Date: 26th January 2015

Dear Marcia McNutt,

This paper compares structural properties of languages with mitochondrial DNA haplogroups in order to investigate whether they have a similar history of movement between populations in Africa and Eurasia. The finding is that genetic variation between populations can explain properties of unrelated languages such as their word orders, phonemes, and how many consecutive consonants languages allow, and this holds after controlling for known language relatedness using several different statistical methods such as sampling and mixed effects models. A particular finding is that complex consonant clusters have travelled by horizontal transfer between thirteen language families of Eurasia and North Africa, with a distribution closely corresponding to the spread of haplogroups such as HV (R2=0.34, p=3.8e-08), and provides evidence for Neolithic Eurasian migration into North Africa not explained by recent Indo-European influence or the spread of Arabic.

Geneticists regularly compare Y chromosome and mitochondrial DNA haplogroups with the spread of language families based on vocabulary (Cavalla-Sforza 2000), but not systematically with structural properties of languages (although Barbieri *et al.*, *Eur. J. Hum. Genet.* 2013 on the spread of clicks in Africa is one exception). Some prior work has established that structural properties in specific families such as Indo-European can be historically informative and correlate with genetic variation (Longobardi and Guardiano, *Lingua* 2009, Colonna *et al.*, *Hum. Hered.* 2010), but there is no published work comparing a large number of language families in order to investigate language contact or distant relatedness across whole continents, and controlling for known language relatedness. Other work has used structural properties to investigate language prehistory in ways analogous to the study of species or genetics (Dunn e*t al*., *Science* 2005, and Atkinson, *Science* 2011), but has not explicitly comparing how well structural properties and genetic variants correlate and what can be inferred from these correlations. Dediu and Ladd (*PNAS* 2007) compare linguistic tone with genetic variants, but in order to make the claim that two specific genes affect the processing of tone in the brain, rather than the claim in this paper that structural properties travel with genetic variants between populations. Forster and Renfrew (*Science* 2000) showed that male migration as shown by Y chromosome haplogroups often correlates with the spread of language families into new regions, while mtDNA haplogroups often represent more ancient settlement in these new regions; this paper therefore tested mtDNA haplogroups in particular in order to show that structural properties can reflect a different side of language history, namely the way that language families moving into a new region can be drastically altered in their structures under the influence of languages which were already spoken there, and hence which may cause mtDNA to correlate especially well with the distribution of structures across families.

It is hoped that these findings will be of interest to readers of Science, because it demonstrates a novel method for investigating language prehistory; and to geneticists and archeologists in particular in integrating their findings not just with language families but with new linguistic databases of structural properties. It opens up a new method in linguistics of explaining properties of languages with reference to genetics, such as why western Eurasian languages have more complex complex consonant clusters than East Asian languages. It also supports a previously suggested but still controversial point in linguistics, that structural properties can be distinctive and stable enough over time to be historically informative, and that they can be used to track population movements in ways analogous to the study of how mitochondrial DNA haplogroups have migrated. This opens up the possibility for linguists to reconstruct the history of how linguistic structures have travelled between populations by horizontal transfer, and not just the way that language families have travelled, as has been the primary focus of research in this area (Dunn *et al.*, *Science* 2005; Bouckaert *et al.*, *Science* 2011).

I confirm that this is an original manuscript which has not been published previously, and is not under consideration elsewhere. This paper has been reviewed by Michael Dunn (Uppsala University), Nicholas Enfield (University of Sydney), Pieter Muysken (Radboud University), Sarah Graham (Max Planck Institute for Psycholinguistics) and Stephen Levinson (Max Planck Institute for Psycholinguistics). Dan Dediu and Seán Roberts at the Max Planck Institute for Psycholinguistics and Raphael Hunsucker, Katrine Smiet, and Eveline Ritten at Radboud University also provided feedback on the study.

I hope that you will find the manuscript suitable for publication. Thank you for your time and consideration, and I look forward to your reply.

Yours sincerely,

Jeremy Collins  
Radboud University, Nijmegen