University of Reading

Environmental conditions do not predict diversification rates in the Bantu languages

Article

Published Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Open Access

Beyer, R., Singarayer, J. S., Stock, J. T. and Manica, A. (2019) Environmental conditions do not predict diversification rates in the Bantu languages. Heliyon, 5 (10). e02630. ISSN 24058440 doi: 10.1016/j.heliyon.2019.e02630 Available at https://centaur.reading.ac.uk/86881/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1016/j.heliyon.2019.e02630

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

University of Reading

Central Archive at the University of Reading

Reading's research outputs online

Heliyon 5 (2019) e02630

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.heliyon.com

Research Article

Environmental conditions do not predict diversification rates in the Bantu languages

Robert Beyer [a,b,*], Joy S. Singarayer [c], Jay T. Stock [b,d,e], Andrea Manica [a]

[a] Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, United Kingdom
[b] PAVE Research Group, Department of Archaeology, University of Cambridge, Cambridge, CB2 3DZ, United Kingdom
[c] Department of Meteorology and Centre for Past Climate Change, University of Reading, Whiteknights campus, PO Box 243, Reading, RG6 6BB, United Kingdom
[d] Department of Anthropology, Western University, London, Ontario, N6A 5C2, Canada
[e] Department of Archaeology, Max Planck Institute for the Science of Human History, Kahlaische Strasse 10. D-07745 Jena, Germany

ARTICLE INFO

Keywords:
Climatology
Ecology
Paleoecology
Population dynamics
Linguistics
Environmental science
Linguistic diversity
Language phylogeny
Palaeoclimate modelling
Environmental risk hypothesis
Isolation by distance

ABSTRACT

The global distribution of language diversity mirrors that of several variables related to ecosystem productivity. It has been argued that this is driven by the size of social networks, which tend to be larger in harsher climates to ensure food security, leading to reduced language divergence. Is this pattern purely synchronic, or is there also a quantifiable relationship between environmental conditions and language diversification over time? We used a spatio-temporal phylogeny of the Bantu language family to estimate local diversification rates at the times and locations of language divergence. We compared these data against spatially-explicit reconstructions of several palaeoclimate and palaeovegetation variables (mean annual temperature and the temperature of the coldest and warmest quarter, annual precipitation and the precipitation of the wettest and driest quarter, growing degree days, the length of the growing season, and net primary production), to investigate a potential link between local environmental factors and diversification rates in the Bantu languages. A regression analysis does not suggest a statistically significant relationship between climatic or ecological variables and linguistic diversification over time. We find a strong positive correlation between pairwise linguistic and geographic distances in the Bantu languages, arguing for a dominant role of isolation as a result of the rapid Bantu expansion that might have overwhelmed any potential influence of local environmental factors.

1. Introduction

The distribution of present-day global language diversity has been shown to correlate strikingly well with several ecologically relevant variables, including climate and altitude (Bentz et al., 2018), biodiversity (Gorenflo et al., 2012; Harmon, 1996; Loh and Harmon, 2005; Moore et al., 2002; Nettle and Romaine, 2000; Sutherland, 2003), the distance from the equator (Bentz et al., 2018; Breton, 1991; Currie and Mace, 2012, 2009; Mace and Pagel, 1995; Nichols, 1999, 1990), and the length of the growing season (Nettle, 1998). A possible causal mechanism underlying these patterns was proposed by Nettle (1998, 1996), who explained a strong correlation between country-level language diversity and the length of the growing season in terms of the environmental risk hypothesis. The environmental risk hypothesis states that intra- or inter-annually fluctuating food supply requires the formation of larger social networks in order to mitigate these variations by means of trading and sharing resources when needed, and thus ensure a continual availability of food. An increasing uncertainty of a stable supply of subsistence products ('ecological risk') is argued to require increasingly larger networks, and the social and economic interactions within those require a common linguistic base. The synchronic studies mentioned above rely on spatial correlations between present-day language diversity and environmental variables. Can environmental conditions also explain the diversification of languages over time in a diachronic approach?

Tracking the diversification of languages through time and space requires temporally and spatially explicit language phylogenies. Reconstructions of such phylogenies have been proposed for the Austronesian languages (Gray et al., 2009), the Indo-European languages (Bouckaert et al., 2012), the Tupi languages (Walker et al., 2012) and the Bantu languages (Grollemund et al., 2015). Among these, the diversification of the Bantu languages is arguably best understood and mapped (Diamond and Bellwood, 2003), and so we focus on this language family

* Corresponding author.
E-mail address: rb792@cam.ac.uk (R. Beyer).

https://doi.org/10.1016/j.heliyon.2019.e02630
Received 23 August 2019; Received in revised form 26 September 2019; Accepted 8 October 2019
2405-8440/© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

here. The expansion of the Austronesian languages is complicated by sea barriers, the location of nodes in the phylogeography of the Indo-European languages are largely uncertain, and the methods used to estimate the timing and location of divergences in the Tupi languages are likely too simplistic.

Grollemund et al. (2015) previously showed that environmental conditions significantly affected the routes and speed of the spatial expansion of Bantu-speaking peoples. Here, we test whether climate and vegetation also affected the rate of diversification of the Bantu language family through time. We use a dated phylogeny to calculate rates of language diversification at the nodes of the phylogenetic tree, and link these to the environmental conditions that were present at the times and locations provided by the phylogeography. Reconstructed environmental variables include mean annual temperature and the temperature of the coldest and warmest quarter, annual precipitation and the precipitation of the wettest and driest quarter, growing degree days, the length of the growing season, and net primary production, thus covering a range of possible drivers behind the ecological risk hypothesis.

2. Materials and methods

2.1. Spatio-temporal language phylogeny and diversification rates

The Bantu languages spread as a result of the expansion south and east of farmers in tropical West Central Africa beginning around 5,000 years ago, and eventually covered most parts of subequatorial Africa, inducing a large-scale cultural shift. The reconstruction and understanding of the Bantu expansion has greatly benefited from the synthesis of linguistic data with genetic, archaeological and biogeographical evidence (Bostoen et al., 2015; Diamond and Bellwood, 2003). Following a spatially-explicit but undated phylogeny (Currie et al., 2012), Grollemund et al. (2015) derived a fully dated tree for 424 Bantu languages using Markov chain Monte Carlo methods and archaeological data to first confine the age ranges of specific nodes before applying a variable-rates molecular clock model to estimate node ages. Based on the geographical locations of extant languages, node-specific latitude and longitude were inferred using a variable-movement rate Brownian motion (see Fig. 1).

We calculate the local rate of language diversification at each node of the phylogenetic tree as follows. Letting $n_1, \ldots, n_N$ and $a_1, \ldots, a_N < 0$denote the tree's internal nodes and their ages, respectively. The clade originating at some node n_i defines a lineage-through-time (LTT) plot, mapping the time interval $[a_i, 0]$ onto the logarithm of the number of ancestral lineages of that clade (Ricklefs, 2007). The latter equals 0 at time a_i and increases towards the present, where it is equal to the logarithm of the total number of extant languages descending from node n_i. The local diversification rate at n_i corresponds to the right derivative of the LTT plot at time a_i. Since the LTT plot is given by a discontinuous step function, we define the node-specific diversification rate by means of the slope of a linear regression passing through the point $(a_i, 0)$of the LTT plot in a suitable interval $[a_i, a_i + \varepsilon]$ (Fig. 2). The value of $\varepsilon = 100$ years was found to be robust, and is used in the analyses.

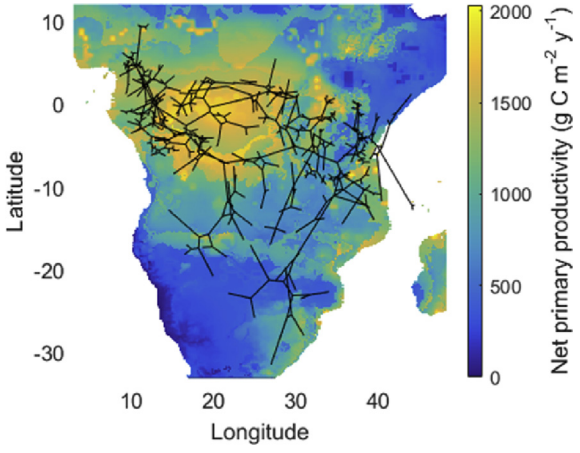

Fig. 1. Phylogeography of the Bantu language family and simulated present-day net primary production.

2.2. Reconstructions of environmental conditions

Environmental variables (listed in Fig. 3) are based on a recent dataset of high-resolution Holocene climate and vegetation reconstructions (Beyer et al., 2019) that is based on downscaled and bias-corrected outputs of the HadCM3 and HadAM3H global climate models (Valdes et al., 2017), and simulations of the Biome4 vegetation model (Kaplan et al., 2003). HadCM3 was previously used to simulate the last glacial cycle in multiple snapshots (see Singarayer and Burrough (2015), for a detailed description of the model setup). We linearly interpolated the available monthly climate data from their 0.5° spatial resolution and 1,000 year temporal resolution to the geographic location and time of each node. In addition to mean annual temperature and mean temperature of the warmest and coldest quarter, annual precipitation and precipitation of the wettest and driest quarter, and net primary production, we also included growing degree days, and the mean length of the growing season in our analysis. Growing degree days (Prentice et al., 1992) were derived using estimates of daily temperature values, which we approximated from the monthly data by means of a cubic spline. The mean length of the growing season, which Nettle (1998) previously linked to global linguistic diversity, is calculated as the number of months when mean temperature lies above 6 °C and is smaller than 0.5 times its total precipitation (in mm).

3. Results

Fig. 3 plots diversification rates of the Bantu family against the environmental conditions present at the time and location of the relevant node of the phylogeny. We do not observe a distinct pattern; linear regressions are weak ($R^2 < 0.01$) and not significant ($p > 0.07$) for all covariates.

These results suggest that if environmental conditions did at all affect the diversification of the Bantu languages, then they did so to a much lesser extent than other factors not considered in our analysis. In addition to, for example, population size (Bromham et al., 2015; Wichmann et al., 2007; Wichmann and Holman, 2009) and sociocultural variables (Gavin et al., 2013) (neither of which we included in our analysis due to lack of data for the relevant time frame), a likely candidate for a more dominant driver of diversification is isolation by distance, given the considerable speed at which the Bantu people spread into distant regions over a very large area, continually resulting in the loss of contact between groups and the collapse of existing social ties. Indeed, we observe a strong relationship between pairwise language divergence and geographical

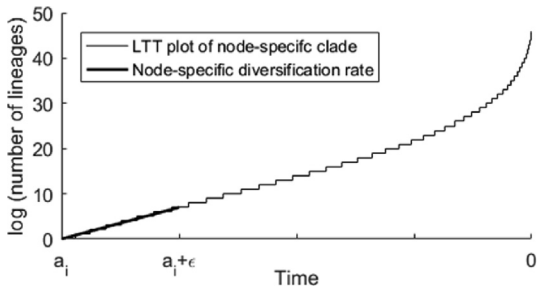

Fig. 2. Calculation of the node-specific diversification rate. The black step function represents the lineage-through-time (LTT) plot corresponding to the clade whose root is given by some node n_i at time a_i. The grey line illustrates how the diversification rate at n_i is defined in terms of the smoothed slope of the LTT plot near time a_i.

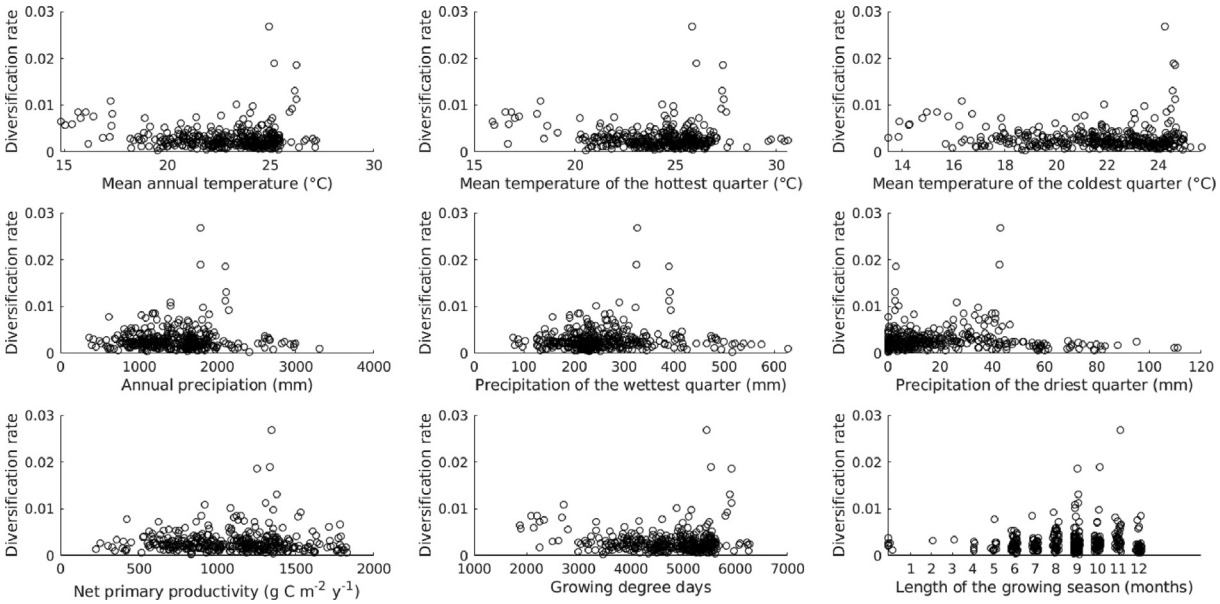

Fig. 3. Node-specific diversification rate against reconstructed local environmental conditions.

distance (Fig. 4).

4. Discussion

Here, we did not identify a significant effect of environmental conditions on the diversification rates of the Bantu languages. This may potentially be due to the rapid nature of the expansion, which could have overwhelmed any possible influence of local environmental factors. Languages that have not undergone such a fast expansion as the one observed for Bantu family might be more informative on the role of the local environment, although suitable spatio-temporal phylogenies on such languages have not yet been derived. A better understanding of the relationship between local environments on linguistic diversity – in combination with existing theory on language evolution (e.g. Perc, 2012; Petersen et al., 2012) – could enhance our understanding of past human migrations and spatial demographics, where archaeological records alone do not provide sufficient information to draw a clear picture. In these cases, the incorporation and synthesis of linguistic and biogeographical data can help to provide additional evidence, test hypotheses and thus complete the picture (Bostoen et al., 2015).

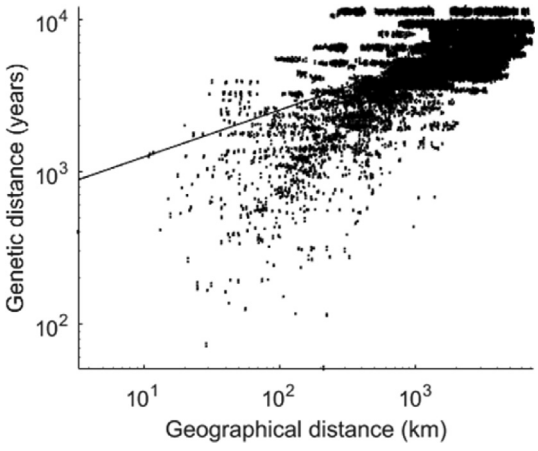

Fig. 4. Language divergence against geographic distance. The figure shows the genetic distances between each pair of tips of the Bantu phylogeny plotted against their geographical distances (In other words, each X and Y value corresponds to the sum of the distances of two languages to their last common ancestor along the phylogentic and phylogeographic tree, respectively.) The linear regression of the log-log scatter plot has $R^2 = 0.36$ and is significant ($p = 0.001$).

Declarations

Author contribution statement

Robert Beyer, Andrea Manica: Conceived and designed the analysis; Analyzed and interpreted the data; Wrote the paper.

Joy S. Singarayer: Analyzed and interpreted the data; Contributed analysis tools or data; Wrote the paper.

Jay T. Stock: Analyzed and interpreted the data; Contributed analysis tools or data; Wrote the paper

Funding statement

Robert Beyer and Andrea Manica were supported by ERC Consolidator Grant 647797 "LocalAdaptation". Robert Beyer and Jay T. Stock were supported by ERC Consolidator Grant 617627 "ADaPt".

Competing Interest Statement

The authors declare no conflict of interest.

Additional Information

Data associated with this study has been deposited at Open Science Framework under the https://doi.org/10.17605/OSF.IO/3HRZS (https://osf.io/3hrzs/).

References

Bentz, C., Dediu, D., Verkerk, A., Jäger, G., 2018. The evolution of language families is shaped by the environment beyond neutral drift. Nat. Hum. Behav. 2, 816–821.

Beyer, R., Krapp, M., Manica, A., 2019. High-resolution Terrestrial Climate, Bioclimate and Vegetation for the Last 120,000 Years. EarthArXiv.

Bostoen, K., Clist, B., Doumenge, C., Grollemund, R., Hombert, J.M., Muluwa, J.K., Maley, J., 2015. Middle to late Holocene paleoclimatic change and the early Bantu expansion in the rain forests of Western Central Africa. Curr. Anthropol. 56, 354–384.

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S.J., Alekseyenko, A.V., Drummond, A.J., Gray, R.D., Suchard, M.A., Atkinson, Q.D., 2012. Mapping the origins and expansion of the Indo-European language family. Science 337, 957–960.

Breton, R.J.L., 1991. Geolinguistics: Language Dynamics and Ethnolinguistic Geography. University of Ottawa Press.

Bromham, L., Hua, X., Fitzpatrick, T.G., Greenhill, S.J., 2015. Rate of language evolution is affected by population size. In: Proc. Natl. Acad. Sci., 112, pp. 2097–2102.
Currie, T.E., Mace, R., 2012. The evolution of ethnolinguistic diversity. Adv. Complex Syst. 15, 1150006.
Currie, T.E., Mace, R., 2009. Political complexity predicts the spread of ethnolinguistic groups. In: Proc. Natl. Acad. Sci., 106, pp. 7339–7344.
Currie, T.E., Meade, A., Guillon, M., Mace, R., 2012. Cultural phylogeography of the Bantu languages of sub-saharan Africa. In: Proc R Soc B, 280, p. 20130695.
Diamond, J., Bellwood, P., 2003. Farmers and their languages: the first expansions. Science 300, 597–603.
Gavin, M.C., Botero, C.A., Bowern, C., Colwell, R.K., Dunn, M., Dunn, R.R., Gray, R.D., Kirby, K.R., McCarter, J., Powell, A., Rangel, T.F., Stepp, J.R., Trautwein, M., Verdolin, J.L., Yanega, G., 2013. Toward a mechanistic understanding of linguistic diversity. Bioscience 63, 524–535.
Gorenflo, L.J., Romaine, S., Mittermeier, R.A., Walker-Painemilla, K., 2012. Co-occurrence of linguistic and biological diversity in biodiversity hotspots and high biodiversity wilderness areas. In: Proc. Natl. Acad. Sci., 109, pp. 8032–8037.
Gray, R.D., Drummond, A.J., Greenhill, S.J., 2009. Language phylogenies reveal expansion pulses and pauses in pacific settlement. Science 323, 479–483.
Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C., Pagel, M., 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. Proc. Natl. Acad. Sci. 112, 13296–13301.
Harmon, D., 1996. Losing species, losing languages: connections between linguistic and biological diversity. Southwest J. Linguist. 15, 89–108.
Kaplan, J.O., Bigelow, N.H., Prentice, I.C., Harrison, S.P., Bartlein, P.J., Christensen, T.R., Cramer, W., Matveyeva, N.V., McGuire, A.D., Murray, D.F., Razzhivin, V.Y., Smith, B., Walker, D.A., Anderson, P.M., Andreev, A.A., Brubaker, L.B., Edwards, M.E., Lozhkin, A.V., 2003. Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections. J. Geophys. Res. Atmospheres 108, 8171.
Loh, J., Harmon, D., 2005. A global index of biocultural diversity. Ecol. Indicat. 5, 231–241.
Mace, R., Pagel, M., 1995. A latitudinal gradient in the density of human languages in north America. In: Proc. R. Soc. Lond. B Biol. Sci., 261, pp. 117–121.
Moore, J.L., Manne, L., Brooks, T., Burgess, N.D., Davies, R., Rahbek, C., Williams, P., Balmford, A., 2002. The distribution of cultural and biological diversity in Africa. In: Proc. R. Soc. Lond. B Biol. Sci., 269, pp. 1645–1653.
Nettle, D., 1998. Explaining global patterns of language diversity. J. Anthropol. Archaeol. 17, 354–374.
Nettle, D., 1996. Language diversity in west Africa: an ecological approach. J. Anthropol. Archaeol. 15, 403–438.
Nettle, D., Romaine, S., 2000. Vanishing Voices: the Extinction of the World's Languages. Oxford University Press.
Nichols, J., 1999. Linguistic Diversity in Space and Time. University of Chicago Press.
Nichols, J., 1990. Linguistic diversity and the first settlement of the new world. Language 66, 475–521.
Perc, M., 2012. Evolution of the most common English words and phrases over the centuries. J. R. Soc. Interface 9, 3323–3328.
Petersen, A.M., Tenenbaum, J.N., Havlin, S., Stanley, H.E., Perc, M., 2012. Languages cool as they expand: allometric scaling and the decreasing need for new words. Sci. Rep. 2, 943.
Prentice, I. Colin, Cramer, W., Harrison, S., Leemans, R., Monserud, R., Prentice, I.C., Harrison, S.P., Monserud, R.A., Solomon, K.R.S.A., Prentice, C., Solomon, S.G., 1992. Special Paper: a Global Biome Model Based on Plant Physiology and Dominance, Soil Properties and Climate.
Ricklefs, R.E., 2007. Estimating diversification rates from phylogenetic information. Trends Ecol. Evol. 22, 601–610.
Singarayer, J.S., Burrough, S.L., 2015. Interhemispheric dynamics of the African rainbelt during the late Quaternary. Quat. Sci. Rev. 124, 48–67.
Sutherland, W.J., 2003. Parallel extinction risk and global distribution of languages and species. Nature 423, 276–279.
Valdes, P.J., Armstrong, E., Badger, M.P.S., Bradshaw, C.D., Bragg, F., Crucifix, M., Davies-Barnard, T., Day, J.J., Farnsworth, A., Gordon, C., Hopcroft, P.O., Kennedy, A.T., Lord, N.S., Lunt, D.J., Marzocchi, A., Parry, L.M., Pope, V., Roberts, W.H.G., Stone, E.J., Tourte, G.J.L., Williams, J.H.T., 2017. The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0. Geosci. Model Dev. 10, 3715–3743.
Walker, R.S., Wichmann, S., Mailund, T., Atkisson, C.J., 2012. Cultural phylogenetics of the Tupi language family in lowland south America. PLoS One 7, e35025.
Wichmann, S., Holman, E.W., 2009. Population size and rates of language change. Hum. Biol. 81, 259–274.
Wichmann, S., Stauffer, D., Schulze, C., Holman, E.W., 2007. Do language change rates depend on population size? Adv. Complex Syst. 11, 357–369.