

## THE LATE PLEISTOCENE HUMAN SPECIES OF ISRAEL

### LES ESPÈCES HUMAINES DU PLÉISTOCÈNE SUPÉRIEUR D'ISRAËL

Milford H. WOLPOFF <sup>1</sup>, Sang-Hee LEE <sup>2</sup>

#### ABSTRACT

The human remains from the Late Pleistocene Mousterian sites in modern day Israel raised the issue of variation for the first time in the history of paleoanthropology. Their current interpretation is both problematic, in that the sources of their variation still remain unresolved, and historic, in that attempts at resolution reflect the currently accepted philosophy of the paleoanthropologists as strongly as they reflect the nature of the data. Today this philosophy can be seen in the penchant of some paleoanthropologists to define species taxa on the basis of consistent differences, no matter how minute. Here, we examine the question of whether the observed variation in the cranial remains from Amud, Qafzeh, Skhul, and Tabun reflects species differences. We try to refute a hypothesis of no difference, and suggest new approaches for examining this phylogenetic question. We report on the distribution of a testing statistic based on the standard error of the slope of regressions relating the measurements common to pairs of specimens. We show that this standard error test of the null hypothesis (STET) has the power to reject a null hypothesis for significant hominid taxonomic differences but does not reject the null hypothesis for the Israeli remains.

*Keywords:* Pleistocene human species, Levant, analysis of variation, STET.

#### RÉSUMÉ

*Pour la première fois en Paléanthropologie, des restes humains provenant de sites d'une même zone géographique peuvent permettre d'étudier l'influence de la variation (crânienne) pour discuter leur statut phylogénétique. Il s'agit des spécimens du Pléistocène supérieur découverts dans les sites moustériens d'Israël. L'interprétation taxonomique de ces vestiges est problématique. En effet, les raisons de cette variation sont toujours inconnues et les tentatives d'explication de ces différences morphométriques reflètent souvent les idées phylogénétiques des chercheurs aussi nettement que la nature des données. Aujourd'hui, cela se traduit par le penchant des paléanthropologues à définir des espèces fossiles sur la base de différences anatomiques. Notre contribution teste l'hypothèse selon laquelle la variation observée entre les restes crâniens d'Amud, Qafzeh, Skhul et Tabun traduit bien des différences d'ordre spécifique. Nous allons essayer de rejeter l'hypothèse de l'absence de*

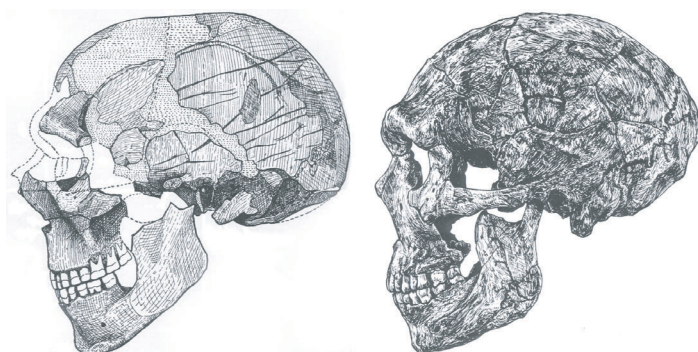
- 
1. Department of Anthropology, University of Michigan, Ann Arbor, MI, USA 48109-1382.
  2. Department of Anthropology, University of California at Riverside, Riverside, CA, USA 92521-0418.

*différence en proposant de nouvelles approches statistiques pour étudier cette problématique phylogénétique. Nous analysons la distribution de la courbe de régression d'une série de mesures communes à des paires de spécimens. Puis nous déterminons le STET (erreur standard de la courbe de régression) à partir de la modification d'une méthode proposée par Thackeray et al. (1997). Nous montrons que ce test (STET) permet de rejeter l'hypothèse lorsque l'on utilise des spécimens présentant de fortes différences taxonomiques mais ne le permet pas lorsque l'on ne considère que les fossiles israéliens.*

Mots-clés : espèces, Pléistocène, Proche-Orient, variation, analyse, statistique, STET.

## INTRODUCTION

Mousterian-associated Late Pleistocene human remains from the Levant of Western Asia provide a unique opportunity to examine the taxonomic implications of human anatomical variation. The earliest large sample is from the Mount Carmel caves, and its analysis first raised the question of whether two distinct species of humans at different stages of evolutionary development, samples from an evolving human clade, or two populations mixed from different regions, accounted for its variation (McCown and Keith, 1939; Dobzhansky, 1944; Thoma, 1962; Howells, 1973; Gould, 1988). The first description of the Mount Carmel remains from Skhul and Tabun (McCown and Keith, 1939) suggested the presence of only one type of human, represented by a wide range of variation, which the authors thought reflected a species "in the throes of evolutionary change." T. McCown and A. Keith concluded that the variability was due to the sampling of this evolving lineage. Much more extreme views have emerged since, and it has been suggested that the variation at Skhul itself reflects different human species (Schwartz and Tattersall, 2000; Tattersall and Schwartz, 1999). A middle view, widely but not universally held, is that the human remains from these caves, and Amud and Kebara, can be pigeonholed into categories of "Neandertal" and "early Modern Human". This often is taken to imply that these categories denote species differences, whether or not they are meant to do so (*fig. 1*).



*Fig. 1 - In many ways, the issue of human variation in Levant crania from the Late Pleistocene is illustrated in this comparison of the "early Modern Human" male Skhul 4 (left, after McCown and Keith, 1939, fig. 194) and Amud the Neandertal male (right, from Larsen et al. 1991). Can we disprove the hypothesis they are males of the same species?*

The explanation of variability is the fundamental problem of all biology. There is, of course, significant variation among the Late Pleistocene Levant human sample. The problem is to explain it. The issue in the Levant case is about the question of when variability can be explained by taxonomy. This is not an insubstantial question, since variation at or above the species level exists in protected gene pools that can be mixed up with each other but cannot mix together, while below the species level mixture is expected and normal, and its effect on the pattern and magnitude of variation is significant.

In this paper we examine the issue of when variation can validly be attributed to taxonomy. We question whether the magnitude and pattern of variation disprove the contention that there is a single species in the Israeli crania from the Late Pleistocene Levant human sample; indeed, whether the combined cranial sample from this region is unduly or unusually variable in comparison with other hominid populations.

## MATERIALS AND METHODS

### Null hypothesis approach

While differences within species can be quite large, it is well known that differences between species can be subtle and minimal, as shown by sibling species (Mayr, 1963), cases of mimicry (Jiggins *et al.*, 2001) and certain other cases of closely related groups (Kimbel and Martin (eds), 1993). One exacerbating aspect of this Late Pleistocene Levant sample is that the similarities between the specimens are not so small, nor are the differences so great, as to allow an obvious interpretation of whether interspecific or intraspecific variation is represented. Moreover, it is unlikely that any test can be expected to provide a statistic to ascertain whether two specimens represent a single species or two. However we can expect to develop tests of the null hypothesis, that a sample of specimens that could belong to the same species actually do so, based on the magnitude and pattern of variation *within* that species. The failure of such tests to provide valid results because of hypothetical differences *between* species that cannot be quantified or observed is beyond the capacity of any statistic to address <sup>3</sup>.

---

3. Discussing this issue, L.C. Aiello *et al.* (2000) suggest that a method of multiple working hypotheses (Chamberlain, 1965) be used instead of trying to reject the null hypothesis. Apart from the epistemological issues, it is challenging to apply such a method when the difficulties of assessing multiple species hypotheses are considered (Milius, 2001).

### Standard error of the bivariate slope

A new test of the null hypothesis was proposed by J.F. Thackeray *et al.* (1995, 1997). They quantified their expectations for conspecific variation by examining the standard error of the LMS regression slope in a series of pairwise comparisons of linear measurements. In these comparisons each specimen is plotted against another with the measurement values of one acting as the x-axis coordinates and the values of the same measurements of the other acting as the y-axis coordinates. It is the dispersion of variables around the regression line that is important for this test, not the slope of the line itself. This modification of Q-mode analysis compares specimens rather than measurements, similar to an approach suggested by C.O. Lovejoy (1979). It was developed and tested on a large sample of 1260 specimens representing 70 extant vertebrate and invertebrate taxa (Thackeray *et al.*, 1995, 1997). Ten disperse measurements of the cranium and mandible were used in an exact randomization approach, comparing males to females known to be within the same species within each of the 70 taxa. The slope for each comparison was calculated, along with its standard error. The test statistic these studies examined is the log of the standard error of the slope.

The standard error of the slope is a measure of the dispersion around the regression line, which is the observation of interest.

“High  $s.e._m$  values relate to high morphological variability when measurements of any two specimens are compared, reflected also by a high degree of scatter of measurements around a regression line. Relatively low  $s.e._m$  values can be expected in situations where there is only a small degree of morphological difference, associated with limited scatter around a regression line, reflecting similarities in shape of two specimens being compared” (Thackeray *et al.*, 1997, p. 196).

The log of the standard error was used as the test statistic because it has a normal distribution for the distribution of values for all 1260 specimens. This statistic reflects the consequences of variation in both size and shape, in that it shows “both geometric and allometric shape similarities between crania rather than just geometric shape similarity” (Aiello *et al.*, 2000, p. 180). Allometry is a function of size.

The mean log of the standard error for the linear regression slope was reported to be  $-1.78$  for the 70 species reference sample, with a standard deviation of  $0.27$ . J.F. Thackeray *et al.* (1997) found that the interval of  $\pm 2$  standard deviations around the mean ( $-1.24$  to  $-2.32$ ) encompasses 95% of the logs of the standard error for the linear regression slope for pairwise comparisons within the 70 species they studied. Subsequent analysis (Aiello *et al.*, 2000) based on 20 measurements resulted in a 95% upper limit for 15 specimens of *Pan* of  $-1.32$ , and a 95% upper limit for 8 non-human primate species (including *Pan*) with sample sizes ranging between 8 and 24, of  $-1.05$ .

Strictly speaking, the lower (more negative) confidence intervals are not relevant since pairwise specimen comparisons with even smaller values are none-the-less within the same species. The confidence interval determination is one-sided, and therefore all the figures given in these papers are somewhat too large—they actually represent the 97.5% confidence interval. A more serious problem, discussed below, lies in the small number of comparisons used for their determinations.

### STET

We have modified the test statistic proposed by J.F. Thackeray and colleagues, and the procedure for using it. There are a number of reasons for this. The argument for using a log transform was not compelling, and we did not want to dampen the effects of larger values of our test statistic without sufficient reason. The test procedure did not take advantage of the significant potential of the approach to compare samples of specimens with uncertain or unknown sex (unless sex determination was the object of the comparison, as in J.F. Thackeray *et al.*, 2000). The original testing was based on the same set of 10 measurements in 70 species, and subsequent comparisons with those results with pairwise analyses of different sets of measurements as dictated by the state of preservation of different specimens (Thackeray *et al.*, 2000; Aiello *et al.*, 2000), raise problems of sample size that must be addressed. Our modifications respond to these concerns, and to the issues raised in the discussion of error detailed below. We call our modified approach the **ST**andard **E**rror **T**est of the null hypothesis of no difference <sup>4</sup> -**STET**.

### Calculation of STET

The linear regression analysis used by J.F. Thackeray and colleagues minimizes the deviation of the dependent variable from the regression line. For cases where the bivariate sample is not symmetric around a linear regression line, the regression of X on Y differs from the regression of Y on X, and the standards errors of the regression slopes differ as well. In our analysis we have no *a priori* reason to choose independent and dependent variables in the pairwise comparisons <sup>5</sup>, but quickly recognized that this choice has some influence on the standard errors for the linear regression slopes that we

---

4. Because it is based on the expected distribution under conspecificity, STET can only be used to refute the hypothesis of no difference. The fundamental precept is asymmetrical: sufficient difference can refute the same species hypothesis but similarity cannot refute the hypothesis of different species.

5. J.F. Thackeray and colleagues always did regressions of females on males.

calculated. One solution to this problem would be to calculate the reduced major axis regression, which minimizes the orthogonal distance of each point from the regression line instead of minimizing the distance along the X axis or the Y axis (the orthogonal distance is the square root of the sum of the squares of the X axis and Y axis deviations). The disadvantage of a reduced major axis approach is that there is no direct way to calculate the standard error of its slope (Sokal and Rohlf, 1981). For this reason we chose a different solution, calculating standard errors of the mean for each comparison ( $s.e._{mx}$  for the linear regression of X on Y and  $s.e._{my}$  for Y on X) and reporting combined value as the square root of the sum of the squares of the two. One could think of STET as a hypotenuse joining the sides of a triangle determined by the two orthogonal standard errors.

$$(1) \text{ STET} = 100[(s.e._{mx})^2 + (s.e._{my})^2]^{1/2}$$

This test statistic is not directly comparable to the standard error based statistics published by J.F. Thackeray *et al.* (1995, 1997), and L.C. Aiello *et al.* (2000). This may not be a problem, however, because as noted below, there are compelling reasons not to make such comparisons.

### Error related to sample size

We were concerned about the influence of sample size on these comparisons, how does sample size influence the magnitude of STET? Such an influence is suggested by the differences in the 95% intervals calculated for the two studies cited above: 70 species using 10 measurements had an upper limit of  $-1.24$ , but 8 species using 20 measurements had an upper limit of  $-1.05$ . One would have expected the larger number of species to encompass a broader range of variation, but it did not. The opposite was the case. We were also concerned because our procedure has the potential for a different number of measurements in every pairwise comparison <sup>6</sup>.

The issue of whether the number of measurements was important was approached by calculating STET values, and serializing pairwise plots of the specimens in order of the values for this statistic. We plotted a sample of 311 pairwise comparisons between crania covering the period from the Pliocene to the “modern” Skhul/Qafzeh sample. For the most part a visual serialization of the dispersion in these plots closely fit the ordering based on STET. The smaller the value for STET, the closer the fit to a straight line. Only a few of the pairwise comparisons did not seriate with the STET values. They stood out as being either much less or much more disperse than their neighbors in the ordering (*fig. 2*). In every case the sample size for these exceptions was less than 40.

---

6. This was not a problem in earlier studies, where the number of measurements was held to a small but constant value.

To further specify the influence of sample size, we plotted the STET values as a function of the number of measurements underlying each comparison, and calculated a LMS regression. The effect of sample size becomes random when the sample size gets larger, but is evident in smaller samples whose values for the statistic were sample size dependent. To find when the sample size is large enough to avoid this effect, we examined the residuals from the regression slopes. For the comparisons of specimens with 40 or more measurements we could discern no relationship between the number of measurements and the size of the residuals. Moreover, none of the residuals exceeded 0.8 standard deviations from the expected value. Therefore, we do not report comparisons for specimens with fewer than 40 measurements in common.

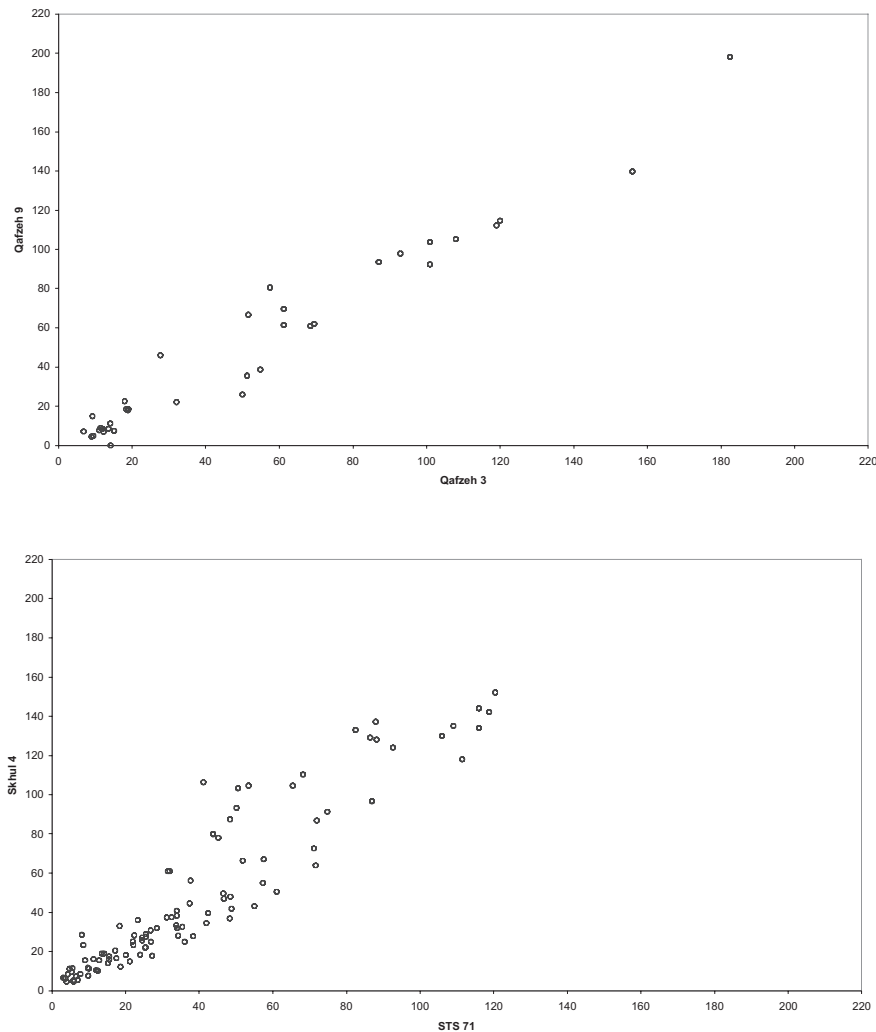


Fig. 2 - Comparison of an insufficient number of measurements (36) in the pairwise plot of Qafzeh 9 and Qafzeh 3 (above) and a large number (101) in the plot of Skhul 4 and STS 71 (below).

The STET values are the same, 5.1 (tables 1 and 3), but the patterns of diversity are quite different. This illustrates the effects of sample size sensitivity.

### Procedure

STET is calculated from pairwise comparisons of crania described by up to 290 homologous linear measurements systematically taken on the original specimens by one of the authors (MHW) over the course of several decades. The actual measurements used, of course, depend on the preservation of the two specimens. These data were recorded for all Plio/Pleistocene hominids older than 25 kya, and include measurements of all the human remains from the Levant. The size of the measurement set reflects the standardized measurements from R. Martin, the Biometrika school, W.W. Howells data set, and other normally used sources, and additional measurements developed to allow comparisons of fragmentary cranial remains too incomplete for standard measurements to be possible. This provides a much larger measurement base, and one particularly designed to maximize comparisons of specimens that are not complete. Our pairwise comparisons were made without regard to the sex of the individuals, so specimens could be included whether or not sex was known, or estimated. While incomplete specimens were used, in all of the cases considered, most (or all) of the vault and at least part of the face was present. We did not consider pairwise comparisons based on single cranial bones, or on specimens missing entire portions (such as the anterior portion missing from MLD 37/38 or the posterior missing from Zuttiyeh). No direct dental measurements were included, although measurements along the palate defined by tooth positions were in the sample pool. We also did not include mandibular measurements as only a few of the crania have associated mandibles.

### THE LEVANT QUESTION

We use STET to test a null hypothesis in a sample reflecting past human variation. The question is whether some of this variation has a phylogenetic basis: are there different human species in the Late Pleistocene of the Levant? The dispersion around the bivariate slopes addresses two issues about variation in size and shape: are the Levant “Neandertals” more different from the Skhul/Qafzeh specimens than the Skhul/Qafzeh specimens differ from each other, and is the combined Levant sample unusually variable in comparison with other single-species samples?

There is good reason to believe STET has the resolution to address this issue. As we noted, STET values are excellent predictors of dispersion for pairwise comparisons (see *tabl. 1–3* for specific values). A low value for STET such as found in the comparison of Skhul 5 and Qafzeh 6 (*fig. 3*) is reflected in a linear array of the measurements for the two specimens that shows little dispersion around a straight line. Marked dispersion is seen at the other extreme, for instance as in the comparison of specimens from two different

hominid genera. The comparison of Skhul 4 and STS 71 (*fig. 2*) has a much higher STET value than even the Levant cranial comparison with the highest value for STET (*fig. 4*).

|          | Skhul 5 | Qafzeh 6 | Skhul 4 | Qafzeh 9 | Qafzeh 3 | Skhul 9 | Amud | Tabun |
|----------|---------|----------|---------|----------|----------|---------|------|-------|
| Skhul 5  |         | 173      | 97      | 128      | 40       | 40      | 166  | 125   |
| Qafzeh 6 | 1.13    |          | 100     | 123      | 41       | 46      | 175  | 129   |
| Skhul 4  | 1.35    | 1.56     |         | 76       | *        | *       | 103  | 75    |
| Qafzeh 9 | 1.84    | 1.56     | 2.27    |          | *        | 42      | 128  | 97    |
| Qafzeh 3 | 2.75    | 2.86     | *       | *        |          | *       | 45   | 40    |
| Skhul 9  | 3.28    | 2.83     | *       | 4.30     | *        |         | 51   | 49    |
|          |         |          |         |          |          |         |      |       |
| Amud     | 1.28    | 1.35     | 1.72    | 2.06     | 2.56     | 2.58    |      | 138   |
| Tabun    | 1.98    | 1.56     | 1.98    | 2.69     | 2.16     | 3.41    | 1.98 |       |

|         | STS 71 | TM 1511 | STS 5 | STS 19 | STW 505 | STS 25 |
|---------|--------|---------|-------|--------|---------|--------|
| STS 71  |        | 70      | 207   | 42     | 93      | 41     |
| TM 1511 | 1.34   |         | 85    | *      | *       | *      |
| STS 5   | 2.19   | 2.48    |       | 59     | 102     | 47     |
| STS 19  | 2.76   | *       | 2.83  |        | *       | *      |
| STW 505 | 2.30   | *       | 3.25  | *      |         | *      |
| STS 25  | 5.54   | *       | 5.62  | *      | *       |        |

*Tabl. 1 - Pairwise comparisons<sup>1</sup> of Levant specimens STET values.*

*Tabl. 2 - Comparisons<sup>1</sup> of Sterkfontein specimens STET values.*

1. Comparisons in the lower left portion of the table show STET, the upper right comparisons show the sample sizes for each comparison. An asterisk means the comparison is based on an insufficient number of measurements.

|         | Skhul 5    | Qafzeh 6   | Qafzeh 9                | Skhul 9   | Qafzeh 3  | Skhul 4    |
|---------|------------|------------|-------------------------|-----------|-----------|------------|
| STS 19  | 3.58 (59)  | *          | *                       | *         | *         | *          |
| TM 1511 | 4.56 (74)  | 4.94 (57)  | 7.28 (40)               | *         | *         | 6.02 (46)  |
| STS 5   | 3.61 (215) | 3.86 (196) | 4.99 (140)              | 4.72 (57) | 4.39 (54) | 4.56 (102) |
| STS 71  | 4.72 (171) | 4.89 (179) | 6.16 (131)              | 9.59 (55) | 4.12 (53) | 5.08 (101) |
| STW 505 | 6.94 (70)  | 6.68 (84)  | 8.80 (55)               | *         | *         | *          |
| STS 25  | 10.87 (45) | *          | 13.15 (39) <sup>2</sup> | *         | *         | *          |

*Tabl. 3 - Comparisons<sup>1</sup> of Sterkfontein with Skhul and Qafzeh specimens STET values.*

1. Comparisons show STET, with the sample sizes for each comparison in parentheses. An asterisk means the comparison is based on an insufficient number of measurements.
2. Although this is just below the cut-off point we are using, the value is consistent with other STS 25 values.

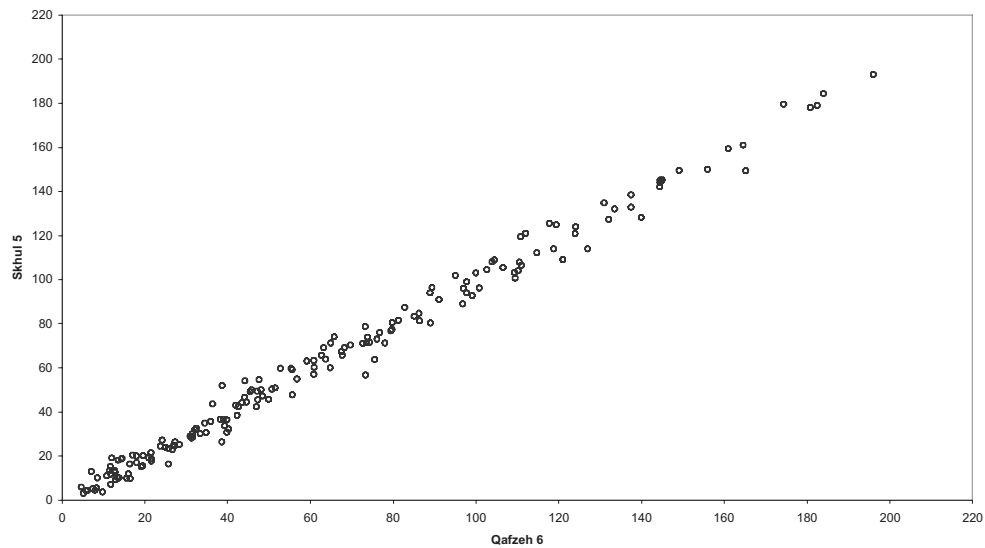


Fig. 3 - Comparison of Skhul 5 and Qafzeh 6 for 173 measurements, two specimens which are visually very similar and widely regarded as representing local populations of the same taxon. STET is 1.13 (tabl. 1), the smallest value for the Levant sample comparisons. There is little dispersion around a straight line in this pairwise plot.

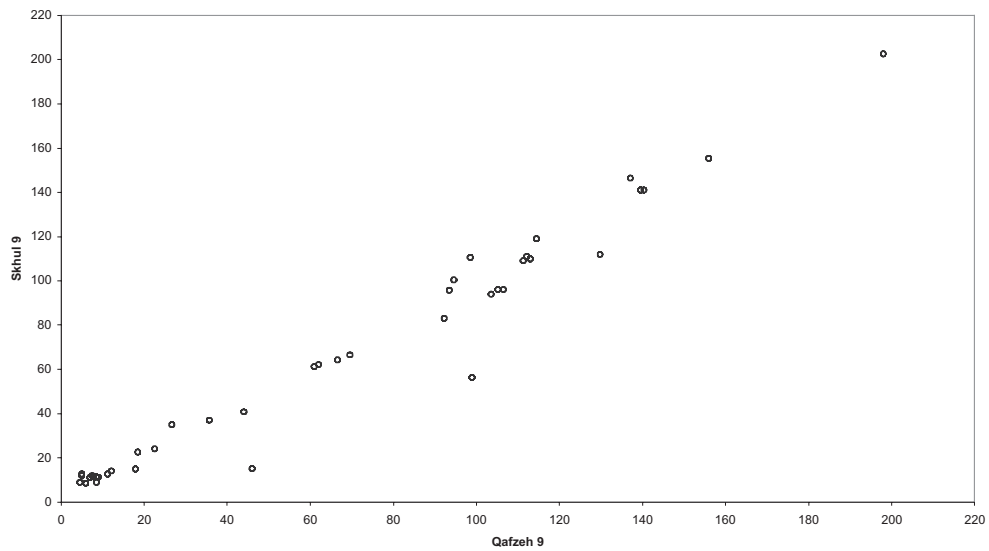
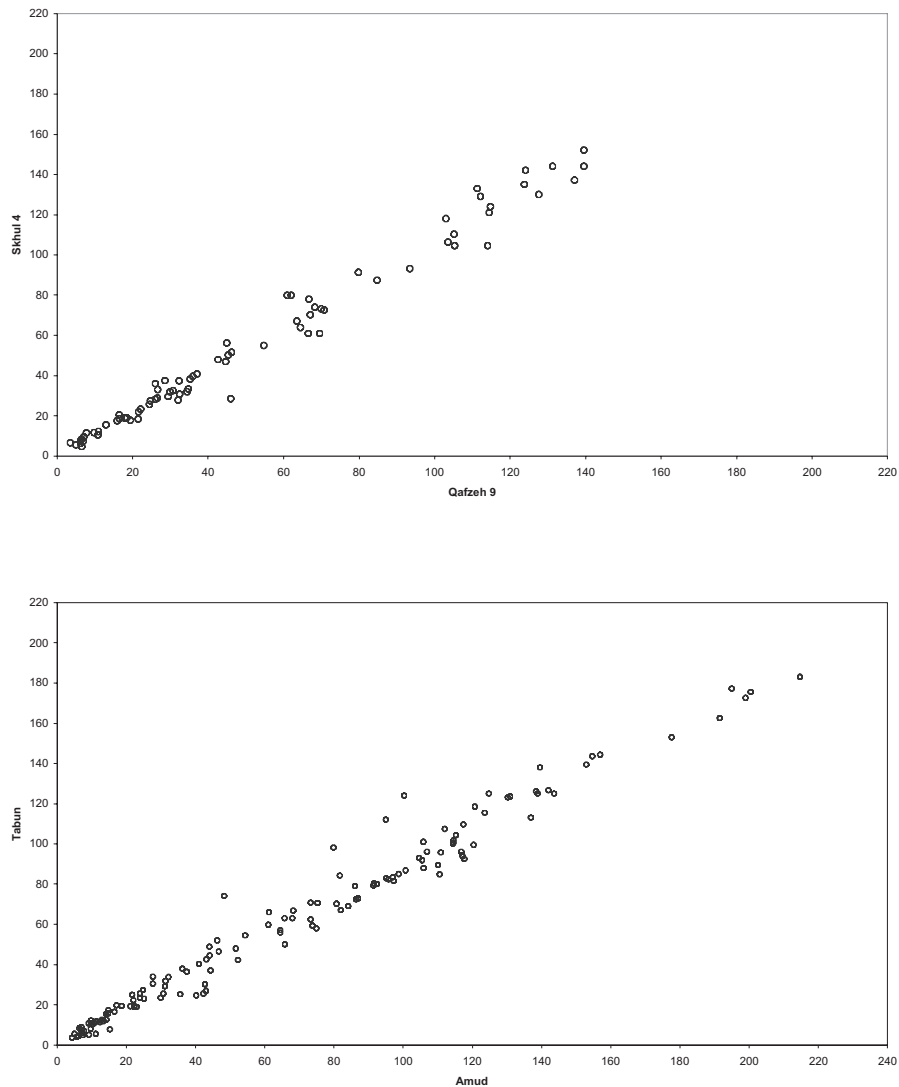


Fig. 4 - Comparison of Skhul 9 and Qafzeh 9 for 42 measurements, the two specimens whose dispersion around the linear slope is maximum for the Skhul/Qafzeh sample. STET is 4.30.

### Differences within the Levant sample

The question of the sources of variation is approached in two different ways. The first addresses differences in the Levant crania: are the Levant Neandertals (Amud and Tabun) more different from the other Levant specimens from Skhul and Qafzeh than the Skhul/Qafzeh remains are from each other (*fig. 5*)? Because species are protected gene pools, such a difference is an expected result if two different species are sampled.



*Fig. 5 - Above is the comparison of Skhul 4 and Qafzeh 9 for 76 measurements, two specimens with a dispersion around the linear slope that is just about average for the Skhul/Qafzeh sample. STET is 2.27. Below is the comparison of Tabun and Amud for 138 measurements. STET is 1.98.*

Fig. 6 shows that the pairwise comparisons of STET values within the Skhul/Qafzeh sample both encompass a broader range and include specimens with more dispersion around the bivariate slope than the pairwise comparisons between the “Neandertals” and the Skhul/Qafzeh remains (also see *tabl. 1*). STET values for all combinations of the Skhul/Qafzeh crania are generally larger than STET values for the “Neandertal” and Skhul/Qafzeh comparisons. We conclude that the Levant “Neandertals” have more similarities in size and shape with the Skhul/Qafzeh crania than the Skhul/Qafzeh crania have similarities to each other. The null hypothesis cannot be rejected.

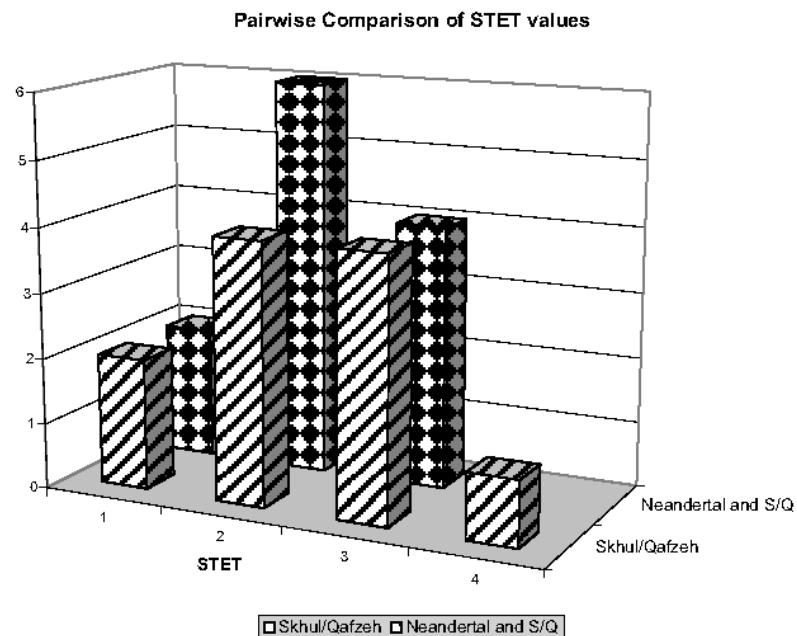


Fig. 6 - The figure shows two different STET comparisons for the Skhul/Qafzeh sample. One of these compares all of the Skhul/Qafzeh specimens to each other, and the other compares each of the two Levantine “Neandertals” with each of the Skhul/Qafzeh specimens. Data are from table 1. On average, the individuals within the Skhul/Qafzeh sample are less like each other than the two Levant “Neandertals” are like the Skhul/Qafzeh specimens, and some of the comparisons between Skhul/Qafzeh specimens are more disperse than any of the comparisons between the Levant “Neandertals” and Skhul/Qafzeh.

### Variation compared

If the Levant sample is comprised of two species, we might expect its variation to resemble other samples comprised of a mixture of species, and be unlike other samples of a single species. To further examine this expectation we used STET in an exact

randomization of pairwise comparisons for the *Australopithecus africanus* crania from the Sterkfontein site and for all the Levant crania (“Neandertals” and Skhul/Qafzeh). We then did a pairwise comparison of this Levant sample with the Sterkfontein sample, to establish the characteristics of a sample with mixed taxa.

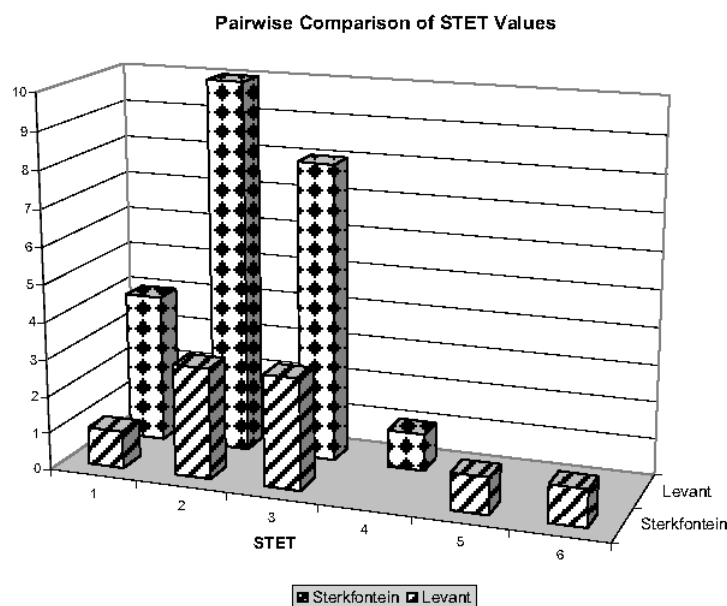
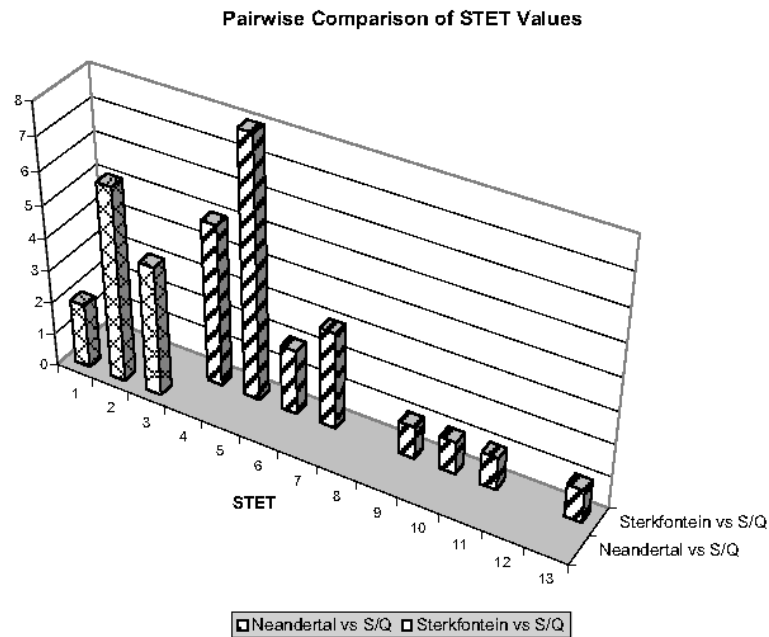


Fig. 7 - The exact randomization of the Levant pairwise distribution, including the Neandertals, compared with the exact randomization of pairwise values for the single Sterkfontein species. To be conservative, STS 19, STW 53 and STW 252 are not included because some authors, although not the present ones, believe they are different taxa. This still leaves a substantial sample size. Sterkfontein has a wider range and more pairwise comparisons that are disperse, compared with the Levantines. The lesser range of variation in these comparisons is not unexpected, under the hypothesis that the Levant sample is a single species, because of the narrower time range these specimens reflect. The comparison does not support the contention that the Levant sample is a mixture of species. See fig. 8 for a comparison of the dispersions within the Levant comparisons with a sample that combines different species.

Sterkfontein crania are more variable in their dispersion around the regression slopes (fig. 7). The mean STET for Sterkfontein is higher than STET for the Levant crania (compare tabl. 1 and 2) and the Sterkfontein sample includes pairwise comparisons with more dispersion around the regression slope than any pairwise comparisons within the Levant sample. Thus we find more size and shape similarities within the Levant sample that mixes “Neandertals” and “early moderns” than we find within the Sterkfontein sample that is drawn from a single species.

When a sample of Neandertals paired with Skhul/Qafzeh is compared with Sterkfontein paired with Skhul/Qafzeh (*fig. 8*), the differences are quite clear and the distributions do not even overlap. The magnitude and pattern of variation in the dispersion statistic unequivocally identifies the sample that mixes two species, and reveals different patterns of relationship in size and shape. One might argue that the putative Levant species are more similar than the two species in this comparison, even that they are so similar that they cannot be clearly distinguished. Perhaps so, but if there is no way to distinguish species differences in a mixed sample, the null hypothesis cannot be disproved for it.



*Fig. 8 - STET distribution for the two elements of the Levant sample compared with a sample known to be comprised of two different species. Here the Skhul/Qafzeh pairwise comparisons with the Levant "Neandertals" are shown along with the Skhul/Qafzeh pairwise comparisons with the Sterkfontein sample of Australopithecus africanus. The ranges do not overlap.*

## DISCUSSION

The Levant sample of human crania from Amud, Qafzeh, Skhul, and Tabun vary, that is not the issue. The question is *how much* they differ, and *why*. The specimens from these sites are thought to represent a single species by some <sup>7</sup> (Arensburg and Belfer-Cohen, 1998; Corruccini, 1992; Kidder *et al.*, 1992; Kramer *et al.*, 2001; Wolpoff, 1999), while others find multiple species at Skhul alone (Schwartz and Tattersall, 2000; Tattersall and Schwartz, 1999). Many simply regard the sample as divided into two taxa, the “Neandertals” of Tabun and Amud and the early moderns of Skhul and Qafzeh, with little or no interbreeding possible (*e.g.* many of the papers in Akazawa *et al.* (ed.), 1998).

But are there Levant “Neandertals” or can the variation be described as T. McCown and A. Keith (1939) contended, as reflecting two distinct extremes that were linked by intermediate features? The extremes they identified were the paleoanthropic (European Neandertal-like), and neanthropic (early modern European-like, often exemplified by Cro Magnon). The woman from Tabun was not initially described as a Neandertal, but as representing the “Neandertal end” of a continuous range of variation at Mount Carmel. For a variety of reasons these caveats and explanations were lost, and with the claim of a substantial time difference between Tabun and Skhul (Howell, 1959), Levantine “Neandertals” came to exist as a distinct taxon. B. Vandermeersch proposed an explanation of *why* the “Neandertals” entered the Levant (1989, 1997). He began with the precept, ultimately proved correct, that the Qafzeh folk were earlier than the “Neandertals” in this region. Emigrations of Neandertals out of Europe were related to the onset of glacial conditions as the Würm began. Neandertal characteristics in the Levant, by this model, are actually European characteristics, expressed in a region that O. Bar-Yosef (1990) once described as a bus terminal for intercontinental migrations. But if these contentions are correct, are some of the Levantines “Neandertals” and others not (the Neandertal *versus* moderns interpretation), or is it as T. McCown and A. Keith described when they had the full Skhul sample before them? Was the Late Pleistocene Levant populated by peoples with a mixed morphology spanning the range between Neandertal (*e.g.* European) and non-Neandertal (*e.g.* Asian and African) poles, and does this imply that the people themselves were of mixed ancestry? Answering these questions impacts on the issue of how many human species are found in these remains.

The discussion has evolved beyond T. McCown and A. Keith’s contributions, with new key specimens (including the male “Neandertal” from Amud), accurate dating, and

---

7. Even without Amud and Qafzeh, T. McCown and A. Keith (1939) interpreted Skhul and Tabun as a population of a species in the process of evolutionary divergence; T. Dobzhansky (1944) considered them hybridized races of the same species.

more sophisticated assessments of anatomical variation. Specimens now ascribed to “Neandertal” and “non-Neandertal” samples are not without systematic differences. For instance, although the Levantine “Neandertals” are virtually the same height as the Skhul/Qafzeh sample—the midsex height means are 165 and 166 cm—relative distal limbs differ. The “non-Neandertals” have brachial and crural indices of 77 and 83, virtually the same as a sample of recent North Africans from Afalou. The “Neandertal” brachial index is even higher than these, 79, but the crural index at 78 differs and is lower (although the same as Skhul 5 and not as low as in the European Neandertals). A recent discussion of other differences can be found in the volume edited by T. Akazawa *et al.* (1998).

Yet, in his analysis of the Amud skull, H. Suzuki (1970) examined the details of the marked similarity he found between Amud with certain Skhul specimens, and concluded that Skhul 4 is more similar to Amud (*fig. 1*) than Skhul 4 is to Skhul 5. B. Arensburg and A. Belfer-Cohen (1998) argued that a combination of modern and archaic anatomical features in the highly variable Levant humans makes use of the term “Neandertal” undesirable. This point is supported by independent multivariate analyses of R. Corruccini (1992) and J. Kidder *et al.* (1992), and by the absence of substantial strength or use-pattern differences in the internal structure of “Neandertal” and “non-Neandertal” humeri reported by S. Ben-Itzhak *et al.* (1988).

A. Kramer *et al.* (2001) undertook a phylogenetic analysis of all the complete or mostly complete earlier Late Pleistocene Levant crania. They examined the question of whether the two “Neandertals” shared a set of uniquely derived traits that clustered them together as a linked anatomical entity (*fig. 9*). Tabun and Amud did not share such a set. Instead, each of the 17 most probable analyses showed that Tabun and Amud were never solely associated with each other. An additional analysis compared the Tabun woman with the other Levant crania, in a pairwise analysis (Wolpoff *et al.*, 2001) of 12 nonmetric features (*fig. 10*). Amud is quite different from Tabun for the diagnostic features used in this analysis; in fact, only two of the seven crania were found to be more different from Tabun than Amud is. These results also suggest that Levantine “Neandertals” are not a diagnosable entity.

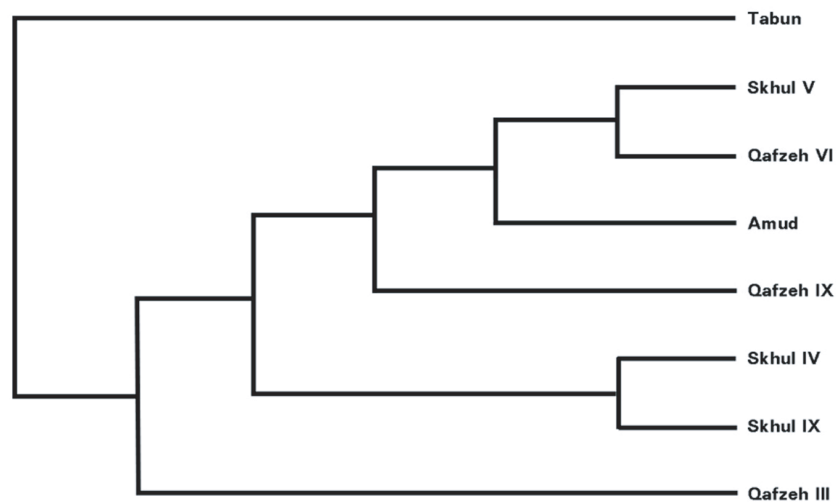


Fig. 9 - Phylogenetic analysis of Levant crania from Kramer et al. (2001). This is the average tree for the 17 best fitting ones.

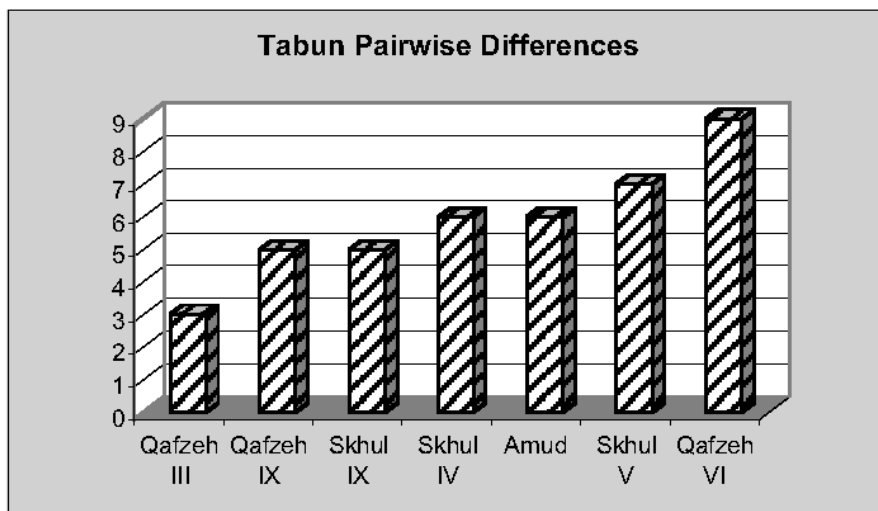


Fig. 10 - Pairwise differences of nonmetric characters, between Tabun and the other Levant crania (from Kramer et al., 2001). These are based on observations of the following 12 features: suprainiac fossa; maximum cranial breadth position; posterior parietal form; digastric sulcus form; occipitomastoid crest size; anterior glenoid slope; postglenoid tubercle; frontal-sagittal keel; lateral frontal boss; glabellar bulge; central occipital torus form; supraorbital torus form. Tabun is most like Qafzeh 3, the only other female in the sample, and is least like Qafzeh 6. Four Skhul/Qafzeh specimens, 2/3 of the sample, are more similar to Tabun than Amud is.

## CONCLUSION

Here we present a new sampling statistic designed to address the null hypothesis, and lay out the procedure for its use. STET describes size and shape variation with the estimates of dispersion around the slopes of pairwise linear regressions. It has the advantages of maximizing fossil data with the comparison of specimens of uncertain or unknown sex and allows pairwise analyses based on different numbers of measurements and different measurements to be combined.

STET was used to examine the null hypothesis for Late Pleistocene crania from Israel, crania often characterized as “Neandertal” and “early modern.” The results of this study reinforce several others, *not* in concluding that the Levant “Neandertals” are the same as the Skhul/Qafzeh specimens, but in concluding that a null hypothesis —taxonomy does not underlie the size and shape variation of the sample— cannot be disproved. The pairwise variation in size and shape within the Skhul/Qafzeh cranial sample is, if anything, more than the variation found in comparing the Levant “Neandertals” with the Skhul/Qafzeh specimens. The pairwise comparisons within the combined Levant sample show less dispersion than comparisons within the Sterkfontein australopithecine sample. However, the pairwise comparisons of the Sterkfontein australopithecines with the Skhul/Qafzeh sample show much more variation than the pairwise comparison of the Levant “Neandertals” with the Skhul/Qafzeh sample.

The Levant sample is the first, but certainly not the only instance when taxonomy has been advanced as an explanation for variation in a sample. These claims are often repeated but rarely subjected to testing. We expect that the plethora of species names that have been advanced to describe variation in the human clade will not hold up to continued scrutiny.

## Acknowledgments

We are pleased to present this essay in honor of Dr. Bernard Vandermeersch, and are deeply grateful to and Dr. Bruno Maureille and Dr. Jaroslav Bruzek for their invitation to contribute to this volume. We thank the curators of the fossil specimens from Amud, Qafzeh, Skhul, Sterkfontein, and Tabun for permission to study the specimens in their care. We appreciate the help from Dr. Karen Rosenberg in editing this paper.

## BIBLIOGRAPHY

- AIELLO (L.C.), COLLARD (M.), THACKERAY (J.F.), WOOD (B.A.) 2000, Assessing exact randomization-based methods of determining the taxonomic significance of variability in the human fossil record, *South African Journal of Science* 96: 179-183.
- AKAZAWA (T.), AOKI (K.), BAR-YOSEF (O.) (eds) 1998, *Neandertals and Modern Humans in Western Asia*, Plenum Press, New York.
- ARENSBURG (B.), BELFER-COHEN (A.) 1998, Sapiens and Neandertals. *Neandertals and Modern Humans in Western Asia*, in T. Akazawa, K. Aoki, O. Bar-Yosef (eds), Plenum Press, New York, p. 311-322.
- BAR-YOSEF (O.) 1990, Mousterian adaptations—a global view, *Quaternaria Nova* 1: 575-591.
- BEN-ITZHAK (S.), SMITH (P.), BLOOM (R.A.) 1988, Radiographic study of the humerus in Neandertals and *Homo sapiens sapiens*, *American Journal of Physical Anthropology* 77, 2: 231-242.
- CHAMBERLAIN (T.C.) 1965, The method of multiple working hypotheses, *Journal of Geology* 5: 837-848.
- CORRUCCINI (R.) 1992, Metrical reconsideration of the Skhul IV and IX and Border Cave 1 crania in the context of modern human origins, *American Journal of Physical Anthropology* 87: 433-445.
- DOBZHANSKY (T.) 1944, On species and races of living and fossil man, *American Journal of Physical Anthropology* 2, 3: 251-265.
- GOULD (S.) 1988, A novel notion of Neanderthal, *Natural History* 97: 16-21.
- HOWELL (F.) 1959, Upper Pleistocene stratigraphy and early man in the Levant, *Proceedings of the American Philosophical Society* 103: 1-65.
- HOWELLS (W.) 1973, *Evolution of the Genus Homo*, Addison-Wesley, Reading.
- JIGGINS (C.D.), NAISBIT (R.E.), COE (R.L.), MALLET (J.) 2001, Reproductive isolation caused by colour pattern mimicry, *Nature* 411: 302-305.
- KIDDER (J.), JANTZ (R.), SMITH (F.) 1992, Defining modern humans: a multivariate approach, in G. Bräuer, F.H. Smith (eds), *Continuity or replacement: controversies in Homo sapiens evolution*, Balkema, Rotterdam, p. 157-177.
- KIMBEL (W.H.), MARTIN (L.B.) (eds) 1993, *Species, Species Concepts, and Primate Evolution*, Plenum, New York.
- KRAMER (A.), CRUMMETT (T.L.), WOLPOFF (M.H.) 2001, Out of Africa and into the Levant: replacement or admixture in Western Asia? *Quaternary International* 75, 1: 51-63.
- LARSEN (C.S.), MATTER (R.M.), GEBO (D.L.) 1991, *Human Origins, The Fossil Record*, Second Edition, Waveland Press, Prospect Heights.
- LOVEJOY (C.O.) 1979, Contemporary methodological approaches to individual primate fossil analysis, in *Environment, Behavior, and Morphology*, Gustav Fischer, New York, p. 229-243.
- MAYR (E.) 1963, *Animal Species and Evolution*, Belknap Press of Harvard University Press, Cambridge.

- MCCOWN (T.D.), KEITH (A.) 1939, *The Stone Age of Mount Carmel: The Fossil Human Remains from the Levallois-Mousterian*, Volume II, Clarendon Press, Oxford.
- MILIUS (S.) 2001, Alarming butterflies and go-getter fish: Overlooked ways to invent new species, *Science News* 160, 3: 42-45.
- SCHWARTZ (J.H.), TATTERSALL (I.) 2000, The human chin revisited: what is it and who has it? *Journal of Human Evolution* 38: 367-409.
- SOKAL (R.R.), ROHLF (F.J.) 1981, *Biometry*, Second Edition, W.H. Freeman, San Francisco.
- SUZUKI (H.) 1970, The skull of the Amud man. in H. Suzuki and F. Takai (eds), *The Amud Man and his Cave Site*, University of Tokyo, Tokyo, p. 123-206.
- TATTERSALL (I.), SCHWARTZ (J.) 1999, Hominids and hybrids: The place of Neanderthals in human evolution, *Proceedings of the National Academy of Sciences USA* 96: 7117-7119.
- THACKERAY (J.F.), BELLAMY (C.L.), BELLARS (D.), BRONNER (G.), BRONNER (L.), CHIMIMBA (C.), FOURIE (H.), KEMP (A.), KRÜGER (M.), PLUG (I.), PRINSLOO (S.), TOMS (R.), VAN ZYL (A.J.), WHITING (M.J.) 1997, Probabilities of conspecificity: application of a morphometric technique to modern taxa and fossil specimens attributed to *Australopithecus* and *Homo*, *South African Journal of Science* 93, 4: 195-196.
- THACKERAY (J.F.), HELBIG (J.), MOSS (S.) 1995, Quantifying morphological variability within extant mammalian species, *Palaeontologia Africana* 31: 23-25.
- THACKERAY (J.F.), MDAKA (S.), NAVSA (N.), MOSHAU (R.), SINGO (S.) 2000, Morphometric analysis of conspecific males and females: an exploratory study of extant primate and extinct hominid taxa, *South African Journal of Science* 96: 534-536.
- THOMA (A.) 1962, Le déploiement évolutif de l'*Homo sapiens*, *Anthropologia Hungarica* 5: 1-111.
- VANDERMEERSCH (B.) 1989, The evolution of modern humans: recent evidence from Southwest Asia, in P. Mellars and C.B. Stringer (eds) *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans*, Edinburgh University Press, Edinburgh, p. 155-164.
- VANDERMEERSCH (B.) 1997, The Near East and Europe: continuity or discontinuity? in G.A. Clark, C.M. Willermet (eds), *Conceptual Issues in Modern Human Origins Research*, Aldine de Gruyter, New York, p. 107-116, and combined bibliography on p. 437-492.
- WOLPOFF (M.H.) 1999, *Paleoanthropology*, Second Edition, McGraw-Hill, New York.
- WOLPOFF (M.H.), HAWKS (J.D.), FRAYER (D.W.), HUNLEY (K.) 2001, Modern human ancestry at the peripheries: a test of the replacement theory, *Science* 291: 293-297.