Marine Micropaleontology 133 (2017) 21–27

Contents lists available at ScienceDirect

Marine Micropaleontology

journal homepage: www.elsevier.com/locate/marmicro

Research paper

Calibration of the repeatability of foraminiferal test size and shape measures with recommendations for future use

Anieke Brombacher[a,*], Paul A. Wilson[a], Thomas H.G. Ezard[a,b]

[a] National Oceanography Centre Southampton, University of Southampton, Waterfront Campus, European Way, Southampton SO14 3ZH, UK
[b] Centre for Biological Sciences, University of Southampton, Life Sciences Building 85, Highfield Campus, Southampton SO17 1BJ, UK

ABSTRACT

The fossil record of planktonic foraminifera is ideally suited to defining stratigraphic age controls and exploring fundamental questions in evolutionary biology due to its excellent preservation potential that yields continuous, high-resolution fossil archives of large numbers of individuals. For full morphometric analyses foraminifera tests are generally mounted, oriented and imaged manually, while data are processed using standard software such as ImageJ or Image Pro. However, manually induced orientation errors are a source of potential bias in trait measurements even when quantified using the same computational subroutine. Here we test the repeatability of four measures of foraminiferal test shape on six morphologically distinct species and present a calibration (power analysis) of the number of individuals needed to determine a given percentage change in these traits. We mounted and measured every individual twice and analysed the difference between the two measurements to determine the effects of small orientation changes on the studied traits. We show that measurements of test area and aspect ratio are statistically indistinguishable between runs for all species studied, and a power law calibration suggests that between 25 and 50 individuals are needed to detect at least a 10% in- or decrease in either trait. However, despite mounting tests on glass slides to clarify perimeter outlines, test perimeter was only repeatable in the spherical species *Orbulina universa*, and test roundness was not repeatable for three out of six studied species. We recommend the use of aspect ratios constructed from lengths and avoidance of perimeters and their dependent metrics to reduce orientation induced bias.

1. Introduction

The planktonic foraminifera bequeath one of the most complete fossil records known to science. The accumulation in deep sea sediments of well-preserved shells of vast numbers of individuals make the planktonic foraminiferal fossil record uniquely suited for both biostratigraphic age controls (Blow, 1969; Bolli et al., 1989; Berggren et al., 1995; Wade et al., 2011), and for answering fundamental questions in evolutionary biology (e.g. Wei and Kennett, 1988; Norris, 1991; Alizon et al., 2008; Hull and Norris, 2009; Ezard et al., 2011). The preservation of complete specimens allows for the construction of multivariate trait datasets, which can be used to distinguish between species in a quantitative manner and pinpoint the exact timing of speciation and extinction (Wei, 1994; Aze et al., 2011; Pearson and Ezard, 2014), and allow for high-resolution reconstructions of species' evolutionary trajectories over millions of years (Kucera and Malmgren, 1998; Wade and Olsson, 2009; Pearson et al., 2014).

The reliability of morphometric records depends on the precision with which individual traits can be measured. A good measurement system allows for rapid processing, is repeatable between runs and produces reliable results. Planktonic foraminifera are most often measured from two-dimensional images taken by a camera attached to a microscope with individual tests mounted in a given orientation, and trait measurements are extracted from the image's 2D test representation. This set up has the potential to introduce bias in two main ways: manually measuring traits on the imaged specimens involves subjectivity, and hand mounting of individual specimens introduces error in the orientation of the tests. The first issue can be addressed using automated specimen detection and trait measurements with fixed magnification and light intensity, which reduces subjective human input. Mesaurement biases induced by mounting errors can be reduced by mounting tests on a rotatable hemispherical stage which is manually adjusted to fine-tune specimen orientation prior to imaging (MacLeod and Carter, 1984; Knappertsbusch, 2007; Knappertsbusch et al., 2009; Pearson and Ezard, 2014). However, specimen adjustment in this setup still relies on subjective human input, and the time consuming nature of adjusting, imaging and analysing each individual separately makes this approach suboptimal for large (> 10,000 specimens) datasets that are

* Corresponding author.
E-mail address: j.brombacher@noc.soton.ac.uk (A. Brombacher).

http://dx.doi.org/10.1016/j.marmicro.2017.05.003
Received 30 January 2017; Received in revised form 15 May 2017; Accepted 23 May 2017
Available online 26 May 2017

increasingly produced (Brombacher et al., 2017; Malmgren and Kennett, 1981; Knappertsbusch, 2000; Pearson and Ezard, 2014; Hsiang et al., 2016). We focus, therefore, on hand-mounted individuals using a fixed stage, but the effects of mounting errors need to be estimated by repeatedly mounting, measuring, remounting and re-measuring individual tests.

Here we test the repeatability of four measures of foraminifera test size and shape: test area, perimeter, aspect ratio and roundness. Test area represents the individual's body size, an ecologically important trait (Hecht, 1976; Schmidt et al., 2004) that is easy to measure and a useful first estimate to distinguish between species. Test perimeter is often used in multivariate morphometric studies to assign the position of landmarks (Wei and Kennett, 1988; Biolzi, 1991; Wei, 1994). Aspect ratio and roundness are two measures of test shape, further enabling species identification as well as quantifying the test area-to-volume ratio, an important measure in terms of metabolic processes. Both metrics are routinely generated by popular software such as ImageJ or Image Pro. Together, these four traits form an important part of describing planktonic foraminifera morphologies. Therefore, quantifying their precision is crucial to the task of interpreting species morphometric records for both biostratigraphic and evolutionary purposes.

2. Material and methods

2.1. Study species

Here we present repeated measurements on the tests of six species of planktonic foraminifera with distinct shell morphologies. All taxonomic descriptions given here are from Kennett and Srinivasan (1983) and references therein.

Orbulina universa (Fig. 1a). The adult stage consists of a single spherical final chamber enveloping the earlier part of the test. In this study only adult tests are used.

Globoconella inflata (Fig. 1b). Low trochospiral tests with a broadly rounded axial periphery and an extraumbilical-umbilical aperture. Chambers more inflated on the umbilical side than the spiral side and increase slowly in size as added.

Globoconella puncticulata (Fig. 1c). Low trochospiral tests with a flattened spiral side, highly vaulted umbilical side and bluntly rounded axial periphery. Chambers are angular and increase slowly in size as added. The extraumbilical-umbilical aperture is a high interiomarginal arch.

Truncorotalia crassaformis (Fig. 1d). Low trochospiral tests with a flat spiral side, strongly convex umbilical side and planoconvex, subacute axial periphery. Chambers are compressed and increase rapidly in size as added. The extraumbilical-umbilical aperture is a low-arched slit bordered by a lip.

Globigerinella siphonifera (Fig. 1e). Adult tests are evolute and planispiral, with a rounded axial periphery and a wide arched interiomarginal aperture. Chambers are globular and increase rapidly in size as added.

Globorotalia tumida (Fig. 1f). Tests are trochospiral and convex, with the spiral side more convex than the umbilical side and a narrow extraumbilical-umbilical aperture. The axial periphery is acute with a heavy keel. The chambers are wedge-shaped and increase rapidly in size as added. The extraumbilical-umbilical aperture is covered by a plate-like lip.

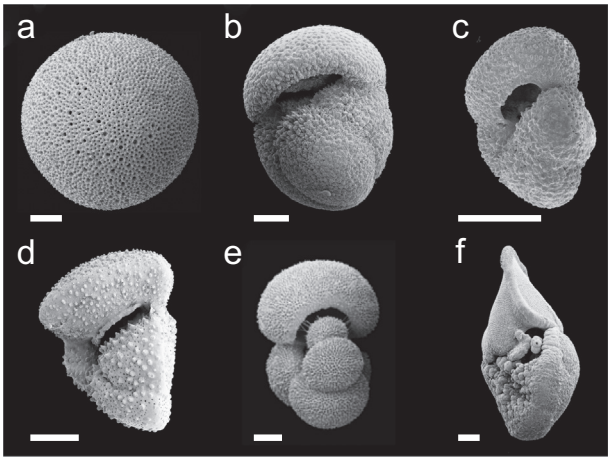


Fig. 1. SEM images of a) *Orbulina universa*, b) *Globoconella inflata*, c) *Globoconella puncticulata*, d) *Truncorotalia crassaformis*, e) *Globigerinella siphonifera* and f) *Globorotalia tumida*. Scale bars represent 100 μm.

2.2. Analysis

Specimens of *O. universa, G. siphonifera* and *G. tumida* were picked from a box core sample collected by the GLObal Warming (GLOW) cruise at station GLOW 5 (− 8.9 °N, 41.5 ′W). Individuals of *G. inflata, G. puncticulata* and *T. crassaformis* were collected from IODP Site U1313 (Leg 306, 41 °N, 32.5 ′W). *G. puncticulata* and *T. crassaformis* were picked from sample 306-U1313C-12H-4, 22–24 cm and specimens of *G. inflata* were picked from sample 306-U1313B-10H-4, 45–47 cm.

We picked and mounted the first 100 specimens encountered of each species. To minimise measurement errors from background imperfections two types of slides were tested using three different adhesives to find the most homogenous background. Gridded cardboard slides allow for easy specimen identification with one individual per numbered square, however small white background imperfections in the cardboard result in parts of the cardboard slide being mistakenly identified as belonging to the foraminifera test (Fig. 2a-c). Transparent glass slides provide a homogenous dark background when illuminated from above (Fig. 2d–f). Pritt stick is easy to apply on both types of slides but leaves opaque strands of glue that are identified as part of the foraminifera test (Fig. 2a, d); transparent particle glue leaves less traces but dries out quickly, allowing too little time to mount tests carefully (Fig. 2b,e). Transparent double-sided sticky tape is easy to apply and does not dry out so quickly (Fig. 2f). Therefore, tests were mounted on glass slides using transparent double-sided tape (Fig. 2f).

Individuals were oriented in side view on a fixed stage. This orientation generally provides the best view of the test aperture and related, ecologically relevant landmarks and is often used in

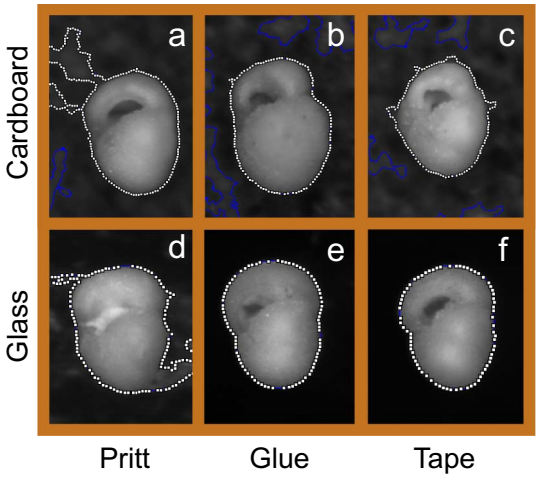
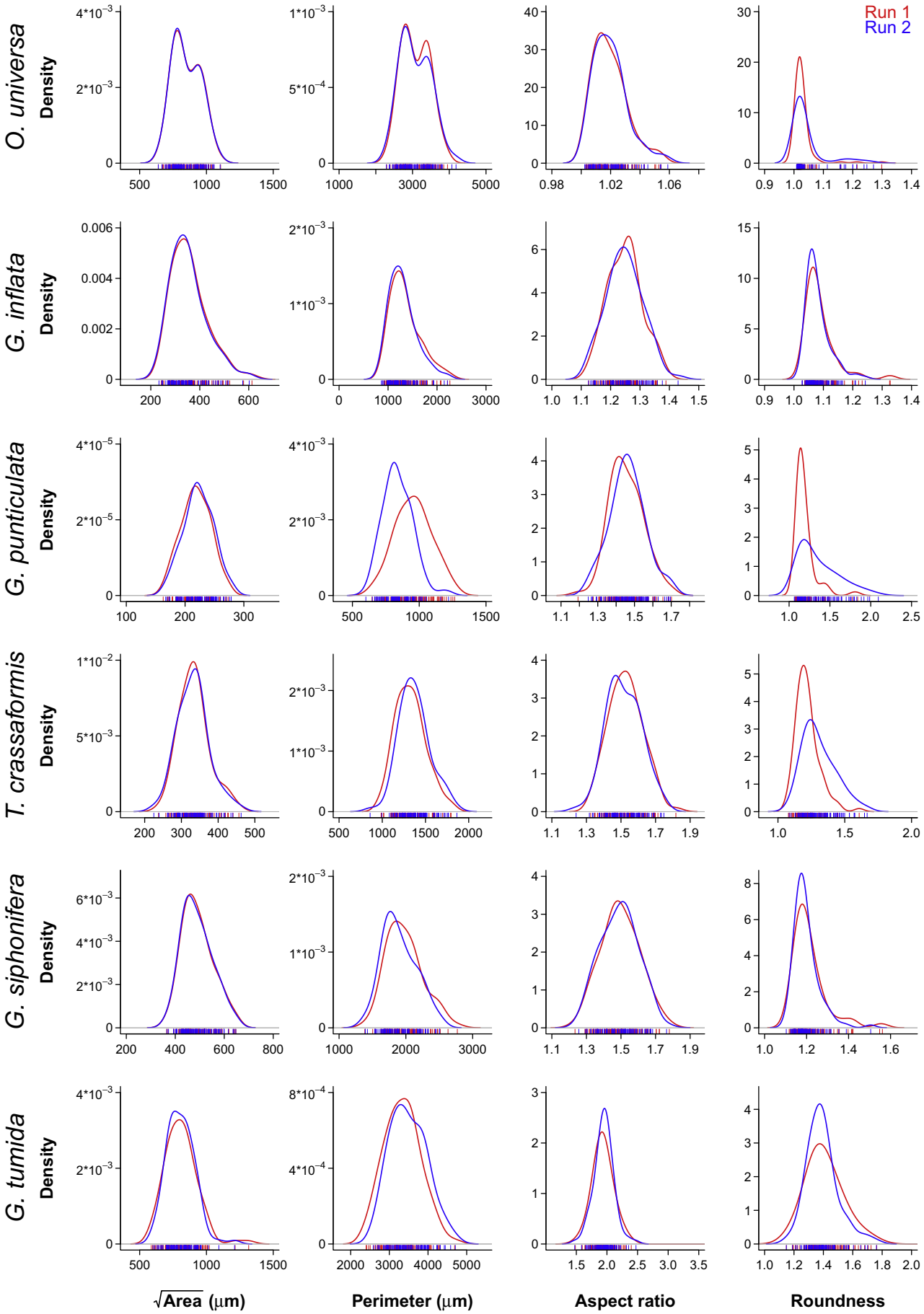


Fig. 2. Foraminifera tests mounted on cardboard (a-c) and glass (d-f) slides using Pritt stick (a,d), transparent particle glue (b,f) and transparent double-sided tape. White dotted outlines indicate the tests as recognised by the Image Pro Premier software. Blue outlines indicate background imperfections also picked up by the software. Object recognition was best using foraminifera mounted on glass slides using double sided tape (f) and therefore this setup was used here to test trait repeatability. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(caption on next page)

Fig. 3. Kernel density plots of first (red) and second (blue) set of morphometric measurements on *Orbulina universa*, *Globoconella inflata*, *Globoconella puncticulata*, *Truncorotalia crassaformis*, *Globigerinella siphonifera* and *Globorotalia tumida*. Rug plots on the horizontal axes indicate individual data points.

morphometric analyses of planktonic foraminifera (Lohmann and Malmgren, 1983; Wei, 1994; Kucera and Malmgren, 1998; Knappertsbusch, 2007; Pearson and Ezard, 2014). However, depending on the acuteness of the axial periphery this is also the least stable position, which potentially increases measurement errors because the test sits obliquely on the slide and more or less of the perimeter is visible. Tests were imaged with fixed light intensity and camera exposure time using an Infinity 3 Lumenera camera mounted on an Olympus SZX10 light microscope, with illumination from above. Test area, aspect ratio (ratio between maximum test height and width), perimeter length and roundness (π*Perimeter^2/Area) were extracted from the images using an automated image analysis macro in the Image Pro Premier software. Individuals were mounted, measured, carefully removed from the slides to avoid damaging the test, and remounted and re-measured once to provide an upper bound on trait repeatability.

Here we study the differences between the first and second measurements of a given trait on the same individuals. Measurements of the first and second run are plotted in Fig. 3 using continuous frequency distributions analogous to histograms (kernel density estimates with a guassian kernel and bandwidth $h = 1.06*sn^{-1/5}$ following Silverman (1986), with s the standard deviation of trait mesurements per species and n the number of analysed individuals). To determine which traits are repeatable in which species we use a Wilcoxon signed-rank test in R (R Core Team, 2013), which is a non-parametric test that compares repeated measurements on a single sample to assess whether their mean ranks differ. If the first and second runs are statistically similar the measurement differences within individuals are expected to be centred around 0 and show little error. Differences significantly deviating from 0 indicate a systematic offset between measurements in the first and second run, implying high sensitivity to small differences in test orientation and low trait repeatability.

If a trait is shown to be repeatable, we determine the number of individuals required to reliably detect a change in a given trait using power analysis (Cohen, 1988). This number is influenced by both the natural variability within a species, with higher variability requiring a larger number of individuals to detect a given trend, as well as variation induced by small mounting errors. Both kinds of variability are present in our dataset, but because it is impossible to perfectly mount specimens it is not possible to separate these effects in our dataset. However, high repeatability suggests that most of the observed variation is due to natural trait variability as opposed to procedural errors, whereas it delimiting the two becomes more troubling when repeatability is low. We apply power analysis with the 'pwr' package in R (R Core Team, 2013). When variation in a population is known, power analysis calculates the sample size required to detect a specified trend (effect size) for a given power (probability of finding a true effect) and significance level (probability of finding that an effect that is not there). We use power analysis to determine the sample size required to detect a trait change of 5, 10, 15, 20, 25 and 30% for varying power values with a significance level set to 0.01.

3. Results & discussion

3.1. Area

Measurements of test area vary little between runs (Fig. 3). Differences between first and second measurements per individual were very small with the Interquartile Range of the differences (the distance between the 25th and 75th percentiles) reaching < 5% away from the species' mean for all studied species except *G. puncticulata* (Fig. 4a), and no significant differences were detected between runs for all studied species (Wilcoxon signed-rank tests, see Fig. 4a for species-specific *p*-values). These results imply that foraminifera body size is a repeatable measure not dependent on small mounting errors. Because of this trait's high repeatability, relatively small size changes can be detected reliably (Fig. 5): only 50 individuals are needed to detect an increase in test area of 10% or larger for power, i.e., the probability of detecting an effect that is present in the data, > 0.9 and a confidence level, i.e. probability of a false positive, of p = 0.01 (Fig. 5). The only exception is *G. inflata*, for which the same number of individuals would only enable a detection of a > 15% change in size. In principle, this lower sensitivity in *G. inflata* can be explained by either higher natural and/or higher mounting-induced variability in this species, but given that this species is relatively easy to mount because of its rounded periphery, and mounting induced errors are very low in *G. inflata* size (Fig. 4a), we conclude that a high natural size variability due to the existence of different morphotypes (Kennett and Srinivasan, 1983) is the most likely explanation for the observed differences.

3.2. Aspect ratio

Measurements of aspect ratio are similar between runs for all species (Fig. 3, third column) and repeated measurements on the same individuals are statistically indistinguishable (Fig. 4c, Wilcoxon signed-rank tests), implying that aspect ratio is a repeatable measure of foraminifera test shape. Results from power analysis show that only 25 individuals are needed to detect a 10% increase in aspect ratio for power > 0.9 and a confidence level of 0.01 for all species (Fig. 5).

3.3. Perimeter

Measurement distributions of test perimeter vary between runs for all species (Fig. 3). The differences between first and second measurements on individuals deviate significantly from 0 in all species except the spherical *O. universa* (Wilcoxon singed-rank tests, see Fig. 4b for species-specific p-values), implying that test perimeter is not a repeatable measure in the other five species. These results underline the need for species-specific error quantification when using this measure, especially when used in a full morphometric approach where landmarks for other traits are assigned at specific points on the test outline. When this approach is used the repeatability of each landmark should be quantified separately, because uncertainty in the perimeter can also influence the repeatability of the individual landmarks and their associated traits.

3.4. Roundness

The similarity of the measurement distribution for roundness varies among the studied species, with most repeatable measurements for *O. universa*, *G. inflata* and *G. siphonifera* (Fig. 3). The individual differences between first and second measurements deviate significantly from 0 in *G. puncticulata, T. crassaformis* and *G. siphonifera*, implying that roundness is not a repeatable trait in these species (Fig. 4d). In *O. universa*, *G. inflata* and *G. tumida*, however, we found no significant mounting-induced errors. For each of these three species, power analysis shows that fewer than 25 individuals are required to detect a 10%

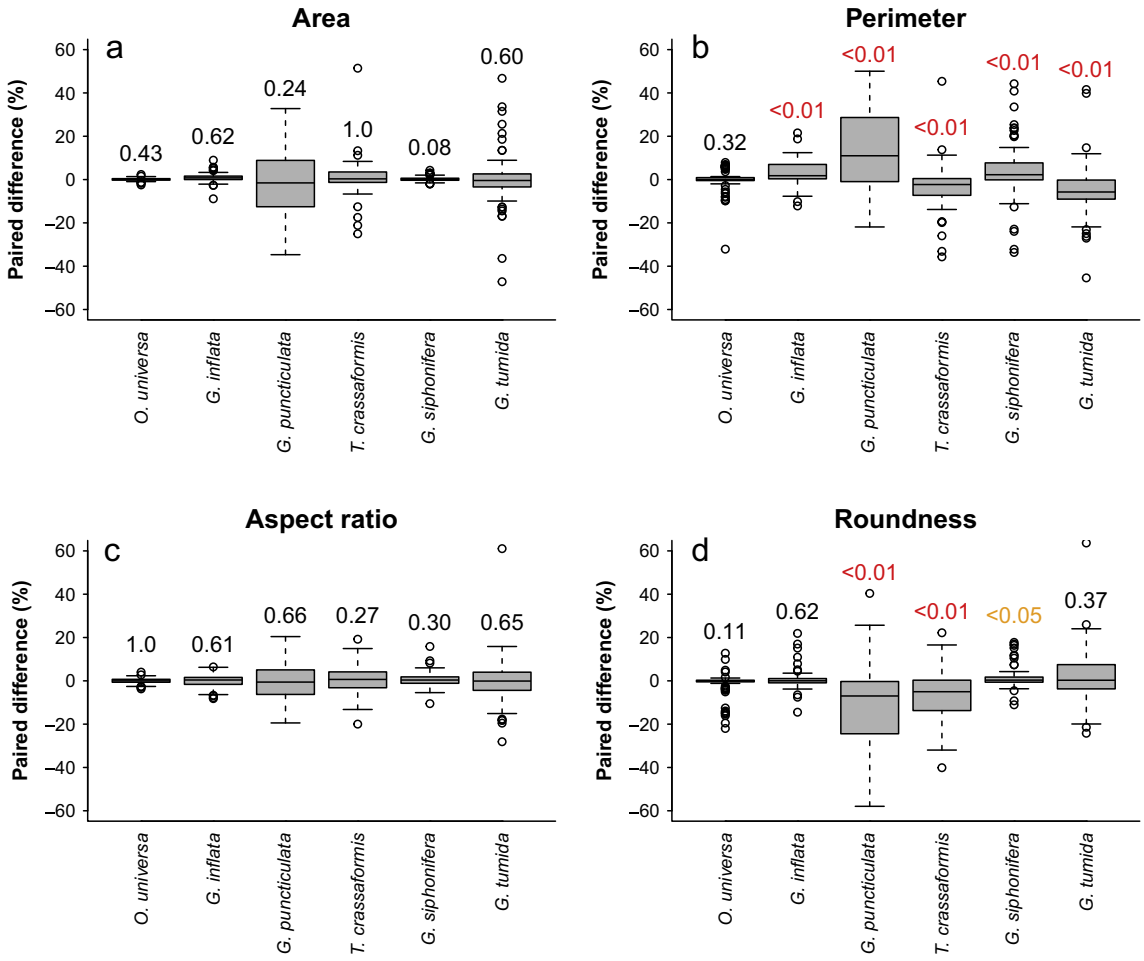


Fig. 4. Barplots of the difference between repeated measurements on a) area, b) perimeter, c) aspect ratio and d) roundness on the same individual (paired difference) expressed as percentage of the individual's trait mean. p-Values of the Wilcoxon signed-rank test performed on subsequent measurements on the same individuals are given for every species, with p-values smaller than 0.01 indicating significant differences between runs shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

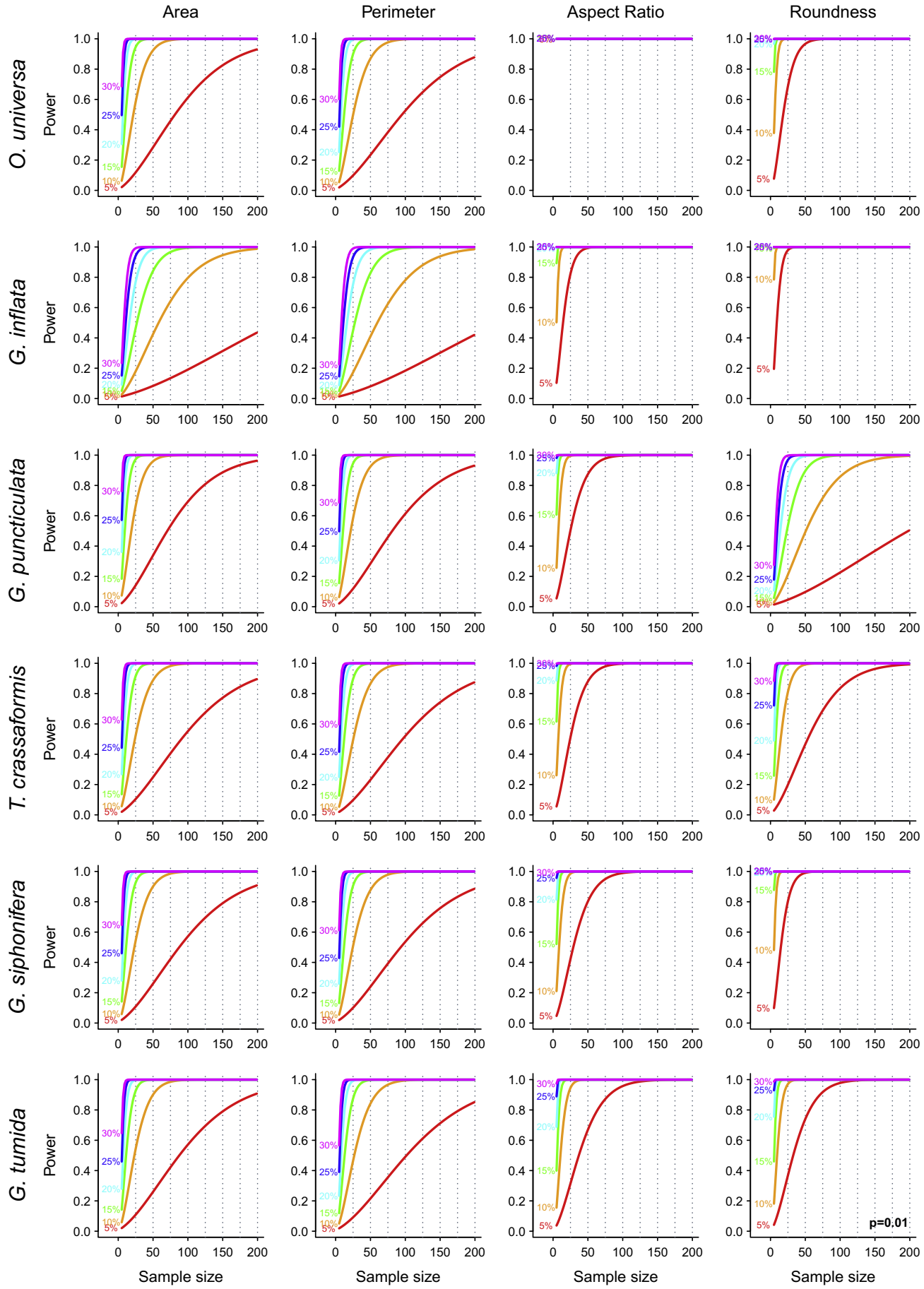
change in roundness with power > 0.9 and a significance level of p = 0.01 (Fig. 5).

The varying degrees of repeatability in roundness could reflect the composite nature of the trait: roundness is defined as π*Perimeter^2/Area. Because test perimeter is greatly influenced by orientation errors, its biases are also expected to influence test roundness. These results imply that extra care should be taken when analysing more complex composite traits, and that the reliability of all separate traits should be quantified prior to interpreting any changes in roundness in the fossil record.

4. Conclusion and recommendations

We report a test of the repeatability of four measures of planktonic foraminifera: size, shape and outline and the sample size required to pick up trends of a given magnitude. We present a novel mounting technique using a glass slides that reduces background imperfections and increases accuracy of trait capture (Fig. 2). Both test area and aspect ratio are repeatable measures of test size and shape, whereas roundness is a repeatable measure for *O. universa*, *G. inflata* and *G. tumida* but not *G. puncticulata*, *T. crassaformis* and *G. siphonifera*, while perimeter is not repeatable for any of our non-spherical species (Fig. 4). Our results underline the need for measurement error quantification in individual species' traits prior to interpreting their morphological records. In particular, test perimeter and the other composite traits it influences should be applied with extreme caution. Work is needed to investigate the repeatability of individual landmarks on test outline before they are applied for evolutionary or biostratigraphical purposes.

Results from the power analyses show that between 25 and 50 individuals are needed to detect a 10–15% change in the repeatable traits, which is well within the scope of most species of planktonic foraminifera. We use a significance threshold of p = 0.01 because of the abundance of the microfossil record: we recommend that micropalaeontologists target lower significance levels (e.g., p < 0.01 rather than p < 0.05), particularly in common species, to reduce the probability of reporting false positive results. The sample size required to detect statistically significant trait changes depends on the magnitude of change, and should therefore be determined at the start of each experiment separately to ensure efficient data collection protocols. Focussing on repeatable traits will also ensure that statistical outputs like effect size, which are arguably more informative than the level of statistical significance in inferring the ecological role of trait changes, can be estimated more accurately.



(caption on next page)

Fig. 5. Calibration of the number of individuals required to detect a given trait change in individual species. Here power is plotted against the number of individuals needed to detect changes in trait values by 5% (red), 10% (orange), 15% (green), 20% (cyan), 25% (blue) and 30% (magenta). The significance level is set to p = 0.01.

Acknowledgments

We thank Bridget Wade and Kirsty Edgar for help with foraminifera taxonomy and species identification and Richard Pearce for his assistance with SEM imaging. We also thank two anonymous reviewers for comments and suggestions that greatly improved this manuscript. AB acknowledges a Johanna M. Resig Foraminiferal Research Fellowship awarded by the Cushman Foundation for Foraminiferal Research. THGE is funded by NERC Advanced Research Fellowship NE/J018163/1. PAW acknowledges a Royal Society Wolfson Research Merit Award. Samples were provided by the GLObal Warming (GLOW) cruise, and Integrated Ocean Drilling Program (IODP), which is sponsored by the US National Science Foundation and participating countries under management of Joint Oceanographic Institutions, Inc.

References

Alizon, S., Kucera, M., Jansen, V.A., 2008. Competition between cryptic species explains variations in rates of lineage evolution. Proc. Natl. Acad. Sci. U. S. A. 105, 12382–12386.

Aze, T., Ezard, T.H., Purvis, A., Coxall, H.K., Stewart, D.R., Wade, B.S., Pearson, P.N., 2011. A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. Biol. Rev. Camb. Philos. Soc. 86, 900–927.

Berggren, W.A., Kent, D.V., Swisher, C.C., Aubry, M.-P., 1995. A revised Cenozoic geochronology and chronostratigraphy. In: Geochronology, Time Scales, and Global Stratigraphic Correlation. 54. pp. 129–212 (SPEM special publication).

Biolzi, M., 1991. Morphometric analyses of the Late Neogene planktonic foraminiferal lineage *Neogloboquadrina dutertrei*. Mar. Micropaleontol. 18, 129–142.

Blow, W.H., 1969. Late Middle Eocene to Recent planktonic foraminiferal biostratigraphy. In: Proceedings of the First International Conference on Planktonic Microfossils. EJ Brill, Leiden, pp. 199–422.

Bolli, H.M., Saunders, J.B., Perch-Nielsen, K., 1989. Planktic Foraminifera, Calcareous Nannofossils and Calpionellids. Cambridge University Press, New York.

Brombacher, A., Wilson, P.A., Bailey, I., Ezard, T.H.G., 2017. The Breakdown of Static and Evolutionary Allometries During Climatic Upheaval. The American Naturalist 190 (in press) (In press at American Naturalist).

Cohen, J., 1988. Statistical Power Analysis for the Behavioural Sciences. Lawrence Earlbaum Associates, Hillside, NJ.

Core Team, R., 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statiastical Computing, Vienna, Austria.

Ezard, T.H.G., Aze, T., Pearson, P.N., Purvis, A., 2011. Interplay between changing climate and species' ecology drives macroevolutionary dynamics. Science 332, 349–351.

Hecht, A.D., 1976. An ecological model for test size variation in recent planktonic foraminifera: applications to the fossil record. J. Foraminifer. Res. 6, 295–311.

Hsiang, A.Y., Elder, L.E., Hull, P.M., 2016. Towards a morphological metric of assemblage dynamics in the fossil record: a test case using planktonic foraminifera. Philos. Trans. R. Soc. B 371, 20150227.

Hull, P.M., Norris, R.D., 2009. Evidence for abrupt speciation in a classic case of gradual evolution. Proc. Natl. Acad. Sci. U. S. A. 106, 21224–21229.

Kennett, J.P., Srinivasan, M.S., 1983. Neogene Planktonic Foraminifera. A Phylogenetic Atlas. Hutchinson Ross Publishing Company, Stroudsburg, Pennsylvania.

Knappertsbusch, M., 2000. Morphologic evolution of the coccolithophorid *Calcidiscus leptoporus* from the Early Miocene to Recent. J. Paleontol. 74, 712–730.

Knappertsbusch, M., 2007. Morphological variability of *Globorotalia menardii* (planktonic foraminifera) in two DSDP cores from the Caribbean Sea and the Eastern Equatorial Pacific. In: Carnets de Géologie/Notebooks on Geology, Brest, Article. 4. pp. 1–34.

Knappertsbusch, M., Binggeli, D., Herzig, A., Schmutz, L., Stapfer, S., Schneider, C., Eisenecker, J., Widmer, L., 2009. AMOR - a new system for automated imaging of microfossils for morphometric analyses. Palaeontol. Electron. 12, 1–20.

Kucera, M., Malmgren, B., 1998. Differences between evolution of mean form and evolution of new morphotypes: an example from Late Cretaceous planktic foraminifera. Paleobiology 24, 49–63.

Lohmann, G., Malmgren, B.A., 1983. Equatorward migration of *Globorotalia truncatulinoides* ecophenotypes through the late Pleistocene: gradual evolution or ocean change? Paleobiology 9, 414–421.

MacLeod, N., Carter, J.L., 1984. A method for obtaining consistent specimen orientations for use in microfossil biometric studies. Micropaleontology 306–310.

Malmgren, B.A., Kennett, J.P., 1981. Phyletic gradualism in a Late Cenozoic planktonic foraminiferal lineage; DSDP Site 284, southwest Pacific. Paleobiology 7, 230–240.

Norris, R.D., 1991. Biased extinction and evolutionary trends. Paleobiology 17, 388–399.

Pearson, P.N., Ezard, T.H.G., 2014. Evolution and speciation in the Eocene planktonic foraminifer *Turborotalia*. Paleobiology 40, 130–143.

Pearson, P.N., Coxall, H.K., Stouge, S., 2014. Origin of the Eocene planktonic foraminifer *Hantkenina* by gradual evolution. Palaeontology 57, 243–267.

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

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. CRC press.

Wade, B.S., Olsson, R.K., 2009. Investigation of pre-extinction dwarfing in Cenozoic planktonic foraminifera. Palaeogeogr. Palaeoclimatol. Palaeoecol. 284, 39–46.

Wade, B.S., Pearson, P.N., Berggren, W.A., Pälike, H., 2011. Review and revision of Cenozoic tropical planktonic foraminiferal biostratigraphy and calibration to the geomagnetic polarity and astronomical time scale. Earth Sci. Rev. 104, 111–142.

Wei, K.-Y., 1994. Stratophenetic tracing of phylogeny using SIMCA pattern recognition technique: a case study of the Late Neogene planktic foraminifera *Globoconella* clade. Paleobiology 20, 52–65.

Wei, K.-Y., Kennett, J.P., 1988. Phyletic gradualism and punctuated equilibrium in the late Neogene planktonic foraminiferal clade *Globoconella*. Paleobiology 14, 345–363.