Interlaboratory comparison study of Mg/Ca and Sr/Ca measurements in planktonic foraminifera for paleoceanographic research

Yair Rosenthal

Institute of Marine and Coastal Sciences, and Department of Geological Sciences, Rutgers, State University of New Jersey, 71 Dudley Road, New Brunswick, New Jersey 08901, USA (rosentha@imsc.rutgers.edu)

Suzanne Perron-Cashman and Caroline H. Lear

Institute of Marine and Coastal Sciences, Rutgers, State University of New Jersey, 71 Dudley Road, New Brunswick, New Jersey 08901, USA

Edouard Bard

CEREGE, Europole de l'Arbois BP80, 13545 Aix–en–Provence cdx 4, France

Stephen Barker

Department of Earth Sciences, University of Cambridge, Cambridge, Downing Street, Cambridge CB2 3EQ, UK

Katharina Billups

College of Marine Studies, University of Delaware, 700 Pilottown Road, Lewes, Delaware 19958, USA

Martha Bryan

Lamont–Doherty Earth Observatory, Palisades, New York 10964, USA

Margaret L. Delaney

Ocean Sciences/Institute of Marine Sciences, University of California, Santa Cruz, California 95064, USA

Peter B. deMenocal

Lamont–Doherty Earth Observatory, Palisades, New York 10964, USA

Gary S. Dwyer

Division of Earth and Ocean Sciences, Nicholas School of the Environment and Earth Sciences, Duke University, Durham, North Carolina 27708, USA

Henry Elderfield

Department of Earth Sciences, University of Cambridge, Cambridge, Downing Street, Cambridge CB2 3EQ, UK

Chris R. German

Southampton Oceanography Center, European Way, Southampton SO14 3ZH, UK

Mervyn Greaves

Department of Earth Sciences, University of Cambridge, Cambridge, Downing Street, Cambridge CB2 3EQ, UK

David W. Lea

Department of Geological Sciences and Marine Science Institute, University of California, Santa Barbara, California 93106, USA

Thomas M. Marchitto Jr.

Institute of Arctic and Alpine Research, University of Colorado, Campus Box 450, Boulder, Colorado 80309, USA

Dorothy K. Pak and Georges L. Paradis

Department of Geological Sciences and Marine Science Institute, University of California, Santa Barbara, California 93106, USA

Ann D. Russell

Department of Geology, University of California, One Shields Avenue, Davis, California 95616, USA

Ralph R. Schneider

Département de Géologie et Océanographie, UMR 5805 EPOC, CNRS/Université de Bordeaux I, 33405 Talence Cedex, France

Kathleen Scheiderich

College of Marine Studies, University of Delaware, 700 Pilottown Road, Lewes, Delaware 19958, USA

Lowell Stott

Department of Earth Sciences, University of Southern California, Los Angeles, California 90089, USA

Kazuyo Tachikawa

CEREGE, Europole de l'Arbois BP80, 13545 Aix-en-Provence cdx 4, France

Eric Tappa and Robert Thunell

Department of Geological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA

Michael Wara

Ocean Sciences/Institute of Marine Sciences, University of California, Santa Cruz, California 95064, USA

Syee Weldeab

Center for Marine Environmental Sciences (MARUM), Department of Geosciences, Bremen University, Klagenfurterstrasse, 28334 Bremen, Germany

Paul A. Wilson

Southampton Oceanography Center, European Way, Southampton SO14 3ZH, UK

[1] Thirteen laboratories from the USA and Europe participated in an intercomparison study of Mg/Ca and Sr/Ca measurements in foraminifera. The study included five planktonic species from surface sediments from different geographical regions and water depths. Each of the laboratories followed their own cleaning and analytical procedures and had no specific information about the samples. Analysis of solutions of known Mg/Ca and Sr/Ca ratios showed that the intralaboratory instrumental precision is better than 0.5% for both Mg/Ca and Sr/Ca measurements, regardless whether ICP-OES or ICP-MS is used. The interlaboratory precision on the analysis of standard solutions was about 1.5% and 0.9% for Mg/Ca and Sr/Ca measurements, respectively. These are equivalent to Mg/Ca-based temperature repeatability and reproducibility on the analysis of solutions of $\pm0.2°C$ and $\pm0.5°C$, respectively. The analysis of foraminifera suggests an interlaboratory variance of about $\pm8\%$ (%RSD) for Mg/Ca measurements, which translates to reproducibility of about $\pm2-3°C$. The relatively large range in the reproducibility of foraminiferal analysis is primarily due to relatively poor intralaboratory repeatability (about $\pm1-2°C$) and a bias (about $1°C$) due to the application of different cleaning methods by different laboratories. Improving the consistency of cleaning methods among laboratories will, therefore, likely lead to better reproducibility. Even more importantly, the results of this study highlight the need for standards calibration among laboratories as a first step toward improving interlaboratory compatibility.

1. Introduction

[2] In the past few years there has been an increasing interest in using foraminiferal Mg/Ca as a proxy for seawater paleotemperatures [*Elderfield and Ganssen*, 2000; *Hastings et al.*, 1998; *Lea et al.*, 1999; *Nürnberg et al.*, 1996; *Rosenthal et al.*, 1997]. The interest in Mg-paleothermometry is primarily due to the fact that, in principle, by making paired measurements of $\delta^{18}O$ and Mg/Ca on the same shells, one can use Mg/Ca to adjust for the temperature-dependency of $\delta^{18}O$ and isolate the $\delta^{18}O$ of seawater [*Mashiotta et al.*, 1999]. Oxygen isotope ratios, recorded in fossil foraminiferal shells, have been paleoceanographers' principal tool for constraining glacial-interglacial variations in sea level and continental ice volume. However, in addition to the climatically driven change of seawater $^{18}O/^{16}O$ composition due to changes in the amount of ^{16}O sequestered in continental ice, records of foraminiferal $\delta^{18}O$ also incorporate the temperature effect on the isotopic fractionation during calcification. The relative contribution of these two factors cannot be separated without an additional independent variable (e.g., an independent estimation of the calcification temperature). While several proxies for sea surface temperatures have been developed (e.g., faunal assemblage analysis, alkenone unsaturation index), Mg-paleothermometry offers a few distinctive advantages since it is measured on the same phase as $\delta^{18}O$. First, it gives the actual temperature at which the foraminifer shell precipitated. Second, combined measurements of foraminiferal $\delta^{18}O$ and Mg/Ca on the same specimens avoid errors potentially introduced by studying different specimens (or different organisms), which might have not lived at the same depth, season or time. Paired $\delta^{18}O$ and Mg/Ca measurements on the same samples allow for reconstructing the temporal relationships between changes in the sea surface temperature (SST) and ice-volume in the past. Depending on the oceanographic context, a record of $\delta^{18}O_{water}$ can provide valuable information about regional paleosalinity or global ice volume.

[3] It is now clear that Mg/Ca paleothermometry has reached a critical stage whereby steps should be taken to assure compatibility between measurements generated by different groups. This study was designed to examine the variability in measurements of Mg/Ca among laboratories. It also included Sr/Ca measurements because of the interest in studying secular variations in seawater Sr/Ca [*Martin et al.*, 1999; *Stoll and Schrag*, 2000; *Stoll et al.*, 1999]. In general, measurements of foraminiferal Sr/Ca seem more consistent than Mg/Ca among different laboratories; however, down core variations in foraminiferal Sr/Ca have significantly lower amplitude than in Mg/Ca. For example, glacial-interglacial variations of planktonic foraminiferal Mg/Ca are typically on the order of 20–25% and only 5–6% for Sr/Ca. The small size of the signal means that it is important also to test the compatibility among laboratories with respect to Sr/Ca. Because most laboratories are measuring both elemental ratios at the same time, we included both Mg/Ca and Sr/Ca in this study.

[4] Differences in elemental ratios measured in foraminiferal shells by different laboratories may be due to the following reasons: (1) offsets due to instrument calibration or incompatibility of the working standards; (2) offsets due to differences in methods applied by individual laboratories for cleaning foraminifera shells; (3) differences due to natural variability in the sample assemblages. The latter primarily reflects differences in the growth period and depth of calcification in the water column of shells of the same species as well as the mixing of shells of different ages after burial in the sediment. The first two items lead to systematic offsets among laboratories whereas the natural variability results in random differences in data produced by different laboratories as well as by each individual laboratory. The current study was designed to assess whether there are any significant offsets in measurements of Mg/Ca and Sr/Ca among laboratories and understand the possible causes of these offsets.

2. Experimental Design

[5] The study included 13 participants from the USA and Europe. Each laboratory received both standard solutions and foraminifera samples and was assigned a random identification number. The identification numbers along with the instruments and cleaning methods used by each laboratory are listed in Table 1. In most cases, laboratories used either inductively coupled plasma - optical emission spectroscopy (ICP-OES) or inductively coupled plasma - mass spectrometry (ICP-MS). The study was designed to assess the significance of systematic errors (i.e., those introduced by instrumental biases and/or methodological differences) relative to the natural variability in foraminiferal Mg/Ca and Sr/Ca.

2.1. Standards

[6] Two different standard solutions were prepared from three new single element standards (High Purity[®]; Ca = 1000 ± 3 ppm, Mg = 1000 ± 3 ppm, Sr = 1000 ± 3 ppm). The mixed standard solutions were prepared by spiking the appropriate volumes (determined gravimetrically) of the Mg and Sr

Table 1. Instrumentation and Cleaning Methods

Laboratory ID	Instrument	Method[a]
108	ICP-OES	1
129	DCP-OES	5
140	ICP-OES	1
142	ICP-OES	1
171	ICP-OES	4
213	ICP-MS	2
360	ICP-OES	2
558	ICP-MS	1
578	ICP-MS	2
628	ICP-OES	1
856	ICP-OES	1
867	ICP-OES	2
881	ICP-OES	1
935	ICP-OES	3

[a] Methods are as follows: (1) "Mg cleaning" method: Rinse (ddH_2O/methanol), oxidation, dilute acid leach (ultrasonication employed for clay removal and leaches); (2) "Cd cleaning" method: Rinse (ddH_2O/methanol), reduction, oxidation, dilute acid leach (ultrasonication employed for clay removal and leaches); (3) rinse (ddH_2O/methanol), oxidation, rinse (ddH_2O/methanol), dry in oven (ultrasonication employed for clay removal and leaches); (4) rinse (ddH_2O/methanol), reduction, oxidation (ultrasonication employed for clay removal and leaches); (5) oxidation with 5% sodium hypochlorite (full strength Clorox), 6x rinse ddH_2O. Oxidation, 30% H_2O_2/0.1M NaOH solution; reduction, anhydrous hydrazine/NH_4OH solution.

standards directly into the primary Ca standard to obtain specific Mg/Ca and Sr/Ca ratios. Each laboratory received the two standard solutions with Ca concentrations of approximately 1000 ± 3 ppm and the following elemental ratios: Standard 1, Mg/Ca = 1.856 ± 0.005 mmol mol^{-1}, Sr/Ca = 1.011 ± 0.003 mmol mol^{-1}; Standard 2, Mg/Ca = 3.682 ± 0.005 mmol mol^{-1}, Sr/Ca = 2.025 ± 0.003 mmol mol^{-1}. The laboratories were instructed to dilute the solutions to the appropriate working concentration and analyze them multiple times as regular samples against their working standards.

2.2. Foraminifera

[7] Foraminifera samples used for this study were picked from surface sediments from different geographical regions. The samples included five species, some are often used for reconstructing upper seawater temperatures: *Globigerinoides sacculifer* (355–425 μm; w/out sac), *G. ruber* (212–300 μm; white variety), *Pulleniatina obliquiloculata* (425–500 μm), *Globigerina bulloides* (300–355 μm), and *Orbulina universa* (425–500 μm). The size fractions are those typically used for Mg/Ca and Sr/Ca measurements.

Region*	Core	Core depth meter	Sample depth, cm	Species	Size fraction, μm	Sample code
LBB	OC205 BC60	1312	10.5	G. ruber	212-300	f
OJP	MW97-20 MC-9E	2015	0.5	G. ruber	212-300	k
OJP	MW91-1 BC56	4401	1	G. ruber	355-425	r
OJP	MW91-1 BC56	4401	1	G. sacculifer	355-425	b
OJP	MW91-1 BC56	4401	1	G. sacculifer	355-425	d
LBB	OC205 BC60	1312	10.5	G. sacculifer	355-425	c
LBB	OC205 BC60	1312	10.5	G. sacculifer	355-425	o
OJP	MW97-20 MC-20D	2958	0.5	G. sacculifer	355-425	l
OJP	MW97-20 MC-9E	2015	0.5	G. sacculifer	355-425	m
OJP	MW97-20 MC-20D	2958	0.5	P. obliquiloqulata	425-500	j
OJP	MW97-20 MC-9E	2015	0.5	P. obliquiloqulata	425-500	g
OJP	MW97-20 MC-9E	2015	0.5	P. obliquiloqulata	425-500	n
LBB	OC205 BC60	1312	10.5	P. obliquiloqulata	425-500	h
LBB	OC205 BC60	1312	10.5	P. obliquiloqulata	425-500	q
N. Atlantic	M39059-2	1605	1	G. bulloides	300-355	e
N. Atlantic	M39059-2	1605	1	G. bulloides	300-355	p
S. Pacific	MD972 108BX	2140	1	G. bulloides	300-355	i
LBB	OC205 BC60	1312	10.5	O. universa	425-500	a

Figure 1. Sample information. Re-picks of the same samples are color-coded (except for black). Asterisks indicate LBB, Little Bahama Bank; OJP, Ontong Java Plateau.

[8] Cores selected for this study represent regions with significantly different surface water temperatures (Figure 1). The cores cover a wide water depth range and thus represent varying preservation states. In general, the selected sites represent four regions: (1) Western subtropical Atlantic samples are from a relatively shallow core (1312 m) from Little Bahama Banks (LBB) characterized by high degree of foraminiferal shell preservation reflecting the calcite oversaturation state of the bottom water [*Bainbridge*, 1981]. Overlying mean annual SST = 26.7°C [*Levitus and Boyer*, 1994]. (2) Western Equatorial Pacific samples are from three cores on Ontong Java Plateau (OJP) covering a water depth range between ∼2 and 4.5 km. This depth transect is characterized by an increasing degree of foraminiferal shell dissolution reflecting the calcite undersaturation state of the bottom water [*Brown and Elderfield*, 1996; *Rosenthal et al.*, 2000]. Overlying mean annual SST = 29.2°C. (3) North Atlantic samples are from a box-core recovered at relatively shallow depth (1605 m) off the coast of Portugal.

Overlying mean annual SST = 17°C. (4) South Pacific samples are from a box-core recovered from the Tasman Plateau (2140 m) south of New Zealand. Overlying mean annual SST = 13°C.

[9] Each individual sample (codes a through r, Figure 1), weighing about 30 mg, was divided into six subsamples of about 5 mg each (marked from 1 to 6; Figure 2). The number of shells in each subsample varied depending on the individual species shell weight (e.g., about 350 shells of *G. ruber* and 150 shells of *G. sacculifer*). To minimize differences due to the natural variability within each subsample, we followed the "crush & split" protocol proposed by [*Boyle*, 1995]: each subsample weighing 5 mg was gently crushed and the broken parts were mixed thoroughly in order to "homogenize" the subsample. Subsequently, the homogenized subsample was split into 8 aliquots using a razor blade. The crushed splits were placed in 8 acid leached 0.5 mL safe-lock Eppendorf® vials. Four of the vials were marked with one random ID (e.g.,
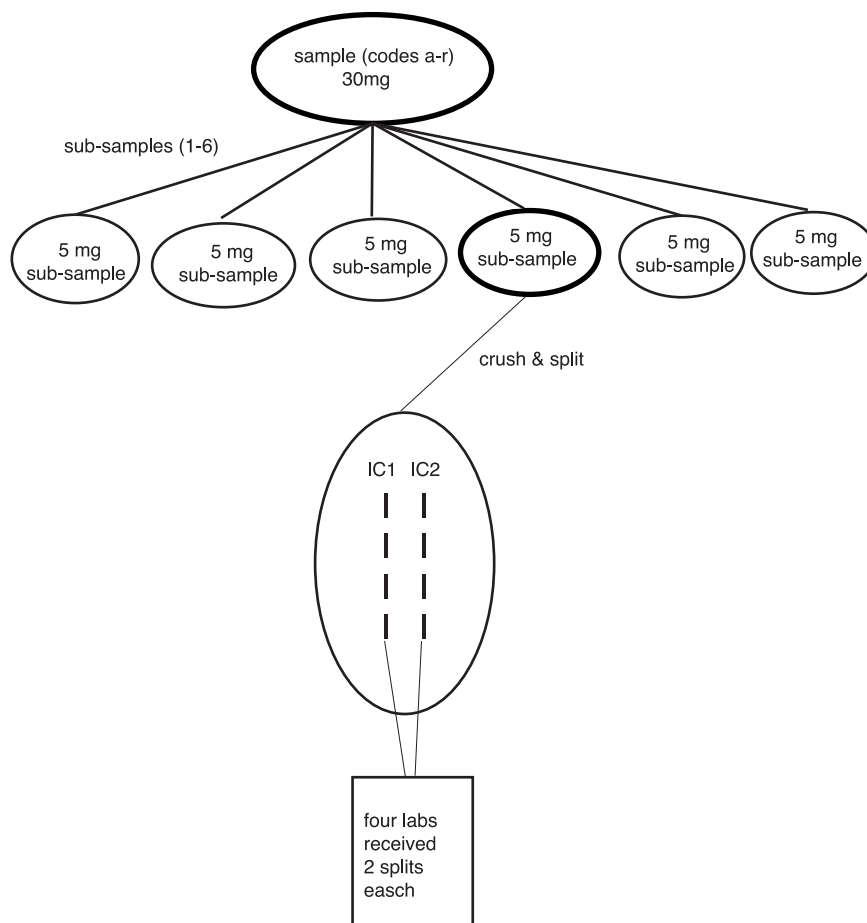
Figure 2. Schematic depiction of the experimental design of this study.

IC22) and the other four vials were marked with another random ID (e.g., IC56). Four laboratories received two splits (e.g., IC22 and IC56) of the same subsample, which are considered true replicates of the same foraminifera shells. Other laboratories received shells from different subsamples, considered here as re-picks of the same sample. The ID numbers were random so sample details were not known to the participating laboratories and thus each vial was treated as an independent sample. A total of 32–34 vials, each containing ~500 μg sample were sent to each laboratory. The laboratories were instructed to follow exactly the same methods of sample cleaning and analytical protocols that they are using in their routine work.

2.3. Statistical Evaluation

[10] In evaluating the results of the interlaboratory comparison study, we generally follow the proce-

dures adopted by previous calibrations studies [*Rosell-Melé et al.*, 2001]. For each individual laboratory we report the means of replicate measurements and, when comparing among laboratories, calculate the mean, median and mode of all measurements. The dispersion of the data is reported in terms of standard deviation (SD 1σ) and relative standard deviation (%RSD, which is equivalent to the coefficient of variation in percentage units). In addition, using analysis of variance (ANOVA), the intralaboratory variance (repeatability, r) and the interlaboratory variance (reproducibility, R) are estimated [*Nilsson et al.*, 1997]. The repeatability is an estimate of the reliability of the procedures used by each individual laboratory. The reproducibility is a measure of comparability of results obtained by different laboratories. The repeatability and reproducibility are calculated from the intralaboratory variance (S_r)

Table 2a. Statistical Results of the Analysis of Two Spiked Solutions for Each Individual Laboratory

Standard	Lab ID	Mg/Ca	SD	Sr/Ca	SD
S1	108[a]	1.729	0.020	1.022	0.005
S2	108[a]	3.663	0.023	2.052	0.005
S1	129	1.791	0.007	1.008	0.009
S2	129	3.573	0.001	1.990	0.023
S1	140	1.861	0.008	1.018	0.006
S2	140	3.761	0.072	2.013	0.007
S1	142	1.972	0.011	1.021	0.002
S2	142	3.848	0.011	2.024	0.003
S1	171	1.898		1.045	
S2	171	3.766		2.097	
S1	558	1.847	0.001	1.012	0.000
S2	558	3.638	0.019	1.999	0.002
S1	578	1.843	0.012	1.009	0.001
S2	578	3.594	0.023	1.995	0.002
S1	628	1.870	0.004	1.028	0.002
S2	628	3.658	0.013	2.041	0.004
S1	856	1.985		0.952	
S2	856	3.663		1.812	
S1	867	1.858	0.008	1.027	0.001
S2	867	3.697	0.009	2.070	0.006
S1	881	1.904	0.066	1.023	0.003
S2	881	3.794	0.049	2.045	0.007
S1	935	1.765	0.024		
S2	935	3.673	0.029		

[a] Results from laboratory #108 were added in proofs. Therefore these results are not included in the figures and the statistical calculations presented in this paper.

and interlaboratory variance (S_R), respectively, at the 95% confidence level where repeatability, r = 2.8 S_r and reproducibility, R = 2.8 S_R.

[11] Thus converted to degrees centigrade, the reproducibility is a measure of the closeness of agreement between two temperature estimates obtained by different laboratories. In other words, temperature estimates that differ from each other by less than the reproducibility of the method are considered the same within the 95% confidence level. It is important, however, to remember that these statistical parameters are valid only when the data are normally distributed.

[12] The current study was designed to assess several parameters that might affect the comparability of Mg/Ca measurements obtained by different laboratories: (1) the degree of intralaborator and interlaboratory analytical precision determined from repeated analysis of spiked standard solutions, (2) repeatability of each laboratory when analyzing splits from the same subsample of foram-

iniferal shells, (3) reproducibility among laboratories when analyzing splits from the same subsample of shells, (4) the contribution of natural variability to the variability in Mg/Ca and Sr/Ca measurements as revealed from the analysis of different subsamples of the same sample.

3. Results and Discussion

3.1. Standards

[13] The analysis of the two spiked solutions is used to evaluate the precision and accuracy of instrumental procedures used by different laboratories, when no sample manipulation is required. The results highlight a few important points:

[14] 1. The intralaboratory analytical precision (%RSD) for Mg/Ca and Sr/Ca measurements is generally better than 0.50% (Tables 2a and 2b). Results from solution analysis suggest that the short-term instrumental precision is similar for both ICP-MS and ICP-OES instruments. Long-term precision cannot be assessed from the current results.

[15] 2. The Mg/Ca data of both spiked solutions show bi-modal distribution (Figure 3). This is

Table 2b. Statistical Results of the Analysis of Two Spiked Solutions

Parameter	Standard 1		Standard 2	
	Mg/Ca	Sr/Ca	Mg/Ca	Sr/Ca
Expected value	1.856	1.011	3.682	2.025
Mean	1.879	1.019	3.744	2.018
Mean[a]	1.867	1.020	3.697	2.023
Median	1.866	1.021	3.731	2.018
Mode	1.868	1.025	3.690	2.052
Intralaboratory SD	0.009	0.003	0.019	0.008
Intralaboratory%RSD	0.5%	0.3%	0.5%	0.4%
Interlaboratory SD	0.063	0.012	0.091	0.037
Interlaboratory%RSD	3.4%	1.2%	2.4%	1.8%
Interlaboratory SD[a]	0.024	0.007	0.062	0.022
Interlaboratory%RSD[a]	1.3%	0.7%	1.7%	1.0%
S_r	0.010	0.004	0.021	0.009
Repeatability r	0.028	0.011	0.059	0.025
RSDr%	1.51%	1.10%	1.58%	1.25%
S_R[a]	0.027	0.008	0.074	0.025
Reproducibilty R*	0.077	0.023	0.208	0.070
RSD$_R$%*	4.1%	2.3%	5.6%	3.5%

[a] Excluding all the following data: Mg/Ca standard 1, all results from laboratories 129,142, 935 and one high Mg/Ca datum from laboratory 881; Mg/Ca standard 2, all results from laboratories 129,142, 935 and the high values from laboratory 140; Sr/Ca both standards, data from laboratory 856. n.d., not determined.
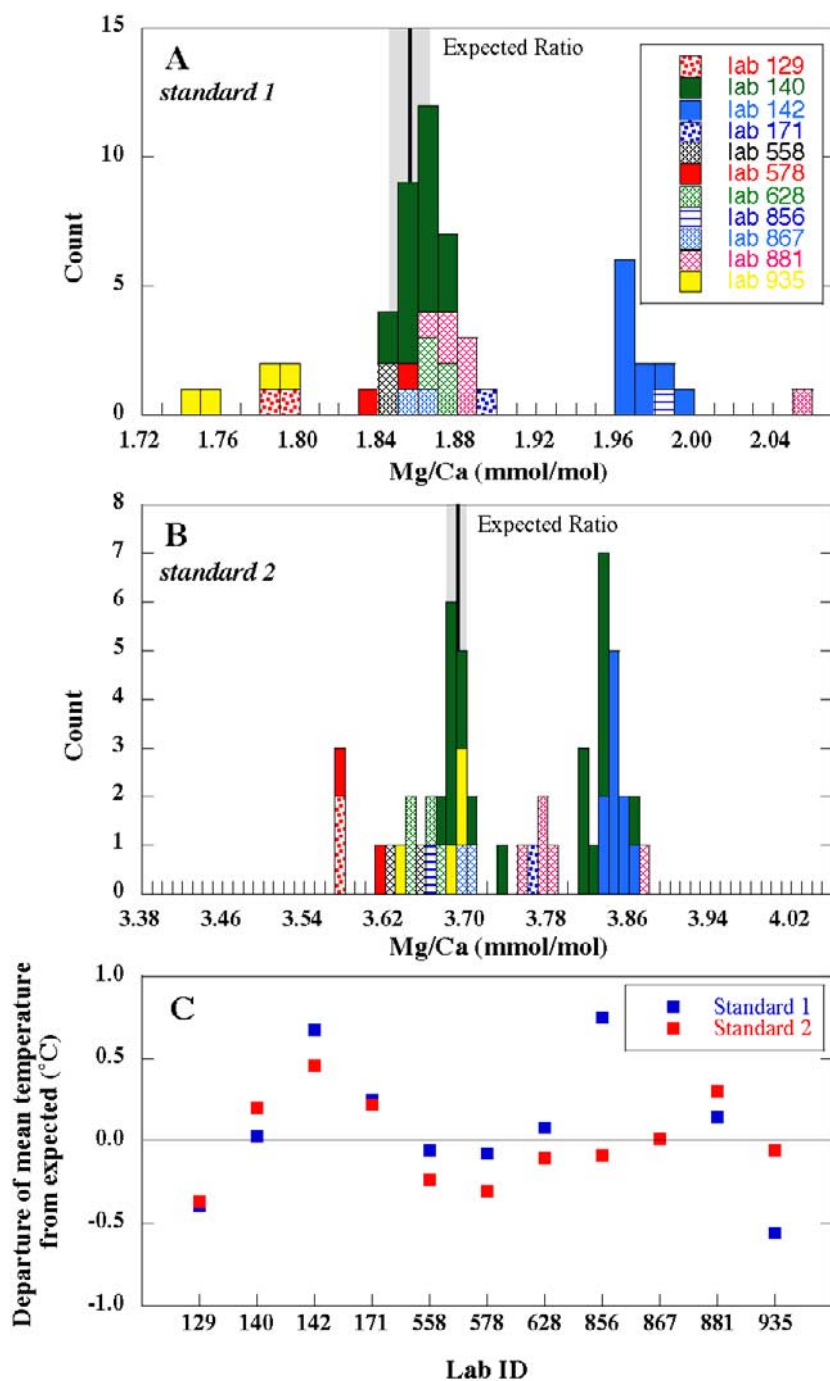
Figure 3. (a, b) Histograms of individual Mg/Ca measurements of spiked solutions 1 and 2, respectively, obtained by each laboratory. (c) The difference between the mean of Mg/Ca measurements and the expected Mg/Ca in terms of temperature (°C), calculated from data in Table 2a, using the calibration: Mg/Ca = 0.38Exp(0.090T) where T is the seawater temperature in °C [*Anand et al.*, 2003]. Note that the Mg/Ca scale and bin size of both Figures 3a and 3b represent the same relative change (±9% and ±0.5%, respectively) with respect to the mean value in both cases. Temperature estimates are based on the calibration of *Anand et al.* [2003].

especially evident for standard 2. In both cases the main mode is close to the expected value, whereas the second mode is higher than the expected value. The bi-modality is not seen in the distribution of the Sr/Ca results (Figure 4). Each of the spiked solutions was sent for analysis in duplicate, 500 μl vials. All the aliquots, however, came from the same batch solutions, and therefore it is unlikely that the bias is due to error in the preparation of the solutions. This is supported by the fact that there is no such bias in the Sr/Ca data. There are two other possibilities to explain the bimodality. First, Mg contamination during handling of one of the spiked solutions might explain the bi-modal results reported by laboratory 140 for standard 2. It is unlikely, however, that contamination can account for the bias in the results reported by other laboratories (e.g., 142). The second possibility is that the offset is due to inaccuracies in the preparation of working standards by some laboratories, poor matching between the calcium concentrations of the working standards and that of the spiked solutions (i.e., matrix effects) or the use of inappropriate calibration methods. For example, it has been reported by laboratory 935 that they did not dilute the spiked solution and therefore, the Ca concentration of these solutions was substantially higher than that of the working standards. Such matrix mismatch can account for the fact that measurements of standard 1 by laboratory 935 gave relatively low Mg/Ca ratios [*Lear et al.*, 2002]. We suspect that this is also the case with laboratory 129. Another example are the Mg/Ca results from laboratory 856. Their reported value for S1 is 7% higher than the expected value, whereas S2 is consistent with the expected value. Unfortunately this laboratory measured each standard only once so it is difficult to ascertain the significance of these values. Taken at face value, however, it suggests a problem with their calibration curve, leading to non-linear offsets. These latter biases can lead to significant offsets in absolute temperature estimates (Figure 3c). That results from foraminiferal analyses obtained by this laboratory are often higher than those obtained by other laboratories is consistent with our suggestion that the bi-modality is due to problems with particular laboratories.

[16] Results of Sr/Ca analysis are almost normally distributed. For both solutions the modes of the measured data deviates (at the 95% level) from the expected values. The discrepancy is likely due to inaccuracy in the estimation of the expected values. Values obtained by laboratory 856 are consistently lower than others, suggesting inaccuracies in their working standards, which lead to significant offsets in foraminiferal analyses as shown below.

[17] 3. The interlaboratory analytical precision (%RSD) is 3.4% and 2.4% for Mg/Ca analysis of solutions 1 and 2, respectively. Likewise, the interlaboratory Sr/Ca precision is 1.2% and 1.8% for the same solutions (Tables 2a and 2b). These RSD's are significantly larger than the intralaboratory precision. We recalculated the mean and standard deviation without outliers. Thus we excluded the following data points: standard 1, all results from laboratories 129,142, 935 and one high Mg/Ca datum from laboratory 881; standard 2, all results from laboratories 129,142, 935 (note that although results from this laboratory fall with all the other data they did not match the solutions matrix with their standards and we decided not to include their data) and the high values from laboratory 140. On the basis of this edited data set, we estimate the Mg/Ca interlaboratory precision on the analysis of the two solutions to about 1.3% and 1.7% for standards 1 and 2, respectively. Likewise, the Sr/Ca interlaboratory precision is ≤1% for both standard solutions (excluding data from laboratory 856 for both standard solutions). The repeatability (r) of the Mg/Ca results for both solutions are on average about 1.5%. The reproducibility (R) is about 4% and 6% for solutions 1 and 2, respectively (Tables 2a and 2b). Converting these values to degrees centigrade, using the recently published calibration equation: Mg/Ca = 0.38Exp(0.090T) where T is the seawater temperature in °C [*Anand et al.*, 2003], we calculate repeatability and reproducibility of ±0.2°C and ±0.5°C, respectively, for the analysis of solutions only.

3.2. Foraminifera

[18] Averages of duplicate measurements of Mg/Ca of individual species from various sites are given in Figures 5 and 6. The Sr/Ca data are given in
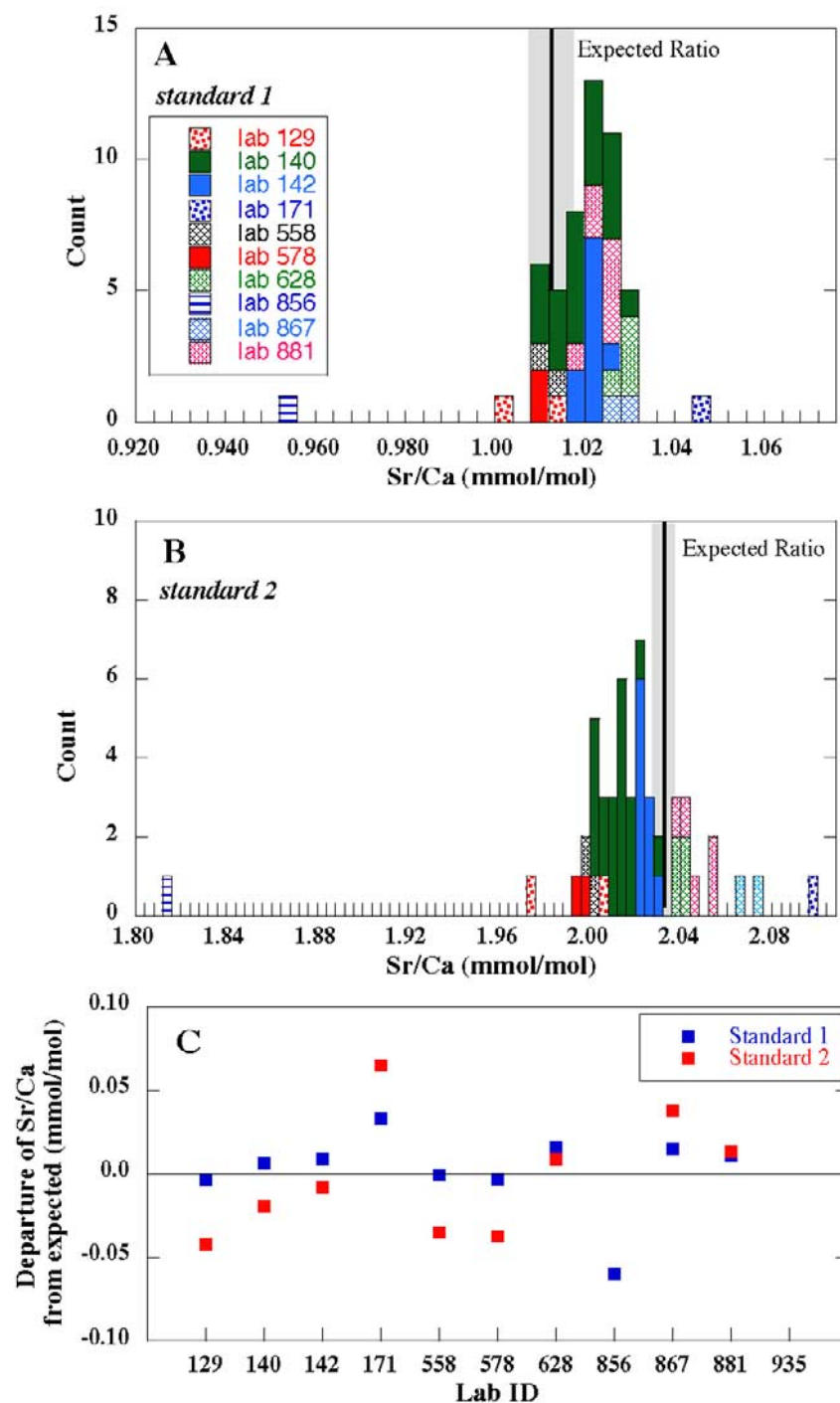
Figure 4. (a, b) Histograms of individual Sr/Ca measurements of spiked solutions 1 and 2, respectively, obtained by each laboratory. (c) The difference between the mean of the Sr/Ca measurements and the expected Sr/Ca in mmol mol^{-1} (Table 2a). Note that the Sr/Ca scale and bin size in both Figures 4a and 4b represent the same relative change ($\pm 15\%$ and $\pm 0.9\%$, respectively) with respect to the mean value in both cases. Also, note that the anomaly of laboratory #856 is larger than -0.10 mmol mol^{-1} and therefore is not shown in Figure 4c.

Samp ID	Laboratory ID													
	108	129*	140	142	171	213	360	558	578	628	856	867	881	935
a	8.34	9.00	9.00	9.15	7.62	8.01	8.02	8.61	8.55	8.33	8.24	7.55	8.25	9.53
b	3.59	5.70	3.69	3.69	3.36	3.21	3.25	3.68	3.58	3.60	3.59	3.52	3.71	3.62
c	4.28	9.87	4.43	4.80					3.99	4.71			4.71	4.46
d	3.55	7.31	5.19	4.09	3.37	3.32	3.15	3.86	3.46	3.56	3.48	3.35	3.80	3.86
e			2.99		2.34	2.40	2.61	3.06	3.32	3.25	2.93	2.75	2.95	
f	4.83	6.38	5.72	5.55	4.52	4.80	4.27	4.86	4.74	5.16	5.24	4.81	5.28	5.42
g	2.67	3.93	2.93	2.94	2.46	2.51	2.53	2.78	2.79	2.86	3.01	2.65	2.69	2.98
h	3.36	9.72	5.54	3.86	3.09	3.05	3.33	3.09	3.55	3.26	3.18	3.36	3.51	2.97
i	1.92	2.07	1.85	1.82	1.57	1.63	1.51	1.98	1.62	1.83	1.85	1.74	1.79	1.85
j	2.67	4.99	3.03	2.45	2.35	2.40	2.07	2.67	2.54	2.70	2.51	2.69	2.75	3.04
k	4.52	5.90	4.75	5.14	4.40	4.16	3.98	4.80	4.34	4.62	4.58	4.49	4.90	4.48
l		7.21	4.32	4.58	4.25	3.92	4.00	4.13	3.93	4.50	4.16	3.97	4.29	4.60
m	4.17	5.49	4.18	4.38	4.25	3.93	3.95	4.14	4.07	4.23	4.49	4.10	4.12	4.28
n	2.49	3.40	3.28	2.79	2.42	2.62	2.45	2.86	2.63	2.50	2.74	2.79	2.79	3.06
o	4.24	6.28	4.48		4.10	4.22	3.86	4.45	4.01	4.54	4.29	4.41	4.72	4.69
p	3.12	7.37	2.87	3.16	2.24	2.85	2.49	4.29	3.10	3.08	2.99	2.76	2.94	3.18
q	3.45	5.04	3.50	3.68	3.36	2.99	3.28	3.40	3.00	3.48	3.51	3.16	3.56	3.77
r	4.09	7.87	4.68	5.81	3.77	3.97	3.61	4.43	4.12	4.23	3.96	4.15	4.23	4.44

Figure 5. Foraminiferal Mg/Ca data. Crush and split subsamples are marked by the same color. Each datum represents the mean of duplicate analyses except for italicized values for which one of the duplicates was considered an outlier (strike-through values are not considered as the two duplicates are considered outliers). Asterisks indicate that all data from this laboratory were not considered in the statistical analysis (see text).

Figures 7 and 8. The complete data set from the analysis of the solutions and foraminifera samples is available in as downloadable appendix at the end of the paper. The results of Mg/Ca and Sr/Ca analysis in foraminifera by different laboratories show large scatter that is significantly greater than the precision of replicate measurements of the same sample by each individual laboratory. The interlaboratory variability is also substantially larger than that obtained from the analysis of spiked solutions. To better quantify the errors associated with these analyses we need to understand the different sources of error. Errors can be divided into systematic and random types. Systematic offsets may be due to the application of different cleaning protocols, contamination from reagents used during the cleaning stage and the analysis (blank problem). As discussed previously, it is possible that some of the offsets are due to the use of inaccurate working standards. Random errors are often caused by contamination from adhering sediments or insufficient matrix matching between samples and standards. Random errors are typically recognized as outliers and are characterized by poor repeatability of replicate analysis of the same sample. At first glance the scatter in the data is quite large (Figures 6 and 8). To better quantify the comparability of Mg/Ca and Sr/Ca measurements by different laboratories we have filtered out data that are clearly biased relative to the entire population (note, however, that all the data are given in the tables). To identify outliers, we used Youden plots in which the means of one sample measured by each laboratory are plotted against the means of another sample. Samples that are offset from the next datum point by more than 3 standard deviations (calculated for the entire sample set) are considered to be outliers. The rejected data include: (1) Both Mg/Ca and Sr/Ca data sets from laboratory 129. Data from this laboratory are substantially higher than obtained by other laboratories and exhibit relatively poor reproducibility. The bias is likely a result of the cleaning procedure used by this laboratory. This
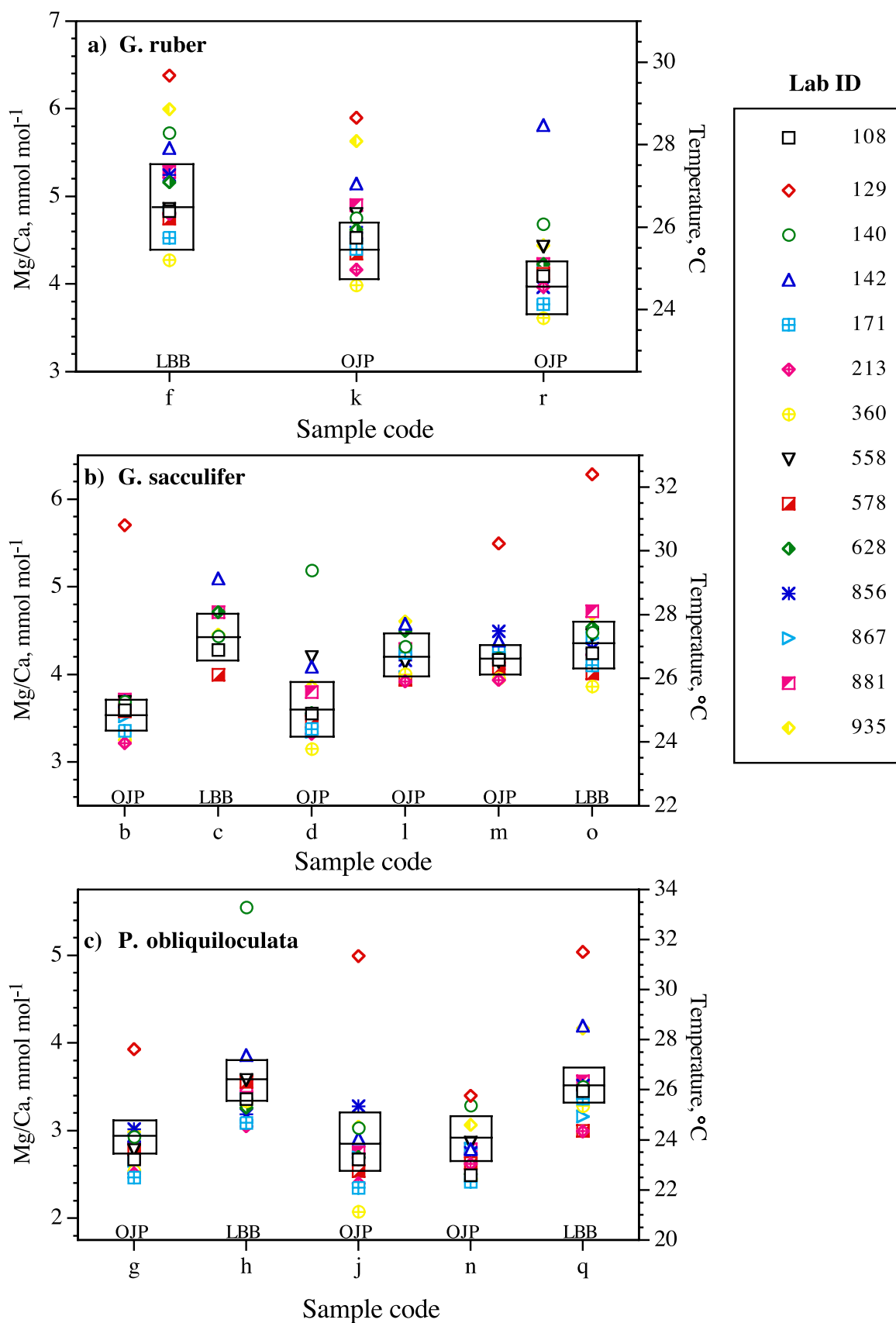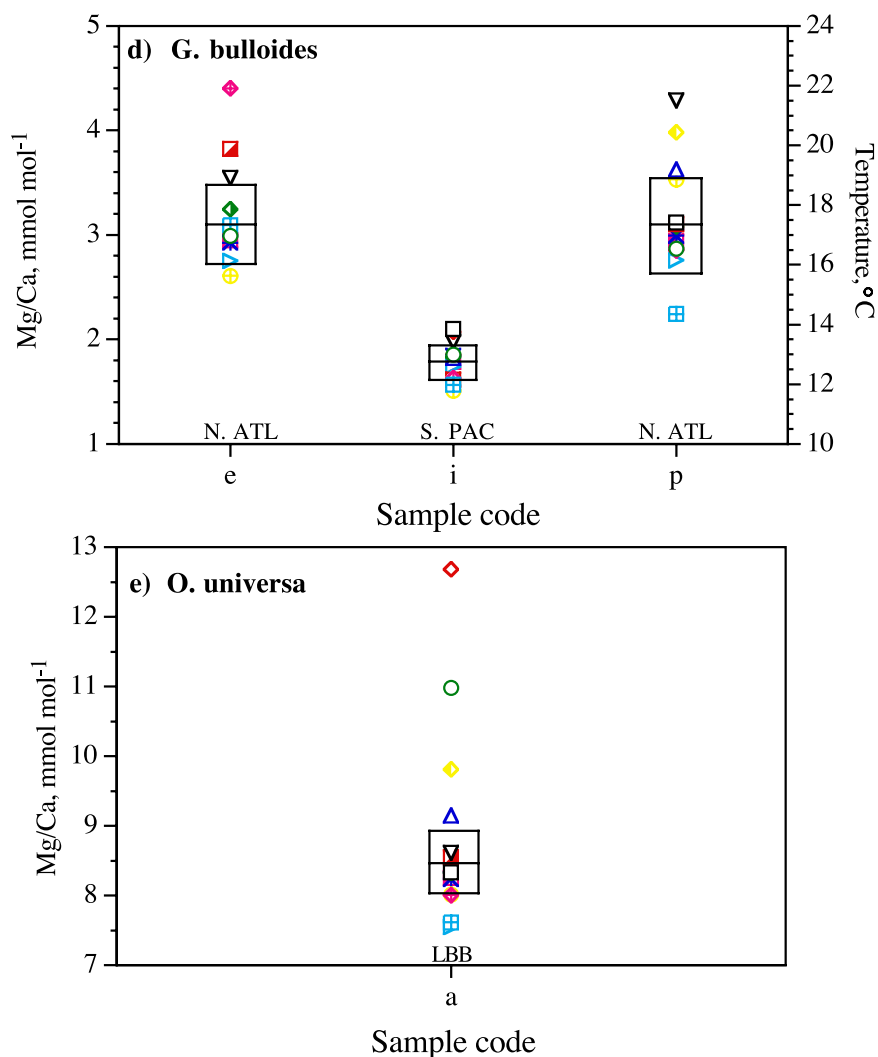
Figure 6.

Figure 6. (continued)

lab tested a simple and rapid cleaning procedure that is clearly inadequate (i.e., oxidation with 5% sodium hypochlorite, followed by multiple rinses with deionized water without ultrasonication); (2) Two Sr/Ca data sets (laboratories 360 and 856) that are significantly lower than the other data sets. The analysis of spiked solutions shows that Sr/Ca values obtained by laboratory 856 are about 10% lower than the average value, which accounts for the bias observed in the foraminifera data. Although laboratory 360 did not analyze the spiked solutions, it is likely that inaccuracies in their working standards cause the observed offset in the foraminiferal data. These offsets can be easily corrected in future work, therefore we do not include them in the statistical evaluation; (3) 28 Mg/Ca data points that are significantly higher than the other results. The outliers constitute about 6% of the Mg/Ca data set (excluding the results from laboratory 129). These high values are probably due to contamination from adhering clays [*Barker et al.*, 2003], which have high magnesium content

Figure 6. Mg/Ca results from the analysis of five species of planktonic foraminifera. Each point represents an average of the analysis of two replicates from the same subsample. Boxes represent the mean and standard deviation (1 S.D.) for each sample, excluding the outliers (see details of the calculation in the text). Temperature estimates are based on the calibration of *Anand et al.* [2003].

ID	108	129*	140	142	171	213	360*	558	578	628	856*	867	881	935
a	nd	1.82	1.47	1.42	1.48	1.37	1.32	1.33	1.37	1.40	1.28	1.42	1.36	nd
b	nd	1.40	1.36	1.37	1.39	1.35	1.31	1.35	1.36	1.38	1.27	1.38	1.39	nd
c	nd	2.00	1.40	1.41				1.38	1.40				1.44	nd
d	nd	1.46	1.39	1.37	1.39	1.35	1.31	1.36	1.35	1.38	1.27	1.36	1.40	nd
e	nd		1.34		1.43	1.35	1.29	1.33	1.36	1.35	1.25	1.35	1.36	nd
f	nd	1.64	1.50	1.49	1.49	1.45	1.40	1.44	1.45	1.47	1.36	1.46	1.49	nd
g	nd	1.40	1.41	1.41	1.42	1.39	1.37	1.39	1.38	1.42	1.33	1.40	1.44	nd
h	nd	2.07	~~1.70~~	1.53	1.57	1.53	1.43	1.49	1.52	1.54	1.41	1.59	1.50	nd
i	nd	1.38	1.35	1.36	1.40	1.35	1.30	1.33	1.35	1.37	1.25	1.35	1.37	nd
j	nd	1.41	1.36	1.39	1.39	1.39	1.33	1.36	1.37	1.40	1.31	1.36	1.40	nd
k	nd	1.49	1.44	1.47	1.52	1.46	1.41	1.38	1.47	1.50	1.38	1.47	1.50	nd
l	nd	1.36	1.34	1.35	1.35	1.32	1.29	1.32	1.33	1.37	1.24	1.35	1.35	nd
m	nd	1.37	1.35	1.36	1.38	1.34	1.29	1.33	1.34	1.37	1.25	1.36	1.37	nd
n	nd	1.39	1.43	1.40	1.43	1.37	1.35	1.40	1.37	1.44	1.31	1.42	1.41	nd
o	nd	1.61	1.34		1.40	1.37	1.32	1.33	1.35	1.37	1.25	1.36	1.37	nd
p	nd	1.37	1.35	1.37	1.43	1.32	1.30	1.30	1.36	1.36	1.25	1.38	1.38	nd
q	nd	1.71	1.48	1.55	1.55	1.49	1.43	1.47	1.51	1.59	1.40	1.54	1.52	nd
r	nd	1.53	1.43	1.46	1.49	1.44	1.38	1.44	1.45	1.48	1.35	1.46	1.47	nd

Figure 7. Foraminiferal Sr/Ca data. Crush & split subsamples are marked by the same color. Each datum represents the mean of duplicate analyses except for italicized values for which one of the duplicates was considered an outlier (strike-through values are not considered as the two duplicates are considered outliers). Asterisks indicate that all data from this laboratory were not considered in the statistical analysis (see text); nd, not determined.

[*Emiliani*, 1955]. The fact that there is significant difference in the number of outliers among different laboratories indicates that some laboratories hold more stringent cleaning protocols and supports our argument that the outliers are likely a result of insufficient cleaning during sample preparation. The low number of Sr/Ca outliers (about 1%) is consistent with the fact that detrital clays are relatively enriched with Mg but not with Sr.

[19] In the following discussions and statistical calculations we ignore all the outliers in order to obtain a better estimate of the community performance. Mean, median, standard deviation (1σ), interlaboratory variance (S_R) and reproducibility (R) values were calculated from individual Mg/Ca and Sr/Ca measurements in foraminifera and are

given in Table 3. In most cases, the results represent 28 individual measurements. The means and standard deviations are also shown in Figures 4 and 5. The standard deviations (1σ) of Mg/Ca measurements obtained by the different laboratories range from ±0.2 to 0.6 mmol mol^{-1} (1σ), or an average of ±8% RSD (Table 3). In comparison, the Sr/Ca variability among laboratories is about ±0.02 to 0.06 mmol mol^{-1}, or an average of ±3% RSD. Using recent field calibrations [*Anand et al.*, 2003], we calculate the errors associated with temperature estimates to be between ±0.3 and 1.3°C (1σ) or better than ±1°C for all samples. Using the same temperature calibration for all the samples, we estimate an interlaboratory reproducibility of 20–30% or 2–3°C. This estimate is somewhat misleading, however. In previous studies, interla-

Figure 8. Sr/Ca results from the analysis of five species of planktonic foraminifera. Each point represents an average of the analysis of two replicates from the same subsample. Boxes represent the mean and standard deviation (1SD) for each sample, excluding the outliers (see details of the calculation in the text).
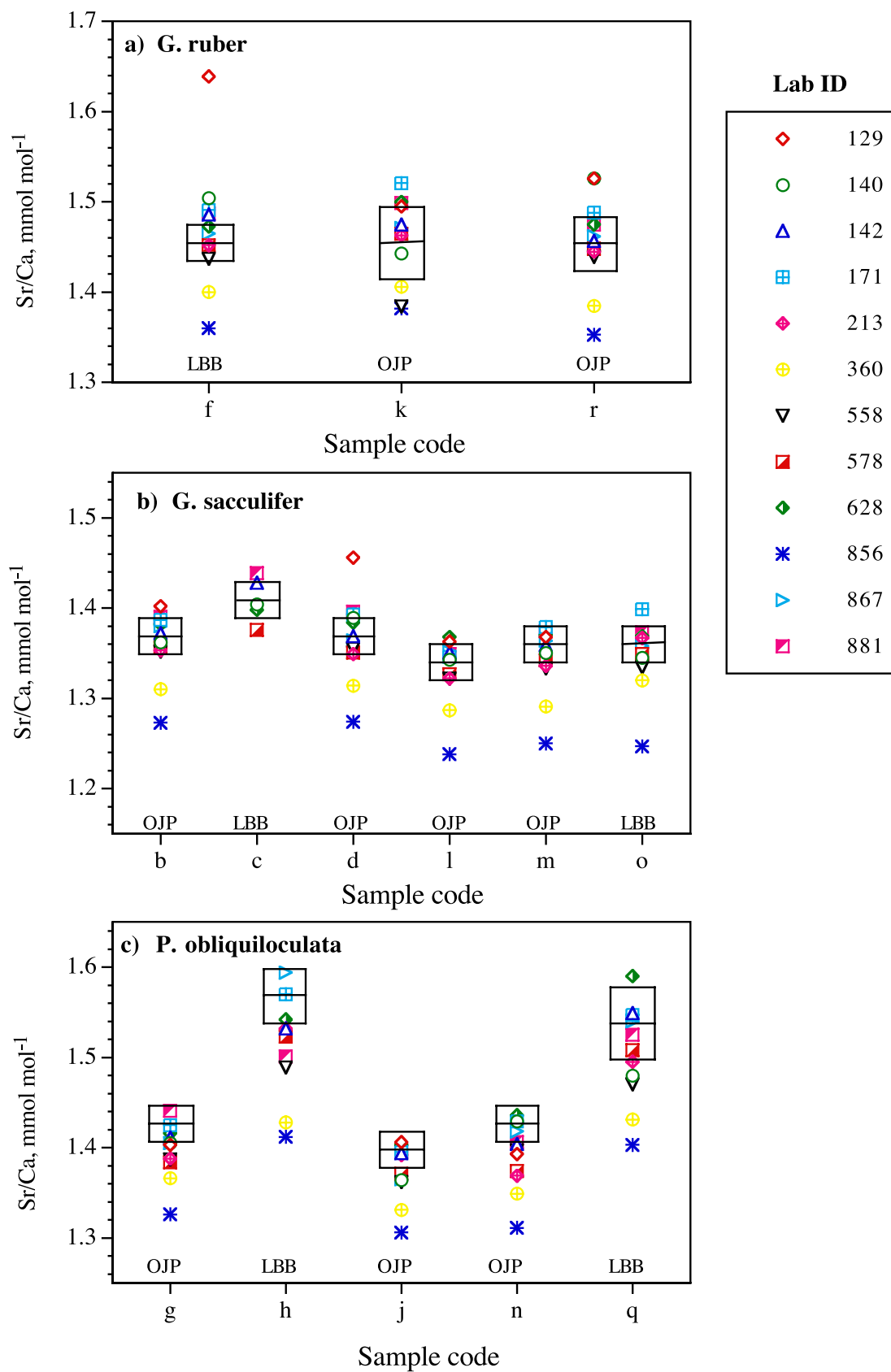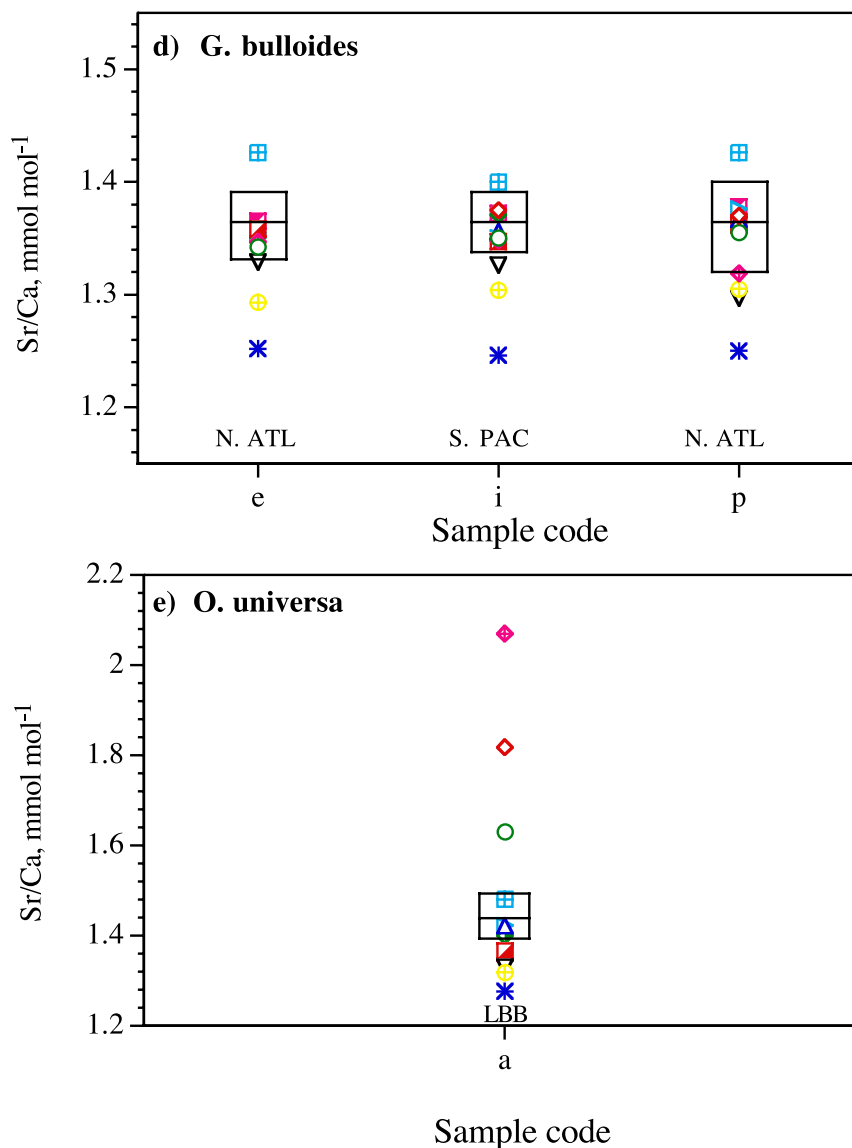
Figure 8.

Figure 8. (continued)

boratory discrepancies were often circumvented by choosing site specific calibrations that yield the correct modern surface temperature when applied to core top samples. These calibrations not only corrected for differences in foraminiferal preservation among sites (see below) but also for poor interlaboratory reproducibility. So, while the interlaboratory reproducibility of Mg/Ca measurements is on the order of 20–30%, in practice the reproducibility of Mg/Ca-derived temperature estimates is significantly better.

[20] The interlaboratory precision for the analysis of foraminiferal samples is about four times worse than that obtained on the spiked solutions. In the following sections we explore the relative contribution of several parameters to the scatter in Mg/Ca data. These include intralaboratory precision, natural variability among samples and cleaning efficacy.

3.3. Sources of Variability in Mg/Ca Analysis

3.3.1. Intralaboratory Variance

[21] To test the precision and repeatability of individual laboratories in the analysis of foraminiferal samples we compare results from the analysis of four replicates from the same samples (Table 4,

Table 3. Interlaboratory Precision (SD) and Reproducibility (R) Mg/Ca and Sr/Ca (in mmol mol^{-1}) for the Analysis of Foraminifera Samples

Sample Code	Mean	Median	SD	RSD%	S_R	R	RSD_R%
			Mg/Ca				
a	8.33	8.18	0.67	8.1%	0.81	2.27	27.2%
b	3.54	3.58	0.21	5.8%	0.25	0.69	19.4%
c	4.46	4.46	0.32	7.3%	0.42	1.18	26.4%
d	3.56	3.51	0.28	8.0%	0.35	0.97	27.2%
e	2.90	2.91	0.28	9.8%	0.36	1.01	34.8%
f	5.00	4.91	0.42	8.4%	0.51	1.43	28.6%
g	2.76	2.68	0.26	9.5%	0.32	0.89	32.2%
h	3.32	3.30	0.29	8.8%	0.36	1.01	30.2%
i	1.76	1.79	0.15	8.5%	0.18	0.51	28.8%
j	2.61	2.57	0.33	12.6%	0.37	1.04	39.6%
k	4.55	4.59	0.34	7.5%	0.41	1.15	25.2%
l	4.22	4.16	0.28	6.6%	0.34	0.94	22.4%
m	4.18	4.16	0.18	4.2%	0.21	0.59	14.2%
n	2.72	2.70	0.29	10.7%	0.34	0.95	35.0%
o	4.33	4.34	0.28	6.4%	0.34	0.94	21.8%
p	2.88	3.01	0.31	10.8%	0.39	1.08	37.4%
q	3.37	3.30	0.31	9.1%	0.37	1.04	30.8%
r	4.14	4.13	0.31	7.4%	0.37	1.04	25.0%
			Sr/Ca				
a	1.38	1.38	0.06	4.6%	0.08	0.22	16.0%
b	1.36	1.36	0.04	2.6%	0.04	0.12	8.9%
c	1.40	1.41	0.03	2.0%	0.04	0.11	8.0%
d	1.37	1.37	0.05	3.3%	0.06	0.16	11.5%
e	1.34	1.35	0.05	3.4%	0.06	0.16	11.7%
f	1.46	1.46	0.04	3.0%	0.05	0.15	10.2%
g	1.40	1.40	0.03	2.4%	0.04	0.11	8.2%
h	1.52	1.52	0.06	3.8%	0.07	.20	13.3%
i	1.34	1.35	0.04	3.0%	0.05	0.14	10.1%
j	1.37	1.37	0.04	2.7%	0.04	0.12	9.1%
k	1.46	1.47	0.05	3.3%	0.06	0.17	11.4%
l	1.33	1.34	0.04	2.7%	0.04	0.12	9.3%
m	1.34	1.35	0.04	2.9%	0.05	0.13	10.0%
n	1.39	1.41	0.04	2.9%	0.05	0.14	9.9%
o	1.35	1.36	0.04	3.1%	0.05	0.15	10.8%
p	1.35	1.36	0.05	3.5%	0.06	0.16	12.1%
q	1.50	1.50	0.06	3.8%	0.07	0.19	12.9%
r	1.44	1.45	0.04	2.9%	0.05	0.14	9.9%

Figure 9). The samples represent four pairs of re-picked samples including b&d, c&o, e&p, h&q (the pairs are color coded in Figure 1). The average intralaboratory precision on the analysis of *G. sacculifer* samples is about ±3% (RSD) and we find no significant difference between the deep Pacific (OJP) and shallow Atlantic sites. The repeatability is accordingly good, about ±1°C. The precision on the analysis of *P. obliquiloculata* from LBB and *G. bulloides* from the North Atlantic is significantly lower, about ±1.5° and 2°C, respectively. Variably high Mg/Ca ratios, found in *G. bulloides* shells from North Atlantic core

tops, have been attributed to insufficient removal of high Mg calcite layers during the cleaning process, although the nature of these has not been fully investigated [*Elderfield and Ganssen*, 2000]. Similar problems have been encountered in studies of other planktonic species from LBB, which might explain the poor repeatability of the *P. obliquiloculata* results (Y. Rosenthal, unpublished data, XXXX).

[22] The precision and repeatability of Sr/Ca results are significantly better than that of Mg/Ca (Table 4). Interestingly, Sr/Ca data from the

Table 4. Intralaboratory Precision (SD) and Repeatability (r) of Mg/Ca and Sr/Ca for Replicate Analyses of Four Pairs of Repicked Foraminifera Samples

Sample	Species	Region	SD	RSD%	Sr	r	RSD_r%
			Mg/Ca				
b&d	sacc	OJP	0.12	3.4%	0.13	0.36	10.3%
c&o	sacc	LBB	0.12	2.8%	0.14	0.39	9.1%
h&q	obliq	LBB	0.16	5.6%	0.19	0.53	18.7%
e&p	bull	N. Atl	0.27	8.0%	0.30	0.84	24.9%
			Sr/Ca				
b&d	sacc	OJP	0.010	0.7%	0.011	0.031	2.3%
c&o	sacc	LBB	0.018	1.3%	0.020	0.056	4.1%
h&q	obliq	LBB	0.010	0.7%	0.011	0.031	2.3%
e&p	bull	N. Atl	0.035	2.3%	0.039	0.109	7.2%

analysis of *G. bulloides* shells also show the largest variance. The intralaboratory relative precision (%RSD) and repeatability of Mg/Ca and Sr/Ca measurements of spiked solutions is essentially the same (Tables 2a and 2b). In contrast, the relative precision (%RSD) and repeatability of Sr/Ca analysis of foraminifera shells is 2–5 times better than that of Mg/Ca, suggesting that the difference in the results is not due to instrumental biases, but rather related to either natural variability in the samples or to the cleaning methods used for preparing samples for the analysis.

3.3.2. Natural Variability

[23] This variability is primarily caused by differences in the depth of calcification of individual shells of the same species in the water column and the mixing of shells of different ages after burial in the sediment. Histograms of all replicate measurements from each laboratory for the analysis of Mg/Ca in individual foraminifera samples are shown in Figure 10. Also, for each site we compare the mean and standard deviation obtained by all participating laboratories (excluding outliers) with the mean and standard deviation obtained from individually

crushed and split subsamples. Analyses carried out on the same subsamples are marked in Tables 4 and 5 by the same color. As shown, there are noticable differences in mean Mg/Ca ratios among re-picked samples. However, within the precision of the data (±1σ) the mean Mg/Ca ratios for individual subsample splits are, in most cases, not significantly different from each other and are statistically the same as the mean of the entire sample population. In some cases, however, the average Mg/Ca value for a subsample (i.e., crushed and homogenized) is significantly different from that obtained for other subsamples. The data also suggest that homogenizing the samples results in greater sample reproducibility than obtained on re-picked samples. These results suggest that although natural variability may sometimes be significant, it cannot account for all of the observed variability among the different laboratories. Consequently, a significant part of the variability must be related to differences in cleaning methods.

3.3.3. Analytical Difference

[24] The analysis of spiked solutions suggests that a significant part of the interlaboratory variability

Figure 9. Comparison of intralaboratory Mg/Ca standard deviations (±1σ) obtained from the analysis of replicates of each sample, calculated in terms of temperature using the calibration of *Anand et al.* [2003]. (a) Sample replicates of *G. sacculifer* from Ontong Java Plateau Core MW91-1 BC56 (4401 meters below seafloor); (b) sample replicates of *G. sacculifer* from Little Bahama Bank Core OC205 BC60 (1312 meters below seafloor); (c) sample replicates of *P. obliquiloculata* from Little Bahama Bank Core OC205 BC60 (1312 meters below seafloor); (d) sample replicates of *G. bulloides* from North Atlantic Core M39059-2 (1605 meters below seafloor). Also compared are the RSD% obtained by laboratories using a "Mg cleaning" (thin line) and "Cd cleaning" (bold line) methods.
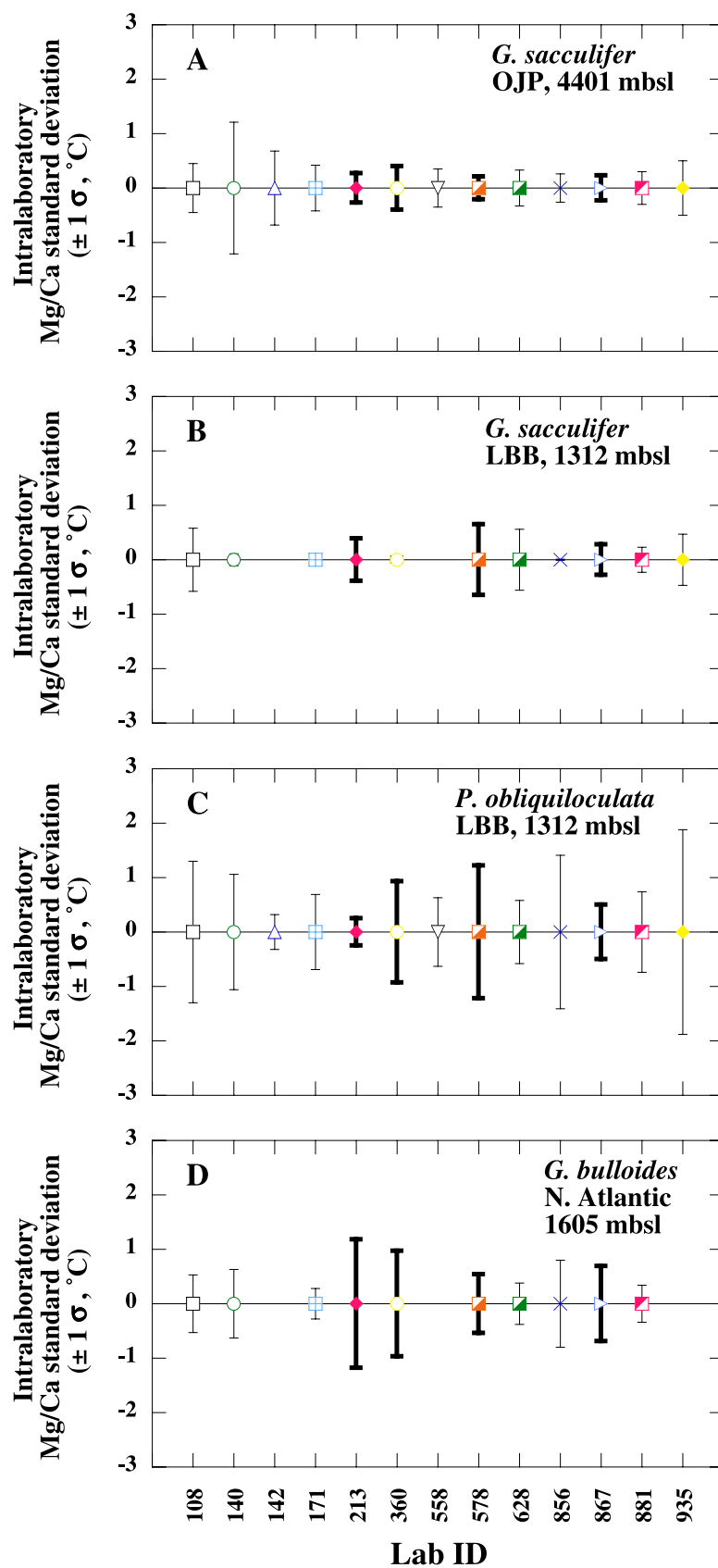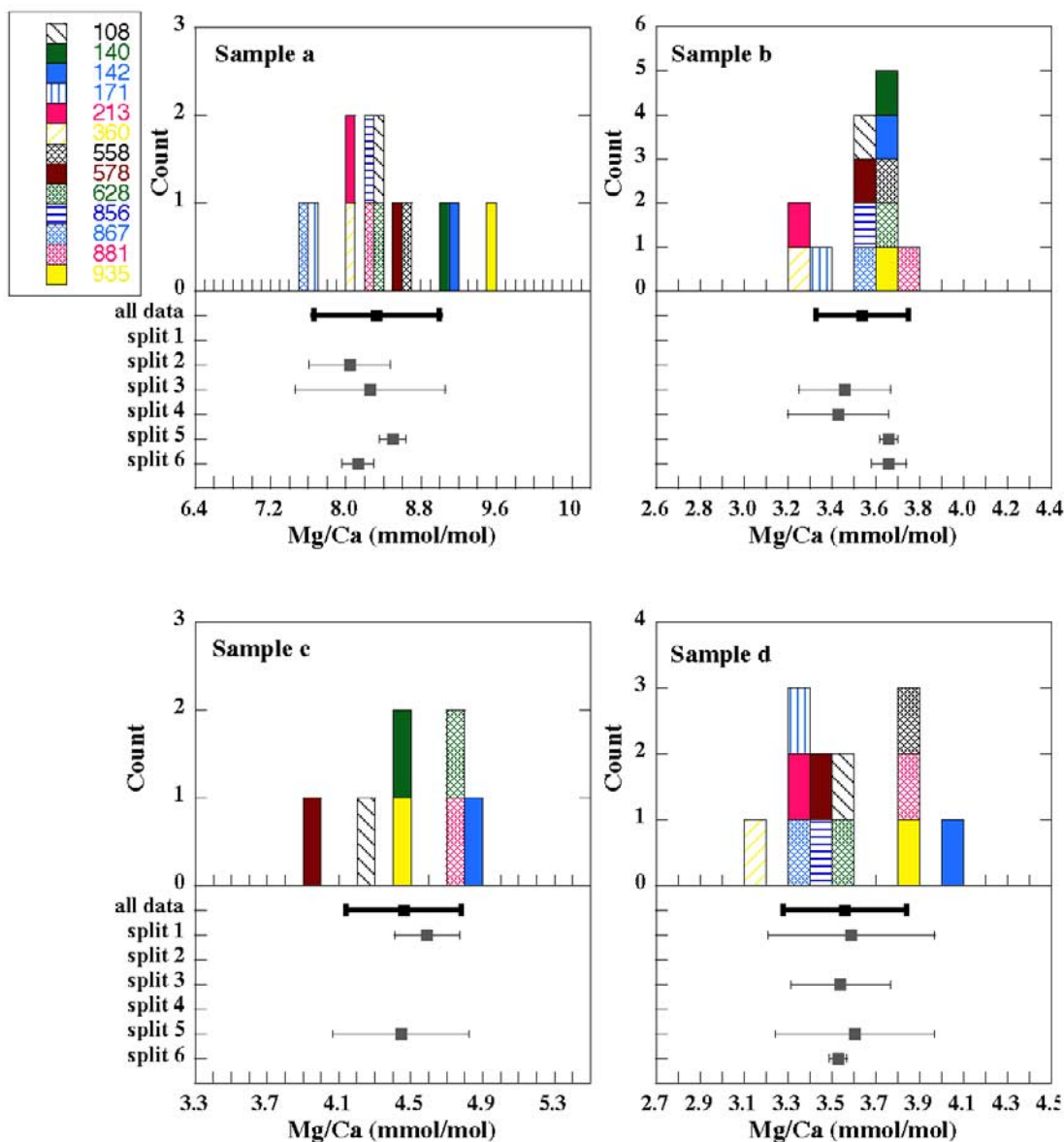
Figure 9.

Figure 10. Histograms of Mg/Ca measurements of each foraminiferal sample (a through r) obtained by different laboratories (top panel). In the bottom panel the mean Mg/Ca and standard deviation (±1σ) for all the laboratories (bold) is compared with the mean ±1σ of "crush & split" subsamples. Note that the range in each figure is 50% of the mean of all data. Bins are all 0.1 mmol/mol. The standard deviations are calculated from the mean of each duplicate sample rather than for each individual analysis, as was done in Table 3.

in both Mg/Ca and Sr/Ca is due to inconsistencies among standards prepared by different laboratories. As suggested above, these inconsistencies account for about 2% (RSD) of the total variability. An additional source of variability in Mg/Ca is the offsets caused by differences in cleaning methods applied by different laboratories. In general, the two most commonly used cleaning protocols are:

(1) "Mg cleaning" protocol, which is a short version of the [*Boyle and Keigwin*, 1985] method. This method includes multiple rinses with distilled deionised water (ddH$_2$O) and methanol in an ultrasonic bath followed by oxidation in a warm (60°C) mixture of 30% H$_2$O$_2$/0.1M NaOH solution and typically a single acid leaching step with 250 μL 0.001N HNO$_3$ ultrasonicated for 30 seconds; and
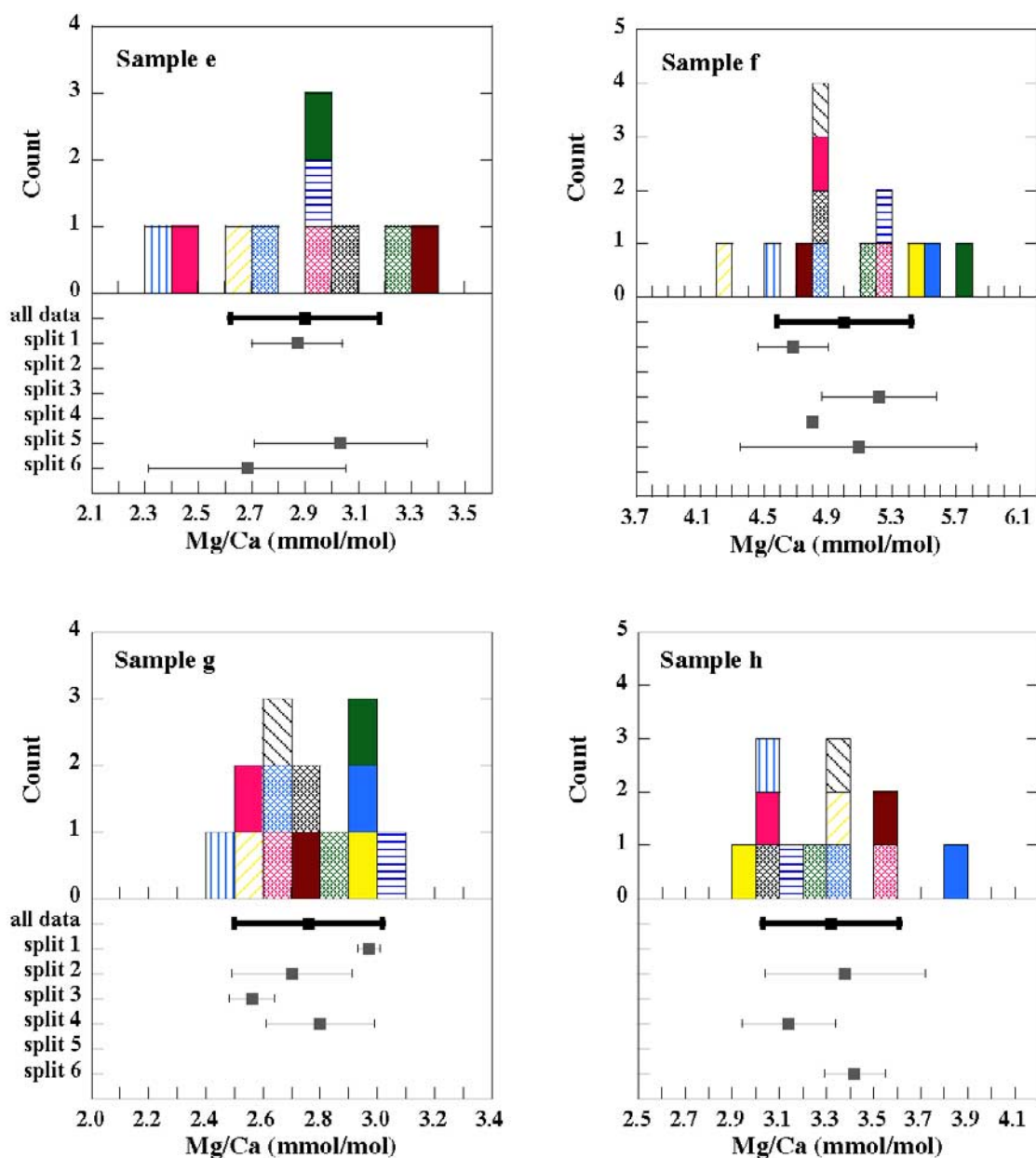
Figure 10.   (continued)

(2) ''Cd cleaning'' method, which is the full version [*Boyle and Keigwin*, 1985; *Boyle and Rosenthal*, 1996]. This protocol includes multiple rinses with ddH$_2$O and methanol in an ultrasonic bath followed by a reduction step in hot solution of anhydrous hydrazine/NH$_4$OH solution, then oxidation in a warm 30%H$_2$O$_2$/0.1M NaOH solution and finally multiple acid leachings (0 to 4 times depending on sample size), each with 250 μL

0.001N HNO$_3$ ultrasonicated for 30 seconds. The ''Cd cleaning'' cleaning protocol includes the reductive step, which was devised to remove Mn- and Fe-oxides adhering to the shells, and is used when other metal ratios (e.g., Cd/Ca, Ba/Ca and U/Ca) are measured in addition to Mg/Ca and Sr/Ca typically by laboratories using ICP-MS. It has been shown that cleaning of planktonic foraminifera leads to a progressive decrease in bulk
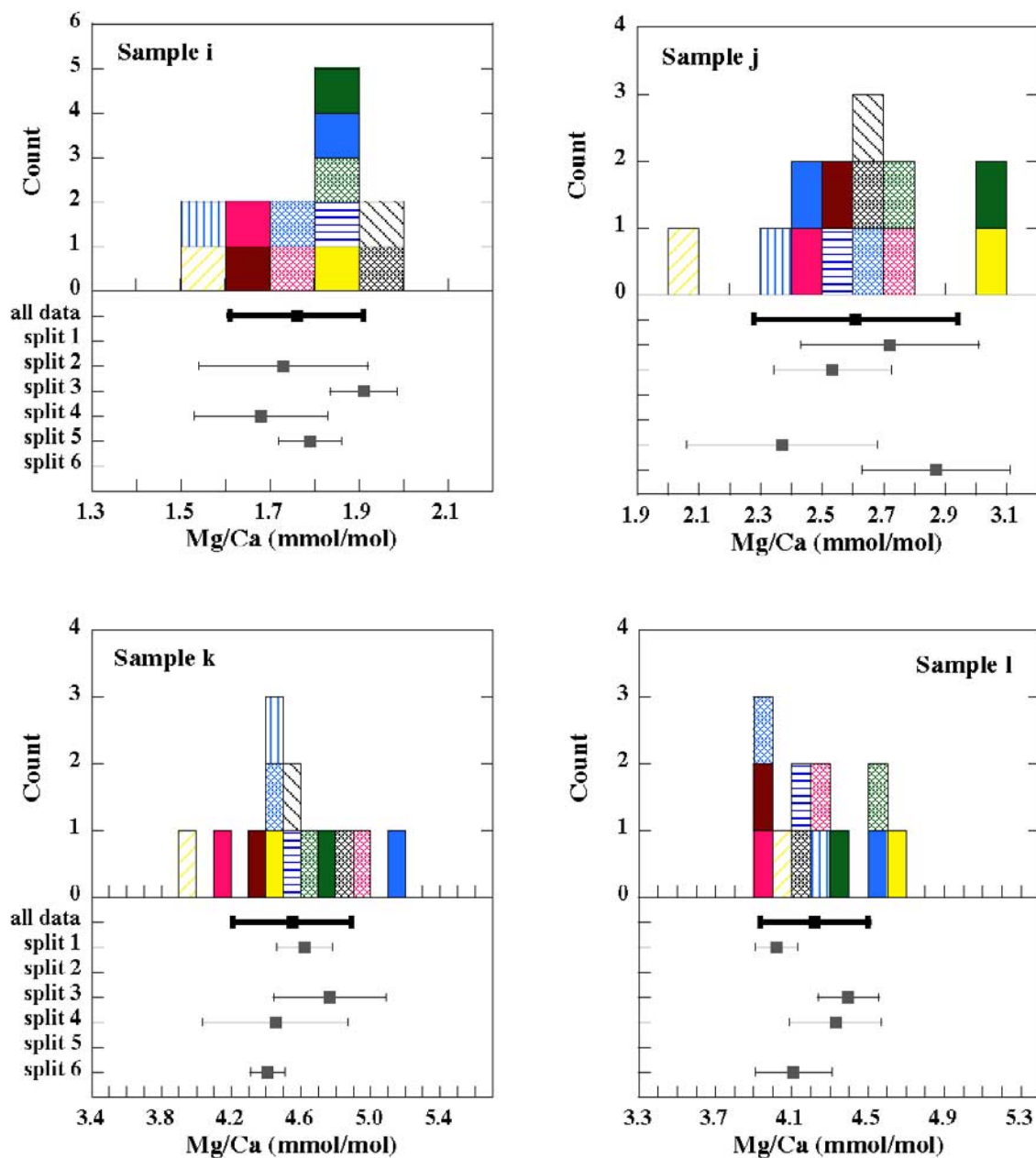
Figure 10. (continued)

shell Mg/Ca ratios; more rigorous cleaning results in lower Mg/Ca ratios [*Hastings et al.*, 1998]. In that study, the authors argued that the reduction step does not have a discernible effect on Mg/Ca and concluded that samples cleaned by either the "Mg method" or "Cd method" should yield statistically similar ratios. However, other experiments on the effects of cleaning on metal ratios in

benthic foraminifera suggested that the addition of reductive step might lead, in some samples, to a significant decrease in Mg/Ca [*Martin and Lea*, 2002]. These authors suggest that the lowering of Mg/Ca values following oxidation and reduction appears to reflect the removal of contaminant Mg associated with remnant organic matter and adsorbed phases. However, given that only one
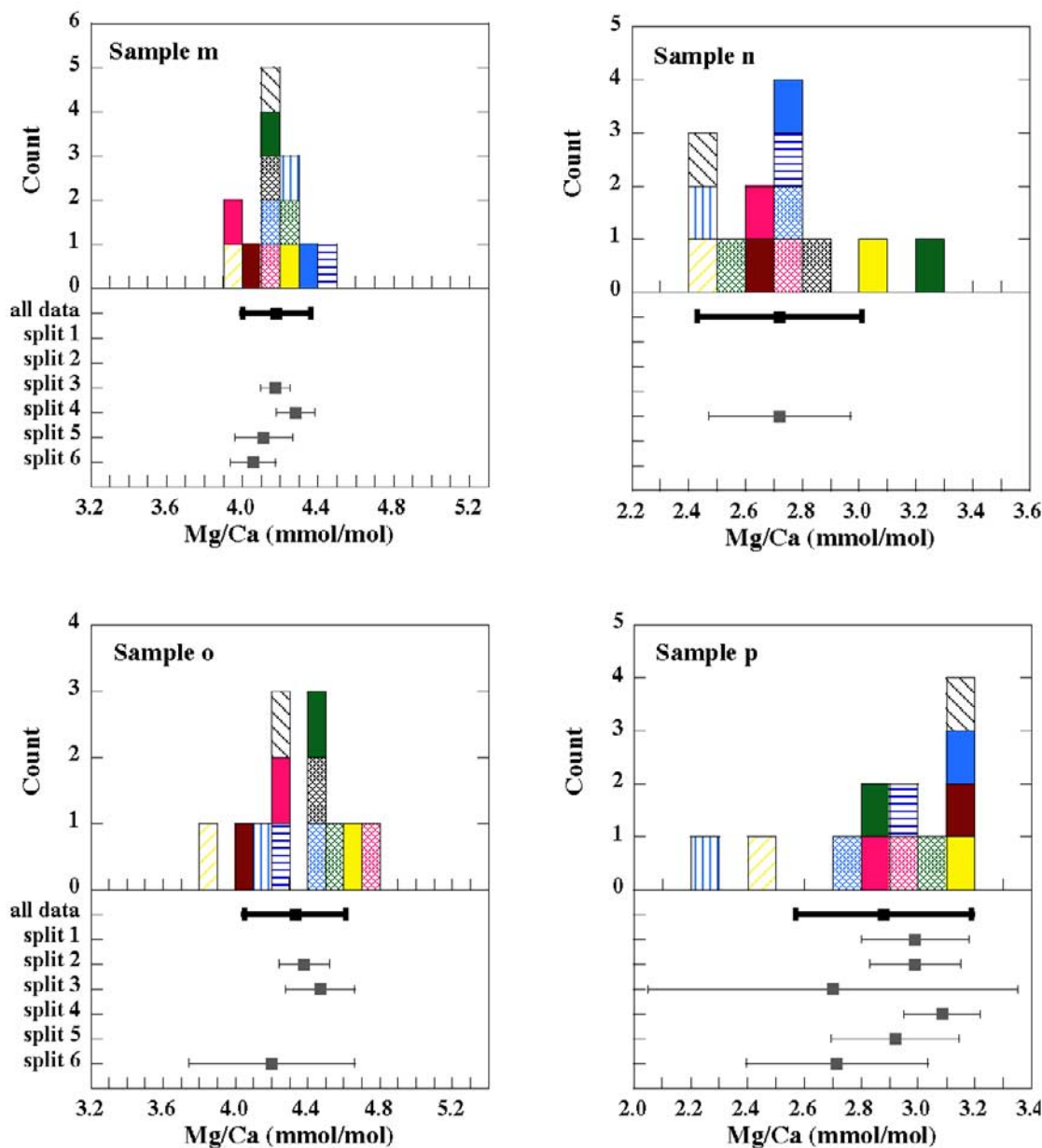
Figure 10. (continued)

out of the two samples showed a decrease in Mg/Ca following the reductive step, it remains debatable whether the difference between samples cleaned with and without reductive treatment is significant. The current study provides a sufficiently large data set to rigorously address this issue.

[25] To test the consistency between the two cleaning methods we compare Mg/Ca results obtained for all species from laboratories using

the "Mg cleaning" protocol with those using the "Cd cleaning" method. We exclude from this analysis the data from *O. universa* because of the relatively high Mg/Ca ratios characteristic of this species (and their large scatter); including these high mean and SD values tends to heavily bias the statistical analysis. Also excluded from this discussion are the results from laboratories whose analysis of spiked solutions showed poor consistency with other data sets. The comparison
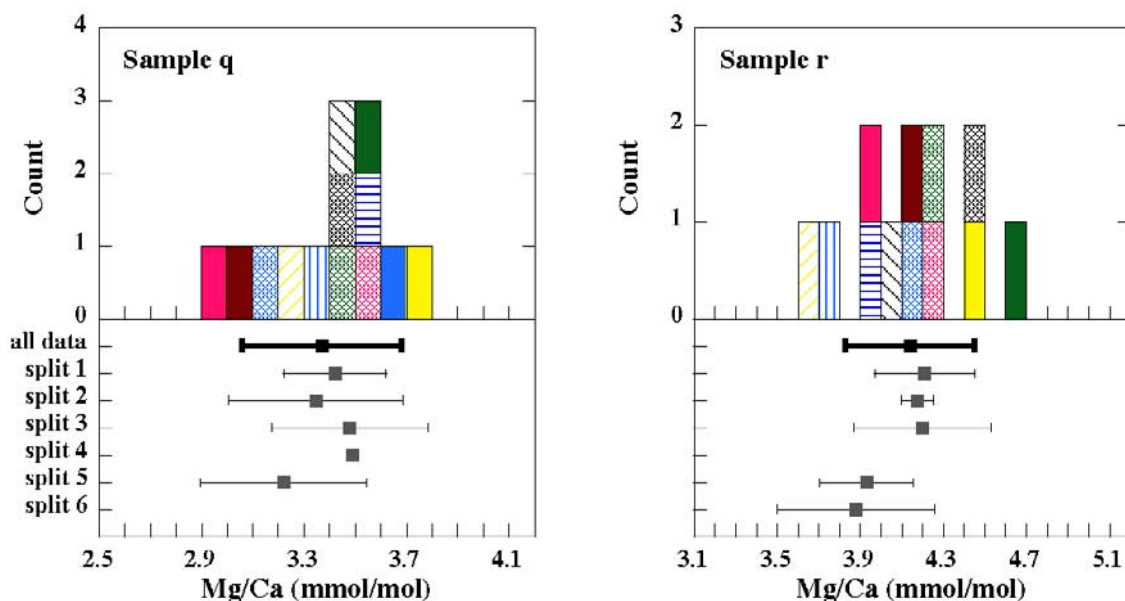
Figure 10. (continued)

shows that data generated by laboratories using the ''Mg cleaning'' method are internally consistent within the analytical errors (Figure 11a). Likewise, data generated only by the ''Cd cleaning'' method are also internally consistent with possibly one exception (Figure 11a); results from

laboratory 360 are lower (up to 7% at the high end) than the other data sets. The offset is not statistically significant, however. Because there are no measurements of the spiked solutions from this laboratory, we cannot evaluate the reasons for this apparent offset.

Table 5. Interlaboratory Variance (SD) and Reproducibility (R) of Mg/Ca Analyses (in mmol mol^{-1}) Obtained Only From Laboratories Using the ''Mg cleaning'' Protocol

| Sample	Mean	Median	SD	RSD%	SR	R	RSD$_R$%
				Mg/Ca			
a	8.46	8.64	0.60	7.1%	0.78	2.18	25.8%
b	3.64	3.66	0.14	3.8%	0.17	0.48	13.1%
c	4.56	4.52	0.27	5.9%	0.32	0.90	19.6%
d	3.73	3.67	0.23	6.2%	0.30	0.84	22.5%
e	3.03	3.03	0.15	5.0%	0.20	0.56	18.5%
f	5.24	5.18	0.34	6.5%	0.42	1.18	22.4%
g	2.85	2.75	0.24	8.4%	0.28	0.78	27.5%
h	3.40	3.47	0.32	9.4%	0.42	1.18	34.6%
i	1.86	1.86	0.09	4.8%	0.12	0.34	18.1%
j	2.71	2.68	0.22	8.1%	0.28	0.78	28.9%
k	4.76	4.74	0.25	5.3%	0.32	0.90	18.8%
l	4.33	4.33	0.21	4.8%	0.27	0.76	17.5%
m	4.24	4.21	0.16	3.8%	0.20	0.56	13.2%
n	2.78	2.74	0.30	10.8%	0.34	0.95	34.2%
o	4.45	4.45	0.19	4.3%	0.25	0.70	15.7%
p	3.01	3.01	0.18	6.0%	0.23	0.64	21.4%
q	3.47	3.44	0.30	8.6%	0.37	1.04	29.9%
r	4.27	4.23	0.26	6.1%	0.33	0.92	21.6%
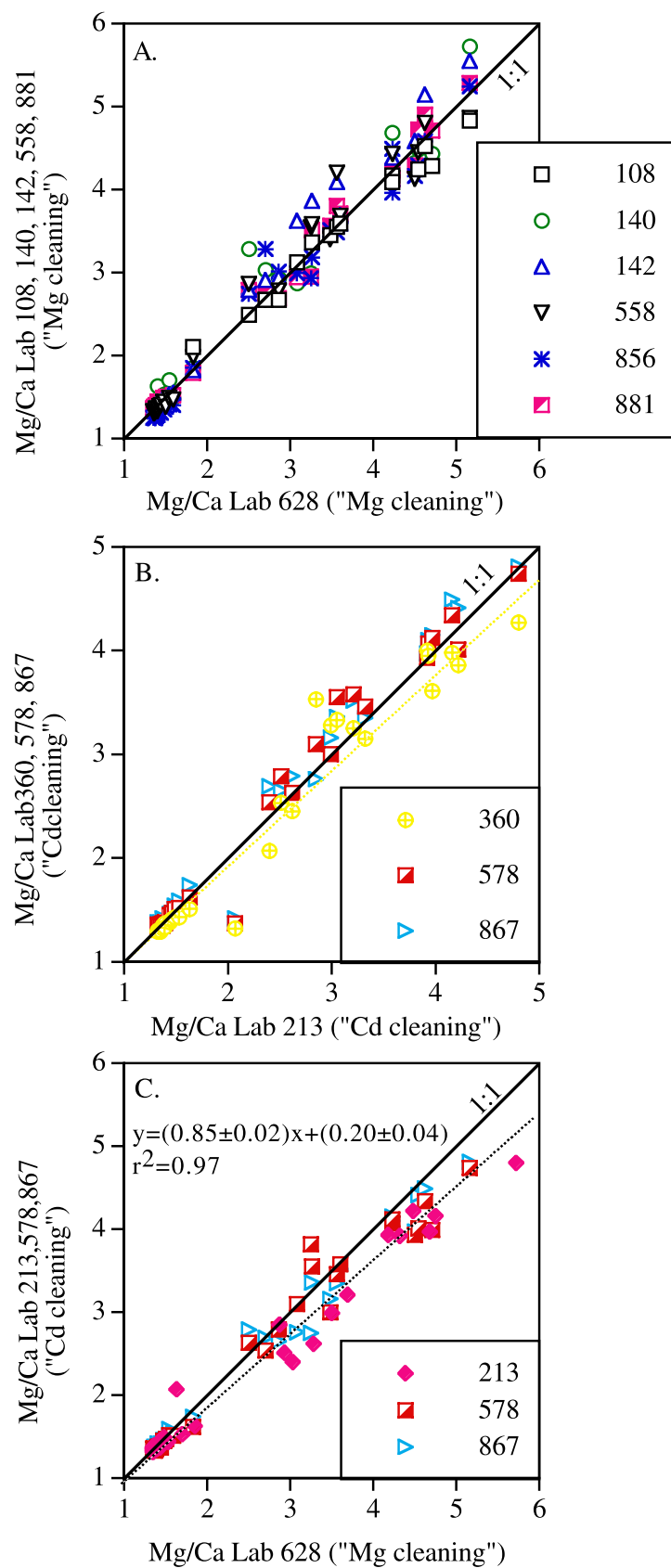
Figure 11.

[26] A regression analysis between the two cleaning methods yield the following relationship: y = (0.85 ± 0.02) x + (0.20 ± 0.04) where y marks results from "Cd cleaning" and x from "Mg cleaning". Note that the intercept is not significantly different from zero. This is somewhat expected since both cleaning methods exhibit similar precision. The results suggest that Mg/Ca ratios obtained from planktonic shells cleaned by the "Cd cleaning" protocol are on average 15% lower than those obtained by the "Mg cleaning" method (Figure 11c). This offset, which is attributable to the addition of reductive step, is a robust result ($r^2 > 0.97$) and is consistent with the results of a recent study by [*Barker et al.*, 2003]. It is also noteworthy that the same slope and intercept are obtained when using either MODEL I or MODEL II regressions. The variance in this estimate is related to several factors such as the preservation state of the shells and the species. *Barker et al.* [2003] demonstrate that the decrease in bulk shell Mg/Ca associated with the reductive step is largely a result of partial dissolution of high-Mg containing layers of the foraminiferal shell; hydrazine reduction is a very corrosive step, leading to a significant dissolution of the calcitic shell. These authors also show that removal of Mn-Fe-oxides is unlikely to explain the observed decrease in Mg/Ca. Partial dissolution of Mg rich layers of the shell is consistent with processes occurring on the seafloor as suggested in several studies [*Brown and Elderfield*, 1996; *Rosenthal et al.*, 2000]. It is also consistent with analyses of planktonic foraminifera using a flow-through dissolution method [*Benway et al.*, 2003; *Haley and Klinkhammer*, 2002].

[27] We use the temperature calibration of [*Anand et al.*, 2003] to estimate the temperature bias due to cleaning effects. Figure 12 shows the difference in temperature estimates of samples cleaned with the

"Cd cleaning" method relative to results obtained by laboratory 628 who used the "Mg cleaning" cleaning method for all the different sites. While there is significant scatter in the data, it is also clear that on average, samples from the "Cd cleaning" method yield temperatures that are about 0.6°C colder than those obtained by the less rigorous cleaning. This difference is significant at the 95% confidence level (±2 SE). In contrast with Mg/Ca, Sr/Ca ratios are not significantly dependent on the choice of cleaning methods.

[28] The current study and that of [*Barker et al.*, 2003] provide strong evidence to suggest that differences in cleaning methods constitute a significant source of interlaboratory variability in Mg/Ca analysis. Thus when applied in paleoceanographic studies cleaning offsets may lead to significant biases in estimating absolute seawater temperature. This is especially the case when samples used in down core studies were cleaned differently than those used for the calibration. The offset does not have, however, a discernible effect on the estimates of relative temperature changes (e.g., glacial-interglacial temperature variability) (Rosenthal-unpublished data). On the basis of this study, there is no clear difference in the intralaboratory %RSD between samples cleaned with the "Mg cleaning" and "Cd cleaning" cleaning (Figure 9). The main issue is, therefore, the interlaboratory reproducibility. In Table 5 we recalculated the interlaboratory standard deviation (SD) and reproducibility (R) of Mg/Ca analyses based only on measurements from laboratories using the "Mg cleaning" protocol. The average standard deviation for all the samples is 6% and the reproducibility is about 22% as compared with average SD of 8% and reproducibility of 28% obtained from measurements using both the "Mg cleaning" and "Cd cleaning" methods. This is a considerable improvement (about

Figure 11. Comparison of Mg/Ca data from the analysis of planktonic foraminifera obtained from different laboratories. Each point represents the mean of the analysis of two replicates from the same subsample. (a) Data from laboratories using the "Mg cleaning" method version. Note that all the data fall on the 1:1 line; (b) Data from laboratories using "Cd cleaning". Note that, except for laboratory 360, all the data fall on the 1:1 line; (c) Comparison of data generated by laboratories using "Cd cleaning" with data from a laboratory using "Mg cleaning" (Lab #628). Note that results obtained from the "Cd cleaning" method are about 15% lower than those obtained from the "Mg cleaning" method.
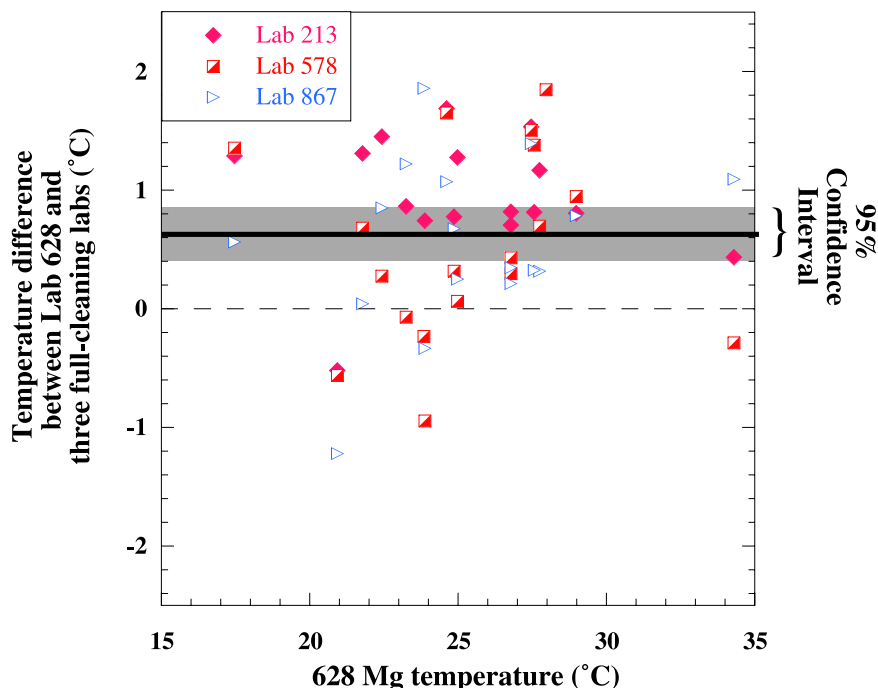
Figure 12. Difference between Mg temperatures from Lab 628, which employed "Mg cleaning" and laboratories 213, 578 and 867, which employed "Cd cleaning" technique. On average, the "Mg cleaning" Mg temperatures are 0.63°C warmer than the "Cd cleaning" Mg temperatures (solid black line). The gray shaded area defines the ±2 S.E. (95% confidence interval) of the data shown. Temperature estimates are based on the calibration of *Anand et al.* [2003].

25%) in interlaboratory performance, which translates to a temperature error of about ±0.7°C rather than the ±1°C obtained previously.

3.3.4. Preservation Effects

[29] Discussion of the alteration of Mg/Ca ratios in planktonic foraminifera due to post-depositional dissolution is beyond the scope of this paper. However, our results show that differences in Mg/Ca due to variable preservation states of the samples significantly exceed the magnitude of analytical differences estimated here. For example, *G. ruber* samples from a shallow core in Little Bahama Bank in the Atlantic Ocean show significantly higher Mg/Ca than samples from the Ontong Java Plateau in the Pacific Ocean although sea surface temperature is significantly warmer in the latter site (26.7° and 28.9° in the LBB and OJP, respectively). In addition, there is a significant difference in Mg/Ca between shallow and deep cores in the OJP. Similar trends are observed for the other planktonic species studied here. Sr/Ca ratios do not exhibit similar depth dependent trends suggesting that post-depositional dissolution has a much smaller effect if any on Sr/Ca ratios in planktonic foraminifera. Using the calibration of *Anand et al.* [2003] we show that dissolution effects may lead to offsets in Mg/Ca-based temperature estimates of up to 4°C. Clearly, corrections for post-depositional-dissolution effects need to be made by either using regional calibrations [*Lea et al.*, 2000] or dissolution corrected calibrations [*Dekens et al.*, 2002; *Rosenthal and Lohmann*, 2002].

4. Conclusions

[30] Results from this study suggest that, in principle, Mg/Ca-based estimates of seawater temperatures are reproducible among laboratories. The few large deviations from the mean population's value are mainly caused by the use of inaccurate working standards or inadequate/inconsistent cleaning. After omitting the errant data sets we

estimate an interlaboratory RSD of about ±8% for Mg/Ca measurements. This suggests an interlaboratory reproducibility of about ±2–3°C. In comparison, the intralaboratory repeatability is about ±1–2°C. Both parameters strongly depend on the type of the sample (i.e., species and degree of preservation). A significant degree of the scatter in the data is caused by the inconsistency in cleaning methods used by the different laboratories. We find a consistent bias, of about 1°C, between the "Cd cleaning" method that includes both oxidative and reductive steps and the "Mg cleaning" version that includes only an oxidation step. These analytical offsets have a significant effect on estimates of absolute temperatures and therefore, on interlaboratory comparability. The offsets, however, do not have a significant effect on estimates of relative down core temperature changes. Obviously, adopting a consistent cleaning method would lead to better reproducibility among laboratories measuring Mg/Ca. In practice, the problem is partly circumvented because laboratories are using local calibration equations to adjust their core top temperature estimates to the overlying observed hydrography. Using this practice, laboratories not only correct for differences in foraminiferal preservation among sites but also inadvertently for poor interlaboratory reproducibility. So in fact, the reproducibility of temperature estimates is often better than that of the Mg/Ca data on which they are based.

[31] In practice, it would be advantageous if all laboratories adopt a single cleaning protocol for Mg/Ca analysis. The results of this study and that of *Barker et al.* [2003] suggest that using the "Mg cleaning" protocol is sufficiently effective in removing all contamination sources. That the "Mg cleaning" protocol is also significantly less laborious than the "Cd cleaning" version further recommends it as a universal method for Mg/Ca measurements. Indeed, many laboratories currently using ICP-OES already use a variation of the short method so this proposal may seem reasonable. However, laboratories studying other trace metal proxies (e.g., Cd/Ca, Ba/Ca, U/Ca) will continue to use the "Cd cleaning" protocol. Thus it is difficult to foresee the adoption of a single method in the

near future. Though a universal method may not be currently feasible, comparability among laboratories can be improved. First, we recommend that working standards be calibrated against independently quantified standards. Second, given the high Mg content of clays we suggest that Al/Ca, Fe/Ca or Ti/Ca be measured along with Mg/Ca to monitor for possible contamination from adhering clays. Thirdly, it becomes apparent that the dilute acid leaching step may have a significant impact on the bulk shell Mg/Ca ratio [*Barker et al.*, 2003]. Therefore it seems that the two cleaning methods could be more consistent if the leaching step in the "Mg cleaning" method is adjusted to compensate for the lack of reductive step (and its concomitant partial dissolution). This, however, needs to be rigorously assessed.

[32] An obvious, but nonetheless important, conclusion of this study is the need for standards calibration among laboratories as a first step toward improving interlaboratory compatibility. Clearly with the increase in number of laboratories involved in Mg/Ca measurements for paleoceanographic studies there is an urgent to develop an agreeable solid standard to be used by all the laboratories much as is done in isotope analyses. It would also be beneficial if new comers consult with more experienced laboratories about cleaning and analytical techniques and carry out their own intercomparison studies before proceeding with paleoceanographic research. Reducing the uncertainties in Mg/Ca-based temperature estimates, both due to interlaboratory inconsistency and dissolution effects, offers the potential of using paired $\delta^{18}O$ and Mg/Ca data for rigorous paleoceanographic reconstructions in a similar manner to that conducted with modern data. Issues raised here should be addressed through further research and perhaps also in a workshop to discuss the different approaches.

Acknowledgments

References

Anand, P., H. Elderfield, and M. H. Conte (2003), Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time-series, *Paleoceanography*, *18*(2), 1050, doi:10.1029/2002PA000846.

Bainbridge, A. E. (1981), *GEOSECS Atlantic Expedition, Vol. 1, Hydrographic Data*, 121 pp., Natl. Sci. Foundation, Washington, D. C.

Barker, S., M. Greaves, and H. Elderfield (2003), A study of cleaning procedures used for foraminiferal Mg/Ca paleothermometry, *Geophys. Geochem. Geosyst.*, *4*(9), 8407, doi:10.1029/2003GC000559.

Benway, H. M., B. A. Haley, G. P. Klinkhammer, and A. C. Mix (2003), Adaptation of a flow-through leaching procedure for Mg/Ca paleothermometry, *Geochem. Geophys. Geosyst.*, *4*(2), 8403, doi:10.1029/2002GC000312.

Boyle, E. A. (1995), Limits on benthic foraminiferal chemical-analyses as precise measures of environmental properties, *J. Foraminiferal Res.*, *25*, 4–13.

Boyle, E. A., and L. D. Keigwin (1985), Comparison of Atlantic and Pacific paleochemical records for the last 250,000 years: Changes in deep ocean circulation and chemical inventories, *Earth Planet. Sci. Lett.*, *76*, 135–150.

Boyle, E. A., and Y. Rosenthal (1996), Chemical hydrography of the South Atlantic during the last glacial maximum: D¹³C vs. Cd, in *The South Atlantic: Present and Past Circulation*, edited by G. Wefer et al., pp. 423–443, Springer-Verlag, New York.

Brown, S. J., and H. Elderfield (1996), Variations in Mg/Ca and Sr/Ca ratios of planktonic foraminifera caused by postdepositional dissolution: Evidence of shallow Mg-dependent dissolution, *Paleoceanography*, *11*, 543–551.

Dekens, P. S., D. W. Lea, D. K. Pak, and H. J. Spero (2002), Core top calibration of Mg/Ca in tropical foraminifera: Refining paleotemperature estimation, *Geochem. Geophys. Geosyst.*, *3*(4), 1022, doi:10.1029/2001GC000200.

Elderfield, H., and G. Ganssen (2000), Past temperature and δ^{18}O of surface ocean waters inferred from foraminiferal Mg/Ca ratios, *Nature*, *405*, 442–445.

Emiliani, C. (1955), Mineralogical and chemical composition of the tests of certain pelagic foraminifera, *Micropaleontology*, *1*, 377–380.

Haley, B. A., and G. P. Klinkhammer (2002), Development of a flow-through system for cleaning and dissolving foraminiferal tests, *Chem. Geol.*, *185*, 51–69.

Hastings, D. W., A. D. Russell, and S. R. Emerson (1998), Foraminiferal magnesium in Globeriginoides sacculifer as a paleotemperature proxy, *Paleoceanography*, *13*, 161–169.

Lea, D. W., T. A. Mashiotta, and H. J. Spero (1999), Controls on magnesium and strontium uptake in planktonic foraminifera determined by live culturing, *Geochim. Cosmochim. Acta*, *63*, 2369–2379.

Lea, D. W., D. K. Pak, and H. J. Spero (2000), Climate impact of late Quaternary equatorial Pacific sea surface temperature variations, *Science*, *289*, 1719–1724.

Lear, C. H., Y. Rosenthal, and N. Slowey (2002), Benthic foraminiferal Mg/Ca-paleothermometry: A revised core-top calibration, *Geochim. Cosmochim. Acta*, *66*, 3375–3387.

Levitus, S., and T. P. Boyer (1994), *World Ocean Atlas 1994 Volume 4: Temperature*, 117 pp., U.S. Dept. of Comm., Washington, D. C.

Martin, P. A., and D. W. Lea (2002), A simple evaluation of cleaning procedures on fossil benthic foraminiferal Mg/Ca, *Geochem. Geophys. Geosyst.*, *3*(10), 8401, doi:10.1029/2001GC000280.

Martin, P. A., D. W. Lea, T. A. Mashiotta, T. Papenfuss, and M. Sarnthein (1999), Variation of foraminiferal Sr/Ca over Quaternary glacial-interglacial cycles: Evidence for changes in mean ocean Sr/Ca?, *Geochem. Geophys. Geosyst.*, *1*, Paper number 1999GC000006.

Mashiotta, T. A., D. W. Lea, and H. J. Spero (1999), Glacial-interglacial changes in subantarctic sea surface temperature and d¹⁸O-water using foraminiferal Mg, *Earth Planet. Sci. Lett.*, *170*, 417–432.

Nilsson, T., R. Ferrari, and S. Facchetti (1997), Interlaboratory studies for the validation of solid-phase microextraction for the quantitative analysis of volatile organic compounds in aqueous samples, *Anal. Chim. Acta*, *356*, 113–123.

Nürnberg, D., J. Bijma, and C. Hemleben (1996), Assessing the reliability of magnesium in foraminiferal calcite as a proxy for water mass temperature, *Geochim. Cosmochim. Acta*, *60*, 803–814.

Rosell-Melé, A., et al. (2001), Precision of the current methods to measure the alkenone proxy UK'37 and absolute alkenone abundance in sediments: Results of an interlaboratory comparison study, *Geophysics Geochemistry Geosystems*, *2*, Paper number 2000GC000141.

Rosenthal, Y., and G. P. Lohmann (2002), Accurate estimation of sea surface temperatures using dissolution-corrected calibrations for Mg/Ca paleothermometry, *Paleoceanography*, *17*(3), 1044, doi:10.1029/2001PA000749.

Rosenthal, Y., E. A. Boyle, and L. Labeyrie (1997), Last glacial paleochemistry and deep water circulation in the Southern Ocean: Evidence from foraminiferal cadmium, *Paleoceanography*, *12*, 778–787.

Rosenthal, Y., G. P. Lohmann, K. C. Lohmann, and R. M. Sherrell (2000), Incorporation and preservation of Mg in *Gs. sacculifer*: Implications for reconstructing sea surface temperatures and the oxygen isotopic composition of seawater, *Paleoceanography*, *15*, 135–145.

Stoll, H. M., and D. P. Schrag (2000), Coccolith Sr/Ca as a new indicator of coccolithophorid calcification and growth rate, *Geochem. Geophys. Geosyst.*, *1*, Paper number 1999GC000015.

Stoll, H. M., D. P. Schrag, and S. C. Clemens (1999), Are seawater Sr/Ca variations preserved in Quaternary foraminifera?, *Geochimica Cosmochimica Acta*, *63*, 3535–3547.