Size-dependent response of foraminiferal calcification to seawater carbonate chemistry

Michael J. Henehan[1], David Evans[1,2], Madison Shankle[1], Janet Burke[1], Gavin L. Foster[3],
Eleni Anagnostou[3], Thomas B. Chalk[3], Joseph A. Stewart[3,4], Claudia H. S. Alt[3,5], Joseph Durrant[3], and
Pincelli M. Hull[1]

[1]Department of Geology and Geophysics, Yale University, 210 Whitney Avenue, New Haven CT-06511, USA
[2]Department of Earth Sciences, University of St. Andrews, Irvine Building, North Street, St. Andrews, Fife, KY16 9AL, UK
[3]Ocean and Earth Science, University of Southampton, National Oceanography Centre Southampton, Southampton, SO14 3ZH, UK
[4]National Institute of Standards and Technology, Hollings Marine Laboratory, 331 Ft. Johnson Road Charleston, SC-29412, USA
[5]Department of Biology, College of Charleston, Charleston SC-29424, USA

Correspondence to: Michael Henehan & David Evans (michael.henehan@yale.edu; de32@st-andrews.ac.uk)

NOTE: Changes in response to Reviewer 1 are marked in red, while changes in response to Reviewer 2 are shown in blue. Line numbers for changes in response to Reviewer 2 are given in our Reply to their comment. For Reviewer 1, see (Line numbers in the resubmitted manuscript, not this annotated version):

Page 4, Line 20

5 Page 7, Line 7, Line 14

Page 8, Lines 10-12, Line 30, 31

Page 9, Line 31 - Page 10, Line 25

Page 10, Lines 28-29, 30-31

Page 13, Line 21, Line 22

10

Abstract.

The response of the marine carbon cycle to changes in atmospheric CO_2 concentrations will be determined, in part, by the relative response of calcifying and non-calcifying organisms to global change. Planktonic foraminifera are responsible for
15 a quarter or more of global carbonate production, therefore understanding the sensitivity of calcification in these organisms to environmental change is critical. Despite this, there remains little consensus as to whether, or to what extent, chemical and physical factors affect foraminiferal calcification. To address this, we directly test the effect of multiple controls on calcification in culture experiments and core-top measurements of *Globigerinoides ruber*. We find that two factors, body size and the carbonate system, strongly influence calcification intensity in life, but that exposure to corrosive bottom waters can overprint
20 this signal post mortem. Using a simple model for the addition of calcite through ontogeny, we show that variable body size between and within datasets could complicate studies that examine environmental controls on foraminiferal shell weight.

1

In addition, we suggest that size could ultimately play a role in determining whether calcification will increase or decrease with acidification. Our models highlight that knowledge of the specific morphological and physiological mechanisms driving ontogenetic change in calcification in different species will be critical in predicting the response of foraminiferal calcification to future change in atmospheric $p\text{CO}_2$.

1 Introduction

Calcium carbonate (CaCO_3) production and transport to the deep ocean (the so-called 'carbonate pump') is one of the most important sinks of carbon, acting across a range of geological timescales (Sigman et al., 2010; Berner and Raiswell, 1983). In the Cenozoic (0-66 Ma), biogenic CaCO_3 production by foraminifera, coccolithophores and coral reef ecosystems comprises the vast majority of marine carbonate production (Ridgwell and Zeebe, 2005). The strength of this carbonate pump can be altered in three principle ways: 1) by changing the efficiency of inorganic and/or organic carbon export and burial, 2) by changing the absolute or relative abundance of calcifying and non-calcifying taxa, and 3) by changes in the calcification efficiency of marine calcifiers. All three factors are thought to be sensitive to environmental conditions (e.g. Beaufort et al., 2011; Barker and Elderfield, 2002), although the exact nature of this environmental sensitivity remains unclear. Here we use a series of culturing experiments to specifically address how pH change can influence the extent to which foraminifera calcify their tests (i.e. their 'calcification intensity').

Metrics of calcification in planktonic foraminifera have already been the subject of much scientific attention because of both the importance of foraminifera to the global carbonate burial flux (32-80% of the total deep marine calcite budget; Schiebel, 2002) and the potential of these metrics to act as proxies for changes in marine carbonate chemistry. In this manuscript we will refer to several types of related (but distinct) metrics that have been used to describe calcification in foraminifera, and so for clarity these are summarised in Table 1. Foraminiferal size-normalised weight (SNW) has variously been suggested as a tracer of both changes in deep sea carbonate dissolution (e.g. Lohmann, 1995; Broecker and Clark, 2001) and as a proxy for changes in the surface ocean carbonate system (e.g. Barker and Elderfield, 2002; Bijma et al., 2002; Moy et al., 2009; Marshall et al., 2013). As a proxy of surface conditions, it is thought that the SNW of foraminiferal tests reflects carbonate ion concentration in the shallow waters inhabited during their lives. Conversely, studies of sea-floor dissolution relate SNW to carbonate diagenesis after death in undersaturated bottom waters. In both cases, studies implicitly assume that either the conditions during life, or conversely, after death, have a relatively minor effect on SNW, or can be accounted for by other means (as discussed by Bijma et al., 2002; Barker and Elderfield, 2002). Whilst both culture and field studies support a surface carbonate system control on foraminiferal calcification (Bijma et al., 2002; Barker and Elderfield, 2002; Russell et al., 2004; Lombard et al., 2010; Marshall et al., 2013), others have observed secondary environmental controls on SNW such as nutrient availability or temperature (e.g. Aldridge et al., 2012; Weinkauf et al., 2016). Furthermore, other studies observe an inverse response of SNW to carbonate system change in some species (Beer et al., 2010b; Foster et al., 2013) - that is, a greater test thickness at lower pH and/or $[\text{CO}_3^{2-}]$.

It is possible that at least some of the discrepancies described above may stem from methodological differences, since foraminiferal SNW has been quantified in a number of different ways (see Beer et al., 2010a; Aldridge et al., 2012; Marshall et al., 2013, for further discussion). Many early studies used sieve-based weight measurements, where SNW is calculated as the measured mass of pooled individuals within a set sieve-size fraction divided by the number of individual tests (e.g. Lohmann, 1995; Broecker and Clark, 2001). However, shell size can vary within a studied sieve range (Beer et al., 2010a). Many later studies circumvented this issue by directly measuring the major axis length (Barker and Elderfield, 2002; Aldridge et al., 2012; Beer et al., 2010a) or cross-sectional area (Beer et al., 2010b; Marshall et al., 2013, 2015; Osborne et al., 2016; Weinkauf et al., 2013, 2016) of each individual within a sampled population. However, as discussed by Weinkauf et al. (2016), an assumption common to most shell-weight studies is that SNW metrics themselves do not vary as a function of size – which is unlikely to be true. The predominant model of foraminiferal biomineralisation posits that every time a new chamber is added, foraminifera thicken the calcite of their previous chambers (e.g. Be and Lott, 1964; Erez, 2003), so that a given chamber will appear increasingly heavily calcified over the course of an individual foraminifera's life. It is therefore possible that size may contribute to variability in calcification responses recorded between and within studies.

Attempts to reconcile these varied experimental and observational SNW data with each other and with foraminiferal biomineralisation models (de Nooijer et al., 2014, and references within) are still broadly lacking. To address this shortfall, we show here how a simple calcification model can be used to provide a theoretical framework for SNW metrics. We then present new observations from core-top measurements and culture experiments in the shallow-dwelling symbiont-bearing species *Globigerinoides ruber* (Henehan et al., 2013; Evans et al., 2016a, b), in the light of this new modelled framework. We discuss the implications of our modelling and empirical observations both for explaining the often conflicting results in previous studies, and for predicting the response of planktonic foraminifera to future global change.

2 Methods

2.1 Culturing

Data from *Globigerinoides ruber* (white) used in this study are collated from numerous experiments across a range of temperature, pH and major ion seawater chemistries, cultured at the Interuniversity Institute of Eilat between January 2010 and November 2013. see Reply to reviewer 1 These cultures include both *sensu stricto* and *sensu lato* morphotypes (Wang, 2000). A detailed description of culturing methods is provided elsewhere (Henehan et al., 2013; Evans et al., 2016a, b). Briefly, for all experiments, foraminifera were towed from the Gulf of Aqaba (Eilat) (depth <20m, temperature 22-24 °C, salinity ∼ 40.4 psu), and cultured in individual 120 mL airtight flasks within temperature-controlled water baths. Illumination was provided by a metal halide lamp (420 W) at levels of ∼200 mmol photons m^{-2}s^{-1} (13 h light:11 h dark), equivalent to irradiance at 15-20 m depth in the open waters of the northern Gulf of Aqaba (Shaked and Genin, 2006). Every 1-2 days, individuals were transferred to a petri dish, measured using an optical micrometer under a Zeiss inverted light microscope, and fed a juvenile brine shrimp.

After gametogenesis, foraminifera were rinsed in deionised water, dried and stored for subsequent analysis.

Here we consider changes in calcification intensity in foraminifera that were originally cultured to investigate two different geochemical proxy systems: boron isotopes and Mg/Ca (Henehan et al., 2013; Evans et al., 2016a, b). The experimental design varies slightly between these two cases, so we discuss any such differences where they arise. In all experiments, culture solution pH was determined using an electrode calibrated against NBS buffers. The same electrode was also used to measure the pH of a range of prepared seawater solutions which were subsequently analysed for Dissolved Inorganic Carbon (DIC) concentrations and Total Alkalinity (TAlk), allowing us to cross-calibrate our electrode-derived pH values against calculated pH on the total scale using CO2sys.m (van Heuven et al., 2011), the constants of Dickson (1990), Lueker et al. (2000), and boron concentrations from Lee et al. (2010). Because pH control for boron isotope experiments was paramount, in these experiments each individual flask pH was measured every 2-3 days, and flasks that experienced pH drift had culture seawater solution replaced from a stock solution stored throughout the experiment in airtight bottles without headspace. Uncertainty on pH is therefore calculated as two standard errors on the mean of each culture flask pHs during cultures. For those experiments intended to test Mg incorporation (Evans et al., 2016a), pH monitoring during culture was less frequent, and this is reflected in a more conservative approximation of pH uncertainty in these cultures. In total, we collate calcification intensity data from 11 separate culture experiments, with temperatures ranging from 22.8 to 27.8 °C, seawater Mg/Ca ratios ranging from 2.17 to 6.25, and pH (total scale) ranging from 7.54 to 8.20. Some estimates of future anthropogenic ocean acidification suggest a pH drop of 0.5 units by 2100 (Caldeira and Wickett, 2003), and so this wide pH range allows us to investigate the possible changes in calcification over the next century.

2.2 Deriving Weights from Cultures

Dried cultured foraminifera were imaged and major and perpendicular axes measured using Macnification software (Orbicule Inc., v2.0). They were then weighed individually on microbalances at Royal Holloway University of London (Exps. DE3 & DE4), Yale University (Exp. MH2), and University of Bristol (Exp. MH1). Mean uncertainty assessed by 2 standard deviations of triplicate measurements of individual foraminifera was $< \pm 1$ μg. While these foraminifera were not ashed to remove any remnant organic matter prior to weighing, previous comparisons of non-ashed sample weights with mass of $CaCO_3$ determined by ICPMS showed no significant difference (Henehan et al., 2013).

Since it is unfeasible to take direct shell weight measurements from live pre-culture foraminifera without harming the organism, we estimate pre-culture shell mass from test size (major and perpendicular axes measured via ocular micrometer) at the beginning of culture, using a locally-defined size-weight relationship for *G. ruber*, as in previous studies (e.g. Kisakürek et al., 2008; Lombard et al., 2010; Henehan et al., 2013). Here we expand on our previous size-weight calibration dataset, combining a total of 205 measurements from individuals of *G. ruber* (see Reply to reviewer 1 mixed morphotypes, ranging in size from 141 - 517 μm) towed from the Gulf of Aqaba (Eilat). The equation of this line ($R^2 = 0.61$) is given in Equation 1

4

below.

$$\text{shell mass (in } \mu\text{g)} = 217 \times (\text{product of axes, in mm}^2)^{1.43} \tag{1}$$

2.3 Defining a calcification metric for cultured foraminifera

Existing metrics for calcification (as discussed above in Sect. 1), cannot be applied directly to laboratory cultures of planktonic
species, as many chambers are precipitated prior to collection. Previous studies have therefore used mean test weight of cultured
foraminifera from a given size range (e.g. Bijma et al., 2002), or made corrections for pre-culture mass and time spent in
culture (Lombard et al., 2010) to describe how calcification responded to culture conditions. Here, we developed a new metric,
calcification intensity (CI; Eq. 2), that accounts for both the size/mass of foraminifera upon collection, as well as the (often
differential) amount of mass added between culture experiments:

$$\text{CI} = \frac{\Delta\text{mass}}{\Delta\text{area}} \tag{2}$$

where Δmass is the difference in mass between the start and end of the culture (expressed in μg) and Δarea is the difference
in product of the major and minor axes (in mm^2) between the start and end of the culture. We quantify calcification in cultured
foraminifera in this simple way because it allows for more complete consideration of mass grown outside of culture, and relies
only on pre- and post-culture dimensions and mass that are routinely measured. We note that for our experiments, pre-culture
mass was estimated from a relationship of size to mass (Eq. 1) constructed from foraminifera towed from in the Gulf of Aqaba
(Eilat) at the same time as our culture experiments. pH measured at the time of sampling these tows was 8.10 ± 0.05 (2se),
approximately at the middle of our range of experimental pH. Thus while the size-mass calibration presented here is suitable for
our investigations, we suggest that future culture studies considering CI should characterise this relationship in the populations
from which their cultured individuals were sampled.

2.4 Ontogenetic modelling of calcification intensity

Planktonic foraminifera are single-celled eukaryotes with calcium carbonate tests that show distinct morphological changes
throughout ontogeny (Schmidt et al., 2013; Brummer et al., 1987). Foraminifera grow by sequentially adding calcium carbonate
chambers along a primary coiling axis, and it is thought that they lay down an extra layer of calcite ('secondary calcite') over
existing chambers when a new chamber is formed (Be and Lott, 1964; Hemleben et al., 1989). Over ontogeny coiling behaviour
often changes, as do the relative size, shape, thickness and perforations (i.e., 'porosity') of the calcium carbonate chambers (e.g.
Schmidt et al., 2013). Finally, when foraminifera reproduce and die, many species (although not *G. ruber*; Caron et al., 1990)
precipitate a thick final layer of carbonate known as gametogenic calcite. Other species appear to secondarily thicken their tests
following precipitation of their final chamber (Spero et al., 2015; Fehrenbacher et al., 2017), although this has not yet been
observed in *G. ruber*. Studies of calcification in foraminifera must therefore disentangle the effects of various environmental
factors from these known ontogenetic phenomena. In open-ocean studies (e.g. sediment trap, core-top), SNW measurements are
commonly taken from within narrow size ranges to minimise the effect of size-dependent variation in calcification (as discussed

in Beer et al., 2010a; Weinkauf et al., 2016). However, normalisation for size effectively assumes that there is no ontogenetic variation in calcification or that all individuals come from the same ontogenetic stage, which is unlikely, particularly when comparing results across different studies. In laboratory cultures, there are additional difficulties in measuring CI as it is generally not possible to select a narrow starting size range (given specimen limitation), and the number of chambers added by each individual over the course of culture experiments is highly variable. Therefore, to provide a quantitative framework for exploring the relationship between CI and ontogeny, and to explore how existing calcification metrics may be biased by the use of large or variable size fractions, we developed an ontogenetic model of CI using empirical observations from *G. ruber*.

2.4.1 Model parameter constraints

To ground the model in empirical observations, morphological measurements of 39 specimens were taken from a natural *G. ruber* population from a Woods Hole Oceanographic Institution core-top sample (sample KC78) from the equatorial Atlantic (5.267°N by 44.133°W, 3273 m water depth; Sun et al., 2006). Selected specimens range in size from 250-600 μm (150-250 μm (n=11), 250-300 μm (n=11), 300-425 μm (n=10), and 425-600 μm (n=7)). Foraminifera were first mounted and imaged using a light microscope and ImageJ (www.imagej.net), so that major and minor axis measurements could be taken from as many chambers as possible, as well as from the whole test. Individual chambers were subsequently broken and removed so that chamber wall cross sections could be imaged and measured. Wall thickness measurements were made away from sutures and chamber apertures, as wall thickness often varies in these regions of the test (Bé, 1980). All reported values are the mean of three replicate measurements, and are given in Supplementary Table 1.

2.4.2 Model description

Building upon the measurements described in Sect. 2.4.1 and other published data (listed in Table 2), we designed a simple calcification intensity model in Matlab that tracks the cumulative calcium carbonate in an idealised individual of *G. ruber* throughout its life cycle. For a full description of the model, see Appendix A (note the Matlab code also accompanies this paper). In short, this model simulates the addition of $CaCO_3$ mass with growth, as scaled to numerous morphological and wall-thickness parameters. Morphological parameters used to determine carbonate content include chamber size, chamber shape, and chamber overlap. Wall thickness parameters include the thickness of the initial chamber wall prior to any subsequent thickening (the 'primary wall'), wall porosity, and the thickness of secondary layers added to preceding chambers. There is no *a priori* knowledge of which (if any) morphological parameter(s) we should expect to vary in response to environmental conditions. To address this, we allowed each parameter to vary randomly within set tolerances and ran the model 10^6 times. With a set of model runs this large, we could identify which factors were systematically linked to change in CI irrespective of concomitant random changes in others. This is an important advantage of this computationally-inexpensive model over more complex and sophisticated foraminiferal growth models (e.g. Berger, 1969; Signes et al., 1993). Another critical aspect of our model is that it captures the non-linear behaviour of calcification with ontogenetic growth. Three variant forms of the model (each ran 10^6 times) are presented here, with wall thickness parameterised either as a function of size (Models 1 and 2) or of

chamber number (Model 3). All model variants are built upon open ocean measurements, and are screened against open-ocean populations to ensure they are representative (see Appendix A for a more detailed treatment).

We use the CI-models for two primary ends. First, we explore the effect of ontogeny on CI given size-dependent and chamber-dependent variation in primary wall thickness. Second, we use the ontogenetic predictions to correct for variable

5   chamber addition in culture, given the importance of secondary calcification on shell weight (e.g. Be and Lott, 1964; Erez, 2003). For instance, a foraminifera possessing 8 chambers that is of similar overall test size to a 7-chambered individual will have layered their previous chambers with calcite more times, leading to higher CI at the same body size. Control for variable chamber addition in culture was achieved by simulating the change in CI with each chamber addition (ΔCI) for a foraminifera growing from a set starting size (a maximum axis of 100 μm) using all feasible shell growth models.

10   ## 2.5   Core-top Sampling

To supplement modelling and culturing, and to examine the extent to which the physiological controls observed in culture are preserved in fossil assemblages, we also examined SNW in core-top *G. ruber*. Core-top sites and locations are given in the Supplementary Information, and span a range of bottom water calcite saturation state ($\Omega_{calcite}$) from 0.78 to 4.06. Pre-industrial surface ocean carbonate system conditions for each site were estimated using modern alkalinity relationships (from Lee et al.,

15   2006), air-sea pCO_2 disequilibrium (from Takahashi et al., 2009; Gloor et al., 2003) and local hydrography (from Takahashi et al., 2009; Garcia et al., 2010) following Henehan et al. (2013). Deep ocean carbonate chemistry and calcite saturation state were calculated from DIC and TAlk estimates at depth from Goyet et al. (2000). Carbonate system calculations were carried out as in Section 2.1 above, with pressure corrections from Millero (1995) according to each core site's bathymetry.

20   For size-normalisation of shell weight measurements, we follow the 'area density' (ρ_A) approach of Marshall et al. (2013), as used elsewhere (Weinkauf et al., 2013, 2016; Marshall et al., 2015; Osborne et al., 2016). From each core-top sample, individual specimens (n = 8-26) of *G. ruber sensu stricto* and *sensu lato* were picked from discrete sieve size fractions, imaged (umbilical side up), and weighed, and their cross-sectional or silhouette area determined using Macnification (Orbicule Inc., v2.0). ρ_A was calculated as mass (in μg) over silhouette area (in mm^2). As in Marshall et al. (2013) and Osborne et al. (2016),

25   the mean ρ_A and silhouette area in each core-top sample was then used in multivariate statistical analysis.

3   Results

3.1   Modelling ontogenetic trends and intrinsic drivers of CI and ρ_A

Model parameter combinations within the prescribed tolerances of the size-weight ratios seen in non-cultured populations of *G. ruber* (white, see Reply to reviewer 1 mixed morphotype) are displayed in Fig. 1, panels a,d,g. Using the validated subsets

30   of each model, we calculated CI as it evolved through the ontogenetic growth of each individual modelled foraminifera (Fig 1, panels b,e,h). All models suggest that CI should increase rapidly with test size up to size of $\sim$0.05 mm^2 (equating to a major

axis of approximately 210 μm). Beyond this size there is considerably more scope for inter-individual (i.e. inter-model run) variability, but most model runs continue to show increasing CI with size throughout the remainder of their ontogeny.

From the perspective of foraminiferal cultures, these models predict a strong dependency of CI on foraminifer size on collection, and also the number of chambers precipitated in culture. For instance, cultures in which foraminifer added three chambers

5 on average are expected to have a higher CI than cultures in which most foraminifera added two chambers. Panels (c), (f) and (i) of Fig. 1 explore this effect of chamber addition on CI. For each model, frequency distributions of modelled change in CI (ΔCI) after two and three chamber additions are shown, relative to a baseline addition of one-chamber and a starting size of 100 μm maximum axis diameter. The median increase of calculated CI with each chamber addition is 17 μg mm^{-2}, and was largely insensitive to the base model – a similar dependency of chamber addition on CI was observed in models 1-3 (Fig. 1c,f,i).

10 Because the mean number of chambers added per individual varied between culture experiments, we therefore normalise our culture CI data for the mean number of chambers precipitated in culture (detailed in Sect. 3.2).

In addition to discerning ontogenetic trends in CI, these models also help to constrain which morphological parameters exert the most coherent control on CI. For almost any given model parameter, the full range of CI observed in culture can be

15 produced given the right combination of other morphological parameters. For example, CIs of between 120-250 μg mm^{-2} can be produced irrespective of the chamber aspect ratio. In other words, whilst a change in chamber aspect ratio can affect CI, any such change can also be offset by a compensatory change in other model parameters. The only exceptions to this in our model are the variation in the coefficients a and b in equation A2 – the parameters that determine the non-linear growth function of wall thickness. Specifically, very low CIs in adult-sized foraminifera were only observed with a and b at their lowermost and

20 uppermost limit respectively (Fig. A2). In contrast to all other parameters, CI change as forced by the shape of the regression between wall thickness and chamber size/number cannot be compensated for by other prescribed ontogenetic parameters.

3.2 Culture results

Consistent with results of modelling in Sect. 3.1, CI in our culture experiments varied with test size and chamber addi-

25 tion. A logarithmic regression of individuals' CI versus their size at the beginning of culture experiments (Fig. 2a) yields an R^2 of 0.37. Once normalised according to the mean number of chambers added during each culture experiment (see Sect. 3.1 and Fig. 1c above), the coherence of this relationship with size becomes stronger, with an R^2 of 0.59 (Fig. 2b). REWORDED, see Reviewer 1: We chose logarithmic regression fits because in most of our models for *G. ruber* CI change through ontogeny approximates to a logarithmic relationship within the range in test size of our cultured *G. ruber* (see Fig.

30 1b).

To investigate the effect of culture conditions on CI, the residuals from the logarithmic regression fit in Fig. 2b (given in Table 3) are compared to environmental parameters (Fig. 3). No significant variation with $\mathrm{Mg/Ca_{sw}}$ was observed (Fig. 3), consistent with previous observations that varying seawater [Mg] does change growth rates in foraminifera (Evans et al., 2015, 2016a). Similarly, no effect of temperature was observed (Fig. 3b). A statistically significant correlation with pH was found (R^2= 0.47,

8

$p = 0.01$; Fig. 3c). Other studies often parameterise carbonate system changes in terms of carbonate ion concentration, $[CO_3^{2-}]$ (e.g. Marshall et al., 2013), or $[DIC]/[H^+]$ (Jokiel, 2011; Bach, 2015). We also tested residual CI against these parameters, using carbonate speciation constants adjusted for changing Mg/Ca_{sw} according to Hain et al. (2015). There was little difference in correlation coefficient in the case of $[DIC]/[H^+]$ ($R^2 = 0.47$ for $[DIC]/[H^+]$ and pH), but in the case of $[CO_3^{2-}]$ correlation was

5   weaker ($R^2 = 0.40$). We opt to primarily focus on pH in figures and discussions here as a less abstracted parameter, but we cannot rule out $[DIC]/[H^+]$ as being the primary driver, as suggested by Bach (2015).

3.3 Core-top results

Core-top results are given in Supplementary Table 2. Multiple linear regression analysis demonstrates that deep ocean carbonate system conditions at the core-top site, test size and morphospecies identity were all statistically significant controls on ρ_A within

10   the set of core-top samples measured here. Together, these three factors could explain 86% of the variance in ρ_A seen in our core-top assemblages (for regression statistics, see Supplementary Table 3). We used the relaimpo R package (Groemping, 2006) to determine relative importance of these factors, and found bottom water pH at the site of deposition to be the strongest determinant of ρ_A (as shown in Fig. 4 a and b, and Supplementary Table 3). Shell size (as parameterised by shell silhouette area; Fig. 4c), and species type within the broader *G. ruber* plexus (Fig. 4b,c) were found to be secondary, but nevertheless

15   significant factors. Where both *G. ruber* sensu stricto and sensu lato (see Reply to reviewer 1 encompassing both *G. elongatus* and *G. pyramidalis*; Aurahs et al. (2011)) were measured at the same site, *G. ruber* sensu stricto displayed significantly lower values of ρ_A (paired t-test; $p = 0.01$). Despite the observations from our culture experiments, estimated surface-ocean pH was not found to be a significant control on ρ_A in the core-top samples, and would be excluded from the model by stepwise parameter reduction according to its Akaike Information Criterion (AIC). We include it here for illustrative purposes only

20   given the focus of our study (Fig. 4), since it has little effect on the overall model fit ($R^2 = 0.86$ in both cases), or the relative importance of other factors in the regression. We note also that we tested other deep and surface water carbonate system parameters in lieu of pH (Ω, $\Delta[CO_3^{2-}]$, etc.), but in all cases pH produced in stronger model fits.

4   Discussion

4.1   Towards understanding the differential response of foraminifera to acidification

25   Previous studies investigating the effect of acidification on marine calcifiers have so far failed to explain the highly divergent responses both within and between the major groups of calcifiers. For example, coccolithophores were more heavily calcified during geological epochs characterised by higher CO_2, and lower pH (e.g. Bolton et al., 2016), even if this result has not always been reproduced in culture (Riebesell et al., 2000; Langer et al., 2009). Additionally, despite observations of decreasing pH and $[CO_3^{2-}]$ negatively impacting calcification in most planktonic and reef-dwelling foraminifera (e.g. de Moel et al., 2009;

30   Marshall et al., 2013; Kuroyanagi et al., 2009) in agreement with results from our cultures, benthic foraminifera became more heavily calcified or exhibited little response over the Palaeocene-Eocene Thermal Maximum (PETM) and ETM2 (Foster et al.,

2013). Examining this finding in the context of our model allows us to investigate the morphological responses to pH change that could produce these patterns, and so begin to form a unifying hypothesis to explain these various apparently contradictory observations. To do this, we must first consider how foraminiferal morphology itself might affect calcification.

5   Our model predicts that calcification intensity – a metric for how heavily calcified cultured foraminifera are – is dependent on foraminifera size on collection as well as the number of chambers precipitated in culture, as we observe (Figs. 1, 2). But beyond this, varying permutations of model parameters (as laid out in Appendix A) reveals that CI is most strongly dependent on covariation of the coefficients that describe the increase in wall thickness with ontogeny (a and b in Eq. A2). It appears that the overall shape of this regression, rather than either of its constituent coefficients alone, is most important in controlling CI.

10   Furthermore, the size-wall thickness coefficients drive CI in models 1 and 3, but varying maximum wall thickness in isolation has no systematic effect (model 2). These observations lead us to hypothesise that it is not simply that large adult foraminifera lay down less calcite in their walls in response to acidification, but rather that the slope of the regression between wall thickness and chamber diameter is shallower. This mechanism, if correct, would produce two physiological responses: (1) lower pH will result in a thinning of shell walls when the foraminifer is larger, and (2) smaller foraminifera will exhibit ambiguous, or even

15   positive, responses to acidification. These hypothesised responses are represented in Fig. 5.

    Although it is perhaps counter-intuitive to envisage such contrasting calcification responses to acidification with size, Note the following two paragraphs have been restructured, following comments from Reviewer 1 and Reviewer 2, to better separate extension to benthics: such a hypothesis may have a mechanistic foundation in the physiology of biomineralisation, when one

20   considers two important observations of calcite precipitation in foraminifera. Firstly, based on test oxygen isotope ratios, it seems that foraminifera precipitate calcite from a species-specific combination of HCO_3^- and CO_3^{2-} (Zeebe, 1999; Uchikawa and Zeebe, 2010). Secondly, small foraminifera may not have large internal calcium and/or carbon pools (Nehrke et al., 2013), whereas large foraminifera do (Erez, 2003; ter Kuile et al., 1987, 1989). This is because smaller foraminifera build chambers that require far less material volumetrically, and so they can potentially source the ions required for calcification on the same

25   timescales as chamber precipitation. In this way, lower pH could have less of an effect on their wall thickness, or even favour more heavily calcified chambers, if associated with a rise in [DIC]. Increased availability of carbon for calcification could promote calcification, decreasing the volume of seawater needing to be cycled to produce a given amount of calcite for chamber formation. Once planktonic and benthic foraminifera reach a certain size, however, the large amounts of $CaCO_3$ required to build a new chamber necessitates prior storage of carbon internally (Erez, 2003; ter Kuile et al., 1987, 1989). Importantly, the

30   efficiency of this internal storage mechanism is thought to be related to the ability of the organisms to raise the pH of vacuolised seawater (Bentov et al., 2009; de Nooijer et al., 2009). Therefore, lower seawater pH acts against the efficiency of this carbon concentrating mechanism, meaning that chamber formation is more difficult.

    Besides our observed response of planktonic foraminiferal wall thickness to changes in seawater carbonate chemistry

35   (Fig. 2), there is other empirical evidence to support our size-dependent model (see Fig. 5). Field studies indicate that large

shallow-water species of planktonic foraminifera respond to decreased pH by producing thinner walls (de Moel et al., 2009). Among benthic foraminifera, culture experiments in *Elphidium* also revealed reduced calcification at low pH (Allison et al., 2010). While there is as yet little data from smaller species of planktonic foraminifera and early ontogenetic stage individuals of larger planktonic foraminiferal species, there are some lines of evidence from small benthic foraminifera that could support

5 the inverse calcification response that we propose. Foster et al. (2013) found that small benthic foraminifera became more heavily calcified during the Palaeocene-Eocene Thermal Maximum (PETM), when ocean pH was lower (Penman et al., 2014). By contrast, during the high pH 'carbonate overshoot' in the aftermath of the event (Penman et al., 2016), these foraminifera displayed thinner walls (Foster et al., 2013). Indeed, while biomineralisation pathways clearly differ, recent work favours a similarly positive response of calcification to increased aqueous pCO_2 with lower pH in the much smaller coccolithophores

10 (Bolton et al., 2016; McClelland et al., 2016).

Size-dependent calcification could yet provide a common mechanism to unify the often variable responses observed in foraminifera to date. However, we note that there may be different ecophysiological factors driving calcification in benthic and planktonic foraminifera, and that it remains untested as to what extent calcification mechanisms in juvenile individuals of large

15 foraminiferal species may be comparable to adult individuals of smaller species. We suggest therefore that this hypothesis requires further investigation. At present, there are insufficient measurements of changes in calcification and morphology through ontogeny to robustly test this model, or indeed to re-interpret existing SNW data with confidence. Comparative CT scans of foraminifera, including examination of individuals grown under different pH conditions, could provide such a test by constraining ontogenetic variation in calcite (as in Schmidt et al., 2013), while more, multi-species culture experiments would

20 no doubt be beneficial.

4.2   Implications for size-normalised weight (SNW) in foraminifera as a proxy

Published investigations using SNW as an environmental proxy commonly assume that the SNW metrics themselves are independent of size or ontogeny. Our modelling approach shows that there is in fact a strong effect of test size on SNW. Modelled area density (ρ_A, a commonly used SNW metric Marshall et al., 2013, 2015; Osborne et al., 2016; Weinkauf et al.,

25 2013, 2016) is shown in Figure 6a as a function of test diameter. Virtually all model runs predict a positive relationship between ρ_A and test size, although the exact nature of this relationship may vary with model parameterisation. This suggests that at least some of the variability between and within published studies could derive from the widely divergent shell sizes and sieve size ranges used (Fig. 6b). Our models similarly reveal a strong size-dependency in volume-normalised approaches, such as that used by Foster et al. (2013) across the PETM. All model runs predict that the foraminiferal calcite volume/total volume ratio

30 exhibits a strong dependence on test size (Supplementary Fig. 2).

It has been suggested, given some size-mass (or volume-mass) relationships in planktonic foraminifera appear approximately linear (Weinkauf et al., 2016), that the use of area and volume-normalised weight to estimate SNW is not complicated by spatial or temporal variations in mean body size. In fact, a linear relationship between mass and area or between mass and volume

11

directly implies that area density *is* size-dependent, given:

$$\text{mass} = k \times \text{mm}^3 \tag{3}$$

Where k is a constant equal to the slope. If volume is approximated from area by raising to the power of $3/2$ (Weinkauf et al., 2016), then:

$$\rho_A = \frac{\text{mass}}{(\text{axis}_1\text{axis}_2)} \quad \approx \quad \frac{k \times (\text{axis}_1\text{axis}_2)^{3/2}}{(\text{axis}_1\text{axis}_2)} \quad \approx \quad k \times (\text{axis}_1\text{axis}_2)^{1/2} \tag{4}$$

which predicts a linear dependence of ρ_A on body size (given aspect ratio in Table 2), such that:

$$\rho_A \approx k \times \sqrt{\frac{\text{axis}_1^{\,2}}{1.16}} \quad \approx \quad \text{axis}_1 \times \frac{k}{\sqrt{1.16}} \tag{5}$$

For *G. ruber*, the slope of this relationship between size and ρ_A approximates to 2.6×10^{-3}. This means that a 100 μm increase in test major axis would lead to an ρ_A increase of $\sim$38 μg mm^{-2} due to the change in body size alone. This response is of the same magnitude as the area density response to carbonate system changes reported in other studies (e.g. Marshall et al., 2013; Osborne et al., 2016). Higher overall ρ_A observed by Marshall et al. (2013) for *T. sacculifer* in larger size fractions may constitute further evidence of a positive size effect on ρ_A. Similarly, we also find that cross-sectional area is a significant predictor of ρ_A in our core-top foraminifera (Fig. 4).

We thus provide theoretical, model and empirical evidence for a strong test size control on size-normalised weight metrics. This size-dependence applies equally to other SNW metrics, as well as the specific metrics directly discussed here. That said, size-dependence of SNW metrics may be variably manifest, and variably problematic, in real-world datasets. One implication of this finding is that datasets from different studies using different foraminiferal size fractions are not directly comparable. Given the range of size-fractions used in SNW studies (Fig. 6b), size and ontogenetic stage may explain some of the discrepancies between findings, particularly for those studies using a wide sieve size fraction (e.g. several hundred μm). The magnitude of increase in SNW (here quantified as ρ_A, Fig. 6a) with size is likely strongest at smaller body sizes ($< \sim$350 μm) - encompassing size fractions commonly used in both modern calibration and down-core studies (Fig. 6b). With these findings in mind, here we make recommendations of best practice for future studies. Firstly, in death assemblages, where post-gametogenic foraminifera will likely have added a similar number of chambers within a full life cycle, the influence of ontogeny and body size on SNW can be minimised by using the narrowest possible size fraction of only post-gametogenic individuals, and reporting mean test size. This may of course be challenging over transient climatic events associated with a shift in body size, such as the PETM – in that case, models of calcification with size and ontogeny are needed. Living assemblages may present other difficulties, as individuals of the same size may have more or fewer chambers, and hence differing amounts of secondary calcification. For pre-gametogenic individuals, then, some estimate of chamber number and overall test size is needed. These additional measurements are necessary because the SNW metrics are inherently dependent on foraminifer size, and because changes in the carbonate system may have a differential effect through ontogeny (as previously suggested by Aldridge et al.,

2012).

Besides the influence of size on SNW, our study highlights other fundamental caveats about the applicability of SNW as a surface-water proxy. Although our culture data supports a primary carbonate system control on calcification, in agreement with other studies (e.g. de Moel et al., 2009; Marshall et al., 2013; Osborne et al., 2016), our core-top samples demonstrate that this may be overwhelmed by dissolution processes at the site of deposition (see Fig. 4). Indeed, while size and morphospecies are preserved as significant controls on SNW (quantified in this case as ρ_A; Fig. 4), any primary signal of surface carbonate chemistry (for a range in pH of 8.09 - 8.21) in our core-top sample set has been entirely lost to dissolution. We therefore urge caution before attributing open-ocean SNW patterns exclusively to either primary or secondary processes. In sites where carbonate saturation at the site of deposition is high, it is probably reasonable to attribute SNW changes to surface conditions (as in Barker and Elderfield, 2002; Marshall et al., 2013; Weinkauf et al., 2013; Osborne et al., 2016, , etc.). At depths approaching the lysocline, our data supports the earlier uses of SNW as a proxy for the deep sea carbonate system (e.g. Lohmann, 1995; Broecker and Clark, 2001). In between these end-member scenarios, it may be difficult to untangle competing effects (Bijma et al., 2002), and so combining shallow-water and deep-water cores is advisable, following Barker et al. (2004).

see Reply to reviewer 1- section removed Incidentally, our finding that *G. ruber* sensu lato in core-tops has significantly higher SNW (ρ_A) than *G. ruber* sensu stricto (Fig. 4) may help to explain the apparently negative correlation between $[CO_3^{2-}]$ and SNW in *G. ruber* observed in tows by Beer et al. (2010b). Beer et al. did not differentiate between species, but it is likely they would have sampled an increasing proportion of higher-SNW sensu lato species (i.e. *G. elongatus* and *G. pyramidalis*) in lower-pH upwelling waters, given these species preference for colder waters (Steinke et al., 2005). This aside, it is also conceivable that these authors may have sampled individuals of similar size, but with a greater number of smaller chambers, as they approached the centre of upwelling. While we cannot confirm these suggestions, taken together, our observations may help to build a more consistent picture of controls on primary calcification intensity in planktonic foraminifera, including the dominance of sedimentary signals in most deep ocean settings.

4.3 Significance for Global Biogeochemical Cycling

Previous investigations into the controls on foraminiferal shell weight have often struggled to determine conclusively which environmental controls, if any, impact foraminiferal calcification, as temperature and $[CO_3^{2-}]$ are often correlated in hydrographic datasets (Aldridge et al., 2012; Marshall et al., 2013). In our cultures where both temperature and carbonate system parameters were varied, we show that the carbonate system is the most likely the driver of CI in *G. ruber* (Fig. 3). Given the significance of planktonic foraminiferal tests to the global pelagic $CaCO_3$ budget (Schiebel, 2002), this finding could therefore have important implications for global carbonate alkalinity fluxes, and projections of the response of biogeochemical cycling to anthropogenic ocean acidification. Within the pelagic realm, foraminiferal calcification reduces TAlk and DIC in a 2:1 ratio, releasing CO_2 and thereby lowering surface ocean pH (e.g. Zeebe and Wolf-Gladrow, 2001). Therefore, it is possible that reduced alkalinity uptake in surface waters may constitute a weak negative feedback on surface ocean acidification. Scaling this in terms of fluxes, considering changes in other calcifying groups like coccolithophorids, and accounting for body size and population size, is beyond the scope of this study. We also note that if the response to acidification is size-related, there may be some role for other environmental parameters that influence body size in foraminifera (e.g. temperature, light; Schmidt et al.,

2004; Lombard et al., 2010) in determining their response to anthropogenic ocean acidification. We suggest that modelling of the pelagic ecosystem, that includes considerations of the physiological costs to calcification in each group of marine calcifiers and the effects on cumulative fitness relative to other non-calcifying groups, may be critical in addressing current shortfalls in prediction of future biogeochemical changes (Mora et al., 2013).

5 Conclusions

In this study we approach the question of environmental controls on changing foraminiferal calcification intensity from multiple perspectives, incorporating observations from culturing and the open ocean with models of ontogenetic growth. Our models for shell growth suggest that calcification intensity (i.e. mass increase per unit size increase) changes as a function of ontogeny and body size in *G. ruber*. This finding provides a theoretical framework for interpreting results from culture experiments. After controlling for size and chamber addition, our culture experiments suggest neither temperature, nor seawater Mg/Ca ratios affect calcification intensity, but that acidification significantly reduces calcification in adult-stage foraminifera, supporting previous open-ocean observations (e.g. Barker and Elderfield, 2002; Marshall et al., 2013). Based on our modelling work, we hypothesise that carbonate chemistry may affect different sized foraminifera differently, with acidification leading to reduced calcification in bigger foraminifera, but conversely exerting little control (or even favouring calcification) in small individuals. While further work is required to test this differential response in larger and smaller individuals, these model results could help to explain a number of published observations, and serve to stress the importance of considering size and ontogeny when studying foraminiferal SNW. Additionally, our core-top results also highlight the central importance of post-mortem dissolution, followed by body size and species ID, in driving SNW in fossil assemblages. While the effects of lower ocean pH upon ecosystem-level biogeochemical fluxes are not yet fully constrained, our findings suggest that production of $CaCO_3$ by large planktonic foraminifera in the pelagic realm will likely be reduced by future anthropogenic ocean acidification.

References

Aldridge, D., Beer, C. J., and Purdie, D. A.: Calcification in the planktonic foraminifera *Globigerina bulloides* linked to phosphate concentrations in surface waters of the North Atlantic Ocean, Biogeosciences, 9, 1725–1739, doi:10.5194/bg-9-1725-2012, http://adsabs.harvard.edu/abs/2012BGeo....9.1725A, 2012.

5 Allison, N., Austin, W., Paterson, D., and Austin, H.: Culture studies of the benthic foraminifera *Elphidium williamsoni*: Evaluating pH, $\Delta[CO_3^{2-}]$ and inter-individual effects on test Mg/Ca, Chemical Geology, 274, 87–93, 2010.

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

10 Bach, L.: Reconsidering the role of carbonate ion concentration in calcification by marine organisms, Biogeosciences, 12, 4939–4951, 2015.

Barker, S. and Elderfield, H.: Foraminiferal Calcification Response to Glacial-Interglacial Changes in Atmospheric CO_2, Science, 297, 833–836, doi:10.1126/science.1072815, http://www.sciencemag.org/cgi/content/abstract/297/5582/833, 2002.

Barker, S., Kiefer, T., and Elderfield, H.: Temporal changes in North Atlantic circulation constrained by planktonic foraminiferal shell weights, Paleoceanography, 19, n/a–n/a, doi:10.1029/2004PA001004, http://doi.wiley.com/10.1029/2004PA001004, 2004.

15 Bé, A.: Gametogenic calcification in a spinose planktonic foraminifer, *Globigerinoides sacculifer* (Brady), Marine Micropaleontology, 5, 283–310, 1980.

Be, A. W. H. and Lott, L.: Shell Growth and Structure of Planktonic Foraminifera, Science, 145, 823–824, doi:10.1126/science.145.3634.823, http://www.sciencemag.org/cgi/doi/10.1126/science.145.3634.823, 1964.

Beaufort, L., Probert, I., de Garidel-Thoron, T., Bendif, E. M., Ruiz-Pino, D., Metzl, N., Goyet, C., Buchet, N., Coupel, P., Grelaud, M., et al.: Sensitivity of coccolithophores to carbonate chemistry and ocean acidification, Nature, 476, 80–83, 2011.

20 Beer, C. J., Schiebel, R., and Wilson, P. A.: Technical Note: On methodologies for determining the size-normalised weight of planktic foraminifera, Biogeosciences, 7, 2193–2198, doi:10.5194/bg-7-2193-2010, http://www.biogeosciences.net/7/2193/2010/, 2010a.

Beer, C. J., Schiebel, R., and Wilson, P. A.: Testing planktic foraminiferal shell weight as a surface water $[CO_3^{2-}]$ proxy using plankton net samples, Geology, 38, 103–106, doi:10.1130/G30150.1, http://geology.gsapubs.org/content/38/2/103.abstract, 2010b.

25 Bentov, S., Brownlee, C., and Erez, J.: The role of seawater endocytosis in the biomineralization process in calcareous foraminifera, Proceedings of the National Academy of Sciences, 106, 21 500–21 504, 2009.

Berger, W. H.: Planktonic foraminifera ; basic morphology and ecologic implications, Journal of Paleontology, 43, 1369–1383, 1969.

Berner, R. A. and Raiswell, R.: Burial of organic carbon and pyrite sulfur in sediments over Phanerozoic time: a new theory, Geochimica et Cosmochimica Acta, 47, 855–862, 1983.

30 Bijma, J., Hönisch, B., and Zeebe, R. E.: Impact of the ocean carbonate chemistry on living foraminiferal shell weight: Comment on "Carbonate ion concentration in glacial-age deep waters of the Caribbean Sea" by W. S. Broecker and E. Clark, Geochemistry Geophysics Geosystems, 3, 1064–1071, doi:2002 10.1029/2002GC000388 [Citation], http://www.agu.org/pubs/crossref/2002/2002GC000388.shtml, 2002.

Bolton, C. T., Hernández-Sánchez, M. T., Fuertes, M.-Á., González-Lemos, S., Abrevaya, L., Mendez-Vicente, A., Flores, J.-A., Probert, I.,
35 Giosan, L., Johnson, J., et al.: Decrease in coccolithophore calcification and CO_2 since the middle Miocene, Nature communications, 7, 2016.

Broecker, W. S. and Clark, E.: Glacial-to-Holocene Redistribution of Carbonate Ion in the Deep Sea, Science, 294, 2152 –2155, doi:10.1126/science.1064171, http://www.sciencemag.org/content/294/5549/2152.abstract, 2001.

Brummer, G.-J. A., Hemleben, C., and Spindler, M.: Planktonic foraminiferal ontogeny and new perspectives for micropalaeontology, Nature, 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), Marine Micropaleontology, 12, 357–381, 1987.

Caldeira, K. and Wickett, M. E.: Oceanography: Anthropogenic carbon and ocean pH, Nature, 425, 365–365, doi:10.1038/425365a, http://www.nature.com/nature/journal/v425/n6956/full/425365a.html, 2003.

Caron, D. A., Roger Anderson, O., Lindsey, J. L., Faber Jr., W. W., and Lin Lim, E.: Effects of gametogenesis on test structure and dissolution of some spinose planktonic foraminifera and implications for test preservation, Marine Micropaleontology, 16, 93–116, doi:10.1016/0377-8398(90)90031-G, http://www.sciencedirect.com/science/article/pii/037783989090031G, 1990.

de Moel, H., Ganssen, G., Peeters, F., Jung, S., Kroon, D., Brummer, G., and Zeebe, R.: Planktic foraminiferal shell thinning in the Arabian Sea due to anthropogenic ocean acidification, Biogeosciences, 6, 1917–1925, 2009.

de Nooijer, L., Spero, H., Erez, J., Bijma, J., and Reichart, G.: Biomineralization in perforate foraminifera, Earth-Science Reviews, 135, 48–58, 2014.

de Nooijer, L. J., Toyofuku, T., and Kitazato, H.: Foraminifera promote calcification by elevating their intracellular pH, Proceedings of the National Academy of Sciences, 106, 15 374–15 378, 2009.

Dickson, A. G.: Thermodynamics of the dissociation of boric acid in synthetic seawater from 273.15 to 318.15 K, Deep Sea Research Part A. Oceanographic Research Papers, 37, 755–766, doi:10.1016/0198-0149(90)90004-F, http://www.sciencedirect.com/science/article/pii/019801499090004F, 1990.

Erez, J.: The Source of Ions for Biomineralization in Foraminifera and Their Implications for Paleoceanographic Proxies, Reviews in Mineralogy and Geochemistry, 54, 115 –149, doi:10.2113/0540115, http://rimg.geoscienceworld.org/content/54/1/115.short, 2003.

Evans, D., Erez, J., Oron, S., and Müller, W.: Mg/Ca-temperature and seawater-test chemistry relationships in the shallow-dwelling large benthic foraminifer *Operculina ammonoides*, Geochimica et Cosmochimica Acta, 148, 325–342, 2015.

Evans, D., Brierley, C., Raymo, M. E., Erez, J., and Müller, W.: Planktic foraminifera shell chemistry response to seawater chemistry: Pliocene-Pleistocene seawater Mg/Ca, temperature and sea level change, Earth and Planetary Science Letters, 438, 139–148, doi:10.1016/j.epsl.2016.01.013, http://adsabs.harvard.edu/abs/2016E%26PSL.438..139E, 2016a.

Evans, D., Wade, B. S., Henehan, M., Erez, J., and Müller, W.: Revisiting carbonate chemistry controls on planktic foraminifera Mg/Ca: implications for sea surface temperature and hydrology shifts over the Paleocene–Eocene Thermal Maximum and Eocene–Oligocene Transition, Clim. Past, 12, 819–835, doi:10.5194/cp-12-819-2016, http://www.clim-past.net/12/819/2016/, 2016b.

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

Foster, L. C., Schmidt, D. N., Thomas, E., Arndt, S., and Ridgwell, A.: Surviving rapid climate change in the deep sea during the Paleogene hyperthermals, Proceedings of the National Academy of Sciences, 110, 9273–9276, 2013.

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Zweng, M. M., Baranova, O. K., and Johnson, D. R.: World Ocean Atlas 2009, Volume 4: Nutrients (phosphate, nitrate, silicate), in: NOAA Atlas NESDIS 71, edited by Levitus, S., p. 398, U.S. Government Printing Office, 2010.

17

Gloor, M., Gruber, N., Sarmiento, J., Sabine, C. L., Feely, R. A., and Rödenbeck, C.: A first estimate of present and preindustrial air-sea CO2 flux patterns based on ocean interior carbon measurements and models, Geophysical Research Letters, 30, 1010, doi:10.1029/2002GL015594, http://adsabs.harvard.edu/abs/2003GeoRL..30a..10G, 2003.

Goyet, C., Healy, R. J., and Ryan, J. P.: Global distribution of total inorganic carbon and total alkalinity below the deepest winter mixed layer
5 depths., Tech. Rep. ORNL/CDIAC-127, NDP-076, Carbon Dioxide Information Analysis Centre, Oak Ridge National Laboratory, U.S., 2000.

Groemping, U.: Relative Importance for Linear Regression in R: The Package relaimpo, Journal of Statistical Software, 17, 1–27, 2006.

Hain, M. P., Sigman, D. M., Higgins, J. A., and Haug, G. H.: The effects of secular calcium and magnesium concentration changes on the thermodynamics of seawater acid/base chemistry: Implications for Eocene and Cretaceous ocean carbon chemistry and buffering, Global
10 Biogeochemical Cycles, p. 2014GB004986, doi:10.1002/2014GB004986, http://onlinelibrary.wiley.com/doi/10.1002/2014GB004986/abstract, 2015.

Hemleben, C., Spindler, M., and Erson, O. R.: Modern planktonic foraminifera, Springer, Berlin, 1989.

Henehan, M. J., Rae, J. W., Foster, G. L., Erez, J., Prentice, K. C., Kucera, M., Bostock, H. C., Martínez-Botí, M. A., Milton, J. A., Wilson, P. A., Marshall, B. J., and Elliott, T.: Calibration of the boron isotope proxy in the planktonic foraminifera *Globigerinoides ruber* for use in
15 palaeo-CO$_2$ reconstruction, Earth and Planetary Science Letters, 364, 111–122, doi:10.1016/j.epsl.2012.12.029, http://www.sciencedirect.com/science/article/pii/S0012821X12007157, 2013.

Henehan, M. J., Foster, G. L., Bostock, H. C., Greenop, R., Marshall, B. J., and Wilson, P. A.: A new boron isotope-pH calibration for *Orbulina universa*, with implications for understanding and accounting for 'vital effects', Earth and Planetary Science Letters, 454, 282 – 292, doi:http://dx.doi.org/10.1016/j.epsl.2016.09.024, http://www.sciencedirect.com/science/article/pii/S0012821X16305040, 2016.

20 Jokiel, P. L.: Ocean acidification and control of reef coral calcification by boundary layer limitation of proton flux, Bulletin of Marine Sciences, 87, 639–657, 2011.

Kisakürek, B., Eisenhauer, A., Böhm, F., Garbe-Schönberg, D., and Erez, J.: Controls on shell Mg/Ca and Sr/Ca in cultured planktonic foraminiferan, *Globigerinoides ruber* (white), Earth and Planetary Science Letters, 273, 260–269, doi:10.1016/j.epsl.2008.06.026, http://www.sciencedirect.com/science/article/pii/S0012821X08004007, 2008.

25 Kuroyanagi, A., Kawahata, H., Suzuki, A., Fujita, K., and Irie, T.: Impacts of ocean acidification on large benthic foraminifers: Results from laboratory experiments, Marine Micropaleontology, 73, 190–195, 2009.

Langer, G., Nehrke, G., Probert, I., Ly, J., and Ziveri, P.: Strain-specific responses of *Emiliania huxleyi* to changing seawater carbonate chemistry, Biogeosciences, 6, 2637–2646, 2009.

Lee, K., Tong, L. T., Millero, F. J., Sabine, C. L., Dickson, A. G., Goyet, C., Park, G.-H., Wanninkhof, R., Feely, R. A., and Key, R. M.:
30 Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans, Geophysical Research Letters, 33, L19 605, doi:10.1029/2006GL027207, http://onlinelibrary.wiley.com/doi/10.1029/2006GL027207/abstract, 2006.

Lee, K., Kim, T., Byrne, R. H., Millero, F. J., Feely, R. A., and Liu, Y.: The universal ratio of boron to chlorinity for the North Pacific and North Atlantic oceans, Geochimica et Cosmochimica Acta, 74, 1801–1811, doi:10.1016/j.gca.2009.12.027, http://www.sciencedirect.com/science/article/pii/S0016703709007789, 2010.

35 Lohmann, G. P.: A model for variation in the chemistry of planktonic foraminifera due to secondary calcification and selective dissolution, Paleoceanography, 10, 445–457, doi:199510.1029/95PA00059, http://www.agu.org/journals/pa/v010/i003/95PA00059/, 1995.

Lombard, F., Da Rocha, R. E., Bijma, J., and Gattuso, J.: Effect of carbonate ion concentration and irradiance on calcification in planktonic foraminifera, Biogeosciences, 7, 247–255, http://adsabs.harvard.edu/abs/2010BGeo....7..247L, 2010.

Lueker, T. J., Dickson, A. G., and Keeling, C. D.: Ocean pCO$_2$ calculated from dissolved inorganic carbon, alkalinity, and equations for K$_1$ and K$_2$: validation based on laboratory measurements of CO$_2$ in gas and seawater at equilibrium, Marine Chemistry, 70, 105–119, doi:10.1016/S0304-4203(00)00022-0, http://www.sciencedirect.com/science/article/pii/S0304420300000220, 2000.

Marshall, B. J., Thunell, R. C., Henehan, M. J., Astor, Y., and Wejnert, K. E.: Planktonic foraminiferal area density as a proxy for carbonate ion concentration: A calibration study using the Cariaco Basin Ocean Time Series, Paleoceanography, 28, 1–14, doi:10.1002/palo.20034, http://onlinelibrary.wiley.com/doi/10.1002/palo.20034/abstract, 2013.

Marshall, B. J., Thunell, R. C., Spero, H. J., Henehan, M. J., Lorenzoni, L., and Astor, Y.: Morphometric and stable isotopic differentiation of *Orbulina universa* morphotypes from the Cariaco Basin, Venezuela, Marine Micropaleontology, doi:10.1016/j.marmicro.2015.08.001, http://linkinghub.elsevier.com/retrieve/pii/S0377839815300013, 2015.

McClelland, H. L. O., Barbarin, N., Beaufort, L., Hermoso, M., Ferretti, P., Greaves, M. J., and Rickaby, R. E. M.: Calcification response of a key phytoplankton family to milennial-scale environmental change., Scientific Reports, 6, 34 263, 2016.

Millero, F. J.: Thermodynamics of the carbon dioxide system in the oceans, Geochimica et Cosmochimica Acta, 59, 661–677, doi:10.1016/0016-7037(94)00354-O, http://www.sciencedirect.com/science/article/pii/001670379400354O, 1995.

Mora, C., Wei, C.-L., Rollo, A., Amaro, T., Baco, A. R., Billett, D., Bopp, L., Chen, Q., Collier, M., Danovaro, R., Gooday, A. J., Grupe, B. M., Halloran, P. R., Ingels, J., Jones, D. O. B., Levin, L. A., Nakano, H., Norling, K., Ramirez-Llodra, E., Rex, M., Ruhl, H. A., Smith, C. R., Sweetman, A. K., Thurber, A. R., Tjiputra, J. F., Usseglio, P., Watling, L., Wu, T., and Yasuhara, M.: Biotic and Human Vulnerability to Projected Changes in Ocean Biogeochemistry over the 21st Century, PLOS Biol, 11, e1001 682, doi:10.1371/journal.pbio.1001682, http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001682, 2013.

Moy, A. D., Howard, W. R., Bray, S. G., and Trull, T. W.: Reduced calcification in modern Southern Ocean planktonic foraminifera, Nature Geosci, 2, 276–280, doi:10.1038/ngeo460, http://dx.doi.org/10.1038/ngeo460, 2009.

Nehrke, G., Keul, N., Langer, G., De Nooijer, L., Bijma, J., and Meibom, A.: A new model for biomineralization and trace-element signatures of Foraminifera tests, Biogeosciences, 10, 6759–6767, 2013.

Osborne, E. B., Thunell, R. C., Marshall, B. J., Holm, J. A., Tappa, E. J., Benitez-Nelson, C., Cai, W.-J., and Chen, B.: Calcification of the planktonic foraminifera *Globigerina bulloides* and carbonate ion concentration: Results from the Santa Barbara Basin, Paleoceanography, p. 2016PA002933, 2016.

Parker, F. L.: Planktonic foraminiferal species in Pacific sediments, Micropaleontology, pp. 219–254, 1962.

Penman, D. E., Hönisch, B., Zeebe, R. E., Thomas, E., and Zachos, J. C.: Rapid and sustained surface ocean acidification during the Paleocene-Eocene Thermal Maximum, Paleoceanography, p. 2014PA002621, doi:10.1002/2014PA002621, http://onlinelibrary.wiley.com/doi/10.1002/2014PA002621/abstract, 2014.

Penman, D. E., Turner, S. K., Sexton, P. F., Norris, R. D., Dickson, A. J., Boulila, S., Ridgwell, A., Zeebe, R. E., Zachos, J. C., Cameron, A., Westerhold, T., and Röhl, U.: An abyssal carbonate compensation depth overshoot in the aftermath of the Palaeocene–Eocene Thermal Maximum, Nature Geoscience, 9, 575–580, doi:10.1038/ngeo2757, http://www.nature.com/doifinder/10.1038/ngeo2757, 2016.

Ridgwell, A. J. and Zeebe, R. E.: The role of the global carbonate cycle in the regulation and evolution of the Earth system, Earth and Planetary Science Letters, 234, 299–315, doi:10.1016/j.epsl.2005.03.006, http://www.sciencedirect.com/science/article/B6V61-4G1WY99-2/2/b3e95243cfd0adbde2b9951a3437aea0, 2005.

Riebesell, U., Zondervan, I., Rost, B., Tortell, P. D., Zeebe, R. E., and Morel, F. M.: Reduced calcification of marine plankton in response to increased atmospheric CO$_2$, Nature, 407, 364–367, 2000.

19

Russell, A. D., Hönisch, B., Spero, H. J., and Lea, D. W.: Effects of seawater carbonate ion concentration and temperature on shell U, Mg, and Sr in cultured planktonic foraminifera, Geochimica et Cosmochimica Acta, 68, 4347–4361, doi:10.1016/j.gca.2004.03.013, http://www.sciencedirect.com/science/article/B6V66-4DKJCJ3-6/2/67be2e25cd29a18c3806626b687539b9, 2004.

Schiebel, R.: Planktic foraminiferal sedimentation and the marine calcite budget, Global Biogeochemical Cycles, 16, 1065, doi:10.1029/2001GB001459, http://www.agu.org/pubs/crossref/2002.../2001GB001459.shtml, 2002.

Schmidt, D. N., Renaud, S., Bollman, J., Schiebel, R., and Thierstein, H. R.: Size distribution of Holocene planktic foraminifer assemblages: biogeography, ecology and adaptation, Marine Micropaleontology, 50, 319–338, 2004.

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, doi:10.1111/pala.12013, http://onlinelibrary.wiley.com/doi/10.1111/pala.12013/abstract, 2013.

Shaked, Y. and Genin, A.: The Israel National Monitoring Program at the Gulf of Eilat (NMP)., Scientific report, Interuniversity Institute, Eilat, 2006.

Sigman, D. M., Hain, M. P., and Haug, G. H.: The polar ocean and glacial cycles in atmospheric CO_2 concentration, Nature, 466, 47–55, 2010.

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

Spero, H. J., Eggins, S. M., Russell, A. D., Vetter, L., Kilburn, M. R., and Hönisch, B.: Timing and mechanism for intratest Mg/Ca variability in a living planktic foraminifer, Earth and Planetary Science Letters, 409, 32 – 42, doi:http://dx.doi.org/10.1016/j.epsl.2014.10.030, http://www.sciencedirect.com/science/article/pii/S0012821X14006578, 2015.

Steinke, S., Chiu, H., Yu, P., Shen, C., Löwemark, L., Mii, H., and Chen, M.: Mg/Ca ratios of two *Globigerinoides ruber* (white) morphotypes: Implications for reconstructing past tropical/subtropical surface water conditions, Geochemistry Geophysics Geosystems, 6, Q11 005, doi:10.1029/2005GC000926, http://www.agu.org/pubs/crossref/2005/2005GC000926.shtml, 2005.

Sun, X., Corliss, B. H., Brown, C. W., and Showers, W. J.: The effect of primary productivity and seasonality on the distribution of deep-sea benthic foraminifera in the North Atlantic, Deep Sea Research Part I: Oceanographic Research Papers, 53, 28–47, 2006.

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C., Delille, B., Bates, N., and de Baar, H. J.: Climatological mean and decadal change in surface ocean pCO_2, and net sea–air CO_2 flux over the global oceans, Deep Sea Research Part II: Topical Studies in Oceanography, 56, 554–577, doi:10.1016/j.dsr2.2008.12.009, http://www.sciencedirect.com/science/article/pii/S0967064508004311, 2009.

ter Kuile, B., Erez, J., and Lee, J.: The role of feeding in the metabolism of larger symbiont bearing foraminifera, Symbiosis, 4, 335–350, 1987.

ter Kuile, B., Erez, J., and Padan, E.: Mechanisms for the uptake of inorganic carbon by two species of symbiont-bearing foraminifera, Marine Biology, 103, 241–251, 1989.

Uchikawa, J. and Zeebe, R. E.: Examining possible effects of seawater pH decline on foraminiferal stable isotopes during the Paleocene-Eocene Thermal Maximum, Paleoceanography, 25, 2010.

van Heuven, S., Pierrot, D., Rae, J. W. B., Lewis, E., and Wallace, D. W. R.: MATLAB Program Developed for CO_2 System Calculations., http://cdiac.ornl.gov/ftp/co2sys/CO2SYS_calc_MATLAB/, 2011.

Wang, L.: 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, 381–394, doi:10.1016/S0031-0182(00)00094-8, http://www.sciencedirect.com/science/article/pii/S0031018200000948, 2000.

Weinkauf, M. F. G., Moller, T., Koch, M. C., and Kučera, M.: Calcification intensity in planktonic foraminifera reflects ambient conditions irrespective of environmental stress, Biogeosciences, 10, 6639–6655, doi:10.5194/bgd-10-6639-2013, http://www.biogeosciences-discuss.net/10/11213/2013/, 2013.

Weinkauf, M. F. G., Kunze, J. G., Waniek, J. J., and Kučera, M.: Seasonal Variation in Shell Calcification of Planktonic Foraminifera in the NE Atlantic Reveals Species-Specific Response to Temperature, Productivity, and Optimum Growth Conditions, PLOS ONE, 11, e0148 363, doi:10.1371/journal.pone.0148363, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0148363, 2016.

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

Zeebe, R. E. and Wolf-Gladrow, D. A.: CO_2 in seawater: Equilibrium, Kinetics, Isotopes., no. 65 in Elsevier Oceanography Series, Elsevier, Amsterdam, 2001.
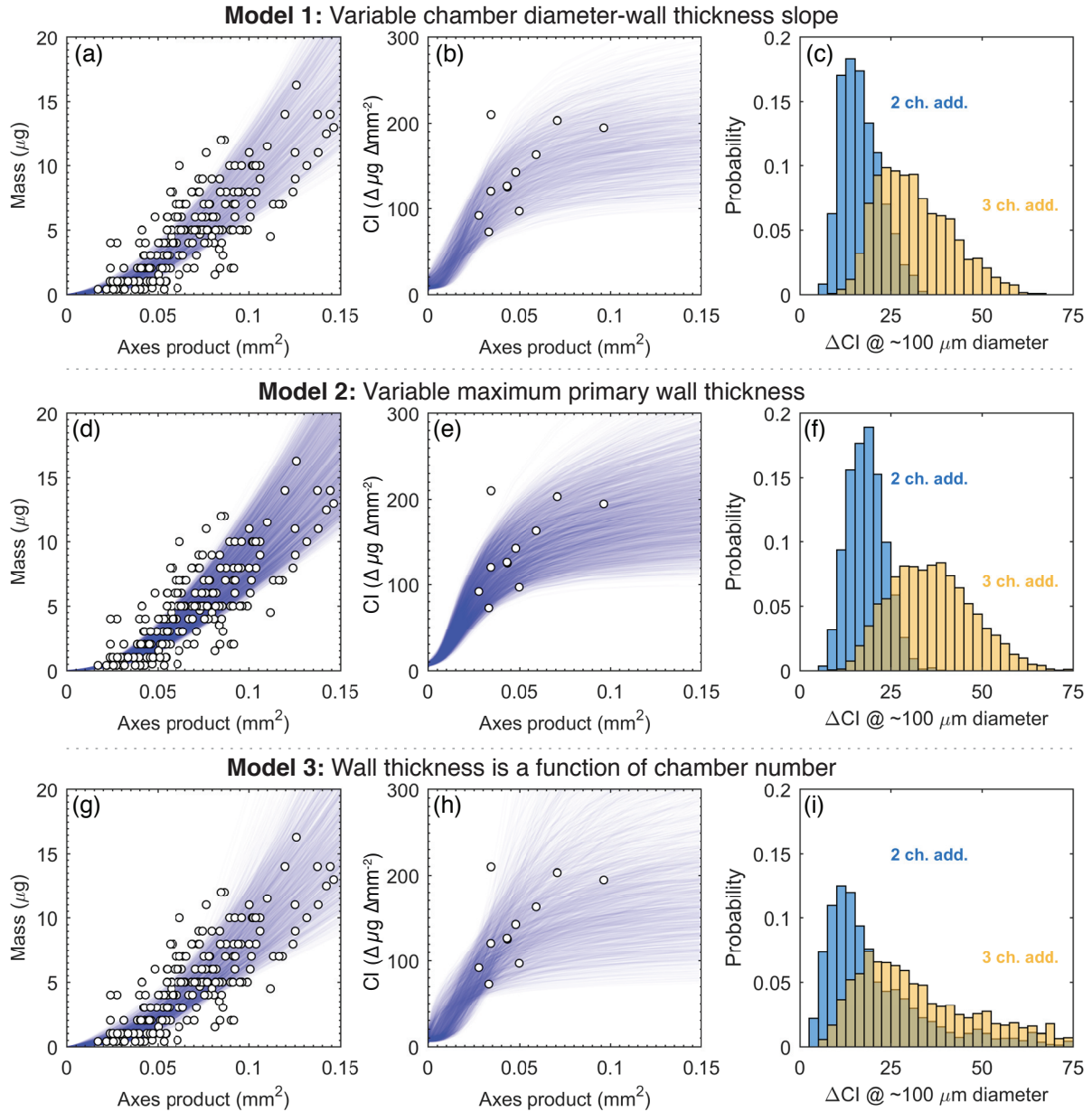
Figure 1. Model output for the three model scenarios described in Sect. 2.4.2. (a) Measured size-mass relationship for *G. ruber* based on towed specimens from the Gulf of Eilat that were not cultured. The residual sum of squares between these data defines which model runs (individual blue lines) are taken as representative of this natural population. (b) CI data of cultured specimens (equation 2), shown in the context of that predicted from the same set of models for the scenario where two chambers were precipitated in culture. All models predict that CI is dependent on body size on collection, which is also our empirical observation. (c) Model CI dependence on the amount of chambers added in culture. Broadly, the more chambers precipitated, the higher the resultant measured CI. Because of this finding, we use the model relationship between CI and the number of chambers added in culture to normalise the CI data shown in panels b.
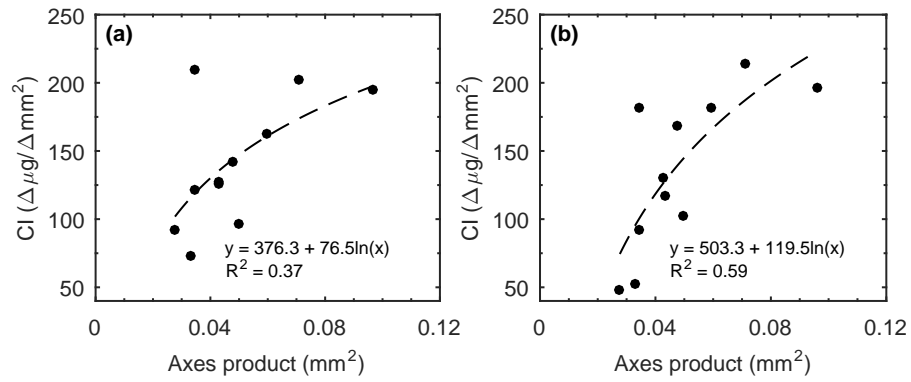
Figure 2. (a) Measured CI as a function of the product of major and minor axes at the start of culture for all experiments. (b) CI data can normalised for the number of chambers added in culture, according to the median CI increase of 17 $\Delta\mu g/\Delta mm^2$ per chamber added from Fig. 1c,f,i. In this case, data were normalised to two chambers added during culture, which was mode for our culture experiments.
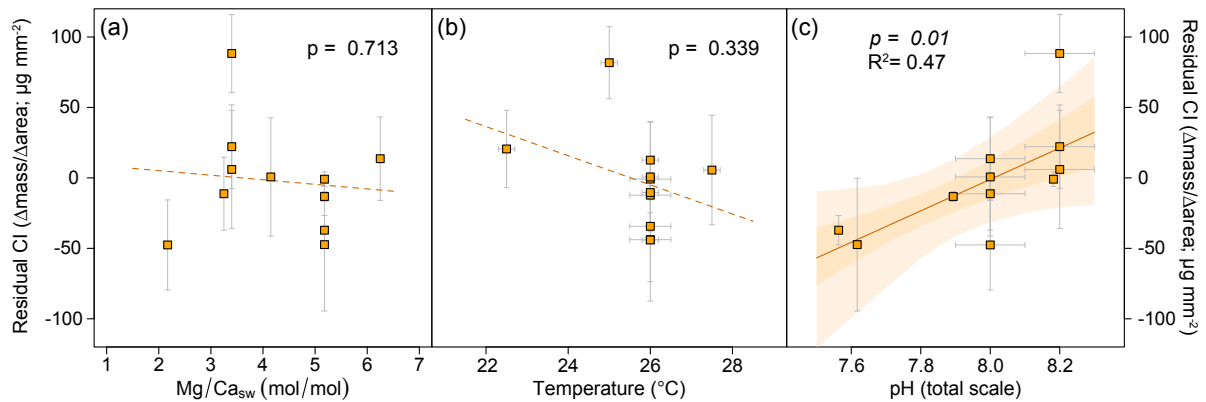
Figure 3. Ambient environmental response of chamber-addition corrected residual CI (change in mass/change in area compared to the logarithmic regression through all experiments shown in Fig. 2, see text for justification). (a) Mg/Ca_{sw}, (b) temperature, (c) pH. Shaded regions are 1σ and 2σ bounds of uncertainty, as calculated via combined wild bootstrap and Monte Carlo analysis, accounting for error in X and Y variables, following Henehan et al. (2016). Dashed lines indicate non-significant relationships, solid regression lines are significant to $\alpha < 0.05$. Shaded regions of uncertainty are 1 and 2 sd of 1000 Monte Carlo linear regression models through randomly simulated datasets sampled within the given 2σ X- and Y-error margins for each sample.
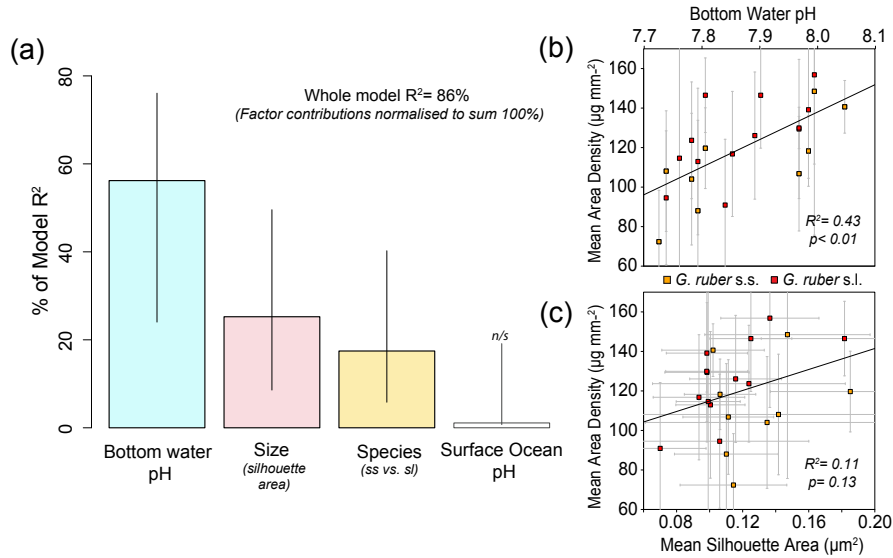
Figure 4. Drivers of ρ_A in core-top *G. ruber* (white). In panel (a), the relative importance of environmental (surface and deep ocean pH) and physiological (size and species) factors in predicing ρ_A are shown. A multiple linear regression model containing bottom water pH (i.e. pH at the site of deposition), test size (i.e. mean silhouette area in the core-top sample) and morphospecies (i.e. *G. ruber* (white) sensu stricto vs. sensu lato (Wang, 2000)) can describe 86% of the variance in ρ_A, with all predictors significant to $p < 0.01$ except surface ocean pH, which is not a significant contributor to the model but is shown here for illustrative purposes only (see also Supplementary Table 3). Of these variables, bottom water pH is the strongest correlate, as determined by the R package *relaimpo* (Groemping, 2006). Uncertainty bounds on the relative contribution of each variable are determined via bootstrapping. Univariate regression lines of ρ_A vs. bottom-water pH and vs. mean test size are shown in panels (b) and (c) respectively. is shown. Error bars on ρ_A are 2 standard deviations of variation within core-tops (panels b and c). Error bars on mean silhouette area (panel c) are 2 standard deviations of silhouette area within a core-top seive size fraction. In each case, *G. ruber* sensu stricto are plotted in orange, and sensu lato in red. Generally heavier values of ρ_A are observed in *G. ruber* sensu lato (panels b, c), constituting the third most important control on ρ_A within our core-top dataset (panel a). Mutliple linear regression statistics are given in Supplementary Table 3. Residual variation in ρ_A around the relationship with bottom water pH is significantly correlated with test silhouette area ($R^2 = 0.35$, $p < 0.01$; see Supplementary Fig. 5). Note, other deep ocean carbonate system parameters (Ω, $\Delta[CO_3^{2-}]$) were also trialled in multiple regression models, but the strongest correlation was observed with pH at each core-top site, and so that variable is preferred here.
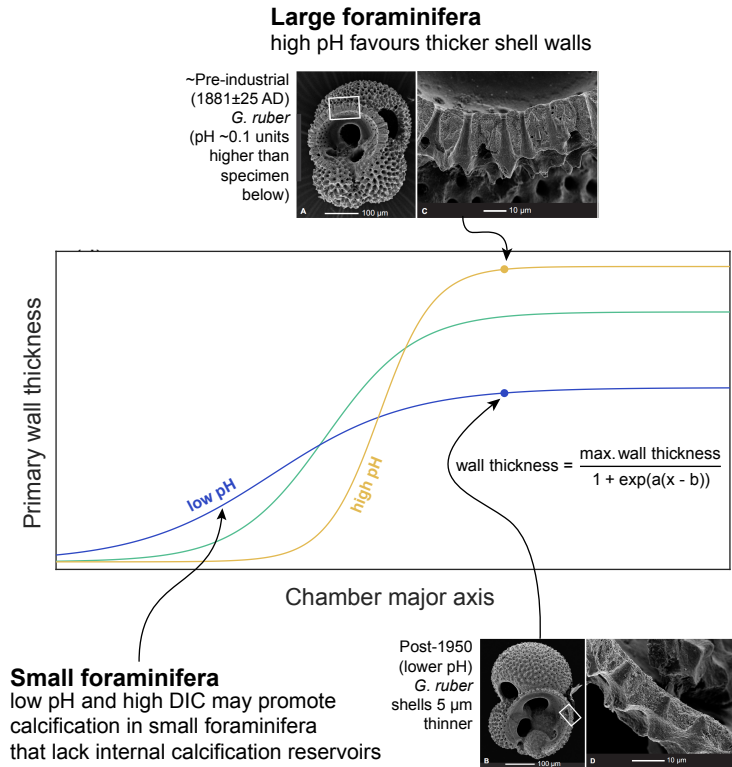
Figure 5. Schematic response of the chamber wall thickness-chamber size slope with carbonate chemistry, based on our observations that (1) CI is controlled by pH in culture, and (2) modelled CI responds principally to the slope of this relationship, where shallower slopes are characterised by lower CI. These findings potentially reconcile the differential response of small and large foraminifera to acidification. For example, *G. ruber* in culture (this study) and the Arabian Sea (de Moel et al., 2009) respond negatively to reduced pH (blue line). See Response to Reveiwers 1 and 2: Benthics removed from figure and instead noted in caption Conversely, if benthic foraminiferal calcification can be considered analogous to planktonic, the positive calcification response small benthic species exhibit to acidification over the PETM and ETM2 (Foster et al., 2013), could support our hypothesised size-dependent calcification response. Images of *G. ruber* from de Moel et al. (2009) reproduced under CC3.0 licence.
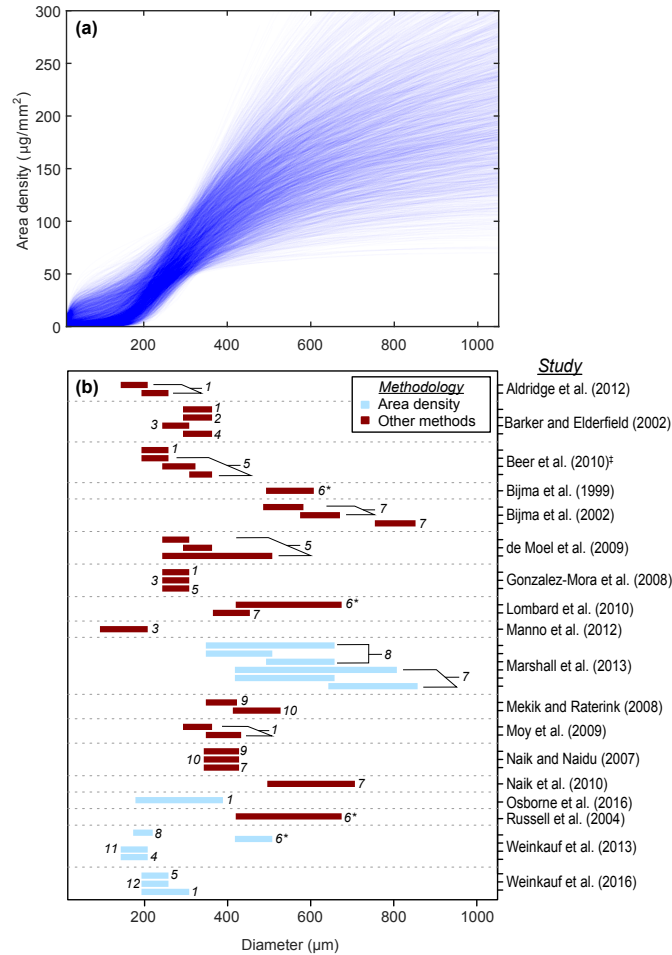
Figure 6. Modelled dependency of area density (ρ_A) on foraminiferal diameter (panel a), for comparison with the wide range of size fractions used in studies of size-normalised weight (panel b). Note we model area density in panel a, and so those studies that also use this exact metric are shaded separately in blue in panel (b). Species numbers are (1) *Globigerina bulloides*, (2) *Truncorotalia truncatulinoides*, (3) *Neogloboquadrina pachyderma*, (4) *Globoconella inflata*, (5) *Globigerinoides ruber* (white), (6) *Orbulina universa*, (7) *Trilobatus sacculifer*, (8) *Globigerinoides ruber* (pink), (9) *Neogloboquadrina dutertrei*, (10) *Pulleniatina obliquiloculata*, (11) *Globorotalia scitula*, (12) *Globigerinoides elongatus*. *Since *O. universa* is spherical, we stress that test size exerts a negligible control on ρ_A, in contrast to other species. [‡] We note that this study uses cross-sectional area in calculating a size-normalised weight, but their approach to normalisation to a set size is different to the ρ_A method outlined in Marshall et al. (2013).

Table 1. Description of important terms used in this paper and the relevant literature.

Term	Shorthand	Meaning	Further reading
Size-normalised weight	SNW	A general term for the mass of foraminiferal tests divided by some metric of test size. The term describes how 'heavily calcified' foraminiferal shells are, but importantly, it does not discern between changes in the degree of calcification during life vs. post-mortem thinning and dissolution. Many methods exist in the literature for normalising test mass to size, each with merits and pitfalls.	e.g. Aldridge et al. (2012); Beer et al. (2010a)
Calcification Intensity	CI	A more specific term under the umbrella of SNW that refers to how thickly the foraminifera calcified in life. For our culture experiments, this is defined more specifically in Eq. 2.	this study, Sect. 2.3
Area Density	ρ_A	A specific metric of size-normalised weight that normalises test mass to cross-sectional area. Since it has been increasingly used in recent studies, we discuss some implications for this particular metric.	e.g. Marshall et al. (2013); Weinkauf et al. (2016)

Table 2. Data-constrained *G. ruber* measurements used to construct the model. †Variable names are those used to construct the Matlab code, available in the supporting material.

Parameter	Variable name†	Value	Definition/notes	Source
Foraminifer aspect ratio	forAspRat	1.16	Test aspect ratio (major/minor axis)	This study
Foraminifer test growth rate	forAxIn	1.19	Fraction increase in test major axis per chamber addition	This study
Number of chambers	noCh	16	To better assess model behaviour 18 chambers were modelled, but only 16 were utilised in calculations and plots	Parker (1962)
Chamber aspect ratio	chAspRat	1.66	Chamber aspect ratio (major/minor axis)	This study
Chamber growth rate	axInOb	1.15	Fraction increase in chamber major axis per chamber addition	This study
Chamber overlap with previous	chCut	45%	Proportion of spheroid chamber	This study
Initial chamber minor axis	inchAx2	5 μm	Proloculus semi axis based on *T. sacculifer*	Schmidt et al. (2013)
Porosity	forPor	4.2%	Percentage chamber wall that is pore space	This study
Maximum primary wall thickness	maxInWall		Saturation point of wall thickness with ontogeny:	
		10.5-19.5 μm	Model 2	de Moel et al. (2009)
		18-22 μm	Models 1 and 3	This study
Secondary calcite layer thickness	secChAd	67%	Secondary calcite added per chamber addition, calculated as a proportion of the mass of all previous chambers	This study

Table 3. Culture experiment details, raw calcification intensity (CI), and residual CI calculated following correction for the observed and modelled relationship between CI and axes product. Note for experiments beginning 'DE', carbonate ion concentrations are calculated with carbonate system constants adjusted for Mg/Ca_{sw} using the MyAMI model (Hain et al., 2015). For MH2-Exp1, carbonate system calculations were adjusted for artificially enhanced seawater boron concentrations. DIC concentrations from experiments beginning 'DE' are from measured composite seawater solutions, and for experiments beginning 'MH' are derived from total alkalinity measured in composite seawater solutions.

Experiment	Temperature (°C)	Mg/Ca_{sw} (mol/mol)	pH (total)	±	DIC (μM)	$[CO_3^{2-}]$ (μM)	CI $\pm$ SE	Initial product of axes (mm^2)	Mean chamber addition	Residual CI
DE3-2-26	26.3	2.17	8.0	0.1	2194	228	194±32	0.096	1.4	-48
DE3-3-26	26.3	3.25	8.0	0.1	2067	216	126±26	0.043	2.5	-11
DE3-4-26	26.3	4.15	8.0	0.1	2057	209	127±42	0.043	2.0	1
DE3-6-26	26.3	6.25	8.0	0.1	2042	201	202±29	0.083	1.8	14
DE4-3-22.5	22.8	3.40	8.2	0.1	2024	281	142±30	0.048	1.2	22
DE4-3-25	25.3	3.40	8.2	0.1	1997	300	209±28	0.034	3.8	88
DE4-3-27.5	27.8	3.40	8.2	0.1	2072	336	163±42	0.059	1.3	6
MH1-HighpH	26.0	5.18	8.182	0.007	1959	297	120±5	0.034	3.5	-1
MH1-MedpH	26.0	5.18	7.893	0.013	1955	164	91±4	0.027	4.2	-13
MH1-LowpH	26.0	5.18	7.564	0.008	1942	79	72±10	0.033	3.0	-37
MH2-Exp1	26.0	5.18	7.617	0.01	2064	94	97±47	0.050	2.2	-47

Appendix A: Model Description and Discussion

A1   Model introduction

A number of excellent models already exist that describe chamber addition and three-dimensional coiling in foraminifera (e.g. Berger, 1969; Signes et al., 1993). The models we present here, by contrast, are more directly focussed on addressing questions of foraminiferal size-normalised shell weight, which can be difficult to address with these more complex types of models. The annotated source Matlab code for the model accompanies this paper. With simplicity in mind, o ur models track the mass of calcium carbonate in a particular chamber, as the foraminifera grows. With each chamber addition, the amount of the total calcium carbonate in the shell is summed up, before the content of the subsequent chamber is calculated. These data can then be normalised to overall test size to calculate metrics of size-normalised weight currently used in palaeoceanographic studies.

The mass of calcium carbonate in a particular chamber is determined from a number of prescribed morphological and wall-thickness parameters, which are derived from empirical observations. Morphological parameters used to determine a chamber's carbonate content include chamber size relative to preceding chambers, chamber aspect ratio (i.e., relative round or ovoid), and relative overlap with previous chambers (effectively hiding part of test). Wall thickness parameters include the thickness of the initial chamber wall (i.e. before subsequent thickening- the 'primary wall'), wall porosity, and the thickness of secondary layers added to preceding chambers. Wall thickness was then scaled with ontogeny either as a function of size or of chamber number (see below), up to a parameterised maximum primary wall thickness that was constrained from observations of *G. ruber* (see Table 2, Supplementary Table 1, and Supplementary Figure 1).

Three distinct models were explored to assess the importance of the choice of growth model (illustrated schematically in Fig. A1):

– **Model 1**. Ontogenetic changes in calcification were modelled as a function of body size, with calcification changes controlled by the shape of the relationship between wall thickness and increasing chamber size.

– **Model 2**. As model 1, except that calcification changes through ontogeny were mainly driven by varying the maximum primary wall thickness attainable in later chambers.

– **Model 3**. As model 1, except that ontogenetic changes in wall thickness were varied as a function of chamber number rather than increasing chamber size.
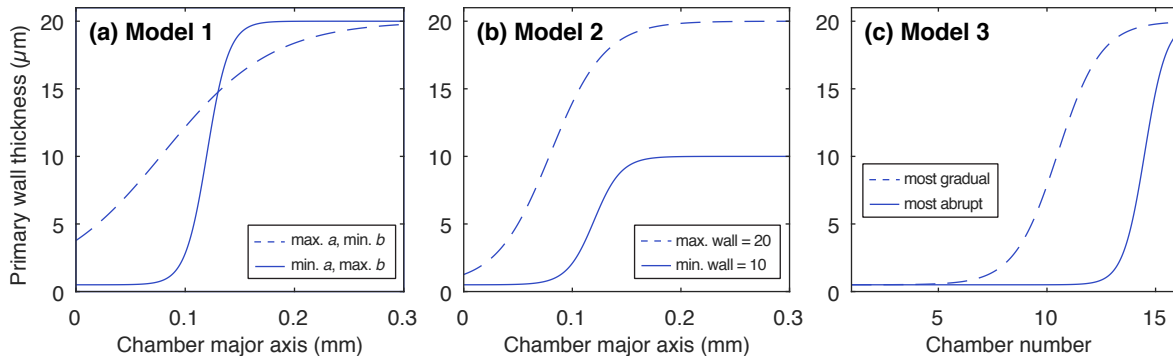

Figure A1. Scaling of wall thickness with chamber size in our models, showing the principle difference between the models described in Sect. 2.4.2. (a) Model 1 was designed to test the effect of varying the slope (parameterised by constants a and b) of this relationship. (b) Model 2 was designed to test the effect of varying the maximum primary chamber wall thickness, while keeping a and b roughly constant. (c) Model 3 parameterises wall thickness as a function of chamber number instead of body size. The two lines in each model show the maximum extent by which this regression was allowed to randomly vary, and therefore delineate model extremes. See equation A2 for the definition of this relationship.

The differences between models appear subtle, but are important. In models 1 and 2, two post-gametogenic foraminifera of different sizes would have a different calcite thickness on their final chamber – the larger individual would have thicker final chamber calcite. By contrast in model 3, two post-gametogenic foraminifera of different sizes, but the *same total chamber number*, would have the same calcite thickness on their final chamber. In a scenario where calcification is limited by availability of the necessary ions rather than energetic cost, model 1 foraminifera might accommodate unfavourable conditions by maintaining chamber dimensions at the expense of wall thickness, whereas model 3 foraminifera would maintain chamber wall thickness at the expense of chamber size and/or shape. Similarly, the difference between models 1 and 2 is subtle but important. Both models explore the role that the relationship between foraminifer size and wall thickness exert on calcification intensity. However, model 1 achieves this mainly through varying the test size at which wall thickness increases begin to rapidly ramp up, whereas model 2 instead varies the maximum primary wall thicknesses attained in later chambers (although we stress that some flexibility was maintained in all parameters). Together, these models cover a range of possible scenarios for parameterising ontogenetic and environmental calcification change, and hence provide a solid framework for understanding the size-dependence of inferred CI and the apparent conflict in previous studies.

A2 Model construction

Each model was constructed on the basis of sequential addition of spheroid chambers of a set porosity. Chamber addition is initialised relative to a specified first chamber (i.e., proloculus) with an average diameter of 10 μm as based on empirical measurements (Table 2). Approximate chamber volume was calculated by modelling chambers as spheroids:

$$\text{chamber volume} = 4/3\pi \times \text{semiaxis}_1 \times \text{semiaxis}_2^2 \tag{A1}$$

Where semiaxis_1 is half the Feret's diameter of the spheroid chamber, and is equal to semiaxis_2 multiplied by the prescribed chamber aspect ratio (Table 2). Chamber addition continued in steps until all chambers were added. *G. ruber* typically adds between 15 and 17 chambers (Parker, 1962), so a terminal chamber count of 16 was used in all models (Table 2).

Specific morphometric measurements from which the ontogenetic model was constructed are given in Supplementary Table 1 and Supplementary Figure 1. These are based on core-top *G. ruber* from the equatorial Atlantic; see main text Section 2.4.1 for full description. Because we observe no significant trend in any of these parameters as a function of size (Supplementary Figure 1), we use the mean of all measurements to derive the numbers stated in Table 2. Porosity measurements were on foraminifera spanning a smaller size range than those from which the other measurements were taken (250-425 μm), so we cannot constrain possible ontogenetic changes in porosity based on this dataset.

The three important features of the model are: 1) ontogenetic growth and parameterisation with random variation; 2) independent variation of test size and chamber size; and 3) the inclusion and effect of non-linear ontogenetic changes. Firstly, with each step in ontogeny (i.e., each chamber addition) the amount of calcium carbonate added is determined as a function of body size in the preceding step (size-models, Models 1 and 2) or chamber number (chamber-model, Model 3) based on the empirical *G. ruber* measurements in Table 2 and Supplementary Table 1. Specifically, on average, each new chamber had a major axis length 15% greater than the preceding one, a chamber aspect ratio of 1.66 (major/minor chamber axis length), an overlap of 45% with previous chambers, and a wall porosity of 4.2%. With each chamber addition, the total mass of carbonate in the foraminifera was determined as a function of the size, shape, and porosity of the newly added chamber (parameters described above), the thickness of calcite wall of the new chamber (which varied ontogenetically according to the three models listed above, described in detail below), and the addition of secondary calcite to the pre-existing test (on average parameterised as 67% of the cumulative carbonate content).

For a given model run, each parameter (i.e., model input) was varied randomly from the mean parameter listed in Table 2 by ±10%. This allowed us to explore parameter space and account for uncertainty in each biometric input. Whilst individual foraminifera may deviate by more than 10% from the population mean for any given parameter (see morphological measurements of *G. ruber* in Supplementary Table 1 and Supplementary Figure 1), we assume that the population mean (the parameter primarily of interest in calcification intensity studies) does not vary beyond this range.

Secondly, to calculate a size-normalised weight as in culture and field studies, the model requires a measure of total foraminiferal size (approximated as the sum of the maximum and perpendicular axes). We modelled this independently as a cumulative function, based on the observed increase in foraminifer diameter per chamber addition from our measurements. The test maximum axis (i.e. maximum Feret's diameter) was increased by 19% on average per chamber addition, with an average test major/minor axis ratio of 1.16. In this way, the foraminiferal test area was not derived directly from the size of the modelled chambers. Instead, chamber size and foraminifer size were allowed to vary independently from each other. A benefit of this approach (besides computational efficiency) is that foraminiferal morphology varies in the third dimension (height), so by modelling chamber size and foraminiferal size independently we effectively are capturing this

third dimensional variation. A limitation of this model structure is that it can produce morphologically impossible scenarios, for example foraminifer with a maximum test Feret's diameter less than that of the final chamber. We therefore used natural population measurements of *G. ruber* to filter model parameter combinations. Specifically, randomly generated model combinations were discarded if they a) produced impossible morphological scenarios, like that described above or b) produced area-mass ratios far outside of that observed in natural *G. ruber* populations in the Gulf of Aqaba (Eilat). In the second case, models falling outside of root mean sq. error of 3.12 of the best-fit regression line of the Gulf of Aqaba populations were discarded. In practical terms, this allows for up to a factor of two change in the size-mass relationship given in Eq. 1. This range in successful models constitute ∼8% of the total runs for each model type.

A third key feature of this model is the inclusion of non-linear ontogenetic changes in calcification. Many aspects of foraminiferal morphology, including chamber dimensions and coiling, primary wall thickness, and porosity are known to change non-linearly over ontogeny. Brummer et al. (1986, 1987) noted distinct phases in foraminiferal morphology correlated with chamber count, a finding supported by new results from Schmidt et al. (2013). For *G. ruber*, we lacked the in-depth ontogenetic measurements available for other modern species (Schmidt et al., 2013), and so we parameterise all non-linear growth functions as non-linear primary wall thickness functions (described below; note for clarity, we define primary wall thickness as the thickness of a given chamber wall after precipitation but prior to any secondary calcification during later chamber additions). We recognise that similar transitions likely occur in other aspects of morphology in step with changes in wall thickness, but since wall thickness is likely volumetrically to be by far the most important (rather than e.g. differential porosity in early chambers), it seems reasonable to essentially treat primary wall thickness as a proxy for all such conflated transitions. Our empirical measurements of *G. ruber* wall thickness (Table 2, Supplementary Table 1) reveal that there is no relationship between wall thickness and test diameter above a diameter of ∼ 250 μm, with wall thickness remaining roughly constant at ∼20 μm. We therefore define three logistic relationships to describe the co-variation of primary chamber wall thickness with foraminifer size (shown schematically above in Fig. A1)), such that:

$$\text{primary chamber thickness} = \frac{\text{maximum primary thickness}}{1 + \exp(a(x - b))} \qquad (A2)$$

where a and b are constants, and x is a measure of test size- either the previous chamber's major axis (models 1 and 2) or chamber number (model 3). The coefficients that define how wall thickness increases over ontogeny describe the shape of a non-linear relationship. The first coefficient (a) controls how tightly curved the regression is; models with more negative values of a have thinner chambers during earlier stages of growth, but then more quickly transition to growing chambers with maximum primary wall thickness. The second coefficient (b) defines the size at which the foraminifera begin to build thicker chambers, i.e. models with lower values of b reach maximum primary wall thickness is reached at a lower chamber diameter (Supplementary Figure 2).

Maximum primary wall thickness in models 1 and 3 is fixed at 20 μm ($\pm$10%) on the basis of our core-top measurements of *G. ruber*. In model 2, to test the influence of this parameter, a and b were constrained to within 10%, but maximum test wall was allowed to vary more widely, between 10-20 μm. In models 1 and 3, the shape of the logistic regression (as described by a and b in Eq. A2 above, and illustrated in Supplementary Figure 2) was allowed to vary considerably as, to our knowledge, it is unconstrained by observation. Limits were set only by the post-hoc screening of models that fell outside of the possible range of natural populations (for example, those in which proloculus wall thickness is greater than the chamber diameter).

Because this simple model can be run quickly, we could allow model parameters to vary randomly (within tolerances) and independently of each other, so as to interrogate the parameters driving CI. As stated in the main text, any individual parameter can drive calcification intensity if varied in isolation and unconstrained by the requirement of an ontogenetic model to approximate size-mass relationships seen in the open-ocean samples. However, when we examine only the subset ($\sim 8\%$) of model permutations that produce realistic size-weight relationships,

5 no significant relationship between CI and any one parameter listed in Table 2 is observed. This is demonstrated in Supplementary Figure 3 by plotting the relationship between modelled CI against all input parameters (at a body size of ~ 0.1 mm^2, for the case of two chamber additions). None of these parameters alone drives CI, which as we discuss in the main text, has the implication that a change in one may be offset by a change in another (that shifts CI in the opposite direction) in order to produce foraminifera that conform – within tolerances – to the observed size-weight relationship for this species. In contrast, the parameterisation (a and b) of the relationship between wall thickness and

10 body size or chamber number unavoidably drive CI irrespective of simultaneous changes in the other parameters. This is the case irrespective of whether primary wall thickness is defined as a function of chamber diameter (model 1) or chamber number (model 3) (Fig. A2, see also Sec. 2.4.2). However, changing the maximum wall thickness that is reached at maturity while keeping a and b fixed to within $\pm 10\%$ (model 2) has no significant effect on CI. The dependence of CI on coefficients a and b in isolation is also far less strong than when both vary in tandem (Figure A2, right hand panels). This suggests it is the overall shape of the regression between wall thickness and ontogeny, rather

15 than just one of its constituent coefficients, which is important in controlling CI – thereby leading us to our hypothesis of size-dependent calcification responses.
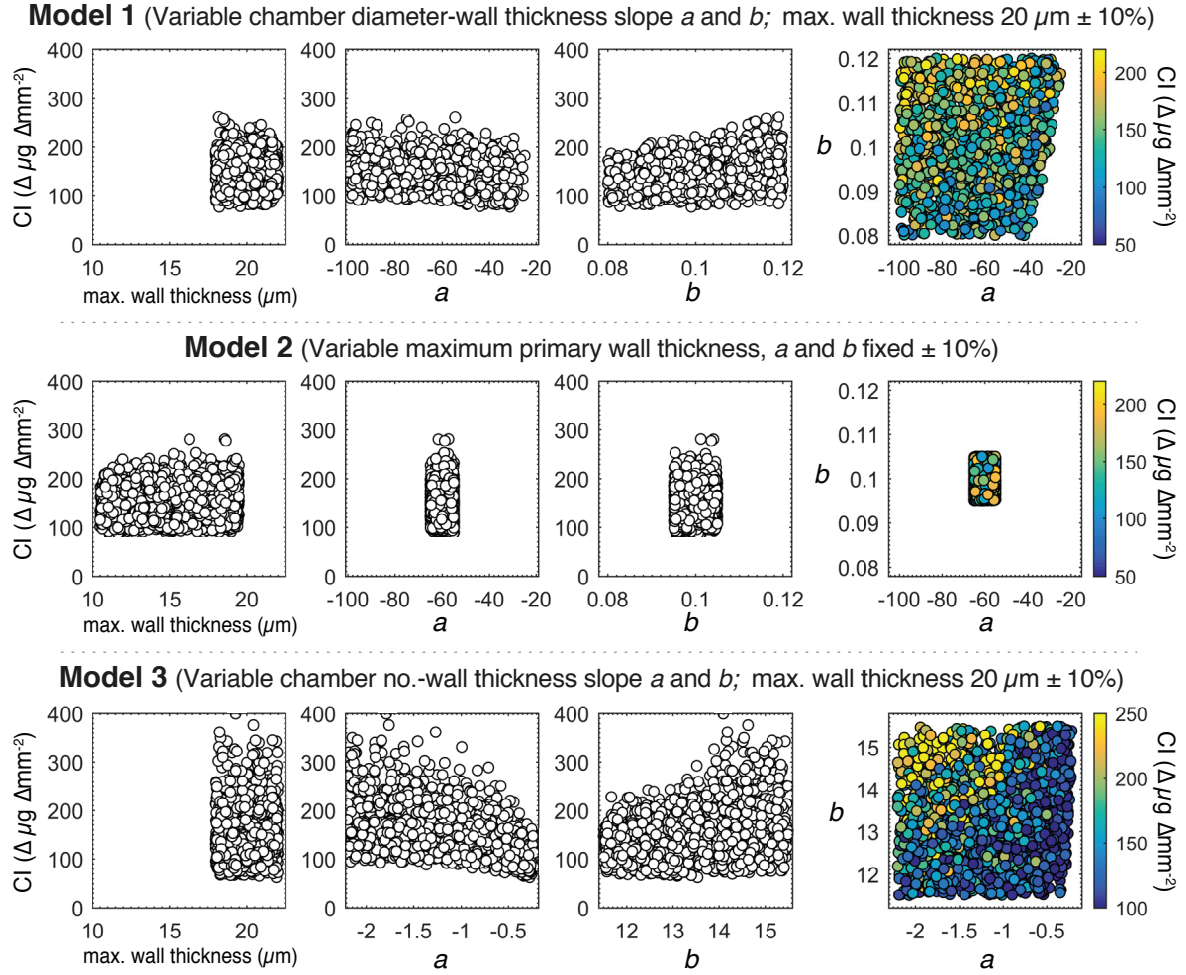
Figure A2. Modelled dependence of CI on morphological parameters within the three model groups for foraminifera of $\sim$0.1 mm^2. No parameter exerts a systematic control on CI in foraminifera of this size with the exception of the coefficients a and b that parameterise the relationship between body size and primary wall thickness (see Eq. A2). In Model 1 (top row), these coefficients are varied, but the maximum primary wall thickness (i.e. before secondar thickening) is kept at 20 ± 2 μm. In model 2, conversely, a and b are constrained, with maximum chamber thickness allowed to vary. In model 3, wall thickness is scaled with at chamber addition step, rather than chamber size. Regardless of base model, varying a and b will drive CI changes (bottom and top row, central pairs). Varying maximum wall thickness does not drive such a relationship (middle row left). When changed together, the a and b explain $\sim$50% of the variance in CI, as illustrated by the coloured plots (models 1 and 3, top- and bottom-right panels).

A3   Model caveats

An inherent limitation in our models is that parameter sets are not utilised if they do not produce realistic mass-size curves matching the natural mass-axes product relationship of *G. ruber* in the Red Sea. However, a substantial relaxation of this tolerance does not significantly

change our results. Moreover, relaxing this tolerance too far would result in mass-size relationships that can no longer be reasonably assumed to represent the species *G. ruber*. Although we note that the tolerances permitted for matching the natural populations are quite large (they allow for change in the size-mass relationship by a factor of two), this natural population itself was sampled from within a narrow range in ambient pH. As a result, we cannot unequivocally rule out a pH-induced change in one of the input parameters in isolation to change CI in response to acidification. That said, to our knowledge there is little empirical support for factors such as chamber aspect ratio or porosity to respond drastically to acidification. In addition, in the open ocean, trade-offs in allocation of calcification resources likely operate that make it difficult for one morphological parameter alone to drive CI. For example, a decrease in the thickness of secondary calcite layering might be compensated for by building smaller chambers so as to ensure structural integrity is maintained. Similarly, the need for cellular defence would preclude an increase in porosity of the magnitude that would be required to effect large changes in CI. Thus we suggest that allowing all parameters to vary randomly and then screening models may be more realistic in reproducing morphological variability in natural populations.

Another potential limitation of this model may stem from our parameterisation of shell thickening as occurring as a series of discrete additions concurrent with each chamber formation (*sensu* Erez, 2003). Emerging findings from species such as *Neogloboquadrina dutertrei* (e.g. Fehrenbacher et al., 2017) suggest that at least some foraminifera may first add chambers until they reach a final test size, and then subsequently thicken all chambers continually over some days prior to gametogenesis. Note that this is distinct from concepts such gametogenic calcite addition or encrustation, and refers specifically to ontogenetic thickening. It has been argued that the same continual thickening processes are observed in *Orbulina universa* (Spero et al., 2015). However, to date, there has been no investigation confirming the existence of such a thickening mechanism in *G. ruber*. Thus, while future work could yet reveal some secondary thickening process at work, at present we lack the observational constraints required to incorporate such a mechanism in our models. In a practical sense, however, we suspect the conceptualisation of shell thickening chosen is not likely to greatly impact our conclusions, for two reasons. Firstly, even with observed end-stage thickening in *N. dutertrei*, older chambers are often more heavily thickened than later chambers (Fehrenbacher et al., 2017), which could in a post-hoc sense result in similarly thickened older chambers as in the models we use here, even if the ontogenetic pathway to achieving this differs. Secondly, and perhaps more crucially, our model ontogenies are screened against size-weight relationships in towed foraminifera, and so by necessity our modelled ontogenetic trends must approximate true physiology. Nonetheless, should secondary thickening be observed in *G. ruber* in future, these sorts of modelling exercises should be revisited.