Evolutionary Ecology Research, 2012, **14**: 1057–1067

Latitude, population size, and the language-farming dispersal hypothesis

Drew H. Bailey[1], Marcus J. Hamilton[2] and Robert S. Walker[3]

[1]*Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA,* [2]*Santa Fe Institute, Santa Fe, New Mexico, USA and* [3]*Department of Anthropology, University of Missouri, Columbia, Missouri, USA*

ABSTRACT

Question: The language-farming dispersal hypothesis postulates that the current biogeographic distribution of global ethnolinguistic diversity is due to prehistoric demic expansions of agricultural populations. Does human population size increase as diet is increasingly based on agriculture, and is this relation moderated by agricultural potential, as predicted by the language-farming dispersal hypothesis? Do these patterns hold constant within and between language families?

Data: We use a global sample of 805 subsistence-level human societies in 160 language families.

Method: We regress population size on latitude, agricultural dependence, and their interaction. We then analyse variation within and between language families.

Conclusions: Our results provide strong support for the language-farming dispersal hypothesis and suggest that while human population sizes increase rapidly with agriculture, much of this effect is driven by latitudinal variation in agricultural potential. The resulting patterns in the ethnographic and ethnolinguistic records reflect competition for space between populations with different subsistence strategies and varying competitive abilities.

Keywords: agriculture, biogeography, humans, language-farming dispersal hypothesis, population size.

INTRODUCTION

The successes of prehistoric human expansions around the world were enabled by cultural adaptation to local environmental conditions (Richerson and Boyd, 2005), first as hunter-gatherers and more recently as agriculturalists. In general, among traditional, non-industrial societies, nomadic hunter-gatherer societies are the least complex human systems and have ethnolinguistic populations that usually number less than a few thousand (Binford, 2001), while the most complex are large-scale agricultural states that often have populations several orders of magnitude larger (Carneiro, 1967; Currie and Mace, 2009; Turchin and Gavrilets, 2009).

Correspondence: D.H. Bailey, Department of Psychology, Carnegie-Mellon University, Pittsburgh, PA 15213, USA. E-mail: drewhalbailey@cmu.edu

Although colonization of the planet by hunter-gatherer populations took place over the last 40,000 years of the Pleistocene epoch, in many parts of the world these societies were subsequently replaced by agricultural societies during the Holocene. The language-farming dispersal hypothesis (Renfrew, 1987, 1989; Bellwood, 1997; Diamond, 1997; Bellwood and Renfrew, 2002; Diamond and Bellwood, 2003) suggests that these replacements were primarily demic events: the growth and divergence of agricultural populations as they expanded out from local centres of domestication, fuelled by increased production from farming, and outcompeted local hunter-gatherer populations. Importantly, these demic events supposedly involved not only the expansion of agricultural technologies and cultural behaviours, but also the languages spoken (and genes carried) by those populations. Thus, the language-farming hypothesis postulates that much of human ethnolinguistic variation we see today has its roots in prehistoric expansions of early farming societies during the Holocene.

The biogeographic distribution of human biocultural diversity follows well-established latitudinal gradients in biodiversity where species richness is highest near the equator and lowest towards the poles (Mace and Pagel, 1995; Rosenzweig, 1995; Nettle, 1998; Cashdan, 2001; Collard and Foley, 2002; Moore *et al.*, 2002; Hawkins *et al.*, 2003). While the ecological and evolutionary mechanisms that drive these latitudinal gradients is a major focus of research in biology, it is clear that in general the biogeography of biodiversity follows planetary gradients in solar radiation and rainfall (Allen *et al.*, 2002). This biological heterogeneity is essential for examining the language-farming hypothesis, as agricultural potential also varies across latitudes, with changing patterns of temperature, rainfall, growing seasons, and soil types. Thus, the effect of agricultural innovations on human population sizes (and densities) is likely strongly mediated by latitude. Indeed, traditionally there are two general themes underlying explanations for variation in human population sizes. The first emphasizes the general role of ecological constraints common to the demographics of all species, particularly energy availability in ecosystems. The second emphasizes behavioural and technological traits unique to humans, such as the invention of agriculture and the ability to manipulate both the community structure of ecosystems and the genetic structure of plants and animals to redirect primary production into consumable, storable foods. However, these two themes are not mutually exclusive. Here we take an integrative approach that considers not only the impact of ecological and economic variables on human population sizes around the world, but also their interactions. Here we test three hypotheses derived from the language-farming dispersal hypothesis: (1) human population sizes will be positively correlated with agricultural reliance, as increased levels of agricultural reliance should increase the harvestable food yield per unit area, thus increasing both total population sizes and local population densities; (2) this effect will vary with the environmental potential for agricultural production, which varies with latitude and associated bioclimatic factors; and (3) the size of populations within language families will be more similar to each other than chance alone if population size is an effect not only of ecological constraints, but also a shared cultural history of agricultural expansion.

METHODS

Variables

We compiled a global dataset of 805 societies with estimates of population size, percent dependence on agriculture, latitudinal occurrence, and basic linguistic taxonomy. Language families ($n = 160$) are from a modern, conservative compilation by Hammarström (2010). Data on the number of speakers for each language are from the most recent (16th) edition of the *Ethnologue* (Lewis, 2009), a comprehensive catalogue of the world's languages, and used as estimates for current population size. These population size estimates are log-tranformed – after adding 1, so that societies with languages that have recently gone extinct have a finite value [when the final model (Table 1) was re-estimated excluding populations with a population size of 0, the interpretations of the parameter estimates were not significantly altered] – to better approximate a normally distributed dependent variable (Fig. 1). Midpoint geographical coordinates of language home ranges are from Global Mapping International (http:www.gmi.org) and are used for absolute latitude. Data on dependence on agriculture were taken from the corrected *Ethnographic Atlas* (Gray, 1999), which approximates the percentage a society depends on agriculture for food. A value of 0 indicates 0–5% dependence on agriculture, a value of 1 indicates 6–15% dependence on agriculture, a value of 2 represents 16–25% dependence on agriculture, and so on with a value of 9 indicating 86–100% dependence on agriculture. Therefore, all populations that occur in the current analyses also appear in the *Ethnographic Atlas*. Descriptive statistics for agriculture, latitude, and population size variables are given in Table 2.

Analyses

We first examine the relationships among agriculture, latitude, and population size. Then, we test whether within- and between-language family components of agriculture and latitude contribute differently to population size. Finally, we examine how agriculture and

Table 1. Maximum likelihood estimates from model effects of latitude and agriculture on population size

	Value	s.e.	d.f.	*t*	*P*
(Intercept)	4.2597	0.2623	799	16.24	<0.0001
Family Lat	−0.0616	0.0121	155	−5.08	<0.0001
Agriculture	−0.0149	0.0382	799	−0.39	0.697
Family Lat^2	0.0011	0.0005	155	2.33	0.021
Within-family Lat	0.0287	0.0124	799	2.31	0.021
Within-family Lat^2	−0.0007	0.0008	799	−0.85	0.397
Family Lat × Agriculture	0.0072	0.0020	799	3.56	<0.0001
Family Lat^2 × Agriculture	−0.0001	0.0001	799	−1.46	0.146
Within-family Lat × Agriculture	−0.0033	0.0024	799	−1.39	0.165
Within-family Lat^2 × Agriculture	0.0006	0.0002	799	3.13	0.002

Note: Between-family effects. The effect of agriculture is non-significant, indicating that, at the equator, increases in agriculture were not associated with increases in population size; the effect of latitude is highly significant, indicating that, for hunter-gatherers, deviation from the equator is associated with smaller population sizes. However, this is qualified by a significant quadratic effect of latitude, indicating that, for hunter-gatherers, population size was smallest at moderate latitudes. The significant interaction between latitude and agriculture indicates that these patterns differed for agriculturalists, such that higher latitudes were associated with larger population sizes.

Within-family effects. These show a different pattern, where, for hunter-gatherers, higher within-family latitudes are associated with increases in population size, and for agriculturalists, relatively higher or lower latitudes are associated with increases in population size. The direction of both of these latitudinal relationships is exactly the opposite as found between language families.

Table 2. Descriptive statistics for latitude, agriculture, and population variables

Variable	mean	S.D.
Latitude (B)	24.17	17.21
Latitude (W)	0	6.08
Agriculture (B)	3.22	2.70
Agriculture (W)	0	1.56
Log Population Size	4.42	1.75

Note: B = between language families, W = within language families.

latitude interact at the within- and between-language family levels to predict population size. Analyses were run in the nlme package (Pinheiro *et al.*, 2012) in R. Family level variables were calculated as a variable's mean across all societies within the language family. Within-family variables were calculated as the difference between a society's value of a variable and the society's family level mean on that variable.

RESULTS

Figure 1 shows a series of frequency distributions characterizing our sample. Figures 1A and B show that both the sizes of populations and language families are unimodal, and approximately lognormal, though with noticeable fluctuations at small sizes. Figures 1C and D show that the distribution of populations and language families across the spectrum of current agricultural reliance is bimodal, with peaks around 0% agriculture (i.e. hunter-gatherer populations) and 60% agriculture. This bimodal pattern indicates that relatively few hunter-gatherer populations incorporate low levels of agricultural products into their diet, and that most subsistence populations are predominantly foragers or farmers. This is likely due to the fact that most (but not all) foraging populations are mobile, whereas most farming populations are sedentary. Moreover, while it is possible for sedentary farming populations to supplement their diet by foraging, it is less viable for mobile hunter-gatherers to incorporate farming into their diet, unless through trading with, or raiding from, farming societies for agricultural products. This suggests that, in effect, foraging and farming are relatively discrete subsistence activities in human societies.

Similarly, Figs. 1E and F show the distribution of populations and families across latitudes is also bimodal, though most clearly among language families. In both distributions, the highest frequencies of populations and language families are found at low latitudes with the second mode at mid latitudes, ~40–45°.

Agriculture and latitude

Agriculture and latitude were negatively correlated both between ($r = -0.52$, $P < 0.0001$) and within ($r = -0.08$, $P = 0.01$) language families, indicating that agriculture reliance increased with increasing distance from the equator with the between-family relationship noticeably stronger. The variance with latitude is higher at the family level than within families (Table 2). One potential explanation for this difference is the more restricted range

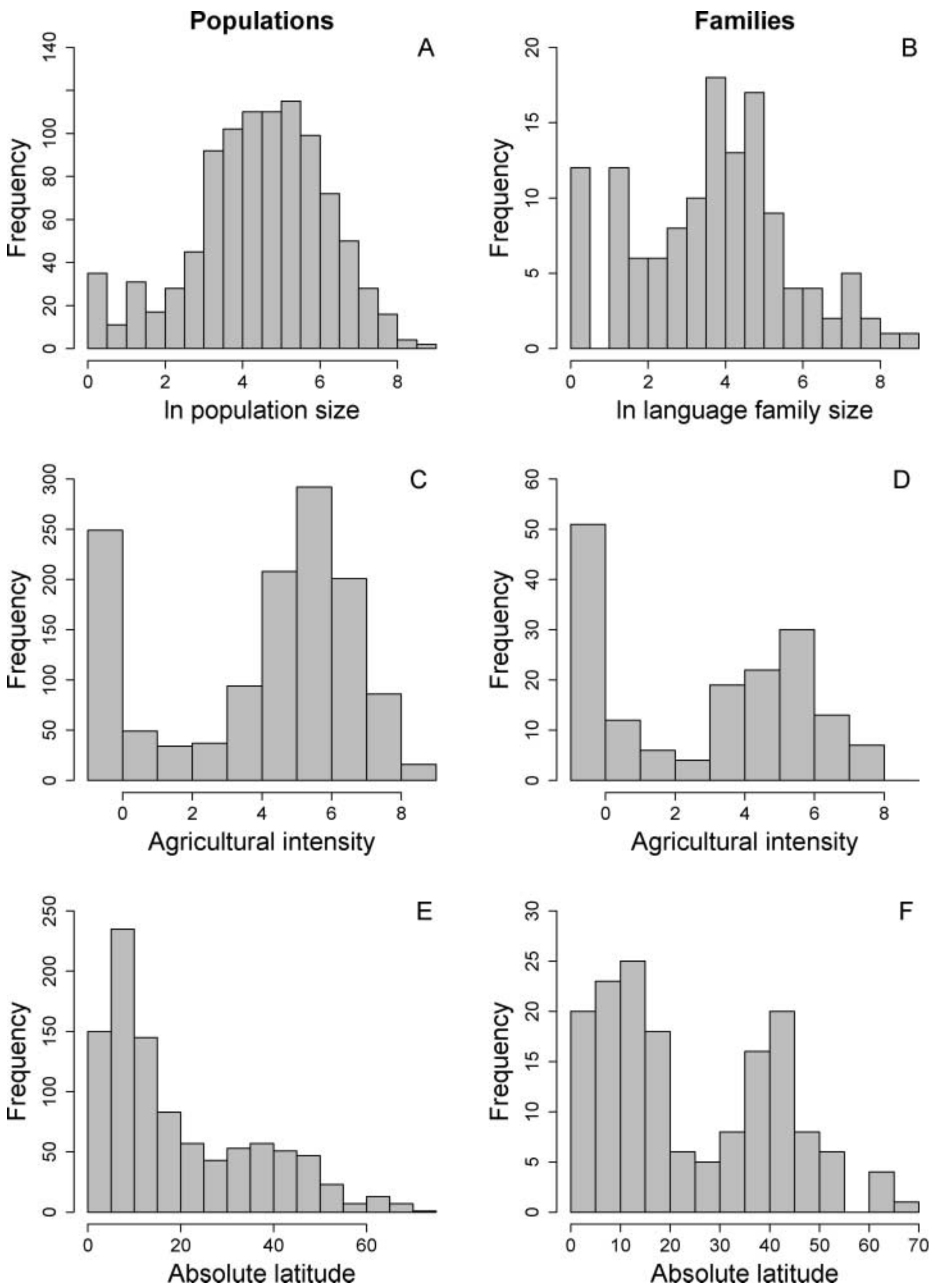

Fig. 1. Histograms of population and family level variables.

of values, which inflates the standard error of the unstandardized regression weight, but does not bias the estimate of the unstandardized regression weight. However, inconsistent with this explanation, the unstandardized regression weights (−0.081 and −0.021 between and within families, respectively), which are not biased by restriction of range, differed far more than their standard errors (0.011 and 0.008 between and within families, respectively), consistent with the hypothesis that different processes underlie covariance between agriculture and latitude between and within language families.

Agriculture, latitude, and population size

A random intercept model of log population size as a function of agricultural reliance with societies nested within language families showed that agriculture was a significant predictor of log population size ($t_{805} = 6.91$, $P < 0.0001$). A second model showed that latitude itself was not a significant predictor of population size ($t_{805} = -0.43$, $P = 0.67$). As we shall see, this is because complex interaction effects between agriculture and latitude confound this analysis.

Because population sizes are smaller at very high latitudes, and because agricultural potential is highest at temperate latitudes, we regressed population size on latitude, agriculture, their interaction, absolute latitude squared, and its interaction with agriculture. The effect of agriculture was non-significant ($t_{801} = 0.29$, $P = 0.77$), indicating that, at the equator, increases in agriculture were not associated with increases in population size. The effect of latitude was highly significant ($t_{801} = -4.22$, $P < 0.0001$), indicating that, for hunter-gatherers, distance from the equator was associated with smaller population sizes. However, this was qualified by a significant quadratic effect of latitude ($t_{801} = 3.27$, $P = 0.001$), indicating that, for hunter-gatherers, population size was smallest at intermediate latitudes. The significant interactions between latitude and agriculture ($t_{801} = 5.18$, $P < 0.0001$) and between latitude squared and agriculture ($t_{801} = -2.18$, $P = 0.03$) indicate that these patterns differed for agriculturalists, such that the highest population sizes were found at temperate latitudes. It should be noted that the interaction between latitude squared and agriculture would probably be larger (and the latitude × agriculture interaction would probably be smaller) if not for the small number of agricultural societies at very high latitudes.

Within- and between-family effects of latitude

Because population size was shown above to relate differently to latitude depending on the level of agricultural reliance and level of analysis, we next constructed a full model that examined the effects of agriculture at the society level and latitude within and across language families, including curvilinear relationships with latitude. The parameter estimates for this model are shown in Table 1. The family level latitude variable, agriculture, and their interaction resemble the parameter estimates described in the model in which latitude was not broken down into within- and between-family components. However, within-family latitudinal effects show a different pattern, where, for hunter-gatherers, higher within-family latitudes are associated with increases in population size, and for agriculturalists, relatively higher or lower latitudes are associated with increases in population size. Note that the direction of both of these latitudinal relationships is exactly the opposite as found between language families.

The aggregate of these complex agricultural and latitudinal interactions are illustrated in the contour plots in the three panels of Fig. 2. Figure 2A illustrates the effects of latitude and agricultural reliance on population size. The relation between population size and agriculture at a particular latitudinal band can be observed by choosing a level of latitude

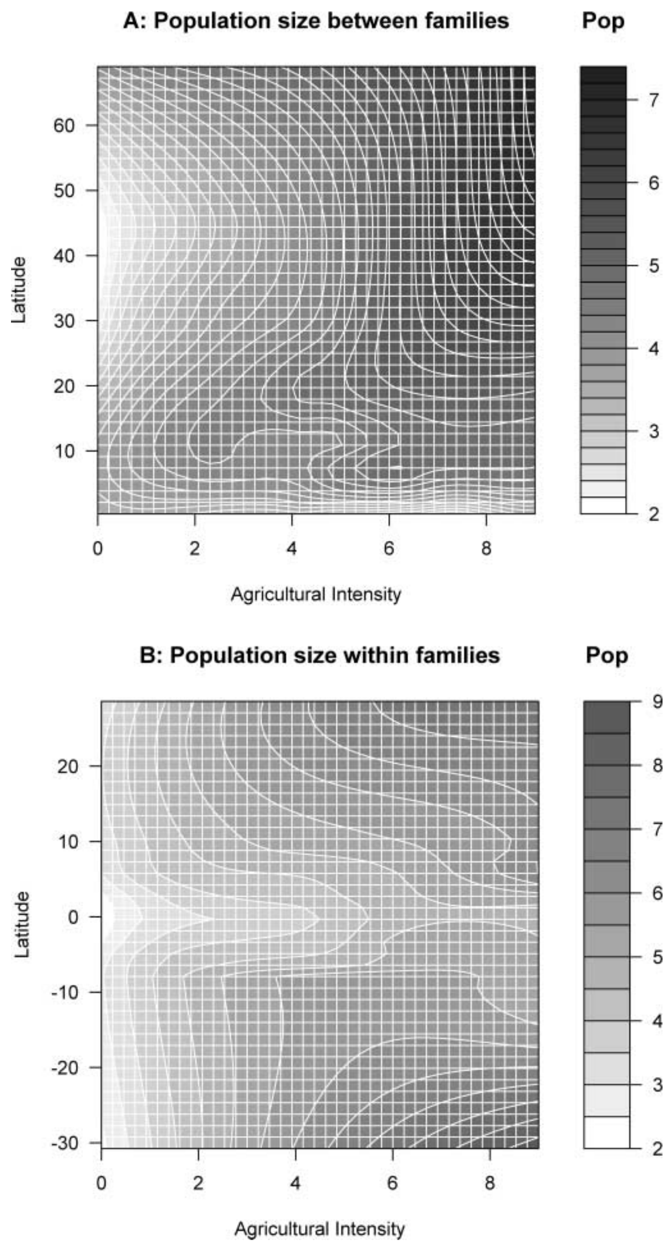

Fig. 2. Contour plots of the relationships between agriculture and latitude between and within language families. The figure illustrates the effects of latitude and agricultural reliance on population size between (A) and within (B) language families. The largest human populations are those temperate populations with greater than 50% reliance on farming. Within families, agricultural societies farthest from their family's centroid are largest. Figures are based on a loess approximation of the surface, with a span value of 0.5. Pop = ln population size.

on the vertical axis of Fig. 2A and moving from left to right along the horizontal axis. For example, at an absolute latitude of 40°, there is a strong relation between agricultural intensity and population size. At absolute latitudes of less than 10°, the relation is not clearly present. There is little systematic effect of latitude on population size: for any given level of agricultural reliance there is no systematic change in population size with increasing latitude. However, moving across the horizontal axis it is clear that population sizes increase rapidly with increasing agricultural reliance, such that the largest human populations are those temperate populations with greater than 50% reliance on farming (in this figure, arctic agricultural populations are also predicted to have large population sizes, but as stated above, this is because these societies do not exist). Within families (Fig. 2b), hunter-gatherer societies are largest when they are farthest away from the equator, from their language family's centroid. For agricultural societies, moving across the vertical axis, those groups farthest from their language family's centroid in either direction are the largest.

DISCUSSION

Our results provide empirical support for the language-farming dispersal hypothesis. Importantly, both the size of ethnolinguistic populations and language families increases exponentially with agricultural reliance, while the level of agricultural reliance itself varies with latitude. Among subsistence societies, the largest populations are those that are primarily agricultural and occur at mid latitudes. We infer that, when confronted with agricultural competitors, hunter-gatherer populations either do not occur at mid latitudes or are very small. However, foraging is a competitively viable option at both low and high latitudes.

The reasons for the competitive advantage of agriculture at mid latitudes are well understood. The potential for agricultural production is maximized at these latitudes due to the length of growing seasons, moderate temperatures, rainfall patterns, and extensive soil development. Not only can agriculture increase the energy intensity of production per unit area in these regions, but the moderate patterns of temperature and rainfall allow the long-term storage of agricultural products, which in turn fuels population growth, the ability to trade with neighbours, and thus the differential accumulation of wealth and power among individuals within societies (Renfrew, 1987, 1989; Bellwood, 1997; Diamond, 1997; Bellwood and Renfrew, 2002; Currie and Mace, 2009). Indeed, the largest politically complex prehistoric states occurred at mid latitudes (Currie and Mace, 2009). At high latitudes average temperatures are too low and growing seasons too short for agriculture, and at low latitudes soil development is often weak, agriculture can be stymied by weeds, herbivores, and diseases, and the ability to store surplus agricultural production in tropical climates is limited, and so the competitive advantage of farming over foraging is reduced in these circumstances.

These general patterns have interesting implications for debates over whether contemporary (or historically known) foraging populations exist only in environments of marginal potential for agriculture (Porter and Marlowe, 2007). Our data suggest that, in general, agricultural populations outcompete foragers at low-to-mid latitudes, i.e. only where the potential for agriculture is highest. However, the competitive advantage of agriculture is greatest (and near absolute) at mid latitudes, whereas at low latitudes foraging populations and agricultural populations can co-exist. Therefore, it is not simply the case that foragers exist only where agricultural populations cannot, but that at different latitudes alternative subsistence strategies have differential competitive advantages: at high latitudes foraging is the only option, at mid latitudes farming is the most successful and competitively dominant strategy, and at low latitudes both foraging and farming can be viable subsistence alternatives.

At a broader scale, our results suggest patterns of variation in human population sizes are driven by complex interactions between ecology and culture. In support of the language-farming dispersal hypothesis, agricultural reliance was positively related to population size. Absolute latitude by itself did not significantly predict population size. However, reliance on agriculture for sustenance significantly moderated this effect, such that latitude was positively related to population size for agriculturalists but negatively related to population size for hunter-gatherer societies.

While latitude and agriculture are strong predictors of population size, as expected, they vary in important ways depending on each other and the level of analysis (within- versus between-language family). Intriguingly, our model indicates that within language families, populations with relatively low levels of agriculture (hunter-gatherers) are largest when they are relatively farther from the equator. Moreover, societies with relatively high levels of agriculture for their language families have the largest populations when they are relatively farther from the central latitude of their language family, again in contrast to relationships across language families. One explanation for these complex patterns is population packing, or niche seeking. Successful societies with flexibility in their subsistence strategies may have migrated to locations where they were more likely to better monopolize resources by at least partially avoiding direct competition from other societies (primarily) within their language family. This effect may be most pronounced for agricultural societies because competition for lands is lowest farthest from the centre of the language family.

Our findings also raise several questions about cultural and genetic evolution, including the possibility of an opportunity for ecologically dependent multi-level selection, as different relations between ecology and population size were found within and between language families.

In a hunter-gatherer society within a family with diversity in the reliance on agriculture, a trait that might increase the likelihood of members of a society to migrate away from the equator (e.g. the personality trait openness to experience) will be selected against at the level of the language family, but will be selected for within language family. Dependence on ecological factors and multi-level selection may account for the maintenance of genetic variation within language families for such traits.

CONCLUSION

Our results strongly support the language-farming dispersal hypothesis, but shed light on the complex interaction among subsistence strategies, latitude, and population size. The spatial heterogeneity in population size we describe here is driven by complex interactions between ecology and economy, and is consistent with patterns found in non-human species: energy availability (whether ecologically or technologically determined) is a critical predictor of human population size. However, these results also reflect humans' unique ability to manipulate their ecology in ways that allow them to thrive under a variety of different conditions.

ACKNOWLEDGEMENTS

We thank Søren Wichmann and Zach Winkler for their help in language matching and Mark Flinn for helpful discussions. Financial support was provided by the Research Board and Arts & Science Alumni Organization Faculty Incentive Grants (University of Missouri) to R.W. M.J.H. is funded by the James S. McDonnell Foundation (grant no. 220020195), the National Science Foundation (grant no. 103522), the Boeing Company (purchase contract no. 525993), the John Templeton Foundation (grant no. 15705), and by a generous gift of the Bryan J. and June B. Zwan Foundation.

REFERENCES

Allen, A.P., Brown, J.H. and Gillooly, J.F. 2002. Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science*, **297**: 1545–1548.

Bellwood, P. 1997. *Prehistory of the Indo-Malaysian Archipelago*. Honolulu, HI: University of Hawaii Press.

Bellwood, P. and Renfrew, C., eds. 2002. *Examining the Farming/Language Dispersal Hypothesis*. Cambridge: McDonald Institute for Archaeological Research.

Binford, L.R. 2001. *Constructing Frames of Reference: An Analytical Method for Archaeological Theory Building Using Ethnographic and Environmental Data Sets*. Berkeley, CA: University of California Press.

Carneiro, R.L. 1967. On the relationship between size of population and complexity of social organization. *Southwest. J. Anthropol.*, **23**: 234–243.

Cashdan, E. 2001. Ethnic diversity and its environmental determinants: effects of climate, pathogens, and habitat diversity. *Am. Anthropol.*, **103**: 968–991.

Collard, I.F. and Foley, R.A. 2002. Latitudinal patterns and environmental determinants of recent human cultural diversity: do humans follow biogeographical rules? *Evol. Ecol. Res.*, **4**: 371–383.

Currie, T.E. and Mace, R. 2009. Political complexity predicts the spread of ethnolinguistic groups. *Proc. Natl. Acad. Sci. USA*, **106**: 7339–7344.

Diamond, J. 1997. *Guns, Germs, and Steel: The Fates of Human Societies*. New York: Norton.

Diamond, J. and Bellwood, P. 2003. Farmers and their languages: the first expansions. *Science*, **300**: 597–603.

Gray, J.P. 1999. A corrected ethnographic atlas. *World Cultures*, **10**: 24–85.

Hammarström, H. 2010. A full-scale test of the language farming dispersal hypothesis. *Diachronica*, **XXVII**: 197–213.

Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guegan, J.F., Kaufman, D.M. *et al.* 2003. Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**: 3105–3117.

Lewis, M.P., ed. 2009. *Ethnologue: Languages of the World*, 16th edn. Dallas, TX: SIL International.

Mace, R. and Pagel, M. 1995. A latitudinal gradient in the density of human languages in North America. *Proc. R. Soc. Lond. B*, **261**: 117–121.

Moore, J.L., Manne, L., Brooks, T., Burgess, N.D., Davies, R., Rahbek, C. *et al.* 2002. The distribution of cultural and biological diversity in Africa. *Proc. R. Soc. Lond. B*, **269**: 1645–1653.

Nettle, D. 1998. Explaining global patterns of language diversity. *J. Anthropol. Archaeol.*, **17**: 354–374.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Development Core Team. 2012. *nlme: Linear and Nonlinear Mixed Effects Models*, R package version 3.1-103. Vienna, Austria: R Foundation for Statistical Computing.

Porter, C.C. and Marlowe, F.W. 2007. How marginal are forager habitats? *J. Archaeol. Sci.*, **34**: 59–68.

Renfrew, C. 1987. *Archaeology and Language: The Puzzle of Indo-European Origin*. London: Jonathan Cape.

Renfrew, C. 1989. The origins of Indo-European languages. *Sci. Am.*, **261**: 106–114.

Richerson, P.J. and Boyd, R. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago, IL: University of Chicago Press.

Rosenzweig, M.L. 1995. *Species Diversity in Space and Time*. Cambridge: Cambridge University Press.

Turchin, P. and Gavrilets, S. 2009. Evolution of complex hierarchical societies. *Social Evol. Hist.*, **8**: 168–198.