A New Fully Automated Approach for Aligning and Comparing Shapes

DOUG M. BOYER,[1]* JESUS PUENTE,[2] JUSTIN T. GLADMAN,[3,4] CHRIS GLYNN,[5] SAYAN MUKHERJEE,[5,6,7] GABRIEL S. YAPUNCICH,[1] AND INGRID DAUBECHIES[7]

[1]Department of Evolutionary Anthropology, Duke University, Durham, North Carolina
[2]Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey
[3]NYCEP, New York Consortium in Evolutionary Primatology, New York, New York
[4]PhD Program in Anthropology, Graduate Center, CUNY, New York, New York
[5]Department of Statistical Science, Duke University, Durham, North Carolina
[6]Department of Computer Science, Duke University, Durham, North Carolina
[7]Department of Mathematics, Duke University, Durham, North Carolina

ABSTRACT

Three-dimensional geometric morphometric (3DGM) methods for placing landmarks on digitized bones have become increasingly sophisticated in the last 20 years, including greater degrees of automation. One aspect shared by all 3DGM methods is that the researcher must designate initial landmarks. Thus, researcher interpretations of homology and correspondence are required for and influence representations of shape. We present an algorithm allowing fully automatic placement of correspondence points on samples of 3D digital models representing bones of different individuals/species, which can then be input into standard 3DGM software and analyzed with dimension reduction techniques. We test this algorithm against several samples, primarily a dataset of 106 primate calcanei represented by 1,024 correspondence points per bone. Results of our automated analysis of these samples are compared to a published study using a traditional 3DGM approach with 27 landmarks on each bone. Data were analyzed with *morphologika*[2.5] and PAST. Our analyses returned strong correlations between principal component scores, similar variance partitioning among components, and similarities between the shape spaces generated by the automatic and traditional methods. While cluster analyses of both automatically generated and traditional datasets produced broadly similar patterns, there were also differences. Overall these results suggest to us that automatic quantifications can lead to shape spaces that are as meaningful as those based on observer landmarks, thereby presenting potential to save time in data collection, increase completeness of morphological quantification, eliminate observer error, and allow comparisons of shape diversity between different types of bones. We provide an R package for implementing this analysis. Anat Rec, 298:249–276, 2015.   © 2014 Wiley Periodicals, Inc.

*Correspondence to: Doug Boyer; Department of Evolutionary Anthropology, Duke University, 130 Science Drive, Box 90383, Durham, NC. Fax: 919-684-8542. E-mail: doug.boyer@duke.edu

INTRODUCTION

As the general theme of this volume is the application of three dimensional geometric morphometrics (3DGM) to functional morphology, there is little need to convince most readers about the importance of morphological studies to evolutionary and developmental biological research. However, the utility of detailed morphological information in such research has become increasingly questioned (see Springer et al. [2013] comment on O'Leary et al. [2013a, b]). Therefore, we emphasize that patterns of phenotypic variation (including morphology) among biological structures form the basis for understanding gene function (e.g., Morgan, 1911; Abzhanov et al., 2006), developmental mechanisms (e.g., Harjunmaa et al., 2012), ecological adaptation (e.g., Losos, 1990; Frost et al., 2003), and evolutionary history (e.g., Leakey et al., 1964; Ostrom, 1975; Gingerich et al., 2001). Given its importance in a diverse set of biological disciplines, we believe that morphological information remains highly relevant to scientific discovery and advancement.

Since the Modern Synthesis of Evolutionary Theory was reached in the 1940s and evolution was appropriately redefined in its most basic population–genetic context, genomic approaches to studying evolution have advanced dramatically. In part, this sea change is a result of increasingly available data and improving computational power. Ever more comprehensive and rapid assessments of genetic variation have been possible as a result (Venter et al., 2001). Since the late 1980s, large-scale automated genomic analyses have flourished; a great deal is now known about and can be inferred from genotypic variation (McVean et al., 2005; Houle et al., 2010). Genetic data are even accessible from remains of extinct organisms such as subfossil lemurs (Orlando et al., 2008) and Neandertals (Green et al., 2010).

The utility of morphology is now questioned, in part, because the ability to analyze morphological data has progressed much more slowly than the ability to analyze genomic data. However, there is a call from some evolutionary biologists for the collection and analysis of high-dimensional phenotypic data (Houle et al., 2010) in an analogous high-throughput and automated fashion. This perspective proposes that the utility and information content of genetic data will only reach its fullest extent once data on associated phenotypes can be analyzed at equivalent rates and scales. Ideally, increasing availability of phenomic data would promote comprehension of how the interaction between phenotypic variation and the environment is mediated by the genome and how selective pressures on the phenome are transferred to the genome. Reflecting the perceived importance of such data, the field of phenomics has recently been defined as

that endeavoring to acquire high-dimensional phenotypic data on an organism-wide scale (Houle et al., 2010). Although phenomics is defined in analogy to genomics, the analogy is misleading in one respect. We can come close to characterizing a genome completely but not a phenome, as the information content of phenomes dwarfs genomes and is heavily influenced by the mode, tempo, duration, and timing of its observation and quantification (Houle et al., 2010).

By itself, variation in morphological structure (a component of phenomic variation) has higher dimensionality than variation in the genome, which makes it exponentially more difficult to quantify in a meaningful way (e.g., Boyer et al., 2011). This is not to say that significant advances in analysis of morphology are impossible or that the field of morphometrics has stagnated. As emphasized and demonstrated by work in this volume, new and more sophisticated approaches are being developed. More sophisticated statistical contexts (Nunn, 2011) are available thanks to improved computing power and flexible open-source coding languages (Orme et al., 2011; R Coding Team, 2012). Additionally, there is growing automation of shape quantification based on new variations of methods for spreading semi-landmarks over a 3D surface model (Bookstein, 1997; Bookstein et al., 1999; Bookstein et al., 2002; Perez et al., 2006; Harcourt-Smith et al., 2008; Mitteroecker and Gunz, 2009). However, 3D shape analyses are generally tied to at least two-user determined landmarks (Polly and MacLeod, 2008), and 3DGM analyses do not appear to be very meaningful without four or more (Gunz et al., 2005; Wiley et al., 2005). As a result, these approaches continue to have many of the same limitations as morphological studies from 30 to 40 years ago. Part of the problem is sample size; in most cases the number of measurements, and the sample sizes per study have changed little (compare Berge and Jouffroy [1986] with Boyer et al. [2013]—though statistical analyses are more sophisticated in the more recent study, there are no substantial differences in measurement complexity or sample sizes in these two studies almost 30 years apart). Other principal limitations to the current traditional approach to morphological studies include: (1) subjectivity/observer-error in interpretation and measurement, (2) time intensiveness for generating large datasets, (3) sparse and potentially incomplete and/or biased representation of specimen morphology and sample variation, and (4) limited accessibility of information encapsulated in morphology due to lack of widespread researcher expertise. All restrictions stem from the necessity that researchers must directly observe, interpret, and actively measure (or mark) every specimen of a study. These limitations likely at least partly explain why genetic data currently provide a more statistically powerful approach to certain evolutionary questions, and

also why questions that can be addressed only by morphology (e.g., what physical traits are functionally beneficial for a certain behavior?) are often less thoroughly examined or appear more controversial despite a long history of analyses.

As discussed by MacLeod et al. (2010), in order to make the study of morphology less of a "cottage industry" and bring it to a new level of objectivity, standardization, efficiency, and accessibility, we should seek more automation in the determination of patterns of morphological similarity and difference. Several researchers (Lohmann, 1983; MacLeod, 1999; Polly and MacLeod, 2008; Sievwright and MacLeod, 2012) have worked to develop techniques that minimize assumptions involved in measuring shape similarity. Initiatives for "automated taxonomy" exist (Weeks et al., 1999; MacLeod, 2007) and have had some degree of success. However, all of these automated approaches require a "dimension reduction" in the initial analytical stages, which still necessitates that researchers make a decision, informed by their understanding of important and "equivalent" morphological features, on how to make that reduction. Most automated work has been carried out on two-dimensional (2D) outlines or raster-photographs. In such cases, the shape of an outline and the images in a photograph are determined by how the researcher orients the camera with respect to the specimen. Even when attempting the "same" view, two different researchers may have systematic error with respect to one another or different levels of random error in setting up specimens for photography. Furthermore, many techniques described as automated, including those for 2D objects, still require direct interaction with the study materials to determine at least one "corresponding point" common to all the shapes of the study sample (see articles in MacLeod, 2007).

Biomedical and neuroscience research pursued by computer scientists has led to some successful automated quantification procedures in 3D (Styner et al., 2006; Paniagua et al., 2012). However, these methods have been designed with a limited range of variation in mind and applied to monospecific samples. Whether these methods would have meaningful success in a sample with more substantial shape diversity among homologous objects is unknown.

In order to begin testing the limits on the degree to which (and the questions for which) shape analysis can be automated toward a scientifically meaningful end, we present a new fully automated algorithm for aligning and placing landmarks comprehensively on digital 3D models of bones. We also provide an R package application to promote the testing of our algorithm and use by other researchers. This method builds conceptually on a previously published approach (Boyer et al., 2011) where it was shown that a superficially similar algorithm can (1) reasonably match corresponding points on different instances of the same bone (represented by different individuals and species), (2) estimate shape differences that allow classification of shapes to species with accuracy comparable to, or better than, user selected landmarks on the same specimens, and (3) allow for the entertainment of different "correspondence hypotheses" based on the morphocline (or "path") that is assumed to connect shapes in the dataset. Operationally, the method of Boyer et al. (2011) finds several hundred candidate alignments between conformally flattened representations of two objects. Each initial

alignment is "improved" using a thin plate spline to align automatically identified extremal points (points of high local curvature—i.e., "type II landmarks"). These mappings are then applied to unflattened versions of the two objects and a continuous Procrustes distance is computed (Lipman and Daubechies, 2010). The mapping that results in the minimum continuous Procrustes distance is treated as the best mapping among the many candidate maps. This minimum distance mapping was found to usually represent a biologically meaningful alignment according to criteria 1 and 2 described above.

Despite its successes, the method presented by Boyer et al. (2011) has several shortcomings: (1) since correspondences used to determine shape differences are purely pairwise and not transitive, there is an inconsistent template for biological correspondence relating all pairs of shapes in the dataset; (2) the conformal flattening procedure of the analysis limits its application to "disc-type" shapes with an open end (like the tooth crowns or ends of long bones of that dataset); and (3) the MATLAB® application for the analysis is difficult to work with, lacks good visualization tools, and does not yield output that can be widely employed in other analytical procedures.

We overcome these limitations in the new algorithm presented here, which we have developed into an R-package called *auto3dgm*. One of the most exciting prospects of *auto3dgm* is its potential to help quantify morphology more comprehensively and equably (if not exhaustively). It has long been acknowledged that measurements of select characters are less meaningful than more comprehensive approaches:

"Direct determination of rate of evolution for whole organisms, as opposed to selected characters of organisms, would be of the greatest value for the study of evolution. Matthew wrote, nearly a generation ago (1914), 'to select a few of the great number of structural differences for measurement would be almost certainly misleading; to average them all would entail many thousands of measurements for each genus or species compared.'" (Simpson, 1944; p. 14)

"Another level of description -of entire surface regions, or of volumetric elements, or of qualitative aspects of structures rather than structures themselves- may in some instances be most meaningful (Roth, 1984, 1991) and bring us closer to identifying the biological processes of interest. Hence the appeal and utility of methods of comparison that interpolate between landmark points, such as D'Arcy Thompson's transformation grids" (Roth, 1993; p. 53)

Matthew's implied perspective was that increasing the number of measurements would be useful (though impractical) and would approach a representation of the "total taxonomic distance." This taxonomic distance is sometimes referred to as "morphological disparity" and may allow meaningful discussion of the amount, rate and pattern of evolution among a sample of species in certain settings. A greater amount of morphological difference between corresponding and homologous structures can be hypothesized to relate to the amount of evolutionary change that has occurred in the compared taxa since they diverged from their common ancestor.

This idea is reflected in the numerical taxonomy movement (Sokal, 1966; Sneath and Sokal, 1973).

A wealth of careful, mathematically-rooted consideration has been aimed at these premises over the years. It has been effectively argued that it is actually impossible to generate a generalized comprehensive view of the total phenetic distance between specimens or taxa (Bookstein, 1980, 1994; MacLeod, 1999). In fact, Bookstein (1991, 1994) argues that morphometrics is purely about documenting covariance among biological forms, stating that morphometric methods are neither suited for "the computation of 'magnitude' of shape change nor for the clustering of individual specimens according to degree of similarity of shape" (Bookstein, 1994; p. 205). MacLeod (1999) explains the insufficiency of morphometrics in this regard, saying: "All morphological disparity estimates published thus far represent indices that are inextricably tied to particular methods of morphological representation and particular scales of morphological assessment", that "it seems…unlikely that a generalized estimate of 'morphological disparity,'…can ever be achieved." and finally that it is imperative that "the morphometrician remembers the domain within which he/she operates is strictly limited" (MacLeod, 1999; p. 134).

We do not suggest the method we present fundamentally resolves any of these issues. It aids in the discussion of morphological disparity because it is more objective and comprehensive in its measurement of shape than previous methods. Though Bookstein (1994) argues that morphometrics must be applied after homology considerations have taken place, we suggest that our method can help identify an "operational homology" or "biological correspondence" (Smith, 1990) more objectively.

Of the various types of homology discussed by evolutionary biologists and paleontologists, it is relevant to review at least three different types here: these include transformational, operational, and taxic homology (Patterson, 1982; Smith, 1990). It would seem that transformational homology is of primary importance in an evolutionary sense. It is similar to Darwinian homology (Simpson, 1961), in which features are considered homologous among several taxa if they are equivalent through "descent with modification" from the common ancestor. This also matches Van Valen's (1982) definition of homology as "continuity of information" through evolution. Of course, comprehension of transformational homology is often fairly elusive, since the morphoclines describing it can be expected to gain accuracy with a more complete fossil record and a more accurate phylogeny of life (Van Valen, 1982).

Operational homology most generally appears to refer to ontologies defining biological correspondence for the sake of measurement, comparison among taxa, and/or as a working hypothesis of transformational homology. What MacLeod (2001; p. 3) describes as "geometric (or morphometric) homology (*sensu* Bookstein, 1991)" of geometric morphometrics can be considered as specific types of operational homologies. In a way, Thompson (1942), as also quoted by Roth (1993), reminds researchers not to forget the distinction between operational homologies and carefully tested hypotheses of transformational homology:

"The morphologist, when comparing one organism with another, describes the differences between them point by point and "character" by "character" and he falls readily into the habit of thinking and talking of evolution as though it had proceeded on the lines of his own descriptions, point by point, and character by character." (Thompson, 1942; p. 1036)

Finally, taxic homology is equivalent to "synapomorphy" or "symplesiomorphy" whereby similarity in morphological form (usually referred to as a "character state") of a transformationally homologous feature exhibited by a taxonomic sample of interest is thought to reflect the inheritance of that "state" from a common ancestor. Whether identified taxic homologies help elucidate phylogenetic relationships depends on whether particular character states have evolved numerous times and exhibit homoplasy, as well as whether perceptions of transformational homology are correct. When discussing features on a finer scale than whole bones or organs, hypotheses of transformational homology are usually difficult to test. When the data necessary for such tests are available (e.g., via a dense fossil record [Van Valen, 1982]) the results can be surprising.

The empirical route to homology hypotheses is a recursive one. Van Valen (1982) says that homology is "more than similarity" which means that assessment of shape similarity *is* involved. Shubin (1994) discusses tests and evaluations of homology hypotheses, saying homology is "only indirectly related to similarity" and that "homologous features may be very dissimilar." But without an *a priori* phylogeny, how does one postulate homology of dissimilar features? In many cases, operational homology hypotheses are qualitatively rooted in geometric similarities even for matching dissimilar features in two taxa. For skeletal elements, operational homology (=topological correspondence) hypotheses are established by researchers physically or conceptually seriating features of specimens into morphoclines. The correspondence among end-members of the morphocline (the humeri of a whale and a bat—for instance) may be un-interpretable next to each other, but will have more definitive operational homologies if they are compared through intermediate forms along a taxonomically rich, seriated sample. Of course, this task is aided by information beyond the geometry of isolated bones: the position and orientation of the bone in the complete skeleton is also known and used (i.e., cues from "type I" landmarks). Different researchers may see and emphasize different aspects of shape, and samples with different taxa will suggest different morphoclines and possibly different patterns of correspondence among end-members. As Roth (1993; p. 53) says "The recognition, and operational definition, of homologous points is a non-trivial problem (Jardine, 1969; Smith, 1990), and one not necessarily with unique solutions." Furthermore, different skeletal element sets from the same taxonomic sample may seriate in morphoclines with different taxonomic orderings. For example, the calcaneus bone of a tarsier has the most extreme form in comparison to any sample of primate species, whereas the astragalus bone of tarsiers can be described as roughly intermediate between that of certain anthropoid and strepsirrhine primates. For a given taxonomic sample, a consideration of which bones arrange in morphoclines with similar orderings of taxa (and thereby present congruent pictures of operational homology) aids in formulating phylogeny hypotheses. Cladistic parsimony analyses can be conceptually related to this practice. Clearly, determination of operational

homology is at least partly based on a qualitative consideration of geometric similarity and morphoclines among samples. Our automated procedure, which considers the total surface of bones and the pattern of distances between them, can be implemented toward this end.

Because *auto3dgm* determines feature correspondence objectively (algorithmically) and more comprehensively, it can assess morphological differences in a way that suffers from less measurement sensitivity. This decreased sensitivity makes the shape quantifications of one bone or "part" more easily generalizable to other parts compared to previous methods (as we will demonstrate with an example). Ultimately, this should allow greater insight into patterns in, and the generation of, morphological disparity through the evolutionary process.

MATERIALS AND METHODS
Institutional Abbreviations

AMNH, American Museum of Natural History, New York, NY; CGM, Egyptian Geological Museum, Cairo, Egypt; DPC, Duke Lemur Center Division of Fossil Primates, Durham, NC; GU, H.N.B Garhwal University, Srinagar, Uttarakhand, India; IGM, Museo Geológico del Instituto Nacional de Investigaciones Geológico-Mineras, Bogotá, Colombia; IRSNB, Institut Royal des Sciences Naturelles del Belgique, Brussels, Belgium; KU, Kyoto University, Kyoto, Japan; MCZ, Museum of Comparative Zoology, Harvard University, Cambridge, MA; MNHN, Muséum National d'Histoire Naturelle, Paris, France; NMB, Naturhistorisches Museum Basel, Basel, Switzerland; NMNH, Smithsonian Institution National Museum of Natural History, Washington, D.C.; NYCEP, New York Consortium in Evolutionary Primatology, New York, NY; SBU, Stony Brook University, Stony Brook, NY; SDNM, San Diego Natural History Museum, San Diego, California; SMM, Science Museum of Minnesota, Minneapolis, MN; UCM, University of Colorado Museum of Natural History, Boulder, CO; UCMP, University of California Museum of Paleontology, Berkeley, California; UK, University of Kentucky, Lexington, KY; UM, University of Michigan, Ann Arbor, Michigan; USGS, U.S. Geological Survey, Denver, Colorado.

Samples

We utilize four samples of surface meshes generated from either microCT or laser scans to test *auto3dgm*. Table 1 is a taxonomic list for each dataset with sample sizes per genus (Supporting Information Tables 1–3 give the specimen numbers for each sample). The first sample includes 106 calcaneal bones of 67 genera, and is the exact sample used by Gladman et al. (2013). We test our method by running the same analyses on this sample as Gladman et al. (2013) and compare the results: *auto3dgm* produces landmark datasets that can be analyzed in a manner identical to traditional user-collected landmark datasets. The second sample is comprised of 80 astragali that we analyze and compare to a subset of 80 calcanei from the first sample. The third sample is 49 distal phalanges representing fossil and extant taxa: it was selected to demonstrate the method on a bone with a "different quality" of shape variation. Distal phalanges are basically cone-shaped with fewer consistent "feature points" than astragali or calcanei. Nonetheless, they

exhibit a range of forms from "blade-like" (falcular) to "spatulate" (unguliform) (Fig. 1). Therefore, each bone is less complex, but the range of variation across the sample remains substantial. The fourth sample also represents astragali and overlaps the second, but includes additional specimens and species (Table 1). This sample is used to demonstrate a semi-supervised alignment procedure of the R-package "Shape_Alignment" that can be applied to samples that will not correctly align with the fully automated procedure of *auto3dgm*.

Sample Processing

Very little pre-processing is required for *auto3dgm*. Surface files should be in the Open file format (.off) and of sufficient resolution to capture all surface features of interest. It should be noted that the .off format is closely related to more widely known Stanford Polygonal Mesh (.ply) format. The free software MeshLab can be used to convert .ply files to .off files, as well as batch converters (see http://www.stat.duke.edu/~sayan/3DGM/index.shtml). If made from CT scans, the surfaces must be carefully checked and cleaned so they have no internal vertices. Virtually no processing is required for laser-scan generated data aside from smoothing the mesh.

The majority of surface files in our datasets were generated by microCT scanning. Details on both laser- and microCT scanning parameters of the astragalus and calcaneus specimens have been reported on previously in appendices and supplementary tables (Boyer and Seiffert, 2013; Boyer et al., 2013). The distal phalanx dataset is new.

auto3dgm Input and Output Files

The method demonstrated here was developed by Puente (2013) as a major component of a PhD thesis and the mathematical details can be found there. Additional technical articles focusing on mathematics are forthcoming (Puente and Daubechies, in prep). The input files for the routine are a set of surface mesh files in .off format. The user must also supply a set of "low resolution" versions of the mesh files that will be used by the algorithm to generate summary images. Downsampling of mesh files can be accomplished with visualization programs such as Meshlab (Cignoni et al., 2012), Avizo (Visualization Sciences Group, 2009), and Geomagic (3D Systems Inc., 2013).

The outputs include (1) an "alignment file," which is a "multi-surface".off file that includes displays of user-supplied low resolution renderings of all specimens shown in the algorithm-determined optimal alignment (Fig. 2); (2) an "MDS file," which is another multi-surface file that embeds the same aligned renderings of specimens in a coordinate space determined by a multi-dimensional scaling (MDS) analysis of the distance matrix of aligned specimens (again for visualization purposes) (Fig. 3); (3) a "scaled".txt file with all of the coordinate data for all specimens scaled to the same centroid size, that can be loaded into, visualized, and analyzed in *morphologika*$^{2.5}$ (O'Higgins and Jones, 2006); (4) an "unscaled".txt file with all of the coordinate data for all specimens at the scale of the original input files which can also be analyzed in *morphologika*$^{2.5}$; and (5) a folder with copies of all the original input files, the coordinates

TABLE 1. Taxonomic samples for this study

Extant samples:

Taxon	Set 1 n Calc.	Set 2 n Ast.	Set 3 n Phal.
Avahi laniger	1	1	–
Microcebus murinus	–	–	1
Cheirogaleus major	1	1	2
Mirza coquereli	–	–	1
Daubentonia madagascariensis	–	1	–
Eulemur fulvus	2	2	1
Hapalemur griseus	3	3	1
Indri indri	2	2	1
Lemur catta	3	3	1
Lepilemur mustelinus	3	3	–
Propithecus verreauxi	2	2	1
Propithecus diadema	1	–	1
Varecia variegata	–	1	1
Galago senegalensis	–	–	2
Otolemur crassicaudatus	–	–	2
Loris tardigradus	–	–	–
Nycticebus coucang	–	–	–
Perodicticus potto	–	–	–
Alouatta seniculus, sp.	4	3	–
Aotus azarae, infulatus, sp.	3	3	2
Ateles paniscus, sp.	3	3	–
Brachyteles arachnoides	1	1	–
Cacajao calvus	2	2	–
Callicebus donaco., moloch	3	3	–
Callimico goeldi	2	2	–
Callithrix jacchus	2	2	–
Cebuella pygmaea	2	3	–
Cebus apella, sp.	2	3	–
Chiropotes satanus, sp.	3	–	–
Leontopithecus rosalia	2	1	–
Pithecia monachus, pithecia	2	0	–
Saguinus midas, mystax, sp.	4	0	–
Saimiri boliviensis, sciureus, sp.	5	0	–
Cercopithecus sp.	2	2	–
Chlorocebus aethiops, cynosuros	2	1	–
Colobus geureza	1	0	–
Erythrocebus patas	1	0	–
Lophocebus albigena	1	0	–
Macaca nigra, tonkeana	2	2	–
Mandrillus sphinx	1	0	–
Nasalis larvatus	1	1	–
Papio hamadryas	2	–	–
Piliocolobus badius	1	–	–
Pygathrix nemaeus	1	–	–
Theropithecus gelada	1	1	–
Trachypithecus obscurus	1	1	–
Gorilla sp.	1	1	–
Hylobates lar	1	1	–

Fossil samples:

Set 4 n Ast.	Taxon	Set 1 Calc. Cat. #	Set 2 Ast. Cat. #	Set 3 Phal. Cat. #
–	Cantius abditus	USGS 6783	USG 21832	–
1	Cantius sp.	USGS 6774	–	–
–	Cantius trigonodus	AMNH 16852	–	–
–	Cantius trigonodus	USGS 21829	–	–
1	Cebupithecia sarmientoi	UCMP 38762*	UCMP 38762*	–
1	Marcgodinotius indicus	GU 709	GU 748	–
1	Mesopithecus pentelici*	MNHN PIK-266	–	–
–	Neosaimiri fieldsi*	IGM-KU 89202	IGM-KU	–
1	Neosaimiri fieldsi*	IGM-KU 89203	–	–
1	Notharctus sp.	AMNH 55061	AMNH 11474	–
–	Notharctus tenebrosus	AMNH 11474	AMNH 129382	AMNH 143612-3
–	Omomyid	AMNH 29164	UM 38321	–
–	Omomys sp.	UM 98604	UM 98648	–
–	Oreopithecus bambolii	NMB 37*	–	–
–	Ourayia uintensis	SDNM 60933	–	–
1	Parapithecid	DPC 15679	DPC 5027	–
1	Parapithecid	DPC 20576	DPC 5416A	–
1	Parapithecid	DPC 2381	DPC 1001	–
1	Parapithecid	DPC 8810	–	–
1	Proteopithecus sylviae	DPC 24776	DPC 22844	–
1	Smilodectes gracilis	AMNH 131763	–	–
–	Smilodectes gracilis	AMNH 131774	–	–
1	Teihardina belgica	IRSNB16786-03	IRSNB16786-01	–
1	Washakius insignis	AMNH 88824	UM 99704	–
1	Carpolestes simpsoni	–	–	UM 101963 (x4)
1	Ignacius clarksforkensis	–	–	UM 82606
–	Plesiadapis churchilli	–	–	SMM P77.33.517
1	Nannodectes intermedius	–	–	USNM 442229
1	Incertae sedis	–	–	6 from UCMP
1	TOTAL fossil N:	24	14	14

TABLE 1. (continued).

Extant	Set 1	Set 2	Set 3	Set 4	Fossil	Set 1	Set 2	Set 3
Taxon	n Calc.	n Ast.	n Phal.	n Ast.	Taxon	Calc. Cat. #	Ast. Cat. #	Phal. Cat. #
Pan troglodytes	2	2	—	—				
Pongo pygmaeus	1	1	—	—				
Symphalangus syndactylus	1	1	—	—				
Tarsius pumilus	—	—	2	—				
Tarsius bancanus	—	—	2	1				
Tarsius spectrum	—	—	2	1				
Tarsius syrichta	—	—	—	1				
Cynocephalus volans	—	—	—	2				
Galeopterus variegatus	—	—	—	1				
Ptilocercus lowii	—	—	—	2				
Tupaia glis	—	—	—	2				
Lepus sp.	—	—	—	2				
Sylvilagus sp.	—	—	—	1				
Ochotona princeps	—	—	—	1				
Erethizon sp.	—	—	—	1				
Coendou prehensilis	—	—	—	1				
Marmota sp.	—	—	—	1				
Sciurus sp.	—	—	—	1				
Aplodontia rufa	—	—	—	1				
Allactaga major	—	—	—	1				
Hemiechinus auritus	—	—	4	1				
Erinaceus europaeus	—	—	3	1				
Erinaceus roumanicus	—	—	4	—				
Chrysochloris asiatica	—	—	—	1				
Crocidura olivieri	—	—	—	1				
Desmana moschata	—	—	—	1				
Solenodon paradoxus	—	—	—	1				
Potos flavus	—	—	—	1				
Arctictis binturong	—	—	—	1				
Nasua narica	—	—	—	1				
Petrodromus tetradactylus	—	—	—	1				
Tenrec ecaudatus	—	—	—	1				
Setifer setosus	—	—	—	1				
Hemicentetes semispinosus	—	—	—	1				
Echinops telfairi	—	—	—	1				
Potamogale velox	—	—	—	1				
TOTAL extant N:	82	66	34	52				

TABLE 2. Comparison between traditional 3DGM of 106 calcanei sample and FAA of this study. See Gladman et al. (2013) for original manual analyses and definition of anatomical terms.

Comparison point	27 landmark—Manual analysis	1,024 landmark—Automated
PC 1% variance	35.9	34.7
PC 2% variance	13.6	13.6
PC 3% variance	9.5	6.7
PC 4% variance	6.7	4.6
Sum PC 1–4	64.9	59.6
PC 1 loadings	Overall width/length proportions with emphasis on distal elongation.	Overall width/length proportions with emphasis on distal elongation.
PC 2 loadings	Position of lateral peak of the peroneal tubercle relative to both ectal and cuboid facets.	(1) Dorsoplantar elevation of the ectal facet's distal margin relative to the calcaneus body; (2) distinctiveness, but not position, of peroneal tubercle.
PC 3 loadings	(1) Proximal segment elongation, shape/orientation of ectal facet, (2) dorsal projection of dorsal heel.	Tradeoff between a prominent proximal plantar heel process and an accentuated angulation at the distal plantar tubercle.
PC 4 loadings	Ectal facet position, curvature, and orientation relative to long axis of the calcaneus.	Proximal elongation and dorsal projection of dorsal heel.

TABLE 3. Correlation (r) and Probability (P) between manual and automated PCs

	Absolute values of linear correlations (r)			
Manual	Automated Pseudolandmarks			
3DGM	PC-1	PC-2	PC-3	PC-4
PC-1	0.96	0.16	0.09	0.07
PC-2	0.11	0.50	0.34	0.28
PC-3	0.15	0.64	0.03	0.18
PC-4	0.01	0.06	0.38	0.32
	Probability of no correlation (P)			
Manual	Automated Pseudolandmarks			
3DGM	PC-1	PC-2	PC-3	PC-4
PC-1	<0.0001	ns	ns	ns
PC-2	ns	<0.0001	0.0004	0.0042
PC-3	ns	<0.0001	ns	ns
PC-4	ns	ns	<0.0001	0.0008

of which have been multiplied by the rotation matrix used in the final alignments.

The purpose of the alignment file is to check for errors generated by the alignment algorithm. If errors are found, we provide functions allowing for a semi-supervised repair, though most likely such errors indicate insufficient degrees of incremental variation in the dataset (i.e., the morphological gaps between a single specimen, or certain groups of specimens, and the rest of the dataset are too large). The purpose of the MDS file is to provide a quick view of the phenetic affinities suggested by the matrix of continuous Procrustes distances between specimens in the analysis. The *morphologika*[2.5] file allows further analyses of the sample of shapes as aligned by the method. Finally, the aligned versions of the input files provide data for users who wish to standardize alignment before taking manual measurements that are sensitive to orientation (like relief indices or other topographic variables measured on teeth [Bunn et al., 2011]), or who wish to use the images for figure generation.

Pseudolandmarks and Alignment

In order to facilitate adoption of this method by 3DGM community, this protocol represents and aligns pairs of surfaces with landmark-like feature points. We say these are "landmark-like" because we represent each bone with same number of points (in this study 1,024 points per bone are used, but the algorithm can be set to use more or fewer), and by the final stage of the algorithmic protocol each point has a fairly consistent biological identity across all bones of the sample. Each of these points is therefore analogous to an observer-placed landmark. On the other hand, they are not identified based on any of the criteria for determining type I, II, or III landmarks (Zelditch et al., 2004), or even semilandmarks (Bookstein, 1997; Mitteroecker and Gunz, 2009), and therefore are dubbed "pseudolandmarks" here. Other recent, fully automated algorithms (Boyer et al., 2011) do not generate a globally consistent mapping of a set number of points across all specimens of a dataset, and this limits their utility for certain applications.

Major Computational Steps

There are at least four important ingredients to the protocol. The first is re-sampling of surface coordinates to a specified standard number of points (Fig. 4). This is done using approaches that evenly spread points over the surface (Eldar et al., 1997). Once a new sample of bones with a standard number of evenly spread coordinates has been generated, the algorithm attempts to align each pair of bones using an iterative closest points (ICP) procedure (Besl and McKay, 1992). We avoid incorrect local minima known to plague ICP by having our algorithm assume that principal axes of variation will tend to be homologous in some sense between bones. After computing the principal axes of variation in points for two surfaces, the algorithm attempts alignments where the first principal axes are aligned in one of two possible ways (Fig. 5). There are a total of eight ways to align the first through third principal axes, and these
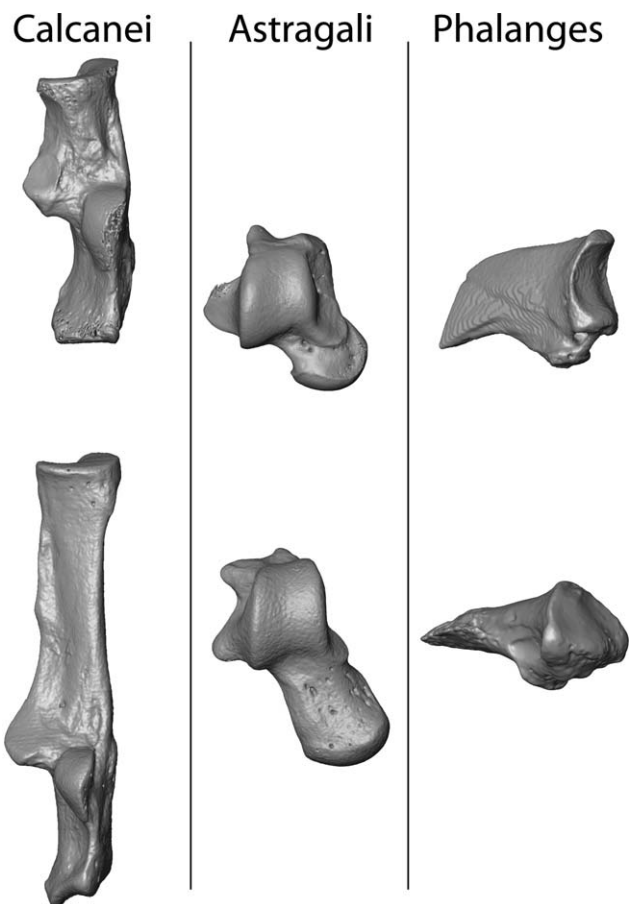
Fig. 1. Bones of the study. This study utilizes scan datasets of three different types of bones. These datasets are chosen to challenge the automatic alignment algorithm we present with a range of geometric properties. The astragalus and calcaneus datasets are samples that represent geometrically complex bones with seemingly modest sample variance, while the distal phalanges are geometrically more simple bones with apparently large inter-bone variance. Analyses include one on a sample of 106 calcanei that is compared with a traditional 3DGM analysis using 27 landmarks by Gladman et al. (2013); one on a sample of 80 calcanei and 80 taxon-matched astragali in a single "mixed-bone" analysis; and one on a sample of 49 distal phalanges (Table 1).

eight possible alignments are our starting points for ICP. They can be run simultaneously, and an approximation of the global minimum Procrustes distance can be found quickly (especially if a low number of pseudolandmarks are used). Of course, a major advantage of the method is the ability to include large numbers of data points on the surface. To resolve the conflict between processing speed and accuracy, our algorithm performs initial alignments with highly down-sampled surfaces using several hundred points (the exact number of pseudolandmarks is a user-defined parameter). Next, more densely sampled surfaces are rigidly transformed to match their down-sampled counterparts, so that only the final "tweaking" of the alignment has to be performed on the full-resolution surface file.

Since the best alignment is found by computing a Procrustes distance, a Procrustes distance matrix is available for computation of a minimum spanning tree (MST) for the sample. The MST connects all cases in the dataset using the shortest edge length possible and is a unique

solution, except in datasets where several cases are exactly equidistant from each other. Though not all points will be connected to their nearest neighbors in such a tree, most or all connections represent a joining of nearest neighbors for one of the cases involved. In datasets with high degrees of shape diversity, it is virtually guaranteed that between certain pairs of bones, the minimum Procrustes alignment will be a biologically meaningless arrangement. However, because the segments of the MST connecting pairs are among the shortest in the distance matrix, they are the most likely to be biologically meaningful and/or precise alignments. Therefore, instead of attempting to directly align pairs of shapes that have a relatively large minimum Procrustes distance separating them, alignments between such pairs are generated by propagating alignments between intermediate shapes following connections of the MST, ultimately allowing very different shapes to be aligned indirectly (Fig. 6).

Parameters That Must be Specified

Before the "automated part" of our algorithm can begin, the user must choose values for three parameters. Varying values of these parameters (see below), improves fidelity, detail, and accuracy of alignment in the one direction, and speed of calculation in the other. It may be possible to determine optimal values for these parameters in more or less general conditions by incrementally modifying them, re-running analyses, and checking the results. We have not yet done this systematically; however, replicate analyses with increasing numbers of final pseudolandmark points yield increasingly consistent results. Using the dataset of 116 teeth of Boyer et al. (2011), we found that the coefficient of determination (r^2) between PC1 scores of replicated analyses was 0.85 for 128 points, 0.92 for 200 points, and 0.95 for 1000 points. On the other hand, the coefficients of determination between PC1 scores of the 128 and 1000 point analyses are 0.82–0.85, while those between 200 and 1000 point analyses are 0.91–0.95. At the very least, these preliminary checks show that increasing the number of final points is desirable and that the results of the method should not be considered deterministic. We were able to determine that the loss of identity between runs happens during the initial step in which pseudolandmarks are spread over each surface. If alignment and pseudolandmark propagation procedures are rerun on the same pseudolandmark set, results are identical for any pseudolandmark point number. Whether manually re-collecting a set of traditional landmarks would yield better or worse correlations is probably dependent on the observer and the diagnostic precision of the landmarks collected.

The parameters to be set include (1) the number of points used to represent shapes in the low resolution version of the alignment; (2) the number of points to represent shapes in the high-resolution, or final version of the alignment; and (3) the number of principal alignments (usually this number is set to the eight possible combinations of the alignments along the first three principal axes, but additional random principal alignments can be chosen). In the first three samples we evaluate in this study, we use the following pairs of point numbers: Calcaneus dataset of 106 specimens: initial = 150 points, final = 1,024 points, 8 principal
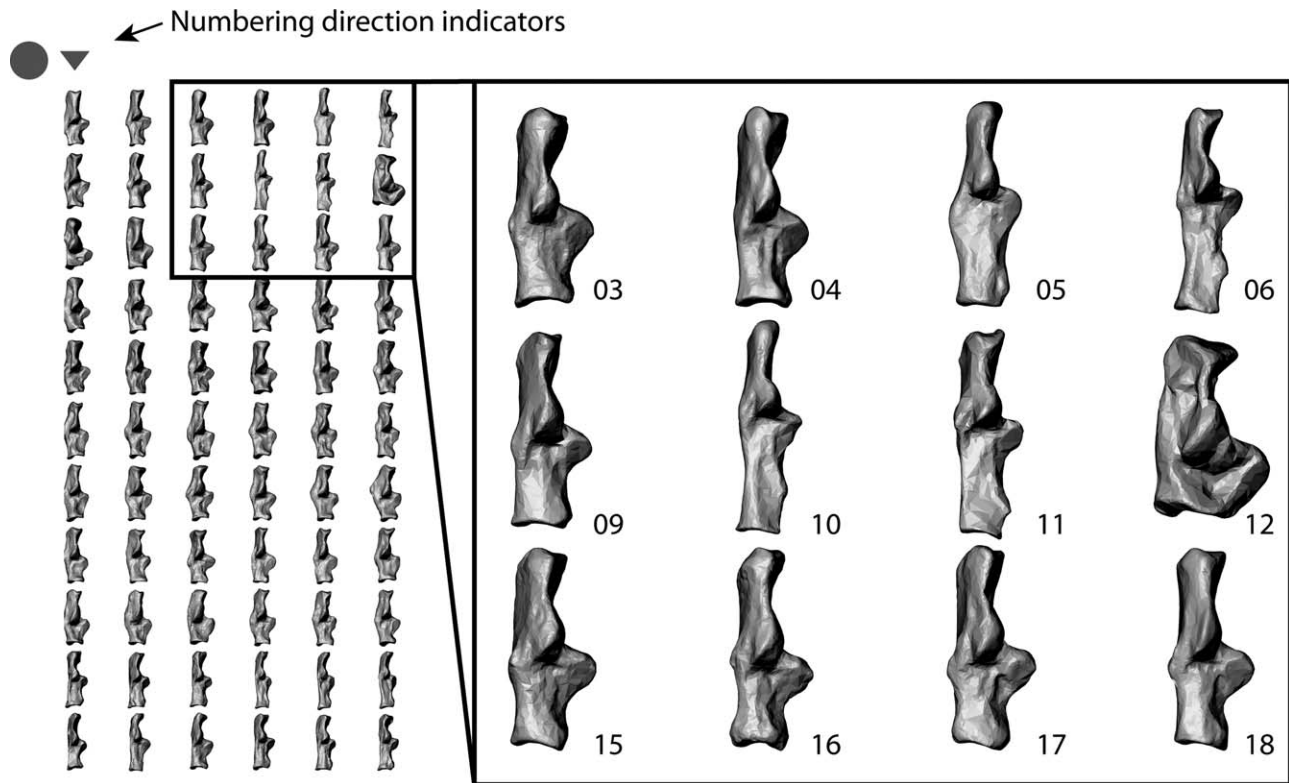
Fig. 2. Example of bones in an alignment file. One of the outputs of the fully automated alignment algorithm is a 3D mesh file that shows all the specimens of the sample aligned. This allows the researcher to quickly survey the results to determine if he/she should proceed with shape analyses based on the implied correspondence. Sometimes one or more bones may be misaligned. If this results the researcher will catch it at this stage: we present several strategies for correcting such misalignments. The "numbering direction indicators" are mesh objects that show where the #1 bone in the spreadsheet is located. The arrow points down column #1, and numbering proceeds down rows. This allows the researcher to match bones in the alignment file with a spreadsheet containing any metadata on the surface files (like taxonomic information).

alignments; paired calcaneus and astragalus datasets: initial = 256 points, final = 1,024 points, 12 principal alignments; distal phalanx dataset: same as for paired astragalus and calcaneus. In the fourth dataset we use far fewer points in order to generate problematic alignments: initial = 32, final = 64, 8 principal alignments.

Fixing Errors in the Alignment Protocol

Because it is sometimes the case that at least one specimen is mapped into the MST with an incorrect alignment, it is important to provide options for correcting the problem.

1. Usually such problems stem from insufficient number of initial points (first parameter above). Thus, the first step is to try re-running the initial steps of the algorithm with slightly greater numbers of points per file. However, the problem can also stem from the lack of an adequately similar partner shape in the dataset (from the perspective of its orientation and articulation in the skeleton). This shape represents an "island shape" for which the best geometric alignment (that with the smallest Procrustes distance) to any other shape is a biologically "incorrect" alignment. This property does not guarantee a bad alignment in the final stage of the algorithmic protocol since it may not

connect to its nearest neighbor in the MST, but usually a bad alignment is expressed nonetheless. However, it is possible that there are still some shapes in the sample with which the island shape(s) will correctly align. We do not currently have an automated protocol for discovering such shapes, if they exist. We have implemented two different protocols for fixing alignment problems. If there is a **single** misaligned shape: We allow the user to display the results of direct alignments of the island shape to each of the other shapes in the sample using the function branch_pw_distances.r in the R-package. If there are n specimens in the sample, this function creates $n - 1$ multi-surface mesh files. There is one file for every corresponding pair between the island shape and the remaining shapes. Even if n is very large, these can be visually scanned quickly to find a correct alignment. Tiling the multiple files in Meshlab or Aviso is one possible way of quickly arriving at the correct alignment when n is large. If the user finds a shape to which the island shape correctly aligns, the MST is recalculated without the island shape, the global alignment of the remaining shapes is double-checked, and the island shape is connected to the new MST through its successfully aligning partner. The analysis is then completed in the usual way. If there are multiple specimens with which the island shape correctly
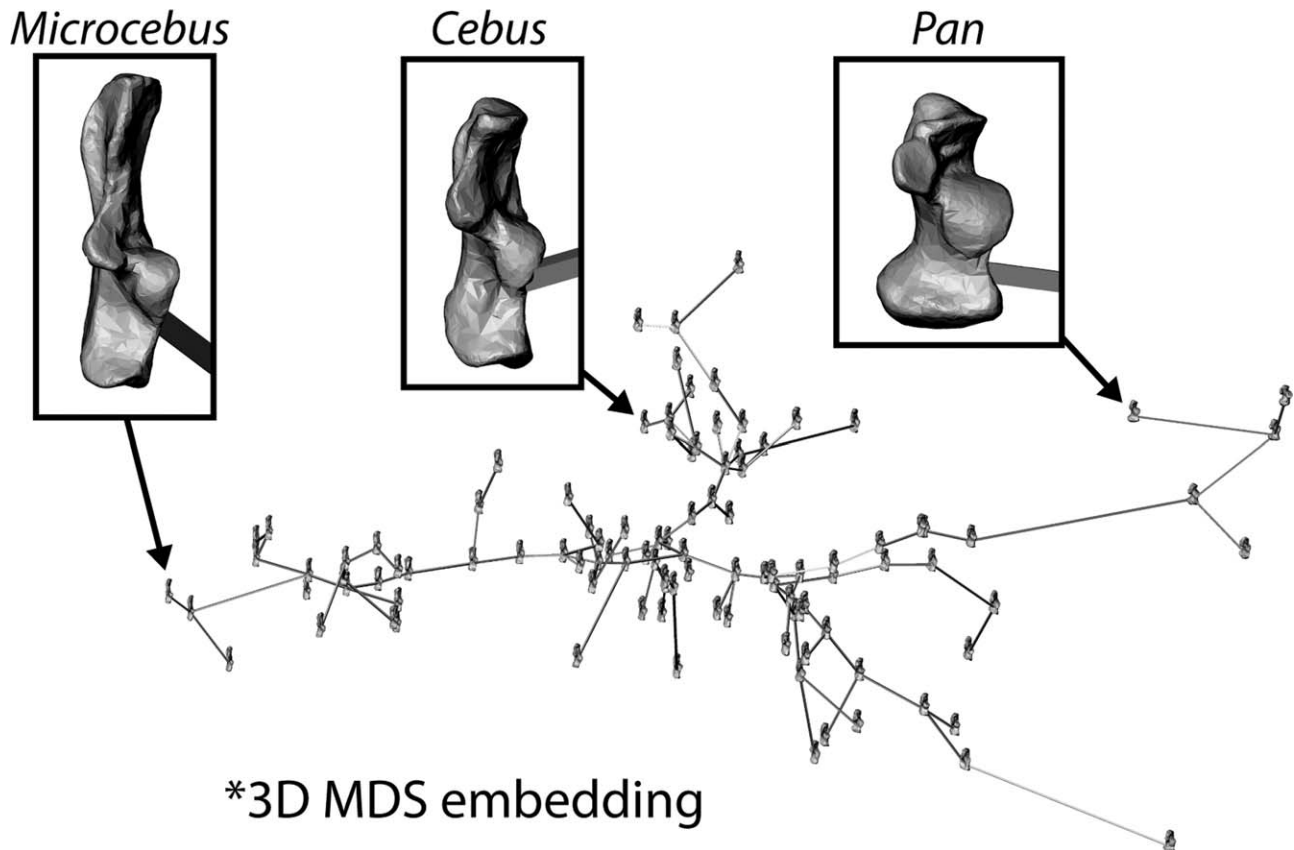
Fig. 3. MDS and MST embedding file. This second output is of the same file type as that in Figure 2. It is however, less essential, because it is not useful for visualizing alignments and the data it presents can be re-calculated by the user later. The file simply displays the bones of the sample with their centroids embedded in the coordinate space of an MDS analysis result that is run on the pairwise distance matrix as determined via the MST. The MST is also shown. The point of this file is to give researchers a quick look at the clustering of their specimens.

aligns, the user can choose which to use as the connecting shape, though it seems logical to choose that with the smallest Procrustes distance to the island shape. The pairwise output files from branch_pw_distances.r orders the shape correspondences by their Procrustes distance. The ordering of correspondence will be in the name of the files for clarity.

2. If there are **multiple** island shapes, a more involved protocol is required, because there may be several groups of consistently aligned shapes (Fig. 7). The general problem is that the analysis may return a result in which certain branches are internally consistent, but are misaligned with respect to other such branches. It is therefore necessary to have a protocol allowing the user to chop apart these branches and stick them back together in a way that ensures a globally consistent alignment. The work-flow described below is provided by the example file "alignFix.r" and is available on the first author's website. Documentation that accompanies "alignFix.r" guides the user through a sample problematic dataset (our dataset 4). Users should then be able to edit the code of "alignFix.r" to suit their datasets.

a. Observe misaligned regions using alignment and map files (Fig. 7A and B) together.

a.i. If only one misaligned file is observed, follow the procedure described above.

a.ii. If more than one misaligned file is observed:

    a.ii.1. Record the names of the misaligned files.

    a.ii.2. View the MDS graph showing the MST connections on points labeled by the file they represent.

b. Using the MST displayed in the map image, figure out how many "groups" of misaligned files exist, and how many specimens in each group, and record this information.

b.i. Specify all "groups greater than 2" (three or more files that are correctly aligned to each other, but not to surrounding shapes) as "groups to analyze separately," since a MST will need to be re-computed within each group.

c. For "b.i.," a separate alignment analysis is run on each group of three or more that were internally consistent and all the necessary information is saved (Fig. 7C).

d. Now the user must decide how to "re-connect" the separate sub-groups.

d.i. First attempt to analyze all of the shapes in non-connected segments of the MST. For
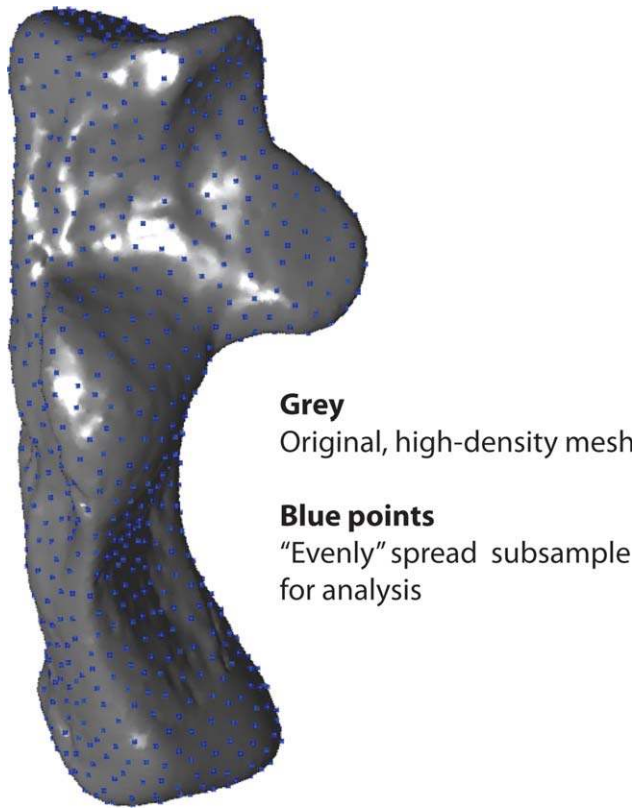
Grey
Original, high-density mesh

Blue points
"Evenly" spread subsample
for analysis

Fig. 4. Down-sampling meshes prior to analysis. The algorithm is run on point clouds represented by a standard number of points specified by the researcher. These points are chosen by the algorithm randomly picking a point on the surface, and then picking another point that is farthest from the first point, then by picking a third point whose position on the surface maximizes the sum distance between it and the two existing points, and so on until the specified number of points is achieved.

example, with four groups (A, B, C, and D), it is possible that only one will end up connecting to the other three through the MST. If both A, C and D connect to B in the original analysis, and are misaligned with respect to B, it is possible that with B excluded, A, C and D will align correctly. If this is true, skip to "d.iv.1" of this description. If not, go to number "d.ii."

d.ii. For cases in which the set of non-connecting groups is still an incorrect alignment, the non-connecting groups should be compared in a pairwise fashion. For instance A–C, A–D, and D–C should each be analyzed separately. It is possible that some of these will have overall correct alignments. If more than two of these are correct, a decision will have to made on which two to merge, since it has already been demonstrated that all three cannot be. We would suggest merging the two that result in the biggest difference in the number of specimens represented in the final two groups, since this makes the subsequent task of searching for a correct alignment between groups that are not correct via their MST easier. At this stage, the goal should be

to merge as many isolated groups together as possible in order to reduce computational demand in the next steps. Ultimately, the user can decide which groups to merge.

d.iii. After managing the isolated but internally consistent segments of the original MST (groups A, C, and D above), the user needs to find a "correct" connection between the isolated groups that were misaligned with respect to each other through the original MST. Some remnant of the original MST will still be preserved, which can be called the "base tree" (group B in our example). Attempting to reconnect the isolated groups to the base tree using the minimum distance pair will likely generate misalignments, since the MST connections were wrong in the original analysis. However, as MST connections often only represent nearest neighbors for one of the two connected cases, there is still a possibility that one of the cases involved in the incorrectly aligning connection between the base tree and another segment was not connected to its nearest neighbor. This makes it important to look at the minimum distance pairs of the isolated groups and the base tree.

d.iv. Assuming the minimum distance pair is still a misalignment, a protocol for checking alignments between particular shapes in each group must be implemented. This again utilizes the function branch_pw_distances.r.

d.iv.1. The user has the option to check all alignments. The output is n x m "summary alignment files" in which n is the number of specimens in one group and m is the number in the other group being searched. Each file shows one shape from the group with n with one of the m specimens of the second group (Fig. 7E). The output files are labeled according to minimum Procrustes distance, so that the first compared specimens are nearest neighbors. The user can then easily identify the correctly aligning pair that also has the minimum Procrustes distance (since there may be more than one correctly aligning pair).

d.iv.2. This process should be repeated for all segments that could not be merged. If there were three remaining segments (e.g., a base tree B, an A–C group and D), there will likely be an option of whether to link each tree to one of two others. We would suggest this linking be done using the option that minimizes the Procrustes distance between the linking pair.

d.iv.3. The user can also opt to only compare specific specimens from one group to specific specimens in the other.

d.iv. Finally, all groups are re-aligned using a tree that represents each separate MST connected along user-specified pathways in "*d.iv.2*" This should result in correct alignments for all bones in the sample (Fig. 7G).

If the user determines that successful alignments between groups of island shapes are impossible, there are two options: (1) remove any island shape groups from the analysis (particularly if their inclusion does not directly
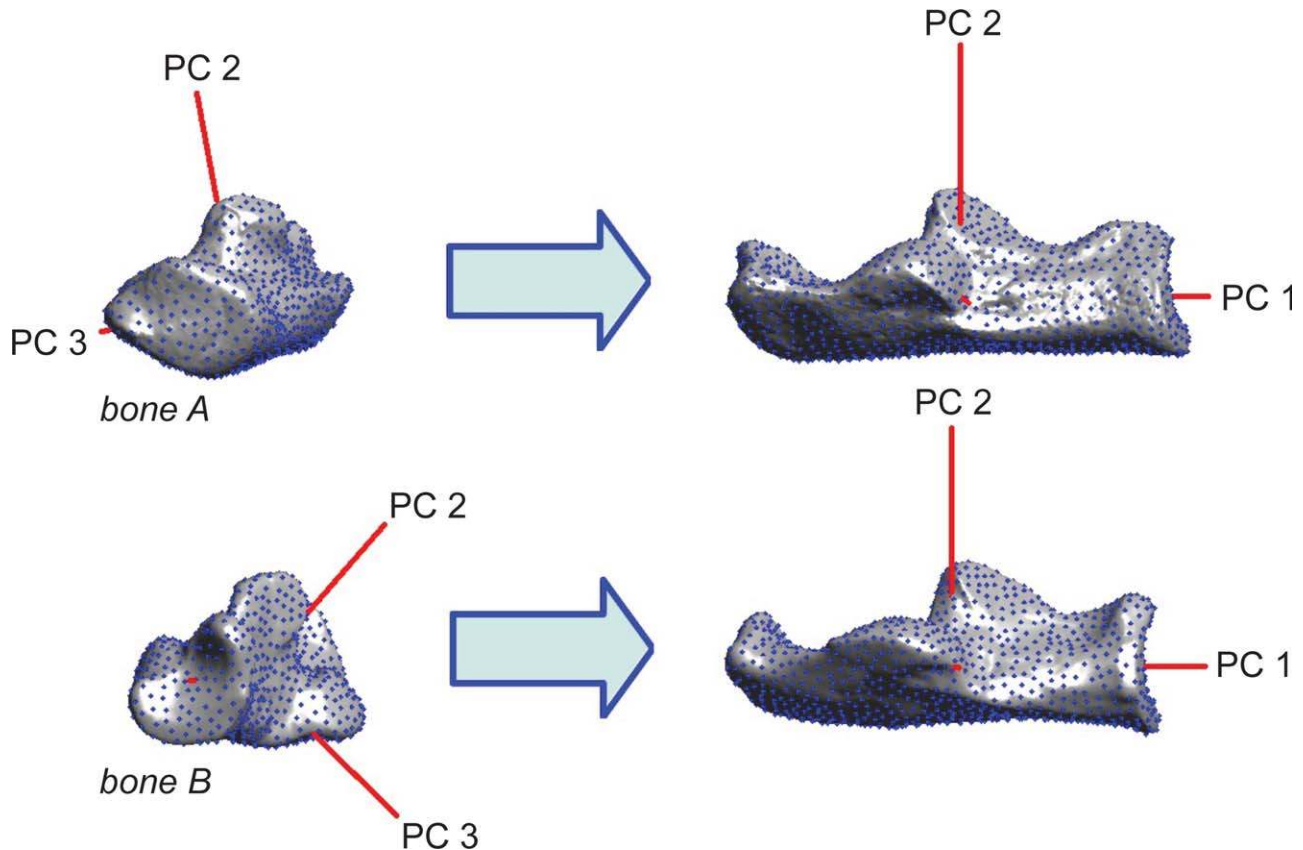
Fig. 5. Principal alignments to improve ICP searches. The best alignment between two bones is almost impossible to find using an ICP approach without any good initial guesses. The problem with supplying an initial guess is that usually this means user intervention is required. Our algorithm supplies at least eight initial guesses without user intervention. It does this by computing the first three principal axes of variance and uses these axes as starting alignments for ICP. The principal axes along which the smallest continuous Procrustes distance between two shapes is found is almost always correct if the shapes are similar. This is a computationally rapid way of solving a complex problem. The algorithm performs better on samples with many incrementally intermediate shapes (see text and Fig. 4). *Red lines* on calcaneal surfaces represent principal axes of point variance. Shapes on left have yet to be aligned, while shapes on the right have been aligned so that their principal axes match.

address the main questions of the analysis); or (2) add more shapes with the hope of bridging distances between island shapes.

Getting the Code for Running Analyses

The R package we developed is called *auto3dgm*. At the time of publication *auto3dgm* has been submitted to CRAN for review, and should ultimately be accessible from their repositories. Until then, *auto3dgm* can be downloaded at www.dougmboyer.com or http://www.stat.duke.edu/~sayan/3DGM/index.shtml. The sample/instructional file for fixing misaligned shapes, alignFix.R, is not part of the R-package itself and will not be available on CRAN. It can, however, be downloaded from the personal websites mentioned above. Documentation for the packages can be found at these sites as well.

Comparison to Results From Traditional Landmarks

In order to maximize our ability to compare and contrast shape information provided by our pseudolandmarks with traditional geometric morphometric datasets, we used the same sample and performed the same analyses on the pseudolandmarked dataset as Gladman et al. (2013) conducted using 27 landmarks and traditional 3DGM techniques.

First, the 3D pseudolandmark coordinate-scaled output file from our algorithm was imported into *morphologika*$^{2.5}$. We then ran a General Procrustes Analysis (GPA) with reflections enabled, followed by a Principal Components Analysis (PCA) with "Full Tangent Space Projection" checked for Calculation Options and "Eigenvalues" and "PC Scores" checked for Printing Output Options. The results were saved as a .csv file that included the PCA output, along with the raw Procrustes distance data in the form of 3D coordinates for each landmarked individual. In *morphologika*$^{2.5}$, the cloud of 1,024 landmarks was visualized and the morphospace of the PC axes was explored. In the traditional 3DGM analysis of this sample, Gladman et al. (2013) added wireframes to the landmarks in order to directly visualize shape changes. Due to the number of pseudolandmarks used by *auto3dgm*, wireframes are currently impractical, but shape changes can easily be observed from transformations of the densely packed pseudolandmarks. All Principal Components (PCs) were examined in *morphologika*$^{2.5}$ by tracking changes in the cloud of
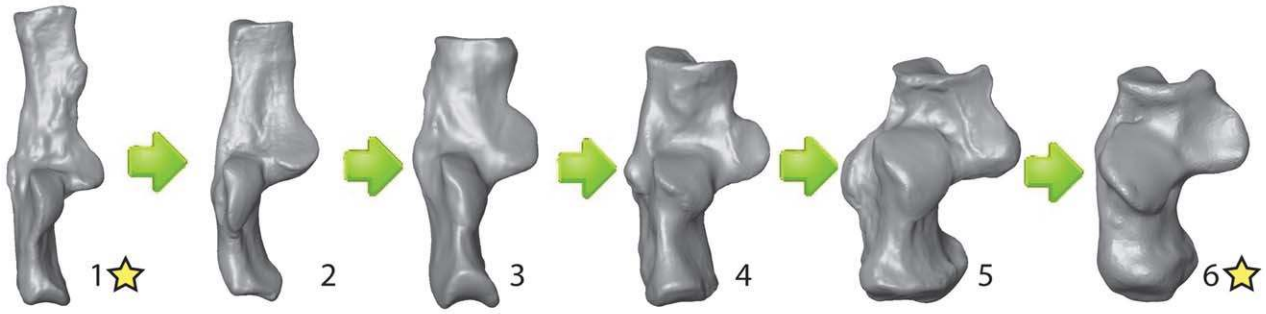
A. Similar objects align correctly
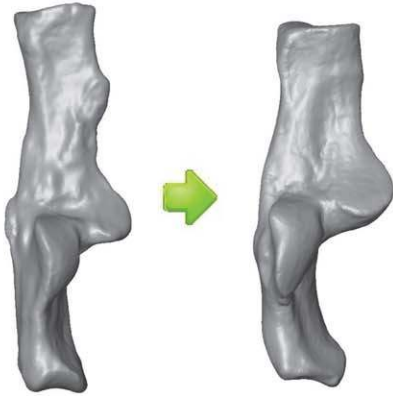
B. Dissimilar objects may not align correctly
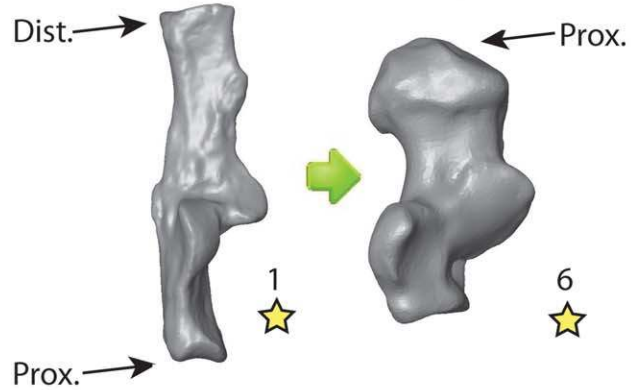
C. Alignments of dissimilar objects are improved if constrained by intermediates

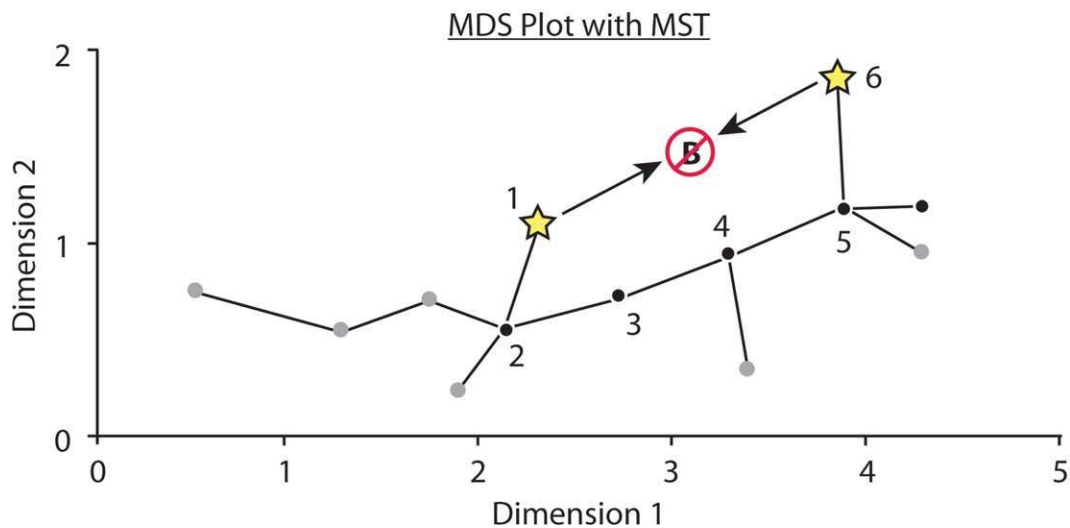## D. Path for intermediate shapes given by minimum spanning tree of pairwise distances

MDS Plot with MST

Fig. 6. Method for successfully aligning disparate shapes. (**A**) the result of applying our version of ICP to two similar shapes. (**B**) the incorrect result that emerges when applying our ICP directly to two dis-similar shapes. In the first stage of the analysis, a pairwise distance matrix is calculated using "direct matches" (even potentially incorrect ones as in B) between all shapes. That distance matrix is used to compute a MST. Because the MST connects only the most similar shapes, these connected pairs almost always represent correct alignments as in "A." (**C**) These connections therefore define a path of intermediates that can be used to figure out the correct alignment between different shapes. (**D**) The MST route is shown graphically.

3D landmarks between the extreme regions of morphospace on each axis. The amount and nature of variation represented by these axes in the 1,024 pseudolandmark dataset was then compared with results from the 27 user-determined landmarks of the Gladman et al. (2013) analyses.

Gladman et al. (2013) also used analyses of "generic" means for cluster analyses in their study of the 106 calcanei sample. They felt that averaging the few individuals for each genus helped control for any extreme variation that might otherwise dominate the small samples being used to represent extant genera. We replicated their approach with the pseudolandmark coordinates here. Extant genera represented by more than one individual were averaged into a single genus representative (Table 1). As in Gladman et al. (2013), fossil individuals were not averaged together in the analyses. Altogether the dataset was reduced from 106 individuals to 67 generic representatives (Table 1).

In order to generate generic means, the matrix of 3D coordinate Procrustes output data (generated in *morphologika*$^{2.5}$) was imported into PAST statistical software (Hammer et al., 2001, 2006). In PAST, all individuals of a single genus were highlighted and averaged using the "Evaluate Expression" function in the "Transform" menu. "Mean (of current column)" was selected in the "Evaluate Expression" menu and then "Compute" in order to change all highlighted rows to the same averaged values. Only one of these newly averaged rows was kept in the dataset to represent a given genus. This technique can be done manually by averaging each X, Y, and Z value separately for each landmark for members of each genus, although with increasingly larger datasets this becomes untenable. Once averaging of the dataset was complete, cluster analyses were run within PAST and then compared with the generic mean analyses of Gladman et al. (2013).

Mixed Bone Analysis

It has been suggested that traditional 3DGM methods could be used to "pool information" from more than one structure (Rohlf, 2002). However, the meaning of results from such an approach could be questioned, since the weight of each structure added will depend on the user's choice of landmarks, as well as the number of landmarks used to represent each bone. Furthermore, since there is no basis for collecting equivalent landmarks across bone types, it has never been possible to include multiple bone types in the same 3DGM analysis using the same landmark template. Our approach with *auto3dgm*, based on spreading landmarks evenly and selecting alignments based on overall geometric similarity, provides a potential solution to this problem and allows simultaneous analysis of multiple types of bones. There are many questions that can be addressed if shape variation can be compared between bone types. For instance, one might wish to ask whether the astragalus has less shape diversity than the calcaneus, due to the former articulating with a greater number of bones and lacking muscular attachments as exhibited by the latter. One might also be interested in investigating whether the degree of overall shape variation is associated with stronger phylogenetic signal (Nunn, 2011) or stronger functional signals. We performed such a "mixed bone" analysis on a sample of 80 astragali and 80 calcanei

representing the same taxa (although sometimes composed of different specimens) and we compare intrinsic levels of overall shape variation.

The basic goal of such an analysis (given the questions above) is to provide a quantitative criterion for comparing size-standardized shape variation between two bones. Since regions on the surface of a calcaneus do not "biologically correspond" in any way to regions on the surface of the astragalus (except in the sense of articulating facets), there is no need to determine a biologically meaningful regional correspondence between them. Therefore, only the most efficient geometric alignment must be established (i.e., the alignment that minimizes the Procrustes distance). However, in a mixed bone analysis, astragali will not only be compared with calcanei, they will also be compared with other astragali. Thus, for some bones in the sample, there *is* a biologically significant alignment that must be discovered before comparisons can be finalized.

To establish a globally transitive pseudolandmark coordinate dataset for a mixed bone sample, we first ran *auto3dgm* on the calcaneus and astragalus datasets separately to produce two sets of globally consistent pseudolandmark datasets. We then performed searches for the alignment and correspondence between an astragalus and calcaneus that exhibited the minimum Procrustes distance among all such pairs in the combined dataset using the branch_pw_distance.r function. In the second step, we were only concerned with distances since no details about the alignment mattered biologically. Once we found the mixed bone pair with the smallest geometric distance separating them, we used that pair to link the MSTs of the initial analyses, creating a mixed-bone, global-correspondence, 3D pseudolandmark dataset. This dataset was imported into *morphologika*$^{2.5}$ and processed with GPA followed by PCA on the entire sample of 80 calcanei and 80 astragali, and then on three sub-samples (the 80 calcanei alone, the 80 astragali alone, and a combined sample of only 40 astragali and taxon-matched 40 calcanei), with results exported as a .csv file, and final analyses performed in PAST like other analyses above.

Procrustes distance matrices for each sample were compared. In particular, we compared the mean and standard deviation for between-bone distances in these different samples, as an estimate of disparity. Distances between each bone of a sample and the mean bone for that sample were used in t-tests comparing calcaneus to astragalus and each of these single bone samples to the combined samples (Table 4).

We used the first two axes of the PCAs in plots and correlation analyses. The goal was to assess how analyzing two different bones together affected the patterns of variance and co-variance of the resulting PC's. We also used PC scores from the combined sample of 160 bones as another way of comparing disparity between the calcaneus and astragalus: After scaling PC scores to the percentage of total sample variance each represents, we plotted the PC scores, measured the area encompassed by the convex hull surrounding data on each element separately, and compared the values for these elements.

We also computed phylogenetic signal on PC scores, as well as on Procrustes distances from the mean. Phylogenetic signal was calculated using *caper* (Orme et al., 2011) in R, and a tree based on v3 of the primate dataset
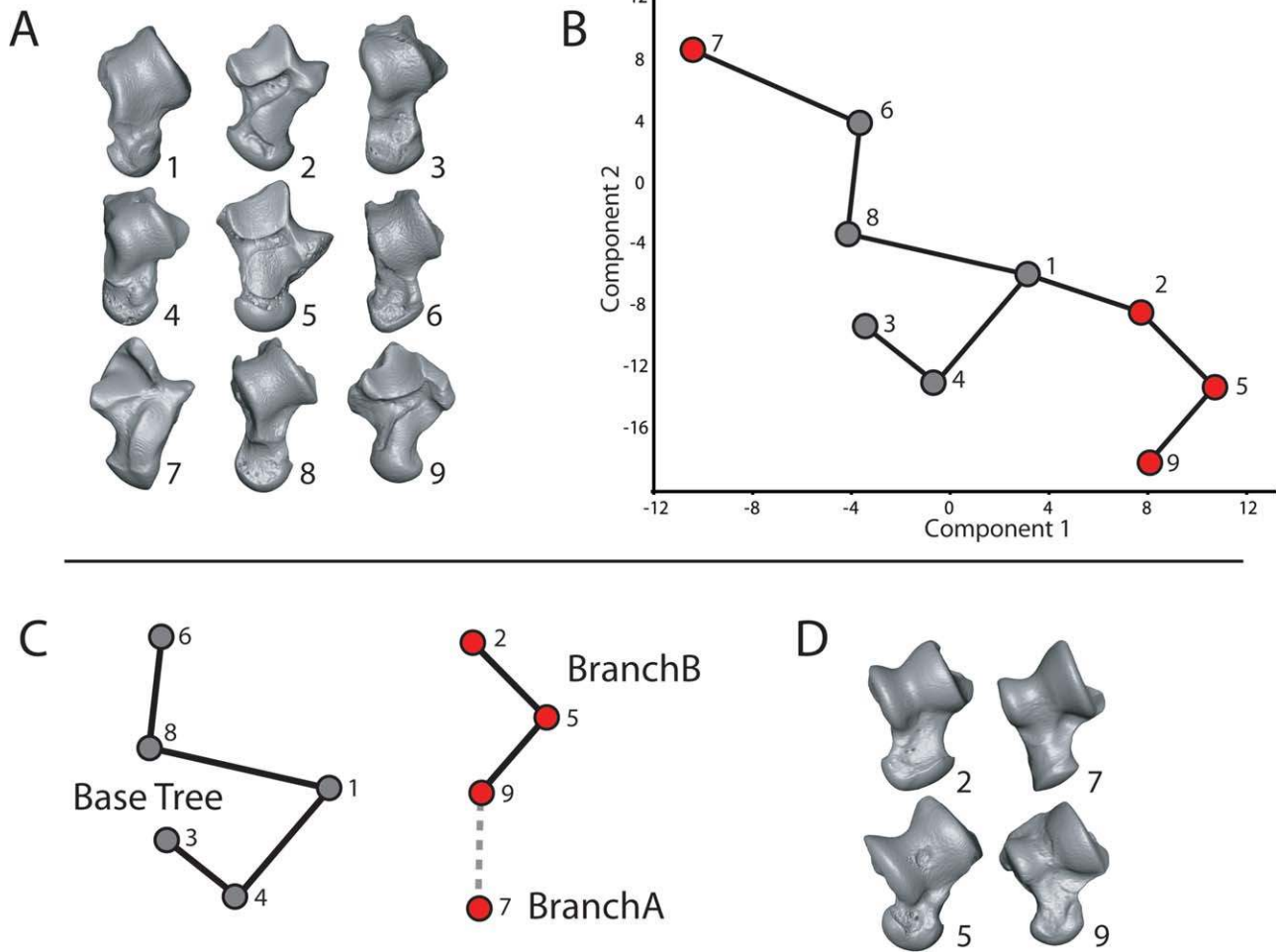
Fig. 7. Schematic of alignFix protocol. (**A**) Visual inspection of initial alignment reveals several specimens are misaligned. (**B**) MST shows misaligned specimens (shown in *red*) can be found on two branches. (**C**) MST is broken into three components representing the base tree (in which all alignments are good), and Branches A and B (the misaligned specimens). (**D**) Unsupervised alignment protocol is performed on originally unconnected branches A and B to determine if global alignment exists for those specimens when base tree specimens are excluded from consideration. Here, we show a successful global alignment. If no such alignment exists, then Branches A and B should be treated separately as if they had been a set connected to each other, as each was to the base tree. (**E**) All misaligned specimens are compared with all specimens in the Base Tree to find the appropriate attachment point (i.e., a pair with a correct alignment). Several example alignments from this exhaustive process are shown here. Pairwise comparisons are visually inspected by the user to find an acceptable alignment with the lowest Procrustes distance between the two specimens. (**F**) The designated pair serves as the connection (*dotted line*) for Branch A + B to the Base Tree. (**G**) Recomputed global alignment using the user determined tree in E reveals all specimens to now align correctly.

from 10k Trees (Arnold et al., 2010). Testing for phylogenetic signal (Pagel's λ) required using generic means of the sample and reduced the sample size from 80 individuals to 43 genus-averaged individuals. Species means were also used for correlation tests.

RESULTS

Alignment Success

Alignment for the calcaneal dataset of 106 bones was successfully accomplished with a low resolution initial alignment of 150 points, and eight principal alignments (Supporting Information Fig. 1). The final high-resolution surface alignment was based on 1,024 points. Successful alignment for the calcaneal dataset of 80 bones was accomplished with a low-resolution initial

alignment of 256 points, eight initial positions based on all possible combinations along three principal axes, and a high-resolution final surface alignment based on 1,024 points. Successful alignment for the astragalar dataset of 80 bones was accomplished with a low-resolution initial alignment of 256 points, 12 initial alignments, and a high-resolution final surface alignment based on 1,024 points (Supporting Information Fig. 2).

The distal phalanx dataset was aligned using a low-resolution initial alignment of 256 points, 12 initial alignments, and a high-resolution final surface alignment based on 1,024 points (Supporting Information Fig. 3). One specimen, UCMP 217919 (a fossil of unknown taxonomic affinities), had an incorrect alignment to its connecting shape in the MST (a tarsier second digit grooming claw, USNM 196477). We identified a correct alignment with SMM P77.33.517, a claw of
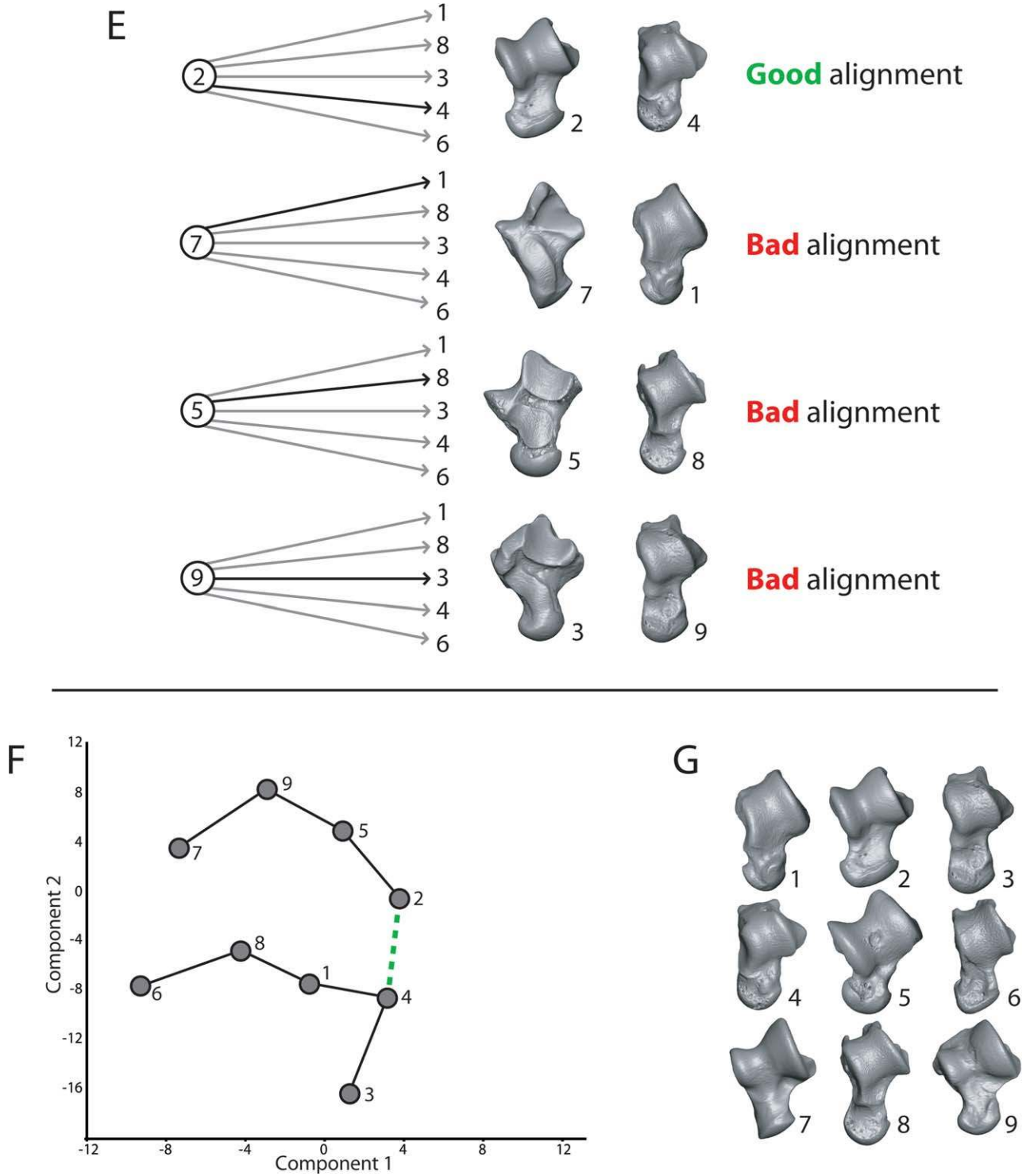
Fig. 7. (Continued)

Plesiadapis churchilli. This is not to say these two bones are very similar. It simply shows that it is usually possible to establish correct alignments for every bone in the sample without manually registering them to each other.

Comparison to Results from Traditional Landmarks

For the PCA of output from *auto3dgm* on individual specimens (N = 106, with no genus-level averaging), the

TABLE 4. Descriptive statistics for three distance matrices from mixed bone analyses

| N = 3,120 | Full distance matrix | | |
	Calc.	Ast.	Mix
Mean	0.18	0.19	0.29
Max	0.40	0.37	0.54
Min	0.05	0.06	0.05
SD	0.06	0.05	0.11

| N = 80 | Dev. from Mean | | |
	Calc.	Ast.	Mix
Mean dev.	0.13	0.13	0.21
Max	0.25	0.27	0.31
Min	0.07	0.07	0.16
SD	0.04	0.03	0.03

t-test (on Dev.)	df	t	P
Ast. vs. Calc.	158	0.50	0.62
Ast. vs. Mix	158	15.16	<0.0001
Calc. vs. Mix	158	14.81	<0.0001

The "Full distance matrix" section uses all 3,120 pairwise distances among the 80 bones in each sample. "Dev. from Mean" represents the distance between the mean bone of the sample and each of the 80 bones comprising it. Thus the number of distances is the same as the sample size. The t-tests compare means between the samples of deviations. "Mix" represents the results of analysis of 40 astragali with 40 taxon-matched calcanei. We do not present the results for the full 160 bone dataset here because the differences in sample size would complicate the meaning of differences in computed statistics relative to the other samples.

first four PC axes account for 59.6% of the total variance. This is very close to that explained in the analysis of the same sample using 27 landmarks by Gladman et al. (2013) (Table 2). Generally speaking, major clades were well separated when plotted in morphospace, as in Gladman et al. (2013) (Fig. 8). Examination of the 3D landmark cloud in *morphologika*[2.5], and the general distribution of specimens in the scatter plots of the PCA morphospace, indicates that PC1 (34.7%) is mostly associated with the overall length and width proportions of the calcanei, with some emphasis on the distal elongation. The distally elongated and narrow-bodied calcanei of omomyiforms and some strepsirrhines dominate one extreme of the PC1 axis, while the distally shorter and wide hominoid calcanei fall on the opposite extreme. This pattern matches well that found by Gladman et al. (2013). Regressing PC1 scores based on manually positioned landmarks against the PC1 scores from analysis of *auto3dgm* output showed high correlations (Table 3). Other axes were more modestly correlated or lacked significant correlations.

Variation found in PC2 (13.6%) captured some aspects of the "flexing" of the calcaneus described by Gladman et al. (2013), although the distribution of the taxa within this PC is not identical to the original description. This PC most notably varies in the position of the distal margin of the ectal facet relative to the body of the calcaneus, either raised dorsally off of the body or sunken plantarly. The hominoids are found on one extreme, with ectal facets that sit atop of the calcaneal body, while pla-

tyrrhines are the most consistent examples of calcanei with ectal facets depressed into the body. Although more difficult to observe directly from the cloud of pseudolandmarks in *morphologika*[2.5], there also seems to be variation in the magnitude, although not the position, of the peroneal tubercle captured in this axis.

The variation found in PC3 (6.7%) also resembles some of the "flexing" that has been previously described, although it also includes new variation not recognized in the previous traditional analyses. On the extremes for this PC axis are the hominoids (excluding hylobatids), which have a pronounced proximal plantar heel process and a dorsal bowing of the body of the calcaneus (giving an un-flexed appearance). At the other extreme are most of the colobines (excluding only *Colobus*), which have no proximal plantar heel process and have a more prominent plantar bowing (flexed appearance) caused, in part, by a more prominent angulation of the body at the distal plantar tubercle. The tradeoff in this axis is between an unflexed calcaneus driven by the presence of a plantar heel and a flexed calcaneus driven by a heightened angle at the distal plantar turbercle.

Finally, similar to PC3 above, PC4 (4.6%) also contributes to variation at the distal plantar tubercle. However, unlike the variation in PC3, the distal plantar tubercle in PC4 only gets larger or smaller in size, and there are no clear changes in the angulation at the tubercle. This PC exhibits variation most notably in the amount of proximal segment elongation and the position of the dorsal heel relative to the ectal facet. While PC1 contained aspects of distal elongation within the larger length and width proportional changes of the calcaneus, PC4 is specifically associated with the elongation of the proximal segment of the calcaneus, measured from the ectal facet to the end of the tuber. Additionally, at the extreme of the PC where the proximal segment is shortest, the dorsal heel is near level with the ectal facet, while at the elongated proximal extreme the heel is sub-level to the ectal facet. The fossil euprimates lie at the extremes for this variation, with omomyiforms exhibiting very low amounts of proximal elongation and the adapiforms in this sample with some of the highest levels.

Cluster analyses of the genus-averaged sample provide another way to compare the results of the analyses of *auto3dgm* generated pseudolandmarks to the results of the traditional landmark analyses reported by Gladman et al. (2013). Though there are many differences when comparing the two analyses by their various dendrograms, there are broad similarities as well (Figs. 9–11). Dendrograms for traditional landmark analysis can be viewed in Gladman et al. (2013: their Figures 9 and 10; pp. 384–386). We detail comparisons for the Neighbor-Joining (NJ) trees here, and note that similar results are obtained from comparisons between the UPGMA and Wards trees (although these latter two clustering algorithms will not be discussed further).

Similarities between the NJ tree based on Gladman et al.'s (2013) landmarks and that based on our pseudolandmark dataset (Fig. 9) include the clustering of adapiforms near the taxon chosen as the tree root, *Marcgodinotius indicus*. Additionally, extant strepsirrhines and omomyids also cluster together in both analyses. Within this cluster there are more detailed similarities: *Lepilemur* + *Ourayia* (SDNM 60933) and Omomyid indet. (AMNH 29164) + *Washakius insignis* (AMNH 88824) form two pairs of nearest neighbors, which form a unitary cluster with
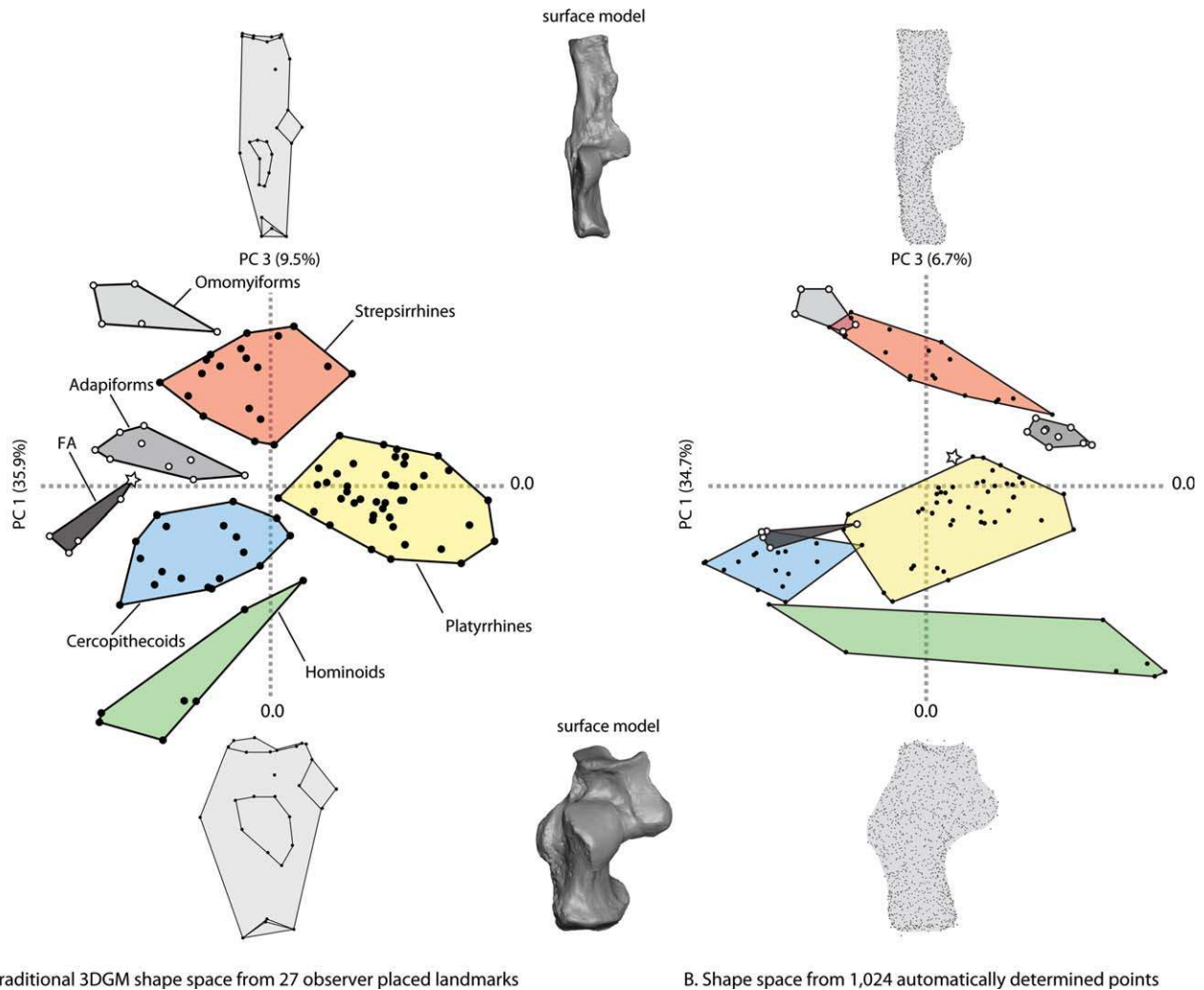
surface model

PC 3 (9.5%)

Omomyiforms

Strepsirrhines

Adapiforms

FA

PC 1 (35.9%)

0.0

Platyrrhines

Cercopithecoids

Hominoids

0.0

surface model

PC 3 (6.7%)

PC 1 (34.7%)

0.0

0.0

A. Traditional 3DGM shape space from 27 observer placed landmarks

B. Shape space from 1,024 automatically determined points

Fig. 8. Shape space of our analysis and comparison to a traditional 3DGM analysis. (**A**) PCA plot of principal component scores 1 and 3 for data from Gladman et al. (2013) based on 27 landmarks of the calcaneus in a sample of 106 bones. (**B**) PCA plot of principal component scores 1 and 3 for the same sample, but as represented by 1,024 pseudolandmark points generated by the algorithm presented here. Both datasets, including our automated output, and that from Gladman et al. (2013) were analyzed with *morphologika*$^{2.5}$. One of the benefits of the output of our algorithm is that it can be analyzed as if it were observer-collected data with traditional statistical software.

Teilhardina (IRSNB 16786-03) and *Omomys* (UM 98604) in both analyses. *Eulemur, Hapalemur,* and *Lemur* form a cluster in both analyses. *Varecia* is external to all members of the strepsirrhine + omomyiform group except *Daubentonia*. All indriids are adjacent to each other. Anthropoids form a unitary cluster separate from non-anthropoids in both analyses, and hominid and pitheciine genera form unitary clusters with respective members of their clades alone (i.e., monophyletic clusters).

Major differences include *Daubentonia* falling outside of all clusters and occupying the position closest to the root in the *auto3dgm* based analyses, whereas in Gladman et al. (2013) it clusters with other strepsirrhines. Adapiforms form a unitary cluster with strepsirrhines and omomyiforms in the *auto3dgm*-based results, whereas in Gladman et al. (2013), adapiforms formed a unitary cluster basal to all other clusters (in the position of *Daubentonia* in the *auto3dgm*-based analysis). In Gladman et al. (2013), the strepsirrhine + omomyiform cluster and the anthropoid cluster group more closely to each other than either does to the adapiform cluster. Though indriids are adjacent in both analyses, they do not form a unitary cluster in the *auto3dgm*-based analysis, and *Propithecus* groups with *Avahi*, rather than with *Indri* as in Gladman et al. (2013). In the *auto3dgm* based analysis, adapiform fossils cluster cleanly by assigned genus with four *Cantius*, two *Smilodectes*, and two *Notharctus* fossils forming three sets of unitary clusters, while in Gladman et al. (2013) these specimens are more mixed. Atelids form a unitary cluster in *auto3dgm* based analysis; in Gladman et al. (2013), they are only adjacent. Hylobatids do not cluster near other hominoids in *auto3dgm* based analysis, whereas hominoids form a unitary cluster in Gladman et al. (2013). *Proteopithecus* (DPC 24776) clusters at the base of a grouping composed primarily of platyrrhines in *auto3dgm*-based analysis, whereas it clusters at the base of, and exclusively with, Fayum parapithecid fossils in Gladman et al. (2013).
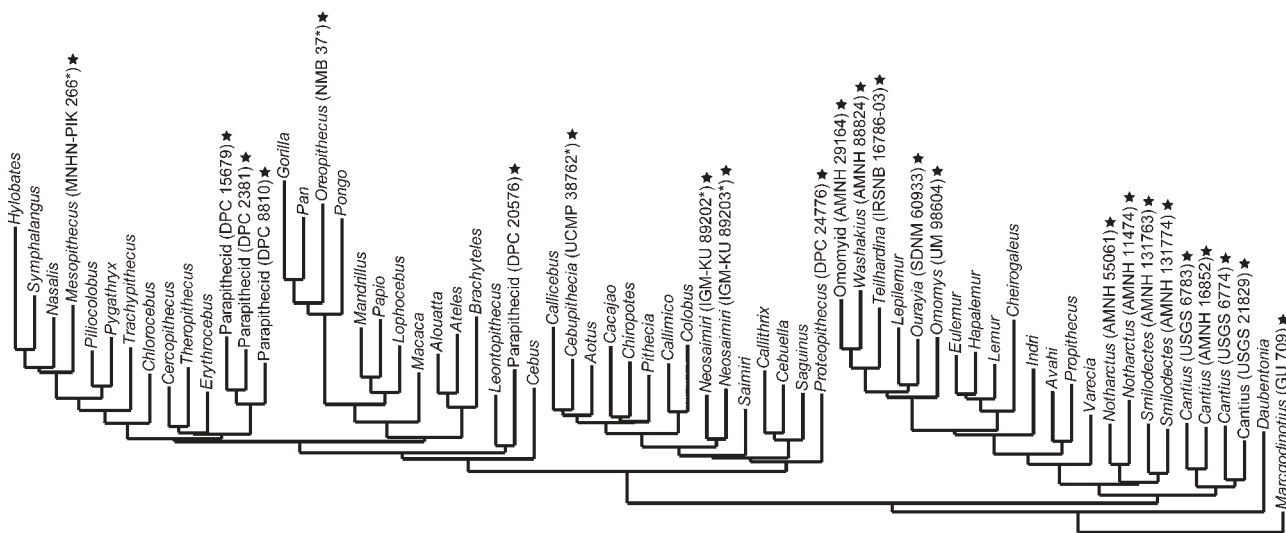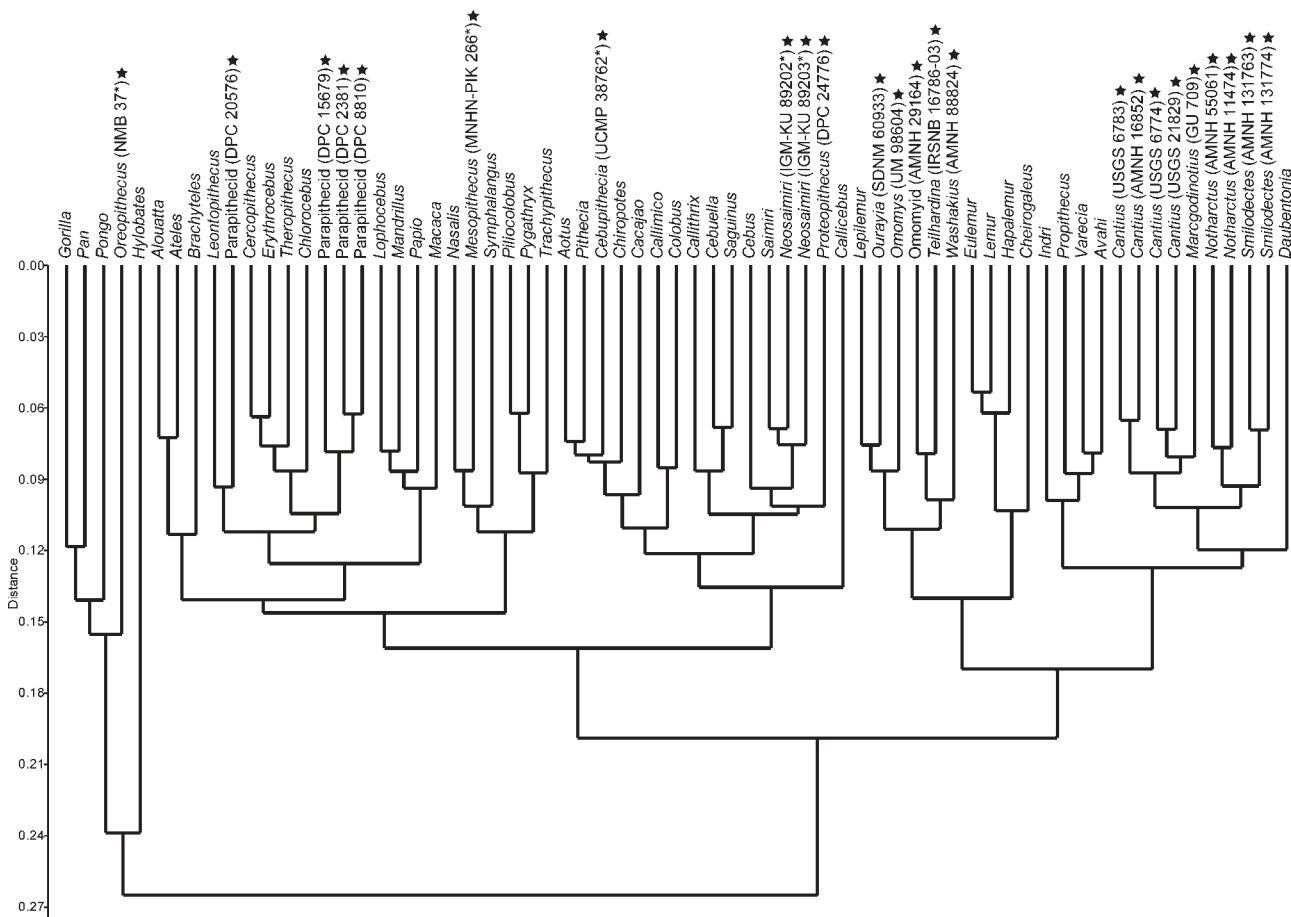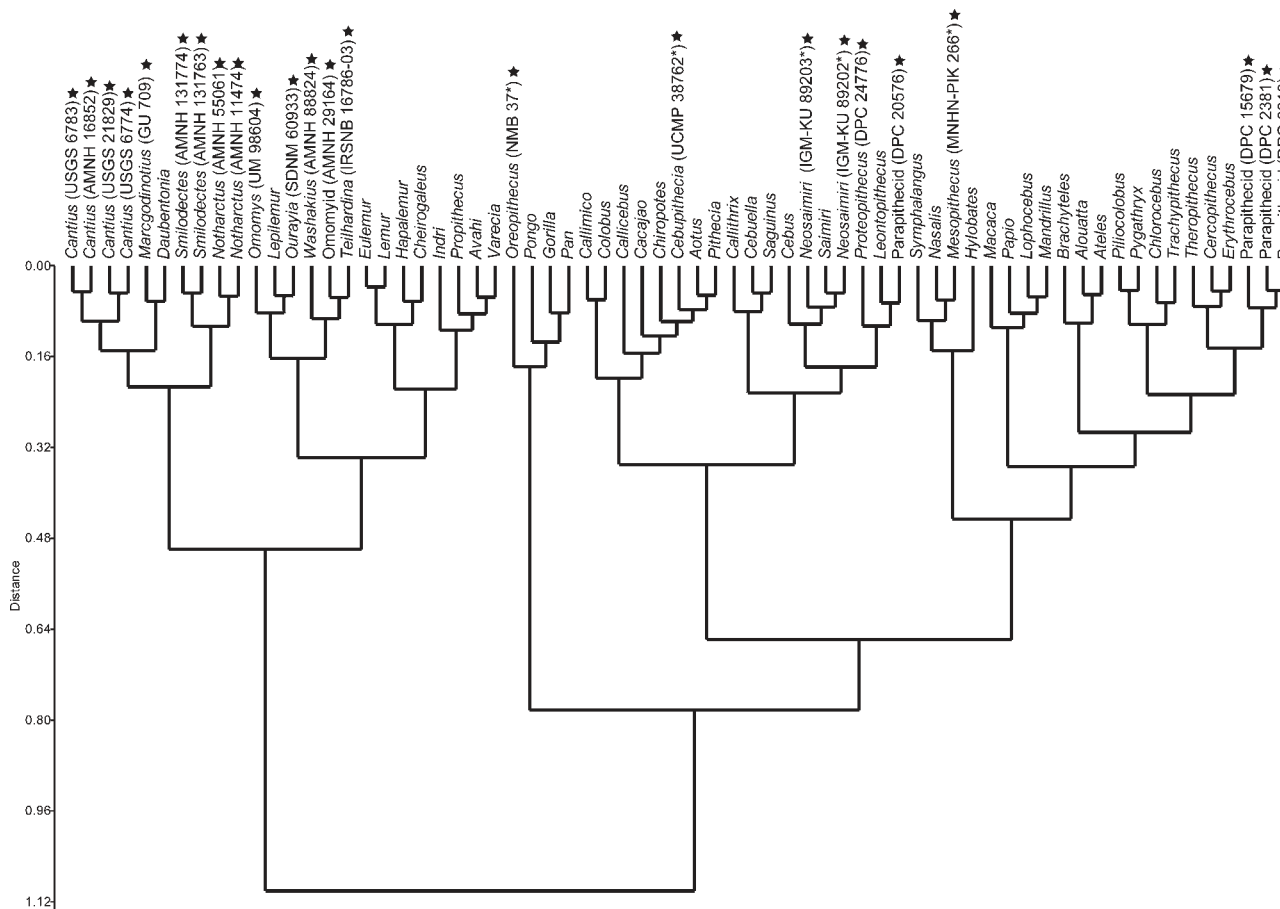
Fig. 9. Neighbor Joining tree. To explore phenetic affinites implied by pseudolandmarks in the calcaneal dataset we averaged coordinate data from individual specimens into species means as described in the text and then performed three types of clustering algorithms, just as was also done by Gladman et al. (2013) for a 27 landmark tradi-tional dataset. The NJ tree requires specification of a root to which nearest neighbors are attached. Fossils were not averaged. Therefore stars and specimen numbers represent individual fossils. These analyses were carried out in PAST (Hammer et al. 2001, 2006).

Fig. 10. UPGMA tree. To explore phenetic affinities implied by pseudolandmarks in the calcaneal data-set we averaged coordinate data from individual specimens into genus means as described in the text and then performed three types of clustering algorithms, just as was also done by Gladman et al. (2013) for a 27 landmark traditional dataset. Fossils were not averaged. Therefore stars and specimen numbers represent individual fossils. These analyses were carried out in PAST (Hammer et al. 2001, 2006).

Fig. 11. Wards tree. To explore phenetic affinities implied by pseudolandmarks in the calcaneal dataset we averaged coordinate data from individual specimens into genus means as described in the text and then performed three types of clustering algorithms, just as was also done by Gladman et al. (2013) for a 27 landmark traditional dataset. Fossils were not averaged. Therefore stars and specimen numbers represent individual fossils. These analyses were carried out in PAST (Hammer et al. 2001, 2006).

Generally speaking, *auto3dgm* based results were less precise when it comes to interpretable clusters of platyrrhines, cercopithecoids, and hominoids compared with the results of Gladman et al. (2013).

Mixed Bone Analysis

Because all bones are first scaled to the same unit centroid size (the square root of the sum of squared distances of all landmarks to the centroid of the object), there is a theoretical maximum distance that can accumulate between any pair of bones, and therefore also among all pairs of bones of a given sample size. Nonetheless, the Procrustes distance for any pair of bones and a sample of any size can also approach zero, meaning that shape diversity can be compared by looking at the mean and variance of distances in the distance matrix.

Interestingly, we found that the mean inter-specimen distance and standard deviation were virtually identical for the calcaneal dataset and astragalus dataset treated separately. On the other hand, the mixed samples (both the full 160 specimen sample, and reduced 80 specimen sample—with 40 of each bone type) showed significantly higher mean distance and distance variance (Table 4). That is, results indicate what might be expected intuitively—that there is greater shape diversity in samples containing two kinds of bones than samples containing one kind of bone. Plotting principal component scores reveals obvious taxonomic and phylogenetic clustering (Fig. 12).

Comparing phylogenetic signal shows consistently (though not significantly) higher estimates of Pagel's lambda in principal component scores of the calcaneus dataset for PCs 1–2 as calculated from both the separate and combined datasets (Table 5). The only exception is that the distance-from-combined-sample-mean dataset ("*mix* MD" in Table 5) for the astragalus had a value of lambda that was higher and more similar to lambda values of the calcaneus datasets. There was extensive correlation among species mean astragalar and calcaneal PC scores if they came from the single, mixed bone PCA, while if separate PCA's were run on the two bones, correlations between astragalar and calcaneal PC scores were less frequently significant (Tables 6 and 7). After plotting the PC scores for all bones of the mixed-bone analysis (Fig. 12A), with PC2 scaled to represent the appropriate total fraction of sample variance
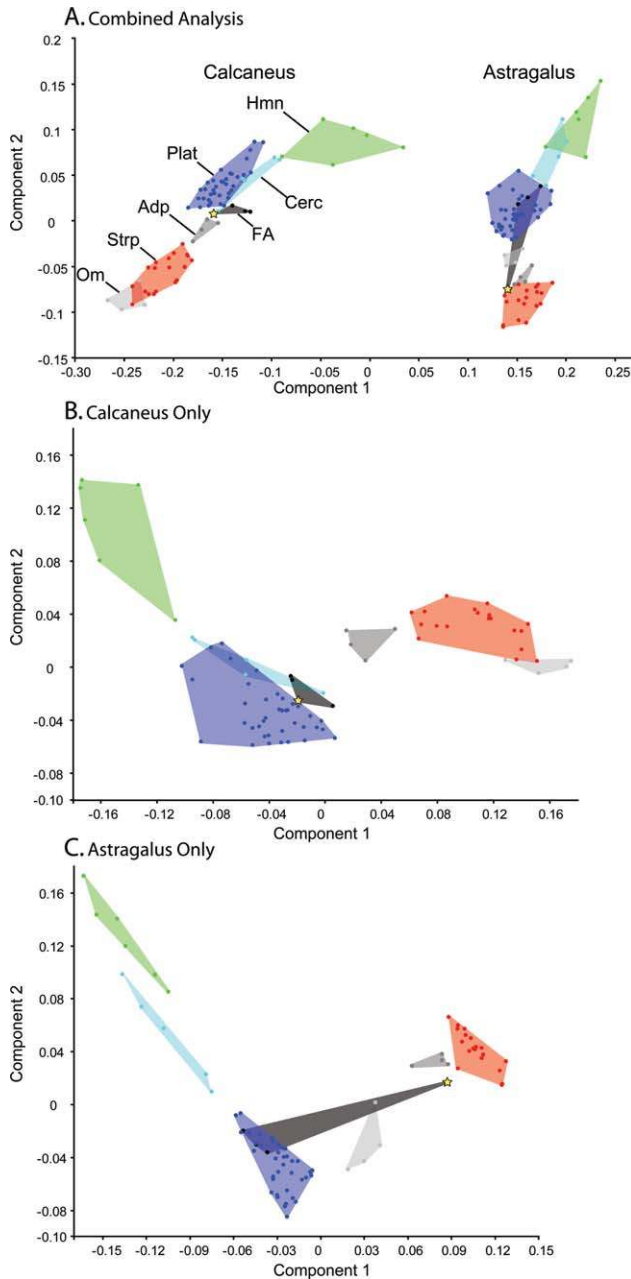
Fig. 12. Mixed bone analyses. (**A**) PCA plot (PC's 1 and 2) of the mixed bone analysis. MST's were established for each bone type independently using our FAA in the way described above with 1,024 pseudolandmark correspondence points for each set. Then we exhaustively computed the minimum Procrustes distance between every pair of astragalus and calcaneus. We used that pair with the smallest distance to connect the calcaneal to the astragalar MST and allow the template to extend between two bones. Then we were able to run GPA and PCA on the mixed bone analysis. (**B**) PCA plot (PC's 1 and 2) for the calcaneus when no astragali are included. **C**, PCA plot (PC's 1 and 2) for the astragalar dataset when no calcanei are included. The star represents the Fayum anthropoid *Proteopithecus*. Note that the there is good phylogenetic correlation with and between bones on the same axes whether the analyses are done on mixed or single bone samples. This is demonstrated quantitatively in Tables 6 and 7.

(PC1 = 61.2% of total variance, PC2 = 7.6%), we computed the plot area of the calcaneus to be 35% larger than that of astragalus. This is somewhat surprising given that results of the distance matrix comparisons (Table 4) could be taken to suggest that they should be equally disparate.

DISCUSSION

Comparisons with *Conventional* 3DGM

We found the degree of similarity between *auto3dgm* based analyses and those performed on the same sample by Gladman et al. (2013) to be surprising. Compared with our analysis using 1,024 automatically determined points, the carefully selected 27 landmarks used by Gladman et al. (2013) showed similar loadings of shape variance on its PC axes, similar variance breakdown on the first several PCs, and even a strong correlation between some of the principal component scores (Table 3). The traditional landmark analysis consolidated slightly more variance in its first 4 PCs, though the differences are more pronounced on PCs 3 and 4. Because there are more PCs for the automated analyses than for the manual one (two orders of magnitude more), it makes sense that the automated method should have a steeper drop-off.

Our automated approach appears more sensitive to errors caused by noise in the surface mesh. This intuitively makes sense and is supported by consideration of some of the clustering "errors" and/or differences between the automated and manual methods. The relatively poor sorting of platyrrhines, hominoids, and cercopithecoids by our automated analysis can be attributed to cases that do not represent mean values, but are the only exemplars of their genus. In particular, the vast majority of catarrhine species in our sample are represented by single specimens, whereas most of our platyrrhines and strepsirrhines are represented by at least two individuals. A single *Colobus* (AMNH 27711) breaks up an otherwise consistent platyrrhine cluster. Though observation of this specimen does not suggest mesh-defects, the bone's lack of any peroneal tubercle projection is anomalous when compared with the prominent peroneal tubercles of all other cercopithecoids in the sample. The lack of a projecting tubercle may give this bone overall length to width proportions that better match the more slender platyrrhines than the more robust cercopithecoids. Perhaps the use of a single point in the 27 landmark analysis to represent the peroneal region reduces the effect of this feature's variance on the pattern of morphological affinities (a feature represented by ~100 points in the automated analysis). Similar problems with other specimens likely indicate that having multiple specimen samples is more important generally with our automated approach.

Aside from anomalous individuals, broken specimens and faulty meshes can be expected to "fool" the analysis. A likely example of this is *Leontopithecus* joining a fossil parapithecid (DPC 20576) among a cluster otherwise represented by cercopithecoids. The fossil is not well preserved in its distal aspect, which likely accentuates the appearance of a strongly sloping lateral border as seen in the callitrichine. It should also be noted however, that Gladman et al. (2013) found that among sampled, extant platyrrhines, *Leontopithecus* has the strongest morphological affinities to cercopithecoids. Both our *auto3dgm*

TABLE 5. Phylogenetic signal in astragalus and calcaneus shape data based on automated analysis of 1,024 pseudolandmarks

Phylogenetic signal

Astragalus				Calcaneus			
Variable	Lambda (CI)	P(0)	P(1)	Variable	Lambda (CI)	P(0)	P(1)
mix PC1	0.884 (0.578, NA)	<0.0001	0.13	*mix* PC1	1.0 (0.924, NA)	<0.0001	1
mix PC2	0.861 (0.623, NA)	<0.0001	0.06	*mix* PC2	1.0 (0.919, NA)	<0.0001	1
mix PC3	0.871 (0.638, NA)	<0.0001	0.06	*mix* PC3	1.0 (0.954, NA)	<0.0001	1
mix MD	1.0 (0.855, NA)	<0.0001	1	*mix* MD	1.0 (0.949, NA)	<0.0001	1
sep PC1	0.862 (0.641, NA)	<0.0001	0.05	*sep* PC1	1.0 (0.945, NA)	<0.0001	1
sep PC2	0.995 (0.856, NA)	<0.0001	0.89	*sep* PC2	1.0 (0.942, NA)	<0.0001	1
sep PC3	0.846 (0.339, 0.985)	0.003	0.01	*sep* PC3	1.0 (0.845, NA)	<0.0001	1
sep MD	0.990 (0.769, NA)	<0.0001	0.91	*sep* MD	1.0 (0.929, NA)	<0.0001	1

"Mix" preceding the variable name indicates that the data were the result of the sequential GPA and PCA on a "mixed" sample of 160 astragali and calcanei. "MD" stands for mean distance and values represent the continuous Procrustes distance of each specimen from the mean shape. $P(0/1)$ stands for the probability of lambda being zero or one.

TABLE 6. Correlations between PC scores of astragalus and calcaneus, and correlations between PC scores of mixed and separate bone analyses

Between bone correlations (comparisons within separate and mixed analyses)

Sep.	Ast.				Sep.	Ast.			
Calc.	1	2	3	MD	Calc.	1	2	3	MD
1	0.86	−0.17	−0.13	–	1	<0.0001	ns	ns	–
2	−0.08	0.86	0.05	–	2	ns	<0.0001	ns	–
3	−0.16	−0.02	0.02	–	3	ns	ns	ns	–
MD	–	–	–	0.57	MD	–	–	–	<0.0001
Mix.	Ast.				Mix.	Ast.			
Calc.	1	2	3	MD	Calc.	1	2	3	MD
1	0.68	0.86	0.57	–	1	<0.0001	<0.0001	<0.0001	–
2	0.40	0.84	0.76	–	2	0.007	<0.0001	<0.0001	–
3	−0.25	−0.76	−0.80	–	3	ns	<0.0001	<0.0001	–
MD	–	–	–	−0.25	MD	–	–	–	ns

Within bone correlations (comparisons between separate and mixed analyses)

Calc.	Mix.				Calc.	Mix.			
Sep.	1	2	3	MD	Sep.	1	2	3	MD
1	−0.93	−0.98	0.93	–	1	<0.0001	<0.0001	<0.0001	–
2	0.43	−0.01	0.23	–	2	0.004	ns	ns	–
3	−0.08	−0.01	−0.05	–	3	ns	ns	ns	–
MD	–	–	–	0.45	MD	–	–	–	0.003
Ast.	Mix.				Ast.	Mix.			
Sep.	1	2	3	MD	Sep.	1	2	3	MD
1	−0.57	−0.98	−0.90	–	1	<0.0001	<0.0001	<0.0001	–
2	0.80	0.26	−0.29	–	2	<0.0001	ns	ns	–
3	−0.10	0.07	−0.11	–	3	ns	ns	ns	–
MD	–	–	–	0.95	MD	–	–	–	<0.0001

Linear correlation (r) values in boxes on the left, (P) values in boxes on the right. These analyses are run on species mean values and N = 43 in all cases.

analyses and those of Gladman et al. (2013) suggest morphological affinities uniting Fayum fossil parapithecids with cercopithecoids.

Comparisons of Morphological Diversity Among Parts (Mixed Bone Analysis)

Our analyses revealed that the astragalus and calcaneus reflect almost identical amounts of shape variation (similar "disparity" as measured with 1,024 evenly distributed points and using the raw distance matrix. See Table 4). This appears to be a meaningful result since the mixed bone samples (which we believe should express greater shape variation) do, indeed, exhibit significantly greater average distances between shapes. However, it is in some contrast to the results that show the calcaneus to take up 35% more plot area than the astragalus. While we acknowledge that this preliminary approach is somewhat crude since it does not account for the variance

TABLE 7. Phylogenetically informed correlations between astragalus and calcaneus variables that resulted from sequential GPA followed by PCA on 1,024 pseudolandmarks per bone

	PGLS correlations						
Test	Lambda (CI)	$P(0)$	$P(1)$	Slope	r Square	P
sep PC1 (ast. vs. calc.)	1.0 (0.946, NA)	<0.0001	1	0.28	0.073	0.05
mix PC1 (ast. vs. calc.)	1.0 (0.924, NA)	<0.0001	1	0.84	0.204	0.0002
sep MD (ast. vs. calc.)	1.0 (0.925, NA)	<0.0001	1	0.1	0.057	0.79
mix MD (ast. vs. calc.)	1.0 (0.952, NA)	<0.0001	1	−0.36	0.074	0.05

"Mix" preceding the variable name indicates that the data were the result of the sequential GPA and PCA on a "mixed" sample of 160 astragali and calcanei. "MD" stands for mean distance and values represent the continuous Procrustes distance of each specimen from the mean shape. $P(0/1)$ stands for the probability of lambda being zero or one.

represented in the remaining variables (a MDS approach could be better since we could probably force more of the variance onto two axes), these remaining PC's explain an increasingly miniscule amount of variation.

Interestingly, the phylogenetic signal for a given bone-type was minimally affected (if at all) by running GPA and PCA on a mixed bone sample versus a single bone-type sample (Table 5). The calcaneus, in obtaining possibly greater morphological disparity than the astragalus, seems to have developed a stronger phylogenetic signal as well (Table 5). This suggests that change in calcaneus has approximated a Brownian motion model along the branches of the primate phylogenetic tree more so than the astragalus. This difference in mode may be explained functionally by noting that the calcaneus comes into (almost) direct contact with the environment (through the skin, etc.) as the heel, and helps comprise a load arm/lever arm pair that experiences functional demands for leaping and other forms of locomotion (Boyer et al., 2013). In contrast, the astragalus is almost completely isolated with no part that touches the ground, and no attaching muscles. Therefore, the astragalus may often be insulated from subtle changes in functional demands and be more likely to experience periods of stasis, whereas the calcaneus probably responds more faithfully to small changes in mechanical environment.

The astragalus has long been noted for its high valence in reflecting systematic relationships, while the calcaneus has appeared less useful. At first pass, this observation seems contradicted by our results. However, if the astragalus has experienced stasis more generally than the calcaneus and developed its comparable morphological variance through more punctuated changes, then the resulting variance may be more clearly associated with more inclusive taxonomic groups (like strepsirrhines, tarsiers, platyrrhines, cercopithecoids, and hominoids) than with species-level differences.

Biological Significance of Automated Pseudolandmarks

The most obvious difference between pseudolandmarks of our method and traditional landmarks is that points associated with a particular feature (e.g., peroneal tubercle), or an articular surface on one bone, may not be located on those features in another bone. This may rub some morphologists the wrong way if they feel that they know that the peroneal tubercle is homologous between two taxa, but the algorithm does not bear this out.

There are several points to be made here. First, as reviewed by MacLeod (2001), Owen's (1846) original definition considered homology as pertaining to "organs" (or we could say "whole bones" here) but did not define mappings of sub-regions therein. In a strict sense, the concept of homology does not apply to features of organs.

Second, the essence of Darwinian homology is that features in different taxa are biologically equivalent if they can be traced to the same feature in a common ancestor through the process of "descent with modification." This is reflected in a more recent definition stating that homology is a "continuity of information" (Van Valen, 1982). This view dictates that the ultimate arbiter of contrasting homology hypotheses is the pattern of transformations that occurred in evolution, but it is rare to obtain data allowing such an empirical test, meaning that researchers should remain open to multiple alternatives.

Third, the critics of the adaptationist program (Gould and Lewontin, 1979) warn us to beware of "spandrels." One can ask whether the feature of interest exists by genetic design or by developmental context. If the peroneal tubercle "exists" as a genetically specified bump on the side of the calcaneus (in the sense that there are gene products that cause the formation of this bump, and variation in the position or size of the tubercle can be explained by these gene products being expressed at different positions, at different concentrations, and/or for different durations along the shaft of the calcaneus), then it follows that this "bump" should be marked with a landmark of the same identity on any bone regardless of where topologically it occurs. However, it seems equally likely that the form of the bony peroneal tubercle is a mechanical and re-modeling consequence of the paths of the peroneal tendons and attachment positions of the retinacular ligaments. In this alternative scenario, representing the position of this bump by the same point regardless of its position on the calcaneus seems misrepresentative. The truth is that the genetic influences and developmental homologies for most features are not known. An informative test of these alternatives (although cruel) would be to remove the tendons at an early stage of development and observe whether and where a peroneal tubercle developed. Even if it were to become known that peroneal tubercle development occurred independent of attaching ligaments and tendons, and the forces they exert, this would only imply evolutionary homology if we assume parsimony in evolution (or Hennig's auxiliary principle) which some researchers are willing to do, but others are not. This also comes down to whether type I or type II landmarks are preferred when the respective criteria suggest different correspondence patterns for a given anatomical region. Finally, in this particular example, there is no widespread

agreement on the evolutionary homology of the peroneal tubercle among primates (Decker and Szalay, 1974).

Variation in features that are plastic and can be modified during life (such as ligament attachment points and articular surface areas and boundary shapes) may be explained by ontogenetic causes. For instance, variation in the development of certain astragalar facets in humans has been explained by different postural tendencies among populations (Barnett, 1954). If we use the distal boundary of the tibial facet as a landmark, this feature point may extend all the way down the astragalar neck in some people, or not approach it at all in others. This would be useful for quantifying variation due to postural differences among humans, but probably not for distinguishing the shape of a human astragalus from a chimpanzee astragalus.

Another argument for adding the use of pseudolandmarks to the morphologist's toolkit is the fact that the research community already accepts similar approaches to shape comparison including Fourier analysis (Rohlf and Archie, 1984), eigenshape analysis (MacLeod, 1999), and eigensurface analysis (Polly and MacLeod, 2008). These methods retain no fidelity to specific landmark-like features. The most significant conceptual difference between our approach and eigensurface analysis is that the anatomical axes must be manually set in the latter. A more practical difference is that eigensurface is restricted to "relief-type" or "disc-type" surfaces, whereas *auto3dgm* can be applied to either disc-type or fully 3D surfaces.

The question of whether points or regions on different instances of the same bone are "equivalent" is ultimately a question about transformational homology. Our method provides an "operational homology" (=topological correspondence). The MST used to link forms can be taken as a hypothesis of transformational homology to be tested. The best answer to whether certain "point features" are equivalent must be answered by assessing whether treating them as such results in phenetic patterns that correlate with independent datasets on phylogenetic relationships or functional capacity. This means that if the utility of automated methods is going to increase, then automated correspondence determinations that are more sensitive to feature points (type II landmarks) must also be developed. This requires algorithms based on "non-area preserving maps." The original work of Boyer et al. (2011) presents such a method but lacks applicability to "full 3D" shapes and does not provide a means for inducing transitivity of comparisons. Different patterns of transformational homology will be implied by different phylogenetic hypotheses, which could be evaluated according to different optimization criteria.

Too Many Variables, Not Enough Specimens?

A major challenge in statistical modeling as applied to molecular biology (Golub et al., 1999), genetics (Patterson et al., 2006), image analysis (Roweis and Saul, 2000), and text analysis (Blei et al., 2003) has been the *large P, small N* setting (Poggio and Smale, 2003; West, 2003) where the number of variables is typically much larger than the number of samples. In statistics, the difficulty of modeling data as the number of variables increases and exceeds the number of observations is often called "the curse of dimensionality," a phrase coined by Bellman with respect to optimization problems (Bellman, 1984). However, many of the great advances in the last 10 years in statistics, machine learning, and applied mathematics are related to the observation that the relevant dimension of the data is not the number of variables, but the number of independent variables (the intrinsic dimension) (Donoho, 2000). For 1,024 landmarks spread on a sample of 80–160 objects, the intrinsic dimensionality will be much lower than the number of landmarks. If the perspective promoted by statisticians dealing with *large P, small N* problems is correct, then the problem of over-determination can be avoided by limiting the number of independent variables generated by data reduction techniques from a landmark dataset with hundreds or thousands of points. The idea that seemingly high-dimensional data have few degrees of freedom, or low intrinsic dimensionality, is central to the methodologies developed in this article.

As a matter of precedent, this philosophy is implicitly acknowledged in articles that use large numbers of evenly (or "optimally") spread semi-landmarks as well as in eigenshape analysis (Polly, 2008; Polly and MacLeod, 2008; Sievwright and MacLeod, 2012). Harcourt-Smith et al. (2008) provides a pertinent example, in which a total of nine user-defined landmarks were used to generate 361 semilandmark points on the talo-tibial facets of a sample with 54 specimens representing three species. Another example is Sievright and MacLeod (2012). These authors used 62 points to represent the dorsal surface of the proximal humerus in a sample of 50 falconiform specimens. They projected their coordinates into tangent space and used principal component analysis to generate projection scores. These mutually orthogonal (independent) projection scores were then used to run a Canonical Variates Analysis (=DFA). They limited the number of principal components used in their analysis to 21 (because this number represented 95% of the total variation in the dataset and was much less than their N = 50). These authors recognize the importance of the number of independent variables, but do not discuss the statistical ramifications of the number of original, yet correlated, variables.

SUMMARY AND CONCLUSION

Greater automation and standardization for morphological studies are needed if morphology is to survive as a branch of phenomics with relevance comparable to genomics. The most important level at which such automation must occur is in determining biological/geometric correspondence between shapes. Past attempts to automate such determinations have suffered from the prospect that computations involved were too time intensive (as well as philosophical arguments against the idea of such an approach). Dimension reduction techniques such as working from photographs and outlines have been applied to circumvent this issue, but an observer is needed to orientate objects before such application, slightly defeating the purpose of automation. Greater computing power and techniques for simplifying the search for alignment and correspondence mapping between 3D digital models are applied here and an R package for implementing this method has been created.

Our analyses show a surprising and reassuring degree of similarity between quantifications based on user-defined landmarks and our *auto3dgm* approach. Although human interaction must occur at several

stages of the analyses to verify that erroneous alignments have not been generated, this approach still represents a step beyond any automation procedures yet applied, because (1) no qualitative decisions about the geometric equivalence of point features are required and (2) protocols for generating alignments and pseudolandmark datasets lack observer error, since the final procedure for the result of the algorithm can be described via the numerical parameter input to the model. At this juncture, it is important to re-iterate that the method of *auto3dgm* was not completely deterministic in the test samples we used even at 1000 points, though the correlations between multiple runs were very high. This means that researchers should, as always, be cautious with the method when it comes to interpreting phenetic affinities of individual specimens.

Very little familiarity with anatomical terminology or features is required. Only a basic ability to visually compare shapes is necessary in *auto3dgm* in order to verify the absence of misalignments. This method has the potential for adoption by geneticists, molecular biologists, and biomedical engineers who may feel uncomfortable about their ability to take measurements with repeated accuracy or with biological significance to their questions of interest.

One of the most exciting capabilities provided by this algorithm is the ability to compare variance magnitude and patterns for different skeletal elements. Our initial experiments with this approach show that two articulating bones of the skeleton have similar levels of morphological diversity with strong covariance, which makes sense developmentally, but the calcaneus has a consistently stronger phylogenetic signal in its variance patterns than the astragalus, which is potentially related to more direct functional demands.

Future work will explore different types of correspondence algorithms with an emphasis on constructing algorithms that can efficiently determine non-area preserving maps (those that mimic user-defined type II landmarks of 3DGM more closely). Of course, developing methods that yield more deterministic results is critical as well. Furthermore, we intend to compare variance levels among different regions of the skeleton with the expectation that patterns of covariance and variance magnitudes will differ more between bones that are far apart from each other on the skeleton and are more likely to have different developmental and historical natural–selective contexts. We recognize that these quantities are still dependent on the sample composition, the parameters of any particular run of *auto3dgm*, any ordination methods that are used, and random differences based on initial pseudolandmark spreading by *auto3dgm*. Nonetheless, we feel that the patterns will be informative for evolutionary questions including those dealing with disparity because the quantification of inter-bone shape distance is objective and more comprehensive using *auto3dgm*, and we have articulated a rational geometric basis for comparing variance between groups of non-homologous elements. A related challenge for large scale automation of morphological analysis is provision of comprehensive samples of digitized bones. Towards that end a new "genbank-like" online archive for 3D bone digitizations called MorphoSource (www. morphosource.org) has been created at Duke University (Boyer et al. 2014; Kaufman et al. 2014). At the time of writing over 1,000 bones have been shared on this site. We encourage readers to investigate and contribute to this resource as its development is as important as methods development for moving the field forward.

ACKNOWLEDGEMENTS

LITERATURE CITED

3D Systems Inc. 2013. Geomagic studio. Rock Hill: 3D Systems Inc.

Abzhanov A, Kuo WP, Hartmann C, Grant BR, Grant PR, Tabin CJ. 2006. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. Nature 442:563–567.

Arnold C, Matthews LJ, Nunn CL. 2010. The 10kTrees website: a new online resource for primate phylogeny. Evol Anthropol 19:114–118.

Barnett CH. 1954. Squatting facets on the European talus. J Anat (London) 88:509–513.

Bellman RE. 1984. Eye of the Hurricane: an autobiography. Singapore: World Scientific Publishing Company.

Berge C, Jouffroy FK. 1986. Morpho-functional study of *Tarsius's* foot as compared to the galagines': what does an "elongate calcaneuys" mean? In: Taub D, King F, editors. Current perspectives in primate biology. New York: Van Nostrand Reinhold. p 146–158.

Besl PJ, McKay ND. 1992. A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 14:239–256.

Blei DM, Ng AY, Jordan MI. 2003. Latent dirichlet allocation. J Mach Learn Res 3:993–1022.

Bookstein FL. 1980. When one form is between two others: an application of biorthogonal analysis. Am Zool 20:627–642.

Bookstein FL. 1991. Morphometric tools for landmark data. Geometry and biology. Cambridge: Cambridge University Press.

Bookstein FL. 1994. Can biometrical shape be a homologous character? In: Hall BK, editor. Homology: the hierarchical basis of comparative biology. San Diego: Academic Press. p 197–227.

Bookstein FL. 1997. Landmark methods for forms without landmarks: localizing group differences in outline shape. Med Image Anal 1:225–243.

Bookstein FL, Schafer K, Prossinger H, Seidler H, Fieder M, Stringer C, Weber GW, Arsuaga J-L, Slice DE, Rohlf FJ, Recheis

W, Mariam AJ, Marcus LF. 1999. Comparing frontal cranial profiles in archaic and modern Homo by morphometric analysis. Anat Rec (New Anat) 257:217–224.

Bookstein FL, Streissguth AP, Sampson PD, Connor PD, Barr HM. 2002. Corpus callosum shape and neuropsychological deficits in adult males with heavy fetal alcohol exposure. Neuroimage 15:233–251.

Boyer DM, Lipman Y, St. Clair E, Puente J, Patel BA, Funkhouser TA, Jernvall J, Daubechies I. 2011. Algorithms to automatically quantify the geometric similarity of anatomical surfaces. Proc Natl Acad Sci 108:18221–18226.

Boyer DM, Seiffert ER. 2013. Patterns of astragalar fibular facet orientation in extant and fossil primates and their evolutionary implications. Am J Phys Anthropol 151:420–447.

Boyer DM, Seiffert ER, Gladman JT, Bloch JI. 2013. Evolution and allometry of calcaneal elongation in living and extinct primates. PLoS One 8(7):e67792.

Boyer DM, Kaufman S, Gunnell GF, Gomes E, Thostenson J. 2014a. MorphoSource: A currently active project-based 3D digital web-accessible data archive for museums and individuals at Duke University USA. In: Mallison H, Vogel J, and Belvedere M, editors. Digital Specimen 2014 - Abstracts of Presentations. p http://www.naturkundemuseum-berlin.de/forschung/tagungen/digitalspecimen-berlin-2014/home/abstracts-of-presentations/.

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.

Cignoni P, Ranzuglia G, Callieri M, Corsini M, Ganovelli F, Pietroni N, Tarini M. 2012. MeshLab Visual Computing Lab - ISTI- CNR. http://meshlab.sourceforge.net/.

Decker RL, Szalay FS. 1974. Origins and function of the pes in the Eocene Adapidae (Lemuriformes, Primates). In: Jenkins Jr. FA, editor. Primate locomotion. New York: Academic Press. p 261–291.

Donoho DL. 2000. High-dimensional data analysis: the curses and blessings of dimensionality. In: American Mathematical Society Conference: Math Challenges of the 21st Century. p 1–32.

Eldar Y, Lindenbaum M, Porat M, Zeevi YY. 1997. The farthest point strategy for progressive image sampling. IEEE Trans Image Process 6:1315.

Frost SR, Marcus LF, Bookstein F, Reddy DP, Delson E. 2003. Cranial allometry, phylogeography and systematics of large bodied papionins (Primates: Cercopithecinae) inferred from geometric morphometric analysis of landmark data. Anat Rec 275A:1048–1072.

Gingerich PD, ul Haq M, Zalmout IS, Khan IH, Malkani MS. 2001. Origin of whales from early artiodactyls: hands and feet of Eocene Protocetidae from Pakistan. Science 293:2239–2242.

Gladman JT, Boyer DM, Simons EL, Seiffert ER. 2013. A calcaneus attributable to the primitive late Eocene anthropoid *Proteopithecus sylviae*: phenetic affinities and phylogenetic implications. Am J Phys Anthropol 151:372–397.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286:531–537.

Gould SJ, Lewontin RC. 1979. The spandrels of San Marco and the Panglossian pradigm: a critique of the adaptationist programme. Proc R Soc London B 205:581–598.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, Hansen NF, Durand EY, Malaspinas A-S, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Ž, Gušic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PLF, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S. 2010. A draft sequence of the neandertal genome. Science 328:710–722.

Gunz P, Mitteroecker P, Bookstein FL. 2005. Semilandmarks in three dimensions. In: Slice DE, editor. Modern morphometrics in

physical anthropology. New York: Kluwer Academic/Plenum Publishers. p 73–98.

Hammer Ø, Harper DAT, Ryan PD. 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontol Electron 4:9. http://palaeo-electronica.org/2001_2001/past/issue2001_2001.htm.

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

Harcourt-Smith WEH, Tallman M, Frost SR, Wiley DF, Rohlf JF, Delson E. 2008. Analysis of selected hominoid joint surfacesusing laser scanning and geometricmorphometrics: a preliminary report. In: Dagosto M, Sargis EJ, editors. Mammalian evolutionary morphology: a tribute to Frederick S. Szalay. New York: Springer Science. p 373–383.

Harjunmaa E, Kallonen A, Voutilainen M, Hämäläinen K, Mikkola ML, Jernvall J. 2012. On the difficulty of increasing dental complexity. Nature 483:324–327.

Houle D, Govindaraju DR, Omholt S. 2010. Phenomics: the next challenge. Nat Rev Genet 11:855–856.

Jardine N. 1969. The observational and theoretical components of homology: a study based on the morphology of the dermal skull-roofs of rhipidistean fishes. Biol J Linnean Soc 1:327–361.

Kaufman S, Boyer DM. 2014. Developing Platforms for Management and Distribution of Digital Specimen Data. In: Mallison H, Vogel J, and Belvedere M, editors. Digital Specimen 2014 - Abstracts of Presentations. http://www.naturkundemuseum-berlin.de/forschung/tagungen/digitalspecimen-berlin-2014/home/abstracts-of-presentations/.

Leakey LSB, Tobias PV, Napier JR. 1964. A new species of the genus *Homo* from Olduvai Gorge. Nature 202:7–9.

Lipman Y, Daubechies I. 2010. The continuous procrustes distance. Technical report. Princeton: Princeton University.

Lohmann GP. 1983. Eigenshape analysis of microfossils: a general morphometric method for describing changes in shape. Math Geol 15:659–672.

Losos JB. 1990. The evolution of form and function: morphology and locomotor performance in west indian *Anolis* lizards. Evolution 44(5):1189–1203.

MacLeod N. 1999. Generalizing and extending the eigenshape method of shape space visualization and analysis. Paleobiology 25:107–138.

MacLeod N. 2001. Landmarks, localization and the use of morphometrics in phylogenetic analysis. In: Adrain JM, Edgecombe GD, Lieberman BS, editors. Fossils, phylogeny, and form. New York: Springer. p 197–233.

MacLeod N. 2007. Automated taxon identication in systematics: theory, approaches, and applications. London: CRC Press, Taylor & Francis Group.

MacLeod N, Benfield M, Culverhouse P. 2010. Time to automate identification. Nature 467:154–155.

McVean G, Spencer CCA, Chaix R. 2005. Perspectives on human genetic variation from the hapmap project. PLoS Genet 1:e54.

Mitteroecker P, Gunz P. 2009. Advances in geometric morphometrics. Evol Biol 36:235–247.

Morgan TH. 1911. The origin of five mutations in eye color in drosophila and their modes of inheritance. Science 33:534–537.

Nunn CL. 2011. The comparative method in evolutionary anthropology and biology. Chicago: University of Chicago Press.

O'Higgins P, Jones N. 2006. Tools for statistical shape analysis. http://sites.google.com/site/hymsfme/resources. UK: Hull York Medical School.

O'Leary MA, Bloch JI, Flynn JJ, Gaudin TJ, Giallombardo A, Giannini NP, Goldberg SL, Kraatz BP, Luo Z-X, Meng J, Ni X, Novacek MJ, Perini FA, Randall ZS, Rougier GS, Sargis EJ, Silcox MT, Simmons NB, Spaulding M, Velazco PM, Weksler M, Wible JR, Cirranello AR. 2013a. The Placental Mammal Ancestor and the Post–K-Pg Radiation of Placentals. Science 339:662–667.

O'Leary MA, Bloch JI, Flynn JJ, Gaudin TJ, Giallombardo A, Giannini NP, Goldberg SL, Kraatz BP, Luo Z-X, Meng J, Ni X, Novacek MJ, Perini FA, Randall ZS, Rougier GS, Sargis EJ, Silcox MT, Simmons NB, Spaulding M, Velazco PM, Weksler M, Wible JR, Cirranello AR. 2013b. Response to comment on "the placental mammal ancestor and the post–k-pg radiation of placentals". Science 341:613.

Orlando L, Calvignac S, Schnebelen C, Douady CJ, Godfrey LR, Hänni C. 2008. DNA from extinct giant lemurs links archaeole-murids to extant indriids. BMC Evol Biol 8:121.

Orme CDL, Freckleton RP, Thomas G, Petzoldt T, Fritz SA, Isaac N, Pearse W. 2011. CAPER: comparative analysis of phylogenetics and evolution in R. R package version 3.1–104.

Ostrom JH. 1975. On the origin of *Archaeopteryx* and the ancestry of birds. Colloq Int CNRS 218:521–532.

Owen R. 1846. Report on the archetype and homologies of the verte-brate skeleton. Report of the British Association for the Advance-ment of Science (Southampton meeting). p 169–340.

Paniagua B, Bompard L, Cates J, Whitaker R, Datar M, Vachet C, Styner M. 2012. Combined SPHARM-PDM and entropy-based particle systems shape analysis framework. In: SPIE Medical Imaging: International Society for Optics and Photonics. p 83170L–83177L.

Patterson C. 1982. Morphological characters and homology. In: Joy-sey KA, Friday AE, editors. Systematics Association Special Vol-ume No. 21, Problems of phylogenetic reconstruction. London and New York: Academic Press. p 21–74.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS Genet 2:e190.

Perez SI, Bernal V, Gonzalez PN. 2006. Differences between sliding semilandmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. J Anat 208:769–784.

Poggio T, Smale S. 2003. The mathematics of learning: dealing with data. Notices Am Math Soc 50:537–544.

Polly PD. 2008. Adaptive zones and the pinniped ankle: a 3D quan-titative analysis of carnivoran tarsal evolution. In: Sargis EJ, Dagosto M, editors. Mammalian evolutionary morphology: a trib-ute to Frederick S. Szalay. Dordrecht, The Netherlands: Springer. p 165–194.

Polly PD, MacLeod N. 2008. Locomotion in fossil Carnivora: an application of eigensurface analysis for morphometric comparison of 3D surfaces. Palaeontol Electron 11:13.

Puente J. 2013. Distances and algorithms to compare sets of shapes for automated biological morphometrics. In: Program in Applied and Computational Mathematics. unpublished: Princeton. p 72.

Puente J, Daubechies I. (in prep). A generalization of procrustes anaylsis.

R Coding Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Rohlf FJ. 2002. Geometric morphometrics and phylogeny. Syst Assoc Spl Vol 64:175–193.

Rohlf FJ, Archie JW. 1984. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera, Culicidae). Syst Zool 3:302–317.

Roth VL. 1993. On three dimensional morphometrics, and on the identification of landmark points. In: Marcus LF, Bello E, Garcia-Valdecasas A, editors. Contributions to morphometrics. Madrid: Museo Nacional de Ciencias Naturales CSIC. p 41–61.

Roweis S, Saul L. 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323–2326.

Shubin NH. 1994. History, ontogeny, and evolution of the archetype. In: Grande L, Riepper O, editors. Homology, the hierarchical basis of comparative biology. New York: The Academic Press, Inc. p 249–271.

Sievwright H, MacLeod N. 2012. Eigensurface analysis, ecol-ogy, and modelling of morphological adaptation in the falco-niform humerus (Falconiformes: Aves). Zool J Linn Soc 165: 390–419.

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

Simpson GG. 1961. Principles of animal taxonomy. New York: Columbia University Press.

Smith GR. 1990. Homology in morphometrics and phylogenetics. In: Rohlf FJ, Bookstein FL, editors. Proceedings of the Michigan Mor-phometrics Workshop. Michigan: The University of Michigan Museum of Zoology. p 325–338.

Sneath PHA, Sokal FJ. 1973. Principles of numerical taxonomy, 2nd ed. San Francisco: Freeman.

Sokal RR. 1966. Numerical taxonomy. Sci Am 215:106–116.

Springer MS, Meredith RW, Teeling EC, Murphy WJ. 2013. Techni-cal comment on "the placental mammal ancestor and the post–k-pg radiation of placentals". Science 341:613.

Styner M, Oguz I, Xu S, Brechbuler C, Pantazis D, Gerig G. 2006. Framework for the statistical shape analysis of brain structures using SPHARM-PDM. Insight J 1071:242–250.

Thompson DA. 1942. On growth and form. Cambridge: University Press. 1116 p.

Van Valen LM. 1982. Homology and causes. J Morphol 173:305–312.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Francesco VD, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji R-R, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang ZY, Wang A, Wang X, Wang J, Wei M-H, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu SC, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers Y-H, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigŏ R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang Y-H, Coyne M, Dahlke C, Mays AD, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. 2001. The sequence of the human genome. Sci-ence 291:1304–1351.

Visualization Sciences Group. 2009. AVIZO, Version 6.0. Burlington: Mercury Computer Systems.

Weeks PJD, O'Neill MA, Gaston KJ, Gauld ID. 1999. Species-identi-cation of wasps using principal component associative memories. Image Vis Comput 17:861–866.

West M. 2003. Bayesian Factor Regression Models in the "Large p, Small n" Paradigm. In: Bernardo JM, Bayarri MJ, Berger OJ, Dawid AP, Heckerman D, Smith AFM, West M, editors. Bayesian statistics. Oxford: Oxford University Press. p 733–742.

Wiley DF, Amenta N, Alcantara DA, Ghosh D, Kil YJ, Delson E, Harcourt-Smith W, Rohlf FJ, St. John K, Hamann B. 2005. Evolu-tionary morphing. Proceed IEEE Visual 2005:8.

Zelditch ML, Swiderski DL, Sheets DH, Fink WL. 2004. Geometric morphometrics for biologists. San Diego: Elsevier Academic Press.