydrothermal (PH) pyrite. (E) and (F) show the feature importance of the Random Forests and XGBoost models, respectively. For Random Forests, the feature importance is decided by the mean decrease in impurity when a feature is used to build a split. For XGBoost, the feature importance is decided by the gain, which is the improvement in accuracy brought by a feature.

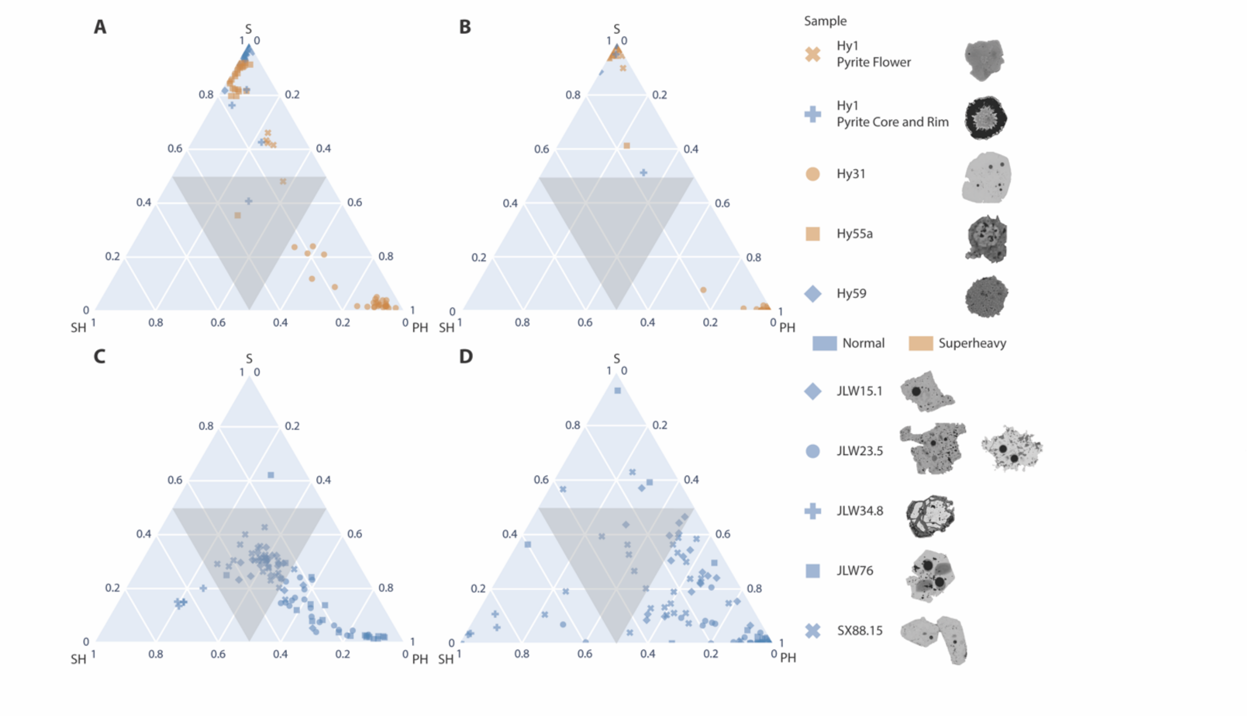


Fig. 3. 3. Ternary plots of model predictions on Hy, JLW, and SX samples.

Model predictions on pyrite analyses of Hy, JLW, and SX samples are expressed as the probability of being identified as sedimentary (S), synsedimentary hydrothermal (SH), or post-sedimentary hydrothermal (PH) pyrite. Each data point represents one single LA-ICP-MS analysis. (A) and (B) are model outputs of Hy samples based on the Random Forests (A) and XGBoost (B) models. (C) and (D) are model outputs of JLW and SX samples based on the Random Forests (C) and XGBoost (D) models. Blue data points represent isotopically normal pyrite and brown data points represent isotopically superheavy pyrite. Superheavy pyrite is defined by $\delta^{34}S$ >50‰, which is substantially higher than that of contemporaneous seawater (*4*). Gray triangular shades in ternary plots represent the ambiguous zone. Pyrite analyses in the ambiguous zone have a less than 50% probability of being identified to any single label.

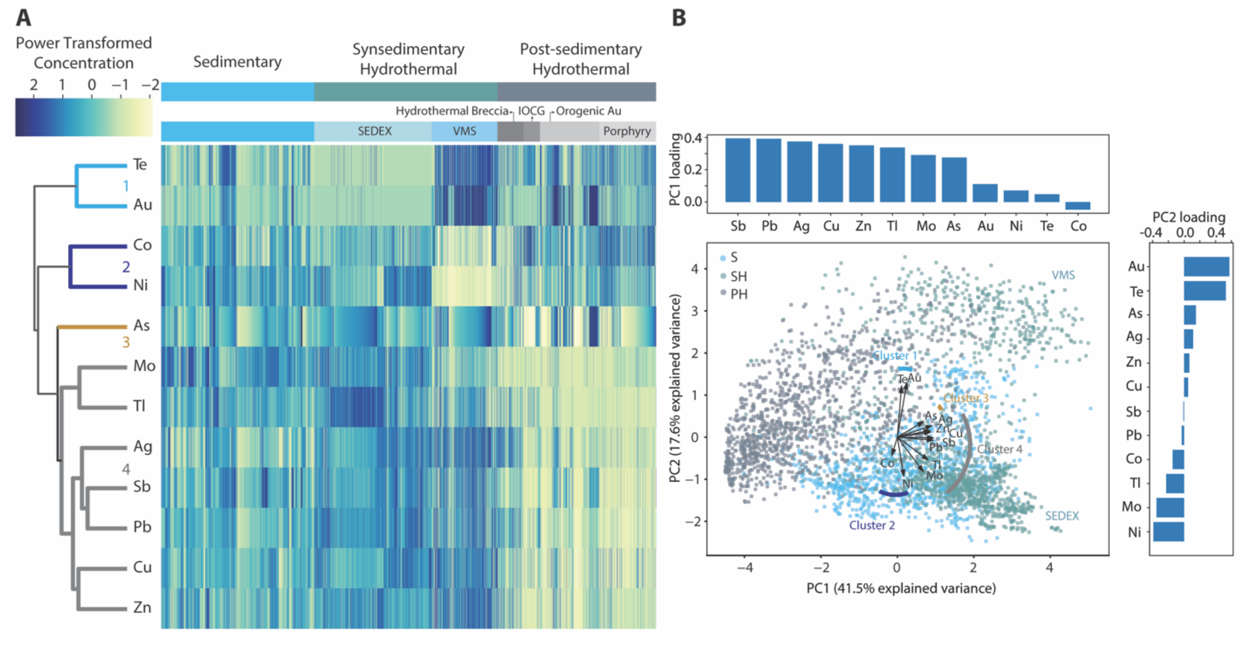


Fig. 3. 4. HCA and PCA for power-transformed pyrite trace element database.

(A) HCA by elements (n = 12). Four elemental clusters were identified and marked in different colors on the dendrogram. Columns (n = 3632) represent individual pyrite analyses, which were grouped by their origins. Each grid represents the power-transformed concentration of a specific trace element from an individual pyrite analysis. (B) PCA biplot of the first two principal components (PC1 and PC2). Loadings for each element in PC1 and PC2 are shown as arrows in the biplot and histograms along the x and y axes. The length and direction of the arrows were decided by the loadings in PC1 and PC2. The arrows were also grouped by elemental clusters from HCA. Analyses of synsedimentary hydrothermal pyrite are divided into two clusters in the biplot: the upper cluster is VMS-type pyrite and the lower cluster is SEDEX-type pyrite.

TABLE

Model	Testing Dataset	Average Precision	Average Recall	Average F1 score
Random Forests	In-sample-locality test	0.91	0.91	0.91
	Out-of-sample-locality test	0.92	0.92	0.92
XGBoost	In-sample-locality test	0.93	0.93	0.93
	Out-of-sample-locality test	0.92	0.93	0.92

Table 3. 1. Supervised machine learning model evaluation.

Simple arithmetic averages of precision, recall, and F1 score across labels (sedimentary, synsedimentary hydrothermal, post-sedimentary hydrothermal) on in-sample-locality test and out-of-sample-locality test datasets. For each label, the precision is calculated by dividing the true positives by the sum of true positives and false positives, recall by dividing the true positives by the sum of true positives and false negatives, and F1 score is the harmonic mean of precision and recall.

CONCLUSION

The Neoproterozoic Era stands as a critical interval in Earth's history, witnessing profound environmental shifts that influenced the evolution of life. Through an integrative exploration of this era, this dissertation has shed new light on the interaction of life and environment in a key chapter of Earth history.

The investigation into the early Tonian Huaibei Group, as described in Chapter 1, uncovered vital insights into oceanic redox structure and nutrient cycles during the early Neoproterozoic. The Fe speciation and nitrogen isotope data indicate ferruginous conditions in shallow marine environments, where nitrate availability was limited due to strong denitrification/anammox in anoxic seawaters. This revelation, complemented by a global compilation of Mesoproterozoic-Neoproterozoic $\delta^{15}N$ data, unveiled a stepwise increase in marine nitrate availability and eukaryotic contribution to sedimentary nitrogen at ~800 Ma. These changes likely stemmed from a confluence of tectonic, oceanic, atmospheric, and biological factors that in combination shaped marine redox conditions and nutrient availability, thereby fostering the ecological ascendancy of eukaryotes.

In Chapter 2, my goal was to establish a robust stratigraphic framework for the early Tonian Period with data from three cratons, enabling us to better understand the spatial and temporal variations in oceanic environments during the Neoproterozoic. The dynamic time warping algorithm successfully aligned $\delta^{13}C_{carb}$ chemostratigraphic data from different sedimentary sections and identified the ca. 920 Ma Majiatun excursion as a potential chemostratigrpahic marker for global correlation. Although its underlying drivers remain unknown, this excursion likely marks the onset of Neoproterozoic carbon cycle volatility.

The novel application of supervised machine learning in Chapter 3 unraveled the origins of Neoproterozoic superheavy pyrite, differentiating sedimentary and hydrothermal pyrite on the basis of trace element concentrations. This analytical technique holds promise in decoding the complexities surrounding sulfur biogeochemical cycles. Furthermore, the study of pyritic rims surrounding fossiliferous chert nodules from the Ediacaran Doushantuo Formation unraveled a complex history, underscoring the key role of early diagenetic pyrite in chert nodule formation and the subsequent overprinting by hydrothermal fluids.

Collectively, these chapters deepen our understanding of the Neoproterozoic Era, highlighting the intertwined relationship between environmental dynamics and the evolution of life. The nuanced insights garnered herein not only enrich our comprehension of Earth's history but also offer valuable perspectives in deciphering the current environmental challenges. As we navigate contemporary environmental threats, this research provides crucial lessons in resilience and adaptation from a pivotal era in our planet's history.

APPENDICES

Supplementary Material for Chapter 1

Supplementary Text

Lip biomarkers and the ecological dominance of eukaryotes

As a proxy for ecological dominance, lipid biomarkers have been used to understand the relative contribution of bacterial versus eukaryotic biomass, especially primary biomass, to sedimentary organic carbon in marine sediments. The dearth of eukaryote-derived steranes, coupled with the dominance of bacterially derived hopanes, in sedimentary rocks older than the Tonian has been taken as evidence for a low ecological abundance of eukaryotes (*1, 2*). The oldest known steranes may be present in rocks as old as 1400 Ma (*3*), but it is not until the Tonian Period when the ratio of steranes to hopanes (S/H) increased by several orders of magnitude and reached up to 0.38 in sedimentary organic matter, an increase that has been interpreted as evidence for an ecological rise of eukaryotic primary producers (*1, 2*). These late Tonian steranes are dominated by cholestane (C_{27}), possibly sourced from red algae (*2*) and/or heterotrophic eukaryotes (*1*). A later switch to a stigmastane (C_{29}) dominance in the Cryogenian is interpreted as evidence for the ecological expansion of green algae (*1*). Although the biomarker record may have been subjected to preservational biases [e.g., the mat-seal effect, which preferentially removes sterols relative to hopanoids from microbial mat communities (*4*)], a ca. 800 Ma rise of a eukaryote-rich marine ecosystem is also supported by other proxies such as zinc isotope data (*5*). Higher sedimentary δ^{66}Zn values in the mid-Neoproterozoic are interpreted as evidence for increasing organic-derived Zn burial, likely related to a larger eukaryotic contribution in the ecosystem due to the higher affinity of eukaryotes for Zn compared with prokaryotes (*5*).

Distribution of diatoms, cyanobacteria, and nitrate in modern oceans

To investigate the distribution patterns of diatoms, cyanobacteria, and nitrate in modern oceans, we analyzed data from NASA Ocean Biogeochemical Model (NOBM) (*6*). NOBM is a 3D model of coupled circulation/biogeochemical/radiative processes in the global ocean (*7*). It has been extensively validated by in-situ and remote sensing data (*7, 8*). In this study, we focused on monthly data between 01/2011 and 12/2015. The dataset covers open ocean areas, with a bottom depth of > 200 m, spanning from –84° to 72° latitude and –180° to 180° longitude in increments of 1.25° longitude by 2/3° latitude. The concentrations of diatoms and cyanobacteria are expressed as units of chlorophyll a (mg chl a/m^3), and the concentration of nitrate is expressed as µM (micro mole/L).

Singular Value Decomposition (SVD) was used to extract major features from five-year monthly data of all three variables. To calculate the time-averaged (01/2011–12/2015) concentrations of the i^{th} major feature, we used the product of the i^{th} left singular vector, the i^{th} singular value, and the mean of i^{th} right singular vector (the temporal variation of the i^{th} feature). The time-averaged concentration maps of the first major feature of the three variables are shown in Fig. 1. Diatoms are enriched in high latitude and upwelling areas, and nitrate exhibits similar distribution patterns. Conversely, cyanobacteria show enrichment in mid-ocean gyres. In summary, diatoms are enriched in areas with high nitrate concentrations whereas cyanobacteria dominate areas with low nitrate concentrations. This pattern is also supported by a linear regression analysis, which shows positive correlation between time-averaged diatom and nitrate data of the first major component, and negative correlation between cyanobacteria and nitrate data (Fig. 1D–E).

Fe speciation of carbonate samples

As discussed in the main text, the iron speciation proxy was initially developed from analysis of fine-grained siliciclastic rocks. While recent work by Clarkson et al. (*9*) demonstrates the usefulness of iron speciation data from carbonate rocks with Fe_T > 0.5 wt%, they focused on the Fe_{HR}/Fe_T ratio. To evaluate the effectiveness of Fe_{py}/Fe_{HR} of carbonate rocks, which is important for distinguishing euxinic and ferruginous conditions, we compiled an iron speciation dataset of carbonate samples (FeT > 0.5 wt%) from modern and ancient anoxic basins (table S1). In this dataset, redox conditions of ancient carbonate samples are inferred from Fe speciation data of stratigraphically adjacent siliciclastic rocks or from other independent redox proxies such as molybdenum enrichment factor. The result (Fig. 4) confirms the validity of the recently proposed threshold (Fe_{py}/Fe_{HR} = 0.6–0.8) (*10*) for the separation of definitely ferruginous and euxinic conditions based on carbonate samples with Fe_T > 0.5 wt%.

Model validation: Reproducing modern ocean $\delta^{15}N$ values

To validate the nitrogen cycle model used in the main text, we first ran a full model that considered all fluxes, including all fluxes depicted in fig. S2, as well as two minor input fluxes (i.e., atmospheric deposition, F_{depo}; and riverine input, F_{river}). This full model, as described below, was able to reproduce modern ocean $\delta^{15}N$ values. We then re-ran a simplified model (fig. S2), with F_{depo} and F_{river} omitted, and the modeling results did not change substantially. Thus, we adopted the simplified model (as described in the main text and depicted in fig. S2) to simulate the early Tonian nitrogen cycle.

At steady state, the input/output fluxes of N_{fixer}/ammonium and $N_{assimilator}$/nitrate should maintain a mass balance. According to the full model, we have

$$F_{modern\_fix} = F_{remin} + F_{fixer\_burial} \quad (1)$$

$$F_{remin} + F_{modern\_depo} + F_{modern\_river} = F_{wcd} + F_{sd} + F_{assimilator\_burial} \quad (2)$$

Modern input and burial fluxes and their nitrogen isotopic compositions are summarized in table S2. As estimated modern input fluxes are available, we defined burial efficiency (k) as the proportion of input fluxes being buried.

$$F_{fixer\_burial} = F_{fix} \times k_{fixer} \quad (3)$$

$$F_{assimilator\_burial} = \left(F_{remin} + F_{depo} + F_{river}\right) \times k_{assimilator} \quad (4)$$

Considering the greater body size of eukaryotic assimilators relative to that of nitrogen fixers in the modern ocean, we assumed that the burial efficiency of assimilator is 2 times higher than that of fixer.

$$k_{assimilator} = 2 \times k_{fixer} \quad (5)$$

In addition, the sedimentary denitrification/anammox flux is estimated to be 2.4 times water column denitrification/anammox flux in the modern ocean (*11*).

$$F_{sd} = 2.4 \times F_{wcd} \quad (6)$$

At steady state, isotope mass balance is also maintained.

$$F_{fix} \times (\delta^{15}N_{atmo} + \varepsilon_{fix}) = F_{remin} \times \delta^{15}N_{fixer/ammonium} + F_{fixer\_burial} \times \delta^{15}N_{fixer/ammonium} \quad (7)$$

$$\begin{aligned} F_{remin} \times \delta^{15}N_{fixer/ammonium} + F_{river} \times (\ \delta^{15}N_{terrestrial} + \varepsilon_{river}) + F_{depo} \times (\ \delta^{15}N_{atom} + \varepsilon_{depo}) \\ = F_{wcd} \times (\delta^{15}N_{assimilator/nitrate} + \varepsilon_{wcd}) + F_{sd} \times (\delta^{15}N_{assimilator/nitrate} \\ + \varepsilon_{sd}) + F_{assimilator\_burial} \times \delta^{15}N_{assimilator/nitrate} \end{aligned} \quad (8)$$

where $\delta^{15}N_{atmo}$, $\delta^{15}N_{terrestrial}$, $\delta^{15}N_{fixer/ammonium}$, $\delta^{15}N_{assimilator/nitrate}$ are nitrogen isotope compositions of atmospheric N_2, terrestrial nitrogen reservoir, N_{fixer}/ammonium reservoir, and $N_{assimilator}$/nitrate reservoir, and ε_{fix}, ε_{river}, ε_{depo}, ε_{wcd}, and ε_{sd} are fractionation factors ($\varepsilon = \delta^{15}N_{product} - \delta^{15}N_{reactant}$) associated with nitrogen fixation, riverine input, atmospheric deposition, and water-column and sedimentary denitrification/anammox. As discussed in the main text, nitrate burial rate is expressed as a proportion of total nitrogen burial rate and is ultimately related to marine nitrate availability ($N_{assimilator}$).

$$f_{assimilator}=F_{assimilator\_burial}/F_{total\_burial} \tag{9}$$

Next, the nitrogen isotope composition of sediments ($\delta^{15}N_{sed}$) can be calculated according to equation (10).

$$\delta^{15}N_{sed}=(1 - f_{assimilator}) \times \delta^{15}N_{fixer/ammonium} + f_{assimilator} \times \delta^{15}N_{assimilator/nitrate} \tag{10}$$

Solving equation (1) to (10) with parameters in table S2, we were able to reproduce the $\delta^{15}N$ of modern oceanic sediments (average 4.8‰, range 2.7‰−6.9‰) with a $f_{assimilator}$ of 0.70, which is consistent with observations of modern seafloor sediments (mode 4‰−6‰, range 2.5‰−16.6‰) (*12*). We also conducted sensitivity tests, and modified the burial efficiency of assimilators (eukaryotes) to be 3 times or 5 times greater than that of fixers (cyanobacteria). The reproduced $\delta^{15}N$ of modern oceanic sediments is 5.3‰ (3.0‰−7.6‰) or 5.8‰ (3.4‰−8.2‰), which is still in good agreement with modern sediment measurements. Finally, we ran the simplified model (fig. S2) and were also able to reproduced the modern values ($\delta^{15}N_{sed}$ = 6.5‰, range 3.9‰−9.1‰; $f_{assimilator}$ = 0.66), suggesting that F_{depo} and F_{river} are negligible. Therefore, we argue that our box model results are robust and accurately captures major processes controlling sedimentary nitrogen isotopic compositions.

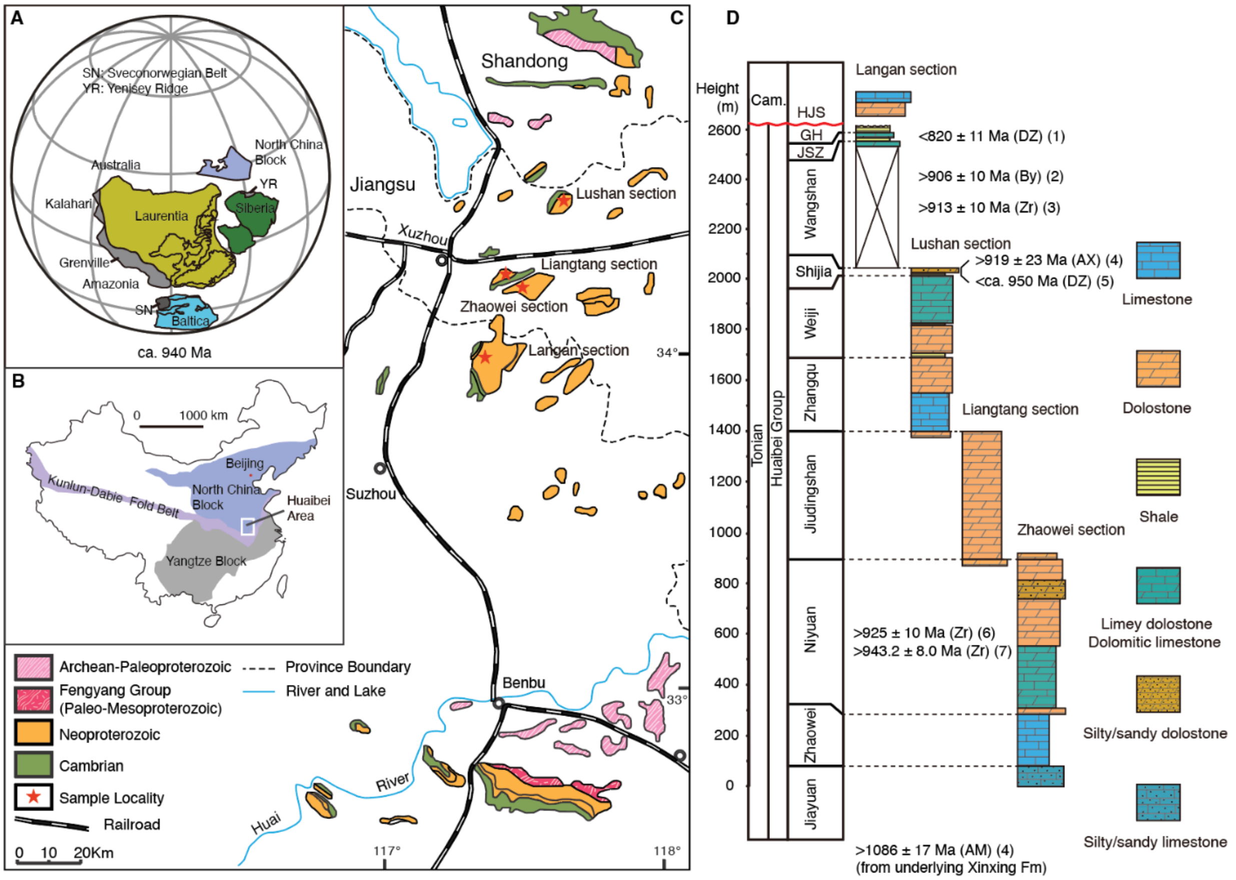


Fig. 1. S1. Paleogeographic and geological maps and stratigraphic column.

(A) Paleogeographic map showing location of North China Block, Laurentia, Baltica, and Siberia at ca. 940 Ma, after Zhao et al. (*13*). (B) Simplified geological map showing location of North China Block. (C) Geological map of the Huaibei and Huainan regions. Sample locations are marked as red stars. (D) Stratigraphic column of middle and upper Huaibei Group in four sampling sections. Modified from Xiao et al. (*14*) and Wan et al. (*15*). References for radiogenic isotope ages: (1) Yang et al. (*16*), (2) Su et al. (*17*), (3) Zhu et al. (*18*), (4) Zhang et al. (*19*), (5) He et al. (*20*), (6) Fu et al. (*21*), (7) Zhao et al. (*13*). HJS−Houjiashan Fm, GH−Gouhou Fm, JSZ−Jinshanzhai Fm, DZ−detrital zircon, By−baddeleyite, Zr−magmatic zircon, AX−authigenic xenotime, AM−authigenic monazite.

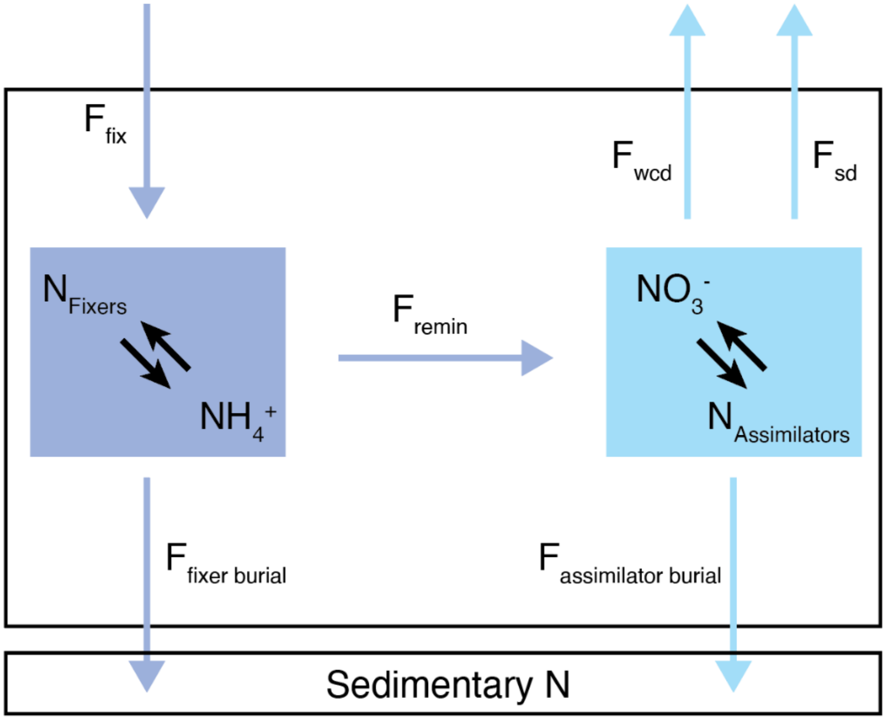


Fig. 1. S2. Simplified nitrogen cycle box model.

Box model was modified after Kipp et al. (*22*). Two major reservoirs are presented in our model. The first reservoir (N_{fixer}/ammonium) includes organic nitrogen of organisms fixing nitrogen (N_{fixer}) and ammonium produced from remineralization of N_{fixer}. The second reservoir ($N_{assimilator}$/nitrate) includes nitrate and organic nitrogen of organisms assimilating nitrate ($N_{assimilator}$). Nitrogen fixation (F_{fix}) is the only input flux considered when simulating Precambrian oceans, and denitrification/anammox in water columns (F_{wcd}) and sediments (F_{sd}) are the major output fluxes, whereas F_{fixer_burial} and $F_{assimilator\_burial}$ are minor output fluxes. Nitrogen is supplied to the nitrate reservoir through remineralization and nitrification (F_{remin}). The sedimentary nitrogen is a mixture of burial fluxes derived from fixers (N_{fixer}) and assimilators ($N_{assimilator}$).

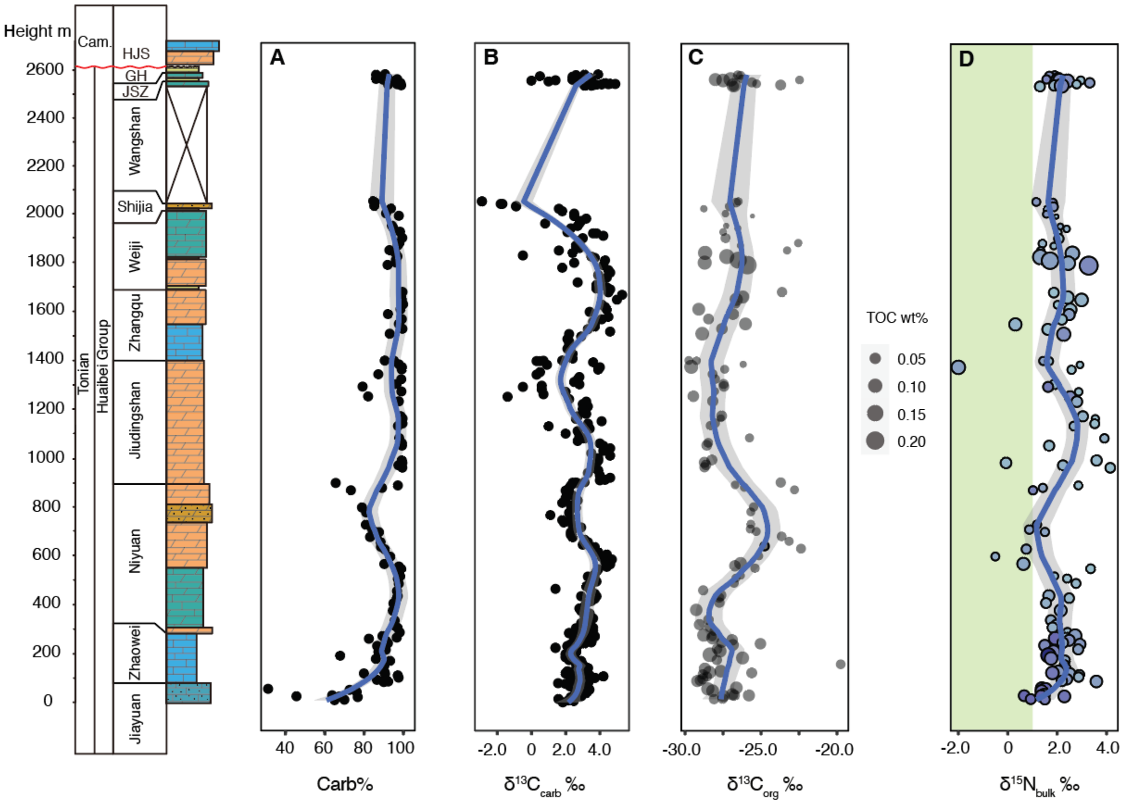


Fig. 1. S3. Geochemical profiles of carbonate content (C_{arb}%) (A), inorganic carbon isotope composition ($\delta^{13}C_{carb}$) (B), organic carbon isotope composition ($\delta^{13}C_{org}$) (C), and bulk nitrogen isotope composition ($\delta^{15}N_{bulk}$) (D) of the Huaibei Group samples.

$\delta^{13}C_{carb}$ data were published in Xiao et al. (*14*). In (C), size of data points is keyed to reflect TOC contents. In (D), color and size of data points are keyed to reflect TN and TOC contents. Blue lines are LOWESS regression fit, with 95% confidence interval (gray shadow) calculated from standard errors.

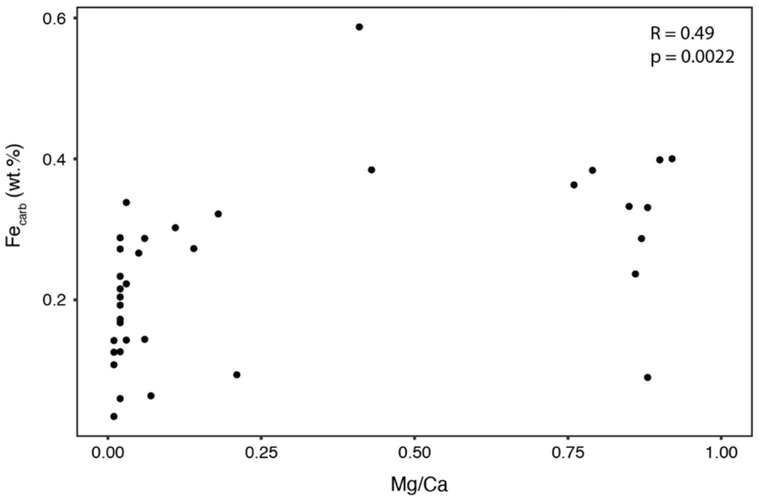


Fig. 1. S4. Cross-plot of Fe_{carb}-Mg/Ca for Huaibei Group samples.

R and p values were calculated using Pearson correlation coefficient. No significant correlation is observed for Huaibei Group samples. Mg and Ca concentrations were published in Zhang et al. (*23*).

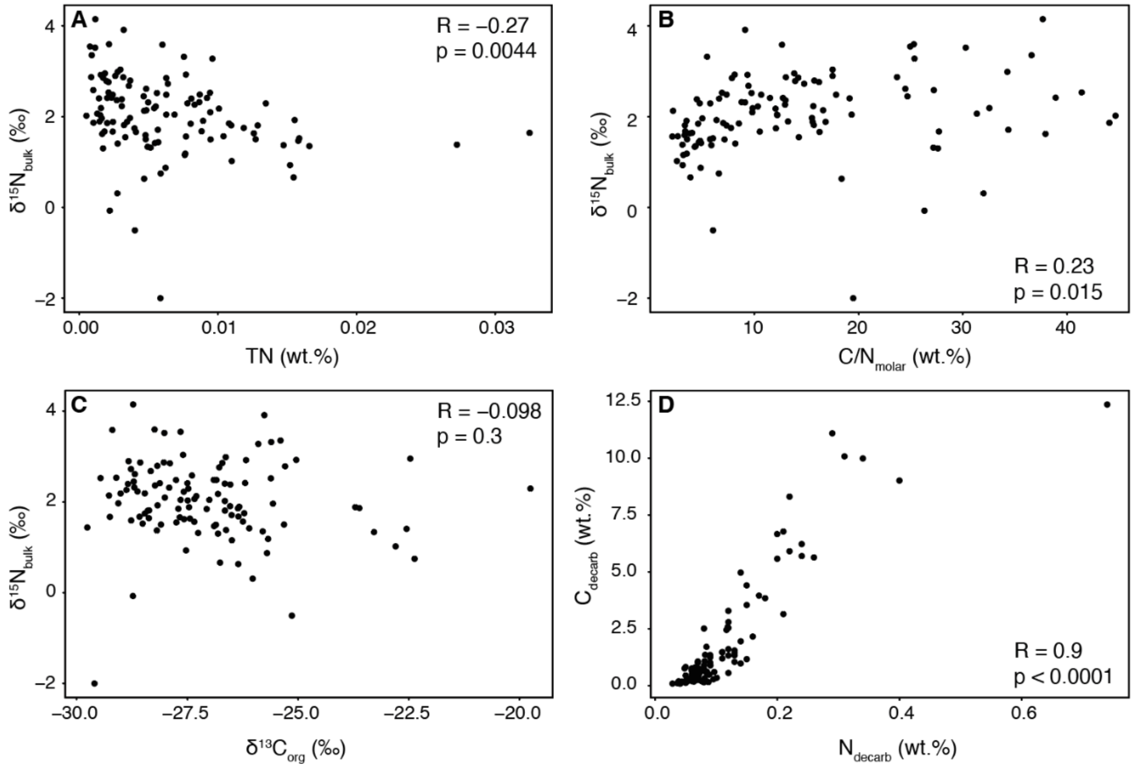


Fig. 1. S5. Cross-plots of $\delta^{15}N_{bulk}$-TN (A), $\delta^{15}N_{bulk}$-C/N_{molar} (B), $\delta^{15}N_{bulk}$-$\delta^{13}C_{org}$ (C), C_{decarb}-N_{decarb} of Huaibei Group samples.

R and p values were calculated using Pearson correlation coefficient. No significant correlation is observed in $\delta^{15}N_{bulk}$ vs. TN, C/N_{molar}, and $\delta^{13}C_{org}$. A strong positive correlation between C_{decarb} and N_{decarb} is observed, suggesting analyzed nitrogen largely originated from organic matter.

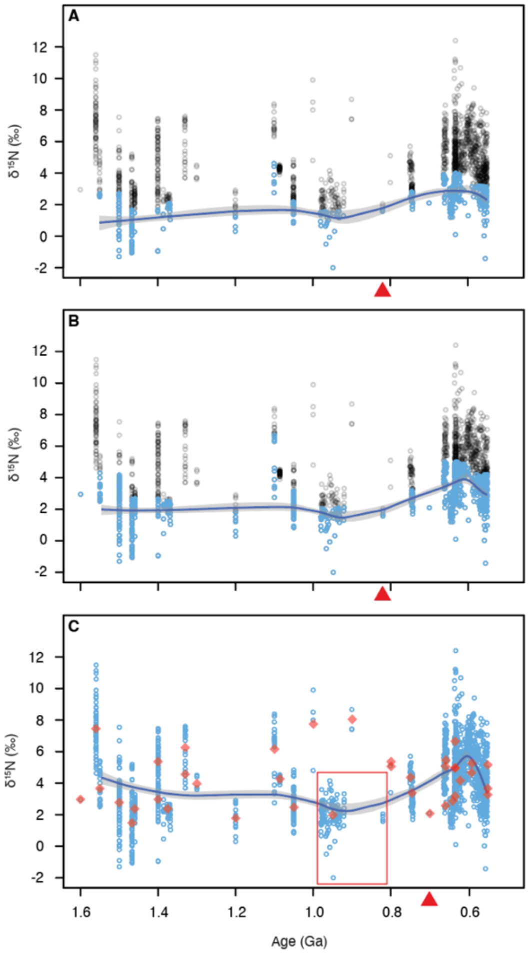


Fig. 1. S6. LOWESS regression fit and changepoint analysis on compiled $\delta^{15}N$ dataset.

(A) LOWESS regression fit and changepoint analysis (red triangle) on lowest 25% data of every 100 Myr time bin. (B) LOWESS regression fit and changepoint analysis (red triangle) on lowest 50% data of every 100 Myr time bin. (C) LOWESS regression fit and changepoint analysis (red triangle) on entire dataset. 95% confidence intervals (gray shade) were calculated from 10,000 bootstrapping experiments. Blue data points represent samples considered in each regression fit analysis. Red diamonds represent average values for each individual dataset (typically from the same stratigraphic unit). Red rectangle highlights the data from the Huaibei Group that fills a critical gap in Proterozoic $\delta^{15}N$ data.

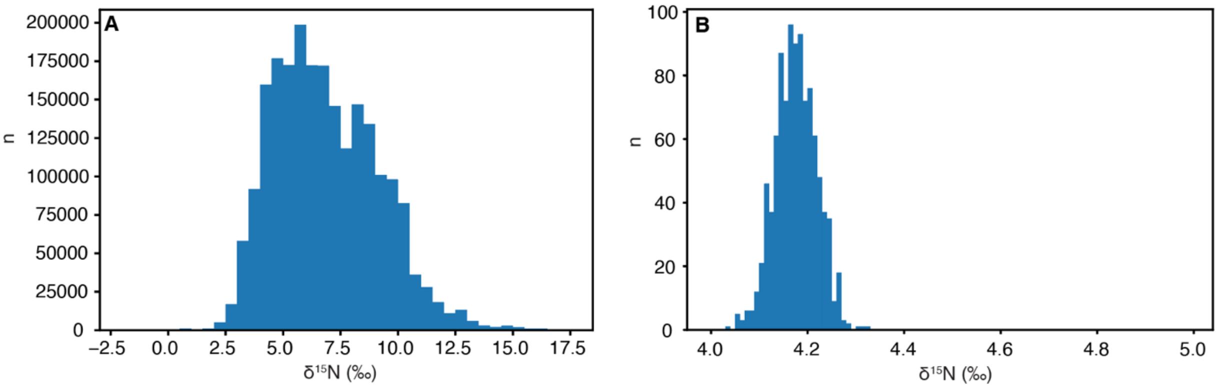


Fig. 1. S7. Modern marine sediment nitrogen isotope composition.

(A) Frequency distributions of modern marine sediment $\delta^{15}N$ values from bootstrapping experiments (1000 bootstrap iterations). (B) Frequency distributions of mean $\delta^{15}N$ values of the lower quartile data (modern marine sediment) from bootstrapping experiments. The results show that the lower quartile has an average $\delta^{15}N$ of 4.18‰, which is substantially higher than any Proterozoic interval that we analyzed. Nitrogen isotope data were compiled by Tesdal et al. (*12*).

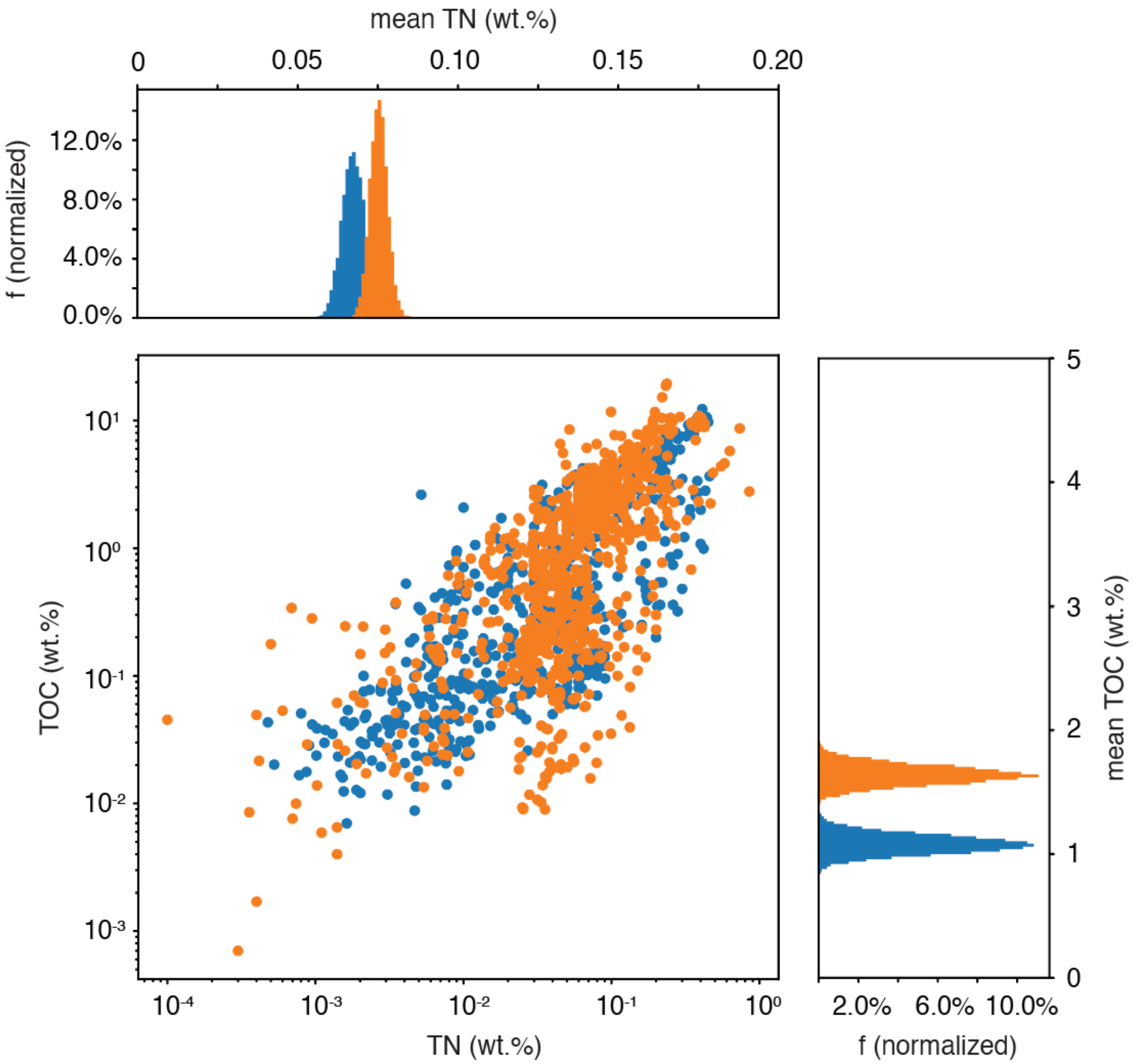


Fig. 1. S8. Comparison of TOC and TN content of compiled sedimentary rock dataset.

Pre-800 Ma (blue) and post-800 Ma (orange) subgroups have broadly similar TOC and TN contents. Cross-plot shows bulk concentrations, and histograms show mean TOC and TN values from bootstrapping experiments (n = 10,000). Note that scales of raw data and resampled means are different.

Reference	**Locality**	**Independent Evidence for Redox Condition**
(*24*)	Cariaco Basin	Water column measurements
(*9*)	Black Sea	Water column measurements
(*25*)	Demerara Rise, Coniacian-Santonian OAE3, Cretaceous	P, Cd, Mo, V, Zn concentrations
(*26*)	Yangtze Platform, early Cambrian	U, V, Mo concentrations
(*27*)	Yangtze Block, Ediacaran-Cambrian	Stratigraphically adjacent siliciclastic rocks
(*28*)	Yanshan Basin, Middle Proterozoic	Mo concentration
(*29*)	Eastern Mediterranean Sapropels, late Quaternary	Fe isotope and Mo concentration

Table 1. S1. Published dataset used in Fe_{py}/Fe_{HR} compilation.

Parameter	Description	Values	Unit	Reference
F_{modern_fix}	Modern nitrogen fixation flux	205	Tg a^{-1}	(*30*)
F_{modern_depo}	Modern atmospheric nitrogen deposition flux	15	Tg a^{-1}	(*30*)
F_{modern_river}	Modern riverine nitrogen input flux	25	Tg a^{-1}	(*30*)
F_{total_burial}	Total burial flux, kept as a constant in the model	25	Tg a^{-1}	(*30*)
ε_{fix}	Fractionation factor associated with nitrogen fixation	−2 to 1	‰	(*22*)
ε_{depo}	Fractionation factor associated with atmospheric deposition	−4	‰	(*30*)
ε_{river}	Fractionation factor associated with riverine input	3	‰	(*30*)
ε_{wcd}	Fractionation factor associated with water column denitrification/anammox	−30 to −22	‰	(*22*)
ε_{sd}	Fractionation factor associated with sedimentary denitrification/anammox	0	‰	(*22*)
$\delta^{15}N_{atmo}$	Atmospheric nitrogen isotope composition	0	‰	(*30*)
$\delta^{15}N_{terrestrial}$	Terrestrial nitrogen isotope composition	3.4	‰	(*30*)

Table 1. S2. Parameters used in box model calculations.

References

1. J. J. Brocks *et al.*, The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* **548**, 578-581 (2017).

2. J. A. Zumberge, D. Rocher, G. D. Love, Free and kerogen‐bound biomarkers from late Tonian sedimentary rocks record abundant eukaryotes in mid‐Neoproterozoic marine communities. *Geobiology* **18**, 326-347 (2020).
3. S. Zhang *et al.*, Eukaryotic red and green algae populated the tropical ocean 1400 million years ago. *Precambrian Research* **357**, 106166 (2021).
4. P. A. Cohen, R. B. Kodner, The earliest history of eukaryotic life: uncovering an evolutionary story through the integration of biological and geological data. *Trends in Ecology & Evolution* **37**, 246-256 (2022).
5. T. T. Isson *et al.*, Tracking the rise of eukaryotes to ecological dominance with zinc isotopes. *Geobiology* **16**, 341-352 (2018).
6. W. Gregg, C. Rousseaux. (Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, USA, 2017).
7. W. W. Gregg, N. W. Casey, Modeling coccolithophores in the global oceans. *Deep Sea Research Part II: Topical Studies in Oceanography* **54**, 447-477 (2007).
8. W. W. Gregg, P. Ginoux, P. S. Schopf, N. W. Casey, Phytoplankton and iron: validation of a global three-dimensional ocean biogeochemical model. *Deep Sea Research Part II: Topical Studies in Oceanography* **50**, 3143-3169 (2003).
9. M. O. Clarkson, S. W. Poulton, R. Guilbaud, R. A. Wood, Assessing the utility of Fe/Al and Fe-speciation to record water column redox conditions in carbonate-rich sediments. *Chemical Geology* **382**, 111-122 (2014).
10. S. W. Poulton, in *Elements in Geochemical Tracers in Earth System Science,* T. Lyons, A. Turchyn, C. Reinhard, Eds. (Cambridge University Press, 2021), pp. 1-28.
11. A. H. Devol, Denitrification, anammox, and N2 production in marine sediments. *Annual review of marine science* **7**, 403-423 (2015).
12. J. E. Tesdal, E. D. Galbraith, M. Kienast, Nitrogen isotopes in bulk marine sediment: linking seafloor observations with subseafloor records. *Biogeosciences* **10**, 101-118 (2013).
13. H. Zhao *et al.*, New geochronologic and paleomagnetic results from early Neoproterozoic mafic sills and late Mesoproterozoic to early Neoproterozoic successions in the eastern North China Craton, and implications for the reconstruction of Rodinia. *GSA Bulletin* **132**, 739-766 (2020).
14. S. Xiao *et al.*, Biostratigraphic and chemostratigraphic constraints on the age of early Neoproterozoic carbonate successions in North China. *Precambrian Research* **246**, 208–225 (2014).
15. B. Wan *et al.*, Repositioning the Great Unconformity at the southeastern margin of the North China Craton. *Precambrian Research* **324**, 1-17 (2019).
16. D.-B. Yang *et al.*, U–Pb ages and Hf isotope data from detrital zircons in the Neoproterozoic sandstones of northern Jiangsu and southern Liaoning Provinces, China: implications for the Late Precambrian evolution of the southeastern North China Craton. *Precambrian Research* **216**, 162-176 (2012).
17. X. Su, P. Peng, S. Foley, W. Teixeira, M.-G. Zhai, Initiation of continental breakup documented in evolution of the magma plumbing system of the ca. 925 Ma Dashigou large igneous province, North China. *Lithos* **384**, 105984 (2021).
18. R.-Z. Zhu *et al.*, Geochronology, geochemistry and petrogenesis of the Laozhaishan dolerite sills in the southeastern margin of the North China Craton and their geological implication. *Gondwana Research* **67**, 131-146 (2019).

19. S. Zhang *et al.*, SIMS Pb-Pb dating of phosphates in the Proterozoic strata of SE North China Craton: Constraints on eukaryote evolution. *Precambrian Research* **371**, 106562 (2022).
20. T. He *et al.*, Measuring the 'Great Unconformity' on the North China Craton using new detrital zircon age data. *Geological Society London Special Publications* **448**, 145-159 (2016).
21. X. Fu *et al.*, New paleomagnetic results from the Huaibei Group and Neoproterozoic mafic sills in the North China Craton and their paleogeographic implications. *Precambrian Research* **269**, 90-106 (2015).
22. M. A. Kipp, E. E. Stüeken, M. Yun, A. Bekker, R. Buick, Pervasive aerobic nitrogen cycling in the surface ocean across the Paleoproterozoic Era. *Earth and Planetary Science Letters* **500**, 117-126 (2018).
23. F. Zhang *et al.*, Uranium isotope evidence for extensive shallow water anoxia in the early Tonian oceans. *Earth and Planetary Science Letters* **583**, 117437 (2022).
24. T. W. Lyons, J. P. Werne, D. J. Hollander, R. Murray, Contrasting sulfur geochemistry and Fe/Al and Mo/Al ratios across the last oxic-to-anoxic transition in the Cariaco Basin, Venezuela. *Chemical Geology* **195**, 131-157 (2003).
25. C. März *et al.*, Redox sensitivity of P cycling during marine black shale formation: dynamics of sulfidic and anoxic, non-sulfidic bottom waters. *Geochimica et Cosmochimica Acta* **72**, 3703-3717 (2008).
26. C. Jin *et al.*, A highly redox-heterogeneous ocean in South China during the early Cambrian (~ 529–514 Ma): Implications for biota-environment co-evolution. *Earth and Planetary Science Letters* **441**, 38-51 (2016).
27. L. Xiang, S. D. Schoepfer, S.-z. Shen, C.-q. Cao, H. Zhang, Evolution of oceanic molybdenum and uranium reservoir size around the Ediacaran–Cambrian transition: Evidence from western Zhejiang, South China. *Earth and Planetary Science Letters* **464**, 84-94 (2017).
28. C. Li *et al.*, Marine redox conditions in the middle Proterozoic ocean and isotopic constraints on authigenic carbonate formation: Insights from the Chuanlinggou Formation, Yanshan Basin, North China. *Geochimica et Cosmochimica Acta* **150**, 90-105 (2015).
29. A. Benkovitz *et al.*, Tracing water column euxinia in Eastern Mediterranean Sapropels S5 and S7. *Chemical Geology* **545**, 119627 (2020).
30. T. Algeo, P. Meyers, R. S. Robinson, H. Rowe, G. Jiang, Icehouse–greenhouse variations in marine denitrification. *Biogeosciences* **11**, 1273-1295 (2014).

Supplementary Material for Chapter 3

Supplementary Text

Decision Tree-based Supervised Machine Learning

Decision Trees are a non-parametric supervised learning method widely used for classification. They are constructed by recursively splitting subsets of the training dataset into two descendant subsets (*1*). Conditions on features of the training dataset are used as splitting criteria to maximize the separation of the dataset (*1*). The constructed hierarchical structure describes the relationship between features and class labels. To reduce the risk of overfitting and to further produce better model performance, decision tree ensembles were developed, which combine trees together to form stronger models. Random Forests and XGBoost are two powerful ensemble methods. Random Forests construct ensembles of independent decision trees using different bootstrap samples of the training dataset. Meanwhile, at every internal node of a tree, the best splitting criterion is chosen among a randomly selected subset of features (*2, 3*). XGBoost, on the other hand, is an implementation of gradient boosting. Instead of building decision trees in parallel, gradient boosting constructs shallow decision trees (trees with only one or two splits) sequentially so that every tree corrects its predecessor's error (*4*). Specifically, every decision tree is trained using the residual errors of the predecessor as labels (*5*).

Geological Settings of Hy, JLW, SX samples

Hy samples were selected from the ZK1105 drill core, which penetrated the Cryogenian Tiesi'ao and Datangpo formations at the Daotuo mine (28°07'04" N, 108°52'26" E), Songtao County, eastern Guizhou Province in South China (*6*). The Tiesi'ao Formation is a glacial diamictite deposited during the early Cryogenian Sturtian glaciation. The overlying Datangpo

Formation was deposited in the slope-basinal facies of a rift basin (*7*) and is subdivided into three members. Member 1 is dominated by black shale with Mn-rich carbonate intervals at the base, Member 2 is dominated by gray shale, and Member 3 consists of siltstone (*6*). The depositional age of the Tiesi'ao and Datangpo formations is constrained by a TIMS zircon U-Pb age of 662.9 ± 4.3 Ma (*8*) and a SIMS zircon U-Pb age of 667.3 ± 9.9 Ma (*9*) from tuff beds in member 1 of the Datangpo Member and a SIMS zircon U-Pb age of 654.5 ± 3.8 Ma (*10*) from an ash bed in the uppermost Datangpo Formation.

JLW and SX samples were collected from the Ediacaran Doushantuo Formation at the Jiulongwan and Sixi sections (30°48'11.6"N, 111°3'16.7"E; 30°45'18.5"N 110°55'44"E; respectively) in the Yangtze Gorges area of South China (*11*). The Doushantuo Formation is generally subdivided into four members. Member I is represented by a ~5-m-thick cap carbonate. Member II consists of alternating black shales and dolostones deposited in a partially restricted marine basin (*12*). Member III consists of dolostone and ribbon rock (alternating limestone and dolostone) deposited at or above the wave base (*11*). Member IV consists of black shales. The ages of the Doushantuo Formation are constrained by zircon TIMS U-Pb ages of 635.2 ± 0.6 Ma and 551.1 ± 0.7 Ma from volcanic ash beds near the base and top of the formation (*13*).

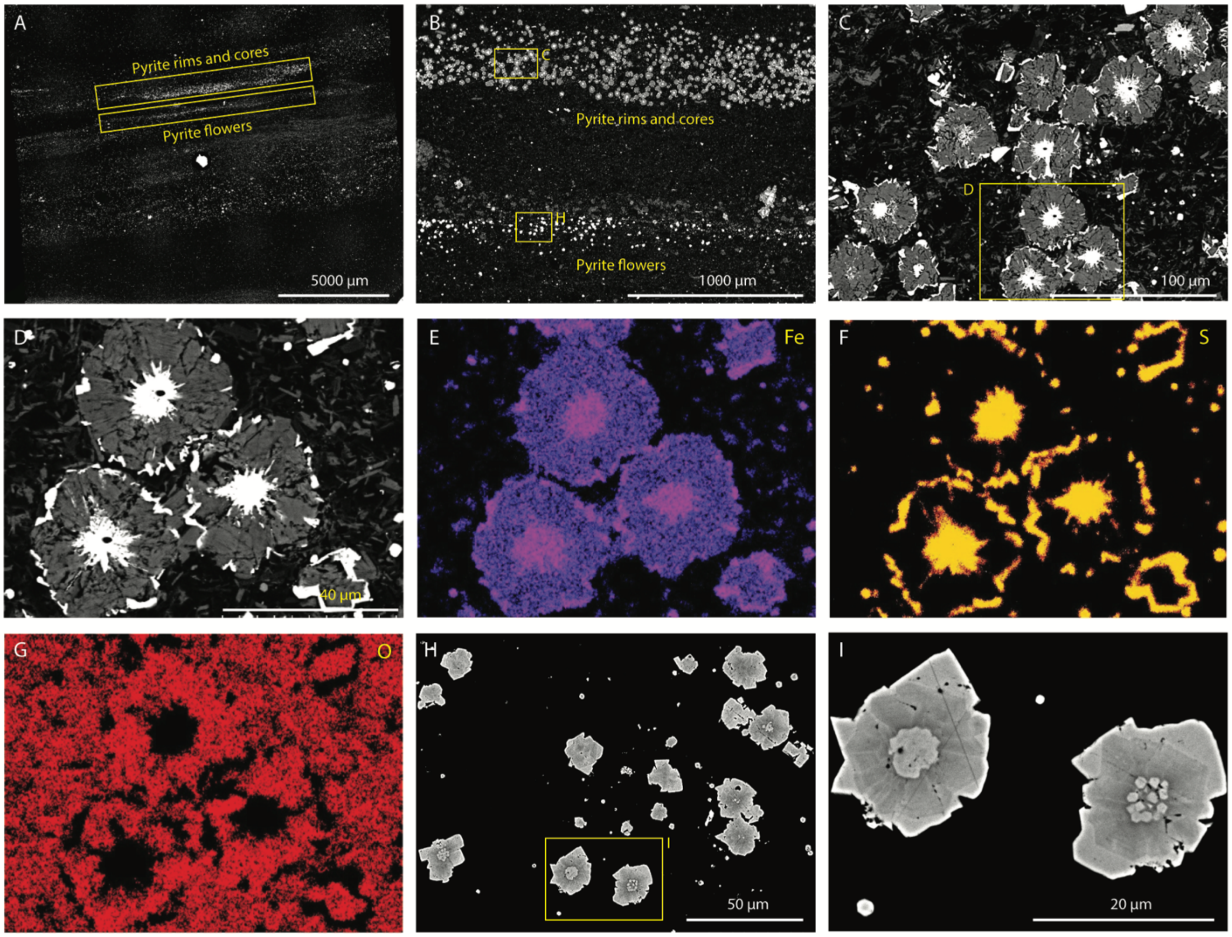


Fig. 3. S1. SEM images and elemental maps of Hy1.

Labeled rectangles mark areas of magnified views. (A) Overall SEM image. (B) Magnification of central upper part of A. (C–G). Pyrite rims and cores with a layer of iron oxide in-between. (E–G) are elemental maps of (D), with elements marked on the top right. (H) and (I). Pyrite flowers. All images and maps are from Cui et al. (*6*). Black ellipses in SEM images (C and D) are SIMS analytical spots.

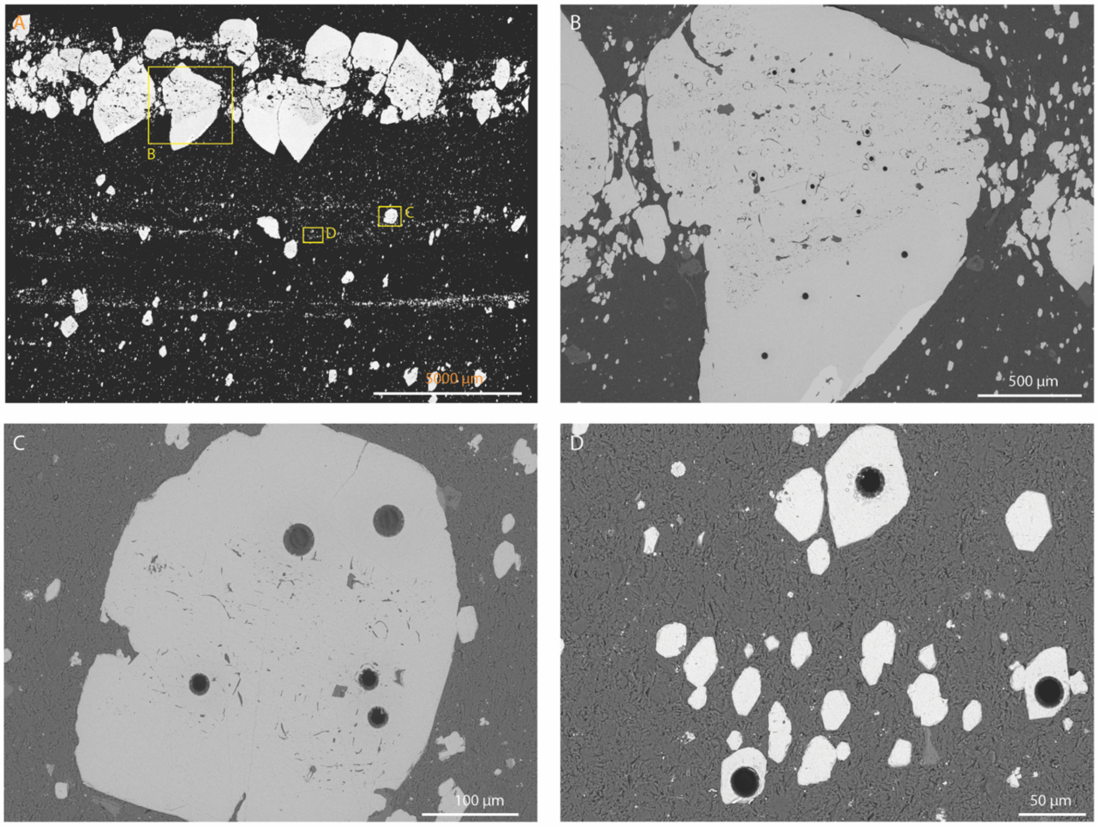


Fig. 3. S2. SEM images of subhedral pyrite grains in Hy31.

Labeled rectangles in (A) mark areas magnified in (B–D). Large pyrite grains in (B) and (C) are composed of micrometer-size sub-grains cemented by later-stage pyrite. A is from Cui et al. (*6*). (B–D) are from this study. Black dots in (B–D) images are LA-ICP-MS analytical spots.

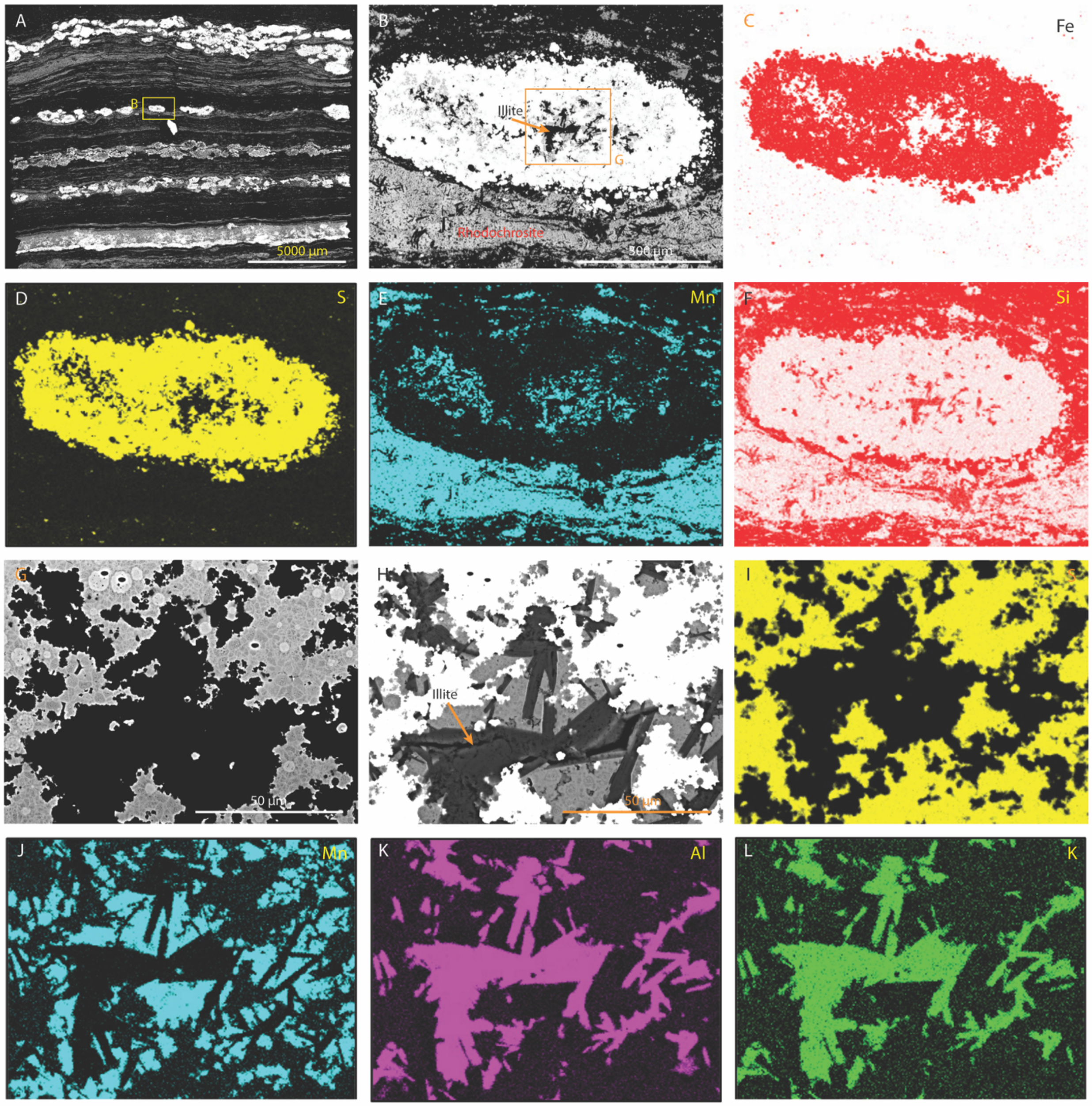


Fig. 3. S3. SEM images and elemental maps of Hy55a.

Nodular and sausage-shaped pyrite structures in Hy55a, interpreted to have replaced the pre-existing rhodochrosite layers (A–F) and lath-shaped illite crystals (G–L). (G) and (H) are the same view with different contrast and exposure to highlight fine structures of pyrite and illite. All images and maps are from Cui et al. (*6*). Black ellipses in (G) are SIMS analytical spots. Labeled rectangles in (A) and (B) mark areas magnified in (B) and (G), respectively.

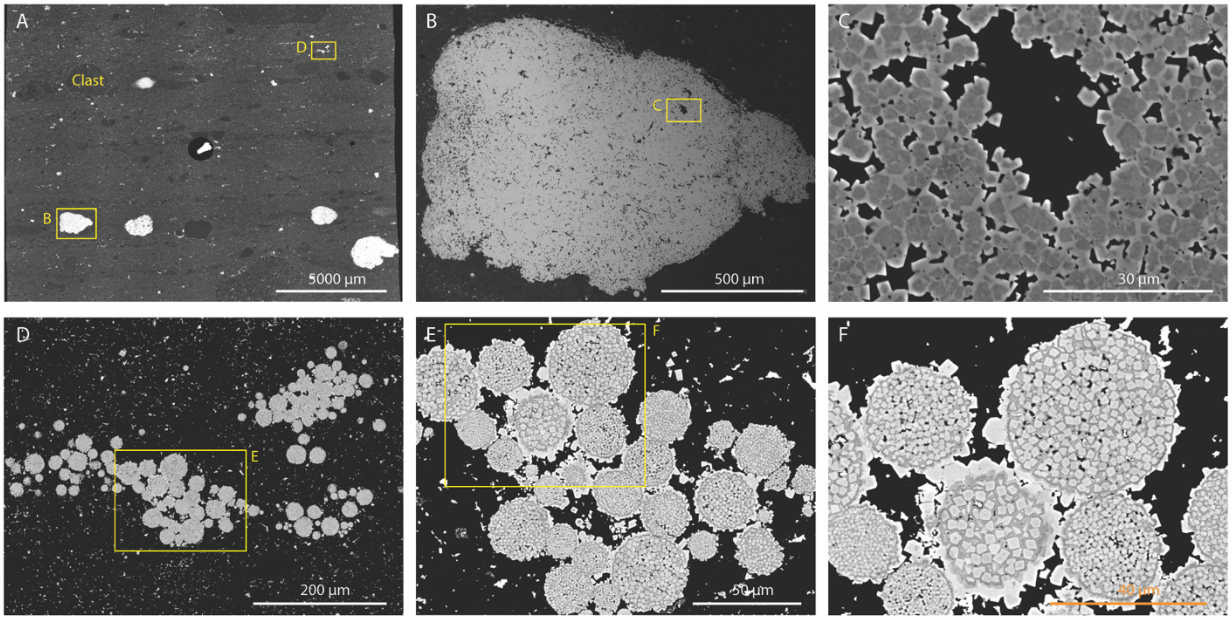


Fig. 3. S4 SEM images of Hy59.

Pyrite nodule (B) and (C) and framboids (D–F) in Hy59. All images and maps are from Cui et al. (*6*). Labeled rectangles mark areas of magnified views.

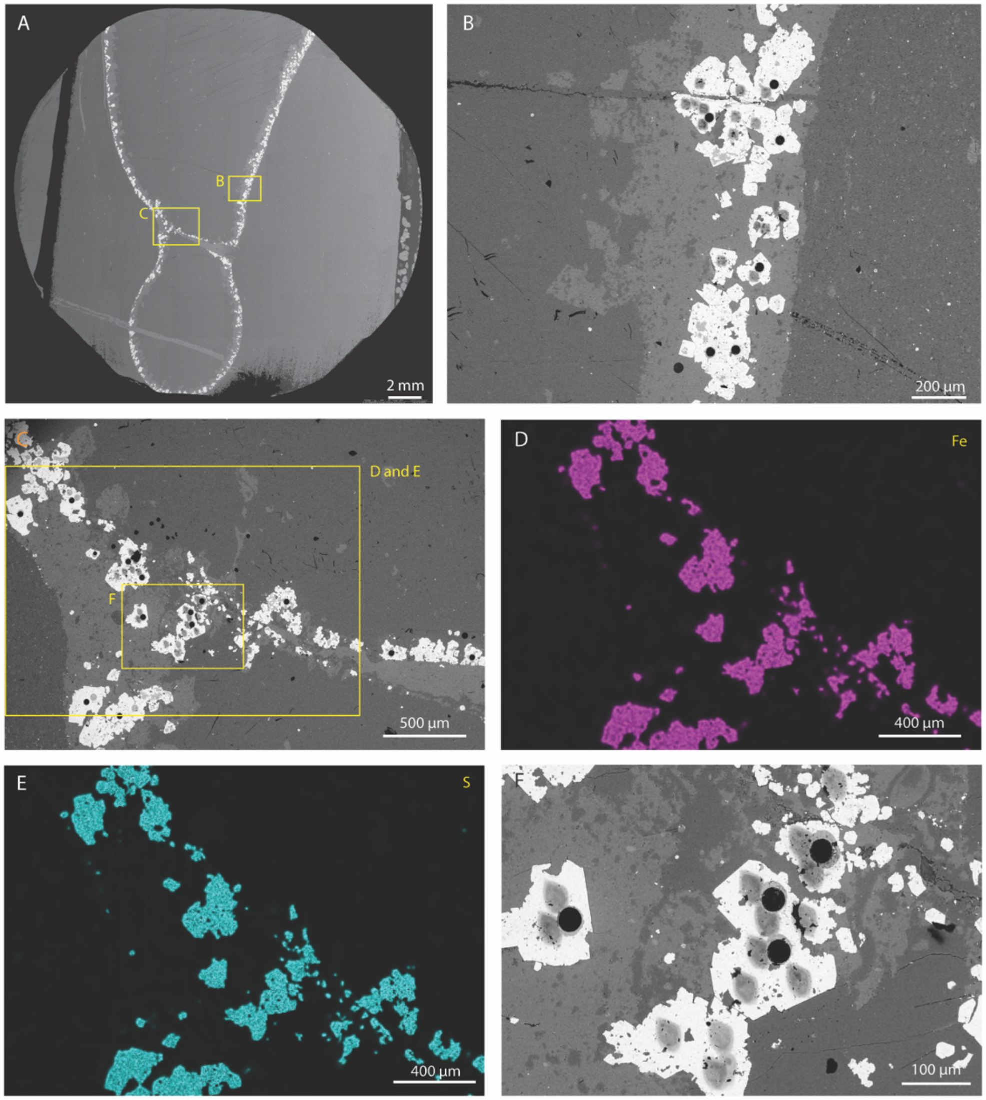


Fig. 3. S5. SEM images and elemental maps of JLW15.1.

Pyritic nodule rim in JLW15.1. (A), (D), and (E) are from Xiao et al. (*11*). (B), (C), and (F) are from this study. Black dots and dark gray shades in SEM images are, respectively, LA-ICP-MS and SIMS analytical spots. Labeled rectangles mark areas of magnified views.

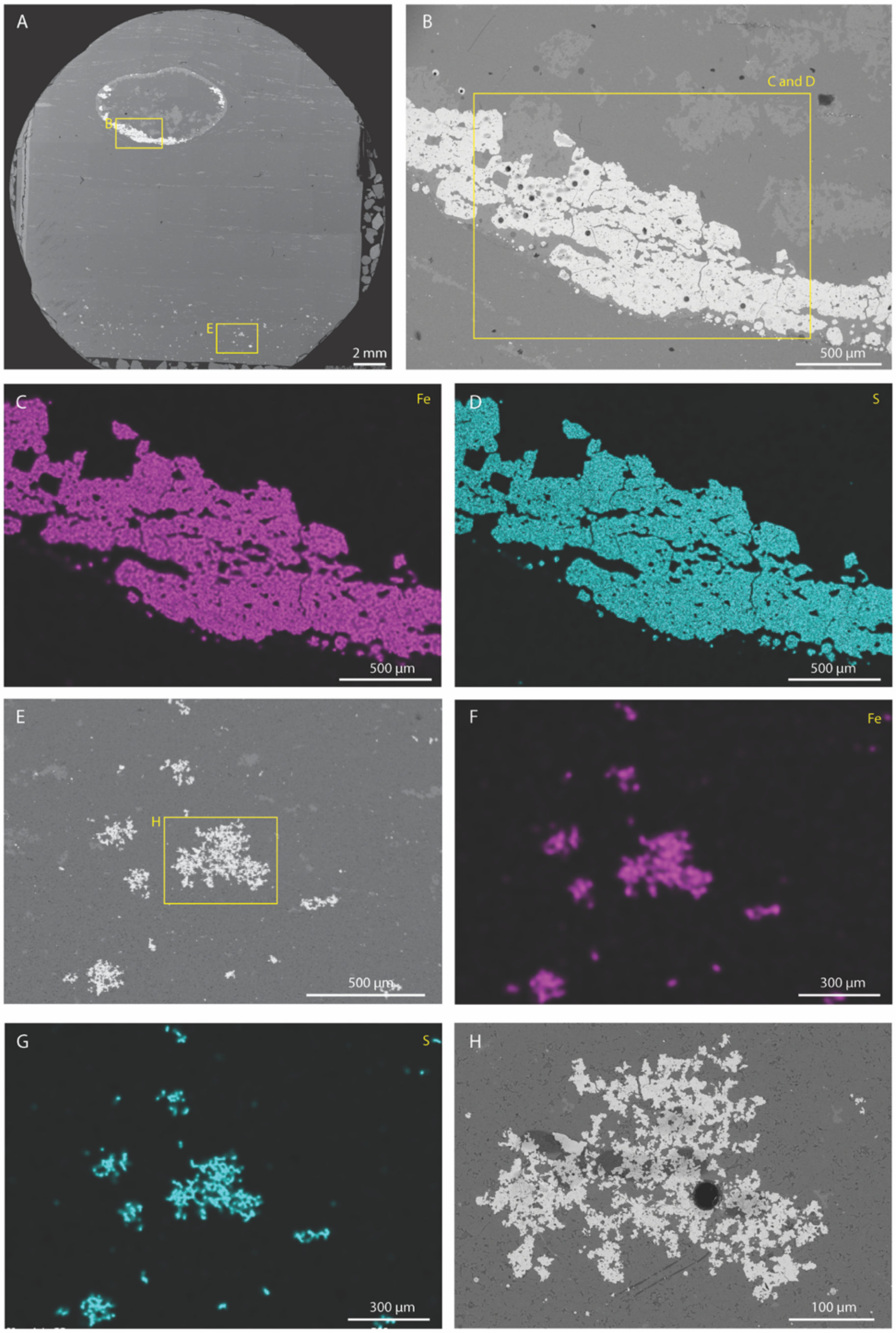


Fig. 3. S6. SEM images and elemental maps of JLW23.5.

Pyritic nodule rim (B–D) and disseminated anhedral pyrite (E–H) in JLW23.5. (A) and (C–G) are from Xiao et al. (*11*). B and H are from this study. Black dots and dark gray shades in SEM images are, respectively, LA-ICP-MS and SIMS analytical spots. Labeled rectangles mark areas of magnified views.

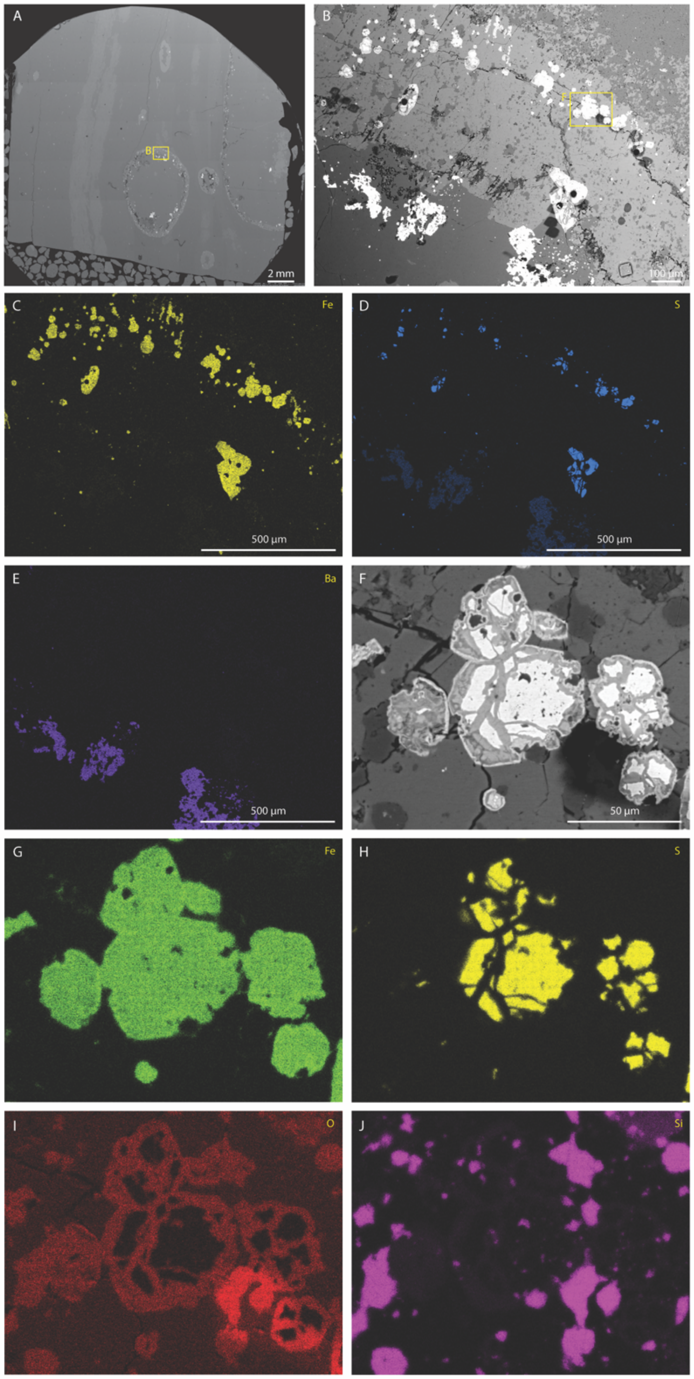


Fig. 3. S7. SEM images and elemental maps of JLW34.8.

Pyritic nodule rim in JLW34.8. Pyrite was partially oxidized (F–J). A is from Xiao et al. (*11*). (B–J) are from this study. Black dots and dark gray shades in SEM images are, respectively, LA-ICP-MS and SIMS analytical spots. (C–E) and (G–J) are elemental maps, with elements marked in the upper right. Labeled rectangles mark areas of magnified views.

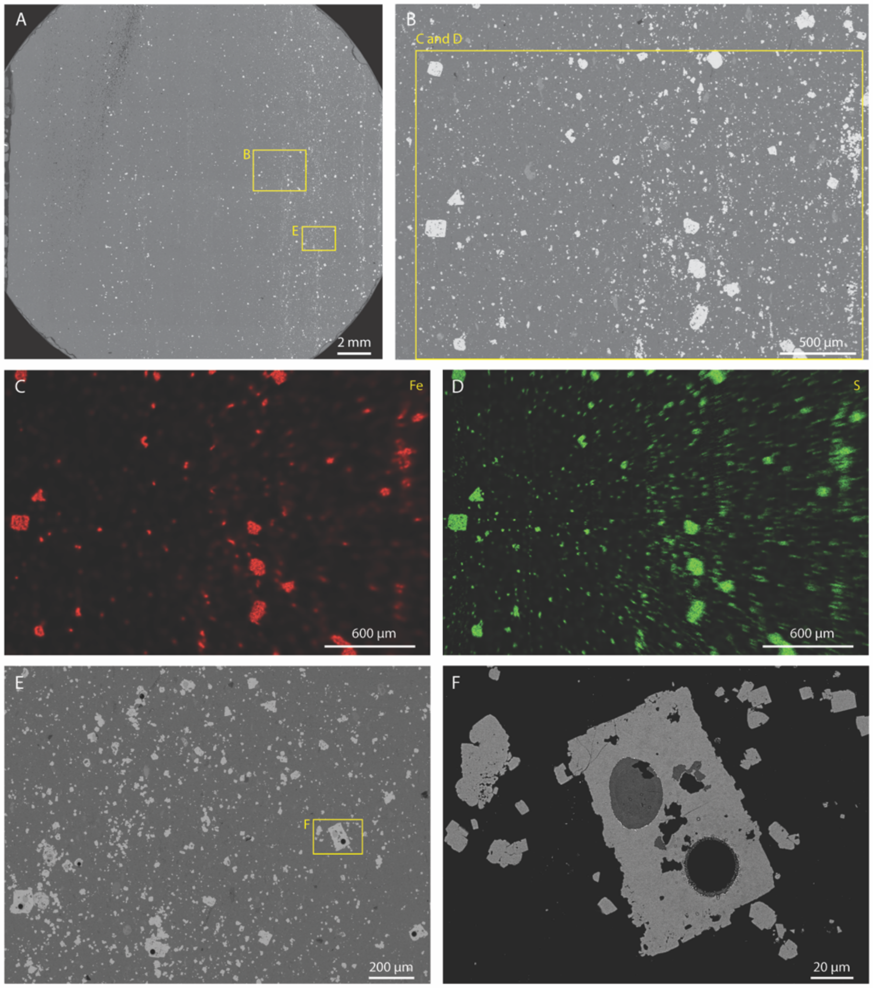


Fig. 3. S8. SEM images and elemental maps of JLW76.

Disseminated euhedral and subhedral pyrite in JLW76. (A–D) are from Xiao et al. (*11*). (E) and (F) are from this study. Black dots and dark gray shades in SEM images are, respectively, LA-ICP-MS and SIMS analytical spots. Labeled rectangles mark areas of magnified views.

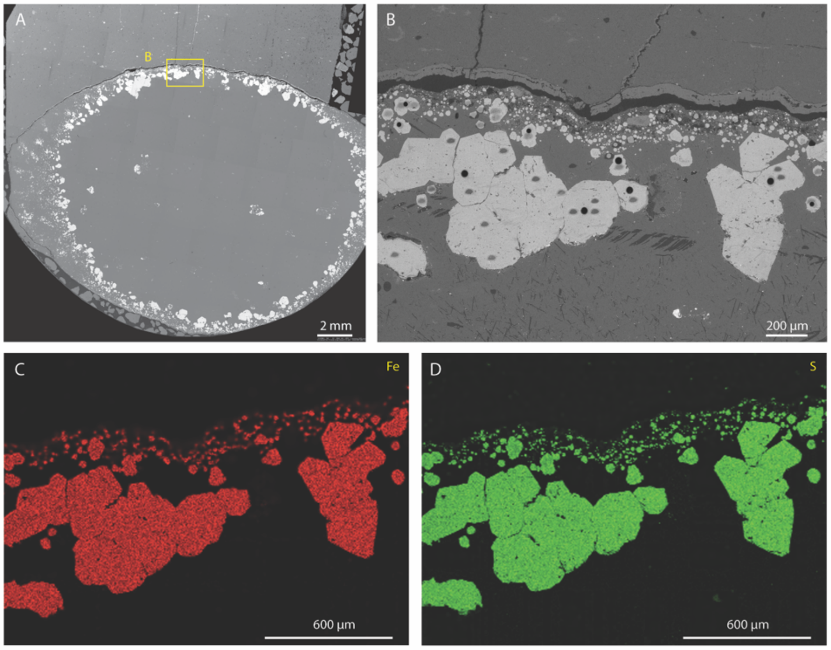
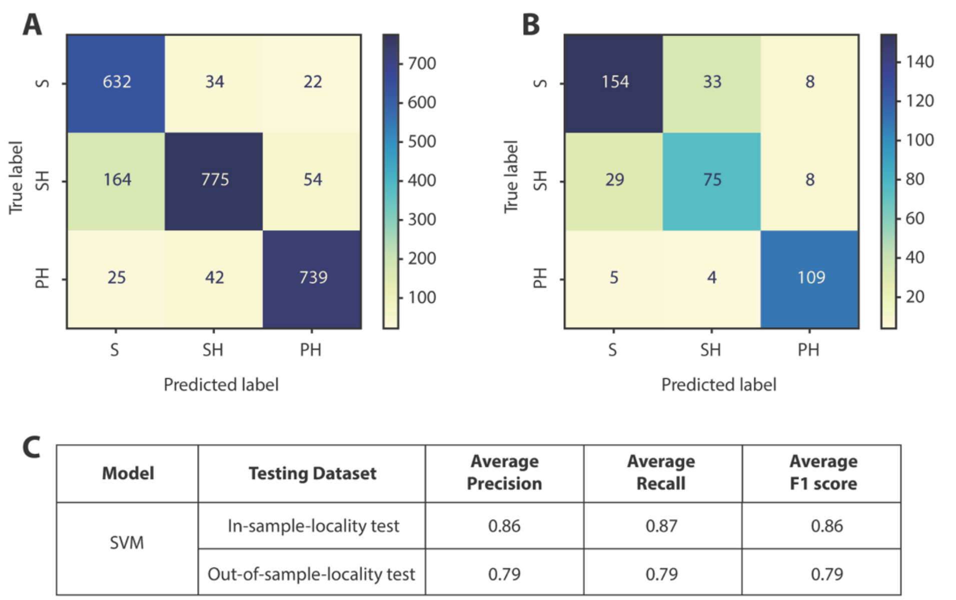


Fig. 3. S9. SEM images and elemental maps of SX88.15.

Pyritic nodule rim in SX88.15. (A), (C), and (D) are from Xiao et al. (*11*). (B) is from this study. Black dots and dark gray shades in (B) are, respectively, LA-ICP-MS and SIMS analytical spots. Labeled rectangle in (A) marks area magnified in (B).



Model	Testing Dataset	Average Precision	Average Recall	Average F1 score
SVM	In-sample-locality test	0.86	0.87	0.86
	Out-of-sample-locality test	0.79	0.79	0.79

Fig. 3. S10. Support Vector Machine (SVM) model performance, with Au and Te included in the analysis.

(A) and (B) are confusion matrixes of the SVM model on in-sample-locality test and out-of-sample-locality test datasets, respectively. Confusion matrixes visualize model performance on identifying sedimentary (S), synsedimentary hydrothermal (SH), and post-sedimentary hydrothermal (PH) pyrite. (C) shows simple arithmetic averages of precision, recall, and F1 score across labels (sedimentary, synsedimentary hydrothermal, post-sedimentary hydrothermal) on in-sample-locality test and out-of-sample-locality test datasets

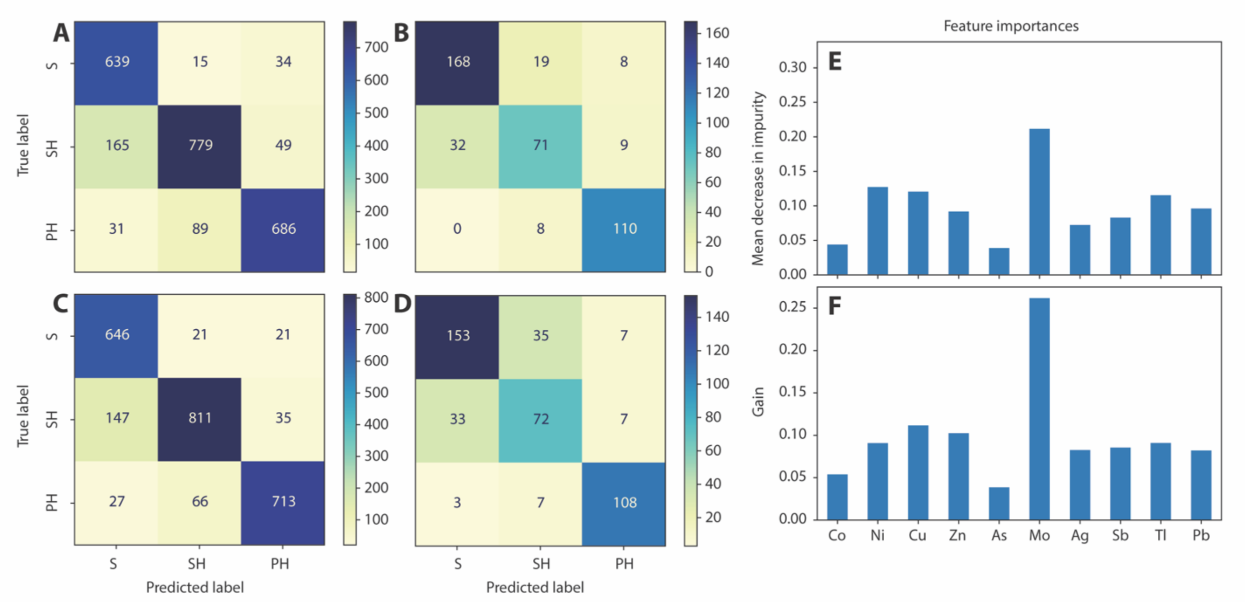


Fig. 3. S11. Supervised machine learning model confusion matrix and feature importance, with Au and Te excluded from the analysis.

(A) and (B) are confusion matrixes of the Random Forests model on in-sample-locality test and out-of-sample-locality test datasets, respectively. (C) and (D) are confusion matrixes of the XGBoost model on in-sample-locality test and out-of-sample-locality test datasets, respectively. Confusion matrixes visualize model performance on identifying sedimentary (S), synsedimentary hydrothermal (SH), and post-sedimentary hydrothermal (PH) pyrite. (E) and (F) show the feature importance of the Random Forests and XGBoost models, respectively. For Random Forests, the feature importance is decided by the mean decrease in impurity when a feature is used to build a split. For XGBoost, the feature importance is decided by the gain, which is the improvement in accuracy brought by a feature.

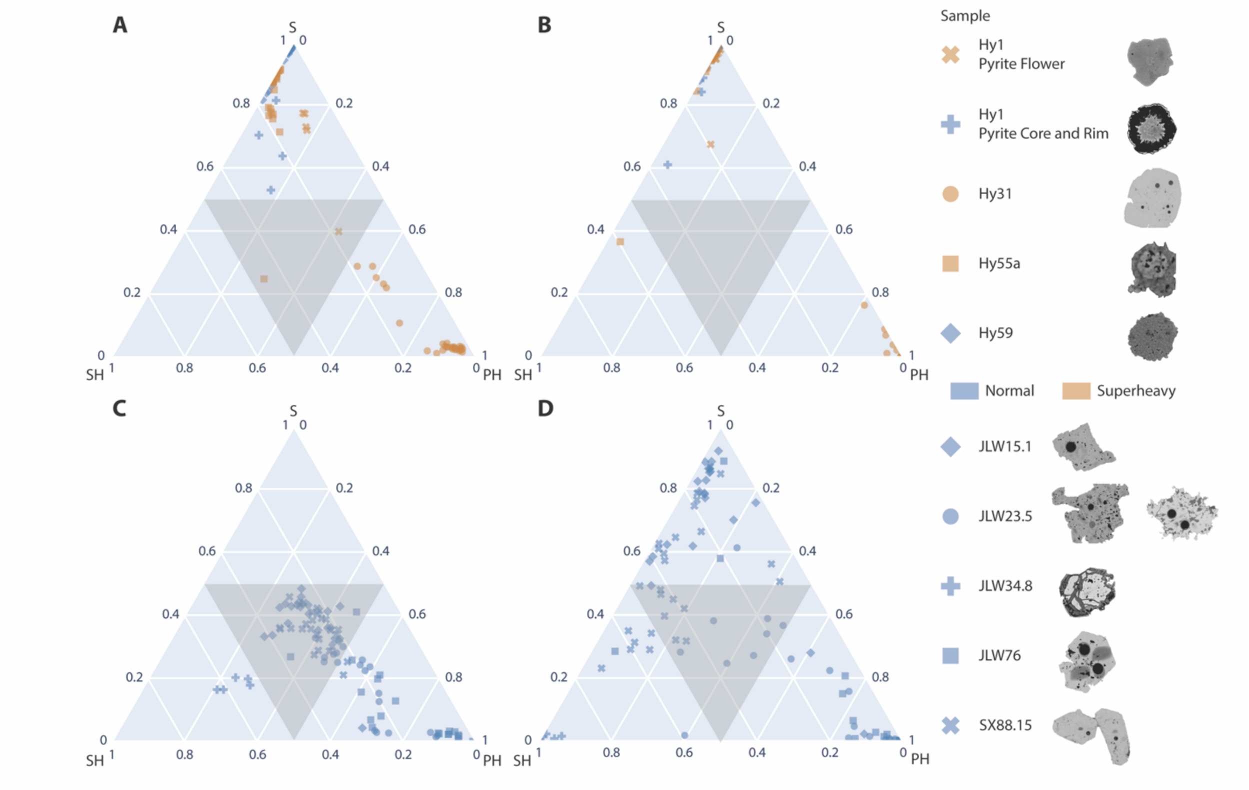


Fig. 3. S12. Ternary plots of model predictions on Hy, JLW, and SX samples, with Au and Te excluded from the analysis.

Model predictions on pyrite analyses of Hy, JLW, and SX samples are expressed as the probability of being identified as sedimentary (S), synsedimentary hydrothermal (SH), or post-sedimentary hydrothermal (PH) pyrite. Each data point represents one single LA-ICP-MS analysis. (A) and (B) are model outputs of Hy samples based on the Random Forests (A) and XGBoost (B) models. (C) and (D) are model outputs of JLW and SX samples based on the Random Forests (C) and XGBoost (D) models. Blue data points represent isotopically normal pyrite and brown data points represent isotopically superheavy pyrite. Superheavy pyrite is defined by $\delta^{34}S > 50‰$, which is substantially higher than that of contemporaneous seawater (*14*). Gray triangular shades in ternary plots represent the ambiguous zone. Pyrite analyses in the ambiguous zone have a less than 50% probability of being identified to any single label.

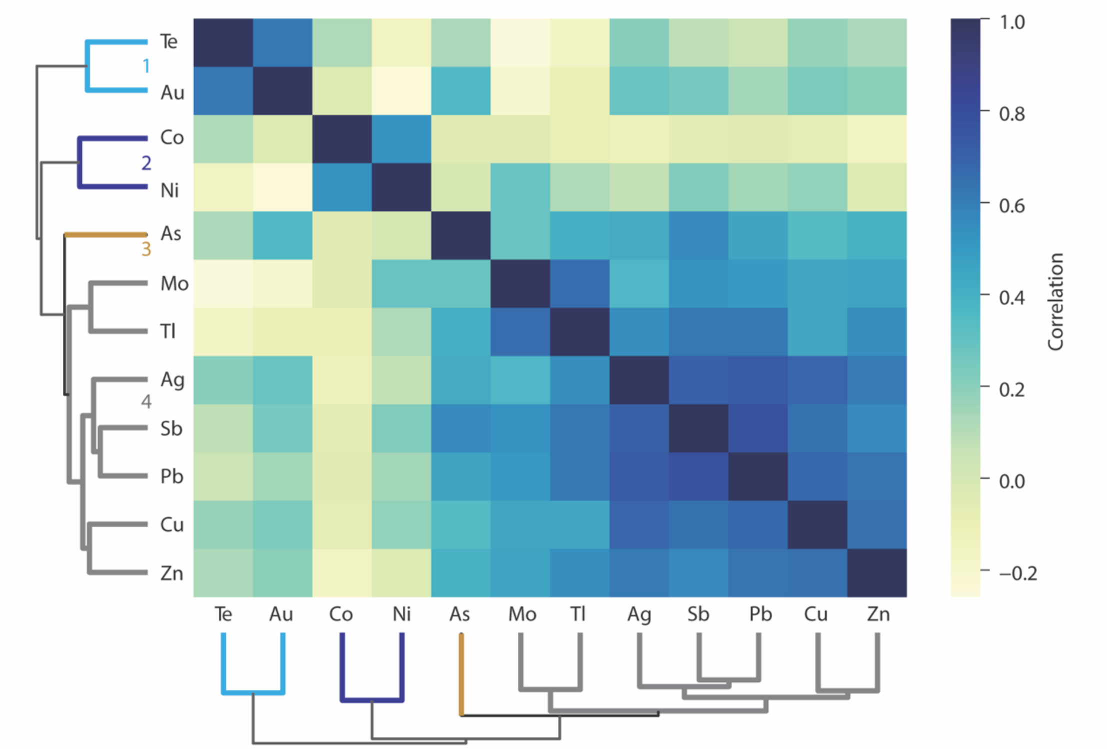


Fig. 3. S13. Correlation matrix of 12 trace elements from the LA-ICP-MS database.

Elements were grouped based on HCA results. Each colored square shows the Pearson's correlation coefficient for a pair of elements.

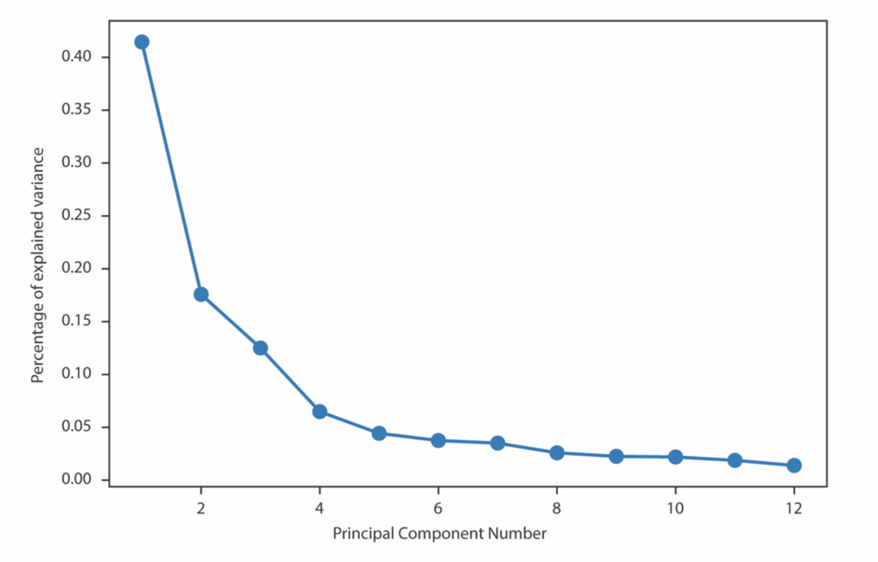


Fig. 3. S14. PCA scree plot.

The plot shows the amount of variation captured by each principal component. The first two principal components (PC1 and PC2) capture 59% of the total variance.

		Unit	**Location**	**Train**	**In-sample-locality test**	**Out-of-sample-locality test**	**References**
Sedimentary	In-sample-locality test	Sedimentary Muderong Shale, Winning Group (125 Ma)	Canarvon Basin, WA, Australia	15	3	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary J3, Oxford (169 Ma)	Volga River, Kazan, Russia	15	14	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Jet Rock, Whitby Mudstone (180 Ma)	Yorkshire, UK	15	13	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Kockatea Shale (252 Ma)	Perth Basin, WA, Australia	15	18	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Woody Island Siltstone, Lower Parmeener Group (297 Ma)	Woody Island, Tas, Australia	15	7	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Linton Formation (310-305 Ma)	Carbondale, USA	15	14	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Yulleroo, Clanmeyer, and Gogo formations (384-355 Ma)	Canning Basin, Australia	15	192	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Canol Formation (387-384 Ma)	Selwyn Basin, Yukon, Canada	15	216	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Alum Shale (490 Ma)	Kallsedet, Western Sweden	15	13	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Salmon River	Arthur River, Tas, Australia	15	1	N/A	Large et al.(*15*);

		Siltstone, Togari Group (540 Ma)					Gregory et al.(*16*)
		Sedimentary lower Kepple Creek Formation, Togari Group (630 Ma)	Smithton, Tas, Australia	15	33	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Doushantuo Formation (635-551 Ma)	Guizhou Province, China	15	1	N/A	Gregory et al.(*17*)
		Sedimentary Aralka Formation (657-630 Ma)	Armadeus Basin, NT, Australia	15	7	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Satka Formation, Burzyun Group (1550 Ma)	Satka, South Urals, Russia	15	14	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Curnamona, NSW, Australia (1690 Ma)	Curnamona, NSW, Australia	15	22	N/A	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Hamersley Group, Mount Bruce Supergroup (2572-2504 Ma)	Hamersley Basin, WA, Australia	15	120	N/A	Large et al.(*15*); Gregory et al.(*16*)
	Out-of-sample-locality test	Sedimentary Posidonia Formation (180 Ma)	Dotternhausen, Germany	N/A	N/A	20	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Upper D unit, Brigadier Formation (202 Ma)	NW Shelf, WA, Australia	N/A	N/A	15	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Railway Shale (late Ordovician)	Bendigo, Vic, Australia	N/A	N/A	15	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Benjamin Limestone, Gordon Group (450 Ma)	Florentine Valley, Tas, Australia	N/A	N/A	13	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Que River Shale, Mt	Bulgobac River, Tas, Australia	N/A	N/A	14	Large et al.(*15*);

		Charter Group (505 Ma)					Gregory et al.(*16*)
		Sedimentary Pertatataka Formation (635-582 Ma)	Armadeus Basin, NT, Australia	N/A	N/A	9	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Broadhurst (upper) Formation, Throssell Group (860 Ma)	Yeneena Basin, WA, Australia	N/A	N/A	15	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Cowrie Siltstone, Rocky Cape Group (1290 Ma)	Tas, Australia	N/A	N/A	28	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Valkyrie Formation, Roper River (1360 Ma)	McArthur Basin, NT, Australia	N/A	N/A	9	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Barney Creek Formation (1640 Ma)	McArthur Basin, NT, Australia	N/A	N/A	15	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Upper Callie Member, Dead Bullock Formation, Tanami Group (1840 Ma)	Tanami, NT, Australia	N/A	N/A	25	Large et al.(*15*); Gregory et al.(*16*)
		Sedimentary Johnson Cairn Formation, Yerrida Group (2170 Ma)	N Yilgarn, WA, Australia	N/A	N/A	17	Large et al.(*15*); Gregory et al.(*16*)
Synsedimentary Hydrothermal	In-sample-locality test	SEDEX Don	Don, Canada	24	97	N/A	Gadd et al.(*18*)
		SEDEX HYC	HYC, Australia	24	316	N/A	Maier(*19*)
		SEDEX Lady Loreta	Lady Loretta, Australia	24	25	N/A	Gregory et al.(*20*)
		SEDEX Pelly North	Pelly North, Canada	24	76	N/A	Gadd et al.(*18*)
		SEDEX XYC	XY deposit, Canada	24	163	N/A	Gadd et al.(*18*)

		VMS Golden Grove	Golden Grove, Australia	12	0	N/A	Belousov et al.(*21*)
		VMS Jaguar Mine	Jaguar, Australia	10	0	N/A	Belousov et al.(*21*)
		VMS Kutlular	Kutlular, Turkey	15	6	N/A	Revan et al.(*22*)
		VMS Kyzilkaya	Kyzilkaya, Turkey	13	0	N/A	Revan et al.(*22*)
		VMS Lahanos	Lahanos, Turkey	15	0	N/A	Revan et al.(*22*)
		VMS Scuddles	Scuddles, Australia	9	0	N/A	Belousov et al.(*21*)
		VMS Yaman-Kasy	Yaman-Kasy deposit, Russia	46	310	N/A	Maslennikov et al.(*23, 24*)
	Out-of-sample-locality test	SEDEX Anniv Central	Anniversary deposit central, Canada	N/A	N/A	44	Gadd et al.(*18*)
		SEDEX Anniv East	Anniversary deposit east, Canada	N/A	N/A	15	Gadd et al.(*18*)
		SEDEX OP	OP, Canada	N/A	N/A	7	Gadd et al.(*18*)
		VMS Chaely	Chaely deposit, Turkey	N/A	N/A	4	Revan et al.(*22*)
		VMS DeGrussa	DeGrussa, Australia	N/A	N/A	32	Belousov et al.(*21*)
		VMS Kilik	Kilik, Ural, Turkey	N/A	N/A	10	Revan et al.(*22*)
Post-sedimentary Hydrothermal	In-sample-locality test	Hydrothermal breccia Menninnie Dam	Menninnie Dam, SA, Australia	30	86	N/A	Gregory et al.(*25*)
		Hydrothermal breccia Telephone Dam	Telephone Dam, WA, Australia	30	46	N/A	Gregory et al.(*25*)
		IOCG Manxman	Manxman, Australia	30	53	N/A	Gregory et al.(*25*)
		IOCG Punt Hill	Punt Hill, Australia	30	7	N/A	Gregory et al.(*25*)
		Orogenic Darlot	Darlot, Australia	4	9	N/A	Belousov et al.(*21*)

		Orogenic East Repulse	East Repulse, Australia	4	41	N/A	Gregory et al.(*26*)
		Orogenic Fortnum	Fortnum, Australia	4	26	N/A	Belousov et al.(*21*)
		Orogenic Golden Mile	Golden Mile, Australia	4	39	N/A	Belousov et al.(*21*)
		Orogenic Granny Smith	Granny Smith, Australia	4	8	N/A	Belousov et al.(*21*)
		Orogenic Lancefield	Lancefield, Australia	4	6	N/A	Belousov et al.(*21*)
		Orogenic Mars Pit	Mars, Australia	4	13	N/A	Belousov et al.(*21*)
		Orogenic Meekatharra Prohibition	Meekathara, Prohibition, Australia	4	6	N/A	Belousov et al.(*21*)
		Orogenic Micky Doolan Meekatharra	Meekathara, Micky Doolan, Australia	4	6	N/A	Belousov et al.(*21*)
		Orogenic Minjar	Minjar, Australia	4	6	N/A	Belousov et al.(*21*)
		Orogenic Nathans Labouchere	Nathans Labouchere, Australia	4	7	N/A	Belousov et al.(*21*)
		Orogenic Paddington	Paddington, Western Australia	4	6	N/A	Belousov et al.(*21*)
		Orogenic Songvang	Songvang, Australia	4	9	N/A	Belousov et al.(*21*)
		Orogenic Sunrise Dam	Sunrise Dam, Australia	4	8	N/A	Belousov et al.(*21*)
		Orogenic Victory	Victory, Australia	4	68	N/A	Gregory et al.(*26*)
		Porphyry Chalkidiki	Chalkidiki, Greece	30	70	N/A	Gregory et al.(*20*)
		Porphyry Cadia	Cadia, Australia	30	286	N/A	Gregory et al.(*20*)
	Out-of-sample-locality test	Orogenic Hill50	Hill 50, Australia	N/A	N/A	22	Belousov et al.(*21*)
		Orogenic Wallaby	Wallaby, Australia	N/A	N/A	23	Belousov et al.(*21*)

		Orogenic Wiluna	Wiluna, Australia	N/A	N/A	62	Belousov et al.(*21*)
		Orogenic Youanmi	Youanmi, Australia	N/A	N/A	11	Belousov et al.(*21*)

Table 3. S1. Summary of LA-ICP-MS pyrite trace element database.

Model	Testing Dataset	Average Precision	Average Recall	Average F1 score
Random Forests	In-sample-locality test	0.85	0.85	0.85
	Out-of-sample-locality test	0.81	0.81	0.81
XGBoost	In-sample-locality test	0.87	0.88	0.87
	Out-of-sample-locality test	0.78	0.78	0.78

Table 3. S2. Supervised machine learning model evaluation, with Au and Te excluded from the analysis.

Simple arithmetic averages of precision, recall, and F1 score across labels (sedimentary, synsedimentary hydrothermal, post-sedimentary hydrothermal) on in-sample-locality test and out-of-sample-locality test datasets are shown in this table. For each label, the precision is calculated by dividing the true positives by the sum of true positives and false positives. The recall is calculated by dividing the true positives by the sum of true positives and false negatives. The F1 score is the harmonic mean of precision and recall.

References

1. L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, Classification and regression trees. Belmont, CA: Wadsworth. *International Group* **432**, 9 (1984).
2. L. Breiman, Random forests. *Machine learning* **45**, 5-32 (2001).
3. P.-N. Tan, M. Steinbach, A. Karpatne, V. Kumar, *Introduction to Data Mining (2nd Edition)*. (Pearson, New York, NY, ed. 2nd, 2018).
4. T. Hastie, R. Tibshirani, J. H. Friedman, J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. (Springer, 2009), vol. 2.
5. T. Chen, C. Guestrin, paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016.
6. H. Cui *et al.*, Questioning the biogenicity of Neoproterozoic superheavy pyrite by SIMS. *American Mineralogist: Journal of Earth and Planetary Materials* **103**, 1362-1400 (2018).
7. C. Wu *et al.*, Nanhuan manganese deposits within restricted basins of the southeastern Yangtze Platform, China: Constraints from geological and geochemical evidence. *Ore Geology Reviews* **75**, 76-99 (2016).
8. C. Zhou *et al.*, New constraints on the ages of Neoproterozoic glaciations in South China. *Geology* **32**, 437-440 (2004).
9. C. Yin *et al.*, SHRIMP II U-Pb zircon date from the Nanhuan Datangpo formation in Songtao county, Guizhou Province. *Acta Geologica Sinica* **80**, 278-285 (2006).

10. S. Zhang, G. Jiang, Y. Han, The age of the Nantuo Formation and Nantuo glaciation in South China. *Terra Nova* **20**, 289–294 (2008).
11. S. Xiao, J. Schiffbauer, K. McFadden, J. Hunter, Petrographic and SIMS pyrite sulfur isotope analyses of Ediacaran chert nodules: Implications for microbial processes in pyrite rim formation, silicification, and exceptional fossil preservation. *Earth and Planetary Science Letters* **297**, 481-495 (2010).
12. M. Ader *et al.*, A multilayered water column in the Ediacaran Yangtze platform? Insights from carbonate and organic matter paired δ^{13}C. *Earth and Planetary Science Letters* **288**, 213-227 (2009).
13. D. Condon *et al.*, U-Pb ages from the Neoproterozoic Doushantuo Formation, China. *Science* **308**, 95-98 (2005).
14. D. A. Fike, A. S. Bradley, C. V. Rose, Rethinking the ancient sulfur cycle. *Annual Review of Earth and Planetary Sciences* **43**, 593-622 (2015).
15. R. R. Large *et al.*, Trace element content of sedimentary pyrite as a new proxy for deep-time ocean–atmosphere evolution. *Earth and Planetary Science Letters* **389**, 209-220 (2014).
16. D. D. Gregory *et al.*, Trace element content of sedimentary pyrite in black shales. *Economic Geology* **110**, 1389-1410 (2015).
17. D. D. Gregory *et al.*, Whole rock and discrete pyrite geochemistry as complementary tracers of ancient ocean chemistry: an example from the Neoproterozoic Doushantuo Formation, China. *Geochimica et Cosmochimica Acta* **216**, 201-220 (2017).
18. M. G. Gadd, D. Layton-Matthews, J. M. Peter, S. J. Paradis, The world-class Howard's Pass SEDEX Zn-Pb district, Selwyn Basin, Yukon. Part I: trace element compositions of pyrite record input of hydrothermal, diagenetic, and metamorphic fluids to mineralization. *Mineralium Deposita* **51**, 319-342 (2016).
19. R. C. Maier, University of Tasmania, (2011).
20. D. D. Gregory *et al.*, Distinguishing ore deposit type and barren sedimentary pyrite using laser ablation-inductively coupled plasma-mass spectrometry trace element data and statistical analysis of large data sets. *Economic Geology* **114**, 771-786 (2019).
21. I. Belousov *et al.*, Pyrite compositions from VHMS and orogenic Au deposits in the Yilgarn Craton, Western Australia: Implications for gold and copper exploration. *Ore Geology Reviews* **79**, 474-499 (2016).
22. M. K. Revan *et al.*, Mineralogy and trace-element geochemistry of sulfide minerals in hydrothermal chimneys from the Upper-Cretaceous VMS deposits of the eastern Pontide orogenic belt (NE Turkey). *Ore Geology Reviews* **63**, 129-149 (2014).
23. V. Maslennikov *et al.*, Chimneys in Paleozoic massive sulfide mounds of the Urals VMS deposits: Mineral and trace element comparison with modern black, grey, white and clear smokers. *Ore Geology Reviews* **85**, 64-106 (2017).
24. V. Maslennikov, S. Maslennikova, R. Large, L. Danyushevsky, Study of trace element zonation in vent chimneys from the Silurian Yaman-Kasy volcanic-hosted massive sulfide deposit (Southern Urals, Russia) using laser ablation-inductively coupled plasma mass spectrometry (LA-ICPMS). *Economic Geology* **104**, 1111-1141 (2009).
25. D. D. Gregory, S. Meffre, R. R. Large, M. Cracknell, S. Kuhn, "Annual report: South Australian pyrite, hematite and magnetite fingerprint database," (Department for Manufacturing, Innovation, Trade, Resources and Energy, Hobart, Australia, 2015).

26. D. D. Gregory *et al.*, Trace element content of pyrite from the kapai slate, St. Ives Gold District, Western Australia. *Economic Geology* **111**, 1297-1320 (2016).