SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Biological anthropology

» Databases

» Bone

A collection of non-human primate computed tomography scans housed in MorphoSource, a repository for 3D data

Lynn E. Copes[1,†], Lynn M. Lucas[1], James O. Thostenson[2], Hopi E. Hoekstra[3] & Doug M. Boyer[4]

A dataset of high-resolution microCT scans of primate skulls (crania and mandibles) and certain postcranial elements was collected to address questions about primate skull morphology. The sample consists of 489 scans taken from 431 specimens, representing 59 species of most Primate families. These data have transformative reuse potential as such datasets are necessary for conducting high power research into primate evolution, but require significant time and funding to collect. Similar datasets were previously only available to select research groups across the world. The physical specimens are vouchered at Harvard's Museum of Comparative Zoology. The data collection took place at the Center for Nanoscale Systems at Harvard. The dataset is archived on MorphoSource.org. Though this is the largest high fidelity comparative dataset yet available, its provisioning on a web archive that allows unlimited researcher contributions promises a future with vastly increased digital collections available at researchers' finger tips.

Design Type(s)	organism part comparison design • species comparison design • observation design
Measurement Type(s)	morphometry
Technology Type(s)	computed tomography scanner
Factor Type(s)	animal body part • life cycle stage
Sample Characteristic(s)	skull • cranium • mandible • femur • humerus • Saimiri oerstedii • Cebus capucinus • Ateles geoffroyi • Trachypithecus cristata • Macaca fascicularis • Euoticus elegantulus • Galago alleni • Cercocebus albigena • Gorilla gorilla gorilla • Pan troglodytes troglodytes • Saguinus • …

[1]Institute of Human Origins and School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, USA. [2]Department of Electrical Engineering, Duke University, Durham, North Carolina, USA. [3]Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts, USA. [4]Department of Evolutionary Anthropology, Duke University, Durham, North Carolina, USA. †Present address: Department of Medical Sciences, Frank H. Netter MD School of Medicine, Quinnipiac University, Hamden, Connecticut, USA. Correspondence and requests for materials should be addressed to L.E.C. (email: lynn.copes@quinnipiac.edu, about conditions of collection and format of the dataset or any technical issues), D.M.B. (email: dmb65@duke.edu, about MorphoSource) or H.E.H. (email: hoekstra@oeb.harvard.edu, about specimens or about use of data for purposes other than that allowed by currently assigned copyright licenses).

Background & Summary

Digital data in comparative morphology

High fidelity, microCT and surface scan renderings of osteological materials and wet specimens have become essential starting points for basic research in many subfields of evolutionary biology over the last two decades. There are at least three reasons for this: strict limitations on what can be measured and how precisely those measurements can be obtained working from physical specimens, the fact that not all morphology can be measured externally, and the fact that museum specimens are often fragile, and frequent handling can damage specimens. CT scans represent a digitization modality that can be used to visualize and quantify internal morphology, and allow for the creation of 3D surface scans from which quantifications of morphological structures can be derived without being constrained or limited by complexity of the desired morphological structure, or absolute size of the specimen. The need for such data is so critical that (1) meeting demands from researchers for borrowing specimens for scanning has presented a significant challenge to museum staff (pers. comm. R. Voss, American Museum of Natural History, Dept. of Mammalogy); (2) the majority of budgets for many dissertation level projects is now targeted for scanning equipment, facility fees, or software; (3) researchers often spend the majority of their time traveling, scanning and processing datasets, which is traded for time that could have been put into research design and/or analysis.

Despite this rush to digitize, comparative morphology is experiencing a crisis as a mode of addressing large-scale evolutionary questions due to the difficulty involved in accruing datasets large enough to have high explanatory power, and the small community of researchers that can participate effectively. This presents a paradox: If so many researchers are putting large efforts into scanning, where are the massive samples? Though a few research groups have managed to generate large samples of scans comprehensively representing diversity in one clade or another[1], this work has been time consuming, and expensive: as a result these scans are not made widely accessible to non-collaborating researchers. This inequality in access to what is now essential, basic data clearly falls short of scientific ideals for meritocracy. Furthermore, a significant component of the unmanageable demand for 3D scan data experienced by museums may represent wasteful recollection of data already held by other research groups.

Comparative morphology can be revitalized by democratizing access to microCT scans of specimens in vouchered in museum collections, broadening the community of researchers who can participate in morphological research, and ultimately allowing it join the ranks of other big data science initiatives. In order for this to happen, an infrastructure of efficiently accessing and distributing large numbers of scans is necessary, and researchers and museums must yield their scan collections to it, either voluntarily or through stricter policies from museums and/or funding agencies. The voluntary option is more preferable, but indicates a different incentive structure. Researchers and museums must be able to directly benefit academically and institutionally from third party use of their scans. MorphoSource, the data archive used for this work, addresses this problem by allowing DOI assignment to individual scans and providing usage statistics on each scan.

Nonhuman primate microCT scans

In this paper we announce an openly accessible microCT data sample that—by itself—can catalyse a small transformation in research on primates because it is the first of its kind. Non-human primate skulls are some of the most frequently examined specimens in natural history museums. Anthropologists and mammalogists alike study aspects of morphology in the skulls of members of the human order Primates to answer comparative, taxonomic, phylogenetic, behavioural, biomechanical, and physiological questions. Two of us (LC and LL) collaborated to collect microCT scans of 431 adult and juvenile non-human primate skulls from the Museum of Comparative Zoology at Harvard University for use in our respective dissertations. We have uploaded the scans to MorphoSource.org (see Data Record), an online repository for 3D data, so others can use the scans for their research projects. This being the first openly accessible primate dataset of its kind, we hope it will encourage other researchers to make their own datasets available in equivalent ways, ideally through MorphoSource as well.

Methods

A total of 431 skulls of adult and juvenile non-human primates housed at the Museum of Comparative Zoology at Harvard University were microCT (μCT) scanned at Harvard's Center for Nanoscale Systems. A femur and humerus from some individuals was also scanned. Adulthood was determined by full eruption of the permanent third molars and canines. Any specimen with signs of bony pathology that might have impacted vault or facial growth was excluded. Specimens listed as captive were also not included. Specimens included in the final analyses came from 59 species representing all major families in the order Primates. The only major groups not included are *Phaner*, *Mirza*, *Allocebus*, and *Cheirogaleus* of the Cheirogaleiidae, *Lepilemur* of the Lepilemuridae, or any genus of Daubentoniidae or Tarsiidae. A list of all available specimens, with scanning parameters, is provided in Table 1 (available online only).

The scanner at CNS is a X-Tek HMXST225 μCT scanner. The X-ray detector panel is a Perkin Elmer 1621, which provides a 2,000 × 2,000 pixel and 16 inch × 16 inch field of view with a 7.5 frames per second readout and a physical pixel size of 200 microns. The X-ray source is an X-Tek Nikon microfocus open tube with both reflection and transmission targets. The energy settings for each scan ranged between

70–90 kV and 90–125 µA, depending on the size of the specimen. The strepsirrhine, platyrrhine, and smaller catarrhine specimens were scanned without filters, while *Pan, Pongo,* and *Gorilla* specimens were scanned with a tungsten target to minimize beam hardening.

Each skull was placed in a foam holder that was then positioned inside the scanner on a rotating platform. The foam held the skull in place while still allowing the X-rays to fully penetrate the specimen without leaving visual artefacts. All crania in this study were scanned at parameters optimum for the highest possible resolution within the time available to capture all samples. All crania were scanned using 1,000–1,500 projections, scan time per specimen ranged from 18–60 min, and cubic voxel dimensions ranged from 18 microns for smaller specimens (e.g., *Microcebus*) to 125 microns for the largest (e.g., *Pongo*) (See Table 1 (available online only) for all sample scanning parameters).

Each scan was saved as DICOM files, zipped using either the built-in zip function in OSX (scans < 4 GB) or 7zip (scans >4 GB) and uploaded to MorphoSource. Other contributors should note that standard archiving software, such as that available in OSX, will appear to successfully archive files larger than 4 GB of material, but the archive will be unusable once uploaded to MorphoSource. Thus, any file collection greater than 4 GB should be zipped with WinZip or 7zip, which will use the appropriate ZIP64 format to do the zipping. Scans can be downloaded directly by any registered user of the site. Because not all zip file extractors are compatible with the ZIP64 format, we recommend any PC users unzip files greater than 4 GB using WinZip, winRAR or 7zip.

Once downloaded, users are free to collect their own data from the scans. Several examples of surface renderings created from the scans are shown in Figure 1.

Data Records
MorphoSource
The microCT data from this project are available through MorphoSource (http://www.MorphoSource.org/). We chose to use MorphoSource because it provides a dynamic archive where microCT datasets continually gain relevance by their incorporation into an ever-expanding digital sample representing collections of multiple researchers and institutions. The site was created to meet the new demand for digital datasets discussed above. Its primary aim is to improve researcher access to relevant comparative samples. MorphoSource is the first project-based data repository for storing, collaborative sharing, and distribution of microCT scans, 3D surface renderings, and 2D digital imagery of specimens. The site has been active since April 2013 (refs 2–4). It currently includes 1,432 registered participants from across the globe and hosts ~8,100 files representing 'raw' microCT volumetric data; mesh files (stl, ply) from laser scans, structured light, photogrammetry, or microCT; and 2D digital photographs. These files represent 2,400 repository-vouchered specimens from 73 institutions. The holdings are growing rapidly. Data on the site are protected by creative commons restrictions as customized by each contributing researcher (data author) according to his/her needs, concerns, or third party agreements (e.g., with museums). Most data published on the site can be immediately downloaded by registered users. Other datasets can be released for download upon request, by data authors who retain rights to grant third party access.

The files associated with the current project can be downloaded with open access and are tagged with creative commons copyright license of CC BY-NC as dictated by the copyright holder, the MCZ. This means the data can be downloaded and re-used for non-commercial academic purposes. These limitations are maintained as a component of the non-negotiable terms of the MCZ, the home repository. This framework serves the interests of both physical repositories (museums) and data authors by tracking use statistics on datasets. Such statistics provide evidence of collection value and magnify impact of researcher-collected data. The current MorphoSource dataset is tagged by 489 digital object identifiers (one for each scan, with some specimens represented by multiple scans). As of 1/22/16, the dataset has been viewed more than 31,800 times, and more than 1800 scans have been downloaded. MorphoSource provides search tools to allow users to find, and batch download the samples most relevant to their research design. MorphoSource is free to users and contributors, and the amount of storage space is not explicitly limited. The network storage is distributed between multiple physical locations as part of Duke University's IT data infrastructure.

Each scan dataset is a 'media record' on MorphoSource. The media record includes the metadata on the scan (Table 1 (available online only)) and data files themselves. Searching MorphoSource by specimen will return media records associated those specimens from our data project, as well as media records from other data projects that included digital imagery for those specimens (because other researchers may have scanned and uploaded other bones of the skeleton for the same specimens). Each media record is assigned a digital object identifier DOI, which represents a permanent, direct link to the data and should be cited in any study that uses the scan (in addition to other details—see below).

Museum of comparative zoology
A copy of the complete microCT dataset is also archived at the MCZ and may be accessible by contacting curators there. As well, 3D pdf files depicting a surface rendering of each skull can be downloaded from the MCZ specimen record pages in the museum's online database, MCZbase. On MorphoSource, users will find a link to each specimen's MCZbase page. Researchers should note that these surface renderings are not necessarily to scale currently (whereas all morphosource records are).

Digitized craniometric data available through Dryad

During the process of CT scanning, we also used a 3D digitizer (MicroScribe G2X) to capture 60+ standard craniometric landmarks from each skull, which are illustrated in Figure 2. The points are available to download freely from Dryad, a non-profit repository for data underlying the international scientific and medical literature (Data Citation 6).
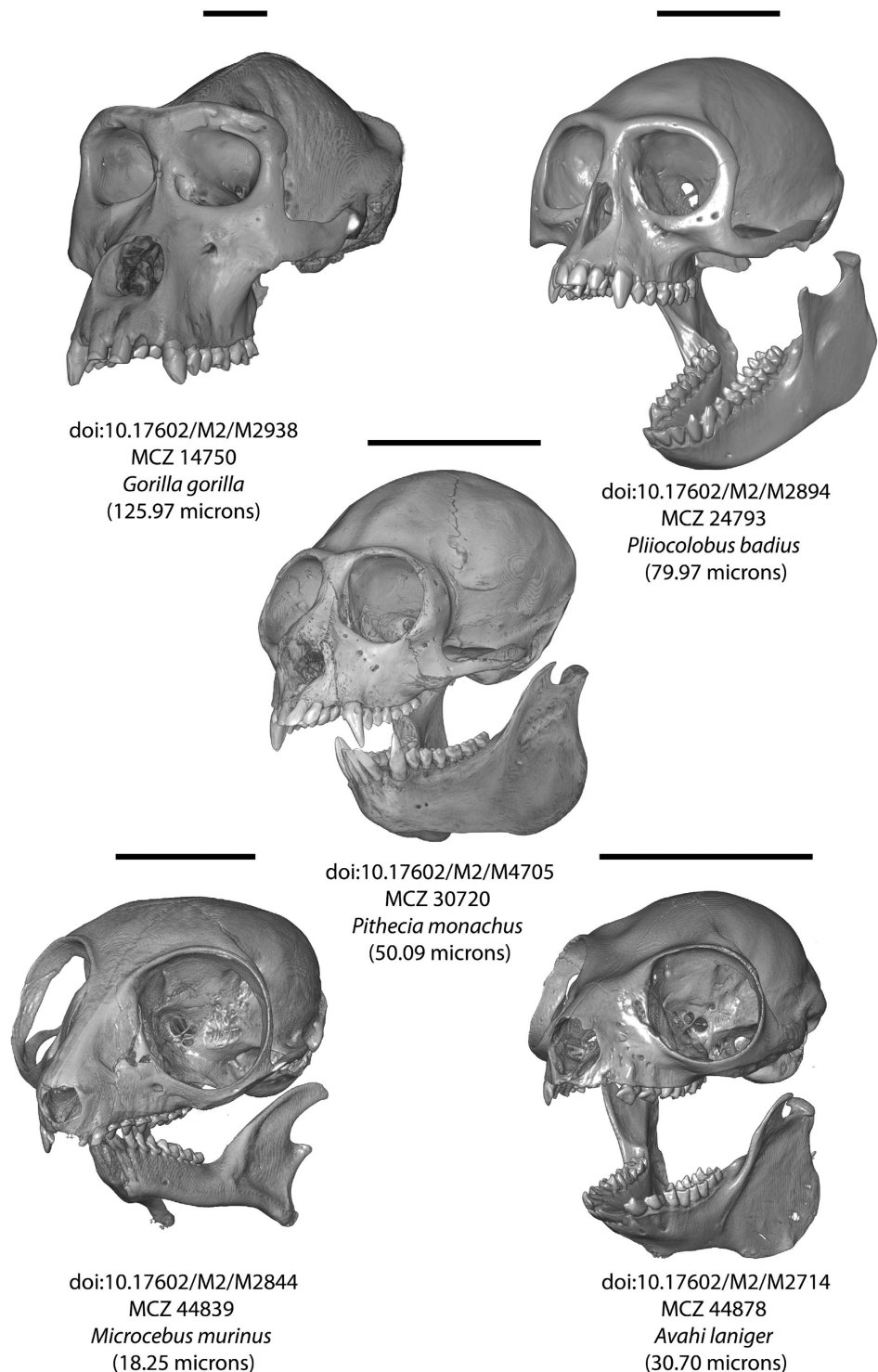

doi:10.17602/M2/M2938
MCZ 14750
Gorilla gorilla
(125.97 microns)

doi:10.17602/M2/M2894
MCZ 24793
Pliiocolobus badius
(79.97 microns)

doi:10.17602/M2/M4705
MCZ 30720
Pithecia monachus
(50.09 microns)

doi:10.17602/M2/M2844
MCZ 44839
Microcebus murinus
(18.25 microns)

doi:10.17602/M2/M2714
MCZ 44878
Avahi laniger
(30.70 microns)

Figure 1. Renderings of voxel data (done with Avizo 8.1) on five of the 489 µCT scan datasets announced in this publication. Note that scans of some specimens include the cranium and the mandible in a single dataset, while for others specimens these elements were scanned separately. The full web links to the DOIs listed in the figure can be found in Data Citations 1–5.
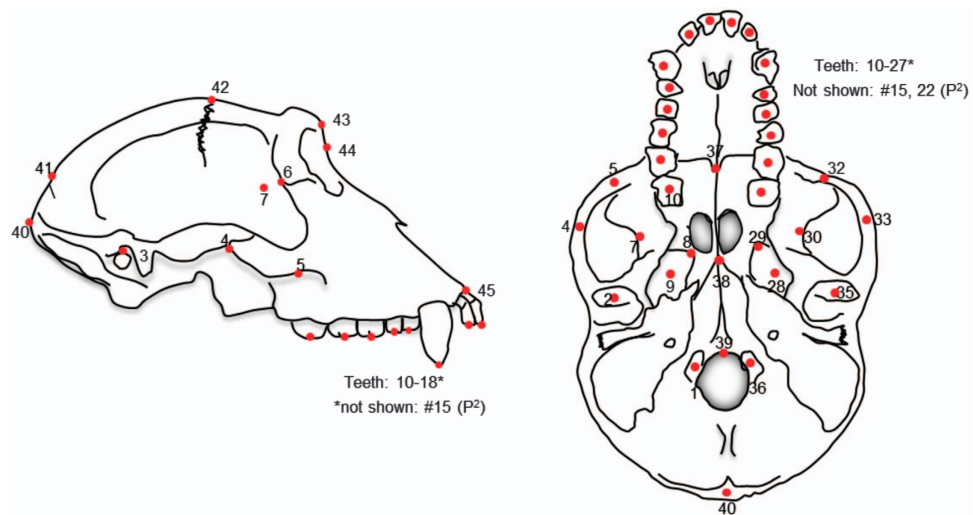
Figure 2. Landmarks digitized on each skull using a MicroScribe. Data are available for download from Dryad (Data Citation 6).

In order to gauge inter- and intra-observer error in the craniometric dataset, some specimens were digitized repeatedly. A single researcher (LL) initially digitized all specimens once. Thirty-one specimens were chosen to be re-digitzed three times each by a second researcher (LC). Seven distances were calculated and compared to measure error. Intraobserver error was nearly always $< 5\%$, but in some cases, interobserver error exceeded 5%, especially in the smaller specimens. In such cases, it is likely that the two independent researchers disagreed on the exact location of certain landmarks. However, we are confident that the 3D landmark data provided for each specimen is reliable and can be used confidently for future work.

Technical Validation

Calibrations of non-metrology specific industrial x-ray CT scanners that guarantee a certain minimum error for a particular machine at all settings do not exist, and almost all existing microCT datasets have been collected with such non-metrology specific machines (i.e., all of the major academic scanning facilities in the US have non-metrology specific units: Penn State's Center for Quantitative Imaging, The University of Texas' High-Resolution X-ray Computed Tomography Facility, American Museum of Natural History's Microcopy and Imaging Facility). Industrial x-ray CT scanners that are used for high accuracy ($< 0.01\%$ error) metrology only calibrate to a single setting at a time (i.e., fixed source, detector and stage settings) with limited ability to scan at different configurations and retain the same accuracy. However, non-metrology scanning facilities (including Harvard's—the facility where the scans were made) do not typically attempt to record error levels in their voxel sizes. It is assumed that error levels are typically around or below 1% for dimensional based measurements. The accuracy of which is largely determined during the instrument's initial installation and any further check of calibration done by the facility or other service personnel. However, assuming adequate preventative maintenance has been administered, different machines with the same components and configurations should produce similar errors of around 1% (Greg Lin, personal communication).

We decided to empirically evaluate the expected error under the range of settings used in this study by analyzing results of scanning calibration balls at different resolutions on Duke University's scanner in the Shared Materials Instrumentation Facility. This scanner (XTH 225ST) is a similar model to the one at Harvard (HMX 225ST). Most importantly they incorporate the same X-ray source (Nikon 225 Reflection tungsten target, with focal spot of 3 um up to 15 W) and detector plate (Perkin Elmer AN1620) with similar range for source-detector distances available (~1.3 m). Therefore our results should correlate with the machine at Harvard as well.

To determine potential error in the scans using a similar scanner, four standard spheres of 3.175 mm ($+/- 1.0\,\mu m$ tolerance) were scanned at voxel resolutions of 5, 6, 7, 8, 9, 10, 15 and 20 μm per voxel, and four standard spheres of 12.7 mm ($+/- 1.0\,\mu m$ tolerance) at voxel resolutions of 30, 40, 50, 60 and 70 μm per voxel on a Nikon XTH 225 ST. Each scan was collected at 150 keV, 80 μA (12 W), 354 ms, 800 projections and a projection average of 1. No filters were used and the projection data was reconstructed into volumetric data via use of Nikon's proprietary *CT Pro 3D* and *CT Agent* software. Digital measurements of the standard spheres were performed in Volume Graphics *VG Studio Max 2.2* along with Volume Graphics's *Coordinate Measurement Module*.

For measurement of the spheres, surfaces of all four (in both sets) were first generated, fit points (>10) were then placed on the surface and an idealized sphere was then fit to these points. Via use of the *Coordinate Measurement Module's* sphere fitting function, diameters of each sphere were recorded, averaged and then compared to the physically measured diameter. The relative percentage error was calculated with the following equation: Relative error % = ((average measurement of sphere diameter from VG—reported value of sphere from manufacturer)/(average measurement of sphere diameter from VG))*100. Table 2 gives data on relative errors. Figure 3 plots these error values against resolution.

These calibrations at different resolutions using spheres of known diameter suggest negligible error in absolute size or shape at the scanning resolutions between 5–70 microns for X-Tek microCT scanners (< 0.2%). A comparison of digital and physical measurements of a skull from Duke's collection, which was measured using calipers, scanned at 109 microns (cubic voxel dimension), and then measured digitally, indicate an error of < 0.9%. The coarsest resolution used for the MCZ specimen scans is 125 microns, and while we do not have data on comparison between physical vouchers and digital avatars at this scale, we are confident that the potential for significant scale error is low.

Usage Notes

Working with the data

The data appear as series of 2D image files, with each image representing a cross section through the specimen. The file format is DICOM, which can be read by freeware ImageJ[5], Avizo or Amira. ImageJ is predominantly useful for viewing the image cross sections. The whole series of files can be loaded into ImageJ by going to 'File -> Import -> Image Sequence'. Often the contrast will look poor on initial opening, but this is merely a default issue. Go to 'Image -> Adjust -> Brightness/Contrast' and click 'Auto' to reset contrast or adjust the sliders to the desired image brightness and contrast. From ImageJ, the files can be re-saved as tiff format (16-bit) for modification or annotation in Adobe Photoshop, Illustrator or an equivalent program.

Three-dimensional visualization of these data is most easily done in Avizo or Amira. Two examples of freeware are Slicer3D or Fiji[6]. Other volume display and manipulation software include VG Studio MAX, Osirix, and Mimix.

Computer requirements

In order to open a particular file in ImageJ your computer should have RAM exceeding the complete sample file size to some degree (the greater the better). As a rule of thumb, one should have twice the amount of RAM installed in his/her computer than the largest file size(s) one wishes to open. Otherwise not all of the images of the scan will open, and any processing will be extremely limited by the computers lack of space to manipulate the data. The minimum requirements for 3D visualization is similar, except more stringent. In this case, it is imperative that a computer is equipped with RAM equalling at least two times the file size for satisfactory processing and results with a 64-bit operating system (OS) installed—in many cases, a computer with at least 12 GB of RAM will suffice. Once this condition is met, the most important components are having a processor with high clock speed (>3 GHz), sufficient number of cores (≥2), and a higher-end graphics card (DDR5 with ≥2 GB).

Voxel Size μm	Relative Error (%)
5.00274654	0.143102011
6.002784351	0.121112981
7.004121857	0.141531687
8.00419928	0.110114834
9.004866272	0.124254864
10.0052979	0.122683947
15.00729757	− 0.009449712
20.00988969	0.099114265
30.01900974	− 0.147461222
40.02104118	− 0.157728707
50.04978064	− 0.236779795
60.04728152	− 0.157728707
70.04123235	− 0.157728707

Table 2. Voxel size and relative error for Nikon, X-Tek XHT 225 ST scanner at Duke University (results comparable for Harvard's X-Tek HMX 225 ST μCT scanner).
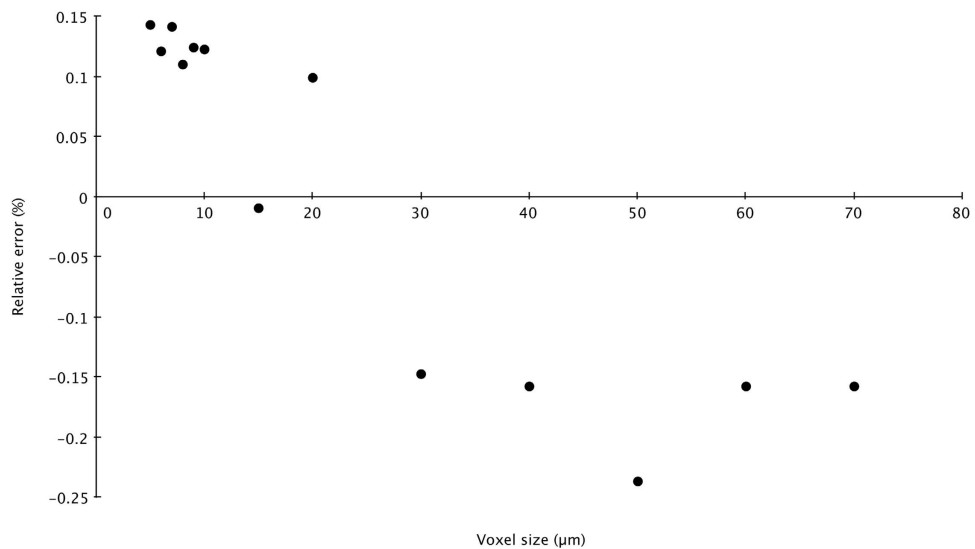
Figure 3. Voxel size versus relative error for Nikon, X-Tek XHT 225 ST scanner at Duke University (results comparable to Harvard's X-Tek HMX 225 ST μCT scanner).

Citation of scans

Any publications using scans from this dataset should do the following: (1) list DOIs of all scans used; and (2) include the following statement in the acknowledgments: 'Lynn Copes, Lynn Lucas, and the MCZ provided access to this [or these] scan[s], originally appearing in Copes and Kimbel[7] and Copes *et al.* (2016), funding for the collection of which was provided by NSF DDIG #0925793, and a Wenner-Gren Foundation Dissertation Grant #8102 (both to Lynn Copes). These scans were downloaded from MorphoSource.org, a web-accessible archive for 3D digital data housed by Duke University.'

References

1. Spoor, F. *et al.* The primate semicircular canal system and locomotion. *Proc. Natl. Acad. Sci. USA* **104,** 10808–10812 (2007).
2. Boyer, D. M., Kaufman, S., Gunnell, G. F., Rosenberger, A. L. & Delson, E. Managing 3D digital data sets of morphology: MorphoSource is a new project-based data archiving and distribution tool. *Am. J. Phys. Anthropol.* **153,** 84A (2014).
3. Boyer, D. M., Kaufman, S., Gunnell, G. F., Gomes, E. & Thostenson, J. MorphoSource: A currently active project-based 3D digital web-accessible data archive for museums and individuals at Due University USA. *Digital Specimen Abstracts of Presentations* (2014).
4. Kaufman, S. & Boyer, D. M. Developing platforms for management and distribution of digital specimen data. *Digital Specimen Abstracts of Presentations* (2014).
5. Schenider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9,** 671–675 (2012).
6. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9,** 676–682 (2012).
7. Copes, L. E. & Kimbel, W. H. Cranial vault thickness in primates: *Homo erectus* does not have uniquely thick vault bones. *J. Hum. Evol.* **90,** 120–134 (2016).

Data Citations

1. Copes, L. E., Lucas, L. M., Thostenson, J. O., Hoekstra, H. E. & Boyer, D. M. *MorphoSource.* http://dx.doi.org/10.17602/M2/M2938 (2015).
2. Copes, L. E., Lucas, L. M., Thostenson, J. O., Hoekstra, H. E. & Boyer, D. M. *MorphoSource.* http://dx.doi.org/10.17602/M2/M2894 (2015).
3. Copes, L. E., Lucas, L. M., Thostenson, J. O., Hoekstra, H. E. & Boyer, D. M. *MorphoSource.* http://dx.doi.org/10.17602/M2/M4705 (2015).
4. Copes, L. E., Lucas, L. M., Thostenson, J. O., Hoekstra, H. E. & Boyer, D. M. *MorphoSource.* http://dx.doi.org/10.17602/M2/M2844 (2015).
5. Copes, L. E., Lucas, L. M., Thostenson, J. O., Hoekstra, H. E. & Boyer, D. M. *MorphoSource.* http://dx.doi.org/10.17602/M2/M2714 (2015).
6. Copes, L. E., Lucas, L. M. & Boyer, D. M. *Dryad Digital Repository.* http://dx.doi.org/10.5061/dryad.dm57j (2015).

Acknowledgements

Author Contributions

L.E.C. scanned specimens, uploaded them to MorphoSource, re-digitized some specimens for error analysis, uploaded the digitized data to Dryad, and helped write the paper. L.M.L. scanned specimens, took initial 3D digitized landmarks on each, and helped write a section of the paper. J.O.T. helped in MorphoSource development, conducted an error/calibration study of the microCT scanner, and helped write a section of the paper. H.E.H. is the curator of mammalogy at the Museum of Comparative Zoology. D.M.B. created MorphoSource, provided technical support for the uploading process, and helped write the paper.

Additional Information

Table 1 is only available in the online version of this paper.

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Copes, L. E. *et al.* A collection of non-human primate computed tomography scans housed in MorphoSource, a repository for 3D data. *Sci. Data* 3:160001 doi: 10.1038/sdata.2016.1 (2016).