

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

Chuan-Chao Wang, Qi-Liang Ding, Huan Tao, Hui Li*

Atkinson (Reports, 15 April 2011, p. 346) reported a declined trend of phonemic diversity from Africa that indicated the African exodus of modern languages. However, his claim was only supported when the phonemic diversities were binned into three or five levels. Analyses using raw data without simplification suggest a decline from central Asia rather than from Africa.

Atkinson (1) analyzed the phoneme numbers of 504 languages around the world and found a strong inverse relationship between the phonemic diversity and distance from an inferred origin in Africa, which supports an African origin of modern languages. Although a statistically significant declined trend of phonemic diversity from Africa can be observed from the analyses of his normalized data set, his conclusion was questionable because of the simplification of the phoneme inventories.

The simplified data used in Atkinson’s analyses were obtained directly from the *World Atlas of Language Structures* (WALS) (2), where the phoneme numbers of the languages were simply binned into three or five levels. However, this kind of simplification of the data lost most information of the phonemic diversity and might have resulted in bias conclusion. For example, the consonant inventory varies from less than 10 to more than 80 among the world languages (3), while only five levels were counted in Atkinson’s analyses.

We collected a new data set of world phonemic diversity, including 579 languages from 95 linguistic families (table S1). The phoneme inventories were displayed without any simplification. To balance among the linguistic families, we excluded 69 samples of some well-studied linguistic families (i.e., Indo-European, Austronesian, and Sino-Tibetan) from our analyses. This made our data comparable to Atkinson’s (table S2). Our analyses were based on the remaining 510 languages.

Judged from the original WALS maps and Atkinson’s normalized data set, the diversities of vowel quality, tone, and consonant (Fig. 1A) all exhibit significant declines from Africa to the rest of the world. However, the declines from Africa will not be that pronounced when the data are not

simplified (the exact counts of vowel qualities, tones, and consonants). Languages from Eurasia show higher diversities of vowel qualities and tones (Fig. 1B). Therefore, we argued that Atkinson’s

statistics were distorted by WALS’s data simplification, which truncated the high ends of the scales (2).

For example, WALS binned the vowel quality inventories into three groups: small (2 to 4 qualities), medium (5 to 6 qualities), and large (7 to 14 qualities). Actually, the basic vowel quality inventory varies from 2 to 20 and is distributed unequally among the geographic regions (4). Most large vowel quality inventories appear in Eurasia, whereas only small inventories can be found in the Americas and Australia. The Germanic languages and the Wu Chinese dialects have the largest vowel quality inventories in the world, mostly larger than 10—for example, the Standard Swedish has at least 16 vowel qualities, and the Dǒndǎc Wu spoken in southern Shanghai has 20 vowel qualities. In contrast, few languages from Africa have more than 10 vowel qualities (Fig. 1B). Therefore, a lower limit of seven qualities for large inventory in WALS’s

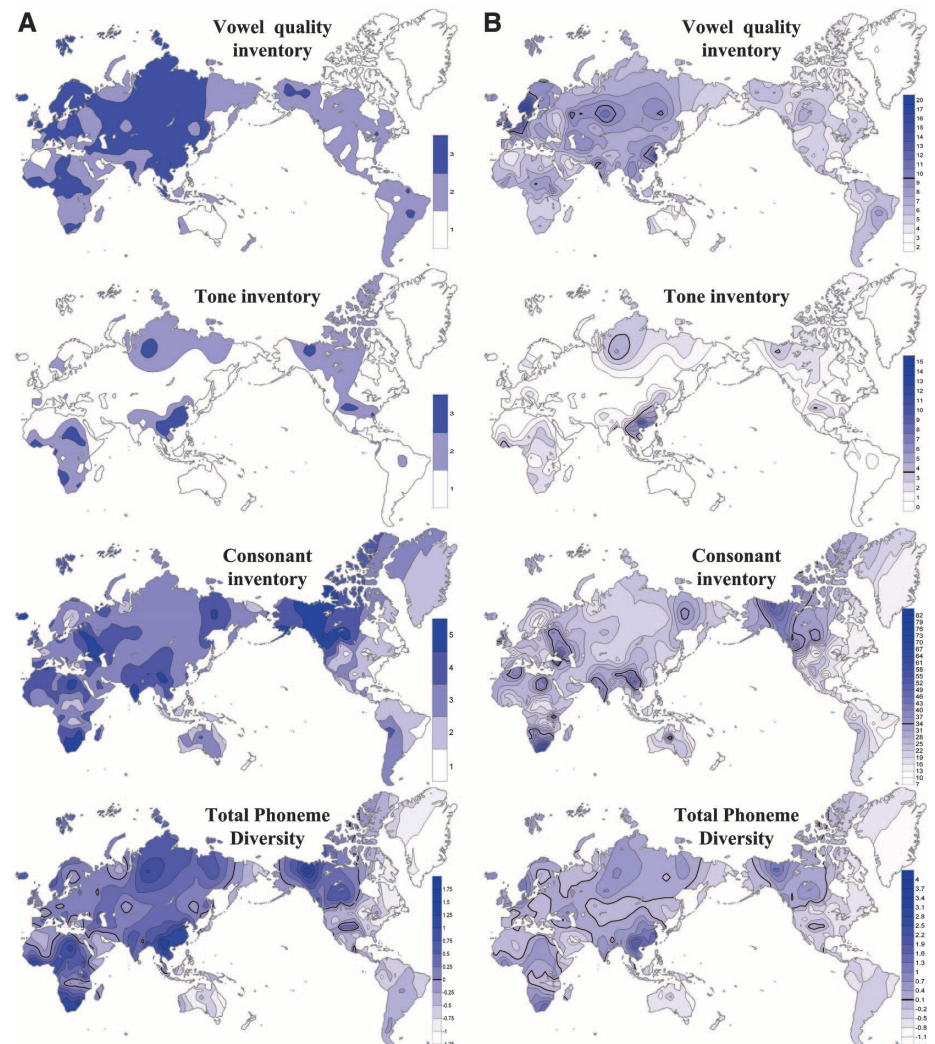


Fig. 1. Geographic distribution of the phonemic diversities of the world’s languages. (A) Simplified phonemic diversities used by WALS. (B) Exact phoneme inventory counts and the corresponding total phoneme diversity.

Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, and Department of Chinese Language and Literature, Fudan University, Shanghai 200433 China.

*To whom correspondence should be addressed. E-mail: lihui.fudan@gmail.com

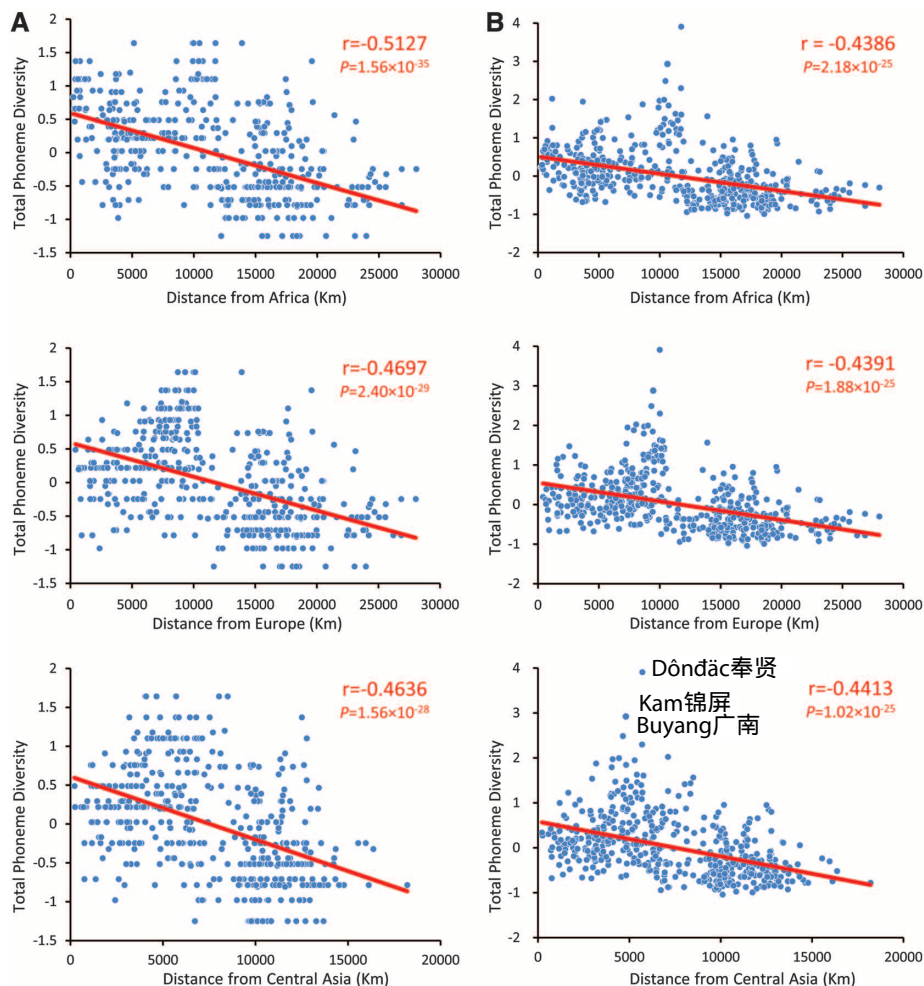


Fig. 2. Correlation between the distance from the best-fit origin of the languages and total phoneme diversity. Red lines are the fitted regression lines. Pearson correlation r and P values are shown in the upper right. **(A)** Estimated from simplified phoneme inventories. **(B)** Estimated from exact phoneme inventory counts.

data set eliminated the difference about the vowel diversity levels between the African and Eurasian languages. Besides the curtness in the vowel inventory counts, Atkinson’s analysis ignored all other phonetic features of the vowels, such as nasalization, diphthong, and length, which vary tremendously among the world’s languages (5).

Similar problems happened to the simplification of tone diversity. Most tonal languages in Africa have less than four tones, whereas most tonal languages in Asia have more than four. The Kam spoken in southwest China has the largest tone inventory (15 tones). The difference

of tone diversity between the two continents is also screened in WALS. Using the raw data of all phonemes of the 510 languages without simplification, we analyzed the total phoneme diversity (table S1). The highest diversity is demonstrably in Asia (Fig. 1B), and the top three languages are Dôngdác (3.91), Kam (2.87), and Buyang (2.49).

We further redid the correlation analyses between the total phoneme diversity and distance from the “best-fit origin.” Different regions were chosen as potential best-fit origins. Interestingly, stronger negative correlation was observed when choosing central Asia (the exact locus was

Ashgabat) (Pearson correlation, $r = -0.4413$) or Europe ($r = -0.4391$) as the origin than choosing Africa ($r = -0.4386$) (Fig. 2B). Moreover, when using mean diversity across language families, central Asia ($r = -0.5503$, $P = 5.08 \times 10^{-5}$) exhibited even stronger correlation than Europe ($r = -0.4945$, $P = 3.54 \times 10^{-4}$) or Africa ($r = -0.5053$, $P = 2.49 \times 10^{-4}$). However, when the WALS simplification was applied to our data set, the strongest negative correlation was still found between the diversity and distance from Africa (Fig. 2A).

To further test the robustness of these findings, we repeated regressions after controlling for modern speaker population size (6). Africa never exhibited strongest correlation unless the data were simplified (table S3). Because there are some clear outliers among the sampled languages, we applied a robust linear model to minimize the possible influence of outliers (7). In this model, the best-fit origin also turned from Africa to Asia when the data was not simplified (table S4).

Thus, we demonstrated that WALS’s data simplification has distorted Atkinson’s results. Apparently, the results without simplification should be more reliable. Therefore, Asia (where the Babel was supposed to be) might be a more appropriate best-fit origin for modern languages if modern languages have a common origin.

References and Notes

1. Q. D. Atkinson, *Science* **332**, 346 (2011).
2. I. Maddieson, in *The World Atlas of Language Structures Online*, M. S. Dryer, M. Haspelmath, Eds. (Max Planck Digital Library, Munich, 2011); chaps. 1, 2, and 13.
3. P. Ladefoged, I. Maddieson, *The Sounds of the World’s Languages* (Blackwell, Oxford, 1996).
4. H. Li, *Commun. Contemp. Anthropol.* **2**, 42 (2008), 10.4236/coca.2008.21007; <http://comonca.org.cn/Abst/2008/007.htm>.
5. International Phonetic Association, *Handbook of the International Phonetic Association* (Cambridge Univ. Press, Cambridge, 1999).
6. L. Excoffier, G. Laval, S. Schneider, *Evol. Bioinform. Online* **1**, 47 (2005).
7. W. N. Venables, B. D. Ripley, *Modern Applied Statistics with S* (Springer, New York, 2002).

Acknowledgments: This work was supported by Shanghai Commission of Education Research Innovation Key Project (11zz04) and Shanghai Professional Development Funding (2010001). Y. Hu helped with some of the statistics.

Supporting Online Material

www.sciencemag.org/cgi/content/full/335/6069/657-c/DC1
Tables S1 to S4
References

3 May 2011; accepted 3 January 2012
10.1126/science.1207846