THE ANATOMICAL RECORD 298:1816–1823 (2015)

Detailed Anatomical Orientations for Certain Types of Morphometric Measurements Can Be Determined Automatically With Geometric Algorithms

DOUG M. BOYER,[1]* JULIA M. WINCHESTER,[2] CHRIS GLYNN,[3] AND JESUS PUENTE[4]

[1]Department of Evolutionary Anthropology, Duke University, Durham, North Carolina
[2]Interdepartmental Program in Anthropological Sciences, Stony Brook University, Stony Brook, New York
[3]Department of Statistical Science, Duke University, Durham, North Carolina
[4]Departamento De Matemáticas, Instituto Tecnológico Autónomo De México, Mexico City, Mexico

ABSTRACT

Morphometric datasets only convey useful information about variation when measurement landmarks and relevant anatomical axes are clearly defined. We propose that anatomical axes of 3D digital models of bones can be standardized prior to measurement using an algorithm that automatically finds a universal geometric alignment among sampled bones. As a case study, we use teeth of "prosimian" primates. In this sample, equivalent occlusal planes are determined automatically using the R-package *auto3dgm*. The area of projection into the occlusal plane for each tooth is the measurement of interest. This area is used in computation of a shape metric called relief index (RFI), the natural log of the square root of crown area divided by the square root of occlusal plane projection area. We compare mean and variance parameters of area and RFI values computed from these automatically orientated tooth models with values computed from manually orientated tooth models. According to our results, the manual and automated approaches yield extremely similar mean and variance parameters. The only differences that plausibly modify interpretations of biological meaning slightly favor the automated treatment because a greater proportion of differences among subsamples in the automated treatment are correlated with dietary differences. We conclude that—at least for dental topographic metrics—automated alignment recovers a variance pattern that has meaning similar to previously published datasets based on manual data collection. Therefore, future applications of dental topography can take advantage of automatic alignment to increase objectivity and repeatability. Anat Rec, 298:1816–1823, 2015. © 2015 Wiley Periodicals, Inc.

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: NSF; Grant number: DDIG BCS-0622544; Grant sponsor: NSF; Grant number: BCS-1304045; Grant sponsor: NSF; Grant number: BCS 1341120; Grant sponsor: Leakey Foundation, Evolving Earth Foundation, American Society of Mammalogists, AAPA Professional Development.

*Correspondence to: Doug M. Boyer, Duke University, Department of Evolutionary Anthropology, 130 Science Drive, Durham, NC 27708. E-mail: doug.boyer@duke.edu

Received 3 April 2015; Revised 11 June 2015; Accepted 16 June 2015.

DOI 10.1002/ar.23202

Published online 30 July 2015 in Wiley Online Library (wileyonlinelibrary.com).

Key words: high throughput phenomics; morphometrics; auto3dgm; tooth; relief index; dental topography

A quantitative and statistical perspective on anatomical variation is critical for making connections between phenomic diversity and underlying environmental and genetic factors (Houle et al. 2010). Quantification is typically done either with measurements in arithmetic units of length, area, and volume or through recording positions of landmark coordinates in either two or three dimensions. Many measurements of the "size" of a feature (in arithmetic units) are defined as the cord distance between two feature points. Alternatively, measurements of size may be defined as distances between one feature point and a given anatomical plane or axis. This latter sort of measurement will be more precise and present less potential for systematic error when there are clear and broadly repeatable means of establishing equivalent anatomical planes or axes among all specimens of a sample. Unfortunately, it is often much more difficult to consistently reproduce measurements that refer to anatomical planes/axes than measurements between two landmarks. Further, only when both measurement landmarks and relevant anatomical axes are clearly defined can morphometric datasets convey useful information about variation.

One potential avenue for improving standardization of measurement is to use automated computational procedures for identifying anatomical equivalence before measurement (Boyer et al., 2015). This takes the task of determining anatomical planes for a sample out of researcher hands—removing subjectivity—and replaces it with a procedure that will produce the same alignment for any researcher who uses it, given the same sample and procedural parameters. Using automated computational procedures, of course, requires working with digitized 3D replicas of anatomical structures. While it is doubtful that automated approaches can be used for capturing all types of variation of interest to researchers, the exact realm of their utility is currently untested.

In this study we recalculate molar relief index (RFI) for a sample of strepsirrhine primates and tarsiers. RFI values for this sample were originally published by Boyer (2008) and Bunn et al. (2011). Calculating RFI requires computing two values: 1) the surface area of the tooth crown and 2) the area of the occlusal footprint of the tooth in a two-dimensional plane (Ungar and M'Kirera F, 2003). The ratio of these measures then gives the final RFI value. RFI has been shown to correlate with dietary preferences as well as relative tooth size in primates (Boyer, 2008; Bunn et al., 2011; Winchester et al., 2014). While tooth crown area is independent of an anatomical reference plane, the area of the occlusal footprint (or shadow) of the tooth can vary based on how the occlusal plane is defined. In previous computations of RFI, researchers orientated each tooth independently to approximate what they deemed to be the occlusal plane. Presumably this introduces both random and (sometimes) systematic observer error (see Boyer, 2008; his Fig. 2), not to mention bias. We minimize observer-based error by using the program *auto3dgm* (Boyer et al., 2015). This program automatically aligns all the digitized teeth of the sample to each other. After this step, the occlusal plane needs only to be defined manually on one specimen as this reference plane can be automatically applied to the rest of the sample. The question we ask is whether the automated alignment by a computer algorithm introduces more error in what we might consider meaningful correspondences of teeth (based on important homologous features) compared to alignment by a human observer for whom the level of systematic error, random error, and bias can vary.

METHODS

Sample Details

The sample represents 146 second mandibular molar teeth of 38 species of primates (Table 1; Supporting Information Table S1). Measurements of surface area and occlusal footprint area were collected from digital models of these teeth in Stanford ply format. Details on the creation of these digital models can be found in previous publications analyzing the sample by Boyer (2008) and Bunn et al. (2011). All digital models as well as further detail on the sample can be downloaded from the MorphoSource database (www.morphosource.org) by browsing for the bibliographic record of the current study. On MorphoSource, more than one mesh file is available for each tooth. This study used the "smoothed and cropped" versions of each file.

Sample Processing

The sample was then aligned using *auto3dgm* (Boyer et al., 2015). This code can be accessed at (https://stat.duke.edu/~sayan/auto3dgm/index.shtml). Alternatively a more powerful MATLAB version of the code can be retrieved from GitHub (https://github.com/trgao10/PuenteAlignment). Parameters used in the alignment include an initial pass using 300 pseudolandmark points and a final pass using 600 points. In brief, the *auto3Dgm* program aligns groups of point clouds (representing bones or teeth) in the following way: (1) The program resamples each digital object to a uniform number of "evenly spread" pseudolandmark points. (2) All point clouds of the sample are scaled to a centroid of 1.0 and centered. (3) The first three principle axes of variation in point distribution are computed for each point cloud. (4) These principle axes are used to define eight initial alignments between pairs of bones (i.e., two bones can have their first principle axes mutually aligned in one of two ways; independently, their second principle axes can be aligned in one of two ways and likewise for their third axes). (5) All pairs of objects are aligned using the iterative closest

TABLE 1. Taxon mean values

Species	N	Diet*	Manual ln(2D area)		Auto ln(2D area)		Manual RFI		Auto RFI	
			mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Arctocebus calabarensis	4	1	2.050	0.126	2.101	0.125	0.594	0.059	0.568	0.058
Galago demidovii	4	1	1.008	0.127	1.072	0.132	0.583	0.015	0.551	0.022
Galago senegalensis	3	1	1.347	0.109	1.407	0.103	0.590	0.017	0.561	0.020
Loris tardigardus	4	1	1.791	0.079	1.837	0.062	0.589	0.017	0.566	0.016
Nycticebus javanicus	3	1	2.017	0.059	2.030	0.065	0.525	0.021	0.518	0.021
Tarsius bancanus	2	1	1.688	0.091	1.761	0.072	0.573	0.011	0.537	0.001
Tarsius spectrum	4	1	1.573	0.029	1.668	0.031	0.570	0.005	0.523	0.018
Tarsius syricta	3	1	1.739	0.023	1.850	0.035	0.560	0.025	0.505	0.017
Tupaia minor	2	1	1.115	0.053	1.186	0.059	0.539	0.063	0.504	0.060
Tupaia glis	11	1	1.831	0.091	1.921	0.081	0.634	0.050	0.589	0.050
Avahi laniger	7	2	2.321	0.041	2.356	0.049	0.568	0.040	0.551	0.039
Cynocephalus variegatus	5	2	1.942	0.103	1.998	0.106	0.587	0.075	0.559	0.097
Cynocephalus volans	3	2	2.526	0.052	2.544	0.010	0.626	0.025	0.617	0.003
Eulemur fulvus	8	2	2.920	0.065	2.922	0.068	0.506	0.027	0.505	0.026
Hapalemur griseus	5	2	2.555	0.062	2.571	0.053	0.488	0.018	0.480	0.020
Indri indri	9	2	3.582	0.100	3.591	0.097	0.478	0.049	0.474	0.046
Lemur catta	6	2	2.748	0.065	2.754	0.057	0.481	0.026	0.478	0.026
Lepilemur sp.	5	2	2.026	0.201	2.068	0.192	0.527	0.021	0.506	0.026
Prolemur simus	2	2	3.148	0.100	3.162	0.120	0.506	0.010	0.499	0.001
Propithecus diadema	4	2	3.394	0.143	3.413	0.156	0.550	0.035	0.541	0.031
Propithecus verreauxi	2	2	3.192	0.007	3.222	0.017	0.551	0.020	0.536	0.025
Galago alleni	3	3	1.762	0.062	1.822	0.057	0.507	0.030	0.477	0.027
Microcebus griseorufus	6	3	0.604	0.068	0.601	0.065	0.470	0.030	0.471	0.009
Mirza coquereli	3	3	1.773	0.011	1.811	0.016	0.473	0.026	0.454	0.028
Nycticebus bengalensis	3	3	2.285	0.052	2.295	0.059	0.490	0.011	0.484	0.011
Nycticebus coucang	3	3	2.079	0.046	2.094	0.055	0.468	0.025	0.461	0.031
Phaner furcifer	3	3	1.421	0.014	1.459	0.081	0.469	0.015	0.450	0.020
Cheirogaleus major	5	4	2.059	0.090	2.044	0.090	0.348	0.016	0.355	0.019
Cheirogaleus medius	4	4	1.693	0.134	1.673	0.143	0.354	0.020	0.364	0.024
Daubentonia madagascarensis	5	4	2.671	0.115	2.659	0.114	0.363	0.011	0.369	0.011
Perodicticus potto	6	4	2.288	0.182	2.283	0.181	0.458	0.025	0.461	0.027
Varecia sp.	8	4	3.275	0.079	3.266	0.080	0.444	0.019	0.449	0.021

*1 = insectivore, 2 = folivore, 3 = omnivore, 4 = frugivore. See Boyer (2008, Table 1) for data and criteria used for diet categorization.

points (ICP) algorithm in each of the eight possible initial alignments. The post-ICP alignment with the smallest Procrustes distance is used as the "final alignment". (6) The Procrustes distance matrix of final alignments is used to define a minimum spanning tree. (7) Point correspondences are propagated through the path defined by the minimum spanning tree, such that most original pairwise alignments are replaced by new alignments implied by this propagation procedure. (8) The Procrustes distance matrix is recomputed, the minimum spanning tree is recomputed and this procedure is reiterated until convergence (usually happens after one iteration). (9) The final rotation matrix for each point cloud is applied to the original input file and the result saved to an output folder. The workflow of this algorithm allows alignment of very different shapes, unlike commercially available auto-alignment procedures in Avizo and Geomagic, which can only be used to align similar objects. For more details on *auto3Dgm* and relevant references see Boyer et al. (2015).

This program uses .off formatted files, so first all files were batch converted from .ply to .off using the open source software tool *meshconv* (Min 2015) using the command line parameters "meshconv.exe <input-filename> –c off –o <outputfilename>," as well as a custom-written python wrapper titled *meshconv-wrapper* (provided by JMW, available on request) to enable batch-processing as meshconv is not capable of this by default.

The aligned sample was checked in Avizo and it was discovered that one *Cynocephalus* tooth (AMNH 120449) failed to achieve the uniform alignment of the rest of the sample. This one tooth was therefore automatically aligned in a pairwise fashion to one of the correctly aligned *Cynocephalus* teeth (AMNH 106286) using the "align surfaces" function in Avizo.

The occlusal plane was manually set for one specimen of *Tarsius*, and the rotation matrix applied to the rest of the sample. The program *MorphoTester* (provided by JMW, available on request) was used to compute tooth crown surface area and 2D occlusal footprint for the entire sample using a batch analysis function.

Data Analysis

Our main questions were whether RFI values as computed from automatically aligned data were as meaningful as RFI values computed from manually orientated data. RFI is the natural log of the ratio of the square root of crown surface area to square root of crown occlusal footprint area (see above and Boyer, 2008). We were particularly interested in whether taxonomic group means and variance parameters changed between methods. We were

TABLE 2. Comparison of values from manually versus automatically orientated teeth using individual specimen data

Var	Test	Type	N	test statistic	*P*	Manual	Auto	
2DA	paired t	Position	146	−8.513 (t)	<0.001	2.2277	2.2568	A+
RFI	paired t	Position	146	8.513 (t)	<0.001	0.5138	0.4993	M+
2DA	Levene	Variance	146	0.128 (W)	0.721	0.7448	0.7298	M+
RFI	Levene	Variance	146	3.842 (W)	0.051	0.0829	0.0713	M+

*Column header abbreviations: Var, variable; N, sample (number of species); *P*, probability; manual, mean or standard deviation (see type) for manually orientated sample; auto, mean or variance (see type) for automatically orientated sample.

TABLE 3. ANOVA and *post hoc* tests (Tukey's Q) on RFI of individual specimens with taxon factor

Set	*df*	*F*	*P*	MSE b	MSE w	Total Sig. Pairs	Between Diets	Within Diets	Fxn ratio
manual	144	23.14	<0.001	0.820	0.130	312	264 (84.6%)	48 (15.4%)	5.5
auto	144	14.67	<0.001	0.568	0.141	264	228 (86.4%)	36 (13.6%)	6.34

*Column header abbreviations: Set, variable set; F, F statistic for ANOVA; MSE b, Mean Squared Error between taxa; MSE w, Mean Squared Error within taxa. Total sig. pairs, the overall number of significant *post hoc* comparisons between taxa (at $P < 0.05$); Fxn ratio, ratio of significant differences between taxa of different diet groups to that number within diet groups, a higher number means a greater proportion of the pairwise differences can be explained by diet. Between diets shows significant *post hoc* comparisons between taxa of different dietary groups by number and percentage of total sig. pairs, while within diets similarly indicates comparisons between taxa in the same dietary group.

also interested in whether dietary-preference group means and variance parameters changed. As a result, we performed statistical analyses on species means and dietary group means as well as on individual specimens.

Analysis of Samples of Individual Specimens. We ran two sets of analyses using individual specimens (N = 146). First, we compared the effects of manual versus automatic alignment treatments on central tendencies (using paired *t*-tests) and variance parameters (using Levene's test) of both natural log of 2D area (2DA) and RFI (Table 2). We chose to use paired t-tests because the groups being compared represent "repeated measures" on the same samples. The chosen test differs from the usual independent samples *t*-test in having a more explicit error structure, which makes it more powerful. An alternative approach would have been to treat groups as independent samples for a more conservative analysis. The alternative approach would address the question of whether actually independent samples—representing similar treatments—would be distinguishable as a result of using the automated versus manual approach. We present an alternative version of our Tables 2 and 4 in Supporting Information. In these alternative tables, results of independent sample *t*-tests replace the results of paired *t*-tests.

For the second set of analyses on individual specimen data, we performed two ANOVAs on RFI using taxonomic groups as the factor. One ANOVA was performed on values derived from manually orientated teeth, while the other was performed on the automatic treatment. For significant ANOVAs, we recorded the number of significant *post hoc* comparisons between taxa (Table 3). For each significant *post hoc* comparison, we noted whether the taxa compared were in the same or different diet groups (Table 3).

Analysis of Samples of Taxonomic Means and Variances. One could envision using additional paired *t*-tests on individual values to compare the mean and variance of each taxonomic subsample. Such analyses would assess which taxa contribute most to any differences in overall sample mean or variance. However, many of the species level samples are very small, meaning those tests would have extremely low power (see Table 1). Therefore, we assessed whether taxonomic means and variance among taxonomic means changed in a significant way at the level of the entire sample of N = 32 taxa. We did this with paired *t*-tests and Levene's tests—the same protocol as for the individual specimen dataset. Running these analyses at the taxon-mean level also addresses concerns arising from the fact that specimens of the same taxon are probably highly interdependent. Such nonindependence could invalidate patterns observed in the individual specimen-tests and Levene's tests if certain taxa tend to differ more or differently between manual and automatic treatments than others. Such interaction effects would be less of a concern if all taxa were represented by the same number of individuals; however, some taxa are represented by many individuals while others are only represented by a few. We also performed the paired *t*-tests and Levene's test on intra-taxon standard deviation as a way of assessing whether the manual and automated approaches differed in their representation of intra-taxonomic variation.

The full complement of 38 species is not represented in the taxon mean data because (1) we had only 1 specimen of *Tupaia montana* and could not compute variance parameters, and (2) we used generic means for *Lepilemur* and *Varecia* due to uncertainty of the most correct species allocation for several individuals in that sample.

Before running *t*-tests or Levene's tests we checked each set of samples to be compared for normality using a number of different normality tests in PAST.exe (Hammer et al., 2006). One variable (intra-taxon standard deviations of RFI) was strongly non-normal; we used Mann-Whitney U for comparing automated and manual samples of this variable. Other variables were highly suggestive of normal distributions (Table 4).

For another set of analyses on species mean data, we regrouped the manually and automatically captured versions of the data by diet preference as defined in Boyer

TABLE 4. Comparison of values from manually versus automatically orientated teeth using taxon averages

Var	I-S val	Test	Type	N	Test statistic	*P*	Manual	Auto	
2DA	stdev	Levene	variance	32	0.011 (W)	0.9180	0.0022	0.0021	M+
RFI	stdev	Levene	variance	32	0.161 (W)	0.6901	0.0003	0.0004	A+
2DA	stdev	paired t	position	32	0.565 (t)	0.5760	0.0806	0.0823	A+
RFI	stdev	M-W U	position	32	507 (U), −0.06042(z)	0.9518	0.0268	0.0267	M+
2DA	means	Levene	variance	32	0.026 (W)	0.9076	0.5170	0.4957	M+
RFI	means	Levene	variance	32	1.149 (W)	0.2880	0.0053	0.0038	M+
2DA	means	paired t	position	32	5.29 (t)	0.0005	2.1383	2.1701	A+
RFI	means	paired t	position	32	5.29 (t)	0.0005	0.5147	0.4989	M+

*Column header abbreviations: Var, variable; I-S val, intra-specific value; N, sample (number of species); *P*, probability; manual, mean or variance (see type) for manually orientated sample; auto, mean or variance (see type) for automatically orientated sample.

TABLE 5. Comparison of species mean values from manually versus automatically orientated teeth by diet group (using taxon averages)

diet	N	var-W	var-P	V-m	V-a	Mean-t	Mean-P	M-m	M-a
Ins	10	0.038	0.8475	0.00093	0.00086	7.67	<0.0001	0.576	0.542
Fol	11	0.154	0.1537	0.00227	0.00189	4.47	0.001	0.534	0.523
Omn	6	0.34	0.573	0.00025	0.00019	2.84	0.037	0.479	0.466
Frug	5	0.096	0.7646	0.00283	0.00258	−4.752	0.01	0.394	0.4

*Column header abbreviations: var-W, W statistic for levene test; Var-P, probability of identity of sample variances associated with W statistic; V-m, variance value of manually orientated sample; V-a, variance value of automatically orientated sample; mean-t, t statistic for paired t-test; mean-*P*, probability of identity of sample viarances associated with t statistic; M-m, mean value of manually orientated sample; M-a, mean value of automatically orientated sample.

TABLE 6. ANOVA and *post hoc* tests (Tukey's Q) on taxon mean RFI with diet group factor (gray cells = nonsignificant *post hoc* tests)

set	*F*	*P*	MSE b	MSE w	Ins v. Fol	Ins v Omn	Ins v Frug	Fol v Omn	Fol v Frug	Omn v Frug
manual	26.13	<0.0001	0.041	0.0016	0.205/*2.86*	0.0006/*6.54*	0.0002/*12.36*	0.066/*3.68*	0.0002/*9.51*	0.0017/*5.82*
auto	19.84	<0.0001	0.027	0.0014	0.746/*1.43*	0.0029/*5.53*	0.0002/*10.38*	0.034/*4.10*	0.0002/*8.95*	0.0097/*4.85*

*set, variable set; *F*, *F* statistic for ANOVA; MSE b, Mean Squared Error between diet groups; MSE w, Mean Squared Error within diet groups. Each diet group *post hoc* comparison gives (*P*-value/Tukey's Q).

(2008); these subsets were also compared for similarity in mean and variance, as before with paired *t*-tests and Levene's tests (Table 5). Finally, two sets of diet group ANOVAs, followed by *post hoc* comparisons, were run on taxon mean RFI—one on manual and one automatic data; the results for manually and automated datasets are reported (Table 6). Software used for these analyses included PAST.exe (Hammer et al., 2006).

RESULTS

Analysis of Samples of Individual Specimens

At the individual specimen level (N = 146), paired 2DA values in the automatic sample are slightly but significantly larger than values in the manually orientated sample, while for RFI, values in the automatic sample are slightly but significantly lower (Fig. 1, Table 2). For RFI, variance in the automatic sample is lower with marginal significance ($P = 0.051$), while for 2DA variances are not significantly different (Table 2).

Individual specimen ANOVAs of RFI with a taxon factor were significant for both manual and automatic treatments (Table 3). The manual treatment exhibits a higher F value. The manual treatment evinces a greater number of significant *post hoc* comparisons, but the automatic treatment has a proportionally greater number of significant *post hoc* comparisons of taxon pairs with different dietary groups (Table 3: Fxn ratio is higher).

Analysis of Samples of Taxonomic Means and Variances

The patterns of significance from taxonomic mean (N = 32) *t*-tests and Levene's tests are the same as for the individual data analyses, with RFI *t*-tests and 2DA t-tests yielding significant results, while Levene's tests yield insignificant results. Similarly, for intra-taxon standard deviations of both RFI and 2DA, both *t*-tests and Levene's tests are nonsignificant (Fig. 1, Table 4).

When comparing diet group subsets of the sample, every *t*-test returned significant differences between the manual and automatic treatments, while none of the Levene's tests were significant (Table 5). In addition, ANOVAs of taxon means separated into diet groups were significant for both manual and automatic treatments (Fig. 2, Table 6). Again, the manual treatment has higher *F*-values. However, for diet-group, taxon-mean

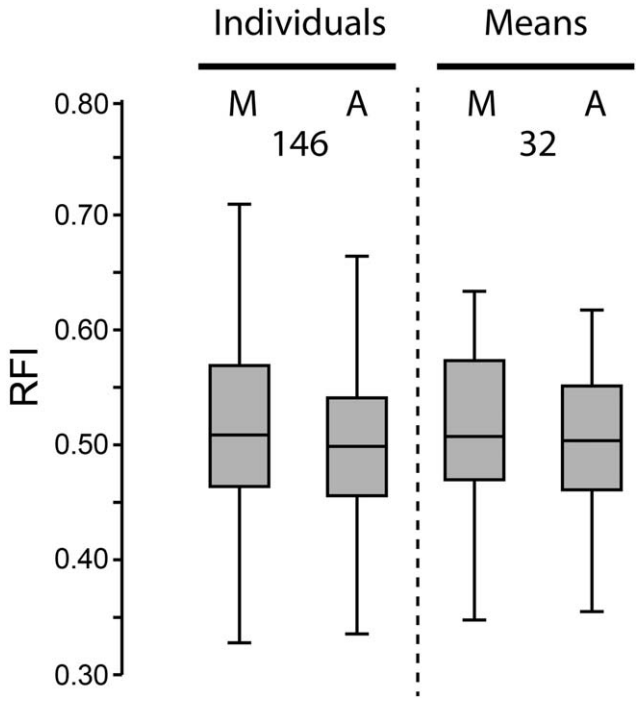

Fig. 1. Box plots comparing values of relief index (RFI) yielded by manual (M) and automated (A) alignments of teeth for occlusal plane area computation. Left two boxes each shows distribution of 146 individual specimen data points. Right two boxes each shows the distribution of 32 taxon mean data points.

ANOVAs, the automatic treatment returns a greater number of significant post hoc comparisons (Table 6).

Finally, replacing the paired t-tests described above with independent samples *t*-tests in all cases yields nonsignificant results (Supporting Information Table S2), except for in one case, the comparison between insectivore mean values between treatments: The manual treatment still yields a significantly higher value.

DISCUSSION AND CONCLUSIONS

Comparisons of datasets produced by manually and automatically aligned teeth reveal miniscule differences (Figs. 1 and 2; Tables 2–6). However, these differences are actually significant for paired samples of 2Da and RFI values at both the individual specimen level and the taxon mean level; the mean value of RFI from automatically orientated specimens is lower (Tables 2 and 4). Such differences—offsets in value between two datasets measuring the same thing—represent systematic error by definition. Therefore, it suggests a slight offset in the calibration of the occlusal plane of the automatically aligned sample relative to that in the manual treatment. It is possible that a slightly different uniform rotation of the entire sample with respect to the occlusal plane would eradicate the differences. However, it is not clear that this would be desirable for the purpose of comparing values by diet-group between the two methods because the signs of differences between diet-group subsamples of manual and automated treatments are not uniform (Table 5, Fig. 2). Therefore, such an adjustment of the occlusal plane for the automated sample—while decreasing the difference between certain subsamples—would accentuate differences between other subsamples. Ultimately this points to an interaction effect in systematic error by diet-group, plus differential sample size among diet-groups as the source of the small positional offset between the two treatments at the level of the overall sample. At this juncture, it is also important to note that when treated as independent samples, all of the significant *t*-test differences between treatments become strongly nonsignificant (Supporting Information Table S2).

Turning to variance parameters (Tables 2 and 4), there are no significant differences for any of the treatments (individual specimens, means, or diet-group subsamples) using either *t*-tests or Levene tests; this result suggests that random errors, or precisions, of the manual and automatic treatments are essentially similar.

Significant differences between manual and automatic treatments also appear when comparing data partitioned by diet groups (Table 5). Diet ecology was the factor of interest when the manual RFI dataset was originally collected. That differences in sample distribution arise when using dietary partitions at least two possible explanations. The first possibility is that observer expectation/aspiration for group differences led to bias that artificially magnified differences between certain groups. Consistent with the expected effects of observer bias, frugivores have lower values in the manual treatment, while insectivores and folivores have higher values in the manual treatment.

The second possibility is that the (usually) greater relief of insectivore and folivore teeth leads to a less anatomically accurate occlusal alignment by *auto3dgm*. Consistent with this hypothesis, the differences between the manual and automated treatments are much stronger for the insectivore and folivore groups than for the omnivore and frugivore groups.

We are not sure how to distinguish between these two possibilities *post hoc*. Keeping in mind that the original goal of RFI was to capture functionally significant variation in tooth form, the most obvious way to resolve the ambiguity discussed above would be to redo the study by orientating the teeth relative to the corpus of the mandible. We have not endeavored to do this because we do not have the access to a dataset that includes mandibles for any of the 146 sampled specimens at this time.

Acknowledging that none of the variance comparisons in Table 4 are significant, we can still ask the question: Does one treatment have results suggestive of more meaningful species distinctions than the other? We can assess this by comparing average intra-taxon standard deviation (rows 3-4 of Table 4): For 2DA, the automated treatment has a slightly higher average intra-taxon standard deviation, which might equate to lower distinctiveness between taxa. However, for RFI, the manual treatment is slightly higher. If we assess the question using the magnitude of inter-specific variance, we find the reverse answer (rows 5–6 of Table 4): For 2DA, the automated treatment has a higher inter-specific variance, suggesting greater distinction between species, while for RFI the manual approach is higher. So the answer to this question is “no”: The signal is mixed.

Results from ANOVAs suggest two apparently contradictory conclusions: (1) that the manual method more powerfully separates among taxonomic and ecological

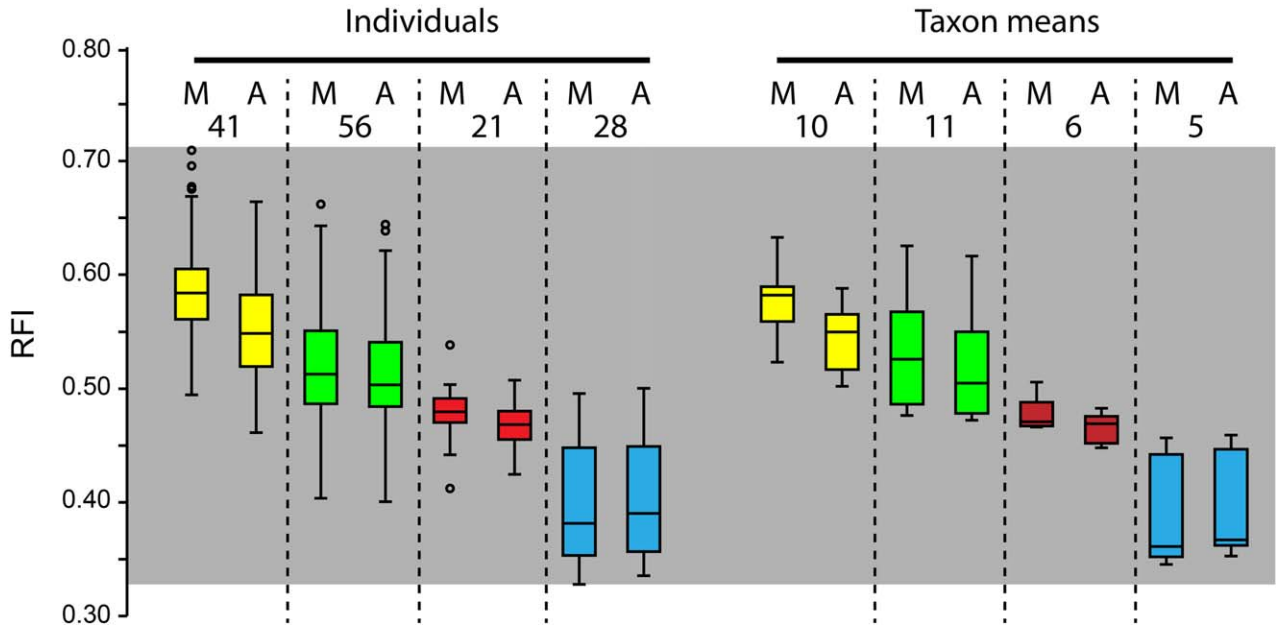

Fig. 2. Box plots comparing values of relief index (RFI) yielded by manual (M) and automated (A) alignments of teeth for occlusal plane area computation. Color codings indicate dietary groupings: yellow, insectivore; green, folivore; red, omnivore; blue, frugivore. Left eight boxes show distributions of 146 individual specimen data points per treatment. Right eight boxes show distributions of 32 taxon mean data points per treatment. The gray interval shows the vertical range of RFI values in the manual treatment.

groups; (2) the pattern of differences yielded by the automated method is more ecologically meaningful. The first conclusion is supported by the following observation: Both types of ANOVAs (Tables 3 and 6) return higher F values when run on data from the manual treatment, indicating less within-group variance compared to between-group variance. This pattern is also generally reflected in numbers of *post hoc* comparisons in Table 3 as well as in the values of Q and associated P values of the *post hoc* tests in Table 6: the values tend to be more extremal (higher for Q and lower for P) in the manual group—meaning there is slightly better distinction between the diet-groups in manually collected data. As for the diet subsample comparisons, either observer bias or imprecision in the automated alignment process may have caused this: we do not have the ability to definitively disentangle these possibilities.

Two observations support the second conclusion from ANOVA. First, is an exception to the pattern of better diet-group distinction in the manual treatment in Table 6: A significant *post hoc* comparison between omnivores and frugivores was recovered in the automated sample but not the manual one (Table 6). If this is not a random occurrence (that is, if it reflects a real geometric difference between the two groups), then it is hard to explain why the manually orientated data did not also show this result. If the observer who orientated these teeth originally (DMB) suffered from unconscious bias due to prior expectations about differences between other groups, this could have hindered accuracy in characterizing differences between omnivores and frugivores, specifically.

The second observation supporting the conclusion that the automated treatment is more ecologically meaningful is that a greater proportion of the significant inter-taxon differences are found between dietarily distinct groups in the automated treatment (Table 3).

In sum, we feel the results of these analyses demonstrate unequivocal benefits to using automatically aligned teeth for computation of RFI. We are unconcerned about our inability to determine whether observer bias or algorithm imprecision is the main explanation for slight differences between the automatic and manual data. If observer bias is the source of differences, then clearly the automated approach is preferable on the grounds that it (1) is not plagued by that phenomenon, (2) is therefore also a better descriptor of the sample variation, (3) recovers a greater number of significantly different groups at the diet group level, and (4) recovers a greater proportion of dietarily distinct inter-taxon differences. On the other hand, if slightly greater algorithmic imprecision explains the differences, the algorithmic approach is still preferred because, unlike manual datasets, the algorithmic level of imprecision can be assumed to be relatively constant. It is very likely that if other researchers reorientated the sample and recollected the same dataset, some of them would generate measures with higher systematic and/or random error than either dataset analyzed here. Therefore, we recommend future applications of RFI and other dental topographic metrics that are affected by orientation of the tooth surface—such as OPC (Evans et al., 2007)—use *auto3dgm* (Boyer et al., 2015) to standardize the orientation of the sample.

While we think *auto3dgm* could be used for aligning other types of bones for measures requiring standardized views, it will not be appropriate for all cases. However, it may sometimes be appropriate to use a few select landmarks to orientate a sample of digital bone models prior to collection of particular linear or angular measures. While

this seems obvious, most researchers either prefer geometric morphometric measures or (if they are more focused on biomechanics) linear, area and angular measures. These methods do not need to be kept separate for all purposes and researchers should be creative about combining them to maximize the benefits of each one in collecting data and measuring biologically meaningful variation.

The most important observation of this study is that overall patterns of variation are extremely similar between manual and automated treatments. This at least confirms that if there is a user bias in the original data set, it is very minimal. At the same time, it suggests that if the algorithmic approach is less precise with regard to the particular manual dataset, the degradation of precision is also very minimal. The existence of such small differences makes the automated approach preferable, again, because it sides steps and obviates concerns about observer bias, or variability in observer error.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the staff and curators at the American Museum of Natural History Mammalogy Dept. (Darin Lunde, Eileen Westwig, Neil Duncan, Dr. Nancy Simmons). Additional collections staff to whom we are indebted include those at the Smithsonian (Linda Gordon and Darin Lunde) and Museum of Comparative Zoology. The curators of the Stony Brook University Anatomy Departments collection must be thanked for access to that collection. The authors thank Siobhan Cooke for access to *Brachyteles* specimens. Jukka Jernvall provided access to *Prolemur simus* specimens for molding. We thank Stefan Judex and Clint Rubin of the Stony Brook University Center for Biotechnology for access to their scanners at the time of original data collection.

LITERATURE CITED

Boyer DM. 2008. Relief index of second mandibular molars is a correlate of diet among prosimian primates and other euarchontan mammals. J Hum Evol 55:1118–1137.

Boyer DM, Puente J, Gladman JT, Glynn C, Muhkerjee S, Yapuncich GS, Daubechies I. 2015. A new fully automated approach for aligning and comparing shapes. Anatomical Record 298:249–276.

Bunn JM, Boyer DM, Jernvall J, Lipman Y, Daubechies I. 2011. Dirichlet normal surface energy of tooth crowns, a new technique of molar shape quantification for dietary inference, compared with previous methods in isolation and in combination. Am J Phys Anthropol 145:247–261.

Evans AR, Wilson GP, Fortelius M, Jernvall J. 2007. High-level similarity of dentitions in carnivorans and rodents. Nature 445:78–81.

Hammer Ø, Harper DAT, Ryan PD. 2006. PAST: paleontological statistics software package for education and data analysis. 1.43 ed. Oslo: University of Oslo.

Houle D, Govindaraju DR, Omholt S. 2010. Phenomics: the next challenge. Nature Reviews Genetics 11:855–856.

Min P. 2015. Meshconv version 1.16 [Software]. Available at: http://www.cs.princeton.edu/~min/meshconv/.

Ungar PS, M'Kirera F. 2003. A solution to the worn tooth conundrum in primate functional anatomy. Proc Natl Acad Sci USA 100:3874–3877.

Winchester JM, Boyer DM, Cooke SE, St.Clair EM, Ledogar JA, Gosselin-Ildari AD. 2014. Dental topography of platyrrhines and prosimians: convergence and contrasts. Am J Phys Anthropol 153:29–44.