Journal of Archaeological Science 31 (2004) 1643–1658

Journal of Archaeological Science

http://www.elsevier.com/locate/jas

Computationally intensive multivariate statistics and relative frequency distributions in archaeology (with an application to the Early Epipaleolithic of the Levant)

Aaron Jonas Stutz[a,*], George F. Estabrook[b]

[a]Museum of Anthropology, University of Michigan, Ann Arbor, MI 48109-1079, USA
[b]Department of Ecology and Evolutionary Biology, University of Michigan, 2039 Natural Sciences Building, Ann Arbor, MI 48109-1048, USA

Received 16 May 2003; received in revised form 30 March 2004

Abstract

Archaeologists seek to analyze patterns of similarity and difference among diverse kinds of assemblages that (1) vary in the number of specimens and (2) have been characterized by standard multi-category frequency distributions. Recent developments in computer simulation methods offer marked improvements in our ability to test statistical hypotheses about variation in relative taxonomic or typological abundance data, drawn from assemblages of variable sizes and diverse archaeological contexts (American Antiquity 66 (2001) 715; Journal of Archaeological Science 30 (2003) 837). In this article we extend the highly flexible and powerful computationally intensive framework for statistical analysis to the multivariate family of methods of analysis of variation (MANOVA) and non-hierarchical cluster analysis. We treat the relative type-frequency distribution as a multivariate quantitative description of the archaeological assemblage. We then introduce two simulation-based computer applications for analyzing variability between groups of assemblages. We utilize the multivariate applications in a case study; we evaluate how standard microlith typological classifications perform in capturing information about technological and formal variability among Early Epipaleolithic microlith assemblages from the Southern Levant. Computationally intensive, simulation-based statistical techniques allow the researcher to custom-tailor the measure of statistical variation and the model of random archaeological record formation relevant to the given problem. We suggest that with simulation-based approaches, the costs of computer programming and processing are far outweighed by the potential for explaining quantifiable variability in material traces of human activity in the past.

Keywords: Statistical analysis; Relative frequencies; Multivariate; Computationally intensive

1. Introduction

In this article we introduce two new computationally intensive statistical approaches for testing hypotheses about a common and theoretically important form of quantitative variation in archaeological, paleofaunal and paleobotanical remains. Archaeological researchers often seek to evaluate patterns of similarity and difference among assemblages that are characterized by *multi-category relative frequency distributions*. Diverse topics – from past climate change to the role of ceramic vessel decoration in the development of social differentiation – may be addressed by examining interassemblage variation in the relative abundance of biological taxa or material culture types [27]. We are particularly interested in testing hypotheses that predict grouping structure within sets of archaeological assemblages characterized by detailed relative frequency distributions. The two statistical applications we present are based on computer simulation, developed especially to deal with multi-category percentage data. Each incorporates different assumptions about our prior knowledge of how the study assemblages might cluster, applying different models of random archaeological

* Corresponding author. Present address: Department of Archaeology and Classical History, Lund University, 223 50 Lund, Sweden.
E-mail address: astutz@umich.edu (A.J. Stutz).

0305-4403/$ - see front matter
doi:10.1016/j.jas.2004.04.005

record formation against which to compare our observed patterns of variability.

The first application evaluates whether *previously defined* clusters of assemblages, each assigned to its group according to criteria chosen by the researcher, are significantly unlikely to have occurred independently within the context of the observed data. By asking whether some a priori hypothesized group structure was underlain by a non-random process in the past, this first approach provides an analogue to standard, parametric multivariate ANOVA techniques for a data format – that of the multi-category relative frequency distribution – that does not satisfy the basic assumptions of MANOVA models (see Ref. [64], pp. 818–819).

The second application may be described as '*k*-means style' cluster analysis with statistical significance levels. The between-cluster distances for the identified best *k*-cluster separation of the observed *n* assemblages are compared to the distribution of between-cluster differences obtained by computer simulation, in which the artifacts or faunal or botanical specimens are repeatedly shuffled among the assemblages. This latter approach allows the archaeologist to treat the results yielded by the cluster analysis algorithm as a testable statistical research question. Does the best *k*-cluster solution represent a grouping structure unlikely to have formed if the artifacts under study had been deposited randomly into different assemblages?

We suggest that the two approaches – the statistical analysis of variation between previously defined *k* groups and the search for statistically significant clusters within a sample of *j* assemblages – can powerfully complement one another in addressing archaeological problems that focus on detailed relative abundance data. In presenting a case study, we apply the computationally intensive techniques to the statistical analysis of microlithic typological variability among Early Epipaleolithic (ca. 23–18 ka, calibrated ^{14}C years) assemblages from the Southern Levant.

2. Background

This article represents the respective archaeological and statistical interests of the authors. It stems in part from a wider archaeological project initiated by AJS to revisit how analysis of technological and formal variation in Epipaleolithic (ca. 23–11 ka, calibrated radiocarbon years) chipped stone assemblages might shed new light on shifting regional patterns of social interaction in the long-term emergence of pre-agricultural sedentism in the Levant [5–7,15,34,38–40,42,43,46]. This article also offers an introduction to computationally intensive simulation approaches to multivariate statistical analysis in archaeology; GFE has long advocated for the application of computationally intensive statistical techniques to fit the particular research questions and data limitations characteristic of different disciplines, from the biological to the social sciences (e.g. Ref. [45]; see Ref. [33]). The term 'computationally intensive' in statistical analysis may be generically defined as using the computer's calculating power in repeatedly simulating datasets that yield probability distributions predicted from whatever assumptions of randomness, and for whatever measures, the researcher might consider interesting, instead of those required by the mathematical methods of standard statistics (Ref. [64], pp. 803–825).

Archaeologists may be variously concerned with Mousterian tool-type counts, taxonomic representation in pollen samples, or decorative motif frequencies in ceramic sherd assemblages. Still, we regularly deal with detailed percentage data. Several statistical models have been developed for evaluating the fit of a given sample's frequency distribution – or patterns of variation among many samples' frequency distributions – to a well-studied underlying population distribution (cf. Ref. [64], Chapters 13 and 17), but we often have insufficient control over the archaeological recovery process to support an important formal requirement. Our samples should be randomly drawn representatives of the artifact populations of theoretical interest (e.g. formal tools produced during the Middle Paleolithic period in the French Perigord region). We note that important previous archaeological applications of computationally intensive methods, including bootstrapping, have dealt with this problem; in the absence of sufficient information about the archaeological population of interest, how do we evaluate the statistical relationship (or lack of it) between an assemblage's size and its typological or taxonomic diversity [13,27,47–49,53]? Indeed, there are technical challenges to working with detailed percentage data drawn from artifact assemblages that vary widely in size and that are collected from heterogeneous archaeological contexts. As recently emphasized by Baxter [13] and Cochrane [27], archaeological percentage data are highly amenable to statistical analyses that utilize computationally intensive techniques, although especially in light of Baxter's discussion of Kintigh's [48] classic simulation approach, we emphasize that certain care must be paid in constructing a model of random assemblage or assemblage-group formation.

We seek to make the case below that marrying multivariate statistical analysis with a computationally intensive approach is especially appropriate for examining the variability among groups of archaeological assemblages described quantitatively by detailed relative frequency data. Indeed, percentage profiles of taxonomic or typological classifications certainly characterize an important form of multidimensional variation in archaeological remains. We recognize that standard multivariate techniques, including cluster analysis, factor analysis, correspondence analysis and parametric MANOVA, have produced interesting insights about archaeological cases in which variability is characterized by detailed relative frequency distributions [34,57] (see Refs. [11,63] for reviews). However, with the exception of the exploratory technique of correspondence analysis [11,63], the commonly employed multivariate methods were not specifically designed with relative abundance data in mind. We argue that processing power in the contemporary personal computer makes possible the use of more reliable, more straightforward, and more archaeology-specific methods to measure the relevant differences among groups of assemblages, as well as their statistical significances.

3. The case study

Our simulation approach to multivariate statistics may be introduced through analyzing microlith type-frequency variability among assemblages dated by absolute or relative means to the early portion of the Epipaleolithic (ca. 23–18 ka, calibrated ^{14}C years) in the Levant. In this article, we examine the following hypothesis outlined by Henry (Ref. [43], pp. 435–437): *during the Early and Middle Epipaleolithic (ca. 23–15 ka), ecological and cultural factors combined to maintain a long-term, archaeologically visible social boundary between human forager networks inhabiting the Mediterranean phytogeographic zone of the Levant and those occupying the more arid Irano-Turanian grassland margins to the east and south*. Henry [43] has suggested this hypothesis based primarily on observations of the formal and technological variability among microlith assemblages from different Epipaleolithic sites in the Levant.

3.1. Henry's social boundary hypothesis in theoretical context

The Epipaleolithic period has been classically defined on lithic techno-typological grounds. In comparison with the preceding and initially somewhat temporally overlapping Late Upper Paleolithic, the Epipaleolithic is characterized by chipped-stone production strategies focused increasingly on microliths (i.e. bladelets intentionally shaped through retouch into 'non-geometric microliths' and segments of bladelets shaped into 'geometric microliths' [16]). As replaceable inserts in compound tools, Epipaleolithic microlithic elements took on diverse forms through multiple techniques for segmenting and retouch [4,5,16,22,34,38–43]. With the adoption of this innovative, flexible technological system, different microlithic forms produced by alternative strategies of bladelet segmentation and shaping could provide identical functions, primarily as inserts in compound tools (cf. Ref. [9]). The same microlithic form could also be used in different activities, as points or lateral barbs in projectiles or as cutting edges in straight sickles or toothed saws (cf. Refs. [8,24,31,67]).

Considering that microlithic technological systems potentially produce multiple 'isochrestic' variants (sensu Refs. [61,62]), many researchers working on the Epipaleolithic have argued that formal and technological variation among microlithic assemblages reflects stylistic differences across long-term social boundaries [4–7,16, 34,38–40,42,43]. Other workers have contended that interassemblage variability in the Epipaleolithic was more likely patterned by different, locally adaptive strategies for managing optimal lithic tool kits within in a mobile foraging economy [3,23,29,55]. In light of this disagreement over the causes of microlith assemblage variability, Henry's specific hypothesis about a persistent Early-Middle Epipaleolithic social boundary between the ecological 'center and edge' in the southern portion of the Levant is especially relevant. Other factors being equal, the techno-typological distinctions between semi-arid steppe and Mediterranean area assemblages could simply be explained by distinct technological solutions adopted by the same group of hunter-gatherers moving between ecozones. However, Henry (Ref. [43], p. 314) emphasizes that his proposed, primarily east–west techno-typological divide is defined in significant part by a distinction in the technology by which bladelets are segmented for shaping into microlithic elements. Debris distinctive to the microburin technique is comparatively common in Early and Middle Epipaleolithic assemblages from the steppic zone, but it is generally rare from sites within the contemporaneous Mediterranean vegetation zone to the west [40,42,43]. Formally identical microlithic elements – including many subtypes within the common Early-Middle Epipaleolithic categories of curved and pointed bladelets, obliquely truncated backed bladelets, and triangles – may be made equally effectively with the microburin technique or with a simple snap and/or abrupt retouch on an anvil (see Refs. [16,41,42]). Thus, as Henry [43] argues, a feature of isochrestic variation in the technological *chaine operatoire* for producing formally similar tools provides one of the most notable archaeological criteria for separating between arid and Mediterranean ecozone assemblages. This distinction would not be so simply explained by locally adaptive technological strategies. In the absence of a geographically stable social boundary, why would mobile Early Epipaleolithic hunter-gatherers have made curved and pointed backed bladelets utilizing the microburin technique when camping in the Wadi Jilat (Eastern Jordan) but have employed only abrupt retouch on an anvil when in the Jordan Valley (e.g. Refs. [36,37,44])? A well-defined social boundary – occurring between mobile foraging groups adapted to different environmental zones – could not only have been expressed in culturally common notions of group identity and territoriality, but also reinforced in many contexts of everyday interaction within the group, by distinctive technological practices in microlith production.

We agree with many aspects of Henry's [43] argument for an Early Epipaleolithic social boundary. In particular, we find his proposition very plausible that archaeologically preserved social boundaries in the Pleistocene could correspond with eco-geographic transition zones. We suggest, though, that there are two reasons for re-analyzing the goodness of fit of the archaeological data to what would be predicted from his hypothesis. First, we make the general point that Henry argues for his model based on general 'eyeball' assessments of technological and type-frequency variation among microlith assemblages, and we offer below a formal statistical evaluation of his claims. Second, we point out that Goring-Morris [38] identifies the Nizzanan archaeological culture as an important subgroup of assemblages with microburin debris and products that date to the end of the Early Epipaleolithic sequence. The Nizzanan microlithic assemblages are described as being dominated quantitatively by microburins and scalene bladelets/triangles. (Some of these assemblages have been labeled as variant A2 of the Geometric Kebaran by Bar-Yosef and Vogel [10] (see Refs. [5,7,22,38–40,42] for techno-typological definitions of Levantine Epipaleolithic archaeological cultures).) Most relevant for testing Henry's eco-social boundary hypothesis, the Nizzanan site distribution transgresses the geographic boundary drawn by Henry. Nizzanan assemblages have been recovered from sites distributed in an arc from the northwestern Negev desert, along the coastal plain and Carmel ridge, and across the Upper Jordan Valley, continuing to the Azraq area (Refs. [38, pp. 198–200, 40]). Stratigraphic data suggest that these microburin-rich assemblages date to a temporal horizon at the very end of the Early Epipaleolithic (post-dating or contemporaneous with deposits containing late Kebaran industries) or the beginning of the Middle Epipaleolithic (pre-dating deposits with classic Geometric Kebaran or Mushabian industries [38,39]). If Goring-Morris has correctly identified the Nizzanan as a significant grouping of Early Epipaleolithic industries, then Henry's proposed long-term social boundary hypothesis requires modification. Instead, distinctive stylistic traditions of microlith production – in significant part defined by the rarity or commonness of the microburin technique – may have dynamically shifted in their geographic distributions, sometimes traversing multiple ecological zones [40].

3.2. Formalizing the hypotheses

Our analytical concern is to account for the causes of variation in Early Epipaleolithic microlith relative type-frequency distributions, and our statistical analysis involves a comparison of the explanatory power of two hypotheses. The first (H_1) is based on Henry's [43] proposed long-term stable social boundary. It is predicted that Early Epipaleolithic microlith assemblages from the semi-arid steppic margin are more typologically similar to one another than to Early Epipaleolithic assemblages from the Mediterranean vegetation zone. The second hypothesis (H_2) is developed from Goring-Morris and Belfer-Cohen's [40] observation that the terminal Early Epipaleolithic Nizzanan assemblage cluster – defined by an eyeball estimation of high richness in microburin debris and triangles or proto-triangles – is distributed across the Mediterranean/Irano-Turanian steppe phytogeographic boundary (see also Ref. [38]). This alternative hypothesis predicts that Early Epipaleolithic assemblages rich in microburin artifacts, falling into what have been labeled as the Nebekian and Nizzanan archaeological cultures, will have microlith relative type-frequency profiles that are more similar to one another than to the profiles of assemblages poor in microburin artifacts, regardless of geographic association.

3.3. The case study dataset

Our case study dataset includes microlithic type counts from 17 archaeological samples, representing stratigraphically distinct assemblages from 13 different Southern Levantine Early Epipaleolithic buried and non-deflated sites (Table 1). The group from the Mediterranean zone is dominated by sites in the Jordan Valley but also includes assemblages recovered from deposits in the Galilee (Hayonim Cave Layer C) and the southern Coastal Plain (Kfar Darom 8). The group from the eastern arid steppic zone includes sites from the Azraq area, dominated by the Early Epipaleolithic Nebekian archaeological culture. In each geographic group is one so-called Nizzanan assemblage: Ein Gev IV from the Mediterranean zone and Wadi Jilat 6/Upper from the semi-arid shrub-grassland. In the case study sample, as shown in Table 1, the characterization of Nebekian and Nizzanan assemblages as 'microburin-rich' is supported by a clear pattern in relative abundance of microburin debris and microlith types defined by use of the microburin technique (la Mouillah points and Qalkan points; see Refs. [42,43]).

In analyzing the published typological data that underlie our study sample, we have sought to maintain the maximum number of typological categories in our study, keeping as much descriptive information as possible. However, we were required to collapse several categories of micropoints and curved backed bladelets into a single category – that of curved and pointed backed bladelets – because not every researcher has followed the same typological subdivision. The resulting number of typological microlith classes is 31 (compare

Table 1
Early Epipaleolithic assemblages from the Southern Levant in the study sample

Assemblage (abbreviated name)	Assemblage (site and/or layer)	Geographic location	Archaeological culture[a]	Microburin index	Approximate age (calibrated ^{14}C dates)
KDr 8[b]	Kefar Darom 8	S. Coastal Plain	Kebaran	0.010	
Hay C[b]	Hayonim Cave/Layer C	W. Lower Galilee	Kebaran	0.002	
EG IV[b]	Ein Gev IV	Upper Jordan Valley	Nizzanan	0.780	
EG III[b]	Ein Gev III	Upper Jordan Valley	Kebaran	0.027	
EG II[b]	Ein Gev II	Upper Jordan Valley	Kebaran	0	
EG I[b]	Ein Gev I/Layers 3–4	Upper Jordan Valley	Kebaran	0	
UR IIa[c]	Urkan-e-Rubb IIa	Middle Jordan Valley	Kebaran	0	18–19 ka
Fa VII[d]	*Wadi Fazael VII*	*Middle Jordan Valley*	*Kebaran*	*0.006*	
Fa IIIA[d]	*Wadi Fazael IIIA*	*Middle Jordan Valley*	*Kebaran*	*0.006*	
Fa IIIB[d]	*Wadi Fazael IIIB*	*Middle Jordan Valley*	*Kebaran*	*0.022*	
WH 26[e]	Wadi Hammeh 26	Middle Jordan Valley	Kebaran	0.131	
Jil 6 U[f]	*Wadi Jilat 6/Upper*	*Azraq Basin*	*Nizzanan*	*0.677*	*18–19 ka*
Jil 6 M[f]	*Wadi Jilat 6/Middle*	*Azraq Basin*	*Nebekian*	*0.644*	
Jil 6 L[f]	*Wadi Jilat 6/Lower*	*Azraq Basin*	*Nebekian*	*0.685*	
Uw 14U[g]	*Uwaynid 14/Upper*	*Azraq Basin*	*Nebekian*	*0.546*	*22–23 ka*
Uw 14M[g]	*Uwaynid 14/Middle*	*Azraq Basin*	*Nebekian*	*0.339*	*22–23 ka*
Uw 18[g]	Uwaynid 18	Azraq Basin	Nebekian	0.541	22–23 ka

The published sources for the archaeological culture designations and type-counts for assemblages are as follows: [a]Goring-Morris and Belfer-Cohen [40] divide Early Epipaleolithic assemblages from the study area into Kebaran, Nebekian, and Nizzanan archaeological cultures; [b]Bar-Yosef [4]; [c]Hovers and Marder [44]; [d]Goring-Morris [68]; [e]Edwards et al. [32]; [f]Garrard et al. [36]; [g]Garrard et al. [37]. The 'microburin index' is calculated as the sum of microburin debris and microburin-based microlith types (la Mouillah and Qalkan points) divided by the sum of all microliths and microburin debris (modified from Refs. [38,42,52]). In this table, adjacent row entries in italics represent assemblages from buried deposits within well-defined multi-component stratified sites (e.g. Jilat 6 Lower/Upper).

with Ref. [34]). The microlith-type relative frequency profiles along with the artifact totals for each assemblage are presented in Table 2.

We note that our sample of assemblages is not complete. At the time of this writing, we have not yet included several Early Epipaleolithic ('Nebekian/Qalkan' and 'Kebaran/Early Hamran' technocomplexes) assemblages from buried contexts in southern Jordan [43,65], as well as two additional buried, non-deflated Nizzanan assemblages from the northwestern Negev in Israel [38]. Adding these assemblages will increase our Early Epipaleolithic Levantine sample, but for this exploratory study, we have employed a still sizeable Early Epipaleolithic sample of microlithic assemblages. While the results of our case study should be considered with this in mind, we stress that our initial aim is to illustrate how computationally intensive statistical applications – with custom-designed measures of interassemblage variability and random assemblage formation – can offer more reliable ways to test predictions derived from current archaeological debates about what caused the variability in Levantine Epipaleolithic chipped-stone tool assemblages.

4. Statistical methods

We aim to analyze patterns of multivariate similarity and difference in microlith type-frequency distributions, in order to evaluate the explanatory power of the two alternative hypotheses defined above (summarized in Table 3). Our main theoretical concern is to allow the archaeological question at hand to shape the statistical methods and arguments. This has two implications for developing computationally intensive approaches for our analysis. First, we need to choose formal measures of similarity and difference between assemblages and groups of assemblages, in order to achieve comparisons that address the archaeological research problem. Second, in investigating how well the alternative hypotheses account for the observed patterns of interassemblage variation, it is also necessary to simulate the variation that would be predicted by relevant, plausible random processes of archaeological record formation.

4.1. Measuring variation in relative frequency profiles between pairs of assemblages

We suggest that a consistent and concise quantitative summary of the difference between any two assemblages' relative frequency profiles can be defined as

$$D_{ij} = \frac{\left(\sum_{m}^{c} \left| f_{mi} - f_{mj} \right| \right)}{2} \quad (1)$$

This pairwise dissimilarity index involves summing the absolute values of the differences between the respective relative frequencies, f, of the m^{th} type in each of assemblages i and j, for each of the c types. The sum of the absolute differences in the relative frequencies can vary

Table 2
Relative type-frequency distributions for microlithic assemblages in the study sample

Microlith type	KDr8	HayC	EGI	EGII	EGIII	EGIV	UR2a	Fa3B	Fa3A	Fa7	WH26	Jil6L	Jil6M	Jil6U	Uw14M	Uw14U	Uw18
Curved pointed bladelet	0.543	0.127	0	0	0.113	0	0.421	0.478	0.349	0.017	0.020	0.598	0.031	0.120	0.230	0.023	0.461
Backed and truncated bladelet	0.098	0.225	0.327	0.622	0.317	0	0.011	0.034	0.251	0.462	0.837	0	0	0	0	0	0
Quasi-trapeze-rectangle	0.009	0	0	0	0.162	0	0	0	0.009	0.017	0	0.141	0.372	0.103	0.596	0.664	0.464
Backed bladelet	0.034	0.056	0.138	0.085	0.085	0.240	0.051	0.101	0.115	0.040	0.092	0.043	0.132	0.072	0.027	0.031	0.019
Retouched bladelet	0.047	0.158	0.125	0.073	0.021	0.024	0.230	0.039	0.101	0.012	0.013	0	0	0	0.005	0.016	0
Truncated bladelet	0.090	0.063	0	0	0	0	0	0.056	0.120	0.087	0.020	0	0	0	0	0	0
Microgravette	0.003	0.028	0.003	0	0.085	0.275	0	0.006	0.002	0.254	0	0	0.016	0.113	0	0	0
Micropoint w/basal truncation	0.081	0.106	0	0	0	0	0.006	0.219	0.002	0	0	0	0	0	0	0	0
La Mouillah point	0	0	0	0	0	0	0	0	0	0	0	0.141	0.380	0.031	0.093	0.148	0.011
Scalene triangle	0.003	0.017	0	0	0	0.178	0	0	0.001	0	0	0	0	0.223	0	0	0
Isosceles triangle	0.006	0.002	0	0	0	0.188	0	0.006	0	0.006	0	0	0	0.182	0	0	0
Pointed backed bladelet	0	0	0.097	0.085	0	0	0.140	0	0	0	0	0.011	0.008	0.031	0	0.008	0.025
Bladelet w/inverse retouch	0.032	0.048	0	0	0	0	0.017	0	0.007	0.069	0	0	0	0	0	0	0
Double-retouched bladelet	0	0.054	0.049	0.012	0.035	0.014	0.017	0.006	0.008	0	0.007	0.022	0	0	0	0	0
Retouched bladelet w/prox. trunc.	0	0	0.120	0.098	0	0	0	0	0	0	0	0	0	0	0	0	0
Proto-trapeze	0.001	0.020	0	0	0	0	0	0.006	0.008	0.012	0	0	0	0	0.044	0.086	0.008
Retouched and truncated bladelet	0	0	0	0	0.169	0.031	0.034	0.034	0	0	0	0	0	0	0	0	0
Bladelet w/alternating retouch	0.015	0.026	0.010	0	0	0.003	0.006	0	0.005	0	0.013	0	0	0	0	0	0
Falita point	0	0	0.102	0.012	0	0.003	0	0.006	0	0	0	0	0	0	0	0	0
Retouched pointed bladelet	0.005	0.019	0.020	0	0	0	0	0.006	0.009	0	0	0	0	0	0	0	0
Double-retouched pointed bladelet	0	0.020	0.003	0.012	0.014	0.024	0	0	0.007	0.012	0	0	0	0	0	0	0
Protolunate	0.003	0.002	0	0	0	0.010	0	0	0.001	0	0	0	0	0.089	0	0	0
Scalene bladelet	0.014	0.011	0	0	0	0	0.006	0	0.002	0	0	0	0	0	0	0	0
Double-backed bladelet	0	0	0	0	0	0	0	0	0	0	0	0.022	0.023	0.027	0.005	0.023	0.011
Trapeze	0.007	0.004	0	0	0	0.003	0.034	0.006	0.002	0.006	0	0	0	0	0	0	0
El Wad point	0	0.013	0.005	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lunate	0.004	0	0	0	0	0	0	0	0	0	0	0.022	0	0.010	0	0	0
Rectangle	0.003	0	0	0	0	0	0.006	0	0.001	0	0	0	0.016	0	0	0	0
Proto-rectangle	0.001	0	0	0	0	0.003	0.011	0	0.001	0.006	0	0	0	0	0	0	0
Qalkan point	0	0	0	0	0	0	0	0	0	0	0	0	0.023	0	0	0	0
Pt. backed bladelet w/trunc. base	0	0	0	0	0	0	0.011	0	0	0	0	0	0	0	0	0	0
Total	951	537	391	82	142	287	178	178	1706	173	153	92	129	292	183	128	360

Microlithic types are listed in descending order of regional abundance. '0' entries indicate that the given type is absent in the assemblage. Assemblage names are abbreviated as in Table 1. See Bar-Yosef [4], Goring-Morris [38], Henry [43], and Tixier [69] for comprehensive technological and formal definitions of the microlith types employed in this analysis.

Table 3
Classifications of the study assemblages into two groups according to alternative hypothesized criteria structuring formal and technological similarity among Early Epipaleolithic microlith assemblages

Assemblage (abbreviated name)	(H_1) Ecological zone (Mediterranean vs. arid steppe)	(H_2) Abundance of microburin debris (rich vs. poor)[a]
KDr 8	Mediterranean	Poor
Hay C	Mediterranean	Poor
EG I	Mediterranean	Poor
EG II	Mediterranean	Poor
EG III	Mediterranean	Poor
EG IV	Mediterranean	Rich (Nizzanan)
UR IIa	Mediterranean	Poor
Fa IIIB	Mediterranean	Poor
Fa IIIA	Mediterranean	Poor
Fa VII	Mediterranean	Poor
WH 26	Mediterranean	Poor
Jil 6 L	Arid steppe	Rich
Jil 6 M	Arid steppe	Rich
Jil 6 U	Arid steppe	Rich (Nizzanan)
Uw 14M	Arid steppe	Rich
Uw 14U	Arid steppe	Rich
Uw 18	Arid steppe	Rich

H_1: assemblages are best separated according to phytogeographic criteria, because environmental factors would have reinforced a social/stylistic boundary between hunter-gatherer communities in the Mediterranean and the arid steppe zones through the Early Epipaleolithic. H_2: assemblages are best separated according to richness or rarity of microburin debris, because this aspect of technological style would have distinguished traditions of microlith production in different hunter-gatherer groups, regardless of their geographic distribution. Note that the microburin-rich Nizzanan assemblages (Ein Gev IV [EG IV] and Wadi Jilat 6 Upper [Jil 6 U]) transgress the key phytogeographic boundary emphasized under H_1.

between 0 (for identical assemblages) and 2 (maximally different ones – if tools of a given type are present in one assemblage, then they are absent in the other). In our analysis, we take the convenient step of dividing the raw dissimilarity level by 2, thus rescaling this value on a scale from $D_{ij} = 0$ for identical assemblages to $D_{ij} = 1$ for completely non-identical ones.

This index of interassemblage dissimilarity is essentially the 'Robinson coefficient of agreement', which was initially developed for the fine seriation of Mesoamerican prehistoric ceramic assemblages (Ref. [59], p. 297; see also Refs. [21,30, pp. 139, 272–276, 51,60]). Our modified 'Robinson-style' index of dissimilarity offers a direct formalization of the common intuitive approach of eyeballing variation between relative-frequency graphs. It also remains methodologically simple; in comparing the i^{th} and j^{th} assemblages, it further allows us to break out and analyze d_{mij}, defined as the absolute difference in relative frequencies for the m^{th} individual type class:

$$d_{mij} = \left| f_{mi} - f_{mj} \right| \tag{2}$$

More importantly, the relative-frequency profile index of dissimilarity, D_{ij} (see Eq. (1)), translates into a fundamental theoretical concern with the *whole* relative-frequency profile, in which a given type contributes to the measurement of interassemblage variability in direct proportion to its global relative abundance in the study sample. The different archaeological models of culture change and cultural variability in the Epipaleolithic – including those that emphasize technological and organizational adaptations to conditions of water, food, and raw material availability (cf. Ref. [55]), as well as those that focus on the social formation of regional groups and inter-group boundaries reinforced by a variety of traditional practices (cf. Refs. [61,62]) – share a basic prediction. Over centuries or millennia of archaeological record formation, the relative frequencies of all microlithic types, more than the presence or absence of certain *fossiles directeurs*, should be of foremost concern in characterizing the techno-typological differences among assemblages (cf. Refs. [17,19,20]). On the one hand, generally common types in the study sample are expected to be more important and more statistically robust markers of variability in those assemblage-formation processes posited by the archaeologist to be theoretically relevant. On the other hand, generally rare types can nevertheless display significant patterns of variation in relative abundance; such *fossiles directeurs*, which distinguish clusters of assemblages, can be discovered analytically by examining between-group variation in d_{mij} for each type category (see Eq. (2)). We note that other approaches to measuring interassemblage difference may be appropriate. Chi-squared distance effectively quantifies the variability among assemblages described by detailed abundance profiles (Ref. [63], pp. 314–318). Also, when we can establish clearer prior expectations about the Epipaleolithic context of production, use and discard of different microlithic forms, it may then be appropriate to standardize the scale of variation in abundance values, treating each type-class as an independent, equally weighted variable. Nevertheless, D_{ij} is simple and straightforward to calculate, and it provides a logical starting point, from important archaeological theoretical perspectives, for a quantitative analysis of Epipaleolithic microlith type-abundance variability.

4.2. Analysis of variation between a priori defined groups

In testing the two proposed hypotheses (see Table 3), we can begin by considering whether the geographic criterion proposed by Henry (H_1: separating Early Epipaleolithic assemblages into Mediterranean vegetation zone [$n_1 = 11$] and steppic margin [$n_2 = 6$] categories) yields a statistically significant intergroup pattern of dissimilarity, and we simulate the random process in which each microlithic assemblage has the same likelihood of being associated with a given geographic zone. Is the observed intergroup difference between the 11 assemblages from the Mediterranean zone and the six assemblages from the semi-arid steppic margin significantly greater than what we would find if we repeatedly randomly drew groups of 11 and six assemblages from our study sample? Second, we follow this analysis of between-group variation by examining the alternative hypothesis (H_2) that Early Epipaleolithic microlithic assemblages observed to be rich in microburin debitage ($n_2 = 7$) are significantly more similar to one another than to assemblages categorized as microburin-poor ($n_1 = 10$). We test this proposition by simulating the random process in which each microlithic assemblage has an equal probability of associating with a high or low level of microburin debris. Is the between-cluster difference exhibited by separating the ten microburin-poor from the seven microburin-rich assemblages significantly larger than that obtained by randomly redrawing groups of ten and seven assemblages? The results of this analysis will provide a preliminary test of Henry's [43] hypothesis (H_1) that a social boundary between the Mediterranean and steppe zones persisted from the Early through the Middle Epipaleolithic.

By repeatedly drawing random combinations of the group sizes defined under each hypothesis, we can employ computer simulation to generate a distribution of random between-group distances in the context of the observed data. Thus, we can obtain an accurate estimate of the probability, P, that our observed intergroup distance is no greater than what we would find by our null hypotheses of random process. We define the multivariate characterization of each observed archaeological assemblage by its relative microlith type-frequency distribution. We calculate the dissimilarity index, D_{ij}, as defined in Eq. (1), for each pair of assemblages. We then determine the distance between groups as the average between-group dissimilarity,

$$\mathrm{avBGD} = \frac{\mathrm{SBGD}}{\mathrm{NBGD}} \tag{3}$$

where SBGD is the sum of dissimilarities between those pairs of assemblages whose members are assigned to different groups, and NBGD is the count of pairs of assemblages with the members in different groups. The measure of avBGD directly provides our test statistic, and using computer simulation, we repeatedly permute group membership 1000 times. Each time, we hold group size constant, and we recalculate the between-group difference value, avBGD, for every random permutation. In this way, we simulate the probability distribution of avBGD, as predicted by the hypothesis of random assemblage grouping. If the average pairwise dissimilarity between the observed groups is equal to or exceeded by that found in the random groups fewer than 50 times, then we may conclude that the observed differences are statistically significant at the 5% level.

As outlined in the previous section, it is of interest to the archaeologist to examine the class-by-class components, d_{mij}, of the overall dissimilarity in relative type-frequency profiles, D_{ij}. We carry out more detailed analyses of between-group variation for individual typological categories. We measure the type-specific dissimilarity in a given pair of assemblages, defined in Eq. (2) as d_{mij}, and for each microlith type, we calculate the average pairwise dissimilarity between assemblages in different groups. Then, we estimate the statistical significance of each type-specific between-group value, using the same repeated permutation procedure described above. Thus, we obtain a breakdown of those tool types that are significantly associated, and those that significantly crosscut, the hypothesized grouping criteria.

4.3. Searching for significant clusters among assemblages characterized by typological count data

The 'a priori groups' analysis of between-group variation described in the previous section tests for statistical significance by simulating a distribution of between-cluster dissimilarities according to specific prior 'null' hypotheses of random archaeological record formation. These hypotheses maintain that each assemblage has an equal likelihood of being assigned to one of the specified clusters, but every cluster has places for only a fixed number of assemblages set by the researcher. This assumes that the archaeologist has independently determined a relevant way to divide the study sample. However, we are also interested in searching more exhaustively for a 'best' way to group assemblages in the study sample and thereby potentially gain considerable new information about multivariate patterns in our detailed relative-frequency profile dataset.

In searching for a 'best' grouping of the study assemblages, only the number of groups, k, would be defined a priori. The number of assemblages in each of the k groups could be left unspecified. Non-hierarchical cluster analysis provides an excellent multivariate framework in which the number of groups may be preset but the size of each group is not defined. We measure the difference between each pair of assemblages according to our dissimilarity metric, D_{ij}, tailored for relative frequency profile data. Then, we select a criterion according to which we seek to identify optimal separation between groups of assemblages; for instance, we could utilize the measurement of between-cluster separation adopted in the a priori groups application and seek the assignment of assemblages into k groups that maximizes the average between-group pairwise difference (avBGD). The optimal k-cluster solution is then sought through a heuristic iterative relocation algorithm [11,30,63]. When we make no prior specification about the size of the groups, there are more than 65,000 ways of dividing our study assemblages ($n = 17$) into two clusters. Out of this set of possible two-cluster solutions, the iterative-relocation non-hierarchical cluster analysis procedure can allow us to look for criteria better than those specified in H_1 (stable eco-geographic social boundary along the Mediterranean-arid steppe transition zone) or H_2 (persistence of microburin technique for distinguishing social groups, despite their shifting geographic distributions) that separate the sampled Early Epipaleolithic Levantine microlith assemblages into two groups.

However, there is reason to remain concerned that *no* grouping criteria can appropriately characterize the study sample. The analysis of variation between a priori groups does not exhaust the set of relevant, plausible random factors of archaeological record formation. Our study assemblages vary markedly from one another in the number of microlithic artifacts they contain (see Table 2). Assemblages could differ in size due to diverse, essentially random variables, including past sedimentation rates as well as contemporary excavation effort. These factors would have operated independently of the variation in cultural patterns that influenced the frequencies at which different kinds of microlithic elements were produced, used, and discarded during the Epipaleolithic period in a given site deposit. In the simplest scenario of random assemblage formation, a particular microlithic form would have been produced and used at identical relative frequency at each site, but during the periods of archaeological deposit formation and recovery, the artifact type would have been essentially sampled randomly from the same underlying regional Early Epipaleolithic microlith-type abundance distribution [13,47–49,53,58]. Given the geological age of the deposits and the patchiness of their contemporary exposure and recovery, we may still not be reasonably sure that the available regional archaeological sample of Levantine Early Epipaleolithic artifacts is representative of the typological diversity of prehistoric anthropogenically produced microlithic elements. In a non-hierarchical cluster analysis of the study sample, the between-group difference measured for best separation of the 17 assemblages into two groups might actually be within the range of variation we would expect if our observed microlith assemblages had been formed by drawing samples of different sizes from an underlying single microlith population. If we found this to be the case, then the archaeological arguments underlying H_1 and H_2 would actually involve reading significant anthropogenic patterns into data that sample a process of random assemblage formation.

Indeed, any cluster analysis algorithm will find an optimal way to group the assemblages or objects under study, even when the variation among them is purely random [1,11,57,63]. Within the context of the observed data, though, we can approach a meaningful statistical test of the simple 'null' model of random microlith assemblage formation. In the non-hierarchical cluster analysis procedure, we can use a computationally intensive approach to simulate randomly generated artifact assemblages with the same sizes as those observed. The simulated hypothetical assemblages are the result of shuffling together all of the tools in the regional study sample and randomly reassigning tools, regardless of their type, to an assemblage until its observed sample size is filled. We treat all possible sets of 'shuffled and re-dealt' assemblages as equally likely; we use computer simulation effectively to sample 1000 such randomly generated assemblage sets. In each randomized sample, the non-hierarchical cluster analysis algorithm searches for the optimal two-group solution. Thus, we simulate the distribution of the measured distances between the best two groups in randomly formed microlith assemblages. If we find that the 'null' hypothesis of random process is rejected at the 5% level – that is, if the random between-group dissimilarity is greater than or equal to the observed between-group dissimilarity in fewer than 50 of the 1000 iterations – then the researcher can proceed to interpret the best groupings in the observed data in the light of other explanatory hypotheses.

In order to analyze the possible arrangements of $n = 17$ assemblages into $k = 2$ groups – whether the assemblages are characterized by observed or simulated sets of artifacts – we begin by calculating D_{ij}, the index of dissimilarity in relative frequency distributions for each pair of assemblages. The matrix of pairwise differences between assemblages consists of dissimilarity index values between 0 and 1. In general, a good grouping entails that pairs of assemblages from within the same group have D_{ij} values closer to 0, while pairs with members in different groups have values closer to 1. A formal measure of the between-group dissimilarity – that is, the overall 'goodness of grouping' – can be described as an optimality criterion (OC). There is a wide variety of ways to define the OC, and different optimality criteria can have substantial effects on the cluster analysis results [14,56]. Commonly employed OCs in non-hierarchical cluster analysis (e.g. minimizing the within-group error sum of squares) tend to define the best grouping as one in which different groups have similar numbers of assemblages and evenly spaced (i.e. spherical) distributions of assemblages [1,63] (cf. Ref. [14]). It is possible, however, to define an OC that tends to find the best grouping in which the preponderance of assemblages are placed in one or a few groups, with the remainder in a much smaller group. Having a range of optimality criteria for determining the best k groups is especially relevant if we do not want to impose restrictions on the size or shape of the groups we aim to find. We note that both prior hypotheses under consideration (see Table 3) specify asymmetric groups, which differ in size by ratios of roughly 1.5:1 to 2:1.

In the search for better grouping criteria in the study assemblage, we have chosen to allow the optimality criterion for determining the best k groups to vary along a spectrum, from favoring even group sizes to emphasizing highly asymmetric group sizes. At one end of the spectrum, goodness of separating n assemblages into k groups can be measured when the optimality criterion of the SBGD, the sum of between group distances, is maximized. SBGD tends to find groupings in which each cluster contains a similar number of assemblages. At the other end of the OC spectrum is a related criterion, that of the average between group distance (avBGD), the measure of between-group dissimilarity that we employ in the a priori groups application (see Eq. (3)). In non-hierarchical cluster analysis, *with no prior specification* of the sizes of each of the k groups, avBGD tends to be maximum when there are one or two large groups, with the remaining, most unique assemblages placed into their own separate groups. Optimality criteria that achieve intermediates between these two extremes can be calculated with an exponential modifier X:

$$\mathrm{OC} = \frac{\mathrm{SBGD}}{(\mathrm{NBGD})^X} \tag{4}$$

$\mathrm{OC} = \mathrm{SBGD}$ when $X = 0$. $\mathrm{OC} = \mathrm{avBGD}$ when $X = 1$. With values of X around 0.5, maximizing the OC tends to find cluster solutions that contain groups of many different sizes. By varying the optimality criterion, we can evaluate how sensitive to OC choice a given cluster analysis result might be.

We run the non-hierarchical cluster analysis, searching for the optimal two-group classification of our study assemblages, 11 different times. In each run, we change the optimality criterion by increasing X an increment of 0.1, from 0 to 1. We begin by seeking the best classification defined by the sum of between-group differences (SBGD, $X = 0$); we conclude by searching for the optimal solution based on the average pairwise between-group difference (avBGD, $X = 1$). For each optimality criterion we adopt, we run the search for the best two-cluster solution in the observed data on 100 different random starts (for summaries of the iterative-relocation procedure, see Refs. [11,63]). Once the solution maximizing the OC value is found, we then simulate 1000 datasets according to the random model of assemblage formation described above, so that we can ascertain the statistical significance of the observed best OC value.

The computationally intensive feature of repeated random assemblage simulation offers a highly useful means of evaluating the statistical significance of grouping structure found among assemblages described quantitatively by their relative frequency distributions. In our case study, this simulation component of the non-hierarchical cluster analysis provides a comprehensive check of our initial analysis of variation among a priori specified groups.

Table 4
Matrix displaying results of the 'a priori groups' analysis of variation when the study assemblages are classified into two groups under H_1 (Mediterranean group, $n_1 = 11$; steppic group, $n_2 = 6$)

	Mediterranean	Steppe
Mediterranean	0.63	0.82
Steppe	1/1000	0.53

The value in the cell below the diagonal shows the simulated significance of the observed between-group dissimilarity (avBGD; see Eq. (3)). This significance value is the number of times out of 1000 reiterations that the avBGD value of the simulated random grouping equaled or exceeded the observed avBGD value. Values of 50/1000 or less are significant at the 5% level. The value in the cell above the diagonal is the average between group dissimilarity, avBGD (see Eq. (3)). avBGD includes the dissimilarities in microlith relative type-frequency distributions between those pairs of assemblages with one member in the row group and one member in the column group. The values in the cells on the diagonal are the average dissimilarities between pairs of assemblages within each group, avWGD.

5. Results

The outcome of applying the 'a priori groups' and cluster analysis applications to our case study is summarized in Tables 4–7 and in Fig. 1. Through the analysis of variation between prior specified groups (see Tables 4 and 5), we can reject the hypothesis of random cluster formation, that the grouping criteria specified under H_1 (geography) and H_2 (microburin technique) are no better than any random grouping factor. When the observed assemblages are repeatedly randomly shuffled into two clusters of the sizes expected according to H_1 ($n_1 = 11$ and $n_2 = 6$), as well as H_2 ($n_1 = 10$ and $n_2 = 7$), only one simulated grouping yields a value of avBGD greater than that observed for either of the groupings favored under the two prior hypotheses. At the same time, the two hypotheses display essentially identical between-group levels of dissimilarity ($\mathrm{avBGD} = 0.82$; see Tables 4 and 5). Within the context of the study sample, the 'a priori groups' analysis of variation between microlith type-frequency assemblages cannot discriminate between H_1 and H_2. This result has one of two likely explanations. The Nizzanan archaeological culture (see Table 1) might display some technological and formal microlithic features that link it to the chronologically earlier Nebekian technocomplex, but other features might tie it to the Kebaran technocomplex. Alternatively, the Nizzanan subsample

Table 5
Matrix displaying results of the 'a priori groups' analysis when the study assemblages are classified into two groups under H_2 (microburin-poor group, $n_1 = 10$; microburin-rich group, $n_2 = 7$)

	MB-poor	MB-rich
MB-poor	0.59	0.82
MB-rich	0/1000	0.63

See Table 4 for explanation of the values in the cells.

Table 6
Category-by-category breakdown of the partitioning of variation in relative microlith type frequencies, according to the 'a priori groups' analysis of variation

	avPWD	H_1: geography		H_2: microburin technique	
		avBGD (H_1)	Sig (H_1)	avBGD (H_2)	Sig (H_2)
Curved pointed bladelet	0.25	0.24	537	0.23	836
Backed and truncated bladelet	0.27	0.29	140	0.32	17
Quasi – trapeze – rectangle	0.23	0.37	0	0.32	2
Backed bladelet	0.06	0.06	426	0.06	143
Retouched bladelet	0.07	0.07	171	0.08	74
Truncated bladelet	0.04	0.04	368	0.04	124
Microgravette	0.08	0.07	869	0.08	282
Micropoint w/basal truncation	0.04	0.04	809	0.04	538
La Mouillah point	0.08	0.13	0	0.12	1
Scalene triangle	0.05	0.05	317	0.06	152
Isosceles triangle	0.04	0.04	609	0.05	141
Pointed backed bladelet	0.04	0.04	612	0.04	605
Bladelet w/inverse retouch	0.02	0.02	741	0.02	480
Double-retouched bladelet	0.02	0.02	380	0.02	361
Retouched bladelet w/proximal truncation	0.02	0.02	1000	0.02	1000
Proto-trapeze	0.02	0.02	60	0.02	110
Bladelet w/alternating retouch	0.03	0.02	665	0.03	742
Retouched truncated bladelet	0.01	0.01	399	0.01	272
Falita point	0.01	0.01	808	0.01	900
Retouched pointed bladelet	0.01	0.01	677	0.01	283
Double-retouched pointed bladelet	0.01	0.01	275	0.01	263
Protolunate	0.01	0.02	290	0.01	141
Scalene bladelet	0.00	0.00	894	0.00	604
Double backed bladelet	0.01	0.02	0	0.02	1
Trapeze	0.01	0.01	370	0.01	473
Lunate	0.00	0.00	1000	0.00	1000
El Wad point	0.00	0.01	119	0.01	132
Rectangle	0.00	0.00	371	0.00	425
Proto-rectangle	0.00	0.00	778	0.00	872
Qalkan point	0.00	0.00	346	0.00	451
Pointed backed bladelet w/basal truncation	0.00	0.00	1000	0.00	1000

For all pairs of assemblages in the study sample, avPWD is the average pairwise dissimilarity, d_{mij}, for the m^{th} type (see Eq. (2)). avBGD is the average d_{mij} value of the m^{th} type for pairs with members in different groups. 'Sig' refers to the number of times out of 1000 reiterations that the avBGD value of the simulated random grouping equaled or exceeded the observed avBGD value. 'Sig' values of 50 or less are significant at the 5% level. Microlith types shown in italics are significant under both H_1 and H_2. Microlith types shown underlined are significant only under H_2.

included here – comprised of Ein Gev IV and Wadi Jilat 6 Upper – is too small to influence the between-group dissimilarities observed between Nebekian (steppic margin) and Kebaran (Mediterranean zone) microlithic assemblages.

The results of the non-hierarchical cluster analysis, which involved relaxing prior constraints on the size of each group, also strongly indicate that we can reject the hypothesis of random microlith assemblage formation at a highly significant level. As illustrated in Fig. 1, we find highly significant two-group solutions across the spectrum of optimality criteria, as defined by the exponential coefficient X (see Eq. (4)). It is unambiguous that microlith types associate with particular assemblages in a highly non-random pattern. However, among the possible two-group classifications of the study sample analyzed (see Table 7), H_1 fails to provide the *best* between-cluster separation across the entire OC spectrum. The separation into microburin-rich and poor assemblages (H_2) was found to have the highest level of between-cluster separation across only a portion of the OC spectrum ($0.69 \leq X \leq 0.90$) that favors clusters of somewhat asymmetric membership sizes. In fact, the most robust two-cluster solution, indicated as solution A in Table 7, was not predicted by either prior hypothesis. This two-cluster arrangement distinguishes the study assemblages, as displayed in Table 7, as follows: group 1 tends to include assemblages relatively rich in backed and truncated bladelets, and group 2 tends to incorporate assemblages comparatively dominated by curved backed bladelets. This separation of assemblages likely follows a chronological trend. As documented at the Fazael sites – all assigned to the Kebaran archaeological culture – the stratigraphic sequence displays a decline through time in the frequency of curved backed bladelets, along with a rise in the proportion of

Table 7
Membership for the best assemblage groupings according to the non-hierarchical cluster analysis application

A. $0 \leq X \leq 0.68$		B. $0.69 \leq X \leq 0.90$		C. $0.91 \leq X \leq 1.00$	
Group I	Group II	Group I	Group II	Group I	Group II
Hay C	KDr 8	KDr 8	EG IV	KDr 8	EG IV
EG I	Fa IIIB	Hay C	Jil 6 U	Hay C	
EG II	UR IIa	EG I	Jil 6 M	EG I	
EG III	Jil 6 L	EG II	Jil 6 L	EG II	
EG IV	Jil 6 M	EG III	Uw 14 M	EG III	
Fa IIIA	Jil 6 U	Fa IIIB	Uw 14 U	Fa IIIB	
Fa VII	Uw 14M	Fa IIIA	Uw 18	Fa IIIA	
WH 26	Uw 14U	Fa VII		Fa VII	
	Uw 18	UR IIa		UR IIa	
		WH 26		WH 26	
				Jil 6 U	
				Jil 6 M	
				Jil 6 L	
				Uw 14 M	
				Uw 14 U	
				Uw 18	

With variation in the criterion for optimizing between-group dissimilarity (see Eq. (4) in the text), three different solutions were found, labeled A–C. Solution B is the two-group arrangement of the study assemblages predicted by hypothesis H_2.

backed truncated bladelets [68]. The result of cluster solution A supports a long-held suggestion that the Kebaran entity can be divided into early and late components, which display a significant shift in the relative abundance of key microlithic forms [4]. Cluster solution A also links the microlith type-frequency distributions from apparent early Kebaran assemblages to those from Nebekian assemblages, which have been dated to the earliest Epipaleolithic (see Table 1). However, this clustering solution cuts across the small Nizzanan subsample. The results may remain substantially sensitive to the number of assemblages sampled, but the finding of more than one highly significant clustering solution, in the scan across the OC spectrum (see Fig. 1), suggests generally that multiple non-random factors of assemblage formation operated at different temporal and geographic scales. The cluster analysis application with simulated statistical significance levels indicates that neither Henry's hypothesis (H_1) nor the alternative (H_2) provides the best single available argument to account for observed variation in Early Epipaleolithic microlith relative type-frequency distributions.

6. Discussion

In carrying out the case study, our approach has been to test statistical hypotheses that predict groupings of microlith assemblages, according to explicit models of how long-term cultural processes of social boundary maintenance structured interassemblage variation in artifact morphology and technology of manufacture. Which statistical methodology we employed depended considerably on how we decided to measure artifact variation. We chose to base our measurement of microlith variability on the standard Levantine Epipaleolithic typological framework. As reviewed above, substantial debate has persisted over which of two main anthropogenic factors predominantly structured technological

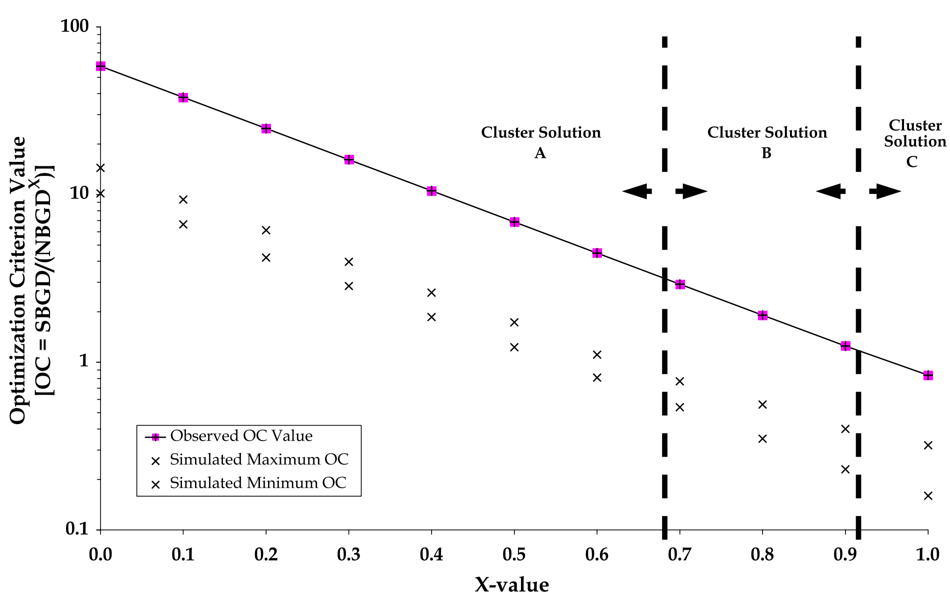

Fig. 1. Observed optimality criterion values compared with the distribution of OC values obtained when the hypothesis of random microlith assemblage formation was simulated. Note that for every value of the exponential factor X – defining a spectrum of OCs from avBGD to SBGD (see Eq. (4)) – the observed OC value far exceeded that of the OC values for all simulated random assemblages. The ranges of X over which each of three different best-grouping solutions were found are indicated as cluster solutions A, B and C. Solution B is that predicted by the prior hypothesis H_2.

and morphological variation in Epipaleolithic microlith assemblages: long-term practices of social boundary formation and maintenance (see Ref. [40]) vs. adaptation to local environmental conditions, involving the joint management of residential mobility, lithic raw material provisioning, and tool-use strategies in foraging and maintenance activities (see Refs. [3,29,55]). In fact, argument has further developed over the basic methodological usefulness of typology itself in characterizing Epipaleolithic chipped stone tool variability [3,23,55]. Despite their widespread application in description and analysis, Clark [23] has argued that standard knapped tool classifications may impose an arbitrary ad hoc pattern on the variability in Epipaleolithic artifact assemblages (see also Refs. [3,55]). In a recent review of chipped stone tool analysis in Middle Paleolithic research, Bisson (Ref. [18], p. 2) has made the relevant, concise observation that typology "inadvertently conceals important relationships between morphology, raw material, function, tool life-history, and possibly stylistic behavior by being an uncontrolled mixture of attributes related to all those factors." Standard lithic typologies are often insufficient to elucidate the many factors (residential mobility patterns, raw material availability and nodule size, blank production techniques, practical tool-use patterns in foraging, woodworking, etc., and social and symbolic constraints on the production of isochrestic variation) that may differentially contribute to the observed archaeological variability [18] (see also Refs. [28,54]). There is sound cause for caution in choosing methodologies for analyzing Levantine Epipaleolithic microlith assemblages. Indeed, an attribute approach that records multiple metric and non-metric features relating to each artifact's form and technique of manufacture could offer a more comprehensive basis for analyzing microlith variability than would a standard tool-type classification (e.g. Refs. [25,26,38]). However, detailed typological profiles from a regional sample of sites are already widely accessible in the literature, offering an immediately available resource for preliminary tests of hypotheses about what caused variation in microlith assemblages.

Seeking to take advantage of these published data, we developed two computationally intensive approaches that use the typological frequency profile format in measuring interassemblage variability and simulate processes of randomness in the formation of that variability. We have tested whether the standard microlith type profile, as a rough-and-ready measure, captures a significant level of variation among assemblage clusters, when clusters are predicted by prior models. If we had found that our predicted assemblage groups were no more distinct than those formed by random processes – that is, if the analytical results did not support our rejecting the null hypotheses of random assemblage formation at the $P \leq 0.05$ significance level – then it would be appropriate to investigate more rigorously whether the standard typology simply muddles the measurement of relevant variation. Such a statistical result could also plausibly indicate that the Early Epipaleolithic microlithic assemblages in the study sample are simply more similar to one another in their technology and morphology than archaeologists have assumed (see Ref. [23]). In fact, results of the analysis of variability between a priori groups of assemblages, as measured by dissimilarities in their type-frequency profiles, identified highly statistically significant differences. Henry's [43] model of long-term social-boundary maintenance between Mediterranean zone and semi-arid steppe forager communities in the Southern Levant – predicting the two-group hypothesis we have labeled H_1 (see Table 3) – explains a significant amount of variation among the 17 sampled Early Epipaleolithic microlith type-frequency distributions. However, the a priori groups analysis further shows that an alternative hypothesis of clustering, H_2, in which assemblages were separated into two groups based on whether or not they featured high amounts of microburin debris (as defined quantitatively above), irrespective of their geographic context, explains a virtually identical proportion of the measured variability. The steppe-zone assemblage group defined by H_1 is a subset of the more geographically dispersed microburin-rich assemblage group designated in H_2.

The cluster analysis results for the two-group solution confirm that Early Epipaleolithic microlith assemblages were formed by highly non-random processes. Yet, the available study sample and analytical approach are insufficient to elucidate precisely the impact of different archaeological structuring processes – including long-term stable social boundary maintenance, shifts in social networks and boundaries, and changes in economic and social role of microlith technology in Epipaleolithic hunter-gatherer communities. The cluster analysis results do illustrate that more than one significant factor shaped microlith type-frequency variation in Early Epipaleolithic assemblages.

From this perspective, Henry's [43] model has been a useful starting point, because it explicitly formulates the relationship between the geographic distributions of microlith types and long-term patterns of social interaction among mobile Early and Middle Epipaleolithic hunter-gatherer groups in the Southern Levant. According to our statistical analytical results, we argue that H_1 – predicted from Henry's [43] scenario of long-term social boundary maintenance along the Mediterranean/semi-arid steppe phytogeographic boundary – does not likely describe the dominant factor shaping microlith type-frequency variation during the Early and Middle Epipaleolithic on a regional Levantine scale. However, we may not exclude phytogeographic variation as one of the significant factors in structuring stable social boundaries between Early and Middle Epipaleolithic forager groups, at least within subregions of the Levant.

Because this article places the emphasis on analytical methodology, we leave a full discussion of the archaeological implications of our results for a subsequent publication, as we continue investigating how formal and technological variation among archaeological microlith assemblages might relate to Levantine Epipaleolithic trends in residential mobility and patterns of social interaction. We argue here that the results of our case study provide the basis for a practical conclusion, one that addresses the role of statistical methodology in problem-oriented Epipaleolithic research. Computationally intensive approaches to multivariate analysis indeed offer a useful way to test theoretically relevant hypotheses about the causes of technological and formal variation in chipped stone tool assemblages, when those assemblages are characterized quantitatively by their standard typological profiles.

7. Conclusion

In this article we have focused on the treatment of detailed relative type-frequency distributions in multivariate statistical analysis. While we have developed a case example involving Levantine Epipaleolithic stone tool typology, we have presented two statistical applications – the analysis of variation between a priori hypothesized groups of assemblages and non-hierarchical cluster analysis with simulation-based significance levels – that could be directly applied to studies of interassemblage variation in ceramic type abundance or in the relative frequencies of faunal or botanical taxa from archaeological contexts. These multivariate applications, available in the PHENCON2 package (http://www-personal.umich.edu/~gfred/), incorporate a computationally intensive methodology. In quantitative studies in archaeology as well as in other disciplines,

> we are frequently required to test for significance in situations where we know little or nothing about the expected distribution of the variables or statistics being tested, or where we know that the data to be analyzed do not meet the assumptions required for the customary statistical tests. (Ref. [64], p. 803)

This statement aptly characterizes a major practical challenge for analyzing and drawing inferences about archaeological assemblage grouping structure from relative frequency profiles. Recently, a growing body of work in quantitative analysis suggests that computationally intensive methods, involving the repeated simulation of comparative datasets according to a randomization model, can offer powerful solutions when no conventional statistical model is appropriate for our data and research questions (Ref. [64], pp. 803–825). In this article we have sought to build on earlier work on computationally intensive approaches to the analysis of variability in archaeological artifact frequency distributions [13,47–49] (see Ref. [27] for a recent application), expanding our scope to the multivariate analysis of grouping structure among assemblages. We have emphasized that with the data-processing power of contemporary personal computers, it is now highly feasible to tailor the computational approach in order to create probability distributions predicted from whatever assumptions of randomness, and for whatever statistical measures, the archaeologist might consider interesting. Computationally intensive approaches can transform what have classically been 'exploratory' multivariate analysis methods [11] into more specifically problem-oriented hypothesis-testing tools.

Hopefully, this establishes greater methodological transparency. In archaeology any multivariate quantitative study aims to compensate for the cost of the mathematical complexity or – as we have stressed in this article – the computational intensity involved in the analytical framework. The payoff ideally takes the form of real insight and clarification about the complex dataset and the archaeological problems connected to it (see Refs. [11,30,63]). With computationally intensive approaches to multivariate statistics, especially, archaeologists can gain a greater amount of control in linking statistical results to particular arguments for explaining variability in the archaeological record.

Acknowledgements

We presented an earlier version of this paper at the annual meeting of the Paleoanthropology Society, Denver, Colorado, in March 2002 [66], and we acknowledge initial support in the form of travel grants to AJS from the Paleoanthropology Society and the Rackham School of Graduate Studies, University of Michigan. This project has grown out of a seminar in multivariate and computationally intensive statistical approaches led by GFE and initially organized by Graciela Cabana and Liz Sobel. We thank Graciela and Liz for their efforts, and we thank the Department of Anthropology, University of Michigan, for providing support for the seminar. We are also especially grateful for the criticism and suggestions provided by the anonymous reviewers and Dr John Grattan.

References

[1] M.S. Aldenderfer, Methods of cluster validation for archaeology, World Archaeology 14 (1982) 61–72.

[3] C.M. Barton, M.P. Neeley, Phantom cultures of the Levantine Epipaleolithic, Antiquity 70 (1996) 139–147.
[4] O. Bar-Yosef, The Epi-paleolithic cultures of Palestine, PhD Thesis, The Hebrew University, 1970.
[5] O. Bar-Yosef, The Epi-paleolithic complexes in the Southern Levant, in: J. Cauvin, P. Sanlaville (Eds.), Préhistoire du Levant: Chronologie et l'Organisation de l'Espace Depuis les Origines jusqu'au VIe Millènaire, Colloques Internationaux du CNRS no. 598, Maison de l'Orient, Lyon, 1981, pp. 389–408.
[6] O. Bar-Yosef, The Late Pleistocene in the Levant, in: O. Soffer (Ed.), The Pleistocene Old World: Regional Perspectives, Plenum Press, New York, 1987, pp. 219–236.
[7] O. Bar-Yosef, A. Belfer-Cohen, The origins of sedentism and farming communities in the Levant, Journal of World Prehistory 3 (1989) 447–498.
[8] O. Bar-Yosef, A.N. Goring-Morris, Geometric Kebaran A occurrences, in: O. Bar-Yosef, J.L. Phillips (Eds.), Prehistoric Investigations in Gebel Maghara, Northern Sinai. Qedem 7, Monographs of the Institute of Archaeology, Hebrew University, Jerusalem, 1977, pp. 115–148.
[9] O. Bar-Yosef, S.L. Kuhn, The big deal about blades: laminar technologies and human evolution, American Anthropologist 101 (1999) 1–17.
[10] O. Bar-Yosef, J.C. Vogel, Relative and absolute chronology of the Epipaleolithic in the south of the Levant, in: O. Aurenche, J. Evin, F. Hours (Eds.), Chronologies du Proche Orient, BAR International Series 379, 1987, 219–246.
[11] M.J. Baxter, Exploratory Multivariate Analysis in Archaeology, Edinburgh University Press, Edinburgh, 1994.
[13] M.J. Baxter, Methodological issues in the study of assemblage diversity, American Antiquity 66 (2001) 715–725.
[14] M.J. Baxter, Statistical modelling of artefact compositional data, Archaeometry 43 (2001) 131–147.
[15] A. Belfer-Cohen, O. Bar-Yosef, Early sedentism in the Near East: a bumpy ride to village life, in: I. Kuijt (Ed.), Life in Neolithic Farming Communities: Social Organization, Identity, and Differentiation, Kluwer Academic/Plenum, New York, 2000, pp. 19–37.
[16] A. Belfer-Cohen, A.N. Goring-Morris, Why microliths? Microlithization in the Levant, in: R. Elston, S.L. Kuhn (Eds.), Thinking Small: Global Perspectives on Microlithic Technologies, Archaeology Papers (AP3A) series, American Anthropological Association, Washington, DC, 2002, pp. 57–68.
[17] L.R. Binford, S.R. Binford, A preliminary analysis of functional variability in the Mousterian of Levallois facies, American Anthropologist 68 (1966) 238–295.
[18] M.S. Bisson, Nineteenth century tools for twenty-first century archaeology? Why the Middle Paleolithic typology of François Bordes must be replaced, Journal of Archaeological Method and Theory 7 (2000) 1–48.
[19] F. Bordes, Mousterian cultures in France, Science 134 (1961) 803–810.
[20] F. Bordes, D. de Sonneville-Bordes, The significance of variability in Palaeolithic assemblages, World Archaeology 2 (1970) 61–73.
[21] G.W. Brainerd, The place of chronological ordering in archaeological analysis, American Antiquity 16 (1951) 301–313.
[22] B.F. Byrd, Late Quaternary hunter-gatherer complexes in the Levant between 20,000 and 10,000 BP, in: O. Bar-Yosef, R.S. Kra (Eds.), Late Quaternary Chronology and Paleoclimates of the Eastern Mediterranean, Radiocarbon, Department of Geosciences, The University of Arizona, Tucson, AZ, 1994, pp. 205–226.
[23] G.A. Clark, Plus Français que les Français, Antiquity 70 (1996) 138–139.
[24] J.D. Clark, J. Phillips, P.S. Staley, Interpretations of prehistoric technology from ancient Egyptian and other sources. Part I: ancient Egyptian bows and arrows and their relevance for African prehistory, Paléorient 2 (1974) 323–388.
[25] A.E. Close, The identification of style in lithic artifacts, World Archaeology 10 (1978) 223–237.
[26] A.E. Close, F. Wendorf, R. Schild, The Afian: A Study of Stylistic Variation in a Nilotic Industry, Southern Methodist University Press, Dallas, TX, 1979.
[27] G.W.G. Cochrane, Artefact attribute richness and sample size adequacy, Journal of Archaeological Science 3 (2003) 837–848.
[28] H. Dibble, N. Rolland, On assemblage variability in the Middle Paleolithic of western Europe: history, perspectives and a new synthesis, in: H. Dibble, P. Mellars (Eds.), The Middle Paleolithic: Adaptation, Behavior and Variability, University of Pennsylvania Museum Press, Philadelphia, PA, 1992, pp. 1–28.
[29] M.L. Donaldson, D.I. Olszewski, A preliminary analysis of the Step C microliths from Tor al-Tareeq, an Epipaleolithic site in the Wadi al-Hasa, Jordan, in: N.R. Coinman (Ed.), Excavations at Middle, Upper and Epipaleolithic Sites. Anthropological Research Papers no. 52, The Archaeology of the Wadi Al-Hasa, West-Central Jordan, 2, Arizona State University, Tucson, AZ, 2000, pp. 315–325.
[30] J.E. Doran, F.R. Hodson, Mathematics and Computers in Archaeology, Harvard University Press, Cambridge, MA, 1975.
[31] P.C. Edwards, Wadi Hammeh 27: an Early Natufian site at Pella, Jordan, in: O. Bar-Yosef, F.R. Valla (Eds.), The Natufian Culture in the Levant, International Monographs in Prehistory, Ann Arbor, MI, 1991, pp. 123–148.
[32] P.C. Edwards, P.G. Macumber, M.J. Head, The Early Epipaleolithic of Wadi al-Hammeh, Levant XXVIII (1996) 115–130.
[33] G.F. Estabrook (nd), Computer Programs of George F. Estabrook. Available from: http://www-personal.umich.edu/~gfred/.
[34] R.O. Fellner, Cultural Change and the Epipaleolithic of Palestine, BAR International Series 599, Tempus Reparatum, Oxford, 1995.
[36] A. Garrard, B. Byrd, A. Betts, Prehistoric environment and settlement in the Azraq Basin: an interim report on the 1984 excavation Season, Levant XVIII (1986) 5–24.
[37] A. Garrard, A. Betts, B. Byrd, C. Hunt, Prehistoric environment and settlement in the Azraq Basin: an interim report on the 1985 excavation season, Levant XIX (1987) 5–25.
[38] A.N. Goring-Morris, At the Edge: Terminal Pleistocene Hunter-Gatherers in the Negev and Sinai, BAR International Series, 1987, 361.
[39] A.N. Goring-Morris, Complex hunter-gatherers at the end of the Paleolithic (20,000–10,000 BP), in: T.E. Levy (Ed.), The Archaeology of Society in the Holy Land, Leicester University Press, London, 1995, pp. 141–168.
[40] A.N. Goring-Morris, A. Belfer-Cohen, The articulation of cultural processes and Late Quaternary environmental changes in Cisjordan, Paléorient 23 (1998) 71–93.
[41] D.O. Henry, The utilization of the microburin technique in the Levant, Paléorient 2 (1974) 389–398.
[42] D.O. Henry, From Foraging to Agriculture: The Levant at the End of the Ice Age, University of Pennsylvania Press, Philadelphia, PA, 1989.
[43] D.O. Henry, Prehistoric Cultural Ecology and Evolution: Insights from Southern Jordan, Plenum Press, New York, 1995.
[44] E. Hovers, O. Marder, Typo-chronology and absolute dating of the Kebaran Complex: implications from the second season of excavation at Urkan E-Rub IIA, Mitekufat Haeven: Journal of the Israel Prehistoric Society 24 (1991) 34–58.
[45] M.L. Julius, G.F. Estabrook, M.F. Edlund, E.F. Stoermer, Recognition of taxonomically significant clusters near the species level, using computationally intense methods, with examples from the *Stephanodiscus niagarae* complex (*Bacillariophyceae*), Journal of Phycology 33 (1997) 1049–1054.
[46] D. Kaufman, Hunter-gatherers of the Levantine Epipaleolithic: the socioecological origins of sedentism, Journal of Mediterranean Archaeology 5 (1992) 165–201.

[47] D. Kaufman, Measuring archaeological diversity: an application of the jackknife technique, American Antiquity 63 (1998) 73–85.
[48] K.W. Kintigh, Measuring archaeological diversity by comparison with simulated assemblages, American Antiquity 63 (1984) 73–85.
[49] K.W. Kintigh, Sample size, significance and measures of diversity, in: R.D. Leonard, G.T. Jones (Eds.), Quantifying Diversity in Archaeology, Cambridge University Press, Cambridge, 1989, pp. 25–36.
[51] D.L. Lehmer, Robinson's coefficient of agreement: a critique, American Antiquity 17 (1951) 151.
[52] A.E. Marks, P.A. Larson Jr., Test excavations at the Natufian site of Rosh Horesha, in: A.E. Marks (Ed.), The Avdat/Aqev Area, Part 2, and the Har Harif, Prehistory and Paleoenvironments in the Central Negev, Israel, II, Department of Anthropology, Institute for the Study of Earth and Man, Southern Methodist University, Dallas, TX, 1977, pp. 191–232.
[53] P.H. McCartney, M.G. Glass, Simulation models and the interpretation of archaeological diversity, American Antiquity 55 (1990) 521–536.
[54] P. Mellars, The Neanderthal Legacy, Princeton University Press, Princeton, NJ, 1996.
[55] M.P. Neeley, C.M. Barton, A new approach to interpreting late Pleistocene microlith industries in southwest Asia, Antiquity 68 (1994) 275–288.
[56] I. Papageorgiou, M.J. Baxter, M.A. Cau, Model-based cluster analysis of artefact compositional data, Archaeometry 43 (2001) 571–588.
[57] D.W. Read, Statistical methods and reasoning in archaeological research: a review of praxis and promise, Journal of Quantitative Anthropology 1 (1989) 5–78.
[58] D. Rhode, Measurement of archaeological diversity and the sample-size effect, American Antiquity 53 (1988) 708–716.
[59] W.S. Robinson, A method for chronologically ordering archaeological deposits, American Antiquity 16 (1951) 293–301.
[60] W.S. Robinson, Robinson's coefficient of agreement: a rejoinder, American Antiquity 18 (1952) 60–61.
[61] J.R. Sackett, Approaches to style in lithic archaeology, Journal of Anthropological Archaeology 1 (1982) 59–112.
[62] J.R. Sackett, Style and ethnicity in archaeology: the case for isochrestism, in: M. Conkey, C. Hastorf (Eds.), Uses of Style in Archaeology, Cambridge University Press, Cambridge, 1990, pp. 32–43.
[63] S. Shennan, Quantifying Archaeology, second ed., Edinburgh University Press, Edinburgh, 1997.
[64] R.R. Sokal, F.J. Rohlf, Biometry: The Principles and Practice of Statistics in Biological Research, third ed., W.H. Freeman, New York, 1995.
[65] M.N. Stevens, Operational sequences at Tor al Tareeq (WHS 1065), an Epipaleolithic site in the Wadi al-Hasa, West-Central Jordan, in: N.R. Coinman (Ed.), Excavations at Middle, Upper and Epipaleolithic Sites. Anthropological Research Papers no. 52, The Archaeology of the Wadi Al-Hasa, West-Central Jordan, 2, Arizona State University, Tucson, AZ, 2000, pp. 281–313.
[66] A.J. Stutz, G.F. Estabrook, A computationally intensive statistical technique for analyzing interassemblage variability in lithic type frequency distributions, with an application to the Epipaleolithic of the Levant. Paper presented at the Paleoanthropology Society Annual Meeting, Denver, 2002. Available from: http://www.paleoanthro.org/abst2002.htm.
[67] S. Yamada, Development of the neolithic: lithic use-wear analysis of major tool types in the Southern Levant, PhD Thesis, Harvard University, 2000.
[68] A.N. Goring-Morris, Late Quaternary sites in Wadi Fazael, Lower Jordon Valley, MA Thesis, The Hebrew University, 1980.
[69] J. Tixier, Typologie de L'epipaléolithique du Maghreb, Mémoire du Centre de Récherches anthropologiques, Préhistoriques et ethnographiques, Alger, 2, Arts et Métiers, Paris, 1963.