

Linguistic diversity of the Americas can be reconciled with a recent colonization

DANIEL NETTLE*

Merton College, Oxford, OX1 4JD, United Kingdom

Communicated by Colin Renfrew, University of Cambridge, Cambridge, United Kingdom, December 21, 1998 (received for review August 14, 1998)

ABSTRACT The Americas harbor a very great diversity of indigenous language stocks, many more than are found in any other continent. J. Nichols [(1990) *Language* 66, 475–521] has argued that this diversity indicates a great time depth of *in situ* evolution. She thus infers that the colonization of the Americas must have begun around 35,000 years ago. This estimate is much earlier than the date for which there is strong archaeological support, which does not much exceed 12,000 years. Nichols' assumption is that the diversity of linguistic stocks increases linearly with time. This paper compares the major continents of the world to show that this assumption is not correct. In fact, stock diversity is highest in the Americas, which are by consensus the youngest continents, intermediate in Australia and New Guinea, and lowest in Africa and Eurasia where the time depth is greatest. If anything, then, after an initial radiation, stock diversity decreases with time. A simple model is outlined that predicts these dynamics. It assumes that early in the peopling of continents, there are many unfilled niches for communities to live in, and so fissioning into new lineages is frequent. As the habitat is filled up, the rate of fissioning declines and lineage extinction becomes the dominant evolutionary force.

The question of when and how the Americas were colonized is one of the most intriguing in human prehistory and continues to generate a huge literature. The discipline that traditionally has dominated this literature is archaeology, for it is only archaeology that can provide direct evidence of prehistoric human presence in an area, and only archaeology that has an absolute method of dating such a presence (1–3). More recently, however, the considerations of archaeologists have been augmented by two other sources of information. The first of these is molecular genetics, which, by assuming approximate constancy in the mutation rate, can provide indirect methods of inferring the date of divergence of human populations (4–9). The second is comparative linguistics. Linguists have used the distribution of pre-Columbian language families in the Americas, in conjunction with inferences about rates and patterns of diversification, to attempt to reach back into the past (10–12). The best known conclusion to be drawn from the linguistic data is that of the innovative study by Johanna Nichols, who infers from the great linguistic diversity of the Americas that the time depth of human habitation must far exceed that accepted by the majority of archaeologists (10). In this paper, I argue that Nichols' assumptions lack empirical validity, and that the very linguistic data she discusses are equally compatible with, if not suggestive of, a recent colonization.

The Colonization of the Americas

The prehistory of the Americas has one reference point we can be absolutely sure of; humans were present in the midlatitudes

of the Americas by 11,200 years B.P. This date is the radiocarbon date of the earliest of numerous sites associated with the Clovis culture, which appears at this time (1, 13, 14). There is a slightly earlier culture in Alaska that represents the probable progenitor (15, 16). An ice-free corridor between the Laurentide and Cordilleran glaciers had opened in the millennia preceding 12,000 years ago, giving access to the North American plains from Alaska, while a land bridge from Siberia to Alaska was still open at this time, being destroyed a few hundred years later by rising sea levels. A consensus thus emerged among most archaeologists that the Clovis people represented the first Americans, having slipped in from Asia in the window between glaciation and the flooding of the land bridge (14).

Archaeological claims for a pre-Clovis presence have been frequent over the decades. Most of the sites proposed have proved problematic with respect to dating or interpretation. Of 50 pre-Clovis candidates identified in 1964, only four were still seriously considered in 1976 and none in 1984 (1). However, other candidates have sprung forward in the meantime. In particular, it has become increasingly clear that human occupation in South America was at least contemporaneous with the earliest Clovis dates (17, 18), or, in the case of the Monte Verde settlement in Chile, around 1,000 years earlier (19).

The South American sites only push the dates for the beginning of colonization back by a few centuries. However, claims of a much greater time depth, stretching back to before the last glacial maximum 20,000 years ago, frequently are made. It is clear that there could have been an earlier entry; the land bridge was open from 30,000 B.P., and anyway may not be a prerequisite, as crossings can be made over sea ice. The ice-free corridor would not be a constraint if the route was coastal, as some have proposed. Nonetheless, the burden of proof must fall on those who argue for a presence substantially earlier, given that we have abundant and direct evidence from Clovis times forward and very little evidence of anything much before.

In the absence of direct material evidence, arguments for an early colonization generally rest on indirect argumentation. For example, Clovis-period sites cover the entire expanse of the Americas almost instantaneously when they start to appear. Some investigators have claimed that demographic expansion across such a large area would take a much longer period (3, 20). Moreover, the fact that Monte Alegre and Monte Verde are different from Clovis and so much further south can be argued to imply that these sites represent a much earlier wave of inhabitants. However, these arguments appear to be invalidated by recent models that assume a density-dependent rate of population growth (and thus predict a rapid rate of increase when the continent is empty) and also factor in the different demographic potential of different habitats (and thus predict an early and dense population relatively far south rather than in the inhospitable north) (21).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

*To whom reprint requests should be addressed. e-mail: daniel.nettle@merton.ox.ac.uk.

Two other sources of ancillary evidence come from outside archaeology: molecular genetics and linguistics. Studies of mitochondrial DNA variation in contemporary and recent native American populations have yielded divergence times that are greater than 11,200 years, and frequently greater than 20,000 years, and these data have been claimed as evidence for a pre-Clovis colonization (4, 5, 6, 9). However, the interpretation of genetic coalescence dates is not straightforward, even leaving aside uncertainties in the mutation rate that greatly alter the dates produced. Coalescence dates and the date of colonization should be expected to coincide only if the colonization event was also an extremely severe population bottleneck; that is to say, the founding population was extremely small. If the founding population was moderately large, it would have brought significant diversity with it, and genetic coalescence would be back somewhere in the history of the founding population in Asia (8, 22). Genetic mismatch distributions showing a population expansion over 20,000 years ago (7) could be consistent with the Clovis chronology if that expansion had begun in the ancestral Asian population. In the absence of information about the number of colonists, and the size of the population from which they were drawn, the genetic evidence therefore is hard to interpret.

The second source of ancillary information is linguistics. Native America was home to a remarkable diversity of languages and language families. The radiation of language families is a signature of past population events, and so the language map can be used as a source of information about prehistory. Nichols (10) applies this reasoning to the problem of the colonization of the Americas and produces a colonization date of 35,000 years ago. This date is in the range of the genetic divergence times. This date could be argued to be a happy convergence of two independent lines of evidence. However, it may well be no more than the multiplication of uncertainties, because the genetic dates are not clearly related to colonization, and the linguistic dates are quite invalid. Not only does the Nichols model, with a small tweaking of parameters, generate any date between 12,000 and 91,000 years ago, but its assumptions lack general support, as I shall show.

The Distribution of Linguistic Diversity

Linguists identify groups of related languages at various taxonomic levels. Nichols surveys the world's linguistic diversity at a level she calls the stock (10, 23). This level is the deepest level reconstructible by the standard comparative method of historical linguistics (the existence of deeper nodes frequently is hypothesized; in no case, however, have they been reconstructed). Nichols argues that such units represent a time depth of divergence of around 5,000–8,000 years. This assertion, however, must be treated with suspicion, because, first, we have no fixed points at all against which to calibrate the date, and second, we do not know whether the rate of linguistic change is constant. It is most likely that it is not (24). It must be

emphasized, in any case, that the stock is defined by a degree of linguistic similarity and not by a known age.

There are around 250 stocks, so defined, in the world (refs. 10 and 23; for a discussion of the problems involved in counting linguistic units and the caveats required, see ref. 25). Their distribution across the continents is given in Table 1. More than 150 stocks are indigenous to the Americas, which makes them exceptionally rich in this type of linguistic diversity. This statement remains true when the sizes of the different continents are taken into account (as in the stock density figures in Table 1).

Nichols argues that linguistic lineages ramify at a roughly constant rate, which she estimates by using recent known families at 1.6 descendants per 5,000 years. By this reasoning, she concludes that if all the languages of the Americas apart from the Na-Dene and Eskimo-Aleut families, which are thought to reflect more recent entries, stem from a single lineage, then at least 50,000 years has elapsed since this lineage began to ramify. She concludes that multiple colonization is more likely than this incredible date, but, even given multiple colonization, she argues that the process must have begun much earlier than the Clovis time horizon, before the last glacial maximum, and perhaps as much as 35,000 years ago.

The underlying logic of Nichols' position is that the diversity of linguistic lineages in a continent increases linearly with time. She does allow that diversity eventually will reach an equilibrium level, but implies that the time depth required for this leveling off to occur is vast. Under the constant rate assumption, the great diversity of the Americas demonstrates an ancient origin. However, it has never been shown that diversity accumulates linearly with time. Theoretically, there is no reason to expect that this will be the case. Divergence of language lineages begins when some demographic or social event leads to the splitting apart of parts of a previously homogeneous community. There is no reason to believe that such events arise at a constant rate in the way that genetic mutations do.

Fortunately, the hypothesis of a constant rate of ramification is a straightforward one to test. We know for each continent the number and density of stocks. We also have archaeological estimates of the date of settlement for all the major continents (Table 1). These variables can be plotted against each other.

The data show that there is no tendency for the number of stocks in a continent to increase with time (Fig. 1). The most ancient continents have no more stocks than the younger ones; in fact, the rank correlation coefficient is negative, though not significantly so ($r_s = -0.46$, $n = 6$, not significant). The most striking feature is the exceptionally high diversity of the Americas, which is not mirrored on any older continent.

The absolute number of stocks is, however, less meaningful than their diversity relative to the size of the continent. Fig. 2 plots the stocks per million square kilometers for the different continents. (Note the logarithmic scale: New Guinea is extreme in diversity given its small size and otherwise would be well out of the distribution). Again, there is no positive

Table 1. Stock diversity and density for the major continental areas of the world

Area	Stocks	Stock density	Languages	Language/stock	Date of colonization
Africa	20	4.4	2,614	130.7	>100
S & SE Asia	10	3.8	1,998	199.8	60
New Guinea	27	227.3	1,109	41.1	50
Australia	15	13.0	234	15.6	50
N. Eurasia	18	3.3	732	40.7	40
Americas	157	27.2	1,219	7.8	12

Oceania is excluded because of small size. Sources: Stocks (23), with the separate entries in the source for North America, South America, and Mesoamerica amalgamated. Stock density (23), converted here from stocks per million square miles to stocks per million square kilometers. Languages (25), the language counts are obtained by summing country totals, and so some languages are counted twice. The figures are therefore slightly inflated, but remain useful in relative terms. Dates of settlement (26, 27), thousands of years ago.

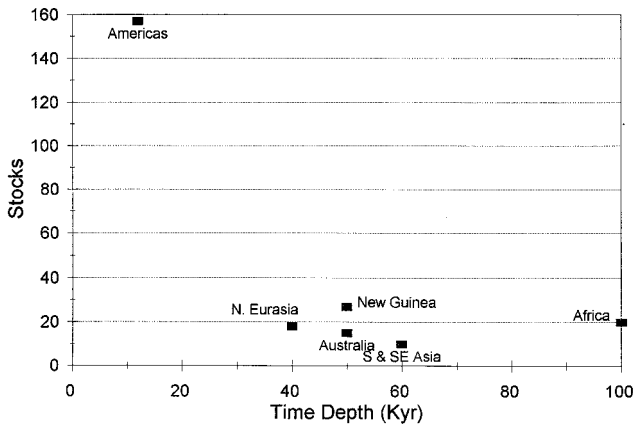


FIG. 1. The number of linguistic stocks against the approximate time depth of human habitation.

relationship between stock density and time ($r_s = -0.20, n = 6$, not significant).

One might argue, however, that controlling for the crude surface area of the continents is artificial. Different continents have different population densities and offer radically different prospects for human habitation. The tropics are home to many more communities per unit area than the extreme latitudes, and coasts are more densely peopled than interiors (23, 25, 28). New Guinea, for example, is very lush and offers ecological niches for well over 1,000 different communities (25). New Guinea's high stock diversity relative to crude surface area therefore is not surprising; an alternative measure is needed that reflects the density of stocks relative to potential for human habitation.

A more appropriate measure of relative linguistic diversity therefore might be the number of stocks divided by the number of human social groups who make a living in an area. The number of social groups can be estimated by the number of spoken languages, assuming a general, though not perfect, coincidence between social-economic and linguistic groupings (25). The best measure of relative stock diversity therefore is the number of languages per stock. Where this number is high there are many languages to each stock, and so relative stock diversity is low. Where the languages per stock figure is low, the average stock consists of just a few sister languages.

The number of languages per stock is strongly and positively related to time depth of habitation ($r_s = 0.84, n = 6, P < 0.05$; Fig. 3). That is, the older a continental population is, the fewer and larger the linguistic stocks it contains. Thus Africa and Eurasia have a few large stocks and the Americas

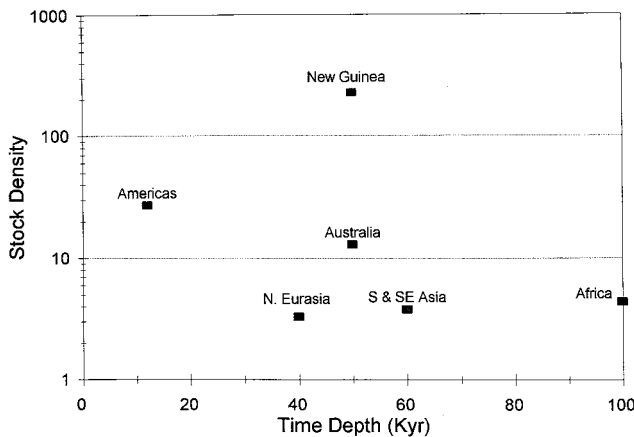


FIG. 2. The density of linguistic stocks (stocks/million square km) against the approximate time depth of human habitation.

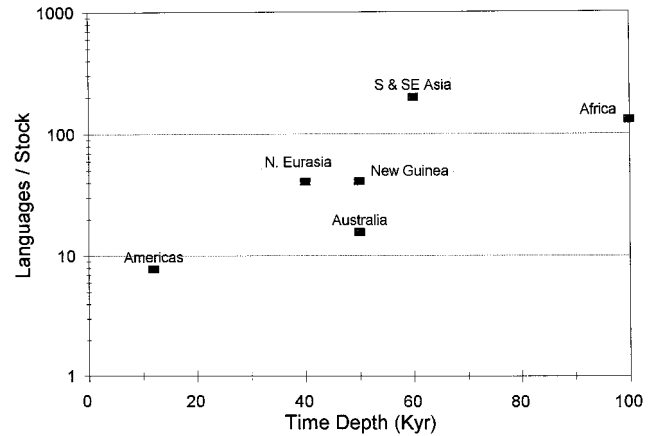


FIG. 3. The number of languages per stock in the major continents, against the approximate time depth of human habitation.

very many smaller ones. New Guinea and Australia are intermediate. If the Americas are excluded from the analysis, there is still an apparent trend toward lower stock density and more languages/stock with increasing time depth, though the data are too few to achieve statistical significance.

These data cannot settle the question of the date of colonization either way. The trend clearly suggests that the Americas are the most recently colonized continents, but cannot adjudicate between a time depth of 12,000 years and one of 20,000 or 30,000. However, it is clear that there is no basis for Nichols' claim that linguistic diversity indicates great time depth. Over the time scales we are examining, diversity of linguistic stocks does not increase with time, but rather seems to decrease. The great diversity of the Americas may not be evidence for their greater-than-realized antiquity, but rather, as Dixon (29) recently has argued, a normal symptom of their youth, and is entirely compatible with the Clovis or any other reasonable chronology.

The Evolution of Diversity

How can we explain the observed patterns of diversity? The long-term tendency appears to be for diversity to decline with time. However, there must be an initial period during which diversity increases. If the Clovis chronology is approximately right, for example, then there has been a period of rapid diversification from the few (perhaps single) incoming lineages in the Americas to the 150 seen in historical times. To reconcile the data with the archaeological chronology, a model is needed that predicts initial rapid increase in diversity, followed by slow decrease.

Dixon (29) provides a framework for constructing such a model. He argues that linguistic ramification occurs during infrequent and short periods of rapid population upheaval, which he calls punctuations. A good example of a punctuation is the development of agriculture in Eurasia, which, by causing a great increase in the population growth rate, sent a few populations rapidly expanding across the continent (30). As they expanded and fragmented, the languages of these populations split and began to diverge, giving the tree-like radiation we see so clearly in the Indo-European language family (31, 32). Between punctuations are long periods when all of the ecological niches of a continent are full, and no community has a sufficient demographic or technological advantage over its neighbors to displace them. During such periods (equilibria, in Dixon's parlance) the splitting off of new lineages is rare.

It is clear that the most significant punctuation there could possibly be is the entry of human beings into a new area. With so much empty habitat, population growth would be rapid, and groups of foragers would spread and fission at a very high rate

as they moved out through the continent (21). Each such split would be associated with the founding of a new linguistic lineage. As the available niches for independent foraging communities began to fill up, the rate of new fissionings would begin to decline. Population growth would slow and anyway would be absorbed into making existing communities larger, as population pressure and competition between groups drove cultural evolution toward more intensive resource extraction (33) and bigger and more complex societies (34). The rate of ramification of new lineages then can be modeled by any function, which at first rises steeply then levels off with time, such as Eq. 1.

$$\Delta S = A/t, \quad [1]$$

where ΔS is the number of new stocks produced, t is the time elapsed in thousands of years, and A is a constant reflecting the size of the land mass. We have to amend this equation slightly to allow for the fact that separate stocks are not produced overnight. Rather, it takes, according to Nichols, 5,000–8,000 years after the branching point for sufficient evolutionary change to accrue for linguists to recognize two lineages as separate stocks (10, 23). Thus the incoming stock remains unitary for 8,000 years, and so the time variable in Eq. 1 must be lagged by this amount.

Dixon emphasizes that extinction is as significant an evolutionary force as ramification. Lineages of languages may become extinct if the communities speaking them suffer natural disasters, disperse to other groups, or are absorbed by expansionary neighbors. Such extinctions are not rare in the ethnographic record (35). Dixon also claims that stocks can disappear from the linguistic record when extensive diffusion and convergence with neighboring languages make their distinct origin impossible to detect. We must assume, then, that per unit time, every stock has a small probability of becoming extinct or disappearing from the record. Setting this at 5%, we have the rate of extinction given by Eq. 2.

$$e = 0.05 S, \quad [2]$$

where S is the number of stocks in the population.

Eqs. 1 and 2 can be used to model the number of stocks, S , in a notional continent over the millennia after colonization. The number of stocks at time point $t + 1$ will be given by Eq. 3.

$$S_{t+1} = S_t + \Delta S - e. \quad [3]$$

Starting with one stock, and the constant A set to 70, the model produces the trajectory against time shown in Fig. 4.

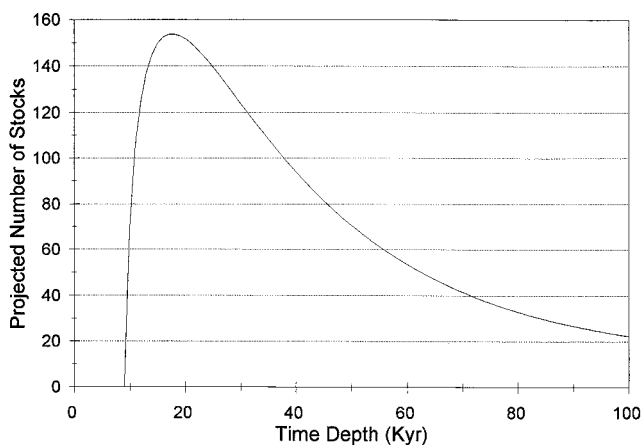


FIG. 4. The expected number of stocks in a continent against time depth of human habitation, using the model described by Eqs. 1–3.

Clearly, this model is simplistic and notional. Furthermore, the specific values have been chosen for illustrative purposes and have no independent justification. However, the precise values are unimportant. Any model in which the rate of ramification is high early in colonization when there are many empty niches, then levels off, while the extinction rate is proportional, will produce the same general pattern: an early steep rise, followed by a gradual decline.

Furthermore, this pattern is the one observed in the data. The linguistic history of the last few millennia in Africa and Eurasia, for example, is clearly one of the absorption of many small, distinct lineages by a few giants such as Bantu and Indo-European, with diversity declining overall (23, 32, 36). It would seem that the Americas in 1492, with their extraordinary stock diversity, were either at the peak or still in the steep rise of Fig. 4. Their linguistic palimpsest is thus exactly what a time depth of 13,000 or 14,000 years would seem, on the basis of this simple model, to predict.

The number of stocks is not the only argument put forward by Nichols in favor of an early date. She also urges the problem of the time required to reach South America in the terminal Pleistocene if something close to the Clovis chronology is correct; this issue, however, has been dealt with elsewhere (21, 37). Using a more specifically linguistic argument, she notes the high level of structural diversity in the languages of Americas, which she again assumes represents a great depth of diversification. The assumption underlying this argument is, though, much the same as that which underlies the argument from the number of stocks—that is, diversity accrues at a constant rate against time. This assumption is unlikely to be true. A very similar line of reasoning to that used for ramification of stocks here could be applied to structural diversification, which would be rapid early in a linguistic radiation as many independently evolving lineages were established and decline as diffusion and extinction began to bite. This reasoning is also consistent with global data; Africa and Eurasia are probably lower in structural diversity than the Pacific and Australia. Thus this argument, too, fails.

Conclusion

The problem of the colonization of the Americas will be definitively answered only by archaeology, because archaeology has direct methods for dating human presence. The purpose of this paper then is not to seek to prove that colonization was late. However, much has been made of the idea that the genetic and linguistic data force a radical revision of the archaeological picture and of the fact that Torroni's genetic and Nichols' linguistic inferences coincide (12). This paper shows that there is no basis for the argument from linguistic diversity for an early date. The linguistic data are quite compatible with any date, including Clovis, that emerges. The model presented here gives one way of reconciling the great linguistic diversity with the shallow time depth that the Americas have if the Clovis chronology, or something only slightly longer, is correct. Given that the genetic evidence is also equivocal, the idea that nonarchaeological considerations make belief in a late colonization untenable must be dismissed.

I thank Colin Renfrew, Robert Foley, Merrillyn Onisko, and two anonymous referees for their comments on an earlier version of this paper.

1. Meltzer, D. (1995) *Annu. Rev. Anthropol.* **24**, 21–45.
2. Dillehay, T. D. & Meltzer, D. J., eds. (1991) *First Americans: Search and Research* (CRC, Boca Raton, FL).
3. Dillehay, T. D., Calderón, G. A., Politis, G. & Beltrao, M. C. (1992) *J. World Prehis.* **6**, 145–204.
4. Torroni, A., Schurr, T. G., Cabell, M. F., Brown, M. D., Neel, J. V., Larsen, M., Smith, D. G., Vullo, C. M. & Wallace, D. C. (1993) *Am. J. Hum. Genet.* **53**, 563–590.

5. Torroni, A., Sukernik, R. I., Schurr, T. G., Starikovskaya, Y. B., Cabell, M. F., Crawford, M. H., Comuzzie, A. G. & Wallace, D. C. (1993) *Am. J. Hum. Genet.* **53**, 591–608.
6. Torroni, A., Neel, J. V., Barrantes, R., Schurr, T. G. & Wallace, D. C. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 1158–1162.
7. Stone, A. C. & Stoneking, M. (1998) *Am. J. Hum. Genet.* **62**, 1153–1170.
8. Ward, R. H., Redd, A., Valenica, D., Frazier, B. L. & Pääbo, S. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 8720–8724.
9. Gibbons, A. (1993) *Science* **259**, 312–313.
10. Nichols, J. (1990) *Language* **66**, 475–521.
11. Gruhn, R. (1988) *Man* **23**, 77–100.
12. Gibbons, A. (1998) *Science* **279**, 1306–1307.
13. Haury, E. H., Sayles, E. B., Wasley, W. W., Antevs, E. A. & Lance, J. F. (1959) *Am. Antiq.* **25**, 2–42.
14. Haynes, C. V. (1969) *Science* **166**, 709–715.
15. Hoffecker, J. F., Powers, W. R. & Goebel, T. (1993) *Science* **259**, 46–53.
16. Powers, W. R. & Hoffecker, J. F. (1989) *Am. Antiq.* **54**, 263–287.
17. Roosevelt, A. C., Dacosta, M. L., Machado, C. L., Michab, M., Mercier, N., Valladas, H., Feathers, J., Barnett, W., Dasilveira, M. I., Henderson, A., et al. (1996) *Science* **272**, 373–384.
18. Sandweiss, D. H., McInnis, H., Burger, R. L., Cano, A., Ojeda, B., Paredes, R., Sandweiss, M. & Glascock, M. D. (1998) *Science* **281**, 1830–1832.
19. Dillehay, T. D. (1996) *Monte Verde: A Late Pleistocene Settlement in Chile; Vol. 2: The Archaeological Context* (Smithsonian Institution, Washington, DC).
20. Whiteley, D. S. & Dorn, R. I. (1993) *Am. Antiq.* **58**, 626–647.
21. Steele, J., Adams, J. & Sluckin, T. (1998) *World Arch.* **30**, 286–305.
22. Pamilo, P. & Nei, M. (1988) *Mol. Biol. Evol.* **5**, 568–583.
23. Nichols, J. (1992) *Linguistic Diversity in Space and Time* (Univ. of Chicago Press, Chicago).
24. Nettle, D. (1999) *Lingua*, in press.
25. Nettle, D. (1999) *Linguistic Diversity* (Oxford Univ. Press, Oxford).
26. Diamond, J. (1997) *Guns, Germs and Steel: The Fates of Human Societies* (Jonathan Cape, London).
27. Roberts, R. G., Jones, R. & Smith, M. A. (1990) *Nature (London)* **345**, 153–156.
28. Mace, R. & Pagel, M. (1995) *Proc. R. Soc. London Ser. B* **261**, 117–121.
29. Dixon, R. M. W. (1997) *The Rise and Fall of Languages* (Cambridge Univ. Press, Cambridge, U.K.).
30. Ammerman, A. & Cavalli-Sforza, L. L. (1973) in *The Explanation of Culture Change: Models in Prehistory*, ed. Renfrew, C. (Duckworth, London), pp. 335–358.
31. Renfrew, C. (1987) *Archaeology and Language: The Puzzle of Indo-European Origins* (Jonathan Cape, London).
32. Renfrew, C. (1991) *Cambridge Archaeol. J.* **1**, 3–23.
33. Cohen, M. N. (1977) *The Food Crisis in Prehistory* (Yale Univ. Press, New Haven, CT).
34. Johnson, A. & Earle, T. (1987) *The Evolution of Human Societies: From Foraging Group to Agrarian State* (Stanford Univ. Press, Stanford, CA).
35. Soltis, J., Boyd, R. & Richerson, P. J. (1995) *Curr. Anthropol.* **36**, 473–494.
36. Diamond, J. (1997) *Nature (London)* **389**, 544–546.
37. Beaton, J. M. (1991) in *First Americans: Search and Research*, eds. Dillehay, T. D. & Meltzer D. J. (CRC, Boca Raton, FL), pp. 209–230.