AUTHORS:
J. Francis Thackeray[1]
Caitlin M. Schrein[2]

AFFILIATIONS:
[1]Evolutionary Studies Institute, School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa

[2]Independent Researcher, Washington, DC, USA

CORRESPONDENCE TO:
Francis Thackeray

EMAIL:
francis.thackeray@wits.ac.za

A probabilistic definition of a species, fuzzy boundaries and 'sigma taxonomy'

Morphometric and genetic analyses of a variety of living and fossil taxa[1-7] support the use of a probabilistic definition of a species in the context of 'sigma taxonomy' (where sigma represents 'S' for spectrum[6,7]), in contrast to alpha taxonomy[8], for which boundaries discriminating species are presumed to be distinct, thus accommodating only rigid, 'either-or' classification schemes.

Recently, integrated taxonomic approaches, involving morphology and genetics, have demonstrated that traditional definitions of species boundaries may require re-evaluation and revision. Integrated analyses of gibbons[9] and giraffes[10], for example, have narrowed boundaries and led researchers to recognise more species than were previously identified. Species identification is complicated in part by the potential for some populations to hybridise and, in the case of living wolf populations[11], genetic analyses have widened boundaries and revealed that there are fewer species – or more hybrids – than previously thought. Analyses of ancient DNA have also exposed hybrids of the past: there is now evidence that populations of elephants and mammoths likely interbred[12] and, of course, Neanderthal DNA is known to be part of the modern human genome to this day, as a result of introgression of Neanderthals and early modern *Homo sapiens* – a relationship hinted at by morphology[13], now confirmed genetically[14].

De Manuel et al.[4] have recently provided genetic evidence of interbreeding between chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) within the last million years. The two groups diverged sometime between 1 and 2 million years ago[5], and most likely interbred during episodic contraction of forests during relatively dry and cool intervals within the Plio-Pleistocene[15]. This finding is consistent with morphometric analyses of homologous pairs of cranial measurements of specimens of *P. troglodytes* and *P. paniscus* showing there is not a clear boundary between the taxa.[6,16,]

In morphometric analyses of hominoid crania[6], homologous cranial measurements of specimens A and B are compared by least squares linear regression analysis. Two 'log sem' statistics are obtained when specimen A is on the *x*-axis and B is on the *y*-axis, and vice versa, where 'log sem' is the log transformed standard error of the *m*-coefficient in equations of the form $y = mx + c$.[6,17] The mean of these two log sem statistics is called M-log sem, where M relates to the difference between log sem values, termed delta log sem.[6] Delta log sem values are small (circa 0.03) when the two specimens are known conspecifics (e.g. two individuals from the same population), and large ($>> 0.03$) when they are of different species of different size and shape. Delta log sem data are assessed in relation to M-log sem values obtained from pairwise comparisons of cranial measurements in regression analyses. Mean log sem values for modern conspecifics tend to show central tendency around a value of -1.61, which Thackeray[18] hypothesised to be an approximation of a biological species constant ($T = -1.61 \pm 0.2$) through geographical space and evolutionary time, associated with a probabilistic definition of a species.

Using homologous pairs of cranial measurements of *P. troglodytes* and *P. paniscus*, Gordon and Wood[16] confirmed that the mean of two log sem values for conspecific specimens tends to approximate an average M-log sem value of -1.61. Remarkably, this applies to specimens of both *P. troglodytes* and *P. paniscus* (based on adult male and female specimens, $n > 1000$ regression analyses).[16] There is no clear boundary between *P. paniscus* and *P. troglodytes* on the basis of log sem values[6,16], which can now be explained in terms of genetic evidence indicating hybridisation within the Plio-Pleistocene[4].

A recent study by Roux et al.[3] sheds light on 'the grey zone of speciation' in living taxa, from a genomics perspective; based on genetic analyses of more than 61 animals, the authors found that the '"grey zone" of speciation, in which taxonomy is often controversial, spans from 0.5% to 2% of net synonymous divergence…'[3]. This range of values is compatible with an approximation of a biological species constant (T) of -1.61 ± 0.2 and lends support to the concept of 'sigma taxonomy'.[7]

In a recent review article, Llamas et al.[1] stated that 'Admixture…blurs the species limits for extinct groups, especially since the morphological identification of hybrids is difficult'. This 'blurring' of species limits, or 'fuzzy boundaries' as A.R. Wallace put it in 1870[19], reflects the concept of 'palaeo-spectroscopy' in hominin evolution, advocated by Thackeray and Odes[20] who conducted a morphometric analysis of early Pleistocene African hominin crania in the context of a statistical (probabilistic) definition of a species.

We propose that a probabilistic definition of a species may be obtained by recognising the 'grey zone' concept, or 'sigma taxonomy'[7], as opposed to 'alpha taxonomy'[8]. We strongly recommend the adoption of a probabilistic definition of a species which has the potential to be applied to fossil hominins[15,20,21] and other taxa.

Acknowledgements

References

1. Llamas B, Willerslev E, Orlando L. Human evolution: A tale from ancient genomes. Philos Trans R Soc B. 2017;372(1713), Art. #20150484. https://doi.org/10.1098/rstb.2015.0484

2. Jónsson H, Schubert M, Seguin-Orlando A, Ginolhac A, Petersen L, Fumagalli M, et al. Speciation with gene flow in equids despite extensive chromosomal plasticity. Proc Natl Acad Sci USA. 2014;111(52):18655–18660. https://doi.org/10.1073/pnas.1412627111

3. Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. Shedding light on the grey zone of speciation along a continuum of genomic divergence. PLoS Biol. 2016;14(12), e2000234, 22 pages. https://doi.org/10.1371/journal.pbio.2000234

4. De Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, et al. Chimpanzee genomic diversity reveals ancient admixture with bonobos. Science. 2016;354(6311):477–481. https://doi.org/10.1126/science.aag2602

5. Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, et al. The bonobo genome compared with the chimpanzee and human genomes. Nature. 2012;486:527–531. https://doi.org/10.1038/nature11128

6. Thackeray JF, Dykes S. Morphometric analyses of hominoid crania, probabilities of conspecificity and an approximation of a biological species constant. HOMO J Comp Hum Biol. 2016;67(1):1–10. http://dx.doi.org/10.1016/j.jchb.2015.09.003

7. Thackeray JF. Sigma taxonomy in relation to palaeoanthropology and the lack of clear boundaries between species. Proc Eur Soc Study Hum Evol. 2015;4:220.

8. Mayr E, Linsley EG, Usinger RL. Methods and principles of systematic zoology. New York: McGraw-Hill; 1953.

9. Fan P-F, He K, Chen X, Ortiz A, Zhang B, Zhao C, et al. Description of new species of *Hoolock* gibbon (Primates: Hylobatidae) based on integrative taxonomy. Am J Primatol. 2017;9999, e222631, 15 pages. https://doi.org/10.1002/ajp.22631

10. Fennessey J, Bidon T, Reuss F, Kumar V, Elkan P, Nilsson MA, et al. Multi-locus analyses reveal four giraffe species instead of one. Curr Biol. 2016;26(18):2543–2549. http://dx.doi.org/10.1016/j.cub.2016.07.036

11. VonHoldt BM, Cahill JA, Fan Z, Gronau I, Robinson J, Pollinger JP, et al. Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. Sci Adv. 2016;2(7), e1501714, 13 pages. http://dx.doi.org/10.1126/sciadv.1501714

12. Callaway E. Elephant history rewritten by ancient genomes. Nature News. 2016 September 16. http://dx.doi.org/10.1038/nature.2016.20622

13. Thackeray JF, Maureille B, Vandermeersch B, Braga J, Chaix R. Morphometric comparisons between Neanderthals and 'anatomically modern' *Homo sapiens* from Europe and the Near East. Ann Transv Mus. 2005;42:47–51.

14. Green RE, Krause J, Briggs AW, Maricic T, Stenzel, U, Kirchner M, et al. A draft sequence of the Neandertal genome. Science. 2010;328(5979):710–722. https://doi.org/10.1126/science.1188021

15. Thackeray JF. *Homo habilis* and *Australopithecus africanus*, in the context of a chronospecies and climatic change. Palaeoecol Afr. 2016;33:53–58.

16. Gordon AD, Wood BA. Evaluating the use of pairwise dissimilarity metrics in paleoanthropology. J Hum Evol. 2013;65465–477. https://doi.org/10.1016/j.jhevol.2013.08.002

17. Dykes SJ, Dykes R. 'Professor regressor': A computer programme for rapid processing of large sets of pairwise regression analyses in palaeontological contexts. Palaeontol Afr. 2015;49:53–59.

18. Thackeray JF. Approximation of a biological species constant? S Afr J Sci. 2007;103:489.

19. Wallace AR. Contributions to the theory of natural selection. London: Macmillan; 1870. p. 161. https://doi.org/10.5962/bhl.title.1254

20. Thackeray JF, Odes E. Morphometric analysis of early Pleistocene African hominin crania in the context of a statistical (probabilistic) definition of a species. Antiquity. 2013;87:1–2. http://antiquity.ac.uk/projgall/thackeray335/

21. Thackeray JF. Probabilities of conspecificity. Nature. 1997;390:30–31. https://doi.org/10.1038/36240