Technical Note: Comparing von Luschan Skin Color Tiles and Modern Spectrophotometry for Measuring Human Skin Pigmentation

Anna K. Swiatoniowski, Ellen E. Quillen, Mark D. Shriver, and Nina G. Jablonski*

Department of Anthropology, The Pennsylvania State University, State College, PA 16802

KEY WORDS reflectometry; conversion; M-index; skin color-matching; historical skin color data

ABSTRACT Prior to the introduction of reflectance spectrophotometry into anthropological field research during the 1950s, human skin color was most commonly classified by visual skin color matching using the von Luschan tiles, a set of 36 standardized, opaque glass tiles arranged in a chromatic scale. Our goal was to establish a conversion formula between the tile-based color matching method and modern reflectance spectrophotometry to make historical and contemporary data comparable. Skin pigmentation measurements were taken on the forehead, inner upper arms, and backs of the hands using both the tiles and a spectrophotometer on 246 participants showing a broad range of skin pigmentation. From these data, a second-order polynomial conversion formula was derived by jackknife analysis to estimate melanin index (M-index) based on tile values. This conversion formula provides a means for comparing modern data to von Luschan tile measurements recorded in historical reports. This is particularly important for populations now extinct, extirpated, or admixed for which tile-based measures of skin pigmentation are the only data available. Am J Phys Anthropol 151:325–330, 2013. © 2013 Wiley Periodicals, Inc.

The description and assessment of skin color has a long history dating back to the late 18th century (Smith, 1787). Prior to the introduction of skin reflectance spectrophotometry (reflectometry) to anthropological field research during the 1950s, the only methods available for quantifying skin color were based on visual matching with standardized, ordinal colors presented on strips of paper or glass tiles. The most widely used were the colored glass tiles developed by Austrian anthropologist, Felix von Luschan (Fig. 1). The goal of this study was the development of a conversion to make skin color data collected using the von Luschan tiles comparable to those gathered using modern reflectometry.

Originally called *Hautfarbertafel*, meaning skin color board or tablet, the von Luschan tiles are a set of 36 standardized, opaque, colored glass tiles that were created in the late 1800s as a visual matching reference scale for recording skin color. An early review of the von Luschan tiles provides a rare but brief description of how to use the tiles (Thomas, 1905). It suggests that the first "six [tiles were] for anemic Europeans, the remainder for all degrees of pigmentation of normal individuals, European or otherwise." However, no further details on method or suggested measurement locations were specified and available data from subsequent researchers indicates that the first six tiles were regularly used to assess nonanemic individuals. In the first half of the 20th century, von Luschan skin color data were collected throughout the world including Africa (e.g., Trevor, 1947; Wells, 1952), Europe (e.g., von Luschan and von Luschan, 1914; Searight et al., 1944), Australia and the South Pacific (e.g., Shapiro, 1933; Birdsell, 1967), the Americas (e.g., Stewart and Strong, 1939; Drusini and Tommaseo, 1981), and admixed populations (e.g., Day, 1932; Little, 1943). The von Luschan tiles provided a device for systematically measuring skin color, but their design presented a number of issues inherent in the tiles including surface imperfections and glare as well as challenges due to human observers such as reproducibility, variation in color perception, and variable lightening conditions (Searight et al., 1944; Harrison, 1957). Despite these limitations, the development of the tiles solved a long-recognized problem in anthropology (Thomas, 1905). With the invention and subsequent widespread use of reflectometers which minimized the impact of human error and lighting conditions, the tiles were largely abandoned (Weiner, 1951; Garn et al., 1956). However, the historical data collected using the von Luschan tiles are irreplaceable.

METHODS

Participants

Participants were recruited at The Pennsylvania State University in State College, PA, and Morehouse and Spelman Colleges in Atlanta, GA. Volunteers were excluded on the basis of pregnancy; being younger than 18 or older than 40; use of sunless tanner or tanning

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Nina Jablonski, Department of Anthropology, 409 Carpenter Building, The Pennsylvania State University, University Park, PA 16802, USA. E-mail: ngj2@psu.edu

Present address of Ellen E. Quillen: Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas, 78227

Received 16 October 2012; accepted 18 March 2013

DOI: 10.1002/ajpa.22274
Published online 30 April 2013 in Wiley Online Library (wileyonlinelibrary.com).

Fig. 1. The von Luschan tiles measure 8 cm by 10 cm. (**a**) Individual tiles are arranged in a semichromatic scale from 1 through 36 with the first 18, lighter colored tiles on one side and the second 18, darker colored tiles on the opposite side. (**b**) The tiles were carried in a protective metal case.

beds within the past 1 or 2 months, respectively; and any medical reason the subject's skin pigmentation might be aberrant (e.g., vitiligo, acanthosis nigricans, eczema, psoriasis). These criteria were meant to minimize pregnancy-, medical-, and age-related changes in skin pigmentation. A total of 246 individuals entered into the study: 188 in State College and 58 in Atlanta. Age and primary ethnic affiliation were self-reported: mean age = 23, African American/Black $N = 66$, Asian or Asian American $N = 14$, European American/White/Not Hispanic or Latino $N = 150$, Hispanic/Latino $N = 8$, "some other race" $N = 8$. All participants were recruited with implied consent under protocols approved by the institutional review boards at the three participating institutions.

Data collection

von Luschan tiles. Two independent observers (AKS and EEQ) matched the skin of each participant to the von Luschan tiles (observers used tiles 1–36 throughout the study) at the forehead, each inner upper arm, and the back of each hand. These locations were chosen to reflect both historical and contemporary skin color study protocols. The oldest studies using the von Luschan tiles generally measured the forehead and/or the backs of the hands; however, by the early 1920s, the inner upper arm became the most common. Other measurement locations used in reflectometry studies such as the back or buttocks were in not included as they were not used in von Luschan tile data collection to the best of our knowledge. As data collection took place at several locations, precautions were taken to maintain consistent lighting during each session and between sessions. Measurements using the von Luschan tiles were always taken with the participant and observer standing away from windows using overhead fluorescent lights. In total, 2,460 measurements using the tiles were taking including both observers and all measurement locations of the body.

Spectrophotometry. The DermaSpectrometer (DermaSpec; Cortex, Hadsund, Denmark) was used to measure reflectance of skin as previously described (Shriver and Parra, 2000). The DermaSpec, a narrow-band spectrophotometer, measures the reflectance of green and red light emitting diodes (LEDs). The red diode is centered on 655 nm and from this the DermaSpec computes the melanin index (M-index), a measure of the concentration of the brown/black pigment melanin. The presence of skin features such as birth marks, freckles, or significant amounts of hair were noted and avoided. All measurements were taken in triplicate and instruments were calibrated before each data collection session (1–4 h in length). The tiles were also measured directly using the DermaSpec to provide a quantitative measurement.

Statistical methods and conversion formulae. Paired *t* tests were used to determine interobserver agreement and for comparison between bilateral measurements. Conversion formulae between the von Luschan tiles and the DermaSpec M-index data were derived with jackknife (leave-one-out) linear and polynomial regressions across all body sites and for each site on the body independently using the "bootstrap" and "plyr" packages (Tibshirani, 2007) in R (2010).

RESULTS

Patterns of tile use

Consistent with previous studies, observers did not use all 36 tiles with equal frequency and some tiles were not used at all (Supporting Information Fig. s1). Many, but not all, tile numbers left unused occurred at the ends of the chromatic scale and the frequency of usage varied across the body. Boxplots showing the distribution of tile use by measurement location are shown in the Supporting Information (Fig. s2). Consecutive tiles are not consistent in chromatic order or degree of color difference, which often made matching difficult (Fig. 1). Tile 1 is a chalky and stark white, tile 2 is a pale yellow, and tile 6 is a dark, orange-like tan. However, tile 7 returns to a pale white and thereafter a general pattern of progressive darkening can be seen through tile 36. Certain tiles appear similar to each other whether consecutively numbered or not (for example, consecutive tiles 11 and 12 and nonconsecutive tiles 3 and 10).

When the tiles were measured directly with the DermaSpec, M-indices generally increase with tile number (Fig. 2). However, the degree and direction of change in M-index was not uniform among consecutive tiles. There is a marked difference in color range between the sides: the lighter half (tiles 1–18) had a range of 44 M-index units, while the darker half (tiles 19–36) had a range of 124 M-index units. Glass has different physical properties from skin so the M-indices obtained from measuring

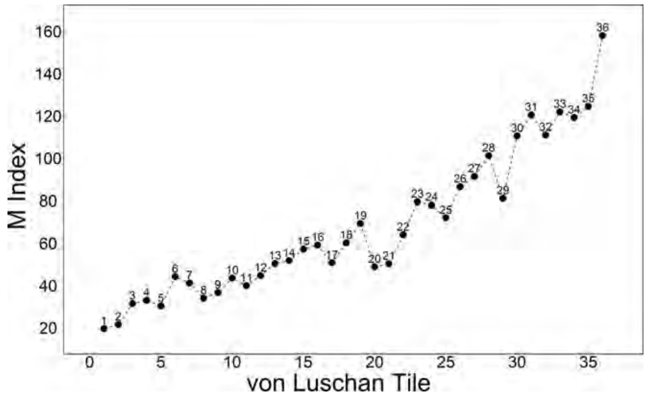

Fig. 2. M-index values of tiles when measured directly with the DermaSpec reflectometer.

the tiles directly with a reflectometer cannot be compared to those obtained from human skin.

Interobserver agreement

Paired t test of tile matching data from the two observers using all measurement sites combined shows an average interobserver difference of 0.95 ($H_a \neq 0$, $P < 2.2 \times 10^{-16}$), or less than one tile, with linear regression coefficients further supporting concordance ($\beta = 0.96$, $R^2 = 0.85$, $P < 0.001$). When the measurement locations were analyzed separately, results were similar with mean differences of 1.24 tiles for forehead, 0.58 tiles for arms, and 1.19 tiles for hands ($P << 0.05$ for all tests). Plots of observations are shown in Supporting Information Figure s3. The average difference between the left and right tile measurements recorded by the same observer is also non-zero for hands (diff = 0.15, $P = 0.0004$) and arms (diff = 0.14, $P = 0.015$). The average difference in M-index values are also significantly different from zero for arms (diff = 0.61, $P = 0.001$) but equal for hands ($P = 0.99$).

Tile numbers and participant M-indices

Both linear and polynomial regression models were considered in assessing the relationship between the tile and M-index measures. The most comprehensive conversion formula is based on the inclusion of all measurement locations to capture the broadest possible distribution of pigmentation levels as the inner upper arm values are lower, on average, than those of the more exposed hands and forehead (Supporting Information Fig. S2). This is particularly important at the darker end of the distribution where there were many fewer observations. The estimated bias derived from the jackknife analysis was less than 0.004 for all coefficients for both the linear and polynomial models. However, the second-order polynomial explained 88% of the variation in the sample compared to 82% of the variation explained by the linear model and the absolute difference between the predicted and observed M indices averaged 3.7 M units for the polynomial and 4.8 M units for the linear model (Supporting Information Table s1). Based on the full data set, the 95% confidence interval for the predicted M-indices spans 19.54 M units using the best conversion formula

$$M = 23.118 - 0.110T + 0.055T^2$$

where M is predicted M-index and T is tile value.

For reference, second-order polynomial regressions were also derived for forehead, arms, and hands separately (Fig. 3, Supporting Information Table s1). Based on the above conversion formula, Pearson's correlations were computed for observed M-index versus predicted M-index for the full data set ($\rho = 0.95$, $P < 2.2 \times 10^{-16}$) as well as each self-reported ethnic group: African American/Black ($\rho = 0.88$, $P < 2.2 \times 10^{-16}$), Asian/Asian American ($\rho = 0.84$, $P < 2.2 \times 10^{-16}$), European American/White ($\rho = 0.28$, $P < 2.2 \times 10^{-16}$), and Hispanic/Latino ($\rho = 0.70$, $P < 2.2 \times 10^{-16}$). The relatively low correlation between observed and predicted M-index values for European Americans is due to this group having the smallest variation in both tile and observed M-index values. Plots of the predicted and observed M-index values for each ethnic group can be seen in Supporting Information Figure s4.

Tile number to M-index conversion for historical data

To test the utility of the conversion formula, we applied it to von Luschan tile data from seven samples from six historical studies and compared the predicted M-indices to modern skin color data collected from similar populations using the DermaSpec. Where available, individual tile values were converted before averaging, but, for several studies, only averaged tile values were published and used for comparison. All studies measured the inner arm with the exception of Day (1932), who measured the forehead. Table 1 shows the averaged tile data, the corresponding predicted M-indices, and the DermaSpec M-index data for each of the seven paired data sets. The difference between the DermaSpec M-index and the predicted M-index values ranges from 0.4 to 13.4 M-index units with an average of 5.2 M-index units (Table 1, Fig. 4). For comparison, the mean test–retest difference for DermaSpec readings in our sample was 1.6 M units. The Pearson's correlation between the DermaSpec M-index data and predicted M-index data from similar populations is $\rho = 0.92$ ($P = 0.004$). The discordance is due in part to the differences between the populations used for the tile data compounded by changing concepts of self-identity over the past several decades (particularly with regard to the African-American samples). Additionally, some variance is introduced by the conversion itself. As expected, the admixed African-American and African-British samples collected by Day and Little have the largest distributions of both tile and M-index values (Fig. 4).

DISCUSSION

While our subject pool included a broad range of pigmentation, it does not represent all global skin color variation and tiles at either end of the von Luschan scale were used less frequently. The extremes of the scale have been used in historical studies including Searight et al. (1944) using tiles 1 and 2 with Scottish inhabitants of the Outer Hebrides and Birdsell (1967) using tiles 6 and tiles 30 to 35 with Northern Australian Aborigines. While the formula allows for conversion of tiles absent from our dataset, care should be taken when considering historical data including tiles that deviate from the basic

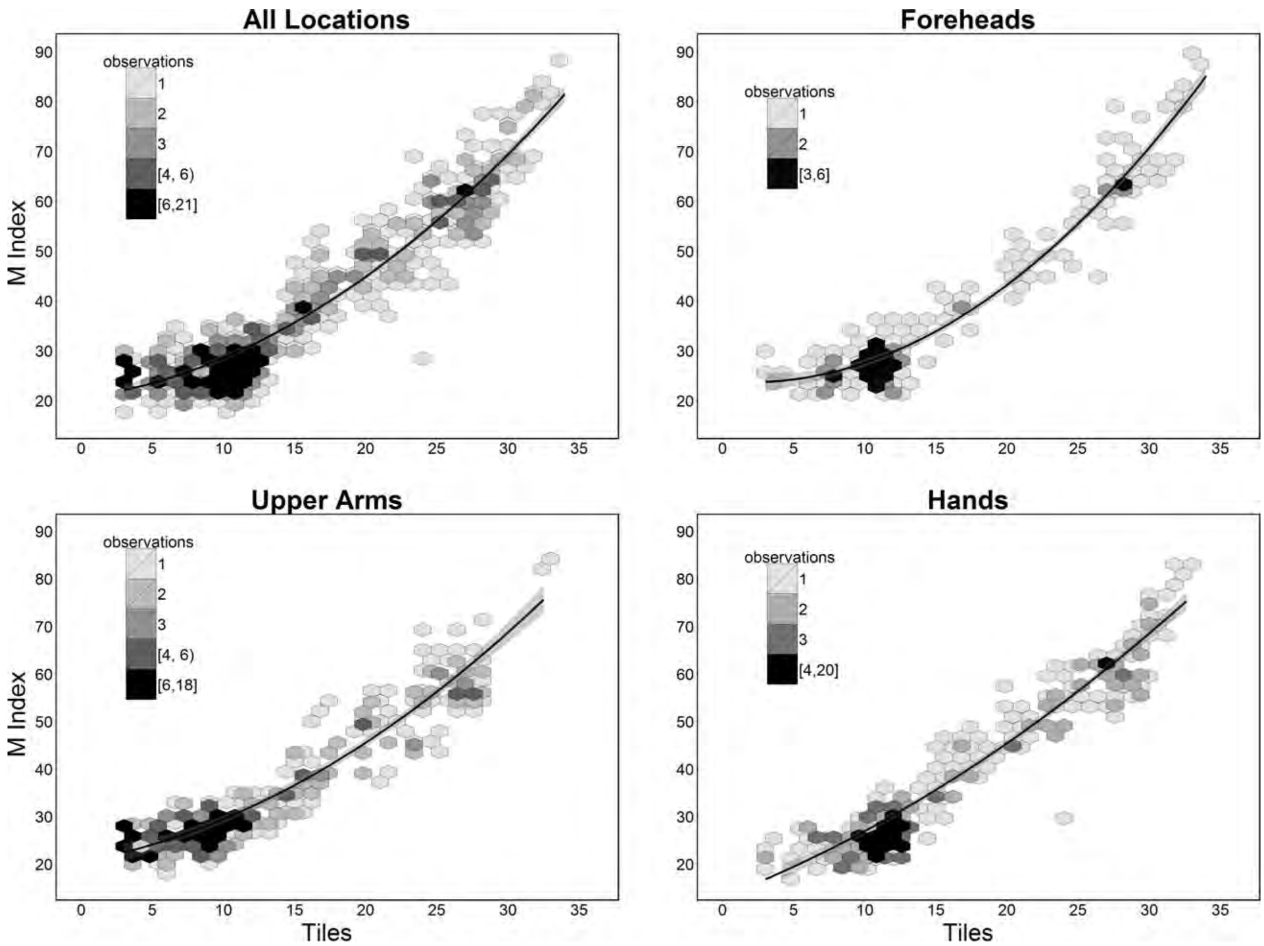

Fig. 3. M-indices corresponding to mean observed tile number per participant for all measurement locations, forehead, inner upper arms, and hands. Darker hexagons indicate more observations. The polynomial predicted values are plotted over the observed data with the 90% confidence interval of predicted values represented by shading.

TABLE 1. Comparison of DermaSpec M-index data to predicted M-index values in similar population samples

Von Luschan tiles					DermaSpec				
Study	Group	N	Tile mean ± SD	Predicted M ± SD	Study	Group	N	Mean ± SD	M-index difference
(a)	Papua New Guinea[a]	51	27.2 ± 2.1	61.0 ± 6.0	(g)	Papua New Guinea[a]	21	67.9 ± 9.4	6.9
(b)	Tamils, India	140	20.0 ± unkn.	42.9 ± unkn.	(h)	India and Pakistan	7	37.1 ± 4.2	5.8
(c)	Gaelic Scots	54	4.8 ± 2.6	24.2 ± 1.3	(i)	Irish, Dublin	146	26.4 ± 2.2	2.2
(d)	British Isles	95	7.5 ± 1.3	25.5 ± 1.0	(i)	Irish, Dublin	146	26.4 ± 2.2	0.9
(e)	"English" (White)	110	9.2 ± unkn.	26.8 ± unkn.	(i)	Irish, Dublin	146	26.4 ± 2.2	0.4
(e)	"F1 Anglo-Negro"	41	14.9 ± unkn.	33.7± unkn.	(j)	1/2 "West African"[b]	16	37.04 ± 6.6	2.3
(f)	"Negro"	460	13.4 ± 7.6	34.8 ± 12.1	(j)	African American	110	48.04 ± 9.9	13.4

Study Abbreviations: (a) Howells, 1933; (b) Bernhard, 1983; (c) Searight, 1944; (d) von Luschan and von Luschan, 1914; (e) Little, 1943; (f) Day, 1932; (g) Norton et al., 2006; (h) Shriver and Parra, 2000; (i) Candille et al., 2012; (j) Shriver et al., 2003 and Shriver (unpublished).
Unkn. = unknown due to limitations of data published (e.g. when only averages of tile data are available).
[a] PNG: New Britain island and the Sepik River and north eastern coastal regions of New Guinea.
[b] Determined from self-reported genealogical ancestry of subject and subject's parents and grandparents.

chromatic order or where the difference between adjacent tiles is considerable (e.g., the chromatic distance between tiles 6 and 7 and between tiles 18 and 19) (Figs. 1 and 2).

Additionally, it should be noted that this conversion formula includes all 36 tiles in contrast with the recommendations by Thomas (1905) and von Luschan and von Luschan (1914), who indicate only tiles 7–36 should be

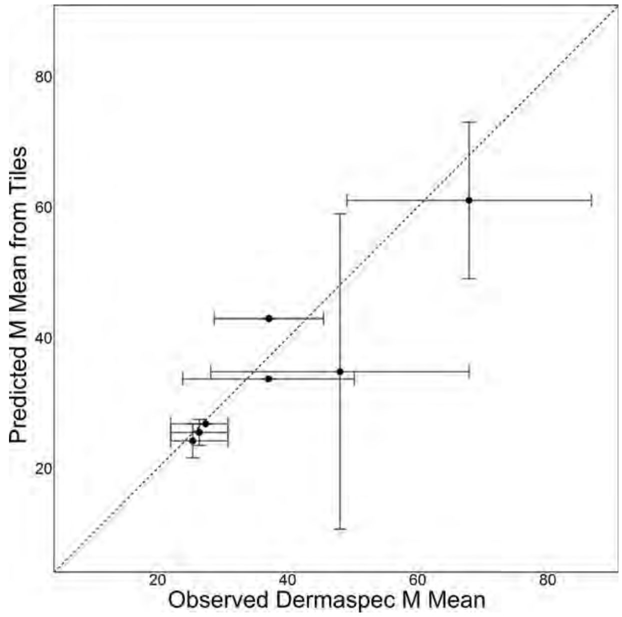

Fig. 4. Predicted mean M-indices from von Luschan tile data, derived using our conversion formula, plotted against mean M-index DermaSpec data in similar populations. Two standard deviations from the means are shown for each data pair. The identity line represents the hypothetical perfect concordance between the observed and predicted M-indices. Observed DermaSpec values have been offset by 1 M unit for the three instances of the Irish data to avoid overlap of whiskers. Data are shown in Table 1.

used on nonanemic individuals. However, we have not yet encountered any other researchers who mentioned adhering to the tiles 1–6 exception. Either these researchers did not know about von Luschan's intended protocol or they used all tiles regardless. As such, our protocol of using all 36 tiles in data collection provides a better approximation of the body of historical tile data.

Because the forehead, hands, and inner upper arms were all used, to different degrees, in studies from the late 19th century into the 1950s, we considered each of these three measurement sites. The means and distributions of M-indices and tiles vary across body sites (see Supporting Information Fig. s2), likely due to variation in sun exposure and anatomical differences in distribution and density of chromophores throughout the body. Conversion formulae were considered for each body site separately as well as for all sites combined and vary little with the exception of the formula based solely on the hand data. However, the formula derived from all measurement locations combined is the best predictor and is recommended for use in converting all data. The variation in formulae derived from the different measurement location is due to the different data distributions and not to actual differences in the quality of the skin at that data site. For example, there is no reason to expect that tile 18, when observed at the hand, represents a different concentration of melanin or different M-index value than when tile 18 is observed at the forehead.

Despite these limitations, the utility of this method is borne out by the greater than 90% concordance between M-index values predicted from historical tile data and modern reflectance measures from similar populations. The Italian anthropologist Renato Biasutti (1967; first published 1941) published a map of global variation in skin color from data collected using the von Luschan tiles dating back to the late 19th century. This map and the data used to create it are records of variation in skin color that existed before wide-spread globalization and cannot be measured again. Large-scale and long-distance migration of peoples from their ancestral environments has resulted in more opportunities for admixture, and the extirpation and extinction of many populations. Thus, the addition of historical data derived from the von Luschan tile-matching studies will add to the modern understanding of global skin color variation and will create opportunities to test and refine hypotheses about the evolution of skin pigmentation.

ACKNOWLEDGMENTS

We thank Genevieve Fisher, Registrar of the Peabody Museum of Archaeology and Ethnology of Harvard University for the loan of the von Luschan tiles, and India Spartz, Senior Archivist of the Peabody Museum for access to the Caroline Bond Day archive. Nina Jablonski thanks the Africana Research Center of The Pennsylvania State University for support her visit to Harvard University.

LITERATURE CITED

Bernhard W. 1983. Ethnogenesis of South Asia with special reference to India. Anthropol Anz 41:93–110.

Biasutti R. 1967. Le razze e I popoli della terra, 4th ed. Torino: Unione Tipografico.

Birdsell J. 1967. Preliminary data on the trihybrid origin of the Australian Aborigines. Archaeol Oceania 2:100–155.

Candille S, Absher DM, Beleza S, Bauchet M, McEvoy B, Garrison NA, Li, JZ, Myers RM, Barsh GS, Tang H, Shriver MD. 2012. Genome-wide association studies of quantitatively measured skin, hair, and eye pigmentation in four European populations. PLoS One 7:e48294.

Dawson JB, Barker DJ, Ellis DJ, Grassam E, Cotterill JA, Fisher GW, Feather JW. 1980. A theoretical and experimental study of light absorption and scattering by in vivo skin. Phys Med Biol 25:695–709.

Day CB. 1932. A study of some Negro-White families in the United States. Cambridge, MA: Peabody Museum of Harvard University.

Diffey BL, Oliver RJ, Farr PM. 1984. A portable instrument for quantifying erythema induced by ultraviolet radiation. Br J Dermatol 111:663–672.

Drusini A, Tommaseo M. 1981. Physical anthropology of Tarahumara Indians of northern Mexico. Anthropol Anz 39: 189–199.

Edwards EA, Duntley SQ. 1939. The pigments and color of living human skin. Am J Anat 65:1–33.

Garn SM, Selby S, Crawford MR. 1956. Skin reflectance studies in children and adults. Am J Phys Anthropol 14:101–117.

Harrison G. 1957. The measurement and inheritance of skin colour in man. Eugen Rev 49:73–76.

Howells W. 1933. Anthropometry and blood types in Fiji and the Solomon Islands. Anthropol Pap Am Mus Nat Hist 33:279–339.

Jablonski NG, Chaplin G. 2000. The evolution of human skin coloration. J Hum Evol 39:57–106.

Kollias N, Seo I, Bargo PR. 2010. Interpreting diffuse reflectance for in vivo skin reactions in terms of chromophores. J Biophotonics 3:15–24.

Kollias N, Stamatas GN. 2002. Optical non-invasive approaches to diagnosis of skin diseases. J Investig Dermatol Symp Proc 7:64–75.

Little K. 1943. Some anthropological characteristics of Anglo-Negro children. J R Anthropol Inst 73:57–73.

Minitab 16 Statistical Software. 2010. [Computer software]. State College, PA: Minitab, Inc. Available at: www.minitab.com.

Norton HL, Friedlaender JS, Merriwether DA, Koki G, Mgone CS, Shriver MD. 2006. Skin and hair pigmentation variation in Island Melanesia. Am J Phys Anthropol 130: 254–268.

R Development Core Team. 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rees JL. 2003. Genetics of hair and skin color. Annu Rev Genet 37:67–90.

Searight HF, Bathurst C, Noone R, Trevor J. 1944. A contribution to the anthropology of the Outer Hebrides. J R Anthropol Inst 74:25–32.

Shapiro H. 1933. The physical characteristics of the Ontong Javanese. Anthropol Pap Am Mus Nat Hist 33: 227–278.

Shriver MD, Parra EJ. 2000. Comparison of narrow-band reflectance spectroscopy and tristimulus colorimetry for measurements of skin and hair color in persons of different biological ancestry. Am J Phys Anthropol 112:17–27.

Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, Pfaff C, Jones C, Massac A, Cameron N, Baron A, Jackson T, Argyropoulous G, Jin L, Hoggart CJ, McKeugue PA, Kittles RA. 2003. Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 112:387–399.

Smith SS. 1787. An essay on the causes of the variety of complexion and figure in the human species. New York: Williams and Whiting.

Stamatas G, Zmudzka B, Kollias N, Beer JZ. 2004. Non-invasive measurements of skin pigmentation in situ. Pigment Cell Res 17:618–626.

Stewart T, Strong W. 1939. Anthropometric observations on the Eskimos and Indians of Labrador. In: Publications of the field museum of natural history. Anthropological series, Vol. 31. Chicago: Field Museum Press.

Thomas N. 1905. *Hautfarbentafel*. by von Luschan. Man 5:160.

Tibshirani R. 2007. Bootstrap: functions for the book "an introduction to the bootstrap." R package version 1.0-22.

Trevor J. 1947. The physical characters of the Sandawe. J R Anthropol Inst 77:61–78.

von Luschan F, von Luschan E. 1914. Anthropologische messungen an 95 Englandern. Z Ethnol 46:58–80.

Weiner J. 1951. A spectrophotometer or measurement of skin colour. Man 51:152–153.

Wells L. 1952. Physical measurements of Northern Bushmen. Man 52:53–56.