



Comment on "Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa" Michael Cysouw, *et al. Science* **335**, 657 (2012); DOI: 10.1126/science.1208841

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this infomation is current as of February 13, 2012):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

http://www.sciencemag.org/content/335/6069/657.2.full.html

Supporting Online Material can be found at: http://www.sciencemag.org/content/suppl/2012/02/09/335.6069.657-b.DC1.html

A list of selected additional articles on the Science Web sites **related to this article** can be found at: http://www.sciencemag.org/content/335/6069/657.2.full.html#related

This article cites 9 articles, 4 of which can be accessed free: http://www.sciencemag.org/content/335/6069/657.2.full.html#ref-list-1

This article has been **cited by** 1 articles hosted by HighWire Press; see: http://www.sciencemag.org/content/335/6069/657.2.full.html#related-urls

This article appears in the following **subject collections:** Evolution http://www.sciencemag.org/cgi/collection/evolution Technical Comments http://www.sciencemag.org/cgi/collection/tech_comment

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2012 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

Comment on "Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa"

Michael Cysouw,¹*† Dan Dediu,^{2,3}† Steven Moran¹

We show that Atkinson's (Reports, 15 April 2011, p. 346) intriguing proposal—that global linguistic diversity supports a single language origin in Africa—is an artifact of using suboptimal data, biased methodology, and unjustified assumptions. We criticize his approach using more suitable data, and we additionally provide new results suggesting a more complex scenario for the emergence of global linguistic diversity.

Recently, Atkinson (1) reported a negative correlation between the size of the phonemic inventory of a language and its geographic distance from western Africa. He proposed that this is the result of a repeated linguistic founder effect accompanying the migration of modern humans out of Africa some 50,000 to 70,000 years ago. According to this proposal,

*To whom correspondence should be addressed. E-mail: cysouw@lmu.de

⁺These authors contributed equally to this work.

the original languages spoken in western Africa would have had a large phoneme inventory, which became reduced during the spread of modern humans over the globe because of imperfect transmission in the small founder populations involved. From a linguistic perspective, this result is surprising and contradicts our intuitions. Although we agree that there is clear nonrandom geographic patterning in the distribution of phoneme inventory sizes among the world's languages, we very much doubt that it shows any detectable remnant of the proposed demographic scenario.

In summary (2), we see the following problems with Atkinson's findings. First, his data are coarse-grained summaries of the UCLA Phonological Segment Inventory Database (UPSID) (3) as reported in the *World Atlas of Language Structures* (WALS) (4). To illustrate our concerns,

BIC+4 'Origins' of various WALS features

we used the original UPSID database (which is freely available online) together with tone data from WALS. Atkinson's WALS-based estimates of phoneme inventory size turn out to be only imperfectly correlated with the actual number of phonemes as specified in UPSID (r = 0.60, P < 2.2×10^{-16}). Specifically, his WALS-based data give unjustified weight to the number of vowels and tones at the expense of the number of consonants, strongly biasing the resulting geographic patterning toward western Africa's having large phoneme inventories (figs. S3 and S4). When the UPSID data are appropriately corrected for speaker community size and linguistic genera through a mixed-effects model, the largest phoneme inventories are actually found in North America (fig. S8).

Second, Atkinson methodologically follows previous work reporting clines of decreasing genetic and phenotypic diversity with increasing distance from Africa (5, 6). Accordingly, he uses the term "phonemic diversity" interchangeably with the more accurate "phoneme inventory size," but this seems misplaced. In the original papers (5, 6) "diversity" refers to variation within populations of individuals, whereas Atkinson's linguistic diversity refers to differences between languages. In fact, the languages of western Africa and New Guinea/Australia in UPSID show the lowest variability in inventory sizes (fig. S5). On the basis of the proposed serial founder effect, low variability might have been expected in New Guinea/Australia, but surely not in the supposed origin in western Africa.

In practice, Atkinson's biologically inspired method searches for the geographic location



Fig. 1. Areas of "origin" of various other inventory-like linguistic characteristics as identified using Atkinson's methodology. Notably, the origins are dispersed over the whole globe and not concentrated in Africa. The dark red area in Africa is the origin of phoneme inventories as proposed by Atkinson. The dark green

area in Africa and the Near East is the corresponding area based on the UPSID phoneme inventory data. The small red area on the eastern tip of New Guinea is the origin for the UPSID phoneme inventory data using a quadratic geographical distance model. Details about the other areas can be found in (2).

¹Research Unit Quantitative Language Comparison, Ludwig Maximilian University, Geschwister Scholl Platz 1, D-80539 Munich, Germany. ²Language and Genetics Department, The Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD, Nijmegen, Netherlands. ³Donders Institute for Brain, Cognition and Behaviour, Kapittelweg 29, 6525 EN, Nijmegen, Netherlands.

minimizing the Bayesian information criterion (BIC) (7) of the regression between phonemic inventory size and geographic distance, including further control variables. Atkinson selects those locations at most four BIC units away from this optimum as having considerable support in being the origin of the expansion. A quick computation shows that this implies accepting models that are at most $e^2 \sim 7.4$ times less likely than the optimal one (2), which strikes us as rather arbitrary. Further, this BIC optimization method necessarily "spreads" any origin across a contiguous geographic region, even in the case of totally random data (fig. S12).

Notwithstanding this criticism, we replicated Atkinson's method using the UPSID data, but instead of a single origin in western Africa, we found two separate "origins," one in eastern Africa and one in the Caucasus (fig. S10). The BIC+4 range of possible origins covers a large area, including also the Middle East and southern Africa. Although this finding does not necessarily contradict an expansion from Africa, it does not provide clear support in its favor, either. Further, adding a quadratic distance factor to the model substantially improves the fit and suggests an alternative origin located in New Guinea with a small phoneme inventory (fig. S10). Even more problematic, when we apply the original method to other inventory-like linguistic characteristics from WALS (Fig. 1), we find origins of global clines all over the world, not just in Africa, and not always corresponding to the highest structural "complexity" (fig. S11). Therefore, the observation of an Africa-based phoneme inventory cline does not generalize to other linguistic characteristics of a similar kind.

Third, Atkinson's explanation crucially depends on a positive correlation between phonemic inventory size and speaker community size, which, unfortunately and contrary to his own claim [see figure S1 in (1)], does not hold for small populations when using UPSID data (r = 0.04, P = 0.64) (fig. S6). This correlation reaches significance at the 5% level only when languages with speaker populations above 10^5 are included, but such large speaker community sizes only arose in the context of agriculture long after the peopling of most of the globe (8).

Fourth, the geographic patterning of tone might be influenced by a genetic bias postdating the out-of-Africa migration by tens of thousands of years (9). Moreover, consonant inventories (and to a lesser extent, vowel inventories) do not seem to be phylogenetically stable enough (10) to conserve the kind of deep signal necessary for the proposed scenario, whereas other, more stable, features show non-African "origins" (fig. S11).

Finally, we believe that Atkinson's interpretation of the reported worldwide cline in terms of a linguistic serial founder effect is problematic because of the extraordinary large amount of horizontal processes affecting language (11, 12) and because the underlying mechanism proposed by Atkinson is linguistically not plausible (13). Further, global clines in linguistics, like in genetics, do not necessarily equate with a serial founder effect and can have other causes (2, 14).

Summarizing, the reported linguistic evidence for an expansion from Africa is unfortunately an artifact of various methodological decisions and biased interpretations. We consider this to be unfortunate, because we would very much welcome any new insights into human prehistory based on geographic patterns of linguistic diversity. In this respect, we applaud Atkinson for further developing this approach (15) and renewing the methodological discussion, because only explicit testing and refutation opens the way for the formulation of more specific hypotheses concerning the identification of possible linguistic signatures of ancient demographic events.

References and Notes

- 1. Q. D. Atkinson, Science 332, 346 (2011).
- 2. Materials and methods are available as supporting material on *Science* Online.
- I. Maddieson, K. Precoda, The UCLA Phonological Segment Inventory Database (1990); http://web.phonetik. uni-frankfurt.de/upsid_info.html.
- M. Haspelmath, M. S. Dryer, D. Gil, B. Comrie, Eds., *The World Atlas of Language Structures* (Oxford Univ. Press, Oxford, 2005).
- 5. A. Manica, W. Amos, F. Balloux, T. Hanihara, *Nature* **448**, 346 (2007).
- L. Betti, F. Balloux, W. Amos, T. Hanihara, A. Manica, Proc. Biol. Sci. 276, 809 (2009).
- 7. G. E. Schwarz, Ann. Stat. 6, 461 (1978).
- S. Mithen, After the Ice (Orion Books, London, 2003).
 D. Dediu, D. R. Ladd, Proc. Natl. Acad. Sci. U.S.A. 104, 10944 (2007).
- 10. D. Dediu, Proc. Biol. Sci. 278, 474 (2011).
- S. G. Thomason, T. Kaufman, Language Contact, Creolization, and Genetic Linguistics (Univ. of California Press, Berkeley, CA, 1988).
- 12. T. F. Jaeger, P. Graff, W. Croft, D. Pontillo, *Linguist. Typol.* 15, 281 (2011).
- 13. C. Bowern, Linguist. Typol. 15, 207 (2011).
- 14. I. Maddieson, T. Bhattacharya, D. E. Smith, W. Croft, *Linguist. Typol.* **15**, 267 (2011).
- 15. J. Nichols, *Linguistic Diversity in Space and Time* (Univ. of Chicago Press, Chicago, 1992).

Acknowledgments: The authors declare no conflict of interests. M.C., D.D., and S.M. designed the research; M.C. and S.M. organized the data; M.C. and D.D. performed the analysis and wrote the paper. M.C. and S.M. are funded by European Research Council starting grant 240816 QuantHistLing.

Supporting Online Material

www.sciencemag.org/cgi/content/full/335/6069/657-b/DC1 Materials and Methods SOM Text

Figs. S1 to S12 Table S1 References

24 May 2011; accepted 3 January 2012 10.1126/science.1208841