RETRACTED: Is photoshop with Qualitative Image Analysis a valid technique for measuring hair morphology? A test using wires of known dimensions

Authors: Glueck, Christopher, and Wilson, James A.

Source: Folia Zoologica, 68(4) : 235-245

Published By: Institute of Vertebrate Biology, Czech Academy of Sciences

URL: https://doi.org/10.25225/fozo.012.2019

Folia Zool. – 68 (4): 235–245 (2019) DOI: 10.25225/fozo.012.2019

RETRACTED: Is photoshop with Qualitative Image Analysis a valid technique for measuring hair morphology? A test using wires of known dimensions

Christopher GLUECK and James A. WILSON*

Department of Biology, University of Nebraska at Omaha, 6001 Dodge Street, Omaha, NE 68182-0040, USA; e-mail: jameswilson@unomaha.edu

Received 28 February 2019; Accepted 5 August 2019

Abstract. This article has been retracted at the request of James A. Wilson, author.

It has been brought to the attention of the author of this work, as well as the Editor-in-Chief of *Journal of Vertebrate Biology*, that there are substantial methodological errors in this paper relating to the measurements of wires, which resulted in inaccurate results. These deficiencies were substantial enough to result in incorrect conclusions regarding the appropriateness of the procedure, which was the focus of the article.

Correction of these errors will result in the conclusions of the work being substantially altered. Therefore a decision has been made to retract publication of this *Journal of Vertebrate Biology* article.

Key words: digital photography, fox squirrel, ImageJ, melanism, *Sciurus niger*, thresholding

Introduction

Hair is prevalent on almost all mammals and mediates a range of physiological and/or behavioural functions. For example, hair is important for maintaining thermoregulation in mammals (Glanville & Seebacher 2010), providing a means of communication between individuals (Wilcox & Larsen 2009), and is involved in a number of behavioural systems (Ahl 1986). Although hair is an important feature of mammals, little is known about the relationship between hair morphology and these functions. Specifically, hair morphology is thought to play a role in thermoregulation (Walsberg 1988) as radiative heat gain is an important component of thermoregulation, even for endotherms (Walsberg & Wolf 1995, Walsberg et al. 1997). In addition to morphology, hair colour has been shown to alter heat absorption rates with darker hairs having increased heat absorption (Walsberg 1983, Armitage 2009).

Hair represents a more responsive component in heat balance, compared to more fixed traits such as body size (Reynolds 1993, Steudel et al. 1993). Many mammals shed their fur twice annually to compensate for warm summers and cold winters. For example, lemmings can effectively double their fur depth and density to increase insulation during the winter (Steudel et al. 1993). However, in smaller animals such as mice (~4× smaller than lemmings), scaling issues may preclude such flexibility (Steudel et al. 1993). Hair morphology can also have negative impacts on thermoregulation. For example, hair has approximately 8× higher thermal conductivity than motionless air (Steudel et al. 1993). As a result, high-density fur may actually increase heat loss compared to a less dense fur coat that traps a boundary layer of air near the skin.

The importance of hair morphology in mammalian thermoregulation is hypothesized to be related to the hair's impact on heat transfer between the animal and the environment (Bejan 1990). Specifically, mathematical modelling shows that the physical dimensions of hair are important variables to consider when examining heat transfer rates and that optimal hair width does not scale linearly with animal size (Bejan 1990). While hair width is certainly an important factor in heat transfer, there are limits to

* *Corresponding Author*

how far hair width can be increased. Hair morphology represents an important trait under the influence of natural selection, which humans have taken advantage of during artificial selection. For example, in tropical and subtropical regions, farmers place a high value on cattle that can cope with heat stress (Olson et al. 2003). Olson et al. (2003) identified a major gene that impacted the hair coat type that is inherited, which allowed cattle breeders to increase the frequency of cows with heat tolerant hair.

In addition to interspecific differences in hair morphology there can be considerable intraspecific differences in hair between individuals or even across the body of each individual. For example, Hetem et al. (2009) found that springboks (*Antidorcas marsupialis)* with three different fur colours had different thermoregulatory behaviours and body temperatures. Specifically, black springboks maintained a higher body temperature during the winter and lower foraging behaviours than white or normal coloured springboks. More recently, Fratto & Davis (2011) found that melanistic (black fur) fox squirrels (*Sciurus niger*) had significantly thinner body hairs and thicker tail hairs than the other fur colours, presumably to offset increased heat absorption due to the darker coloured fur.

Human hair also exhibits a wide range of physical characteristics including fibre shape, curvature, kink, and colour (Restano et al. 2001, Lasisi et al. 2016, De La Mettrie et al. 2019). Within humans there is also variation in the cross-sectional shape of hair follicles, both in terms of width and shape (round *vs.* oval). Specifically, Asian populations have thicker cross-sectional area (Lasisi et al. 2016) and straighter hair (Fujimoto et al. 2008), whereas humans of African descent have a more elliptical cross-sectional shape and greater curvature (Lasisi et al. 2016). However, given these general differences among the major human groups, De La Mettrie et al. (2019) warn that these broad classifications do not adequately represent the full variation in human hair and that eight categories of hair type might represent a better classification scheme.

Early studies on hair morphology in humans used microscopy to allow the measurement of key characteristics such as the curvature of human hair. Hrdy (1973) suggested that the curvature of a single hair shaft can be measured by mounting the hair between two glass slides, thereby reducing any three-dimensional curvature to an easily measured two dimensions. Lasisi et al. (2016) took this methodology further and successfully analysed digital scans of hairs using ImageJ (https://imagej.nih.gov). It should be noted that for hair curvature, Lasisi et al. (2016) digitally traced the curvature of each hair by hand and used these digital "curve data" to perform a modified analysis to that described by Hrdy (1973). In addition, unlike Davis (2010), Lasisi et al. (2016) made the precise methodology of their digital curvature analysis available as supporting documentation in their publication.

Davis (2010) and Fratto & Davis (2011) continued the use of digital imagery to offset the use of classic microscopy in measuring hair morphology. The study by Fratto & Davis (2011) was particularly noteworthy because they used digitized image of hairs and a computer program that automatically measured the hairs (Davis 2010, Fratto & Davis 2011). The new technique, originally proposed by Davis (2010), provided hair size and colour information that could be relevant for thermoregulation between different squirrel colour morphs (melanistic *vs.* normal). The technique was introduced with the goal of reducing the time needed to manually measure hairs with a microscope, and thus promoting the study of hair morphology (Davis 2010).

The technique proposed by Davis (2010) automated the process of measuring the general size characteristics of a small number of hairs at a time and involved mounting ten sample hairs onto a white notecard. These cards would then be scanned into a digital file at maximum resolution (i.e. 1200 dpi) for the scanner that was used (Davis 2010). The image file generated could then be loaded in Adobe Photoshop, and the software extension FoveaPro (currently Qualitative Image Analysis 64) could be used to automatically measure all hairs on the card (Davis 2010).

Along with the introduction of this novel technique, Davis (2010) claimed that sample cards can be prepared such that "hairs need not be straight, but should not overlap" and that sample preparation requires only minimal care because "each card took approximately two minutes to make" (Davis 2010). In his initial study Davis (2010) took 120 deer hairs and made 12 sample cards, with 10 hairs per card, and the time required to process those cards "was less than five minutes."

While the concept behind the Davis (2010) technique is innovative, it may require a more complex understanding of the imaging software than was indicated in his study. In addition, Davis (2010) does not report whether any validation was performed to verify that the image analysis program was performing as expected. In addition, the methodology outlined in

Davis (2010) is lacking in important details that make replication difficult at best.

Unlike using microscopes to directly measure individual hairs, image-processing software measures the pixels of a scanned image that represent the actual hair. This is an important distinction because the process of digitizing the hair may lead to unknown errors that could render image analysis of hair measurements inaccurate. For example, the resolution of a scan determines the number of pixels (often listed as dots per inch, dpi) in a given area of the image, and it is these pixels that are used to calculate hair dimensions. Therefore, if the scanned image resolution is not adequate then small differences in hair width, for example, might not be detectible.

Davis (2010) presented two possible image analysis programs for analysing hairs: 1) ImageJ, and 2) Qualitative Image Analysis 64 (QIA-64 2016; referred to as "Fovea-Pro" in Davis 2010). QIA-64 was designed to run as an add-on to Adobe Photoshop and had the ability to automate measurements in scanned images. This study will only focus on QIA-64 because it was the only software explicitly used by Davis (2010) and Fratto & Davis (2011) to measure hair morphology. Additionally, ImageJ does not provide many of the important hair measurements useful to biologists.

QIA-64 is capable of detecting features of interest (e.g. individual hairs) through a process called thresholding, which changes pixels with a colour or darkness above a defined threshold into black features (i.e. hair) pixels or background (white) pixels. This process essentially isolates the hair from the background, allowing it to be measured as one continuous feature. However, problems can occur if imperfections in the card sample (e.g. bubbles in the tape or shadows created during scanning) create additional dark pixels. These can then incorrectly be added to the dimensions of the actual hair and lead to measurement error. Therefore, control of the thresholding process may provide a more accurate representation of the hair and lead to more accurate measurements.

This study tested the accuracy of the digital analysis technique used by Davis (2010) by exploring three related methods for measuring hair length and width. Davis (2010) relied entirely on animal hairs of unknown size to test the technique, whereas this study focused on validation using objects of known length and width. We predicted that the indirect nature of the image analysis methods being tested would not generate accurate measurements, but might still show general width trends between the wires studied. We also predicted that, contrary to what is claimed by Davis (2010), the orientation and curvature of a hair when placed on the sample card is an important consideration when using QIA-64.

Material and Methods

Experimental design

This study attempted to use the methods described by Davis (2010), although not all the methodological details were clear. For example, Davis (2010) implied that he measured an entire card containing multiple hairs at once, and may not have isolated and measured each hair individually. We did not test the difference between analysing each hair/wire sample in a separate data file and analysing multiple hairs in a single data file. Theoretically the hair/wire samples on a card should be far enough apart that they would not generate false images that would influence other hair samples. For this study, we chose to isolate each wire individually and analyse it as a separate data file. While this may not be the same procedure as Davis (2010), it does allow us to more properly identify how well the QIA-64 program functions under the different sample conditions (width, orientation, curvature).

For this study experimental cards were created by mounting various samples of wire or thread onto a blank notecard. Each wire was measured in such a way that a text file was generated for every wire in the sample, as opposed to generating only one data file for the four wires on each notecard. Samples were mounted using ultra-clear tape, as described by Davis (2010). Experimental treatments consisted of five sample widths (Standard 1-5) ranging from 0.65 to 0.15 mm arranged in one of two card Layouts (A, B). Layout A represented straight samples mounted at different angles while Layout B consisted of samples that were curved or bent in various ways. This produced a 5 × 2 factorial experimental design with Standard as the first factor and Layout as the second factor.

Sample preparation

Sample wires/threads were obtained from a craft shop and were labelled as Standard 1 through Standard 5. Standard 1 was 0.65 mm dark brown, waxed cord. Standard 2 was 0.51 mm dark brown, vinyl coated bead wire. Standard 3 was 0.36 mm black, nylon-coated stainless steel wire. Standard 4 was 0.20 mm silver, bead wire. Finally, Standard 5 was 0.15 mm black, thermally bonded bead-weaving thread and was of a similar width as the deer hairs used by Davis (2010). All wires were cut to be the same

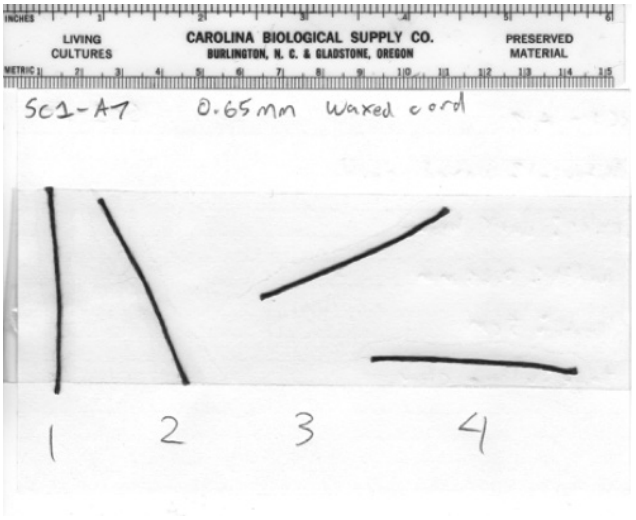

Fig. 1. Sample scan of a straight wire card (Layout A). This card has wire samples from Standard 1 (0.65 mm). Wire orientations moving across the card from left to right were: Wire 1 (Vertical), Wire 2 (~ 65 degrees), Wire 3 (~ 35 degrees), and Wire 4 (Horizontal).

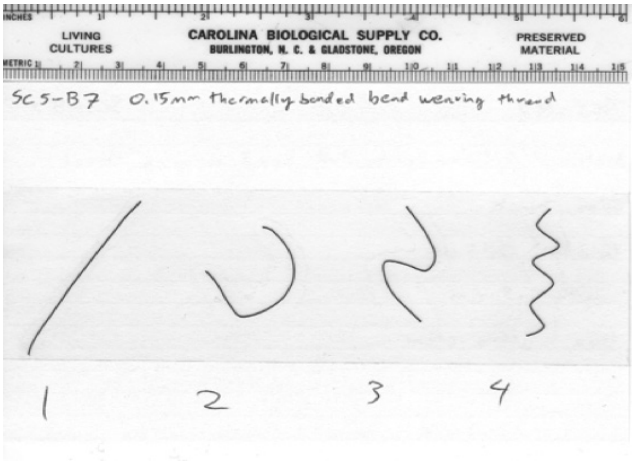

Fig. 2. Sample of a curvy wire card (Layout B). This card has wires from Standard 5 (0.15 mm). Within Layout B the first wire is straight, but the remaining wires exhibit varying degrees of curvature, the second wire is "U-shaped", the third wire is "S-shaped", and the fourth wire is "curly".

length, 50 mm, and taped onto a white 101 × 152 mm notecard with ultra-clear tape. Each card consisted of a series of four wires placed in four orientations based on one of two layouts, labelled A or B. Each wire was arranged on the card such that no wires touched, with enough space in between to allow individual wires to be isolated and saved as individual digital files.

Layout A comprised straight wires placed in four different angles on the notecard (Fig. 1). Wire orientations moving across the card from left to right were: Wire 1 (Vertical), Wire 2 (~ 65 degrees), Wire 3 (~35 degrees), and Wire 4 (Horizontal). Layout B was used to identify any effect of sample curvature in measuring length and width. Within Layout B the first wire was straight, but the remaining wires exhibit varying degrees of curvature: the second wire was U-shaped, the third wire was S-shaped, and the fourth wire was "curly" (containing multiple bends; Fig. 2).

For each of the Standards (1-5), ten copies of both Layout A and Layout B cards were prepared. This resulted in the creation of 100 cards, with a total of 400 wire samples. Each card was scanned, with a ruler included in each scan, using a flatbed scanner at the highest resolution possible for the scanner (1200 dpi, like Davis 2010). Scanned images were imported into Adobe Photoshop and the analysis of each scanned wire was performed using the QIA-64 add-on to Adobe Photoshop.

Thresholding algorithm

Thresholding is a process whereby the program identifies pixels in the image that are sufficiently dark (i.e. above the darkness threshold) to be considered part of the sample (wire or hair). Pixels above the threshold are counted as a "feature" (wire or hair) whereas pixels below it are considered as background and set as white pixels. QIA-64 provides an automatic thresholding feature that uses an algorithm and statistical tests to define the image contents automatically. According to the QIA-64 manual, the default thresholding option was originally developed to distinguish printed text from a scanned document. However, Davis (2010) used this option to identify individual hairs from the white notecard background. In addition to the default automatic thresholding algorithm, QIA-64 provides seven alternative algorithms that could potentially enhance the thresholding of particular hair or wire samples. This study explored two thresholding algorithms: automatic (used by Davis 2010) and the Johannsen algorithm. The Johannsen algorithm was used as an alternative thresholding algorithm in this study because preliminary testing showed that it performed well in defining some of the wires from Standard 1.

In this study each wire sample was analysed by QIA-64 using three different methods (Methods 1-3). First, wire samples were analysed using a method (Method 1) that was thought to be most similar to Davis (2010). Method 1 used the default automatic thresholding taking any pixel that the program considered as a sample, regardless of whether it actually was part of the sample. This resulted in potential over-inclusivity of pixels as "features". Method 2, like Method 1, used the default automatic thresholding technique, but allowed the experimenter to manually select which pixels would be considered as part of the sample.

Finally, Method 3 used the Johannsen algorithm, as a potential alternative to the default automatic thresholding algorithm. The details of each method are described below.

Image analysis Method 1

Method 1 was designed to match the methodology of Davis (2010) as closely as possible, although Davis did not explicitly state exactly which options were used. First, images were calibrated to a 1 mm scale using the ruler that was included with each scan. Second, using the program's selection tool menu, a digital box was drawn to encompass each individual wire's image ensuring that no other wires were included in the sample image. QIA-64 add-on places a new menu bar within the original Adobe Photoshop menu bar called "Filter". All QIA-64 functions can be found under the Filter top bar menu, including the thresholding options. The third step was to select the Bilevel Thresholding option from the QIA-V Thresholding menu with the Automatic option selected. Following the completion of the thresholding step, "Measure All Features" was selected from the QIA-VIII feature measurements menu. This option automatically included all pixels that the thresholding algorithm had determined to be a "feature", or part of the wire sample. Finally, this option prompts the user to create and save the image as a new file.

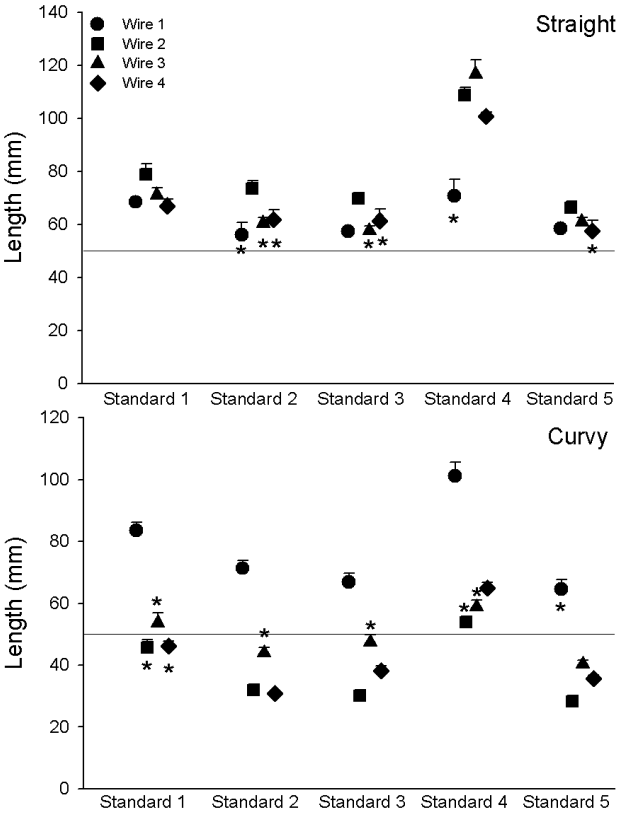

Fig. 3. Mean length (mm) for Layout A (straight) and Layout B (curvy) wires using Method 1 (Davis 2010). Standards 1-5 represent wire thickness from 0.65 to 0.15 mm. Specific wire orientation can be referenced in Figs. 1 and 2. Wires with asterisks were not significantly different than the expected length of 50 mm, indicated by the line.

Image analysis Method 2

Method 2 was designed to refine the process used by Davis (2010) by manually selecting the individual features used to create the digital wire sample. It was thought that Method 1 might generate a large number of inappropriate pixels resulting in an overestimation of the wire size. The main difference between Methods 1 and 2 was that under the QIA-VIII feature measurements menu Method 2 used the "Select Features" option following the automatic thresholding step instead of the "Measure All Features" option. The "Select Features" option allowed individual pixels to be manually selected for inclusion in the wire sample image. For the purposes of this study, feature selection was done using the Mean Fibre Width category, which grouped features into columns based on the detected Mean Fibre Width. When wire or hair samples are scanned the vast majority of the sample is contained in a single, large "feature", but several other smaller features (dark shadows, tape bubbles, or dust particles) may also meet the thresholding criteria and will be listed as unique "features". In many cases, all but one feature would be labelled as having a very small width (i.e. not part of the wire). Items with widths near zero were assumed to be imperfections in the scan and were eliminated from the analysis.

Image analysis Method 3

Within the QIA-64 program there are several additional thresholding algorithms that can be used to choose which pixels to include in sample analysis. Method 3 was designed to test one of these to determine if there was any improvement in the quality of data compared to the default automatic option. The Johannsen algorithm was selected after some preliminary testing on the Standard 1 wires. Features were selected for using the Mean Fibre Width category, as in Method 2, thus difference between Methods 2 and 3 was the use of the Johannsen thresholding algorithm.

Once all three analysis methods were completed, data from the variables of interest (length, Mean Fibre Width, and Inscribed Radius) for each wire were compiled. Wire length was measured as the sum of all pixels along the long axis. However, the values used to measure width – Mean Fibre Width and Inscribed Radius – were measured from the widest feature

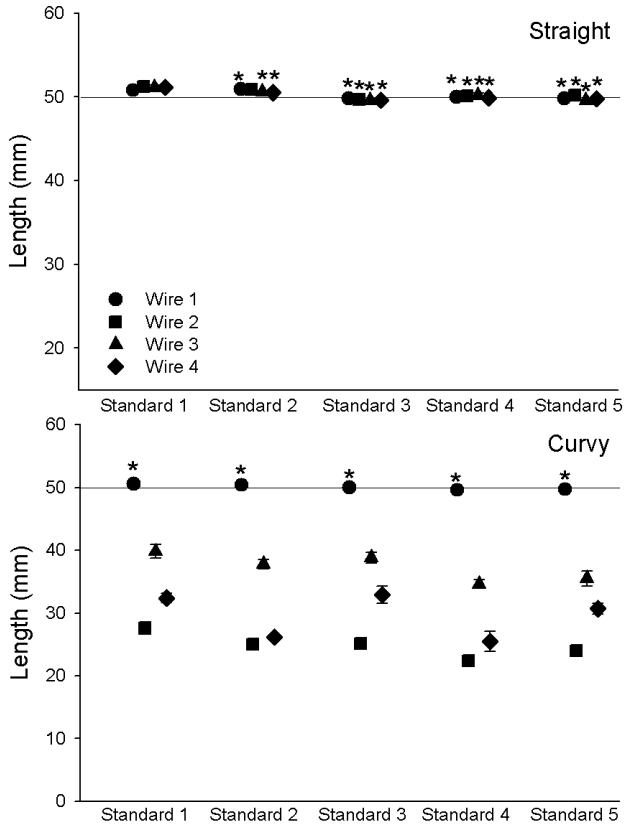

Fig. 4. Mean length (mm) for Layout A (straight) and Layout B (curvy) wires using Method 2. Standards 1-5 represent wire thickness from 0.65 to 0.15 mm. Specific wire orientation can be referenced in Figs. 1 and 2. Wires with asterisks were not significantly different than the expected length of 50 mm, indicated by the line.

present in the data set. Mean Fibre Width was a simple measure of the length of all the pixels spanning the width of the wire. However, the Inscribed Radius was a measure that determined the size (radius) of a circle that would fit inside the pixels that were determined to be the wire's width.

Statistical tests

Statistical analysis of the data generated by QIA-64 was performed using SAS. Identifying whether the measurement procedure correctly captured the true length or width was performed using Bonferroni adjusted paired t-tests (significant *P*-value 0.00125). For any t-test with a *P*-value less than 0.00125 the QIA-64 measurement was considered significantly different from the actual measurement. A 5 × 2 factorial ANOVA (Standard × Layout) was used to determine if layout and wire orientation had an effect on measurement accuracy for each of the three methods. Finally, regression analysis was used to determine if there was a difference in each method's accuracy across a range of widths. All values reported are mean ± standard error.

Results

Wire length

Length estimates varied according to analysis method with Method 1 performing worse than Methods 2 and 3 (Figs. 3-5). Method 1 only produced accurate length estimates for 15 (37.5 %) of the 40 samples (Fig. 3). Of these, seven were from Layout A (straight wires) and eight were from Layout B (curvy wires). In addition, estimates from Method 1 were not consistent for most wire orientations and showed significant differences in measures for Layout A ($F_3 = 9.81$, $P < 0.001$) and Layout B ($F_3 = 58.95$, $P < 0.001$).

However, Method 2 produced correct length estimates for 21 (52.5 %) of the 40 wire samples (Fig. 4). Here, 16 correct length estimates were from Layout A (straight) and five from Layout B (curvy). All of the Layout B wires that were estimated accurately were from the same layout position: the straight wire. All of the estimates for Layout A were accurate ($F_3 = 0.07$, $P = 0.97$). However, for the curvy samples (Layout B), estimates were inaccurate for all wire orientations ($F_3 = 1173.08$, $P < 0.001$).

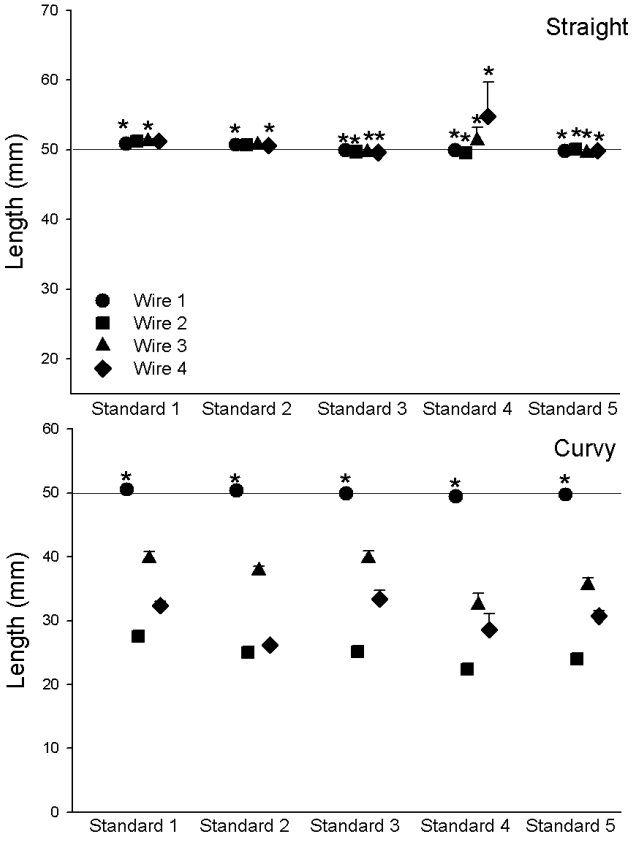

Fig. 5. Mean length (mm) for Layout A (straight) and Layout B (curvy) wires using Method 3. Standards 1-5 represent wire thickness from 0.65 to 0.15 mm. Specific wire orientation can be referenced in Figs. 1 and 2. Wires with asterisks were not significantly different than the expected length of 50 mm, indicated by the line.

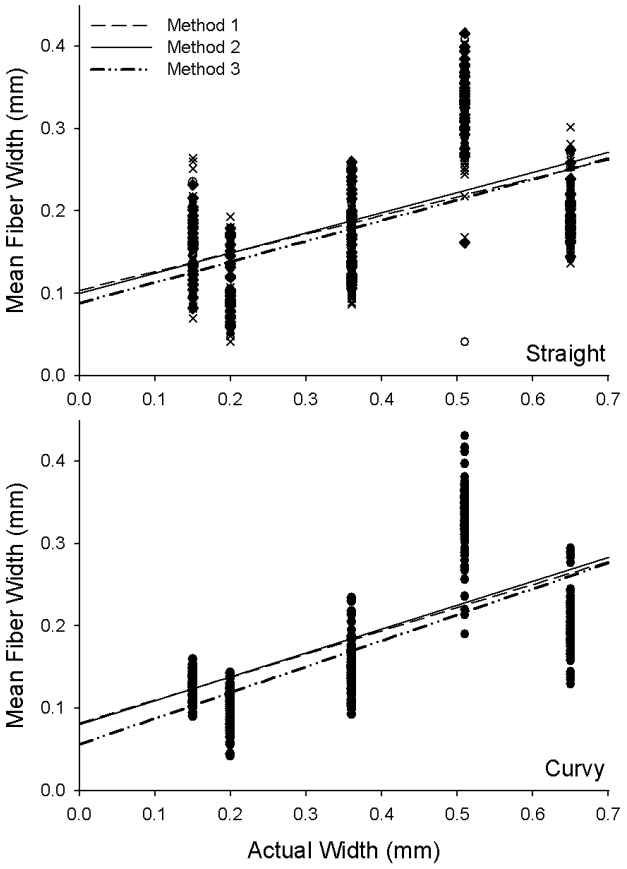

Fig. 6. Mean Fibre Width (mm) measured from the QIA-64 analysis for Layout A (straight) and Layout B (curvy) wires compared to the actual width of the wires. Data are from all three methods of analysis.

Using Method 3, 21 (52.5 %) of the 40 wire layouts were accurately estimated (Fig. 5). From these, 16 were from Layout A (straight) and five were from Layout B (curvy). As with Method 2, all of the estimates were accurate for Layout A ($F_3 = 0.74$, $P = 0.53$) but inaccurate for Layout B ($F_3 = 422.46$, $P < 0.001$).

Methods 2 and 3 were consistently more accurate than Method 1. In addition, the most accurately measured wires were all straight. Length was consistently underestimated for curved wires, the "U-shaped" wire being the most underestimated across all lengths, followed by the curly wire and the "S" shaped wire.

Mean fibre width

Width estimates using the Mean Fibre Width algorithm tended to be underestimated for both Layouts (Fig. 6), although estimates were accurate for 0.15 mm and close to accurate for 0.20 mm wires. There was no significant difference in the estimate accuracy among methods either for straight wires ($F_2 = 2.64$, $P = 0.07$), Method 1 (slope = 0.23, intercept = 0.10, $R^2 = 0.25$), Method 2 (slope = 0.24, intercept = 0.10, $R^2 = 0.28$), or Method 3 (slope = 0.25, intercept = 0.09, $R^2 = 0.28$) or curvy wires (Layout B; $F_2 = 0.49$, $P = 0.61$), Method 1 (slope = 0.28, intercept = 0.08, $R^2 = 0.40$), Method 2 (slope = 0.29, intercept = 0.08, $R^2 = 0.41$), and Method 3 (slope = 0.32, intercept = 0.06, $R^2 = 0.42$).

Wire orientation had a significant effect for width estimates of straight wires using Method 1(Layout A; $F_3 = 3.14$, $P = 0.025$), but not for curvy wires (Layout B; $F_3 = 0.16$, $P = 0.92$). The same was true for Method 2: straight wires (Layout A; $F_3 = 2.96$, $P = 0.03$), curvy wires (Layout B; $F_3 = 0.15$, $P = 0.93$) and Method 3: straight wires (Layout A; $F_3 = 10.11$, $P = 0.001$), curvy wires (Layout B; $F_3 = 0.75$, $P = 0.52$).

Inscribed radius

Width estimates using the Inscribed Radius algorithm performed better than the Mean Fibre Width algorithm, but lost accuracy with the thinner wires (0.15 and 0.20 mm, Fig. 7). Estimates were similar for both straight and curvy wires. Estimates among methods were not identical for straight wires, ($F_2 = 3.19$, $P = 0.04$) with Method 3 having a slightly lower intercept than Methods 1 and 2; Method 1 (slope =

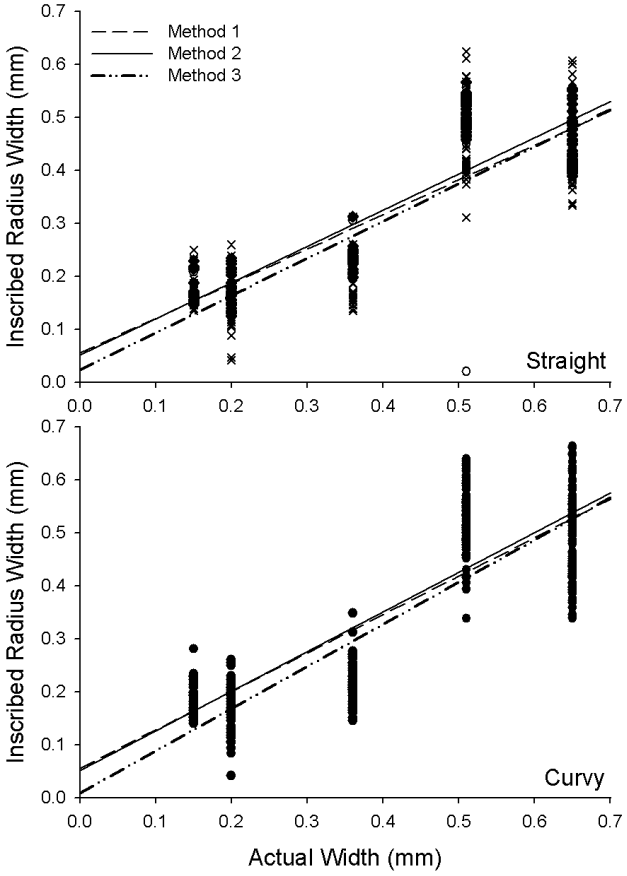

Fig. 7. Mean Inscribed Radius (mm) measured from the QIA-64 analysis for Layout A (straight) and Layout B (curvy) wires compared to the actual width of the wires. Data are from all three methods of analysis.

0.65, intercept = 0.06, R^2 = 0.74), Method 2 (slope = 0.68, intercept = 0.05, R^2 = 0.78), and Method 3 (slope = 0.70, intercept = 0.02, R^2 = 0.71). There was no difference among methods in the estimates for curvy wires (F_2 = 1.33, P = 0.26), Method 1 (slope = 0.72, intercept = 0.06, R^2 = 0.75), Method 2 (slope = 0.75, intercept = 0.05, and R^2 = 0.76), Method 3 (slope = 0.80, intercept = 0.01, R^2 = 0.72). Wire orientation did not affect estimates for either straight or curvy wires using Method 1, straight (Layout A; F_3 = 1.60, P = 0.19), curvy (Layout B; F_3 = 2.07, P = 0.10) or 2, straight (Layout A; F_3 = 1.16, P = 0.33), curvy (Layout B; F_3 = 2.01, P = 0.11). Orientation did affect estimates using method 3 for straight wires (Layout A; F_3 = 4.07, P = 0.007) but not for curvy wires (Layout B; F_3 = 1.90, P = 0.13).

Discussion

The length measurements made using QIA-64 were substantially improved when specific features (i.e. pixels belonging to the wire of interest) within the program were manually selected. For example, wires from Layout A tended to be overestimated using Method 1, the automated method used by Davis (2010). However, Methods 2 and 3 showed most Layout A wires to be measured reasonably accurately. Since automatic thresholding was used for both Method 1 and Method 2, the increased accuracy in Method 2 can be attributed to the use of manual feature selection. This procedure eliminates incidental features (shadows, tape bubbles, or other mistakes) captured by the automatic thresholding process. Estimates using Method 1 also varied with wire orientation. For example, two wires of equal length were estimated at 70.8 mm when vertical, but 116.6 mm at a ~35 degree angle. However, using Method 2 there was < 1 mm difference in length between these same two wires.

Davis (2010) suggested that the hairs measured did not need to be straight in order to be measured accurately. However, this study found that length measurements for all samples of curved wires (Layout B) were less accurate than straight wires, indicating that curvature is an important consideration when preparing samples. The increase in errors associated with curved wires is likely a result of the way the program determines the ends of the sample. Specifically, the QIA-64 manual states that the length is a measure of the distance between the two outermost pixels, and it might be difficult for the program to determine where the ends of the wire are in a "U-shaped" wire. This finding might explain why estimates for the "U-shaped" wire were consistently shorter than for the "S-shaped" wire, and why estimates were only accurate for the straight wires.

Method 1 did produce some of accurate estimates for the curved wires but this might be a result of chance rather than an indication of accuracy in the method. Specifically, when all features were automatically selected in straight wires when using Method 1, there was a tendency to overestimate the length of each wire. However, when measuring curved wires (Layout B) the program typically appears to underestimate curved wire length. In some samples, the algorithm might arrive at a more correct estimate when it combines both over and underestimating errors, effectively cancelling out these two errors.

Defining the width of each wire using QIA-64 is more complicated than measuring length because there is not a specific data category defined within the program that explicitly matches what would traditionally be defined as a width measurement. Instead, the two closest width data categories within QIA-64 were Mean Fibre Width and Inscribed Radius. Mean Fibre Width is an indirect width measurement based on a digital skeleton of the wire sample that is built by the program from the scanned image. However, the width of the wire skeleton may not be representative of the actual width of the sample due to variability in pixel darkness associated with the scanning process. Similarly, the Inscribed Radius measurement finds the largest circle that can be drawn within the wire feature (encompassing the width of the pixels comprising the wire sample). The program then calculates the radius of that circle, and converts that to an indirect measure of width.

The Mean Fibre Width measurements did a relatively poor job of reflecting the changes in wire width across the five standards used in this study, regardless of which method was used. Method 2 performed the best, though it only accurately measured four (10%) of the 40 possible wire types. For all three methods, only the thinnest wire diameters (Standard 5, 0.15 mm) were measured accurately. All other standards were underestimated, with the thickest sample (Standard 1, 0.65 mm) being the most underestimated. Regression analysis showed there was no difference in estimates of mean fibre width among methods, suggesting that neither feature selection nor the use of an alternative algorithm improved accuracy in comparison with the default.

We found significant orientation effects for Mean Fibre Width measurements of the straight wires (Layout A). This finding suggests that the angle that

each wire was placed on the card can impact Mean Fibre Width measurements. The Mean Fibre Width measurement takes an average of the skeleton width along the length of the feature, but as the wire changes its angle of orientation on the card the generated medial axis for the skeleton also shifts, resulting in an error in the construction of the skeleton by the program. These errors, associated with layout differences, were not seen when using the Inscribed Radius method. Additionally, we suspected that if the wire's angle orientation affected width measurements, then we would also expect that wire curvature would also impact width measurements. However, this was not the case as width measures using Fibre Width or Inscribed Radius were not impacted by wire curvature (Layout B). Regardless, these results leave little confidence that Fibre Width could be reliably used to determine if two wires of a similar size had a significantly different width.

Width measurements using the Inscribed Radius were more accurate than the Mean Fibre Width measurements. All Methods produced accurate estimates for either 16 (40.0 %) or 17 (42.5 %) of the 40 possible wire types. Inscribed Radius was expected to do a better job of showing trends in width because it takes measurements directly from the feature, and not from a skeletonized version of the feature. However, estimate accuracy was still < 50%. There was no clear effect of wire layout to explain the inaccuracies because the wires that were estimated accurately were a mix of straight and curved wires. This finding implies that, with respect to the Inscribed Radius, wire orientation and shape is only important for ensuring an accurate length, and has little effect on width measures.

The capacity of the program to accurately measure wire width was sporadic; standards 2 (0.51 mm), 4 (0.20 mm), and 5 (0.15 mm) were most commonly estimated accurately. Standard 1 (0.65 mm), the widest wire, had just two accurate estimates using Method 2 and 3, and none with Method 1. Standard 3 (0.36 mm) had no wires accurate estimate by any method. It is possible that another wire characteristic is responsible for affecting accuracy in width measurements, and that is the colour of the wire. Although we did not analyse the wires by colour, it is possible that colour may affect how dark they are when scanned and this may influence the thresholding process and generate different digital widths. In future the effects of sample colour should be tested, especially since many mammals have hair with bands of different colour along the same hair shaft.

Two other potential width measures from the program were briefly explored (but not included in this study): breadth and circumscribed radius. Circumscribed radius differs from Inscribed Radius, by finding the smallest circle that can completely contain the entire width of the wire, while Breadth attempts to measure a line that is drawn across the width of the wire. However, when the program draws this imaginary line it is "not necessarily perpendicular to the length of the wire" (i.e. not the actual width) and might introduce error because the line is drawn at an angle across the wire width thereby making the wire wider than it actually is. As a result, the actual breadth measurements included values that are far larger than the actual wire width. For example, one of the straight wires from Standard 5 (0.15 mm), the thinnest wire used, had a breadth measurement of 5.6 mm, considerably larger than the actual width.

The inability to identify an ideal way to measure wire width is puzzling because Davis (2010) seemed to have no problem determining hair width, even on non-straight hair samples. Davis (2010) explains that the width measurement used was one that took an "average width of the entire shaft". According to the QIA-64 manual, the only width measurement that did this was Mean Fibre Width. This study clearly shows that Mean Fibre Width is a rather poor indicator of actual wire width, and this raises some questions about the width measurements presented by Davis (2010). However, it is important to acknowledge that there are some differences between the three methods explored in this study and the image analysis technique introduced by Davis (2010). The exact wording used in the methods of Davis (2010) makes it difficult to be certain that Mean Fibre Width was used (as it was in this study), but it seems that Davis (2010) may have viewed width and Mean Fibre Width as equivalent terms. In his paper, Davis (2010) only uses the term width, which he defines as the average width over the entire shaft, a term that does not exactly match terms used by QIA-64 (formerly FoveaPro). Similarly, if Fratto & Davis (2011) also used Mean Fibre Width as an equivalent measure to width, it would call into question the significant hair width differences that they documented.

This study was designed to better understand the image analysis technique established by Davis (2010) and used by Fratto & Davis (2011). Our results only support using QIA-64 to measure the length of straight hairs with manually selected features after thresholding. The use of this software to measure width shows considerable error and does not seem

to provide reliable estimates. QIA-64 offers a full set of features that may be capable of refining this process into one that is more viable than what has been presented. However, it seems that the technique introduced by Davis (2010) is not, when used as described, capable of accurately and reliably measuring sample width. However, the core idea of the technique is still appealing because it offers a raid, automated way to study small changes or differences in hair morphology, once the difficulties identified in the present study are addressed.

Several other factors that may be important in the thresholding process were not investigated in this study, including sample colour, changes in width along the sample, and using stereo-microscope photographs of the sample to scan instead of scanning the hair or wire sample directly. Mammalian hair has a wide variety of colours among species, as well as colour differences between individuals of a particular species. Light and dark colouration, ranging from amelanistic (white) to melanistic (black) colouration, can be found within many mammal species (Caro 2005, 2009). Individuals often possess hair colour variation that is due to genetics (Graf et al. 2005), sex (Bradley & Mundy 2008), or age (Ross & Regan 2000). In addition, many mammals possess hair that changes colour along the shaft of individual hair follicles. For example, many mammals have agouti hair banding where hairs have distinctive light and dark stripes along the shaft (Nakayama et al. 2010) or have lighter or darker hair tips (Rhoad 1936). Even among humans there is some variation in individual hair colour on the scalp, and in some humans, heterochromia can be found (Restano et al. 2001).

Given the high frequency of variable hair colour among mammals there may be some effect of hair colour on the ability of automated programs, like the thresholding in QIA-64, to accurately measure the width of hair shafts. Davis (2010) did address this potential problem and suggested that his methodology can be altered to accommodate hairs that are lighter, darker, or have light/dark bands. Specifically, Davis (2010) recommended using an appropriately coloured background during the scanning phase (e.g. dark background for light hairs) and manually digitizing colour band sections from individual hairs. In fact, Davis (2010) used 20 deer hairs to manually isolate and measure the length of colour bands along hairs. With further validation of specific methods that can be used on different coloured hair, this technique should work on a variety of hair colours.

In addition to differences in colour, many mammals have hair that widens in the centre of the shaft and tapers at both ends (Pocock & Jennings 2006), which could lead to errors in width measurement if the location of measurement is not standardized. In response to this, Davis (2010) decided to measure the width along the entire shaft and produce an average width value for each hair. Hair can also be of different cross-sectional shapes (round *vs.* oval; Lasisi et al. 2016) which may lead to different amounts of shadowing during the scanning process. If different shadows are created during scanning this may produce errors in feature selection during thresholding. Currently the role that cross-sectional shape plays on digitizing a sample is unknown. One potential method that may reduce this error is using stereomicroscopy to generate a photograph that is then digitized for use with QIA-64.

Future studies should seek to improve the image analysis process that has been explored in this study. We believe that the technique outlined in Davis (2010) is worth pursuing, but that significant problems currently exist. Since this study has raised questions about the results presented by both Davis (2010) and Fratto & Davis (2011), we suggest that their results need further confirmation. More importantly, any research involving QIA-64 needs to include validation against samples of known length and width. Of the three methods investigated in this study, the manual selection of thresholding features offers the most useful refinement to the procedure presented by Davis (2010), particularly for improving length estimates. The angle that the straight wires were placed on the notecards had only a minor impact on the length estimates but wire curvature was shown to have a significant impact. Finally, all three hair methods failed to accurately estimate width, even when multiple data categories were examined.

Literature

Ahl A.S. 1986: The role of vibrissae in behavior: a status review. *Vet. Res. Commun. 10: 245–268.*

Armitage K.B. 2009: Fur color diversity in marmots. *Ethol. Ecol. Evol. 21: 183–194.*

Bejan A. 1990: Theory of heat transfer from a surface covered with hair. *J. Heat Trans.-T. ASME 112: 662–667.*

Bradley B.J. & Mundy N.I. 2008: The primate palette: the evolution of primate color. *Evol. Anthropol. 17: 97–111.*

Caro T. 2005: The adaptive significance of coloration in mammals. *BioScience 55: 125–136.*

Caro T. 2009: Contrasting coloration in terrestrial mammals. *Philos. Trans. R. Soc. Lond. B 364: 537–548.*

Davis A.K. 2010: A technique for rapidly quantifying mammal hair morphology for zoological research. *Folia Zool. 59: 87–92.*

De La Mettrie R., Saint-Léger D., Loussouarn G. et al. 2019: Shape variability and classification of human hair: a worldwide approach. *Hum. Biol*. *79: 265–281*.

Fratto M.A. & Davis A.K. 2011: Do black-furred animals compensate for high solar absorption with smaller hairs? A test with a polymorphic squirrel species. *Curr. Zool. 57: 731–736.*

Fujimoto A., Kimura R., Ohashi J. et al. 2008: A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum. Mol. Genet*. *17: 835–843.*

Glanville E.J. & Seebacher F. 2010: Advantage to lower body temperatures for a small mammal (*Rattus fuscipes*) experiencing chronic cold. *J. Mammal. 91: 1197–1204.*

Graf J., Hodgson R. & van Daal A. 2005: Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. *Hum. Mutat*. *25: 278–284*.

Hetem R., Witt B., Fick L. et al. 2009: Body temperature, thermoregulatory behavior and pelt characteristics of three colour morphs of springbok (*Antidorcas marsupialis*). *Comp. Biochem. Physiol. 152: 379–388.*

Hrdy D. 1973: Quantitative hair form variation in seven populations. *Am. J. Phys. Anthropol*. *39: 7–17*.

Lasisi T., Ito S., Wakamatsu K. & Shaw C.N. 2016: Quantifying variation in human scalp hair fiber shape and pigmentation. *Am. J. Phys. Anthropol*. *160: 341–352.*

Nakayama K., Shotake T., Takeneka O. & Ishida T. 2010: Variations in the coding region of the agouti signaling protein gene do not explain agouti/non-agouti phenotypes in macaques. *J. Mamm. Evol*. *17: 211–214.*

Olson T.A., Lucena C., Chase C.C. & Hammond A.C. 2003: Evidence of a major gene influencing hair length and heat tolerance in *Bos taurus* cattle. *J. Anim. Sci. 81: 80–90.*

Pocock M.J.O. & Jennings N. 2006: Use of hair tubes to survey for shrews: new methods for identification and quantification of abundance. *Mammal Rev. 36: 299–308.*

QIA-64 2016: Image analysis manual for Quantitative Image Analysis 64. *Reindeer Graphics, Asheville, North Carolina.*

Restano L., Barbareschi M., Cambiaghi S. et al. 2001: Heterochromia of the scalp hair: a result of pigmentary mosaicism? *J. Am. Acad. Dermatol*. *45: 136–139.*

Reynolds P.S. 1993: Effects of body size and fur on heat loss of collared lemmings, *Dicrostonyx groenlandicus*. *J. Mammal. 74: 291–303.*

Rhoad A.O. 1936: The silver gray color in Indian cattle. *J. Hered. 27: 113–118.*

Ross C. & Regan G. 2000: Allocare, predation risk, social structure and natal coat color in anthropoid primates. *Folia Primatol. 71: 67–76.*

Steudel K., Porter W.P. & Sher D. 1993: The biophysics of Bergmann's rule: a comparison of the effects of pelage and body size variation on metabolic rate. *Can. J. Zool. 72: 70–77.*

Walsberg G.E. 1983: Coat color and solar heat gain in animals. *BioScience 33: 88–91.*

Walsberg G.E. 1988: The significance of fur structure for solar heat gain in the rock squirrel, *Spermophilus variegatus*. *J. Exp. Biol. 138: 243–257.*

Walsberg G.E., Tracy R.L. & Hoffman T.C.M. 1997: Do metabolic responses to solar radiation scale directly with intensity of irradiance? *J. Exp. Biol. 200: 2115–2121.*

Walsberg G.E. & Wolf B.O. 1995: Effects of solar-radiation and wind-speed on metabolic heat-production by two mammals with contrasting coat colors. *J. Exp. Biol. 198: 1499–1507.*

Wilcox J.T. & Larsen B.N. 2009: A group defense incident involving juvenile striped skunks, *Mephitis mephitis*. *Can. Field-Nat. 122: 80–82.*