Analysis of spheniscid humerus and tarsometatarsus morphological variability using DAISY automated image recognition.

Stig A. Walsh[1,2], Norman MacLeod[1] and Mark O'Neill[3]

[1]The Natural History Museum, Department of Palaeontology, Cromwell Road, London, UK, SW7 5BD. s.walsh@nhm.ac.uk
[2]School of Earth and Environmental Sciences, Portsmouth University, Burnaby Road, Portsmouth, UK, PO1 3QL.
[3]Centre for Neuroecology, Henry Wellcome Building, University of Newcastle upon Tyne, Newcastle upon Tyne, UK, NE1 7RU.

ABSTRACT - Despite a long history of research, relationships within fossil and extant Sphenisciformes remain unclear. This is largely because most fossil species were described on the basis of either the tarsometatarsus or humerus. Neither of these elements is particularly phylogenetically informative, and the extent of intraspecific morphological variation also remains unknown. Herein we investigate a new approach – the use of artificial neural-net (ANN) technology – to determine whether either of these elements can be reliably used to identify extant species. The DAISY ANN system was able to recognise most species from either tarsometatarsal or humerus morphology, but its success rate improved when the species training sets were combined into generic groups, indicating the need for larger image libraries. Our preliminary results suggest that these elements can allow reliable identifications for most taxa, but that the tarsometatarsus is on the whole a better element for this purpose. These results also demonstrate the potential for artificial neural-net technology to address problems in avian taxonomy.

Keywords: *Spheniscidae, morphology, intraspecific variability, automated identification, DAISY.*

L'analyse de la variabilité morphologique d'humérus et de tarsometatarsus de Spheniscidae employant DAISY a automatisé l'identification d'image - En dépit d'une longue histoire de recherche, les rapports entre les Sphenisciformes fossiles et actuels demeurent peu clairs. C'est en grande partie parce que la plupart des espèces fossiles ont été décrites sur la base du tarsométatarse ou de l'humérus. Ni l'un ni l'autre de ces éléments n'est particulièrement instructif du point de vue de la phylogénie, et l'ampleur de la variation morphologique intraspécifique demeure également inconnue. Nous avons essayé une nouvelle approche - l'utilisation de la technologie de réseau neural artificiel (ANN) - pour étudier si ces éléments peuvent être employés de façon fiable pour identifier des espèces actuelles. Le programme DAISY ANN a pu identifier la plupart des espèces, mais son taux de succès est meilleur quand les espèces formant des ensembles ont été combinées dans des groupes génériques, indiquant le besoin de plus grandes bibliothèques d'images. Nos résultats préliminaires suggèrent que ces éléments peuvent permettre des identifications fiables pour la plupart des taxons, mais que le tarsométatarse est dans l'ensemble un meilleur élément à cette fin. Ces résultats prometteurs montrent le potentiel de la technologie ANN pour l'étude des problèmes de la taxonomie avienne.

Mots clés: *Spheniscidae, morphologie, variabilité intraspécifique, identification automatisée, DAISY.*

INTRODUCTION

With a fossil record extending at least as far back as the early Palaeocene (Slack et al., 2006), Sphenisciformes (penguins) has a fossil record longer than many extant avian clades. The first fossil penguin remains were described by Huxley in 1859 and, by now, one would expect this group's early evolution, systematics and historical diversity pattern to be well understood. This is not the case. Two main impediments to fossil spheniscid research exist. The first of these is that, while their remains are not uncommon in some strata, their stratigraphic and geographic distribution is somewhat disjunct (Fordyce & Jones, 1990). New discoveries, particularly in New Zealand and South America (e.g., Slack et al., 2006; Walsh & Suarez, 2006; Acosta Hospitaleche et al., 2007), may in time help to resolve this aspect of uncertainty.

A second and potentially more serious problem is that the majority of remains come from nearshore, relatively high-energy sequences, where skulls and articulated or asso-

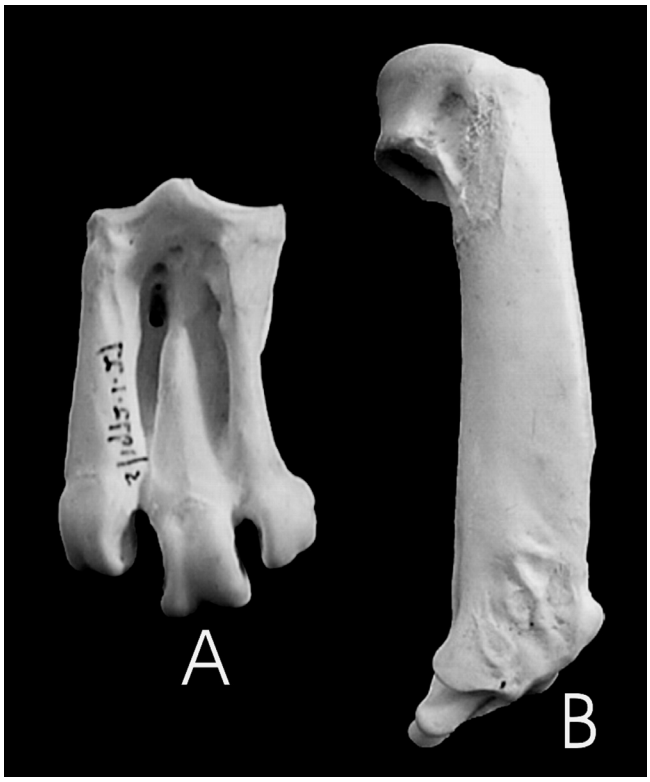

Figure 1 - Spheniscid postcranial skeletal elements considered in this study: A, right tarsometatarsus in dorsal view; B, right humerus in caudal view. Both specimens from *Spheniscus magellanicus*.

ciated remains are rather rare. In fact, most described species are based on only a single postcranial element, usually either the humerus or tarsometatarsus (Fordyce & Jones, 1990; fig. 1). These robust elements are distinctive in Sphenisciformes, but exhibit few synapomorphies, even when well-preserved (Simpson, 1946; Zusi, 1975; Fordyce & Jones, 1990). Further, in osteological collections of extant taxa, those synapomorphies that are present often prove variable within a single species (Simpson, 1946; Zusi, 1975; Walsh, unpublished data), making phylogenetic analysis problematic. The paucity of well-defined characters included within these elements has meant that most fossil species have been recognised on aspects of overall morphology rather than by clear apomorphies. Considering the conservative nature of spheniscid postcrania, and that the morphology of these elements may be highly variable within extant species, this raises the important question of how reliable are the humerus and tarsometatarsus for fossil spheniscid identification?

Here, we present preliminary results from the first direct attempt to determine whether spheniscid tarsometatarsus and humerus morphological intraspecific variability is sufficiently restricted to allow reliable taxonomic identification. To accomplish this task objectively we employ a new technique currently being developed at the Natural History Museum, London, involving a prototype artificial neural-net image identification system. This study is, so far as we are aware, the first to use such an approach to address problems of morphological variability in relation to species identification.

RATIONALE OF STUDY

Morphological variability in the humerus and tarsometatarsus of fossil spheniscids has been difficult to test due to insufficient numbers of specimens. However, osteological collections of extant species are plentiful, and should provide a reasonable approximation of the characteristic variability of these skeletal elements in extinct forms. A multivariate morphometric approach would be the most appropriate for addressing this problem in detail. Although Myrcha et al. (2002) used simple linear distance measurements and ratios of tarsometatarsi to revise the Eocene spheniscid species of Antarctica, only Livezey (1989) has attempted to use multivariate geometric morphometric techniques on penguin bones. Livezey's (1989) study used distance-based variables (e.g., relative flatness of humerus = least diaphysial width/maximum diaphysial width) and succeeded in providing a measure of sexual dimorphism in living species. In addition, this study identified allometric trends in the wing and hind limb in fossil and living species. However, Livezey's study was not concerned with intraspecific morphological variability. More importantly, the distance-based variables employed would have been incapable of addressing aspects such as element shape, the position, size and shape of homologous features (e.g., tuberosities, fossae and foramina), or the sectional shape of the element.

Several suitable techniques capable of analysing these factors are currently available. In these, landmarks are chosen that characterise the shape and size of the specimen, and the position of any other features within the specimen outline. Data are usually collected from digital images of the specimens and used to create linear models of shape variability in authoritatively identified 'training sets' of each class (in this case, species of penguin). Analyses using linear distances or constellations of superimposed landmarks can provide some measure of how the element shape and position of landmarks change between species, whereas analysis of an element's outline – or of any curve within that object outline – focuses on variability of overall element shape (e.g., eigenshape analysis; MacLeod, 1999). A combination of these techniques would provide a good assessment of the variability in shape and position of key features. However, even in combination, these methods are less suitable to address the more subtle differences in three-dimensional morphology that an expert palaeoornithologist takes into account when identifying a spheniscid humerus or tarsometatarsus (e.g., on the tarsometatarsus: shape and development of the m. tibialis tuberosity or the relative 'roundness' of the metatarsals and the depth of the inter-metatarsal sulci).

One alternative approach addresses this problem

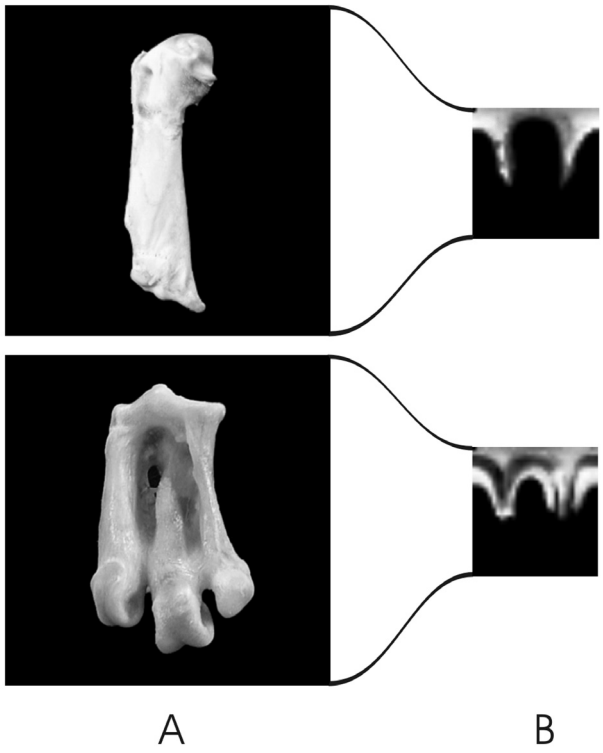

Figure 2 - Transformation performed by DAISY during training set construction. A, Cartesian format 500 x 500 pixel TIF images are loaded and then subsampled to B, 32 x 32 polar format thumbnails.

by combining information about 3-D morphology, outline shape, and relative size, shape and position of features (e.g., tuberosities) contained in digital images of specimens themselves. This method requires only a desktop computer and a digital camera, and is a fast and cost-effective way to analyse 3-D morphology.

Originally developed as an image-recognition system for entomology, the Digital Automated Identification SYstem (DAISY) has proven remarkably successful with a wide range of object types, from microfossils to dinosaur teeth (Gaston & O'Neill, 2004; MacLeod et al., 2007). Put simply, DAISY is an artificial neural net (ANN) capable of grouping images of objects together based on each object's spectrum of pixel brightness values. The 'neurons' of this system are software switches designed to respond to the strength of input signals. These artificial neurons are assigned numerical weights that amplify or diminish the strength of the input image signal based on differences between collections of authoritatively identified image-training sets. DAISY incorporates a variant of the Lucas *n*-tuple nearest neighbour classifier and plastic self-organising mapping (PSOM), making DAISY capable of self-learning. As such, the DAISY approach incorporates an aspect of artificial intelligence.

DAISY accepts authoritatively identified, colour or greyscale, uncompressed images in the tagged image file format (TIF). When the system builds the training set, the images are histogram normalized, sub-sampled from 500 x 500 pixels to a 32 x 32 pixel grid thumbnail, and transformed from a Cartesian to an equivalent polar format in which the pixels of the image are represented by their angular deviation and distance from the central point (fig. 2a, b). The histogram normalisation removes strong variations in brightness and contrast, whereas the thumbnail subsampling massively reduces the computing requirements. The polar transformation is intended to downgrade information about object outline shape to focus the system on patterning internal to the object.

As these transformations are performed, DAISY builds a discriminant space for each species using the PSOM/*n*-tuple classifier. Identifications are then made by determining the *n*-fold nearest-neighbour coordination between a new image and the training set ordination. Identifications are of three kinds, depending on the level of confidence. The highest is a *coordination* identification, in which the position of an unknown image in the DAISY ordination is given relative to the number of nearest neighbours of the same class. The lowest coordination number is set by default at three nearest neighbours, below which the identification falls to the next level, the *sill* identification. These identifications indicate the proximity of the unknown image to the edge of a species cluster within the training set ordination. The lowest support is the *majority vote*, in which the class of each object closest to the unknown image is listed, and the level of certainty for the identification given in per cent. Below this level DAISY is unable to identify the unknown image.

DAISY uses the same information (patterns of light and shade) available to a palaeoornithologist comparing published photographs of fossil specimens. The DAISY system is currently set to ignore scale and, most importantly, must therefore make its identifications based solely on morphological information contained within an image. Subjective 'operator bias' and 'guessing' are thus eliminated from the analysis. Provided the aspects of interspecific variability captured by the penguin bone images is sufficiently structured to allow DAISY to cluster images of the same class, DAISY should be capable of making accurate identifications. If within-groups (intraspecific) variability is strong within the training-set images, either no clustering will occur, or the clusters will grade into each other. In either case, DAISY will be unable to make consistent positive identifications.

Our specific objectives are therefore to (1) determine whether the intraspecific variability present in the humerus and tarsometatarsus of extant spheniscid species is great enough to cause the DAISY ANN to fail in routine identification tasks at the specific and generic level; (2) use DAISY's success rate to indicate which element is most useful for identification purposes, and (3) discover which of the two standard views of these bones contains the greatest amount of relevant identification information.

	Humerus cranial	Humerus caudal	Tarsomet. dorsal	Tarsomet. plantar
Aptenodytes forsteri	7	6	10	8
Aptenodytes patagonicus	6	8	9	9
Aptenodytes (combined)	13	14	19	17
Eudyptes crestatus	15	9	12	3
Eudyptes (combined)	19	9	12	3
Eudyptula minor	8	6	6	4
Pygoscelis adeliae	9	15	17	10
Pygoscelis antarctica			4	3
Pygoscelis papua	8	8	7	6
Pygoscelis (combined)	19	23	28	19
Spheniscus demersus	6	9	11	8
Spheniscus humboldti	4	6		3
Spheniscus magellanicus	4	6	8	5
Spheniscus (combined)	14	21	19	16

Table 1 – Image numbers tested in each class. Note that for the combined genus class *Eudyptes* and *Pygoscelis* in the cranial humerus training set, the figure is higher than for *Eudyptes crestatus* because it was possible to include images for *E. sclateri* and *E. pachyrhinchus*, and for *Pygoscelis* because *P. antarctica* was included. These species were not included in the species-level training set due to insufficient numbers.

METHODS AND MATERIALS

A hand-held Fujifilm S3000 digital camera was used to acquire 345 RGB colour images at 72 dpi. All images were of extant taxa comprising ten species in five genera from 65 individual skeletons held in the collections of the Natural History Museum, Tring. Due to lack of suitable specimens *Megadyptes antipodes*, *Eudyptes chrysocome*, *E. robustus E. chrysolophus*, *E. pachyrhynchus*, *E. sclateri* and *Spheniscus mendiculus* were not included in this study, and *Eudyptula minor albosignata* was included as *Eu. minor*. However, where training sets of species were combined to provide training sets of genera, it was sometimes possible to incorporate images of some of the above species (*E. pachyrhynchus* and *E. sclateri*) that comprised too few images to be included in the species-level tests. Due to the condition of some specimens, some species were possible to image in one view, but not another. Consequently, numbers of images between views of the two elements are not consistent. Details of numbers of images used for each species, genus, element and view are given in Table 1.

Each specimen was imaged in two views—caudal and cranial for the humerus; dorsal and plantar for the tarsometatarsus—on a matte black background to maximise contrast between the specimen and its background. All specimens were illuminated consistently from the top left for right-hand specimens, and from the top right for left-hand specimens in order to minimize extraneous image variation caused by shadowing. Of the 345 images originally taken, 56 were rejected due to quality. The remaining 289 images were digitally maximised for brightness and contrast, and corrected for artifacts (e.g., accession numbering, wire connectors, adherent soft tissue) using Corel Photo-Paint 11.0. For consistency, each specimen image was segmented from the original background, inserted onto a black 72 dpi 500 x 500 pixel grid, and, where necessary, reorientented to a standard north-south pose. For right-hand standardisation, images of left-hand elements were mirrored using Corel Photopaint's 'flip horizontally' tool. Each image was then resampled to 8-bit greyscale to minimise pattern interference from variable staining of the bone surface.

Each image in the training set was then named according to taxon, and assigned a sequential number for cross reference to the raw image library. Identification using the DAISY approach is dependent on the accuracy of the original identifications. Since the majority of spheniscid specimens held at the Natural History Museum, Tring, are from wild or captive individuals identified *in vivo*, the collection identifications were regarded as accurate. After the images were loaded the training set ordination was constructed using the DAISY build tool. Consistency of the training set ordination for each element and view at species and genus level was achieved by using the DAISY jack-knife test function. This function employs sequential cross-validation testing of each image against the rest of the training set ordination to determine which images in the training set pass and which fail. By highlighting images that fail, this test is often useful for determining the reason for the failure, since obvious differences from the rest of the class are often apparent when viewing the image itself. Overall pass/fail results from each jack-knife test were used as a measure of the success of the system for each element and view. Breakdown of the cross-validation results for each species in each element and view was used to determine how well each species performed as a

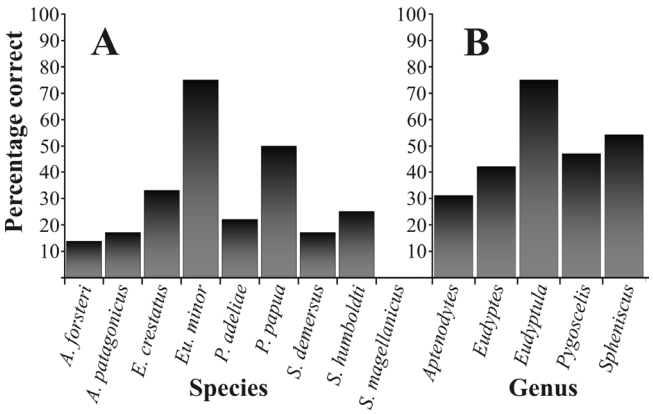

Figure 3 - Results of the cranial humerus test in terms of percentage correct. A, species-level test; B, genus-level test.

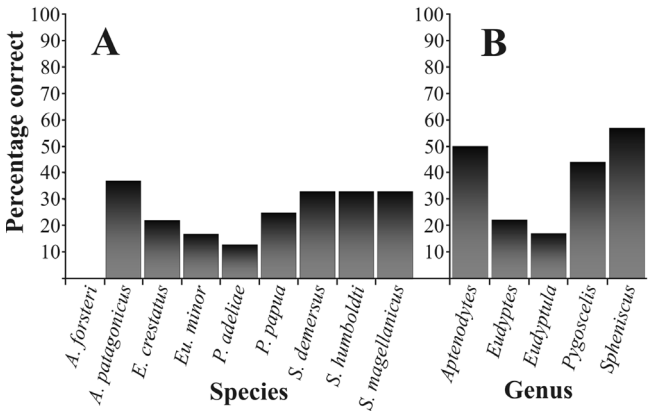

Figure 4 - Results of the caudal humerus test in terms of percentage correct. A, species-level test; B, genus-level test. Note the poorer overall result compared with the cranial humerus test, suggesting less identification-related information is contained in this view of the spheniscid humerus.

subject for identification purposes. For this study, only identifications at the coordination and sill level were accepted.

As a control condition, the sequential numbering of each image in the training set for the plantar view of the tarsometatarsus was reassigned according to a random number set (created by the random number generation function in Microsoft Excel 2000). This randomised training set was tested for species partitioning in the same way as the other sets, with the rationale that no groupings should be present due to the increased within-group variation.

RESULTS

Humerus – cranial view. At the species level (fig. 3a) jack-knife testing revealed clear but weak groupings of taxa, with over 31 per cent correct identifications. It must be remembered that although the identification failures are high, the clustering is clearly much better than random (mean coordination of four nearest neighbours). The best results were obtained for *Eu. minor*, with 75 per cent correct; *S. demersus* performed the poorest with no correct identifications. At the genus level (fig. 3b) the results were improved greatly, with 47 per cent correct identifications. *Eudyptula* again achieved 75 per cent correct identifications, with *Spheniscus* managing 54 per cent correct. The poorest performing genus was *Aptenodytes* with only 31 per cent correct.

Humerus – caudal view. The species-level test (fig. 4a) performed worse overall than the species-level test of the cranial view, with only 23 per cent correct. Mean coordination level was lower at three nearest neighbours, indicating a weaker grouping of images. *A. forsteri* performed the poorest, with no correct identifications, and *Eu. minor* also performed poorer than in the cranial view test, with only 17 per cent correct. *Aptenodytes patagonicus* performed the best, with 37 per cent correct. Each of the *Spheniscus* species achieved 33 per cent correct. When the species images were combined as genera the overall results (fig. 4b) rose to 44 per cent correct, and a mean coordination of 3.75. The best performing genera were *Spheniscus* and *Pygoscelis* with 57 per cent and 50 per cent respectively. This time *Eudyptula* was the worst with only 17 per cent correct.

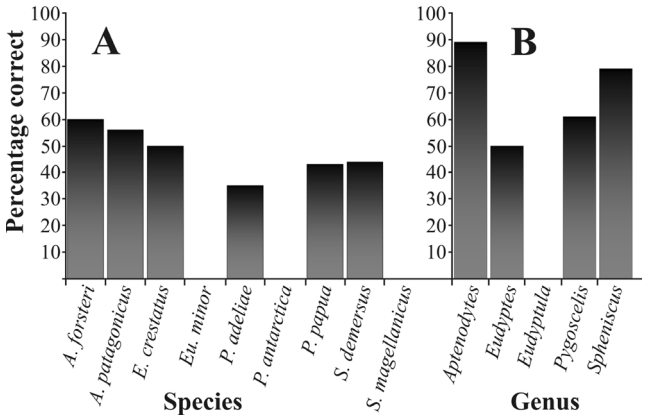

Figure 5 - Results of the dorsal tarsometatarsus test in terms of percentage correct. A, species-level test; B, genus-level test. These results are noticeably better than for either of the humerus views, although three species and one genus failed to achieve any correct identifications.

Tarsometatarsus – dorsal view. With an overall correct identification rate of 37 per cent at the species level (fig. 5a), the dorsal view performed better than either view of the humerus. However, the mean coordination of three nearest neighbours suggests that this result is attributable to relatively strong clustering in some species, but no clustering in others. For instance, *Aptenodytes forsteri* achieved 60 per cent correct, with *A. patagonicus*, *P. adeliae*, *P. papua*, *E. crestatus* and *S. demersus* all achieving between 35 per cent and 55 per cent correct. Conversely, *Eu. minor*, *S. magellanicus* and *P. antarctica* acheieved no correct identifications.

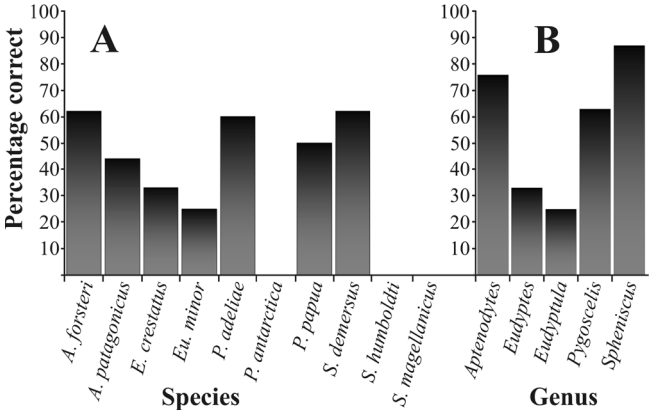

Figure 6 - Results of the plantar tarsometatarsus test in terms of percentage correct. A, species-level test; B, genus-level test. This element and view performed the best overall in this study, suggesting that the plantar view of the spheniscid tarsometatarsus contains a relatively good amount of identification information. However, species of *Spheniscus* may not be best separated by either view of this element.

At the genus level (fig. 5b), the overall pass was yet further improved at 65 per cent correct, with a far higher mean coordination of five nearest neighbours. *Aptenodytes* achieved 89 per cent correct, with *Spheniscus* (79%), *Pygoscelis* (60%) and *Eudyptes* (50%) all performing very respectably. Only *Eudyptula* failed to achieve any correct identifications.

Tarsometatarsus – plantar view. The species-level test (fig. 6a) achieved an overall correct identification rate of 42 per cent, with a mean coordination of 3.3 nearest neighbours. Individual species also achieved better results than the dorsal view test, with *A. forsteri*, and *S. demersus* joint best at 63 per cent, closely followed by *P. adeliae* with 60 per cent correct. *Spheniscus humboldti*, *S. magellanicus* and *P. antarctica* performed the worst with no correct identifications. The remaining species achieved between 25 per cent and 50 per cent correct. At the genus level (fig. 6b), 69.5 per cent correct identifications were achieved overall, with a mean coordination of almost five nearest neighbours. No genus failed to achieve correct identifications, although *Eudyptula* performed the poorest with 25 per cent correct. The highest was *Spheniscus* at 87 per cent correct, but the remaining genera performed well with *Aptenodytes* at 76 per cent, *Pygoscelis* at 63 per cent and *Eudyptes* at 33 per cent correct.

As expected, the randomised training set for the species-level test of the plantar view of the tarsometatarsus produced no correct identifications, and hence showed no clustering. This result supports the partitions recovered within the other four training set conditions.

DISCUSSION AND CONCLUSIONS

These results suggest a reasonable morphological separation of humeri and tarsometatarsi in most spheniscid species and genera. Consequently, it would appear that there is enough tonal information in monochrome images of these skeletal elements for basic species recognition, and that the intraspecific variability is not in principal an overriding impediment to achieving consistent identifications.

The tarsometatarsus is clearly the better element for identification purposes at both specific- and generic-levels, particularly for species of *Spheniscus* and *Aptenodytes*. However, although the humerus is a poorer element overall for identification, it is particularly effective at identifying *Eudyptula*. We suspect the humerus results relate to the distinctively narrow and straight humerus of that genus. However, the results for the tarsometatarsus may be more complex. While the broad and short tarsometatarsus of species of *Aptenodytes* is distinctive relative to that of the other species, the tarsometatarsus of *Spheniscus* does not appear to be particularly different from the other genera.

These observations suggest that DAISY is indeed using patterns of light and shade within the images, but that the object outline and overall shape remain important identification criteria. Investigation of these possibilities is possible by comparing the results achieved through performing an outline analysis (such as eigenshape analysis) of the elements, and a Procrustes-type landmark analysis in which the main areas of light and shade are delimited by the position of the landmarks. We are currently investigating this possibility using the same image library as analysed in the present study.

While *S. demersus* is slightly easier to separate from the other *Spheniscus* species, *S. demersus*, *S. humboldti* and *S. magellanicus* are all apparently difficult to separate using either of these elements. This result is at least consistent with the cytochrome *b* sequence-based phylogeny of Davis and Renner (2003), where all *Spheniscus* species appear closely related and probably represent a very recent radiation (Davis & Renner, 2003; p. 35; see also Bertelli & Giannini, 2006). If so, recently described fossil evidence that places the earliest known fossil representative of *Spheniscus* in the Miocene of South America (Walsh & Hume, 2001; Stucchi, 2002; Stucchi et al., 2003; Bertelli & Giannini, 2006; Göhlich, 2007) would suggest that representatives of the other Recent genera should be expected in even older sediments.

Separation of genera using these elements is better supported. This is most likely because grouping the species images into genera resulted in creation of larger image training sets. The DAISY system would have failed more often in this test if the interspecific variability had caused the generic clusters to interdigitate within the generic ordination. As a proof-of-concept test of the DAISY software's ability to address the morphological variation problem in the tarsometatarsus and humerus of penguins, low numbers of images were adequate. Although disadvantaged by this small training set and absence of information about relative size of specimens (size being an important key to identification for a human expert, who also has access to written descriptions),

the success of the system indicates a full test at the specific level will be worthwhile. This is currently underway, but will require further images of existing taxa, and training sets of images for species that were not included in the current test. From previous experience with other object types we believe that a training set of around 30 images per species will be sufficient for this purpose.

Overall, the cranial view of the humerus and the plantar view of the tarsometatarsus are the most useful for identification purposes, suggesting that there is proportionately more diagnostic information available in these views. This result may be useful in a taxonomic sense, in that it potentially indicates that characters that can be coded for cladistic analysis are visible in this view. While extraction of ordinal data from these elements for numerical phylogenetic analysis is clearly problematic, the fact remains that this material is all that is available. Note that, apart from simple presence/absence observations, much of the ordinal character information routinely used in cladistic analyses involves subjective coding of continuously changing variables into discrete states (for detailed discussion see MacLeod, 2001). As an example within the Spheniscidae, the coding of a multistate character such as 'nuchal crest absent (0), poorly developed (1), well developed (2)' is only possible with reference to other taxa that bear nuchal crests, and through preconceived ideas as to where on the continuous scale of crest development a given taxon lies. Coding of such a character will thus depend on the experience and viewpoint of the observer, and is inherently affected by subjectivity.

We believe that studies employing morphometric analysis and similar approaches to the DAISY ANN can identify and refine aspects of morphology for phylogenetic analysis. Indeed, such an approach for these elements is perhaps the only way to extract the detailed phylogenetic information necessary to make a long overdue review of fossil and extant Sphenisciformes possible.

Regardless, the results of this preliminary study show that the DAISY ANN clearly has potential as a powerful tool for taxonomic problems, and we expect ANN technology will play an important role in 21st century taxonomy.

ACKNOWLEDGEMENTS

We thank J. Cooper and R. Prys-Jones (Natural History Museum, Tring, UK) for access to osteological material and useful discussion, and U. Göhlich (Natural History Museum, Vienna, Austria) and M. Knoll (University of Portsmouth, UK) for help in improving this manuscript.

REFERENCES

Acosta Hospitaleche, C., Tambussi, C., Donato, M. & Cozzuol, M. 2007. A new Miocene penguin from Patagonia and its phylogenetic relationships. *Acta Paleontologica Polonica*, 52 (2): 299–314.

Bertelli, S. & Giannini, N. P. (2005) A phylogeny of extant penguins (Aves: Sphenisciformes) combining morphology and mitochondrial sequences. *Cladistics*, 21: 209–239.

Davis, L. S. & Renner, M. 2003. *Penguins*.T. & A. D. Poyser, London.

Fordyce, R. E. & Jones, C. M. 1990. Penguin history and new fossil material from New Zealand; pp. 419-446 *In* Davis, L. S. & Darby, J. T. (eds) *Penguin Biology*. Academic Press, San Diego.

Gaston, K. J. &. O'Neill, M. A. 2004. Automated image recognition: why not? *Philosophical Transactions of the Royal Society of London, Series B* 359: 655–667.

Göhlich, U. B. 2007. The oldest fossil record of the extant penguin genus Spheniscus-a new species from the Miocene of Peru. *Acta Paleontologica Polonica*, 52 (2): 285–298.

Huxley, T. H. 1859. On a fossil bird and a fossil cetacean from New Zealand. *Quarterly Journal of the Geological Society of London*, 15: 670-677.

Livezey, B. C. 1989. Morphometric patterns in Recent and fossil penguins (Aves, Sphenisciformes). *Journal of the Zoological Society of London*, 219: 269-307.

Macleod, N. M. 1999. Generalizing and extending the eigenshape method of shape space visualisation and analysis. *Paleobiology* 25 (1): 107-138.

Macleod, N. M. 2001. Landmarks, localization and the use of morphometrics in phylogenetic analysis. pp. 197-233 *In* Edgecome, G., Adrian, J. & Lieberman, B. (eds) *Fossils, phylogeny and form: an analytical approach*. Kluver Academic, New York.

MacLeod, N., O'Neill, M. A. & Walsh, S. A. 2007. Comparison between morphometric and artificial neural net approaches to the automated species-recognition problem in systematics. pp. 37-62 *In* Curry, G. B. and Humphries, C. J. (eds) *Biodiversity Databases: Techniques, Politics and Applications.* Systematics Association Special Volume Series 73. CRC Press, London.

Myrcha, A., Jadwiszczak, P., Tambussi, C. P., Noriega, J. I., Gaździcki, A., Tatur, A. & del Valle, R. 2002. Taxonomic revision of Eocene Antarctic penguins based on tarsometatarsal morphology. *Polish Polar Research*, 23: 5-46.

Simpson, G. G. 1946. Fossil penguins. *Bulletin of the America Museum of Natural History*, 87: 1-99.

Slack, K. E., Jones, C. M., Ando, T., Harrison, G. L., Fordyce, R. E., Arnason, U. & Penny, D. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Molecular Biology and Evolution*, 23: 1144-1155

Stucchi, M. 2002. Una nueva especie de *Spheniscus* (Aves: Spheniscidae) de la formación Pisco, Perú. *Boletín de la Sociedad Geológica del Perú*, 94: 17-24.

Stucchi, M., Urbina, M. & Giraldo, A. 2003. Un nueva especie de Spheniscidae del Mioceno tardío de la formación Pisco, Perú. *Bulletin de l'Institut Français*

d'Études Andines, 32 (2): 361-375.
Walsh, S. A. & Hume, J. P. 2001. A new Neogene marine avian assemblage from north-central Chile. *Journal of Vertebrate Paleontology*, 21: 484-491.
Walsh, S. A. & Suarez, M. 2006. New penguin remains from the Pliocene of northern Chile. *Historical Biology*, 18: 115-126.
Zusi, R. L. 1975. An interpretation of skull structure in penguins. pp. 59-84 *In* Stonehouse, B. (ed.) *The biology of penguins*, Macmillan, London.