

Studying Language Evolution in the Age of Big Data

Tanmoy Bhattacharya^{a,b,*}, Damián E. Blasi^{d,e}, William Croft^g, Michael Cysouw^h, Daniel Hruschkaⁱ, Ian Maddieson^g, Lydia Müller^j, Nancy Retzlaff^{j,c}, Eric Smith^{a,k,l}, Peter F. Stadler^{a,l,j,*}, George Starostin^m, Hyejin Youn^{a,n}

^a*Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA*

^b*Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545-0285, USA*

^c*Max Planck Institute for Mathematics in the Sciences, D-04103 Leipzig, Germany*

^d*University of Zürich, Zürich, Switzerland*

^e*Max Planck Institute for the Science of Human History, Jena, Germany*

^f*Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany*

^g*Linguistics, University of New Mexico, Albuquerque, NM 87131-0001, USA*

^h*Forschungszentrum Deutscher Sprachatlas, Philipps Universität Marburg, Marburg, Germany*

ⁱ*School of Human Evolution and Social Change, Arizona State University, Tempe AZ 85287, USA*

^j*Bioinformatics Group, Department of Computer Sciences, University of Leipzig, Leipzig 04107, Germany*

^k*Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA 22030, USA*

^l*Ronin Institute, Montclair, NJ 07043*

^m*Centre for Comparative Linguistics, Russian State University for the Humanities, Moscow, Russia*

ⁿ*The Institute for New Economic Thinking at the Oxford Martin School, Oxford, OX2 6ED, UK*

1 ACKNOWLEDGEMENTS This manuscript grew out of a working group held at the Santa Fe Institute in spring 2014. This work was supported in part by the National Science Foundation (no. SMA-1312294) and the CABDyN (to HJ) and a PPP grant from the German Academic Exchange Service (to PFS). In addition, we would like to acknowledge the valuable comments from anonymous referees to a previous version of this manuscript.

*Correspondence should be addressed to TB and PFS. emails: tanmoy@santafe.edu and stadler@bioinf.uni-leipzig.de.