University of Southampton

Faculty of Environmental and Life Sciences

School of Ocean and Earth Science

New methodologies to reduce measurement bias in high-throughput LA-ICP-MS and X-ray CT, and applications to planktonic foraminifera

by

Alex Searle-Barnes

ORCID ID 0000-0003-0389-7717

Thesis for the degree of Doctor of Philosophy

July 2025

University of Southampton

Abstract

Faculty of Environmental and Life Sciences

School of Ocean and Earth Science

Member of Staff for the Degree of Doctor of Philosophy by Publication

A Commentary of Published Work

by

Alex Searle-Barnes

The work I present here has been published in peer-reviewed journals in accordance to the Member of Staff for the Degree of Doctor of Philosophy by publication pathway.

This commentary presents advancements in the study of foraminifera by integrating high-resolution imaging, innovative data processing methodologies, and geochemical analysis to investigate environmental, ecological, and evolutionary dynamics preserved in marine sediment cores. I present ten recommendations for X-ray microfocus computed tomography (µCT) imaging to enhance the volumetric analysis of individual foraminifera, enabling precise characterisation of internal structures such as chamber volumes, shapes, and pore distributions. These structures form the basis for linking morphology to environmental gradients and understanding species-specific adaptations to oxygen minimum zones.

Geochemical analysis using laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) revealed ontogenetic trends in chamber wall thickness, demonstrating that earlier chambers are thicker due to secondary calcite overgrowth and more representative of environmental conditions compared to the penultimate and final chambers. Results suggest that Mg/Ca ratios in earlier chambers correlate strongly with $\delta^{18}O$, offering more reliable palaeoclimate reconstructions.

I present the LABLASTER package for the R statistical environment to automate data reduction workflows, improving accuracy, transparency, and reproducibility in geochemical analyses. By algorithmically detecting laser ablation endpoints in time-resolved LA-ICP-MS data, LABLASTER ensures consistent targeting of biologically relevant signals, facilitating more precise extraction of geochemical data from heterogenous bioarchives. These methodological advances establish a robust framework for studying morphological and geochemical signals in foraminifera and other carbonate-shelled organisms, providing new insights into past environments and evolutionary dynamics.

Table of Contents

Table of Tables

Table of Figures

Bibliography of Published Publications

1. Brombacher, A., Searle-Barnes, A., Mulqueeney, J. M., Standish, C. D., Milton, J. A., Katsamenis, O., Watson, R. A., Trueman, C. N., Sinclair, I., Wilson, P. A., Foster, G. L. & Ezard, T. H. G. 2025. Detecting environmentally dependent developmental plasticity in the fossil record. *Proceedings of the National Academy of Sciences*., 112, e2421549122. https://doi.org/10.1073/pnas.2421549122

2. Brombacher, A., Searle-Barnes, A., Zhang, W. & Ezard, T. H. G. 2022. Analysing planktonic foraminiferal growth in three dimensions with foram3D: an R package for automated trait measurements from CT scans. *J. Micropalaeontol.,* 41, 149-164. https://doi.org/10.5194/jm-41-149-2022

3. Kearns, L. E., Searle-Barnes, A., Foster, G. L., Milton, J. A., Standish, C. D. & Ezard, T. H. G. 2023. The Influence of Geochemical Variation Among *Globigerinoides ruber* Individuals on Paleoceanographic Reconstructions. *Paleoceanography and Paleoclimatology,* 38, e2022PA004549. https://doi.org/10.1029/2022pa004549

4. Lin, H., Zhang, W., Mulqueeney, J. M., Brombacher, A., Searle-Barnes, A., Nixon, M., Cai, X. & Ezard, T. H. G. 2024. 3DKMI: A MATLAB package to generate shape signatures from Krawtchouk moments and an application to species delimitation in planktonic foraminifera. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210x.14388

5. Mulqueeney, J. M., Searle-Barnes, A., Brombacher, A., Sweeney, M., Goswami, A. & Ezard, T. H. 2024. How many specimens make a sufficient training set for automated three-dimensional feature extraction? *Royal Society Open Science,* 11, rsos. 240113. https://doi.org/10.1098/rsos.240113

6. Searle-Barnes, A., Brombacher, A., Katsamenis, O. L., Rankin, K., Mavrogordato, M. & Ezard, T. H. G. 2025. Ten recommendations for scanning foraminifera by X-ray computed tomography, *Journal of Micropalaeontology*, 44, 107-117. https://doi.org/10.5194/jm-44-107-2025

7. Searle-Barnes, A., Milton, J. A., Standish, C. D., Foster, G. L. & Ezard, T. H. G. 2023. Laser ablation mass spectrometry blast through detection in R. *Rapid Communications in Mass Spectrometry,* 37, e9533. https://doi.org/10.1002/rcm.9533

8. Zhang, W., Ezard, T., Searle-Barnes, A., Brombacher, A., Katsamenis, O. & Nixon, M. Towards understanding speciation by automated extraction and description of 3D foraminifera stacks. 2020 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), 2020. IEEE, 30-33. http://dx.doi.org/10.1109/SSIAI49293.2020.9094611

Research Thesis: Declaration of Authorship

Print name: Alex Searle-Barnes

Title of thesis: A Commentary of Published Work

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as described in the Bibliography of Published Publications.

Signature: ..... Date: .....

Acknowledgements

I would like to thank everyone who has supported me throughout the seven years of my research both academically and socially.

First, I would like to thank Thomas Ezard. From the start you embedded me as a research technician into your group, you guided me through my PhD and we navigated this PhD by publication pathway, and your unwavering support has been instrumental to my journey. Your invaluable guidance, expertise, and encouragement have shaped my development as a scientist, and your mentorship has allowed me to flourish as an independent researcher. Thank you for your endless enthusiasm and encouragement to keep going, for the opportunities you opened to share this work with others at conferences and demonstrating to the next generations. As the principle with PISTON, you provided me with the employment and financial stability made this research possible – a rare benefit.

To Anieke for your warm welcoming and patience while I learn everything I know about foraminifera. I was a complete novice, and now I present this work with your solid foundations. For your guidance and encouragement throughout, I am deeply grateful. Your advice and your willingness to share your knowledge has been a cornerstone of my development as a scientist. Thank you to my PISTON colleagues who contributed to the very multidisciplinary nature and success of this project.

To Gavin Foster and Paul Wilson for your technical expertise and explanations have been integral to the progress of my research.

To the entire 2017 – 2024 Palaeoclimate and Palaeoceanography group for your company, lunchtime discussions, and even more pub-time discussions. We have travelled across continents as friends and scientists, sharing an equal amount of wine and laughter, as well as inspiration that have shaped me as a researcher. Your companionship has made this journey unforgettable, and I am deeply grateful for the moments we've shared both inside and outside the lab. A special thank you to Chloe Jones, James Mulqueeney, Tereza Kunkelova, Alex Nicol-Harper, Lorna Kearns, Thomas Chalk, and Elwyn de la Vega.

Thank you to Andy Milton, Bastian Hambach, Megan Wilding, Orestis Katsamenis, Katy Rankin, and Matt Cooper for your laboratory support, and the many weeks shared with you in the lab and keeping everything running.

I wish to thank my mum, dad, and family for their consistent and unwavering support. For providing me a home, a place to rest, and plenty of distractions. To Julie Seager-Smith for your company and rescue pickup from Gatwick Airport from Forams 2023 conference.

And finally, thank you Jack Kitchen, for sharing it all with me.

Definitions and Abbreviations

µCT	X-ray microfocus computed tomography
LABLASTER	Laser Ablation BLASt Through Endpoint in R. An open-source software package for the R environment for statistical computing
LA-ICP-MS	Laser ablation inductively coupled plasma mass spectrometry
Mg/Ca	Magnesium in mmol divided by calcium in mol
mol	SI unit of mole
mmol	One thousandth of the SI unit of moles
SIRMS	Stable isotope ratio mass spectrometry
$\delta^{13}C$	A measure of the ratio of the stable isotopes carbon-13 to carbon-12 in a sample, expressed in per mil (‰) relative to the Vienna Pee Dee Belemnite (VPDB) standard.
$\delta^{18}O$	A measure of the ratio of the stable isotopes oxygen-18 to oxygen-16 in a sample, expressed in per mil (‰) relative to the Vienna Pee Dee Belemnite (VPDB) standard.

Chapter 1 Introduction

Planktonic foraminifera are unicellular zooplankton distributed throughout the world's oceans as a key resource in understanding the Earth's climate system. Foraminifera grow connected chambers throughout their life with those in the final whorl accessible for trace element analysis. The thickness of a foraminifera chamber wall can vary substantially even within an individual as some have thick solid chamber walls, while others have a highly porous structure as a result of species-level morphological variations, biological controls and environmental influences.

Fossilised foraminifera tests are often recovered from drilling into ocean floor sediments, lake beds and the earth's crust, giving plenty of opportunities for the mixing of fossil specimens from multiple time intervals into a single assemblage, which can obscure fine-scale temporal patterns and blur signals of evolutionary, ecological, and environmental change (Hunt, 2004). Traditional approaches to fossilised foraminiferal analysis often rely on bulk sample measurements that average geochemical or morphological signals across many individuals and time intervals. These averaging masks the ontogenetic, ecological, and evolutionary nuances that are important factors that influence the incorporation of trace elements into the test during growth that we use to infer life histories and past climatic conditions. New studies often compare their results with published literature, creating inconsistencies when compounded averages are used as benchmarks for interpretation (Franke and Donner, 2019). To address these limitations, I avoid averaging across individuals or life stages to mature understandings and to untangle the influences and interactions between biology, environment, and climate and the measured geochemical signals.

I use planktic foraminifera to study this variability of phylogeny and geochemistry between individuals and within species to test ecology and evolution theories and expand palaeoclimate reconstructions in novel ways. Working with high temporal resolution sediment core samples and at individual foraminifera chamber resolution, my analyses reveal taxonomic and geochemical changes occurring on geological and ecological timescales for the first time, bridging the palaeontological-ecological gap in more deeply geochemical ways than ever before. My results are pertinent for answering questions about ecological evolution, earth dynamics and understanding climate change.

Early approaches to studying geochemistry of planktonic foraminifera tests required up-to 400 individuals per stable isotope measurement (Emiliani, 1955) and about ten individuals for trace element measurement (Bender *et al.*, 1975). Such bulk measurements aggregate signals across multiple individuals, averaging out critical differences that could reveal ontogenetic changes, species-specific adaptations, or environmental responses. Since then, analytical instrumentation sensitivity has improved, allowing recent studies to measure both stable isotopes and trace elements using only a single specimen (Anand *et al.*, 2003). Similarly, whole-test analyses integrate geochemical signals from chambers formed at different life stages, obscuring the unique environmental conditions each chamber experienced during its formation. Further instrumentation improvements have allowed me to increase the spatial resolution to repeated measurements of single chambers of individual foraminifers that now reveal the nuanced geochemical signals to accurately resolve fine-scale temporal, ecological, and evolutionary dynamics.

Through my work, I have µCT scanned 3,000 individual *Menardella* (also called *Globorotalia*) lineage planktonic foraminifer from ODP Site 925. This is the world's largest data set of single genus planktonic foraminifer, consisting of 600,000 X-ray CT images, 40 million LA-ICP-MS trace metal data points and 1,250 stable isotope ratios.

Generating such a large dataset necessitated careful planning to ensure study reproducibility and efficient sample throughput, all within the constraints of laboratory availability. Equally important was the foresight to maximise the future uses of this dataset through open-access archiving and publication. These practical considerations drove the development of novel methodological improvements I present in my publications. I developed streamlined sample mounting techniques, advanced data processing algorithms, and automation tools to enhance sample throughput, ensure consistent reproducibility, and reduce analytical uncertainty.

Together, these advancements mature and progress our understanding of foraminifera life history, developmental plasticity, evolutionary changes, and how the relationships between embedded geochemical signals and climate change vary. Applying these advancements to such a large dataset enhances the explanatory power of mathematical models, providing confidence in their ability to uncover patterns and mechanisms driving ecological and environmental change. By integrating advanced imaging techniques, algorithmic data

processing, and individual chamber resolution analyses, my work pushes the boundaries of traditional foraminifera research.

Here, I present a portfolio of peer-reviewed published manuscripts in accordance with the member of staff PhD submission by published work route. These published papers take a multi-proxy approach bridging palaeontological, ecological, geochemical, and computer science disciplines to untangle the compounded environmental and biological signals in foraminifera shells to address challenges around evolution and climate change. My portfolio documents my novel analytical developments, and through those methodological refinements, the new insights and interpretations I have been able to draw on how to extract biological "vital effects" from faithful recording of the Earth System.

All the code discussed here, and further data reduction stages were written in Bash, Batch, and R and are available to download from my GitHub[1]. On the conclusion of the NERC large grant research project *PISTON: Does Developmental Plasticity Influence Speciation?*[2], I will make all data collected available at the British Oceanographic Data Centre[3].

[1] Alex Searle-Barnes software repository GitHub https://alexsb1.github.io
[2] PISTON: Does Developmental Plasticity Influence Speciation? https://www.southampton.ac.uk/oes/research/projects/piston-does-developmental-plasticity-influence-speciation.page
[3] British Oceanographic Data Centre repository https://www.bodc.ac.uk

Chapter 2 Nature of the Research

Individual analysis of fossilised planktic foraminifera tests using X-ray computed tomography (µCT), laser ablation inductively coupled plasma mass spectrometry, stable isotope ratio mass spectrometry, and all combinations thereof has become a popular technique across multiple earth- and life-science researchers (Elderfield and Ganssen, 2000; Elderfield *et al.*, 2002; Eggins *et al.*, 2003; Johnstone *et al.*, 2010; Katz *et al.*, 2010; Pearson, 2012; Schmidt *et al.*, 2013; Fehrenbacher *et al.*, 2015; Spero *et al.*, 2015; Caromel *et al.*, 2016; Hsiang *et al.*, 2019; Burke *et al.*, 2020; Zhang *et al.*, 2020; Kearns *et al.*, 2023; Fehrenbacher *et al.*, 2024; Jana *et al.*, 2024; Mulqueeney *et al.*, 2024).

These techniques directly measure the chemistry and morphology of the foraminifer test, which are all related to indirect measurements of the environmental and biological factors experienced during its lifetime (Spero *et al.*, 1997; Elderfield and Ganssen, 2000; Eggins *et al.*, 2003; Fehrenbacher *et al.*, 2024). The results from these proxy measurements reveal insights into oceanic temperatures, ice volume changes, ecological niches, and species-specific vital effects, advancing reconstructions of paleoenvironmental conditions and the evolutionary history of marine organisms (Shackleton, 1967; Zachos *et al.*, 2001; Rosenthal and Lohmann, 2002; Anand *et al.*, 2003; Barker *et al.*, 2005; Regenberg *et al.*, 2009; Johnstone *et al.*, 2011; Pearson, 2012; Schmidt *et al.*, 2013; Kendall *et al.*, 2020; Rosenthal *et al.*, 2022).

Given the astonishing insight these state-of-the-art techniques provide, there is an opportunity for researchers to reimagine how we design sampling protocols to unlock new scientific questions and thus research horizons. By integrating variables such as ontogenetic stage, geochemical proxies (e.g., Mg/Ca, $\delta^{18}O$), and morphological parameters, and by specifying workflows that combine high-resolution imaging techniques (e.g., X-ray CT) with geochemical analyses (e.g., LA-ICP-MS), I achieved a more nuanced understanding of foraminiferal biology and environmental interactions. Such integrative approaches mature identifying adaptive strategies, speciation mechanisms, and environmental responses across temporal and spatial scales. Here, I focus on the individual-level measurements to examine ontogenetic (developmental) stages, species-specific adaptations, and environmental responses.

2.1 X-ray Computed Tomography

X-ray microfocus computed tomography (µCT) is a popular, non-destructive technique to study internal structures and features that are otherwise hidden from view (Withers *et al.*, 2021). High-quality images and three-dimensional visualisations from µCT scans enable detailed manual and automated analyses of internal morphologies that have long been associated with forensic evolutionary divergence (Huber *et al.*, 1997; de Vargas *et al.*, 1999). X-ray µCT is a tool that reveals detailed internal morphology and structures, such as chamber wall thickness and chamber shapes, that provides additional details to distinguish the true identity of otherwise cryptically similar species based exclusively on external visible morphologic descriptions. Chamber volumes, growth rates, porosity, surface areas, and wall thickness are functional traits that reveal ecological and evolutionary patterns and can be quantified using µCT (Speijer *et al.*, 2008; Briguglio *et al.*, 2011; Schmidt *et al.*, 2013; Brombacher *et al.*, 2022). These traits offer insights into the ontogenetic development of foraminifera, their calcification processes, and adaptations to environmental pressures. Functional traits are an additional tool to distinguish the true identity of otherwise cryptically similar species based exclusively on external visible morphologic descriptions.

X-ray µCT is a three-dimensional (3D) imaging technology that unlocks new horizons for understanding environmental, ecological and evolutionary dynamics by revealing internal views compared to the surface level of traditional light microscopy. High-resolution µCT imaging, 3D reconstruction algorithms and volume data analysis software enable researchers to quantify morphological changes at both intra- and interspecific intervals (Schmidt *et al.*, 2013; Caromel *et al.*, 2016; Burke *et al.*, 2020), extract ontogenetic trajectories (Brombacher *et al.*, 2022), identify the extent of dissolution (Johnstone *et al.*, 2010; Iwasaki *et al.*, 2015), measure shell wall thicknesses as a proxy for dissolution (Fox *et al.*, 2020) and correlated changing geochemical signals with subspecies taxonomic variation (Kearns *et al.*, 2023), all while preserving the physical integrity of the specimens.

More recently, µCT images have gained importance as training datasets for machine learning and image classification tools that aim to automate species identification and quantify key taxonomic and ecological features (Hsiang *et al.*, 2019; Marchant *et al.*, 2020; Mulqueeney *et al.*, 2024). With resolutions capable of capturing fine-scale morphological details, µCT imaging reveals previously inaccessible structural variations, such as chamber volumes, wall thickness, and ontogenetic features, which provide insights into evolutionary patterns and

functional morphologies. Powerful models to detect speciation changes require larger data sets, which require more specimens to be scanned by µCT. I present an advanced high-throughput mounting and imaging workflow, which enables throughput with a more consistent µCT image and reconstruction. By training algorithms on these consistent µCT data, I demonstrate high-throughput approaches to feature extraction, enhancing reproducibility, and improving efficiency of morphological analyses across large sample sizes.

In this study, I developed novel protocols with µCT to generate new three-dimensional volumes of individual foraminifera to (a) measure the test wall thickness at the locations of LA-ICP-MS target, (b) measure the number of chambers and their volumes through ontogeny and at pre- and post-speciation between lineages, and (c) generate new images for training algorithms for feature extraction and species classification.

2.2 Laser Ablation Inductively Coupled Plasma Mass Spectrometry

Laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) is a powerful analytical tool to quantify the elemental composition of a wide variety of natural and anthropogenic materials. In LA-ICP-MS, a laser beam is focussed to the surface of a target and then pulsed to ablate the sample. The ablated particles are then transported into an inductively coupled plasma ionisation source, where they are converted into ions and subsequently analysed in a mass spectrometer, which detects elements based on their mass-to-charge ratio. This process creates a time-resolved isotopic or elemental profile that can capture data with high spatial resolution. LA-ICP-MS has a reputable history in analysing biogenic carbonates, including foraminifera which act as geochemical archives (Eggins *et al.*, 2003; Hathorne *et al.*, 2003; Fehrenbacher *et al.*, 2015), coral skeletons (Chalk *et al.*, 2021) and molluscs (Schöne *et al.*, 2011). These biogenic carbonates retain elemental signatures from their lifetime that can be used as proxy measurements, offering insights into past environmental conditions and evolutionary responses to long-term climate change.

As foraminifera grow, trace elements from the surrounding seawater are incorporated into their tests in proportions influenced by both environmental conditions and biological mechanisms (Chave, 1954). Variations in these trace element concentrations reflect environmental changes and biological processes that foraminifera experience during their lifespans. The elemental signatures preserved within the carbonate test serve as geochemical

archives, allowing researchers to reconstruct past climates, analyse taxonomic distinctions among cryptic species, and trace evolutionary responses to climatic fluctuations. These applications make LA-ICP-MS a critical tool in palaeoceanography and evolutionary biology, as demonstrated by its use in reconstructing past sea surface temperatures through Mg/Ca ratios (e.g. Anand *et al.* (2003); Barker *et al.* (2005)), identifying cryptic species using elemental distinctions such as B/Ca offsets (e.g. Osborne *et al.* (2020)), and tracing evolutionary adaptations in response to climatic shifts by examining ontogenetic geochemical variability (e.g. Hathorne *et al.* (2003); Hupp and Fehrenbacher (2024)).

One of the most widely used temperature proxies in palaeoceanography is the Mg/Ca ratio in foraminifera calcite. This ratio increases with temperature, as magnesium substitutes for calcium in calcite in temperature-dependent proportions. Since magnesium incorporation in calcite is primarily temperature-driven, Mg/Ca ratios in foraminiferal shells increases with temperature. Some extant species of foraminifera have been calibrated to provide quantitative sea surface temperature reconstructions (Anand *et al.*, 2003; Rosenthal *et al.*, 2022). For extinct species, there are no direct calibrations between Mg/Ca and temperature.

The Mg/Ca ratio additionally reflects aspects of population structure, depth habitat, and microenvironmental conditions (Barker *et al.*, 2005). Non-thermal influences in the Mg/Ca ratio include the calcifying water's salinity and pH (Gray and Evans, 2019), which can be influenced by the foraminifer's respiration (Eggins *et al.*, 2004), and biological mechanisms (Eggins *et al.*, 2003).

These differences in depth habitats and microenvironments, along with species-specific adaptations can cause Mg/Ca ratio disequilibrium, resulting with the measured Mg/Ca ratio being offset from that expected purely through thermodynamic control (Elderfield *et al.*, 2002; Spero *et al.*, 2015). Variations in Mg/Ca ratios then can indicate different thermal or chemical microhabitats occupied by foraminiferal species, contributing to our understanding of their ecological niches and habitat preferences. These factors highlight the complexity of using Mg/Ca as a direct temperature proxy, especially without a calibration specific to the population or habitat being studied.

Measuring trace elements in foraminifera tests at individual chamber resolution is essential for distinguishing between biological and environmental changes that occur over an individual's lifetime and over geological timescales. These variations help reveal ecological

and evolutionary differences within a population and disentangle biological effects, erroneously termed "vital effects" from environmental signals.

Vital effects is a catch-all term that refer to the biological processes, such as species-specific calcification mechanisms, ontogenetic changes, and physiological stress responses, that cause deviations from the thermodynamically expected relationship between geochemical signals (e.g., Mg/Ca, $\delta^{18}O$) and environmental variables (e.g., temperature, salinity) (Erez, 1978; Spero and Williams, 1989; Spero *et al.*, 1991; Hönisch *et al.*, 2003; Birch *et al.*, 2013; Hupp and Fehrenbacher, 2024). For example, differences in the rate of calcification or metabolic activity can influence the incorporation of magnesium into calcite, creating species-dependent offsets in Mg/Ca ratios (Elderfield *et al.*, 2002). Similarly, isotopic fractionation during calcification, influenced by respiration and photosynthesis, can alter $\delta^{18}O$ values, skewing the palaeoclimate reconstructions if not properly accounted for (Zeebe, 1999). By isolating individual chambers, I tease apart the compounding of vital effects from environmental signals, improving understanding of the interplay between biology and environment in shaping foraminiferal geochemistry.

Recent advancements in LA-ICP-MS instrumentation have made it possible to collect trace element-to-calcium (TE/Ca) ratios with precision comparable to traditional solution-based ICP-MS, but with simpler sample preparation and increased throughput (Fehrenbacher *et al.*, 2020). The laser-ablation approach enables high spatial resolution measurements, avoiding the averaging effect seen in solution-based ICP-MS and allowing for higher spatial resolution measurement that avoids the heterogeneity averaging that occurs in solution based ICP-MS.

I used LA-ICP-MS to target six individual chambers (with repeats) in the final whorl of *M. menardii*, *M. limbata*, *M. exilis* and *M. pertenuis* foraminifera to measure Mg/Ca in the test. I then compared how Mg/Ca varied (a) through an individual's ontogeny (adult life stage), (b) among individuals, and (c) before and after a speciation event, to reveal patterns in Mg/Ca that can mature our understanding of both ecological adaptation and evolutionary processes.

2.3 Stable Isotope Ratio Mass Spectrometry Analysis

Oxygen stable isotope ratios in foraminiferal calcite are an established proxy for reconstructing past oceanic conditions that providing insight into changes in global climate and ice volume over geological timescales (Hoogakker *et al.*, 2024). This temperature-dependent incorporation of ^{18}O in foraminiferal calcite and measured as $\delta^{18}O$ values are widely used as proxies for past ocean temperatures (Zachos *et al.*, 2001; Lisiecki and Raymo, 2005).

The $\delta^{18}O$ value of isotopes in foraminifera shells changes with both seawater temperature and the volume of global ice. This isotopic ratio shifts due to the preferential evaporation and precipitation of lighter ^{16}O isotopes at high latitudes, especially during colder glacial periods. As ice sheets grow, they store large quantities of ^{16}O, resulting in an enrichment of ^{18}O in the remaining seawater. When temperatures rise, these ice sheets melt releasing ^{16}O-rich water back into the oceans and decreasing the ^{18}O proportion.

As the lighter ^{16}O isotope is preferentially incorporated into the calcite of foraminifera at higher temperatures, the $^{18}O/^{16}O$ ratio in the foraminiferal test becomes lower in warmer water conditions. Conversely, in colder waters, more ^{18}O is incorporated into the calcite, resulting in higher $^{18}O/^{16}O$ ratios (Emiliani, 1955; Shackleton, 1967).

While $\delta^{18}O$ and temperature calibrations have been proposed for *M. menardii* (Farmer *et al.*, 2007) these are generic values and not specific to the study time period of 5.65 million years ago (Ma) to 2.91 Ma. Here, I present oxygen isotope values in $\delta^{18}O$ notation and infer environmental and climatic shifts from Ceara Rise in the Equatorial Atlantic (ODP Site 925), a region where isotopic signals are reliably preserved due to stable deposition and good calcite preservation (Curry *et al.*, 1994).

How $\delta^{18}O$ varies by species and geological timespans is one tool in disentangling the effects and responses that climate has on new species appearances and evolution.

2.4 Uncertainties using geochemical proxies

Each of the proxy measurements summarised above have associated weaknesses and cause offsets between the recorded value from the *true environmental* value they intend to represent (Katz *et al.*, 2010). Proxy offsets, which are discrepancies between these measured and expected values, can stem from biological and environmental factors that alter the chemical composition of foraminiferal tests. Here, I categorise these sources of uncertainty into biological and environmental groups.

The principle biological uncertainty has been referred to as "vital effects", which are physiological processes unique to each species that cause deviations in the elemental or isotopic composition of the test from what would be expected based solely on external environmental conditions (Schiebel and Hemleben, 2017; Fehrenbacher *et al.*, 2024). The ion composition within their calcifying fluid can result in an apparent Mg/Ca ratio or $\delta^{18}O$ value that does not directly correspond to the ambient seawater conditions (Katz *et al.*, 2010). Vital effects are species specific and can vary through different life stages and over geological timescales.

Species specific Mg/Ca and $\delta^{18}O$ offsets are influenced by the environment the foraminifer occupies during its life (Anand *et al.*, 2003; Erez, 2003). An individual may occupy different habitats at different life stages and incorporate a range of geochemical signals in its test. For geochemical analysis using the entire test, each of these signals are compounded during measurement, resulting with an averaged life history (Eggins *et al.*, 2003). Shell morphology, chamber wall thickness, and calcification processes also differ across species and through geological time, inherently causing species-specific differences in their calcification biology (Schmidt *et al.*, 2004). Laminations of new calcite growth upon old results in a banding of trace element ratios that reflect the life stage when the calcite was laid (Hathorne *et al.*, 2009; Dueñas-Bohórquez *et al.*, 2011; Fehrenbacher *et al.*, 2017). When measured by LA-ICP-MS, the laser ablates through each of these laminations with the detector recording the varying trace element compositions. With our LA-ICP-MS setup and data resolutions, an averaging of these lamination compositions was observed.

Some planktonic foraminifera host photosynthetic symbionts that affect the carbon and oxygen isotopic and trace element composition of the host's calcite (Wolf-Gladrow *et al.*, 1999; Erez, 2003; Eggins *et al.*, 2004). Symbionts influence the microenvironment around the

foraminifer, altering concentrations of CO_2 and other elements, which in turn can impact the $\delta^{18}O$ values in the foraminiferal test (Wolf-Gladrow *et al.*, 1999; Katz *et al.*, 2010). Symbiont-bearing foraminifera typically exhibit different isotopic signatures than symbiont-free species due to these localised microenvironments. The studied *Menardella* species are suggested to be symbiont-bearing, and therefore their isotopic measurements may reflect not only external seawater conditions but also the internal effects of symbiont photosynthesis, leading to additional uncertainty when interpreting stable isotope values as proxies for environmental parameters.

Variability of trace element ratios and isotopic values within fossilised individual foraminifera tests are caused by differences in growth, ontogeny, and post-deposition test preservation (Wolf-Gladrow *et al.*, 1999). To identify the variation of Mg/Ca ratios within a chamber, I repeated LA-ICP-MS measurements at multiple locations on the same chamber for three chambers in the final whorl. Our results are encouraging evidence that the differences in Mg/Ca ratios across chambers are because of environmental differences, rather than analytical uncertainty.

Diagenesis of fossilised foraminiferal tests buried in sediment can alter the original geochemical composition of the calcite through recrystallisation or chemical exchange with pore waters. Recrystallisation of the test can exchange Mg/Ca ratios and $\delta^{18}O$ values from the time of calcite growth to that of the pore water post-deposition, leading to discrepancies between the fossilised proxy measurements and the original environmental signals (Katz *et al.*, 2010).

Post depositional dissolution removes calcium carbonate from the test, potentially biasing the remaining test's isotopic and elemental composition (Rosenthal and Lohmann, 2002) and seen first in thin walls, such as internal chamber septums. The dissolution of septums reduces the count of individual chambers when measured by X-ray CT scan images, which is a key taxonomic metric to distinguish between species. New mineral deposits form on the test surface post-mortem introduce additional overgrowth material that may not represent the original geochemical signals, and as such are another source of measured Mg/Ca and $\delta^{18}O$ offsets from the *true environmental* values.

Following results and suggestions from Kearns *et al.* (2023) that a focus for palaeoceanography and palaeoclimate research must be on increasing sample sizes for LA-ICP-MS work to achieve representative sample sizes is to include >55 individuals from each

species. The unprecedent scale of my integrated dataset I present here includes 306 *M. menardii*, 747 *M. limbata*, 549 *M. exilis*, and 245 *M. pertenuis* to be consistently sufficiently large to infer statistical support for systemic species variation using mathematical models. The novel outputs from my published work include methodological improvements for µCT scanning and trace element measuring for such a quantity of individuals, open-source software to process the volume of data reliably and repeatably, powerful multidisciplinary integrated datasets as the foundation of models that explain morphometrical (shape, size, chamber volumes, and wall thickness) and geochemical variations (Mg/Ca and $\delta^{18}O$) within- and between- individuals. I demonstrate how my results answer questions about ecological evolution, developmental plasticity and relationships to climate.

Chapter 3 Coherence between each published research article

I present here a collection of published research articles that develop and mature the scientific understanding of the ecological, evolutionary, and environmental dynamics affecting four *Menardella* species of foraminifera. Each research article uses the same fossilised tests of individual *M. menardii*, *M. limbata*, *M. exilis*, and *M. pertenuis* from ODP Site 925.

Through direct measurements of geochemical and morphological values collected from (a) individual chamber volumes and wall thickness measurements from X-ray micro-computed tomography (µCT), (b) Mg/Ca ratios in the final whorl using laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS), and (c) whole test $\delta^{18}O$ values. I combined these together in *Laser ablation mass spectrometry blast through detection in R* and *The Influence of Geochemical Variation Among Globigerinoides ruber Individuals on Paleoceanographic Reconstructions* to build powerful models to disentangle the interplay between ontogenetic development, species-specific biological effects, and climate-driven changes over evolutionary timelines. Each direct chemical and morphological measurement made is associated with a substitute proxy measurement for the environment and life history experienced by the individual foraminifer during its lifetime. This multi-proxy, multidisciplinary approach allows us to disentangle complex signals embedded in the foraminifera test, providing new insights that link foraminiferal morphology and chemistry with their environmental context.

I questioned if it was possible to estimate a foraminifer's test thickness at the location of laser ablation without the need of additional tools. One output from my R package LABLASTER that I developed to optimise LA-ICP-MS data processing to reduce measurement uncertainty by minimising systematic biases is measuring the time taken to fully ablate through the test wall and described in *Laser ablation mass spectrometry blast through detection in R* (Searle-Barnes et al., 2023). By combining the measured time to ablate through the test wall with manually measured test wall thicknesses from the µCT scan volumes as described in *Ten recommendations for scanning foraminifera by X-ray computed tomography*, I demonstrate the thickness of the test can be estimated with mixed-effect models using only the time

taken to fully ablate. Quantifying this correlation between time taken to ablate and chamber wall thickness provides researchers with an additional tool to estimate the test thickness at the time of LA-ICP-MS and without the need to additionally µCT the specimen.

This body of work integrates high-resolution µCT, LA-ICP-MS, and stable isotope ratio analysis to investigate morphological and geochemical variability in foraminifera. By combining these methods, I quantified chamber volumes, growth rates, and wall thicknesses with precision and linked these morphological metrics to geochemical signatures at individual chamber resolution. This approach enabled the identification of ontogenetic stages within *Menardella* lineage species and the detection of species-specific Mg/Ca patterns that distinguish biological "vital effects" from environmental influences across individuals and ontogenetic stages. The resulting large, detailed, multidisciplinary dataset bridges gaps between palaeoclimate, palaeontological and ecological research, revealing relationships between environmental variability, ecological adaptation, and evolutionary processes. My findings contribute to more precise climate reconstructions and a deeper understanding of the ecological and evolutionary dynamics of foraminifera.

Chapter 4 Methods

I washed and selected 1268 individual foraminifera from the *Menardella (Globorotalia)* lineage comprised of 261 *Menardella menardii*, 500 *Menardella limbata*, 376 *Menardella exilis* and 131 *Menardella pertenuis* as described by Kennett and Srinivasan (1983). These were collected from the Ceara Rise in the Equatorial Atlantic region at Ocean Drilling Program (ODP) Site 925, which comprised Hole 925B (4°12.248'N, 43°29.349'W), Hole 925C 20 (4°12.256'N, 43°29.349'W), and Hole 925D (4°12.260'N, 43°29.363'W). See Curry *et al.* (1994) for more details. The samples used spanned 5.65 million years ago (Ma) to 2.91 Ma (Wilkens *et al.*, 2017).

I sampled pre- and post-speciation events for the *M. menardii – M. limbata*, the *M. limbata – M. exilis* and *M. exilis – M. pertenuis* splits, following the speciation dates as described by Aze *et al.* (2011) in addition to every ~90 Kya.

Each foraminifer was picked from a >63 µm size fraction then individually cleaned by ultrasonification in ethanol for 3 – 5 seconds followed by a wash in Milli-Q water to remove surface and embedded particulates as described by Eggins *et al.* (2003). The foraminifera selected from ODP Site 925 between 30 – 214 metres composite depth were excellently preserved. This site's relatively shallow depth of ~3040 meters contributes to favourable preservation conditions, as it lies above the carbonate compensation depth, reducing dissolution rates (Chaisson and Pearson, 1997; Cornuault *et al.*, 2023).

I systematically identify each chamber within a foraminifer by counting the number of chambers from last grown chamber in the final whorl; where f-0 is the final chamber, f-1 is the penultimate, f-2 the antepenultimate, and so on.

4.1 X-ray micro-CT scanning

I developed a novel mounting methodology to load individual foraminifer into a clear plastic straw sandwiched between slices of phenolic foam and layered into a single vertical stack. This stacked layering prevents the foraminifera from touching one another within the straw, ensuring that each individual is fully scanned without missing sections due to overlapping. This arrangement simplifies the separation of individuals in reconstructed µCT volumes during post-processing, enhances automation, and supports the high-throughput nature of the workflow. Each straw contained up to 20 slices, grouped in blocks of five with each block being separated by slice of lower-density foam. The addition of the lower-density foam provided a visual safe space to cut the straw to aid the recovery each foraminifer from the foam before subsequent analysis by LA-ICP-MS.

In advance of scanning, I optimised the tungsten source power, beam filter, and optical and geometric magnification to balance throughput and brightness of the foraminifera to set a standard operating method. I scanned each foraminifer by micro-focus computed tomography using a Zeiss Xradia Versa 510 X-ray microscope CT Scanner (Carl Zeiss Microscopy GmbH, Germany).

Our standard operating method consisted of a beam peak voltage of 110 kV and 10 W power pre-filtered through a 0.15 mm SiO_2 ceramic filter to absorb and attenuate the lower-energy X-ray photons from the spectrum for the reduction of beam hardening artefacts. A voxel resolution of 1.75 µm was achieved through geometric magnification, a 4x optical magnification and 2x detector binning. Each foraminifer was mapped as an individual region of interest (ROI) and was radiographed with a 1.3 s exposure time for 1011 projections.

To emphasise the contrast between the foraminifera from the phenolic foam I applied a custom byte-scaling (-0.04 – 0.1) on the greyscale range within the Zeiss Reconstructor software. Additionally, a beam hardening correction algorithm to reduce beam hardening artefacts was applied before the projections were reconstructed into a stack of 16-bit tiff images.

4.2 Individual chamber trace element ratios by Laser Ablation Inductively Coupled Plasma Mass Spectrometry

I divided the 1268 foraminifer into batches of approximately 120 individuals and mounted them onto a glass slide umbilical side up and held in place using double sided adhesive tape. Each batch was analysed separately for practical reasons across multiple days, and its data was processed accordingly.

Trace element (TE)/Ca ratios were analysed using a New Wave UP193 laser ablation system (ArF source, 30 µm spot diameter, fluence of 0.73 J cm^{-3} and 5 Hz pulse rate) coupled to an Agilent 8900 triple quadrupole inductively coupled plasma (ICP-QQQ) mass spectrometer in single quadrupole mode using a Ar and He gas mixture (90%:10%, 900 mL min^{-1}) at the University of Southampton. During analysis the isotopes ^{7}Li, ^{9}Be, ^{11}B, ^{24}Mg, ^{27}Al, ^{43}Ca, ^{44}Ca, ^{55}Mn, ^{65}Cu, ^{66}Zn, ^{88}Sr, ^{137}Ba and ^{238}U were measured as a time resolved acquisition; only Ca and Mg data are presented here.

Prior to an analysis session, the mass spectrometer and laser system were tuned for optimal sensitivity, stability, and low oxide formation. Using the laser software, foraminifer and reference material locations were mapped and recorded to form a programmed sequence. Ten replicates each of NIST 610 and 612 reference glasses, United States Geological Survey MACS-3 (a pressed carbonate pellet) and Japanese Geological Survey's JCp-1 (a Porites sp. coral reference material) were analysed at the start, end and in batches throughout each session between 9 – 11 foraminifer (about 80 – 100 spots) as followed from Kearns *et al.* (2023). For each foraminifer chambers f-0, f-1 and f-3 were mapped in duplicate and chambers f-5, f-7 and f-9 were mapped once, where f-0 is notation for the final chamber. Each laser spot pattern was sequenced with a 30 s warmup, 50 s laser pulse and 30 s washout.

The LA-ICP-MS time-resolved acquisition data was processed in R (Team, 2010) with each step repeated for each isotope. The first 15 seconds were used to calculate the background signal with the median subtracted from the signal counts. The signal counts were truncated to include only the timestamps when the laser was turned on. Internal corrections using Ca wt% were calculated using the concentration values 8.15 wt% of NIST 610, 8.5 wt% of NIST612, 37.7 wt% of MACS-3 and 37.5 wt% of JCp-1 (see Jochum *et al.* (2019) and Jochum *et al.* (2011)). Influential outlier values were detected using Cook's Distance and anomalous values beyond a threshold were removed (0.001 for standards or 0.1 for foraminifer). The

endPoint function from the lablaster package (Searle-Barnes *et al.*, 2023) was applied across all foraminifer (dt = 10, smoothing = 10) to detect when the laser had fully ablated through the test and therefore the timestamp when the recorded signal was no longer representing the desired target. Only the data points between 15 seconds elapsed and the timestamp LABLASTER returned were kept as this is when the laser was pulsing and on-target. A moving window followed the progress of the analytical session and was used to select the nearest 100 signal counts of each reference material for constructing calibration curves that regressed the measured signal of NIST 610, NIST612, MACS-3 and JCp-1 against their literature concentration values (Jochum *et al.*, 2011; Jochum *et al.*, 2019).

Trace element (TE) to calcium ratios, including Mg/Ca ratio (mmol/mol) were calculated and the median taken. For contamination control, I checked Mg/Ca medians lay between 0 and 15 mmol/mol and removed 1326 chamber locations that fell outside this range out of 12285 chambers measured in total.

4.3 Measuring of chamber volumes and test thickness

Four *M. menardii*, eight *M. limbata*, two *M. exilis* and five *M. pertenuis* individuals were analysed manually within Dragonfly (2022 version 2022.1, Object Research Systems, Montreal, Canada). The tiff image stack was loaded with the x, y and z planes aligned to the laser shot locations.

4.3.1 Chamber volumes

Chamber volumes were manually measured as individual regions of interest. The internal area of each chamber for each µCT image slice was manually marked and then integrated into a 3D volume and assigned as a region of interest. Chamber boundaries were determined by changes in the bounding calcite density.

4.3.2 Test thickness

The thickness of the test at the laser shot locations were measured using the ruler tool.

I estimated the test thickness at the location where the laser was ablating for every chamber measured by LA-ICP-MS. By multiplying the time taken to fully ablate through the test and the depth ablated by each pulse of the constant power laser, I estimate the test thickness in µm.

4.4 Whole test $\delta^{18}O$ ratios by stable isotope ratio mass spectrometry

Of the foraminifera that were measured for trace elements, 1160 of these were also measured for $\delta^{13}C$ and $\delta^{18}O$ stable isotope ratios at the SEAPORT Stable Isotope Laboratory, University of Southampton. Single specimen analysis was performed on 218 menardii, 460 limbata, 356 exilis and 216 pertenuis individuals using a Kiel IV carbonate device coupled to a MAT-253 isotope ratio mass spectrometer. Replicates of reference materials NBS18 and NBS19 and in-house quality control materials GS1 and SHK were analysed at the start, end and throughout each session between 8 – 10 foraminifera.

The long-term analytical precision of the setup is reported as the 1σ estimate from the repeat measurements of the in-house quality control GS1 and were estimated as ±0.04 ‰ for $\delta^{13}C$ and ±0.06 ‰ for $\delta^{18}O$.

All results were standardised to Vienna Pee Dee Belemnite (VPDB) using a two-point calibration between NBS19 and NBS18 and as deviations in parts per thousand (‰). I present only $\delta^{18}O$ in this collection of papers.

4.5 Data processing

All the data processing and statistical analysis was performed in the R environment for statistical computing (Team, 2010) with additional packages used for data wrangling and for calculating mixed effect models (Wickham, 2007; Bates *et al.*, 2015; Wickham *et al.*, 2019; Searle-Barnes *et al.*, 2023).

Chapter 5 Novel findings

5.1 Laser ablation mass spectrometry blast through detection in R

I developed the LABLASTER (Laser Ablation BLASt Through Endpoint in R) package that imports a single time resolved LA-ICP-MS analysis, then detects when the laser has ablated through the carbonate as a function of change in signal over time, and outputs key summary statistics. This was in response to a need for precise and efficient LA-ICP-MS data reduction analyses where the analysed target test thickness was variable and unknown in advance as existing setups required a fixed integration time. Existing data processing workflows often introduced uncertainty due to their inability to isolate periods when the laser is precisely focused on the target and when the laser had fully ablated through the target. I present the first open-source R package, with an unrestricted use licence, that automates the selection of desired data during data reduction workflows by automating the detection of when the laser has ablated through a sample using a smoothed time-series, followed by removal of off-target data points. This software improves repeatability during the data reduction stages by eliminating human subjectivity and biases determining when the laser has fully ablated through the test wall, and provides clear, reproducible workflows with its thoroughly documented algorithm. The functions are flexible and adjust dynamically to maximise the duration of the desired geochemical target signal making this package applicable to a wide range of heterogenous bioarchives. This method eliminates extraneous values recorded when the laser beam was not focused on the desired area of the test, resulting in more accurate and reproducible geochemical measurements. Notably, keeping only the data points that are measured when the laser is on-target reduces the uncertainty for a calculated average due to narrowing of data point variance. Figure 5.1 shows this processing by LABLASTER algorithm to remove systemic biases in the retrieved signal due particularly to the period of after the laser has fully ablated the test wall. Algorithmically identifying this point provides consistency by removing subjective human judgment and biases, supports high-throughput analyses across large datasets, and improves comparability and reproducibility when used across studies.

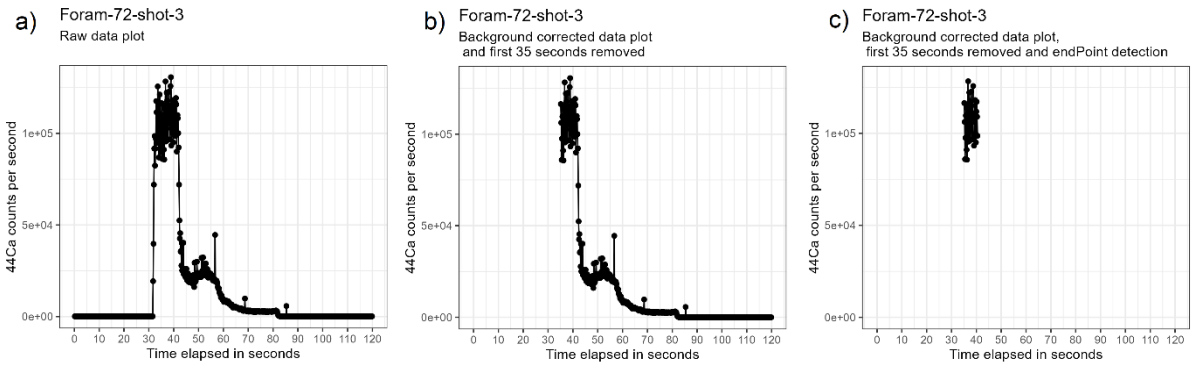


Figure 5.1: The processes implemented in the endPoint function in a time resolved acquisition of Foram-72-shot-3 as a visualisation of the returned data frame $df. Panel (a) visualises the entire raw ^{44}Ca data collected; panel (b) visualises the background corrected ^{44}Ca with the first 35 seconds removed as this was before the laser was turned on and the ablated material is still travelling through the system and this is the data frame passed into the endPoint function; plot (c) visualises the target data retained after running the endPoint function, with both the first 35 seconds and post endPoint detection ^{44}Ca data removed.

By applying this function during LA-ICP-MS data reduction of a *M. exilis* foraminifer, I demonstrated that the inferred median ^{24}Mg/^{44}Ca ratio decreased from 4.37 ± 2.10 mmol/mol (when processed without the function, n=286) to 2.46 ± 1.35 mmol/mol (when processed with the function, n=32). This reduction in uncertainty of approximately 36% demonstrates the efficacy of LABLASTER in improving precision in LA-ICP-MS measurements.

By combining manually measured test wall thicknesses from µCT scans at the same locations as laser ablation during LA-ICP-MS and the time taken to fully ablate identified by LABLASTER during data reduction, I show using a generalised linear model that test wall thickness can be estimated solely based on the laser ablation time. The model explains 50% of the variation in thickness predictions when LABLASTER is used to determine the ablation time. Test wall thickness serves as a useful morphological indicator, and this study demonstrates that it can now be estimated during routine LA-ICP-MS data reduction workflows, eliminating the need for time-consuming manual measurements using traditional methods or X-ray CT scans.

Fehrenbacher *et al.* (2020) demonstrated the compatibility of laser ablation and solution-based ICP-MS results for foraminiferal Mg/Ca and Sr/Ca analyses. LABLASTER enables higher-

throughput processing and improved precision in LA-ICP-MS workflows, while maintaining the flexibility to target specific regions of interest within foraminiferal tests. This is particularly useful for studies of ontogeny, where chamber-by-chamber geochemical variability provides insights into growth and environmental conditions.

LABLASTER is novel through its ability to accommodate variable laser duration and mass spectrometer recording times, a feature not typically present in setups that enforce fixed parameters in time-resolved analyses (TRA). By retaining only data from targeted regions of interest, the package not only reduces error but also enhances the repeatability of results in studies using LA-ICP-MS. The ability to predict the test wall thickness based on laser ablation time provides a new tool for uncovering links between morphology and geochemistry, and without the need for multiple analytical instrumentation. LA-ICP-MS results in collaboration with LABLASTER contributes to a deeper understanding of ontogenetic and environmental influences in geochemical signals within foraminifera.

Using LA-ICP-MS and the LABLASTER R package, and manual measurements from X-ray CT, I demonstrate that that approximately 50% of the chamber wall thickness can be inferred from the time required to ablate through it. I show that chambers f-5 and f-7 are notably thicker (Figure 5.2), likely due to secondary calcite overgrowth, and that these chambers contribute substantially to the whole-test stable isotope composition due to their greater mass.

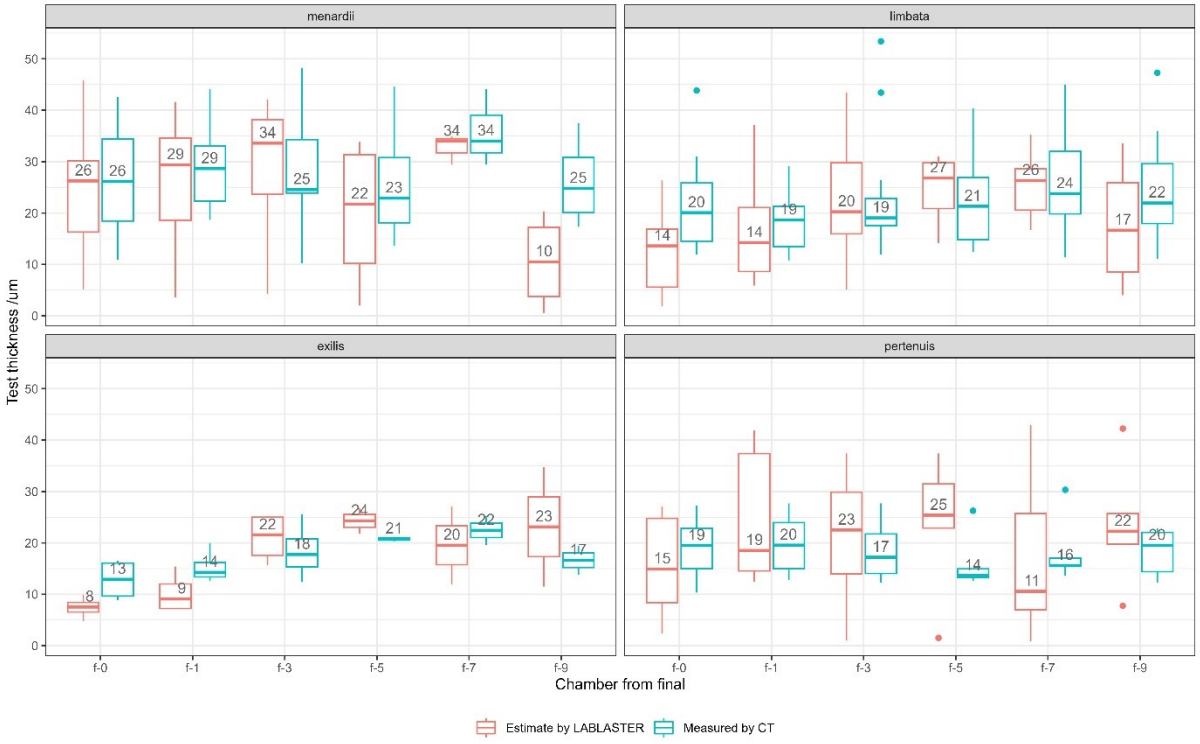


Figure 5.2: Test wall thicknesses for the chambers 0, 1, 3, 5, 7 and 9 from final chamber grown for each of the species *M. menardii*, *M. limbata*, *M. exilis*, and *M. pertenuis*. I show the median and median absolute deviation (as error bars) for the estimated thickness using a 200 nm depth per laser pulse and the LABLASTER package (in red) and the same foraminifer were used when measuring the test thickness using X-ray CT scan data (in teal). The median test thicknesses for both estimated and measured are rounded and displayed to whole numbers and are similar in value.

5.2 Ten recommendations for scanning foraminifera by X-ray computed tomography

I present here a novel method of sample preparation that doesn't include glue, gel, or solvents, which could affect the accuracy of image segmentation. I describe mounting foraminifers within a rigid plastic straw encased within malleable phenolic resin foam to isolate the samples from dust and making an easily transportable unit without the risk of losing the specimens ready for µCT imaging (Figure 5.3). This is in contrast to literature methods (Johnstone *et al.*, 2011; Görög *et al.*, 2012; Schmidt *et al.*, 2013; du Plessis *et al.*, 2017; Hipsley *et al.*, 2020), which include mounting foraminifer specimens on individual wooden toothpicks or pipette tips (Kendall *et al.*, 2020; Vanadzina and Schmidt, 2022), keeping the specimen exposed to surroundings, dust, and increasing the risk of loss. The vertical stacking of specimens, with an alternate coloured- and density- foam markers placed every five specimens, ensures all specimens are accounted for during scanning and provides safe cutting points for efficient retrieval from the straw. This arrangement minimises beam hardening artefacts by aligning each specimen individually with the X-ray beam, and the ability to recover specimens after scanning allows for their reuse in subsequent analyses, maximising resource efficiency and preserving the integrity of the specimens. This paper describes the X-ray µCT setup used throughout this PhD and for the NERC *PISTON* project.

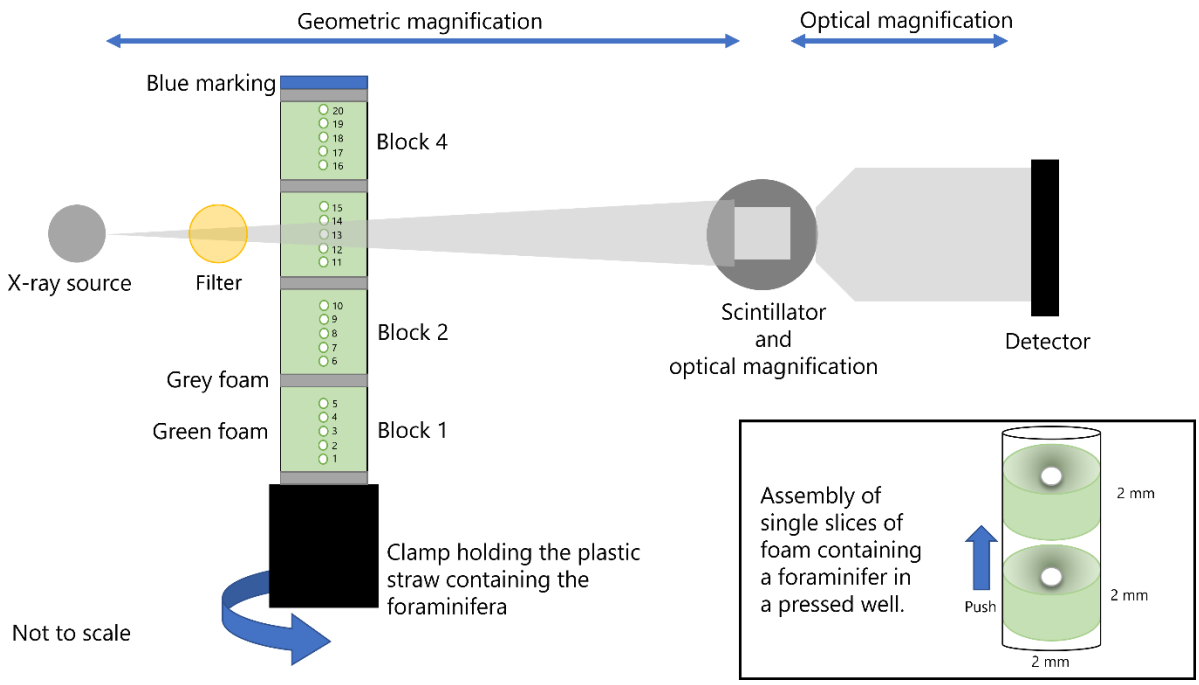


Figure 5.3: Our XRadia Versa 510 setup used throughout the presented case study. Key instrument parts shown are a tungsten X-ray source, a scintillator with optical magnification and a charged coupled device (CCD) detector. The X-ray beam, shown as grey shading, passes through a 0.15 mm SiO_2 ceramic filter before a plastic straw packed with foraminifers sandwiched between green foam and with grey foam after every five specimens to build blocks of specimens and safe areas to cut for specimen retrieval and a blue marking to indicate the top of the straw to ensure the scan data corresponds to the expected specimen. Individual specimens and each block are captioned in the figure. The straw is clamped to a stage and is incrementally rotated (360° / number of projections) after each radiograph is taken.

Through varying the primary X-ray parameters of µCT imaging that a researcher can optimise, I demonstrate the impact to throughput, signal-to-noise ratio and cost requirements to generate three dimensional (volumetric) datasets. I showcase the effect of varying the X-ray beam power and energy, detector binning, number of projections, and exposure times had on the captured image quality of a single *Dentoglobigerina altispira* planktonic foraminifer in Table 5.1 and Figure 5.4. In our case study the highest beam power resulted in the widest contrast between the subject of interest and the background, allowing

for easiest threshold-based segmentation of the object and aiding computers in automated feature extraction.

Table 5.1: Instrument parameters for repeated scans of a single *Dentoglobigerina altispira* foraminifer with the achieved resolution and resulting file size.

Figure 5.4 Panel	Magnification	Exposure (sec)	Projections	Energy (kV)	Binning	Power (W)	Scan Time (hours)	File size (kb)	Resolution (nm)
(a)	20x	15	701	40	4	3	3.5	463	1370
(b)	20x	24	1101	60	2	5	8.5	1857	685
(c)	20x	12	1101	80	2	7	4.5	1857	685

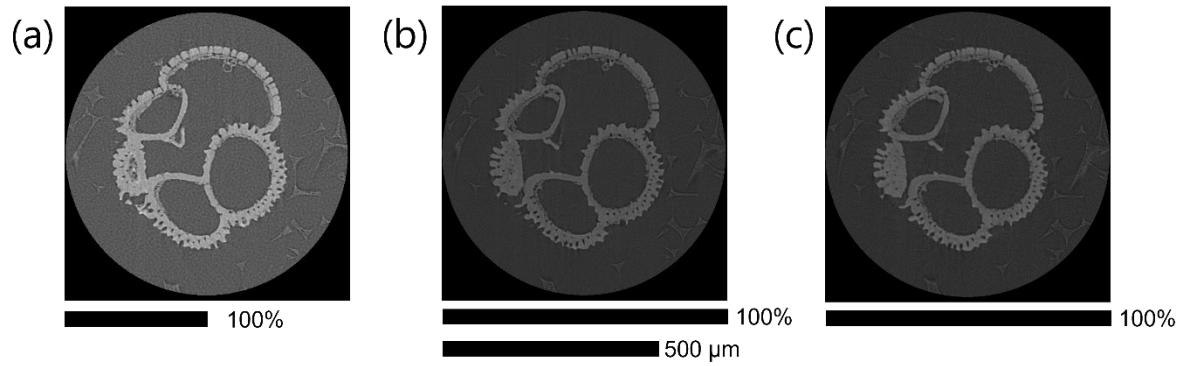


Figure 5.4: The same *Dentoglobigerina altispira* foraminifer was repeatedly scanned with varying parameters and a single slice from the reconstructed tiff stacks are presented to visually compare the effects of resolution, contrast and artefacts. A scale bar shows the relative width of each square image and indicates that image A is presented twice as large to make the 500 µm scale bar applicable to all images. (A) 15 s exposure, 40 kV and 4x binning achieving 1370 nm resolution; (B) 24 s exposure, 60 kV and 2x binning achieving 685 nm resolution; (C) 12 s exposure, 80 kV and 2x binning achieving 685 nm resolution.

Table 5.2: Comparison of mean brightness values of air, foam and calcite (the foraminifer shell) using values from the histograms of Figure 5.4 A, B and C.

Figure	Mean air brightness	Mean foam brightness	Mean calcite brightness	Brightness separation between foam and calcite	Separation between foam and calcite

A	23051	26018	40853	14835	57%
B	13075	15167	22358	7191	47%
C	12795	14534	23353	8820	61%

I developed and implemented the sample preparation methodology with the aim of consistent throughput imaging and reproducibility across research experiments. I designed the experimental framework to test and optimise µCT scanning parameters, using multiple configurations to assess their effects on resolution, signal-to-noise ratio, and segmentation accuracy. I generated and analysed datasets that demonstrated the advantages of high beam power in achieving maximum contrast between foraminiferal calcite and the surrounding phenolic resin foam. Having applied my novel methods across 3,000 individuals — far exceeding the 55 specimens suggested by Kearns *et al.* (2023) — my work reveals hidden morphological features such as chamber volumes, tropospirality, and growth curves that uncover how these traits vary across species and within lineages. My results highlight the plasticity of foraminiferal growth patterns and provide novel insights into speciation events, linking ontogenetic development to broader evolutionary processes.

With a successful methodology that produces µCT scan images that are useful across research domains and that follows the FAIR principles of findability, accessibility, interoperability, and reusability of data (Wilkinson *et al.*, 2016), this dataset has been used and re-used across multiple studies, such as exploring foraminifera test pore distribution variation (Abu Bakar *et al.*, 2023), calculating chamber volumes through manual and automatic methods, and as training data sets for machine learning algorithms (Mulqueeney *et al.*, 2024).

5.3 The Influence of Geochemical Variation Among *Globigerinoides ruber* Individuals on Paleoceanographic Reconstructions

This paper investigates the geochemical and morphological variation among the *sensu stricto*, *sensu lato* and *sensu lato extreme* subspecies of the extant planktic foraminifera *Globigerinoides ruber*. Our experimental design distinguishes between subspecies and intraspecific variability as well as the repeatability of laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS). I show that geochemical variability in Mg/Ca ratios is driven by differences in subspecies depth habitat and that ontogenetic trends in Mg/Ca ratios are evident in the final whorl, with the final chamber consistently showing depleted Mg/Ca. These ontogenetic trends are not driven by individual chamber or test size. The Mg/Ca value variance among individuals is ~100 times higher than the variance among repeated laser spot analyses of single chambers, directing laboratory protocols towards the need to sample ecologically and environmentally homogeneous samples. Our results emphasise that I can use LA-ICP-MS to quantify how individual variability aggregates to bulk results, and highlights that, with sufficient sample sizes, it is possible to reveal how intraspecific variability alters geochemical inference.

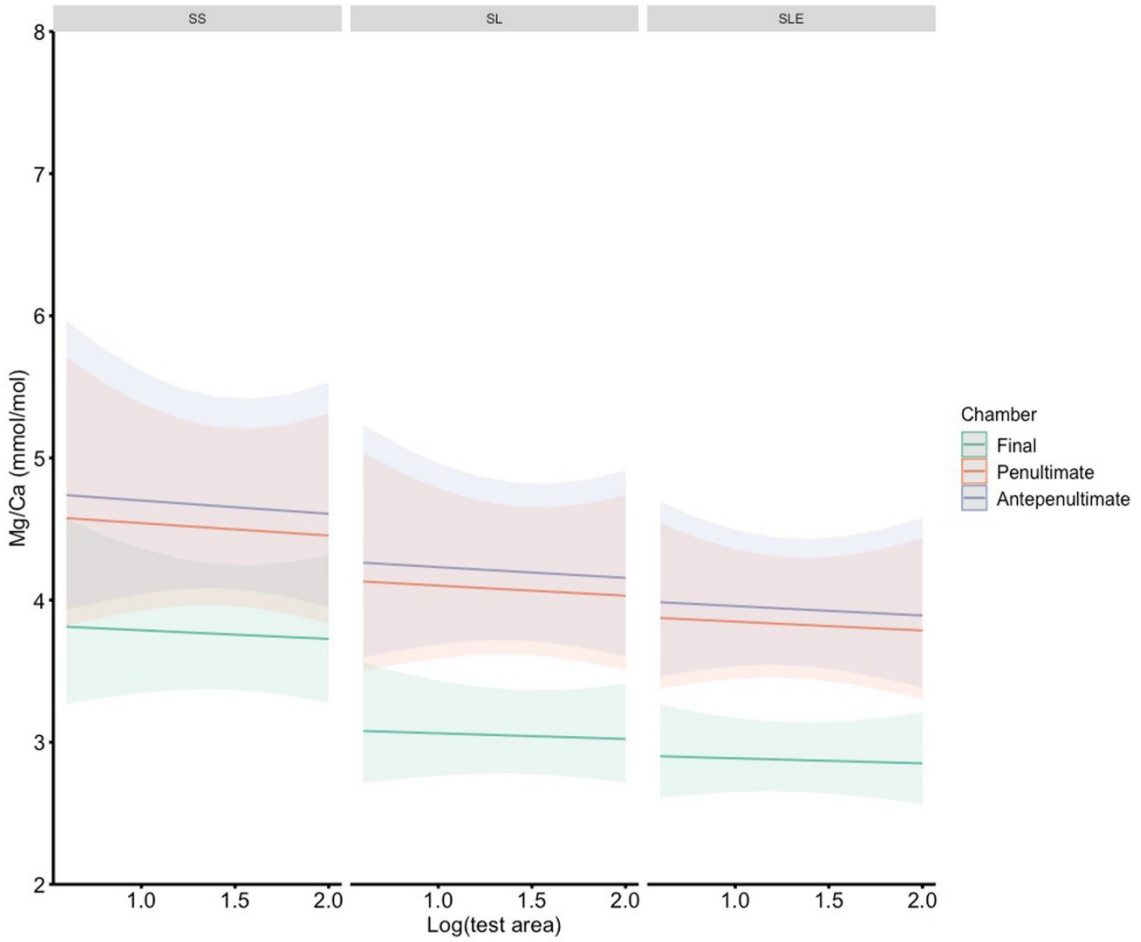


Figure 5.5: Reproduced from Kearns, L. E., Searle-Barnes, A., Foster, G. L., Milton, J. A., Standish, C. D., & Ezard, T. H. G. (2023). The influence of geochemical variation among *Globigerinoides ruber* individuals on paleoceanographic reconstructions. *Paleoceanography and Paleoclimatology*, 38.

I contributed to this work through data collection and the development of novel data reduction techniques. New methods improved the precision of chamber-specific median Mg/Ca ratio estimates by reducing uncertainty ranges compared to traditional and existing software approaches. I demonstrated that the observed geochemical variability originates at the individual level, rather than being an artifact of batch processing or laser spot variability. I present with increased precision that the median Mg/Ca ratios is systematic at the subspecies level with different Mg/Ca ratios through the final whorl. I observed an overall pattern of final chamber depletion in all subspecies (Figure 5.5) with the antepenultimate and penultimate chambers showing similar values.

By revealing the influence of intraspecific variation and subspecies-specific depth habitats on Mg/Ca ratios, this work advances the precision and reliability of geochemical proxies in reconstructing past ocean conditions. This research identifies that each subspecies has similarities in Mg/Ca ratio patterns suggesting that they are ecologically or biologically driven geochemical signatures. As foraminifera grow, older chambers are overprinted with the calcite of newly formed chambers (Hemleben *et al.*, 1989) increasing chamber thickness and adding new Mg/Ca bands. Intratest variability in high and low Mg bands may explain the variation we see if such banding was subspecies specific and a geochemical "vital effect". Additionally, I show the measured variation of Mg/Ca ratios across chambers is a true signal within the individual, rather than batch or shot.

5.4 Contributions to software packages

I manually labelled µCT scan images for individual chamber volumes of the 50 individual specimens that were used to build and test the foram3D algorithm. I assisted in developing the analytical pipeline, validating the algorithms for trait extraction by comparing to my manual dataset, and applying the package to large datasets to assess its utility.

I demonstrate the benefits of integrating high-throughput X-ray CT and automated data processing into micropalaeontological research. The foram3D package is a computational tool that produces standardised and reproducible results from foraminifera morphology measurements. By enabling researchers to study foraminiferal traits in three dimensions, foram3D bridges the gap between traditional two-dimensional microscopy and the demand for richer datasets to investigate growth patterns, ontogeny, and evolutionary transitions.

5.4.1 Analysing planktonic foraminiferal growth in three dimensions with foram3D: an R package for automated trait measurements from CT scans

We present foram3D, an R package designed to automate the extraction and measurement of planktonic foraminiferal morphological traits from X-ray micro-computed tomography (µCT) scans. This package enables high-throughput analysis of three-dimensional traits, such as chamber volumes, test surface area, and growth rates, while preserving spatial context.

This innovation provides new opportunities to study ontogeny, evolutionary processes, and environmental influences in foraminifera using standardised and reproducible methods.

5.4.2 3DKMI: A MATLAB package to generate shape signatures from Krawtchouk moments and an application to species delimitation in planktonic foraminifera

This paper describes the development and application of an open-source MATLAB package called 3DKMI, which utilises Krawtchouk moment invariants to generate 3D shape signatures from planktonic foraminifera X-ray CT scans. These shape signatures provide a mathematically robust and rotation-invariant way to quantify morphological traits, facilitating species delimitation, and addressing challenges in cryptic speciation. By combining advanced computational methods with ecological and evolutionary questions, this study demonstrates the potential of 3DKMI for automated species identification and morphological analysis. 3DKMI integrates high-throughput and automated methods into the morphological variability through evolutionary processes by foraminifera. By providing a quantitative method to delineate species boundaries and track morphological evolution through geological time, I aid the identification of cryptic species of foraminifera.

5.4.3 How many specimens make a sufficient training set for automated three-dimensional feature extraction?

This paper determines the minimum number of µCT training images required to produce accurate volumetric and shape data from internal and external structures of planktonic foraminifera by balancing sufficient specimen variability for robust model training with the practical constraints of data acquisition. I used Dice scores and principal component analysis to evaluate the accuracy of each neural network to identify thresholds where additional specimens provide diminishing returns in improving model accuracy.

This work has implications for disciplines like evolutionary biology and palaeontology, where 3D morphological data are expanding traditional light-based microscopy understanding form and function. With the recent rise in using machine learning for automated 3D feature extraction, there is currently a lack of clear guidance on sample size requirements. By

addressing this gap, this paper matures our understanding of model generalisation and minimising resource expenditure in biological imaging studies.

5.5 Towards understanding speciation by automated extraction and description of 3D foraminifera stacks

This conference proceeding describes a systematic investigation into the number of specimens required to create a sufficient training set for automated three-dimensional (3D) feature extraction in micropaleontology. In the field of computer science, conference proceedings are valued full-length published articles for cutting-edge research communication and recognised for their academic contribution. Publishing in the *Institute of Electrical and Electronics Engineers*, where foraminiferal research is not typically discussed, shows the multidisciplinary nature of my work converging advanced imaging, algorithm development, morphometric and ontogenetic changes, and geochemistry.

The study uses X-ray CT scans of fifty planktonic foraminifera, comprising 4 *M. menardii*, 17 *M. limbata*, 18 *M. exilis* and 11 *M. pertenuis* specimens to evaluate the efficiency and accuracy of automated workflows across varying training set sizes. I quantified the trade-offs between sample size, measurement precision, and computational resources required to discriminate between these species using their morphological differences. The results of this paper show that automated segmentation yields equivalent results to manual segmentation when extracting three-dimensional features from images, and drastically reduces the time required. Most of the automated accuracy improvement occurs between training sets of one to eight specimens. I discuss the need of balancing comprehensive training sets with the practical limitations of specimen availability and computational resources. As morphology is a key taxonomic indicator, these computer vision feature extraction algorithms will be able to aid researchers identify species of previously species-ambiguous or cryptic specimens.

I provided the X-ray CT scan images and manually labelled the 50 individual specimens used to build and test the convolutional neural network. I contributed to drafting sections of the manuscript and interpreting how the results inform best practices in manual and automated feature extraction for future studies utilising machine learning in foraminiferal research.

5.6 Detecting environmentally dependent developmental plasticity in the fossil record

This paper describes the environmental influence on developmental and evolutionary processes in planktonic foraminifera. By integrating µCT with LA-ICP-MS, this study reconstructs ontogenetic trajectories and calcification environments for the *M. limbata*, *M. exilis*, and *M. pertenuis* species described earlier. I link somatic growth rates and shell morphology to calcification temperature, inferred from Mg/Ca ratios in the foraminifera shells, to uncover how environmentally associated developmental plasticity shapes evolutionary outcomes.

I measured test wall Mg/Ca (mmol/mol) ratios across six chambers in the final whorl for each species (Figure 5.6). The Mg/Ca ratio was highest in the final chamber (f-0) for each species, with a general trend of decreasing Mg/Ca ratios earlier in ontogeny. I identify both intra-individual (the final chambers exhibit elevated Mg/Ca ratios than earlier grown chambers, except for *M. pertenuis*), and species-specific variation (that *M. menardii* has the lowest Mg/Ca ratios) across the four species studied.

I identified that for the four species measured, the Mg/Ca ratio of chamber f-5 exhibits the greatest sensitivity to temperature changes (Figure 5.6), as evidenced by the steepest slope in the relationship between Mg/Ca and $\delta^{18}O$ across a range of -2‰ to 2‰ $\delta^{18}O$ values. That chamber f-5 responds more markedly to variations in temperature makes it a key target for chamber specific LA-ICP-MS trace element measurements. The pronounced Mg/Ca ratio sensitivity highlights its potential utility in reconstructing paleoenvironments with greater accuracy. The similarity in Mg/Ca and $\delta^{18}O$ correlation slopes across earlier chambers suggests that these patterns are unlikely to be biologically controlled and may instead be driven by thermodynamic processes.

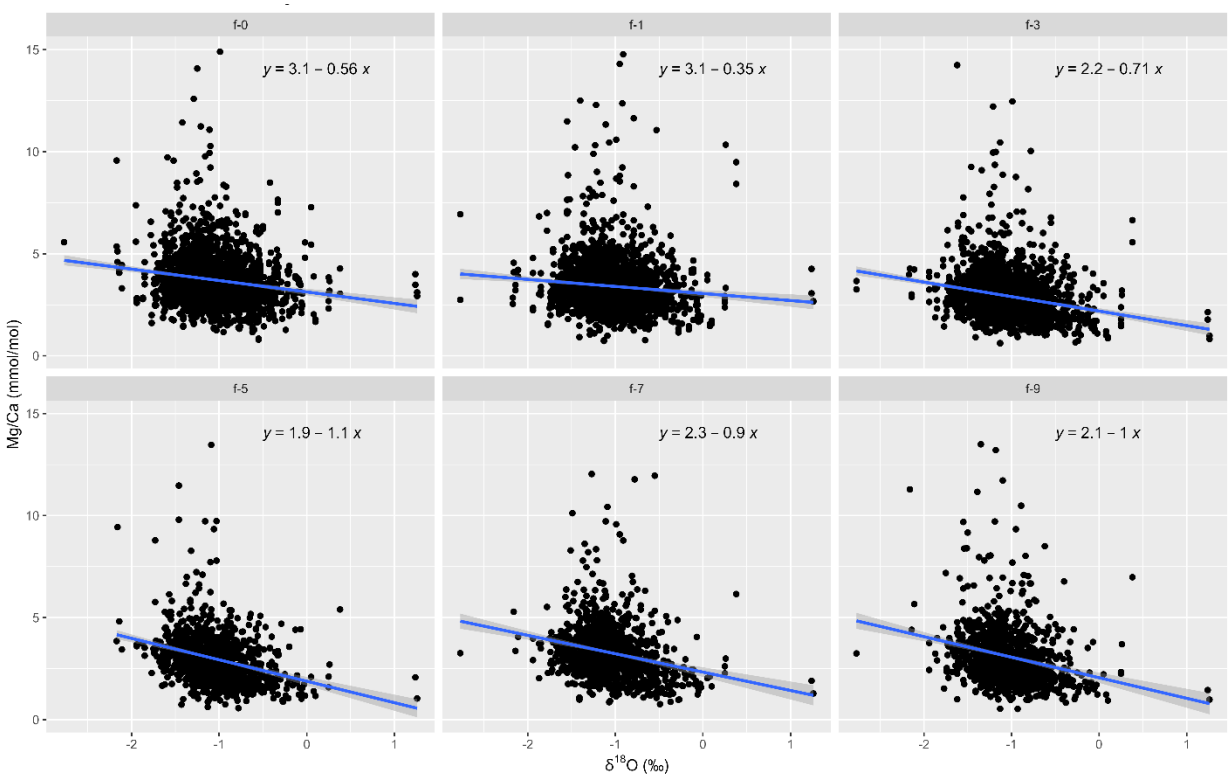


Figure 5.6: Individual chamber Mg/Ca (mmol/mol) and whole test $\delta^{18}O$ values for 1159 foraminifera with linear regression lines. The linear regression line equations show earlier grown chambers have steeper gradients (and therefore strongest correlation between both axis), with the steepest at chamber f-5. Chambers grown towards the end of life have regression lines with shallower gradients, with chamber f-1 the flattest. Line equation for chamber (f-0) $y = 3.1 - 0.56x$, (f-1) $y = 3.1 - 0.35x$, (f-3) $y = 2.2 - 0.7x$, (f-5) $y = 1.9 - 1.1x$, (f-7) $y = 2.3 - 0.9x$, and (f-9) $y = 2.1 - 1x$.

Chamber specific predicted log(Mg/Ca) values with fixed effects confidence intervals for each chamber when $\delta^{18}O$ is -2‰, 0‰, and 2‰ as estimated by a linear mixed model are shown in Figure 5.7. The overlapping error bars for each chamber estimate with all other chambers in the panel suggests that under these conditions (likely warmer temperatures), the Mg/Ca ratios across different chambers do not show significant variation, implying a uniform response across the test. This uniformity contrasts with the more distinct differences observed at -1‰ and 0‰ $\delta^{18}O$ values, where environmental conditions have a more varied impact on different parts of the test. At 0‰ and 2‰ $\delta^{18}O$ values, the final (f-0) and penultimate (f-1) chambers show elevated Mg/Ca ratio estimates compared to earlier chambers in the same panel. In panels (b) and (c), we identify two groupings: chambers f-0

and f-1 with elevated Mg/Ca ratio estimates and chambers f-3, f-5, f-7, and f-9 with lower estimates. The error bars between groups are separate and non-overlapping, suggesting a significant difference in the range between earlier and later grown chambers.

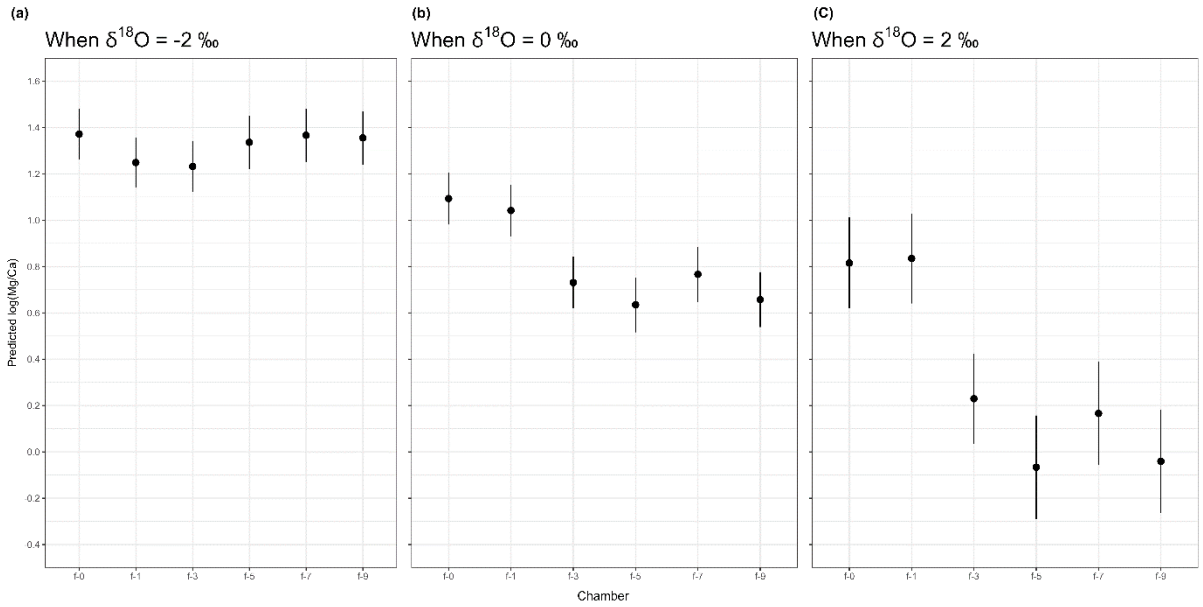


Figure 5.7: Mg/Ca log prediction fixed effects confidence intervals for each chamber when $\delta^{18}O$ is (a) -2‰, (b) 0‰ and (c) 2‰.

I built two models to test if Mg/Ca and $\delta^{18}O$ were directly correlated and where in the foraminifer's ontogeny the individual chamber Mg/Ca ratios most closely correlate with the measured whole test $\delta^{18}O$ values (Figure 5.8). The $\delta^{18}O$~MgCa (red) model predicts the scaled $\delta^{18}O$ ratio using the scaled logarithm of the Mg/Ca ratio and includes random effects for species and age to account for the variability due to the four species and geological ages of the individual foraminifers. The MgCa~$\delta^{18}O$ (teal) model predicts the scaled logarithm of the Mg/Ca ratio using the scaled $\delta^{18}O$ values and includes the same random effects.

I show the mean model estimates when Mg/Ca is the response variable (teal) are slightly stronger than their equivalent $\delta^{18}O$ response model (red), with both model means within the error bar ranges for each chamber. Chamber f-5 demonstrates the strongest correlation between chamber resolution Mg/Ca and whole test $\delta^{18}O$, with a scaled response of -2.9 for the Mg/Ca model (teal). With chamber f-5 showing the strongest correlation between Mg/Ca ratios and whole test $\delta^{18}O$ values, we suggest that this chamber is most representative of the

entire fossilised test. The penultimate (f-1) chamber shows the weakest correlation strength approaching -0.07 (0 = no correlation) and the final chamber also demonstrating weaker correlations. The earlier grown chambers f-9, f-7, f-3 show stronger correlations, however the error bars for these chambers with each model do overlap (as do f-5 and f-9) — suggesting it is not possible to distinguish between model strengths but are substantially separated from the final and penultimate chambers, which themselves have overlapping error bars.

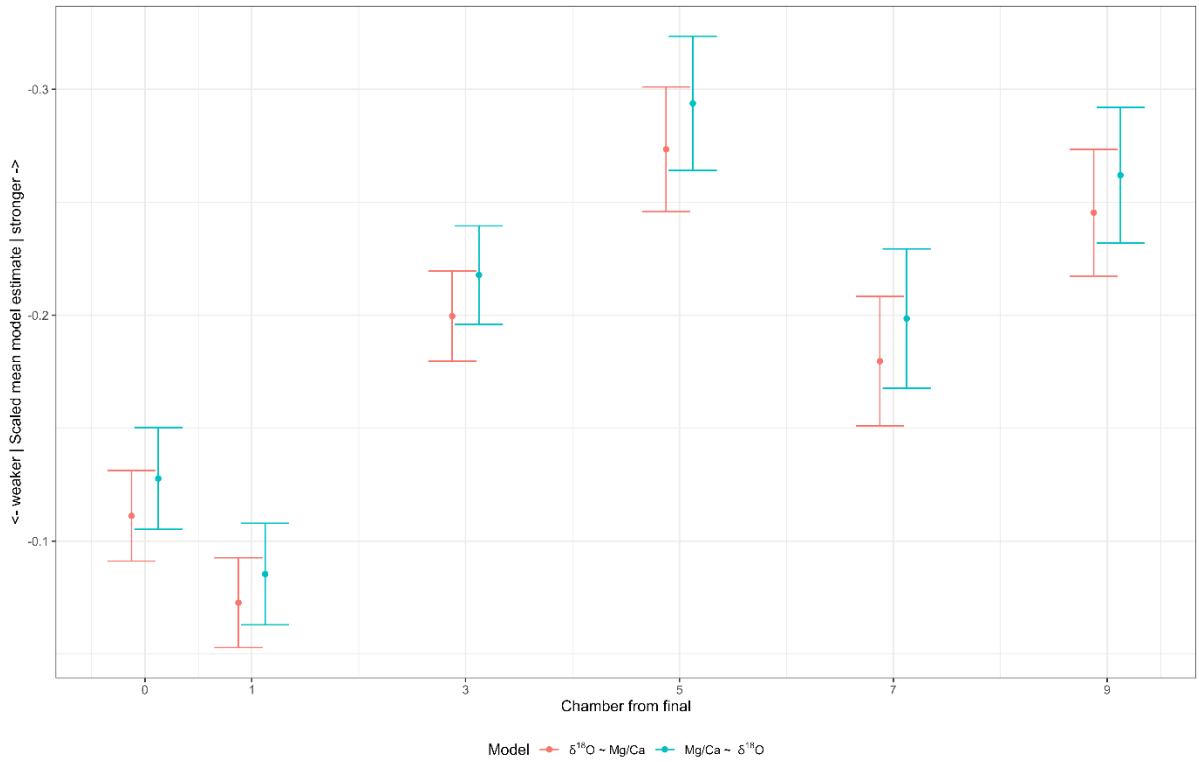


Figure 5.8: Comparison of scaled mean model estimates by measured chambers from final (f-0, f-1, f-3, f-5, f-7, and f-9). I ran the model in two directions: $\delta^{18}O$ ~ Mg/Ca (in red) as stable isotope response to trace elements and as Mg/Ca ~ $\delta^{18}O$ (in teal) as trace elements response to stable isotopes to identify if the model estimates were similar. Chamber f-5 shows the strongest correlation between scaled mean estimates. The inverse relationship between Mg/Ca and $\delta^{18}O$ causes the scaled mean model estimates to become more negative as the correlation becomes stronger.

I compare the chamber specific Mg/Ca ratios using this study's data set with whole test Mg/Ca ratios from literature across a range from -2.5‰ to 2.5‰ $\delta^{18}O$ values (Figure 5.9). We identified that chamber f-5 was the closest match to Anand *et al.* (2003) and Regenberg *et al.*

(2009) warm published values between -0.9‰ to 2.5‰ $\delta^{18}O$. Chambers f-0 and f-1 showed almost no correlation to any of the presented literature values.

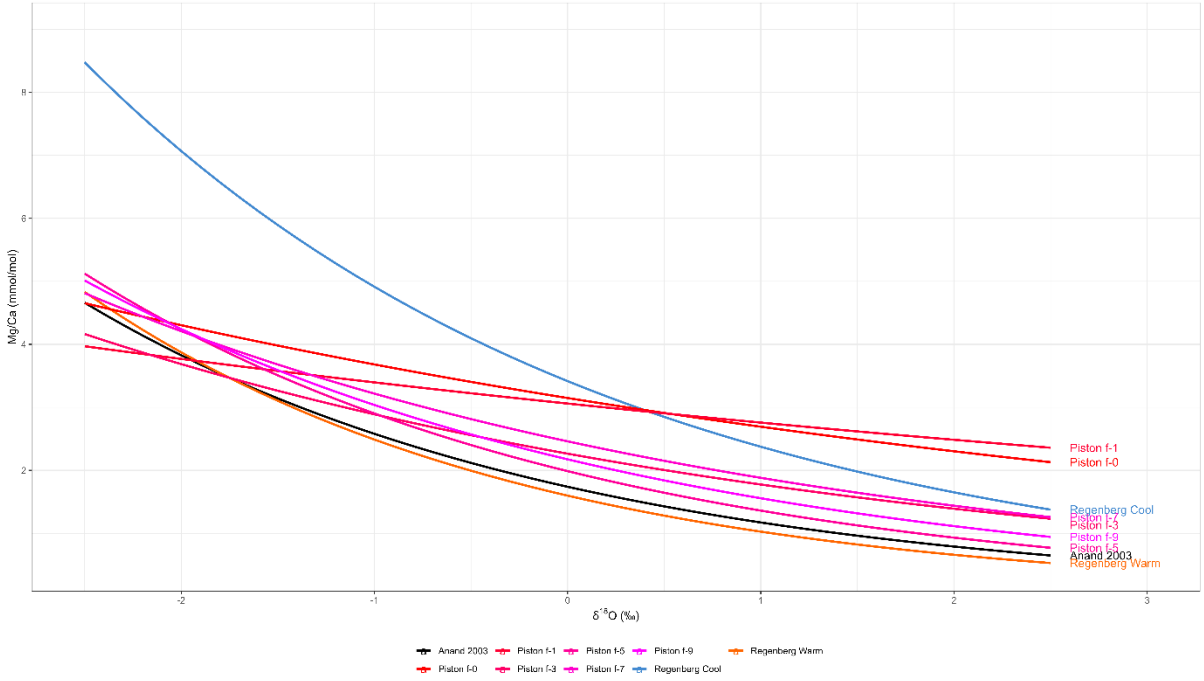


Figure 5.9: Comparison of Mg/Ca ratios against $\delta^{18}O$ values from this study and literature. This study is labelled "Piston" and coloured from red (chamber f-0) to pink (chamber f-9) where the trace elements were measured at individual chambers using LA-ICP-MS. Values from Anand *et al.* (2003) ten planktonic species calibration are presented in black where the trace elements were measured as whole test solution based ICP-MS. Planktonic warm species (orange) and cold species (blue) palaeothermometry calibrations are from Regenberg *et al.* (2009).

Chapter 6 Conclusion

1. New high-throughput workflows

My development of novel high-throughput workflows for scanning foraminifera by X-ray computed tomography has provided an unprecedented ability to extract functional ecological traits with fine resolution. These workflows link geochemistry with biology by revealing detailed morphological traits, such as pore distribution, and their relationship with environmental factors like oxygen minimum zones, temperature, and lived environment. By integrating morphology and environmental data, I deepen understanding of ecological adaptations and functional biology, offering novel insights into how foraminiferal traits respond to environmental gradients.

2. Reduce analytical uncertainty with new LA-ICP-MS data reduction software

LABLASTER is novel, open-source software for the R environment integrates improved methodological and statistical approaches into an accessible platform for the scientific community. By automating the detection of laser ablation endpoints and improving reproducibility, LABLASTER reduces subjectivity and analytical errors. Its open-source nature encourages collaboration and adaptation, enabling broader use of high-precision methodologies in geochemistry and further advancing open science principles in palaeoceanographic research. When LABLASTER is used during data process and analysis, the outputs reduce trace element ratio uncertainties by 60%, improving the precision of geochemical measurements. This reduction enhances the reliability of individual measurements, allowing for more accurate reconstructions of environmental conditions. By addressing measurement uncertainty, these approaches establish a new benchmark for trace element analysis, strengthening confidence in geochemical proxies and their ecological and climatic interpretations.

3. Ontogenetic Chamber Thickness Changes

My portfolio of papers demonstrates that foraminiferal chamber wall thickness decreases through ontogeny, with earlier-grown chambers being consistently thicker than the penultimate and final chambers in all measured species. These findings illuminate patterns of calcification during growth and provide a vital context for interpreting morphological variability within and across species. I reveal connections between chamber development (growth rates, chamber volumes, morphometrics) with environmental proxies (Mg/Ca and $\delta^{18}O$) to refine an understanding of foraminiferal life history.

4. Geochemical Mg/Ca $\delta^{18}O$ correlations in earlier chambers

Combining my methodological developments and unprecedentedly large dataset of individual specimen stable isotope ratios ($\delta^{18}O$) and individual chamber trace element (Mg/Ca) that go back further in ontological growth and for more individuals provides enhanced resolution for detecting geochemical variations across life stages and building powerful mathematical models. I evidence the relationship between Mg/Ca ratios and $\delta^{18}O$ values in the fossilised tests of four foraminifera species (*M. menardii*, *M. limbata*, *M. exilis*, and *M. pertenuis*) across ontogeny and geological time to better understand their response to environmental and climate change. My findings show that earlier chambers are characterised by thicker secondary calcite layers, contribute disproportionately to whole-test $\delta^{18}O$, and are most sensitive to temperature changes and exhibit the strongest correlations between Mg/Ca and $\delta^{18}O$. These results mature detailed chamber-level analyses into palaeoclimate studies, improving our ability to reconstruct past environmental conditions with greater precision and untangling so catch-all "vital effects".

5. Unsuitability of Final Chambers for Palaeoclimate Reconstructions

I show that the final and penultimate chambers to be the least accurate recorders of ambient environmental conditions, making them unsuitable for palaeoclimate reconstructions. This insight challenges the conventional focus on the final chamber for geochemical analysis and highlights the need for integrating ontogenetic perspectives when selecting chambers for analysis. By identifying the limitations of these chambers, this work directs future studies

toward more reliable portions of the foraminiferal test, advancing the field of palaeoceanography and improving palaeoclimate reconstruction accuracy.

List of References

Abu Bakar, M. H. E. b., Lowery, C. & Kearns, L. Investigating the Application of X-Ray Micro-CT Scan to Study 3D Morphological Variability of Guembelitria cretacea During the Post Cretaceous-Paleogene (K-Pg) Extinction Event. AGU Fall Meeting, 2023 San Francisco.

Anand, P., Elderfield, H. & Conte, M. H. (2003) Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography,* **18**.

Aze, T., Ezard, T. H., Purvis, A., Coxall, H. K., Stewart, D. R., Wade, B. S. & Pearson, P. N. (2011) A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. *Biological Reviews,* **86,** 900-927.

Barker, S., Cacho, I., Benway, H. & Tachikawa, K. (2005) Planktonic foraminiferal Mg/Ca as a proxy for past oceanic temperatures: a methodological overview and data compilation for the Last Glacial Maximum. *Quaternary Science Reviews,* **24,** 821-834.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software,* **67,** 1 - 48.

Bender, M. L., Lorens, R. B. & Williams, D. F. (1975) Sodium, Magnesium and Strontium in the Tests of Planktonic Foraminifera. *Micropaleontology,* **21,** 448-459.

Birch, H., Coxall, H. K., Pearson, P. N., Kroon, D. & O'Regan, M. (2013) Planktonic foraminifera stable isotopes and water column structure: Disentangling ecological signals. *Marine Micropaleontology,* **101,** 127-145.

Briguglio, A., Metscher, B. & Hohenegger, J. (2011) Growth rate biometric quantification by X-ray microtomography on larger benthic foraminifera: three-dimensional measurements push nummulitids into the fourth dimension. *Turkish Journal of Earth Sciences,* **20,** 683-699.

Brombacher, A., Searle-Barnes, A., Zhang, W. & Ezard, T. H. G. (2022) Analysing planktonic foraminiferal growth in three dimensions with foram3D: an R package for automated trait measurements from CT scans. *J. Micropalaeontol.,* **41,** 149-164.

Burke, J. E., Renema, W., Schiebel, R. & Hull, P. M. (2020) Three-dimensional analysis of inter- and intraspecific variation in ontogenetic growth trajectories of planktonic foraminifera. *Marine Micropaleontology,* **155,** 101794.

Caromel, A. G. M., Schmidt, D. N., Fletcher, I. & Rayfield, E. J. (2016) Morphological Change During The Ontogeny Of The Planktic Foraminifera. *J. Micropalaeontol.,* **35,** 2-19.

Chaisson, W. P. & Pearson, P. (1997) Planktonic foraminifer biostratigraphy at Site 925: Middle Miocene-Pleistocene. *Proceedings of the Ocean Drilling Program, Scientific Results,* **154,** 3-31.

Chalk, T. B., Standish, C. D., D'Angelo, C., Castillo, K. D., Milton, J. A. & Foster, G. L. (2021) Mapping coral calcification strategies from in situ boron isotope and trace element measurements of the tropical coral Siderastrea siderea. *Scientific Reports,* **11,** 472.

Chave, K. E. (1954) Aspects of the Biogeochemistry of Magnesium 1. Calcareous Marine Organisms. *The Journal of Geology,* **62,** 266-283.

Cornuault, P., Westerhold, T., Pälike, H., Bickert, T., Baumann, K. H. & Kucera, M. (2023) Nature and origin of variations in pelagic carbonate production in the tropical ocean since the mid-Miocene (ODP Site 927). *Biogeosciences,* **20,** 597-618.

Curry, W., Shackleton, N. & Richter, C. 9. LEG 154 SYNTHESIS[1]. Proceedings of the Ocean Drilling Program: Initial report, 1994. The Program, 421.

de Vargas, C., Norris, R., Zaninetti, L., Gibb, S. W. & Pawlowski, J. (1999) Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proceedings of the National Academy of Sciences,* **96,** 2864-2868.

Dragonfly (2022). Dragonfly. 2022.1 ed. http://www.theobjects.com/dragonfly: Object Research Systems

du Plessis, A., Broeckhoven, C., Guelpa, A. & le Roux, S. G. (2017) Laboratory x-ray micro-computed tomography: a user guideline for biological samples. *GigaScience,* **6**.

Dueñas-Bohórquez, A., da Rocha, R. E., Kuroyanagi, A., de Nooijer, L. J., Bijma, J. & Reichart, G.-J. (2011) Interindividual variability and ontogenetic effects on Mg and Sr incorporation in the planktonic foraminifer Globigerinoides sacculifer. *Geochimica et Cosmochimica Acta,* **75,** 520-532.

Eggins, S., De Deckker, P. & Marshall, J. (2003) Mg/Ca variation in planktonic foraminifera tests: implications for reconstructing palaeo-seawater temperature and habitat migration. *Earth and Planetary Science Letters,* **212,** 291-306.

Eggins, S. M., Sadekov, A. & De Deckker, P. (2004) Modulation and daily banding of Mg/Ca in Orbulina universa tests by symbiont photosynthesis and respiration: a complication for seawater thermometry? *Earth and Planetary Science Letters,* **225,** 411-419.

Elderfield, H. & Ganssen, G. (2000) Past temperature and δ18O of surface ocean waters inferred from foraminiferal Mg/Ca ratios. *Nature,* **405,** 442-445.

Elderfield, H., Vautravers, M. & Cooper, M. (2002) The relationship between shell size and Mg/Ca, Sr/Ca, δ18O, and δ13C of species of planktonic foraminifera. *Geochemistry, Geophysics, Geosystems,* **3,** 1-13.

Emiliani, C. (1955) Pleistocene temperatures. *The Journal of geology,* **63,** 538-578.

Erez, J. (1978) Vital effect on stable-isotope composition seen in foraminifera and coral skeletons. *Nature,* **273,** 199-202.

Erez, J. (2003) The Source of Ions for Biomineralization in Foraminifera and Their Implications for Paleoceanographic Proxies. *Reviews in Mineralogy and Geochemistry,* **54,** 115-149.

Farmer, E. C., Kaplan, A., de Menocal, P. B. & Lynch-Stieglitz, J. (2007) Corroborating ecological depth preferences of planktonic foraminifera in the tropical Atlantic with the stable oxygen isotope ratios of core top specimens. *Paleoceanography,* **22**.

Fehrenbacher, J., Marchitto, T. & Spero, H. J. (2020) Comparison of Laser Ablation and Solution-Based ICP-MS Results for Individual Foraminifer Mg/Ca and Sr/Ca Analyses. *Geochemistry, Geophysics, Geosystems,* **21,** e2020GC009254.

Fehrenbacher, J. S., Hupp, B. N., Branson, O., Evans, D., Foster, G. L., Glock, N., Thirumalai, K. & Wycech, J. (2024) INDIVIDUAL FORAMINIFERAL ANALYSES: A REVIEW OF CURRENT AND EMERGING GEOCHEMICAL TECHNIQUES. *Journal of Foraminiferal Research,* **54,** 312-331.

Fehrenbacher, J. S., Russell, A. D., Davis, C. V., Gagnon, A. C., Spero, H. J., Cliff, J. B., Zhu, Z. & Martin, P. (2017) Link between light-triggered Mg-banding and chamber formation in the planktic foraminifera Neogloboquadrina dutertrei. *Nature Communications,* **8,** 15441.

Fehrenbacher, J. S., Spero, H. J., Russell, A. D., Vetter, L. & Eggins, S. (2015) Optimizing LA-ICP-MS analytical procedures for elemental depth profiling of foraminifera shells. *Chemical Geology,* **407-408,** 2-9.

Fox, L., Stukins, S., Hill, T. & Miller, C. G. (2020) Quantifying the Effect of Anthropogenic Climate Change on Calcifying Plankton. *Scientific Reports,* **10,** 1620.

Franke, J. G. & Donner, R. V. (2019) Correlating paleoclimate time series: Sources of uncertainty and potential pitfalls. *Quaternary science reviews,* **212,** 69-79.

Görög, Á., Szinger, B., Tóth, E. & Viszkok, J. (2012) Methodology of the micro-computer tomography on foraminifera. *Palaeontologia Electronica,* **15**.

Gray, W. R. & Evans, D. (2019) Nonthermal Influences on Mg/Ca in Planktonic Foraminifera: A Review of Culture Studies and Application to the Last Glacial Maximum. *Paleoceanography and Paleoclimatology,* **34,** 306-315.

Hathorne, E. C., Alard, O., James, R. H. & Rogers, N. W. (2003) Determination of intratest variability of trace elements in foraminifera by laser ablation inductively coupled plasma-mass spectrometry. *Geochemistry, Geophysics, Geosystems,* **4**.

Hathorne, E. C., James, R. H. & Lampitt, R. S. (2009) Environmental versus biomineralization controls on the intratest variation in the trace element composition of the planktonic foraminifera G. inflata and G. scitula. *Paleoceanography,* **24**.

Hemleben, C., Spindler, M. & Anderson, O. R. (1989) Modern Planktonic Foraminifera. *Modern Planktonic Foraminifera.***,** 363.

Hipsley, C. A., Aguilar, R., Black, J. R. & Hocknull, S. A. (2020) High-throughput microCT scanning of small specimens: preparation, packing, parameters and post-processing. *Scientific Reports,* **10,** 13863.

Hönisch, B., Bijma, J., Russell, A. D., Spero, H. J., Palmer, M. R., Zeebe, R. E. & Eisenhauer, A. (2003) The influence of symbiont photosynthesis on the boron isotopic composition of foraminifera shells. *Marine Micropaleontology,* **49,** 87-96.

Hoogakker, B., Ishimura, T., de Nooijer, L., Rathburn, A. & Schmiedl, G. (2024) A review of benthic foraminiferal oxygen and carbon isotopes. *Quaternary Science Reviews,* **342,** 108896.

Hsiang, A. Y., Brombacher, A., Rillo, M. C., Mleneck-Vautravers, M. J., Conn, S., Lordsmith, S., Jentzen, A., Henehan, M. J., Metcalfe, B., Fenton, I. S., Wade, B. S., Fox, L., Meilland, J., Davis, C. V., Baranowski, U., Groeneveld, J., Edgar, K. M., Movellan, A., Aze, T., Dowsett, H. J., Miller, C. G., Rios, N. & Hull, P. M. (2019) Endless Forams: >34,000 Modern Planktonic Foraminiferal Images for Taxonomic Training and Automated Species Recognition Using Convolutional Neural Networks. *Paleoceanography and Paleoclimatology,* **34,** 1157-1177.

Huber, B. T., Bijma, J. & Darling, K. (1997) Cryptic speciation in the living planktonic foraminifer Globigerinella siphonifera (d'Orbigny). *Paleobiology,* **23,** 33-62.

Hunt, G. (2004) Phenotypic Variation in Fossil Samples: Modeling the Consequences of Time-Averaging. *Paleobiology,* **30,** 426-443.

Hupp, B. N. & Fehrenbacher, J. S. (2024) INTRATEST TRACE ELEMENT VARIABILITY IN POLAR AND SUBPOLAR PLANKTIC FORAMINIFERA: INSIGHTS INTO VITAL EFFECTS, ONTOGENY, AND BIOMINERALIZATION PROCESSES. *Journal of Foraminiferal Research,* **54,** 355-374.

Iwasaki, S., Kimoto, K., Sasaki, O., Kano, H., Honda, M. C. & Okazaki, Y. (2015) Observation of the dissolution process of Globigerina bulloides tests (planktic foraminifera) by X-ray microcomputed tomography. *Paleoceanography,* **30,** 317-331.

Jana, D., Torres, M., Evans, K., Jayan, A. K. & Thirumalai, K. (2024) Paired X-Ray Micro CT Scanning and Individual Foraminifera Isotopic Analysis Reveal (De)coupled Changes in Carbonate Preservation and Temperature. *Paleoceanography and Paleoclimatology,* **39,** e2023PA004821.

Jochum, K. P., Garbe-Schönberg, D., Veter, M., Stoll, B., Weis, U., Weber, M., Lugli, F., Jentzen, A., Schiebel, R., Wassenburg, J. A., Jacob, D. E. & Haug, G. H. (2019) Nano-Powdered Calcium Carbonate Reference Materials: Significant Progress for Microanalysis? *Geostandards and Geoanalytical Research,* **43,** 595-609.

Jochum, K. P., Weis, U., Stoll, B., Kuzmin, D., Yang, Q., Raczek, I., Jacob, D. E., Stracke, A., Birbaum, K., Frick, D. A., Günther, D. & Enzweiler, J. (2011) Determination of Reference Values for NIST SRM 610–617 Glasses Following ISO Guidelines. *Geostandards and Geoanalytical Research,* **35,** 397-429.

Johnstone, H. J. H., Schulz, M., Barker, S. & Elderfield, H. (2010) Inside story: An X-ray computed tomography method for assessing dissolution in the tests of planktonic foraminifera. *Marine Micropaleontology,* **77,** 58-70.

Johnstone, H. J. H., Yu, J., Elderfield, H. & Schulz, M. (2011) Improving temperature estimates derived from Mg/Ca of planktonic foraminifera using X-ray computed tomography–based dissolution index, XDX. *Paleoceanography,* **26**.

Katz, M., Cramer, B., Franzese, A., Hönisch, B., Miller, K., Rosenthal, Y. & Wright, J. (2010) Traditional and Emerging Geochemical Proxies in Foraminifera. *Journal of Foraminiferal Research,* **40,** 165-192.

Kearns, L. E., Searle-Barnes, A., Foster, G. L., Milton, J. A., Standish, C. D. & Ezard, T. H. G. (2023) The Influence of Geochemical Variation Among Globigerinoides ruber Individuals on Paleoceanographic Reconstructions. *Paleoceanography and Paleoclimatology,* **38,** e2022PA004549.

Kendall, S., Gradstein, F., Jones, C., Lord, O. T. & Schmidt, D. N. (2020) Ontogenetic disparity in early planktic foraminifers. *J. Micropalaeontol.,* **39,** 27-39.

Kennett, J. P. & Srinivasan, M. S. (1983). Neogene Planktonic Foraminifera: A Phylogenetic Atlas. Hutchinson Ross.

Lisiecki, L. E. & Raymo, M. E. (2005) A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography,* **20**.

Marchant, R., Tetard, M., Pratiwi, A., Adebayo, M. & de Garidel-Thoron, T. (2020) Automated analysis of foraminifera fossil records by image classification using a convolutional neural network. *J. Micropalaeontol.,* **39,** 183-202.

Mulqueeney, J. M., Searle-Barnes, A., Brombacher, A., Sweeney, M., Goswami, A. & Ezard, T. H. G. (2024) How many specimens make a sufficient training set for automated three-dimensional feature extraction? *Royal Society Open Science,* **11,** rsos.240113.

Osborne, E. B., Umling, N. E., Bizimis, M., Buckley, W., Sadekov, A., Tappa, E., Marshall, B., Sautter, L. R. & Thunell, R. C. (2020) A Sediment Trap Evaluation of B/Ca as a Carbonate System Proxy in Asymbiotic and Nondinoflagellate Hosting Planktonic Foraminifera. *Paleoceanography and Paleoclimatology,* **35,** e2019PA003682.

Pearson, P. N. (2012) Oxygen Isotopes in Foraminifera: Overview and Historical Review. *The Paleontological Society Papers,* **18,** 1-38.

Regenberg, M., Steph, S., Nürnberg, D., Tiedemann, R. & Garbe-Schönberg, D. (2009) Calibrating Mg/Ca ratios of multiple planktonic foraminiferal species with δ18O-calcification temperatures: Paleothermometry for the upper water column. *Earth and Planetary Science Letters,* **278,** 324-336.

Rosenthal, Y., Bova, S. & Zhou, X. (2022) A User Guide for Choosing Planktic Foraminiferal Mg/Ca-Temperature Calibrations. *Paleoceanography and Paleoclimatology,* **37,** e2022PA004413.

Rosenthal, Y. & Lohmann, G. P. (2002) Accurate estimation of sea surface temperatures using dissolution-corrected calibrations for Mg/Ca paleothermometry. *Paleoceanography,* **17,** 16-1-16-6.

Schiebel, R. & Hemleben, C. (2017). Planktic foraminifers in the modern ocean. Springer.

Schmidt, D. N., Rayfield, E. J., Cocking, A. & Marone, F. (2013) Linking evolution and development: Synchrotron Radiation X-ray tomographic microscopy of planktic foraminifers. *Paleontology,* **56,** 741-749.

Schmidt, D. N., Thierstein, H. R. & Bollmann, J. (2004) The evolutionary history of size variation of planktic foraminiferal assemblages in the Cenozoic. *Palaeogeography, Palaeoclimatology, Palaeoecology,* **212,** 159-180.

Schöne, B. R., Zhang, Z., Radermacher, P., Thébault, J., Jacob, D. E., Nunn, E. V. & Maurer, A.-F. (2011) Sr/Ca and Mg/Ca ratios of ontogenetically old, long-lived bivalve shells (Arctica islandica) and their function as paleotemperature proxies. *Palaeogeography, Palaeoclimatology, Palaeoecology,* **302,** 52-64.

Searle-Barnes, A., Milton, J. A., Standish, C. D., Foster, G. L. & Ezard, T. H. G. (2023) Laser ablation mass spectrometry blast through detection in R. *Rapid Communications in Mass Spectrometry,* **37,** e9533.

Shackleton, N. (1967) Oxygen isotope analyses and Pleistocene temperatures re-assessed. *Nature,* **215,** 15-17.

Speijer, R. P., Van Loo, D., Masschaele, B., Vlassenbroeck, J., Cnudde, V. & Jacobs, P. (2008) Quantifying foraminiferal growth with high-resolution X-ray computed tomography: New opportunities in foraminiferal ontogeny, phylogeny, and paleoceanographic applications. *Geosphere,* **4,** 760-763.

Spero, H. J., Bijma, J., Lea, D. W. & Bemis, B. E. (1997) Effect of seawater carbonate concentration on foraminiferal carbon and oxygen isotopes. *Nature,* **390,** 497-500.

Spero, H. J., Eggins, S. M., Russell, A. D., Vetter, L., Kilburn, M. R. & Hönisch, B. (2015) Timing and mechanism for intratest Mg/Ca variability in a living planktic foraminifer. *Earth and Planetary Science Letters,* **409,** 32-42.

Spero, H. J., Lerche, I. & Williams, D. F. (1991) Opening the carbon isotope 'vital effect' black box, 2, Quantitative model for interpreting foraminiferal carbon isotope data. *Paleoceanography,* **6,** 639-655.

Spero, H. J. & Williams, D. F. (1989) Opening the carbon isotope “vital effect” black box 1. Seasonal temperatures in the euphotic zone. *Paleoceanography,* **4,** 593-601.

Team, R. D. C. (2010) R: A language and environment for statistical computing. *(No Title)*.

Vanadzina, K. & Schmidt, D. N. (2022) Developmental change during a speciation event: evidence from planktic foraminifera. *Paleobiology,* **48,** 120-136.

Wickham, H. (2007) Reshaping Data with the reshape Package. *Journal of Statistical Software,* **21,** 1 - 20.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L. & Hester, J. (2019) Welcome to the Tidyverse. *Journal of open source software,* **4,** 1686.

Wilkens, R. H., Westerhold, T., Drury, A. J., Lyle, M., Gorgas, T. & Tian, J. (2017) Revisiting the Ceara Rise, equatorial Atlantic Ocean: isotope stratigraphy of ODP Leg 154 from 0 to 5 Ma. *Climate of the Past,* **13,** 779-793.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., ’t Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. & Mons, B. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data,* **3,** 160018.

Withers, P. J., Bouman, C., Carmignato, S., Cnudde, V., Grimaldi, D., Hagen, C. K., Maire, E., Manley, M., Du Plessis, A. & Stock, S. R. (2021) X-ray computed tomography. *Nature Reviews Methods Primers,* **1,** 18.

Wolf-Gladrow, D. A., Bijma, J. & Zeebe, R. E. (1999) Model simulation of the carbonate chemistry in the microenvironment of symbiont bearing foraminifera. *Marine Chemistry,* **64,** 181-198.

Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *science,* **292,** 686-693.

Zeebe, R. E. (1999) An explanation of the effect of seawater carbonate concentration on foraminiferal oxygen isotopes. *Geochimica et Cosmochimica Acta,* **63,** 2001-2007.

Zhang, W., Ezard, T., Searle-Barnes, A., Brombacher, A., Katsamenis, O. & Nixon, M. Towards understanding speciation by automated extraction and description of 3D foraminifera stacks. 2020 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), 2020. IEEE, 30-33.

J. Micropalaeontology, 41, 149–164, 2022
https://doi.org/10.5194/jm-41-149-2022

Journal of Micropalaeontology
Open Access

Analysing planktonic foraminiferal growth in three dimensions with foram3D: an R package for automated trait measurements from CT scans

Anieke Brombacher[1], **Alex Searle-Barnes**[1], **Wenshu Zhang**[2], **and Thomas H. G. Ezard**[1]

[1]School of Ocean and Earth Science, University of Southampton, Southampton, SO14 3ZH, United Kingdom
[2]Department of Computer Science, Cardiff Metropolitan University, Cardiff, CF5 2YB, United Kingdom

Correspondence: Anieke Brombacher (anieke.brombacher@soton.ac.uk)

Received: 10 June 2022 – Revised: 31 August 2022 – Accepted: 30 September 2022 – Published: 1 November 2022

Abstract. Foraminifera are one of the few taxa that preserve their entire ontogeny in their fossilised remains. Revealing this ontogeny through micro-computed tomography (CT) of fossil planktonic foraminifera has greatly improved our understanding of their life history and allows accurate quantification of total shell volume, growth rates and developmental constraints throughout an individual's life. Studies using CT scans currently mainly focus on chamber size, but the wealth of three-dimensional data generated by CT scans has the potential to reconstruct complete growth trajectories. Here we present an open-source R package to analyse growth in three-dimensional space. Using only the centroid xyz coordinates of every chamber, the functions determine the growth sequence and check that chambers are in the correct order. Once the order of growth has been verified, the functions calculate distances and angles between subsequent chambers, determine the total number of whorls and the number of chambers in the final whorl at the time each chamber was built, and, for the first time, quantify trochospirality. The applications of this package will enable repeatable analysis of large data sets and quantification of key taxonomic traits and ultimately provide new insights into the effects of ontogeny on evolution.

1 Introduction

Foraminifera are one of the few taxa that preserve an entire ontogeny in their fossilised remains. This preservation enables reconstructions of ontogenetic trajectories in deep time, a feature of particular interest to studies investigating the influence of development on long-term evolutionary change (Brombacher et al., 2022a). Developmental plasticity can influence phenotype frequency in a population through a process called genetic accommodation (West-Eberhard, 2003). A new environmental cue will cause a plastic trait to be expressed in a novel way, and if this new phenotype has a positive effect on fitness, it will likely be selected for, increasing the frequency of both the phenotypic and genetic components (West-Eberhard, 2005, 2003). Developmental plasticity has been argued to both drive and inhibit evolutionary innovation, but the majority of research is based on theoretical models (Dewitt et al., 1998; Murren et al., 2015; Price et al., 2003) and/or modern populations (e.g. Pigliucci et al., 2006; Beldade et al., 2011; Moczek et al., 2011; Pfennig et al., 2010) generally limited to a handful of generations. Fossils contain information on both macroevolutionary transitions and microevolutionary change, but the lack of data on juvenile states makes reconstructing developmental trajectories difficult and so limits our ability to infer the role of developmental plasticity on macroevolution. Planktonic foraminifera could shed new light on the developmental drivers of evolutionary change. Freely distributed methodological tools would facilitate this contribution. Here we present a new, open-source R package that automatically analyses three-dimensional foraminiferal growth trajectories from micro-computed tomography (CT) scans. Our package enables fuller use of the incredible richness from x-ray CT and thus more comprehensive understanding of the role ontogeny plays in determining the size and shape of adult forms.

In the last decade, the application of micro-CT scanning to fossil planktonic foraminifera has greatly improved our un-

derstanding of their life history. This non-destructive, fully volumetric method has provided new insights in calcification (Iwasaki et al., 2019b; Todd et al., 2020; Fox et al., 2020) and dissolution (Iwasaki et al., 2019a, 2015; Johnstone et al., 2010, 2011) and allowed more accurate quantification of total shell volume (Speijer et al., 2008; Brigulgio et al., 2011; Brombacher et al., 2018; Zarkogiannis et al., 2019, 2020; Kendall et al., 2020; Burke et al., 2020), growth trajectories (Speijer et al., 2008; Brigulgio et al., 2011; Schmidt et al., 2013; Caromel et al., 2016, 2017; Kendall et al., 2020; Burke et al., 2020; Vanadzina and Schmidt, 2022) and shell density (Duan et al., 2021). Studies of growth, however, remain largely restricted to univariate analyses of growth rates (Caromel et al., 2016, 2017; Schmidt et al., 2013; Kendall et al., 2020; Speijer et al., 2008; Brigulgio et al., 2011; Burke et al., 2020; Vanadzina and Schmidt, 2022) or shell volume and aspect ratio at different ontogenetic stages (Caromel et al., 2016, 2017; Schmidt et al., 2013; Kendall et al., 2020). The wealth of three-dimensional data generated by CT scans has the potential to reconstruct growth trajectories in three dimensions, for example by embedding the data in three-dimensional growth as demonstrated by Caromel et al. (2017) and Morard et al. (2019). Multivariate growth trajectories could substantially improve our understanding of developmental constraints on planktonic foraminifera growth and form, but this aspect of CT scans remains largely unused.

We present an open-source R package (available on GitHub, https://github.com/AniekeBrombacher/foram3D, last access: 31 August 2022) that automatically analyses planktonic foraminifera growth in three-dimensional space using the chamber centroid xyz coordinates (Fig. 1). The functions calculate distances and angles between centroids to arrange chambers in order of growth, calculate distances and angles between subsequent chambers, determine the total number of whorls and the number of chambers in the final whorl at the time each chamber was built, and, for the first time, quantify trochospirality. Included in the package are example data sets of *Menardella limbata* and *Trilobatus sacculifer* specimens to illustrate function usage for contrasting morphologies. The goal of this study and the intended applications of this package are envisaged to enable repeatable analysis of large data sets and ultimately provide new insights into the effects of developmental processes on the evolution of planktonic foraminifera.

2 Functions

The functions of the foram3D package calculate distances and angles between centre points of individual chambers. Chamber centres can be represented by centroids (the geometric centre of a three-dimensional object) or the centre of mass (the point where the entire mass of an object is concentrated). The examples in this study were reconstructed by filling in chambers on individual CT-scan image slices, creating objects of uniform density (see Caromel et al., 2016, 2017; Kendall et al., 2020; Schmidt et al., 2013 for detailed methodology), for which centroid and centre of mass coordinates are identical (but note that these coordinates can differ for studies assigning different densities to chamber walls and cavities). No species-specific morphological assumptions are made. We illustrate the package functions using example *Menardella limbata* and *Trilobatus sacculifer* specimens.

2.1 Chamber ordering

The order of growth is essential for analyses of ontogeny. Current CT image analysis relies on the manual detection of chambers in individual slices, which are then put together to calculate chamber volume and centroid coordinates (e.g. Caromel et al., 2016, 2017; Schmidt et al., 2013; Kendall et al., 2020; Speijer et al., 2008; Brigulgio et al., 2011; Burke et al., 2020; Vanadzina and Schmidt, 2022). With chamber centroids plotted in three-dimensional space (see the code in the Supplement, lines 37–57 for interactive examples), foraminifera growth spirals are relatively easy to recognise, and chamber numbers can be assigned manually. However, ordering chambers this way is a time-consuming process that is prone to manual error, particularly for the smallest, earliest chambers. A repeatable automated process would speed up the process and reduce error.

2.1.1 Algorithm

Foraminifera grow in approximate logarithmic spirals with the distance between subsequent chambers increasing roughly exponentially (Signes et al., 1993). Starting from the earliest chambers, the algorithm finds the closest unassigned chamber to every chamber to determine the order of growth (see Fig. 2). The algorithm starts by finding the starting point of the spiral. The second chamber has always been found to be the smallest chamber (Brummer et al., 1986, 1987; Huang, 1981; Sverdlove and Be, 1985), so chamber 2 is assigned first (Figs. 2a, 3a). The first chamber can then be found by looking for the chamber closest to chamber 2 (Fig. 2b). The first and second chamber calcify together (Davis et al., 2020; Takagi et al., 2020) forming a flat plane as internal chamber wall, putting the centroid of the first chamber closer to the centroid of the second chamber than the third chamber. Chamber 3 is the next closest centroid to chamber 2 (Fig. 2c). From there on, when the first i chambers have been assigned, chamber $i+1$ is the closest remaining chamber to chamber i (Fig. 2d–f). This step repeats until all chambers are assigned.

Occasionally, the proloculus is not present in fossil specimens, for example due to dissolution of the earliest chambers (Johnstone et al., 2010; Iwasaki et al., 2015). In that case, the algorithm takes the smallest chamber as the starting point. This version of the algorithm is less reliable than the version with the proloculus, as size does not always increase between subsequent chambers. For example, in Fig. 3a, if only

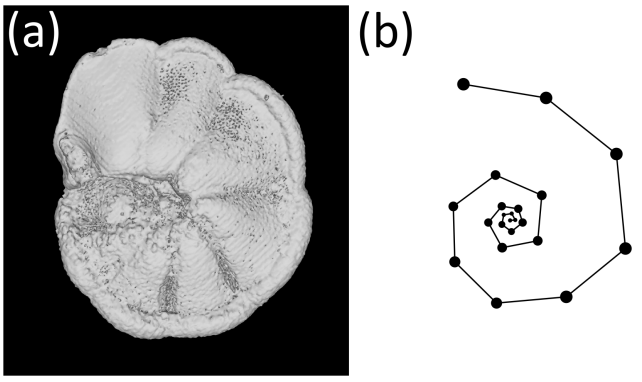


Figure 1. The exemplar *Menardella limbata* specimen released with the package. **(a)** CT scan of the exemplar *M. limbata* specimen. **(b)** Chamber centroids and growth spiral of the exemplar *M. limbata* specimen. Code to generate an interactive figure of **(b)** can be found in the code in the Supplement, lines 44–50.

chambers 15–22 were preserved, the algorithm would assign chamber 16 as the earliest chamber (rather than chamber 15). Generally, chamber size alone is not a reliable means to determine the order of growth. A number of species, such as the *Menardella* lineage, show reduced growth in the final chambers; small bullae on top of the final chamber are common, and occasionally unusually small chambers are found in random locations in the ontogeny (Duan et al., 2021). These deviations of the growth trajectory can have a profound effect on the resulting chamber order (Fig. 3b).

2.1.2 Usage

The function requires vectors with chamber x-, y- and z-centroid coordinates and chamber volume (code in the Supplement, lines 63–89). Its default version assumes that the proloculus is present but can be changed to proloculus=FALSE if necessary. It returns a data frame with the original data ordered by the newly assigned chamber numbers. See below for an example of the function usage in R. See Sect. A1 in Appendix A for function output.

```
with(Mlimbata, order.chambers(x=Cen
troidx, y=Centroidy, z=Centroidz,
V=Volume, proloculus=TRUE))
```

2.2 Checking the chamber order

Once chambers have been ordered (either by algorithm or manually), the R package contains a function that checks the chamber order and identifies unusual growth patterns to identify incorrectly ordered specimens. In foraminifera specimens with three or more chambers per whorl, the angle between subsequent chambers is $\geq 60°$. Therefore, sequences with angles smaller than 60° could be an indication of incorrectly ordered chambers. Common reasons for unusual sequences include manual error, faulty data or disrupted growth. The algorithm flags up any specimens that need additional inspection.

2.2.1 Algorithm: angles

Angles between subsequent chambers are calculated using the chamber.angles() function (code in the Supplement, lines 97–104). It requires vectors with chamber x-, y- and z-centroid coordinates and returns a vector with the angle every chamber makes with its previous two chambers (in degrees). Note that angles are generated from the third chamber onwards because the first and second chamber sit on a straight line.

2.2.2 Usage

The function requires vectors with chamber x-, y- and z-centroid coordinates. It returns a vector with the angle each chamber makes relative to the previous two previous chambers (see the code in the Supplement, lines 97–104). Note that at least three chambers are required for angle determination; therefore, no angles can be calculated for chambers 1 and 2. Code shown below is for the exemplar *M. limbata* specimen. See Fig. 5 and Sect. A2 for function output. See the code in the Supplement, line 104, for the output for *T. sacculifer*.

```
with(Mlimbata, chamber.angles(x=Cen
troidx, y=Centroidy, z=Centroidz))
```

2.2.3 Algorithm: chamber order check

To detect incorrectly ordered specimens, the algorithm calculates all angles α_i that chamber i makes with the previous two chambers $i-1$ and $i-2$ (Fig. 4). Angle α_3 is discarded if the proloculus is known to be present: the first three chambers form a triangle, and α_3 therefore forms an angle smaller than 60° with the proloculus and deuteroconch. In specimens where the first few chambers are absent, for example due to dissolution, all chamber angles are considered. Specimens with all angles larger than 60° are labelled as correctly ordered (Fig. 6a). Any specimens with one or more angles smaller than 60° are flagged to the user as potentially incorrectly ordered specimens.

Specimens with angles smaller than 60° either represent true negatives with one or more incorrectly ordered chambers (Fig. 6b) or specimens that experienced unusual growth resulting in sharp angles (Fig. 6c). Unusual growth can include a bulla or a smaller-than-expected chamber elsewhere in the sequence. Any manually ordered specimens flagged as incorrectly ordered can be run through the chamber-ordering algorithm to attempt to correct to sequence. If the sequence still returns sharp angles after ordering, the specimen needs to be inspected manually. In a preliminary analysis of 60 specimens, manual inspection was necessary for 5 % of the

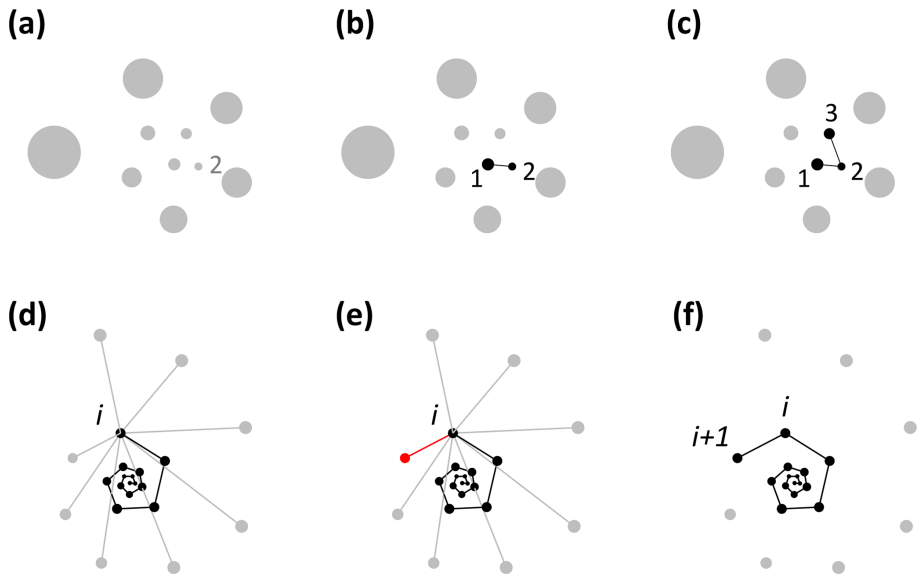


Figure 2. Visualisation of the chamber-ordering algorithm in the exemplar *Menardella limbata* data released with the package. **(a–c)** Discs represent chamber centroids of the earliest chambers, with disc size indicating chamber size. **(a)** The smallest chamber is chamber 2. **(b)** The closest chamber to chamber 2 is chamber 1. **(c)** The next closest chamber to chamber 2 is chamber 3. **(d–f)** Finding chamber $i + 1$ when chambers 1 to i have been assigned. Black and grey dots represent centroids of ordered and unordered chambers, respectively. **(d)** Calculation of the distance from chamber i to all remaining chambers. **(e)** The closest unassigned chamber (red) to chamber i is chamber $i + 1$ **(f)**.

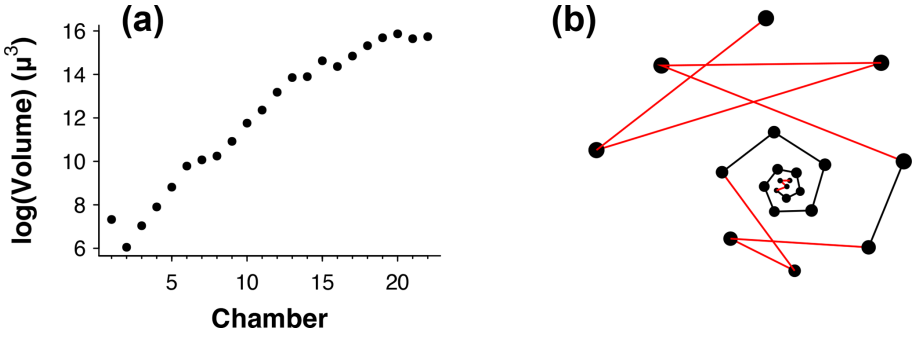


Figure 3. **(a)** Chamber size through ontogeny and **(b)** chambers ordered (incorrectly) by size only from the exemplar *Menardella limbata* data released with the package. The code in the Supplement, lines 67–80, can be used to generate an interactive plot of **(b)**.

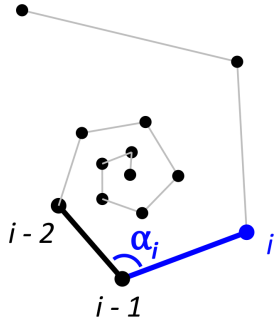


Figure 4. Chamber i is added at angle α_i relative to the previous two chambers.

specimens. In all cases unusual growth was the cause of the ordering issues.

2.2.4 Usage

The function requires vectors with chamber x-, y- and z-centroid coordinates. The default option assumes that the proloculus is present, which can be changed to proloculus=FALSE if necessary. It returns a vector with either "correct order" if all angles are larger than 60° or "check chamber order" if at least one chamber angle is smaller than 60°. See the code in the Supplement, lines 67 and 74, for examples. See Sect. A3 for function output.

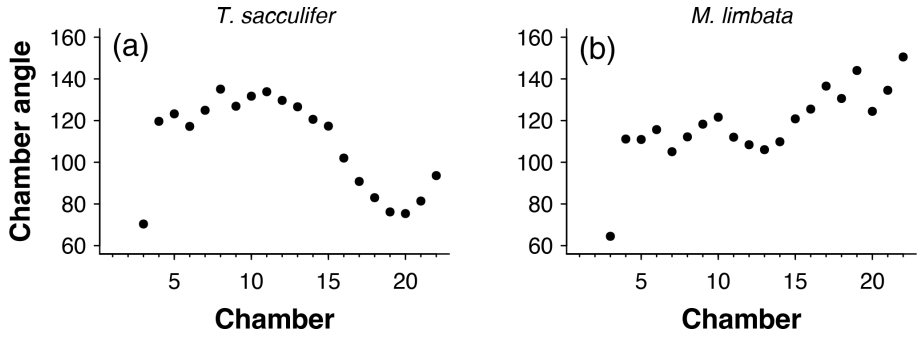


Figure 5. Angles between subsequent chambers for **(a)** *Trilobatus sacculifer* and **(b)** *Menardella limbata*.

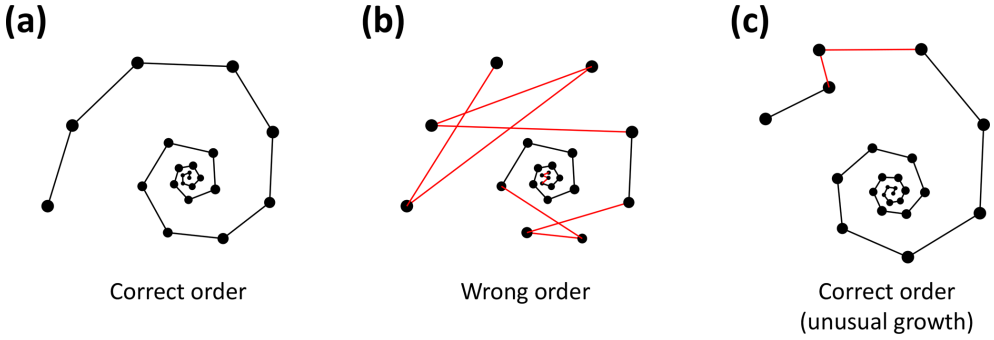


Figure 6. Three examples of correct and incorrect chamber sequences using the exemplar *Menardella limbata* data released with the package. Specimen **(a)** is a correctly ordered specimen with all angles larger than 60°. Specimen **(b)** is not ordered correctly and contains multiple angles much smaller than 60° (red lines) as a result. Specimen **(c)** is correctly ordered but contains a sharp angle due to unusual growth of the penultimate chamber and is therefore flagged as an individual that needs closer inspection (note that the flagged angle appears larger than 60° due to two-dimensional representation).

```
with(Mlimbata, check.chamber.order(x=Cen
troidx, y=Centroidy, z=Centroidz,
proloculus=TRUE))
```

2.3 Number of chambers in the final whorl

The number of chambers in the final whorl is a frequently used diagnostic feature in taxonomy and can help distinguish between species and genera (e.g. Kennett and Srinivasan, 1983; Olsson et al., 1999; Pearson et al., 2006; Wade et al., 2018). Traditionally this number is determined from scanning electron microscope (SEM) or light microscope images and only describes the visible outer whorl. In contrast, CT scans also reveal all inner chambers. Using chamber centroid coordinates from CT data, we can determine the number of chambers in the final whorl at the time each chamber was built (Fig. 7). These new data could provide new insights into when ontogeny differences in whorl size between species and genera first arise. Manual determination of the number of chambers in the final whorl can be ambiguous, resulting in the inclusion of partial chambers in the final number (e.g. *Globoconella inflata* is described as having 3–3.5 chambers in the final whorl). These manual results are subjective and depend on visual volume determination of a partially hidden chamber. The function presented here uses distance between chambers to identify which chamber in the previous whorl is closest to the chamber under consideration. Therefore, the output is always in integers and no incomplete parts of chambers are considered.

2.3.1 Algorithm

To calculate the number of chambers in the final whorl at the time chamber i was built (cfw_i), the algorithm searches for the closest chamber in the whorl immediately below chamber i. This chamber marks the end of the previous whorl (see Fig. 8a). Depending on the chamber configuration, the closest chamber to chamber i can be either chamber $i-1$ or the chamber in the whorl below. The algorithm selects the chamber with the shortest distance to chamber i that is not chamber $i-1$. If the closest non-neighbouring chamber to chamber i is chamber 1, chamber i is in the first and only whorl at time i (see Fig. 8b). In this case, $\mathrm{cfw}_i = i$. If the final chamber of the previous whorl is not the proloculus, $\mathrm{cfw}_i = i - \#$ final chamber of previous whorl (Fig. 8c, d).

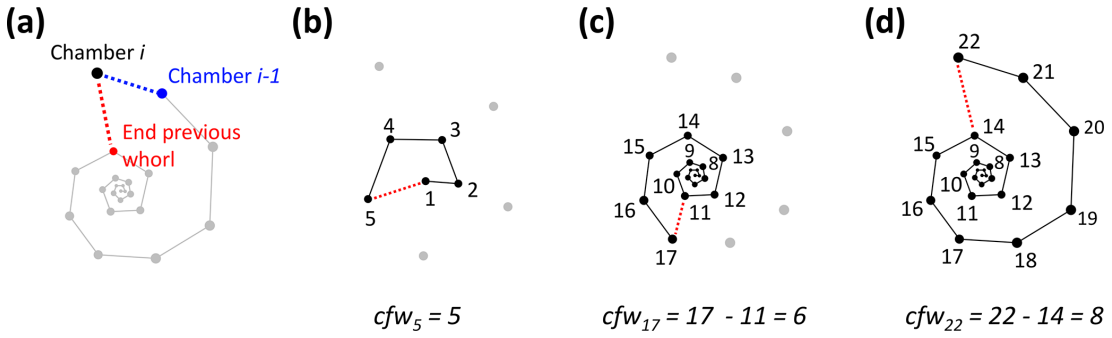


Figure 7. Determining the number of chambers in the final whorl for *Menardella limbata* at the time chamber i was built (cfw_i). **(a)** The two closest chambers to chamber i are chamber $i-1$ (blue) and the final chamber of the previous whorl (red). **(b)** If the closest non-neighbouring chamber is the proloculus, the chamber forms part of the first whorl and $cfw_i = i$. **(c, d)** If the closest chamber in the whorl below is not the proloculus, $cfw_i = i -$ final chamber of previous whorl.

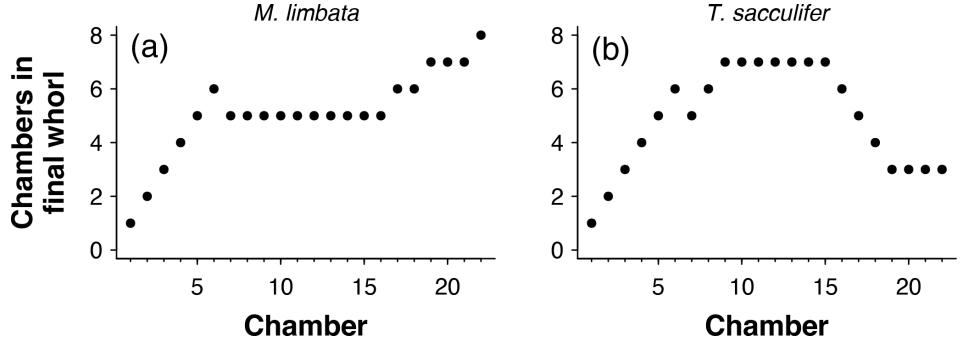


Figure 8. The number of chambers in the final whorl at the time each chamber was built using the exemplar **(a)** *Menardella limbata* and **(b)** *Trilobatus sacculifer* data.

2.3.2 Usage

The function requires vectors with chamber x-, y- and z-centroid coordinates. It returns a two-column data frame with the closest non-neighbouring chamber to each chamber and the number of chambers in the final whorl at the time each chamber was built (see the code in the Supplement, lines 115–122). Results shown are for the exemplar *M. limbata* specimen. See Sect. A4 for function output. See the code in the Supplement, line 122 for the output for *T. sacculifer*.

```
with(Mlimbata, chambers.in.whorl(x=Cen
troidx, y=Centroidy, z=Centroidz))
```

2.4 Number of whorls

Taxonomic descriptions often include the number of whorls typically found in adult specimens (e.g. Kennett and Srinivasan, 1983; Olsson et al., 1999; Pearson et al., 2006; Wade et al., 2018). Three to four whorls are common in planktonic foraminifera, whereas benthic foraminifera, especially larger species, can have many more (Holbourn et al., 2013). Differences in the numbers of whorls among specimens can help species identification. A change in the number of whorls through time can point to an altered developmental history, such as paedomorphosis (reduced development) or peramorphosis (extended development).

To determine the total number of whorls, all chambers need to be visible. Determining the number of whorls manually can introduce bias similar to counting the number of chambers in the final whorl by eye. By letting an algorithm find the closest chamber in the whorl below individual chambers from CT scan data, the number of whorls can be determined objectively.

2.4.1 Algorithm description

The algorithm starts with the proloculus and makes its way up the logarithmic spire from there. The first whorl consists of all chambers whose closest non-neighbouring chamber is the proloculus (Fig. 9a). The second whorl starts with the first chamber whose closest chamber in the whorl below is not the proloculus (chamber 7 in Fig. 9a). Subsequently, whorl $i+1$ starts with the first chamber whose closest chamber in the whorl below is in whorl i (Fig. 9b). The last whorl is often not complete (e.g. Pearson et al., 2006) (Fig. 9c). The algorithm checks if the last whorl is complete by comparing the number of chambers in the last whorl as determined algorith-

mically (counted forward from the proloculus onwards) with the number of chambers in the final whorl at the time the final chamber was built (counted backwards, starting with the final chamber). If these numbers are not the same, all chambers in the last whorl are flagged as being part of an incomplete whorl. For example, the specimen in Fig. 9c had eight chambers in its final whorl at the time the final chamber was built, but only five of these were part of the actual last whorl.

2.4.2 Usage

The function requires vectors with chamber x-, y- and z-centroid coordinates. It returns a two-column data frame with the whorl number of each chamber and whether the whorl is complete (code in the Supplement, lines 124–130). All whorls but the final whorl are complete by definition, whereas the final whorl is either complete or incomplete. See Sect. A5 for function output.

```
with(Mlimbata, whorl(x=Cen
troidx, y=Centroidy, z=Centroidz))
```

2.5 Coiling direction

Coiling direction is an important trait for biostratigraphy, with changes in the dominant coiling direction marking biostratigraphic horizons (e.g. Wade et al., 2011). For some species the coiling direction is the primary means of identification, such as in *Neogloboquadrina pachyderma* and *N. incompta* (Darling et al., 2006). Although the coiling direction of adult individuals can easily be determined by eye, the coiling of earlier, hidden ontogenetic stages is more difficult to analyse. The R package includes a function that determines the coiling direction throughout ontogeny. This could prove useful to determine the exact moment of coiling change in the rare planktonic foraminifera species that build whorl-enveloping chambers such as *Orbulinoides beckmanni* or more likely in streptospirally coiled benthic foraminifera species.

2.5.1 Algorithm

To determine the coiling direction of chamber i we compare the vertical direction of growth to the cross product of the vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ between chambers $i-2$, $i-1$ and i. The cross product creates a vector perpendicular to both $\boldsymbol{v}$ and $\boldsymbol{w}$, with the direction of $\boldsymbol{v} \times \boldsymbol{w}$ determined by the direction of travel from $\boldsymbol{v}$ to $\boldsymbol{w}$ (Fig. 10). Clockwise travel as viewed from the umbilical side (i.e. sinistral coiling) produces a cross-product vector pointing away from the vertical direction of growth, whereas anticlockwise travel as viewed from the umbilical side (i.e. dextral coiling) produces a vector pointing in the same direction as the direction of vertical growth (give or take minor deviations). Therefore, if the angle between $\boldsymbol{v} \times \boldsymbol{w}$ and the direction of growth is smaller than 90°, the coiling direction is anticlockwise or dextral, whereas an angle larger than 90° indicates clockwise or sinistral coiling.

2.5.2 Usage

The function requires vectors with chamber x-, y- and z-centroid coordinates. It returns a vector with the coiling direction of each chamber relative to the two previous chambers (see the code in the Supplement, lines 132–138). Note that at least three chambers are required for coiling direction determination; therefore, no coiling direction can be established for chambers 1 and 2. Also note that the earliest chambers in the juvenile stage grow nearly planispirally, the smallest deviations from which can result in an apparent change of coiling direction. See Sect. A6 for function output.

```
with(Mlimbata, coiling.direction(x=Cen
troidx, y=Centroidy, z=Centroidz))
```

2.6 Trochospirality

The height of the trochospire is another trait used to distinguish among species (Fabbrini et al., 2021; Biolzi, 1991) and genera (Gradstein and Waskowska, 2021; Lipps, 1966; Pearson and Coxall, 2014). Specimens are qualitatively described as either planispiral (Fig. 11a) or having a low or high trochospire (Fig. 11b, c), but there is currently no direct way to quantify the height or angle of the trochospire. Quantifying the trochospire can help determine the exact moment of evolutionary divergence of two closely related species through statistical analyses (e.g. Pearson and Ezard, 2014). The height of the trochospire has also been noted to change through ontogeny: growth is nearly planispiral in the juvenile stage but increases as individuals mature (Caromel et al., 2016; Apthorpe, 2020; Kendall et al., 2020; Poole and Wade, 2019; Morard et al., 2019). The timing of change could be used to determine the transition from the juvenile to the neanic stage, which is currently only described qualitatively through manual inspection (Brummer et al., 1986, 1987).

2.6.1 Algorithm

Shells growing in approximate logarithmic spirals, such as ammonoids and gastropods, are traditionally described using "Raupian" parameters (Raup, 1966, 1967) that quantify growth relative to the coiling axis. The translation rate or growth in the direction of the coiling axis for every ontogenetic step can theoretically be used to calculate trochospirality in foraminifera (Morard et al., 2019; Caromel et al., 2016). This method depends on correct identification of the coiling axis. Ammonoids and gastropods grow continuously and roughly isometrically, making the coiling axis easy to identify. Foraminifera, however, grow in discrete steps, and growth patterns change through ontogeny (Signes et al., 1993). Therefore, the coiling axis of a foraminifera specimen

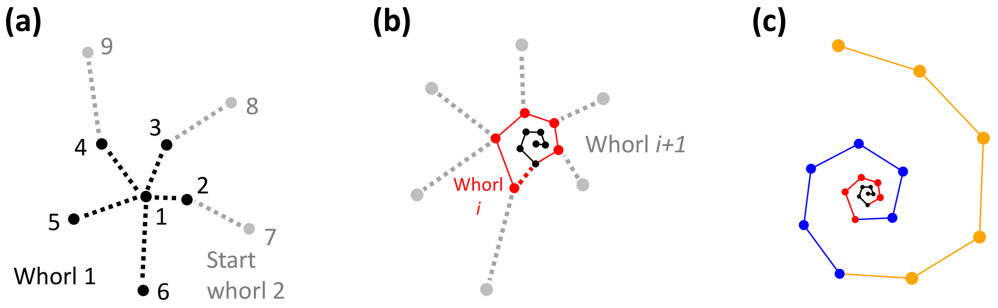


Figure 9. Determining the total number of whorls for the example *Menardella limbata* specimen. **(a)** The first whorl consists of all chambers whose closest chamber is the proloculus (black). Whorl 2 (grey) starts with the first chamber whose closest neighbour in the whorl below is not the proloculus. **(b)** Whorl $i+1$ (grey) starts with the first chamber that has a chamber from whorl i (red) as its closest non-neighbouring chamber. **(c)** All three complete whorls (black, red, blue) and the fourth incomplete whorl (orange).

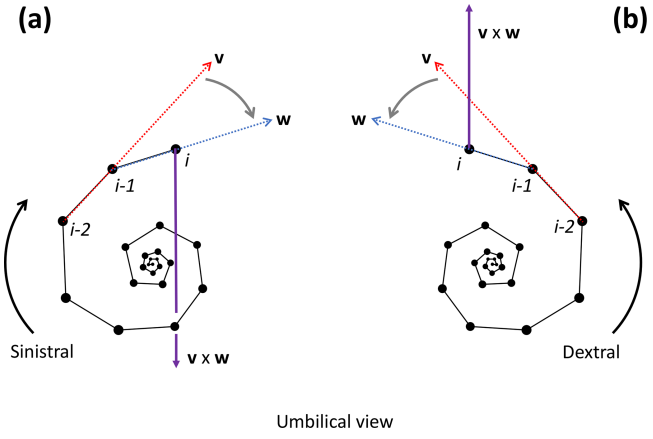


Figure 10. Determination of the coiling direction of chamber i for the example *Menardella limbata* specimen. **(a)** If the direction of change from vector $\boldsymbol{v}$ through chambers $i-2$ and $i-1$ (red) and vector $\boldsymbol{w}$ through chambers $i-1$ and i (blue) is clockwise as viewed from the umbilical side (i.e. sinistral coiling), the cross-product $\boldsymbol{v} \times \boldsymbol{w}$ (purple) points towards the proloculus in the opposite direction to the direction of growth. **(b)** If the direction of change from $\boldsymbol{v}$ to $\boldsymbol{w}$ is anticlockwise as viewed from the umbilical side (i.e. dextral coiling) $\boldsymbol{v} \times \boldsymbol{w}$ aligns with the direction of growth. Note that the orientation of the specimen does not influence the coiling direction outcome of the algorithm: if the same specimens were viewed from the spiral side, both the cross product and the direction of growth would flip by 180°, maintaining their similar or opposite relationship necessary to determine the coiling direction.

changes through ontogeny and is impossible to determine mathematically. Approximations of the coiling axis would need to be done by hand, potentially introducing bias and influencing any resulting calculations of trochospirality.

Alternatively, we can define the trochospirality of chamber i as the angle between chamber i and the plane defined by the previous three chambers (Fig. 11). This method does not rely on a coiling axis and only requires chamber centroid coordinates. Specimens following a perfect logarithmic spiral will have a constant trochospirality value throughout ontogeny. However, natural growth is rarely perfect. The trochospirality of a specimen can be defined as the average trochospirality τ_i of all chambers, and deviations in τ_i can be used to assess developmental plasticity and ontogenetic constraints (Fig. 12).

2.6.2 Usage

The function requires vectors with chamber x-, y- and z-centroid coordinates. It returns a vector with the trochospirality angle of each chamber, relative to the plane formed by the previous three chambers (see the code in the Supplement, lines 106–113). Note that at least three previous chambers are required to form a plane. Therefore, no trochospirality angle exists for the first three chambers. See Sect. A7 for function output.

```
with(Mlimbata, trochospirality(x=Cen
troidx, y=Centroidy, z=Centroidz))
```

3 Package applications and future work

For ease of use, all trait functions have been combined into a single function: foram.growth.3D(). This function requires vectors with chamber number, centroid coordinates and volumes and returns a data frame with the original data, plus angles between subsequent chambers, trochospirality, chambers in the final whorl at the time each chamber was built, the whorl number each chamber belongs to, whether the whorl is complete, coiling direction, and chamber order (see the code in the Supplement, lines 140–150). See Sect. A8 for function output.

```
with(Mlimbata, foram.growth.3D(n=Cham
ber, x=Centroidx, y=Centroidy,
z=Centroidz, proloculus=TRUE))
```

Plotting combinations of traits against each other can reveal ontogenetic trajectories and trait covariation patterns through ontogeny. The example data show stable chamber

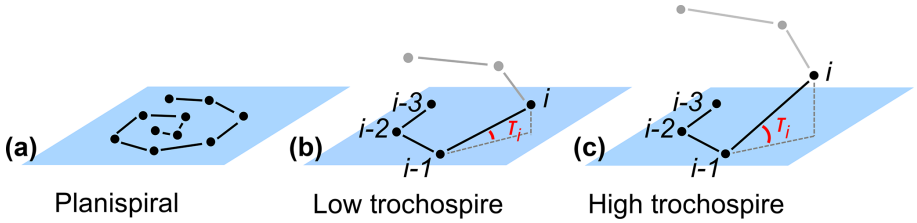


Figure 11. Theoretical examples of **(a)** planispiral growth and **(b, c)** low and high trochospires, respectively. Planes represent the growth plane of the entire specimen in the case of planispiral growth **(a)** or the plane defined by chambers $i-3$, $i-2$ and $i-1$ for non-planispiral growth **(b, c)**; τ_i is the trochospire angle of chamber i relative to the plane defined by the previous three chambers.

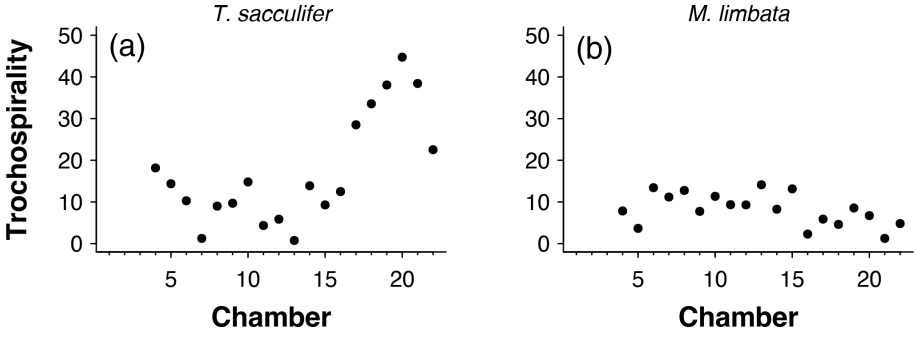


Figure 12. Trochospirality per chamber for the exemplar data of **(a)** *Trilobatus sacculifer* and **(b)** *Menardella limbata*.

angles until around chamber 14, when angles increase with chamber number (Fig. 13, far left column). Trochospirality and growth rate decrease around the same time, with the number of chambers in the final whorl increasing several chambers later. The remaining columns in Fig. 13 show the correlations between all remaining trait combinations. Strong covariations between some traits are expected, such as whorl number and chamber volume (Fig. 13, bottom row, second from right), because chambers in earlier whorls are typically smaller. Others, such as the covariation between the number of chambers in the final whorl and chamber volume (Fig. 13, bottom row, third from right), are not immediately obvious and could point to a size threshold for whorl expansion. Similarly, chamber angles increase at the same time the rate of chamber growth starts to decrease, several chambers before the number of chambers in the final whorl increases, suggesting a switch away from isometric growth that leads to a reduced chamber-by-chamber growth rate and thus increasing numbers of chambers in the final whorl.

4 Prospects for future work

We present a new R package to automatically reconstruct foraminifera growth trajectories from chamber x, y and z centroid coordinates. The package functions arrange chambers in order of growth, calculate distances and angles between chambers, determine the total number of whorls and the number of chambers in the final whorl at the time each chamber was built, and, for the first time, quantify trochospirality. Applied to large numbers of specimens through multiple lineages and evolutionary transitions, the foram3D package functions can shed new light on ontogenetic trajectories and developmental constraints through time. They can be used to quantify ontogenetic stages, for example through changes in trochospirality in the juvenile and neanic stages or the increase in chambers in the final whorl for the neanic to adult transition. Additionally, they can be used to quantify species differences, for example through average trochospirality, angles between chambers in the final whorl and the total number of whorls. This will be particularly valuable for determining the exact moment of the origination of a new species, which is often difficult to identify by eye. Our package enables multivariate analyses of the incredible richness from x-ray CT data and could substantially improve our understanding of developmental constraints on planktonic foraminifera growth and form.

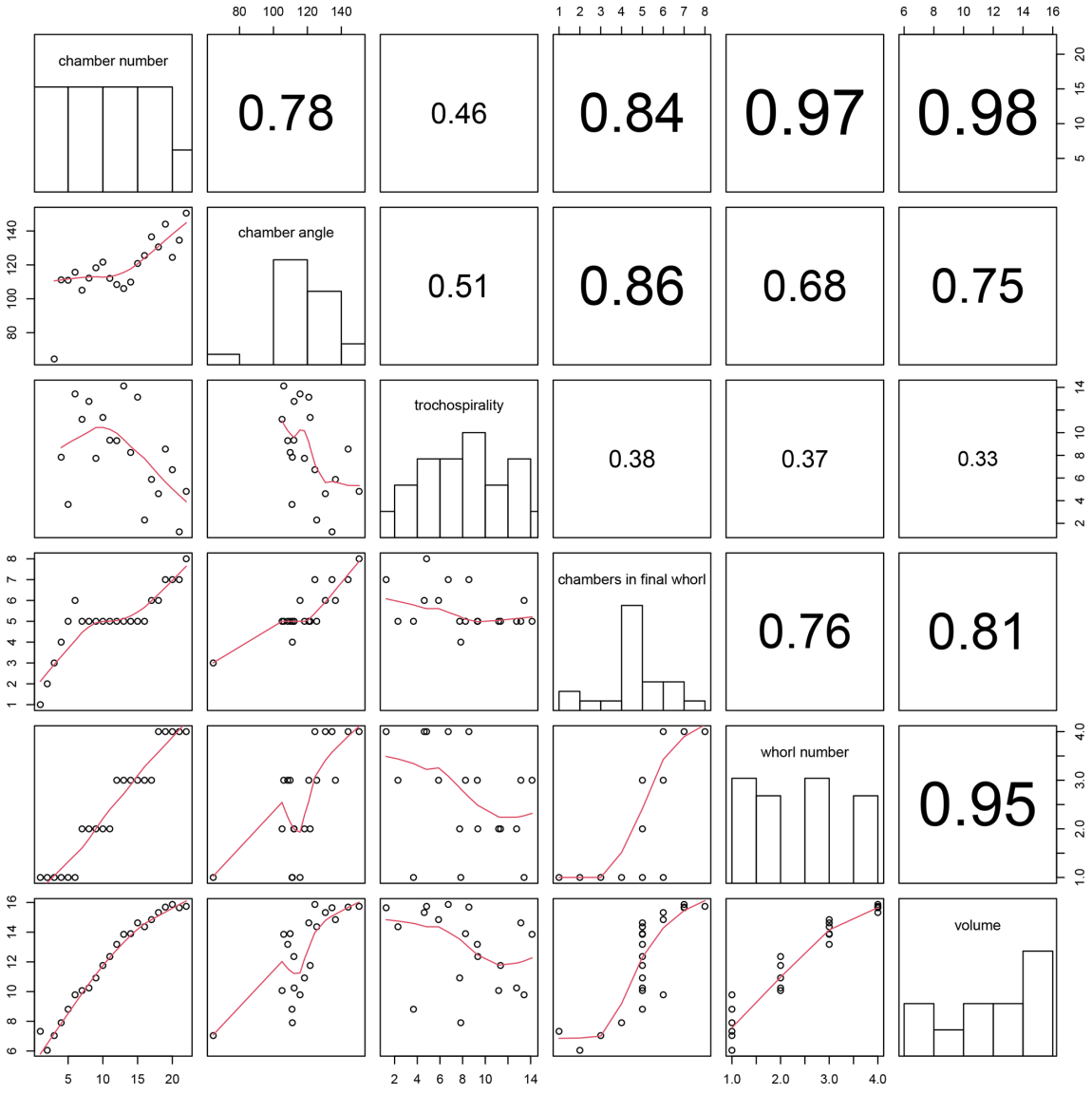


Figure 13. Correlation between all combinations of traits calculated for the example *Menardella limbata* specimen by the foram3D package. Plots on the diagonal show histograms of the trait data. Plots below the diagonal show all combinations of traits plotted against each other, with red lines representing splines. R^2 values of all trait correlations are shown in the panels above the diagonal. The column on the far left shows all trait values through ontogeny. The next column shows all remaining traits plotted against chamber angle.

Appendix A

"NA" stands for "not available" and indicates an empty entry.

A1 Output of the function that orders chambers in the direction of growth

```
with(Mlimbata, order.chambers(x=Centroidx, y=Centroidy, z=Centroidz, V=Volume,

proloculus=TRUE))
```

	x centroid	y centroid	z centroid	Volume	Chamber number
1	790.2737	947.1173	453.7312	1519.0677	1
2	800.3547	954.0144	450.9993	423.1689	2
3	801.7997	944.2201	464.2364	1139.3008	3
4	785.5543	936.5239	468.6528	2707.1956	4
5	770.2097	941.8412	454.1172	6727.2998	5
6	779.0149	952.0797	431.4037	17740.5406	6
7	810.6712	947.0431	431.7093	23588.9512	7
8	819.5385	925.7222	454.7856	28146.1542	8
9	792.6812	907.2282	473.0768	55277.7883	9
10	748.8339	906.8281	458.0008	127824.1210	10
11	743.5526	931.0394	400.9859	233106.3626	11
12	815.3179	941.4148	371.8330	530149.1992	12
13	876.3621	888.9579	429.9081	1040610.1960	13
14	818.6441	827.8897	498.2896	1085341.3140	14
15	710.8669	810.0553	481.1860	2250536.7620	15
16	651.7275	847.0710	393.0151	1730125.8720	16
17	675.3247	899.9679	290.0133	2798247.4690	17
18	798.6004	930.2500	228.4772	4510008.8980	18
19	949.4519	901.2303	231.3315	6496217.0310	19
20	1036.1169	801.7007	365.4462	7777235.9650	20
21	982.8341	716.6077	509.5566	6215726.6070	21
22	863.8481	657.6218	603.0626	6834399.4760	22

A2 Output of the function calculating angles between subsequent chambers

```
with(Mlimbata, chamber.angles(x=Centroidx, y=Centroidy, z=Centroidz))
```

```
 [1]          NA          NA   64.47616  111.14269  110.93628  115.67051  105.06948
 [8]  112.18464  118.27829  121.63305  112.03424  108.40934  106.04217  109.85145
[15]  120.83569  125.51267  136.54085  130.60246  144.03099  124.42830  134.53817
[22]  150.52445
```

A3 Output of the function checking the chamber order

```
with(Mlimbata, check.chamber.order(x=Centroidx, y=Centroidy, z=Centroidz,

proloculus=TRUE))
```

```
 [1] "Correct" "Correct" "Correct" "Correct" "Correct" "Correct" "Correct"
 [8] "Correct" "Correct" "Correct" "Correct" "Correct" "Correct" "Correct"
[15] "Correct" "Correct" "Correct" "Correct" "Correct" "Correct" "Correct"
[22] "Correct"
```

A4 Output of the function determining the number of chambers in the final whorl at the time chamber *i* was built

```
with(Mlimbata, chambers.in.whorl(x=Centroidx, y=Centroidy, z=Centroidz))
```

```
##        Closest chamber  Chambers in whorl
## [1,]                 1                 1
## [2,]                 1                 2
## [3,]                 1                 3
## [4,]                 1                 4
## [5,]                 1                 5
## [6,]                 1                 6
## [7,]                 2                 5
## [8,]                 3                 5
## [9,]                 4                 5
## [10,]                5                 5
## [11,]                6                 5
## [12,]                7                 5
## [13,]                8                 5
## [14,]                9                 5
## [15,]               10                 5
## [16,]               11                 5
## [17,]               11                 6
## [18,]               12                 6
## [19,]               12                 7
## [20,]               13                 7
## [21,]               14                 7
## [22,]               14                 8
```

A5 Output of the function determining which whorl each chamber belongs to

```
with(Mlimbata, whorl(x=Centroidx, y=Centroidy, z=Centroidz))
```

```
##     Whorl number  Whorl complete?
## 1              1         complete
## 2              1         complete
## 3              1         complete
## 4              1         complete
## 5              1         complete
## 6              1         complete
## 7              2         complete
## 8              2         complete
## 9              2         complete
## 10             2         complete
## 11             2         complete
## 12             3         complete
## 13             3         complete
## 14             3         complete
## 15             3         complete
## 16             3         complete
## 17             3         complete
## 18             4       incomplete
## 19             4       incomplete
## 20             4       incomplete
## 21             4       incomplete
## 22             4       incomplete
```

A6 Output of the function determining coiling direction at the time chamber *i* was built

```
with(Mlimbata, coiling.direction(x=Centroidx, y=Centroidy, z=Centroidz))
```

```
## [1]   NA         NA         "sinistral"  "sinistral"  "sinistral"  "dextral"
## [7]   "dextral"  "dextral"  "dextral"    "dextral"    "dextral"    "dextral"
## [13]  "dextral"  "dextral"  "dextral"    "dextral"    "dextral"    "dextral"
## [19]  "dextral"  "dextral"  "dextral"    "dextral"
```

A7 Output of the function calculating trochospirality

```
with(Mlimbata, trochospirality(x=Centroidx, y=Centroidy, z=Centroidz))
```

```
## [1]          NA        NA         NA  7.835840  3.664881  13.421233  11.177728
## [8]   12.761478  7.737775  11.345215  9.337931  9.295332  14.123845   8.254441
## [15]  13.135575  2.295534   5.876539  4.614813  8.554413   6.731464   1.251507
## [22]   4.820587
```

A8 Output of the function combining the outputs of all functions described above

```
with(Mlimbata, foram.growth.3D(n=Chamber, x=Centroidx, y=Centroidy, z=Centroidz,

proloculus=TRUE))
```

```
##    Chamber number  Centroid x  Centroid y  Centroid z  Chamber angle
## 1               1    790.2737    947.1173    453.7312             NA
## 2               2    800.3547    954.0144    450.9993             NA
## 3               3    801.7997    944.2201    464.2364       64.47616
## 4               4    785.5543    936.5239    468.6528      111.14269
## 5               5    770.2097    941.8412    454.1172      110.93628
## 6               6    779.0149    952.0797    431.4037      115.67051
##    Trochospirality  Chambers in final whorl  Whorl number  Whorl complete?
## 1               NA                        1             1         complete
## 2               NA                        2             1         complete
## 3               NA                        3             1         complete
## 4         7.835840                        4             1         complete
## 5         3.664881                        5             1         complete
## 6        13.421233                        6             1         complete
##    coiling direction  chamber order
## 1               ⟨NA⟩        Correct
## 2               ⟨NA⟩        Correct
## 3          sinistral        Correct
## 4          sinistral        Correct
## 5          sinistral        Correct
## 6            dextral        Correct
```

Code and data availability. The R package foram3D is freely available for download directly into R from Anieke Brombacher's GitHub page (https://github.com/AniekeBrombacher/foram3D, last access: 31 August 2022; https://doi.org/10.5281/zenodo.7252765, Brombacher et al., 2022b), which contains all functions and data used in this paper. An example user guide to R scripts, containing information on how to download the package into R, generate interactive figures and use code is available in the Supplement.

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/jm-41-149-2022-supplement.

Author contributions. AB designed the functions and wrote the code. AB, ASB, WZ and THGE tested and helped improve the functions and code. AB drafted the manuscript, and ASB, WZ and THGE provided comments on earlier drafts.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We would like to thank Marisa Sweeney for testing the package functions and nine anonymous attendees of the Micropalaeontological Society Foraminifera Festival 2021 that provided constructive comments on earlier versions of the R package through an online poll. We would also like to thank two anonymous reviewers for their constructive comments on an earlier version of this paper.

Financial support. This research has been supported by the Natural Environment Research Council (grant nos. NE/J018163/1 and NE/P019269/1).

Review statement. This paper was edited by Sev Kender and reviewed by two anonymous referees.

References

Apthorpe, M.: Middle Jurassic (Bajocian) planktonic foraminifera from the northwest Australian margin, J. Micropalaeontol., 39, 93–115, https://doi.org/10.5194/jm-39-93-2020, 2020.

Beldade, P., Mateus, A. R. A., and Keller, R. A.: Evolution and molecular mechanisms of adaptive developmental plasticity, Mol. Ecol., 20, 1347–1363, https://doi.org/10.1111/j.1365-294X.2011.05016.x, 2011.

Biolzi, M.: Morphometric analyses of the Late Neogene planktonic foraminiferal lineage *Neogloboquadrina dutertrei*, Mar. Micropaleontol., 18, 129–142, https://doi.org/10.1016/0377-8398(91)90009-U, 1991.

Brigulgio, A., Metscher, B., and Hohenegger, J.: Growth rate biometric qunatification by X-ray microtomography on larger benthic foraminifera: three-dimensional measurements push Nummulitids into the fourth dimension, Turk. J. Earth Sci., 20, 683–699, https://doi.org/10.3906/yer-0910-44, 2011.

Brombacher, A., Elder, L. E., Hull, P. M., Wilson, P. A., and Ezard, T. H. G.: Calibration of test diameter and area as proxies for body size in the planktonic foraminifer *Globoconella puncticulata*, J. Foramin. Res., 48, 241–245, https://doi.org/10.2113/gsjfr.48.3.241, 2018.

Brombacher, A., Schmidt, D. N., and Ezard, T. H. G.: Developmental plasticity in deep time: a window to population ecological inference, Paleobiology, https://doi.org/10.1017/pab.2022.26, online first, 2022a.

Brombacher, A., Searle-Barnes, A., Zhang, W., Ezard, T.H.G.: AniekeBrombacher/foram3D: Release for publication (v1.1.0), Zenodo [code, data set], https://doi.org/10.5281/zenodo.7252765, 2022b.

Brummer, G.-J. A., Hemleben, C., and Spindler, M.: Planktonic foraminiferal ontogeny and new perspectives for micropalaeontology, Nature, 319, 50–52, https://doi.org/10.1038/319050a0, 1986.

Brummer, G.-J. A., Hemleben, C., and Spindler, M.: Ontogeny of extant spinose planktonic foraminifera (Globigerinidae): A concept exemplified by *Globigerinoides sacculifer* (Brady) and *G. ruber* (d'Orbigny), Mar. Micropaleontol., 12, 357–381, https://doi.org/10.1016/0377-8398(87)90028-4, 1987.

Burke, J. E., Renema, W., Schiebel, R., and Hull, P. M.: Three-dimensional analysis of inter- and intraspecific variation in ontogenetic growth trajectories of planktonic foraminifera, Mar. Micropaleontol., 155, 1–12, https://doi.org/10.1016/j.marmicro.2019.101794, 2020.

Caromel, A. G., Schmidt, D. N., Fletcher, I., and Rayfield, E. J.: Morphological change during the ontogeny of the planktic foraminifera, J. Micropalaeontol., 35, 2–19, https://doi.org/10.1144/jmpaleo2014-017, 2016.

Caromel, A. G. M., Schmidt, D. N., and Rayfield, E. J.: Ontogenetic constraints on foraminiferal test construction, Evol. Dev., 19, 157–168, https://doi.org/10.1111/ede.12224, 2017.

Darling, K. F., Kucera, M., Kroon, D., and Wade, C. M.: A resolution for the coiling direction paradox in *Neogloboquadrina pachyderma*, Paleoceanography, 21, 1–14, https://doi.org/10.1029/2005pa001189, 2006.

Davis, C. V., Livsey, C. M., Palmer, H. M., Hull, P. M., Thomas, E., Hill, T. M., and Benitez-Nelson, C. R.: Extensive morphological variability in asexually produced planktic foraminifera, Sci. Adv., 6, 1–7, https://doi.org/10.1126/sciadv.abb8930, 2020.

DeWitt, T. J., Sih, A., and Sloan Wilson, D.: Costs and limits of phenotypic plasticity, Trends Ecol. Evol., 13, 77–81, https://doi.org/10.1016/S0169-5347(97)01274-3, 1998.

Duan, B., Li, T., and Pearson, P. N.: Three dimensional analysis of ontogenetic variation in fossil globorotaliiform planktic foraminiferal tests and its implications for ecology, life processes and functional morphology, Mar. Micropaleontol., 165, 1–9, https://doi.org/10.1016/j.marmicro.2021.101989, 2021.

Fabbrini, A., Zaminga, I., Ezard, T. H. G., and Wade, B. S.: Systematic taxonomy of middle Miocene *Sphaeroidinellopsis*

(planktonic foraminifera), J. Syst. Palaeontol., 19, 953–968, https://doi.org/10.1080/14772019.2021.1991500, 2021.

Fox, L., Stukins, S., Hill, T., and Miller, C. G.: Quantifying the effect of anthropogenic climate change on calcifying plankton, Sci. Rep. UK, 10, 1–9, https://doi.org/10.1038/s41598-020-58501-w, 2020.

Gradstein, F. and Waskowska, A.: New insights into the taxonomy and evolution of Jurassic planktonic foraminifera, Swiss Journal of Palaeontology, 140, 1–12, https://doi.org/10.1186/s13358-020-00214-8, 2021.

Holbourn, A., Henderson, A. S., and Macleod, N.: Atlas of benthic foraminifera, Wiley-Blackwell, London, UK, 656 pp., ISBN 9781118452493, 2013.

Huang, C.-Y.: Observations on the interior of some late Neogene planktonic foraminifera, J. Foramin. Res., 11, 173–190, https://doi.org/10.2113/gsjfr.11.3.173, 1981.

Iwasaki, S., Kimoto, K., Sasaki, O., Kano, H., Honda, M. C., and Okazaki, Y.: Observation of the dissolution process of *Globigerina bulloides* tests (planktic foraminifera) by X-ray microcomputed tomography, Paleoceanography, 30, 317–331, https://doi.org/10.1002/2014pa002639, 2015.

Iwasaki, S., Kimoto, K., Okazaki, Y., and Ikehara, M.: Micro-CT scanning of tests of three planktic foraminiferal species to clarify dissolution process and progress, Geochem. Geophy. Geosy., 20, 6051–6065, https://doi.org/10.1029/2019gc008456, 2019a.

Iwasaki, S., Kimoto, K., Sasaki, O., Kano, H., and Uchida, H.: Sensitivity of planktic foraminiferal test bulk density to ocean acidification, Sci. Rep. UK, 9, 1–9, https://doi.org/10.1038/s41598-019-46041-x, 2019b.

Johnstone, H. J. H., Schulz, M., Barker, S., and Elderfield, H.: Inside story: An X-ray computed tomography method for assessing dissolution in the tests of planktonic foraminifera, Mar. Micropaleontol., 77, 58–70, https://doi.org/10.1016/j.marmicro.2010.07.004, 2010.

Johnstone, H. J. H., Yu, J., Elderfield, H., and Schulz, M.: Improving temperature estimates derived from Mg/Ca of planktonic foraminifera using X-ray computed tomography–based dissolution index, XDX, Paleoceanography, 26, 1–17, https://doi.org/10.1029/2009pa001902, 2011.

Kendall, S., Gradstein, F., Jones, C., Lord, O. T., and Schmidt, D. N.: Ontogenetic disparity in early planktic foraminifers, J. Micropalaeontol., 39, 27–39, https://doi.org/10.5194/jm-39-27-2020, 2020.

Kennett, J. P. and Srinivasan, M. S.: Neogene planktonic foraminifera. A phylogenetic atlas, Hutchinson Ross Publishing Company, Stroudsburg, Pennsylvania, ISBN 9780879330705, 1983.

Lipps, J. H.: Wall structure, systematics, and phylogeny studies of Cenozoic planktonic foraminifera, J. Paleontol., 40, 1257–1274, 1966.

Moczek, A. P., Sultan, S., Foster, S., Ledon-Rettig, C., Dworkin, I., Nijhout, H. F., Abouheif, E., and Pfennig, D. W.: The role of developmental plasticity in evolutionary innovation, P. Roy. Soc. B Biol. Sci., 278, 2705–2713, https://doi.org/10.1098/rspb.2011.0971, 2011.

Morard, R., Füllberg, A., Brummer, G. A., Greco, M., Jonkers, L., Wizemann, A., Weiner, A. K. M., Darling, K., Siccha, M., Ledevin, R., Kitazato, H., de Garidel-Thoron, T., de Vargas, C., and Kucera, M.: Genetic and morphological divergence in the warm-water planktonic foraminifera genus *Globigerinoides*, PLoS One, 14, e0225246, https://doi.org/10.1371/journal.pone.0225246, 2019.

Murren, C. J., Auld, J. R., Callahan, H., Ghalambor, C. K., Handelsman, C. A., Heskel, M. A., Kingsolver, J. G., Maclean, H. J., Masel, J., Maughan, H., Pfennig, D. W., Relyea, R. A., Seiter, S., Snell-Rood, E., Steiner, U. K., and Schlichting, C. D.: Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity, Heredity, 115, 293–301, https://doi.org/10.1038/hdy.2015.8, 2015.

Olsson, R. K., Hemleben, C., Berggren, W. A., and Huber, B. T.: Atlas of Paleocene Planktonic Foraminifera, Smithsonian Institution Press, Washington, DC, https://doi.org/10.5479/si.00810266.85.1, 1999.

Pearson, P. N. and Coxall, H. K.: Origin of the Eocene planktonic foraminifer *Hantkenina* by gradual evolution, Palaeontology, 57, 243–267, https://doi.org/10.1111/pala.12064, 2014.

Pearson, P. N. and Ezard, T. H. G.: Evolution and speciation in the Eocene planktonic foraminifer *Turborotalia*, Paleobiology, 40, 130–143, https://doi.org/10.1666/13004, 2014.

Pearson, P. N., Olsson, R. K., Huber, B. T., Hemleben, C., and Berggren, W. A. (Eds.): Atlas of Eocene planktonic foraminifera, Vol. 41, Cushman Foundation for Foraminiferal Research Special Publication, 9–211, 213–375, 377–411, 413–459, 461–507, ISBN 9781970168365, 2006.

Pfennig, D. W., Wund, M. A., Snell-Rood, E. C., Cruickshank, T., Schlichting, C. D., and Moczek, A. P.: Phenotypic plasticity's impacts on diversification and speciation, Trends Ecol. Evol., 25, 459–467, https://doi.org/10.1016/j.tree.2010.05.006, 2010.

Pigliucci, M., Murren, C. J., and Schlichting, C. D.: Phenotypic plasticity and evolution by genetic assimilation, J. Exp. Biol., 209, 2362–2367, https://doi.org/10.1242/jeb.02070, 2006.

Poole, C. R. and Wade, B. S.: Systematic taxonomy of the *Trilobatus sacculifer* plexus and descendant *Globigerinoidesella fistulosa* (planktonic foraminifera), J. Syst. Palaeontol., 17, 1989–2030, https://doi.org/10.1080/14772019.2019.1578831, 2019.

Price, T. D., Qvarnstrom, A., and Irwin, D. E.: The role of phenotypic plasticity in driving genetic evolution, P. R. Soc. B, 270, 1433–1440, https://doi.org/10.1098/rspb.2003.2372, 2003.

Raup, D. M.: Geometric analysis of shell coiling: general problems, J. Paleontol., 40, 1178–1190, 1966.

Raup, D. M.: Geometric analysis of shell coiling: coiling in Ammonoids, J. Paleontol., 41, 43–65, 1967.

Schmidt, D. N., Rayfield, E. J., Cocking, A., and Marone, F.: Linking evolution and development: Synchrotron Radiation X-ray tomographic microscopy of planktic foraminifers, Palaeontology, 56, 741–749, https://doi.org/10.1111/pala.12013, 2013.

Signes, M., Bijma, J., Hemleben, C., and Ott, R.: A model for planktic foraminiferal shell growth, Paleobiology, 19, 71–91, https://doi.org/10.1017/S009483730001232X, 1993.

Speijer, R. P., Van Loo, D., Masschaele, B., Vlassenbroeck, J., Cnudde, V., and Jacobs, P.: Quantifying foraminiferal growth with high-resolution X-ray computed tomography: New opportunities in foraminiferal ontogeny, phylogeny, and paleoceanographic applications, Geosphere, 4, 760–763, https://doi.org/10.1130/ges00176.1, 2008.

Sverdlove, M. S. and Be, A. W.: Taxonomic and ecological significance of embryonic and juvenile plank-

tonic foraminifera, J. Foramin. Res., 15, 235–241, https://doi.org/10.2113/gsjfr.15.4.235, 1985.

Takagi, H., Kurasawa, A., and Kimoto, K.: Observation of asexual reproduction with symbiont transmission in planktonic foraminifera, J. Plankton Res., 42, 403–410, https://doi.org/10.1093/plankt/fbaa033, 2020.

Todd, C. L., Schmidt, D. N., Robinson, M. M., and De Schepper, S.: Planktic foraminiferal test size and weight response to the Late Pliocene environment, Paleoceanogr. Paleocl., 35, 1–15, https://doi.org/10.1029/2019pa003738, 2020.

Vanadzina, K. and Schmidt, D. N.: Developmental change during a speciation event: evidence from planktic foraminifera, Paleobiology, 48, 120–136, https://doi.org/10.1017/pab.2021.26, 2022.

Wade, B. S., Pearson, P. N., Berggren, W. A., and Pälike, H.: Review and revision of Cenozoic tropical planktonic foraminiferal biostratigraphy and calibration to the geomagnetic polarity and astronomical time scale, Earth-Sci. Rev., 104, 111–142, https://doi.org/10.1016/j.earscirev.2010.09.003, 2011.

Wade, B. S., Olsson, R. K., Pearson, P. N., Huber, B. T., and Berggren, W. A.: Atlas of Oligocene planktonic foraminifera, Cushman Foundation for Foraminiferal Research, London, UK, ISBN 9781970168419, 2018.

West-Eberhard, M. J.: Developmental plasticity and evolution, Oxford University Press, New York, ISBN 9780195122343, 2003.

West-Eberhard, M. J.: Developmental plasticity and the origin of species differences, P. Natl. Acad. Sci. USA, 102, 6543–6549, https://doi.org/10.1073/pnas.0501844102, 2005.

Zarkogiannis, S. D., Antonarakou, A., Tripati, A., Kontakiotis, G., Mortyn, P. G., Drinia, H., and Greaves, M.: Influence of surface ocean density on planktonic foraminifera calcification, Sci. Rep.-UK, 9, 533, https://doi.org/10.1038/s41598-018-36935-7, 2019.

Zarkogiannis, S. D., Fernandez, V., Greaves, M., Mortyn, P. G., Kontakiotis, G., and Antonarakou, A.: X-ray tomographic data of planktonic foraminifera species *Globigerina bulloides* from the Eastern Tropical Atlantic across Termination II, GIGAbyte, https://doi.org/10.46471/gigabyte.5, 2020.

PNAS RESEARCH ARTICLE | EVOLUTION

Detecting environmentally dependent developmental plasticity in fossilized individuals

Anieke Brombacher[a,b,c,1], Alex Searle-Barnes[a], James M. Mulqueeney[a,d], Christopher D. Standish[a], J. Andy Milton[a], Orestis L. Katsamenis[e,f], Richard A. Watson[g], Clive Trueman[a], Ian Sinclair[e,h], Paul A. Wilson[a], Gavin L. Foster[a], and Thomas H. G. Ezard[a]

Affiliations are included on p. 7.

Edited by Nils Stenseth, Universitetet i Oslo, Oslo, Norway; received October 20, 2024; accepted June 10, 2025

The fossil record provides the most powerful evidence of large-scale biodiversity change on Earth, but it does so at coarse and often idiosyncratic temporal scales. One critical problem that arises concerns the evolutionary consequences of individual environmental experience. Individuals respond to their environment instantaneously, whereas the resolution of most fossil records aggregates multiple paleoenvironments over time scales beyond individual lifespans. Therefore, the presence of phenotypic plasticity in deep time and the extent of its influence on macroevolution remain poorly understood. Using coupled computed tomography and laser ablation inductively coupled plasma mass spectrometry protocols, we studied the environmental dependence of developmental trajectories across three sister species of macroperforate planktonic foraminifera. A foraminiferal shell preserves all stages of the individual's ontogeny, as well as the environmental state experienced throughout its lifetime. Generalized additive mixed effect (GAMM) models show that somatic growth rates differ among the three *Menardella* species and that these are inversely correlated with calcification temperature, as reconstructed from Mg/Ca measurements through ontogeny. This environmental dependence varies among species: The thermal sensitivity of individual chamber-to-chamber growth rates of *Menardella limbata* and *Menardella pertenuis* is double that seen in *Menardella exilis*. In contrast, no such environmental signal was recovered for architectural shape traits. Our integrated approach is widely applicable and demonstrates that detecting developmental plasticity in the fossil record is feasible. Extrapolating these techniques in deep time promises to revolutionize our understanding of the ways in which environmentally associated trait variation drove the diversification of life on Earth.

developmental plasticity | reaction norm | deep time | planktonic foraminifera

Significance

Traits are determined by internal factors such as genetics and plastic responses to the external environment. Being able to distinguish which factor most affected the evolution of each trait frames the capacity for organisms to adapt to a changing environment. Most fossils do not carry detailed information on environmental change during their lifetime, but here we study plankton that preserve both their developmental morphological state and their environmental habitat in their shells. We identify environmentally associated trait changes within fossilized individuals from millions of years ago. Our approach can be readily applied to other organisms and help us understand the relative roles of genetically and environmentally associated variation in the origin of new species.

Phenotypic variation forms the basis for micro- and macroevolutionary analysis of the fossil record. Phenotypes form through a combination of genetic and nongenetic components; many genotypes are capable of producing different phenotypes depending on environmental conditions and cues (1). While the fossil record represents an excellent resource for documenting the generation and proliferation of new biodiversity on the largest scales, the common lack of temporal and spatial resolution and the paucity of sufficiently detailed information on the experienced environment have prevented attempts to discriminate how much of the variation in form is environmentally associated (2, 3). Many lines of evidence have been used to speculate that environmentally induced phenotypic plasticity, the capacity of a single genotype to give rise to multiple phenotypes, introduces new variation and therefore has the potential to promote speciation (4, 5). There is growing evidence that phenotypic variation shapes species divergence in deep time (6), but the lack of empirical data means that the role of phenotypic plasticity in macroevolutionary change remains unknown (7–9).

Ecophenotypic, i.e. environmentally associated, variation is often invoked to explain morphological variation in the fossil record (3), but widely accepted evidence is currently lacking. Phenotypic plasticity in modern analogue species is sometimes used to argue for similarly plastic traits in fossil ancestors (e.g., refs. 10, 11–13), but most fossil species do not have modern analogues and, in those that do, the same phenotypic differences can be both environmentally and genetically controlled (14, 15). A high degree of morphological variation within a species (16–21), or morphological variation among populations in different environmental settings (13, 16, 22–26) has often been ascribed to ecophenotypic variation in the fossil record but a purely genetic basis for this variation cannot be ruled out from studying morphology alone (27–29). To quantify the potential adaptability of form in fossil species and determine how much of that versatility flows from environmental influence (3), it is necessary to calculate how much of that variability is environmentally associated.

Author contributions: T.H.G.E. designed research; A.B., A.S.-B., and J.M.M. performed research; C.D.S., J.A.M., O.L.K., R.A.W., C.T., I.S., P.A.W., and G.L.F. contributed new reagents/analytic tools; A.B. analyzed data; and A.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: Anieke.Brombacher@noc.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2421549122/-/DCSupplemental.

Published July 3, 2025.

In extant populations, hypotheses concerning the balance between plastic and genetic trait changes are tested either through experimental manipulation of genetically identical individuals or by gathering repeated measurements on the same individuals experiencing different environments through their lifetimes (30, 31). In the fossil record, the most obvious limiting factor in establishing the extent to which variability results from plastic or genetic properties is the lack of coupled contemporaneous environmental data at sufficient fine-scale temporal resolution. Changes in ontogenetic trajectories (16, 21, 26, 32, 33) or the degree of asymmetry in ontogenetic trajectories (24, 34) have been argued to indicate environmentally associated variation in deep time, but these approaches cannot rule out purely genetic variation. To partition trait variation into genetic and environmentally associated components, we need detectable morphological and environmental signals at biologically relevant temporal resolutions.

Here, we map life history and shell architecture traits of individuals in three species of planktonic foraminifera (*Menardella limbata, Menardella exilis,* and *Menardella pertenuis*, *SI Appendix*, Fig. 1) to the environments experienced by individuals during their life. Each foraminifer grows by adding a new calcite chamber to its overall shell as a discrete ontogenetic stage every few days (Fig. 1). We develop repeatable, high-resolution laboratory-based protocols to map morphological developmental trajectories extracted from X-ray microcomputed tomography (CT, Fig. 1) (35) to the changing environment through the life of each fossil individual as recorded in the calcite Mg/Ca ratios and measured through laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) (36, 37). We compare individual ontogenetic trajectories from developmental stage t to stage $t + 1$ to ambient temperature recorded in the calcite Mg/Ca ratios at stage t. We then aggregate these repeated morphology-environment associations into generalized additive mixed models to build environmental reaction norms in deep time and determine how much trait variation is best explained by an individual's response to the environmental changes it experienced millions of years ago.

1. Results

Menardella pertenuis reaches a larger final size than its ancestor *M. limbata* (two-sample t test, t = 2.48, df = 16.82, P = 0.012, Fig. 2*A*) with two more chambers on average (two-sample t test, t = 3.37, df = 22.47, P = 0.0014). *Menardella exilis* is smaller (two-sided t test, t = 4.03, df = 28.10, $P < 0.001$) and has one less chamber on average (two-sided t test, t = 2.13, df = 43.98, P = 0.019) than *M. limbata*. All three species are characterized by high growth rates early in ontogeny, which decrease later in life (Fig. 2*B*). Growth rates of descendants *M. exilis* and *M. pertenuis* remain lower than those of *M. limbata* early in ontogeny and decrease at a later stage and at larger shell volume than their ancestor. Angles between subsequent chambers do not change detectably through ontogeny for *M. exilis* but increase later in life for *M. limbata* and *M. pertenuis* (Fig. 2*C*). Trochospirality does not differ detectably throughout ontogeny for all studied species (Fig. 2*D*).

Variation in chamber-to-chamber growth rates through an individual's life is best explained by the generalized additive mixed model (GAMM) as a species-specific ontogenetic response to the Mg/Ca temperature proxy (model 7, Table 1); models that instead include internal constraints such as ontogeny (models 2 and 3) and cumulative volume (models 4 and 5) have markedly lower support. The linear mixed effect components of the GAMM showed no detectable impact of species-specific somatic growth rates versus the nested null model, unlike Model 6, which incorporated a general environmental response to ln(Mg/Ca) across all species, and Model 7, which incorporated both the general linear relationship between ln(Mg/Ca) and the nonlinear species-specific environmental response (Table 1). This species-specific environmental response model also outperformed the general environmental response model. Within an individual's life, each unit increase in ln(MgCa), which corresponds to a temperature change of ~6.4 ± 0.9%/°C (38), is associated with an average change in chamber-to-chamber growth rate of −0.0795 (SE = 0.0283, t = −2.81, $P < 0.01$). The species-specific environmental model explained ~9% more variation than the general environmental response model (adjusted r^2 of 0.537 and 0.491, respectively). The higher Akaike weights of models that link ambient temperature at stage t to growth from stage t to $t + 1$ in our fossil individuals (Table 1) is evidence that these species differ not only in their overall morphology (Fig. 2) but how the environment shapes each individual's growth trajectory (Fig. 3).

With respect to the architectural traits, chamber-to-chamber angles within individuals are affected by the volume of the existing

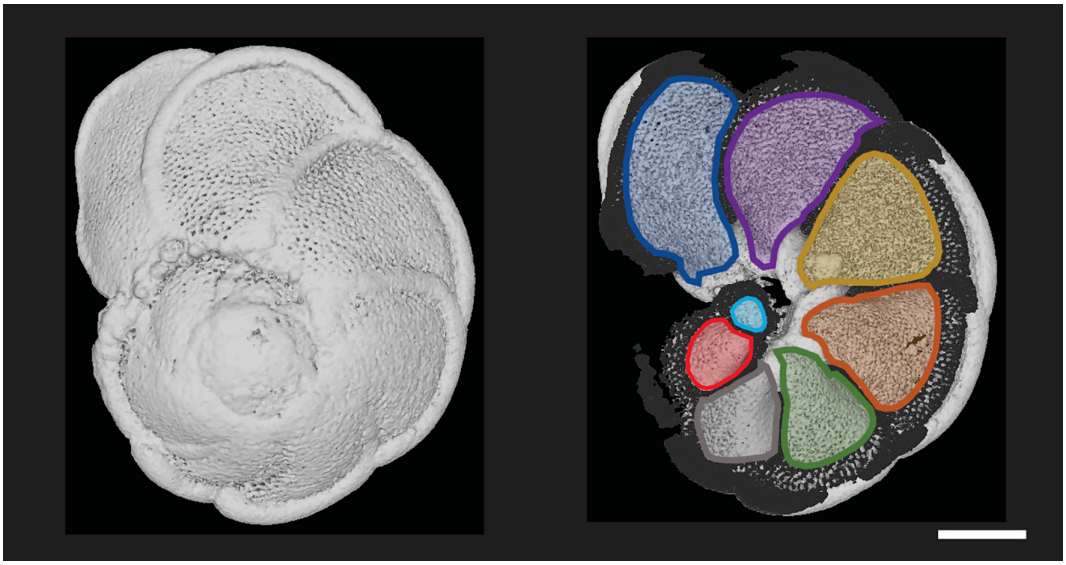


Fig. 1. CT scan of *Menardella limbata* specimen (*Left*) and a 3D reconstruction of a cross-section of the same specimen with individual chambers highlighted in different colors (*Right*). The scale bar represents 100 μm. Visualized in Dragonfly version 2021.3 (Object Research Systems, Canada).

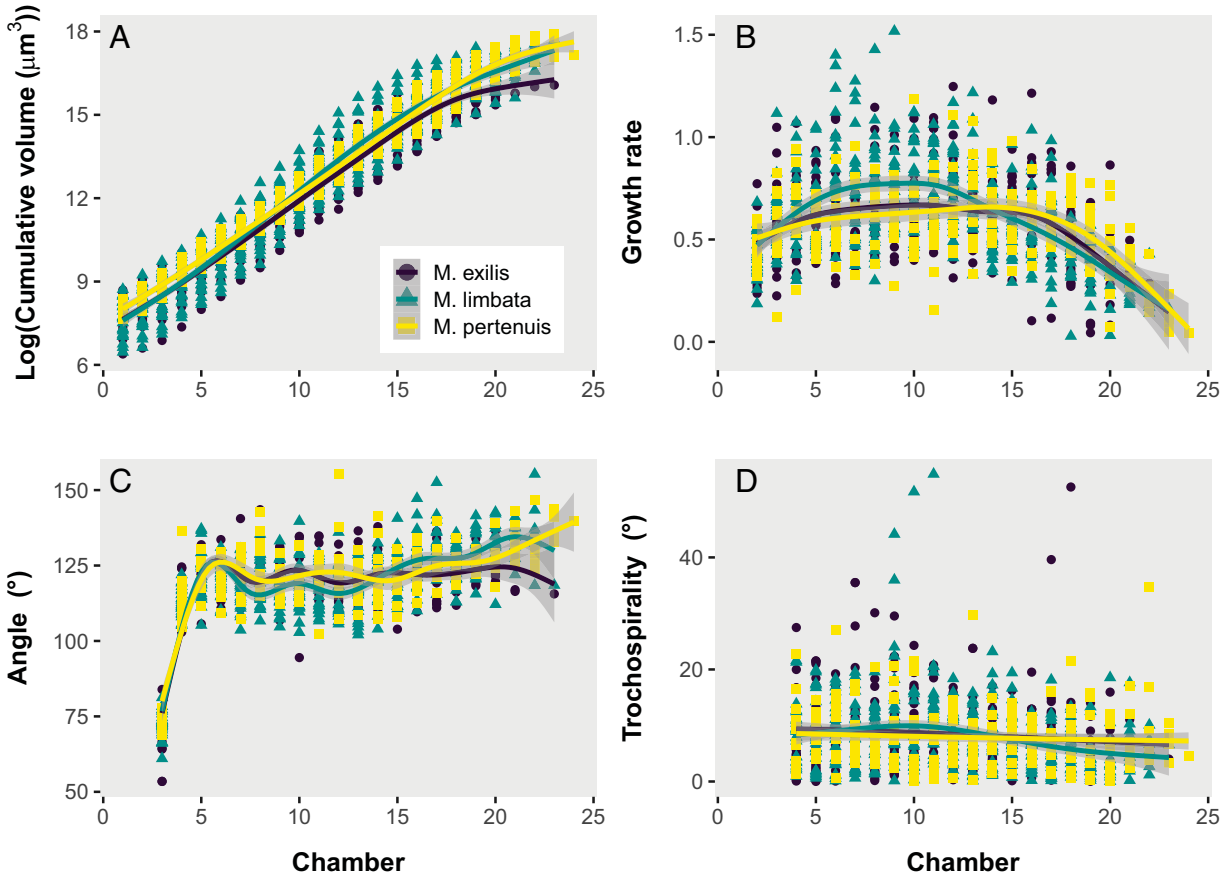


Fig. 2. Cumulative volume (*A*), growth rates (*B*), chamber angles (*C*), and trochospirality (*D*) throughout ontogeny for *Menardella limbata*, *Menardella exilis,* and *Menardella pertenuis*. Smooth lines represent generalized additive model results, with shaded areas for the models' 95% CI.

shell at the time the chamber was built (i.e., stage *t*), not detectably by ambient temperature at stage *t*, and the nonlinearity of the relationships varies among species (Table 2). Trochospirality is also affected by the volume of the existing shell at the time the chamber was built, not detectably by ambient temperature at stage *t* and without any clear species-specific relationships (Table 3).

In all species, somatic growth rate slows toward the terminal (reproductive) life stage but the rate of slowing depends on species-specific responses to temperature as documented by ln(Mg/Ca) of the focal chamber. Growth rates among species are most similar at lowest temperatures (lowest Mg/Ca values). Model-averaged predictions show that chamber-to-chamber growth rates in *M. exilis* are less affected by temperature changes through life than either *M. limbata* or *M. pertenuis*. *Menardella limbata* and *M. pertenuis* exhibit lower somatic growth when temperatures are higher (Fig. 3). The LA-ICP-MS measurements are highly repeatable: The uncertainty associated with repeated measurements of the same chamber is five orders of magnitude smaller than that associated with repeated measurements across chambers of the same individual ($3.88*10^{-6}$ and 0.06, respectively), which implies that within-chamber Mg/Ca variation is negligible compared to among-chamber, i.e. through ontogeny, Mg/Ca variation.

While differential environmental dependence of *M. limbata* and *M. pertenuis* is not detectable through ln(Mg/Ca) reaction norm differentiation, *M. pertenuis* grows a calcite shell that is less than 90% of the thickness of either *M. exilis* or *M. limbata* (gamma generalized linear model with inverse link function, $\beta = 0.00396$, SE = 0.00137, $P < 0.05$; Fig. 4). Together these variations indicate environmentally dependent developmental plasticity throughout the life of these organisms as individuals within the three species respond to their environments in detectably different ways.

2. Discussion

We show a reconstruction of species-specific environmental dependencies throughout the life cycle of fossil individuals. Despite differences in chamber-to-chamber growth rates among species, ontogenetic stage was not the best predictor for growth rate variation (Table 1). Instead, the Mg/Ca temperature proxy for each ontogenetic stage *t* markedly improved predictions of chamber-to-chamber growth from stage *t* to stage *t* + *1* (Fig. 3). While the overall dependence of growth on temperature agrees with patterns seen in biogeographic size distributions (39, 40) and in physiological models (41), we detect weaker environmental sensitivity in *M. exilis* than either *M. limbata* or *M. pertenuis* (Fig. 3).

In contrast to growth rates, the parameters describing the shape of the shell growth spiral show very little response to external influences (Tables 2 and 3). Angles between subsequent chambers depend mainly on the total volume of the shell at the time the chamber was built, with specific relationships varying among species. For *M. limbata* and *M. pertenuis*, angles remain stable through the first half of ontogeny, followed by a gradual increase in the second half of their life. One explanation of this result is linked to a change in growth rate, with relatively smaller chambers completing a smaller section of the whorl and so increasing angles between chambers, but the model including growth rates as a response variable does not explain the data as well as the model with cumulative volume (Table 2). This suggests that, regardless

Table 1. Generalized additive mixed effect (GAMM) model results of Menardella chamber growth rate response to internal and external processes

Model	Degrees of freedom	r^2	AIC	ΔAIC	Akaike weight
m7: species-specific responses to Mg/Ca	**9**	**0.538**	**−403.3**	**0**	**0.959**
m6: similar response to Mg/Ca across all species	8	0.491	−396.1	7.17	0.027
m8: similar response to temperature with ontogenetic differences across all species	9	0.492	−394.6	8.67	0.013
m2: similar response to chamber size across all species	7	0.484	−389.1	14.16	0.001
m3: species-specific responses to chamber size	8	0.485	−388.2	15.1	0.001
m4: similar response to cumulative volume across all species	7	0.371	−291.3	111.93	0
m5: species-specific responses to cumulative volume	8	0.377	−290.2	113.1	0
m1: fixed differences across species, without change through ontogeny	5	0	−111.5	291.76	0
m0: random effects only	5	0	−111.5	291.76	0

Log likelihoods that feed into AIC scores are presented based on maximum likelihood calculations from the linear mixed effect part of the GAMM to facilitate comparisons among fixed effects; random effects for all models were specimen and laser blast (where the same chamber was repeatedly measured to assess instrument consistency). Adjusted r2 calculations are drawn from the generalized additive model part of the GAM. ΔAIC represents the difference between AIC and the set's minimum AIC. The best performing model based on lowest AIC and highest Akaike weight is indicated in bold. All models include autocorrelation within specimens, and specimens and laser shots as random effects. The best performing model includes a general linear relationship with ln(Mg/Ca), as a linearized proxy for temperature of the calcification environment, and a species-specific nonlinear spline between Mg/Ca and species classification.

of temperature-associated growth variation, individuals adjust chamber their shape to maintain their overall spiral shape.

Planktonic foraminifera are known to show temperature-dependent chamber-to-chamber growth rates (42, 43). They reproduce at the end of their life by transforming all cell cytoplasm into gametes (36) and individuals must grow large enough to produce enough gametes quickly enough to reduce the chance of death before reproduction. Growth is limited by the energetic cost of calcification (41). This trade-off between costs and benefits of high growth rates results in different life-history trajectories among species: While *M. limbata* and *M. exilis* have indistinguishable environmental sensitivities, the descendant *M. pertenuis* reaches the same overall size using <90% calcite in its shell. The canonical view of *Menardella* depth habitat through ontogeny includes reproduction in the pycnocline (~30 to 40 m water depth depending on region), followed by downward migration to up to 400 m water depth, and ending with mature specimens ascending back to the pycnocline for reproduction (44). Water temperature decreases with depth, so this would suggest a strong relationship between ontogenetic stage and Mg/Ca, but we do not see this in our data (Table 1, model 7). Meilland, Siccha, Kaffenberger, Bijma, and Kucera (45) also reported that more than half of the planktonic foraminifera they encountered in plankton tows did not follow classic depth habitat patterns, suggesting large variation in depth habitat within ontogenetic stages. If temperature is indeed the largest determinant of chamber-to-chamber growth rate, then variation in depth habitat increases variation in calcification temperature. Given chamber-to-chamber growth is the major determinant of population growth (46), the temperature dependence we report here is closely linked to mean population fitness for this life history.

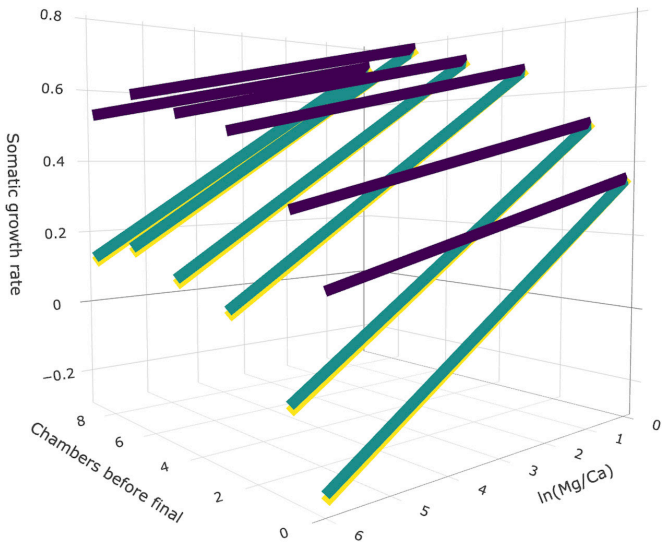


Fig. 3. Generalized additive mixed model results for somatic growth rates, plotted against both chamber number (ontogenetic stage) and Mg/Ca. Note that "chambers before final" indicates ontogenetic growth stage counting back because individuals grow various numbers of chambers (e.g., Fig. 2) and the experimental control is to count back from the final (terminal) growth stage.

Our framework for detecting deep time environmentally associated changes in morphological traits throughout the life cycle of fossilized individuals shows that studying developmental plasticity in the fossil record is feasible. Our observations open a door to help determine whether environmentally associated phenotypic plasticity is involved in driving macroevolutionary change (3). Saulsbury et al. (47) showed how the internal variational properties predict individual growth rates more strongly than external abiotic factors across 195 species of living marine bivalves, while Palmer, Moss, Surge, and Turek (48) showed that mid- to high-latitude bivalves were more impacted by abiotic changes than low latitude individuals. Neither of these studies were able to break down the growth rate analysis into discrete ontogenetic stages because they lacked the contemporaneous environmental resolution we show here. Our approach supersedes these previous approaches because it allows us to detect the existence of environmentally associated developmental plasticity (Fig. 3). Given model predictions that increasing developmental plasticity would increase mean population fitness (46), future investigations might investigate whether species tolerances to temperature are due to changing environmental associations within individuals through ontogeny, or rather across individuals through populations, and whether that environmental association changes before, during, or after speciation.

Levis and Pfennig (7) proposed four criteria to investigate plasticity-led evolution: the opportunity to (1) detect environmentally dependent plastic traits in ancestral and derived

Table 2. GAMM model results of Menardella chamber angle in response to internal and external processes

Model	Degrees of freedom	r^2	AIC	ΔAIC	Akaike weight
m5: species-specific responses to cumulative volume	**8**	**0.27**	**3,126**	**0**	**0.916**
m4: similar response to cumulative volume across all species	7	0.236	3,130.8	4.78	0.084
m7: species-specific response to Mg/Ca	9	0.174	3,158.1	32.11	0
m3: species-specific responses to chamber size	8	0.098	3,158.7	32.66	0
m2: similar response to chamber size across all species	7	0.122	3,159.6	33.58	0
m8: similar response to temperature with ontogenetic differences across all species	9	0.096	3,160.7	34.66	0
m6: similar response to Mg/Ca across all species	8	0.12	3,161.6	35.57	0
m9: similar response to growth rate across all species	7	0.147	3,177.4	51.36	0
m10: species-specific response to growth rate	8	0.147	3,179.4	53.36	0
m1: fixed differences across species, without change through ontogeny	7	0.059	3,202.5	76.47	0
m0: random effects only	5	0	3,212.2	86.22	0

Log likelihoods that feed into AIC scores are presented based on maximum likelihood calculations from the linear mixed effect part of the GAMM to facilitate comparisons among fixed effects; random effects for all models were specimen and laser blast (where the same chamber was repeatedly measured to assess instrument consistency). Adjusted r2 calculations are drawn from the generalized additive model part of the GAM. ΔAIC represents the difference between AIC and the set's minimum AIC. The best performing model based on lowest AIC and highest Akaike weight is indicated in bold. All models include autocorrelation within specimens, and specimens and laser shots as random effects. The best performing model includes a general and species-specific nonlinear spline as a function of cumulative shell volume to the time of calcification.

lineages and (2) reconstruct reaction norms through time with (3) a sufficient sample size to detect an increase in trait variation following environmental changes and (4) sufficient resolution to track mean changes in the focal trait. Our approach, using high-resolution fossil records that preserve ontogenetic stages and environmental variation at the time of growth, fulfills these criteria and can be readily extended to other calcifying marine organisms that preserve both environmental and ontogenetic information in their skeletons. Ammonoids are a prime example, preserving both ontogeny and environmental information (e.g., ref. 49). Additionally, corals and bivalves have long been used to reconstruct past sea surface temperature, salinity, and pH (e.g., refs. 50, 51). Combining their growth layers with the environmental information stored within them can reveal valuable information on their potential for adaptation in changing environments. Finally, fossil bryozoans, colonial marine organisms that have existed since the Ordovician, contain information on both their life history and environment (52), and are known to exhibit developmental plasticity in response to changing environmental conditions (53).

If trait-based approaches are to provide a more mechanistic route to understanding species vulnerabilities (54, 55), then generalizing the contributions of environmentally associated trait variation is an achievable goal to span both the differences in temporal scope among data (56) and the timeframes of fossil preservation and necessary conservation action (54). The ability to extract the lived environment of organisms in time-averaged fossil samples is not only reassuring for the adequacy of temporal age models but also provides a quantitative metric of the sources of the morphological variation (27–29). The Levis and Pfennig (7) criteria were not set out explicitly for paleontological systems, but the acceleration of geochemical and imaging technologies, alongside the integration of methodological advancements used in this study, now provide exciting opportunities to investigate how environmentally associated trait variation contributed to the generation, proliferation, and extinction of life on Earth.

Table 3. GAMM model results of Menardella trochospirality in response to internal and external processes

Model	Degrees of freedom	r^2	AIC	ΔAIC	Akaike weight
m4: similar response to cumulative volume across all species	**7**	**0.036**	**2,947.6**	**0**	**0.615**
m5: species-specific responses to cumulative volume	8	0.043	2,949.3	1.71	0.262
m9: similar response to growth rate across all species	7	0.023	2,952.6	5.07	0.049
m0: random effects only	5	0	2,954.5	6.97	0.019
m10: species-specific response to growth rate	8	0.023	2,954.6	7.07	0.018
m2: similar response to chamber size across all species	7	0.006	2,955.9	8.35	0.009
m6: similar response to Mg/Ca across all species	8	0.006	2,955.9	8.33	0.01
m1: fixed differences across species, without change through ontogeny	7	0.01	2,956.2	8.64	0.008
m3: species-specific responses to chamber size	8	0.006	2,957.9	10.35	0.003
m7: species-specific response to Mg/Ca	9	0.006	2,957.9	10.33	0.004
m8: similar response to temperature with ontogenetic differences across all species	9	0.006	2,957.9	10.33	0.004

Log likelihoods that feed into AIC scores are presented based on maximum likelihood calculations from the linear mixed effect part of the GAMM to facilitate comparisons among fixed effects; random effects for all models were specimen and laser blast (where the same chamber was repeatedly measured to assess instrument consistency). Adjusted r2 calculations are drawn from the generalized additive model part of the GAM. ΔAIC represents the difference between AIC and the set's minimum AIC. The best performing model based on lowest AIC and highest Akaike weight is indicated in bold. All models include autocorrelation within specimens, and specimens and laser shots as random effects. The best performing model includes a general nonlinear spline as a function of cumulative shell volume to the time of calcification.

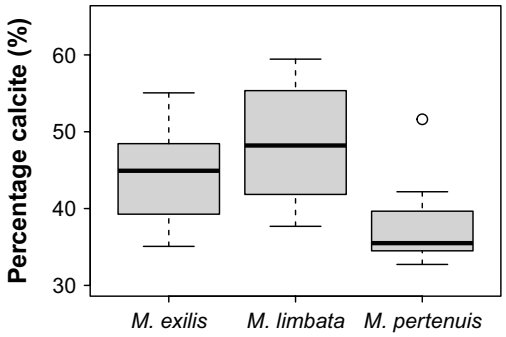


Fig. 4. Percentage calcite of the total shell volume for *Menardella limbata*, *Menardella exilis,* and *Menardella pertenuis*.

3. Methods

3.1. Study Species. We studied the Neogene planktonic foraminifera species *Menardella limbata* and its two descendent species *Menardella exilis* and *Menardella pertenuis* (*SI Appendix*, Fig. 1), which originated in the Pliocene, ~4.5 and ~3.5 Mya (57). All three species are exclusively found in tropical and warm subtropical waters. Their shells are characterized by lenticular, low trochospiral tests with a prominent keel (58). *Menardella exilis* and *M. pertenuis* can be distinguished from *M. limbata* by a thinner, "shinier" shell and lower trochospire, with *M. pertenuis* additionally possessing a thin plate extending over the umbilicus and more chambers in the final whorl than *M. exilis* (58). All three species became extinct between 2.6 and 2.2 Ma (57), shortly after Earth had cooled sufficiently to promote the waxing and waning of large ice sheets on North America and Eurasia for the first time in the Cenozoic era (59, 60).

3.2. Study Site. Ocean Drilling Program (ODP) Site 925 is located in the western equatorial Atlantic at a water depth of 3,040 m (4° 12′ N, 43° 29′ W), well above the local carbonate lysocline (61). The age model for Site 925 was recently updated by high-resolution benthic oxygen isotope stratigraphy (62). The site sits in the core of the biogeographic range of all three study species, which are found in high abundance in Pliocene sediments (57). The first appearance dates of *M. exilis* and *M. pertenuis* are well documented at Site 925 (57) and agree with global low-latitude biostratigraphic schemes (63, 64).

We analyzed five time slices: one before the origination of *M. exilis* and *M. pertenuis*, one during the speciation of *M. exilis*, one after the origination of *M. exilis* but before the origination of *M. pertenuis*, one during the speciation of *M. pertenuis*, and one when all three species had become established. For every time slice, we analyzed 3 to 8 specimens of all species occurring in that time slice. Sediment samples were dry-sieved over a >150 μm mesh sieve and divided using a microsplitter. Specimens were selected for CT scanning through assignment of random numbers to avoid further biases.

3.3. Micro-CT Scanning and 3D Shell Model Reconstruction. To maximize throughput, prior to CT scanning specimens were stacked one on top of the other separated by ~2 mm layers of polyurethane foam in transparent plastic drinking straws (2 mm diameter). The foam and plastic straw have significantly lower attenuation properties compared to foraminiferal calcite which makes it easy to exclude from the CT data. This allowed individual identification for subsequent geochemical work and programming of a multiscan session. Scans were conducted at the μ-VIS X-ray Imaging Centre, University of Southampton, UK by a Versa 510 X-ray microscope (Carl Zeiss Microscopy GmbH, Germany) using a tungsten transmission target, a peak voltage of 110 kV at 10 W and a 0.15 mm SiO2 glass filter to reduce beam hardening artifacts. Two-stage magnification was achieved by utilizing a 4× objective. To increase the signal-to-noise ratio, the detector was binned twice, resulting in a voxel resolution of 1.75 μm. Each scan was focused on a single specimen and imaged using 1,011 projections at 1.3 s exposure time per radiograph projection. The projection data were then reconstructed using a sharp reconstruction filter with a beam hardening correction of 1.0 to further reduce beam hardening artifacts into a 16-bit tiff stack [32 to 16-bit window: (−0.04, 0.10)] characterized by an average dimension of 992 × 1,015 pixels for each 2D slice using the Zeiss XM FBP Reconstruction software (Carl Zeiss Microscopy GmbH, Germany).

3.4. Three-Dimensional Growth Analyses. We analyzed three-dimensional growth trajectories of all scanned foraminifera focusing on ecologically relevant life-history traits (46, 65). We use chamber number as an indicator of ontogenetic stage and calculate size at each ontogenetic stage *t* as the sum of the volumes of all chambers present at the time chamber *t* was built. To reconstruct individual chambers, all specimen CT scans were processed using Dragonfly [66 (version 2021.3)]. Each specimen's volume data were imported into Dragonfly where the XY, XZ, and YX planes were manually aligned to the specimen's proloculus centroid. Using both the smart grid tool and manual brush we labeled each chamber's internal area on a slice-by-slice basis (Fig. 1). All ROIs assigned to a single chamber were combined into a multi-ROI layer, for which Dragonfly calculated the volume and the *x*, *y,* and *z* centroid coordinates. Ten specimens were analyzed by two members of the team to assess repeatability; the mean proportion of variation explained by a regression across the two repeats of chamber volume was 0.977 (range across specimens from 0.912 to 0.999). All final data presented come from a single operator who repeated all measurements for the fullest consistency.

Chamber-to-chamber (somatic) growth rates were calculated as the increase in total shell volume for every chamber added relative to the size at the previous chamber addition. To analyze spiral growth in three dimensions, we extracted centroid coordinates of all chambers from the Dragonfly specimen reconstructions. Using the foram3D package (65) in R, version 4.2.1 (R Core 67) we calculated the angle a chamber makes with the previous two chambers, and trochospirality at each ontogenetic step, where trochospirality is the angle a chamber is added at relative to the plane spanned by the previous three chambers [see figure 11 in Brombacher, Searle-Barnes, Zhang, and Ezard (65)]. Changes in the angle of new chamber addition point to changes in the number of chambers in the final whorl, a key diagnostic feature for *Menardella* species, whereas trochospirality is known to change with ontogenetic stages such as the transition from juvenile to neanic and could thus be used to quantify growth stages (68–74).

3.5. Laser Ablation Inductively Coupled Plasma Mass Spectrometry. We use Mg/Ca ratios in the foraminiferal chamber calcite as a proxy for temperature at the time of chamber formation as foraminiferal calcite Mg/Ca ratios increase with ambient temperature (75, 76). Although specific temperature-Mg/Ca relationships vary among species (75) and these curves are impossible to parameterize experimentally for extinct species, for the purpose of this study, the positive correlation between temperature and Mg/Ca is sufficient to investigate the effect of temperature on ontogenetic trajectories within individuals of our sister species.

Mg/Ca ratios were determined using laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS). Following CT scanning, specimens were mounted on glass slides using double-sided adhesive tape. Trace elements in the foraminiferal calcite were analyzed using an Agilent (Agilent Technologies Inc., CA, USA) 8900 triple quadrupole inductively coupled plasma (QQQ-ICP) mass spectrometer coupled to an Elemental Scientific Lasers (Bozeman, MT, USA) NWR193 excimer laser ablation system at the School of Ocean and Earth Science, University of Southampton. The following isotopes were acquired in time resolved mode: ^{7}Li, ^{9}Be, ^{11}B, ^{24}Mg, ^{27}Al, ^{43}Ca, ^{44}Ca, ^{55}Mn, ^{65}Cu, ^{66}Zn, ^{88}Sr, ^{137}Ba, and ^{238}U (only ^{24}Mg and ^{44}Ca are presented here). The final chamber (f), penultimate chamber (f-1), and chamber f-3 were ablated in two separate locations to quantify within-chamber variation (see also Kearns et al. 2023). Chambers f-5, f-7, and f-9 were ablated in a single location due to limited area.

Each ablation was performed using a 30 μm diameter spot, with a repetition rate of 5 Hz at a fluence of 0.73 J/cm^{2}. Each acquisition was preceded by a 15-s gas blank/laser warmup and was followed by a 30-s "wash" (to permit residual material to leave the system). Ten replicates of NIST 612 and 610 reference glasses, a pressed powder pellet of MACS-3 (United States Geological Survey), and a pressed powder pellet of JCp-1 (Porites sp. coral, Japanese Geological Survey) were analyzed in batches throughout each analysis session after every 80 to 100 analyses.

Each time-resolved analysis was background subtracted using the gas blank measured immediately before each sample (or standard). Sample and standard data were internally normalized using Ca wt% using the following concentrations: 37.7 wt% MACS-3, 37.5 wt% JCp-1, 8.5 wt% NIST 612, and 8.15 wt% NIST 610 (77, 78). Elemental concentrations in the chamber walls were determined using a calibration derived from NIST 612, NIST 610, and MACS-3 (reference concentrations from ref. 77). Following ref. 79, a pressed powder pellet of JCp-1 was

analyzed as an unknown to assess accuracy (using the same calibration strategy). Average concentrations in JCp-1 were determined as follows: Mg was 836 ± 124 ppm (vs. 867 ± 23 ppm in ref. 77) and Ca was 38.1 ± 3.8% Ca [vs. 37.5 ± 2.4% by Sekimoto et al. (80)]. After calibration, our NIST glass calibrations thus predict trace elements within carbonate target material within the specified uncertainty limits. Trace element ratios were extracted using the lablaster R package (81) and presented as the median of all on-target time slices, applying the endPoint function (dt = 10, smoothing = 10) to identify when the laser fully penetrated the chamber wall.

3.6. Statistical Analyses. To investigate the effects of internal and external constraints on *Menardella* growth rates and morphology within and among species, we employed generalized additive mixed models (GAMMs) using the "mgcv" package (82) in R [version 4.2.1 R Core Team (67)]. GAMMs consist of two parts: a linear mixed effect part and a generalized additive model part. Growth rates, trochospirality, and chamber angles were modeled as potentially nonlinear spline relationships against: i) chamber number (ontogenetic stage), ii) cumulative volume (to investigate for size thresholds for changes in growth rates), and iii) Mg/Ca ratios as indicator of temperature of the calcification environment during the lived life-cycle of the now fossil individual. The models for trochospirality and chamber angles additionally included growth rates as fixed effect to investigate possible dependency of architectural traits on growth rates. Specimen ID code and laser shots (in the case of multiple shots of the same chamber) were included as random effects in all models to facilitate model comparisons using Akaike Information Criterion values (AIC) (83). The species-level splines were specified using an "interactive" fixed effect between the explanatory variables above and the species classification. Splines were penalized using the default settings. If models with the species-specific spline performed better than those without, then there is evidence that the effect of environment (via the Mg/Ca proxy) on ontogenetic trajectories differs among species. The best performing model of those fitted has the lowest AIC, which combines deviance explained and the number of parameters used, and thus the highest Akaike weight (83). All R code used to fit the models and output results tables are available as a *SI Appendix*.

Data, Materials, and Software Availability. Ontogenetic and environmental data in a .xlsx file R code for the statistical analyses in a .R file data have been deposited in Figshare (https://doi.org/10.6084/m9.figshare.27260970) (84).

ACKNOWLEDGMENTS. We thank two anonymous reviewers for their constructive feedback on an earlier version of this manuscript. This study was funded by the Natural Environment Research Council (NE/P019269/1). We acknowledge μ-VIS X-ray Imaging Centre (https://muvis.org), part of the National Facility for laboratory-based X-ray CT (nxct.ac.uk–EPSRC: EP/T02593X/1) at the University of Southampton for the provision of the μCT imaging, processing, and data management infrastructure.

Author affiliations: [a]School of Ocean and Earth Science, University of Southampton, Southampton SO14 3ZH, United Kingdom; [b]Department of Earth and Planetary Sciences, Yale University, New Haven, CT 06511; [c]Division of Science and Technology, National Oceanography Centre, Southampton SO14 3ZH, United Kingdom; [d]Department of Life Sciences, Natural History Museum, London SW7 5BD, United Kingdom; [e]μ-VIS X-ray Imaging Centre, Faculty of Engineering and the Environment, University of Southampton, Southampton SO17 1BJ, United Kingdom; [f]Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom; [g]Electronics and Computer Science, University of Southampton Highfield Campus, Southampton SO17 1BJ, United Kingdom; and [h]Engineering Materials, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom

1. R. C. Lewontin, *The Genetic Basis of Evolutionary Change* (Columbia University Press, New York, USA, 1974).
2. G. J. Vermeij, Biological versatility and Earth history. *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1936–1938 (1973).
3. A. C. Love *et al.*, Evolvability in the fossil record. *Paleobiology* **48**, 1–24 (2022).
4. M. J. West-Eberhard, Developmental plasticity and the origin of species differences. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6543–6549 (2005).
5. M. J. West-Eberhard, *Developmental Plasticity and Evolution* (Oxford University Press, New York, 2003).
6. A. Holstad *et al.*, Evolvability predicts macroevolution under fluctuating selection. *Science* **384**, 688–693 (2024).
7. N. A. Levis, D. W. Pfennig, Evaluating "Plasticity-First" evolution in nature: Key criteria and empirical approaches. *Trends Ecol. Evol.* **31**, 563–574 (2016).
8. N. A. Levis, D. W. Pfennig, Plasticity-led evolution: A survey of developmental mechanisms and empirical tests. *Evol. Dev.* **22**, 71–87 (2020).
9. D. J. Futuyma, Evolutionary biology today and the call for an extended synthesis. *Interface Focus* **7**, 1–13 (2017).
10. S. J. Hageman, M. M. Bayer, C. D. Todd, Partitioning phenotypic variation: Genotypic, environmental and residual components from bryozoan skeletal morphology. *J. Nat. Hist.* **33**, 1713–1735 (1999).
11. D. H. Geary, An unusual pattern of divergence between two fossil gastropods: Ecophenotypy, dimorphism, or hybridization? *Paleobiology* **18**, 93–109 (1992).
12. E. M. Standen, T. Y. Du, H. C. Larsson, Developmental plasticity and the origin of tetrapods. *Nature* **513**, 54–58 (2014).
13. S. C. Antón, C. W. Kuzawa, Early homo, plasticity and the extended evolutionary synthesis. *Interface Focus* **7**, 1–15 (2017).
14. D. M. Raup "Approaches to morphologic analysis" in *Models in Paleobiology*, T. J. M. Schopf, Ed. (Freeman & Cooper, San Francisco, USA, 1972), pp. 28–44.
15. K. F. Darling, M. Kucera, D. Kroon, C. M. Wade, A resolution for the coiling direction paradox in *Neogloboquadrina pachyderma*. *Paleoceanography* **21**, 1–14 (2006).
16. P. M. Sander, N. Klein, Developmental plastcity in the life history of a prosauropod dinosaur. *Science* **310**, 1800–1802 (2005).
17. T. M. Cronin, C. E. Schneider, Climatic influences on species: Evidence from the fossil record. *Trends Ecol. Evol.* **5**, 275–279 (1990).
18. G. Fryer, P. H. Greenwood, J. F. Peake, Punctuated equilibria, morphological stasis and the palaeontological documentation of speciation: A biological appraisal of a case history in an African lake. *Biol. J. Linn. Soc.* **20**, 195–205 (1983).
19. G. Fryer, P. H. Greenwood, J. F. Peake, The demonstration of speciation in fossil molluscs and living fishes. *Biol. J. Linn. Soc.* **26**, 325–336 (1985).
20. A. Kin, Early Maastrichtian ammonites and nautiloids from Hrebenne, southeast Poland, and phenotypic plasticity of *Acanthoscaphites tridens* (Kner, 1848). *Cretaceous Res.* **31**, 27–60 (2010).
21. B. M. Gee, Y. Haridy, R. R. Reisz, Histological skeletochronology indicates developmental plasticity in the early Permian stem lissamphibian *Doleserpeton annectens*. *Ecol. Evol.* **10**, 2153–2169 (2020).
22. G. P. Lohmann, B. A. Malmgren, Equatorward migration of *Globorotalia truncatulinoides* ecophenotypes through the Late Pleistocene: Gradual evolution or ocean change? *Paleobiology* **9**, 414–421 (1983).
23. J. F. Dynowski, J. H. Nebelsick, Ecophenotypic variations of *Encrinus liliiformis* (Echinodermata: Crinoidea) from the middle Triassic Muschelkalk of Southwest Germany. *Swiss J. Palaeontol.* **130**, 53–67 (2010).
24. N. Schlüter, Ecophenotypic variation and developmental instability in the Late Cretaceous echinoid *Micraster brevis* (Irregularia; Spatangoida). *PLoS One* **11**, 1–26 (2016).
25. R. R. Schoch, Life cycles, plasticity and palaeoecology in temnospondyl amphibians. *Palaeontology* **57**, 517–529 (2014).
26. S. Sanchez, A. de Ricqlès, R. Schoch, J. S. Steyer, Developmental plasticity of limb bone microstructural organization in Apateon: Histological evidence of paedomorphic conditions in branchiosaurs. *Evol. Dev.* **12**, 315–328 (2010).
27. A. M. Lister, "Phenotypic plasticity in the fossil record" in *Phenotypic Plasticity and Evolution*, D. W. Pfennig, Ed. (CRC Press, Boca Raton, 2021), chap. 11, pp. 267–297.
28. D. Jablonski, Developmental bias, macroevolution, and the fossil record. *Evol. Dev.* **22**, 103–125 (2020).
29. I. S. C. Jackson, Developmental bias in the fossil record. *Evol. Dev.* **22**, 88–102 (2020).
30. V. Oostra, M. Saastamoinen, B. J. Zwaan, C. W. Wheat, Strong phenotypic plasticity limits potential for evolutionary responses to climate change. *Nat. Commun.* **9**, 1–11 (2018).
31. A. Charmantier *et al.*, Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* **320**, 800–803 (2008).
32. A. L. A. Johnson, Detection of ecophenotypic variation in fossils and its application to a Jurassic scallop. *Lethaia* **14**, 277–285 (1981).
33. N. C. Hughes, Morphological plasticity and genetic flexibility in a Cambrian trilobite. *Geology* **19**, 913–916 (1991).
34. M. Webster, Morphological homeostasis in the fossil record. *Semin. Cell Dev. Biol.* **88**, 91–104 (2019).
35. D. N. Schmidt, E. J. Rayfield, A. Cocking, F. Marone, Linking evolution and development: Synchrotron radiation X-ray tomographic microscopy of planktic foraminifers. *Palaeontology* **56**, 741–749 (2013).
36. C. Hemleben, M. Spindler, O. R. Anderson, *Modern Planktic Formaminifera* (Springer, New York, 1989).
37. S. Eggins, P. De Deckker, J. Marshall, Mg/Ca variation in planktonic foraminifera tests: Implications for reconstructing palaeo-seawater temperature and habitat migration. *Earth Planet. Sci. Lett.* **212**, 291–306 (2003).
38. W. R. Gray, D. Evans, Nonthermal influences on Mg/Ca in planktonic foraminifera: A review of culture studies and application to the last glacial maximum. *Paleoceanogr. Paleoclimatol.* **34**, 306–315 (2019).
39. A. D. Hecht, An ecologic model for test size variation in recent planktonic foraminifera: Applications to the fossil record. *J. Foraminiferal Res.* **6**, 295–311 (1976).
40. S. Renaud, D. N. Schmidt, Habitat tracking as a response of the planktic foraminifer Globorotalia truncatulinoides to environmental fluctuations during the last 140 kyr. *Mar. Micropaleontol.* **49**, 97–122 (2003).
41. M. Grigoratou *et al.*, A trait-based modelling approach to planktonic foraminifera ecology. *Biogeosciences* **16**, 1469–1492 (2019).
42. F. Lombard *et al.*, Modelling planktic foraminifer growth and distribution using an ecophysiological multi-species approach. *Biogeosciences* **8**, 853–873 (2011).
43. F. Lombard, L. Labeyrie, E. Michel, H. J. Spero, D. W. Lea, Modelling the temperature dependent growth rates of planktic foraminifera. *Mar. Micropaleontol.* **70**, 1–7 (2009).
44. R. Schiebel, C. Hemleben, Modern planktic foraminifera. *Paläontol. Z.* **79**, 135–148 (2005).
45. J. Meilland, M. Siccha, M. Kaffenberger, J. Bijma, M. Kucera, Population dynamics and reproduction strategies of planktonic foraminifera in the open ocean. *Biogeosciences* **18**, 5789–5809 (2021).
46. A. Brombacher, D. N. Schmidt, T. H. G. Ezard, Developmental plasticity in deep time: A window to population ecological inference. *Paleobiology* **49**, 259–270 (2023).

47. J. Saulsbury *et al.*, Evaluating the influences of temperature, primary production, and evolutionary history on bivalve growth rates. *Paleobiology* **45**, 405–420 (2019).
48. K. L. Palmer, D. K. Moss, D. Surge, S. Turek, Life history patterns of modern and fossil *Mercenaria* spp. from warm vs. cold climates. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **566**, 110227 (2021).
49. J. A. Sessa *et al.*, Ammonite habitat revealed via isotopic composition and comparisons with co-occurring benthic and planktonic organisms. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15562–15567 (2015).
50. D. M. Thompson, Environmental records from coral skeletons: A decade of novel insights and innovation. *WIREs Clim. Change* **13**, 1–40 (2021).
51. N. J. De Winter *et al.*, Absolute seasonal temperature estimates from clumped isotopes in bivalve shells suggest warm and variable greenhouse climate. *Commun. Earth Environ.* **2**, 1–8 (2021).
52. A. M. Smith, "Palaeonenvironmental interpretations using bryozoans: A review" in *Marine Palaeoenvironmental Analysis from Fossils*, D. W. J. Bosence, P. A. Allison, Eds. (The Geological Society, Geological Society London, 1995), **vol. 83**, pp. 231–243.
53. C. Lombardi, S. Cocito, M. C. Gambi, P. D. Taylor, Morphological plasticity in a calcifying modular organism: Evidence from an *in situ* transplant experiment in a natural CO2 vent system. *R. Soc. Open Sci.* **2**, 1–12 (2015).
54. W. Kiessling, J. A. Smith, N. B. Raja, Improving the relevance of paleontology to climate change policy. *Proc. Natl. Acad. Sci. U.S.A.* **120**, 1–7 (2023).
55. A. D. Barnosky *et al.*, Merging paleobiology with conservation biology to guide the future of terrestrial ecosystems. *Science* **355**, 1–10 (2017).
56. M. E. Clapham, Conservation evidence from climate-related stressors in the deep-time marine fossil record. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 1–7 (2019).
57. W. Chaisson, P. N. Pearson, "Planktonic foraminifer biostratigraphy at site 925: Middle miocene-pleistocene" in *Proceedings of the Ocean Drilling Program, Scientific Results*, N. J. Shackleton, Ed. (1997), vol. 154, pp. 3–31.
58. J. P. Kennett, M. S. Srinivasan, *Neogene Planktonic Foraminifera. A phylogenetic atlas* (Hutchinson Ross Publishing Company, Stroudsburg, Pennsylvania, 1983).
59. I. Bailey *et al.*, An alternative suggestion for the Pliocene onset of major Northern Hemisphere glaciation based on the geochemical provenance of North Atlantic Ocean ice-rafted debris. *Quat. Sci. Rev.* **75**, 181–194 (2013).
60. J. F. Spray *et al.*, North Atlantic evidence for a unipolar icehouse climate state at the Eocene-Oligocene transition. *Paleoceanogr. Paleoclimatol.* **34**, 1124–1138 (2019).
61. W. B. Curry, N. J. Shackleton, C. Richter, "Expedition 154 Scientists, Site 925" in *Proceedings of the Ocean Drilling Program, Initial Reports* (1995), vol. 154.
62. R. H. Wilkens *et al.*, Revisiting the Ceara Rise, equatorial Atlantic Ocean: Isotope stratigraphy of ODP leg 154 from 0 to 5 Ma. *Clim. Past* **13**, 779–793 (2017).
63. B. S. Wade, P. N. Pearson, W. A. Berggren, H. Pälike, Review and revision of Cenozoic tropical planktonic foraminiferal biostratigraphy and calibration to the geomagnetic polarity and astronomical time scale. *Earth-Sci. Rev.* **104**, 111–142 (2011).
64. D. J. King, B. S. Wade, R. D. Liska, C. G. Miller, A review of the importance of the Caribbean region in Oligo-Miocene low latitude planktonic foraminiferal biostratigraphy and the implications for modern biogeochronological schemes. *Earth-Sci. Rev.* **202**, 1–27 (2020).
65. A. Brombacher, A. Searle-Barnes, W. Zhang, T. H. G. Ezard, Analysing planktonic foraminiferal growth in three dimensions with foram3D: An R package for automated trait measurements from CT scans. *J. Micropalaeontol.* **41**, 149–164 (2022).
66. Object Research Systems (ORS) Inc, *Dragonfly* (Montreal, Canada, 2022).
67. R. C. Team, *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2023).
68. G.-J.A. Brummer, C. Hemleben, M. Spindler, Planktonic foraminiferal ontogeny and new perspectives for micropalaeontology. *Nature* **319**, 50–52 (1986).
69. G.-J.A. Brummer, C. Hemleben, M. Spindler, Ontogeny of extant spinose planktonic foraminifera (Globigerinidae): A concept exemplified by *Globigerinoides sacculifer* (Brady) and *G. ruber* (d'Orbigny). *Mar. Micropaleontol.* **12**, 357–381 (1987).
70. A. G. Caromel, D. N. Schmidt, I. Fletcher, E. J. Rayfield, Morphological change during the ontogeny of the planktic foraminifera. *J. Micropalaeontol.* **35**, 2–19 (2016).
71. M. Apthorpe, Middle Jurassic (Bajocian) planktonic foraminifera from the northwest Australian margin. *J. Micropalaeontol.* **39**, 93–115 (2020).
72. S. Kendall, F. Gradstein, C. Jones, O. T. Lord, D. N. Schmidt, Ontogenetic disparity in early planktic foraminifers. *J. Micropalaeontol.* **39**, 27–39 (2020).
73. C. R. Poole, B. S. Wade, Systematic taxonomy of the *Trilobatus sacculifer* plexus and descendant *Globigerinoidesella fistulosa* (planktonic foraminifera). *J. Syst. Palaeontol.* **17**, 1989–2030 (2019).
74. R. Morard *et al.*, Genetic and morphological divergence in the warm-water planktonic foraminifera genus Globigerinoides. *PLoS One* **14**, 1–30 (2019).
75. D. W. Lea, T. A. Mashiotta, H. J. Spero, Controls on magnesium and strontium uptake in planktonic foraminifera determined by live culturing. *Geochim. Cosmochim. Acta* **63**, 2369–2379 (1999).
76. D. Nürnberg, J. Bijma, C. Hemleben, Assessing the reliability of magnesium in foraminiferal calcite as a proxy for water mass temperature. *Geochim. Cosmochim. Acta* **60**, 803–814 (1996).
77. K. P. Jochum *et al.*, High-resolution Mg/Ca measurements of foraminifer shells using femtosecond LA-ICP-MS for paleoclimate proxy development. *Geochem. Geophys. Geosyst.* **20**, 2053–2063 (2019).
78. K. P. Jochum *et al.*, Determination of reference values for NIST SRM 610–617 glasses following ISO guidelines. *Geostand. Geoanal. Res.* **35**, 397–429 (2011).
79. L. E. Kearns *et al.*, The influence of geochemical variation among *Globigerinoides ruber* individuals on paleoceanographic reconstructions. *Paleoceanogr. Paleoclimatol.* **38**, 1–20 (2023).
80. S. Sekimoto *et al.*, Neutron activation analysis of carbonate reference materials: Coral (JCp-1) and giant clam (JCt-1). *J. Radioanal. Nucl. Chem.* **322**, 1579–1583 (2019).
81. A. Searle-Barnes, J. A. Milton, C. D. Standish, G. L. Foster, T. H. G. Ezard, Laser ablation mass spectrometry blast through detection in R. *Rapid Commun. Mass Spectrom.* **37**, 1–8 (2023).
82. S. N. Wood, Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**, 3–36 (2011).
83. A. F. Zuur, E. N. Ieno, N. Walker, A. A. Saveliev, G. M. Smith, "Mixed effect models and extensions in ecology with R" in *Statistics for Biology and Health*, M. Gail, K. Krickeberg, J. M. Samet, A. Tsiatis, W. Wong, Eds. (Springer-Verlag, New York, NY, 2009).
84. A. Brombacher *et al.*, Data from 'Detecting environmentally dependent developmental plasticity in fossilised individuals'. Figshare. https://doi.org/10.6084/m9.figshare.27260970. Deposited 19 May 2025.

Paleoceanography and Paleoclimatology

RESEARCH ARTICLE
10.1029/2022PA004549

The Influence of Geochemical Variation Among *Globigerinoides ruber* Individuals on Paleoceanographic Reconstructions

L. E. Kearns[1,2], A. Searle-Barnes[1], G. L. Foster[1], J. A. Milton[1], C. D. Standish[1], and T. H. G. Ezard[1]

[1]School of Ocean and Earth Science, University of Southampton, National Oceanographic Centre Southampton, Southampton, UK, [2]Now at Jackson School of Geosciences, Institute for Geophysics, University of Texas at Austin, Austin, TX, USA

Key Points:

- *Globigerinoides ruber* shows large variability in Mg/Ca ratios that is partially explained by the presence of subspecies
- Within-specimen variation through ontogeny explains more variation in Mg/Ca ratios than differences among subspecies
- Subspecies differences in *G. ruber* geochemistry are driven by depth habitat differences

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

L. E. Kearns,
lorna.kearns@austin.utexas.edu

Citation:

Kearns, L. E., Searle-Barnes, A., Foster, G. L., Milton, J. A., Standish, C. D., & Ezard, T. H. G. (2023). The influence of geochemical variation among *Globigerinoides ruber* individuals on paleoceanographic reconstructions. *Paleoceanography and Paleoclimatology*, *38*, e2022PA004549. https://doi.org/10.1029/2022PA004549

Received 12 SEP 2022
Accepted 6 MAR 2023

Abstract Variation among individuals within species is a biological precondition for co-existence. Traditional geochemical analysis based on bulk averages facilitates rapid data gathering but necessarily means the loss of large amounts of potentially crucial information into variability within a given sample. As the sensitivity of geochemical analysis improves, it is now feasible to build sufficiently powerful datasets to investigate paleoclimatic variation at the level of individual specimens. Here, we investigate geochemical and morphological variation among the sensu stricto, sensu lato and sensu lato extreme subspecies of the workhorse extant planktic foraminifera *Globigerinoides ruber*. Our experimental design distinguishes between subspecies and intraspecific variability as well as the repeatability of laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS). We show that geochemical variability in Mg/Ca ratios is driven by differences in subspecies depth habitat and that ontogenetic trends in Mg/Ca ratios are evident in the final whorl, with the final chamber consistently showing depleted Mg/Ca. These ontogenetic trends are not driven by individual chamber or test size. The Mg/Ca value variance among individuals is ~100 times higher than the variance among repeated laser spot analyses of single chambers, directing laboratory protocols towards the need to sample ecologically and environmentally homogeneous samples. Our results emphasize that we can use LA-ICP-MS to quantify how individual variability aggregates to bulk results, and highlights that, with sufficient sample sizes, it is possible to reveal how intraspecific variability alters geochemical inference.

Plain Language Summary Planktic foraminifera are small marine life forms whose fossil remains are used to reconstruct past climates and ocean conditions. As with all living things, each individual foraminifer varies from one to the next. Understanding this variation among individuals is important for many branches of science. Here, we studied variation within one of the most used species of foraminifera for paleoceanographic reconstructions, *Globigerinoides ruber*. We show that the observed geochemical variation is organized within biologically meaningful groups and among life stages. We need the bigger sample sizes we use here to pick apart individual variation and its drivers as future reconstructions of past climates need to consider variation among individuals to get a fuller understanding of the co-evolution of life and the planet.

1. Introduction

Many paleoceanographic studies rely on the accurate identification of foraminifera as these organisms form a key archive of oceanic conditions. Despite genetic advances (André et al., 2014; Morard et al., 2015, 2019), foraminiferal taxonomy remains reliant on morphological comparisons. Paleoceanographic studies ubiquitously use geochemical proxies on morphologically identified species to infer the ambient environment (Zachos et al., 2001). While the environmental controls of geochemical variance such as temperature and salinity are well studied (Bemis et al., 2000; Bijma et al., 1999; Epstein et al., 1951; Pearson, 2012; Urey et al., 1951), variation as a result of biological factors including physiology and life history (Ezard et al., 2015) are commonly grouped in the catch-all term "vital effects" (Epstein et al., 1951; Urey et al., 1951). These "vital effects" have been increasingly well documented (Bemis et al., 1998, 2000; Birch et al., 2013; Hönisch et al., 2003; Spero et al., 1991; Spero & Williams, 1989) and understood in an evolutionary context (Edgar et al., 2017). Yet, many studies still assume that variability among individuals is statistical noise to be averaged away. To demonstrate that there is biological and ecological meaning in this noise, we designed experimental protocols to deconstruct intraspecific

(within-species) variability of laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) measurements into meaningful differences among subspecies, size classes, and repeated LA-ICP-MS measurements. Our goal is to ascertain the magnitude of variation among individuals relative to the analytical uncertainty of single-specimen LA-ICP-MS analyses, as well as amongst subspecies within the workhorse extant planktic foraminifera *Globigerinoides ruber* (*G. ruber*).

The dismissal of intraspecific variation is a common thread in ecology. The mean of a population is habitually used as a sufficient descriptor of the whole population contrary to mounting evidence that intraspecific variability impacts community stability through resource partitioning (Jung et al., 2010), competition (Bolnick et al., 2011; Hart et al., 2016), and abiotic tolerances (Bolnick et al., 2011). Applying a mean-field approximation to foraminifera might be sufficient for biological factors common among individuals such as seasonality resulting from synchronous reproduction (Bijma et al., 1990). For many other factors, however, the evidence for commonality among individuals is not robust due to the structure of variation among- and within-individuals.

Within-species variation can be apportioned into systemic and non-systemic variation. The latter encompasses random noise; there may be mechanistic reasons why two individuals express a different geochemical signature, but these reasons are often not detectable at the given scale of analysis and are prone to random variation during the measurement process. The systemic reasons could be continuous, such as geographical space (Darling & Wade, 2008) or vertical depth habitat (Norris, 2000) or discrete, such as the presence/absence of symbionts (Ezard et al., 2015; Spero & Lea, 1993). Subspecies, another source of systematic variation, form a supposedly discrete taxonomic rank below the species whose utility in evolutionary biology remains contentious (Phillimore & Owens, 2006). Subspecies' existence could potentially bias paleoceanographic studies in systemic ways if the subspecies are genetically, vertically, seasonally, or geographically distinct (Lazarus, 1983; Mayr, 1942; Norris, 2000; Seears et al., 2012). Co-existence of subspecies at a given location obviates traditional geographical criteria for subspecies delimitation (Mayr & Ashlock, 1991) and makes nuanced taxonomic identification increasingly important. Measuring, understanding, and accounting for systemic subspecies variability should be a priority to reduce errors and reveal potential bias in geochemical proxies (Antonarakou et al., 2015; Sadekov et al., 2008).

Technological advances have popularized high-resolution, single specimen measurements and intratest analysis, previously thought impossible (Emiliani, 1954). Single specimen analysis has revealed size dependent trends in stable isotopes (Kelly et al., 1997), unlocked the ability to circumvent post depositional diagenetic issues for paleotemperature reconstructions (Aze et al., 2014), reconstructed high-frequency climate signals (Ford et al., 2015; Sadekov et al., 2013), further constrained seasonal paleoclimate variability (Metcalfe et al., 2019), revealed the presence of intraindividual chamber difference in biomineralization-related Mg-banding (Eggins et al., 2003; Sadekov et al., 2008, 2009), and, through an innovative use of culturing experiments, begun to illuminate the physiological drivers of such banding (Fehrenbacher et al., 2017; Spero et al., 2015). Despite instrumental improvements, our understanding of intraspecific variation has not progressed substantially because the few studies that have investigated these issues often display conflicting results, in part due to smaller sample sizes as the new technology came onstream.

One of the most studied and contentious species in this field is *G. ruber*, one of the workhorses of paleoceanography. Despite numerous studies across all the major ocean basins (Antonarakou et al., 2015; Aurahs et al., 2011; Kawahata, 2005; Kuroyanagi et al., 2008; Kuroyanagi & Kawahata, 2004; Lynch-stieglitz et al., 2015; Mohtadi et al., 2009; Naik, 2016; Numberger et al., 2009; Sadekov et al., 2008; Steinke et al., 2010; Thirumalai et al., 2014; Wang, 2000), the degree of intraspecific variation in geochemical and morphological space, and its consequent impact of variation on paleoceanographic reconstructions, remains ambiguous. The main issues include insufficient sample size (Löwemark et al., 2005; Lynch-stieglitz et al., 2015; Wang, 2000) and a focus on bulk morphotype analysis that averages away individual variability (Antonarakou et al., 2015; Löwemark et al., 2005; Lynch-stieglitz et al., 2015; Mohtadi et al., 2009; Wang, 2000). One potential reason for this ambiguity is the existence of subspecies variation. *G. ruber* has between three (Parker, 1962) and eight (Robbins & Healy-Williams, 1991) named morphotypes along a morphocline (Kontakiotis et al., 2017; Robbins & Healy-Williams, 1991), with the two most abundant and morphologically distinct (sensu stricto (SS), sensu lato (SL)) commonly separated. The more compressed sensu lato extreme (SLE) is additionally separated in some other studies (Antonarakou et al., 2015; Carter et al., 2017; Kontakiotis et al., 2017; Löwemark et al., 2005; Lynch-stieglitz et al., 2015; Mohtadi et al., 2009; Wang, 2000). If geochemical variation is systemic across these

subspecies because they inhabit different depth habitats, for example, then any cross-study comparisons may not be starting from common units of reference.

Understanding whether variation is structured systemically requires individual-based data. Whilst Groeneveld et al. (2019) studied 451 individuals in 4 species, the largest isotope study of this type to date, their investigations of *G. ruber* analyzed these individuals in pairs. The largest *G. ruber* single-specimen LA-ICP-MS investigation to date analyzed 60 individuals (Naik, 2016). Using LA-ICP-MS on 264 G. *ruber* individuals we generated 1,860 Mg/Ca chamber level measurements with statistical replication at chamber, individual and spot level to tease apart potential ecologically driven variability. We demonstrate that: (a) intraspecific variability in geochemical space is structured around subspecies classification, but (b) within-specimen variation through ontogeny explains more variation in Mg/Ca ratios than differences among subspecies; (c) there is a no detectable impact of specimen size on Mg/Ca ratios within restricted size fractions; and (d) Mg/Ca ratios are a reliable and repeatable measurement when determined using LA-ICP-MS.

2. Material and Methods

2.1. Material and Regional Setting

This study uses material collected during the Paleogene GLObal Warming events, "GLOW" cruise offshore Tanzania in the Western Indian Ocean (Kroon and the Shipboard Scientific Party, 2010). All sites on the cruise collected material that sat well above the carbonate compensation depth [CCD; 3,500–4,500 m (Bickert, 2009)] and lysocline [3,330 m (Belyaeva & Burmistrova, 1984; Ivanova, 2009]; and was composed of more than 30% clay leading to excellent well preserved planktic foraminifera (Kroon and the Shipboard Scientific Party, 2010). For this study, we focus on box core material collected at GLOW station 8 (from here on referred to as GLOW 8 (9 21' 25.20" S; 40 35' 27.60" °E) with a box corer with a diameter of 30 cm and height of 55 cm. GLOW 8 was located in the Southern Seagap, a topographic high between the Davis ridge and the Tanzanian coastline at a depth of 2,420 m. The box corer penetrated to a depth of 47 cm; for this study, we use the top 1 cm of the box core sample, which was separated whilst onboard (Kroon and the Shipboard Scientific Party, 2010). The estimated sedimentation rate in this area is ∼2 cm/kyr (Kroon and the Shipboard Scientific Party, 2010), which, alongside the presence of living benthic foraminifera at the sediment water interface in other cores in the area (Birch et al., 2013), implying we have Late Holocene planktic foraminifera.

The Western Indian Ocean is influenced by the Northeast Madagascan Current (NEMC) and movement of the Inter Tropical Convergence Zone (ITCZ). In Northern Tanzania the seasonal shifts of the ITCZ creates two monsoonal periods with heavy, prolonged rains from March to May (south east monsoon) and shorter rains between October and December (McClanahan, 1988). The coastal waters of North East Africa experience different hydrographic changes through the seasons that lead to different ecosystems from North to South (McClanahan, 1988). North of 4°S, ecosystems benefit from cooler, nutrient-rich waters as a result of seasonal reversals in the wind and current direction and subsequent variability in thermocline depth and nutrient availability resulting in high planktic productivity (McClanahan, 1988). In contrast, the GLOW 8 material used in this study sits at ∼9°S where waters are warm and low in nutrients resulting in a predominance of coral reefs and high benthic productivity (McClanahan, 1988). Today the study area consists of warm surface waters with a max seasonal surface temperature (SST) variation of 5°C (August (∼25–26°C); February (30°C) (Birch et al., 2013; Damassa et al., 2006; McClanahan, 1988)) that remains stratified all year round (Birch et al., 2013) with small variations in surface mixed layer (SML) salinity (<0.5PSU) throughout the year (Birch et al., 2013; Damassa et al., 2006; McClanahan, 1988). Full conductivity, temperature, and depth (CTD) profiles collected at GLOW 5 (8 54' 6.01" °S; 41 29' 42.25" °E) ∼111 km southwest and GLOW 2 (10 54' 6.01" °S; 41 29' 42.25" °E) ∼198 km southeast of GLOW 8 show that the thermocline sits at ∼40 m with the turbidity maximum at ∼137 m indicating a maximum depth for symbiont hosting foraminifera (Supplementary material of Birch et al., 2013). In addition, CTD data show that salinity varies between ∼34.8 and ∼35.4 PSU in the upper 200 m of the two GLOW sites.

2.2. Morphological Analysis

The material was sieved to 250–355 μm size fraction, and the first 150 individuals of each *G. ruber* morphotype (Figure 1a) were picked for analyses and given an individual ID including a number 1–150 and letters referring to

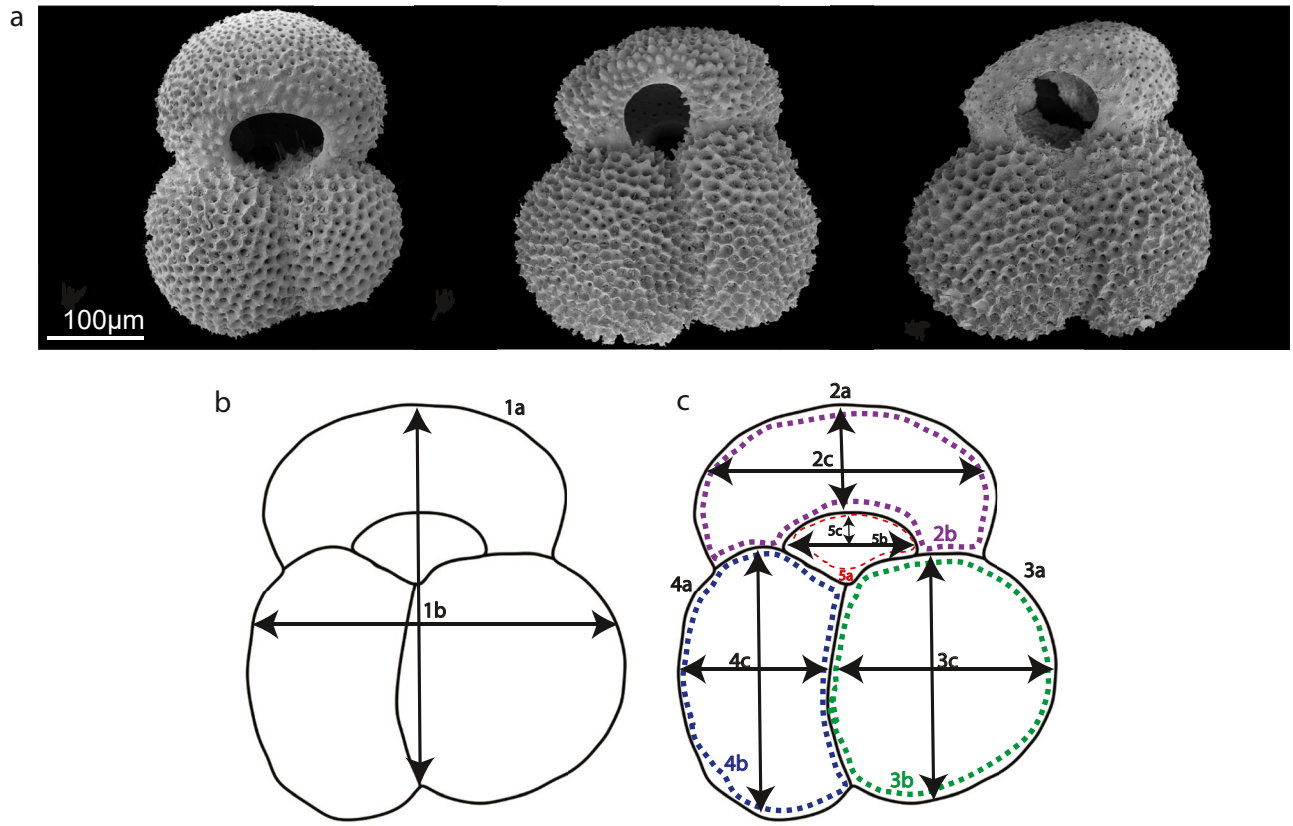


Figure 1. (a) Scanning electron microscope images of the morphotypes of *Globigerinoides ruber* used in this study. Left to right: SS, SL and SLE. (b) Pre-selected traits measured automatically: 1a-test area and 1b-test aspect ratio. (c) Ontogenetic traits measured manually: 2a-final chamber area, 2b-final chamber perimeter, 2c-final chamber aspect ratio, 3a-penultimate chamber area, 3b-penultimate chamber perimeter, 3c-penultimate chamber aspect ratio, 4a-antepenultimate chamber area, 4b-antepenultimate chamber perimeter, 4c-antepenultimate aspect ratio, 5a-aperture perimeter, 5b-aperture width, and 5c-aperture height.

morphotype (sensu stricto: SS, sensu lato: SL and sensu lato extreme: SLE). Due to the observed morphological variation in *G. ruber* (Kontakiotis et al., 2017; Robbins & Healy-Williams, 1991), we only picked individuals, which were the most morphologically distinct in the final chamber as this is where the most visual difference was evident (Figure 1a). Any intermediate morphologies were ignored. The specimens were mounted on glass slides in groups of 20 and orientated with the aperture facing upwards (Brombacher et al., 2017). Images of the final whorl of each sample were captured using an Infinity 3 Lumenera camera mounted on an Olympus SZX10 microscope, illuminated from above (Brombacher et al., 2017). Pre-selected traits (Figure 1b) were measured using automated image analysis macro within Image Pro 9.1 Premier software (Table S1 in Supporting Information S1). Ontogenetic traits (Figure 1c), beyond the scope of the imaging macro, were measured manually in the same software (Table S1 in Supporting Information S1).

2.3. Geochemical Analysis

2.3.1. LA-ICP-MS Trace Element Analysis

For LA-ICP-MS, 100 individuals from each subspecies already measured morphologically were selected. For each subspecies 100 unique random numbers, relating to the previously assigned morphological ID, were sampled without replacement using a random number function in the R environment (Version 4.0.3; R Core Team, 2020). The corresponding individuals were then removed from the slides and placed into separate vials. Individuals were then cleaned by ultrasonification in methanol for 5–6 s followed by two washes in Milli-Q water (Eggins et al., 2003). The samples were then dried overnight in a 50°C oven before being remounted as described previously.

Trace element (TE)/Ca ratios were analyzed using a New Wave UP193 ArF laser ablation system coupled to a Thermo Fisher Scientific X-Series II ICP mass spectrometer at the University of Southampton. During analysis

the isotopes ^{24}Mg, ^{43}Ca, ^{44}Ca, ^{55}Mn, ^{88}Sr, ^{66}Zn, and ^{137}Ba were measured as a time resolved acquisition; only Ca and Mg data are presented here in the main text whilst the other TE/Ca can be found in Table S2 in Supporting Information S1. Individual chambers in the final whorl were ablated in triplicate with each time resolved analysis set to 60 s; wall penetration was achieved prior to the end of each analysis to ensure that the full wall thickness was sampled.

Prior to an analysis session, the mass spectrometer and laser system was tuned for optimal sensitivity, stability, and low oxide formation. Each laser analysis used a spot diameter of 30 μm, a repetition rate of 5 Hz, and fluence 0.3 J/cm^2. Using the laser software, sample and standard locations were mapped and recorded to form a programmed sequence. Ten replicates each of NIST 612 and 610 reference glasses were analyzed in batches throughout each session, before and after the analysis of 9–11 individuals (about 80–100 spots). To assess accuracy of the analytical method, the Japanese Geological Survey's JCp-1 (Porites sp. Coral reference material) was analyzed in a parallel session. Each time resolved analysis was background subtracted using the gas blank measured immediately before each sample or standard. Sample and standard data were internally corrected using ^{43}Ca wt.%. This correction was performed on individual time slices of approximately 0.1 s/slice. The following concentration values were used for the internal correction: NIST 612: 8.5 wt.%, NIST 610: 8.15 wt.%, JCp-1: 37.5 wt.% (see Jochum et al., 2019; Jochum et al., 2011). Following blank and internal correction, a calibration based on averaged measurements of NIST 612 and NIST 610 was determined using elemental concentration values from Jochum et al. (2011) and Jochum et al. (2019). We also processed JCp-1 as an unknown using the same calibration curve as the foraminifera to back calculate its concentration and compare to its published value (Jochum et al., 2019). JCp-1 was measured as 836 ± 124 ppm ^{24}Mg (vs. 867 ± 23 ppm by Jochum et al. (2019)) and as 38.1 ± 3.8% (vs. 37.5 ± 2.4% by Sekimoto et al. (2019)). Our NIST glass calibrations thus also predict trace elements within carbonate material within specified uncertainty limits. Influential outlier values were detected using Cook's Distance and anomalous values beyond a threshold were removed (0.0001 for standards; 0.2 for foraminiferas). For each foraminifer depth profile, the processing algorithm identifies when the laser fully penetrates the chamber wall. The raw ^{43}Ca signal is smoothed using a moving average of 5 data points. The rate of change of the smoothed signal with respect to a time step of 20 observations (7.44 s) was calculated and the largest signal drop identified as the point when the laser had ablated through the calcite wall and thus no longer recorded the geochemical signal (Figure S2 in Supporting Information S1). The function then removes the rows of data that occur between the final laser pulse with high signal intensity and the maximum rate of change (Figure S2 in Supporting Information S1). The corresponding time stamp for this final laser pulse is returned and any rows of data after this are removed from further consideration. A short laser profile with a low number of data points can be caused by a false positive detection, such as where the laser has encountered a contaminant, that is, clay or something else that is not calcite that causes the ^{43}Ca signal to decrease rapidly and trigger the end-point detection algorithm. Such profiles were flagged in the data reduction procedure and removed. The mean number of data points per profile was 20 (7.4 s).

Trace element (TE) to calcium ratios, including Mg/Ca ratio (mmol/mol), were calculated for each time slice and a median taken. Three locations per chamber were analyzed. To manually screen for contaminants, we looked at other indicative TE/Ca ratios against Mg/Ca. Nine profiles were removed using Mn/Ca and a cut-off of 1 mmol/mol (Figure S3a in Supporting Information S1). For further contamination control, we also remove 29 (~1%) additional spots with an unrealistic mean Mg/Ca >10 mmol/mol. The mean spot-average across all foram analyses was 4.36 mmol/mol with a back-transformed 95% confidence interval of (4.19, 4.53). In total, we present results from 1,860 analyses of 263 individuals, which represents 5,323 μm^3 of material ablated, which is the largest study on *G. ruber* to date.

The exponential relationship between Mg/Ca ratios and temperature (Lea et al., 1999; Rosenthal et al., 1997) has resulted in many Mg/Ca based temperature calibrations (e.g., Anand et al., 2003; Bolton et al., 2011; Dekens et al., 2002; Elderfield & Ganssen, 2000; Tierney et al., 2019). To allow for comparisons between other studies we use Equation 1 where T is temperature, b = 0.38 and m = 0.09 based on core top calibrations using *G. ruber* (Dekens et al., 2002; Lea et al., 2000). Mg/Ca data and temperature conversion can be found in and Table S2 in Supporting Information S1.

$$\mathrm{Mg/Ca} = b\exp(\mathrm{m}(T)) \quad (1)$$

2.3.2. Stable Isotope Analysis

Stable isotope measurements were obtained from 89 individuals (SS n = 35, SL n = 23, SLE n = 31; Table S3 in Supporting Information S1), previously measured for trace elements by LA-ICP-MS, using a Thermo

Fisher Scientific Kiel IV carbonate device coupled to a MAT253 stable isotope ratio mass spectrometer at the SEAPORT Stable Isotope Laboratory, University of Southampton. Samples were placed into individual vials and measured against the global reference standards NBS19 and NBS18 as well as an in-house quality control standard (GS1). Long-term analytical precision (1σ) was based on the repeat analysis of GS1 and estimated as ±0.09‰ for $\delta^{18}O$ and ±0.05‰ for $\delta^{13}C$. All results were standardized to Vienna Pee Dee Belemnite (VPDB) using a two-point calibration between NBS19 and NBS18. Calcification temperature was calculated using the general calibration of (Erez & Luz, 1983):

$$T°C = 16.998 - 4.52\,(\delta c - \delta w) + 0.028(\delta c - \delta w)^2 \quad (2)$$

with a δw of 0.47‰ VSMOW (Birch et al., 2013) converted into VPDB using the correction of −0.22‰ (Bemis et al., 1998; Friedman & O'Neil, 1977; Pearson, 2012). Although this conversion was calibrated using *Trilobatus sacculifer,* we employ it here to allow direct comparisons to other studies based on this material.

2.4. Statistical Analysis

The mean spot-averaged Mg/Ca ratios in Section 2.3.1 exaggerate the precision achievable because each analysis is not independent due to nested pseudoreplication at spot, individual and subspecies levels. To remove this pseudoreplication, understand the influence of subspecies and morphological variation on Mg/Ca ratios and ascertain the relative amounts of variance explained by spot, individual and subspecies levels, we used generalized linear mixed effect models implemented in the lme4 package (Bates et al., 2015). Mixed effects models comprise random and fixed effects. Fixed effects are experimentally determined and of direct interest in hypothesis testing (Bolker et al., 2009) for patterns of variation common to all experimental units; random effects are selected from a larger population but thought of as uncontrollable "nuisance" parameters systematically obscuring the signal held by the fixed effects (Bolker et al., 2009; Gillies et al., 2006). The "vital effects" that alter each individual's geochemical composition are examples of random effects, while an overall temperature gradient would be a fixed effect. Variation among individuals means we cannot assume each ablation represents an independent sample: as an example, the Mg/Ca ratio difference from penultimate to final chamber is likely to be more similar in the same individual than from one individual to another. Our experimental design is stratified random sampling because we targeted sufficient numbers of individuals in each subspecies. We do therefore assume that individuals were sampled independently within their subspecies. In all models, individual test ID, repeat laser spot number (spot) and analysis batch (batch) were classified as random effects (Bates, 2005) because, while we know that individuals, batches and spots could differ amongst each other, we do not have a systemic hypothesis for how they differ; chamber and morphotypes were categorized as fixed effects and Mg/Ca ratio as the dependent response variable because we expect certain relationships with increasing chamber size through ontogeny and because we anticipate that subspecies-specific offsets are possible. We consider subspecies a fixed effect following the recommendations of Bolker et al. (2009) given low numbers of categorical levels. Our different statistical models were compared using Analysis of Variance amongst competing models to test model fit assuming a Gamma error distribution for the generalized linear model (inverse link function to transform the mean of the data; variance increases as a quadratic function of the mean). The best fit model was decided by likelihood ratio tests and the Akaike Information Criterion (AIC, Burnham & Anderson, 2002), which summarizes model fit as a compromise between variance explained and parameters used.

All data processing and statistical analysis was carried out in the R environment (version 4.0.3; R Core Team, 2020). Scripts are provided as Supplementary Material.

3. Results

3.1. Morphological Analysis

To test the null hypothesis that the morphotypes of *G. ruber* are morphologically indistinguishable, a principal component analysis (PCA) was conducted using all traits in Figures 1b and 1c. To reduce the number of components, a Horn's Parallel analysis was conducted (Peres-Neto et al., 2005) using the "paran" package (Dinno, 2018). This analysis indicated the first three principal components explaining ∼71% of cumulative variance should be retained (Figure 2, Table S4 in Supporting Information S1). When these components are plotted against each other, the SS and SL subspecies occupy a homogenous morphospace (Figure 2). While PC1 does not separate

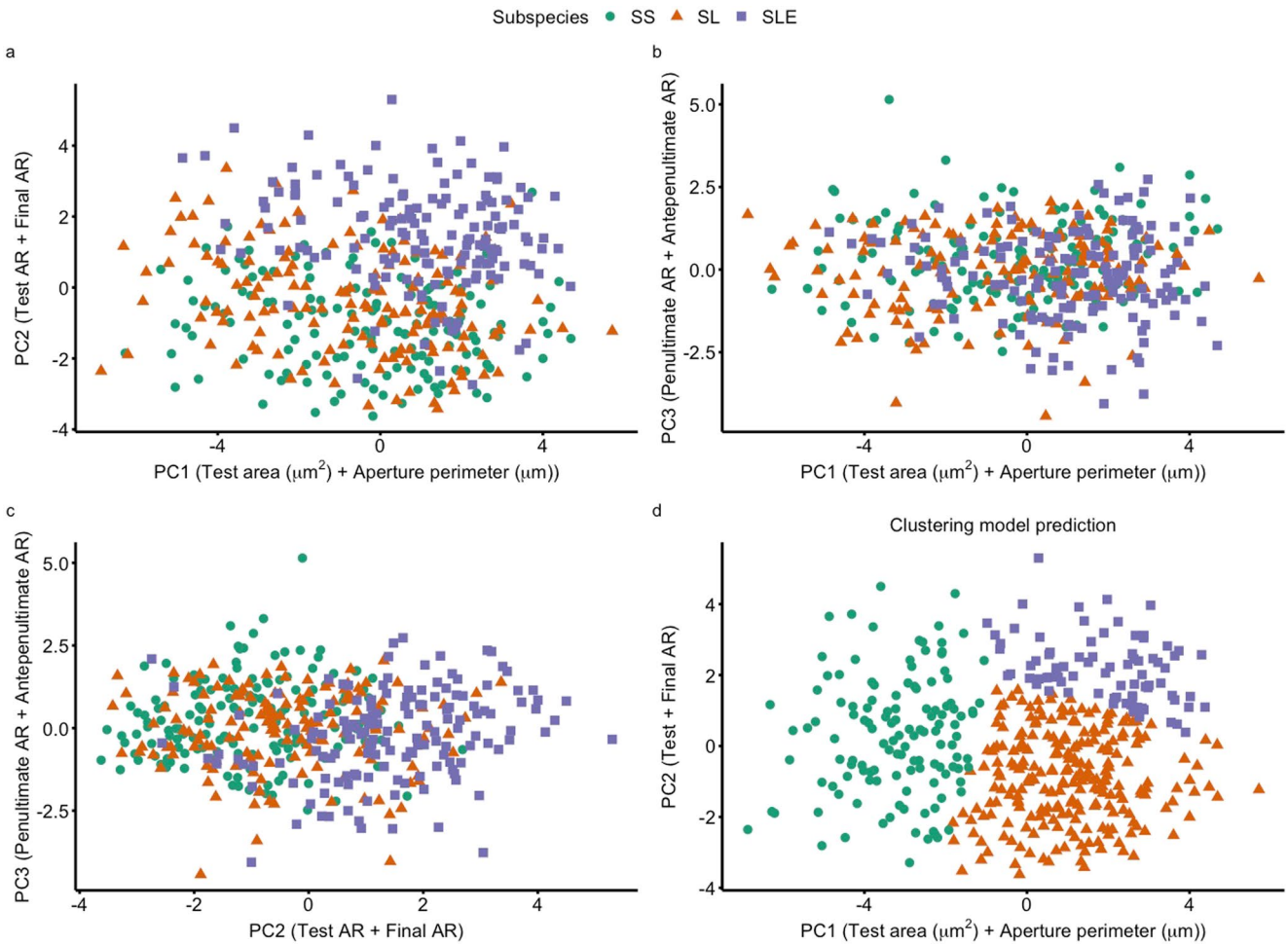


Figure 2. Scatter plot of PCA results showing the first three principal components that explain 71% of the cumulative variance. Principal component 1 (PC1) is loaded highly by test area and perimeter aperture; principal component 2 (PC2) is loaded highly by final chamber aspect ratio and test aspect ratio; and principal component 3 (PC3) is loaded highly by penultimate and antepenultimate aspect ratios. (a) PC1 versus PC2, (b) PC1 versus PC3, (c) PC2 versus PC3, (d) PC1 versus PC2 with symbols and colors representing subspecies as identified by our clustering model when forced to identify three groups. Note the differences between (a) and (d). Loadings of principal components can be found in Table S5 in Supporting Information S1. Predicted clusters can be found in Table S7 in Supporting Information S1.

any subspecies effectively (Figure 2, Figures 3a and 3b), SLE shows a degree of offset in principal component 2 (Figure 2), which is loaded highly by the aspect ratio of the final chamber and the whole test (Figures 3c and 3d, Table S5 in Supporting Information S1).

Following the PCA, a cluster analysis was conducted using the package "mclust" (Scrucca et al., 2016) on the three retained principal components. This package tests a finite number of models to determine the most supported statistical model among those considered and identifies the optimal number, size, shape, and orientation of clusters needed to explain the data using Schwarz's Bayesian information criterion (BIC) (Schwarz, 1978). The BIC considers the statistical fit of a model as well as the number of parameters the model has to determine the posterior model probability (Wintle et al., 2003); in the "mclust" package, a larger BIC value indicates a better model fit (Fraley & Raftery, 1996). To reject our null hypothesis of no sub-species variability, the "mclust" analysis must indicate the data are better represented by more than one cluster and thus different morphotypes should occupy statistically distinct clusters within the morphospace. Clustering analysis shows that the morphospace described by PC1 and PC2 is best represented by two clusters (Figure S4–S5 in Supporting Information S1), though these do not correspond tightly to our subspecies classification (Table S6 in Supporting Information S1). When the method is forced to identify three groups (Figure 2d) subspecies are split across all three groups (Table S7 in Supporting Information S1) and do not correspond to morphologically meaningful units. Therefore, we reject the null hypothesis of no morphological variability in *G. ruber* but conclude the variability identified is not taxonomically aligned to established subspecies definitions.

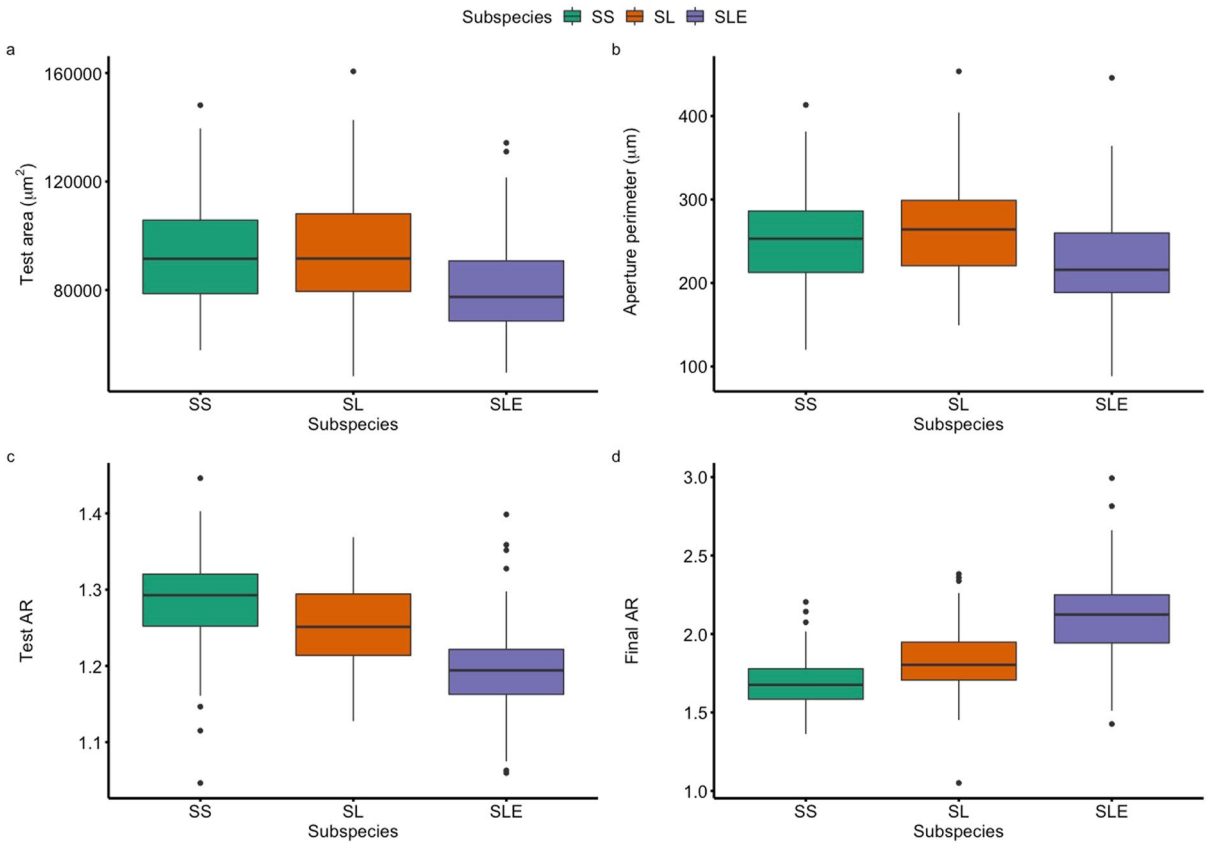


Figure 3. The highest loading traits from PC1 and PC2 (Figure 2a) shown as boxplots. The box represents the interquartile range whilst the whiskers show 1.5*interquartile range. Black circles represent outliers exceeding the 1.5*interquartile range. Black line represents the mean.

3.2. Trace Element Subspecies Variability

The median Mg/Ca ratios (± Median Absolute Deviation, MAD) for *G. ruber* show substantive interspecific differences (SS = 4.74 ± 1.28 mmol/mol, SL = 4.20 ± 1.37 mmol/mol and SLE = 3.64 mmol/mol ± 1.24 mmol/mol). To investigate subspecies trace element variability, we conducted a one-way analysis of variance (ANOVA) on the Mg/Ca dataset that detectable differences amongst subspecies ($F_{2,1857} = 112.9$, $p < 0.01$). A subsequent post-hoc TUKEY HSD test revealed a detectable significant ($p < 0.01$) differences in the mean Mg/Ca ratios between all subspecies. The SL and SLE morphotypes are depleted by 0.48 mmol/mol and 1.11 mmol/mol, respectively, when compared to SS. In addition, SLE is depleted in Mg/Ca ratios compared to SL by 0.63 mmol/mol.

To investigate factors driving interindividual Mg/Ca separation in *G. ruber*, generalized linear mixed effect models were constructed with a gamma link function to account for the positively skewed distribution of the response variable Mg/Ca. The best fitting model assumed that the relationship between test size and Mg/Ca was constant across subspecies whilst the Mg/Ca relationship with chamber position was variable among subspecies. We constructed two sets of models; one set used test size (Figure 1b) to investigate whole test impacts on Mg/Ca ratios whilst the other set used chamber size (chamber area labeled 2a, 3a, 4a in Figure 1c) to investigate ontogenetic impacts on Mg/Ca ratios.

3.2.1. Modeled Drivers of Mg/Ca

Though our morphological analysis showed that morphological traits, such as test area, did not correspond to subspecies (Section 3.1), test size is known to be a controlling variable in stable isotope values but is less well understood as a potential driver of Mg/Ca ratios. To understand whether test size is a driving variable of Mg/Ca ratios we built separate models with test size (test area) and chamber size (chamber area) effects. The best models, chosen through AIC, retained both whole test and chamber size effects. Results are reported here with a coefficient (β), standard error (s.e.), t statistic (t) and p value (p) associated with these coefficients; full model outputs

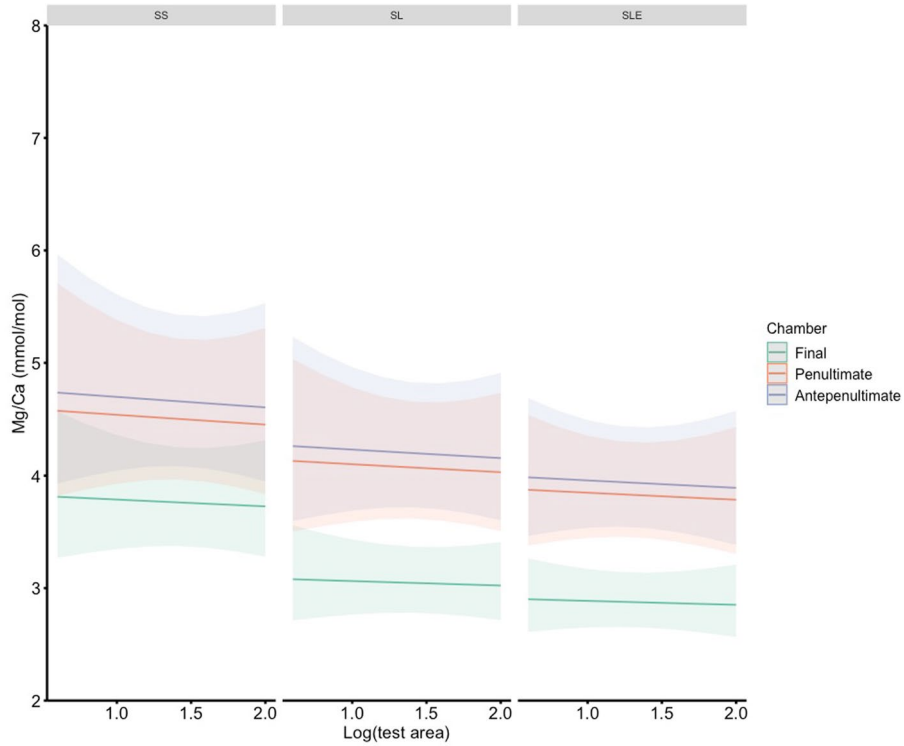


Figure 4. Model predictions of the relationship between Mg/Ca, Subspecies classification (SS = Sensu stricto, SL = Sensu lato, and SLE = Sensu lato extreme) and test size separated by chamber position represented by the color of dot/line with 95% confidence intervals. Panel (a) presents predictions against log transformed test size separated by sub-species and panel (b) separated by mean-centered log test size. Note particularly how test size has no relationship with Mg/Ca.

are available in Supporting Information S1. The generalized linear mixed model found significant evidence that chamber position (final, antepenultimate and penultimate) and subspecies impacted spot-averaged Mg/Ca ratios (Table S8 in Supporting Information S1). Whilst the best fitting models retained both test and chamber size, we find that test size and chamber size are not statistically significant drivers of Mg/Ca ratios over the range in size examined here (Figure 4, Table S8–S9 in Supporting Information S1). Subsequent discussions of results will focus on the model including test area as this was the highest loading morphological trait from principal component analysis (Figure 2).

We detected a strong impact of chamber position where the Mg/Ca ratios differed between the final and penultimate chamber ($\beta = -0.044$, s.e. = 0.004, $t = -9.995$, $p < 0.001$; the coefficients is on the scale of the inverse link function, hence a negative value means a positive relationship through the untransformed data) and between the final and antepenultimate chamber ($\beta = -0.051$, s.e. = 0.004, t = −11.668, $p < 0.001$), that is, a mean enrichment of 0.044 mmol/mol and 0.051 mmol/mol, respectively, in the penultimate and antepenultimate chamber compared to the final chamber (Figure 4, Table S8 in Supporting Information S1). This enrichment differs amongst subspecies such that depletion in the penultimate and antepenultimate chamber is stronger in SL than SS (penultimate: $\beta = -0.039$, s.e. = 0.007, $t = -5.805$, $p < 0.001$, antepenultimate: $\beta = -0.039$, s.e. = 0.007, $t = -5.571$, $p < 0.001$; in Supporting Information S1). This pattern of depletion is stronger still in SLE (penultimate: $\beta = -0.043$, s.e. = 0.007, $t = -5.916$, $p < 0.001$, antepenultimate: $\beta = -0.042$, s.e. = 0.007, $t = -5.953$, $p < 0.001$; in Supporting Information S1). This interaction supports the conclusion that chamber position is differentially influential amongst the subspecies in driving Mg/Ca trends. These trends are not consistent amongst subspecies and are most visible when comparing at chamber resolution rather than the whole specimen (Figure 4).

These interdependent relationships among chamber position and subspecies are present after controlling for so-called "random effects." We included individual test ID, Batch, and Spot number as random effects to test the repeatability of the LA-ICP-MS method on repeated measurements of the same chamber (Spot) in the same individual (test ID) across 2 days (Batch). Compared to residual standard deviation of 0.049, the variance explained

by individual, batch and spot were 0.001, 0.00005, and 0.00001, respectively (Figure S6 in Supporting Information S1). This means that, on top of the systemic "fixed effect" variation explained in the previous paragraph, 100 times more variation (0.001/0.00001) can be explained by intraspecific variability among individuals than by variability among repeated laser spots and 20 times more variation (0.001/0.00005) can be explained by intraspecific variability among individuals than by batches run on different days (Figure S6 in Supporting Information S1). These random effect variances are a formal treatment of the variance disaggregation, but the rank orders match raw calculations: Spot MAD is the smallest variation (0.46 mmol/mol), followed by Batch MAD (0.56 mmol/mol) then individual MAD (1.04 mmol/mol). We therefore conclude that LA-ICP-MS is a highly repeatable technique for extracting trace element signatures given that spot-average Mg/Ca ratios do not vary substantially within a given chamber [Figure S7 in Supporting Information S1; (i.e., Bolton et al., 2011; Sadekov et al., 2008)].

3.2.2. What Is the Optimum Number of Individuals Needed to Separate Subspecies in LA-ICP-MS Mg/Ca Analysis?

To determine the optimum number of individuals needed to detect subspecies differences in Mg/Ca, we conducted a rarefaction subsampling experiment using the best supported model from the previous subsection, and a simplified model without the subspecies effect. The difference in AIC scores between these two models indicates the improvement in model fit by considering systemic variation amongst subspecies. Each subspecies was subsampled at random in increments of 5 up to a maximum of 80. The process was repeated 50 times to create 800 model comparisons in total. We either record the ΔAIC values for the two models with and without subspecies, or N/A if a model failed to converge due to insufficient sampling coverage. As subsample size increased, so too did the ΔAIC (Figure 5a) while the percentage of models failing to converge decreased (Figure 5b). Once ∼55 individuals from each subspecies were sampled (Figure 3b), the ΔAIC was consistently sufficiently large to infer statistical support for systemic subspecies variation with most models converging. This finding illustrates that a focus for paleoceanography and paleoclimate research must be on increasing sample sizes for LA-ICP-MS work to achieve representative sample sizes as used here. Representative sample sizes will vary between species and system investigated, particularly as a function of the homogeneity of the material being sampled.

3.3. Stable Isotope Variability

Stable isotope analysis was conducted on a subset of 89 individual specimens that had previously undergone LA-ICP-MS analysis using the methods described in Section 2.3. In contrast to trace element analysis, no interindividual differences in carbon or oxygen isotopes were detected (ANOVA: Carbon $F(2,86) = 1.403$, $p = 0.252$, Oxygen $F(2,86) = 1.292$, $p = 0.280$) (Figures 6a and 6b). A conversion of oxygen isotope values to temperature using Equation (2) shows no detectable difference ($F(2,86) = 1.305$, $p = 0.277$) between the mean of SS compared to SL and SLE (SS = 24.35°C, SL = 23.25°C, SLE = 23.20°C). The range of calculated temperatures within each subspecies is however large, with all subspecies showing ranges of 9–12.5°C between the highest and lowest calculated temperatures. The relationship between size and stable isotopes differs amongst subspecies (Figures 6c and 6d, Table S10–11 in Supporting Information S1). SS and SLE show large variation with size in both oxygen and carbon; sensu lato shows a positive relationship, clearest in carbon isotopes, between stable isotopes and test area (Figures 6c and 6d).

3.4. Stable Isotope Versus Mg/Ca

In Figure 7, we plot $\delta^{18}O$ and $\delta^{13}C$ values against median Mg/Ca ratios of all spots per individual which, despite final chamber depletion, is a good representative of past ocean conditions (Fehrenbacher et al., 2020; Rustic et al., 2021). There is no detectable correlation between stable isotopes and Mg/Ca ratios (Figure 7; Table S12–S17 in Supporting Information S1) in all subspecies. In oxygen isotopes, we observe that as $\delta^{18}O$ values become more negative Mg/Ca ratios increase in SS and SL which would be expected (Figure 7a); however, this relationship is much weaker in SLE. In $\delta^{13}C$ values there is no observable pattern between $\delta^{13}C$ and Mg/Ca ratios, most likely because of a narrow range of $\delta^{13}C$ values.

4. Discussion

We show substantial variability in *G. ruber* Mg/Ca ratios that is partially explained by the presence of subspecies and differential Mg/Ca signatures through the ontogeny of those subspecies. We find no clear relationship between the morphological traits measured in this study and subspecies identification used but do detect

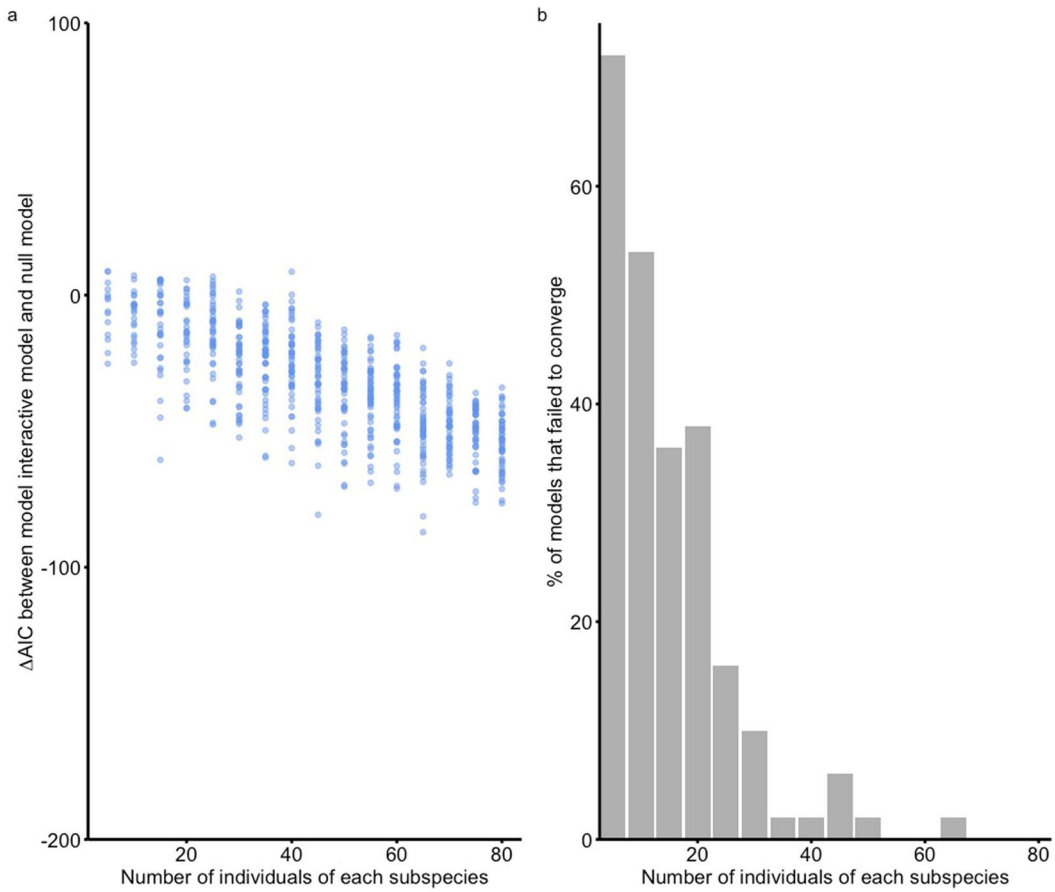


Figure 5. Results from rarefaction subsampling experiment. (a) The change in AIC between the fully interactive model which included subspecies as a fixed effect and the null model which did not include subspecies (in Supporting Information S1). Each subspecies sample size underwent 50 iterations. As the number of individuals from each subspecies increases, so does the support for the interactive model (larger negative values indicating larger difference in support between the models). (b) The percentage of models that failed to converge due to insufficient sample size and/or anomalously high values of Mg/Ca.

geochemical variation aligned to subspecies classification. The final chamber was consistently depleted in Mg/Ca ratios compared to the other two chambers in the final whorl. Intraspecific variation through ontogeny explains more variation in Mg/Ca ratios than differences among subspecies, which implies we need large sample sizes (>55) when performing LA-ICP-MS to have a realistic expectation of detecting such differences. Variation among individuals explains two orders of magnitude more variability in Mg/Ca ratios than variation amongst shots, implying that Mg/Ca ratios are a reliable and repeatable measurement when determined using LA-ICP-MS.

4.1. Morphological Variability in *G. ruber* Subspecies

Inter- and intraspecific morphological variation is abundant amongst the three *G. ruber* subspecies (Figures 1–3), yet the general taxonomic importance of such sub-species variability is disputed (Phillimore & Owens, 2006). The number of distinct morphological groups identified in *G. ruber* ranges from three (Parker, 1962) to eight (Robbins & Healy-Williams, 1991) along a morphocline. In this study we found that morphological variation does exist (Figures 1 and 3) but does not correspond to taxonomically informative units (Figure 2, Table S4 in Supporting Information S1), similar to other studies that used a greater range of morphological measurements (Numberger et al., 2009). We observed some separation in the aspect ratio of the final chamber and test aspect ratio that separated a large proportion of the SLE morphotype (PC2: Figure 2, 3c, and 3d). The separation observed due to test and final chamber aspect ratio, which corresponds to test and chamber compression, respectively, are commonly used morphological differences for morphotype identification (Aurahs et al., 2011; Carter

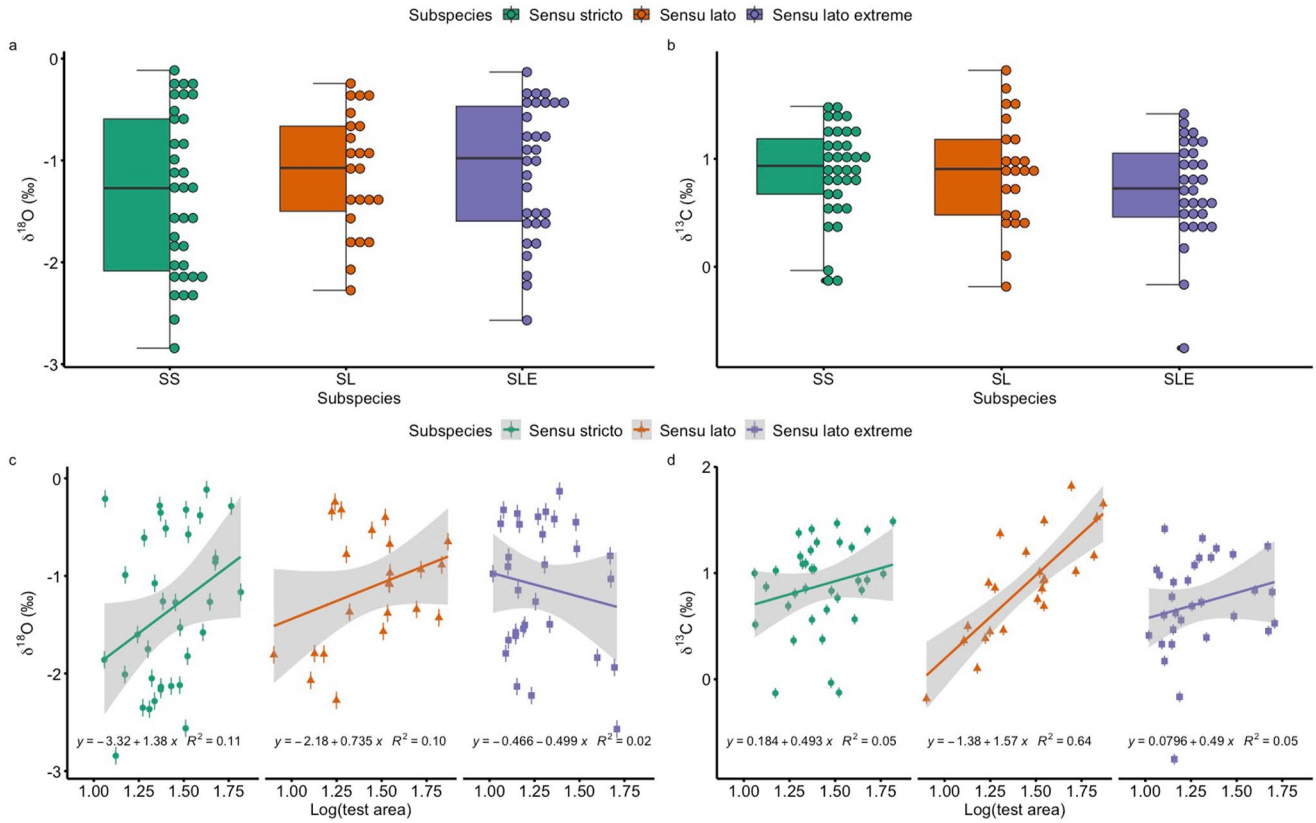


Figure 6. Stable oxygen (a), (c) and carbon (b), (d) isotope composition from analysis of 89 individuals, sample numbers differ between subspecies with SS = 35, SL = 23, SLE = 31. In panels (a)–(b), the box represents the interquartile range whilst the whiskers show 1.5*interquartile range. Black circles represent outliers exceeding the 1.5*interquartile range. Black line represents the mean. Individual dots represent all individual specimens sampled. The among-individual distribution of stable isotope measurements reinforces the importance of considering sufficient sample sizes for intraspecific variability. Panels c–d show oxygen and carbon isotopes of individual foraminifera versus test area (size) separated into subspecies. Bars represent 0.09‰ analytical precision of oxygen measurements (c) and 0.05‰ of carbon measurements (d). The lines show linear regression of isotope value against log(size) with gray representing 95% confidence intervals.

et al., 2017; Kontakiotis et al., 2017; Kuroyanagi et al., 2008; Lin et al., 2004; Löwemark et al., 2005; Steinke et al., 2010; Wang, 2000). Based on our analysis and observations we recommend researchers focus on the final chamber, particularly the degree of compression to separate the subspecies of *G. ruber,* with compression of the final chamber increasing along the morphological cline from SS to SLE (Figures 1a and 3d). We suggest that future morphological studies should focus on final chamber traits as well as investigating morphological differences in the 3D space using µ-CT scanning. Furthermore, we also suggest that future studies measure traits that have a known functional role and can be linked to the environment such as pore size which is linked to gas exchange (Bé, 1968; Burke et al., 2018; Constandache et al., 2013; Kearns et al., 2021).

4.2. Geochemical Variability

Geochemical variability in *G. ruber* does exist at this Indian Ocean site and is the result of interspecific variability. *G. ruber* SS shows systematically higher mean Mg/Ca ratios compared to SL and SLE of 0.48 ± 0.18 and 1.11 ± 0.17 mmol/mol, respectively, based on test-averaged mean Mg/Ca ratios. When test-averaged mean Mg/Ca ratios are converted to temperature using Equation 1 this equates to a 1.34 ± 0.47°C difference between SS and SL and a 3.01 ± 0.43°C difference between SS and SLE.

The temperature differences we find between SS and SL are slightly higher than those found in a study of the Indian Ocean and Western Pacific (0.91 ± 0.75°C; Steinke et al. (2005)). Our Mg/Ca derived temperature difference between SS and SL is smaller than previously found in the Gulf of Mexico (~3°C; Antonarakou et al., 2015), but the SL morphological concept used by Antonarakou et al. (2015) is similar to our SLE here (Figure 1). Based on this observation and published confidence intervals, our 3.01°C difference between SS and

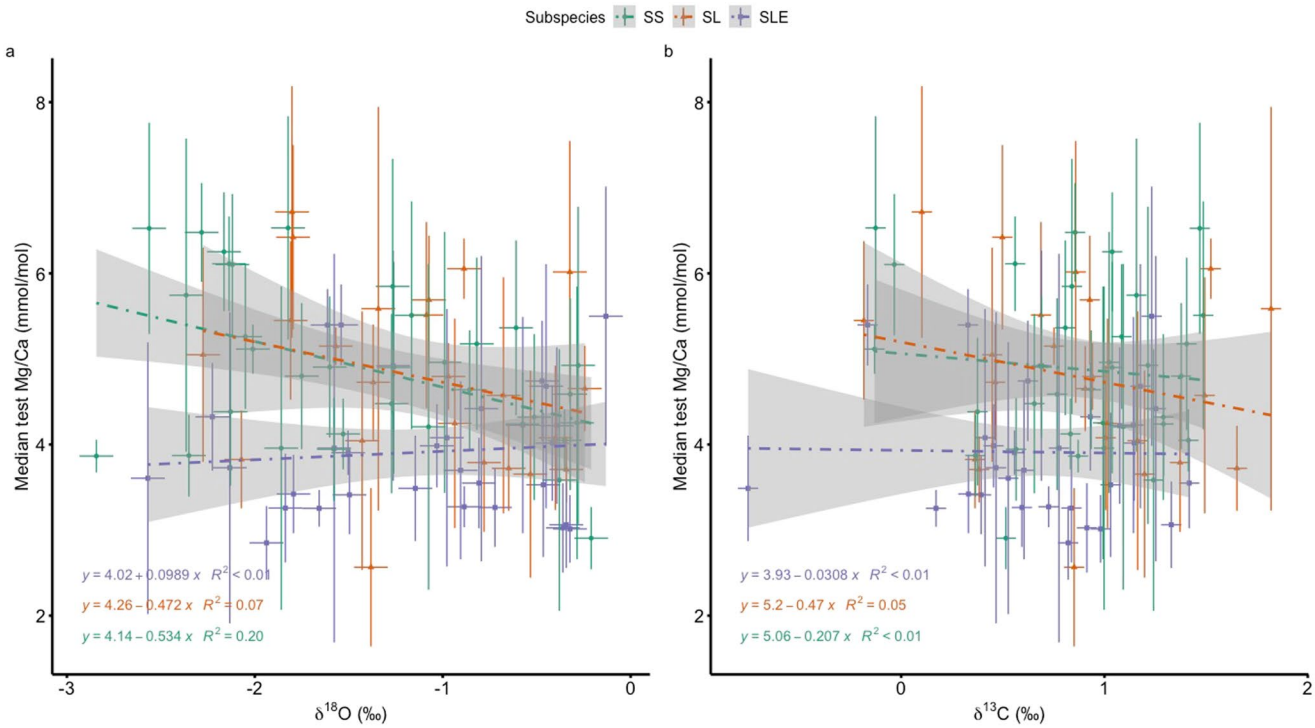


Figure 7. Stable isotopes oxygen (a) and carbon (b) values plotted against the median test Mg/Ca ratios. Each point represents the median chamber value in the *y*-axis against whole specimen $\delta^{18}O$ and $\delta^{13}C$ values with error bars on the *x*-axis representing 0.09‰ analytical precision in oxygen isotopes (a) and 0.05‰ analytical precision in carbon isotopes (b). On the *y*-axis error bars represent median average deviation (MAD) between the spots across the whole test. Where no error bar is present in the *y*-axis the median value is based on only one spot on the corresponding chamber.

SLE in this study is within the range of that between SS and SL in the Gulf of Mexico (Antonarakou et al., 2015). Detailed taxonomic work is clearly fundamental for unbiased geochemical inference.

Despite our trace element results, the $\delta^{18}O$ and $\delta^{13}C$ values do not show statistically significant differences amongst subspecies (Figure 6). The lack of significant isotopic differences among subspecies (Section 3.3) is similar to some studies (Lynch-stieglitz et al., 2015; Mohtadi et al., 2009; Thirumalai et al., 2014) but not others (Antonarakou et al., 2015; Carter et al., 2017; Löwemark et al., 2005; Steinke et al., 2010; Wang, 2000). In those studies that found a significant difference between SS and SL, the smallest mean difference observed over glacial-interglacial cycles is similar to the mean $\delta^{18}O$ difference of ~0.2‰ we observe (Carter et al., 2017; Steinke et al., 2010; Wang, 2000). However, despite similar values, we failed to statistically separate subspecies in stable isotope space (Figure 6). We do however see a large degree of variability resulting in a higher $\delta^{18}O$ mean (−1.21‰) and lower $\delta^{13}C$ (0.82‰) mean than has been recorded in other sites from this cruise (Birch et al., 2013) and another regionally similar study (Fallet et al., 2010), though both those studies were based on bulk isotopes and not individual analyses. When converted to temperature using Equation 2, our $\delta^{18}O$ results produce mean temperatures in all subspecies to be below 25°C (SS = 24.35°C, SL = 23.25°C and SLE = 23.20°C). Based on CTD profiles from the area (Birch et al., 2013; Kroon and the Shipboard Scientific Party, 2010) our $\delta^{18}O$ values suggest all subspecies lived below the base of the SML at ~40m (Birch et al., 2013). The CTD profiles of DIC $\delta^{13}C$ show that there is a weak $\delta^{13}C$ gradient in the upper 100 m of the water column, potentially explaining the narrow range and lack of statistical significance we observe in our $\delta^{13}C$ samples. The range in $\delta^{13}C$ that we do observe probably reflects individual variation in calcification, respiration, and photosynthesis (Henehan et al., 2017; Lombard et al., 2009) and the influence of these processes on the $\delta^{13}C$ of foraminifera (Zeebe. 1999).

4.2.1. Is Trace Element Variability in *G. ruber* Ecologically Driven?

We have shown intraspecific variation in *G. ruber* is systematic. *G. ruber* is restricted to the upper water column (Anand & Elderfield, 2005; Dekens et al., 2002) and is not thought to migrate vertically during life (Aurahs et al., 2011; Tolderlund & Bé, 1971). Despite being used to indicate SML temperatures, *G. ruber* does have

a broad ecological niche extending below the base of the SML to the base of the deep chlorophyll maximum (DCM) (Lončarić et al., 2006; Peeters et al., 2002).

Separation of subspecies in geochemical space has been inferred previously to result from subspecific specific depth habitats (Antonarakou et al., 2015; Carter et al., 2017; Kawahata, 2005; Kuroyanagi & Kawahata, 2004; Löwemark et al., 2005; Numberger et al., 2009; Steinke et al., 2005; Wang, 2000). Our Mg/Ca results agree with this inference with SLE and SL calcifying in an ∼3°C and ∼1.3°C respectively cooler, deeper part of the upper water column compared to SS. Subspecies calcification depth variability would explain the large ecological niche found in other studies (Lončarić et al., 2006; Peeters et al., 2002). Based on CTD profiles from the area (Birch et al., 2013; Kroon and the Shipboard Scientific Party, 2010), our Mg/Ca temperature calibrations would place SS slightly above the base of the SML at ∼40m (Birch et al., 2013) with SLE inhabiting the upper thermocline below the SML but well above the deep chlorophyll maximum at ∼95m (Birch et al., 2013) and SL in-between SS and SLE.

Another potential influence of subspecies temperature variability is the seasonal preference of subspecies. The annual seasonal SST range in this area is ∼5°C (Birch et al., 2013; Damassa et al., 2006; McClanahan, 1988), which is larger than 1–3°C range in Mg/Ca derived temperature we observe among subspecies. Seasonality would imply that SS preferentially live during the summer months when SST is ∼30°C whilst SLE lives during the winter when SST decreases to around 25°C (Birch et al., 2013). SL calculated temperature sits between these two seasonal temperatures suggesting a deeper depth habitat rather than seasonal preference. Preferential seasonality has been proposed as a potential driver in other ocean basins (Antonarakou et al., 2015; Sadekov et al., 2008) but sediment trap studies in Java (Mohtadi et al., 2009), South China Sea (Lin et al., 2004) and the Gulf of Mexico (Thirumalai et al., 2014) as well as plankton tow studies in the Arabian Sea (Peeters et al., 2002) suggest no seasonal or monsoonal preference of *G. ruber* subspecies.

We therefore conclude that our subspecies differences are due to depth habitat preferences with SLE and SL inhabiting a cooler, deeper part of the water column compared to SS. Therefore, the intraspecific variability in *G. ruber* is ecologically meaningful and studies that use a mean-based approach without separating between subspecies are removing ecologically meaningful data whilst potentially influencing paleoclimatic reconstructions. However, rather than being a hindrance, this depth habitat separation of subspecies in *G. ruber*, if consistent across ocean basins, could be used to reconstruct SML changes through time.

4.2.2. Drivers of Stable Isotopes

Whilst our Mg/Ca ratios are comparable across studies (Section 4.2), our stable isotope results are not and furthermore do not align with our Mg/Ca ratios. Additionally, though a weak linear relationship does exist between $\delta^{18}O$ values and Mg/Ca ratios (Figure 7a), indicating a similar depth habitat control on both proxies the relationship is far from perfect and not statistically significant, implying further factors are acting on one or both proxies. The imperfect correlations we see between stable isotopes and Mg/Ca ratios (Figure 7) make it difficult to untangle any secondary controls on Mg/Ca, $\delta^{18}O$, and $\delta^{13}C$. Data from across a wider size range may result in a larger signal-to-noise ratio providing valuable into the various environmental and physiological controls on these proxies and should therefore be the focus of future studies.

4.2.3. Ontogenetic Variability of Mg/Ca Ratios

Subspecies variability (Section 4.2.1) is complicated further by ontogenetic changes (Figure 2). Test heterogeneity of Mg is present in all species of foraminifera and thought to be at least partly decoupled from temperature (Spero et al., 2015). Intratest variability is not unexpected, given the nature of the water column and planktic lifestyle of these foraminifera (Pracht et al., 2019). Ontogenetic variability has been observed in other studies of planktic foraminifera (Anand & Elderfield, 2005; Bolton et al., 2011; Dueñas-Bohórquez et al., 2011; Sadekov et al., 2008). Here, we show for the first time that interchamber variation in *G. ruber* Mg/Ca is systematic at the subspecies level with different Mg/Ca ratios through the final whorl. We observed an overall pattern of final chamber depletion in all subspecies (Figure 4) with the antepenultimate and penultimate chambers showing similar values (Figure 4). The subspecies similarities of these patterns suggest that they are ecologically or biologically driven geochemical signatures.

Final chamber depletion is observed in various planktic foraminifera (*T. sacculifer* (Dueñas-Bohórquez et al., 2011; Hemleben et al., 1989); *G. bulloides* (Anand & Elderfield, 2005; Marr et al., 2011), and *G. ruber* (Bolton et al., 2011; Sadekov et al., 2008)). This pattern is often thought to be a byproduct of the life cycle of

planktic foraminifera and the deepening depth habitat during the terminal growth stages (Bijma et al., 1990; Pracht et al., 2019) and therefore linked to temperature changes in the water column. However, observations of final chamber depletion in cultured, lab-grown *T. sacculifer* (Dueñas-Bohórquez et al., 2011) and plankton tow samples (Bolton et al., 2011) suggest that this pattern may not be environmentally linked. Furthermore *G. ruber,* as previously noted, is not thought to migrate vertically during life (Aurahs et al., 2011; Meilland et al., 2019; Tolderlund & Bé, 1971).

Interspecific chamber heterogeneity could be a result of the biomineralization process. In symbiont bearing planktic foraminifera, like *G. ruber*, diurnal changes in the biological activity of algal symbionts have been hypothesized to contribute to Mg/Ca banding within the chamber walls (Eggins et al., 2004; Fehrenbacher et al., 2017; Sadekov et al., 2005). As foraminifera grow, older chambers are overprinted with the calcite of newly formed chambers (Hemleben et al., 1989) increasing chamber thickness and adding new Mg/Ca bands. Intratest variability in high and low Mg bands may explain the variation we see if such banding was subspecies specific and a geochemical "vital effect." The number of high Mg bands diminishes through the final whorl of *G. ruber* with the final chamber made up of only low Mg bands (Sadekov et al., 2005). Although this banding is not thought to impact overall test signal (Holland et al., 2020), it could influence chamber specific signals. The observed absence of high Mg bands in the final chamber of *G. ruber* specimens (Sadekov et al., 2005) may also explain the apparent depletion of the final chamber in this study and others (Anand & Elderfield, 2005; Bolton et al., 2011; Dueñas-Bohórquez et al., 2011; Sadekov et al., 2008). The experimental design of our study meant that an investigation of intratest and interspecific banding differences, while of fundamental interest, is not possible given the nature of our data collection. Chamber heterogeneity comparisons between test averaged LA-ICP-MS and solution ICP-MS have shown that averaging Mg/Ca across chambers is a good indicator of past ocean conditions (Fehrenbacher et al., 2020; Rustic et al., 2021). We therefore recommend that studies using LA-ICP-MS should measure all chambers in the final whorl (Fehrenbacher et al., 2020; Rustic et al., 2021).

4.2.4. Influence of Chamber and Test Size on Mg/Ca Ratios

Test size is recognized as an influential driver of stable isotope variability in foraminifera (Elderfield et al., 2002; Ezard et al., 2015; Friedrich et al., 2012; Spero, 1998; Spero & Lea, 1996). To minimize such effects, stable isotope analyses are typically conducted on narrow size fractions. A similar practice has been applied to trace elements (Cléroux et al., 2008; Elderfield et al., 2002; Friedrich et al., 2012; McConnell & Thunell, 2005; Ni et al., 2007) with the larger size fraction often recommended (Elderfield et al., 2002). In this study, we picked *G. ruber* individuals from a narrow size fraction (250–355 μm) and found no impact of size on Mg/Ca ratios (Figure 4; Table S7 in Supporting Information S1). Therefore, the use of narrow size fractions remains a good mitigation technique in trace element analysis to avoid test size effects.

The effects of individual chamber size on Mg/Ca ratios have never been investigated until now. We found no detectable impact of chamber size on chamber Mg/Ca ratios of *G. ruber* (Section 3.2.1; Table S8 in Supporting Information S1). The final chamber depletion we consistently observe in all subspecies is not therefore the result of smaller final chambers as often found in "kummerform" individuals (Berger, 1969; Olsson, 1973).

5. Conclusions and Paleoceanographic Implications

Mg/Ca ratios vary systematically due to the presence of three subspecies that live at different depth habitats within the surface mixed year. Using LA-ICP-MS at chamber-by-chamber resolution, we have shown that individual foraminiferal analysis can be used effectively to reveal ecologically meaningful information that is useful for paleoceanographic reconstructions. Although such fine scale measurements do generate statistical noise, meaningful patterns can still be found when sufficient samples of individuals are used. To avoid signal mixing, we recommend (a) paleoceanographic studies should favor *G. ruber* SS as the subspecies that best represents the SML; (b) the continued selection of the narrowest size fraction for LA-ICP-MS analysis of Mg/Ca ratios until other studies find a qualitatively similar result; (c) the continued sampling from the narrowest size fraction possible for stable isotope analysis; (d) to determine test Mg/Ca final whorl chamber resolution LA-ICP-MS should be used; and (e) the need for larger sample sizes than routinely used to sufficiently represent a wider homogeneous population. Through this analysis we have also demonstrated and discussed how more work is needed to understand the drivers of Mg incorporation in *G. ruber* (and other foraminifera) and how this signal is influenced by environmental and biological variables. Better understanding of the biological, ecological, and environmental

causes of intraspecific variability would unlock the full potential of measurements, integrating single-specimen and time-averaged bulk measurements, to reveal past climate change at all levels of temporal granularity from days to millions of years.

Data Availability Statement

Geochemical and morphological data generated as part of this study as well as code to reproduce our results and supplementary figures are available on Figshare via Kearns et al. (2022a) and Kearns et al. (2022b). All analysis and figure creation were done in the R environment (version 4.0.3; R Core Team, 2020).

Acknowledgments
This work was funded by the Natural Environment Research Council NE/P019269/1 to THGE. THGE was also funded by Fellowship NE/J018163/1; LEK was part-funded by a University of Southampton Institute of Life Sciences' PhD Studentship and Natural Environment Research Council NE/L002531/1. We thank Ryan Glaubke and two anonymous reviewers for their comments on this manuscript, and three anonymous reviewers for their comments on a previous version of this manuscript. We also thank Dr Anieke Brombacher for helping with taxonomic and morphometric measurements and Dr Max Holmström helping with stable isotope measurements.

References

Anand, P., & Elderfield, H. (2005). Variability of Mg/Ca and Sr/Ca between and within the planktonic foraminifers *Globigerina bulloides* and *Globorotalia truncatulinoides*. *Geochemistry, Geophysics, Geosystems*, *6*(11). https://doi.org/10.1029/2004GC000811

Anand, P., Elderfield, H., & Conte, M. H. (2003). Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography*, *18*(2). https://doi.org/10.1029/2002PA000846

André, A., Quillévéré, F., Morard, R., Ujiié, Y., Escarguel, G., de Vargas, C., et al. (2014). SSU rDNA divergence in planktonic foraminifera: Molecular taxonomy and biogeographic implications. *PLoS One*, *9*(8), e104641. https://doi.org/10.1371/journal.pone.0104641

Antonarakou, A., Kontakiotis, G., Mortyn, P. G., Drinia, H., Sprovieri, M., Besiou, E., & Tripsanas, E. (2015). Biotic and geochemical (δ18O, δ13C, Mg/Ca, Ba/Ca) responses of *Globigerinoides ruber* morphotypes to upper water column variations during the last deglaciation, Gulf of Mexico. *Geochimica et Cosmochimica Acta*, *170*, 69–93. https://doi.org/10.1016/j.gca.2015.08.003

Aurahs, R., Treis, Y., Darling, K., & Kucera, M. (2011). A revised taxonomic and phylogenetic concept for the planktonic foraminifer species *Globigerinoides ruber* based on molecular and morphometric evidence. *Marine Micropaleontology*, *79*(1–2), 1–14. https://doi.org/10.1016/j.marmicro.2010.12.001

Aze, T., Pearson, P. N., Dickson, A. J., Badger, M. P. S., Bown, P. R., Pancost, R. D., et al. (2014). Extreme warming of tropical waters during the Paleocene-Eocene thermal maximum. *Geology*, *42*(9), 739–742. https://doi.org/10.1130/G35637.1

Bates, D. (2005). Fitting linear mixed models in R. *R News*, *5*(1), 27–30.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bé, A. W. H. (1968). Shell porosity of recent planktonic foraminifera as a climatic index. *Science*, *161*(3844), 881–884. https://doi.org/10.1126/science.161.3844.881

Belyaeva, N. V., & Burmistrova, I. (1984). Critical levels of carbonates accumulation in the Indian Ocean. *Doklady of the USSR Academy of Sciences Earth Science Section*, *277*, 652–655.

Bemis, B. E., Spero, H. J., Bijma, J., & Lea, D. W. (1998). Reevaluation of the oxygen isotopic composition of planktonic foraminifera. *Paleoceanography*, *13*(2), 150–160. https://doi.org/10.1029/98pa00070

Bemis, B. E., Spero, H. J., Lea, D. W., & Bijma, J. (2000). Temperature influence on the carbon isotopic composition of *Globigerina bulloides* and *Orbulina universa* (planktonic foraminifera). *Marine Micropaleontology*, *38*(3–4), 213–228. https://doi.org/10.1016/S0377-8398(00)00006-2

Berger, W. H. (1969). Kummerform foraminifera as clues to oceanic environments. *AAPG Bulletin*, *53*(3), 706. https://doi.org/10.1306/5D25C6C1-16C1-11D7-8645000102C1865D

Bickert, T. (2009). Carbonate compensation depth. In V. Gornitz (Ed.), *Encyclopedia of paleoclimatology and ancient environments* (pp. 136–138). Springer Netherlands. https://doi.org/10.1007/978-1-4020-4411-3_33

Bijma, J., Erez, J., & Hemleben, C. (1990). Lunar and semi-lunar reproductive cycles in some spinose planktonic foraminifera. *Journal of Foraminiferal Research*, *20*(2), 117–127. https://doi.org/10.2113/gsjfr.20.2.117

Bijma, J., Spero, H. J., & Lea, D. W. (1999). Reassessing foraminiferal stable isotope geochemistry: Impact of the oceanic carbonate system (experimental results). *Use of Proxies in Paleoceanography*, (pp. 489–512). https://doi.org/10.1007/978-3-642-58646-0_20

Birch, H., Coxall, H. K., Pearson, P. N., Kroon, D., & O'Regan, M. (2013). Planktonic foraminifera stable isotopes and water column structure: Disentangling ecological signals. *Marine Micropaleontology*, *101*, 127–145. https://doi.org/10.1016/j.marmicro.2013.02.002

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Bolnick, D. I., Amarasekare, P., Araujo, M. S., Burger, R., Levine, J. M., Novak, M., et al. (2011). Why intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, *26*(4), 183–192. https://doi.org/10.1016/j.tree.2011.01.009

Bolton, A., Baker, J. A., Dunbar, G. B., Carter, L., Smith, E. G. C., & Neil, H. L. (2011). Environmental versus biological controls on Mg/Ca variability in *Globigerinoides ruber* (white) from core top and plankton tow samples in the southwest Pacific Ocean. *Paleoceanography*, *26*(2), 1–14. https://doi.org/10.1029/2010PA001924

Brombacher, A., Wilson, P. A., & Ezard, T. H. G. (2017). Calibration of the repeatability of foraminiferal test size and shape measures with recommendations for future use. *Marine Micropaleontology*, *133*, 21–27. https://doi.org/10.1016/j.marmicro.2017.05.003

Burke, J. E., Renema, W., Henehan, M. J., Elder, L. E., Davis, C. V., Maas, A. E., et al. (2018). Factors influencing test porosity in planktonic foraminifera. *Biogeosciences*, *15*(21), 6607–6619. https://doi.org/10.5194/bg-15-6607-2018

Burnham, K., & Anderson, D. (2002). Model selection and multi-model inference.

Carter, A., Clemens, S., Kubota, Y., Holbourn, A., & Martin, A. (2017). Differing oxygen isotopic signals of two *Globigerinoides ruber* (white) morphotypes in the East China Sea: Implications for paleoenvironmental reconstructions. *Marine Micropaleontology*, *131*, 1–9. https://doi.org/10.1016/j.marmicro.2017.01.001

Cléroux, C., Cortijo, E., Anand, P., Labeyrie, L., Bassinot, F., Caillon, N., & Duplessy, J.-C. (2008). Mg/Ca and Sr/Ca ratios in planktonic foraminifera: Proxies for upper water column temperature reconstruction. *Paleoceanography*, *23*(3), n/a–n/a. https://doi.org/10.1029/2007PA001505

Constandache, M., Yerly, F., & Spezzaferri, S. (2013). Internal pore measurements on macroperforate planktonic Foraminifera as an alternative morphometric approach. *Swiss Journal of Geosciences*, *106*(2), 179–186. https://doi.org/10.1007/s00015-013-0134-8

Damassa, T. D., Cole, J. E., Barnett, H. R., Ault, T. R., & McClanahan, T. R. (2006). Enhanced multidecadal climate variability in the seventeenth century from coral isotope records in the Western Indian Ocean. *Paleoceanography*, *21*(2). https://doi.org/10.1029/2005PA001217

Darling, K. F., & Wade, C. M. (2008). The genetic diversity of planktic foraminifera and the global distribution of ribosomal RNA genotypes. *Marine Micropaleontology*, *67*(3–4), 216–238. https://doi.org/10.1016/j.marmicro.2008.01.009

Dekens, P. S., Lea, D. W., Pak, D. K., & Spero, H. J. (2002). Core top calibration of Mg/Ca in tropical foraminifera: Refining paleotemperature estimation. *Geochemistry, Geophysics, Geosystems*, *3*(4), 1–29. https://doi.org/10.1029/2001gc000200

Dinno, A. (2018). Paran: Horn's test of principal components/factors.

Dueñas-Bohórquez, A., da Rocha, R. E., Kuroyanagi, A., de Nooijer, L. J., Bijma, J., & Reichart, G. J. (2011). Interindividual variability and ontogenetic effects on Mg and Sr incorporation in the planktonic foraminifer *Globigerinoides sacculifer*. *Geochimica et Cosmochimica Acta*, *75*(2), 520–532. https://doi.org/10.1016/j.gca.2010.10.006

Edgar, K. M., Hull, P. M., & Ezard, T. H. G. (2017). Evolutionary history biases inferences of ecology and environment from δ13C but not δ18O values. *Nature Communications*, *8*(1), 1106. https://doi.org/10.1038/s41467-017-01154-7

Eggins, S., De Deckker, P., & Marshall, J. (2003). Mg/Ca variation in planktonic foraminifera tests: Implications for reconstructing palaeo-seawater temperature and habitat migration. *Earth and Planetary Science Letters*, *212*(3–4), 291–306. https://doi.org/10.1016/S0012-821X(03)00283-8

Eggins, S. M., Sadekov, A., & De Deckker, P. (2004). Modulation and daily banding of Mg/Ca in *Orbulina universa* tests by symbiont photosynthesis and respiration: A complication for seawater thermometry? *Earth and Planetary Science Letters*, *225*(3), 411–419. https://doi.org/10.1016/j.epsl.2004.06.019

Elderfield, H., & Ganssen, G. (2000). Past temperature and δ18O of surface ocean waters inferred from foraminiferal Mg/Ca ratios. *Nature*, *405*(6785), 442–445. https://doi.org/10.1038/35013033

Elderfield, H., Vautravers, M., & Cooper, M. (2002). The relationship between shell size and Mg/Ca, Sr/Ca, D 18 O, and D 13 C of species of planktonic foraminifera. *Geochemistry, Geophysics, Geosystems*, *3*(8), 1–13. https://doi.org/10.1029/2001GC000194

Emiliani, C. (1954). Depth habitats of some species of pelagic Foraminifera as indicated by oxygen isotope ratios. *American Journal of Science*, *252*(3), 149–158. https://doi.org/10.2475/ajs.252.3.149

Epstein, S., Buchsbaum, R., Lowenstam, H. A., & Urey, H. C. (1951). Carbonate-water isotopic tempreature scale. *Bulletin of the Geological Society of American society of America*, *62*(April), 417–426.

Erez, J., & Luz, B. (1983). Experimental paleotemperature equation for planktonic foraminifera. *Geochimica et Cosmochimica Acta*, *47*(6), 1025–1031. https://doi.org/10.1016/0016-7037(83)90232-6

Ezard, T. H. G., Edgar, K. M., & Hull, P. M. (2015). Environmental and biological controls on size-specific ^{13}C and ^{18}O in recent planktonic foraminifera. *Paleoceanography*, *30*(3), 151–173. https://doi.org/10.1002/2014PA002735

Fallet, U., Brummer, G.-J., Zinke, J., Vogels, S., & Ridderinkhof, H. (2010). Contrasting seasonal fluxes of planktonic foraminifera and impacts on paleothermometry in the Mozambique channel upstream of the Agulhas current. *Paleoceanography*, *25*(4). https://doi.org/10.1029/2010PA001942

Fehrenbacher, J., Marchitto, T., & Spero, H. J. (2020). Comparison of laser ablation and solution-based ICP-MS results for individual foraminifer Mg/Ca and Sr/Ca analyses. *Geochemistry, Geophysics, Geosystems*, *21*(12), e2020GC009254. https://doi.org/10.1029/2020GC009254

Fehrenbacher, J. S., Russell, A. D., Davis, C. V., Gagnon, A. C., Spero, H. J., Cliff, J. B., et al. (2017). Link between light-triggered Mg-banding and chamber formation in the planktic foraminifera *Neogloboquadrina dutertrei*. *Nature Communications*, *8*(May), 1–10. https://doi.org/10.1038/ncomms15441

Ford, H. L., Ravelo, A. C., & Polissar, P. J. (2015). Reduced El Niño–southern oscillation during the last glacial maximum. *Science*, *347*(6219), 255–258. https://doi.org/10.1126/science.1258437

Fraley, C., & Raftery, A. E. (1996). Model-based methods of classification: Using the mclust software in chemometrics. *Journal of Statistical Software*, *18*(6), 1–13. https://doi.org/10.18637/jss.v018.i06

Friedman, I., & O'Neil, J. R. (1977). Compilation of stable isotope fractionation factors of geochemical interest. In M. Fleischer (Ed.), *Data of geochemistry* (pp. 1–12). U. S. Government Printing Office.

Friedrich, O., Schiebel, R., Wilson, P. A., Weldeab, S., Beer, C. J., Cooper, M. J., & Fiebig, J. (2012). Influence of test size, water depth, and ecology on Mg/Ca, Sr/Ca, δ 18O and δ 13C in nine modern species of planktic foraminifers. *Earth and Planetary Science Letters*, *319–320*, 133–145. https://doi.org/10.1016/j.epsl.2011.12.002

Gillies, C. S., Hebblewhite, M., Nielsen, S. E., Krawchuk, M. A., Aldridge, C. L., Frair, J. L., et al. (2006). Application of random effects to the study of resource selection by animals. *Journal of Animal Ecology*, *75*(4), 887–898. https://doi.org/10.1111/j.1365-2656.2006.01106.x

Groeneveld, J., Ho, S. L., & Mohtadi, M. (2019). Deciphering the variability in Mg/Ca and stable oxygen isotopes of individual foraminifera Paleoceanography and Paleoclimatology (pp. 755–773). https://doi.org/10.1029/2018PA003533

Hart, S. P., Schreiber, S. J., & Levine, J. M. (2016). How variation between individuals affects species coexistence. *Ecology Letters*, *19*(8), 825–838. https://doi.org/10.1111/ele.12618

Hemleben, C., Spindler, M., & Anderson, O. R. (1989). *Modern planktonic foraminifera*. Springer Verlag.

Henehan, M. J., Evans, D., Shankle, M., Burke, J. E., Foster, G. L., Anagnostou, E., et al. (2017). Size-dependent response of foraminiferal calcification to seawater carbonate chemistry. *Biogeosciences*, *14*(13), 3287–3308. https://doi.org/10.5194/bg-14-3287-2017

Holland, K., Branson, O., Haynes, L. L., Hönisch, B., Allen, K. A., Russell, A. D., et al. (2020). Constraining multiple controls on planktic foraminifera Mg/Ca. *Geochimica et Cosmochimica Acta*, *273*, 116–136. https://doi.org/10.1016/j.gca.2020.01.015

Hönisch, B., Bijma, J., Russell, A. D., Spero, H. J., Palmer, M. R., Zeebe, R. E., & Eisenhauer, A. (2003). The influence of symbiont photosynthesis on the boron isotopic composition of foraminifera shells. *Marine Micropaleontology*, *49*(1–2), 87–96. https://doi.org/10.1016/S0377-8398(03)00030-6

Ivanova, E. V. (2009). Paleoceanography of the northern Indian ocean: Linkages to monsoon and global thermohaline paleocirculation. In *The global Thermohaline Paleocirculation* (pp. 107–145). Springer Netherlands. https://doi.org/10.1007/978-90-481-2415-2_5

Jochum, K. P., Garbe-Schönberg, D., Veter, M., Stoll, B., Weis, U., Weber, M., et al. (2019). Nano-powdered Calcium carbonate reference materials: Significant progress for microanalysis? *Geostandards and Geoanalytical Research*, *43*(4), 595–609. https://doi.org/10.1111/ggr.12292

Jochum, K. P., Weis, U., Stoll, B., Kuzmin, D., Yang, Q., Raczek, I., et al. (2011). Determination of reference values for NIST SRM 610–617 glasses following ISO guidelines. *Geostandards and Geoanalytical Research*, *35*(4), 397–429. https://doi.org/10.1111/j.1751-908X.2011.00120.x

Jung, V., Violle, C., Mondy, C., Hoffmann, L., & Muller, S. (2010). Intraspecific variability and trait-based community assembly. *Journal of Ecology*, *98*(5), 1134–1140. https://doi.org/10.1111/j.1365-2745.2010.01687.x

Kawahata, H. (2005). Stable isotopic composition in the subtropical Gtobigerinoi des ruber (white) in the North Pacific identified. *Paleontological Research*, *9*(1), 27–35. https://doi.org/10.2517/prpsj.9.27

Kearns, L., Searle-Barnes, A., Foster, G. L., Milton, J., Standish, C., & Ezard, T. (2022a). Data sets for 'The influence of geochemical variation among *Globigerinoides ruber* individuals on paleoceanographic reconstructions' (Version 0). [Dataset]. figshare. https://doi.org/10.6084/m9.figshare.21071203

Kearns, L., Searle-Barnes, A., Foster, G. L., Milton, J., Standish, C., & Ezard, T. (2022b). R Code (.Rmd) for 'The influence of geochemical variation among *Globigerinoides ruber* individuals on paleoceanographic reconstructions' (Version 0). [Software]. figshare. https://doi.org/10.6084/m9.figshare.21071224

Kearns, L. E., Bohaty, S. M., Edgar, K. M., Nogué, S., & Ezard, T. H. G. (2021). Searching for function: Reconstructing adaptive niche changes using geochemical and morphological data in planktonic foraminifera. *Frontiers in Ecology and Evolution*, *9*, 1–17. https://doi.org/10.3389/fevo.2021.679722

Kelly, D. C., Bralower, T. J., & Zachos, J. C. (1997). Evolutionary consequences of the latest Paleocene thermal maximum for tropical planktonic foraminifera. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *141*(1–2), 139–161. https://doi.org/10.1016/S0031-0182(98)00017-0

Kontakiotis, G., Antonarakou, A., Mortyn, P. G. G., Drinia, H., Anastasakis, G., Zarkogiannis, S., & Möbius, J. (2017). Morphological recognition of *Globigerinoides ruber* morphotypes and their susceptibility to diagenetic alteration in the eastern Mediterranean Sea. *Journal of Marine Systems*, *174*, 12–24. https://doi.org/10.1016/j.jmarsys.2017.05.005

Kroon and the Shipboard Scientific Party. (2010). Tropical temperature history during Paleogene global warming (GLOW) events NIOZ site Survey cruise Report.

Kuroyanagi, A., & Kawahata, H. (2004). Vertical distribution of living planktonic foraminifera in the seas around Japan. *Marine Micropaleontology*, *53*(1–2), 173–196. https://doi.org/10.1016/j.marmicro.2004.06.001

Kuroyanagi, A., Tsuchiya, M., Kawahata, H., & Kitazato, H. (2008). The occurrence of two genotypes of the planktonic foraminifer *Globigerinoides ruber* (white) and paleo-environmental implications. *Marine Micropaleontology*, *68*(3–4), 236–243. https://doi.org/10.1016/j.marmicro.2008.04.004

Lazarus, D. (1983). Paleontological society speciation in pelagic Protista and its study in the planktonic microfossil record: A review speciation in pelagic Protista and its study in the planktonic microfossil record: A review. *Paleobiology*, *9*(4), 327–340. https://doi.org/10.1017/s0094837300007806

Lea, D. W., Mashiotta, T. A., & Spero, H. J. (1999). Controls on magnesium and strontium uptake in planktonic foraminifera determined by live culturing. *Geochimica et Cosmochimica Acta*, *63*(16), 2369–2379. https://doi.org/10.1016/S0016-7037(99)00197-0

Lea, D. W., Pak, D. K., & Spero, H. J. (2000). Climate impact of late quaternary quatorial Pacific Sea surface temperature variations. *Science*, *289*(5485), 1719–1724. https://doi.org/10.1126/science.289.5485.1719

Lin, H. L., Wang, W. C., & Hung, G. W. (2004). Seasonal variation of planktonic foraminiferal isotopic composition from sediment traps in the South China Sea. *Marine Micropaleontology*, *53*(3–4), 447–460. https://doi.org/10.1016/j.marmicro.2004.08.004

Lombard, F., Erez, J., Michel, E., & Labeyrie, L. (2009). Temperature effect on respiration and photosynthesis of the symbiont-bearing planktonic foraminifera *Globigerinoides ruber*, *Orbulina universa*, and *Globigerinella siphonifera*. *Limnology & Oceanography*, *54*(1), 210–218. https://doi.org/10.4319/lo.2009.54.1.0210

Lončarić, N., Peeters, F. J. C., Kroon, D., & Brummer, G.-J. A. (2006). Oxygen isotope ecology of recent planktic foraminifera at the central Walvis Ridge (SE Atlantic). *Paleoceanography*, *21*(3). https://doi.org/10.1029/2005PA001207

Löwemark, L., Hong, W.-L., Yui, T.-F., & Hung, G.-W. (2005). A test of different factors influencing the isotopic signal of planktonic foraminifera in surface sediments from the northern South China Sea. *Marine Micropaleontology*, *55*(1–2), 49–62. https://doi.org/10.1016/j.marmicro.2005.02.004

Lynch-stieglitz, J., Polissar, P. J., Jacobel, A. W., Hovan, S. A., Pockalny, R. A., Lyle, M., et al. (2015). Glacial-interglacial changes in central tropical Pacific surface seawater property gradients. *Paleoceanography*, *30*(5), 423–438. https://doi.org/10.1002/2014PA002746

Marr, J. P., Baker, J. A., Carter, L., Allan, A. S. R., Dunbar, G. B., & Bostock, H. C. (2011). Ecological and temperature controls on Mg/Ca ratios of *Globigerina bulloides* from the southwest Pacific Ocean. *Paleoceanography*, *26*(2), 1–15. https://doi.org/10.1029/2010PA002059

Mayr, E. (1942). *Systematics and the origin of species*. Columbia University Press.

Mayr, E., & Ashlock, P. D. (1991). *Principles of systematic biology*. McGraw-Hill.

McClanahan, T. (1988). Seasonality in East Africa's coastal waters. *Marine Ecology Progress Series*, *44*, 191–199. https://doi.org/10.3354/meps044191

McConnell, M. C., & Thunell, R. C. (2005). Calibration of the planktonic foraminiferal Mg/Ca paleothermometer: Sediment trap results from the Guaymas Basin, Gulf of California. *Paleoceanography*, *20*(2). https://doi.org/10.1029/2004PA001077

Meilland, J., Siccha, M., Weinkauf, M. F. G., Jonkers, L., Morard, R., Baranowski, U., et al. (2019). Highly replicated sampling reveals no diurnal vertical migration but stable species-specific vertical habitats in planktonic foraminifera. *Journal of Plankton Research*, *41*(2), 127–141. https://doi.org/10.1093/plankt/fbz002

Metcalfe, B., Feldmeijer, W., & Ganssen, G. M. (2019). Oxygen isotope variability of planktonic foraminifera provide clues to past upper ocean seasonal variability. *Paleoceanography and Paleoclimatology*, *34*(3), 374–393. https://doi.org/10.1029/2018PA003475

Mohtadi, M., Steinke, S., Groeneveld, J., Fink, H. G., Rixen, T., Hebbeln, D., et al. (2009). Low-latitude control on seasonal and interannual changes in planktonic foraminiferal flux and shell geochemistry off south Java: A sediment trap study. *Paleoceanography*, *24*(1). https://doi.org/10.1029/2008PA001636

Morard, R., Darling, K. F., Mahé, F., Audic, S., Ujiié, Y., Weiner, A. K. M., et al. (2015). PFR ^{2}: A curated database of planktonic foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography, and evolution. *Molecular Ecology Resources*, *15*(6), 1472–1485. https://doi.org/10.1111/1755-0998.12410

Morard, R., Füllberg, A., Brummer, G. J. A., Greco, M., Jonkers, L., Wizemann, A., et al. (2019). Genetic and morphological divergence in the warm-water planktonic *foraminifera* genus *Globigerinoides*. *PLoS One*, *14*(12), e0225246. https://doi.org/10.1371/journal.pone.0225246

Naik, S. S. (2016). A comparison of Mg/Ca ratios in *Globigerinoides ruber* (white): Sensu stricto versus a mixture of genotypes. *Journal of the Geological Society of India*, *87*(3), 323–326. https://doi.org/10.1007/s12594-016-0399-4

Ni, Y., Foster, G. L., Bailey, T., Elliott, T., Schmidt, D. N., Pearson, P., et al. (2007). A core top assessment of proxies for the ocean carbonate system in surface-dwelling foraminifers. *Paleoceanography*, *22*(3). https://doi.org/10.1029/2006PA001337

Norris, R. D. (2000). Paleontological society pelagic species diversity, biogeography, and evolution pelagic species diversity, biogeography, and evolution. *Paleobiology*, *26*(4), 236–258. https://doi.org/10.1017/s0094837300026956

Numberger, L., Hemleben, C., Hoffmann, R., Mackensen, A., Schulz, H., Wunderlich, J. M., & Kucera, M. (2009). Habitats, abundance patterns and isotopic signals of morphotypes of the planktonic foraminifer *Globigerinoides ruber* (d'Orbigny) in the eastern Mediterranean Sea since the Marine Isotopic Stage 12. *Marine Micropaleontology*, *73*(1–2), 90–104. https://doi.org/10.1016/j.marmicro.2009.07.004

Olsson, R. K. (1973). What is a kummerform planktonic foraminifer? *Journal of Paleontology*, *47*(2), 327–329.

Parker, F. L. (1962). Planktonic foraminiferal species in Pacific sediments. *Micropaleontology*, *8*(2), 219. https://doi.org/10.2307/1484745

Pearson, P. N. (2012). Oxygen isotopes in foraminifera: Overview and historical review. *Paleontological Society Papers*, *18*, 1–38. https://doi.org/10.1017/s1089332600002539

Peeters, F. J. C., Brummer, G. J. A., & Ganssen, G. (2002). The effect of upwelling on the distribution and stable isotope composition of Globigerina bulloides and *Globigerinoides ruber* (planktic foraminifera) in modern surface waters of the NW Arabian Sea. *Global and Planetary Change*, *34*(3–4), 269–291. https://doi.org/10.1016/S0921-8181(02)00120-0

Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis*, *49*(4), 974–997. https://doi.org/10.1016/j.csda.2004.06.015

Phillimore, A. B., & Owens, I. P. F. (2006). Are subspecies useful in evolutionary and conservation biology? *Proceedings of the Royal Society B: Biological Sciences*, *273*(1590), 1049–1053. https://doi.org/10.1098/rspb.2005.3425

Pracht, H., Metcalfe, B., & Peeters, F. J. C. (2019). Oxygen isotope composition of the final chamber of planktic foraminifera provides evidence of vertical migration and depth-integrated growth. *Biogeosciences*, *16*(2), 643–661. https://doi.org/10.5194/bg-16-643-2019

R Core Team. (2020). R: A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing. [Software]. Retrieved from https://www.R-project.org/

Robbins, L. L., & Healy-Williams, N. (1991). Toward a classification of planktonic foraminifera based on biochemical, geochemical, and morphological criteria. *Journal of Foraminiferal Research*, *21*(2), 159–167. https://doi.org/10.2113/gsjfr.21.2.159

Rosenthal, Y., Boyle, E. A., & Slowey, N. (1997). Temperature control on the incorporation of magnesium, strontium, fluorine, and cadmium into benthic foraminiferal shells from the Little Bahama Bank: Prospects for thermocline paleoceanography. *Geochimica et Cosmochimica Acta*, *61*(17), 3633–3643. https://doi.org/10.1016/s0016-7037(97)00181-6

Rustic, G. T., Polissar, P. J., Ravelo, A. C., & DeMenocal, P. (2021). Relationship between individual chamber and whole shell Mg/Ca ratios in *Trilobatus sacculifer* and implications for individual foraminifera palaeoceanographic reconstructions. *Scientific Reports*, *11*(1), 463. https://doi.org/10.1038/s41598-020-80673-8

Sadekov, A., Eggins, S. M., & Deckker, P. D. (2005). Characterization of Mg/Ca distributions in planktonic foraminifera species by electron microprobe mapping. *Geochemistry, Geophysics, Geosystems*, *6*(12). https://doi.org/10.1029/2005GC000973

Sadekov, A., Eggins, S. M., De Deckker, P., & Kroon, D. (2008). Uncertainties in seawater thermometry deriving from intratest and intertest Mg/Ca variability in *Globigerinoides ruber*. *Paleoceanography*, *23*(1), 1–12. https://doi.org/10.1029/2007PA001452

Sadekov, A., Eggins, S. M., De Deckker, P., Ninnemann, U., Kuhnt, W., & Bassinot, F. (2009). Surface and subsurface seawater temperature reconstruction using Mg/Ca microanalysis of planktonic foraminifera *Globigerinoides ruber*, *Globigerinoides sacculifer*, and *Pulleniatina obliquiloculata*. *Paleoceanography*, *24*(3), 1–17. https://doi.org/10.1029/2008PA001664

Sadekov, A. Y., Ganeshram, R., Pichevin, L., Berdin, R., McClymont, E., Elderfield, H., & Tudhope, A. W. (2013). Palaeoclimate reconstructions reveal a strong link between El Niño-Southern Oscillation and Tropical Pacific mean state. *Nature Communications*, *4*(1), 2692. https://doi.org/10.1038/ncomms3692

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, *8*(1), 289–317. https://doi.org/10.32614/rj-2016-021

Seears, H. A., Darling, K. F., & Wade, C. M. (2012). Ecological partitioning and diversity in tropical planktonic foraminifera. *BMC Evolutionary Biology*, *12*(1), 54. https://doi.org/10.1186/1471-2148-12-54

Sekimoto, S., Homura, Y., Ho, V. D., Inagaki, M., Shirai, N., & Ohtsuki, T. (2019). Neutron activation analysis of carbonate reference materials: Coral (JCp-1) and giant clam (JCt-1). *Journal of Radioanalytical and Nuclear Chemistry*, *322*(3), 1579–1583. https://doi.org/10.1007/s10967-019-06885-z

Spero, H. J. (1998). Life history and stable isotope geochemistry of planktonic foraminifera. *Paleontological Society Papers*, *4*, 7–36. https://doi.org/10.1017/s1089332600000383

Spero, H. J., Eggins, S. M., Russell, A. D., Vetter, L., Kilburn, M. R., & Hönisch, B. (2015). Timing and mechanism for intratest Mg/Ca variability in a living planktic foraminifer. *Earth and Planetary Science Letters*, *409*, 32–42. https://doi.org/10.1016/j.epsl.2014.10.030

Spero, H. J., & Lea, D. W. (1993). Intraspecific stable isotope variability in the planktic foraminifera *Globigerinoides sacculifer*: Results from laboratory experiments. *Marine Micropaleontology*, *22*(3), 221–234. https://doi.org/10.1016/0377-8398(93)90045-Y

Spero, H. J., & Lea, D. W. (1996). Experimental determination of stable isotope variability in *Globigerina bulloides*: Implications for paleoceanographic reconstructions. *Marine Micropaleontology*, *28*(3–4), 231–246. https://doi.org/10.1016/0377-8398(96)00003-5

Spero, H. J., Lerche, I., & Williams, D. F. (1991). Opening the carbon isotope vital effect black box, 2, Quantitative model for interpreting foraminiferal carbon isotope data. *Carbon Isotopic Variability in These Species*, *6*(6), 639–655. https://doi.org/10.1029/91pa02022

Spero, H. J., & Williams, D. F. (1989). Opening the carbon isotope "vital effect" black box 1. Seasonal tempratures in the euphotic zone. *Paleoceanography*, *4*(6), 593–601. https://doi.org/10.1029/pa004i006p00593

Steinke, S., Chiu, H. Y., Yu, P. S., Shen, C. C., Löwemark, L., Mii, H. S., & Chen, M. T. (2005). Mg/Ca ratios of two *Globigerinoides ruber* (white) morphotypes: Implications for reconstructing past tropical/subtropical surface water conditions. *Geochemistry, Geophysics, Geosystems*, *6*(11), 1–12. https://doi.org/10.1029/2005GC000926

Steinke, S., Groeneveld, J., Johnstone, H., & Rendle-Bühring, R. (2010). East Asian summer monsoon weakening after 7.5 Ma: Evidence from combined planktonic foraminifera Mg/Ca and δ18O (ODP Site 1146; northern South China Sea). *Palaeogeography, Palaeoclimatology, Palaeoecology*, *289*(1–4), 33–43. https://doi.org/10.1016/j.palaeo.2010.02.007

Thirumalai, K., Richey, J. N., Quinn, T. M., & Poore, R. Z. (2014). *Globigerinoides ruber* morphotypes in the Gulf of Mexico: A test of null hypothesis. *Scientific Reports*, *4*, 1–7. https://doi.org/10.1038/srep06018

Tierney, J. E., Malevich, S. B., Gray, W., Vetter, L., & Thirumalai, K. (2019). Bayesian calibration of the Mg/Ca paleothermometer in planktic foraminifera. *Paleoceanography and Paleoclimatology*, *34*(12), 2005–2030. https://doi.org/10.1029/2019PA003744

Tolderlund, D. S., Bé, A. W. H., & Be, A. W. H. (1971). Seasonal distribution of planktonic foraminifera in the Western North Atlantic. *Micropaleontology*, *17*(3), 297. https://doi.org/10.2307/1485143

Urey, H. C., Lowenstam, H. A., Epstein, S., & McKinney, C. R. (1951). Measurement of paleotemperatures and temperatures and the southeastern United States. *Bulletin of the Geological Society of America*, *62*(April), 399–416. https://doi.org/10.1130/0016-7606(1951)62[399:MOPATO]2.0.CO;2

Wang, L. (2000). Isotopic signals in two morphotypes of *Globigerinoides ruber* (white) from the South China Sea: Implications for monsoon climate change during the last glacial cycle. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *161*(3–4), 381–394. https://doi.org/10.1016/S0031-0182(00)00094-8

Wintle, B. A., McCarthy, M. A., Volinsky, C. T., & Kavanagh, R. P. (2003). The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology*, *17*(6), 1579–1590. https://doi.org/10.1111/j.1523-1739.2003.00614.x

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, *292*(5517), 686–692. https://doi.org/10.1126/science.1059412

Zeebe, R. E. (1999). An explanation of the effect of seawater carbonate concentration on foraminiferal oxygen isotopes. *Geochimica et Cosmochimica Acta*, *63*(13), 2001–2007. https://doi.org/10.1016/S0016-7037(99)00091-5

References From the Supporting Information

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2019). lme4: Linear mixed-effects models using "Eigen"*and S4*. Retrieved from https://cran.r-project.org/package=lme4

Fraley, C., Raftery, A. E., & Scrucca, L. (2019). Mclust: Gaussian mixture modelling for model-based clustering, classification, and density estimation. Retrieved from https://CRAN.R-project.org/package=mclust

Kassambara, A. (2020). Ggpubr: "ggplot2" based Publication Ready plots. Retrieved from https://CRAN.R-project.org/package=ggpubr

Wickham, H. (2019). tidyverse: Easily Install and Load the "Tidyverse. Retrieved from https://CRAN.R-project.org/package=tidyverse

Xie, Y. (2020). Knitr: A general-Purpose Package for Dynamic Report Generation in R. Retrieved from https://CRAN.R-project.org/package=knitr

Zhu, H. (2019). kableExtra: Construct Complex Table with "kable" and Pipe Syntax. Retrieved from https://CRAN.R-project.org/package=kableExtra

Received: 11 January 2024 | Accepted: 26 June 2024

DOI: 10.1111/2041-210X.14388

Methods in Ecology and Evolution

APPLICATION

3DKMI: A MATLAB package to generate shape signatures from Krawtchouk moments and an application to species delimitation in planktonic foraminifera

Huahua Lin[1] | Wenshu Zhang[1,2] | James M. Mulqueeney[3,4] | Anieke Brombacher[4,5] | Alex Searle-Barnes[4] | Mark Nixon[1] | Xiaohao Cai[1] | Thomas H. G. Ezard[4]

[1]Electronics and Computer Science, University of Southampton, Southampton, UK

[2]Creative Computing Institute, University of the Arts London, London, UK

[3]Life Sciences, Natural History Museum, London, UK

[4]Ocean and Earth Science, University of Southampton Waterfront Campus, National Oceanography Centre Southampton, Southampton, UK

[5]Earth and Planetary Sciences, Yale University, New Haven, Connecticut, USA

Correspondence
Huahua Lin
Email: huahua.lin@soton.ac.uk

Funding information
Natural Environment Research Council, Grant/Award Number: NE/P019269/1

Handling Editor: Gustavo Burin

Abstract

1. The rapid and repeatable characterization of individual morphology has advanced automated taxonomic classification. The most direct study of evolutionary processes is, however, not from taxonomic description, but rather of the evolution of the traits that comprise individuals and define species. Repeatable signatures of individual morphology are crucial for analysing the response to selection at scale, and thus tracking evolutionary trajectories through time and across species boundaries.
2. Here, we introduce our 3DKMI—an open-source MATLAB package designed for the study of morphology using three-dimensional (3D) Krawtchouk moment invariants. The volumetric features derived from the 3D images remain stable under translation, scaling and rotation and, for an image of size $128 \times 128 \times 128$ can be computed in less than 0.1 s.
3. We applied our package as a case study on a collection of 300 X-ray computed tomography scans of planktonic foraminifera specimens across five species to (1) assess the invariance of the features under different transformations and (2) analyse morphological differences among species based on the extracted characteristics.
4. We show that 3DKMI has the capacity to efficiently and repeatedly characterize the signatures of individual morphology. In the future, we hope that the 3D feature extraction technique 3DKMI will be widely applied to digital collections to advance research in ecology and evolution.

KEYWORDS
3D Krawtchouk moments, classification, computer vision, feature extraction, machine learning, planktonic foraminifera, shape, tomography

1 | INTRODUCTION

Feature extraction in computer vision has emerged as a powerful tool in the field of evolutionary biology, revolutionizing the way researchers study and analyse complex biological systems (Porto & Voje, 2020; Wäldchen & Mäder, 2018). By employing various algorithms and techniques, computer vision systems can provide a high-level understanding of visual data (Porto et al., 2021) to automatically detect and analyse morphological features, such as the shape (Cooney et al., 2017; Spampinato et al., 2010), size (Alsmadi et al., 2010; Hsiang et al., 2019), and colour patterns (Alsmadi et al., 2011; Cooney et al., 2019) of organisms. Through extracting and quantifying these features, researchers can conduct comparative analysis across different individuals, species, populations and communities, providing valuable insights into the evolutionary processes and adaptations that shape the diversity of life on Earth.

One common application of feature extraction is the problem of automatic classification of individuals to species (Fassnacht et al., 2016; Spampinato et al., 2010). The individuals grouped into self-similar clusters, and then often mapped back to manual taxonomic classifications as species, comprise distinctive sets of genetic (Pons et al., 2006; Talavera et al., 2013) or phenotypic traits (Pearson & Ezard, 2014). Bridging the gap from traditional taxonomy to automated computer vision pivots on the characterization of those sets of phenotypic traits that define species, but assessing which traits carry the most weight in determining species classifications is often subsumed within a "black box" rather than the core output of model classifications.

Deeper morphological understanding promises substantial potential to improve our understanding of phenotypic evolution (Felice et al., 2018). However, direct morphological evidence of how biological diversity came to exhibit its strikingly discontinuous pattern is not complete (Rolland et al., 2023). Rapid and repeatable methods for morphometric data extraction would therefore boost the study of phenotypic evolution. Geometric morphometric approaches have become the gold standard in evolutionary biology due to the limitations of linear methods in distinguishing complex morphologies (Viacava et al., 2023; Zelditch et al., 2012), providing fuller insights into how morphology relates to underlying functions (Fabre et al., 2014) and developmental processes (Goswami et al., 2014), while also offering less biased descriptors (Zelditch et al., 2012). Despite recent advances, such as enhancing shape characterization with sliding semi-landmarks (Bardua et al., 2019), manual work is susceptible to observer bias, lacks repeatability and is time-consuming. While using automated approaches including automated landmarking (Devine et al., 2020) and landmark-free methods (Mulqueeney et al., 2024) may help overcome these issues, they still require homology for either analysis or alignment, and thus become less effective as more disparate shapes are compared (Collyer et al., 2015).

Image moments, initially conceptualized by Hu (1962), serve as mathematical descriptors representing the spatial distribution of points in an image, widely employed to encapsulate global features such as an object's shape, thus eliminating the necessity for landmarking. Following this seminal work, a variety of new moments (Mukundan et al., 2001; Teague, 1980; Yap et al., 2003) based on different polynomials rapidly emerged. Among them, Krawtchouk moments stand out because the basis functions are orthogonal in the discrete domain of the image coordinate space, thus minimizing feature redundancy and eliminating the need for numerical approximation (Yap et al., 2003). Other advantages of using Krawtchouk moments for feature description over deep learning methods lie in their simplicity, efficiency and interpretability, providing a straightforward mathematical framework for representing image features without requiring comparisons of homologous points, extensive computational resources, or large amounts of labelled data.

Here, we describe our developed 3DKMI package, a MATLAB package designed for the automatic generation of shape features. Our package implements the code for computing three-dimensional (3D) Krawtchouk moment invariants, derived from discrete and orthogonal 3D Krawtchouk polynomials. These invariants can consistently describe the individual regardless of translation, scaling and rotation. Unlike previous studies (Benouini et al., 2018; Sit et al., 2019), our novel in-house experiment focuses on a 3D dataset of planktonic foraminifera species. To the best of our knowledge, we are the first to make this code publicly available (https://github.com/krawtchouk/Krawtchouk-Moment-Invariants) and apply it to a 3D ecological dataset, complemented by a quick-start guide and examples available at http://krawtchouk.github.io.

2 | PACKAGE FEATURES

3DKMI is written in MATLAB, version R2022b, with backward compatibility. The program's dependencies include the "Symbolic Math Toolbox" and the "Image Processing Toolbox". Below, we present the two primary functions—readImage and get3DKMI—within the package.

2.1 | readImage: Construct a 3D array

The function readImage takes two input parameters: the path where the 3D image data is stored along with the desired image dimensions dim in integer format. It produces a corresponding 3D array that represents the input image and is called with the following command:

```
readImage(path, dim)
```

The 3D image will be resized to the given desired dimensions dim so that it can be fitted into a cube formed by the desired dimensions without disrupting its geometry. It is important to use consistent desired dimensions dim when comparing the features extracted from different 3D images.

The path accommodates two distinct scenarios: (i) it can either point to a folder containing a sequence of two-dimensional (2D) image slices, where each is stored as for example an individual .png file or (ii) it can direct to a 3D image file. In the former scenario, the method processes these image slices sequentially, arranging them based on the

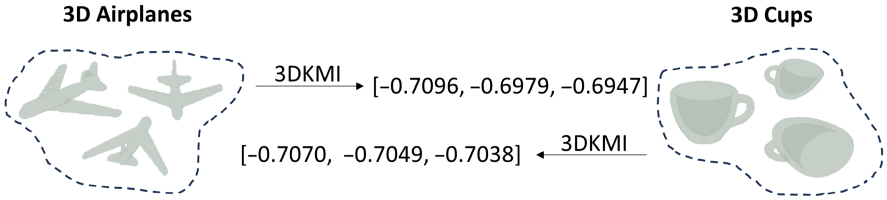


FIGURE 1 Two different types of three-dimensional (3D) shapes and their corresponding features represented by 3D Krawtchouk moments.

filename order (e.g. 0.png, 1.png, …, n.png). This aggregation results in an image stack, ultimately manifested as a 3D array representation. In the latter scenario, our current implementation supports reading .im and .stl formats for 3D image files and generating voxelized data. Reading .im files follows the process provided on McGill's 3D shape benchmark website (Siddiqi et al., 2008). For .stl files, the existing MATLAB built-in functions can only extract the vertex and face information from the image. Therefore, we employ a voxelization method to further derive the 3D array (Adam, 2023). If users find that their file types are unsupported, they can generate their own 3D arrays representing images and then directly use the get3DKMI function.

2.2 | get3DKMI: Feature extraction

The function get3DKMI takes a 3D array arr as input, along with the desired dimensions dim in integer format and produces 3D Krawtchouk moment invariants as output. The function is called using the following command:

get3DKMI(arr, dim, varargin)

where varargin allows the function to accept up to two input parameters: POI (point-of-interest) and/or the order of the 3D Krawtchouk moments. Notably, providing a POI as input implies that the user expects to extract localized features from the image, whereas the absence of this input results in the extraction of features from the image as a whole.

The distinction between local and global feature extraction lies in how weights are assigned to individual pixels within the image. The weight function in Yap et al. (2003) and Sit et al. (2019) was introduced to create a more stable set of polynomials, but we found it to be an integral part of extracting local features around the POI. In the case of local extraction, those pixels around the POI can hold greater significance than the rest of the region by using the weight function. The user can also select the moment's order. The order of moments denotes the mathematical degree of the moments being computed, with different orders capturing distinct aspects of an image's characteristics. A higher order yields a more detailed description of the local or global image, though it requires more computational resources. The length of the output invariants depends on the order of the 3D Krawtchouk moments.

The parameter dim is also included in get3DKMI in case the users themselves generate the 3D array that needs to be resized. Additionally, the spherical shape of the interested local region can be preserved when the image is in the 3D cubic formed by the desired dimensions dim. Once the image dimensions change, the relative positions of the POI will be recalculated accordingly to ensure the features are still extracted around the interested region. However, the size of the image influences the program's computation time, that is larger dimensions result in increased time spent on matrix computations. Our benchmark computation time is provided in Section 3.2. We suggest users experiment with various values according to their hardware capabilities. Detailed mathematical instructions for obtaining invariants through Krawtchouk moments can be found on our GitHub website http://krawtchouk.github.io.

2.3 | How 3DKMI works

To illustrate the functionality of 3DKMI, we selected two distinct 3D shapes from the McGill 3D shape benchmark (Siddiqi et al., 2008): an airplane and a cup as shown in Figure 1. We employed our package to extract their global features, and the calculated second-order invariant values are [−0.7096, −0.6979, −0.6947] for the airplanes and [−0.7070, −0.7049, −0.7038] for the cups. Despite the substantial dissimilarity between the shapes of the airplane and the cup, the invariant values exhibit a small absolute difference between the calculated invariant values of the two shapes, which is a characteristic of the 3D Krawtchouk moments. Note that, at the scale of the feature space generated by the 3D Krawtchouk moments, that small absolute difference is not small for 3DKMI and is sufficient to determine the difference between three airplane images and three cup images in Figure 1. The feature distance between the original airplane image and the transformed airplane image is smaller than that between the original airplane image and the cup image. Likewise, the values of the local features exhibit similar behaviour, allowing us to view the local region as a novel shape. In short, 3DKMI can represent each different 3D shape with a unique signature.

3 | RESULTS: APPLICATIONS ON SPECIES DIFFERENCES IN PLANKTONIC FORAMINIFERA

3.1 | Data

Modern high-resolution X-ray computed tomography (CT) makes it possible for us to observe and analyse the internal and external structures of planktonic foraminifera that are less than 1 mm in

diameter. We compiled a dataset of 300 individuals over five species from Ocean Drilling Program Site 925, located on the Ceara Rise in the Equatorial Atlantic, and Integrated Ocean Drilling Program Site U1313 within the North Atlantic gyre; for site-specific oceanographic settings, see Brombacher et al. (2021) and references therein. The five species are *Globigerinoidesella fistulosa*, *Menardella exilis*, *Truncorotalia crassaformis*, *Sphaeroidinella dehiscens* and *Trilobatus sacculifer*, with some sample images shown in Figure 2. Each species has a sample size of 50, except for *Menardella exilis*, which has an additional 50 samples for efficiency benchmarking. One of us (A.B.) generated taxonomic labels through key diagnostic features, including the shape and arrangement of the chambers, the shape and position of the apertures, and other morphological features of the shells. Each individual was randomly drawn, within the strata of its species classification, from a larger sample of 1226 specimens.

For each individual, approximately 300 2D image slices were obtained by CT scanning. As our emphasis was solely on morphology, we obtained binary volume representations of each specimen from these slices by applying some pre-processing steps such as cropping and segmentation (Figure 3). Global threshold values for segmentation are manually determined to remove background noise while preserving shape integrity. This step can be automated by using threshold values selected automatically by, for example Otsu's method (Otsu, 1975).

3.2 | Implementation efficiency

We utilized 100 specimens from the species *M. exilis* to benchmark the efficiency of 3DKMI in feature extraction, running on a laptop equipped with an AMD Ryzen 5 5600H CPU and 16 GB of RAM. For parameter selection, each image was resized to $128 \times 128 \times 128$ (i.e. dim is set to 128) and the order of moments was fixed to 2. The entire workflow after pre-processing for these 100 *M. exilis*, including image reading and feature extraction, took ~58 s, with the majority of the time dedicated to reading slices to form the 3D array data. The average runtime for each sample was 0.48 s for forming the 3D array and 0.09 s for calculating the 3D Krawtchouk moment invariants.

3.3 | Translation, scale and rotation invariance

In this experiment, we investigated the property of the 3D Krawtchouk moment invariants on the dataset under different geometric transformations, including translation, scaling, and rotation. The 3D Krawtchouk moment invariants were computed up to the second order. For comparison, the 3D images were resized to cubes of the same dimension, that is $128 \times 128 \times 128$. Both global and local features were tested. In our case, we selected the image's centroid as the POI.

One specimen from the species *M. exilis* was randomly selected for demonstration purposes. Three transformations were applied separately to analyse their impact on invariants. We first investigated the translation invariance by shifting the object within the cube with a step of (1, 1, 1) using translation vectors ranging from (−10, −10, −10) to (10, 10, 10). To assess the scale invariance, we applied scaling factors from 0.7 to 1.3 with a step size of 0.05 to the object. Finally, the rotation invariance was explored by rotating the object around each axis with rotation angles ranging from 0° to 90°, at intervals of 10°. To preserve the object's integrity, we scaled down any transformed object that exceeds the dimensions of the cube.

Figure 4 depicts the mean squared error (MSE) between the 3D Krawtchouk moment invariants of the original object and the ones under translation, scaling and rotation. The overall results in

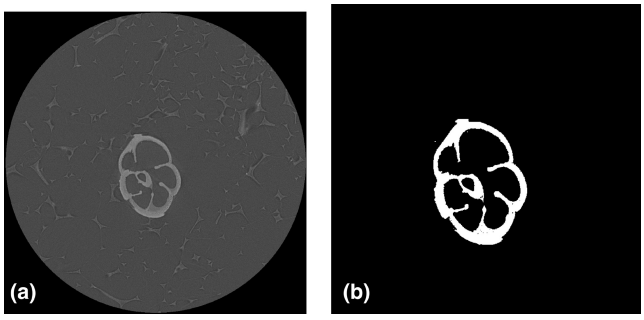


FIGURE 3 (a) A two-dimensional (2D) slice of a single *Menardella exilis* specimen reconstructed by X-ray computed tomography (Zeiss Reconstructor v14) and (b) the same 2D slice represented as a binary image after cropping and segmentation using a manually specified threshold value.

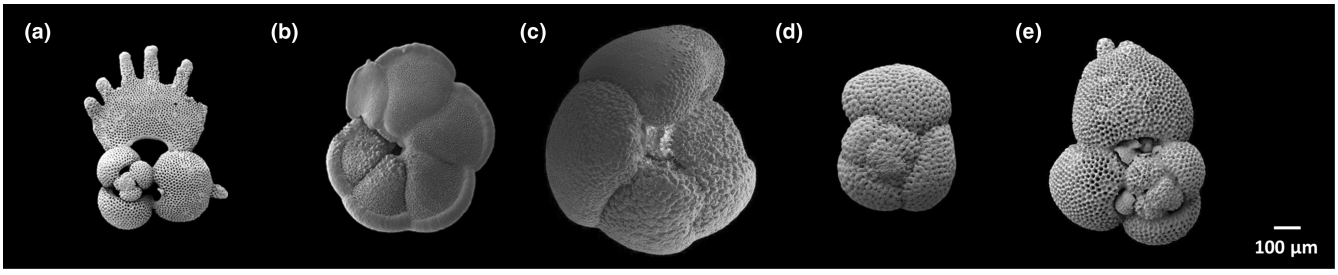


FIGURE 2 Example images of the five different species of planktonic foraminifera used in this application. (a) *Globigerinoidesella fistulosa* (Poole & Wade, 2019), (b) *Menardella exilis* (Woodhouse et al., 2023), (c) *Truncorotalia crassaformis* (Lam & Leckie, 2020a), (d) *Sphaeroidinella dehiscens* (Lam & Leckie, 2020b) and (e) *Trilobatus sacculifer* (Poole & Wade, 2019). The scalar bar for images is 100 µm.

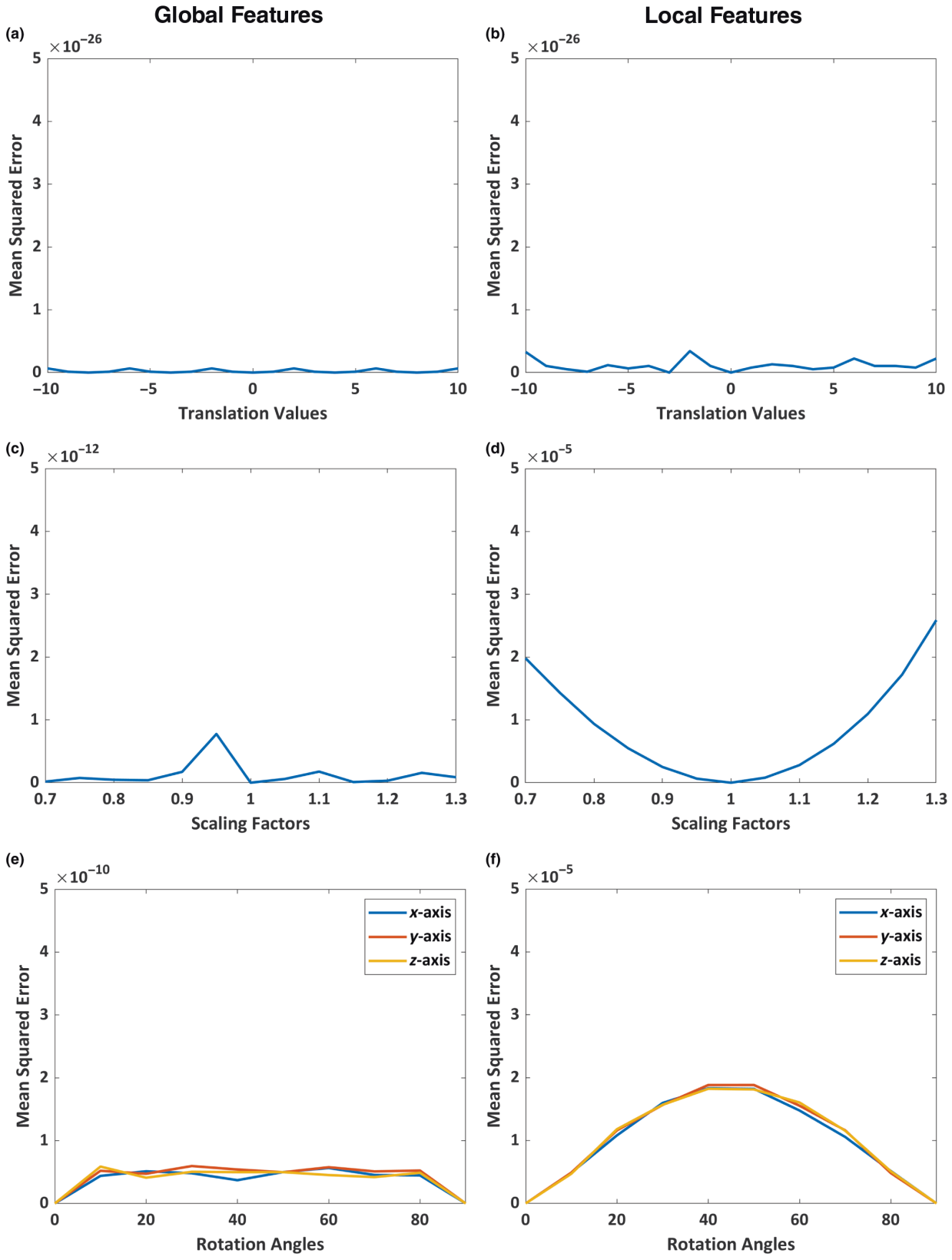


FIGURE 4 Invariance performance validation of 3DKMI. The left and right columns show the mean squared error between the original object and the transformed ones regarding global features and local features, respectively. Three transformations are involved in total: (a, b) when the object is shifted (translation); (c, d) when the size of the object changes (scale); and (e, f) when the object rotates about the *x*-axis, *y*-axis and *z*-axis (rotation). All transformed objects still fit the size of the original image. Note the different scales on the *y*-axes.

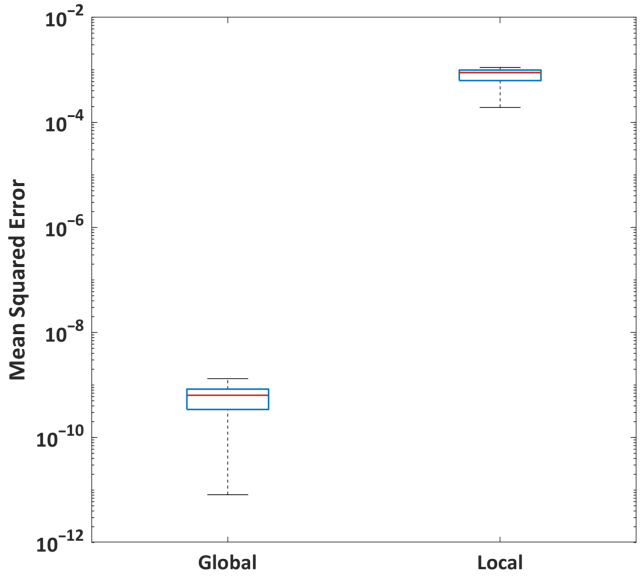


FIGURE 5 Invariance performance validation of 3DKMI when three random transformations (including translation, scaling, and rotation) are performed simultaneously. The mean squared error (MSE) between the 3D Krawtchouk moment invariants of the original object and the transformed one comes from 50 random transformations; the median MSE using global features is $6e^{-10}$ while the one using local features is $8e^{-4}$.

Figure 4 indicate consistently low errors regardless of the different geometric transformations applied to the given object. For the comparison between the two feature extraction manners, that is global and local, the error of the former is always smaller than that of the latter. In the global scaling and rotation experiments, the MSE values are relatively high compared to the translation experiments because when a 3D image is scaled and rotated, if the voxels in the transformed image are in non-integer positions, interpolation will be used to map the voxels in the original image to their new positions in the transformed image. This small change will slightly affect the extracted features, but the MSE values are still less than $1e^{-10}$. The experimental results also show that local feature extraction is most sensitive to scaling because when the object is scaled, the weight function remains to assign higher weights to the same size region of interest. Additionally, as shown in the right panel of Figure 4f, the error exhibits symmetry when the object is rotated from 0° to 90° around the axis, with the largest error occurring at 45° of rotation. This angle of rotation causes the object to exceed the cube dimension the most, and the object is thus at the same time scaled down the most to ensure its integrity.

Furthermore, we randomly and simultaneously applied the aforementioned three transformations to validate the invariance property of 3D Krawtchouk moment invariants under more

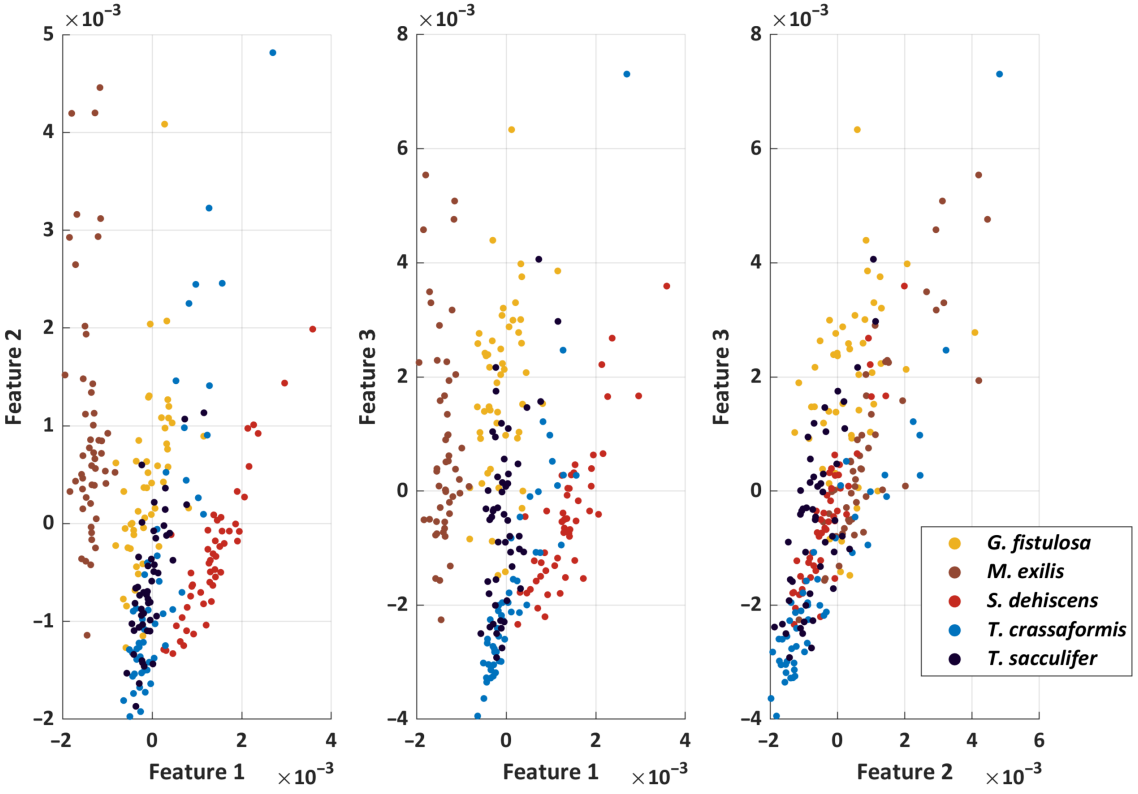


FIGURE 6 Two-dimensional projections on each axis of the Krawtchouk moment invariants for 250 specimens represented by 3DKMI.

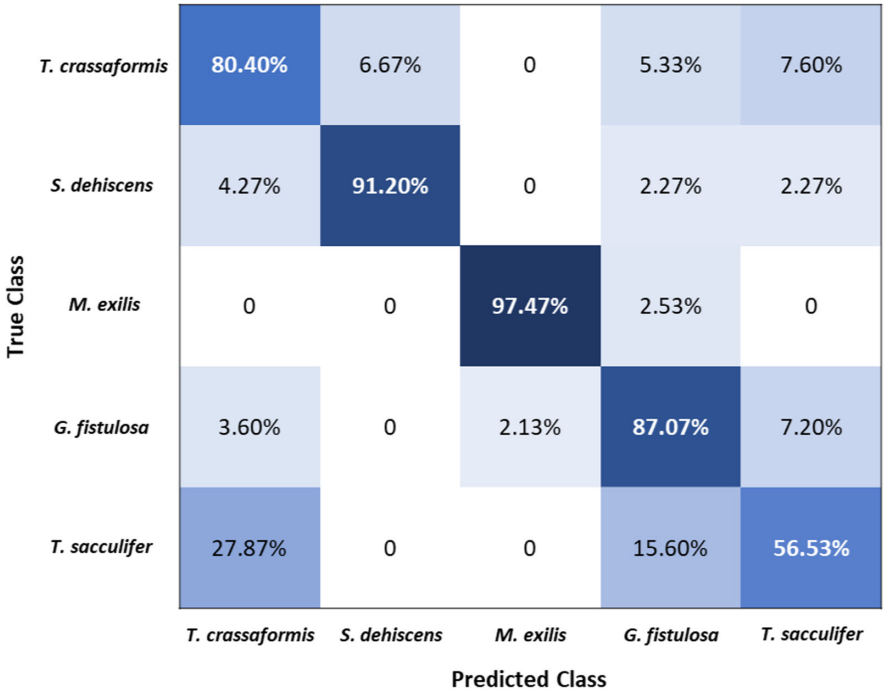


FIGURE 7 The mean confusion matrix of the classification performance from the 50 randomized partitions using the features extracted by 3DKMI. It shows that the samples belonging to the species *Menardella exilis* were nearly perfectly distinguished from the samples of other species, whereas the samples of the species *Trilobatus sacculifer* were not.

challenging conditions. Figure 5 summarizes the MSE between the 3D Krawtchouk moment invariants of the original object and the ones under random transformations. The variation in error arises from 50 random transformations, which were tested separately for both global and local settings. Despite the application of random transformations to the given object, both global and local features exhibit low error. However, it is worth noting that the error associated with local features is higher, with a median of $8e^{-4}$ compared to $6e^{-10}$ observed with global features. This difference is still related to the sensitivity of local features to scaling.

3.4 | Classification

We now show 3DKMI's power in exploiting the signatures of individual morphology to classify various species. All images were resized uniformly to a resolution of $128\times128\times128$ and were described globally using the second order of the moments. Figure 6 shows the Krawtchouk moment morphospace filled by the randomly selected 250 individuals across the five species.

Subsequently, the formed dataset of these three features was partitioned into training and test sets, representing 70% and 30% of the data, respectively. The support vector machine (SVM) classifier, a classic method that aims to maximize the margin between classes and is robust to outliers, was employed to perform the classification task using the features extracted by 3DKMI. In detail, the SVM was fitted to the training set initially to learn the underlying patterns within the data, and then the trained SVM was applied to predict the classes of each sample in the test set. In experiments with 50 random partitions, classification accuracies on the test set ranged from 76% to 91%, with a median accuracy of 83%. Similar results were obtained with a k-nearest neighbours classifier ($k=5$). Figure 7 shows the mean confusion matrix generated from these 50 partitions, revealing the percentage of both correct and incorrect predictions for each class. By analysing the confusion matrix, it is evident that the classifier almost perfectly predicted samples belonging to the species *M. exilis*, and yet exhibited some confusion when classifying samples from the species *T. sacculifer*. *G. fistulosa* evolved directly from the *T. sacculifer* group of morphospecies in the late Pliocene (King et al., 2020); these two species differ through the existence of elongated protuberances in the final life stages. While some confusion in classification is unsurprising given that *T. sacculifer* and *T. crassaformis* share similar aspect ratios and our random sampling regime did not prioritize the most obvious protuberances, a major source of this confusion is the preservation (or not) of internal structures (Figure 3).

4 | CONCLUSIONS

Morphometric characterizations of individuals have emerged as a crucial tool to quantitatively compare biological forms. Automatic techniques that can effectively derive repeatable features to represent individuals have thus assumed significant roles in ecology and evolution. In this study, we developed a MATLAB package, 3DKMI, designed for automated feature extraction from any 3D image. Our package relies on the utilization of 3D Krawtchouk moment invariants, which are well-established tools in computer vision as they are renowned for being repeatable and transformation invariant, making them ideal for shape representation. Our experiments on this planktonic foraminifera dataset suggest that 3D Krawtchouk moment invariants can differentiate among traditionally delimited

species (Figures 6, 7), and the performance would increase further by incorporating other traits, particularly size, which is a clear differentiator between *T. crassaformis* and *T. sacculifer*, when building morphospaces.

The 3DKMI package possesses many distinctive advantages. It requires only a consumer-grade laptop for extremely fast and reproducible implementation and is based on an interpretable mathematical framework that efficiently computes the geometric features of an object directly from the input of a single image without the need to spend time and resources training a deep learning model on a dataset. The extracted features are repeatable and independent of observer bias. 3DKMI is also part of a growing effort to develop new methods that eliminate the reliance on landmarks and homology, contributing new translation, scaling, and rotation invariant tools for fast and repeatable extraction of phenotypic data from 3D volumetric images. We expect that 3DKMI will greatly expand the scope of morphometric research by enabling rapid comparisons of continually more disparate shapes.

AUTHOR CONTRIBUTIONS

Huahua Lin, Wenshu Zhang, Mark Nixon, Xiaohao Cai and Thomas H. G. Ezard conceived the ideas and designed the methodology; Anieke Brombacher, Alex Searle-Barnes and James M. Mulqueeney collected the data; Huahua Lin and Wenshu Zhang analysed the data; Huahua Lin led the writing of the manuscript, which was iterated with Xiaohao Cai and Thomas H. G. Ezard. All authors contributed critically to the drafts and gave final approval for publication.

FUNDING INFORMATION

This work was funded by the Natural Environment Research Council Award NE/P019269/1.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14388.

DATA AVAILABILITY STATEMENT

Data available from the Dryad Digital Repository https://doi.org/10.5061/dryad.66t1g1k93 (Lin et al., 2024).

ORCID

Huahua Lin https://orcid.org/0000-0002-1600-6293
Wenshu Zhang https://orcid.org/0000-0003-0776-7709
James M. Mulqueeney https://orcid.org/0000-0003-3502-745X
Anieke Brombacher https://orcid.org/0000-0003-2310-047X
Alex Searle-Barnes https://orcid.org/0000-0003-0389-7717
Mark Nixon https://orcid.org/0000-0002-9174-5934
Xiaohao Cai https://orcid.org/0000-0003-0924-2834
Thomas H. G. Ezard https://orcid.org/0000-0001-8305-6605

REFERENCES

Adam, A. (2023). *Mesh voxelisation*. https://www.mathworks.com/matlabcentral/fileexchange/27390-mesh-voxelisation. MATLAB Central File Exchange.

Alsmadi, M. K., Omar, K. B., & Noah, S. A. (2011). Fish classification based on robust features extraction from color signature using back-propagation classifier. *Journal of Computer Science*, *7*(1), 52–58.

Alsmadi, M. K., Omar, K. B., Noah, S. A., & Almarashdeh, I. (2010). Fish recognition based on robust features extraction from size and shape measurements using neural network. *Journal of Computer Science*, *6*(10), 1088–1094.

Bardua, C., Felice, R. N., Watanabe, A., Fabre, A.-C., & Goswami, A. (2019). A practical guide to sliding and surface semilandmarks in morphometric analyses. *Integrative Organismal Biology*, *1*(1), obz016.

Benouini, R., Batioua, I., Zenkouar, K., Najah, S., & Qjidaa, H. (2018). Efficient 3D object classification by using direct krawtchouk moment invariants. *Multimedia Tools and Applications*, *77*, 27517–27542.

Brombacher, A., Wilson, P. A., Bailey, I., & Ezard, T. H. (2021). The dynamics of diachronous extinction associated with climatic deterioration near the neogene/quaternary boundary. *Paleoceanography and Paleoclimatology*, *36*(6), e2020PA004205.

Collyer, M. L., Sekora, D. J., & Adams, D. C. (2015). A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity*, *115*(4), 357–365.

Cooney, C. R., Bright, J. A., Capp, E. J., Chira, A. M., Hughes, E. C., Moody, C. J., Nouri, L. O., Varley, Z. K., & Thomas, G. H. (2017). Mega-evolutionary dynamics of the adaptive radiation of birds. *Nature*, *542*(7641), 344–347.

Cooney, C. R., Varley, Z. K., Nouri, L. O., Moody, C. J., Jardine, M. D., & Thomas, G. H. (2019). Sexual selection predicts the rate and direction of colour divergence in a large avian radiation. *Nature Communications*, *10*(1), 1773.

Devine, J., Aponte, J. D., Katz, D. C., Liu, W., Vercio, L. D. L., Forkert, N. D., Marcucio, R., Percival, C. J., & Hallgrímsson, B. (2020). A registration and deep learning approach to automated landmark detection for geometric morphometrics. *Evolutionary Biology*, *47*(3), 246–259.

Fabre, A.-C., Goswami, A., Peigné, S., & Cornette, R. (2014). Morphological integration in the forelimb of musteloid carnivorans. *Journal of Anatomy*, *225*(1), 19–30.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, *186*, 64–87.

Felice, R. N., Randau, M., & Goswami, A. (2018). A fly in a tube: Macroevolutionary expectations for integrated phenotypes. *Evolution*, *72*(12), 2580–2594.

Goswami, A., Smaers, J. B., Soligo, C., & Polly, P. D. (2014). The macroevolutionary consequences of phenotypic integration: From development to deep time. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *369*(1649), 20130254.

Hsiang, A. Y., Brombacher, A., Rillo, M. C., Mleneck-Vautravers, M. J., Conn, S., Lordsmith, S., Jentzen, A., Henehan, M. J., Metcalfe, B., Fenton, I. S., Wade, B. S., Fox, L., Meilland, J., Davis, C. V., Baranowski, U., Groeneveld, J., Edgar, K. M., Movellan, A., Aze, T., … Hull, P. M. (2019). Endless forams: >34,000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks. *Paleoceanography and Paleoclimatology*, *34*(7), 1157–1177.

Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, *8*(2), 179–187.

King, D. J., Wade, B. S., Liska, R. D., & Miller, C. G. (2020). A review of the importance of the caribbean region in oligo-miocene low latitude

planktonic foraminiferal biostratigraphy and the implications for modern biogeochronological schemes. *Earth-Science Reviews*, *202*, 102968.

Lam, A. R., & Leckie, R. M. (2020a). Late neogene and quaternary diversity and taxonomy of subtropical to temperate planktic foraminifera across the kuroshio current extension, northwest pacific ocean. *Micropaleontology*, *66*(3), 177–268.

Lam, A. R., & Leckie, R. M. (2020b). Subtropical to temperate late neogene to quaternary planktic foraminiferal biostratigraphy across the kuroshio current extension, shatsky rise, northwest pacific ocean. *PLoS One*, *15*(7), e0234351.

Lin, H., Zhang, W., Mulqueeney, J. M., Brombacher, A., Searle-Barnes, A., Nixon, M., Cai, X., & Ezard, T. H. G. (2024). Data from: 3DKMI: A MATLAB package to generate shape signatures from Krawtchouk moments and an application to species delimitation in planktonic foraminifera. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.66t1g1k93

Mukundan, R., Ong, S., & Lee, P. A. (2001). Image analysis by tchebichef moments. *IEEE Transactions on Image Processing*, *10*(9), 1357–1364.

Mulqueeney, J. M., Ezard, T. H., & Goswami, A. (2024). Assessing the application of landmark-free morphometrics to macroevolutionary analyses. *bioRxiv*. 2024-04.

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, *11*(285–296), 23–27.

Pearson, P. N., & Ezard, T. H. (2014). Evolution and speciation in the eocene planktonic foraminifer turborotalia. *Paleobiology*, *40*(1), 130–143.

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., & Vogler, A. P. (2006). Sequence-based species delimitation for the dna taxonomy of undescribed insects. *Systematic Biology*, *55*(4), 595–609.

Poole, C. R., & Wade, B. S. (2019). Systematic taxonomy of the *Trilobatus sacculifer* plexus and descendant *Globigerinoidesella fistulosa* (planktonic foraminifera). *Journal of Systematic Palaeontology*, *17*(23), 1989–2030.

Porto, A., Rolfe, S., & Maga, A. M. (2021). Alpaca: A fast and accurate computer vision approach for automated landmarking of three-dimensional biological structures. *Methods in Ecology and Evolution*, *12*(11), 2129–2144.

Porto, A., & Voje, K. L. (2020). Ml-morph: A fast, accurate and general approach for automated detection and landmarking of biological structures in images. *Methods in Ecology and Evolution*, *11*(4), 500–512.

Rolland, J., Henao-Diaz, L. F., Doebeli, M., Germain, R., Harmon, L. J., Knowles, L. L., Liow, L. H., Mank, J. E., Machac, A., Otto, S. P., Pennell, M., Salamin, N., Silvestro, D., Sugawara, M., Uyeda, J., Wagner, C. E., & Schluter, D. (2023). Conceptual and empirical bridges between micro- and macroevolution. *Nature Ecology & Evolution*, *7*, 1181–1193.

Siddiqi, K., Zhang, J., Macrini, D., Shokoufandeh, A., Bouix, S., & Dickinson, S. (2008). Retrieving articulated 3-D models using medial surfaces. *Machine Vision and Applications*, *19*, 261–275.

Sit, A., Shin, W.-H., & Kihara, D. (2019). Three-dimensional krawtchouk descriptors for protein local surface shape comparison. *Pattern Recognition*, *93*, 534–545.

Spampinato, C., Giordano, D., Di Salvo, R., Chen-Burger, Y.-H. J., Fisher, R. B., & Nadarajan, G. (2010). Automatic fish classification for underwater species behavior understanding. *Proceedings of the First ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams*. New York, NY, USA, pp. 45–50.

Talavera, G., Dincă, V., & Vila, R. (2013). Factors affecting species delimitations with the gmyc model: Insights from a butterfly survey. *Methods in Ecology and Evolution*, *4*(12), 1101–1110.

Teague, M. R. (1980). Image analysis via the general theory of moments. *Journal of the Optical Society of America*, *70*(8), 920–930.

Viacava, P., Blomberg, S. P., & Weisbecker, V. (2023). The relative performance of geometric morphometrics and linear-based methods in the taxonomic resolution of a mammalian species complex. *Ecology and Evolution*, *13*(3), e9698.

Wäldchen, J., & Mäder, P. (2018). Machine learning for image based species identification. *Methods in Ecology and Evolution*, *9*(11), 2216–2225.

Woodhouse, A., Procter, F. A., Jackson, S. L., Jamieson, R. A., Newton, R. J., Sexton, P. F., & Aze, T. (2023). Paleoecology and evolutionary response of planktonic foraminifera to the mid-pliocene warm period and plio-pleistocene bipolar ice sheet expansion. *Biogeosciences*, *20*(1), 121–139.

Yap, P.-T., Paramesran, R., & Ong, S.-H. (2003). Image analysis by krawtchouk moments. *IEEE Transactions on Image Processing*, *12*(11), 1367–1377.

Zelditch, M. L., Swiderski, D. L., & Sheets, H. D. (2012). *Geometric morphometrics for biologists: A primer*. Academic Press.

How to cite this article: Lin, H., Zhang, W., Mulqueeney, J. M., Brombacher, A., Searle-Barnes, A., Nixon, M., Cai, X., & Ezard, T. H. G. (2024). 3DKMI: A MATLAB package to generate shape signatures from Krawtchouk moments and an application to species delimitation in planktonic foraminifera. *Methods in Ecology and Evolution*, *15*, 1940–1948. https://doi.org/10.1111/2041-210X.14388

ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Mulqueeney JM, Searle-Barnes A, Brombacher A, Sweeney M, Goswami A, Ezard THG. 2024 How many specimens make a sufficient training set for automated three-dimensional feature extraction? *R. Soc. Open Sci.* **11**: rsos.240113.
https://doi.org/10.1098/rsos.240113

Received: 18 January 2024
Accepted: 25 April 2024

Subject Category:
Organismal and evolutionary biology

Subject Areas:
artificial intelligence, evolution, palaeontology

Keywords:
deep learning, data augmentation, image segmentation, planktonic foraminifera, feature extraction

Author for correspondence:
James M. Mulqueeney
e-mails: j.m.mulqueeeney@soton.ac.uk; james.mulqueeney1@nhm.ac.uk

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7288922.

THE ROYAL SOCIETY PUBLISHING

How many specimens make a sufficient training set for automated three-dimensional feature extraction?

James M. Mulqueeney[1,2], Alex Searle-Barnes[1], Anieke Brombacher[1], Marisa Sweeney[1], Anjali Goswami[2] and Thomas H. G. Ezard[1]

[1]School of Ocean & Earth Science, National Oceanography Centre Southampton, University of Southampton Waterfront Campus, Southampton, UK
[2]Department of Life Sciences, Natural History Museum, London, UK

JMM, 0000-0003-3502-745X; AS-B, 0000-0003-0389-7717; AB, 0000-0003-2310-047X; AG, 0000-0001-9465-810X; THGE, 0000-0001-8305-6605

Deep learning has emerged as a robust tool for automating feature extraction from three-dimensional images, offering an efficient alternative to labour-intensive and potentially biased manual image segmentation methods. However, there has been limited exploration into the optimal training set sizes, including assessing whether artficial expansion by data augmentation can achieve consistent results in less time and how consistent these benefits are across different types of traits. In this study, we manually segmented 50 planktonic foraminifera specimens from the genus *Menardella* to determine the minimum number of training images required to produce accurate volumetric and shape data from internal and external structures. The results reveal unsurprisingly that deep learning models improve with a larger number of training images with eight specimens being required to achieve 95% accuracy. Furthermore, data augmentation can enhance network accuracy by up to 8.0%. Notably, predicting both volumetric and shape measurements for the internal structure poses a greater challenge compared with the external structure, owing to low contrast differences between different materials and increased geometric complexity. These results provide novel insight into optimal training set sizes for precise

image segmentation of diverse traits and highlight the potential of data augmentation for enhancing multivariate feature extraction from three-dimensional images.

1. Introduction

Three-dimensional imaging techniques, such as X-ray micro-computed tomography (micro-CT), have revolutionized the characterization of both internal and external structures of diverse objects. With the ability to generate high-resolution images, researchers can visualize and quantify intricate three-dimensional features with wide-ranging applications [1,2]. However, objective extraction of these features remains a major challenge.

Currently, the prevailing approach to extracting three-dimensional features from image data involves manual segmentation, which is a labour-intensive [3] and subjective process that lacks reproducibility [4]. This manual approach limits the study of a large number of samples and the exploration of complex hypotheses. As the acquisition of high-resolution scans has increased steadily [5,6], there is a pressing need to enhance the efficiency of this important processing step.

Machine learning techniques, particularly deep learning, and convolutional neural networks (CNNs) offer a promising solution for automating image segmentation. These methods can potentially accelerate this processing step to deliver accurate and repeatable results, while being accessible to various research fields [7,8]. However, despite their merits, a fundamental trade-off exists between the quantity of samples necessary for generating accurate neural networks and the time-consuming, subjective nature of manual segmentation for evaluation. The specific number of required samples is likely to vary across datasets and depends on the traits being extracted. Hence, there is an imperative to establish the minimum number of manually segmented specimens needed to train these neural networks and to understand how they vary in different trait extraction scenarios.

Strategies to generate training sets that reduce manual processing while maintaining performance are required. Smart interpolation, whereby pre-segmented slices and the volumetric image data are used to predict segmentation across the entire specimen, is one suggested approach [9,10], but this remains time-consuming for creating large training sets. Data augmentation, which artificially expands the size of the training set without collecting new data, can mitigate overfitting and bolster the accuracy of CNNs during training [11,12]. Consequently, data augmentation techniques may allow smaller training sets to achieve equivalent accuracy levels of their larger counterparts.

In this study, we aim to determine the minimum number of images needed to train a neural network to produce segmentation data that is statistically indistinguishable from manually generated data. To do so, we use a dataset of computed tomography (CT) scans of planktonic foraminifera from the genus *Menardella*. We assess the efficacy of each training set in extracting volumetric and shape data for the external calcite and internal chamber space of selected specimens (see figure 1).

Additionally, we introduce a novel three-dimensional data augmentation technique to bolster training sets by generating six different orientations of each specimen through rotation. This comparison serves to assess how data augmentation strategies can improve training sets to achieve accurate and efficient three-dimensional feature extraction.

2. Material and methods

2.1. Data collection

Fifty planktonic foraminifera, comprising 4 *Menardella menardii*, 17 *Menardella limbata*, 18 *Menardella exilis* and 11 *Menardella pertenuis* specimens, were used in our analyses (electronic supplementary material, figures S1 and S2). The taxonomic classification of these species was established based on the analysis of morphological characteristics observed in their shells. In this context, all species are characterized by lenticular, low trochospiral tests with a prominent keel [13]. Discrimination among these species is achievable, as *M. limbata* can be distinguished from its ancestor, *M. menardii*, by having a greater number of chambers and a smaller umbilicus. Moreover, *M. exilis* and *M. pertenuis* can be discerned from *M. limbata* by their thinner, more polished tests and reduced trochospirality. Furthermore, *M. pertenuis* is identifiable by a thin plate extending over the umbilicus and possessing a greater number of chambers in the final whorl compared with *M. exilis* [13].

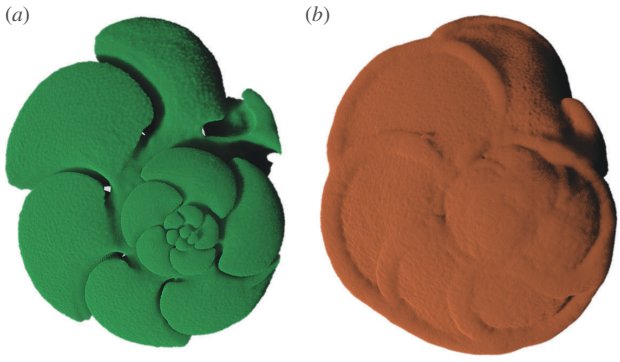


Figure 1. Three-dimensional models of the (*a*) internal and (*b*) external structure of planktonic foraminifera generated using manual segmentation in Dragonfly v. 2021.3 (Object Research Systems, Canada).

The samples containing these individuals and species spanned 5.65 to 2.85 Ma [14] and were collected from the Ceara Rise in the Equatorial Atlantic region at Ocean Drilling Program (ODP) Site 925, which comprised Hole 925B (4°12.248' N, 43°29.349' W), Hole 925C (4°12.256' N, 43°29.349' W) and Hole 925D (4°12.260' N, 43°29.363' W). See Curry *et al*. [15] for more details. This group was chosen to provide inter- and intra-specific species variation, and to provide contemporary data to test how morphological distinctiveness maps to taxonomic hypotheses [16].

The non-destructive imaging of both internal and external structures of the foraminifera was conducted at the μ-VIS X-ray Imaging Centre, University of Southampton, UK, using a Zeiss Xradia 510 Versa X-ray tomography scanner. Employing a rotational target system, the scanner operated at a voltage of 110 kV and a power of 10 W. Projections were reconstructed using Zeiss Xradia software, resulting in 16-bit greyscale .tiff stacks characterized by a voxel size of 1.75 μm and an average dimension of 992 × 1015 pixels for each two-dimensional slice.

2.2. Generation of training sets

We extracted the external calcite and internal cavity spaces from the micro-CT scans of the 50 individuals using manual segmentation within Dragonfly v. 2021.3 (Object Research Systems, Canada). This step took approximately 480 min per specimen (24 000 min total) and involved the manual labelling of 11 947 two-dimensional images. Segmentation data for each specimen were exported as multi-label (three labels: external, internal and background) 8-bit multipage .tiff stacks and paired with the original CT image data to allow for training (see figure 2).

The 50 specimens were categorized into three distinct groups (electronic supplementary material, table S1): 20 training image stacks, 10 validation image stacks and 20 test image stacks. From the training image category, we generated six distinct training sets, varying in size from 1 to 20 specimens (see table 1). These were used to assess the impact of training set size on segmentation accuracy, as determined through a comparative analysis against the validation set (see §2.3).

From the initial six training sets, we created six additional training sets through data augmentation using the NumPy library [17] in Python. This augmentation method was chosen for its simplicity and accessibility to researchers with limited computational expertise, as it can be easily implemented using a straightforward batch code. This augmentation process entailed rotating the original images five times (the maximum amount permitted using this method), effectively producing six distinct three-dimensional orientations per specimen for each of the original training sets (see figure 3). The augmented training sets comprised between 6 and 120 .tiff stacks (see table 1).

2.3. Training the neural networks

CNNs were trained using the offline version of Biomedisa [10], which utilizes a three-dimensional U-Net architecture [18]—the primary model employed for image segmentation [19]—and is optimized using Keras with a TensorFlow back end. We used patches of size 64 × 64 × 64 voxels, which were then scaled to a size of 256 × 256 × 256 voxels. This scaling was performed to improve the network's ability to capture spatial features and mitigate potential information loss during training. We trained three networks for each of the training sets to check the extent of stochastic variation in the results [20].

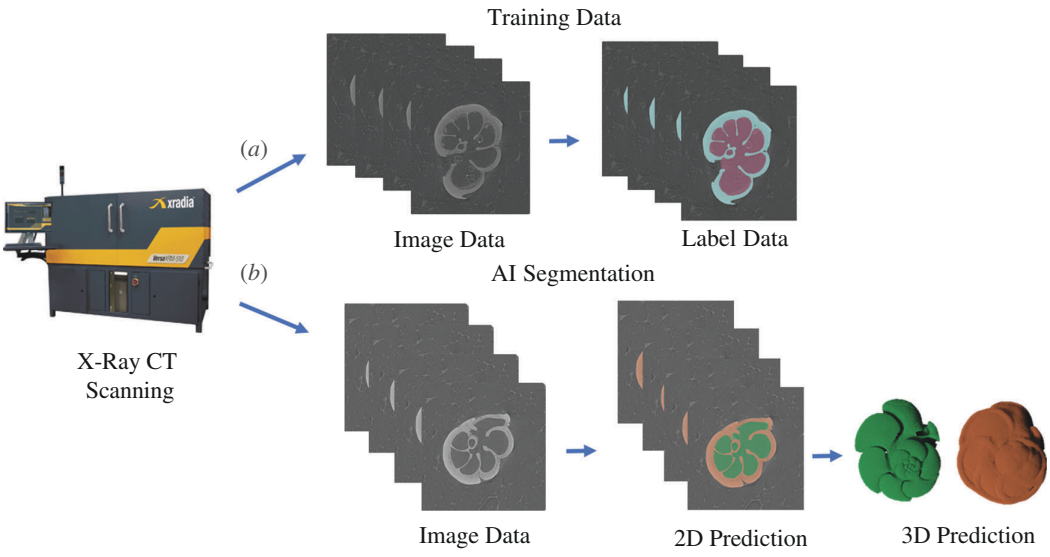


Figure 2. Workflow for producing training data and applying a deep CNN to perform automated image segmentation. The workflow includes (*a*) the creation of training data for the input into Biomedisa and (*b*) an example application of the trained CNN to automate the process of generating segmentation (label) data.

Table 1. Design of the training sets detailing the name of the training set, the data type (original or augmentation), the number of specimens in each training set, the number of two-dimensional images available for training and the approximate time to perform the manual segmentation to create the training set.

name of training set	data type	number of specimens	no. of foreground images (XY)	approx. duration of segmentation (mins)
T0_AI_1_Images	original	1	241	480
T1_AI_2_Images	original	2	390	960
T2_AI_4_Images	original	4	1028	1920
T3_AI_8_Images	original	8	2187	3840
T4_AI_16_Images	original	16	3571	7680
T5_AI_20_Images	original	20	4251	9600
Aug_T0_AI_1_Images	augmentation	1	4480	480
Aug_T1_AI_2_Images	augmentation	2	8812	960
Aug_T2_AI_4_Images	augmentation	4	18 120	1920
Aug_T3_AI_8_Images	augmentation	8	36 398	3840
Aug_T4_AI_16_Images	augmentation	16	71 100	7680
Aug_T5_AI_20_Images	augmentation	20	88 432	9600

To train our models in Biomedisa, we used a stochastic gradient descent with a learning rate of 0.01, a decay of 1×10^{-6}, momentum of 0.9 and Nesterov momentum enabled. A stride size of 32 pixels and a batch size of 24 samples per epoch were used alongside an automated cropping feature, which has been demonstrated to enhance accuracy [21]. The training of each network was performed on a Tesla V100S-PCIE-32GB graphics card with 30 989 MB of available memory. All the analyses and training procedures were conducted on the high-performance computing system at the Natural History Museum, London.

To measure network accuracy, we used the Dice similarity coefficient (Dice score), a metric commonly used in biomedical image segmentation studies [22,23]. The Dice score quantifies the level of overlap between two segmentations, providing a value between 0 (no overlap) and 1 (perfect match). For two segmentations, X and X' consisting of n labels, the Dice score is defined as

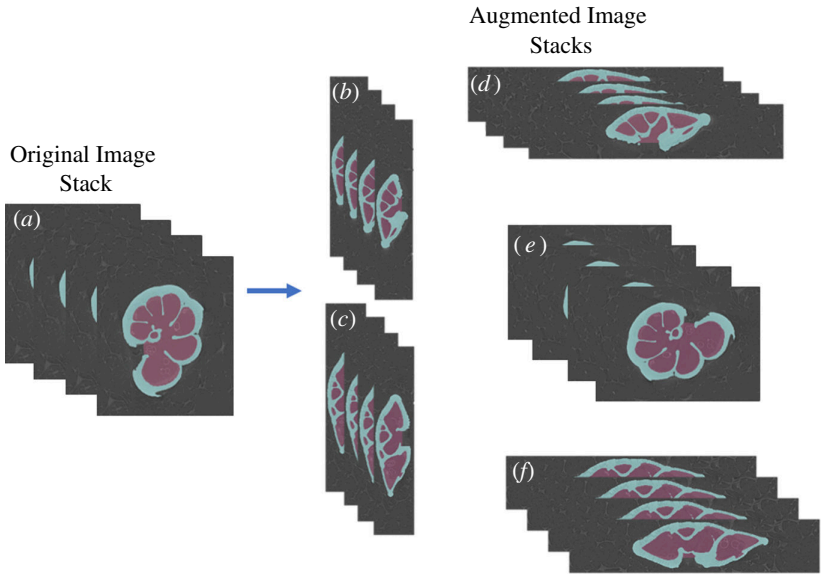


Figure 3. Rotation of the original image to form the new augmentation training data. The original image in an (*a*) *xyz* orientation is rotated into five other three-dimensional planes: (*b*) *yzx*, (*c*) *zyx*, (*d*) *xzy*, (*e*) *yxz* and (*f*) *zxy* orientations. These are then all paired together and used in training.

$$\text{Dice} = \frac{2\sum_{i=1}^{n}(X_i \cap X_i')}{|X| + |X'|},$$

where $|X|$ and $|X'|$ are the total number of voxels of each segmentation, respectively, and X_i is the subset of voxels of X with label i.

We conducted experiments to evaluate the potential efficiency gains of using an early stopping mechanism within Biomedisa. After testing a variety of epoch limits, we opted for an early stopping criterion set at 25 epochs, which was found to be the lowest value as to which all models trained correctly for every training set. By 'trained correctly' we mean if there is no increase in Dice score within a 25-epoch window, the optimal network is selected, and training is terminated. To gauge its impact of early stopping on network accuracy, we compared the results obtained from the original six training sets under early stopping with those obtained on a full run of 200 epochs.

2.4. Evaluation of feature extraction

We used the median accuracy network from each of the 12 training sets to produce segmentation data for the external and internal structures of the 20 test specimens. The median accuracy was selected as it provides a more robust estimate of performance by ensuring that outliers had less impact on the overall result. We then compared the volumetric and shape measurements from the manual data with those from each training set. The volumetric measurements were total volume (comprising both external and internal volumes) and percentage calcite (calculated as the ratio of external volume to internal volume, multiplied by 100).

To compare the shape, mesh data for the external and internal structures was generated from the segmentation data of the 12 training sets and the manual data. Meshes were decimated to 50 000 faces and smoothed before being scaled and aligned using Python and generalized Procrustes surface analysis [24], respectively. Shape was then analysed using the landmark-free morphometry pipeline, as outlined by Toussaint *et al*. [25]. We used a kernel width of 0.1 mm and noise parameter of 1.0 for both the analysis of shape for both the external and internal data, using a Keops kernel (PyKeops; https://pypi.org/project/pykeops/) as it performs better with large data [25]. The analyses were run for 150 iterations, using an initial step size of 0.01. The manually generated mesh for the individual st049_bl1_fo2 was used as the atlas for both the external and internal shape comparisons.

3. Results

3.1. Early stopping versus 200 epochs

A likelihood ratio test found no detectable difference between the observed correlation of the six original training sets under early stopping within 25 epochs and the full run of 200 epochs (analysis of variance, $F_{4,\ 5} = 0.0424$, $p > 0.05$), with a strong correlation between the early stopping and full run values (see figure 4; $R^2 = 0.9997$, $p < 0.001$). Consequently, we report results using the early stopping criterion for all subsequent tests.

3.2. Network accuracy

The Dice scores of both the original and augmented datasets increased with the number of training images (figure 5; electronic supplementary material, table S2). The improvement in accuracy from 1 specimen to 20 specimens was 12.34% and 4.67% for the original and augmented datasets, respectively. Most of this improvement occurred between training sets of one to eight specimens. Dice scores varied across all training sets (analysis of variance, $F_{11,\ 24} = 92.84$, $p < 0.001$). The augmented training data resulted in higher mean Dice scores compared with their original counterparts, especially at lower specimen numbers (figure 5). Increasing specimen numbers increased the Dice score for the augmented data ($\beta = 0.314$ on the scale of the logit link function from a generalized linear model with a quasibinomial error structure, s.e. = 0.027, $p < 0.001$), but did so faster for the original data ($\beta = 0.211$ on the scale of the logit link function, s.e. = 0.031, $p < 0.001$) albeit from a much lower initial baseline ($\beta = -0.633$ on the scale of the logit link function, s.e. = 0.0059, $p < 0.001$). The model explains 96% of the variation in the data.

3.3. Volumetric comparison

We identified a significant overall correlation between the manually generated and network-generated total volumes for the 20 test images ($F_{1,\ 216} = 51\ 223$, $p < 0.001$; figure 6; electronic supplementary material, table S3), but uncovered significant differences in the degree of these correlations across training sets ($F_{11,\ 216} = 19.660$, $p < 0.001$). Here, the R^2 values, serving as indicators of goodness of fit, spanned from 0.9505 for the training set with 1 specimen to 0.9998 for that containing 20 specimens with data augmentation—a difference of 0.0493.

For percentage calcite, the correlation between the manually generated and network-generated values was also significant ($F_{1,\ 216} = 1093.2$, $p < 0.001$), and again significant differences were noted in the degree of correlation across training sets ($F_{11,\ 216} = 21.367$, $p < 0.001$). Notably, the deviation in the gradients from a perfect fit was more pronounced for the lower training sets for percentage calcite than for the total volumes. This difference was reflected in the R^2 values, which ranged from 0.0608 for the training set with 1 specimen to 0.9607 for the 20-specimen set with image augmentation—a much greater difference of 0.8999.

3.4. Shape comparison

We expanded the analysis to compare shape estimates between the manually derived data and the network data for both external and internal structures. In total, 16 530 control points were generated for the external structure, while 17 325 control points were generated for the internal structure. These data points were subsequently reduced to five principal axes which display 100% of the total variation through non-linear kernel principal component analysis (kPCA) [26], using 1000 iterations. This dimensionality reduction technique works by allowing the separation of nonlinear data by making use of kernels and projecting data into higher dimensional space where it becomes linearly separable.

Both the internal and external shape data exhibited strong and statistically significant correlations between the manually generated and network-generated shape values across all three axes, indicating consistency in their shape estimates (figure 7; table 2; electronic supplementary material, tables S4 and S5). In the analysis of the external structure, the high correlation between the manually and network-generated data across all training sets was maintained (table 2). When performing kPCA on the external shape results, we found that PC1 controlled 36.4% of the total variation, followed by PC2

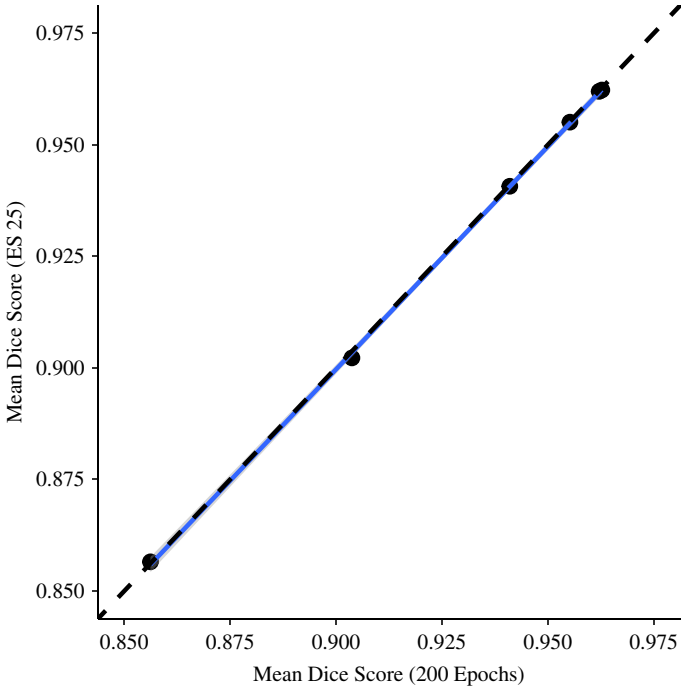


Figure 4. Comparison of Dice scores obtained from early stopping within 25 epochs and a full run of 200 epochs for each of the six original image training sets shows no statistical difference. The correlation coefficient (r^2) between the two sets of scores displayed is 0.9997.

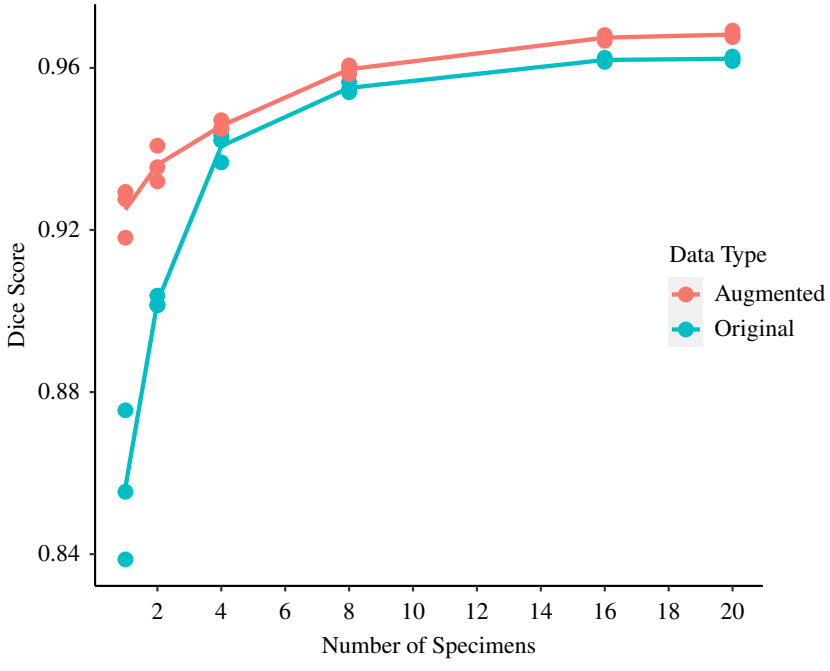


Figure 5. Comparison of segmentation accuracy for Biomedisa automated segmentation using average Dice scores calculated for validation data generated from 10 specimens. The plot shows increasing the training set size and implementing data augmentation improves network accuracy. All models were trained using early stopping within 25 epochs.

with 19.2%, and PC3 with 17.9%, resulting in the first three axes controlling 73.5% of the total variation (electronic supplementary material, table S6). By contrast, the internal structures displayed substantial variations across the training sets as the number of images used increased from 1 to 20 across all three axes (table 2), indicating differences among the accuracy of different training sets. In this analysis of internal structure, PC1 accounted for 37.6% of the total variation, followed by PC2 with 19.7% and PC3 with 18.1%, resulting in the first three axes controlling 75.4% of the total variation (electronic

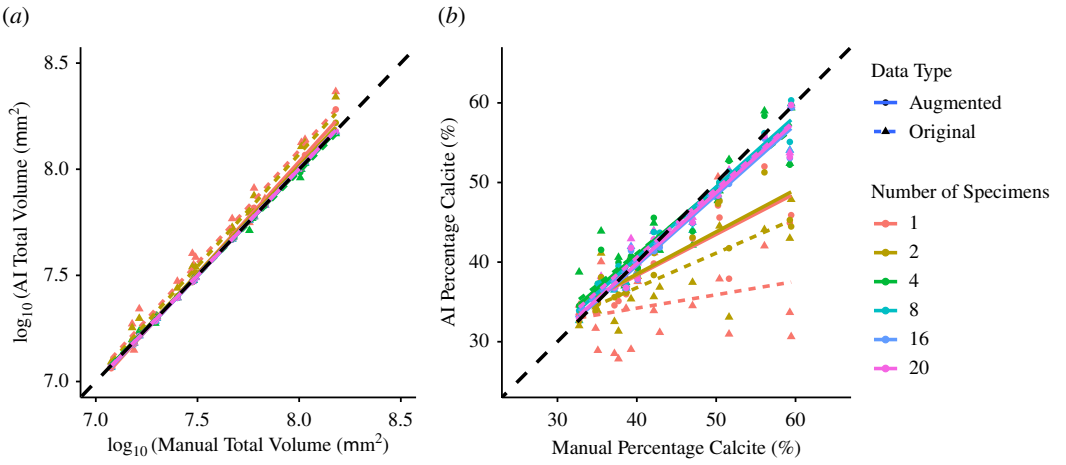


Figure 6. Comparison of (*a*) total volumes (internal and external combined) and (*b*) percentage calcite obtained from manual segmentation compared with those obtained using AI segmentation using the different training sets consisting of different numbers of training images. The predictions of total volume are generally more accurate than percentage calcite.

supplementary material, table S7). Owing to the high proportion of variance controlled by these PC axes, we use these for our primary comparisons.

4. Discussion

In this study, we observed a positive correlation between network accuracy and the quantity of training images, with data augmentation proving an effective tool for enhancing performance, especially for smaller training sets. External structures were extracted relatively accurately in the smaller training sets, unlike the internal structures, and size was more straightforward to extract than shape. Noting that some traits will always be more difficult to extract accurately than others, we discuss key concepts for the ongoing implementation of automated feature extraction in the biological sciences.

4.1. Impact of early stopping

The early stopping feature substantially reduced training time while effectively preserving network accuracy (electronic supplementary material, table S2). These qualities made the early stopping feature particularly valuable for larger training sets, and when applying data augmentation techniques.

Beyond improving time efficiency, the integration of early stopping features also serves to mitigate generalization errors and overfitting [27,28]. As a result, the application of early stopping features is widely applied when training deep learning models [29,30]. Our findings strongly endorse the use of these early stopping features in future research using three-dimensional image segmentation.

4.2. Impacts of training set size and data augmentation

Our findings reaffirm the principle that expanding the training set leads to the production of better deep learning models [31,32], albeit with diminishing returns as accuracy approaches 100% [33]. The expansion of available training data plays a crucial role in reducing generalization error and thus, in facilitating the achievement of optimal accuracy levels [34]. This is reflected in the stability of accuracy observed across epochs in larger training datasets, which are less affected by noise (electronic supplementary material, figures S3–S14).

Importantly, the minimum number of training images to achieve accurate results and the degree to which increasing the training set size enhances network accuracy remains task-specific and context-dependent [35]. The choice of the network architecture may vary in its suitability for segmenting specific material or structures [36], meaning training sets must carefully be selected. Notably, in this study, only a small number of individuals were required to achieve high accuracy scores, with eight specimens being required to achieve the goal of 95% accuracy (with and without augmentation) when comparing the manually and network-generated data. This achievement can be attributed to

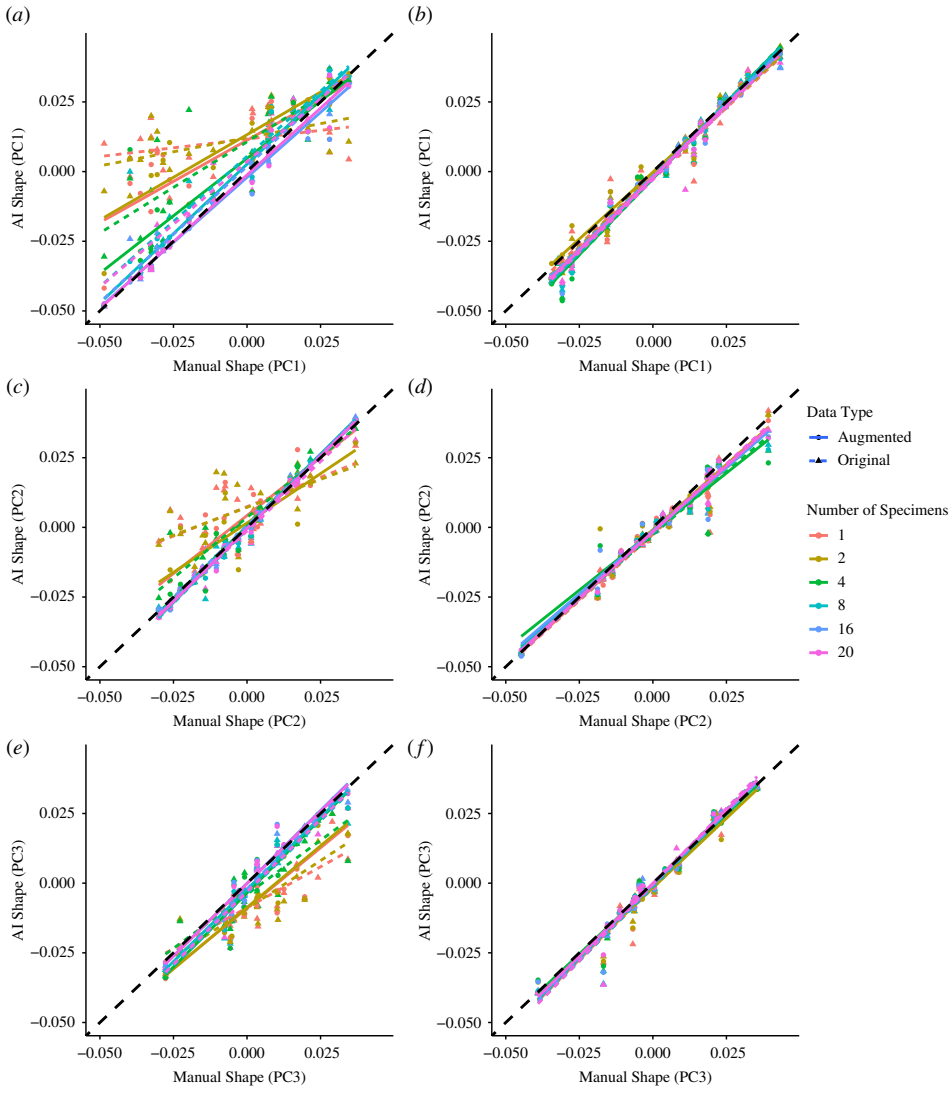


Figure 7. Comparison of shape for the internal (*a, c* and *e*) and external (*b, d* and *f*) structures across PC axes 1 (*a, b*), 2 (*c, d*) and 3 (*e, f*) obtained from manual segmentation compared with those obtained using AI segmentation using the different training sets consisting of different numbers of training images. The figure demonstrates that external calcite is easier to predict than the internal chamber space.

Table 2. There was no quantitative difference between manually generated and network-generated shape estimates for internal and external structures (left as *F* statistics on 1 and 216 degrees of freedom), but clear differences in predictive ability between internal and external shape estimates as the numbers of training images increase (right as *F* statistics on 11 and 216 degrees of freedom). *F* statistics are presented to our significant figures; values in bold indicate $p < 0.05$ whereas non-bold is $p > 0.05$.

	***F*-statistic between manual and network predictions**		***F*-statistic among different numbers of training images**	
axis	internal	external	internal	external
PC1	**548.4**	**7918**	**5.442**	0.5695
PC2	**1002**	**3474**	**3.893**	0.2491
PC3	**1208**	**3886**	**6.997**	0.3501

the wealth of two-dimensional slices per specimen and the ready availability of high-quality segmentation data. Without these attributes, a larger number of individuals for training would probably be necessary, resulting in delayed increases in accuracy. Consequently, the generation of high-quality

training data emerges as a factor more crucial than its sheer quantity [37], with any errors or inconsistencies within the training set likely to manifest in the deep learning models.

Data augmentation is a valuable tool for enhancing network accuracy, addressing limitations stemming from insufficient training data availability. Our results strongly support that data augmentation can effectively boost model accuracy [38,39]. Although variable by feature, we have demonstrated that it is possible to train a network with as few as one or two specimens to extract segmentation data almost as accurately as eight specimens using this augmentation approach. As a result, fewer images are needed for training and thus we can significantly reduce the time spent manually segmenting. Moreover, its application in this context is straightforward and universally applicable to three-dimensional image datasets.

4.3. Extracting phenotypic traits

The methodology presented in this study represents a significant advancement in the extraction of phenotypic traits from organisms such as planktonic foraminifera, offering notable improvements in processing times and repeatability compared with other established methods [40,41]. Our approach exhibits remarkable efficiency, culminating in the generation of results for a single specimen in approximately a minute, thus facilitating the execution of large-scale studies with unprecedented speed. Moreover, our methodology excels in repeatability, as it mitigates the impact of fluctuations in thresholding or greyscale values, ensuring more consistent and reliable data extraction. Crucially, our approach extends the scope of trait extraction beyond percentage calcite, encompassing additional parameters such as overall size (internal + external volumes) and shape, which remain elusive with alternative methodologies. This multi-faceted advancement underscores the versatility and efficacy of our approach in unravelling the intricacies of foraminiferal morphology.

Our results suggest extracting the internal structure of planktonic foraminifera poses a greater challenge compared with extraction of the external structure, evidenced in the measures of percentage calcite and the shape of the internal structure (see figures 6b and 7). This discrepancy in difficulty can probably be attributed to material homogeneity, which is reflected in low contrast differences within CT scans [42,43]. This challenge is particularly prominent in the case of planktonic foraminifera, where the internal structure contains sedimentary infill and nannofossil ooze with densities similar to external calcite [44,45]. The quantity of this infill is likely to vary among specimens, ranging from abundant to absent, while the presence of external calcite remains consistently detectable. Consequently, to alleviate the impact of irrelevant features, such as sediment infill, on trait extraction, we need training sets that encompass a broader array of images, representing a fuller sample of the entire image population compared with datasets without these elements. This can be achieved by either increasing the size of the training set or employing data augmentation techniques, thereby facilitating more accurate predictions.

Further enhancements in the accuracy of selected networks and mitigations of irrelevant features can be achieved through the application of post-processing tools. These tools facilitate the removal of connected components and the application of smoothing methods to eliminate noise from deep learning outputs [46,47], both of which hold the potential to yield more precise segmentation data.

4.4. Implications for future studies

There is a growing demand in the field of phenomics for the integration of machine learning techniques to perform image segmentation, driven by the increasing abundance of available data [48]. Unlike manual segmentation, which is inherently prone to observer bias [4], deep learning methods offer a robust means to ensure greater consistency across measurements [49], facilitating efficient and scalable comparative studies. The complexity and size of phylogenies [50] and birth–death models [51] are only getting larger, with the use of continuous traits [52] and geometric morphometric approaches [53] now commonplace. Thus, there is a need to develop morphometric approaches that can keep pace with these advances.

The application of automated segmentation and the data augmentation techniques applied here can be used universally to improve data acquisition. They integrate seamlessly with widely used segmentation software like Dragonfly, Avizo/Amira, ImageJ and MITK, so can be incorporated into existing workflows. They also accommodate various input sources, encompassing images obtained through X-ray synchrotron microscopy [54], soft microscopy [55] and magnetic resonance imaging [56].

This versatility empowers the workflow to conduct large-scale analyses efficiently and with a high level of accuracy within reasonable timeframes. With the provision of suitable input data [57], these trained networks become accessible to anyone for segmentation, regardless of their expertise in the morphology of the selected specimen.

The ability to extract different three-dimensional features using deep learning remains likely to be context specific, with a variety of features such as the presence of artefacts and irregularities [58], broken or incomplete specimens [59] and low-contrast images [47] each requiring individual solutions to overcome. Improvement in the quality of image data acquisition and selection of specimens can aid in mitigating some of these factors [6]; however, these must coincide with high-quality training sets.

The degree of tolerable error is also influenced by the quantification required. Here, our results demonstrate that extracting the total volume was easier than percentage calcite and shape. The variation in total volume is much greater and is less complicated to measure, and thus is less confounded by inaccurate measurements than more plastic traits. The landmark-free method applied here, for instance, requires the generation of meshes, and scaling and alignment of the data, which further produces error. Researchers should thus consider their intended measurements when building training data and then during the actual training of their neural networks.

5. Conclusion

We provide compelling evidence that leveraging deep learning for automated segmentation yields equivalent results to manual segmentation when extracting three-dimensional features from images and also drastically reduces the time required. Data augmentation emerged as a powerful tool, elevating network accuracy and enhancing the efficacy of smaller training sets. However, it is imperative to acknowledge that the accuracy of feature extraction is contingent on the specific traits targeted, with eight specimens being required here to extract all of the desired traits within levels of 95% accuracy in this case study. The observed differences for internal and external traits extracted underscore the crucial need for thoughtful consideration when training deep learning networks for specific applications.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. All computer code and final raw data used in the paper is available at [60]. All other data is available in the Dryad Repository [61].

Supplementary material is available online [62].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. J.M.M.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, software, validation, visualization, writing—original draft, writing—review and editing; A.S.-B.: data curation, investigation, methodology, software, writing—review and editing; A.B.: data curation, investigation, methodology, validation, writing—review and editing; M.S.: data curation, methodology, writing—review and editing; A.G.: conceptualization, funding acquisition, project administration, resources, software, supervision, validation, writing—original draft, writing—review and editing; T.H.G.E.: conceptualization, data curation, funding acquisition, project administration, resources, software, supervision, validation, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. J.M.M was supported by the Natural Environmental Research Council (grant no. NE/S007210/1). J.M.M, A.S.-B, A.B, M.S and T.H.G.E were also supported through the Natural Environmental Research Council (grant no. NE/P019269/1). A.G was supported by Leverhulme Trust grant RPG-2021-424 and European Research Council grant STG-2014-637171.

Acknowledgements. The authors would like to thank Philipp D. Lösel whose advice on performing the AI network training was invaluable, and all the other members of the Goswami and Ezard Labs who generously provided ideas for the concept of the paper. The authors acknowledge the μ-VIS X-ray Imaging Centre at the University of Southampton for the provision of the X-ray tomographic imaging facilities. We are also grateful for the thoughtful comments from the two anonymous reviewers and the editors.

References

1. Speijer RP, Van Loo D, Masschaele B, Vlassenbroeck J, Cnudde V, Jacobs P. 2008 Quantifying foraminiferal growth with high-resolution X-ray computed tomography: new opportunities in foraminiferal ontogeny, phylogeny, and paleoceanographic applications. *Geosphere* **4**, 760–763. (doi:10.1130/GES00176.1)
2. Cnudde V, Boone MN. 2013 High-resolution X-ray computed tomography in geosciences: a review of the current technology and applications. *Earth Sci. Rev*. **123**, 1–17. (doi:10.1016/j.earscirev.2013.04.003)
3. Hughes EC, Edwards DP, Bright JA, Capp EJR, Cooney CR, Varley ZK, Thomas GH. 2022 Global biogeographic patterns of avian morphological diversity. *Ecol. Lett*. **25**, 598–610. (doi:10.1111/ele.13905)
4. Joskowicz L, Cohen D, Caplan N, Sosna J. 2019 Inter-observer variability of manual contour delineation of structures in CT. *Eur. Radiol*. **29**, 1391–1399. (doi:10.1007/s00330-018-5695-5)
5. Maire E, Withers PJ. 2014 Quantitative X-ray tomography. *Int. Mater. Rev.* **59**, 1–43. (doi:10.1179/1743280413Y.0000000023)
6. Withers PJ *et al*. 2021 X-ray computed tomography. *Nat. Rev. Methods Primers* **1**, 18. (doi:10.1038/s43586-021-00015-4)
7. Yang X, Yu L, Li S, Wang X, Wang N, Qin J, Ni D, Heng PA. 2017 Towards automatic semantic segmentation in volumetric ultrasound. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2017: 20th Int. Conf., Quebec City, QC, Canada,11–13 September, 2017, Proceedings, Part I 20*, pp. 711–719. Springer. (doi:10.1007/978-3-319-66182-7_81)
8. Yan K, Bagheri M, Summers RM. 2018 3D context enhanced region-based convolutional neural network for end-to-end lesion detection. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st Int. Conf., Granada, Spain, 16–20 September, 2018, Proceedings, Part I*, pp. 511–519. Springer. (doi:%2010.1007/978-3-030-00928-1_58)
9. van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T, scikit-image contributors. 2014 scikit-image: image processing in Python. *PeerJ* **2**, e453. (doi:10.7717/peerj.453)
10. Lösel PD *et al*. 2020 Introducing Biomedisa as an open-source online platform for biomedical image segmentation. *Nat. Commun.* **11**, 5577. (doi:10.1038/s41467-020-19303-w)
11. Krizhevsky A, Sutskever I, Hinton GE. 2017 ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90. (doi:10.1145/3065386)
12. Taqi AM, Awad A, Al-Azzo F, Milanova M. 2018 The impact of multi-optimizers and data augmentation on TensorFlow convolutional neural network performance. In *2018 IEEE Conf. on Multimedia Information Processing and Retrieval (MIPR), Miami, FL*, pp. 140–145. IEEE. (doi:10.1109/MIPR.2018.00032)
13. Berggren WA, Kennett JP, Srinivasan MS. 1985 Neogene planktonic foraminifera: a phylogenetic atlas. *Micropaleontology* **31**, 94. (doi:10.2307/1485586)
14. Wilkens RH, Westerhold T, Drury AJ, Lyle M, Gorgas T, Tian J. 2017 Revisiting the Ceara Rise, equatorial Atlantic Ocean: isotope stratigraphy of ODP Leg 154 from 0 to 5 Ma. *Clim. Past* **13**, 779–793. (doi:10.5194/cp-13-779-2017)
15. Curry WB, Shackleton NJ, Richter C. 1995 Leg 154. Synthesis. *Proceedings ODP, Initial Reports* **154**, 421–442.
16. Pearson PN, Ezard THG. 2014 Evolution and speciation in the Eocene planktonic foraminifer *Turborotalia*. *Paleobiology* **40**, 130–143. (doi:10.1666/13004)
17. van der Walt S, Colbert SC, Varoquaux G. 2011 The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng* **13**, 22–30. (doi:10.1109/MCSE.2011.37)
18. Ronneberger O, Fischer P, Brox T. 2015 U-net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th Int. Conf., Munich, Germany*, vol. **Part III 18**, pp. 234–241, Springer. (doi:10.1007/978-3-319-24574-4_28)
19. Siddique N, Paheding S, Elkin CP, Devabhaktuni V. 2021 U-Net and its variants for medical image segmentation: a review of theory and applications. *IEEE Access* **9**, 82031–82057. (doi:10.1109/ACCESS.2021.3086020)
20. Desai AD, Gold GE, Hargreaves BA, Chaudhari AS. 2019 Technical considerations for semantic segmentation in MRI using convolutional neural networks. *arXiv Preprint.* (doi:10.48550/arXiv.1902.01977)
21. Lösel PD, Monchanin C, Lebrun R, Jayme A, Relle JJ, Devaud JM, Heuveline V, Lihoreau M. 2022 Natural variability in bee brain size and symmetry revealed by micro-CT imaging and deep learning. *PLoS Comput. Biol*. **19**, e1011529. (doi:10.1371/journal.pcbi.1011529)
22. Maier-Hein L *et al*. 2018 Why rankings of biomedical image analysis competitions should be interpreted with care. *Nat. Commun*. **9**, 5217. (doi:10.1038/s41467-018-07619-7)
23. Bertels J, Eelbode T, Berman M, Vandermeulen D, Maes F, Bisschops R, Blaschko MB. 2019 Optimizing the Dice score and Jaccard index for medical image segmentation: theory and practice. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd Int. Conf., Shenzhen, China*, vol. **Part II**, pp. 92–100, Springer. (doi:10.1007/978-3-030-32245-8_11)
24. Pomidor BJ, Makedonska J, Slice DE. 2016 A landmark-free method for three-dimensional shape analysis. *PLoS One* **11**, e0150368. (doi:10.1371/journal.pone.0150368)
25. Toussaint N *et al*. 2021 A landmark-free morphometrics pipeline for high-resolution phenotyping: application to a mouse model of Down syndrome. *Development* **148**, dev188631. (doi:10.1242/dev.188631)
26. Schölkopf B, Smola A, Müller KR. 1997 Kernel principal component analysis. In *Int. Conf. on Artificial Neural Networks, Berlin, Germany*, pp. 583–588. Springer. (doi:10.1007/BFb0020217)

27. Caruana R, Lawrence S, Giles C. 2000 Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. In *Advances in neural information processing systems*, vol. **13**, pp. 402–408,
28. Zhang T, Yu B. 2005 Boosting with early stopping: convergence and consistency. *Ann. Statist.* **33**, 1538–1579. (doi:10.1214/009053605000000255)
29. Zhang YD, Hou XX, Chen Y, Chen H, Yang M, Yang J, Wang SH. 2018 Voxelwise detection of cerebral microbleed in CADASIL patients by leaky rectified linear unit and early stopping. *Multimed. Tools Appl*. **77**, 21825–21845. (doi:10.1007/s11042-017-4383-9)
30. Amyar A, Modzelewski R, Li H, Ruan S. 2020 Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: classification and segmentation. *Comput. Biol. Med.* **126**, 104037. (doi:10.1016/j.compbiomed.2020.104037)
31. Bardis M, Houshyar R, Chantaduly C, Ushinsky A, Glavis-Bloom J, Shaver M, Chow D, Uchio E, Chang P. 2020 Deep learning with limited data: organ segmentation performance by U-Net. *Electronics* **9**, 1199. (doi:10.3390/electronics9081199)
32. Narayana PA, Coronado I, Sujit SJ, Wolinsky JS, Lublin FD, Gabr RE. 2020 Deep-learning-based neural tissue segmentation of MRI in multiple sclerosis: effect of training set size. *J. Magn. Reson. Imaging* **51**, 1487–1496. (doi:10.1002/jmri.26959)
33. Kavzoglu T. 2009 Increasing the accuracy of neural network classification using refined training data. *Environ. Model. Softw.* **24**, 850–858. (doi:10.1016/j.envsoft.2008.11.012)
34. Ying X. 2019 An overview of overfitting and its solutions. *J. Phys. Conf. Ser.* **1168**, 022022. (doi:10.1088/1742-6596/1168/2/022022)
35. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. 2014 Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958.
36. Fang Y, Wang J, Ou X, Ying H, Hu C, Zhang Z, Hu W. 2021 The impact of training sample size on deep learning-based organ auto-segmentation for head-and-neck patients. *Phys. Med. Biol.* **66**, 185012. (doi:10.1088/1361-6560/ac2206)
37. Nikolados EM, Wongprommoon A, Aodha OM, Cambray G, Oyarzún DA. 2022 Accuracy and data efficiency in deep learning models of protein expression. *Nat. Commun*. **13**, 7755. (doi:10.1038/s41467-022-34902-5)
38. Pereira S, Pinto A, Alves V, Silva CA. 2016 Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imaging* **35**, 1240–1251. (doi:10.1109/TMI.2016.2538465)
39. Wang G, Li W, Ourselin S, Vercauteren T. 2019 Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th Int. Workshop, BrainLes 2018, held in conjunction with MICCAI 2018*, *Granada, Spain*, vol. **Part II 4**, pp. 61–72, Springer. (doi:10.1007/978-3-030-11726-9_6)
40. Todd CL, Schmidt DN, Robinson MM, De Schepper S. 2020 Planktic foraminiferal test size and weight response to the late pliocene environment. *Paleoceanogr. Paleoclimatol.* **35**, e2019PA003738. (doi:10.1029/2019PA003738)
41. Kimoto K, Horiuchi R, Sasaki O, Iwashita T. 2023 Precise bulk density measurement of planktonic foraminiferal test by X-ray microcomputed tomography. *Front. Earth Sci.* **11**, 1184671. (doi:10.3389/feart.2023.1184671)
42. Furat O, Wang M, Neumann M, Petrich L, Weber M, Krill III CE, Schmidt V. Machine learning techniques for the segmentation of tomographic image data of functional materials. *Front. Mater.* **6**. (doi:10.3389/fmats.2019.00145)
43. Alathari T. 2015 *Feature extraction in volumetric images*. PhD thesis, University of Southampton, Southampton, UK.
44. Zarkogiannis SD *et al*. 2020 An improved cleaning protocol for foraminiferal calcite from unconsolidated core sediments: HyPerCal—a new practice for micropaleontological and paleoclimatic proxies. *J. Mar. Sci. Eng.* **8**, 998. (doi:10.3390/jmse8120998)
45. Zarkogiannis SD, Antonarakou A, Fernandez V, Mortyn PG, Kontakiotis G, Drinia H, Greaves M. 2020 Evidence of stable foraminifera biomineralization during the last two climate cycles in the tropical Atlantic Ocean. *J. Mar. Sci. Eng.* **8**, 737. (doi:10.3390/jmse8100737)
46. Larrazabal AJ, Martinez C, Ferrante E. 2019 Anatomical priors for image segmentation via post-processing with denoising autoencoders. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd Int. Conf.*, *Shenzhen, China*, vol. **Part VI 22**, pp. 585–593, Springer. (doi:10.1007/978-3-030-32226-7_65)
47. Tuladhar A, Schimert S, Rajashekar D, Kniep HC, Fiehler J, Forkert ND. 2020 Automatic segmentation of stroke lesions in non-contrast computed tomography datasets with convolutional neural networks. *IEEE Access* **8**, 94871–94879. (doi:10.1109/ACCESS.2020.2995632)
48. Lürig MD, Donoughe S, Svensson EI, Porto A, Tsuboi M. 2021 Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. *Front. Ecol. Evol.* **9**, 148. (doi:10.3389/fevo.2021.642774)
49. Willers C *et al*. 2021 The impact of segmentation on whole-lung functional MRI quantification: repeatability and reproducibility from multiple human observers and an artificial neural network. *Magn. Reson. Med*. **85**, 1079–1092. (doi:10.1002/mrm.28476)
50. Vernygora OV, Simões TR, Campbell EO. 2020 Evaluating the performance of probabilistic algorithms for phylogenetic analysis of big morphological datasets: a simulation study. *Syst. Biol*. **69**, 1088–1105. (doi:10.1093/sysbio/syaa020)
51. Šmíd J, Tolley KA. 2019 Calibrating the tree of vipers under the fossilized birth-death model. *Sci. Rep*. **9**, 5510. (doi:10.1038/s41598-019-41290-2)
52. Parins-Fukuchi C. 2018 Use of continuous traits can improve morphological phylogenetics. *Syst. Biol*. **67**, 328–339. (doi:10.1093/sysbio/syx072)
53. Smith UE, Hendricks JR. 2013 Geometric morphometric character suites as phylogenetic data: extracting phylogenetic signal from gastropod shells. *Syst. Biol*. **62**, 366–385. (doi:10.1093/sysbio/syt002)
54. van de Kamp T *et al*. 2022 Evolution of flexible biting in hyperdiverse parasitoid wasps. *Proc. R. Soc. B* **289**, 20212086. (doi:10.1098/rspb.2021.2086)
55. Erozan A, Lösel P, Heuveline V, Weinhardt V. 2023 ACSeg: automated 3D cytoplasm segmentation in soft X-ray tomography. *Preprint Series of the Engineering Mathematics and Computing Lab* (doi:10.11588/emclpp.2023.2.94947)

56. Lösel P, Heuveline V. 2017 A GPU based diffusion method for whole-heart and great vessel segmentation. In *Reconstruction, Segmentation, and Analysis of Medical Images: First Int. Workshops, RAMBO 2016 and HVSMR 2016, held in conjunction with MICCAI 2016*, *Athens, Greece*, vol. **1**, pp. 121–128, Springer. (doi:10.1007/978-3-319-52280-7_12)
57. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak J, van Ginneken B, Sánchez CI. 2017 A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88. (doi:10.1016/j.media.2017.07.005)
58. Das S, Nayak GK, Saba L, Kalra M, Suri JS, Saxena S. 2022 An artificial intelligence framework and its bias for brain tumor segmentation: a narrative review. *Comput. Biol. Med.* **143**, 105273. (doi:10.1016/j.compbiomed.2022.105273)
59. Zhang W, Ezard T, Searle-Barnes A, Brombacher A, Katsamenis O, Nixon M. 2020 Towards understanding speciation by automated extraction and description of 3D foraminifera stacks. In *2020 IEEE Southwest Symp. on Image Analysis and Interpretation (SSIAI), Albuquerque, NM, USA*, pp. 30–33. IEEE. (doi:10.1109/SSIAI49293.2020.9094611). https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=9092968.
60. Mulqueeney JM, Searle-Barnes A, Brombacher A, Sweeney M, Goswami A, Ezard THG. 2024 How many specimens make a sufficient training set for automated three-dimensional feature extraction? *bioRxiv* https://github.com/JamesMulqueeney/Automated-3D-Feature-Extraction
61. Mulqueeney JM. 2024 Data from: How many specimens make a sufficient training set for automated three dimensional feature extraction?. Dryad Digital Repository. (doi:10.5061/dryad.1rn8pk12f)
62. Mulqueeney JM, Searle-Barnes A, Brombacher A, Sweeney M, Goswami A, Ezard T. 2024 Supplementary material from: How many specimens make a sufficient training set for automated 3D feature extraction?. Figshare. (doi:10.6084/m9.figshare.c.7288922)

Received: 10 March 2023 | Revised: 28 April 2023 | Accepted: 28 April 2023

DOI: 10.1002/rcm.9533

RESEARCH ARTICLE

Rapid Communications in Mass Spectrometry WILEY

Laser ablation mass spectrometry blast through detection in R

Alex Searle-Barnes | James A. Milton | Christopher D. Standish | Gavin L. Foster | Thomas H. G. Ezard

Ocean and Earth Science, University of Southampton, Southampton, UK

Correspondence
Alex Searle-Barnes, Ocean and Earth Science, University of Southampton, European Way, Southampton SO14 3ZH, UK.
Email: c.j.a.searle-barnes@soton.ac.uk

Funding information
Natural Environment Research Council, Grant/Award Number: NE/P019269/1

Rationale: Organisms that grow a hard carbonate shell or skeleton, such as foraminifera, corals or molluscs, incorporate trace elements into their shell during growth that reflect the environmental change and biological activity they experienced during life. These geochemical signals locked within the carbonate are archives used in proxy reconstructions to study past environments and climates, to decipher taxonomy of cryptic species and to resolve evolutionary responses to climatic changes.

Methods: Here, we use laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) as a time-resolved acquisition to quantify the elemental composition of carbonate shells and skeletons. We present the LABLASTER (Laser Ablation BLASt Through Endpoint in R) package, which imports a single time-resolved LA-ICP-MS analysis, then detects when the laser has ablated through the carbonate as a function of change in signal over time and outputs key summary statistics. We provide two examples within the package: a fossil planktic foraminifer and a tropical coral skeleton.

Results: We present the first R package that automates the selection of desired data during data reduction workflows. This is achieved by automating the detection of when the laser has ablated through a sample using a smoothed time series, followed by removal of off-target data points. The functions are flexible and adjust dynamically to maximise the duration of the desired geochemical target signal, making this package applicable to a wide range of heterogenous bioarchives. Visualisation tools for manual validation are also included.

Conclusions: LABLASTER increases transparency and repeatability by algorithmically identifying when the laser has either ablated fully through a sample or across a mineral boundary and is thus no longer documenting a geochemical signal associated with the desired sample. LABLASTER's focus on better data targeting means more accurate extraction of biological and geochemical signals.

1 | INTRODUCTION

Laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) is a powerful analytical tool to quantify the elemental composition of a wide variety of natural and anthropogenic materials. A laser beam is focussed to the surface of a target and then pulsed to ablate the sample. Particles from the ablated sample are subsequently transported into an inductively coupled plasma ionisation source then to a mass spectrometer for detection based on the mass-to-charge ratio, which can be converted into a time-resolved isotopic or

elemental profile. LA-ICP-MS has become increasingly popular with biogenic carbonates, including foraminifera,[1–3] coral skeletons[4] and molluscs,[5] all of which act as archives of geochemical signals that can be used as proxy measurements to both reconstruct past environments and study the evolutionary response to long-term climate change.

Recent instrumentation advances enable LA-ICP-MS to collect trace element to calcium ratio (TE/Ca) results comparable to traditional solution based ICP-MS but with simpler sample preparation and higher throughput.[6] This solid sample and laser approach also allows for higher spatial resolution measurement of the sample and avoids the averaging of heterogeneity that occurs in solution-based ICP-MS. Nevertheless, there is no package for the R Environment for Statistical and Graphical Computing[7] that combines high data throughput with the additional nuance that laser ablation data processing requires to keep the maximum amount of relevant data. To fully leverage the gains of LA-ICP-MS, any software must be flexible enough to handle nonhomogeneous samples.

Some ready-to-use free computer packages exist to process LA-ICP-MS data such as elementR[8] or the discontinued LAICPMS,[9] both of which use the R environment, and LATools,[10] which uses the Python environment. ElementR provides a point-and-click graphical user interface that slows data reduction throughput while giving the user fine control over the data integration period. TERMITE[11] is not a package per se but is optimised for repeatable data reduction of homogenous samples, where the data integration period must be adjusted individually for each measurement and therefore requires manual validation.

These software packages provide a general end-to-end workflow to process experimental data into results rather than specialising on a particular data reduction step. In comparison, the LABLASTER package presented here contains a function that specialises in identifying when the laser is no longer recording the geochemical target of interest and is therefore designed for high-throughput processing that does not require user interaction once configured. A variety of endpoint detection mechanisms are used in the literature, including fixed time stamps,[2] k-means clustering,[8] analyte signal below a given threshold,[12] the mid-point between high and background signal counts[10] and even manual identification when the complexity of the samples is too great.[10] Here, we fit a function over a first derivative to calculate the rate of signal change. As LA-ICP-MS increases in popularity and experiments become more complex, there is a need for repeatable algorithmic protocols that can deal with heterogeneous samples or where repeat measurements may have different integration times.

Each discipline using LA-ICP-MS tends to measure samples that have different matrixes and properties, for example polished rock sections, pellets or carbonate shells. The examples presented here have been tailored to the field of ecology and evolution with a fossilised planktic foraminifer and the field of paleoclimate geochemistry with a tropical coral skeleton. The LABLASTER package will work with any sample that the laser may ablate through (or past) and hit an undesired target of different elemental composition. The foraminifera example here demonstrates how LABLASTER can be used to meet the specific needs of micro-palaeontologists, whose data are often skewed and highly variable.[8]

At any given time, a single species may contain individuals that have highly variable trace element signals within their shells due to the presence of subspecies, ontogenetic trends and/or lived environment. Comparisons of geochemical signals within a species can untangle and quantify the magnitude of such variability, as exemplified by Kearns et al., who used Mg/Ca ratios of the extant foraminifera species *Globigerinoides ruber* measured by LA-ICP-MS to distinguish between depth habitats and thus inhabited niches of the distinct subspecies. Analysing trace element signals is one technique to identify temperature-differentiated niches that underpin many theoretical ecological, evolutionary and eco-evolutionary models.[13]

Here, we (1) improve current processing capabilities by dynamically identifying the end of the sample of a carbonate subject and (2) implement this improved processing in the first freely available software to automate data extraction of a time-resolved elemental depth profile. As demonstrated in the examples below, the end of the sample may be the maximum depth at a single spot location for a shell or a pore in a linear profile of a coral skeleton, but any nonhomogeneous target sample is generally applicable.

An automated laser ablation setup often requires a constant firing time to be programmed into the controlling computer, with no regard for the heterogeneity or variation in thickness of the target. When samples are porous or have changes in mineralogy or variation in thicknesses within a single analytical session while using a consistent laser pulsing time, there is inevitably a chance that the laser will move across a mineral boundary or ablate through the entire depth of the sample, and thus the recorded data will not be restricted to only the area of interest. Any elemental measurement recorded after the laser has ablated through the sample is not of the target, and it should be removed before subsequent statistical analysis. Because the time taken to ablate through a sample is not consistent, such corrections can be made manually on an ad hoc basis, but additional manual handling would be time-consuming, laborious and prone to subjective differences among operators. There are clear methodological benefits from the development of a repeatable workflow.

The LABLASTER package works alongside elementR or TERMITE application or can be run as a standalone process within bespoke scripts, providing a flexible and versatile methodological improvement for heterogeneous samples that treats each sample individually to optimise signal–noise ratios. LABLASTER can batch process within a workflow, is customisable to the sensitivity for endpoint detection and does not require a point-and-click user interface. These features offer a higher throughput for data reduction compared to manual or alternative software methods and maximise retainment of on-target data for subsequent analysis.

2 | MATERIALS AND METHODS

The LABLASTER package contains one function named endPoint() that calculates four items and four example data sets to illustrate its

use, including the foraminifer and coral examples presented here. The package requires R (≥3.5.0)[7] and has a number of dependencies as it calls functionality from the stats,[7] smooth,[14] ggplot2, magrittr[15] and scales[16] packages.

2.1 | endPoint detection

The main function of the LABLASTER package is to detect when the laser has either ablated through a target (e.g. a carbonate shell or coral skeleton) or across a mineral boundary in a transect. In the following sections, we illustrate this behaviour on a planktic foraminifer and a tropical coral as case studies.

Identifying the time range in which the laser is ablating the target is essential for accessing and correlating the relevant data within the recorded time series. The LABLASTER package assumes the data frame supplied begins with the laser in focus of the desired target and the endPoint() function determines the time stamp when the laser has ablated through the sample or across a boundary where the isotope signal changes rapidly. Keeping only the data between the start time and end time focusses subsequent analysis on only relevant target data.

2.1.1 | Algorithm

The endPoint() function requires a single data frame containing a minimum of two columns: a column containing a time index and a column containing isotope signal counts. This data frame is supplied to the function as the argument **detectDf** with the time index column specified as **timeCol** and the isotope signal column specified as **signalCol**. Any element that is abundant within the target but is scarce in the surrounding medium could be used to detect the ablate through endpoint time. In the case study below we use 'Ca44' containing the ^{44}Ca isotope counts per second as the signalCol as this is abundant within the foraminifera test and therefore shows high counts when the laser is focussed and on target, but once the laser has ablated through the chamber wall, it becomes defocussed, and therefore the measured counts decrease. The function expects for the first row of the supplied data frame to have the laser on-target.

The algorithm of the endPoint function first calculates a simple moving average over the signalCol isotope counts, with the number of points in the moving average controlled by the **smoothing** argument. This smoothed isotope count signal is then scaled between 0 and 1 to provide consistency within subsequent algorithm steps, making this function more generic and applicable to all time-resolved acquisition mass spectrometry data. A larger amount of smoothing reduces the variance between adjacent data points, averaging out any signal spikes and therefore flattening the signal against time curve. A flatter time-resolved profile curve gives greater distinction between the elevated signal of desired target ablation and the lower signal of the undesired under- or adjacent target surface.

Using smoothing to de-spike and flatten the signal during the time of elevated signal reduces the magnitude of variance over time and therefore reduces the likelihood of a false-positive detection in a change of ablated material composition, which is used as part of the detection in later steps of the algorithm. Over-smoothing the signal reduces the distinction from a sudden signal drop to a shallower gradual drop between the higher on-target and the lower off-target signal intensities that the algorithm uses to detect that the laser is no longer ablating the desired target, causing a delayed detection. Over-smoothing can be identified manually using the visualisation tools by comparing the black-scaled smoothed signal and blue-scaled rate of signal change (dy/dt) lines.

Next, the algorithm calculates the number of data points per time step in the data frame supplied. The temporal resolution of the data can affect how sensitive the algorithm is to the rate of scaled signal change. The algorithm uses a moving window to calculate the rate of change in signal against time. Higher temporal resolution data can result in a smoother decline in isotope signal across the default time window, causing a delayed endpoint detection, therefore using a wider time window captures a larger magnitude signal drop. The algorithm uses the largest magnitude of negative rate of isotope signal change to identify that the laser is no longer well focussed on the desired target. Without a rapid signal drop, for example, if the laser did not fully ablate through the target or no mineral boundary was crossed, then the algorithm will return the final observation in the provided data frame and a warning message is displayed to encourage use of the manual validation tools to check the results. The number of data points used as the width of the moving window is controlled by the ***dt*** argument. The minimum number of data points required to detect the endpoint is $dt + 2$.

Once the algorithm has identified the largest magnitude of negative rate of smoothed signal change, the corresponding time stamp is identified. As the largest signal drop occurs shortly after the laser has ablated through the sample or crossed a boundary, it is necessary to also remove the data that occurs between the final data point when the laser was on target and the largest signal drop. This elapsed time is calculated by dividing the moving window width by the number of data points per time step and is subtracted from the signal change time stamp. This earliest time stamp is when the laser was last ablating on the desired target and is returned as a numeric value in the returned data frame.

```
library(lablaster)
endPoint(detectDf, timeCol = "Time", signalCol = "Ca44", smoothing = 5,
         dt = 10, profile = "FALSE", timeUnits = "seconds")
```

2.2 | Visualisation of the blast through algorithm

The endPoint function additionally provides visualisation tools into the mechanism of the algorithm. The smoothed scaled signal, rate of change and inferred bounds on the geochemical target are useful for

diagnosing the real-life efficacy. A ggplot2 object is generated when the **profile** argument is set to TRUE. If the profile argument is TRUE, then the time units of the analysis are also required and specified with the **timeUnits** argument.

2.3 | Returned values

The endPoint function returns a single data frame containing the values calculated.

$startTime contains the earliest time step in the supplied time-resolved acquisition as a numerical value.

$endTime contains the last time step before the laser ablated through the carbonate shell as described above.

$df contains the supplied data frame with the raw data but containing only the rows that occur that are identified as on-target and three additional columns that were calculated by the algorithm of (1) the smoothed signalCol, (2) scaled rate of change and (3) the scaled smoothed signalCol.

$profile contains a visualisation of the endpoint mechanism as described above. This is only available if a profile was generated using profile = 'TRUE'.

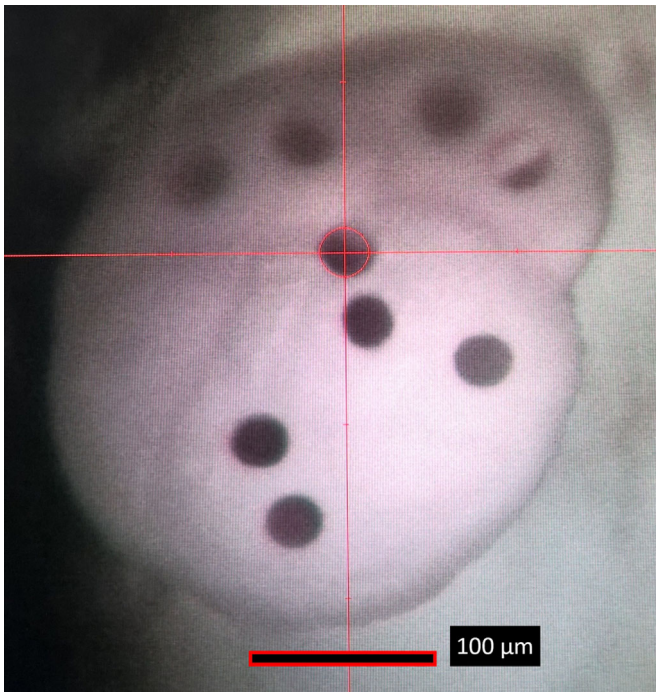


FIGURE 1 LA-ICP-MS holes from each shot are visible in the laser system software. [Color figure can be viewed at wileyonlinelibrary.com]

3 | CASE STUDY

3.1 | A foraminifer setup

3.1.1 | Experimental setup

A single planktonic foraminifer case study is included in the package as a worked example. Planktonic foraminifera are unicellular zooplankton distributed throughout the world's oceans as a key resource in understanding the Earth's climate system.[2,3] Foraminifera grow connected chambers throughout their life with those in the final whorl accessible for LA-ICP-MS analysis[17] (Figure 1). The thickness of a foraminifera chamber wall can vary substantially even within an individual. Some have thick solid chamber walls, while others have a highly porous structure as a result of species-level morphological variations, biological controls and environmental influences. This natural variability is the motivation for our development of automated processing methods and thus better control of geochemical 'vital effects'.[18] The case study is from the antepenultimate chamber of *Menardella exilis* foraminifera 72, identified hereon as 'Foram-72-shot-3'. The aim of this experiment was to quantify how Mg/Ca ratios vary between and within individual foraminifera.

The experimental setup is described fully in Kearns et al. Briefly, major and trace elements in the foraminifera test were analysed using a New Wave UP193 laser ablation system (ArF source, 30 µm spot diameter, fluence of 0.73 J/cm^3 and 5 Hz pulse rate) coupled to an Agilent 8900 triple quadrupole inductively coupled plasma (ICP-QQQ) mass spectrometer in single quadrupole mode using a He and Ar gas mixture (900 mL/min) at the University of Southampton. Each laser spot pattern was sequenced with a 30 s warmup, 50 s laser pulse and 30 s washout. The default values within this endPoint function are based on outputs from this setup.

Initial processing of the time-resolved acquisition data was performed within R.[7]The first 30 s was used to calculate the background signal and then removed. The next 5 s was removed as the isotopic signal began to rise due to the laser firing, but the ablated material was still travelling through the system. As our washout time was enough to purge the system after each analysis, the duration of this signal rise was regular and reproducible.

3.1.2 | endPoint detection function

The data frame containing the background-corrected remainder of the acquisition (Figure 2B) was supplied to the endPoint function. With the first 35 s removed, we ensured that the laser was firing on the foraminifera's outer test wall and the ablated material was reaching the mass detector for the first row in the data frame supplied. In this example, we use the ^{44}Ca isotope measurements due to the high signal:noise ratio and the abundance of ^{44}Ca within the foraminifera test, which provides greater distinction between higher signal counts for the duration when the laser was on target and lower signal counts when the laser was ablating the glass slide mount. The endPoint function was used to dynamically identify the time step when the laser had fully ablated through the foraminifera

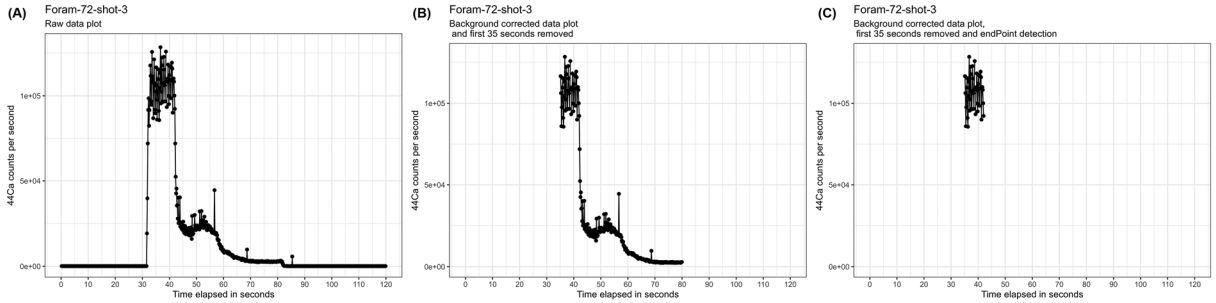


FIGURE 2 The processes implemented in the endPoint function in a time-resolved acquisition of Foram-72-shot-3 as a visualisation of the returned data frame $df. (A) shows the entire raw ^{44}Ca data collected, (B) shows the background-corrected ^{44}Ca with the first 35 s removed as this was before the laser was turned on and the ablated material is still travelling through the system and this is the data frame passed into the endPoint function, and (C) shows the target data retained after running the endPoint function, with both the first 35 s and post endPoint detection ^{44}Ca data removed.

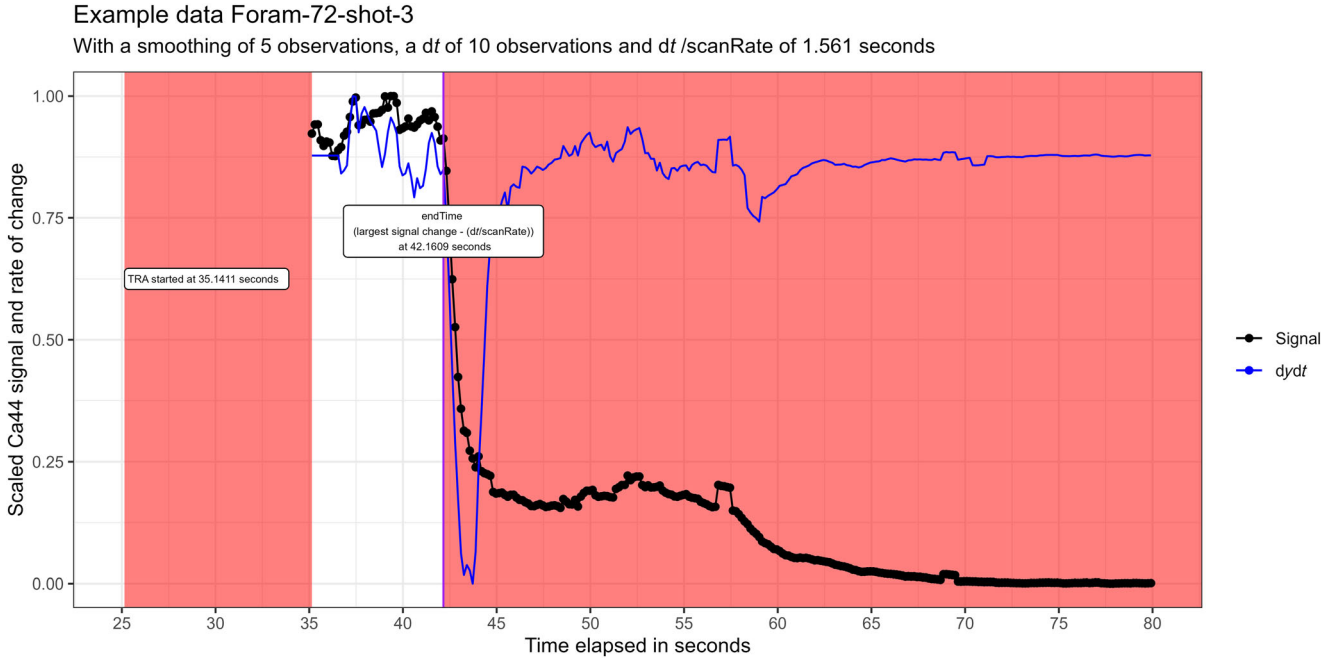


FIGURE 3 A visualisation of the endPoint function for the Foram-72-shot-3 case study, showing the scaled smoothed ^{44}Ca signal in black, the scaled rate of signal change in blue and the shaded red areas that identify the data point rows that are removed in the returned data frame. [Color figure can be viewed at wileyonlinelibrary.com]

test and to keep only the observations while the laser was ablating the target (Figure 2C).

```
endPoint(detectDf = foram72shot3, dt = 10, smoothing = 5, timeCol =
        "Time", signalCol = "Ca44", profile = "TRUE",  timeUnits =
        "seconds")
```

Figure 3 shows the manual validation tools to check that the endPoint function had successfully identified when the laser had ablated through the entire depth of the test wall. The black line is a scaled transformation of the smoothed signal specified in signalCol, the blue line is the scaled rate of change in signal over the moving time window and the shaded red areas identify the data points removed from the returned data frame, $df, as these exceed the endpoint blast through detection time stamp.

3.1.3 | Comparison of endpoint detection by varying the smoothing and d*t* arguments

We repeated the endpoint detection of the Foram-72-shot-3 case study specifying alternative smoothing and d*t* arguments (Figure 4). In comparison to Figure 3, which uses the default smoothing and d*t*

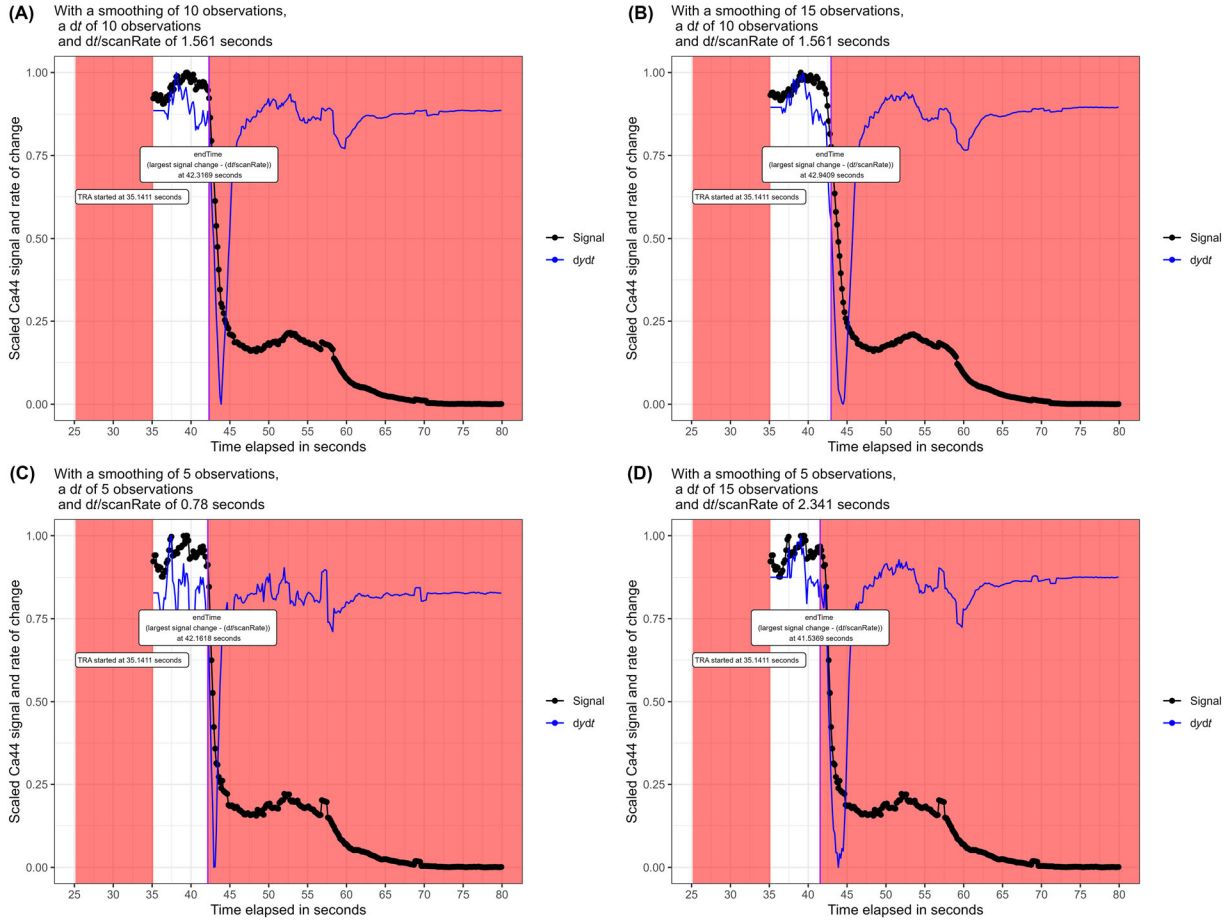


FIGURE 4 Comparison of endpoint detection of Foram-72-shot-3 by varying the smoothing and d*t* arguments. Using the default values of smoothing = 5 and d*t* = 10 the endpoint is identified as 42.0 s, in comparison to (A) smoothing = 10 and d*t* = 10 results with endpoint at 42.3 s, (B) smoothing = 15 and d*t* = 10 results with endpoint at 43.6 s, (C) smoothing = 5 and d*t* = 5 results with endpoint at 42.3 s, and (D) smoothing = 5 and d*t* = 15 results with endpoint at 41.7 s. [Color figure can be viewed at wileyonlinelibrary.com]

values, a larger smoothing value delays the endpoint detection, whereas larger d*t* values decrease the number of data points that are considered on-target. The user should use the visualisation tools to confirm the endPoint() function is optimised for any given application by keeping the maximum on-target data. Excessive smoothing returns additional rows of data when the laser is no longer on-target or an excessive d*t* returns a premature endpoint time stamp and on-target data are unnecessarily removed. In our experience, and with this particular application, examining five profiles was sufficient to identify the optimal smoothing parameter.

3.1.4 | Comparison without using endPoint detection

Foraminifera are archives of geochemical signals that are used in proxy measurement reconstructions. Sea surface temperatures can be reconstructed with calibrated equations using a ratio of magnesium to calcium as an input, making this Mg/Ca (mmol/mol) a popular geochemical measurement.[3,6] Our particular research concerns using trace element ratios in foraminifera to explore speciation events, tracking how new species and subspecies occupy the environment differently to their ancestors.[18,19] As a result, we typically analyse many individual foraminifera with multiple laser spots on each chamber (e.g. in Kearns et al.).

When processed without identifying the endpoint, the analysis time was between 35 and 80 s elapsed resulting with the median $^{24}Mg/^{44}Ca$ ratio of Foram-72-shot-3 as 4.37 ± 2.10 mmol/mol. In contrast, when processed with the endPoint() function, the analysis time was between 35 and 41 s elapsed resulting with the median of 2.46 ± 1.35 mmol/mol. While the signal counts of both Mg and Ca decrease as more time elapses due to laser defocussing and ablation through the carbonate structures (Figure 3), their ratio increases in this case. Lower absolute counts also increase uncertainty in the calculated ratios; hence, the use of the endPoint() function increases both accuracy (by focussing on the geochemical target) and precision through using higher-quality sample data for statistical analysis.

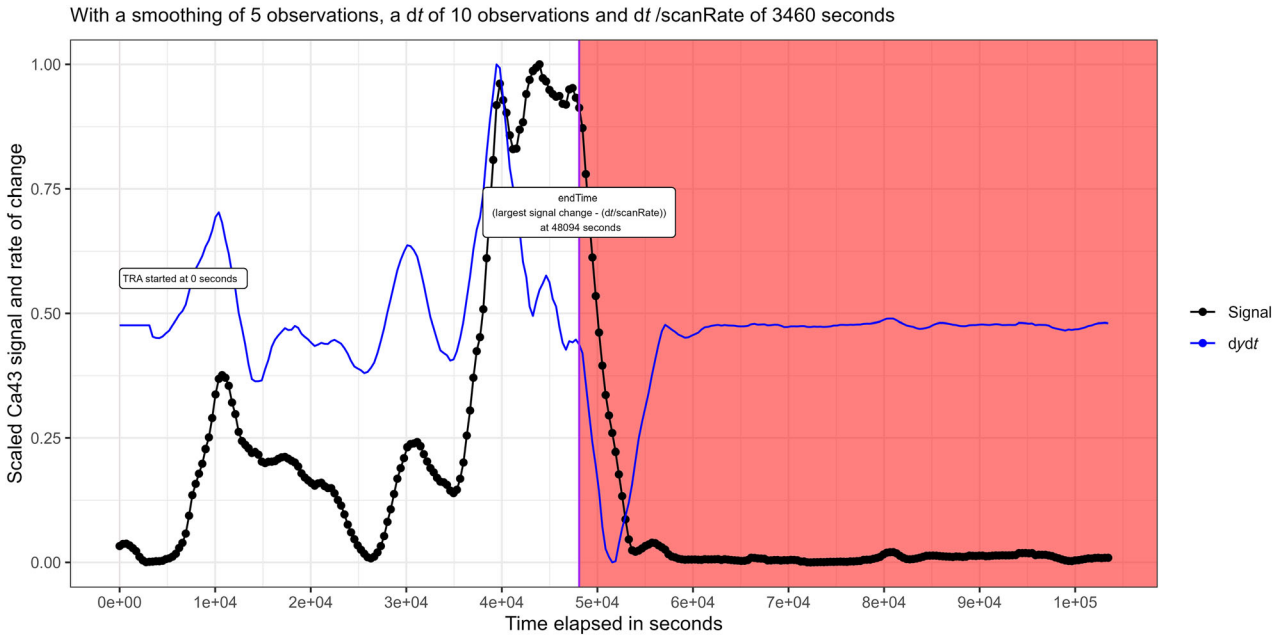


FIGURE 5 A visualisation of the endPoint function for the Coral-6 case study, showing the scaled smoothed ^{43}Ca signal in black, the scaled rate of signal change in blue and the shaded red areas identifying the data point rows that are removed in the returned data frame. [Color figure can be viewed at wileyonlinelibrary.com]

3.2 | A coral example

3.2.1 | Experimental setup

A transect across a sample of tropical coral *Acropora polystoma*, cultured in artificial seawater in a closed coral mesocosm at the Coral Reef Laboratory, National Oceanography Centre, Southampton,[20] is also included in the package as an example of an ablation line and is identified hereon as 'coral-6'. As polished coral sections can be irregular in shape, thickness and porosity, it is possible for the laser to ablate fully through at a thinner location or traverse a boundary between the sample and its mounting resin. In this example, the ablation line passes over a thinner section of coral and consequently the laser fully ablates through before the end of the analysis time.

3.2.2 | endPoint detection function

This coral section has bands of higher and lower calcium concentration across the laser transect and consequently fluctuating higher and lower counts for the calcium signal are observed. In this example, we selected the d*t* and smoothing parameters to de-spike these real signal fluctuations and using the included manual visualisation tool (Figure 5) to validate our selection identified the moment when the laser fully ablated through the coral. In this analysis, ^{43}Ca was the calcium isotope measured and hence used for endpoint detection.

```
endPoint(detectDf = coral6, dt = 10, smoothing = 5, timeCol =
        "Time", signalCol = "Ca43", profile = "TRUE",  timeUnits =
        "seconds")
```

4 | CONCLUSION

While LA-ICP-MS has many benefits over traditional solution-based ICP-MS methods for measuring major and trace elements in various carbonate objects, current analytical setups can be inflexible with regards to laser firing duration. Our freely available LABLASTER package implements methodological improvements to refine analytical workflows. We plan to extend endPoint() functionality to also include a startPoint() equivalent to detect when the laser starts pulsing on the desired target.

The endPoint function that implements these improvements is compatible for use within loops, aiding high-throughput and repeatable data cleaning of carbonate materials that are ubiquitously used in past climate reconstructions, geochemical ecology and evolution studies, and subspecies taxonomic distinctions.

AUTHOR CONTRIBUTIONS

Alex Searle-Barnes designed the functions and wrote the code. James A. Milton and Gavin L. Foster provided methodological support and laboratory analysis. Christopher D. Standish provided the coral data for additional algorithm testing and example data published within the LABLASTER package. Thomas H. G. Ezard validated the functions and code. Alex Searle-Barnes drafted the manuscript,

and James A. Milton, Gavin L. Foster, Christopher D. Standish and Thomas H. G. Ezard provided comments on earlier drafts.

ACKNOWLEDGMENTS

This work was supported by Natural Environment Research Council award NE/P019269/1.

CONFLICT OF INTEREST STATEMENT

The authors do not declare a conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in lablaster at https://github.com/alexsb1/lablaster.

ORCID

Alex Searle-Barnes https://orcid.org/0000-0003-0389-7717
Christopher D. Standish https://orcid.org/0000-0002-9726-295X

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/rcm.9533.

REFERENCES

1. Fehrenbacher JS, Spero HJ, Russell AD, Vetter L, Eggins S. Optimizing LA-ICP-MS analytical procedures for elemental depth profiling of foraminifera shells. *Chem Geol*. 2015;407-408:2-9. doi:10.1016/j.chemgeo.2015.04.007
2. Hathorne EC, Alard O, James RH, Rogers NW. Determination of intratest variability of trace elements in foraminifera by laser ablation inductively coupled plasma-mass spectrometry. *Geochem Geophys Geosyst*. 2003;4(12). doi:10.1029/2003GC000539
3. Eggins S, De Deckker P, Marshall J. Mg/Ca variation in planktonic foraminifera tests: Implications for reconstructing palaeo-seawater temperature and habitat migration. *Earth Planet Sci Lett*. 2003;212(3-4):291-306. doi:10.1016/S0012-821X(03)00283-8
4. Chalk TB, Standish CD, D'Angelo C, Castillo KD, Milton JA, Foster GL. Mapping coral calcification strategies from in situ boron isotope and trace element measurements of the tropical coral *Siderastrea siderea*. *Sci Rep*. 2021;11(1):472. doi:10.1038/s41598-020-78778-1
5. Schöne BR, Zhang Z, Radermacher P, et al. Sr/Ca and Mg/Ca ratios of ontogenetically old, long-lived bivalve shells (*Arctica islandica*) and their function as paleotemperature proxies. *Palaeogeogr Palaeoclimatol Palaeoecol*. 2011;302(1-2):52-64. doi:10.1016/j.palaeo.2010.03.016
6. Fehrenbacher J, Marchitto T, Spero HJ. Comparison of laser ablation and solution-based ICP-MS results for individual foraminifer Mg/Ca and Sr/Ca analyses. *Geochem Geophys Geosyst*. 2020;21:e2020GC009254. doi:10.1029/2020GC009254
7. R Core Team. 2022. R: a language and environment for statistical computing. https://www.R-project.org: R Foundation for Statistical Computing.
8. Sirot C, Ferraton F, Panfili J, Childs A-R, Guilhaumon F, Darnaude AM. elementr: an R package for reducing elemental data from LA-ICPMS analysis of biological calcified structures. *Methods Ecol Evol*. 2017;8(12):1659-1667. doi:10.1111/2041-210X.12822
9. Rittner M, Müller W. 2D mapping of LA-ICPMS trace element distributions using R. *Comput Geosci*. 2012;42:152-161. doi:10.1016/j.cageo.2011.07.016
10. Branson O, Fehrenbacher JS, Vetter L, Sadekov AY, Eggins SM, Spero HJ. LAtools: a data analysis package for the reproducible reduction of LA-ICPMS data. *Chem Geol*. 2019;504:83-95. doi:10.1016/j.chemgeo.2018.10.029
11. Mischel SA, Mertz-Kraus R, Jochum KP, Scholz D. TERMITE: an R script for fast reduction of laser ablation inductively coupled plasma mass spectrometry data and its application to trace element measurements. *Rapid Commun Mass Spectrom*. 2017;31(13):1079-1087. doi:10.1002/rcm.7895
12. Fricker MB, Kutscher D, Aeschlimann B, et al. High spatial resolution trace element analysis by LA-ICP-MS using a novel ablation cell for multiple or large samples. *Int J Mass Spectrom*. 2011;307(1-3):39-45. doi:10.1016/j.ijms.2011.01.008
13. Ward BA, Collins S. Rapid evolution allows coexistence of highly divergent lineages within the same niche. *Ecol Lett*. 2022;25(8):1839-1853. doi:10.1111/ele.14061
14. Svetunkov I, Petropoulos F. Old dog, new tricks: a modelling view of simple moving averages. *Int J Prod Res*. 2018;56(18):6034-6047. doi:10.1080/00207543.2017.1380326
15. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. *J Open Source Softw*. 2019;4(43):1686. doi:10.21105/joss.01686
16. Wickham H., Seidel D. Scales: Scale functions for visualization. 2022. https://scales.r-lib.org
17. Lea DW, Mashiotta TA, Spero HJ. Controls on magnesium and strontium uptake in planktonic foraminifera determined by live culturing. *Geochim Cosmochim Acta*. 1999;63(16):2369-2379. doi:10.1016/S0016-7037(99)00197-0
18. Kearns LE, Searle-Barnes A, Foster GL, Milton JA, Standish CD, Ezard THG. The influence of geochemical variation among *Globigerinoides ruber* individuals on paleoceanographic reconstructions. *Paleoceanogr Paleoclimatol*. 2023;38:e2022PA004549. doi:10.1029/2022PA004549
19. Steinke S, Chiu H-Y, Yu P-S, et al. Mg/ca ratios of two *Globigerinoides ruber* (white) morphotypes: Implications for reconstructing past tropical/subtropical surface water conditions. *Geochem Geophys Geosyst*. 2005;6(11). doi:10.1029/2005GC000926
20. D'angelo C, Wiedenmann J. An experimental mesocosm for long-term studies of reef corals. *J Mar Biol Assoc UK*. 2012;92(4):769-775. doi:10.1017/S0025315411001883

How to cite this article: Searle-Barnes A, Milton JA, Standish CD, Foster GL, Ezard THG. Laser ablation mass spectrometry blast through detection in R. *Rapid Commun Mass Spectrom*. 2023;37(15):e9533. doi:10.1002/rcm.9533

J. Micropalaeontology, 44, 107–117, 2025
https://doi.org/10.5194/jm-44-107-2025

Journal of Micropalaeontology
Open Access

Ten recommendations for scanning foraminifera by X-ray computed tomography

Alex Searle-Barnes[1], Anieke Brombacher[1,2], Orestis Katsamenis[3,4], Kathryn Rankin[3], Mark Mavrogordato[3], and Thomas Ezard[1]

[1]Ocean and Earth Science, University of Southampton, European Way, Southampton, SO14 3ZH, UK
[2]National Oceanography Centre, European Way, Southampton, SO14 3ZH, UK
[3]Micro-VIS X-ray Imaging Centre, University of Southampton, SO17 1BJ, UK
[4]Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, UK

Correspondence: Alex Searle-Barnes (c.j.a.searle-barnes@soton.ac.uk)

Received: 25 October 2024 – Revised: 3 February 2025 – Accepted: 12 February 2025 – Published: 29 April 2025

Abstract. Marine sediment cores uniquely provide a temporally high-resolution and well-preserved archive of foraminifera fossils, which are essential for understanding environmental, ecological, and evolutionary dynamics over geological timescales. Foraminifera preserve their entire ontogeny in their fossilized shells, and much of this life history remains hidden from view under a light microscope. X-ray microfocus computed tomography (μCT) imaging of individual foraminifera reveals internal chambers and pores that are traditionally hidden from view. Their volume, shape, and growth form foundations of oceanographic and environmental research.

Here, we present a set of 10 recommendations for the preparation and scanning of individual fossilized foraminifera using glue-, gel-, and solvent-free methods. We focus on the primary X-ray parameters of μCT imaging that a researcher can optimize according to their throughput, signal-to-noise ratio, and cost requirements to generate three-dimensional (3D; volumetric) datasets. We showcase the effect of these parameters on image quality through repeated scans on a single planktonic foraminifer that varied the X-ray beam power and energy, detector binning, number of projections, and exposure times.

In our case study, the highest beam power resulted in the widest contrast between the subject of interest and the background, allowing the easiest threshold-based segmentation of the object and aiding computers in automated feature extraction.

The values of these parameters can exhibit significant variability across individuals, based on the specific needs of the study, the equipment used, and the unique attributes of the samples under consideration. Our motivation with this paper is to share our experience and offer a foundation for similar studies.

1 Introduction

X-ray microfocus computed tomography (μCT) is a popular, non-destructive technique to study internal structures and features hidden from external view. High-quality images are convenient for manual analysis of internal morphology that has long been associated with forensic evolutionary divergence (Huber et al., 1997; de Vargas et al., 1999). In recent times, these images are valuable inputs for training machine learning and image classification tools, which are being developed to automate species identification, feature identification, and quantification of key taxonomic features (Mulqueeney et al., 2024; Marchant et al., 2020; Hsiang et al., 2019).

X-ray CT is powerful three-dimensional (3D) imaging technology that opens new horizons for understanding environmental, ecological and evolutionary dynamics by revealing internal views compared to the surface level of traditional light microscopy. High-resolution microfocus X-ray CT (μCT) imaging, 3D reconstruction algorithms, and volume data analysis software enable researchers to quantify morphological changes at both intra- and interspecific inter-

vals (Caromel et al., 2016; Burke et al., 2020; Schmidt et al., 2013), extract ontogenetic trajectories, identify the extent of dissolution (Johnstone et al., 2010; Iwasaki et al., 2015), measure shell wall thicknesses as a proxy for dissolution, and correlate changing geochemical signals with subspecies taxonomic variation (Kearns et al., 2023), all without damaging the specimens.

Fossilized foraminifera tests preserve the individual organism's entire ontogeny alongside the environmental conditions during the period of calcification. Brummer et al. (1987) describe the five stages of foraminifera growth from the proloculus through the juvenile and neanic stages before maturing in the terminal reproduction stage. Using X-ray CT and post-processing software, researchers can calculate chamber growth rates and aspect ratios as one technique to identify when ontogenetic transitions occur, especially in the earlier chambers that are hidden from a traditional light microscope view. Computer software packages aid the quantification of three-dimensional growth trajectories using chamber centroid coordinates and can be applied quickly across large datasets. The resulting 3D models are also useful teaching aids that show in detail the wall texture, porosity, and every grown chamber that Brummer described.

Various methods of sample preparations for scanning and the corresponding scanner configurations have been described (Görög et al., 2012; Hipsley et al., 2020; du Plessis et al., 2017; Briguglio et al., 2011; Vanadzina and Schmidt, 2022; Johnstone et al., 2010; Gooday et al., 2018). Methods in the literature include mounting foraminifer specimens on individual wooden toothpicks or pipette tips (Kendall et al., 2020; Vanadzina and Schmidt, 2022), keeping the specimen exposed to surroundings and dust and increasing the risk of loss. Table 1 compares the sample mounting methods, the time taken to scan, the post-reconstruction segmentation of individual foraminifera, and the achieved resolution of benchtop X-ray CT setups for these references that used laboratory benchtop instrumentation.

Here, we present 10 recommendations to optimize the throughput of X-ray computed tomographic scanning of planktic foraminifera. One of these recommendations is a sample mounting method which does not include glue, gel, or solvents that could interfere with accurate image segmentation, facilitating structured data storage and resulting in a relational database. In our case study, woven throughout this article, we housed 20 foraminifera encased within malleable phenolic resin foam (florist mounting foam) within a rigid plastic straw, isolating the samples from dust and making an easily transportable unit without the risk of losing the specimens.

Similar principles are applicable to other fossils (e.g. other protist shells, coral fragments, and otoliths) that researchers may need to scan using a laboratory-based X-ray CT setup. Our method presents a standardized operating procedure for micro-CT analysis that is advantageous for research projects that contain more individual specimens than can be scanned in a single session and prioritizes consistency within and between scanning sessions. From the case study we present here, the resulting data facilitate greater reproducibility and reusability in collaboration amongst the morphologically and geochemically inclined members of the micropalaeontological community.

2 Laboratory CT scanning

Laboratory X-ray micro-CT scanners function by emitting a focused X-ray beam through a specimen to a detector. X-rays are generated when a heated cathode filament ejects electrons that accelerate and hit a tungsten anode, producing X-ray photons through bremsstrahlung radiation or characteristic emission. The X-ray beam passes through the specimen, with attenuation proportional to the sample's density, and is measured by the detector, which translates the energy into greyscale values reflecting density variations. In a benchtop setup, the X-ray source and detector are fixed, while the specimen rotates incrementally, capturing two-dimensional (2D) projections from various angles. These projections are then back-projected and processed to reconstruct a 3D volume. Geometric magnification, influenced by the distances between the X-ray source, specimen, and detector, affects image resolution, with small focal spots minimizing geometric blurring.

We used a ZEISS Xradia 510 Versa X-ray micro-CT benchtop scanner in the case study presented here. The X-ray beam was pre-filtered through a 0.15 mm SiO_2 ceramic filter to absorb and attenuate the lower-energy X-ray photons from the spectrum, resulting in a beam that is enriched with higher-energy photons, improving the contrast and reducing beam-hardening artefacts in the final image as discussed below. This instrument features a two-stage magnification: the sample is enlarged through geometric magnification, then the beam hits a scintillator that converts X-rays into visible light, which is then magnified through an optical lens. This visible light then hits the detector, and a projection is recorded. This setup enabled us to achieve a spatial resolution of 990 nm, which is detailed enough to resolve the smallest of chambers and higher than would have been possible on a benchtop CT scanner without the second magnification stage. The Versa system was a more costly option than simpler benchtop laboratory CT-scanners but less expensive than synchrotron time.

3 The 10 recommendations

3.1 Preparation

3.1.1 Specimen selection and cleaning

Selecting which specimens to scan can take some consideration. As dissolution starts at the smallest internal chambers of a foraminifer, it may not be possible to judge the

Table 1. Comparisons for mounting methods, the time taken to scan, the post-reconstruction segmentation of individual specimens, and the achieved resolution of benchtop X-ray CT setups for scanning foraminifera where available.

Reference	Mounting	Scan time	Image segmentation	Voxel resolution (µm)
Briguglio et al. (2011)	Pipette tip or LEGO round brick (1 × 1)	Not reported	Manual separation of individual foraminifera from single reconstruction	∼ 4
Johnstone et al. (2010)	Multiple specimens glued to the sample holder	50 min for 10 individuals	Manual separation of individual foraminifera from single reconstruction	7
Görög et al. (2012)	Multiple specimens glued in place on celluloid film	30 min per individual	Manual separation of individual foraminifera from single reconstruction	1–10
du Plessis et al. (2017)	Florist mounting foam material (phenolic resin)	Did not scan foraminifera		
Gooday et al. (2018)	Individuals immersed in water in straight-sided polypropylene jar	37 min per individual	One foraminifer per reconstruction	13
Hipsley et al. (2020)	Multiple specimens wrapped individually in tissue, then inserted into a straw, itself inserted into a centrifuge tube	Did not scan foraminifera	Manual separation of individual fossils from single reconstruction	
Vanadzina and Schmidt (2022)	Multiple specimens mounted on a 45° cut pipette tip held with double-sided sticky tape	20 min for 14 individuals	Manual separation of individual foraminifera from single reconstruction	2.33–2.38
This study	Multiple specimens stacked laterally, avoiding overlaps	1 h per foraminifera	One foraminifer per reconstruction	0.685–1.370

preservation state before taking the scan (Johnstone et al., 2010; Iwasaki et al., 2015). If the study is to research well-preserved foraminifera, we suggest selecting a larger number of individuals to have enough non-dissolved specimens. The loss of internal septa (dividing walls between chambers) hinders chamber identification, which reduces the usefulness of the scan data for studying their life history. Stratified random sampling within a "best-preservation" category will facilitate representative environment, ecological, and evolutionary interpretations of the data.

This is the stage to check the specimen is free from surface debris and embedded particulates. In our case study, we followed the ultrasonication in ethanol method described in Eggins et al. (2003).

Identifying each individual specimen throughout the process from selection through scanning to data processing requires a systematic naming scheme to assign unique identifiers and a database, which can be as simple as a spreadsheet.

3.1.2 Mounting in a straw

Our method suggests mounting specimens in a single vertical stack within a rigid, clear plastic drinking straw encased within malleable phenolic resin foam (florist mounting foam), isolating the sample from dust and making an easily transportable unit (Figs. 1 and 2). The straw and foam keep the specimen securely in place during scanning, which is essential for collecting high-quality scans and is readily removed in post-processing due to the density difference between materials. Selecting a straw with the narrowest diameter that will accommodate the largest specimen will minimize the distance between the source beam and the specimen when mounted within the scanner, which will maximize geometric magnification of the specimen and beam intensity at the detector while also reducing the necessary image exposure time.

In an example setup including 20 individual planktic foraminifera, where the diameter of the largest specimen is

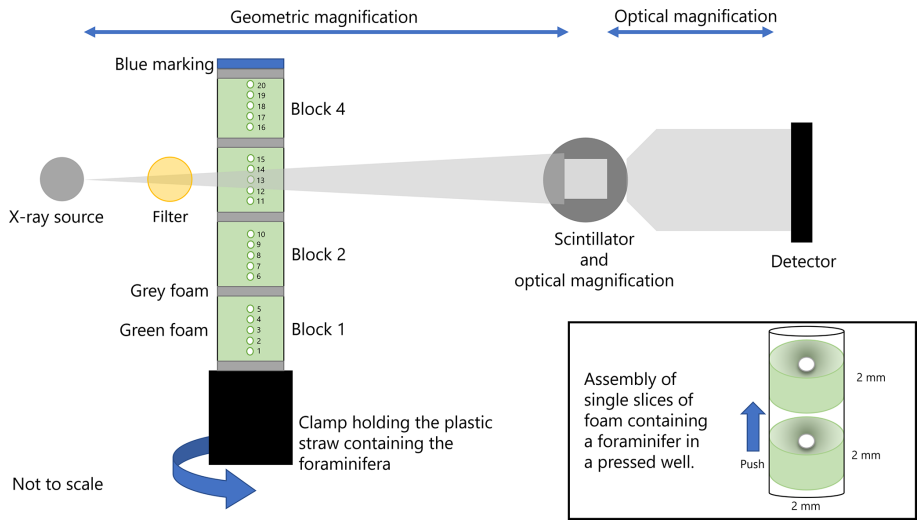


Figure 1. Our Xradia 510 Versa setup used throughout the presented case study. Key instrument parts shown are a tungsten X-ray source, a scintillator with optical magnification, and a charged coupled device (CCD) detector. The X-ray beam, shown as grey shading, passes through a 0.5 mm SiO_2 ceramic filter before entering a plastic straw packed with foraminifera sandwiched between green foam and with grey foam after every five specimens to build blocks of specimens and safe areas to cut for specimen retrieval and a blue marking to indicate the top of the straw to ensure the scan data correspond to the expected specimen. Individual specimens and each block are captioned in the figure. The straw is clamped to a stage and is incrementally rotated after each radiograph is taken. The box insert shows a schematic during the assembly of a straw by pushing individual foam slices with a pressed well containing a single foraminifer within the middle of the straw. Each foam slice is pushed up the straw until the slice meets the foam above, without crushing the foam, to maintain separation between individual specimens.

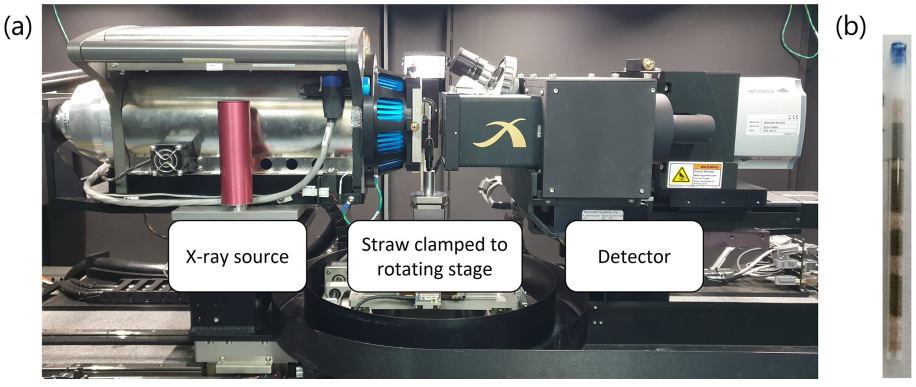


Figure 2. Case study setup showing **(a)** the X-ray source, a straw clamped to a rotating stage, and a scintillator with 4× optical magnification lens in front of the detector and **(b)** the same straw as constructed before clamping to the rotating stage, with the blue marking indicating the top of the straw.

~ 1 mm, we selected a 2 mm diameter wide and 10 cm tall straw. A taller straw when clamped at the bottom may be less stable at the top, causing motion artefacts in the images of the highest mounted specimens. Using the empty straw, we pressed out a 2 mm diameter disc of grey foam to mark the top of the stack and used a rod with the same internal diameter as the straw to push this foam slice to about 1 cm from the top. Next, we marked a 2 mm diameter disc of green foam and made a small well in the centre of the disc to hold a specimen. The slices of green foam should be sufficiently thick that no two specimens will touch, as this can cause automated algorithms in the post-processing stage to misassign the specimen identifier with the scan data. A specimen is easier to retrieve from a thicker slice of foam after scanning, as there is more foam surface area to grip while removing the slice from the straw. The well in the foam slice should be in the middle so that, when the specimen is imaged in 360°, the centre of rotation is also in the middle. Having the centre of rotation in the middle of the straw narrows the field of view required by the detector, thereby reducing the number of radiograph projections required and consequently the time taken to scan the specimen. The well acts as a nest for

the specimen but should be narrow enough to maintain a tight fit.

Once a specimen was transferred into the well, the slice was pushed up against the previous foam slice to secure the specimen in place. We repeated this process until there was a stack of five foraminifera making the first block. We then pressed a 2 mm disc of grey foam and pushed this up the straw to separate the end of the first block and the start of the second block. The slices of grey foam that separate the layers of green foam that make up the blocks are added, as they help the scanner operator navigate the straw under X-ray when locating the specimens during setup. In our example, the grey foam occurs after every five specimens, offering a secondary check that all five specimens in the block have been accounted for before progressing, and provides an area in the straw that is safe to cut after scanning to aid the retrieval of the specimens from the straw. By stacking the specimens vertically, the X-ray beam has only one specimen to pass through before reaching the detector, which reduces the opportunities for beam-hardening artefacts when compared to horizontal stacking.

We repeated building blocks of five foraminifera, separated by foam slices, until all 20 specimens had been loaded. Separating specimens like this facilitates keeping track of individuals using the location of each specimen within the straw and is an important detail in the relational database. We used a naming scheme that includes the straw number, the block number, and the foam slice number to identify each individual specimen. These unique identifiers aid the scan operator to locate individual specimens within a straw under X-ray and without the need of a visual map of each specimen location in the straw.

For continuity between scans, each specimen should be assigned its own tomography point and be the exclusive subject within it.

3.2 Scanning parameters

3.2.1 Purpose of the study

All the guidelines presented in this paper impact scanning time and the resulting image quality; this trade-off defines the optimization process.

The choice of scanning parameters is dependent on the purpose of the study. A large study with many individual specimens may prioritize minimizing the time taken for setup and scanning, encouraging an increased throughput. Other studies may prioritize image contrast, 3D volume resolution, or the standardization of parameters.

In our case study, we selected green and grey foams that both had lighter, but different, densities to the specimen to ensure a separation in greyscale values in the reconstructed images. A wider separation between specimen and support makes region-of-interest separation by grey value thresholding simpler during post-processing. Table 2 quantifies the effects of exposure time, the number of projections, the beam energy and power, and the amount of detector binning on the resulting radiographs (Fig. 3). We summarize the effects of increasing and decreasing each X-ray CT scanning parameter, as discussed in suggestions 4–8, in Table 4.

3.2.2 Field of vision

The highest spatial resolution is achieved by the specimen filling the detector's entire field of view. Filling the field of view is achieved by magnifying the specimen through a geometric magnification and, on Versa-like systems, an additional optical magnification stage. Filling the field of view utilizes more of the detector pixels to capture information from the specimen, contributing to a finer spatial sampling resulting in a detailed image and consequently a higher resolution.

Higher geometric magnification is achieved by decreasing the distance between the source and the specimen while increasing the distance between the specimen and the detector. Increasing the distance between the specimen and the detector provides a longer distance for the X-rays to diverge, leading to a larger area covered by the X-ray beam and greater magnification.

While widening the distance between the specimen and the detector increases magnification and spatial resolution, the image brightness reduces because the detected beam intensity is inversely proportional to the square of the distance from the source. To counter the reduced brightness caused by increased distance, the beam power could be increased, the exposure time could be lengthened to capture more X-rays, or greater detector binning could be applied to increase sensitivity.

By using more of the sensor to capture the foraminifera, there are fewer pixels to capture the surrounding foam or background noise, helping to enhance the signal-to-noise ratio and the selective allocation of pixels. The expanded coverage of the sensor by the foraminifera in each image facilitates enhanced artefact removal or reduction during 3D reconstruction, attributed to the increased number of overlapping pixels in each projection. Image artefacts that may occur because of undersampling and because of underutilizing the sensor include streaks, rings, and blurring of density boundaries (such as the edge of the denser foraminifer calcite and lower-attenuating foam).

By stacking specimens vertically within the plastic straw, we remove the horizontal overlapping of specimens and therefore the chance of beam hardening to occur, which can cause artefacts. By reducing the horizontal overlapping, greater magnification can be achieved to fill the field of vision, further increasing the spatial resolution. Avoiding the X-ray beam passing through multiple specimens in the same path reduces the chance of beam-hardening artefacts, which reduce image quality and can cause some automated 3D reconstructions to fail.

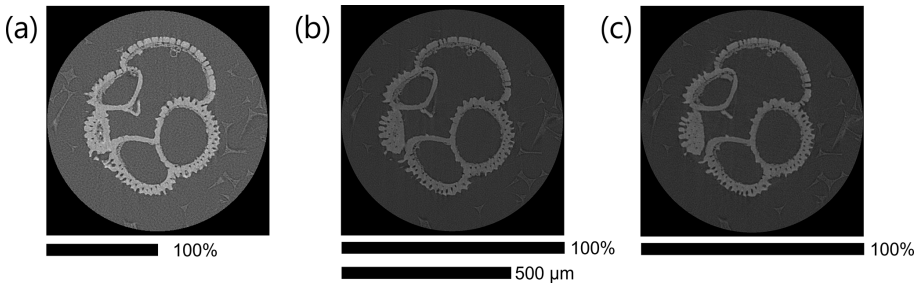


Figure 3. The same *Dentoglobigerina altispira* foraminifer was repeatedly scanned with varying parameters, and a single slice from the reconstructed .tiff stacks is presented to visually compare the effects of resolution, contrast, and artefacts. A scale bar shows the relative width of each square image and indicates that image **(a)** is presented twice as large to make the 500 µm scale bar applicable to all images. **(a)** 15 s exposure, 40 kV, and 4× binning achieving 1370 nm resolution; **(b)** 24 s exposure, 60 kV, and 2× binning achieving 685 nm resolution; **(c)** 12 s exposure, 80 kV, and 2× binning achieving 685 nm resolution.

Table 2. Instrument parameters for repeated scans of a single *Dentoglobigerina altispira* foraminifer with the achieved resolution and resulting file size.

Figure 3 panel	Optical magnification	Exposure (s)	Projections	Energy (kV)	Binning	Power (W)	Scan time (h)	File size (kb)	Resolution (nm)
a	20×	15	701	40	4	3	3.5	463	1370
b	20×	24	1101	60	2	5	8.5	1857	685
c	20×	12	1101	80	2	7	4.5	1857	685

3.2.3 Beam power

Increasing the beam power emits more photons at a higher energy. Denser areas within the specimen cause more attenuation to the X-ray beam, so photons require greater energy to pass through these areas. Higher-energy photons penetrate the specimen further, resulting in higher detection intensity and less difference in attenuation between different parts of the specimen, which lead to lower contrast. A compromise between sufficient power to penetrate the densest area of the sample and sufficient contrast between regions of interest is needed to capture a useful radiograph. In our example of a foraminifer specimen (Fig. 3), the beam power was enough to penetrate the denser shell wall, while retaining enough contrast to separate the calcite test from the primary organic membrane layer and surrounding phenolic foam.

Benchtop X-ray CT sources often produce polychromatic X-ray beams, which contain a range of photon energies. Placing a specific filter, such as doped glass, between the beam source and the straw filter selectively attenuates lower-energy X-rays while allowing higher-energy X-rays to pass through, resulting in a narrower range of higher-energy photons passing through the sample. Including a source filter can reduce beam-hardening artefacts in the radiograph while using a higher beam power to reduce the necessary exposure time and increase signal intensity.

3.2.4 Exposure time

Lengthening the exposure time increases the number of photons to hit the detector and increases the signal-to-noise ratio, giving a brighter, sharper, and less grainy image. Multiplying the exposure time by the number of projections needed provides an estimate for the time taken to complete a scan of the specimen. An exposure time that is too short causes an underexposed image that has a lower signal-to-noise ratio, causing image grain or making use of the entire available greyscale range. An overexposed image saturates the detector and appears whitewashed with a loss of contrast between differing densities in the X-ray beam path. Overexposure is an indication that the beam power is too high or that the exposure time is too long.

3.2.5 Detector binning

Binning a charged coupled detector (CCD) groups adjacent pixels in the imaging array and sums the signal in each pixel into a single grouped value. These grouped pixels increase the signal intensity proportionally to the spatial area of the binning applied, improving the quality of X-ray radiographs by increasing the signal-to-noise ratio and improving the detection of weaker signals, which results in brighter radiographs, which in turn allows shorter exposure times and thus faster scan acquisition times.

A drawback of binning is that, through the combination of adjacent pixels, the spatial resolution is reduced by the recip-

rocal of the amount of binning applied squared (see also Irie et al., 2022); Fig. 3 and Table 3 show the effect of doubling the detector binning.

3.2.6 Scanning time

In a high-throughput setup, both the time taken to set up an automated sequence of tomography points and the scan acquisition time must be considered. Retaining the same scanning parameters for each tomography point minimizes individual specimen setup time, assuming common settings from a suitable compromise for all specimens being scanned. The scan acquisition time is dependent on the beam power, exposure time, number of projections, and detector binning, which all impact the image quality, signal-to-noise ratio, contrast, and resolution. In our case study, with a field of view of 1 mm, we used an X-ray source excitation voltage of 110 kV and a power of 10 W, a 2× spatial binning, and an exposure time of 1.3 s for each of the 1101 projections to achieve a 1 h scan time. The field of view was consistently set for the largest specimen in the set, which we note sacrificed some potential resolution for smaller specimens but was faster to set up and made each scan directly comparable – if it looked bigger, it was bigger.

3.2.7 Specimen reconstruction

Each projection is captured on a detector, and the brightness at each pixel is converted into numbers on a greyscale value between black and white using mathematical transformations and materials of known density for calibration.

Software-based reconstruction filters improve the image quality by reducing streak, bloom, and ring artefacts, along with beam-hardening effects. In our case study, we applied a sharp reconstruction filter with a beam-hardening correction of 1.0 to further reduce beam-hardening artefacts through mathematical transformations.

The centre of rotation of the axial slices (looking down at the sample from the top) is identified algorithmically; once identified, the individual projections are stitched together around the centre of rotation to construct a 3D volume.

As the projection images contained only a single specimen within the surrounding foam, these projections were free from distracting additional regions of higher density, which was optimal for speed and accuracy for the reconstruction algorithms. The resulting reconstructed 3D volume featured a centred specimen with brighter grey values surrounded by darker foam values with sufficient contrast to use a simple brightness thresholding to separate the two. With each volume containing only one specimen and sufficient contrast, we were encouraged to automate the segmentation process and computerized feature identification, resulting in faster throughput of image reconstructions and post-processing analysis of the data because there is no need to separate specimens as unique regions of interest and reduce human-induced uncertainties.

Some reconstruction software allows the application a custom byte-scaling range, which sets the minimum and maximum greyscale values used in the reconstructed scan volume. Byte-scale ranges are bounds to the linear scaling of the original pixel values to fit within the custom range. Selecting a custom byte-scaling range that bounds the specimen's greyscale range results in an image with improved contrast and brightness that enhances the specimen and diminishes speckles from the foam. Unlike dynamic byte-scaling ranges that are derived from the underlying 2D images, which themselves are influenced by the variable instrument parameters and performance, setting custom values provides control for greyscale range consistency within the 3D volume. We consistently applied a byte-scaling range (−0.04 to 0.10) across all our reconstructed .tiff images to make the brightness within each image comparable between scans.

4 Output files and post-processing

The reconstructed three-dimensional volume is a numerical representation of the specimen and the surrounding foam and straw. This volume must be saved to a computer file before it can be used outside of the reconstructor software. Various file types are available for saving 3D volume data, with common options including raw data files or uncompressed image stacks, both of which are open and free formats. An image stack comprises a sequence of 2D images visualizing grey values that depict 3D slices in the axial view. Numerous image processing software options facilitate tasks such as noise reduction, contrast adjustment, and filtering, followed by region-of-interest segmentation and volumetric analysis. In line with open-access research principles advocated by Wilkinson et al. (2016), we encourage the use of any suitable file format (such as .tiff or .raw) that aligns with principles of accessibility, interoperability, and reusability for reconstructed micro-CT data.

We chose to export our 3D volume reconstructions as image stacks of .tiff files for wide compatibility with each software package used in our end-to-end methodology and to simplify data sharing with non-micro-CT experts, collaboration between other researchers, and long-term archival of reconstructed CT data. We used ORS Dragonfly (2022) to manually segment regions of interest for bespoke research questions and for labelling individual images to use as inputs for more automated feature extraction as training data in other projects (e.g. Mulqueeney et al., 2024).

The accuracy of threshold-based segmentation from reconstructed .tiff image stacks is influenced by the scanning parameters, which determine contrast, brightness separation, and the presence of artefacts. Higher X-ray energy (e.g. 80 kV) improves material differentiation by increasing the attenuation contrast between calcite and the surround-

Table 3. Comparison of mean brightness values of air, foam, and calcite (the foraminifer shell) using values from the histograms of Fig. 3a–c.

Figure	Mean air brightness	Mean foam brightness	Mean calcite brightness	Brightness separation between foam and calcite	Separation between foam and calcite
3a	23 051	26 018	40 853	14 835	57 %
3b	13 075	15 167	22 358	7191	47 %
3c	12 795	14 534	23 353	8820	61 %

Table 4. Summary of the effects of increasing and decreasing each X-ray CT scanning parameter, as discussed in suggestions 4–8.

Suggestion	Name	Effect of increasing	Effect of decreasing
4	Field of vision	Larger sample area captured but lower geometric magnification and reduced resolution.	Higher geometric magnification, improving resolution but capturing a smaller area.
5	Beam power	Greater penetration through dense materials and higher signal-to-noise ratio but increased potential for beam hardening artefacts.	Lower penetration and reduced contrast for dense features but minimizes artefacts in soft materials.
6	Exposure time	Higher signal-to-noise ratio and improved image quality but longer scan times and potential for sample drift.	Faster scans but noisier images with reduced contrast and detail.
7	Detector binning	Increases signal per pixel, reducing noise and improving contrast, but lowers spatial resolution.	Higher spatial resolution but increased noise and reduced contrast in low-signal regions.
8	Scanning time	More projections captured, improving reconstruction accuracy and reducing artefacts, but longer processing time.	Faster scanning but fewer projections, increasing reconstruction artefacts and reducing detail.

ing medium, resulting in clearer segmentation boundaries. Lower-energy scans (e.g. 40 kV) enhance internal contrast within the foraminifer shell but reduce the separation between foam and calcite, making threshold selection more challenging. Exposure time and detector binning also impact segmentation precision; longer exposures (24 s) and lower binning (2×) produce higher-resolution images with well-defined edges, facilitating more accurate segmentation. In contrast, shorter exposures (15 s) and higher binning (4×) introduce increased noise and reduced boundary sharpness, complicating the differentiation of materials. The number of projections and total scan time further affect reconstruction quality, with 1101 projections reducing streak artefacts and improving segmentation fidelity compared to scans with 701 projections. Figure 3 is a demonstration of how the scanning recommendations presented here can be optimized for achieving reliable threshold-based segmentation of foraminiferal calcite. Reliable segmentation encourages future developments in automatic and programmatic feature extraction from the reconstructed .tiff images for increased automation and reduction in manual processing time.

5 Comparison of scanning parameters

To demonstrate the impact of adjusting the number of radiograph projections (guidelines 2, 3, 4, 6, 8), beam energy and power (5), exposure time (6), and detector binning (7) on the reconstructed 3D image slice, we repeatedly scanned without filtration the same planktonic foraminifer with four parameter permutations. Table 2 lists each scanning parameter, the resulting scan time, and the file size of a single slice from the reconstructed .tiff image stack. All of these slice images are presented in Fig. 3 and have been reconstructed using the same byte-scaling and resized for ease of comparison. The intensity at each brightness (greyscale) value for each of the slice images (Fig. 3a–c) is shown in Fig. 4 and given in Table 3 to illustrate the link between the scanning parameters,

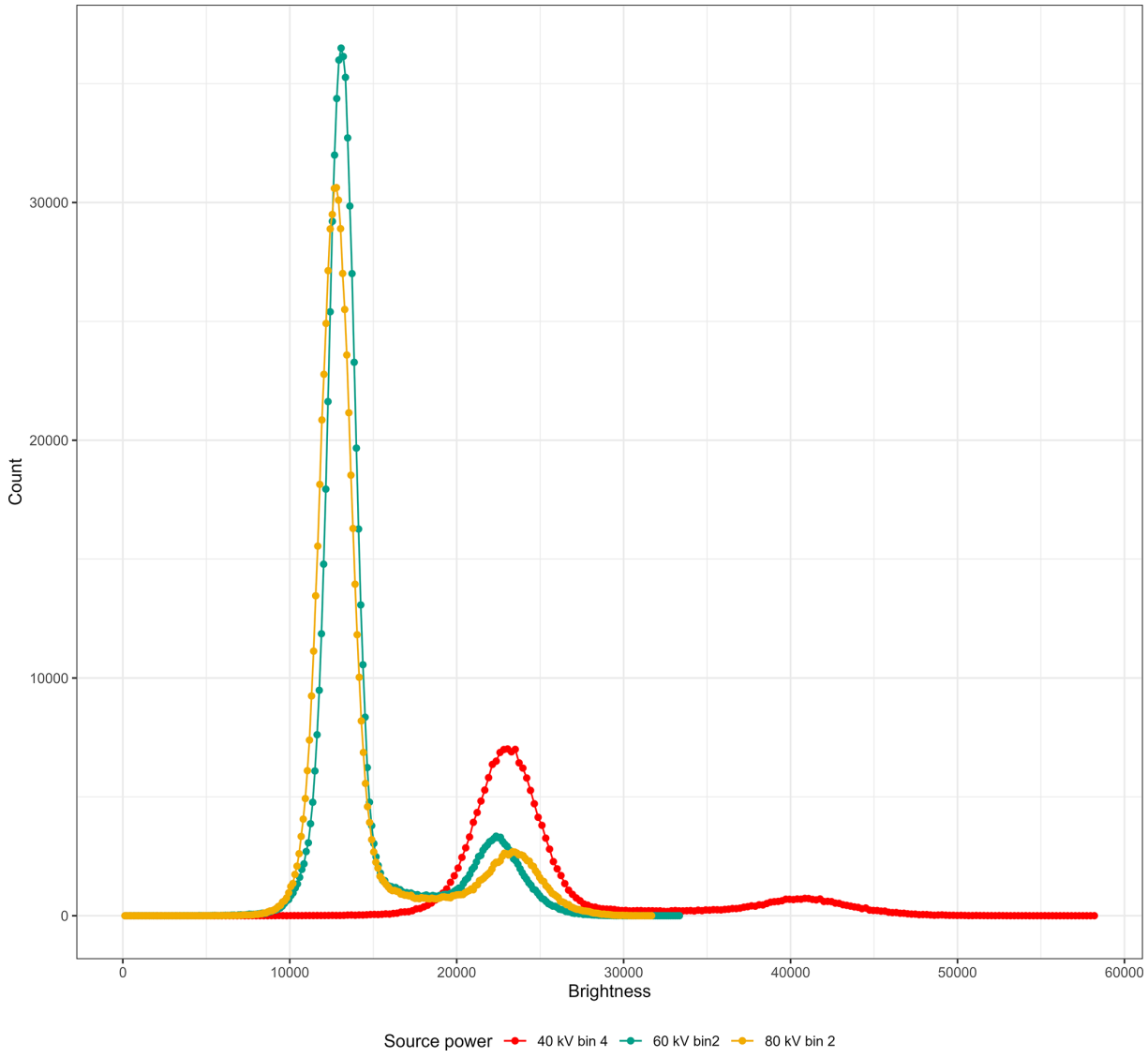


Figure 4. X-ray intensity (counts) by brightness (greyscale) value for the same *Dentoglobigerina altispira* foraminifer embedded in foam within a plastic straw scanned under varying X-ray source power. Scan parameters are given in Table 2, and each is coloured here as **(a)** 40 kV source power and detector binned 4× in red, **(b)** 60 kV source power and detector binned 2× in green, and **(c)** 80 kV source power and detector binned 2× in yellow.

the separation between region-of-interest peaks, and the final reconstructed .tiff image.

A lower-power X-ray beam has a lower signal-to-noise ratio because of the reduced photon intensity and can cause photon starvation in extreme cases. This can lead to grainy or noisy images, making it more challenging to discern fine details such as the edges of the surrounding foam. Figure 3a uses a higher binning to improve detector sensitivity and increase exposure. The combination of highest binning and fewest projections produced the fastest scan time and smallest file size. The resulting image shows a visibly sharp boundary between the foraminifer and the surrounding foam, with the widest absolute brightness separation (contrast) between the two materials of the scans presented here.

Using Dragonfly, we identified the lower and upper brightness bounds for the air, foam, and calcite in each scan (Fig. 3 and Table 3). We identified the identity of each of the three peaks in the image histogram using the identified lower and upper bounds then calculated the brightness separation between the foam and the calcite peaks. While all the scans presented here have sufficient brightness separation (contrast) for isolating the foraminifera calcite shell from the surrounding foam and air using threshold segmentation, a wider contrast aids both scientists and computers in feature extraction.

Having kept the magnification, beam filter, and image processing consistent, the calcite brightness is a consequence of altering one or more of the magnification, exposure, projection, energy, and power parameters. In the presented case study, we note that Fig. 3a, with the setup of the fewest projections and the lowest beam energy and power (guideline 5) with the largest detector binning (guideline 7), results in the fastest scan time (guideline 8) and exhibits the widest contrast between foam and calcite (14 835 grey value or 57 % increase). On the other hand, Fig. 3c, with the setup of the highest beam source power (guideline 5) of the study and with more projections (guideline 4), results in an intermediate scan time with a higher spatial resolution and exhibits the widest relative contrast (8820 grey value or 61 % increase).

We highlight the impact that detector binning has on the level of detail captured in the imaging process, with higher binning values generally resulting in a coarser spatial resolution. The spatial resolution for scan A employs a 4× detector binning configuration, leading to a proportionate reduction in resolution to 1370 nm compared to scans B and C, which employ a 2× detector binning setup, achieving a 685 nm resolution.

The interconnected nature of the beam energy and power, detector binning, and exposure time has been presented here in a case study to optimize each parameter for a single foraminifer and is summarized in Table 4. While scanning a single isolated specimen will likely need one set of parameter values, larger studies involving multiple individuals will mean additional considerations should be made regarding total scan time and consistency across specimens. The parameter values needed to obtain a suitable image vary as widely as the range of specimens that can be scanned by micro-CT as a tool to answer research questions – be they environmental, ecological, or evolutionary.

Code and data availability. The original .tiff images (μ-CT X-ray slice reconstructions), the csv files describing the images, and an R script to process the data and create Fig. 4 are available under a permissive license at https://github.com/alexsb1/Ten_recommendations (Searle-Banes, 2025).

At the end of this research grant, all CT scan data and scripts will be archived with a DOI at the British Oceanographic Data Centre, Southampton, UK (https://www.bodc.ac.uk/, British Oceanographic Data Centre, 2025).

All data will be publicly available and licensed for reuse.

Author contributions. THGE and OLK conceived the project along with IS and MM. ASB and OLK developed X-ray scanning methodology presented as the case study, with additional scans by KR that are presented in Fig. 3. AB picked individual foraminifera and prepared the straws for scanning. ASB drafted the paper, and AB, KR, OLK, and THGE provided comments on earlier drafts.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. The authors acknowledge the μ-VIS X-ray Imaging Centre (https://muvis.org, last access: 21 February 2025), part of the National Facility for laboratory-based X-ray CT (https://nxct.ac.uk/, last access: 21 February 2025) – EPSRC: EP/T02593X/1), at the University of Southampton for provision of tomographic imaging facilities and acknowledge the work of Ian Sinclair for his contributions to the experimental design.

We thank Janet Burke and the anonymous reviewer for their comments on a previous version of this paper.

Financial support. This research has been supported by the Natural Environment Research Council (grant no. NE/P019269/1).

Review statement. This paper was edited by Laia Alegret and reviewed by Janet Burke and one anonymous referee.

References

Briguglio, A., Metscher, B., and Hohenegger, J.: Growth rate biometric quantification by X-ray microtomography on larger benthic foraminifera: three-dimensional measurements push nummulitids into the fourth dimension, Turk. J. Earth Sci., 20, 683–699, https://doi.org/10.3906/yer-0910-44, 2011.

British Oceanographic Data Centre: Marine data sharing and preservation, managed & operated by the National Oceanography Centre, https://www.bodc.ac.uk/ (last access: 12 April 2025), 2025.

Brummer, G.-J. A., Hemleben, C., and Spindler, M.: Ontogeny of extant spinose planktonic foraminifera (Globigerinidae): A concept exemplified by Globigerinoides sacculifer (Brady) and G. Ruber (d'Orbigny), Mar. Micropaleontol., 12, 357–381, https://doi.org/10.1016/0377-8398(87)90028-4, 1987.

Burke, J. E., Renema, W., Schiebel, R., and Hull, P. M.: Three-dimensional analysis of inter-and intraspecific variation in ontogenetic growth trajectories of planktonic foraminifera, Mar. Micropaleontol., 155, 101794, https://doi.org/10.1016/j.marmicro.2019.101794, 2020.

Caromel, A. G. M., Schmidt, D. N., Fletcher, I., and Rayfield, E. J.: Morphological Change During The Ontogeny Of The Planktic Foraminifera, J. Micropalaeontol., 35, 2–19, https://doi.org/10.1144/jmpaleo2014-017, 2016.

de Vargas, C., Norris, R., Zaninetti, L., Gibb, S. W., and Pawlowski, J.: Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic

provinces, P. Natl. Acad. Sci. USA, 96, 2864–2868, https://doi.org/10.1073/pnas.96.6.2864, 1999.

Dragonfly: Comet Technologies Canada Inc, Dragonfly [code], https://dragonfly.comet.tech/ (last access: 12 April 2025), 2022.

du Plessis, A., Broeckhoven, C., Guelpa, A., and le Roux, S. G.: Laboratory x-ray micro-computed tomography: a user guideline for biological samples, GigaScience, 6, 1–11, https://doi.org/10.1093/gigascience/gix027, 2017.

Eggins, S., De Deckker, P., and Marshall, J.: Mg/Ca variation in planktonic foraminifera tests: implications for reconstructing palaeo-seawater temperature and habitat migration, Earth Planet. Sc. Lett., 212, 291–306, https://doi.org/10.1016/S0012-821X(03)00283-8, 2003.

Gooday, A. J., Sykes, D., Góral, T., Zubkov, M. V., and Glover, A. G.: Micro-CT 3D imaging reveals the internal structure of three abyssal xenophyophore species (Protista, Foraminifera) from the eastern equatorial Pacific Ocean, Sci. Rep., 8, 12103, https://doi.org/10.1038/s41598-018-30186-2, 2018.

Görög, Á., Szinger, B., Tóth, E., and Viszkok, J.: Methodology of the micro-computer tomography on foraminifera, Palaeontol. Electron., 15, 1–15, https://doi.org/10.26879/261, 2012.

Hipsley, C. A., Aguilar, R., Black, J. R., and Hocknull, S. A.: High-throughput microCT scanning of small specimens: preparation, packing, parameters and post-processing, Scientific Reports, 10, 13863, https://doi.org/10.1038/s41598-020-70970-7, 2020.

Hsiang, A. Y., Brombacher, A., Rillo, M. C., Mleneck-Vautravers, M. J., Conn, S., Lordsmith, S., Jentzen, A., Henehan, M. J., Metcalfe, B., Fenton, I. S., Wade, B. S., Fox, L., Meilland, J., Davis, C. V., Baranowski, U., Groeneveld, J., Edgar, K. M., Movellan, A., Aze, T., Dowsett, H. J., Miller, C. G., Rios, N., and Hull, P. M.: Endless Forams: $>$ 34,000 Modern Planktonic Foraminiferal Images for Taxonomic Training and Automated Species Recognition Using Convolutional Neural Networks, Paleoceanogr. Paleoclimatol., 34, 1157–1177, https://doi.org/10.1029/2019PA003612, 2019.

Huber, B. T., Bijma, J., and Darling, K.: Cryptic speciation in the living planktonic foraminifer Globigerinella siphonifera (d'Orbigny), Paleobiology, 23, 33–62, https://doi.org/10.1017/S0094837300016638, 1997.

Iwasaki, S., Kimoto, K., Sasaki, O., Kano, H., Honda, M. C., and Okazaki, Y.: Observation of the dissolution process of Globigerina bulloides tests (planktic foraminifera) by X-ray microcomputed tomography, Paleoceanography, 30, 317–331, https://doi.org/10.1002/2014PA002639, 2015.

Irie, M. S., Spin-Neto, R., Borges, J. S., Wenzel, A., and Soares, P. B. F.: Effect of data binning and frame averaging for micro-CT image acquisition on https://doi.org/10.1038/s41598-022-05459-6, 2022.

Johnstone, H. J. H., Schulz, M., Barker, S., and Elderfield, H.: Inside story: An X-ray computed tomography method for assessing dissolution in the tests of planktonic foraminifera, Mar. Micropaleontol., 77, 58–70, https://doi.org/10.1016/j.marmicro.2010.07.004, 2010.

Kearns, L. E., Searle-Barnes, A., Foster, G. L., Milton, J. A., Standish, C. D., and Ezard, T. H. G.: The Influence of Geochemical Variation Among Globigerinoides ruber Individuals on Paleoceanographic Reconstructions, Paleoceanogr. Paleoclimatol., 38, e2022PA004549, https://doi.org/10.1029/2022PA004549, 2023.

Kendall, S., Gradstein, F., Jones, C., Lord, O. T., and Schmidt, D. N.: Ontogenetic disparity in early planktic foraminifers, J. Micropalaeontol., 39, 27–39, https://doi.org/10.5194/jm-39-27-2020, 2020.

Marchant, R., Tetard, M., Pratiwi, A., Adebayo, M., and de Garidel-Thoron, T.: Automated analysis of foraminifera fossil records by image classification using a convolutional neural network, J. Micropalaeontol., 39, 183–202, https://doi.org/10.5194/jm-39-183-2020, 2020.

Mulqueeney, J. M., Searle-Barnes, A., Brombacher, A., Sweeney, M., Goswami, A., and Ezard, T. H. G.: How many specimens make a sufficient training set for automated three-dimensional feature extraction?, Roy. Soc. Open Sci., 11, rsos.240113, https://doi.org/10.1098/rsos.240113, 2024.

Schmidt, D. N., Rayfield, E. J., Cocking, A., and Marone, F.: Linking evolution and development: Synchrotron Radiation X-ray tomographic microscopy of planktic foraminifers, Palaeontology, 56, 741–749, https://doi.org/10.1111/pala.12013, 2013.

Searle-Banes, A.: Ten_recommendations, GitHub [data set], https://github.com/alexsb1/Ten_recommendations (last access: 12 April 2025), 2025.

Vanadzina, K. and Schmidt, D. N.: Developmental change during a speciation event: evidence from planktic foraminifera, Paleobiology, 48, 120–136, https://doi.org/10.1017/pab.2021.26, 2022.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data, 3, 160018, https://doi.org/10.1038/sdata.2016.18, 2016.

TOWARDS UNDERSTANDING SPECIATION BY AUTOMATED EXTRACTION AND DESCRIPTION OF 3D FORAMINIFERA STACKS

Wenshu Zhang[⋆] *Thomas Ezard*[†] *Alex Searle-Barnes*[†] *Anieke Brombacher*[†]
Orestis Katsamenis[†] Mark Nixon[†]

[⋆] Cardiff Metropolitan University [†] University of Southampton

ABSTRACT

The sheer volume of 3D data restricts understanding of genetic speciation when analyzing specimens of planktonic foraminifera and so we develop an end-to-end computer vision system to solve and extend this. The observed fossils are planktonic foraminifera, which are single-celled organisms that live in vast numbers in the world's oceans. Each foram retains a complete record of its size and shape at each stage along its journey through life. In this study, a variety of individual foraminifera are analyzed to study the differences among them and compared with manually labelled ground truth. This is an approach which (i) automatically reconstructs individual chambers for each specimen from image sequences, (ii) uses a shape signature to describe different types of species. The automated analysis by computer vision gives insight that was hitherto unavailable in biological analysis: analyzing shape implies understanding spatial arrangement and this is new to the biological analysis of these specimens. By processing datasets of 3D samples containing 9GB of points, we show that speciation can indeed now be analyzed and that automated analysis from morphological features leads to new insight into the origins of life.

Index Terms— automated shape measurement, machine learning, 3D Krawtchouk moments, understanding speciation, foraminifera

1. INTRODUCTION

Foraminifera are single-celled biomineralizing organisms that live in the ocean. They are very small, typically up to a millimetre in diameter. As a rich marine resource, their Calcium Carbonate shells are ubiquitously retained in sediments in the form of fossils. The internal structure of these fossils is an increasingly important material for systematically studying evolution, because it enables the reconstruction of the growth rate of dead individuals and thus better biological insight into the selection pressures experienced as these organisms lived [1]. The high-resolution micro-CT scan technology enables the non-destructive observation of internal structures.

To investigate how variations among individuals affect evolution, biologists need to collect large numbers of specimens to measure the internal dimensions of these fossils in order to analyze the factors that lead to evolution over a long timeline. Traditional methods require an experienced taxonomist to manually label each chamber of foram for processing. Although accurate data can be obtained in this way, this approach is obviously time-consuming and labour intensive for research projects that demand analysis of thousands of samples. In our work we are using computer vision to enable understanding of variation by using automated measurement of shape of geological specimens, thereby giving new insight concerning spatial layout and arrangement.

In this work, we propose an end-to-end framework to avoid manual effort and remove subjectivity as much as possible. The framework takes a computed tomographic (CT) scanned image sequence as input. The output is the detected chambers with corresponding measurements (see Fig. 1). We use manually labelled images as ground-truth to train a neural network for foram chamber prediction. We reconstruct the space inside each chamber as a surface mesh based on the detected chamber contours, which helps measuring the each chamber's size, volume and elongation. This framework will help biologists to prepare the required measurements with reduced manual effort. The results are reasonable and reproducible. We also propose a new shape signature for foraminifera to perform species recognition.

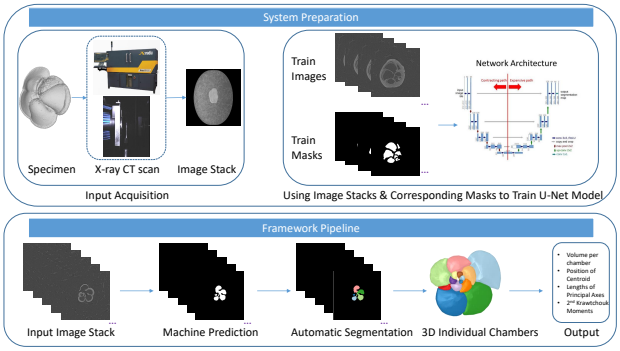


Fig. 1: Framework of the end-to-end automatic foraminifera measurement.

2. RELATED WORK

In this section, we will introduce the case studies of evolution using foraminifera. Followed by that, leveraging computer vision techniques to benefit inter-disciplinary research works will be discussed.

2.1. Speciation

Speciation encapsulates the processes by which new and distinct biological species form. Studying speciation requires data that reveals how differences among individuals become differences among species. The need to study variation mandates large datasets gathered using repeatable protocols that minimise subjective treatments of successive data points.

Dr. Zhang did this work when she was a Research Fellow at University of Southampton. She is currently a Lecturer at Cardiff Metropolitan University.

Under favourable geological conditions, huge numbers of planktonic foraminifera accumulate in continuous sediments and enable study of millions of years of evolutionary history from a single location. The maturity of the group's fossil taxonomy and the abundance of fossil material allows ancestor-descendant relationships to be hypothesized with a high degree of confidence [1]. Multivariate statistical procedures including Gaussian Additive Mixture Models can then be applied to large data sets spanning speciation events to provide quantitative, non-subjective evidence for when the naming of new species corresponds to diagnostic features supporting the emergence of distinct biological forms [2].

2.2. Computer Vision for Foraminifera Study

Image based research A team of researchers at North Caroline State University has conducted a series of works on foraminifera identification. [3] introduced a machine learning method to segment foraminifera images initially targeted at identification. They captured 16 images for each (individual) sample under different illuminations to use the variation for coarse edge map generation. They also trained a convolutional neural network (CNN) and random forest to extract features and refine edge maps to obtain the final chamber segmentation. [4] trained and tested different machine learning algorithms by using the dataset mentioned above. They compared automatic foraminifera identification performance with that of human expertise thus validating results that show the foraminifera can indeed be identified. They have listed more detailed comparison results in [5].

Volume Based Research Although the machine performed competitive results in automated foram identification, it cannot provide a quantitative analysis for each specimen. High resolution X-ray scanning enables a new way of investigating quantification of foraminiferal interiors. [6] explained how X-ray CT scanning with a submicron resolution helps observe inner structure of foram. In [7, 8], the three dimensional reconstructions of the large benthic foraminifera have been introduced. This helps biologists to profile their biometrics and reveal complex shape informations. This type of foram has an obvious chamber shape in most cases. Therefore the reconstructions can be easily done by the popular software like Avizo or ImageJ, with user annotations.

2.3. Moment Based Shape Features

There is an abundance of methods of object recognition which have been well studied. Among them, moment-based techniques are widely used to generate scale, translation and rotation independent features [9]. Krawtchouk moments were first introduced for image analysis by [10], and are based on a set of discrete orthogonal polynomials. This avoids any numerical approximation compared with other traditional moments like Zernike and Legendre moments. According to performance, 3D Krawtchouk moments have been proposed for content-based search and retrieval in [11]. Others have carefully examined their accuracy and efficiency on 3D object analysis and recognition [12]. Recent research [13] on protein local surface shape comparison first employed 3D Krawtchouk moments on irregular shapes instead of well-known 3D shape database. The experimental results in [13] showed that a shape descriptor consists of lower 3D Krawtchouk moments could present a promising protein recognition performance.

2.4. Image Segmentation

As mentioned before, currently foram processing requires much human effort. An automatic individual inner chamber reconstruction is much desired. Normally, segmenting and numbering chambers in each slice is the most intensive work. Fortunately, there has been the rapid development of machine learning techniques, especially on applications of biomedical image segmentation [14, 15]. These examples successfully pioneered computer aided cross-disciplinary research. U-Net is one of the most popular convolutional networks for biomedical image segmentation . The network has excellent performance on segmentation tasks even when trained using few images. It inspires our study as only fewer than 1000 manually labelled images are available for training a network.

3. METHODOLOGY

We propose an end-to-end framework of automatic measuring foraminiferal chambers in this paper. The framework takes X-ray CT scanned image sequences as input and returns individual chamber shape measurements. Importantly, we seek to isolate distinct subcomponents (growth chambers) within each sample (individual) to provide a more coherent representation of the functional constraints of the image and organism under investigation. The detailed structure is demonstrated in Fig. 1.

3.1. Preparation

In our study, we have a high resolution image stack for each specimen shape reconstruction and measurement. To avoid manual labelling, we train a neural network to automatically predict possible chambers. The trained data is randomly chosen from the prepared image stacks. The corresponding masks used for training are manually labelled. The manual processing for one specimen can take up to one day. Therefore we only have several image stacks processed for training. U-Net [16] is then selected as a segmentation network in this paper since it requires only a few images. The trained model is employed to generate image masks to initially segment possible chamber regions from the original images.

3.2. Chamber Tracing

As a single-cell organism, all the chambers in foraminifera share the same corridor to pass food, which will form openings between adjacent chambers when observing from a 2D image. In this situation, the trained network cannot provide a precise prediction. The connected region will need an additional partition to get individual chamber contour.

Assuming all the chambers have been segmented before the current image slice, we can place the centroids of each chamber from last one as seed points to initialize segmentation. For each pixel inside the connected area, we calculate a geodesic distance to every seed point. Those pixels with the shortest distance to the same seed point will be clustered together. The results are shown in Fig. 2.

3.3. Shape Measurement

To investigate evolutionary rates and patterns on multi-million-year time scales, a large number of statistics on quantifying foraminiferal growth is necessary. The fundamental information includes growth rates as represented by volume and equivalent radius (principal axis length). In this paper, we use lower order of Geometric Moments

(a) (b) (c) (d) (e)

Fig. 2: Further segmentation on predicted mask. Use the centroids of previously segmented chambers (a) as the initial seed points on current mask (b) to run partition (c). Use the centroids of current segmented chambers (d) to divide next mask (e).

to capture basic geometric information per chamber for individual specimen.

We use the image masks obtained from section 3.1 to define a binary volume function $f(x, y, z)$ on each specimen (Voxelization). After that, we compute Krawtchouk polynomials and apply them on the aforementioned volume function to calculate 3D weighted Krawtchouk moments $\bar{Q}_{nml}$ of $f(x, y, z)$ up to order $(n + m + l)$ (see Eq. 1 introduced by [11]). This translation, rotation and scaling independent moments will form a shape signature and be discussed in next section.

$$\bar{Q}_{nml} = \sum_{x=0}^{N}\sum_{y=0}^{M}\sum_{z=0}^{L} f(x, y, z) \cdot \bar{K}_n(x; p_x, N) \cdot \bar{K}_m(y; p_y, M) \cdot \bar{K}_l(z; p_z, L). \quad (1)$$

4. EXPERIMENT AND DISCUSSION

4.1. Compare Manual Labelling & Machine Learning Prediction

Normally, there is no standard rule for segmenting individual chambers inside the foram, as long as the connected space has been isolated. We compared the chamber contours from manual labelling with machine prediction, with examples of both in Fig. 3a and 3b.

The main limitation of this work is loss of volume due to imperfect chamber mask prediction. There is an average loss of 15.5% in our experiment, which compares with a previous difference analysis [6] which observed a similar error rate of 15%. It is of course not known from which source the differences derive: the manual or the automated labelling. However, the performance of deep learning prediction relies on the quality of scanned specimens. There are two situations in which the volume loss increases dramatically. The first one is when the samples have a broken surface, there is no chance of generating a correct prediction (see Fig. 4a and 4b). Second is when there are too many sediment particles inside the chambers, the prediction generates a shrunk chamber mask (see Fig. 4c and 4d). It is very tricky to study the correct growth rates from these imcomplete samples. When holes occured on the largest chamber, we truncated the last one as some research believes that the last two chambers will not affect growth rate analysis and can confuse taxonomic identification [17]. And for samples with large sediment data inside chambers, future work will enable the framework with image shadow removal to improve the quality of input.

4.2. Growth Pattern

To understand the biological question of how do differences among individuals make differences among species, we will not only analyze the growth pattern of a single specimen but also compare the patterns as groups to study speciation. Fig. 3c shows the chamber by chamber growth pattern of individuals from three different species, *M.pertenius*, *M.exilis* and *M.limbata* respectively.

According to the observed data, the largest size of the three species is not detectably different (Wilcox Rank Sum Tests on chamber Q: *M.pertenius* vs *M.exilis* W = 144, p = 0.594; *M..limbata* vs *M.exilis* W = 103, p = 0.287; *M.limbata* vs *M.pertenius* W = 94, p = 0.336), but this masks clear differences in their growth rates: the smallest chambers of *M.exilis* are smaller than those of *M.pertenius* (coef = 0.0042, s.e. = 0.0002, p<0.05) and *M.limbata* (coef = 0.0054, s.e. = 0.0002, p<0.01). There are no detectable differences in linear growth rates (coef = $1e^{-5}$, s.e. = $3e^{-5}$, p>0.05 for *M.pertenius* & *M.exolis*) but *M.limbata* has a faster quadratic growth rate (coef = 0.042, s.e. = 0.016, p<0.01) than *M.exilis*. The differences among growth rates and in the earliest lifestages emphasizes the need to study developmental sequences to reveal the structure of variation among individuals.

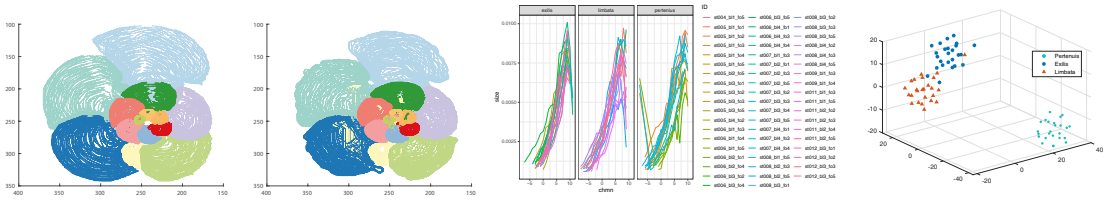
4.3. Shape Signature

The chamber's growth rate is a fundamental property for distinguishing different species because it represents changes in the organism's pace of life. However, we try to investigate the intrinsic shape information of forams during the evolution. We proposed a shape signature for foraminifera to describe the appearance of shape. The 3D Krawtchouk moments up to second order are collected as the shape signature. The binary volume function $f(x, y, z)$ is defined on a grid of $512 \times 512 \times 512$ for every specimen in order to normalize them under the same scale. Because we introduced translation, rotation and scale independent 3D Krawtchouk moments, the specimens have the identical zero order and first order of 3D Krawtchouk moments. Thus we omit them to only compare the second order of 3D Krawtchouk moments in the experiment.

There are 75 specimens randomly chosen from our dataset to validate the hypothesis and 25 of each species *M.pertenius*, *M.exilis* and *M.limbata*. In Fig. 3d, the t-SNE algorithm [18] is used to reduce the signature dimension and visualize the clusters. According to this, the *M.pertenius* has a clear difference on shape appearance compared to its ancestor *M.exilis* and *M.limbata*. Meanwhile, the boundary between clusters *M.exilis* and *M.limbata* is not well defined. There are only few samples falling into incorrect clusters. Because the species label is manually assigned to each specimen, human bias will be one of the reasons of generating this problem.

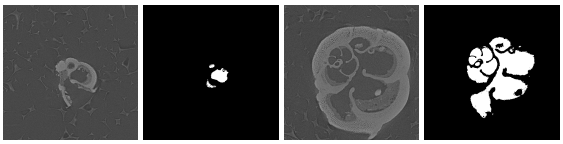
5. CONCLUSION

This paper is an introductory work in understanding the important factors affecting speciation by leveraging computer vision algorithms to reconstruct the internal features of computed tomographic scans. We proposed an end-to-end framework of automatic measurement of foraminiferal shape, by bringing the vast armoury of computer vision to bear on an evolutionary question, does developmental plasticity inhibit speciation? The experimental results reveal the individual life cycle as well as the shape patterns differ between species. In addition, the reproducible shape measurements and signatures reduced human bias. Our protocol provides an efficient access to generate and analyze biological trait based on a large number of samples. Our automated processes provide for analysis of spatial arrangement which has already given new insight into growth processes. As this information is new to biologists, we await with interest their development and analysis of this new information. Future work will focus on improve the performance of



(a) (b) (c) (d)

Fig. 3: Experiment & Discussion. (a) Chamber contours generated from manual labelling. (b) Chamber contours generated from machine learning prediction. (c) Growth patterns of *M.exilis*, *M.limbata* and *M.pertenius*. (d) Dimension reduction and visualization of foram shape signature.



(a) (b) (c) (d)

Fig. 4: Example of imperfect samples. (a) Broken sample. (b) Incorrect prediction. (c) With sediment. (d) Incorrect prediction.

machine learning and expand the number of specimens and diversity of morphological forms as broadly as possible.

6. ACKNOWLEDGEMENT

This work was funded by the Natural Environment Research Council award NE/P019269/1. The authors thank μ-VIS X-ray Imaging Centre at University of Southampton for supporting micro-CT scanning of foreminifera.

7. REFERENCES

[1] T Aze, T HG Ezard, A Purvis, H K Coxall, D RM Stewart, B S Wade, and P N Pearson, "A phylogeny of cenozoic macroperforate planktonic foraminifera from fossil data," *BIOL REV*, vol. 86, no. 4, pp. 900–927, 2011.

[2] P N Pearson and T HG Ezard, "Evolution and speciation in the eocene planktonic foraminifer turborotalia," *Paleobiology*, vol. 40, no. 1, pp. 130–143, 2014.

[3] Q. Ge, B. Zhong, B. Kanakiya, R. Mitra, T. Marchitto, and E. Lobaton, "Coarse-to-fine foraminifera image segmentation through 3D and deep features," in *SSCI*. IEEE, 2017, pp. 1–8.

[4] B. Zhong, Q. Ge, R. Kanakiya, R. Mitral T. Marchitto, and E. Lobaton, "A comparative study of image classification algorithms for foraminifera identification," in *SSCI*. IEEE, 2017, pp. 1–8.

[5] R Mitra, TM Marchitto, Q Ge, B Zhong, B Kanakiya, MS Cook, JS Fehrenbacher, JD Ortiz, A Tripati, and E Lobaton, "Automated species-level identification of planktic foraminifera using convolutional neural networks, with comparison to human performance," *MAR MICROPALEONTOL*, vol. 147, pp. 16–24, 2019.

[6] D. Van Loo R. Speijer, B. Masschaele, J Vlassenbroeck, V Cnudde, and P Jacobs, "Quantifying foraminiferal growth with high-resolution x-ray computed tomography: New opportunities in foraminiferal ontogeny, phylogeny, and paleoceanographic applications," *Geosphere*, vol. 4, no. 4, pp. 760–763, 2008.

[7] A. Briguglio, J. Hohenegger, and G. Less, "Paleobiological applications of three-dimensional biometry on larger benthic foraminifera: a new route of discoveries," *J FORAMIN RES*, vol. 43, no. 1, pp. 72–87, 2013.

[8] W. Renema and L. Cotton, "Three dimensional reconstructions of nummulites tests reveal complex chamber shapes," *PeerJ*, vol. 3, pp. e1072, 2015.

[9] RJ Prokop and AP Reeves, "A survey of moment-based techniques for unoccluded object representation and recognition," *CVGIP*, vol. 54, no. 5, pp. 438–460, 1992.

[10] P-T Yap, R Paramesran, and S-H Ong, "Image analysis by Krawtchouk moments," *IEEE T. IMAGE PROCESS*, vol. 12, no. 11, pp. 1367–1377, 2003.

[11] A Mademlis, A Axenopoulos, P Daras, D Tzovaras, and M G Strintzis, "3D content-based search based on 3D Krawtchouk moments," in *3DPVT'06*. IEEE, 2006, pp. 743–749.

[12] A Mesbah, M El Mallahi, Z Lakhili, H Qjidaa, and A Berrahou, "Fast and accurate algorithm for 3D local object reconstruction using Krawtchouk moments," in *ICMCS*. IEEE, 2016, pp. 1–6.

[13] A Sit, W-H Shin, and D Kihara, "Three-dimensional Krawtchouk descriptors for protein local surface shape comparison," *Pattern Recognition*, vol. 93, pp. 534–545, 2019.

[14] X Yang, L Yu, S Li, X Wang, N Wang, J Qin, D Ni, and P-A Heng, "Towards automatic semantic segmentation in volumetric ultrasound," in *MICCAI*. Springer, 2017, pp. 711–719.

[15] K Yan, M Bagheri, and R M Summers, "3D context enhanced region-based convolutional neural network for end-to-end lesion detection," in *MICCAI*. Springer, 2018, pp. 511–519.

[16] O Ronneberger, P Fischer, and T Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.

[17] S Spezzaferri, M Kucera, P N Pearson, B S Wade, S Rappo, C R Poole, R Morard, and C Stalder, "Fossil and genetic evidence for the polyphyletic nature of the planktonic foraminifera globigerinoides, and description of the new genus trilobatus," *PLoS One*, vol. 10, no. 5, pp. e0128108, 2015.

[18] L Maaten and G Hinton, "Visualizing data using t-sne," *JMLR*, vol. 9, no. Nov, pp. 2579–2605, 2008.