Why does Ross Cunningham, a statistician at the Australian National University, have a new marsupial species named after him?

A Possum's Tale — How Statistics Revealed a New Mammal Species

Peter Hall

In the United States the only naturally occurring marsupial, or pouched mammal, is the possum. In Australia, however, marsupial species abound and are the dominant native mammals. The best known Australian marsupials in North America are koalas and kangaroos and, thanks to Warner Brothers, the tasmanian devil. Yes, the Tasmanian Devil really exists; it is relatively common in the Australian island state of Tasmania. It earns its odd name from its fearsome teeth and the unnerving noise it makes when disturbed — a cross between a whistle and a growl.

However, the marsupial that many Australians know best is the possum. Possums have adapted extraordinarily well to the urban Australian lifestyle, and from some viewpoints are Australia's answer to North America's raccoons. They are nocturnal, they spend much of their lives in trees, and the common species of brushtail possum can survive quite happily on a diet of vegetable scraps. Many Australians regard them as pests; there are companies that specialize in removing possums from the roof spaces of houses, where they enjoy warmth and protection, and returning them to their native environment.

So, we Australians tend to think we know most things that are important about possums. We live with them in quarters that are sometimes closer than we would wish, so it's hard to believe we could overlook an entire species. Yet that is exactly what has happened. For more than two centuries — the period for which Australia has been settled by its nonindigenous inhabitants, and during which we have assiduously classified and catalogued the island continent's unique fauna — we have lived with an unnoticed species of possum. Even more remarkably, it took a very talented statistician, not a zoologist, to realize it was there, happily living in the forests of southeastern Australia.

***Trichosurus cunninghamii*, a species of possum named for statistician Ross Cunningham of Australia.**

The statistician is associate professor Ross Cunningham of the Australian National University, after whom the new species has been named: *Trichosurus cunninghamii*. Associate professor David Lindenmayer, an ANU environmental scientist with whom Cunningham has worked for many years, notes that while any discovery of a new species is significant, finding a new mammal, especially one as large as a possum, is nothing short of remarkable. "Every so often we come up with a new species of frog or reptile. But it's extremely unusual to discover a new mammal. This really charismatic animal has gone unnoticed for 200 years."

True, *T. cunninghamii* is something of a timorous beastie. It's rather shy, it lives in forests away from human habitation, and its diet consists principally

***T. caninus*, the mountain brushtail possum.**

of plants, ferns and fungi. Nevertheless, it is quite bold when captured. It's "feisty, it has a bit of spunk about it," says Lindenmayer.

Why wasn't the possum species noticed before? Members of the species have been observed, and trapped, many times over many years, by trained zoologists. However, in obvious physical respects *T. cunninghamii* resembles its apparently more numerous relative, *T. caninus*, once commonly called the mountain brushtail possum. Conventional wisdom had it that there was just one possum species, referred to as *T. caninus*, that inhabited eastern Australia's lush eucalyptus forests, living in tree hollows and dining on leaves in forest canopies up to 250 feet above the ground.

***T. cunninghamii*, the short ear possum.**

Cunningham showed the fallacy of this assumption. He has a statistician's intuition for biological data, the result of a quarter-century spent collaborating with biologists researching Australian animals and birds. He's no back-room data analyst, however; he's pursued rare birds across coral reefs, dugongs (large marine mammals) in tropical oceans, and now possums in forests of mountain ash, to ensure he understands the science as well as the intricacies of sampling.

In fact, Cunningham has a long-standing interest in the spatial statistical variation of bird and animal populations, in relation to the sizes and shapes of animals, for example — their morphological characteristics, to use the correct scientific term. He helped produce the Atlas of Australian Birds, a database showing the distribution of many different bird species across the continent, and has long been involved in the analysis of ecological data. He knew that when an ANU team of biologists began catching possums to study a parasite that afflicts them, the opportunity for doing more should be seized.

Indeed, the ANU's possum-trapping program, encompassing the states of Victoria, New South Wales, and Queensland, offered an unrivaled opportunity for gathering data that explored issues in which Cunningham was interested. So he prevailed upon his biologist colleagues to make many more measurements than they had planned, in order that their research on the parasite might be conducted concurrently with an analysis of the spatial variation of forest-dwelling possums.

Before Cunningham poked his statistician's nose into biologists' business, no one had seriously considered gathering data on the spatial distribution of the morphology of *T. caninus*. Experience with the koala, and with the common urban-living brushtail possum, led biologists to expect they would find a "cline" (or gradual spatial variation) of morphology within a single forest-possum species, not unlike the gradual geographic variation of some human physical characteristics.

Cunningham always tries to be involved in organizing the collection of data, and indeed with data collection itself, well before any analysis starts. This project was no exception. The fact that the principal possum dataset is so devoid of confounding and other obfuscating aspects owes a great deal to his experience and his involvement at an early stage. Moreover, the multivariate data that are pivotal to this research are remarkably close to being normally distributed, a tribute to careful planning which ensured that, as nearly as possible, the data were homogeneous within groups. Indeed, Cunningham's possum dataset is fast becoming a classic for training ANU statistics students in both exploratory and confirmatory data analysis, for example in teaching them how to use Cook and Weisberg's statistical package, Arc (Cook and Weisberg, 1999).

The possums were trapped at two sites in Victoria, three in New South Wales and two in Queensland. Figure 1 shows the geographic distribution. No harm was done to the animals; they were sedated after capture, before being measured, and carefully released (at the place of capture) after the effects of the sedative had worn off. A small piece of fur was shaved from each animal, and the skin there was marked so that recaptures could be immediately recognised.

The common brushtail possum, a species quite different from its mountain cousins, is well-known to be partial to apples. So apples were used to entice the forest dwellers, too, into traps.

Eight physical measurements were made of each animal: the length of its head, the width of its skull, eye size, ear length, total body length (the distance from the tip of the possum's nose to the end of its uncurled tail), tail length, chest girth, and pes (or foot) length. Age was determined too, and only data from older possums were used in the analysis described below, to minimize potential confounding of age and body shape. Fur color was noted, but subsequent analysis showed that it has little connection to species type. Other data, including weight, sex, tooth wear, etc, were recorded, but will not be discussed below. They can be unreliable in studies of the intrinsic variation of morphological features. For example, variations in weight can often be explained by dietary differences.

An exploratory study of the bivariate distributions of physical characteristics revealed distinct clusters for some pairs of attributes; see the matrix of scatterplots in Figure 2. Examination of the origin of the data in the clusters showed that one cluster consisted largely of data from the two Victorian sampling sites (numbered 1 and 2 in Figure 1) and the other of data from the remaining five sites. Motivated by these results, Cunningham used canonical variate analysis, or CVA, to study in more detail the variation of the eight morphological characteristics among the seven groups of data, i.e., the respective datasets from the seven trapping sites.

The technique of CVA (see, e.g., Blackith and Reyment, 1971) quantifies the relationship among groups of multivariate data by finding successive linear combinations, of components of the data vectors, that maximize distances among the groups. Equivalently, the ratio of the between-groups to the within-groups sum of squares is maximized, producing versions of Fisher's linear discriminant function as the successive linear combinations. (In the case of normal data, and assuming the two populations under study have identical covariance matrices, the likelihood ratio rule for allocating a particular data value to one population or another reduces to an inequality for a certain linear formula in the datum. Fisher's linear discriminant function results when the unknown population means and variances in this formula are replaced by their sample counterparts.)

Applying CVA to the groups, Cunningham found that the first variate emphasized ear, pes, and tail length:

$$\text{first variate} = -0.586 \times (\text{ear length}) - 0.302 \times (\text{pes length}) + 0.451 \times (\text{tail length}) + \text{smaller-order terms.} \quad (1)$$

Remarkably, this single variate accounts for 89% of variation in external morphology in the seven groups. The next two account for a total of 6% of variability, and the remainder give only 5%.

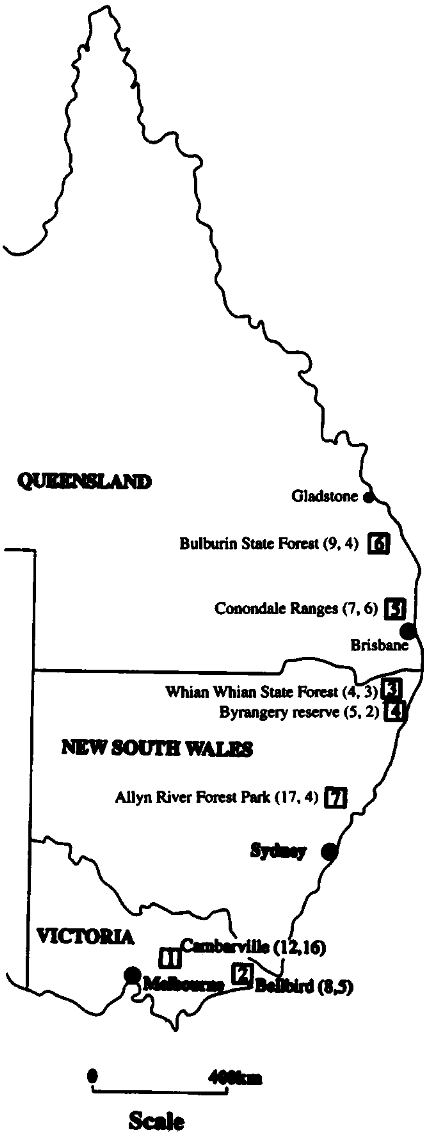

Figure 1. Geographic distribution of seven sites for sampling forest-dwelling possums. The number assigned to each site corresponds to the chronological order in which possums were sampled. Values in parentheses are the numbers of animals at each site (males, females).

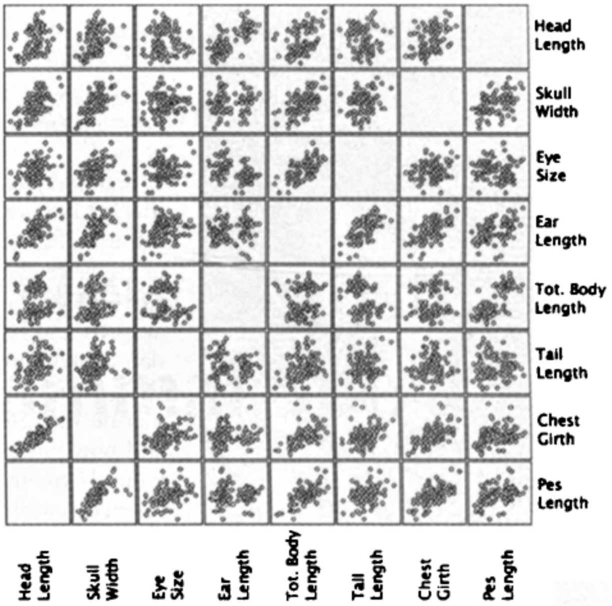

Figure 2. Scatterplot matrix. Results of eight physical measurements are shown in the form of bivariate plots. The order of the plots in the 8 x 8 matrix is the same as those of the eight measurements in Table 1 and described in the text: head length, skull width, eye size, ear length, total body length, tail length, chest girth, and pes length.

The second canonical variate is dominated by body length, which makes hardly any contribution to the first variate:

$$\text{second variate} = -0.279 \times (\text{total body length}) - 0.104 \times (\text{chest girth}) + \text{smaller-order terms.} \quad (2)$$

Table 1 gives coefficients of the first and second variates in greater detail, as they would appear in linear functions of the eight morphological measurements. Figures in parentheses are coefficients of the morphological variables after the latter have been standardized for scale, and so provide a better indication of the relative significance or importance of different measurements. However, it can be seen that this does not affect the choice of dominant terms; they are still those given on the right-hand sides of (1) and (2).

Figure 3 plots the seven within-group means, and confidence regions for those means, for the first and second canonical variates represented at (1) and (2) respectively. In the direction of the first canonical variate there is a marked separation between data from the two Vic-

Table 1— Coefficients for first and second canonical variates.

Variable	First canonical variate coefficients	Second canonical variate coefficients
Head length	.151 (0.438)	-.058 (-.170)
Skull width	.027 (0.070)	-.040 (-.105)
Eye size	.056 (0.054)	-.059 (-.087)
Length of the ear conch	-.586 (-0.819)*	.044 (.061)
Total body length	-.107 (-0.346)	-.279 (-.903)*
Tail length	.451 (0.721)*	.090 (.143)
Chest girth	-.091 (-0.158)	-.104 (.182)
Pes length	-.302 (-0.696)*	.036 (.083)

The seven groups to which canonical variates analysis was applied are the datasets of possums trapped at the seven respective sites shown in Figure 1. The first column in the table lists the eight morphological variables. The first values in the second column (respectively, in the third column) are the coefficients of the linear form in the eight morphological variables that defines the first (second) canonical variate. Values in parentheses are coefficients scaled by the within-group standard deviations. Asterisks denote those measures with the largest coefficients and which best discriminate between the two clusters.

torian trapping sites, numbers 1 and 2, and the other five sites, suggesting at least two quite different populations.

Since the first canonical variate is dominated by the lengths of the ear, pes, and tail, these attributes can be expected to provide the best discriminators between the two populations. In particular, possums from Victoria, in the south, are characterized by longer ears ($p < .001$) and longer pes ($p < .001$) than the animals found at other locations. This in itself provides a clue that the differences

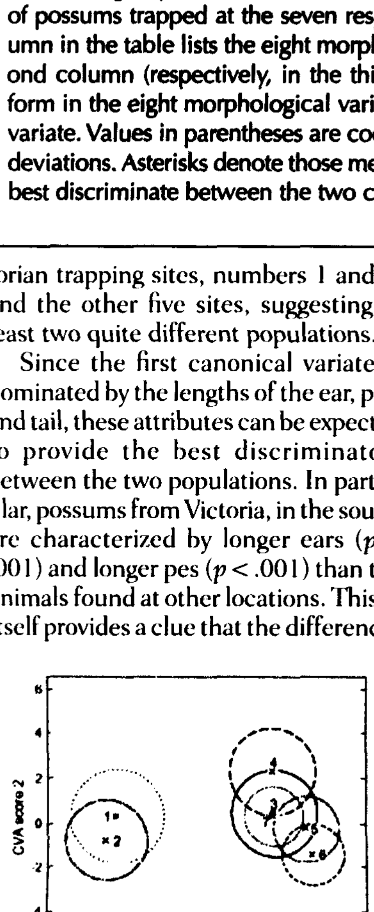

Figure 3. Clusters. Horizontal and vertical axes show the first and second variates in formulae (1) and (2), respectively. Numbers inside the discs refer to the different sampling sites shown in Figure 1. Each disc is centered at the variate mean for the corresponding sampling site, and the discs themselves show extremities of 95% confidence regions for disc centers.

might be unexpected. Allen's rule of morphological variation (Allaby, 1985) states that the lengths of appendages should tend to decrease with increasing latitude; here ear and pes length are increasing.

Tail length is admittedly an exception, but arguably not an important one. Allen's rule is based on the fact that a warm-blooded animal loses heat through protruding body parts, such as its ears, feet, and tail, and so in cooler climates (which tend to be those at greater latitude) those body parts should be smaller. The two sites in Victoria that are "outliers" in Figure 3 certainly have colder average temperatures than any of the other five. However, a possum would lose little heat through its thickly fur-covered tail, where there is only minor blood flow. There would certainly be much less heat loss there than through its bare ears and feet. Therefore Allen's rule would not apply with much effect to a possum's tail.

Figure 4 illustrates the three characteristics (ear, pes, and tail length) for a possum. The animal represented in the left-hand column lives in a warmer, northern climate, such as those at sampling sites 5–7 in Figure 1. The right-hand column shows a possum from the relatively cool south, living at sites 1 or 2. Note the bare skin on ears and feet, and the thick fur on the tail.

Using the three dominant features from the first canonical variate, Cunningham calculated the linear combination of the three principal morphological variables that best divided the data into two clusters:

$$\text{possum score} = 24 - 0.571 \times (\text{ear length}) - 0.149 \times (\text{pes length}) + 0.341 \times (\text{tail length}). \quad (3)$$

A forest-dwelling possum can be classified as a resident of Victoria if its "possum score" is negative; otherwise it lives in New South Wales or Queensland. Classification of the dataset, using this discriminator, is perfect, although the data themselves were used to produce the score function. More to the point, a box plot of possum scores shows clear separation between sites 1 and 2 on the one hand, and the other five sites on the other. Similarly, in a box plot of measurements of ear length (the main variable comprising the first canonical variate) for each of the seven groups, there is complete separation between the plot for any one of the first two sites and that for any one of the last five.

Analyses of these data were first published by Lindenmayer, Viggers, Cunningham and Donnelly (1995), where the findings were reported cautiously. The clear separation of the data into two clusters was noted, as too was the fact that the body measurements that characterize the clusters are "contrary to the expected thermoregulatory response... associated with increasing latitude, as predicted by Allen's rule." The authors concluded by suggesting that "factors other than climate have influenced the morphology of T. caninus."

At this point, tantalized by morphological differences that could not be explained adequately on environmental grounds, and suspecting taxonomic variation at which they could only guess, Cunningham and his coworkers turned to a genetic study to shed more light on the riddle. Blood samples had been taken from the possums at the time of capture and were used to extract DNA for analysis.

The results were stark. They showed unequivocal genetic differences between northern and southern populations of what hitherto had been assumed was a single species of possum. Specifically, the northern and southern forms

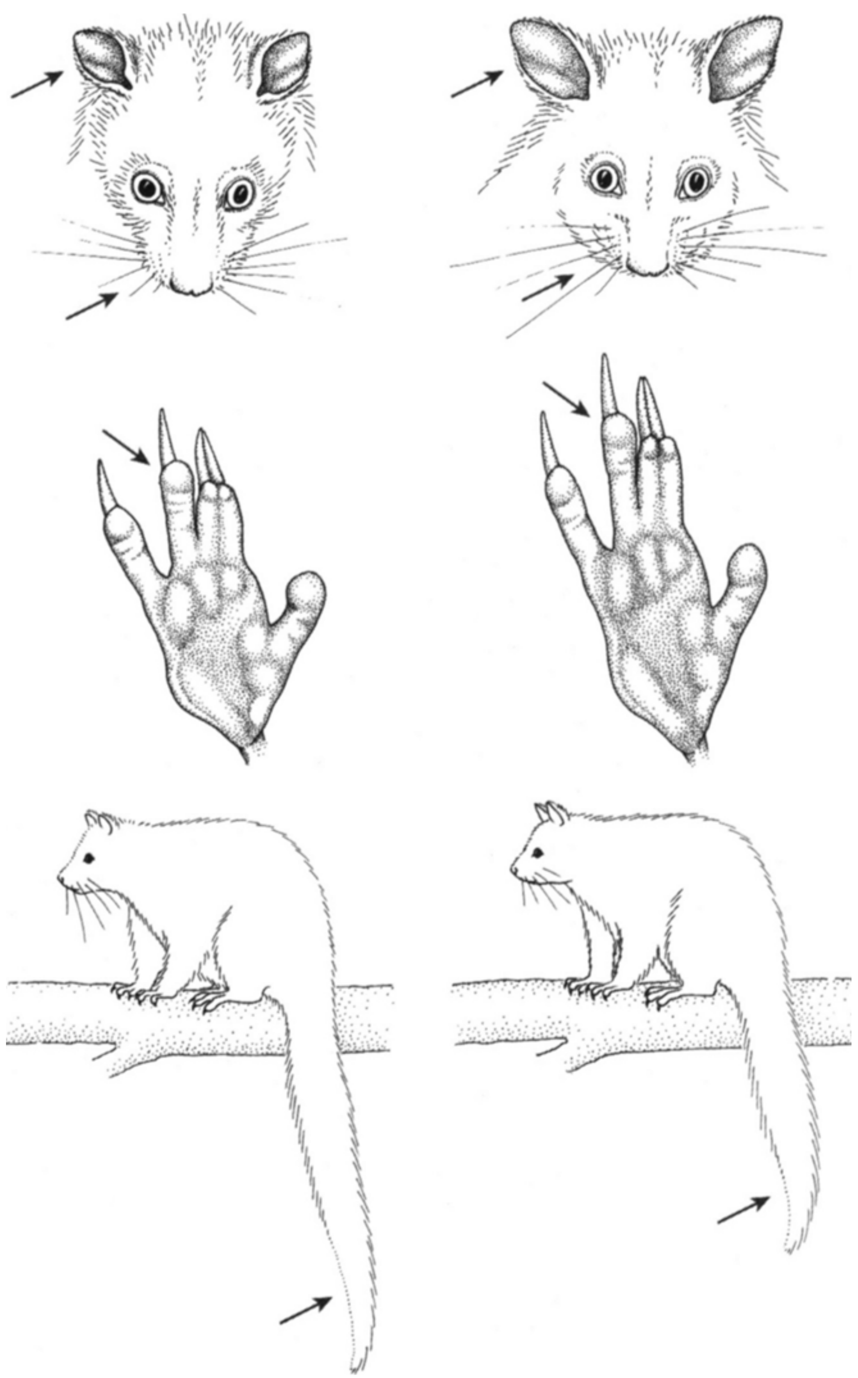

Figure 4. Ear, pes, and tail length. The first column shows characteristics typical of *T. caninus*, the northern-dwelling possum. This animal tends to have shorter ears, a shorter pes (foot), and a longer head and tail than its southern counterpart, *T. cunninghamii*, indicated in the second column.

differed genetically from one another by 2.8%; that is, in a comparison based on the variability of alleles of a gene, the northern and southern types of possum were found to differ in 2.8% of places. By way of comparison, humans and chimpanzees differ by about 1% on the same scale.

The significance of the genetic difference between possum populations was placed well beyond doubt by extensive further genetic analysis and computer simulation. The distance of 2.8% is about half the value it would have assumed if either cluster of animals was compared to a sister species, *T. vulpecula*. Therefore the northern and southern possums do not converge to a shared ancestral node with the sister species; they are closer to one another than to *T. vulpecula*. It was concluded that "*T. caninus* is most likely composed of two taxonomic forms" (Lindenmayer, Dubach and Viggers, 2002).

Cunningham conducted a Procrustcan analysis to compare the genetic and morphological differences among the seven sets of data from different trapping sites. Using concepts and methods introduced by Gower (1971, 1975), involving affine transformations of each dataset, he showed that the agreement between cluster patterns defined by morphological and genetic measurements, respectively, was remarkably close.

There is no pre-determined and universally accepted set of rules for defining a species or subspecies. Criteria tend to differ according to the methods used for comparison, and also with the definition of "species" that is employed; see Davis (1996) and Burgman and Lindenmayer (1998). However, the morphological and genetic differences between the clusters of forest-dwelling possums are so marked, and in such concordance with one another, that the conclusion that the possums belong to two different species is impossible to escape.

The concise geographic distributions of the two species are not yet known. Further information will have to await the analysis of data from a more ambitious trapping program, which is already underway. Nevertheless it is clear that in at least the forests of Victoria, in Australia's southeast, *T. cunninghamii* is reasonably common. The species is apparently not in any danger of extinction, although now that it has been recognized it can be seen to be under potential threat in some respects — for example, from commercial logging that threatens to destroy part of its habitat.

The apparently more numerous northern variety is now referred to by the old name for both species, *T. caninus*. This recognizes the fact that the original classification was based on a northern specimen captured near the Hunter River, at a point approximately 80 miles north of Sydney. That trapping occurred some 170 years ago (Ogilby, 1835; Thomas, 1888), indicating just how long these species of possum have been studied.

Increasingly, the common name for *T. cunninghamii* is "mountain possum" or "bobuck," the latter being an aboriginal name for a possum that lives in mountain forests. The earlier-recorded species, whose habitats in New South Wales and Queensland are predominantly not in mountainous regions but in coastal forests, has become "the short-eared possum," acknowledging its most distinguishing feature — or equivalently, of the dominant term in Cunningham's formula (1) for the first canonical variate, or in his linear classifier (3). Thus, the old species' new common name might fairly be said to derive directly from a statistical analysis. The scientific name for the new species is that of the statistical analyst himself.

Forest-dwelling possums of either species are reasonably gregarious, at least among their fellows; they "move house" at least 20 times each year. However, they probably never figured on spending so much time in the company of a remarkable, determined but unassuming statistician who has caused part of Australia's record of its marsupial species to be rewritten.

Acknowledgements. I'm grateful to Ross Cunningham, Rob Hyndman, David Lindenmayer and Jeff Wood for helpful comments.

References

Allaby, M. (1985). *The Oxford Dictionary of Natural History*. Oxford University Press, Oxford.

Blackith, R.E. and Reyment, R.A. (1971). *Multivariate Morphometrics*. Academic Press, New York.

Burgman, M.A. and Lindenmayer, D.B. (1998). *Conservation Biology for the Australian Environment*. Surrey Beatty, Chipping Norton, UK.

Cook, R.D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. Wiley, New York.

Davis, J. (1996). "Phylogenetics, molecular variation, and species concepts." *Bioscience* 46, 502–511.

Gower, J.C. (1971). "Statistical methods for comparing different multivariate analyses of the same data." *Mathematics in the Archaeological and Historical Sciences* Eds. F.R. Hodson, D.G. Kendall and P. Tautu, pp.138–149. Edinburgh University Press, Edinburgh.

Gower, J.C. (1975). "General Procrustes analyses." *Psychometrika* 40, 33–51.

Lindenmayer, D.B., Dubach, J., and Viggers, K.L. (2002). "Geographic dimorphism in the mountain brushtail possum (Trichosurus caninus) — the case for a new species." *Aust. J. Zool.*, in press.

Lindenmayer, D.B., Viggers, K.L., Cunningham, R.B., andDonnelly, C.F. (1995). "Morphological variation among populations of the mountain brushtail possum *Trichosurus caninus* Ogilby (phalangeridae: marsupialia)." *Aust. J. Zool.* 43, 449–458.

Ogilby, W. (1835). *Proc. Zool. Soc. Lond.* 1835, part III, 191–192.

Thomas, O. (1888). *Catalogue of the Marsuplia and Monotrema in the Collection of the British Museum* (Natural History). Trustees of the Museum, London.