### Testing Models of Early Paleoindian Colonization and Adaptation Using Cladistics

5

Briggs Buchanan and Mark Collard

Clovis and related cultures from the Early Paleoindian period (ca. 11,500–10,500 radiocarbon years B.P. [RCYBP]) represent the first well-documented indications of human occupation in North America. Currently, there is considerable debate regarding the origins of these cultures. The traditional model developed to account for them has been challenged, and a number of alternative hypotheses have been proposed. Here we report the results of a study in which various models that potentially account for the origins of Paleoindian groups were tested by applying the cladistic method of phylogenetic reconstruction to a data set derived from characteristics of Early Paleoindian projectile points.

#### MODELS OF PALEOINDIAN COLONIZATION OF NORTH AMERICA

The most widely accepted hypothesis for the origin of Paleoindians is commonly referred to as the "Clovis-First model." It holds that groups migrated to North America via Beringia, the landmass between Siberia and Alaska that was exposed by sea level regression during glacial intervals (Hopkins et al. 1982). These groups are believed to have traveled through Beringia as they pursued large game migrating east on a quest for forage. Once in Alaska, people could have gained entry to the Great Plains via the ice-free corridor that is hypothesized to have been open soon after 12,000 RCYBP (Catto 1996; White et al. 1985; but see Arnold 2002; Mandryk 2004; Mandryk et al. 2001).

According to the Clovis-First model, after reaching the Great Plains the Paleoindians spread rapidly across the continents, eventually reaching the tip of South America in little more than a few centuries (Fiedel 2000). Apparent similarities among Clovis and contemporaneous sites situated across the diverse environmental regions of North America provide the basis for a model of Paleoindians as highly mobile, wide-ranging biggame hunters (Haynes 1980; Kelly and Todd 1988; West 1996). Initial discoveries of Clovis projectile points, primarily on the High Plains and in the Southwest, were associated with mammoth remains (Cotter 1937, 1938; Figgins 1933; Haury 1953; Haury et al. 1959; Sellards 1938), prompting researchers to suggest that Early Paleoindians specialized in the hunting of big game. The rapid spread of Clovis hunters coincided with the extinction of Late Pleistocene megafauna, leading a number of researchers to support the "blitzkrieg" hypothesis, which causally links the spread of Paleoindian hunters with the demise of the megafauna (Alroy 2001; Fiedel and Haynes 2004; Haynes 2002; Martin 1967).

Competing theories of the peopling of the Americas propose early inland or coastal migrations from northeast Asia (e.g., Bryan 1969, 1978; Dixon 2001; Fladmark 1979; Gruhn 1988), and even the long-held notion of a trans-Atlantic voyage from Europe has been resurrected recently (Bradley and Stanford 2004; but see Straus 2000; Straus et al. 2005). Other hypotheses regard regional manifestations of Clovis as indications of time depth and initial colonization of those regions (Bonnichsen and Schneider 1999; Mason 1962; O'Brien and Wood 1998).

#### CLADISTICS IN ANTHROPOLOGICAL RESEARCH

First presented coherently in the 1950s (Hennig 1950, 1965, 1966), cladistics is now the dominant method of phylogenetic reconstruction used in biology. Based on a null model in which new taxa arise from the bifurcation of existing ones, cladistics defines phylogenetic relationship in terms of relative recency of common ancestry. Two taxa are deemed to be more closely related to one

#### 60 Cultural Transmission and Archaeology: Issues and Case Studies

another than either is to a third taxon if they share a common ancestor that is not also shared by the third taxon. The evidence for exclusive common ancestry is evolutionarily novel, or "derived," character states: Two taxa are inferred to share a common ancestor to the exclusion of a third taxon if they exhibit derived character states that are not also exhibited by the third taxon.

In its simplest form, cladistic analysis involves four steps. First, a character state data matrix is generated. This shows the states of the characters exhibited by each taxon. Second, the direction of evolutionary change among the states of each character is established. Several methods have been developed to facilitate this, including communality analysis (Eldredge and Cracraft 1980), ontogenetic analysis (Nelson 1978), and stratigraphic-sequence analysis (Nelson and Platnick 1981). Currently, the favored method is outgroup analysis (Arnold 1981), which entails examining a close relative of the ingroup taxa to determine which character states are derived (those found only in the ingroup) and which are ancestral (those found in both the ingroup and the outgroup). Having determined the probable direction of change for the character states, the third step is to construct a branching diagram of relationships for each character. This is done by joining the two most-derived taxa by two intersecting lines and then successively connecting the other taxa according to how derived they are. Each group of taxa defined by a set of intersecting lines corresponds to a "clade," and the diagram is referred to as a cladogram, or "phylogenetic tree." The fourth step is to compile an ensemble tree from the character trees. Ideally, the distribution of the character states among the taxa will be such that all the character trees imply relationships among the taxa that are congruent with one another. Normally, however, a number of the character trees will suggest relationships that are incompatible. This problem is overcome by generating an ensemble tree that is consistent with the largest number of characters and therefore requires the smallest number of ad hoc hypotheses of character change, or "homoplasies," to account for the distribution of character states among the taxa. Recently, a number of studies have appeared in which cladistic methods have been applied to cultural data in order to shed light on historical events (e.g., Collard and Shennan 2000; Foley 1987; Foley and Lahr 1997, 2003; Gray and

Jordan 2000; Holden 2002; O'Brien and Lyman 2003a, 2005; O'Brien et al. 2001; O'Brien et al. 2002; Rexová et al. 2003; Robson-Brown 1996; Shennan and Collard 2005; Tehrani and Collard 2002). These studies suggest that cladistics can be a useful tool for tackling certain cultural evolutionary problems. Most significant, cladistics offers a well-understood model that can be fitted to linguistic and material culture data sets in a straightforward manner. Where the fit between a cultural data set and the tree model is close, we can invoke the principle of parsimony and legitimately conclude that the similarities and differences among the cultural units are primarily the result of branching. Conversely, where there are numerous homoplasies and the fit between a cultural data set and the tree model is poor, we can justifiably infer that borrowing or convergent evolution played a more important role in generating the similarities and differences among the cultural units. The instances of homoplasy can then be investigated with biological phylogenetic methods that are not based on the bifurcating-tree model (Bryant et al. 2005; Greenhill and Gray 2005; Hendy and Penny 1992; Hurles et al. 2003; see also ch. 4).

#### **CLADISTICS AND PROJECTILE POINTS**

Cladistic methods have previously been applied to Paleoindian projectile points by O'Brien and colleagues (O'Brien and Lyman 2003a; O'Brien et al. 2001; O'Brien et al. 2002), who focused on points from the southeastern United States. Because the majority of Paleoindian points from this region are surface finds without associated dates, temporal relationships among traditional point types remain largely unknown, and the ordering of types in the region has been developed largely by borrowing temporal sequences from other parts of the country (primarily the Plains and Southwest). With this in mind, the primary goal of the analyses conducted by O'Brien and colleagues was to construct phylogenetic hypotheses that shed light on the evolution of projectile points in the region.

O'Brien et al. (2001) recorded three qualitative and five quantitative characters on a sample of 621 projectile points representing a range of traditional projectile point types, including the wellknown Clovis, Dalton, and Cumberland points. They then subjected the specimens to paradigmatic classification (Dunnell 1971) in order to cluster them into classes with unique combinations of character states. This resulted in 491 classes. O'Brien et al. discarded classes with less than four specimens in order to minimize the impact of idiosyncratic behaviors. This reduced the number of classes to 17, and it is these that were used as taxa in the subsequent phylogenetic analyses. O'Brien and Lyman (2003a) later extended the study by increasing and decreasing the number of taxa in an effort to assess the reliability of the clades formed using 17 taxa. They found that many of the clades remained intact when taxa were removed, although in the majority of cases several most-parsimonious trees were found.

Following from the work of O'Brien et al., we used cladistic methods to infer a model of the evolutionary changes in points recovered from across North America. We then compared these models with models of cultural transmission, routes of population dispersion, and adaptation to regional environments.

#### **Materials and Methods**

We used assemblages rather than paradigmatic classes as taxa. We did this because the goal of our study is to shed light on human population history, and we think that assemblages are better proxies for past populations than classes obtained through paradigmatic classification. Using assemblages as taxa could potentially be problematic if some of the assemblages were created through multiple occupations over long periods of time, since it would increase the chances of conflating multiple technological lineages within a single assemblage. However, we suspect that the likelihood of different technological lineages occurring

Table 5.1. Projectile-Point Assemblages from Early Paleoindian Sites Included in the Analysis by State or Province, Region, and the Number of Complete or Mostly Complete Points Analyzed in Each Assemblage.

Site Site	State or Province	Region <i>a</i>	Number of Points in Analysis
Anzick	Thirty-se TM landmark	NP	1991; O'Brien et al. 2006; Willia (1991)
Blackwater Draw	NM 🕤	SPoliton	However, neither regionacts a well-docu
Bull Brook I	MA	NE	39
Bull Brook II	MA	NE	the maximum 2 inder of ordered states -
Butler	MI	GL	and incoming all 4 although an and all a statistical of the statistica
Cactus Hill	VA	MA	be appreciable on method retains the
Colby	WY	NP	er of states and the shirty of suprabilitiers
Debert	NS	NE	ien chemicters 6 e treatéd és multistate
Dent	СО	NP	section and size 2. The method pounds data
Domebo	OK	SP	ot assemblage 4 which are treated here as
Drake	СО	NP	say in the 13 of state in the single charge
East Wenatchee	WA	NW	Konaxampla II unallogant leogth (OL)
Gainey -	MI market MI	GL	fan Blashman Honsy New Mexico, byr
Gault	TX	SP	o strong Thoug 2 monoto by the assemint
Kimmswick	MO	MC	automation and an analysis of the second second
Lamb	NY	GL	ative sample of 5 attack district by the dif-
Lehner	AZ	SW	10 Internet in the second s
Miami	TX	SP	an in a start and a subject of the start and a start
Murray Springs	AZ	SW	beening a second of the second s
Naco	AZ	SW	and the define in 8 in set of the
Rummells-Maske	IA	MC	10
Shoop	PA	MA	Mile and a state of the state o
Simon	ID	NW	
Vail	ME	NE	16
Whipple	NH	NE	and a standard the 2 but may red

Note: See Buchanan (2005) for assemblage references.

<sup>a</sup> Regional abbreviations: GL = Great Lakes, MA = Midatlantic, MC = Midcontinent, NE = Northeast,
 NP = Northern Plains, NW = Northwest, SP = Southern Plains, SW = Southwest.

in the assemblages is low given the overall short time span represented by the Early Paleoindian period.

We used 216 projectile points from 25 assemblages. An important objective in selecting the specimens was to sample Early Paleoindian point assemblages from across North America. Another consideration was that reliable chronological information should be available in the form of absolute or relative ages, association of points with extinct Pleistocene megafauna, or diagnostic points that have been securely dated to the Early Paleoindian period at other sites. Some level of published information regarding excavation or collection procedures and provenience also had to be available for an assemblage to be sampled.

Details of the assemblages are given in Table 5.1. In terms of regional coverage, perhaps the most noteworthy omissions are the Southeast and the Great Basin. Both regions contain points that may date to the Early Paleoindian period (Great Basin Stemmed in the Great Basin and a variety of types in the Southeast, including Cumberland, Redstone, and Quad [Beck and Jones 1997; Bryan 1991; O'Brien et al. 2001; Willig 1991]). However, neither region has a well-documented,

radiometrically dated, single-component Early Paleoindian site. Hence, they could not be included in the study.

To capture the fine details of shape variation among the projectile points we employed a digitizing method developed by Buchanan et al. (2007; see also Buchanan 2005). Based on recent work in biology (e.g., Adams et al. 2004; Bookstein 1982; Richtsmeier et al. 2002; Rohlf and Marcus 1993), the method entails photographing each specimen with a digital camera and then using a digitizing pad and software package to record a series of landmarks around its edges. Thereafter, Euclidean distances between pairs of landmarks that define what are deemed to be key aspects of projectile point form are computed. We elected to employ this method rather than the conventional caliper-based approach to measuring projectile points because it is both more precise and more accurate. In addition, the method allows area and other variables that are difficult or impossible to record with mechanical measuring devices (e.g., Dibble and Chase 1981) to be determined relatively easily.

Thirty-seven landmarks were recorded, and 11 interlandmark distances (hereafter "characters")

Table 5.2. Characters Used in Morphometric Analyses of Projectile Points.

Characters	Description where the proop With collespond (CTW nor and Lymm 2003ar dW area
PA	Square root of projectile-point area. Calculated as the area enclosed by the 32 landmarks outlining
EL	Average of right and left edge-boundary lengths. Edge-boundary length is calculated as the sum of interlandmark distances along the 13 landmarks that define each edge.
TB	Average of the right and left distances from the tip landmark to each of the basal landmarks.
TW	Average of character TB to the position of the maximum edge inflection along each projectile point edge.
BL	Average of the right and left distances between the position of the maximum edge inflection and the tip landmark.
MW	Average of the right and left distances between the positions of the maximum edge inflections to the midline (character ML).
BB	Base boundary length. Calculated as the sum of the interlandmark distances along the nine landmarks that define the basal concavity situated between the two basal landmarks.
LB	Base linear length. Calculated as the distance between the two basal landmarks.
ML	Midline length. Calculated as the distance from the tip landmark to the midpoint of the basal concavity (character BB).
OL	Overall length. Calculated as the distance from the tip landmark to the midpoint of the segment between the basal landmarks (character LB).
BW	Basal width at one-third the total length above the basal landmarks.
LT il data in events (e.g.	Average of the right and left distances from basal landmarks to the position at one-third the total length along the opposite edge boundaries.



were computed from the Cartesian coordinates of the landmarks. The characters are listed in Table 5.2 and depicted in Figure 5.1. They were designed to capture the complexity of specimen shape (character PA) as well as the main elements of point form (characters El through LT). Three characters (TW, BL, and MW) describe aspects of width, and four characters describe aspects of length (EL, TB, ML, and OL).

To carry out a cladistic analysis using morphometric characters it is necessary first to adjust each character to counter the confounding effects of size and then to convert the size-adjusted data into discrete character states (e.g., Collard and Wood 2000; Strait and Grine 2004). Accordingly, the 12 characters were transformed to "size-free" residuals by regressing each character on the first principal component derived from all the characters (Strauss 1985). The first principal component is a metavariable that expresses the multidimensional quality of size (Bookstein 1982; Rohlf and Bookstein 1987) and therefore is used as the generalized size estimate in the regression (Strauss 1985). Although a ratio-based size-adjustment method (dividing each character by the geometric mean of all the characters) is commonly used in biological anthropology (e.g., Collard and Wood

Figure 5.1. Image of Clovis point from Blackwater Draw, New Mexico, showing locations of 11 of the 12 characters (point area not shown) and locations of the three landmarks. Character initials: EL, edge boundary length; TB, tip landmark to base landmark; TW, width of tip to base length to maximum inflection position; BL, blade length; MW, maximum width; BB, base boundary length; LB, linear measure of base; ML, midline length; OL, overall length; BW, basal width across first third of point; and LT, length from base to one-third along opposite edge.

2000; Lycett and Collard 2005), this method does not account for allometric (size-related shape) differences. Analyses carried out prior to the study described here (Buchanan 2005) indicate that the projectile point data exhibit considerable allometric variation. Thus, the regression-based method was deemed to be more appropriate than the ratio-based method.

The size-corrected data were converted into discrete character states using Thiele's (1993) gap-weighting method. This method uses the following formula to compute range-standardized data:

#### $xs = (x - \min/\max - \min)^* n$

where x is the assemblage mean for a character and n is the maximum number of ordered states allowable in the cladistic programs (26 character states). The gap-weighting method retains the rank order of states and the sizes of gaps between states when characters are treated as multistate and ordered for analysis. The method permits data from point assemblages, which are treated here as analogous to taxa, to be coded into single character states. For example, overall point length (OL) is coded for Blackwater Draw, New Mexico, by taking the mean OL of the points in the assemblage (n = 24) and subtracting the minimum OL for the entire sample of points divided by the difference between the maximum and minimum OL for the entire sample of points and multiplying the product by 26. The resulting value is rounded to an integer and coded using a symbol.

In addition to the 11 size-free characters based on interlandmark distances and area, we recorded four of the characters employed by O'Brien et al. (2001)—overall base shape, outer tang angle, tang-tip shape, and presence/absence of fluting. The 16 characters were then employed in four sets of analyses. In the first, the data were subjected to the permutation-tail probability (PTP) test, in which a data set is reshuffled multiple times and the length of the most-parsimonious phylogenetic tree is computed after each permutation. Thereafter, the length of the most-parsimonious tree obtained from the unpermuted data is compared with the distribution of lengths of the most-parsimonious trees yielded by the permutations. If the original tree is shorter than 95 percent or more of the trees derived from the permutations, then the data set is considered to contain a phylogenetic signal. The PTP test was carried out using PAUP\* (Swofford 1998), and the data set was reshuffled 10,000 times.

In the second set of analyses, the data set was subjected to parsimony analysis, which identifies the tree(s) requiring the smallest number of ad hoc hypotheses of character state change to account for the distribution of character states among the projectile point assemblages. This analysis was also carried out in PAUP\*. The heuristic search routine was used to find the shortest tree(s). Heuristic search can handle large data sets using trial-and-error branch swapping starting from randomly devised trees to find locally optimal trees, but it does not guarantee finding the minimum-length tree (Kitching et al. 1998). We treated our metric characters as ordered, meaning they are allowed to change only in single steps forward or backward along the character-transformation series. Thus, for example, a change from state 1 to state 2 entails only one step, whereas a change from state 1 to state 6 involves five steps. We elected to treat the characters in this way because it seems reasonable to assume that change in point morphology was most likely gradual rather than abrupt. We considered three of the four categorical characters (base shape, tang-tip shape, and presence of fluting) as unordered (transformation between any two states entails the same cost). The fourth categorical character, outer tang angle, was treated as an ordered character.

Our approach to outgroup selection differed from the approach employed by O'Brien et al. (2001). Rather than using occurrence seriation, we selected an outgroup based on the published radiocarbon assays associated with the 25 assemblages. Considerations of the reliability of the dates (e.g., Bonnichsen and Will 1999; Haynes et al. 1984; Levine 1990; Taylor et al. 1996) suggest that the assemblage from Blackwater Draw has the oldest reliable minimum age of the assemblages. It should be noted that the use of Blackwater Draw as an outgroup does not imply that this site represents the *initial* source of a colonizing population for the Americas. Rather, it implies that because of its early age, Blackwater Draw retains the most ancestral character states of the assemblages in the study. It should also be noted that because radiocarbon assays are not associated with all the assemblages in our analysis, it is possible that Blackwater Draw may not be the oldest taxon in our sample.

To evaluate how well the most-parsimonious trees derived from the parsimony analysis explain the distribution of similarities and differences in the data set, we used two goodness-of-fit statistics, the consistency index (CI) and the retention index (RI). The CI assesses homoplasy as a fraction of the character change on a tree, and the RI measures the amount of similarity in a data set that can be retained as homologies on a tree. Unlike the CI, the RI does not reflect the presence of uninformative derived characters (those occurring in only a single taxon). The RI also is not sensitive to the number of characters or taxa employed and can therefore be compared among data sets (Collard et al. 2006; Sanderson and Donoghue 1989).

In the third set of analyses, the data set was subjected to the phylogenetic bootstrap, which is a method for assessing the confidence interval associated with a given clade (Cole et al. 2002; Collard and Wood 2000; Felsenstein 1985b). Bootstrap analysis provides a conservative test of the level of support in the data for individual clades of a tree. Using PAUP\*, 1,000 matrices were derived from the character state matrix by resampling with replacement. The new matrices were subjected to parsimony analysis, and a consensus tree of the most-parsimonious trees was computed using a confidence region of 50 percent for specific branches, following Holden (2002). Again, the Blackwater Draw assemblage was designated as the outgroup.

In the fourth set of analyses, the mostparsimonious and consensus bootstrap trees were compared with trees constructed to reflect several hypotheses that account for Clovis and related cultures, including cultural diffusion and environmental adaptation as well as several possible colonization routes. The strength of fit between the observed and hypothetical trees was measured using the Kishino-Hasegawa test (Kishino and Hasegawa 1989). Tree lengths, together with the accompanying statistics (standard deviation and t statistic) describing tree topologies, are used to measure the degree of fit between the best-fit tree and the hypothetical trees. By definition the observed tree based on the data is the best-fit tree against which explanatory trees representing the various hypotheses are compared. Trees that provide a good fit to the observed tree are considered good explanations for the data, whereas trees with increasingly greater tree lengths are considered relatively worse explanations (Jordan and Shennan 2003). The degree of statistical difference in tree topologies is determined by assessing the number of steps it takes the explanatory tree (these trees are not all fully resolved and may have several "bushes" showing unresolved splits) to fit the general topology of the observed tree. The Kishino-Hasegawa test provides a p value (based on a two-tailed t-test) with which to assess the significance of fit between the observed and explanatory trees.

The cultural diffusion hypothesis posits that technological attributes horizontally transmitted across neighboring groups will create a pattern of increasing change in point form over space (Guglielmino et al. 1995; Jordan and Shennan 2003). Accordingly, the tree representing the cultural diffusion hypothesis was constructed on the basis of geographic proximity. Specifically, it was derived from straight-line distances between sites measured in ArcView (Buchanan 2005).

The environmental adaptation hypothesis predicts that the point assemblages should cluster according to inferred environmental similarities

and differences. The tree was constructed according to the distribution of sites within eight broad paleoenvironmental regions, which were defined using paleovegetation reconstructions and general physiography (Table 5.3). The paleoenvironmental reconstruction is derived from the work of Adams and others (Adams and Faure 1997; Steele et al. 1998), who have synthesized data on the structure and physiognomy of the paleovegetative cover of North America during the Late Pleistocene. Estimates of temperature for this period were obtained from publications of the COHMAP working group (COHMAP 1988; Kutzbach et al. 1993). These paleovegetative and paleoclimate data were used in conjunction with physiographic provinces defined by Hunt (1967) in order to incorporate structural and topographic relief along with aspects of latitude into the defined regions. Next, paleoenvironmental reconstructions based on these data were used to estimate net primary production (nPP) for each region from modern estimates (Melillo et al. 1993). Annual nPP is the net amount of carbon captured by land plants through photosynthesis each year and was included as a proxy measure of the contribution of terrestrial game to the diet (Kelly 1995). Ignoring the contribution of marine animals was deemed to be reasonable based on reconstructions of Paleoindian subsistence behavior (Cannon and Meltzer 2004). The nPP estimates were used to classify the paleoenvironments of the defined regions and structure the hypothetical tree used in the hypothesis testing.

Trees representing colonization sequences were designed to replicate the entry models of Anderson

Table 5.3. Paleoenvironmental Regions Used in the Analysis. Physiographic Description, Paleoenvironmental Biome, and Net Primary Production (nPP) Estimates Are Listed for Each Paleoenvironmental Region.

Paleoenvironmental Region	Description	Biome	nPP
Midcontinent	Central Lowland–Interior Low Plateaus	Prairie	335
Midatlantic	Coastal Plain and Piedmont	Spruce forest	238
Northwest	Intermontane–Columbia Plateau	Semi-desert and mountain mosaic	230
Northern Plains	Interior Plains-Great Plains Province-northern	Dry steppe	214
Southern Plains	Interior Plains-Great Plains Province-southern	Dry steppe	214
Great Lakes	Great Lakes Section	Spruce forest	173
Northeast	New England Province	Parkland to tundra	147
Southwest	Southern Basin and Range-Sonoran Desert Section	Semi-desert	129

and Gillam (2000) relative to possible migration routes used by Early Paleoindians. The least-cost pathways outlined by Anderson and Gillam are based on the assumption that travel was limited by minimum cost and not distance, a reasonable assumption given a colonizing population. The three entry models presented by Anderson and Gillam-the ice-free corridor, the Northwest Coast, and the Isthmus of Panama-were transformed into hypothetical trees using the primary and regional least-cost routes to derive the sequence of events for each scheme (c.f. Gray and Jordan 2000). The pathways were used to order and branch the assemblages employed to construct the hypothetical trees representing each entry model. In cases where particular sites used in the analysis were not connected by pathways shown in the three models, the relationships within each region were estimated according to the location of the site in relation to the defined pathways.

The ice-free corridor model begins on the western coast of Alaska and enters the contiguous United States near the North Dakota–Montana border. The pathway branches first at the Anzick site followed by a long branch connecting the remainder of the assemblages in the Northern Plains and the Northwest regions (Figure 5.2). The next branch connects to the Southern Plains assemblages followed by the Midcontinent, Great Lakes, and then the bifurcation between the Southwest and the Midatlantic and Northeast regions. The Northwest Coast model begins at the mouth of the Columbia River, and the primary pathway moves from west to east with north-south movement to the Southern Plains and the Southwest (Figure 5.3). The Isthmus of Panama model posits that Paleoindian groups bypassed the United States and colonized North America from the south (Figure 5.4). The Isthmus of Panama model differs from the ice-free corridor model in that the first branch is a long one connecting the Southern Plains and Southwest regions; the remaining pathways are similar to those of the ice-free corridor model but in reverse.

#### RESULTS

1000

The permutation-tail probability test indicated that the original most-parsimonious tree was significantly shorter than the 1,000 permuted trees (p = .0001). The permuted trees ranged in length from 507 to 567 (mode = 541), whereas the shortest trees had lengths of 401. Thus, the PTP test

Figure 5.2. Ice-free corridor model showing least-costsolution pathways for North America with glacial (light gray) and pluvial lake (dark gray) boundaries reconstructed for 12,000 **RCYBP** (adapted from Anderson and Gillam 2000). *The large black triangle* represents the hypothesized starting point of entry; black dots are locations of Paleoindian sites identified in Anderson and Gillam's (2000) analysis. Sites discussed are labeled with initials (AZ, Anzick; BD, Blackwater Draw; BB I&II, Bull Brook I and Bull Brook II; CO, Colby; DB, Debert; DE, Dent; EW, East Wenatchee; GN, Gainey; KM, Kimmswick; LE, Lehner; SH, Shoop; VA, Vail).



Figure 5.3. Northwest Coast model showing leastcost-solution pathways for North America with glacial (light gray) and pluvial lake (dark gray) boundaries reconstructed for 12,000 RCYBP (adapted from Anderson and Gillam 2000). The large black triangle represents the hypothesized starting point of entry; black dots are locations of Paleoindian sites identified in Anderson and Gillam's (2000) analysis. Sites discussed are labeled with initials (AZ, Anzick; BD, Blackwater Draw; BB I&II, Bull Brook I and Bull Brook II; CO, Colby; DB, Debert; DE, Dent; EW, East Wenatchee; GN, Gainey; KM, Kimmswick; LE, Lehner; SH, Shoop; VA, Vail).

Figure 5.4. Isthmus of Panama model showing least-cost-solution pathways for North America with glacial (light gray) and pluvial lake (dark gray) boundaries reconstructed for 12,000 RCYBP (adapted from Anderson and Gillam 2000). The large black triangle represents the hypothesized starting point of entry; black dots are locations of Paleoindian sites identified in Anderson and Gillam's (2000) analysis. Sites discussed are labeled with initials (AZ, Anzick; BD, Blackwater Draw; BB I&II, Bull Brook I and Bull Brook II; CO, Colby; DB, Debert; DE, Dent; EW, East Wenatchee; GN, Gainey; KM, Kimmswick; LE, Lehner; SH, Shoop; VA, Vail).

Construction of the second s second s Second secon second sec suggested that a phylogenetic signal is present in the data set.

The parsimony analysis returned four equally parsimonious trees. All but one character (fluting) was parsimony informative. The most-parsimonious trees were 401 steps long and had CIs of .56 and RIs of .67. They disagreed regarding the relationships of the Debert and Vail assemblages



Early Paleoindian-period projectile point data (tree lengths = 401; CI = .56; RI = .67).

(Figures 5.5–5.6). In tree A, Debert is the sister taxon of a clade comprising Colby, East Wenatchee, Gault, Murray Springs, Vail, and Whipple, and Vail is positioned as the sister taxon of a clade consisting of Colby, East Wenatchee, Gault, Murray Springs, and Whipple. Within the latter clade, East Wenatchee and Whipple form a clade that is the sister taxon of a clade comprising



Figure 5.6. Maximum-parsimony trees C and D inferred using the heuristic search algorithm in PAUP\* for the Early Paleoindian-period projectile point data (tree lengths = 401; CI = .56; RI = .67).

70 Cultural Transmission and Archaeology: Issues and Case Studies



Figure 5.7. Strict-consensus tree of four maximum-parsimony trees inferred using the heuristic search algorithm in PAUP\* for the Early Paleoindian-period projectile point data (tree length = 401; CI = .56; RI = .67).

Colby, Gault, and Murray Springs, and Murray Springs is the sister taxon of a clade consisting of Colby and Gault. In tree B, Debert and Vail form a clade that is the sister taxon of a clade comprising Colby, East Wenatchee, Gault, Murray

Springs, and Whipple, and East Wenatchee and Whipple form a clade that is the sister taxon of a clade consisting of Colby, Gault, and Murray Springs. Within the latter clade, Murray Springs is located as the sister taxon of a clade compris-



Figure 5.8. Bootstrap 50 percent majority-rule tree for the Early Paleoindian-period projectile point data.

#### Testing Models of Early Paleoindian Colonization / Buchanan & Collard 71



Figure 5.9. Hypothetical "cultural affinity" tree representing the nearest-neighbor straight-line distances among assemblages.

ing Colby and Gault. In tree C, Vail is the sister taxon of a clade comprising Colby, Debert, East Wenatchee, Gault, Murray Springs, and Whipple. Debert is positioned as the sister taxon of a clade consisting of Colby, East Wenatchee, Gault, Murray Springs, and Whipple. Within the latter clade, East Wenatchee and Whipple form a clade that is the sister taxon of a clade comprising Colby, Gault, and Murray Springs, and Murray Springs is the sister taxon of a clade consisting of Colby and Gault. In tree D, Debert, East Wenatchee, Vail, and Whipple form one clade, and Colby, Gault, and Murray Springs form a second. Within the clade comprising Debert, East Wenatchee, Vail, and Whipple, Vail is positioned as the sister taxon of the other three assemblages, and Debert is located as the sister taxon of Whipple and East Wenatchee. Within the clade comprising Murray Springs, Colby, and Gault, Murray Springs is positioned as the sister taxon of East Wenatchee and Whipple.

A strict-consensus tree of the four most-parsimonious trees is presented in Figure 5.7. This tree consists of three main clades. The largest clade comprises Bull Brook I, Cactus Hill, Colby, Debert, East Wenatchee, Gault, Lamb, Lehner, Murray Springs, Rummells-Maske, Shoop, Vail, and Whipple. The second-largest clade consists of Butler, Domebo, Gainey, Kimmswick, Miami, and Naco. The remaining clade comprises Anzick, Bull Brook II, Dent, Drake, and Simon. None of these is represented entirely by a single site type (kill, camp, or cache) or paleoenvironment.

The 1,000-replication bootstrap analysis supported the general structure of the most-parsimonious trees (Figure 5.8). The support for the clades ranged from 52 to 84 percent. Clades with high bootstrap support (>70 percent) included one comprising East Wenatchee and Whipple and one comprising Miami and Gainey. Additionally, a clade formed by Shoop, Rummells-Maske, Lehner, Cactus Hill, Bull Brook I, Vail, Murray Springs, Gault, Colby, East Wenatchee, Whipple, and Debert appeared in 84 percent of the bootstrap trees. However, eight assemblages-Bull Brook II, Butler, Simon, Naco, Dent, Drake, Domebo, and Anzick-were included in clades in less than 50 percent of the bootstrap samples and therefore are shown in the bootstrap tree as undifferentiated from the outgroup. This suggests that interpretations based on the topology of the consensus tree should be considered with caution. With this in mind, both the equally parsimonious trees and the bootstrap tree were tested against the hypothetical trees.

The Kishino-Hasegawa tests involving the



Figure 5.10. Hypothetical "environmental adaptation" tree representing the nearest-neighbor distances of the environmental proxy measures among assemblages.

most-parsimonious trees, and the trees representing the cultural diffusion (Figure 5.9), environmental adaptation (Figure 5.10), and colonization hypotheses (Figures 5.11–5.13), indicated that all the hypothetical trees are significantly different from the four most-parsimonious trees (Table 5.4). Therefore, based on the p values, none of the hypotheses are as well supported as the most-parsimonious tree. Comparison of tree lengths provides another way to assess the fit of the hypothetical trees to the observed trees (Jordan and Shennan 2003).

DOMESTIC THE CAREES STREPTICE AT THE TOOLET STREET	Show a most similar bas	— Blackwater Draw
outoning mouth out to a torough to torough out being	dano p mor organ i ph	— Anzick
mous trees (Figure 2.8). The support for the	t a clade comprising	Cactus Hill
clades ranged from 52 to 84 percent. Claces with	av Springs and Murray	Shoop
high bootstrap support (>70 percent) included one	In of a clode consisting of	
comprising East Wenatchee and Wildpple and one	P. Deberi, Earl	Bull Brook II
comuting Mianu and Gainey. Additionally, a	hipple form one clade,	
clade formed by Sheep, Rummells Masler,	Auray Springs form a sec-	Vail
-Lehrer, C ettis Hill, Bull Brook I, Vail, Nurray	amprising Debert, East	Naco
Springs, Cauit, qolby, East Wenatchee, Whipple,	inpit, vail is positioned	Murray Springs
and Deber appeared in 84 percent of the boot-	cother three assemblages,	Lehner
strap trees How yes, eight assemblages-Bull	the sister taxon of	Lamb
Brook II, Butler Simor Mico, Derf. Drake,	tchee. Within the clade	Gainey
-Domebo, and Arzick-were included in clades in	iner Colby, and Gault,	
less than 50 percent of the buotstrap samples and	ioned as the sister taxon	Kimmswick
therefore file shown in the bootstran tree as undif-	Whipple.	— Miami
ferentiated from the outerout. This succests that	ee of the four most-parsi-	Gault
<sup>23</sup> aterpretations based on the choleov of the con-	ad in Finner 5.7 This line	- Dent
sensus tree should be considered with daution	tangent and and and and	- Drake
Which is a subbid be considered with participation	the second second second second	Colby
	Real And And And	East Wenatchee
i nees and me boorsteab nee were tested against	, USHID, JOHN, LOHID,	— Simon

Figure 5.11. Hypothetical tree representing the ice-free corridor, Northern Plains colonization route.

Testing Models of Early Paleoindian Colonization / Buchanan & Collard 73



Using this method, the ice-free corridor model has the shortest tree length compared with the observed tree and can be considered the best explanation for the projectile point data. The other hypothetical trees are increasingly longer and poorer fits to the observed tree. The

molecular deteristitutionibile bedrounierte efforgen isomeshave a steptificant medinarpellit dribippin ber vesitific anasier of taka und the dispression far Saltizzon and fiberachine's silp 89) rog ëssion far mulantie (filof). 56 demired from the profectized pointification greater friftilit costected suffreential pointification greater fills and take profectized pointification greater of the interpreter of boundplacy are the formation stationaria estantication and the prophery pointification stationaria estantication and the prophery are the formation stationaria estantication and the prophery are filled to a stationaria estantication and the prophery are the pomentari biogram (and the prophery filled to a stationaria are the pomentari biogram (and the prophery filled to a stationaria are the pomentari biogram (and the propher) and the stationaria are the pomentari biogram (and the propher) and the stationaria estationaria and the propher estationaria contrastication and the propher of the stationaria and the propher (the pomentari biogram (and the propher) and the propher of the stationaria and the propher of the stationaria and the propher of the propher of the stationaria and the propher of the propher of the stationaria and the propher of the

Northwest Coast, Isthmus of Panama, and geographic-proximity models are 3, 10, and 17 steps longer than the ice-free corridor model, respectively. The environmental adaptation model is the least likely explanation, as it is 68 steps longer than the ice-free corridor model and 290 steps Blackwater Draw Cactus Hill

her maion

Shoop Bull Brook I Bull Brook II Whipple Debert Vail Anzick Dent Drake Colby East Wenatchee Simon Lamb Butler Gainey Rummells-Maske Kimmswick Miami Gault Domebo Naco **Murray Springs** Lehner

Figure 5.13. Hypothetical tree representing the Isthmus of Panama colonization route.

 Table 5.4. Results of the Kishino-Hasegawa Tests Comparing the Four Equally Parsimonious

 Projectile-Point Cladograms to Hypothetical Cladograms.

Cladogram	Length	Difference	s.d. (diff)	t Min	р
Point size-free Cladogram 1	401	(best)		Ga	qui
II doc21 Hall	401	0	2.92	0.00	1.0000
3.qinW	401	0	2.92	0.00	1.0000
4 0000	401	0	5.27	0.00	1.0000
Ice-free corridor entry	623	222	65.35	3.40	0.0040 a
Northwest Coast entry	626	225	66.04	3.41	0.0039 a
Isthmus of Panama entry	633	232	63.78	3.64	0.0024 a
Geographic proximity	640	239	67.15	3.56	0.0029 a
Environmental adaptation	691	290	65.40	4.43	0.0005 a

<sup>a</sup> Significantly different from the observed best-point size-free cladogram.

#### longer than the observed tree.

Comparing the bootstrap tree with the hypothetical trees indicates that the hypothetical trees also are significantly different from the bootstrap tree at the .05 level (Table 5.5). Again using tree length to determine which model is the best fit, the ice-free corridor model exhibits the least number of steps. Comparisons of the hypothetical trees with the bootstrap tree exhibit less divergence compared with the previous tests because of the shorter length of the bootstrap tree, but the relationships among the hypothetical trees remain the same.

In sum, the results of both sets of tests indicate that the ice-free corridor entry is at present the best explanatory model for the size-free projectile point data.

#### DISCUSSION

The PTP test returned a significant p value, which suggests that the data set contains a phylogenetic signal. The goodness-of-fit indexes associated

with the most-parsimonious trees are consistent with this. The RIs are .67, which compares favorably with RIs derived from other cultural data sets and also with RIs yielded by biological data sets. For example, the mean RI obtained by Collard et al. (2006) from 21 cultural data sets is .59. The mean RI they obtained from 21 biological data sets is .61. An RI of .67 indicates that the Paleoindian projectile point data set has a comparatively high number of similarities that can be interpreted as shared derived character states. The CIs associated with the most-parsimonious trees are .56. Analyses carried out by Sanderson and Donoghue (1989) using 60 morphological and molecular data sets from a wide variety of organisms showed a significant inverse relationship between the number of taxa and CI. Using Sanderson and Donoghue's (1989) regression formula, the CI of .56 derived from the projectile point data is greater than the expected value of .48 for 25 taxa. This indicates that the projectile point data set contains a low level of homoplasy

Table 5.5. Results of the Kishino-Hasegawa Tests Comparing the Bootstrap-Supported Projectile-Point Cladogram to Hypothetical Cladograms.

Super-					iev 1
Cladogram	Length	Difference	s.d. (diff)	t Run	р
Bootstrap cladogram	476	(best)		Mia	mi h
Ice-free corridor entry	623	147	60.21	2.44	0.0275 a
Northwest Coast entry	626	150	60.80	2.47	0.0261 a
Isthmus of Panama entry	633	157	58.21	2.70	0.0166 a
Geographic proximity	640	164	61.46	2.67	0.0175 a
Environmental adaptation	691	215	58.17	3.70	0.0022 a

<sup>a</sup> Significantly different from the observed best-point size-free cladogram.



Figure 5.14. Strict-consensus maximum-parsimony tree for the Early Paleoindian-period projectile point data with regional labels (GL, Great Lakes; MA, Midatlantic; MC, Midcontinent; NE, Northeast; NP, Northern Plains; NW, Northwest; SP, Southern Plains; SW, Southwest).

compared with biological data sets with comparable dimensions.

Perhaps the most noteworthy aspect of the results of the bootstrap analysis is the low level of variability in support of different clades. Some clades of the 50 percent majority-rule bootstrap tree were supported by a large number of the replicates. For example, a clade consisting of Debert, Whipple, East Wenatchee, Colby, Gault, Murray Springs, Vail, Bull Brook I, Cactus Hill, Lehner, Rummells-Maske, and Shoop occurred in 84 percent of the replicates, and a clade comprising Miami and Gainey appeared in 82 percent of the replicates. However, four clades were found in less than 60 percent of the 1,000 bootstrap samples, and eight assemblages could not be separated from the outgroup. One possible explanation for the interclade variability in bootstrap support is that the homoplasies are strongly clustered in certain parts of the data set. Another possibility is that certain characters are much more prone to homoplastic change than others.

Interpreting the topology of the most-parsimonious and bootstrap trees is not straightforward. Examination of the clades supported in the bootstrap tree shows that most of the assemblages generally are not clustered according to paleoenvironmental region. The consensus tree exhibits some limited regional grouping within clades, but overall its topology does not appear to be explicable in terms of paleoenvironmental regions (Figure 5.14). Site type also does not appear to correlate with the clades on the bootstrap or consensus trees. Although site type can be a categorization that masks variation in activities conducted at particular sites, we used it as a coarse measure of the inferred primary uses of sites-for example, whether a site was used to cache raw materials, to kill game, or as a habitation. Cache assemblages (Anzick, East Wenatchee, Drake, Lamb, Rummells-Maske, and Simon) occur in two of the three largest clades in the consensus tree along with assemblages from habitation and kill sites. In the bootstrap tree, cache assemblages occur undifferentiated from the outgroup and clustered with a habitation assemblage in one of the clades. Assemblages from kill sites (Blackwater Draw, Colby, Dent, Domebo, Kimmswick; Lehner, Miami, Murray Springs, Naco, and Vail [the Murray Springs and Vail assemblages are associated with both kill and habitation activities; Gramly 1982; Haynes and

Huckell 2007]) occur in all three of the major clades in the consensus tree and in all of the wellsupported clades in the bootstrap tree along with habitation and cache assemblages. Similarly, assemblages from habitation sites (Bull Brook I and II, Butler, Cactus Hill, Debert, Gainey, Gault, Murray Springs, Shoop, Vail, and Whipple) cooccur with kill assemblages in all three major clades in the consensus tree and in all of the wellsupported clades in the bootstrap tree.

Hypothesis testing was used to compare the observed trees (the four most-parsimonious trees and the bootstrap tree) with hypothetical trees constructed to represent models of geographic proximity, environmental adaptation, and entry/migration. Results of the hypothesis testing suggest that the best-fit explanatory tree based on tree length is the ice-free corridor model, which proposes that Early Paleoindian groups entered continental North America between the Laurentide and Cordilleran ice sheets. However, it should be noted that support for the ice-free corridor model in the present study does not exclude the possibility that pre-Clovis groups may have already inhabited North America. Rather, what it suggests is that, if there were pre-Clovis populations in North America, it is unlikely that those in the Mid-Atlantic region (such as at Cactus Hill in Virginia) and the Northeast (such as at Meadowcroft Rockshelter in Pennsylvania) contributed to Early Paleoindian cultural variation. According to our analyses, Early Paleoindian cultures had their roots in the Northern Plains, which implies that any pre-Clovis populations in the Mid-Atlantic and Northeast regions were either replaced by or absorbed into a population expanding out of the Northern Plains.

A model invoking the movement of Early Paleoindian groups using a similar point style and dispersing across the continent is by no means novel, nor is locating the center of this dispersion in the north, close to the hypothesized opening of an ice-free corridor. However, our analysis provides a formal test of the predictions of some of the competing models in the debate concerning the modes and adaptations of Early Paleoindians and attempts to formulate some of these arguments in more explicit terms through the use of biological phylogenetics and the implementation

Naco, and Vail (the Murray Springs and with assemblages are associated with both kill and minimison activities Gramiv 1982. Havies and

of formal hypothesis testing. Even so, given the discrepancy in fit between the observed and hypothetical trees, we believe that these results may be refined with further work. Accordingly, in the next phase of our research we will explore the impact of various methodological choices made in the course of this analysis. We intend to focus on taxon formation, character selection, and outgroup selection. With respect to taxon formation, we plan to examine the effects of using different numbers of taxa as well as explore the effects of configuring taxa differently. Examination of character selection will focus on weighting or using only those characters that exhibited high retention and consistency indexes in this initial analysis to construct trees. We will also evaluate alternative outgroups for character polarization. Using cladistic methods to define cultural phylogenies is a complex process that involves the iterative practice of continually refining methods in order to improve the accuracy of results, but it is this transparent methodological process that is one of the major strengths of cladistics (see ch. 4).

#### ACKNOWLEDGMENTS

For access to collections we thank Joanne Dickenson, John Montgomery, C. Vance Haynes Jr., Mike Jacobs, Dennis Stanford, Peggi Jodry, Lynette Miller, Laura Phillips, Deborah Baroff, David Keenlyside, Malinda Blustain, Donald Slater, Mary Lou Curran, Karen Kramer, Melanie Tossell, William Eldridge, Bruce Bourque, Paula Work, Richard Boisvert, John Cordell, George Oberst, Phyllis Oppenheim, Donald Simons, Wayne Rummells, and Richard Maske. Financial support was provided to Buchanan through a Research, Project, and Travel Grant from the Office of Graduate Studies, University of New Mexico, and a Doctoral Dissertation Improvement Grant from the National Science Foundation. Buchanan's work on this chapter also was supported by a National Science Foundation International Research Fellowship and a SSHRC Fellowship. Collard is supported by SSHRC and by the Canada Research Chairs Program. We thank James Boone, Bruce Huckell, Michael O'Brien, Lawrence Straus, and Richard Strauss for their comments on earlier drafts. Last, we thank David Kilby for sharing numerous photos of points.

nious and bootstrap trees is not straightforward. Examination of the clades supported in the bootstrain tree shows that most of the assemblages.

# Cultural Transmission and Archaeology

# **Issues and Case Studies**

Edited by Michael J. O'Brien



# **Cultural Transmission and Archaeology**

## **ISSUES AND CASE STUDIES**

Michael J. O'Brien

The Society for American Archaeology, Washington, D.C. 20002 Copyright © 2008 by the Society for American Archaeology All rights reserved. Published 2008 Printed in the United States of America

M

Library of Congress Card Number: 2008921985 ISBN No. 0-932839-33-9

society for American Archaeolog