Exploring Genetic and Linguistic Phylogenies as a Framework for Describing Family Behaviour

This paper aims to explore the use of phylogenetic supertrees to identify group-level variability in indicators of family behaviour. A supertree constructed based on genetic and linguistic data (Duda and Zrzavy 2016) describes the evolutionary history of ethnic and linguistic groups. Using Gaussian process regression, the genetic and linguistic evolutionary similarities between ethnic groups is represented in a covariance matrix that indicates each group's pairwise phylogenetic distance on the supertree relative to each other. We then match ethnic groups as described by the Demographic and Health Surveys (DHS) to their respective branches on the supertree. Finally, we compare the phylogenetic distances to similarities in family behaviours—also obtained from the DHS—such as age at first childbearing, to see if groups closer to each other on the supertree also display more similar family behaviours. Doing so, we hope to discover genetic and/or linguistic bases of family behaviours.

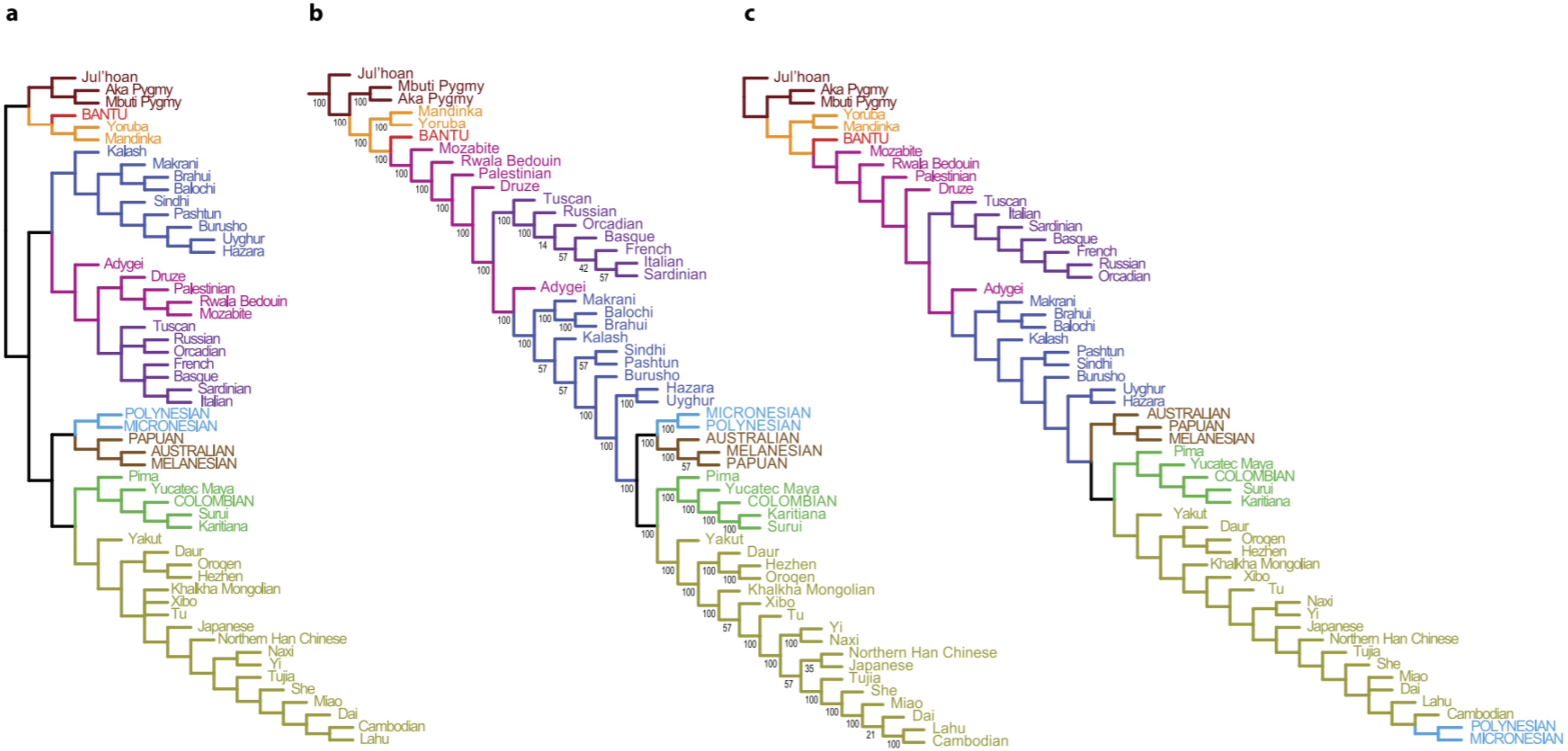
Joan Ryan, University of Pennsylvania

The supertree, or taxonomic congruence, approach was adopted by Duda and Zrzavy (2016) to construct a comprehensive hierarchical topology that summarized existing phylogenetic hypotheses between linguistic and ethnic groups. As the constructed supertree dataset was unpublished by the authors, we generated our own covariance matrices that indicated how close each ethno-linguistic group is to another on the supertree. We referred to the visual supertrees presented in the authors' paper, and manually matched groups as coded in the DHS datasets for all countries, to the ethno-linguistic groups presented in the supertree diagrams (Appendix A). Doing so, we were able to place each group on its respective branch, and then determine the pairwise patristic distance between two groups by counting the number of apomorphic step changes that separated them. Adoting Gaussian process regression (McElreath 2015), we can then build a covariance matrix (pg. 416) that represents the groups' pairwise distances.

We further plan to compare these covariances against indicators of each group's aggregated indicators of family behaviour that are clustered by similarity. These indicators include five categories:

1) family formation, 2) gender relations, 3) household structure, 4) reproduction, and 5) decisions of timing marriage and birth. The specific measures within each category can be found in Appendix B. These measures were aggregated from individual-level data within the DHS, under the Global Family Change project (National Science Foundation #1729185). We hypothesize that similarities in family behaviour will correlate to smaller pairwise patristic distances in the supertree, suggesting some genetic and linguistic relationships between family behaviour. On top of the theoretical contribution towards understanding determinants of demographic behaviour, we also hope to add to the body comparative demographic research done on a global scale.

Appendix A: Visual representation of supertree formulated by Duda and Zrzavy (2016)



1.A	1.B	1.C	1.D
2.A	2.B	2.C	2.D
3.A	3.B	3.C	3.D
4.A	4.B	4.C	4.D

1.A	**1.B**	**1.C**	**1.D**
2.A	**2.B**	**2.C**	**2.D**
3.A	**3.B**	**3.C**	**3.D**
4.A	**4.B**	**4.C**	4.D

1.A	1.B	1.C	1.D
2.A	2.B	2.C	2.D
3.A	3.B	3.C	3.D
4.A	4.B	4.C	**4.D**

Appendix B: Measures of family behaviour

Categories	**Measure**
Family	Prevalence of cohabitation
Family	Prevalence of divorce
Family	Prevalence of marriage
Family	Prevalence of remarriage
Gender	Female headship (non-single mothers)
Gender	Prevalence of hypogamus couples
Gender	Prevalence of hypogamus couples among heterogamus
Gender	Prevalence of paid work (married women)
Gender	Prevalence of paid work (mothers in couple)
Gender	Prevalence of paid work (mothers)
Gender	Women without sons / without daughters
Gender	Sex ratio at birth (girls/boys)
Household	Prevalence childless households
Household	Prevalence of complex households
Household	Prevalence of nuclear households
Household	Prevalence of three-generation households
Household	Prevalence of unipersonal households
Household	Prevalence ons single-mother households
Reproduction	Net Reproduction Rate
Reproduction	Parity progression from 0 to 1
Reproduction	Parity progression from 1 to 2
Reproduction	Parity progression from 2 to 3
Reproduction	Parity progression from 3 to 4
Reproduction	Parity progression from 4 to 5
Reproduction	Prevalence of childlessness
Reproduction	Satisfied need for contraception
Timing	Mean age at first birth
Timing	Mean age at first marriage
Timing	Mean age at last birth
Timing	Mean time between marriage and first birth
Timing	Proportion of premarital first births