A User Guide for Choosing Planktic Foraminiferal Mg/Ca-Temperature Calibrations

Yair Rosenthal1,2, Samantha Bova1,3, and Xiaoli Zhou1,4

1Department of Marine and Coastal Sciences, Rutgers, State University of New Jersey, New Brunswick, NJ, USA, 2Department of Earth and Planetary Sciences, Rutgers, State University of New Jersey, New Brunswick, NJ, USA, 3Now at, Department of Geological Sciences, San Diego State University, San Diego, CA, USA, 4Now at, State Key Laboratory of Marine Geology, Tongji University, Shanghai, China

Abstract Although foraminiferal magnesium/calcium (Mg/Ca) measurements are now widely used for reconstructing sea surface temperature (SST), there is uncertainty about the fidelity of different calibrations of this proxy. Whereas single-variable calibrations suggest temperature sensitivity of about 9–10% per °C, multivariable calibrations suggest ~6% per °C, with additional dependence on salinity and pH. Here, we apply five calibrations to six published Mg/Ca records of Globigerinoides ruber, a planktic foraminifer typically used for reconstructing low latitude SST during the Pleistocene. Reconstructed SST by the different calibrations, spanning the past 250,000 years, can be offset by a few degrees, possibly reflecting variable dissolution or hydrographic effects. However, for 4 out of 5 calibrations, the reconstructed temperature anomalies yield estimates that are consistent within the calibrations’ uncertainty (<±1 °C), despite the fundamental differences in temperature sensitivity among the equations. We further propose a new seawater Mg/Ca record for the late Neogene and show that the same consistency holds for longer time scales (~4 My) independently of the choice of the calibration or which seawater Mg/Ca record is used. These comparisons attest to the robustness of the calibrations despite all the confounding nonthermal effects, and offer an empirical basis for researchers and reviewers to judge the records without any prejudice about which calibration is the “best” and evaluate their uncertainties.

Plain Language Summary The ratio of magnesium to calcium (Mg/Ca) in the shells of foraminifera (single-celled marine organisms) is sensitive to temperature, and therefore is used for reconstructing seawater temperatures. However, salinity and the pH exert additional influence. Recently, multivariable equations, which account for these nonthermal effects, have been developed to replace the simpler calibrations that account only for temperature variability. By applying the different calibrations to the same Mg/Ca records in several cores, we find that the inconsistencies among the calibrations are mostly below ±1 °C, despite the additional effects of salinity and pH on Mg/Ca-derived temperatures. We further show that these results may be applicable for at least the past 4 million years.

1. Introduction

Reconstructions of the past temperature evolution of the ocean are key for understanding Earth’s climatic history. Since the publication of the first calibrations for planktic (Elderfield & Ganssen, 2000; Lea et al., 1999; Nürnberg, 1995; Nürnberg et al., 1996) and benthic foraminifera (Lear et al., 2002; Rosenthal et al., 1997) more than 20 years ago, foraminiferal magnesium/calcium ratios (Mg/Ca) have become a major tool in the paleo-proxy arsenal, allowing the reconstruction of sea surface (e.g., de Garidel-Thoron et al., 2005; Lea et al., 2000a), thermocline (e.g., Ford et al., 2015; Hollstein et al., 2018), and deep ocean temperatures (e.g., Elderfield et al., 2012; Lear et al., 2000; Sosdian & Rosenthal, 2009; Woodard et al., 2014). Furthermore, the combination of Mg/Ca and oxygen isotope (δ18O) measurements, done on foraminiferal shells, has offered a new way for studying long-term changes in Earth’s climatic history, including the Cenozoic climate history and evolution of the cryosphere (e.g., Lear et al., 2000, 2004), ocean heat content (Rosenthal et al., 2013, 2017; Woodard et al., 2014), hydroclimate variability (e.g., Hollstein et al., 2018; Oppo et al., 2009), changes of climate modes (e.g., Marchitto et al., 2010; Moffa-Sánchez et al., 2014; Morley et al., 2014), climate sensitivity (e.g., Elderfield et al., 2012; Irvalı et al., 2020), and more. As application of Mg/Ca paleothermometry reconstructions has increased over the past two decades, it has also become clear that although temperature is likely the dominant control on foraminiferal Mg/Ca, other variables need to be considered in the interpretation of foraminiferal Mg/Ca records. These include analytical...
considerations related to the way samples are cleaned prior to analysis (Barker et al., 2003; Martin & Lea, 2002; Rosenthal et al., 2004), diagenetic effects related to dissolution (e.g., Regenberg et al., 2006, 2014; Rosenthal et al., 2000), secondary overgrowth (Kozdon et al., 2013), ontogenic effects (e.g., Elderfield et al., 2002; Friedrich et al., 2012) as well as nonthermal effects on the incorporation of Mg into the calcitic lattice including salinity and pH (Kisakurek et al., 2008; Lea et al., 1999). Another confounding factor is the long-term variation (>1 Myr) in seawater concentrations of Mg and Ca, and the effect on foraminiferal Mg/Ca (Evans & Müller, 2012). These secondary effects have led to some confusion among users and eroded confidence among reviewers.

As mentioned above, analytical considerations and postdepositional diagenetic effects on planktic foraminiferal Mg/Ca have been previously discussed at length elsewhere, and therefore will not be covered in detail here. Likewise, the role of nonthermal effects on planktic foraminiferal Mg/Ca have been studied both using live culture experiments (Kisakurek et al., 2008; Lea et al., 1999) and field calibrations based on sediment trap (Gray et al., 2018) and core-top samples (Hönisch et al., 2013), leading to the formulation of multivariable calibrations, which ascribe different sensitivities to the Mg/Ca-temperature dependence. While the sensitivities vary among species, there is a significant difference between the two classes of calibrations as well. For example, the single-variable calibrations for the surface-dwelling species *Globigerinoides ruber*, that are often used for sea surface temperature (SST) reconstructions attribute all the change in Mg/Ca to temperature with sensitivity of about 9–10% per °C (Anand et al., 2003; Dekens et al., 2002; Elderfield & Ganssen, 2000; Rosenthal & Lohmann, 2002). The multivariable calibrations, however, consider temperature sensitivity of about 6–7% per °C (e.g., Gray & Evans, 2019; Gray et al., 2018; Saenger & Evans, 2019; Tierney et al., 2019). At face value, this difference may result in significantly different temperature estimates derived from foraminiferal Mg/Ca records. For example, it has been suggested that estimating the LGM-Holocene SST changes in the equatorial Pacific based on a Mg/Ca sensitivity of 6% per °C for *G. ruber*, without accounting for changes in surface salinity and pH, would result in overestimation of the anomaly by about 1.5 °C (or an apparent warming of ~4 °C) instead of ~2.5 °C between the LGM and Holocene (e.g., Gray & Evans, 2019).

The aim of this paper is to evaluate the efficacy of different planktic Mg/Ca calibrations for reconstructing SSTs. We do that by comparing SST reconstructions based on single-variable (temperature) calibrations with multivariable (*T*, Salinity, pH) calibrations applied to the same Mg/Ca data sets spanning both the late Pleistocene glacial-interglacial variability and the long-term changes since the mid-Pliocene.

2. Methodology

Temperature records presented here are based on the original Mg/Ca data sets from the published papers. To each data set, we applied all or some of the following calibrations:

2.1. Single-Variable Calibrations

Anand et al. (2003) multispecies:

1. \( \text{Mg/Ca} = 0.38 \exp(0.09T) \)

Dekens et al. (2002) using the core's water depth to apply a constant correction for foraminiferal shell dissolution:

1. \( \text{Mg/Ca} = 0.38 \exp[0.09(\text{SST} - 0.61(\text{core depth km}))] \) for *G. ruber* in the Atlantic
2. \( \text{Mg/Ca} = 0.38 \exp[0.09(\text{SST} - 0.61(\text{core depth km})) - 1.6^\circ C] \) for *G. ruber* in the Pacific and Indian
3. \( \text{Mg/Ca} = 0.37 \exp[0.09(T - 0.36(\text{core depth km})) - 2.0^\circ C] \) for *Trilobatus sacculifer* in the Pacific

Dekens et al. (2002), which includes a time-dependent correction for dissolution using the bottom water \([\Delta CO_3^{2-}]\) record as a measure of calcite saturation where \([\Delta CO_3^{2-}] = [CO_3^{2-}] - [CO_3^{2-}]_{sat}\) and \([\Delta CO_3^{2-}] > 0\) marks saturated water and \([\Delta CO_3^{2-}] < 0\) marks unsaturated water with respect to calcite (Broecker & Peng, 1982):

1. \( \text{Mg/Ca} = 0.33 \exp[0.09 \times (\text{SST} + 0.042[\Delta CO_3^{2-}])] \) for *G. ruber* in all oceans
Table 1

| Site            | Lat  | Long | Depth | SST  | Sal | ∑CO₂ | Alk |
|-----------------|------|------|-------|------|-----|------|-----|
| GeoB1523-1     | 3.8  | −41.6| 3,292 | 27.6 | 36  | 2,000| 2,350|
| ODP site 999   | 12.0 | −78.8| 1,800 | 27.6 | 36.5| 2,000| 2,350|
| ODP site 806   | 2.0  | 159.2| 2,500 | 29   | 34  | 2,000| 2,300|
| TR163-19       | 2.3  | −91.0| 2,348 | 26   | 34  | 2,000| 2,300|
| WIND28K        | −10.1| 51.8 | 4,157 | 27   | 35  | 2,000| 2,300|
| MD03-2707      | 2.5  | 9.4  | 1,295 | 27.5 | 33  | 1,900| 2,200|

*This is the effective depth based on the sill depth to the basin. Actual depth is 2,830 m.

2.2. Multivariable Calibrations

1. Gray and Evans (2019): Mg/Ca = exp(0.036(S − 35) + 0.064SST − 0.87(pH − 8) − 0.03) for multispecies

For records of the past 150 kyr, temperatures can be calculated either with the Matlab code of Gray and Evans (2019) or using the online program developed by Will Gray (https://willygray.shinyapps.io/mgcarbv1/), which includes built-in modules for estimating salinity and pH. We note that both programs yield identical results when fed with the same data set. However, as discussed below, the Matlab code offers more flexibility. Estimates of pH variability can be done either using planktic δ13B isotope data when available from the same or a close by site or based on the atmospheric pCO₂ record. It is noteworthy that the program offers a few calibrations including a multispecies and *G. ruber* (white). Here, we use the multispecies calibration, as the latter consistently yields temperatures that are ∼1 °C colder. It is noteworthy that the Gray and Evans (2019) calibration does not include a dissolution correction and hence yields apparently cooler SSTs. If a dissolution correction is necessary, it needs to be applied before applying the data to the program, which has not been done here.

1. Tierney et al. (2019): a Bayesian calibration of Mg/Ca known as BAYMAG (https://github.com/jesstierney/BAYMAG). The calibration is based on similar sensitivities as in Gray and Evans (2019). Calculation of in situ salinity and pH is built into the calibration package. The program also includes optional corrections for dissolution and interlaboratory cleaning biases; the latter were not applied as we use the same Mg/Ca data for all the calibrations thus avoiding interlaboratory inconsistencies.

Initial hydrographic conditions, including salinity, surface alkalinity, pCO₂ disequilibrium, or pH required for both programs are taken from various published data sets (Levitus & Boyer, 1994); Geochemical Sections (GEOSECS, 1999) and World Ocean Circulation Experiment (WOCE Data Products Committee, 2002) and from Takahashi et al. (2014). Sensitivity tests that we ran with these programs suggest that errors in the initial condition can propagate to a few tenths of a degree.

1. Saenger and Evans (2019): Mg/Ca = 0.685exp(0.058SST) + 0.928 Ω(deep) for *G. ruber*

This equation includes time-dependent corrections for dissolution using changes in bottom water calcite saturation (Ω<sub>deep</sub>).

Here, we compare the degree of disagreement among the different calibrations, when applied to the same data sets in different sites (Table 1), relative to uncertainties of the calibrations. We compare both the absolute SST estimates and the temperature anomalies relative to the core-top temperature. The 1 SD errors (1 SD) on these calibrations as stated in the published papers are similar within ±1–1.5 °C. Specifically, Dekens et al. (2002) and Anand et al. (2003) ±1.2–1.4°; Gray and Evans (2019) ±1 °C; Tierney et al. (2019) ±1.5 °C. Thus, the pooled 1 SD error on these calibrations is about ±1.4–1.8 °C. However, better uncertainty estimates can be obtained from the built-in Monte Carlo iterations in the multivariable program. For the single-variable calibrations, uncertainties from Monte Carlo estimates can be obtained using the PSU Solver program of Thirumalai et al. (2016).

3. Results and Discussion

Below we compare the single-variable multispecies equations of Dekens et al. (2002) and Anand et al. (2003) with the recently published multivariable equations. We first look at the *G. ruber* (white) Mg/Ca record from the western tropical Atlantic (GeoB1523-1; 3.83°N, 41.62°W, 3,292 m; Henehan et al., 2013, supporting information) for the past 30 ky, discussed in Gray and Evans (2019; Figure 2 in that paper). Following their rationale, we compare in Figure 1a the temperature anomalies relative to the core-top temperature (∆T), calculated assuming a temperature sensitivity (∆(Mg/Ca)/∆T) of 6% per °C without, and with corrections for salinity and pH effects. Following Gray and Evans (2019), the correction for the pH effect is done either using a planktic δ13B record.
from the same core (Henehan et al., 2013) or estimates based on ice record of atmospheric $p$CO$_2$. As shown in the original paper, whereas $\Delta T$ estimates for the last $\sim$15 kyr show no significant difference among the calibrations, when using a temperature sensitivity of 6% per °C the LGM estimates differ by about 1 ± 0.4 °C between the corrected (δ$^{11}$B or $p$CO$_2$) and uncorrected records. In Figure 1b, we compare the corrected records published in Gray and Evans (2019) with a $\Delta T$ record calculated with the Anand et al. (2003) multispecies calibration, which assumes a temperature sensitivity $\Delta$(Mg/Ca)/$\Delta T$ of 9% per °C. Calculated this way, the difference among the three records is <±1 °C throughout the past 25 kyr (i.e., all three calibrations yield consistent temperatures within the uncertainty of the methods). Noticeably, the records based on the Anand et al. (2003) calibration and the record estimated with the Gray and Evans (2019) $p$CO$_2$ calibration are identical and the Tierney et al. (2019) and Gray and Evans (2019) δ$^{11}$B corrected calibrations yield ∼0.7 °C lower estimates during the LGM (Figure 1b).

Next, we compare SST records from the Caribbean Sea (ODP site 999; 12°N, 78.8°W; 2.83-km water depth) generated by different calibrations applied to the same G. ruber (white) Mg/Ca data for the last 130 kyr (Schmidt et al., 2004a, 2004b; data: https://www.ncdc.noaa.gov/access/paleo-search/study/2602; Figure 2). Because of the shallow sill depth (∼1.8 km) of the Caribbean Basin, the cores are bathed by Atlantic intermediate water, and are well above the lysocline with modern bottom water [CO$_3^{2-}$] concentrations of >30 μmol/kg, exceeding the threshold for foraminifer shell dissolution of 21.3 ± 6 μmol/kg (Regenberg et al., 2014). Down core changes in bottom water [CO$_3^{2-}$], obtained from the measurements of B/Ca in the benthic foraminifer Planulina wuellerstorfi from a nearby core (V28-122, 12°N, 79°W, 3.620-m water depth; Yu et al., 2010), show that [CO$_3^{2-}$] was consistently above 30 μmol/kg with higher saturation during the glacial interval, suggesting good shell preservation throughout the entire record. Thus, for this site, we avoid any dissolution correction. We compare the calibrations of Anand et al. (2003), which is identical to Dekens et al. (2002) without the dissolution correction, Tierney et al. (2019), Gray and Evans (2019) corrected both with atmospheric $p$CO$_2$ and a planktic δ$^{13}$B record from the same site (Foster, 2008a; data: https://doi.pangaea.de/10.1594/PANGAEA.716665) and keeping Ω$_{\text{deep}}$ at a constant modern value (∼1.5).

Comparing the SST records from the five calibrations we find strong similarity between the Anand et al. (2003) and Tierney et al. (2019) records, with core-top estimates consistent with the modern SST (Figure 2a). The multispecies calibration of Gray and Evans (2019), using both $p$CO$_2$ and δ$^{13}$B corrections, yields 2–3 °C lower SST estimates. Using the G. ruber (white) instead of the multispecies calibration in the Gray and Evans (2019) program results in even lower temperature estimates (not shown). At this site, although we have not used the

Figure 1. (a) Comparison of temperature anomalies ($\Delta T$) calculated from G. ruber (W) Mg/Ca record from the western tropical Atlantic (GeoB1523-1; 3.83°N, 41.62°W) using the Gray and Evans (2019) calibration with and without corrections for salinity and pH effects (dashed line and open circles); (b) same as (a) but with records calculated with the Tierney et al. (2019) and Anand et al. (2003). Note that while the calibrations of Anand et al. (2003) and Dekens et al. (2002) assume a temperature sensitivity of ($\Delta$(Mg/Ca)/$\Delta T$) of 9% per °C the multivariable calibrations assume a temperature sensitivity of 6%.
Figure 2. Comparison of temperature reconstructions calculated from G. ruber (W) Mg/Ca record from the Caribbean Sea (ODP site 999; 12°N, 78.8°W; 2,830 m water depth) using five different calibrations. (a) Sea surface temperature (SST) reconstructions. Note that because the core is bathed with over-saturated waters, no dissolution correction is applied in any of the calibrations. Black square marks modern SST; (b) same as (a) for temperature anomalies, ΔT, without Saenger and Evans (2019) calibration.

The dissolved corrected calibration of Dekens et al. (2002) (neither by depth nor by [ΔCO$_2$]), the calibrations of Anand et al. (2003) and of Tierney et al. (2019) yield the warmest SST. Therefore, the observed offsets in absolute SST among the various calibrations cannot simply be attributed to the dissolution correction. The offsets, therefore, must arise from other aspects of the calibrations. However, when comparing ΔT, we find better consistency among these four calibrations throughout the past 130 kyr. The difference between the Anand et al. (2003) calibration and either the Gray and Evans (2019; both using δ$^{11}$B and pCO$_2$ corrections) or Tierney et al. (2019) calibrations is ≤0.5 °C throughout most of the record except to a short interval centered around 110 ka where the difference approaches 1 °C. The Saenger and Evans (2019) SST record yields a cooling of ~8 °C at the LGM relative to the core top, which is unrealistic and therefore we do not include it in the ΔT figure (Figure 2b).

In Figure 3, we apply the same calibrations as done in Figure 2 to a 160 kyr G. ruber (white) Mg/Ca record from ODP site 806 (0.2°N, 159.2°E, 2.5-km water depth; Lea et al., 2000b; data: https://www.ncei.noaa.gov/access/paleo-search/study/2540) in the western Pacific warm pool (WPWP). Note, however, that because there is no planktic δ$^{11}$B record from this site, we only use the Gray and Evans (2019) pCO$_2$ correction (Figure 3). We also include the calibrations of Dekens et al. (2002) and Saenger and Evans (2019), which include time-dependent corrections for dissolution using either the bottom water [ΔCO$_2$] record (Dekens et al., 2002) or changes in bottom water calcite saturation (Ω$_{calc}$; Saenger & Evans, 2019). The bottom water [ΔCO$_2$] values for this site were obtained from the measurements of B/Ca in the benthic foraminifer P. wuellerstorfi (Kerr et al., 2017a; data: https://doi.pangaea.de/10.1594/PANGAEA.892517). Since the Mg/Ca and B/Ca data were not measured on the same samples, we resampled both data sets at 1-kyr resolution and then used the [ΔCO$_2$] data to correct the SST record. This was done in other cores where both data sets are available. Today, the site is at a depth close to the lysocline but the [ΔCO$_2$] reconstruction suggests that the site experienced better foraminifer shell preservation during the glacial interval (Kerr et al., 2017b). For the calibration of Saenger and Evans (2019), down core estimates of Ω$_{calc}$ for dissolution correction were estimated using the [CO$_2$]$_{calc}$ ([CO$_2$]$_{calc}$ = [CO$_2$]$_{calc}$ + [CO$_2$]$_{bio}$) seawater [Ca$^{2+}$] = 10.3 mmol/kg, and estimates of the saturation coefficient K'$_{sat}$ considering G-IG bottom water temperature change of 4-2 °C. As in the Caribbean site, we find significant differences among SST estimates generated by the various calibrations: the Dekens et al. (2002) and Tierney et al. (2019) calibrations give realistic core-top values (28–29 °C), whereas the Gray and Evans (2019) and Saenger and Evans (2019) estimates are significantly colder (25 °C). Comparison of the ΔT records shows that the Dekens et al. (2002) and Gray and Evans (2019) calibrations yield statistically identical results. The Tierney et al. (2019) calibration suggest ~1 °C lower SST during the LGM than the other records, which is within the calibration uncertainty. Again, the ΔT record calculated with the Saenger and Evans (2019) calibration yields unrealistic LGM cooling (Figure 3c). We find similarly large deviations in all sites considered here and therefore do not show the Saenger and Evans (2019) estimates for the remaining sites.

Next, we apply the calibrations to core TR163-19 (2°16′N, 90°57′W, 2,348 m) in the eastern equatorial Pacific (Lea et al., 2000b; data https://www.ncei.noaa.gov/access/paleo-search/study/2540). Since there are no benthic foraminiferal B/Ca measurements from this site, bottom water [ΔCO$_2$] values were estimated from the measurements of B/Ca in the benthic foraminifer P. wuellerstorfi in core TT013-PC72 located in the central equatorial Pacific (0°6.82′N, 139°24.08′W, 4.3 km; Kerr et al., 2017b). Because the core is deeper than TR163-19, we adjusted the saturation [ΔCO$_2$] for the depth difference assuming constant [CO$_2$] concentration at both depths. The Dekens et al. (2002) depth corrected record gives core-top temperature of ~26 °C, close to the modern SST, whereas the other equations yield colder temperatures (Figure 4a). When comparing the records, the calibrations

ROSENTHAL ET AL. 5 of 17
of Dekens et al. (2002); both depth and $[\Delta \text{CO}_2]^{-2}$ corrected) and Gray and Evans (2019) $p\text{CO}_2$ corrected yield practically identical anomalies, whereas the calibration of Tierney et al. (2019) yields up to $\pm 1^\circ$C greater negative anomalies with the largest during the last deglaciation (Figure 4b).

In Figure 5, we apply the calibrations to core WIND28K in the western Indian Ocean (10° 09.23ʹS, 9° 23.68ʹE, 1,295 m water depth; Weldeab et al., 2007a; data: https://doi.pangaea.de/10.1594/PANGAEA.610271). This deep site is bathed by undersaturated water (Kerr et al., 2017b), with poor foraminiferal preservation. Changes in bottom water $[\Delta \text{CO}_2]$ values for this site were obtained from the measurements of B/Ca in the benthic foraminifer P. wuellerstorfi (Kerr et al., 2017b). Both the $[\Delta \text{CO}_2]$ reconstruction (Kerr et al., 2017b) and the foraminifer shell weight record suggest relatively constant preservation state during the past $\sim 70$ kyr (Kiefer et al., 2006b).

As above, the SST records show large offsets due to the different dissolution corrections but practically identical $\Delta T$ for most of the time except for the late deglaciation-early Holocene, when the Gray and Evans (2019) and Tierney et al. (2019) show $0.7^\circ$C higher than the Dekens et al. (2002) depth corrected record and the Dekens et al. (2002) $[\Delta \text{CO}_2]$ corrected record shows $-0.6^\circ$C lower $\Delta T$ (Figure 5b).

Lastly, we apply the calibrations of Dekens et al. (2002) depth corrected, Gray and Evans (2019) $p\text{CO}_2$ corrected, and Tierney et al. (2019) to the Mg/Ca record from core MD03-2707 located in the Gulf of Guinea in the eastern equatorial Atlantic (2° 30.11ʹN, 9° 23.68ʹE, 1,295m; Weldeab et al., 2007a). Surface salinity at this site is strongly influenced by freshwater outflow from the Niger and Sanaga Rivers, experiencing large seasonal salinity fluctuations at present (>5 salinity units). Long-term changes of similar magnitude likely occurred throughout the past 155 kyr due to the latitudinal migrations of the monsoon rain belt (Weldeab et al., 2007b). Because of the large freshwater inputs from the river resulting in large salinity and likely pH gradients, it offers another test to assess the influence of nonthermal effects on SST estimates. Currently there is no down core record of $[\Delta \text{CO}_2]$ estimates from nearby cores so we do not include the Dekens et al. (2002) $[\Delta \text{CO}_2]$ corrected temperature record in this comparison. Evidently, all three calibrations generate realistic core-top temperatures and the SST and $\Delta T$ anomalies are consistent within $<\pm 1^\circ$C throughout the 155 kyr record, including the last two terminations (Figures 6c and 6d). The largest offsets ($>2^\circ$C) are found during the deglaciations, when changes in riverine inputs and hence the salinity and pH at the core site and a short interval centered around 105 ka, where there is a $\sim 1.5^\circ$C offset between the Dekens et al. (2002) and the Tierney et al. (2019) and Gray and Evans (2019) calibrations (Figure 6). The deglacial offsets might be expected due to large changes in riverine input and hence the salinity and pH at the core site. Likewise, the 105 ka event coincides with Greenland stadial cold event, when there is evidence from planktic Ba/

Ca for a decrease in riverine flow causing a relative increase in surface salinity (Weldeab et al., 2007b), which is likely not accurately parameterized in the multivariable equations. We tested this by comparing to runs of the Gray and Evans (2019) calibration (Figure 7). The first run uses the prescribed settings for open ocean sites and assumes a random alkalinity range of $\pm 25$ to $+74$ μmol/kg around the modern value for each site and a constant $\Delta p\text{CO}_2$ ($\pm 40$ μatm) around the surface CO$_2$ pressure. This parametrization, which was applied for all the records above, yields $\sim 4^\circ$C change between the LGM and early Holocene. In the second run, we used the surface salinity record for this site, obtained from foraminifera Ba/Ca measurements (Weldeab et al., 2007b). In this case, the resulting LGM-early Holocene amplitude is $\sim 8^\circ$C, which is unreasonable. The large difference is because along with the offshore salinity gradient, the carbonate system parameters (DIC, ALK, and pH) change, which also need to be accounted for. We attempted to account for these variations by scaling the with the salinity record,
which improved the SST record but still resulted in 6 °C amplitude difference. Clearly, the offsets among the reconstructions closely covary with the salinity record (Figure 7c) and may be related to errors in the Ba/Ca salinity reconstruction, the parameterized carbonate system in the program, or both, which highlights the potential uncertainties in the various calibrations. We are not certain about the parameterization in BAYMG but we are impressed with the general consistency among the calibrations for most of the records. Nevertheless, it seems that caution should be taken when evaluating sites that are heavily influenced by local conditions (e.g., near rivers delta, especially events on millennial scales that may not be very well simulated by these programs).

### 3.1. Glacial-Interglacial Variability

Recent multivariable calibrations of planktic foraminiferal Mg/Ca thermometry (e.g., Gray & Evans, 2019; Gray et al., 2018; Saenger & Evans, 2019; Tierney et al., 2019), have concluded that the temperature sensitivity is about 6% rather than 9% per °C, as suggested initially by single-variable calibrations (e.g., Anand et al., 2003; Dekens et al., 2002). The latter calibrations have demonstrated additional nonthermal effects due to changes in the pH and salinity of the water. Consequently, applying only the different temperature sensitivities (i.e., 6 versus 9% per °C) to any Mg/Ca record would result in a significant bias in the reconstructed temperatures (e.g., Gray & Evans, 2019). Such a comparison ignores, however, the contribution of nonthermal effects to the temperature estimates. The temperature reconstructions, generated by different calibrations, that were applied to the same Mg/Ca offer an empirical test to assess the biases among these records. Applying the different calibrations to the same Mg/Ca data sets avoids the uncertainty associated with interlaboratory biases from cleaning and other analytical biases (Greaves et al., 2008; Rosenthal et al., 2004).

Comparing the different calibrations highlights some robust features. First, different calibrations can lead to very different absolute SST estimates. This may partially be due to the dissolution corrections, or absence of any correction, applied in each calibration (Dekens et al., 2002; Tierney et al., 2019) or to local hydrographic influences that are not very well represented in any of the equations as discussed for site MD03-2707 in Figure 6. When comparing the ΔT anomalies, however, we find, with the exception of the Saenger and Evans (2019) calibration, that all calibrations yield similar temperature estimates that are largely consistent within ~±1 °C and mostly less than that (Figures 1–6), which is within the pooled uncertainty of all the calibrations of ±1.4–1.8 °C. A closer inspection of the records shows that the largest differences among the calibrations is often found during glacial-interglacial transitions when major changes in greenhouse gases and hydrographic conditions occur, but the difference among the records is often ~±1 °C or less (Figure 8 and Table 2). Admittedly, this is a large uncertainty when amplitude of the entire glacial-interglacial signal is ~3 °C (i.e., 25% uncertainty), but we note that the magnitude of this bias is often the same as the offsets among the multivariable records. Therefore, these biases cannot simply be attributed to the different temperature sensitivities of the calibrations because it is also apparent when comparing results from the Tierney et al. (2019) and Gray and Evans (2019) calibrations, both with similar sensitivities to temperature, salinity, and pH. As shown above, uncertainties surrounding the nonthermal corrections might introduce systematic biases related to the additional influence of pH and salinity, which may be critical in coastal sites but less in open ocean environments. Changes in the shells’ preservation may also add some uncertainty.

Early core-top calibrations have suggested a large sensitivity to salinity (e.g., Arbuszewski et al., 2010; Mathien-Blard & Bassinot, 2009). However, reanalysis of the data suggests a much smaller dependence on salinity (Dai et al., 2019; Gray et al., 2018; Hertzig & Schmidt, 2013; Khider et al., 2015), consistent with culture lab experiments (e.g., Allen et al., 2016; Kisakurek et al., 2008), collectively suggesting 3–5% increase in Mg/Ca
per salinity unit. A larger role of the pH dependence is suggested by a more recent calibration (−5% to −9% per pH unit; Gray & Evans, 2019). The 90 ppmV lower than pre-industrial atmospheric pCO$_2$ should have increased surface ocean pH by ~0.1 units everywhere in the ocean (e.g., Foster, 2008b).

If that was the case, we would expect the response to lowering atmospheric pCO$_2$ to influence all the records in a similar way, which is not the case. So regional and local factors including changes in temperature, salinity, productivity, upwelling, and rivers outflow of freshwaters could have a stronger effect on local pH, explaining some of the differences among the calibrations. For example, while the glacial ocean salinity increased by ~3% relative to the Holocene due to a ~120 m drop in sea level, surface water more likely responded to climate induced changes in the location and intensity of rainfall, riverine flow, and evaporation (e.g., Gibbons et al., 2014; Weldeab et al., 2007b). Similarly, surface pH response to local forcings (e.g., upwelling or river inflow) could offset the global atmospheric effect.

The differences in SST reconstructions might also be related to different dissolution corrections applied (or not) in each calibration. Multiple lines of evidence demonstrate lowering of planktic foraminiferal Mg/Ca associated with the depth-related decrease in calcite saturation (e.g., Regenberg et al., 2014; Rosenthal et al., 2000; Sadekov et al., 2010), which can compromise the fidelity of Mg/Ca-temperature estimates. Several solutions have been suggested to correct for this diagnostic loss. These include corrections based on the core depth (Dekens et al., 2002), the bottom water saturation defined by the bottom water [ΔCO$_2^{-}$] (Dekens et al., 2002) or Ω$_{dep}$ (Saenger & Evans, 2019; Tierney et al., 2019), or a correction based on size normalized shell weight (Rosenthal & Lohmann, 2002). While the modern dependence between Mg/Ca and [ΔCO$_2^{-}$] can be determined from core-top bathymetric transects, assessing temporal changes in bottom water saturation is still a challenge for paleotemperature reconstructions. Most studies assume that dissolution primarily biases absolute temperature estimates but has only negligible effect on the down core anomalies and therefore apply a constant depth correction (Dekens et al., 2002) or no correction (Gray & Evans, 2019).

Yet this supposition has not been validated and there are reasons to believe that changes in shells preservation may also affect the temperature anomaly records. A previous study demonstrates glacial-interglacial changes in individual shell weight in response to changes in atmospheric pCO$_2$; more calcified (i.e., heavier shells) are associated with lower glacial pCO$_2$ (Barker & Elderfield, 2002), which potentially can affect either the coprecipitation of trace elements in the shell, their loss due to preferential dissolution, or both. We tested these by comparing the depth-related trends in individual shell weight and Mg/Ca in the G. ruber shells from the Sierra Leone Rise, in the eastern equatorial Atlantic, between LGM and late Holocene (LH). The data show a clear difference in the trends of individual shell weight loss between the LGM and LH (Figure 9), suggesting that despite their initial heavier shells, glacial specimens tend to have a stronger decrease in shell weight, likely due to stronger vertical chemical (e.g., [ΔCO$_2^{-}$]) gradients in the ocean at that time (Boyle & Keigwin, 1987). There is also a small change in the trend of Mg/Ca loss associated with the shell thinning (Figure 9). Although the significance of this change cannot be determined with the available data, it serves as a cautionary tale that dissolution might add some uncertainty to the ΔT estimates and may explain the differences in LGM-LH estimates among the calibrations. This can explain why at some sites the difference between the Gray and Evans (2019) and Tierney et al. (2019) calibrations is larger than the difference from the Dekens et al. (2002) calibration (Figure 8). Likewise, the proxies for bottom water saturation may also have relatively large errors contributing to the overall uncertainty of the Mg/Ca-temperature calibrations. Nonetheless, despite all uncertainties we find good consistency among the calibrations within their uncertainties.

---

**Figure 5.** Comparison of temperature reconstructions calculated from G. ruber (W) Mg/Ca record from the western Indian Ocean (core WIND28K, 10°09.23’S, 51°46.15’E, 4,157-m water depth) using four different calibrations. (a) Sea surface temperature (SST) reconstructions. Black square marks modern SST; (b) ΔT reconstructions.
3.2. Long-Term Changes

The residence times of Ca and Mg in the ocean are about 1 and 13 Myr, respectively (Broecker & Peng, 1982). Therefore, on longer than 1 Myr time scales, variations in the seawater Mg/Ca need to be considered when using foraminiferal Mg/Ca to reconstruct ocean temperatures (e.g., Evans & Müller, 2012). A discussion of the various solutions proposed for these corrections is beyond the scope of this paper. Here, we only evaluate the implications of using different calibrations (single-variable versus multivariable calibrations) and different seawater Mg/Ca records on SST reconstructions for the past 4 Myr applied to the T. sacculifer Mg/Ca record from ODP site 806 on the Ontong Java Plateau (Wara et al., 2005).

All the available reconstructions of past seawater Mg/Ca suggest a decreasing trend through the Neogene but differ in the magnitude of the change. A large part of the reconstructions is based on Mg and Ca measurements in different archives, including fluid inclusions in halite crystals (Lowenstein et al., 2014), fossil echinoderms (Dickson, 2002), cold calcite veins precipitated in ocean ridge flank basalts (Coggon et al., 2010), and ancient corals (Gothmann et al., 2015). Averaging these data sets, Tierney et al. (2019) generated a seawater Mg/Ca record suggesting a modest decrease of ~15% over the past 5 Myr. Two other reconstructions based on Ca isotope and Mg measurements in pore waters and sediments (Fantle & DePaolo, 2006) and the comparison between SST reconstructions from planktic Mg/Ca and the organic biomarker TEX86 (Evans et al., 2016a) suggest a stronger decrease of ~25% through the past 5 Myr. The SST reconstruction with the latter seawater correction published by Evans et al. (2016a) shows progressively warmer temperatures in the past, both because of the strong decrease in seawater Mg/Ca and the possible change in the Mg/Ca-temperature sensitivity in response to the change in seawater Mg/Ca.

Here, we propose a new record of seawater Mg/Ca for the Neogene based on the highly resolved [Ca\(^{2+}\)] reconstruction from planktic foraminiferal Na/Ca (Zhou et al., 2021) and the [Mg\(^{2+}\)] measurements in halite fluid inclusions (Brennan et al., 2013). There are only a few measurements of [Mg\(^{2+}\)] during the Neogene but given that the oceanic residence time of Mg is ~13 Myr as compared with ~1 Myr for Ca, changes on a time scale of <10 Myr should be governed by [Ca\(^{2+}\)] variability. Our new Neogene record (Table S1 in Supporting Information S1; Rosenthal et al., 2022) is consistent with pore fluid data, suggesting ~25% lower sea water Mg/Ca during the Pliocene than at present (Figure 10). It is noteworthy, however, that both estimates are not statistically different for the past ~4 Myr and the low slope in the Tierney et al. (2019) record is largely driven by the scatter in the coral data for this period.

A recent study assessed the fidelity of Mg/Ca-SST reconstructions from the WPWP for the past ~6 Myr by comparing Mg/Ca and clumped isotope (Δ\(^{17}\)) temperature estimates from mixed-layer planktic foraminifer T. sacculifer from IODP site U1488 (02°02.59’N, 141°45.29’E, 2,604-m water depth; Meinicke et al., 2021). For the comparison, they used the Gray and Evans (2019) calibration with salinity estimates derived from sea level change (Rohling et al., 2014) and the seawater Mg/Ca record of Tierney et al. (2019). Both proxies consistently show no discernible cooling trend from the mid-Pliocene to present in contrast to the reconstruction based on the organic proxy of TEX86 (Zhang et al., 2014). Therefore, we further test whether the choice of calibration or seawater Mg/Ca record can account for this discrepancy. We use the Mg/Ca records of T. sacculifer from ODP site 806 (Wara et al., 2005). We first correct the measured foraminiferal Mg/Ca for the change in seawater Mg/Ca ratio based on the reconstructions of Tierney et al. (2019) and this study using the following relationship from Evans and Müller (2012):

![Figure 6. Comparison of temperature reconstructions calculated from G. ruber (white) Mg/Ca record from the Gulf of Guinea in the eastern equatorial Atlantic (core MD03-2707; 2°30.11’N, 9°23.68’E, 1,295 m) using three different calibrations. (a) Sea surface temperature (SST) reconstructions. Black square marks modern SST; (b) ΔT reconstructions; (c) Termination I; (d) Termination II.](image-url)
was set to 0.4 following the recommendation in that paper. We then apply the corrected Mg/Ca ratios to the two calibration equations of Anand et al. (2003) and Gray and Evans (2019). The comparison in Figure 11 shows that Mg/Ca records corrected for long-term changes in seawater Mg/Ca yield SST temperatures in the mid-Pliocene warm period (MPWP) that are ~1 °C warmer than the uncorrected record, regardless of the choice of the seawater Mg/Ca record used for the correction (Figure 11b). The Gray and Evans (2019) calibration with an additional salinity correction yields temperature estimates for the MWP that are 2 °C higher than the uncorrected record with higher glacial-interglacial variability. Given the uncertainties in the long-term salinity reconstruction and error propagation, it would be reasonable to conclude that the three corrected records agree within ~±2 °C for the long-term trend and ±1 °C for the glacial-interglacial variability.

Regardless of the corrections, the Mg/Ca-derived records do not support the cooling trend from the mid-Pliocene to present suggested by TEX86 (Zhang et al., 2014). A possible caveat, however, might be the observation that the Mg/Ca-temperature sensitivity decreases with the change in seawater Mg/Ca, which would lead to greater cooling in the Mg/Ca records (Evans et al., 2016a). Since the ΔST data do not support this (Figure 11; Meinicke et al., 2021), we have not implemented this adjustment but further studies should assess the observation.

$$\left( \frac{Mg}{Ca} \right)_{corr} = \left( \frac{Mg}{Ca} \right)_{meas} \times \left( \frac{5.3}{\left( \frac{Mg}{Ca} \right)_{SW}} \right)^H$$

The power coefficient $H$ was set to 0.4 following the recommendation in that paper. We then apply the corrected Mg/Ca ratios to the two calibration equations of Anand et al. (2003) and Gray and Evans (2019). The comparison in Figure 11 shows that Mg/Ca records corrected for long-term changes in seawater Mg/Ca yield SST temperatures in the mid-Pliocene warm period (MPWP) that are ~1 °C warmer than the uncorrected record, regardless of the choice of the seawater Mg/Ca record used for the correction (Figure 11b). The Gray and Evans (2019) calibration with an additional salinity correction yields temperature estimates for the MWP that are 2 °C higher than the uncorrected record with higher glacial-interglacial variability. Given the uncertainties in the long-term salinity reconstruction and error propagation, it would be reasonable to conclude that the three corrected records agree within ~±2 °C for the long-term trend and ±1 °C for the glacial-interglacial variability.

Regardless of the corrections, the Mg/Ca-derived records do not support the cooling trend from the mid-Pliocene to present suggested by TEX86 (Zhang et al., 2014). A possible caveat, however, might be the observation that the Mg/Ca-temperature sensitivity decreases with the change in seawater Mg/Ca, which would lead to greater cooling in the Mg/Ca records (Evans et al., 2016a). Since the ΔST data do not support this (Figure 11; Meinicke et al., 2021), we have not implemented this adjustment but further studies should assess the observation.

Figure 7. (a) Comparison of the sea surface temperature (SST) record from MD03-2707 calculated with Gray and Evans (2019) calibration based on the atmospheric pCO₂ record (blue line as in Figure 6a) with the Gray and Evans (2019) calibration based on the atmospheric pCO₂ and the salinity record for this site (gray line) Black square marks modern SST; (b) temperature difference (ΔT) between the two SST records; (c) Ba/Ca-derived salinity record for this site (Weldeab et al., 2007). Note the close inverse correlation between the ΔT and salinity record.

4. Final Remarks

The influence of nonthermal effects on planktic foraminiferal Mg/Ca are well documented in both culture experiments and field calibrations. These include primary effects on the calcification process, including changes in salinity and pH as well as postmortem effects primarily due to partial dissolution of the shells in the sediments, and biases due to variable cleaning procedures during sample preparation. Indeed, recent multivariable calibrations, which also account for salinity and pH dependencies, suggest that the Mg/Ca-temperature dependency is significantly lower (~6% per °C) than suggested initially by the single-variable equations (~9% per °C), thereby questioning the application and accuracy of single-variable calibrations. Here, we have tested this argument by systematically applying all the calibrations to Mg/Ca records spanning different time scales and hydrographic conditions. Our assessment demonstrates that reconstructed surface temperature anomaly records, derived from applying different calibrations to the same Mg/Ca measurements of G. ruber, yield estimates that are consistent within ~±1 °C, despite the fundamental differences among the equations. The apparent consistency is likely because of the dependence of the carbonate system dissociation coefficients and hence parameters (pH, CO₂⁻) on temperature and salinity, which lead to fortuitous but nonetheless tight consistency among the equations.

The largest discordance, albeit still within ±1 °C, may appear during climate transitions, when changes in temperature are associated with major changes in salinity and pH. A better consistency is found during interglacials, and could even be improved if the seasonal component of the record is removed (see extended material in Bova et al. (2021)). Differences in estimates of absolute SST by the different calibrations are likely attributable to the degree of dissolution correction applied (or not) in each calibration and not to the difference in temperature sensitivities. Better estimates of either in situ salinity or pH variability, using Ba/Ca or another independent salinity proxy, planktic foraminifera δ¹³B as a pH proxy, and benthic foraminifera B/Ca as a dissolution proxy could improve the accuracy of the records, but also may impart larger errors as seen in the record of MD03-2707.
Although the plethora of calibrations and reports of potential problems with the proxy have eroded confidence in the Mg/Ca-temperature proxy, the results of our study demonstrate the robustness of the proxy, at least for the periods tested here. The tolerance for errors depends, however, on the scientific question, and in specific cases it would be beneficial to compare the resulting $\Delta T$ produced using the different calibrations. It is also noteworthy that the pre-exponent constant can be adjusted so the estimated core-top temperature matches the modern hydrographic temperature, which practically account for the absolute SST offsets. Because this does not affect the temperature sensitivity, the temporal variability (e.g., glacial-interglacial) remains unchanged and is practically identical to the $\Delta T$ records. Therefore, researchers should be encouraged to publish core-top data along with any down core data. For records extending before the late Pleistocene, uncertainties in salinity and pH, and hence temperature estimates based on multivariable calibrations, are larger and it is more difficult to evaluate the fidelity of the calibrations. For long time scales, there is a growing data set of changes in surface pH based on $\delta^{11}B$ measurements in planktic foraminifera (e.g., Sosdian et al., 2018) that can be coupled with experimental data on Mg/Ca-pH sensitivity and applied to the multivariable calibrations to improve the accuracy of Mg/Ca-derived SST estimates (e.g., Evans et al., 2016b; Sosdian & Lear, 2020). Recent studies, however,

Figure 8. Comparison of temperature anomalies estimated by the different calibrations for the LGM to late Holocene (LH) (a) and LGM to early Holocene (EH) (b). Note that except for a few cases (discussed in the text) all the calibrations yield consistent estimates within $<1$ °C. Data used for these figures are shown in Table 1.
suggest that DIC may exert significantly larger influence on Mg/Ca in some planktic foraminifera (e.g., *Orbulina universa*), whereas in others (e.g., *G. ruber*) pH seems to better describe the carbonate system dependency (Holland et al., 2020). Furthermore, Holland et al. (2020) suggest that each parameter does not affect foraminiferal Mg/Ca in isolation, and consequently proposed a multivariable calibration for *O. universa* and *G. ruber*.

**Table 2**

**LGM–Holocene Temperature Differences**

| Calibration                        | GeoB1523 | ODP999 | ODP806 | TR163-19 | WIND28K | MD03-2707 |
|------------------------------------|----------|--------|--------|----------|---------|-----------|
| Dekens 2002 depth correction       | 2.5      | 2.1    | 2.3    | 2.5      | 1       | 1.3       |
| Gray and Evans $pCO_2$ correction  | 2.3      | 2.1    | 3.3    | 2.5      | 0.6     | 0.8       |
| Gray and Evans d11B correction     | 3.0      | 2.2    |        |          |         |           |
| Tierney 2019 BAYMAG                | 3.4      | 2.3    | 3.2    | 3.5      | 1.5     | 0.9       |
| Dekens 2002 $\Delta CO_2$ correction | 2.6      | 2.4    | 1.3    |          |         |           |
| Saenger and Evans (2019)           |          |        |        |          | 5.4     |           |

| Calibration                        | GeoB1523 | ODP999 | ODP806 | TR163-19 | WIND28K | MD03-2707 |
|------------------------------------|----------|--------|--------|----------|---------|-----------|
| Dekens 2002 depth correction       | 2.3      | 2.2    | 2.9    | 2.9      | 2.4     | 3.3       |
| Gray and Evans $pCO_2$ correction  | 2.3      | 2.2    | 2.2    | 3.4      | 2.4     | 3.8       |
| Gray and Evans d11B correction     | 3.0      | 2.2    |        |          |         |           |
| Tierney 2019 BAYMAG                | 3.3      | 2.4    | 4      | 4.2      | 3.4     | 3.9       |
| Dekens 2002 $\Delta CO_2$ correction | 3.1      | 2.8    | 1.7    |          |         |           |
| Saenger and Evans (2019)           |          |        |        |          | 6.6     |           |

**Figure 9.** Bathymetric transects of (a) individual shell weight and (b) Mg/Ca in *G. ruber* (white, 212–300 μm) samples from cores recovered at different depths on the Sierra Leone Rise in the eastern equatorial Atlantic during the Endeavor 66 cruise. The transects show depth dependent decreases for core-top (orange symbols) and LGM (blue symbols) samples. The cores’ location and associated hydrographic information can be found at Rosenthal and Lohmann (2002).
Mg/Ca that in addition to temperature requires knowledge of the contemporaneous seawater calcium, surface water dissolved inorganic carbon (DIC), or [H$^+$] concentrations. The correction for changes of seawater Mg/Ca follows the approach suggested in Hasiuk and Lohmann (2010) and Evans and Müller (2012). Reconstructions of seawater [Ca$^{2+}$] are available from fluid inclusions at low resolution (Lowenstein et al., 2014) and from more recently derived from foraminiferal Na/Ca at high resolution (Zhou et al., 2021).

Other confounding factors include the change in seawater Mg/Ca and a possible change in temperature sensitivity. A recent study shows, however, that Mg/Ca-derived temperatures from the mixed-layer foraminifer, T. sacculifer, are consistent with paired estimates from clumped isotopes, thus further supporting the use of Mg/Ca thermometry for the late Neogene (Meinicke et al., 2021). We further show that for the past 4 Myr that this consistency is independent of the choice of calibration, single variable or multivariable, or the seawater Mg/Ca record, yielding estimates that agree within an uncertainty of about ±1 °C. The advantage of using this species is the apparent lack of pH dependency, which is not the case for G. ruber.

The purpose of this paper is not to point out the best calibration but rather to provide an empirical basis for researchers and reviewers to judge the records without any prejudice about which calibration is the “best.” Fundamentally, the multivariable calibrations more accurately described the dependencies of foraminiferal Mg/Ca on the various parameters. But, in practice, the consistency among the calibrations, despite fundamentally different dependencies, suggests that the interdependency among variables at the ocean surface, mainly temperature, salinity, and pH, act to generate an apparent temperature sensitivity that is consistent with the original single-variable calibration. That is not entirely surprising given the climatic dependency among variations in atmospheric $p$CO$_2$, ocean temperatures, and surface pH. Nonetheless, we cannot assume that this is the case all the time, especially over geological time scales. For these time scales, it would require independent estimates of pH from B isotope measurements in planktic foraminifera (e.g., Leutert et al., 2020; Sosdian & Lear, 2020). However, for extinct species we do not have any constraints on their sensitivity to pH and assume it is similar to their modern analogs. Given the uncertainties in reconstructing salinity and pH over these time scales, and the interspecific variability in their response to pH changes, it would be difficult to a-priori reject the use of single-variable equations unless there is evidence otherwise. Indeed, as shown here, empirical validations and the comparison to other proxies in the same archives (e.g., $\Delta$47) provide confidence that, in most cases, Mg/Ca provides robust temperature estimates regardless of the calibration used.
Data Availability Statement

The data on which this article is based were previously published and are publicly available in supporting information for Henehan et al. (2013), Schmidt et al. (2004b), Lea et al. (2000b), Kiefer et al. (2006a), Weld-eab et al. (2007a), supporting information for Wara et al. (2005), Table 1 in supporting information for Yu et al. (2010), Foster (2008a), and Kerr et al. (2017a). The seawater Mg/Ca data set is archived at NOAA (Rosenthal et al., 2022b). Multivariate calibration programs used in this paper are available from the original publications at: Gray and Evans (2019): R and Matlab codes are available on GitHub: https://github.com/willyrgray/MgCaRB for R, https://github.com/dbjevans/MgCaRB for Matlab or online at: https://willyrgray.shinyapps.io/mgcarbv1/; Tierney et al. (2019) BAYMG: https://github.com/jesstierney/BAYMAG.

Acknowledgments
This project was partly funded by the NSF Grant OCE-1834208 to YR. The authors thank Jennifer Hertzberg and David Evans for their helpful reviews. The authors thank David Evans and Will Gray for helping with the parameterization of their calibration program and comments that improved the paper.

References

Allen, K. A., Honisch, B., Eggins, S. M., Haynes, L. L., Rosenthal, Y., & Yu, J. (2016). Trace element proxies for surface ocean conditions: A synthesis of culture calibrations with planktic foraminifera. Geochimica et Cosmochimica Acta, 193, 197–221. https://doi.org/10.1016/j.gca.2016.08.015
Anand, P., Elderfield, H., & Conte, M. H. (2003). Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time-series. Paleoceanography, 18(2), 1050. https://doi.org/10.1029/2002PA000846
Arbuszewski, J., deMenocal, P., Kaplan, A., & Farmer, C. E. (2010). On the fidelity of shell-derived δ¹⁸O seawater estimates. Earth and Planetary Science Letters, 300, 185–196. https://doi.org/10.1016/j.epsl.2010.10.035
Barker, S., & Elderfield, H. (2002). Foraminiferal calcification response to glacial-interglacial changes in atmospheric CO₂. Science, 297, 833–836. https://doi.org/10.1126/science.1072815
Barker, S., Greaves, M., & Elderfield, H. (2003). A study of cleaning procedures used for foraminiferal Mg/Ca paleothermometry. Geochemistry, Geophysics, Geosystems, 4(9), 8407. https://doi.org/10.1029/2003GC000559
Bova, S., Rosenthal, Y., Liu, Z., Godad, S. P., & Yan, M. (2021). Seasonal origin of the thermal maxima at the Holocene and the last interglacial. Nature, 589(7843), 548–553. https://doi.org/10.1038/s41586-020-03155-x
Sadekov, A. U., Eggins, S., Klikhammer, G. P., & Rosenthal, Y. R. (2010). Effects of seafloor and laboratory dissolution on the Mg/Ca composition of Globigerinoides sacculifer and Orbulina universa tests—A laser ablation ICPMS microanalysis perspective. *Earth and Planetary Science Letters*, 292, 312–324. https://doi.org/10.1016/j.epsl.2010.01.039

Saenger, C. P., & Evans, M. N. (2019). Calibration and validation of environmental controls on planktic foraminifera Mg/Ca using global core-top data. *Paleoceanography and Paleoclimatology*, 34, 1249–1270. https://doi.org/10.1029/2018PA002557

Schmidt, M. W., Spero, H. J., & Lea, D. W. (2004a). Links between salinity variation in the Caribbean and North Atlantic thermohaline circulation. *Nature*, 428, 160–163. https://doi.org/10.1038/nature02346

Schmidt, M. W., Spero, H. J., & Lea, D. W. (2004b). Retrieved from https://www.ncei.noaa.gov/access/paleo-search/study/2602

Sosdian, S., & Rosenthal, Y. (2009). Deep sea temperature and ice volume changes across the Pliocene-Pleistocene climate transitions. *Science*, 323, 306–310. https://doi.org/10.1126/science.1169938

Sosdian, S. M., Greenop, R., Hain, M. P., Foster, G. L., Pearson, P. N., & Lear, C. H. (2018). Constraining the evolution of Neogene Ocean carbonate chemistry using the boron isotope pH proxy. *Earth and Planetary Science Letters*, 498, 362–376. https://doi.org/10.1016/j.epsl.2018.06.017

Sosdian, S. M., & Lear, C. H. (2020). Initiation of the Western Pacific warm pool at the middle Miocene climate transition? *Paleoceanography and Paleoclimatology*, 35, e2020PA003920. https://doi.org/10.1029/2020PA003920

Takahashi, T., Sutherland, S. C., Chipman, D. W., Goddard, J. G., Ho, C., Newberger, T., et al. (2014). Climatological distributions of pH, pCO₂, total CO₂, alkalinity, and CaCO₃ saturation in the global surface ocean, and temporal changes at selected locations. *Marine Chemistry*, 164, 95–125. https://doi.org/10.1016/j.marchem.2014.06.004

Thirumalai, K., Quinn, T. M., & Marino, G. (2016). Constraining past seawater δ¹⁸O and temperature records developed from foraminiferal geochemistry. *Paleoceanography*, 31, 1409–1422. https://doi.org/10.1002/2016PA002970

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

Wara, M. W., Ravelo, A. C., & Delaney, M. L. (2005). Permanent El Nino-like conditions during the Pliocene warm period. *Science*, 309, 758–761. https://doi.org/10.1126/science.1112596

Weldeab, S., Lea, D. W., Schneider, R. R., & Andersen, N. (2007a). Retrieved from https://www.ncei.noaa.gov/access/paleo-search/study/5596

Yu, J., Foster, G. L., Elderfield, H., Broecker, W. S., & Clark, E. (2010). An evaluation of benthic foraminiferal B/Ca and δ³⁴B for deep ocean carbonate ion and pH reconstructions. *Earth and Planetary Science Letters*, 292(1), 114–120. https://doi.org/10.1016/j.epsl.2010.02.029

Zhang, Y. G., Pagani, M., & Liu, Z. (2014). A 12-million-year temperature history of the tropical Pacific ocean. *Science*, 344(6179), 84–87. https://doi.org/10.1126/science.1246172

Zhou, X., Rosenthal, Y., Haynes, L., St, W., Evans, D., Huang, K.-F., & Erez, J. (2021). Planktic foraminiferal Na/Ca: A potential proxy for seawater calcium concentration. *Geochimica et Cosmochimica Acta*, 305, 306–322. https://doi.org/10.1016/gca.2021.04.012