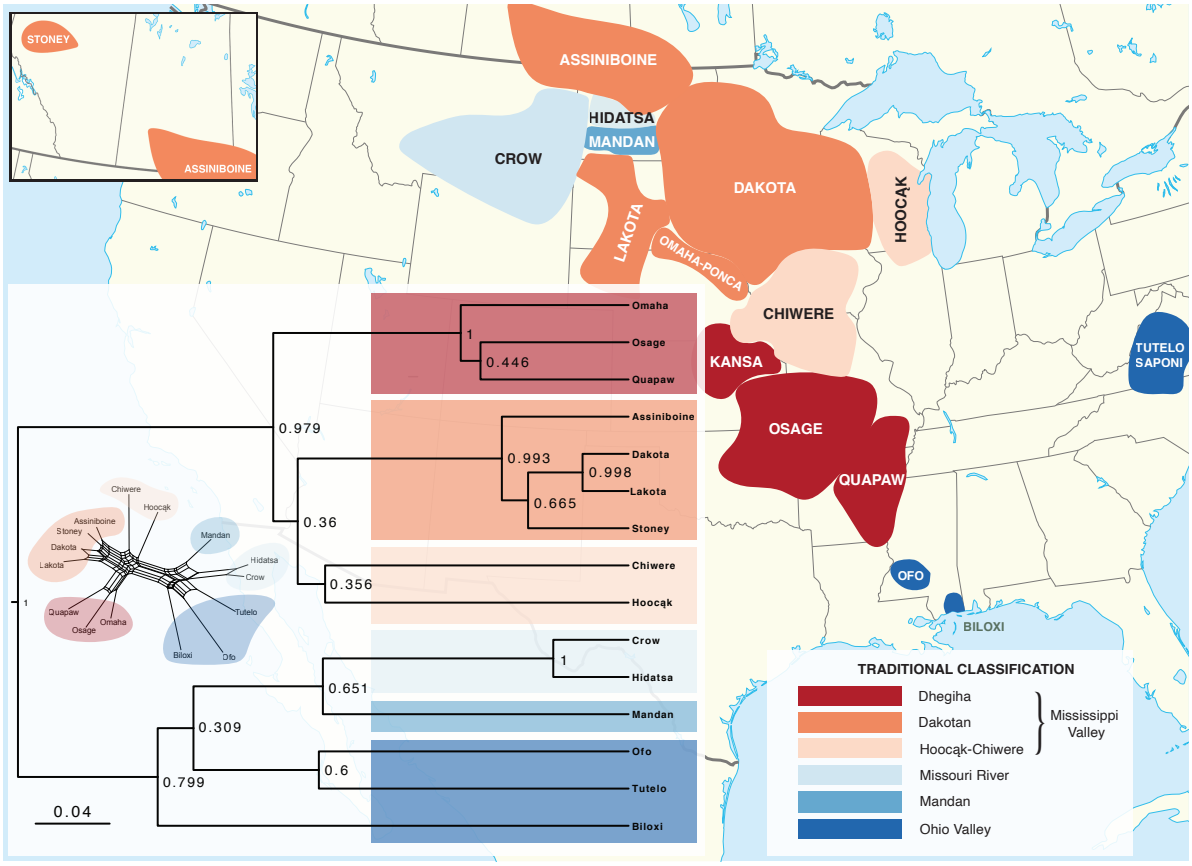
A Bayesian phylogenetic classification of the Siouan family using typological traits

(Workshop: Recent advances in computational historical linguistics; Format: Paper presentation)

Typological data have been used in a few computational phylolinguistic studies (e.g. Dunn et al., 2005, 2008; Sicoli and Holton, 2014). However, their use has been controversial as some argue that the signal produced is not genealogical but geographical (Donohue and Musgrave, 2007; Donohue et al., 2008, 2011). Moreover, Greenhill et al. (2020) consider lexical cognate data as the gold standard for use in computational phylolinguistics (e.g. subgrouping, dating). Despite such pessimism, I argue that there is still value in the use of typological traits in computational phylolinguistics, particularly for inferring the internal classification of language families. In this study, data was binarily coded for fifteen Siouan language varieties, a language family of North America, based on typological features drawn from WALS (Dryer and Haspelmath, 2013) and the modified version of traits listed in Sherzer (1976) used by Sicoli and Holton (2014). Both WALS and Sherzer (1976) contain interdependent and redundant features that were omitted insofar as they could be identified resulting in 248 characters per taxon. Following Bryant et al. (2005), I employed a network analysis and determined that the data was sufficiently tree-like (δ = 0.303, Q-residual = 0.026). I ran the analysis in BEAST 2.6.6 for 10 million steps with a 1k sampling frequency and 25% burn-in. The results are shown below using a binary CTMC+Γ, birth-death prior, and relaxed clock. Traditional subgroups, Missouri, Mississippi, Dakotan, and Dhegihan, were recovered with high posterior of $>$ 0.8. Ohio was grouped with Missouri and Mandan despite their geographical distance, and neither Missouri nor Mandan were grouped with Mississippi despite their geographical proximity. Thus, the results suggest that typological traits can be used reliably to infer language family subgroups.



References

Bryant, D., Filimon, F., and Gray, R. D. (2005). Untangling our past: languages, trees, splits and networks. In Mace, R., Holden, C. J., and Shennan, S. J., editors, *The evolution of cultural diversity: a phylogenetic approach*, pages 77–93. London, UK: UCL Press.

Donohue, M. and Musgrave, S. (2007). Typology and the linguistic macrohistory of Island Melanesia. *Oceanic linguistics*, pages 348–387.

Donohue, M., Musgrave, S., Whiting, B., and Wichmann, S. (2011). Typological feature analysis models linguistic geography. *Language*, pages 369–383.

Donohue, M., Wichmann, S., and Albu, M. (2008). Typology, areality, and diffusion. *Oceanic Linguistics*, pages 223–232.

Dryer, M. S. and Haspelmath, M., editors (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Dunn, M., Levinson, S. C., Lindström, E., Reesink, G., and Terrill, A. (2008). Structural phylogeny in historical linguistics: Methodological explorations applied in Island Melanesia. *Language*, pages 710–759.

Dunn, M., Terrill, A., Reesink, G., Foley, R. A., and Levinson, S. C. (2005). Structural phylogenetics and the reconstruction of ancient language history. *Science*, 309(5743):2072–2075.

Greenhill, S. J., Heggarty, P., and Gray, R. D. (2020). Bayesian phylolinguistics. In Janda, R., Joseph, B., and Vance, B., editors, *The Handbook of Historical :inguistics, Volume II*, pages 226–253. Wiley Online Library.

Sherzer, J. (1976). *An Areal-Typological Study of American Indian Languages North of Mexico.* North-Holland Publishing Company.

Sicoli, M. A. and Holton, G. (2014). Linguistic phylogenies support back-migration from Beringia to Asia. *PLoS One*, 9(3):e91722.