A Bayesian phylogenetic study of the Siouan language family using typological data

Edwin Ko

University of California, Berkeley

41st Siouan and Caddoan Languages Conference, 2021

Acknowledgments

I acknowledge with respect that the University of Berkeley, California resides on the traditional, ancestral, and unceded land of the Ohlone people.

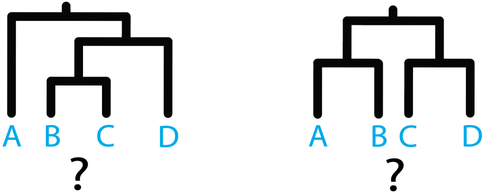
My sincerest thanks and gratitude to John Boyle, Andrew Garrett, David Kaufman, Rory Larson, Tyler Lemon, Sarah Lundquist, Julie Marsault, Armik Mirzayan, and Corey Roberts.

Table of Contents

Table of Contents

Introduction

- Reconstructing the linguistic history of a language family involves making inferences based on available information.
- Because we do not know what the true history is, there is a degree of uncertainty associated with our inferences.
- When there are many possible hypotheses, it is important to quantify these uncertainties to determine the most likely ones.
  - Sound change: $*ch > k$ or $*k > ch$?
  - Subgrouping:



Computational phylogenetic methods

- Determining the internal subgrouping of any language family is a non-trivial and computationally-intensive task.
  - There are approximately 17 Siouan languages which amount to 2.6×10^{42} possible trees.
- Bayesian phylogenetic methods allow us to estimate the possible trees that have the highest likelihood given the available data.
- While these tools can be very useful, the results are only as good as the data and model assumptions that are employed.

Data used in phylolinguistic research

- Most phylolinguistic studies use lexical data for classification of the following types (Chang et al. 2015):
  - Cognate: Proto-Siouan *$ahpá$ > Crow *apá* 'nose', Biloxi *pá* 'head' (Rankin et al. 2015)
  - Root-meaning: 'nose'
    1. Crow *apé*, Hidatsa *abá* (Boyle & Gwin 2006:70)
    2. Mandan *páaxu* (Kasak 2019:201, Ex.3.35e), Lakota *p̌hasú* (Ullrich 2019)
- Other studies have incorporated typological, structural data, although this has been controversial.
  - Dunn et al. (2005) use computational phylogenetic methods with typological features to argue for a shared historical association between Austronesian and Papuan languages.

Typological data: phylogeny or geography?

- The Austronesian-Papuan controversy in a nutshell:
  - Typological features can detect a geneological signal!
    - Dunn et al. 2005, 2007, 2008, Dunn 2009
  - Wrong, typological features detect a geographical signal!
    - Donohue & Musgrave 2007, Donohue et al. 2008, 2011
- Sicoli and Holton (2014) also used typological features to infer the true of the Dene-Yeneseian macro-family.
- It is still unclear how reliable typological features are in inferring the true phylogenetic tree.
  - Typological features are often thought of as being easily diffusable across geographical space (e.g. Holman et al. 2007).

Purpose of this talk

Research questions

1. Can typological/structural data be used to detect a phylogenetic signal or does it indicate a geographical signal?
2. How do the results compare with previously proposed classifications of the Siouan language family?

Table of Contents

Typological data used in previous studies

- Dunn et al. 2008: 115 binary features, selected (i) "to provide broad coverage," (ii) "to distinguish between the languages of Island Melanesia," and (iii) "on which Austronesian and Papuan languages generally diverge" (Dunn et al. 2008:730)
- Sicoli & Holton 2014: 116 binary features, Sherzer's (1976) *An areal-typological study of American Indian languages north of Mexico*
  - Yanovich (2020) argues against Sicoli & Holton: there is too little data and thus the inferences are not robust

Lessons learned

- Incorporate more data and check for robustness of inferences

Data used in this study

- World Atlas of Language Structures (WALS; Dryer & Haspelmath 2013) has been used in phylogenetic studies:
  - Distance-based: Wichmann & Saunders 2007, Donohue et al. 2011, 2012, Greenhill et al. 2010
  - Character-based: Wichmann & Saunders 2007, Dediu 2011, Maurits & Griffiths 2014
- Sixteen Siouan languages: Crow, Hidatsa, Mandan, Quapaw, Osage, Omaha, Ho-chunk, Ioway-Oto, Chiwere, Assiboine, Lakota, Dakota, Stoney, Biloxi, Ofo, and Tutelo
- WALS has many missing information and several inaccuracies:
  - Hidatsa: Incorrectly coded as having noun-demonstrative order based on Matthews's 1965 *Hidatsa Syntax*
  - Osage: Incorrectly coded as having nominal plural citing Quintero (1997:33) – there is no mention of plurality on p.33
    - Likely a typo but p.330–339 show plural *api* following verbs

Data used in this study

- Using most of the features from Sherzer and WALS, I coded from scratch employing more recent and reliable sources, and cross-checking with other Siouanists (thanks again!).
- 258 binary features (153 WALS, 105 Sherzer):
  - 189 features are (potentially) parsimony-informative (i.e. have different values for at least two languages).
    - Only Crow and Hidatsa lack nasal vowels.
    - Only Quapaw lacks /u/ (Rankin 2005:463).
  - 69 features known to have uniform values across all languages
    - All Siouan languages have a mid or mid-high vowel.
    - No Siouan languages employ plural particles on nominals.

Feature type	Percent
Morphological	∼50%
Phonological	∼38%
(Morpho)syntactic	∼8%
(Lexico)semantic	∼4%

Coding the data

- Binary coding for presence ('1') or absence ('0') of features
  - Missing data is coded as missing ('?').
- Features that have multiple values become separate features
  - `Negative Morphemes` has four values (affix/clitic, particle, double, auxiliary word) is converted to four distinct features
  - If a language has a negative affix, then it likely does not also have a negative auxiliary word.
- Issue of interdependent data:
  - Unfortunately, this is common practice even with lexical cognacy data that use binary coding which violates an assumption of independence with Bayesian methods.
  - It would be ideal to use mutli-state characters.

Table of Contents

NeighborNet: How tree-like is the data?

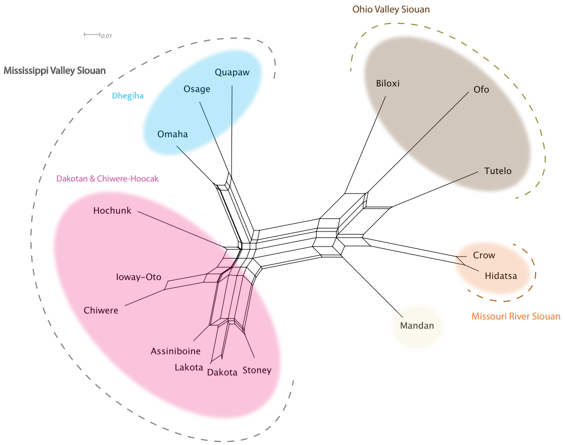
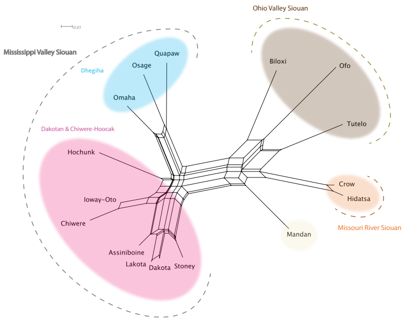


Figure: Splits graph using NeighborNet (Brant & Moulton 2004). Boxes and reticulations (i.e. web-like patterns) indicate conflicting signals.

NeighborNet: How tree-like is the data?



Language	δ-score	Q-residual
Crow	0.227	0.0185
Hidatsa	0.234	0.0204
⇒ **Mandan**	**0.274**	**0.0232**
Lakota	0.221	0.0165
Dakota	0.232	0.0154
Assiniboine	0.242	0.0193
Stoney	0.281	0.0219
⇒ **Chiwere**	**0.273**	**0.0232**
Ioway-Oto	0.271	0.0193
⇒ **Hochunk**	**0.329**	**0.0226**
⇒ **Osage**	**0.306**	**0.0334**
Omaha	0.252	0.0194
Quapaw	0.257	0.0185
Tutelo	0.248	0.0228
⇒ **Biloxi**	**0.273**	**0.0248**
⇒ **Ofo**	**0.304**	**0.0387**
Average	0.264	0.0224

- δ-scores and Q-residuals: 0 (less conflict) to 1 (more conflict), where conflict represents more sharing of traits with other languages.

Comparison of δ-scores and Q-residuals

- How do the δ-score and Q-residual compare with those reported for other language groups?

Language group	δ-score	Q-residual	Data type	Source
Siouan	0.264	0.0224	Typological/Structural	—
Dene-Yeneseian	0.367	0.0492	Typological/Structural	Sicoli & Holton 2014
Austronesian	0.44	0.0354	Typological/Structural	Greenhill et al. 2017
Indo-European	0.23	0.003	Lexical	Gray et al. 2010
Polynesian	0.41	0.020	Lexical	Gray et al. 2010
Ainu	0.25	0.01	Lexical	Lee & Hasegawa 2013
(Mainland) Japanese	0.39	0.002	Lexical	Lee & Hasegawa 2014
Ryukyuan	0.23	0.004	Lexical	Lee & Hasegawa 2014
Chapacuran	0.262	0.016	Lexical	Birchall et al. 2016
Austronesian	0.38	0.0062	Lexical	Greenhill et al. 2017
Dravidian	0.30	0.0069	Lexical	Kolipakam et al. 2018
Tai	0.2808	0.04088	Lexical	Dockum 2018
Turkic	0.34	0.001	Lexical	Savelyev & Robbeets 2020

Key takeaway

The splits graph, δ-score, and Q-residual for the Siouan data is well within the range of what is considered **tree-like**.

Estimating the true tree

- The main goal is to obtain a sample of trees (not just one tree) that explains the data relatively well.
- To do this, the algorithm (Monte Carlo Markov Chain) searches the space of all possible trees step-by-step locating the trees that best fit the data.
- I ran the analysis in `BEAST 2.6.3` (Bouckaert et al. 2019) using 10 million steps (generations) with a 1,000 sampling frequency and 25% burn-in resulting in a total of 7,500 trees.
  - This process was repeated two additional times to checked to ensure the results are similar across the three independent runs.

Assessing performance: Effective sample size (ESS)

- We want to have a good sample of trees, but how can we tell if the sample is sufficient?
- Using `Tracer` (Rambaut et al. 2018), ESS values over 625 are considered to indicate sampling independence (Fabreti & Hoehna 2021).

Statistic	ESS
Posterior	5114
Likelihood	3403
Prior	2265
treeLikelihood.wals-sherzer	3403
TreeHeight.t:tree	3832
gammashape.s:wals-sherzer	709

...

Assessing performance: "Fuzzy caterpillars"

- We also want to know if the model converged; that is, did it end in a state of equilibrium?
- We can also look for "fuzzy caterpillars" in the trace.

Figure: Woolly Bear Caterpillar (from Herald Times Reporter)

Assessing performance: “Fuzzy caterpillars”

- We also want to know if the model converged; that is, did it end in a state of equilibrium?
- We can also look for “fuzzy caterpillars” in the trace.

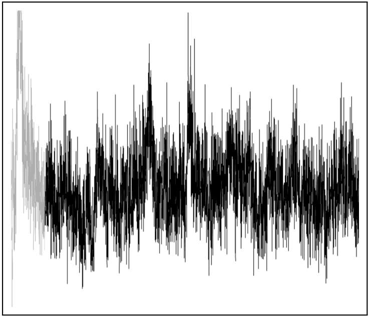


Figure: Non-fuzzy caterpillar trace (Source: Taming the Beast)

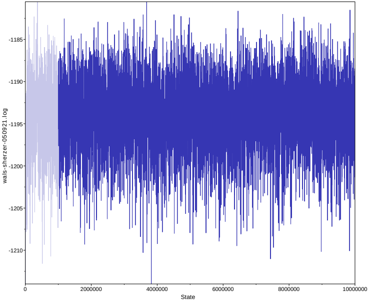


Figure: Fuzzy caterpillar trace of posterior probability

Table of Contents

Summary tree: Maximum clade credibility tree

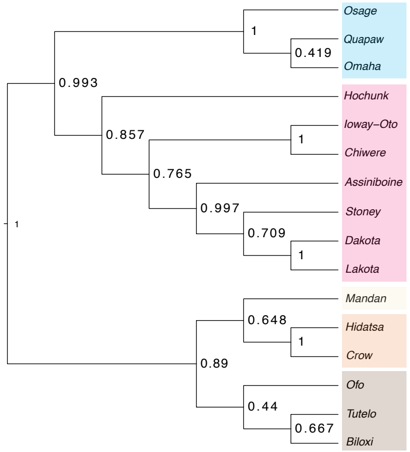


Figure: Maximum clade credibility tree. Values indicate relative frequency of the sampled trees that contain the particular branching.

Visualizing all trees: DensiTree

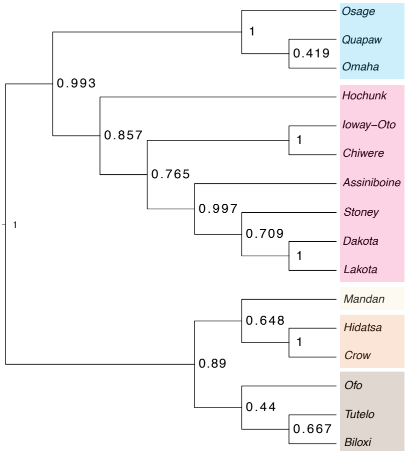


Figure: Maximum clade credibility

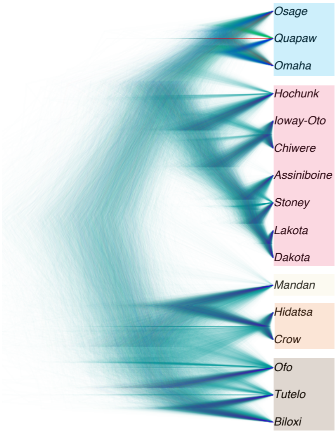


Figure: DensiTree (Bouckaert 2010, Bouckaert & Heled 2014)

Table of Contents

Comparison with Rankin's (2010) proposed tree

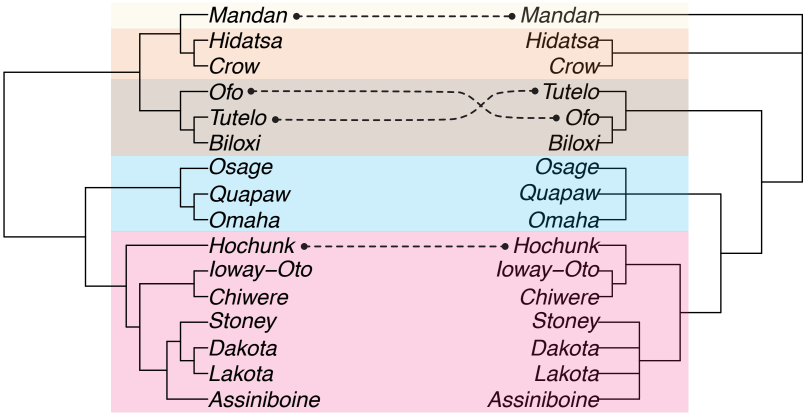


Figure: Comparison between current analysis (left) and the tree proposed by Rankin (2010; right). Dashed lines indicate sites of divergence.

Evidence of a geographical signal?

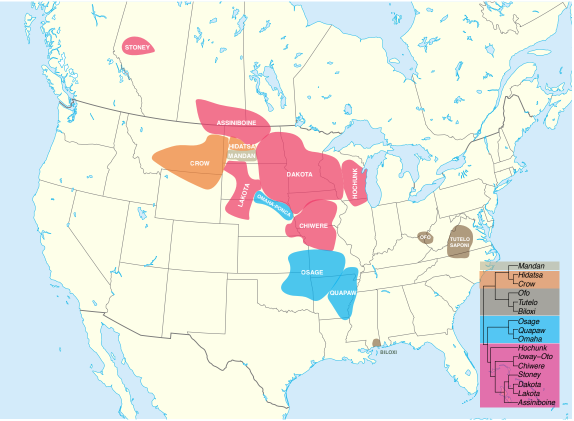


Figure: Map of selected Siouan languages (adapted from Wikimedia Commons). *Disclaimer:* This map is a rough approximation of the language communities and their geographical locations.

A qualitative assessment: Non-contiguous languages

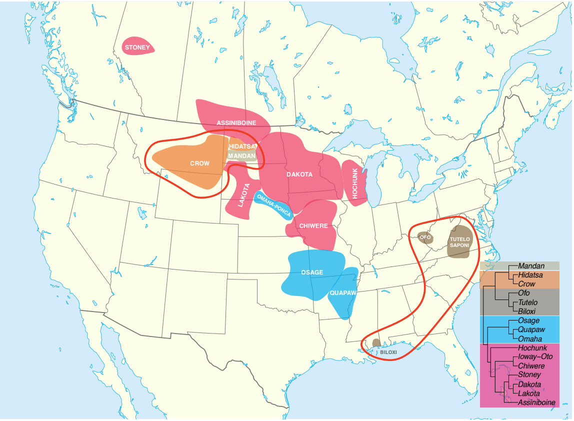


Figure: Map of selected Siouan languages (adapted from Wikimedia Commons). *Disclaimer:* This map is a rough approximation of the language communities and their geographical locations.

A qualitative assessment: Contiguous languages

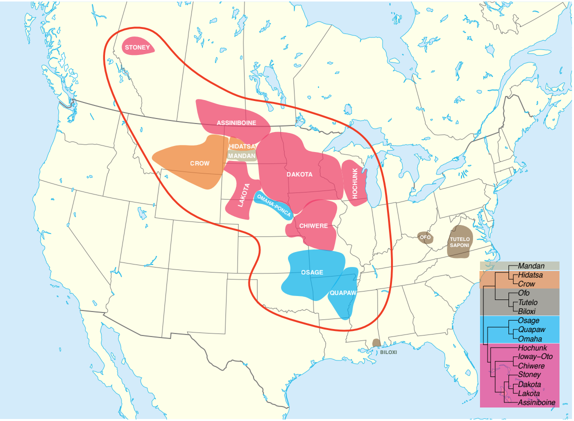


Figure: Map of selected Siouan languages (adapted from Wikimedia Commons). *Disclaimer:* This map is a rough approximation of the language communities and their geographical locations.

Table of Contents

Future directions

- Identifying the features responsible for the subgroupings
  - Phonological features:
    - Higher-level subgroupings
    - Mandan ⇒ Crow, Hidatsa
  - Morphological features:
    - Lower-level subgroupings
    - Mandan ⇒ Mississippi Valley
- Quality-checking the data again (and again)
  - If any linguists would be willing to take a look at (a subset of) the data, I would greatly appreciate it!
- Incorporating Catawba and Yuchi typological data for inferring deeper historical relations
  - Catawba and Yuchi are grouped with the Siouan languages suggesting perhaps that deeper time depths increases the potential for conflicting signals.

Future directions

- Comparing analyses with lexical data (e.g. Kasak, n.d.)
    - Data in the *Comparative Siouan Dictionary* need to be checked thoroughly
- Checking the results with other linguistic (e.g. shared innovations) and historical evidence
    - Thoughts on the possibility of grouping Missouri River (Crow and Hidatsa) with Ohio Valley (Biloxi, Tutelo, and Ofo)?
    - Some potentially shared innovations:
        1. Loss of glottalized consonants
        2. Collapse of the distinction arrive here/there
        3. Emergence of distinct nominal and verbal conjunctions
    - Impressionistically, the Missoui River and Ohio Valley Siouan subgroups appear quite distinct.

Ahóo!

Thank you for listening!

References I

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., et al. (2019).
BEAST 2.5: An advanced software platform for bayesian evolutionary analysis.
PLoS computational biology, 15(4):e1006650.

Bouckaert, R. R. and Heled, J. (2014).
DensiTree 2: Seeing trees through the forest.
BioRxiv, page 012401.

Boyle, J. P. and Gwin, A. (2006).
English-Hidatsa and Hidatsa-English wordlist.
Ms. Mandaree, ND.(Preliminary version).

Bryant, D. and Moulton, V. (2004).
Neighbor-net: an agglomerative method for the construction of phylogenetic networks.
Molecular biology and evolution, 21(2):255–265.

Chang, W., Hall, D., Cathcart, C., and Garrett, A. (2015).
Ancestry-constrained phylogenetic analysis supports the indo-european steppe hypothesis.
Language, pages 194–244.

Dediu, D. (2011).
A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone.
Proceedings of the Royal Society B: Biological Sciences, 278(1704):474–479.

Dockum, R. (2018).
Phylogeny in Phonology: How Tai Sound Systems Encode Their Past.
In *Proceedings of the Annual Meetings on Phonology*, volume 5.

References II

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. (2009).
Contact and phylogeny in Island Melanesia.
Lingua, 119(11):1664–1678.

Dunn, M., Foley, R., Levinson, S., Reesink, G., and Terrill, A. (2007).
Statistical reasoning in the evaluation of typological diversity in Island Melanesia.
Oceanic Linguistics, pages 388–403.

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.

References III

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.

Fabreti, L. G. and Hoehna, S. (2021).
Convergence Assessment for Bayesian Phylogenetic Analysis using MCMC simulation.
bioRxiv.

Greenhill, S. J., Atkinson, Q. D., Meade, A., and Gray, R. D. (2010).
The shape and tempo of language evolution.
Proceedings of the Royal Society B: Biological Sciences, 277(1693):2443–2450.

Holman, E. W., Schulze, C., Stauffer, D., and Wichmann, S. (2007).
On the relation between structural diversity and geographical distance among languages: observations and computer simulations.

Kasak, R.
A Computational Approach to Siouan Phylogenetics.
Manuscript.

Kasak, R. (2019).
Affix ordering and templatic morphology in Mandan. PhD Dissertation, Yale university.

Matthews, G. H. (1965).
Hidatsa syntax.
Number 3. Mouton.

Maturana, P., Brewer, B. J., Klaere, S., and Bouckaert, R. R. (2019).
Model selection and parameter inference in phylogenetics using nested sampling.
Systematic biology, 68(2):219–233.

References IV

Maurits, L. and Griffiths, T. L. (2014).
Tracing the roots of syntax with Bayesian phylogenetics.
Proceedings of the National Academy of Sciences, 111(37):13576–13581.

Quintero, C. F. (1997).
Osage phonology and verbal morphology.
University of Massachusetts Amherst.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018).
Posterior summarization in Bayesian phylogenetics using Tracer 1.7.
Systematic biology, 67(5):901.

Rankin, R. L. (2005).
Quapaw.
In Hardy, H. K. and Scancarelli, J., editors, *Native languages of the Southeastern United States*, Studies in the anthropology of North American Indians, pages 454–498. University of Nebraska Press, Lincoln.

Rankin, R. L. (2010).
The place of Mandan in the Siouan language family.
In *30th annual Siouan and Caddoan Languages Conference. Northeastern Illinois University: Chicago, IL*.

Rankin, R. L., Carter, R. T., Jones, A. W., Koontz, J. E., Rood, D. S., and Hartmann, I. (2015).
Comparative Siouan dictionary.
Leipzig: Max Planck Institute for Evolutionary Anthropology. Available on May, 5:2018.

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

References V

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

Ullrich, J. F. (2019).
New Lakota dictionary online.
Bloomington, Lakota Language Consortium (https://www. lakotadictionary. org/phpBB3/nldo. php).

Wichmann, S. and Saunders, A. (2007).
How to use typological databases in historical linguistic research.
Diachronica, 24(2):373–404.

Yanovich, I. (2020).
Phylogenetic linguistic evidence and the Dene-Yeniseian homeland.
Diachronica, 37(3):410–446.

Seeing the forests for the trees

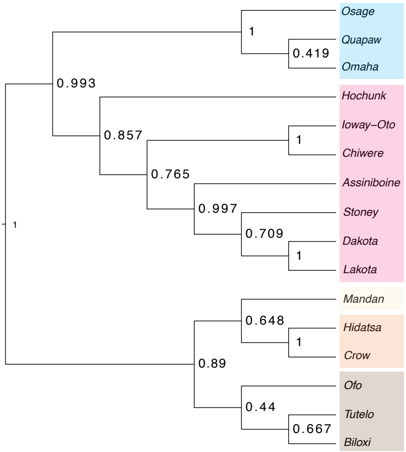


Figure: Maximum clade credibility

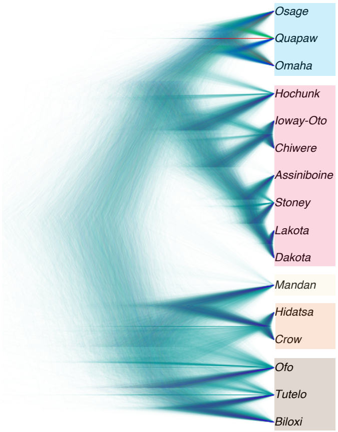


Figure: DensiTree (Bouckaert 2010, Bouckaert & Heled 2014)

Model comparison

- Eighteen models with different settings were considered and `ctmc-bd-relax-gam` fits the data best.
- Marginal likelihood was estimated using the Nested Sampling algorithm (Maturana et al. 2019) using 12 particles.

Analysis	Substitution model	Tree prior	Clock model	Marginal log-likelihood	Bayes factor
⇒ **ctmc-bd-relax-gam**	**CTMC + γ**	**Birth-death**	**Relaxed**	**-1247.3**	—
ctmc-yule-relax-gam	CTMC + γ	Yule (pure birth)	Relaxed	-1247.9	1.2
ctmc-bdsky-relax-gam	CTMC + γ	Birth-death skyline	Relaxed	-1249.4	4.2
cov-yule-strict	Covarion	Yule (pure birth)	Strict	-1251.1	7.6
ctmc-yule-strict-gam	CTMC + γ	Yule (pure birth)	Strict	-1251.2	7.8
ctmc-bdsky-strict-gam	CTMC + γ	Birth-death skyline	Strict	-1251.8	9.0
ctmc-bd-strict-gam	CTMC + γ	Birth-death	Strict	-1253.7	12.8
cov-yule-relax	Covarion	Yule (pure birth)	Relaxed	-1256.1	17.6
cov-bd-relax	Covarion	Birth-death	Relaxed	-1256.1	17.6
cov-bd-strict	Covarion	Birth-death	Strict	-1256.7	18.8
cov-bdsky-strict	Covarion	Birth-death skyline	Strict	-1258.5	22.4
cov-bdsky-relax	Covarion	Birth-death skyline	Relaxed	-1258.7	22.8
ctmc-yule-strict	CTMC	Yule (pure birth)	Strict	-1284.5	74.4
ctmc-yule-relax	CTMC	Yule (pure birth)	Relaxed	-1285.5	76.4
ctmc-bd-relax	CTMC	Birth-death	Relaxed	-1286.0	77.4
ctmc-bdsky-relaxed	CTMC	Birth-death skyline	Relaxed	-1288.1	81.6
ctmc-bdsky-strict	CTMC	Birth-death skyline	Strict	-1290.9	87.2
ctmc-bd-strict	CTMC	Birth-death	Strict	-1293.3	92.0

Note: Interpreting BF: 1–2: weak, 2–6: positive, 6–10: strong, >10: very strong.

Model comparison

- Four other randomly-selected models `cov-yule-relax`, `ctmc-yule-relax-gam`, `ctmc-bdsky-strict-gam`, and `cov-bdsky-relax` produced similar tree topologies suggesting that the analysis is robust to the choice of tree priors (see Yanovich 2020).

Key takeaways

- There is **some evidence for `ctmc-bd-relax-gam`** to explain the data better than other models.
- The **dataset is sufficient enough in size** to make robust inferences about most likely trees given the data.

WALS features (1/3) – non-binary

1. Consonant Inventories
2. Vowel Quality Inventories
3. Consonant-Vowel Ratio
4. Voicing in Plosives and Fricatives
5. Voicing and Gaps in Plosive Systems
6. Uvular Consonants
7. Glottalized Consonants
8. Lateral Consonants
9. The Velar Nasal
10. Vowel Nasalization
11. Front Rounded Vowels
12. Syllable Structure
13. Tone
14. Absence of Common Consonants
15. Presence of Uncommon Consonants
16. Exponence of Selected Inflectional Formatives
17. Locus of Marking in the Clause
18. Locus of Marking in Possessive Noun Phrases
19. Locus of Marking: Whole-language Typology
20. Zero Marking of A and P Arguments
21. Prefixing vs. Suffixing in Inflectional Morphology
22. Case Syncretism
23. Syncretism in Verbal Person/Number Marking
24. Number of Genders
25. Sex-based and Non-sex-based Gender Systems
26. Systems of Gender Assignment
27. Coding of Nominal Plurality
28. Occurrence of Nominal Plurality
29. Plurality in Independent Personal Pronouns
30. Associative Plural
31. Definite Articles
32. Definite Affix
33. Indefinite Articles
34. Indefinite Affix
35. Inclusive/Exclusive Distinction in Independent Pronouns
36. Inclusive/Exclusive Distinction in Verbal Inflection
37. Distance Contrasts in Demonstratives
38. Pronominal and Adnominal Demonstratives

WALS features (2/3) – non-binary

39. Third Person Pronouns and Demonstratives
40. Gender Distinctions in Independent Personal Pronouns
41. Politeness Distinctions in Pronouns
42. Indefinite Pronouns
43. Intensifiers and Reflexive Pronouns
44. Person Marking on Adpositions
45. Number of Cases
46. Position of Case Affixes
47. Comitatives and Instrumentals
48. Ordinal Numerals
49. Numeral Classifiers
50. Conjunctions and Universal Quantifiers
51. Position of Pronominal Possessive Affixes
52. Possessive Classification
53. Adjectives without Nouns
54. Noun Phrase Conjunction
55. Nominal and Verbal Conjunction
56. Perfective/Imperfective Aspect
57. The Past Tense
58. The Future Tense
59. The Perfect
60. Position of Tense-Aspect Affixes
61. The Morphological Imperative
62. The Prohibitive
63. Imperative-Hortative Systems
64. Semantic Distinctions of Evidentiality
65. Coding of Evidentiality
66. Verbal Number and Suppletion
67. Order of Subject, Object and Verb
68. Order of Adposition and Noun Phrase
69. Order of Genitive and Noun
70. Order of Adjective and Noun
71. Order of Demonstrative and Noun
72. Order of Numeral and Noun
73. Order of Relative Clause and Noun
74. Order of Degree Word and Adjective

WALS features (3/3) – non-binary

75. Position of Polar Question Particles
76. Position of Interrogative Phrases in Content Questions
77. Relationship between the Order of Object and Verb and the Order of Adposition and Noun Phrase
78. Relationship between the Order of Object and Verb and the Order of Relative Clause and Noun
79. Relationship between the Order of Object and Verb and the Order of Adjective and Noun
80. Alignment of Case Marking of Full Noun Phrases
81. Alignment of Case Marking of Pronouns
82. Alignment of Verbal Person Marking
83. Expression of Pronominal Subjects
84. Verbal Person Marking
85. Third Person Zero of Verbal Person Marking
86. Order of Person Markers on the Verb
87. Reciprocal Constructions
88. Passive Constructions
89. Antipassive constructions
90. Applicative constructions
91. Nonperiphrastic Causative Constructions
92. Negative Morphemes
93. Polar Questions
94. Predicative Adjectives
95. Zero Copula for Predicate Nominals
96. 'Want' Complement Subjects
97. Hand and Arm
98. Finger and Hand
99. Numeral Bases
100. Green and Blue
101. Red and Yellow
102. M-T Pronouns
103. M in First Person Singular
104. N-M Pronouns
105. M in Second Person Singular
106. Position of Negative Word With Respect to Subject, Object, and Verb

Sherzer features (1/3) – binary

1. Three vowel
2. 1-1-1
3. 2-1
4. Four vowel
5. 2-2
6. 2-1-1
7. 1-2-1
8. Five vowel
9. 3-2
10. 3-1-1
11. 2-2-1
12. Six vowel
13. 2-2-2
14. 2-3-1
15. 3-2-1
16. Seven vowel
17. 2-2-2-1
18. 3-3-1
19. Voiceless vowel
20. Nasal vowel
21. not a,e,i,o,u
22. vowel length contrast
23. mid or mid-high vowel
24. one stop series: voiceless
25. two stop series: voiceless/voiced
26. two stop series: voiceless/glottalized
27. three stop series: voiceless/voiced/glottalized
28. four stop series
29. glottalized stop series
30. labial stop present
31. c/tʃ
32. k/č
33. k/q
34. either k/č or k/q
35. t
36. q
37. kw
38. qw

Sherzer features (2/3) – binary

39. one fricative series: voiceless
40. two fricative series: voiceless/voiced
41. two fricative series: voiceless/glottalized
42. three fricative series: voiceless/voiced/glottalized
43. glottalized fricatives
44. pharyngeal fricatives
45. labial fricative
46. θ
47. ð
48. s/ʃ
49. z
50. x
51. xw
52. x.
53. x.w
54. ɣ
55. ɣw
56. h
57. hw
58. l
59. ɬ
60. tɬ
61. tɬ'
62. dl
63. l'
64. ɬ'
65. ly
66. ɬy
67. voiceless nasal
68. glottalized nasal
69. ɲ
70. ŋ
71. r
72. voiceless r
73. glottalized r
74. r/l
75. voiceless semivowel

Sherzer features (3/3) – binary

76. glottalized semivowel
77. possessive pronouns independent morpheme
78. alienable/inalienable possession?
79. reduplication = distributive or plual
80. reduplication = diminutive
81. augmentative-diminutive consonant symbolism
82. masculine/feminine gender
83. animate/inanimate gender
84. plural in pronouns
85. inclusive/exclusive plural in pronouns
86. dual in pronouns
87. dual in nouns
88. inclusive/exclusive dual in pronouns
89. demonstratives for visible/invisible objects
90. numerals classified by form or shape of object
91. locative prefixes
92. locative suffifxes
93. locative prepositions
94. locative postpositions
95. nominal incorporation
96. subject person marker prefixes
97. subject person marker suffixes
98. subject person markers independent pronouns
99. reduplication in verb = distribution, repetition
100. reduplication in verb = diminutive
101. evidential or source of information marked
102. instrumental markers
103. locative-directional markers
104. locative-directional markers prefix
105. locative-directional markers suffix