



## **Outline shape analysis of penguin humeri: a robust approach to taxonomic classification**

Piotr Jadwiszczak

Faculty of Biology, University of Białystok, Białystok, Poland

### **Аннотация**

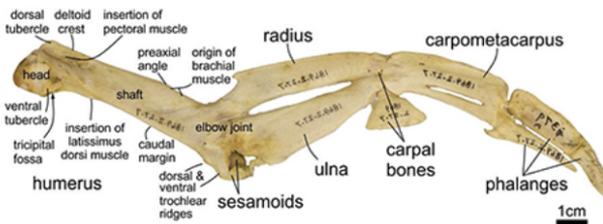
Humeri have been useful bones in taxonomic determinations of extinct penguins. In the context of neontological taxonomic studies, however, their potential remains unsatisfactorily explored. Here, the variation of the overall closed-outline shape of 60 humeri, assignable to five genera of extant penguins, was investigated. A set of normalized outlines was quantified via elliptical Fourier analysis and subjected to linear discriminant analysis on principal component scores extracted from harmonic coefficients. These geometric representations proved to be a source of easily extractable genus-level taxonomic information. The constructed model provided meaningful discrimination between taxa: the first two linear discriminants captured almost 90% of between-group variance. A cross-validation method based on jackknifing yielded 93% correct identifications, and statistically significant differences between group centroids were also detected (multivariate analysis of variance,  $p < 0.05$ ). Predictions of genus membership for the intentionally noisy test data (20 outlines) were accurate in 80% of cases.

**Ключевые слова:** multivariate ordination, Extant Sphenisciformes, genus-level systematics, elliptical Fourier harmonics



## Introduction

Penguins (Aves, Sphenisciformes) are flightless seabirds from the Southern Hemisphere and are arguably the most specialized wing-propelled divers (see Storer 1960). Distally to the shoulder articulation, their pectoral limbs show considerably restricted joint mobility, and relevant musculature is reduced as well (Raikow et al. 1988; Louw 1992). Additionally, all the wing bones are flattened dorsoventrally (Fig. 1; Storer 1960; Stephan 1979; Livezey 1989). In the earliest (Paleocene) penguins, this shape was pronounced primarily in the humerus or upper arm bone (Slack et al. 2006, fig. 1; Ksepka & Ando 2011; Mayr et al. 2020, figs. 1, 3 and 4). This evolutionary trend continued and as early as the Eocene it encompassed the remaining part of the forelimb skeleton (e.g., Jadwiszczak 2010, fig. 2; Jadwiszczak 2012, fig. 2).



**Fig. 1** Wing skeleton of *Spheniscus magellanicus* (mirrored right pectoral limb, ventral view); specimen NHM(T)UK 1869.2.27.7 from the Natural History Museum at Tring (UK).

Recent penguins are characterized by relative consistency in intraspecific proportions, showing low intraspecific variation, although some interspecific differences appear to be statistically significant (Livezey 1989). Considering individual wing bones, the largest of them—the humeri—proved to be very useful specimens for taxonomic determinations, mostly in palaeontology (Simpson 1946, 1971; Livezey 1989; Acosta Hospitaleche et al. 2006, 2019; Chávez Hoffmeister 2014; Thomas et al. 2019; and references therein). Shufeldt (1901: 401), a prolific student of avian osteology, claimed that the modern penguin humeri “apart from size, are characteristically alike in all the species.” This is an obvious oversimplification, and even automated image recognition (involving an artificial neural-net system) can be quite effectively utilized to identify most of the modern Sphenisciformes at the genus level (Walsh et al. 2008). This approach requires obtaining photographs of consistently illuminated bones and can be also applied to tarsometatarsi. The procedure concentrates on patterning internal to



the specimen, not its outline shape. Outline shape may also constitute a source of easily extractable taxonomic information.

In the context of a two-dimensional projection of any three-dimensional object, shape is defined by its outline: "a curve around the perimeter of an object" (Zelditch et al. 2004: 420). Thus, the broad definition of shape can be expressed in terms of curvature, but the raw geometric information contents usually ought to be preprocessed by rooting out location, scale and rotational effects (Kendall 1977; Zelditch et al. 2004; Dryden & Mardia 2016). Within geometric morphometrics, curvature cannot always be captured effectively by a set of landmarks (e.g., Zelditch et al. 2004). Another approach—a curve-fitting method—appears to be more natural and mathematically appealing. This can be conveniently achieved by means of Fourier analysis, particularly elliptical Fourier analysis (Giardina & Kuhl 1977; Kuhl & Giardina 1982). Such an approach has so far been very rarely used in analyses involving penguin bones (e.g., Acosta Hospitaleche & Di Carlo 2010; Jadwyszczak & Mörs 2019), and even more rarely for studying their humeri (e.g., a tightly focused study reported by Acosta Hospitaleche et al. [2013]).

The objective of this article is to investigate the overall closed-outline shape of the latter-day penguin humeri for the presence of a genus-level taxonomic signal, using a series of thoroughly calibrated procedures. The sequence encompasses elliptical Fourier analysis supplemented by unsupervised and supervised multivariate ordination methods with validation as well as significance testing.

## Methods

The source material (a training set in terms of the machine learning approach) used for building a model comprised 60 humeri, represented by standardized ventral-view photographs. The orientation terms used in descriptions of humeri in this article follow those adopted by, among others, Schreiweis (1982), Jadwyszczak (2012) and Chávez Hoffmeister (2014). Bones were assignable to five extant genera of Sphenisciformes: *Aptenodytes*, *Eudyptes*, *Eudyptula*, *Pygoscelis* and *Spheniscus*. Only the rarest modern genus of penguins, a currently monotypic *Megadyptes*, was not included. This was because of the lack of available specimens. All the pictures were taken, in a consistent way, by the author using a Nikon D5100 camera and a Nikkor AF-S DX Micro 40-mm lens. The original specimens are permanently deposited at the Natural History Museum at Tring, UK



and the Professor Andrzej Myrcha University Center of Nature, University of Białystok, Poland (a set accompanying the collection of fossil penguins, curated by the author). Supplementary Tables S1 and S2 comprise details of the specimens used in this work (also as a testing set).

The photographs of humeri were converted to solid black silhouettes against a white background (mirrored if necessary), saved as jpg files with a resolution of 300 dots per inch and gathered in a single directory. The silhouette extraction, geometric morphological analyses of closed outlines of shapes and subsequent ordination/discrimination, statistical and validation procedures were conducted using Momocs (Bonhomme et al. 2014), an R (R Core Team 2019) package for modern two-dimensional morphometrics. Its latest version (1.3.0) is available from GitHub (<https://momx.github.io/Momocs/>).

Outline coordinates were extracted using the `import_jpg` function. The resulting list was subsequently combined with a character-valued vector (genus-level assignments of outlines) encoded as a factor. This newly created Out-class object (Momocs terminology) was visually inspected by the available `panel` method. In the next step, the outlines were normalized using the `coo_center/coo_scale/coo_alignxax/coo_slidedirection("right")/coo_untiltx` sequence of functions. The set was once again visually inspected (`stack` method) for the overall quality as well as for the homology of the first point of the outline (Supplementary Fig. S1).

The comprehensive quantification of outlines was achieved via elliptical Fourier analysis (Giardina & Kuhl 1977; Kuhl & Giardina 1982). The essence of this approach to curve fitting is decomposing the closed outline into a harmonic series, with each harmonic characterized by four coefficients. Such descriptors can be used in further analyses. In the analysis reported here, harmonic power calibration was performed using `calibrate_harmonicpower_efourier` and `calibrate_reconstructions_efourier` functions. This step allowed determination of the appropriate number of harmonics needed to achieve 99.9% of the total harmonic power. Next, Fourier analysis was performed using the `efourier` function. This step was followed by PCA (using the `PCA` function; `plot_PCA` for visualization) to ensure both the dimensionality reduction and, more importantly, orthogonality. The number of components to retain (an initial selection) was determined via `scree_plot`, `scree_min` and `bstick` functions, the last one from the `vegan` package (Oksanen et



al. 2019). In order to calculate, and plot, shape variation along PC axes, the `Pccontrib` function was used.

The discriminant model was built via LDA (using the `LDA` function; `plot_LDA` and `MSHAPES + plot` for visualizations). However, the ultimate number of PCs used as input was based on a simulation that indicated the highest attainable accuracy at the genus-level classification, devoid of the risk of entering flattening and instability zones (suggesting possible overfitting). The accuracy was measured using the `CV.correct` component of the LDA-function output object. The final leave-one-out cross-validation result and other classification metrics were obtained using the `classification_metrics` function.

Additional verification was performed using a testing set of 20 outlines, not used during building the model, extracted from photographs taken by the author and other available images (enumerated in Supplementary Table S2). This was intentionally an eclectic, noisy collection, comprising standardized and non-standardized pictures, images representing extant and extinct members of five analysed genera as well as *Megadyptes waitaha* (an extinct member of the sixth extant genus). The testing set was subjected to the procedure described above; however, ordination analyses were performed using other specialized functions: `rePCA` and `reLDA`. They dealt with new data in the context of the previously built model.

In order to investigate the statistical significance of the centroid/generic differences, MANOVA was performed using the `MANOVA` and `PW_MANOVA` functions.

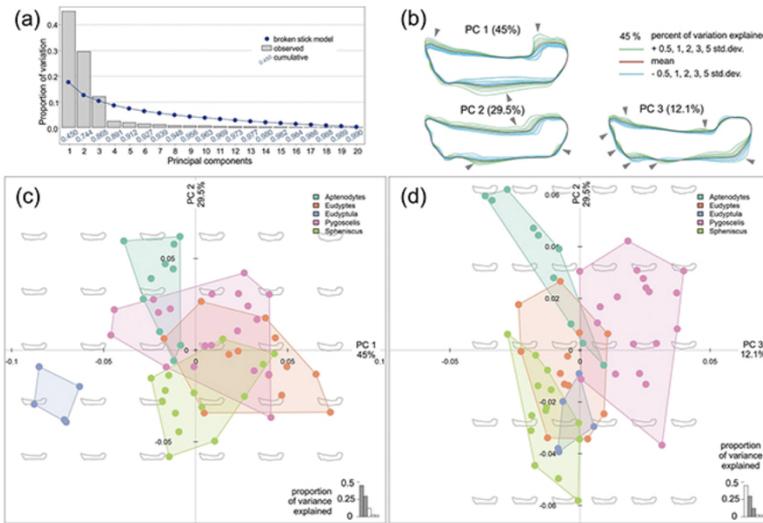
## Results and discussion

The harmonic power calibration revealed that 21 four-coefficient harmonics for each uploaded and normalized outline (made of several thousand x-y coordinates) were required to achieve 99.9% of the total power. Such a range was used in the elliptical Fourier analysis because, with a slight loss of power, a very substantial reduction in dimensionality was gained. Reconstructed shapes, compared visually with the maximal fit, confirmed the benefits of the trade-off (Supplementary Fig. S2).

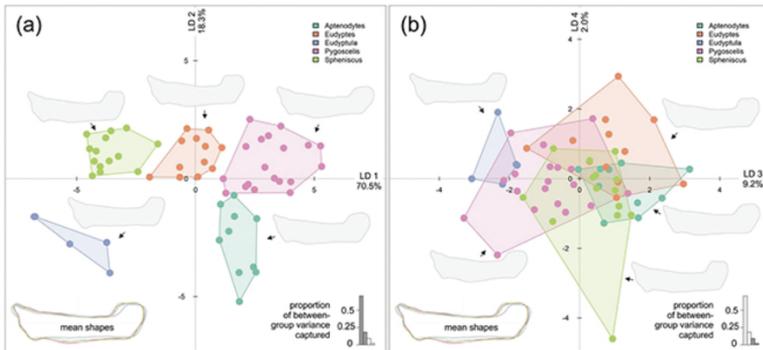
PCA performed on the matrices of harmonic coefficients (values in Supplementary Table S3) demonstrated that the cumulative proportion of variance explained by the first three (out of 60) orthogonal PCs exceeded 0.8 (Fig. 2a).



The first PC captured the distinction between outlines based on the caudal expansion of the humeral head (especially the ventral tubercle) strongly associated with the overall convexity of the cranial margin and (to a lesser extent) caudal expansion of the ventral trochlear ridge (for relevant visualizations, see Fig. 2b; score plots are presented in Fig. 2c, d). The geometric trend, expressed along the second PC axis, was a function of the degree of the overall expansion of the caudal bone margin (in the case of the rim of the tricipital fossa profile, at the cost of the shaft length) and development of the cranial/pre-axial angularity. These trends were inversely related to distinctness of the proximal notch (located just proximal to the dorsal tubercle). The third PC captured disparities primarily in the humeral-head outline rotation relative to the main humeral axis, which can be quantified in terms of the neck-shaft angle. Moreover, the larger the angle was, the more prominent (caudally) was the latissimus dorsi scar and the dorsal trochlear ridge, and the more caudally deflected was the distal third of the bone.



**Fig. 2** Results of PCA performed on matrices of harmonic coefficients: (a) summary of variance explained and comparison with the broken-stick model; (b) shape variation along the first three PC axes; (c) PCA score plot of the first two PCs; and (d) PCA score plot of the second and third PCs. Parts (c) and (d) were plotted with morphospace.



**Fig. 3** Results of LDA performed on PC scores: (a) LDA score plot of the first two LDs and (b) LDA score plot of the third and fourth LDs. Supplemented with mean shapes.

It should also be noted that, although not used for this purpose, PCA, via a gradient along the first PC axis, separated the Fourier-transformed outlines into two distinct (unequal) groups (Fig. 2c). Those from the smaller set have been known to represent *Eudyptula*, a genus of the smallest extant penguins (e.g., Williams 1995). Of course, the above unsupervised ordination method is label agnostic per se (e.g., Quinn & Keough 2002).

Using the 80% or 0.8 threshold criterion together with the results of comparing the observed distribution of variation explained by ordered PCs with the broken stick model (Fig. 2a) would suggest the legitimacy of limiting the number of PCs used in further analyses to



the first three. However, in order to use them as an input in a taxonomy-supervised analysis, such a number of components, despite meeting the above criteria, would not be enough. LDA, simulated for a different ranges of PCs, revealed that the lowest number of them, ensuring the highest accuracy at the genus-level discrimination and not located within no-growth and instability zones (indicative of overfitting), was six (Supplementary Fig. S3).

This optimal value was utilized to derive the full discriminant model, which resulted in a very meaningful ordination (Fig. 3). The coefficients presented in Table 1 determine the linear combinations of PCs forming the discriminants. In all, four of the latter were needed to capture all the between-group variance. This is what can be expected because for k-groups one needs at most k–1 classifiers to differentiate between them (Izenman 2013), unless the number of variables is lower than k–1, which is not the case here (Quinn & Keough 2002).

**Table 1.** Coefficients of LDs.

	LD1	LD2	LD3	LD4
PC1	29.3304 9	36.346 86	19.7191 72	9.152457 1
PC2	73.8473 0	-25.658 54	11.4719 78	-5.41300 99
PC3	96.2968 0	30.033 84	-59.230 354	-0.32523 07
PC4	43.9422 3	-92.421 67	-19.827 980	103.6260 402
PC5	-133.94 916	45.654 18	-73.320 259	10.51868 01
PC6	-69.714 17	37.005 39	2.75769 9	81.23694 58
Proportion of trace <sup>a</sup>	0.7054	0.1826	0.0920	0.0200



**Table 1.** Coefficients of LDs.

	LD1	LD2	LD3	LD4
<sup>a</sup> Between-group variance.				

The first two LD axes captured as much as 89% of between-group variance. The first discriminant represented the axis along which the Eudyptula–Spheniscus group was clearly separated from the Aptenodytes–Pygoscelis cluster. Eudyptes was located between these subsets, slightly overlapping with both Aptenodytes and Spheniscus. The gradient along the second axis testified to the perfect discrimination between Aptenodytes–Eudyptula and the remaining genera (Fig 3a). The third discriminant fully separated Eudyptula from Aptenodytes, Eudyptes and Spheniscus. The ordering along the fourth discriminant axis did not reveal any conspicuous genus-level groups (Fig. 3b).

The leave-one-out/jackknifed cross-validation yielded an impressive overall classification accuracy of 0.933 (or 56/60), and accuracies were also high at the class level (consult diagonal values in Supplementary Fig. S4). The kappa metric of 0.913 meant that the reported classifier (Table 1) was much superior to a random-chance classifier.

MANOVA performed on PC scores (six components were retained) revealed statistical significance of differences between (at least some) group/genus centroids (Pillai's trace = 2.4089, approximated  $F_{24, 212} = 13.371$ ,  $p < 2.2 \times 10^{-16}$ ). This result ought to be considered in terms of the first LD function (Quinn & Keough 2002). The results of such analyses, repeated in a post hoc (pairwise) setting, testified to the statistical significance of differences between all compared pairs, also after Holm's correction for multiple simultaneous comparisons (Supplementary Table S4).

Importantly, the observed pattern of outline-shape discrimination is in line with the phylogeny of present-day penguins. To be more precise, the trichotomy Eudyptula–Spheniscus, Eudyptes and Aptenodytes–Pygoscelis (within a gradient along the first LD axis; Fig. 3a) also reflects the degree of evolutionary affinity between these groups (see Gavryushkina et al. 2017, fig. 3). The contrast between Aptenodytes–Eudyptula and other genera (along the second axis) coincides in turn with the dichotomy between extremely large/small versus medium-sized body dimensions (Williams 1995; Jadwiszczak & Mörs 2019). A discriminant analysis conducted by Livezey (1989),



referred to as canonical analysis, based on 42 skeletal variables—linear measurements from skeletons of virtually all extant penguin species—also resulted in a clear-cut separation of genera (except for *Megadyptes*). However, the first axial gradient (canonical variate 1) did not separate more distantly related genera as nicely as the analysis presented in the current work. *Eudyptes* overlapped considerably with *Spheniscus* and *Pygoscelis*, and the relevant discrimination was attained along the second axis (canonical variate 2). On the other hand, an ordering in accordance with body size, reflected along the first axis (Livezey 1989), appeared to be more natural than that reported here—the former ranged from *Eudyptula* to *Aptenodytes*, with medium-sized penguins caught between these extrema.

At the genus level, the reliability of the humerus-based taxa identification method described in the current work outperforms that involving automated image recognition and artificial neural networks (Walsh et al. 2008). The overall accuracy of classification (calculated via cross-validation) amounted to 47% (“cranial” view) in the experiment by Walsh et al. (2008), whereas that reported here exceeded 90%, also for each studied genus separately (Supplementary Fig. S4). In both studies, *Eudyptula* achieved the highest score (75 and 100% of correct identifications, respectively).

Three-dimensional landmark-based analyses recently conducted by Thomas et al. (2019) for all six modern genera yielded an unweighted genus-level average of 71.3% correct classifications (74.6% without *Megadyptes*). The relevant percentage ranged from 0% (*Spheniscus*) to 100% (*Eudyptula*). In addition to the latter taxon, *Aptenodytes* and *Eudyptes* also exceeded the 90% threshold. In contrast to these results, correct classification scores for *Pygoscelis* and *Spheniscus* exceeded 90% in the current study. The only scores plot presented by Thomas et al. (2019, fig. 5), although based on PCs (an unsupervised ordination), resembles in its message that by Livezey (1989, fig. 9) and slightly more distantly Fig. 3a in this study (both resulting from a supervised method).

Considering predictions of genus membership for the test data (Supplementary Table S2), only four (21.1%) out of 19 specimens (representing genera from the training set) were misclassified (Supplementary Table S5). Importantly, all outline shapes for “problematic” specimens were extracted from a set of photographs published in external sources (Supplementary Table S2), hence not taken in a standardized way. Three fossil specimens were classified correctly (Supplementary Table S5). The posterior probabilities



calculated for an additional (20th) humerus (Supplementary Table S5), a specimen assignable to *Megadyptes*—the only extant genus not represented in the training set—strongly suggested its affinity with *Eudyptes*. This result is in line with ordination spaces obtained by Livezey (1989) and Thomas et al. (2019), where both genera overlapped. Moreover, these taxa are actually closely related (e.g., Gavryushkina et al. 2017).

## Conclusions

A strong genus-level taxonomic correspondence is present in overall outline shapes of present-day penguin humeri. Elliptical Fourier transformation followed by LDA on PC scores, extracted from harmonic coefficients, proved to be a promising tool for addressing problems in classifying skeletons of extant Sphenisciformes. Its usefulness in penguin palaeontology is yet to be assessed.

## Acknowledgements

The author wants to thank two anonymous reviewers for their valuable comments that helped improve this article.

## References

1. Acosta Hospitaleche C., Degrange F.J., Tambussi C.P., Corrado N. & Rustán J.J. 2006. Evaluación de los caracteres del húmero de los pingüinos actuales y fósiles para su uso con fines sistemáticos. (Evaluating the characteristics of humeri of current and fossil penguins for systematic purposes.) *Ornitología Neotropical* 17, 81-94.
2. Acosta Hospitaleche C. & Di Carlo U. 2010. The coracoids in functional and morphological studies of penguins (Aves, Spheniscidae) of the Eocene of Antarctica. *Rivista Italiana di Paleontologia e Stratigrafia* 116, 23-34, doi: [10.13130/2039-4942/5938](https://doi.org/10.13130/2039-4942/5938).
3. Acosta Hospitaleche C., Griffin M., Asensio M., Cione A.L. & Tambussi C. 2013. Middle Cenozoic penguin remains from the Patagonian Cordillera. *Andean Geology* 42, 490-503, doi: [10.5027/andgeoV42n1-a0n](https://doi.org/10.5027/andgeoV42n1-a0n).



4. Acosta Hospitaleche C., Jadwyszczak P., Clarke J.A. & Cenizo M. 2019. The fossil record of birds from the James Ross Basin, West Antarctica. *Advances in Polar Science* 30, 251–273, doi: [10.13679/j.advps.2019.0014](https://doi.org/10.13679/j.advps.2019.0014).
5. Bonhomme V., Picq S., Gaucherel C. & Claude J. 2014. Momocs: outline analysis using R. *Journal of Statistical Software* 56, 1–24, doi: [10.18637/jss.v056.i13](https://doi.org/10.18637/jss.v056.i13).
6. Chávez Hoffmeister M. 2014. Phylogenetic characters in the humerus and tarsometatarsus of penguins. *Polish Polar Research* 35, 469–496, doi: [10.2478/popore-2014-0025](https://doi.org/10.2478/popore-2014-0025).
7. Dryden I.L. & Mardia K.V. 2016. *Statistical shape analysis*. Chichester: Wiley.
8. Gavryushkina A., Heath T.A., Ksepka D.T., Stadler T., Welch D. & Drummond A.J. 2017. Bayesian total evidence dating reveals the recent crown radiation of penguins. *Systematic Biology* 66, 57–73, doi: [10.1093/sysbio/syw060](https://doi.org/10.1093/sysbio/syw060).
9. Giardina C.R. & Kuhl F.P. 1977. Accuracy of curve approximation by harmonically related vectors with elliptical loci. *Computer Graphics and Image Processing* 6, 277–285, doi: [10.1016/S0146-664X\(77\)80029-4](https://doi.org/10.1016/S0146-664X(77)80029-4).
10. Izenman A.J. 2013 *Modern multivariate statistical techniques: regression, classification, and manifold learning*. New York, NY: Springer, doi: [10.1007/978-0-387-78189-1\\_8](https://doi.org/10.1007/978-0-387-78189-1_8).
11. Jadwyszczak P. 2010. New data on the appendicular skeleton and diversity of Eocene Antarctic penguins. In D. Nowakowski (ed.): *Morphology and systematics of fossil vertebrates*. Pp. 44–50. Wrocław: DN Publishers.
12. Jadwyszczak P. 2012. Partial limb skeleton of a “giant penguin” *Anthropornis* from the Eocene of Antarctic Peninsula. *Polish Polar Research* 33, 259–274, doi: [10.2478/v10183-012-0017-0](https://doi.org/10.2478/v10183-012-0017-0).
13. Jadwyszczak P. & Mörs T. 2019. First partial skeleton of *Delphinornis larseni* Wiman, 1905, a slender-footed penguin from the Eocene of Antarctic Peninsula. *Palaeontologia Electronica* 22.2 .32A, 1–31, doi: [10.26879/933](https://doi.org/10.26879/933).
14. Kendall D.G. 1977. The diffusion of shape. *Advances in Applied Probability* 9, 428–430, doi: [10.2307/1426091](https://doi.org/10.2307/1426091).
15. Ksepka D.T. & Ando T. 2011. Penguins past, present, and future: trends in the evolution of the Sphenisciformes. In G.



- Dyke & G. Kaiser (eds.): Living dinosaurs: the evolutionary history of modern birds. Pp. 155–186. Chichester: Wiley.
16. Kuhl F. & Giardina C. 1982. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* 18, 236–258, doi: [10.1016/0146-664X\(82\)90034-X](https://doi.org/10.1016/0146-664X(82)90034-X).
  17. Livezey B.C. 1989. Morphometric patterns in recent and fossil penguins (Aves, Sphenisciformes). *Journal of Zoology* 219, 269–307, doi: [10.1111/j.1469-7998.1989.tb02582.x](https://doi.org/10.1111/j.1469-7998.1989.tb02582.x).
  18. Louw G.J. 1992. Functional anatomy of the penguin flipper. *Journal of the South African Veterinary Association* 63, 113–120.
  19. Mayr G., dePietri V.L., Love L., Mannering A.A., Bevitt J.J. & Scofield R.P. 2020. First complete wing of a stem group sphenisciform from the Paleocene of New Zealand sheds light on the evolution of the penguin flipper. *Diversity* 12, article no. 46, doi: [10.3390/d12020046](https://doi.org/10.3390/d12020046).
  20. Oksanen J., Guillaume Blanchet F., Friendly M., Kindt R., Legendre P., McGlenn D., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Szoecs E. & Wagner H. 2019. *Vegan: community ecology package*. R package version 2.5-6. Accessed on the internet at <https://CRAN.R-project.org/package=vegan>.
  21. Quinn G.P. & Keough M.J. 2002. *Experimental design and data analysis for biologists*. Cambridge: Cambridge University Press.
  22. Raikow R.J., Bicanovsky L. & Bledsoe A.H. 1988. Forelimb joint mobility and the evolution of wing-propelled diving in birds. *Auk* 105, 446–451, doi: [10.1093/auk/105.3.446](https://doi.org/10.1093/auk/105.3.446).
  23. R Core Team 2019. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
  24. Schreiweis D.O. 1982. A comparative study of the appendicular musculature of penguins (Aves: Sphenisciformes). *Smithsonian Contributions to Zoology* 341, 1–46, doi: [10.5479/si.00810282.341](https://doi.org/10.5479/si.00810282.341).
  25. Simpson G.G. 1946. Fossil penguins. *Bulletin of the American Museum of Natural History* 87, 1–99.
  26. Simpson G.G. 1971. Review of fossil penguins from Seymour Island. *Proceedings of the Royal Society of London B* 178, 357–387, doi: [10.1098/rspb.1971.0070](https://doi.org/10.1098/rspb.1971.0070).



27. Shufeldt R.W. 1901. Osteology of the penguins. *Journal of Anatomy and Physiology* 35, 390–404, doi: <https://doi.org/10.1086/277965>.
28. 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, doi: [10.1093/molbev/msj124](https://doi.org/10.1093/molbev/msj124).
29. Stephan B. 1979. Vergleichende Osteologie der Pinguine. (Comparative osteology of the penguin.) *Mitteilungen aus dem Zoologischen Museum in Berlin* 55, 3–98.
30. Storer R.W. 1960. Evolution in diving birds. *Proceedings of XII International Ornithological Congress* 2, 694–707.
31. Thomas D.B., Ksepka D.T., Holvast E.J., Tennyson A.J.D. & Scofield P. 2019. Re-evaluating New Zealand's endemic Pliocene penguin genus. *New Zealand Journal of Geology and Geophysics*, doi: [10.1080/00288306.2019.1699583](https://doi.org/10.1080/00288306.2019.1699583).
32. Walsh S.A., MacLeod N. & O'Neill M. 2008. Analysis of spheniscid humerus and tarsometatarsus morphological variability using DAISY automated image recognition. *Oryctos* 7, 129–136.
33. Williams T.D. 1995. *The penguins*. Oxford: Oxford University Press.
34. Zelditch M.L., Swiderski D.L., Sheets H.D. & Fink W.L. 2004. *Geometric morphometrics for biologists. A primer*. Burlington: Elsevier Academic Press.