

https://dspace.org/

Probabilistic sigma taxonomy of Darwins finches Galpagos

John Thackeray

https://hdl.handle.net/10539/41769
WOS

Downloaded from a DSpace repository. https://dspace.org/

Zoomorphology (2024) 143:489–493
https://doi.org/10.1007/s00435-024-00650-x

RESEARCH

Probabilistic sigma taxonomy of Darwin's finches (Galápagos)

Francis Thackeray[1]

Received: 7 January 2024 / Revised: 10 February 2024 / Accepted: 10 February 2024 / Published online: 9 April 2024
© The Author(s) 2024

Abstract
A study is undertaken using anatomical measurements of specimens attributed to six species of *Geospiza*, the ground finches from the Galápagos archipelago. In a demonstration of method, a probabilistic approach associated with "sigma taxonomy" is adopted to assess the probability that pairs of specimens are or are not conspecific. We use a definition of a species based on morphometric analyses of the kind previously undertaken on extant vertebrate taxa (including mammals, birds and reptiles), using pairwise comparisons of anatomical measurements in regression analyses of the form $y = mx + c$ from which the log-transformed standard error of the m-coefficient is calculated ("log sem"). The latter statistic is a reflection of variability in morphology. There is a high probability that at a species level, specimens attributed to *G. magnirostris* are different from those attributed to *G. fulginosa*, *G. difficilis* or *G. scandens*. Results of this study, using probabilistic sigma taxonomy, confirm the refutation of a single species hypothesis. In addition, we apply the log sem method to demonstrate that in case of comparisons between *G. fortis* and *G. scandens* (which are known to hybridise), there is a high probability that they are not different at a species level.

Keywords Darwin's Finches · *Geospiza* · Sigma Taxonomy · Species definition · Galápagos

Introduction

Taxonomy was a vexing question for Charles Darwin, especially with regard to barnacles. He recognised that it was easy to classify such fauna into one or other species when only a few specimens were available, but as sample sizes increased the boundaries between species tended to break down (Thackeray 1999). For the classification of finches which he collected on the Galápagos archipelago, Darwin (1837) turned to John Gould (1837). As an ornithologist, the latter was an alpha taxonomist relying on the Linnaean assumption that specimens can be pigeon-holed into one or other distinct species. Darwin (1845) stated that the finches were "related to each other in the structure of their beaks, short tails, form of body and plumage. There are thirteen species, which Mr. Gould has divided into four subgroups". However, he went on to say "the most curious fact is the perfect gradation in the size of the beaks in the different species of *Geospiza*". Such observations would have raised the question as to whether there were clear boundaries, and whether particular specimens represented "varieties" as opposed to distinct species. Indeed hybridisation is known to occur in Galápagos finches (eg. Enbody et al 2023; Grant and Grant 1996, 2008, 2014; Lack 1947; Lamichhaney et al 2018), as in a great many other taxa (Thackeray and Schrein 2017), such that a definition of a species becomes an important issue.

We adopt here what we have called "sigma taxonomy", where "sigma" is the Greek letter S (Σ) for spectrum, associated with a probabilistic definition of a species applicable in cases where there are not necessarily clear boundaries (Thackeray 2018; Thackeray and Schrein 2017). It is defined as "The classification of taxa in terms of probabilities of conspecificity, without assuming distinct boundaries between species" (Thackeray 2018), as opposed to alpha taxonomy which does make that assumption (Mayr et al 1953).

The statistical method behind sigma taxonomy has been described by Thackeray and Odes (2013) and by Thackeray and Dykes (2016). The fundamental issue is "what is the probability that two specimens do or do not belong to the same species?". As an example based on illustrations (Fig. 1), the method is demonstrated here initially with

✉ Francis Thackeray
mrsples@global.co.za

[1] Evolutionary Studies Institute, University of the Witwatersrand, Private Bag 3, Johannesburg, South Africa

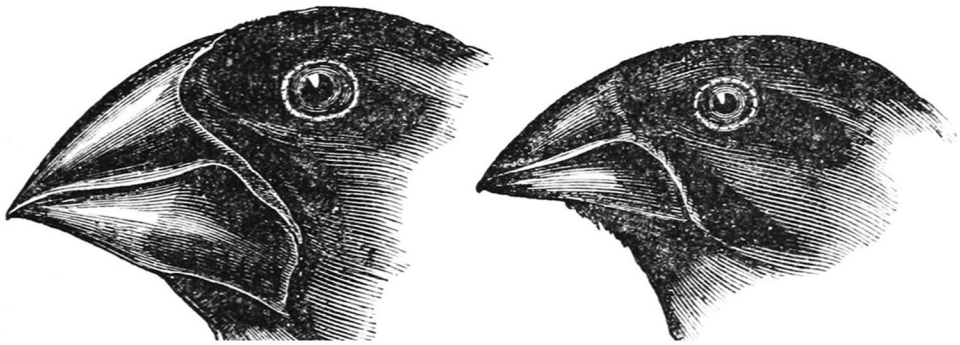


Fig. 1. *Geospiza magnirostris* Gould, 1837 (left) and *G. fortis* Gould, 1837 (right), as published by Darwin (1845)

regard to two finches, *Geospiza magnirostris* and *Geospiza fortis*, classified by Gould (1837) and to this day generally accepted as distinct species. They were illustrated in Darwin's (1845) "*Journal of researches into the natural history and geology of the countries visited during the voyage of H.M.S. Beagle round the world, under the Command of Capt. Fitz Roy*". *G. magnirostris* is larger than *G. fortis*. Both are widespread granivores, and with its relatively large beak the former feeds on large seeds, whereas the latter consumes flowers as well as seeds. Morphological variability of beaks is under genetic control (Abzanhov et al, 2004).

Method, with an example of two "types" of finches

Linear measurements are obtained from anatomical elements as in the case of museum specimens studied by Thackeray et al (1997). Measurements are subjected to pairwise comparisons, using least squares linear regression to quantify the degree of scatter around a regression line of the form $y = mx + c$, where m is the slope and c is the intercept.

In a study of measurements of pairs of specimens of the same (extant) species of many vertebrates, Thackeray et al. (1997) reported central tendency of the log-transformed standard error of the m-coefficient, known as "log sem" which is a measure of the degree of scatter around the regression line, reflecting variability in shape. Central tendency of log sem has also been found using larger samples, including birds, associated with a mean log sem value of -1.61 (Thackeray 2007). The mean log sem value of -1.61 ± 0.1 has been recognized as a typical degree of intraspecific variation in extant species, as described by Thackeray and Dykes (2016). The remarkable consistency of this statistic is reflected by the following sets of data for extant and extinct vertebrate taxa (conspecific pairs):

- − 1.61 (Crania of mammals, birds, reptiles etc.) (Thackeray 2007)
- − 1.61 (Crania: female-female comparisons of *Pan paniscus*) (Gordon and Wood 2013)
- − 1.62 (Crania: male-male comparisons of *Pan paniscus*) (Gordon and Wood 2013)
- − 1.61 (Crania: female-male comparisons of *Pan paniscus*) (Gordon and Wood 2013)
- − 1.62 (Crania: female-female comparisons of *Pan troglodyes*) (Gordon and Wood 2013)
- − 1.60 (Crania: male-male comparisons of *Pan troglodytes*) (Gordon and Wood 2013)
- − 1.60 (Crania: female-male comparisons of *Pan troglodytes*) (Gordon and Wood 2013)
- − 1.61 (Crania: *H. sapiens, P. troglogdytes, P. paniscus, Gorilla gorilla (*Thackeray and Dykes 2016*)*
- − 1.62 (Molars: *H. sapiens, P. troglogdytes, P. paniscus, Gorilla gorilla* (Dykes 2014)
- − 1.61 (Molars: *A. africanus*, *A. afarensis*, *H. habilis*, *H. erectus*, *P. robustus, P. boisei* (Dykes 2014)

When Thackeray and Dykes (2016) confirmed a mean log sem value of − 1.61 from a study restricted to extant hominoids, using data published by Gordon and Wood (2013), the standard deviation was 0.1 (n = 8,072 regressions). The same mean log sem value of -1.61 had been reported by Thackeray (2007) for a greater diversity of fauna (n > 70 species).

A mean log sem value of T = − 1.61 with a standard deviation of 0.1 has been proposed as a probabilistic definition of a species, applicable to anatomical measurements of a diversity of fauna (Thackeray and Dykes 2016). As an example of method, it is used here as a frame of reference reflecting a typical degree of anatomical variation in extant vertebrate species, for purposes of comparison of iconic sketches of two "types" of finches (Fig. 1).

In this case, one reference point is the centre of the eye (point O), while a second landmark is the tip of the beak (E). OE is a reference line for others separated at intervals of 10 degrees. Thus line OD is ten degrees below the reference line OE. Likewise, lines OF, OG and OH are separated at 10 degree intervals relative to each other, clockwise above reference line OE. 20 measurements are obtained, clockwise from OD through to OW. This radial method has previously been used by Braun et al (2004).

Two sets of measurements are obtained, one for *G. magnirostris* and another for *G. fortis.* Two regression lines of the form $y = mx + c$ can be obtained. In the case where

measurements of *G. magnirostris* are on the x-axis and those of *G. fortis* are on the y-axis, we have the following result:

$$y = 0.683x + 7.400 \text{ (log } sem = -1.498$$

In the case where measurements of *G. magnirostris* are on the y-axis and those of *G. fortis* are on the x-axis, we have this result:

$$y = 1.399x - 7.414 \text{ (log } sem = -1.187)$$

The mean log sem for the two regression equations is − 1.342. This is outside the upper 95% confidence limit above the mean log sem value of T = − 1.61 (± 0.1) for reference specimens representing the typical degree of variation in extant vertebrate species (Thackeray and Dykes 2016).

One must go further to assess probability of conspecificity. As discussed by Thackeray and Dykes (2016), the difference between two log sem values in comparisons of conspecific pairs is designated "delta log sem" and is typically equal to 0.03 (the mean delta log sem obtained from more than 8,000 regressions for conspecific pairs). If the difference exceeds this typical value in the case of any pairwise comparison, and if the log sem is outside the upper 95% confidence limit of − 1.61 (± 0.1), this would indicate that the two specimens being compared have a high probability of being different at a species level. In our example here, the difference between the two log sem values (− 1.498 and − 1.187) is 0.31. This greatly exceeds the value of 0.03, and together with the mean log sem value of − 1.342 we may infer that there is a high probability that our specimens attributed to *G. magnirostris* and *G. fortis* represent differences at the level of species in terms of probabilistic sigma taxonomy (Thackeray 2018; Thackeray and Schrein 2017).

To apply the same technique to actual material we use measurements of specimens attributed to six species of *Geospiza*, namely *G. magnirostris*, *G. fortis*, *G. fulginosa*, *G. difficilis*, *G. scandens* and *G. conirostris*. For purposes of demonstration of method, we use lengths of wing, tail, culmen, gonys, depth of bill at base, width of mandible at base, tarsus and middle toe with claw, from a large database curated by the California Academy of Sciences (Swarth 1931; Lack 1945). We selected 240 specimens (20 males and 20 females of each species) from 10 islands of the Galápagos archipelago (Isabela, Santiago, Santa Cruz, Genoseva, Darwin, Wolf, Marchena, Pinta, Fernandina and Pinzon). We recognize that the sample sizes are small, but this study of specimens from only 10 islands serves primarily as a demonstration of method, applied for the first time to Darwin's finches.

We used a computer program for analyzing large data sets, focusing on log sem calculations (Dykes and Dykes 2015).

Results

Log sem and delta log sem values for pairwise comparisons are given in Table 1, based on specimens attributed to each of the six species, as identified in the catalogue of the California Academy of Sciences. In each case, the mean log sem values are circa − 1.90, and the corresponding delta log sem values are circa 0.03 (based on a total of 4,560 regression analyses).

There is a high probability of conspecificity in these comparisons because the mean log sem values of circa − 1.90 are well within the upper 95% confidence limit associated with the probabilistic definition of a species (T = − 1.61 ± 0.1, n > 8,000 regressions for conspecific vertebrate taxa), and also because the mean delta log sem values are circa 0.03, corresponding to the value published by Thackeray and Dykes (2016) for conspecific taxa.

The McKAY-ZINK (MZ) single species hypothesis

McKay and Zink (2015) suggested that there is only a single species (*G. magnirostris*) on the Galápagos archipelago, associated with a high degree of introgression in the context of inter-island travel. This "MZ" hypothesis has since been refuted by Zink and Vázquez-Miranda (2019). Expectations of the MZ hypothesis would have been the following: that

Table 1 Results of pairwise comparisons of measurements of six species of *Geospiza*

Intraspecific pairwise comparisons, specimens attributed to *G. magnirostris*
Mean log sem = − 1.866 ± 0.192 (n = 760 pairwise regressions)
Delta log sem = 0.039. 20 females, 20 males
Intraspecific pairwise comparisons, specimens attributed to *G. fortis*
Mean log sem = − 1.895 ± 0.186 (n = 760 pairwise regressions)
Delta log sem = 0.036. 20 females, 20 males
Intraspecific pairwise comparisons, specimens attributed to *G. fulginosa*
Mean log sem = − 1.949 ± 0.194 (n = 760 pairwise regressions)
Delta log sem = 0.037. 20 females, 20 males
Intraspecific pairwise comparisons, specimens attributed to *G. difficilis*
Mean log sem = − 1.951 ± 0.212 (n = 760 pairwise regressions)
Delta log sem = 0.046. 20 females, 20 males
Intraspecific pairwise comparisons, specimens attributed to *G. scandenss*
Mean log sem = − 1.926 ± 0.163 (n = 760 pairwise regressions)
Delta log sem = 0.023. 20 females, 20 males
Intraspecific pairwise comparisons, specimens attributed to *G. conirostris*
Mean log sem = − 1.831 ± 0.168 (n = 760 pairwise regressions)
Delta log sem = 0.032. 20 females, 20 males

for pairwise comparisons with specimens generally understood to be representative of *G. magnirostris*, mean log sem values would fall within the uppermost 95% confidence limit for what is defined as a species in terms of sigma taxonomy ($T = -1.61 \pm 0.1$), and secondly that delta log sem values would be ≤ 0.03 (Thackeray and Dykes 2016). As indicated below, these conditions do not apply in three instances.

It is not the case in the comparison of samples attributed to *G. magnirostris* and *G. fulginosa* (mean log sem = − 1.263, delta log sem = 0.270, n = 50 regressions); (2) nor is it the case for a comparison of samples attributed to *G. magnirostris* and *G. difficilis* (mean log sem = − 1.305, delta log sem = 0.135; n = 50 regressions); (3) nor is it the case for the comparison of samples attributed to *G. magnirostris* and *G. scandens* (mean log sem = − 1.358, delta log sem = 0.122; n = 50 regressions). In terms of sigma taxonomy there is a high probability that the samples being compared do not represent a single species of *G. magnirostris*.

These results serve to confirm the conclusion reached by Zink and Vázquez-Miranda (2019), refuting the single species hypothesis.

It can be noted that in the case of a comparison between *G. scandens* and *G. fortis*, a mean log sem value of − 1.560 and a mean delta log sem value of only 0.023 are obtained (n = 50 regressions). These results point to the probability that *G. fortis* and *G. scandens* are not different at a species level, using $T = -1.61 \pm 0.1$ as a probabilistic definition of a species. Indeed, a recent paper by Enbody et al (2023) indicates that the two taxa can hybridise, a fact which is not surprising in the context of log sem statistics. Furthermore, when UPGMA was used to generate a phenetic tree for six species of *Geospiza*, based on log sem values (Thackeray 2022), *G. fortis* and *G. scandens* grouped together. Thackeray's phenetic tree corresponds closely to a phylogeny obtained by Burns et al. (2014) and Reaney et al. (2020) based on genetic data.

Conclusions

Our approach serves as one way of addressing Darwin's appeal in 1859 (in the last chapter of *The Origin of Species*) for the "amount" of variation in a species to be assessed. This is indeed necessary in the context of the question as to how many species of *Geospiza* exist on the Galápagos archipelago. In confirmation of the conclusion reached by Zink and Vázquez-Mirana (2019), it has been possible to refute a single species hypothesis on the basis of our quantitative approach based on assessment of the probability that any two specimens are or are not conspecific, using a definition of a species in the case of morphometric analyses (Thackeray and Dykes 2016). At a species level, specimens attributed to *G. magnirostris* have a high probability of being different from those attributed to *G. fulginosa*, *G. difficilis* and *G. scandens*. By contrast, our results point to the probability that *G. fortis* and *G. scandens* are not different at a species level, using $T = -1.61 \pm 0.1$ as a probabilistic definition of a species.

In a demonstration of method, our analysis indicates the potential of the log sem approach, combined with delta log sem values in the context of sigma taxonomy and a probabilistic definition of a species (Thackeray and Dykes 2016).

Acknowledgements Anatomical measurements used in this study are from collections of *Geospiza* in the California Academy of the Sciences, and we are grateful to Moe Flannery for permission to make use of them. Ashley Reaney, Peter and Rosemary Grant, Arkhat Abzanhov, Jack Dumbacher, Bob Zink, Tim Crowe, Marine Cazenave, Caitlin Schrein, Peter Knox-Shaw and two anonymous referees are thanked. Richard Dykes kindly assisted with statistical analyses. I dedicate this paper to the late Sue Dykes who encouraged me to undertake "log sem" analyses of anatomical measurements of ground finches from the Galápagos archipelago, in relation to a probabilistic definition of a species.

Author contributions J.F.T. is the sole author.

Funding Open access funding provided by University of the Witwatersrand. University of the Witwatersrand.

Data availability Anatomical measurements used in this study are from collections of *Geospiza* in the California Academy of the Sciences.

Declarations

Conflict of interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ (2004) Bmp4 and morphological variation of beaks in Darwin's finches. Science 305:1462–1465

Braun S, Thackeray JF, Loots M (2004) A morphometric technique to assess probabilities of conspecificity in extant primates and Plio-Pleistocene hominids. Ann Transvaal Museum 41:93–95

Burns KJ, Shultz AJ, Title PO, Mason NA, Barker FK, Klicka J et al (2014) Phylogenetics and diversification of tanagers (Passeriformes: Thraupidae), the largest radiation of Neotropical songbirds. Mol Phylogenet Evol 75:41–77. https://doi.org/10.1016/j.ympev.2014.02.006

Darwin C (1837) Remarks upon the habits of the genera *Geospiza, Camarhynchus, Cactornis,* and *Certhidea* of Gould. Proc Zool Socety Lond 5:49

Darwin C (1845) Journal of researches into the natural history and geology of the countries visited during the voyage of H.M.S. In: Fitz RN (ed) Beagle round the world, under the Command of Capt, 2nd edn. John Murray, London

Dykes SJ (2014) A morphometric analysis of hominin teeth attributed to different species of *Australopithecus*, *Paranthropus* and *Homo*. University of the Witwatersrand, Johannesburg (Masters dissertation)

Dykes SJ, Dykes RD (2015) 'Professor Regressor': a computer programme for rapid processing of large sets of data for pairwise regression analyses in palaeontological contexts. Palaeontol Afr 49:53–59

Enbody ED, Sendell-Price AT, Sprehn CG, Rubin CJ, Visscher PM, Grant BR, Grant PR, Andersson L (2023) Community-wide genome sequencing reveals 30 years of Darwin's finch evolution. Science 381:6665. https://doi.org/10.1126/science.adf6218

Gordon AD, Wood BA (2013) Evaluating the use of pairwise dissinilarity metrics in paleoanthropology. J Hum Evol 65:465–477. https://doi.org/10.1016/j.jhevol.2013.08.002

Gould J (1837) Remarks on a group of Ground Finches from Mr Darwin's collection, with characters of the new species. Proc Zool Soc Lond 5:4–7

Grant BR, Grant PR (1996) High survival of Darwin's Finch hybrids: effects of beak morphology and diets. Ecology 77(3):500–509. https://doi.org/10.2307/2265625

Grant PR, Grant BR (2008) How and Why Species Multiply: The Radiation of Darwin's Finches. Princeton University Press, Princeton

Grant PR, Grant BR (2014) 40 years of evolution: Darwin's finches on Daphne Major Island. Princeton University Press

Lack D (1945) The Galapagos finches (Geospizinae). A study in variation. Occas Pap Calif Acad Sci 21:1–58

Lack D (1947) Darwin's Finches. Cambridge University Press, New York

Lamichhaney S, Han F, Webster MT, Andersson L, Rosemary Grant B, Grant PR (2018) Rapid hybrid speciation in Darwin's finches. Science 359:224–228. https://doi.org/10.1126/science.aao4593

Mayr E, Linsley EG, Usinger RL (1953) Methods and Principles of Systematic Zoology. McGraw-Hill Book Co, New York

McKay BD, Zink RM (2015) Sisyphean evolution in Darwin's finches. Biol Rev 90:689–698. https://doi.org/10.1111/brv.12127

Reaney AM, Bouchenak-Khelladi Y, Tobias JA, Abzhanov A (2020) Ecological and morphological determinants of evolutionary diversification in Darwin's finches and their relatives. Ecol Evol 10(24):14020–14032. https://doi.org/10.1002/ece3.699411

Swarth HS (1931) The avifauna of the Galápagos Island. Occas Pap Pf Calif Acad Sci 18:1–299

Thackeray JF (1999) Alternative views on hominid diversity in the context of Darwin's assessment of barnacles. In: Anderson JM (ed) Towards Gondwana alive Promoting biodiversity and stemming the sixth extinction. National Botanical Institute Gondwana Alive Society, Pretoria, pp 100–109

Thackeray JF (2007) Approximation of a biological species constant? S Afr J Sci 103:489. http://www.scielo.org.za/pdf/sajs/v103n11-12/a0210312.pdf

Thackeray JF (2018) Alpha and sigma taxonomy of *Pan* (chimpanzees) and Plio-Pleistocene hominin species. S Afr J Sci 114(11/12):1–2. https://sajs.co.za/article/view/5823

Thackeray JF (2022) Morphometric ('log sem') analysis of anatomical measurements of Galápagos finches (Geospiza), chimpanzees (Pan) and Plio-Pleistocene hominins (Paranthropus, Australopithecus and early Homo). S Afr J Sci. https://doi.org/10.17159/sajs.2022/11913

Thackeray JF, Odes E (2013) Morphometric analysis of early Pleistocene African hominin crania in the context of a statistical (probabilistic) definition of a species. Antiquity 87. http://antiquity.ac.uk/projgall/thackeray335/

Thackeray JF, Dykes S (2016) Morphometric analyses of hominoid crania, probabilities of conspecificity and an approximation of a biological species constant. Homo J Comp Human Biol. https://doi.org/10.1016/j.jchb.2015.09.003

Thackeray JF, Schrein CM (2017) A probabilistic definition of a species, fuzzy boundaries and 'sigma taxonomy.' South African J Sci 113(5/6):1–2. https://doi.org/10.17159/sajs.2017/a0206

Thackeray JF, Bellamy CL, Bellars D, Bronner G, Bronner L, Chimimba C, Fourie H, Kemp A, Krüger M, Plug I, Prinsloo S, Toms R, Van Zyl AJ, Whiting MJ (1997) Probabilities of conspecificity: application of a morphometric technique to modern taxa and fossil specimens attributed to *Australopithecus* and *Homo*. S Afr J Sci 93:195–196

Zink RM, Vázquez-Miranda H (2019) Species limits and phylogenomic relationships of Darwin's Finches remain unresolved: Potential consequences of a volatile ecological setting. Syst Biol 68:347–357. https://doi.org/10.1093/sysbio/syy073

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.