Integrative Organismal Biology

A Journal of the Society for Integrative and Comparative Biology

academic.oup.com/icb

OXFORD
UNIVERSITY PRESS

ARTICLE

Mammal Molar Size Ratios and the Inhibitory Cascade at the Intraspecific Scale

N. S. Vitek,[1],[*],[†] C. C. Roseman[‡] and J. I. Bloch[†]

*Department of Ecology & Evolution, Stony Brook University, 632 Life Sciences Building, Stony Brook, NY 11794-5245, USA; [†]Florida Museum of Natural History and Department of Biology, University of Florida, Gainesville, FL 32611, USA; [‡]Department of Evolution, Ecology, and Behavior, University of Illinois, Champaign, IL 61820, USA

[1]E-mail: natasha.vitek@stonybrook.edu

Synopsis Mammalian molar crowns form a module in which measurements of size for individual teeth within a tooth row covary with one another. Molar crown size covariation is proposed to fit the inhibitory cascade model (ICM) or its variant the molar module component (MMC) model, but the inability of the former model to fit across biological scales is a concern in the few cases where it has been tested in Primates. The ICM has thus far failed to explain patterns of intraspecific variation, an intermediate biological scale, even though it explains patterns at both smaller organ-level and larger between-species biological scales. Studies of this topic in a much broader range of taxa are needed, but the properties of a sample appropriate for testing the ICM at the intraspecific level are unclear. Here, we assess intraspecific variation in relative molar sizes of the cotton mouse, *Peromyscus gossypinus*, to further test the ICM and to develop recommendations for appropriate sampling protocols in future intraspecific studies of molar size variation across Mammalia. To develop these recommendations, we model the sensitivity of estimates of molar ratios to sample size and simulate the use of composite molar rows when complete ones are unavailable. Similar to past studies on primates, our results show that intraspecific variance structure of molar ratios within the rodent *P. gossypinus* does not meet predictions of the ICM or MMC. When we extend these analyses to include the MMC, one model does not fit observed patterns of variation better than the other. Standing variation in molar size ratios is relatively constant across mammalian samples containing all three molars. In future studies, analyzing average ratio values will require relatively small minimum sample sizes of two or more complete molar rows. Even composite-based estimates from four or more specimens per tooth position can accurately estimate mean molar ratios. Analyzing variance structure will require relatively large sample sizes of at least 40–50 complete specimens, and composite molar rows cannot accurately reconstruct variance structure of ratios in a sample. Based on these results, we propose guidelines for intraspecific studies of molar size covariation. In particular, we note that the suitability of composite specimens for averaging mean molar ratios is promising for the inclusion of isolated molars and incomplete molar rows from the fossil record in future studies of the evolution of molar modules, as long as variance structure is not a key component of such studies.

The study of mammalian tooth crowns is necessarily the study of interrelated modules that covary because of genetic, developmental, and functional relationships (van Valen 1962; Klingenberg 2008; Grieco et al. 2013). Individual traits on a tooth crown can not only covary with each other, but also with those of other teeth (van Valen 1962; Kangas et al. 2004; Carter and Worthington 2016). In this light, studying phenotypic covariation is a promising path forward to understanding the evolution of this complex set of biological structures (Kurtén 1967; Renaud et al. 2017; Delgado et al. 2019).

Within the mammalian molar module, one of the primary patterns of covariation is the relative sizes of each molar crown (Butler 1939; Kavanagh et al. 2007; Halliday and Goswami 2013; Evans et al. 2016; Hlusko et al. 2016). The inhibitory cascade model (ICM) proposes a developmental and genetic link underlying this suite of traits (Kavanagh et al. 2007). Under the strict form of the ICM as well as

proposed modifications to the model, molar crown size proportions are related to the self-regulation of activator and inhibitor molecules influencing the final size of sequentially developing tooth germs, resulting in a mesiodistal trade-off in size (Kavanagh et al. 2007; Young et al. 2015).

One confusing observation in previous studies is that the ICM does not appear to scale neatly with increasing biological scale (Kavanagh et al. 2007; Bernal et al. 2013; Roseman and Delezene 2019). The model explains variation at the intra-individual organ scale in developing tooth germ explants of *Mus musculus*. There, different amounts of exposure to M_1 inhibitor resulted in different sizes of M_2 and M_3 (Kavanagh et al. 2007). The difference between the resulting tooth size ratios between experimental treatments follows the predictions of the ICM (Kavanagh et al. 2007). At a larger clade scale, the ICM can describe differences between species in murid rodents, though it does not always adequately describe variation in other clades (Polly 2007; Labonne et al. 2012; Halliday and Goswami 2013; Carter and Worthington 2016; Couzens et al. 2016). These results imply that at an intermediate within-species or between-population scale relative molar sizes should also follow predictions of the ICM. Studies published thus far analyzing this biological scale in primates do not meet these predictions (Bernal et al. 2013; Roseman and Delezene 2019). However, it is unclear if primates in particular are poor models for studying the ICM or if the ICM fails at this intermediate scale more broadly across Mammalia.

The between-individual, between-population, and between-sister-species biological scales are critical for understanding the evolutionary processes underlying molar variation. Such studies must meet key sampling criteria. To study modern mammals, researchers must make new collections of large numbers of specimens or rely on the prescience of previous collectors for natural history museums to collect adequate samples of multiple populations of single species (Holmes et al. 2016). It is not yet clear what constitutes adequate sample size for intraspecific study of the ICM.

The fossil record is an additional, promising resource with which to study this biological scale because it is the only record capable of sampling the same species over long time scales (10^4–10^6 years). It also contains irreplaceable evidence of how this system might be affected by shifts in climate, ecology, and evolution, including intervals of mass extinction and major transitions in the evolution of clades. Mammalian molars make up a large component of

the fossil record, but additional challenges have prevented its use for this purpose. Samples of fossils are always time-averaged to some degree, and if too large a span of evolutionary change is averaged into a sample then results may not be valid (Barnosky 1990; Roy et al. 1996). Comparing variance of samples through the coefficient of variation (CV) is the primary method for assessing this possibility, but it is not clear if the rule of thumb that species CV < 10 applies to molar ratio values (Barnosky 1990; Plavcan and Cope 2001). In addition, fossils are usually incomplete and highly fragmented. Finding one complete molar row of a species is uncommon, but possible (Labonne et al. 2012; Halliday and Goswami 2013). Finding a large sample size of complete molar rows of a single species limits study to very common species or extraordinarily well-preserved fossil Lagerstätten (Briggs 2001; Lyson et al. 2019).

The purpose of this study is to characterize additional intraspecific properties of the mean, variance, and covariance of molar ratio phenotypes to facilitate future intraspecific study of the ICM. To achieve this goal, we document standing levels of variation and sensitivity to sample size in order to provide recommendations to future studies in terms of what kinds of samples and measurements might be appropriate for addressing the ICM at this scale. We also assess the amount of error introduced by creating point estimates of molar ratios from a composite of isolated molars. Although such a practice may seem ridiculous from the viewpoint of modern specimens where complete mandibles or hemi-mandibles are the norms rather than the exception, it would greatly benefit studies of the zooarcheological and paleontological records where large samples of complete molar rows of a single species are rare but large samples of isolated teeth are more common. Standard practice in previous paleontological studies has been to exclude samples of isolated molars due to concern of intraspecific variation (Halliday and Goswami 2013). That practice has limited other studies of the ICM in the fossil record in the past (Halliday and Goswami 2013; Schroer and Wood 2015), but the limitation may be more perceived than real. In particular, we compare this practice to the alternative approach of limiting studies to the small number (N < 5) of complete molar rows that might be available for any given species in the fossil record (Labonne et al. 2012). Future studies of molar proportions and their relationships to developmental and genetic modules require an evaluation of sample suitability for the research question. To aid in this assessment, we document standing levels of

variation in modern populations and species, as well as a test of the assumption that isolated molars may not provide an accurate estimate of population-level phenotype. By characterizing these ratio phenotypes in terms of their intraspecific variation and analyzing the degree to which sampling affects estimated properties of these phenotypes, we aim to help develop best practices for furthering our understanding of how developmental and genetic processes scale with changes in magnitude of evolution. To achieve this goal, we use samples from the published literature as well as a new sample of measurements of the cotton rat, *Peromyscus gossypinus*, which may be more likely to follow predictions of the ICM than the previously studied primates by virtue of its identity as a cricetid rodent. With these samples, we address four questions:

(1) Are particular measurement types better for studying molar size ratios in the ICM, as has been proposed (Hlusko et al. 2016)?

(2) What is the approximate amount of standing variation in molar size ratios at the population level and species level?

(3) How sensitive are estimates of the mean and error to changes in sample size?

(4) Does the use of composite molar rows instead of complete molar rows change the estimated amount of variation or mean ratio for a species or population?

Answers to these questions will help guide future research of covarying morphotypes, such as better understanding how well the ICM functions as a line of least resistance (Schluter 1996; Young et al. 2015). If the model meets the expectations of lines of least resistance, it may serve as a framework for connecting evolutionary observations at a microevolutionary (within species) scale to those at a macroevolutionary (between species or clades) scale. If it does not, then the ICM serves as an equally interesting avenue of research to understand how potentially disparate biological processes operating at different biological scales can produce similar morphological patterns.

Predictions of ICM

In its most specifically developed form, the ICM predicts the results of a relationship between activator and inhibitor where the effect of activator and inhibitor relative to each other is constant (Kavanagh et al. 2007; Young et al. 2015). That constant is summarized by the ratio of M_2 crown size relative to that of M_1 (Kavanagh et al. 2007; Hlusko et al.

2006). Across the molar row, the constant is a simplification of a linear relationship between activator and inhibitor described by the Equation $1+[(a - i)/i](x - 1)$, where a = activator, i = inhibitor, and x = tooth position (Young et al. 2015).

How best to measure tooth size is under some debate (Renvoisé et al. 2009; Hlusko et al. 2016; Gomes Rodrigues et al. 2017). A common, simple way to measure size is the rectangular estimate of crown area based on length and width (Wilson et al. 2012; Bernal et al. 2013; Halliday and Goswami 2013; Schroer and Wood 2015; Asahara et al. 2016; Evans et al. 2016; Roseman and Delezene 2019). More precise, outline-based measurements of two-dimensional crown area reduce the amount of error introduced into analyses to varying degrees (Kavanagh et al. 2007; Renvoisé et al. 2009; Carter and Worthington 2016; Gomes Rodrigues et al. 2017). However, quantifying size in terms of molar area measured in any way may necessarily subject the ICM to pleiotropic genetic influence of body size because of genetic correlations between buccolingual width and body size (Hlusko et al. 2006). A proposed more direct measure of a single genetic module that follows some ICM-like developmental pathways is the molar module component (MMC). The MMC is measured by mesiodistal length, which should follow the same limited sets of relationships among molar sizes as the ICM assuming that the length of the M_2 is predictable from any given M_1/M_3 ratio. This assumption holds for the primates in which it has been studied (Hlusko et al. 2016). If the MMC is a more precise way of capturing the ICM, length measurements should match ICM predictions more closely than area measurements for the same specimens because there should be reduced pleiotropy and therefore lower likelihood of overprinting of other genetic modules (Hlusko et al. 2016; Monson et al. 2019).

The ICM and possibly the MMC are each one specific model among a more general class of ascending-descending (AD) models in which "segments interact in a constant direction … but the magnitude of the effect varies between any given pair of segments," such that differences in phenotype are compounding and cumulative even if that the degree of compounding is non-linear across the tooth row (Young et al. 2015, 6). Put another way, teeth have shared covariance with a ratcheting effect (Kavanagh et al. 2007; Polly 2007). The space contained in the AD model space is mathematically equivalent to the "ICM-consistent" space of Polly (2007). As such, the AD model still largely applicable to most groups of extant mammals, including

metatherians with more than three molars (Polly 2007). It is appealing in that it provides a well-defined morphospace but the biological meaning is less clear than the ICM because the AD model lacks the explicit links to a describable activator-inhibitor ratio that gives the ICM so much predictive power. If a sample fits the AD model but not the ICM, it is difficult to determine whether a nonlinear effect of activator to inhibitor underlies the results, or if additional interactions help to determine the phenotype (Bernal et al. 2013; Carter and Worthington 2016; Roseman and Delezene 2019). More generally, a trade-off size cascade of the ICM and AD model have been proposed to be a developmental line of least resistance (Young et al. 2015). Whether general or specific, the value of this class of models is that they create a suite of predictions against which observations can be tested and used to learn about both evolutionary rules and their exceptions (Roseman and Delezene 2019).

Materials and methods

Materials

To evaluate patterns of variation within extant populations and species, we combined existing documentation of molar ratios and their variance with new measurements. Although there is a rich publication record of population- or species-level statistics on individual molar sizes, those records were not suitable because we could not estimate the variance structure of the ratios of molar sizes from the summary statistics themselves. Instead, we were able to use two types of literature-based data: (1) means and standard deviations (SDs) of molar size ratios for populations or (2) sizes of each of the three molars reported at the individual level for multiple individuals within a species. To be included in analyses of intraspecific variation, either kind of data type needed to be identified to the species level or below. Both one-dimensional length and two-dimensional area measurements were included. Length and area measurements were kept separated throughout the analytical process. Where sex and locality data were published, they were included. Fossilized specimens were removed to avoid circularity in results and the possibility of including over-averaged datasets (Labonne et al. 2012). A review of literature related to the inhibitory cascade recovered six publications that included sufficient data to evaluate intraspecific variation in molar ratios (Labonne et al. 2012; Asahara 2014, 2017; Asahara and Nishioka 2017; Monson et al. 2019; Roseman and Delezene 2019). The resulting combined datasets consisted of 1593

length entries for 47 species (Monson et al. 2019), 706 individual-level (Labonne et al. 2012; Roseman and Delezene 2019), and 72 population-level (Asahara 2014, 2017; Asahara and Nishioka 2017) area measurements for 35 species. The combined datasets are reported in Supplementary Tables S1 and S2.

The two datasets (length and area measurements) derived from the literature had little overlap at the species level and unknown overlap at the individual level because not all datasets included specimen numbers. In order to account for potential differences in results due to measurement dimensionality (length vs. area), we assembled a new dataset of each kind of measurement from identical specimens from a single species. These specimens all consisted of complete, erupted molar rows of the cricetid cotton mouse *P. gossypinus* that were opportunistically collected for another study. Although *P. gossypinus* is part of a different clade from the murid *M. musculus*, the species with which the ICM was originally developed, it is broadly morphologically and ecologically similar to *M. musculus* in terms of being a small-bodied myomorph rodent and a dietary generalist with third molars that are smaller than second molars, which are in turn smaller than first molars (Wolfe and Linzey 1977). *Peromyscus gossypinus* lacks a fourth premolar, unlike other cricetid rodents that do not adhere to predictions of the ICM potentially due to the influence of the deciduous fourth premolar tooth bud on the growth of the first molar (Viriot et al. 2002; Labonne et al. 2012). Based on these characteristics, we considered it reasonable to hypothesize that the molars of *P. gossypinus* would follow the predictions of the ICM. The dataset of *P. gossypinus* contains 70 specimens sampled from four states across the species' range (Table 1). The geographic sampling was intended to capture first-order variation between mainland phylogeographic groups within the species (Boone et al. 1999; Beckmann 2011). Specimens of both sexes were included and pooled into a single sample because *P. gossypinus* is not significantly sexually dimorphic in size and because sex data were not available for all specimens to test for potential effects of sexual dimorphism directly (Dewsbury et al. 1980).

Specimens were scanned using micro-computed tomography (μCT) on a Nikon XTH 225 ST at 9.99–15.74 μm voxel resolution, 100–140 kV voltage, 132–198 uA amperage, and exposure time of 500 ms across 1800 projections. Scans and 3D surfaces are reposited on MorphoSource (https://www.morphosource.org). Specimen numbers and MorphoSource media IDs are listed in Supplementary Table S3.

Table 1 Comparison of molar size ratios between samples of *P. gossypinus* from different localities across the range of the species

Sample	N	Measurement	Teeth compared	Median ratio value	LA	OK	TN	FL
Louisiana	20	Length	M_3/M_1	0.678	—	*1*	*1*	*1*
			M_2/M_1	0.801	—	*1*	*1*	*0.486*
		Area	M_3/M_1	0.596	—	*1*	*1*	*1*
			M_2/M_1	0.817	—	*1*	*0.718*	*1*
Oklahoma	20	Length	M_3/M_1	0.664	239	—	*1*	*1*
			M_2/M_1	0.803	225	—	*1*	*1*
		Area	M_3/M_1	0.604	197	—	*1*	*1*
			M_2/M_1	0.809	232	—	*0.114*	*1*
Tennessee	10	Length	M_3/M_1	0.665	123	102	—	*1*
			M_2/M_1	0.815	81	68	—	*0.076*
		Area	M_3/M_1	0.595	115	114	—	*1*
			M_2/M_1	0.857	64	47	—	*0.114*
Florida	20	Length	M_3/M_1	0.669	208	172	84	—
			M_2/M_1	0.789	265	225	156	—
		Area	M_3/M_1	0.591	215	209	82	—
			M_2/M_1	0.82	234	211	153	—

Upper triangle (italics) contains *P*-values from a pairwise Mann–Whitney U-test with Bonferroni correction. Lower triangle (plain type) contains U values from the same tests.

FL, Florida; LA, Louisiana; *N*, sample size; OK, Oklahoma; TN, Tennessee.

Three-dimensional surfaces of each molar row were created in Avizo 9.1.1 (FEI Visualization Science Group, Berlin), then measured in MeshLab v.2016.12 by two observers. Each measurement was taken in triplicate. Measurement error, including both error within and between observers, was evaluated using an analysis of variance (ANOVA) approach to calculate percent repeatability (Yezerinac et al. 1992; Sokal and Rohlf 1995; Roseman and Delezene 2019). Repeatability was, on average, between 93.1% for length measurements and 95.9% for area measurements. Measurement replicates were subsequently averaged into a single measurement for each variable for downstream analyses.

Phenotype choice

Two types of measurements have been proposed for comparing relative molar size, each with a slightly different biological interpretation (Kavanagh et al. 2007; Hlusko et al. 2016). The first type, molar crown area, was initially measured in developing teeth and a range of murid rodents, establishing a connection between developmental processes and phenotypic variation between species (Kavanagh et al. 2007). The simplest estimation of crown area takes the product of the maximum mesiodistal length and buccolingual width (Kavanagh et al. 2007; Bernal et al. 2013; Halliday and Goswami 2013; Schroer and Wood 2015; Asahara et al. 2016;

Evans et al. 2016; Roseman and Delezene 2019). This metric often overestimates two-dimensional surface area, but the degree to which this difference adds error to the results varies among clades (Hlusko et al. 2002; Kavanagh et al. 2007; Renvoisé et al. 2009; Carter and Worthington 2016). We chose to use the rectangular estimate of crown surface area because it allowed us to incorporate previously published measurements and use a standard metric to compare across a broader taxonomic dataset (Asahara 2014; Roseman and Delezene 2019). Unlike mammals, such as arvicoline rodents, for which the rectangular estimate is significantly problematic, none of the mammalian teeth we measured had high degrees of infolding or emargination along the margin, which increases the likelihood that a rectangular estimate is suitable in this case (Hlusko et al. 2002; Kavanagh et al. 2007).

The second suggested measurement type is the MMC or relative mesiodistal length of each molar crown without the incorporation of buccolingual width (Hlusko et al. 2016; Monson et al. 2019). Specifically, the metric focuses on the first and third molars with the assumption that the second molar occupies one-third of total molar row length (Hlusko et al. 2016). The MMC metric was developed in order to find a phenotypic trait that more directly captured a specific patterning mechanism, or a suite of covarying traits with a shared genetic basis

that is relatively free from indirect selective influences through pleiotropy (Hlusko et al. 2016). The exact genetic basis of the MMC and the exact connection between the MMC and the developmental processes of the ICM remain unclear (Hlusko et al. 2016; Monson et al. 2019).

Phenotype comparison

Comparison of the two metrics, area and length, derived from the same sample may help clarify whether the MMC is simply a more accurate form of the ICM phenotype or whether the two ratio patterns are influenced by related but distinct processes (Hlusko et al. 2016). To test the performance of length and area measurements in the *P. gossypinus* dataset against specific ICM predictions, we used reduced major axis (RMA) regression (Sokal and Rohlf 1995; Kavanagh et al. 2007) and Bayesian Markov Chain Monte Carlo (MCMC) model fitting (Hadfield 2010; Roseman and Delezene 2019). Note that we distinguish the ICM model from the broader AD model space (Polly 2007; Young et al. 2015), and do not test for fit within AD model space because of the limited developmental or modular inferences that can be drawn from patterns that fit the AD model (see Predictions of ICM). We tested the following specific predictions, developed based on previous study of the ICM:

(1) RMA regression between M_2 and M_3 size, when both are scaled to M_1 size, should explain a significant amount of variation within the sample (Kavanagh et al. 2007; Bernal et al. 2013).

(2) The same RMA regression should result in a linear model not significantly different from the equation $M3 = 2M_2 - M_1$ (Kavanagh et al. 2007; Young et al. 2015).

(3) The linear model of the ICM should be able to predict the ratio of M_3 to M_1 size based on the ratio of M_2 to M_1 size (Kavanagh et al. 2007; Polly 2007) following the equation:

$$\mu\left(\frac{M_3}{M_1}\right) = 2\mu\left(\frac{M_2}{M_1}\right) - 1 \qquad (1)$$

(4) The size of M_2 should be one-third the size of the complete molar row (Kavanagh et al. 2007; Young et al. 2015; Hlusko et al. 2016).

(5) Variance of M_3 size standardized to the total molar row size should equal the variance of similarly standardized M_1 size (Young et al. 2015).

(6) Variance–covariance structure of the molar sizes should be predictable through the following three equations (Roseman and Delezene 2019):

$$\sigma^2(M_3) = 4\sigma^2(M_2) + \sigma^2(M_1) - 4\sigma(M_1, M_2) \qquad (2)$$
$$\sigma(M_1, M_3) = 2\sigma(M_1, M_2) - \sigma^2(M_1) \qquad (3)$$
$$\sigma(M_2, M_3) = 2\sigma^2(M_2) - \sigma(M_1, M_2) \qquad (4)$$

To test Predictions 1 and 2, we used the package *lmodel2* to perform RMA regression and estimate a confidence interval (CI) for the slope and intercept of the model using permutation. To test Predictions 3–6, we fit multiple linear response models using the *MCMCglmm* package (Hadfield 2010). Following previous use of these Bayesian MCMC model fitting (Roseman and Delezene 2019), all traits were modeled as Gaussian, and for all traits, priors were set to be a non-informative inverse Wishart prior for residual covariance with covariances set to zero. Burn-in was set to five hundred thousand iterations of the Markov chain. After discarding the burn-in, we sampled every thousandth iteration 1000 times, for a total of 1000 samples over a period of 1 million iterations (Roseman and Delezene 2019). The 1000 samples form a posterior distribution of estimates of the true molar size, relative sizes, variances, and covariances within *P. gossypinus*. In order to account for measurement error within these models, we multiplied the variance within the resulting posterior sample by the repeatability for the appropriate measurement.

To compare the posterior distribution of observed population values to the theoretical values predicted by the ICM, we used the posterior distribution and mathematical expectations of Predictions 3–6, including Equations (1)–(4) to create theoretical posterior distributions. For example, we applied the posterior distribution of M_2/M_1 sizes to Equation (1) to create a distribution of predicted M_3/M_1 sizes, then compared that predicted theoretical distribution to the observed distribution of M_3/M_1 sizes, then compared the theoretical and observed predictions. To make these comparisons in a standardized system that can be compared with results from other studies, we divided each observed value by its theoretical value, then took the mode and 95% highest posterior density (HPD) intervals for each parameter (Roseman and Delezene 2019). If the 95% HPD interval encompasses one, then we consider the prediction to be consistent with the ICM because there is no difference between the theoretical and observed values. The parameters we analyzed in this way are mean M_3 size as predicted by the ICM linear predictor (Prediction 3), the mean M_2 size relative to the

total molar row (Prediction 4), variance of M_3 relative to the total molar row (Prediction 5), variance of M_3 (Prediction 6), covariance between M_1 and M_3 (Prediction 6), and covariance between M_2 and M_3 (Prediction 6).

Length and area measurements were evaluated separately for each prediction. If either length or area fit model predictions better, we used that single measurement type in downstream analyses. If both or neither was a good fit, then we continued to analyze both measurements downstream to allow the greatest amount of comparison with past and future studies.

Levels of standing variation

To describe and compare standing levels of variation in molar proportions, we used the CV (Simpson 1944; Gingerich 1974). CVs are less commonly calculated for ratio values, in part because it is difficult to interpret any comparison between CVs based on different traits with different variational properties (Pearson 1897; Atchley et al. 1976; Lande1977). However, comparison of CVs calculated from the same trait appears to be a reasonable use of the statistic (Lande 1977). In this study, we do not directly compare CVs across different scales (e.g., length vs. area) or between different traits. Factors such as sexual dimorphism and geographic variation can inflate levels of total-species variation and become conflated with other sources of variation (Plavcan and Cope 2001). However, the nature of most of the data used in this study did not allow for testing of whether species-level pooling results in greater variation than sampling sexes or regions separately. There were two exceptions in our dataset, Roseman and Delezene's (2019) tooth measurements of apes and the tooth measurements of *P. gossypinus* generated for this study.

In the two cases where we could test for potential inflation of variance due to pooling, we used the sign test to investigate whether the CV of each locality or sex subsample was less than the respective CV for the pooled, species-level sample. In order to further determine whether subsamples could be pooled at the species level, we performed a pairwise Mann–Whitney U-test between samples from different states or of different sexes (Asahara 2014). Pairwise tests were corrected using the Bonferroni correction. If populations of *P. gossypinus* were not significantly different from one another, then the sample was pooled at the species level for downstream analysis. If ratio values were significantly different between sexes of primate species, sexes were treated as

separate samples. Where different summary values were recorded for different localities, localities were treated as separate populations (Asahara 2014). Where no locality or sex data were reported for samples, a species was pooled into a single sample. Summary statistics for each applicable ratio value for each sample were calculated and are provided in Supplementary Tables S1 and S2.

Simulating sample sizes

In order to evaluate the minimum sample size necessary to estimate mean and variance structure of molar proportions, we first resampled the *P. gossypinus* dataset without replacement to simulate different sample sizes $2 < N < 70$. Ten thousand resampled pseudoreplicates were created at each N, and mean and variance of each ratio were calculated from each pseudoreplicate. We considered a sample size adequate if 95% of pseudoreplicates produced a mean and variance that fell within the 95% CI of the complete $N = 70$ sample. Based on the results from *P. gossypinus*, we checked to see if any other species in the published datasets had likely been sampled enough to estimate true mean and variance. For those species, we performed the same resampling analysis in order to check for consistency of results across species.

Simulating use of composite molars

To assess the accuracy of using isolated molars to calculate molar ratios, we used the pooled sample of *P. gossypinus*. To simulate the error resulting from the creation of composite molar rows from averages of isolated teeth, we created 50,000 datasets of simulated samples of isolated teeth. To create each simulated sample, we resampled with replacement N teeth of the same tooth position from any of the 70 specimens of *P. gossypinus*, where N is the sample size of a single tooth position, ranged from 1 to 50. In order to create a sample-level point estimate of molar size ratios, we took the mean of each of those N specimens of each molar position to calculate M_1, M_2, and M_3 crown areas constituting a single composite molar row. For example, a simulation at $N = 2$ would take an average of two "isolated" M_1s, an average of two "isolated" M_2s, and an average of two "isolated" M_3s, then calculate a size ratio of the three means.

To estimate variation for each simulated sample, we further resampled one molar from each tooth position in the simulated sample and used that set of three teeth to calculate ratios. We repeated that process 1000 times to create a "pseudosample"
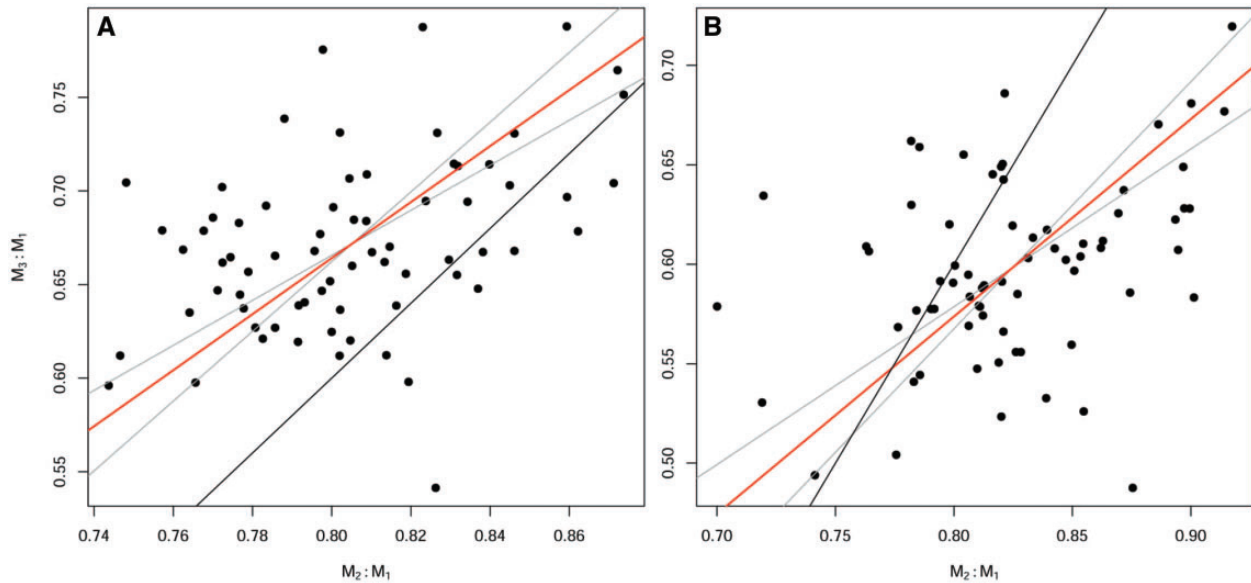
Fig. 1 Plot of relative molar size of specimens of *P. gossypinus* in terms of (**A**) length and the MMC; (**B**) area and the ICM. Red line surrounded by gray CIs indicates the best fit RMA regression line explaining variance in the data. Black line indicates the regression line predicted by the MMC and ICM, respectively.

distribution of ratios around the mean value for the composite tooth row, then calculated the SD of that pseudosample. The entire process repeated 999 more times to produce 1000 estimates of the mean and SD of molar ratio values sampled at N teeth per tooth position. This process mimics that resampling methods that would be available to paleontologists trying to estimate a CI based on a collection of N isolated teeth. We compared those estimates to the 95% CI interval of the true mean and SD of ratios calculated from complete molar rows of *P. gossypinus*. If <5% of the 1000 estimated means for a given N fell outside of those CIs, then we considered a composite sample of at least N teeth per tooth position to adequately approximate a point estimate of the mean and SD.

All analyses were conducted in R 3.5.3 (R Core Team 2015). All resampling was conducted with 1000 pseudoreplicates (Kowalewski and Novack-Gottshall 2010). Code is reposited on GitHub (project name "molar_ratio_sampling") with a snapshot archived on FigShare (doi: 10.6084/m9.figshare.12597596).

Results

Length or area

Both length and area measurements met few of the predictions of the ICM (Figs. 1 and 2). Although the cluster of values for specimens overlap with the ICM linear model, sample-specific linear models accounted for relatively small amounts of variation between molar ratios (Fig. 1, Table 2). CIs of slopes and intercepts for those sample-specific models did not include values matching the ICM.

For length measurements of *P. gossypinus*, the only parameter that met ICM expectations was covariance between size of M_1 and M_3, although the credible interval is so wide as to make this result meaningless (Fig. 2A, Supplementary Tables S4 and S5). Variances and covariances between raw lengths were lower than predicted by the ICM, but relative variance of M_3 size was higher than expected. That is, its standardized variance was greater than the standardized variance of M_1 size. The size of M_3 was larger than expected, and concordantly the size of the M_2 in the context of the complete molar row was smaller than expected.

Results for molar crown area measurements were qualitatively similar to length measurements in terms of variance and covariance structure between raw measurements (Fig. 2B). One major difference is that relative variance of M_3 size met expectations of the ICM. In terms of means of raw and relative size, M_3 was too small and M_2 was too large relative to the size of the molar row, opposite the results for length measurements of the same teeth.

Standing variation in proportions

For all ratios, there was no difference in proportions between any of the locality-level samples of *P. gossypinus* (Table 1, $P > 0.114$) and CV of a pooled sample was not consistently higher than those of locality-

Table 2 Comparison of molar size ratios between samples of *P. gossypinus* from different localities across the range of the species

Measurement	Slope				Intercept			
	Value	95% CI		Prediction	Value	95% CI		Prediction
		Lower	Upper			Lower	Upper	
Length	1.497	1.203	1.863	2	−0.533	−0.828	−0.297	−1
Area	0.993	0.793	1.242	2	−0.22	−0.426	−0.056	−1

Prediction refers to the values predicted if sample met the expectations of the MMC (length) or ICM (area).
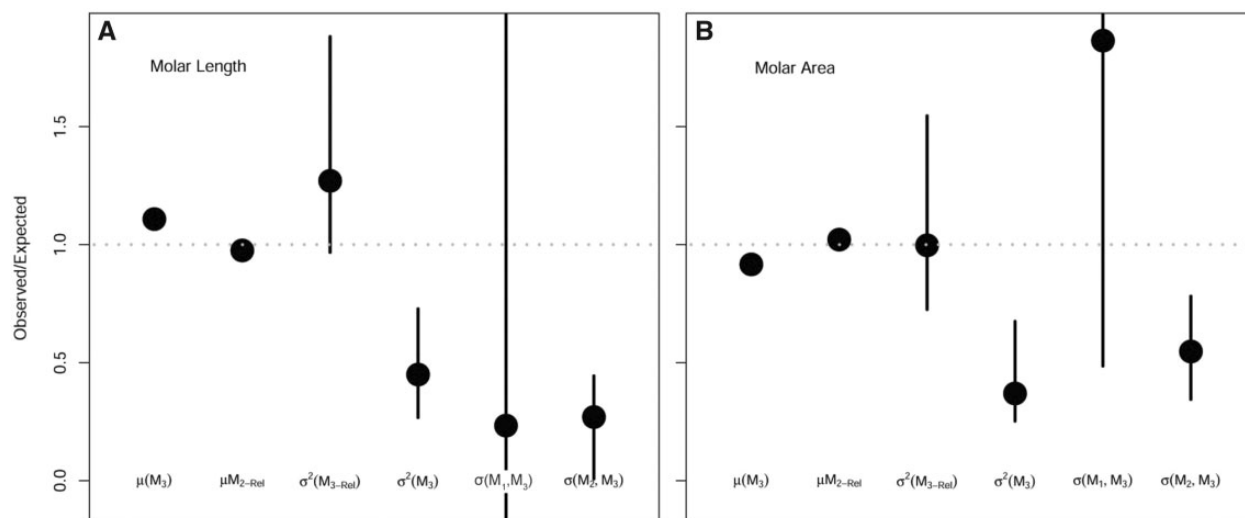

Fig. 2 Plot comparing (**A**) MMC and (**B**) ICM predictions to direct estimates of six variables based on a sample of *P. gossypinus* following Roseman and Delezene (2019). The *y* axis measures the proportion of direct estimates fulfilled by predictions. Values higher than one are over-predicted by models. Values less than one are under-predicted. Black points indicate the mode of the posterior distribution and black lines indicate the 95% posterior credible interval. Values on the *x* axis, from left to right, indicate mean M_3 size ($\mu[M_3]$), mean size of M_2 as a proportion of total molar row size (μM_{2-Rel}), variance of relative M_3 size ($\sigma^2[M_{3-Rel}]$), variance of the absolute M_3 size ($\sigma^2[M_3]$), covariance of M_1 and M_3 ($\sigma[M_1, M_3]$), covariance of M_2 and M_3 ($\sigma[M_2, M_3]$).

level samples (Fig. 3A, $P > 0.5982$). The specimens were subsequently pooled into a single sample for further analysis. Primates had significantly different ratio values by sex (Table 3), though CVs of pooled samples were not consistently higher than single-sex samples (Fig. 3B, $P > 0.298$). Primate data from Roseman and Delezene (2019) were separated by sex in downstream analyses.

Ratios measuring relative M_3 area were notably high in Carnivora, but these samples all represent populations of the raccoon dog *Nyctereutes procyonoides* in which some ratio values are 0 because the M_3 is absent (Fig. 4, Supplementary Table S1; Asahara 2014). When samples of *N. procyonoides* are excluded from the dataset, CVs of ratios from area measurements are lower and approximately equivalent to values derived from length measurements and CV was generally similar across species (Fig. 4; Supplementary Table S2).

Suitable sample size

Estimates of mean ratios values were less sensitive to sample size than SD of the same ratios in *P. gossypinus* (Fig. 5, Supplementary Table S6). At $N = 2$ specimens or greater, 95% of sample means was within the HPD of the mean of the species for all ratio measurements. SD was more sensitive to sample size, with average SD at low sample sizes ($N < 3$) underestimating the true SD (Fig. 5). Depending on ratio type, at least $N = 33$–42 specimens were necessary to estimate an SD that was likely to fall within the 95% HPD of the true SD.

Eighteen other samples had at least 43 specimens for length measurements and three samples had at least 43 specimens for area measurements. Applying the same analyses to these samples yielded similar results, although five samples required 50 or more individuals before SD were likely to fall within the 95% HPD (Supplementary Table S7).
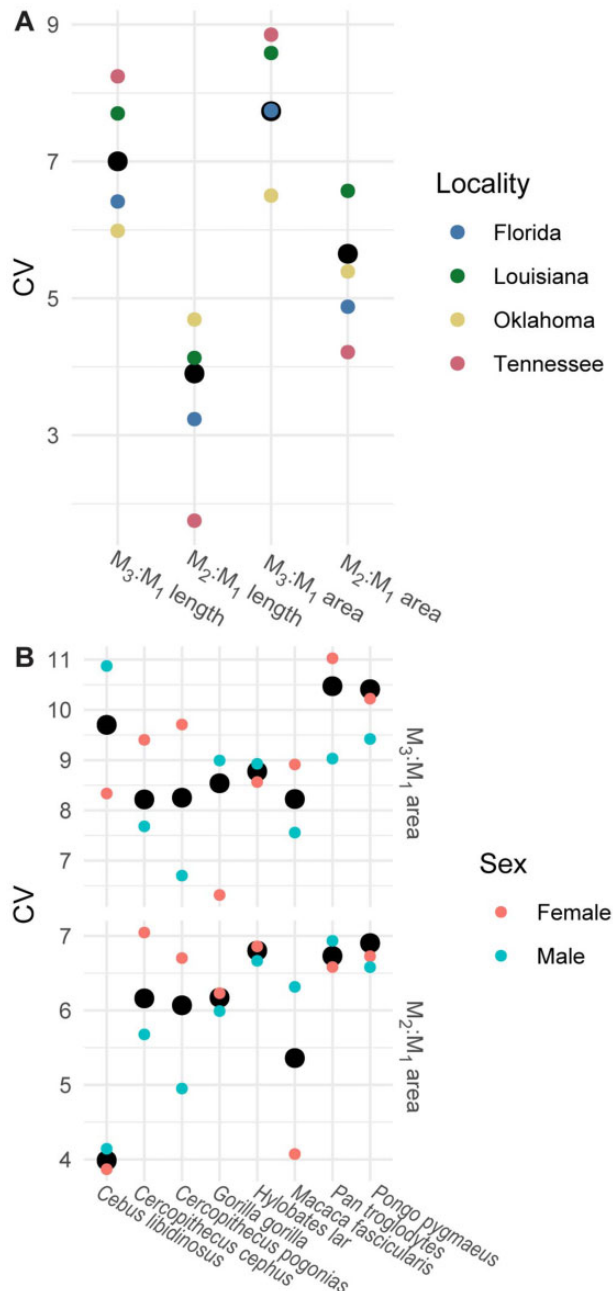
	Female		Male			
$M_3:M_1$ area	N	Median	N	Median	P	U
Cebus libidinosus libidinosus	30	0.589	34	0.58	0.888	521
Cercopithecus cephus cephus	23	1.168	48	1.162	0.869	566
Cercopithecus pogonias grayi	27	1.27	32	1.242	0.322	498
Gorilla gorilla gorilla	32	1.149	45	1.206	0.005	448
Hylobates lar carpenteri	16	1.122	16	1.115	0.539	111
Macaca fascicularis fascicularis	28	1.514	33	1.576	0.177	368
Pan troglodytes troglodytes	43	1.013	31	1.074	0.041	480
Pongo pygmaeus	16	1.015	16	1.116	0.032	71
$M_2:M_1$ area						
Cebus libidinosus libidinosus	30	0.884	34	0.879	0.995	511
Cercopithecus cephus cephus	23	1.295	48	1.278	0.483	610
Cercopithecus pogonias grayi	27	1.377	32	1.345	0.052	560
Gorilla gorilla gorilla	32	1.238	45	1.253	0.035	516
Hylobates lar carpenteri	16	1.131	16	1.173	0.184	92
Macaca fascicularis fascicularis	28	1.387	33	1.356	0.504	509
Pan troglodytes troglodytes	43	1.112	31	1.134	0.3	571
Pongo pygmaeus	16	1.117	16	1.168	0.08	81

U- and P-values come from pairwise Mann–Whitney U-tests with Bonferroni correction.

Fig. 3 CV of samples of (**A**) *P. gossypinus* from individual localities (small colored circles) or (**B**) species of primate subdivided into sexes (small colored circles) in comparison to a pooled sample (large black circles).

Use of composite molar rows

Mean length ratio values calculated from composite molar rows were within the 95% CI of the true mean 95% of the time even at sample size of $N = 2$ (Fig. 6, Table 4). At $N = 4$ or greater, 95% of estimated means was within the 95% CI of the true mean regardless of measurement type. The true variance was

not recovered in composite samples regardless of sample size. Estimated variance was frequently greater than the true variance, especially as sample size increased (Fig. 6, Table 4).

An alternative way of visualizing these results is to compare the distribution of the composite-based estimates of mean ICM values to the distribution of complete specimens in ratio ICM ratio space (Fig. 7). At $N = 1$, a collection of composite estimates occupies a greater range of morphospace than the true population sample (Fig. 7). At $N = 5$, the composite estimate is already notably reduced to approximately the extent of the true population range, and by $N = 10$ the range of estimates of the mean was more precise than an estimate than many estimates would be based on a single, complete molar row.

Discussion

Our intraspecific sample of *P. gossypinus* failed to meet the predictions of the ICM, similar to previous findings in primates. The choice of measuring molar size as a length or an area did not affect prediction fit. Overall, the two measurements provided similar results, but differed in details that might lead to
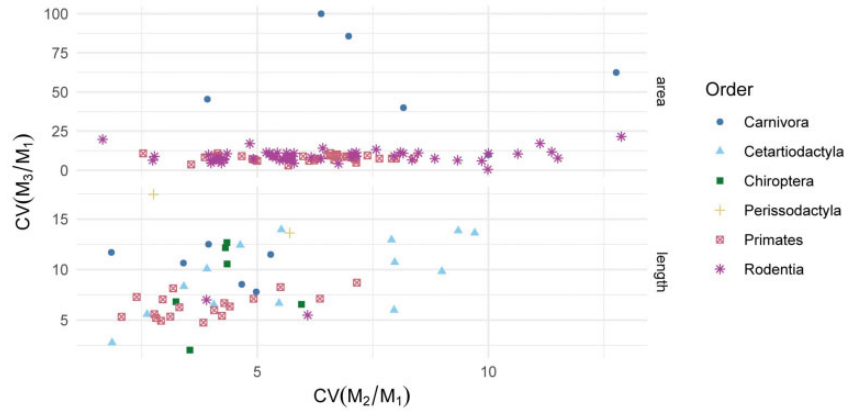
Fig. 4 CVs of the ratios calculated to study the inhibitory cascade (top) or MMC (bottom).
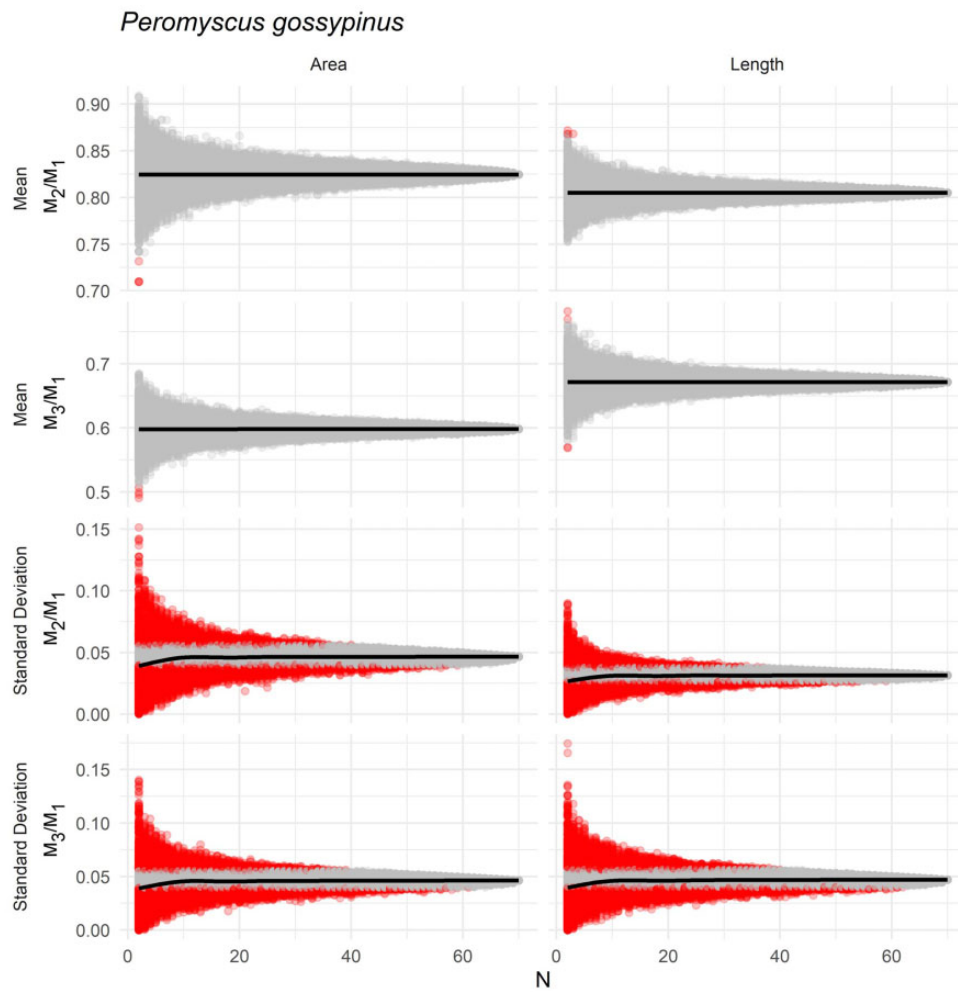

Fig. 5 Relationship between sample size and estimates of mean and SD of ratios calculated to study the ICM (area) or MMC (length) based on resampling complete molar rows of *P. gossypinus*. Each gray point is one of 10,000 pseudoreplicates sampled at *N* sample size. Black line indicates the mean value calculated from pseudoreplicates. Red points indicate pseudoreplicate samples whose mean or SD is outside the 95% CI around the observed value calculated for the complete $N = 70$ sample.

Table 4 Proportion of samples of composite molar rows that resulted in statistics outside of 95% highest posterior density (HPD)

	Area				Length			
	M_2/M_1		M_3/M_1		M_2/M_1		M_3/M_1	
N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	0.237	1	0.148	1	0.117	1	0.04	1
2	0.108	0.813	0.066	0.787	0.055	0.78	0.005	0.834
3	0.066	0.866	0.034	0.75	0.036	0.738	0	0.705
4	0.05	0.927	0.012	0.755	0.011	0.74	0	0.668
5	0.027	0.959	0.011	0.829	0.009	0.792	0	0.577
6	0.015	0.981	0.007	0.854	0.006	0.835	0	0.521
7	0.008	0.989	0.006	0.884	0.001	0.858	0	0.469
8	0.003	0.992	0.001	0.914	0.001	0.895	0	0.427
9	0.005	0.996	0	0.92	0	0.892	0	0.421
10	0.002	0.999	0	0.945	0	0.911	0	0.399
11	0.001	0.999	0	0.947	0	0.94	0	0.362
12	0.001	1	0	0.964	0	0.947	0	0.309
13	0.003	1	0	0.968	0	0.944	0	0.311
14	0	1	0	0.977	0	0.958	0	0.3
15	0.001	1	0	0.981	0	0.96	0	0.268
16	0	1	0	0.983	0	0.969	0	0.25
17	0	1	0	0.982	0	0.982	0	0.236
18	0	1	0	0.993	0	0.979	0	0.21
19	0	1	0	0.991	0	0.983	0	0.229
20	0	1	0	0.993	0	0.989	0	0.207
21	0	1	0	0.992	0	0.99	0	0.183
22	0	1	0	0.995	0	0.989	0	0.162
23	0	1	0	0.999	0	0.997	0	0.159
24	0	1	0	1	0	0.993	0	0.14
25	0	1	0	0.997	0	0.996	0	0.149
26	0	1	0	0.999	0	0.998	0	0.14
27	0	1	0	0.998	0	0.997	0	0.113
28	0	1	0	0.999	0	0.996	0	0.129
29	0	1	0	1	0	0.997	0	0.111
30	0	1	0	1	0	0.999	0	0.106
31	0	1	0	1	0	0.999	0	0.101
32	0	1	0	1	0	0.998	0	0.096
33	0	1	0	1	0	0.997	0	0.096
34	0	1	0	1	0	0.999	0	0.087
35	0	1	0	0.998	0	1	0	0.064
36	0	1	0	0.999	0	1	0	0.081
37	0	1	0	1	0	0.999	0	0.074
38	0	1	0	1	0	0.999	0	0.075
39	0	1	0	1	0	1	0	0.076
40	0	1	0	1	0	1	0	0.061
41	0	1	0	1	0	0.999	0	0.071
42	0	1	0	1	0	1	0	0.056

Table 4 Continued

	Area				Length			
	M_2/M_1		M_3/M_1		M_2/M_1		M_3/M_1	
N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
43	0	1	0	1	0	1	0	0.065
44	0	1	0	1	0	0.999	0	0.062
45	0	1	0	1	0	1	0	0.058
46	0	1	0	1	0	1	0	0.043
47	0	1	0	1	0	1	0	0.047
48	0	1	0	1	0	1	0	0.042
49	0	1	0	1	0	1	0	0.051
50	0	1	0	1	0	1	0	0.057

N, the number of specimens sampled to calculate average size of an individual molar in composite row.

different conclusions about why the observed distribution of ratios did not fit the ICM. For example, if we are using tooth crown area (Fig. 2A), we might look for a reason why M_3 size was smaller than expected, but if using tooth crown length we would look for a cause of an apparently opposite result of M_3 size larger than expected (Fig. 2B). These results alone do not clarify which measurement is preferable.

Although these ratio phenotypes cannot be fully explained by the ICM, they show promise for future studies addressing the source and history of their covariation. Use of relatively small samples of either complete or composite molar rows is adequate to characterize average molar ratios. Rare or fragmentary records of species, extant or extinct, may be more suitable for study of the molar module than previously considered. Mean ratio values are much less sensitive to sample size than estimates of variance, matching general statistical observations (Gotelli and Ellison 2004).

Assembling datasets with the requisite number of specimens to study variance structure will likely require pooling multiple localities and potentially sexes. Pooling increases the likelihood that additional sources of variation, such as sexual dimorphism and genetic differentiation, may influence results (Albrecht and Miller 1993). For example, some species have documented differences in molar ratio values between populations (Asahara 2014), and these differences may influence estimates of variance structure within the species. To address whether pooling may be the source of increased variability, the standard variability metric CV is often calculated and compared against a standard (Barnosky 1990; Plavcan and Cope 2001). Standing species-level CV

for tooth sizes, both in terms of lengths and areas, are already well documented (Wallace 1968; Gingerich and Ryan 1979; Gingerich and Winkler 1979; Sakai 1981; Williamson 2001; Tornow et al. 2006; Natsume et al. 2008). Based on such documentation, an accepted diagnostic principle has emerged that measurements of size of molars themselves are variable within a species up to about CV = 10% or less (Gingerich 1981; Plavcan and Cope 2001). Higher variability is usually interpreted as a sign of over-pooling or the presence of more than one species in the sample (Plavcan and Cope 2001). Variability in ratios of molar sizes is less well understood, and based on results of this study the same recommended standard for molar sizes too low for molar ratio values. Ratios of molar size appear to be more variable than the sizes of the molars themselves, probably because of the way in which error propagates through the calculations.

Recommendations for future studies

Based on the results of this study, a minimum of 40–50 specimens from a single species is required to accurately study variance structure of relative molar sizes. A far smaller sample, as small as $N = 2$ is adequate for characterizing the mean relative sizes of molars in complete tooth rows. Use of averages of samples of isolated molars to calculate molar ratios within ICM is acceptable with sample sizes of at least four specimens per tooth position, assuming the samples are not overly time-averaged for the question of interest (Kowalewski and Bambach 2008). We consider a CV >15% for either the M_3:M_1 or M_2:M_1 ratio in a sample of pooled specimens an indicator that the sample may be overly averaging in some way, such as over time, space, or sex. This
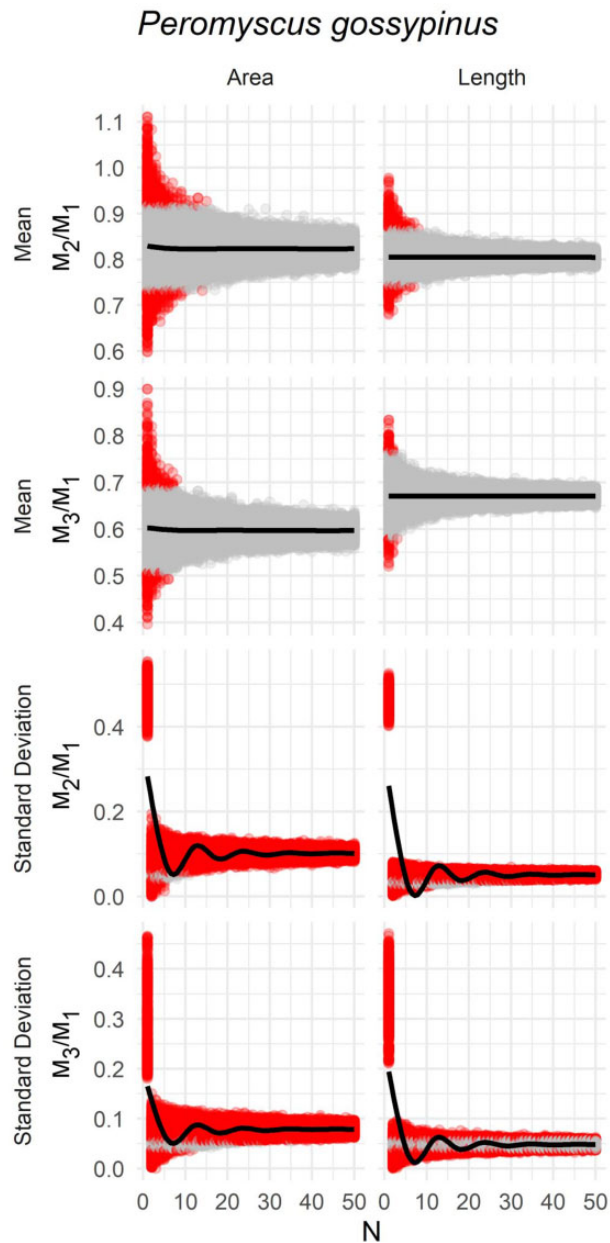
Fig. 6 Relationship between sample size and estimates of mean and SD of ratios calculated to study ICM (area) or MMC (length) based on resampling individual molars from molar rows of *P. gossypinus* to make composite molar rows. Each gray point is one of 10,000 pseudoreplicates sampled at N sample size. Black line indicates the mean value calculated from pseudoreplicates. Red points indicate pseudoreplicate samples whose mean or SD is outside the 95% CI around the observed value calculated for the complete $N = 70$ sample.

Fig. 7 Comparison of occupation of ICM (area) or MMC (length) ratio morphospace by complete versus simulated molar rows derived from the same sample of *P. gossypinus*. Large red squares represent 70 complete molars rows. Small blue circles represent point estimates of ratios calculated by taking the mean of 1, 2, 5, or 10 isolated molars per tooth position per pseudosample.

approach requires a careful choice of how samples are pooled into groups, as well as a consideration of variance in the sample (Kidwell and Holland 2002), but with those caveats the practice allows the ICM to be studied in systems that were previously considered unapproachable. In cases where tooth rows are frequently incomplete, such as in the fossil record,

bootstrap resampling of collections of at least five molars per tooth position can produce CIs that are comparable to the natural variability within a species (Fig. 7). Isolated molars are not suitable for estimating variance structure of ratio values.

Remaining challenges to understanding the ICM

The ICM has been proposed to be a line of developmental least resistance in evolution (Young et al.

2015). Such a line, if comparable, should behave similarly to a genetic line of least resistance, or a linear model of multiple traits that describes the maximum vector of genetic variation within a population (Schluter 1996). Genetic lines of least resistance scale from populations to species, such that patterns of evolution between closely related species bear resemblance to primary patterns of variation within a single species, and statistically speaking the vectors of maximum variance between the two scales of evolution covary (Schluter 1996; Marroig and Cheverud 2010). The behavior of the ICM is unusual from this perspective in that even though the vector of maximum ratio variance between an inbred laboratory strain of *M. musculus* is statistically indistinguishable from a between-species vector of maximum ratio variance in murid rodents, variation documented in this study and others at the intermediate level of standing variation in natural populations does not follow the same vector (Roseman and Delezene 2019).

One necessary test unable to be conducted in this study is whether variation in wild population of *M. musculus* itself, the original subject of ICM studies, follows the ICM. Results of that test would help establish whether the ICM truly does not hold at intermediate biological levels, contrary to the concept of a line of least resistance, or if a wider range of species than previously appreciated do not follow the ICM. Notably, the sample of *P. gossypinus* occupies the same region of ICM morphospace as the ICM linear model, meaning that if the species had been studied in terms of a species-level means it would be consistent with the results of Kavanagh et al. (2007).

The ability to incorporate a greater proportion of the fossil record into studies of relative molar sizes may help explain how certain clades came to be "exceptions" to the ICM. For example, extant equids and ursids not only do not follow the ICM, but also fail to meet the predictions of more general AD models (Polly 2007). Genetic hypotheses have been proposed to explain this deviation in ursids (Asahara et al. 2016) but fossils of extinct stem and crown-ursids may be useful to understand when a shift away from AD model space occurred, what phenotypes the shift is associated with, and what the timing of such shifts can explain about overcoming developmental constraints of patterning mechanisms.

Acknowledgments

Funding

Supplementary data

Supplementary data available at *IOB* online.

References

Albrecht GH, Miller JMA. 1993. Geographic variation in primates: a review with implications for interpreting fossils. In: Kimbel WH, Martin LB, editors. Species, species concepts, and primate evolution. New York (NY): Plenum Press. p. 123–61.

Asahara M. 2014. Evolution of relative lower molar sizes among local populations of the raccoon dog (*Nyctereutes procyonoides*) in Japan. Mamm Study 39:181–4.

Asahara M. 2017. Geographic variation of absolute and relative lower molar sizes in two closely related species of Japanese field mice (*Apodemus speciosus* and *Apodemus argenteus*: Muridae, Rodentia). Zoolog Sci 34:26–34.

Asahara M, Nishioka Y. 2017. Geographic variation of absolute and relative lower molar sizes in the Japanese Macaque (*Macaca fuscata*: Primates, Mammalia). Zoolog Sci 34:35–41.

Asahara M, Saito K, Kishida T, Takahashi K, Bessho K. 2016. Unique pattern of dietary adaptation in the dentition of Carnivora: its advantage and developmental origin. Proc R Soc Lond B Biol Sci 283:20160375.

Atchley WR, Gaskins CT, Anderson D. 1976. Statistical properties of ratios. I. Empirical results. Syst Zool 25:137–48.

Barnosky AD. 1990. Evolution of dental traits since latest Pleistocene in meadow voles (*Microtus pennsylvanicus*) from Virginia. Paleobiology 16:370–83.

Beckmann SM. 2011. The effects of climatic and geographic events on the cotton mouse (*Peromyscus gossypinus*). [dissertation]. [Coral Gables (FL)]: University of Miami. p. 274.

Bernal V, Gonzalez PN, Perez SI. 2013. Developmental processes, evolvability, and dental diversification of New World Monkeys. Evol Biol 40:532–41.

Boone JL, Smith MH, Laerm J. 1999. Allozyme variation in the Cotton Mouse (*Peromyscus gossypinus*). J Mammal 80:833–44.

Briggs DEG. 2001. Exceptionally preserved fossils. In: Briggs DEG, Crowther PR, editors. Palaeobiology II. Malden (MA): Blackwell Publishing. p. 328–32.

Butler PM. 1939. Studies of the mammalian dentition.– Differentiation of the post-canine dentition. Proc Zool Soc Lond B 109:1–36.

Carter KE, Worthington S. 2016. The evolution of anthropoid molar proportions. BMC Evol Biol 16:110.

Couzens AMC, Evans AR, Skinner MM, Prideaux GJ. 2016. The role of inhibitory dynamics in the loss and reemergence of macropodoid tooth traits. Evolution 70:568–85.

Delgado MN, Pérez-Pérez A, Galbany J. 2019. Morphological variation and covariation in mandibular molars of platyrrhine primates. J Morphol 280:20–34.

Dewsbury DA, Baumgardner DJ, Evans RL, Webster DG. 1980. Sexual dimorphism for body mass in 13 taxa of muroid rodents under laboratory conditions. J Mammal 61:146–9.

Evans AR, Daly ES, Catlett KK, Paul KS, King SJ, Skinner MM, Nesse HP, Hublin J-J, Townsend GC, Schwartz GT, et al 2016. A simple rule governs the evolution and development of hominin tooth size. Nature 530:477–80.

Gingerich PD. 1974. Size variability of the teeth in living mammals and the diagnosis of closely related fossil species. J Paleontol 48:895–903.

Gingerich PD. 1981. Variation, sexual dimorphism, and social structure in the early Eocene horse Hyracotherium (Mammalia, Perissodactyla). Paleobiology 7:443–55.

Gingerich PD, Ryan AS. 1979. Dental and cranial variation in living Indriidae. Primates 20:141–59.

Gingerich PD, Winkler DA. 1979. Patterns of variation and correlation in the dentition of the red fox, Vulpes vulpes. J Mammal 60:691–704.

Gomes Rodrigues H, Lefebvre R, Fernández-Monescillo M, Quispe BM, Billet G. 2017. Ontogenetic variations and structural adjustments in mammals evolving prolonged to continuous dental growth. R Soc Open Sci 4:170494.

Gotelli NJ, Ellison AM. 2004. A primer of ecological statistics. Sunderland (MA): Sinauer Associates.

Grieco TM, Rizk OT, Hlusko LJ. 2013. A modular framework characterizes micro- and macroevolution of old world monkey dentitions. Evolution 67:241–59.

Hadfield JD. 2010. MCMC Methods for multi-response beneralized linear mixed models: the MCMCglmm R package. J Stat Softw 33:1–22.

Halliday TJ, Goswami A. 2013. Testing the inhibitory cascade model in Mesozoic and Cenozoic mammaliaforms. BMC Evol Biol 13:79.

Hlusko LJ, Lease LR, Mahaney MC. 2006. Evolution of genetically correlated traits: tooth size and body size in baboons. Am J Phys Anthropol 131:420–7.

Hlusko LJ, Schmitt CA, Monson TA, Brasil MF, Mahaney MC. 2016. The integration of quantitative genetics, paleontology, and neontology reveals genetic underpinnings of primate dental evolution. Proc Natl Acad Sci U S A 113:9262–7.201605901.

Hlusko LJ, Weiss KM, Mahaney MC. 2002. Statistical genetic comparison of two techniques for assessing molar crown size in pedigreed baboons. Am J Phys Anthropol 117:182–9.

Holmes MW, Hammond TT, Wogan GOU, Walsh RE, LaBarbera K, Wommack EA, Martins FM, Crawford JC, Mack KL, Bloch LM, et al 2016. Natural history collections as windows on evolutionary processes. Mol Ecol 25:864–81.

Kangas AT, Evans A, Thesleff I, Jernvall J. 2004. Nonindependence of mammalian dental characters. Nature 432:211–4.

Kavanagh KD, Evans AR, Jernvall J. 2007. Predicting evolutionary patterns of mammalian teeth from development. Nature 449:427–32.

Kidwell SM, Holland SM. 2002. The quality of the fossil record: implications for evolutionary analyses. Annu Rev Ecol Syst 33:561–88.

Klingenberg CP. 2008. Morphological integration and developmental modularity. Annu Rev Ecol Evol Syst 39:115–32.

Kowalewski M, Bambach RK. 2008. The limits of paleontological resolution. In: Harries PJ, editor. High-resolution approaches in stratigraphic paleontology. 1st ed. Dordrecht: Springer Netherlands. p. 1–48.

Kowalewski M, Novack-Gottshall PM. 2010. Resampling methods in paleontology. In: Alroy J, Hunt G, editors. Quantitative methods in paleobiology. Paleontological Society Short Course, October 30, 2010. p. 19–54.

Kurtén B. 1967. Some quantitative approaches to dental microevolution. J Dent Res 46:817–28.

Labonne G, Laffont R, Renvoise E, Jebrane A, Labruere C, Chateau-Smith C, Navarro N, Montuire S. 2012. When less means more: evolutionary and developmental hypotheses in rodent molars. J Evol Biol 25:2102–11.

Lande R. 1977. On comparing coefficients of variation. Syst Zool 26:214–7.

Lyson TR, Miller IM, Bercovici AD, Weissenburger K, Fuentes AJ, Clyde WC, Hagadorn JW, Butrim MJ, Johnson KR, Fleming RF, et al 2019. Exceptional continental record of biotic recovery after the Cretaceous–Paleogene mass extinction. Science 366:eaay2268.

Marroig G, Cheverud J. 2010. Size as a line of least resistance II: direct selection on size or correlated response due to constraints?. Evolution

Monson TA, Boisserie J-R, Brasil MF, Clay SM, Dvoretzky R, Ravindramurthy S, Schmitt CA, Souron A, Takenaka R, Ungar PS, et al 2019. Evidence of strong stabilizing effects on the evolution of boreoeutherian (Mammalia) dental proportions. Ecol Evol 9:7597–612.ece3.5309.

Natsume A, Koyasu K, Oda S, Nakagaki H, Kawai T, Hanamura H. 2008. Tooth size variability and relevance of numerical variation in the Japanese serow. Arch Oral Biol 53:95–8.

Pearson K. 1897. On a form of spurious correlation which may arise when indices are used in the measurement of organs. Proc R Soc Lond B Biol Sci 60:498.

Plavcan JM, Cope DA. 2001. Metric variation and species recognition in the fossil record. Evol Anthropol 10:204–222.

Polly PD. 2007. Development with a bite. Nature 449:413–415.

R Core Team. 2015. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Renaud S, Alibert P, Auffray J-C. 2017. Impact of hybridization on shape, variation and covariation of the mouse molar. Evol Biol 44:69–81.

Renvoisé E, Evans AR, Jebrane A, Labruère C, Laffont R, Montuire S. 2009. Evolution of mammal tooth patterns: new insights from a developmental prediction model. Evolution 63:1327–1340.

Roseman CC, Delezene LK. 2019. The inhibitory cascade model is not a good predictor of molar size covariation. Evol Biol 46:229–238.

Roy K, Valentine JW, Jablonski D, Kidwell SM. 1996. Scales of climatic variability and time averaging in Pleistocene biotas: implications for ecology and evolution. Trends Ecol Evol 11:458–463.

Sakai E. 1981. A study on variation in dentition of Japanese shrew mole, *Urotrichus talpoides* (Talpidae, Insectivora). Jpn J Oral Biol 23:750–789.

Schluter D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766–74.

Schroer K, Wood B. 2015. Modeling the dental development of fossil hominins through the inhibitory cascade. J Anat 226:150–162.

Simpson GG. 1944. Tempo and mode in evolution. New York (NY): Columbia University Press.

Sokal RR, Rohlf FJ. 1995. Biometry: the principles and practice of statistics in biological research. 3rd ed. New York: W.H. Freeman & Co.

Tornow MA, Ford SM, Garber PA, de sa Sauerbrunn E. 2006. Dentition of moustached tamarins (*Saguinus mystax mystax*) from Padre Isla, Peru, part 1: quantitative variation. Am J Phys Anthropol 130:352–363.

van Valen L. 1962. Growth fields in the dentition of *Peromyscus*. Evolution 16:272–277.

Viriot L, Peterková R, Peterka M, Lesot H. 2002. Evolutionary implications of the occurrence of two vestigial tooth germs during early odontogenesis in the mouse lower jaw. Connect Tissue Res 43:129–33.

Wallace JT. 1968. Analysis of dental variation in wild-caught California house mice. Am Midl Nat 80:360–380.

Williamson TE. 2001. Meniscotherium mass-death assemblages. In: Gunnell GF, editor. Eocene biodiversity. Boston (MA): Springer US. p. 335–60.

Wilson LAB, Madden RH, Kay RF, Kay RF, Sánchez-Villagra MR. 2012. Testing a developmental model in the fossil record: molar proportions in South American ungulates. Paleobiology 38:308–21.

Wolfe JL, Linzey AV. 1977. *Peromyscus gossypinus*. Mamm Species 70:1–5.

Yezerinac SM, Lougheed SC, Handford P. 1992. Measurement error and morphometric studies: statistical power and observer experience. Syst Biol 41:471–482.

Young NM, Winslow B, Takkellapati S, Kavanagh K. 2015. Shared rules of development predict patterns of evolution in vertebrate segmentation. Nat Commun 6:6690.