

The Phylogenetic Status of Pacific Coast Athabaskan: A Computational Assessment

JUSTIN SPENCE

University of California, Davis

1 Introduction

The Athabaskan language family, one of the most geographically widespread in North America, is canonically discussed according to its three non-contiguous regional concentrations: Northern Athabaskan in Alaska and western Canada, Apachean in the southwestern United States, and Pacific Coast Athabaskan (PCA) in northern California and southwestern Oregon. The historical significance of these regional groupings is unclear, however (Krauss 1973, Krauss and Golla 1981). The PCA languages are a case in point: while they are uncontroversially divided into an Oregon subfamily that includes Tututni, Tolowa, and Galice (inter alia), and a California subfamily consisting of Hupa, Mattole, Wailaki, and Kato, there is disagreement in the literature about whether PCA as a whole is a meaningful phylogenetic unit that reflects a higher-order historical reality within the family (Hoijer 1960, 1962; Golla 2011:69, 257). The present study considers the status of the Pacific Coast Athabaskan languages, both in relation to each other and in relation to the rest of the family, by applying computational approaches to phylogenetic inference adapted from the biological sciences that have invigorated historical linguistic research over the past decade. These methods have shed light on previously intractable problems, and in some cases sparked new controversies, in language families as diverse as Indo-European (Gray and Atkinson 2003; Atkinson and Gray 2006; Nakhleh et al. 2005; Bouckaert et al. 2012), Bantu (Holden and Gray 2006, Marten 2006), Austronesian (Bryant 2006, Dunn et al. 2008), and Pama-Nyungan (Bowerman and Atkinson 2012).

While the main goal of this study is to explore the status of PCA through computational analysis of lexical data, the discussion is set against a backdrop of broader theoretical and methodological interest. Previous research has suggested that applying branching tree-like models to the Athabaskan family has not been successful. Krauss (1973) in particular has argued that patterns of cross-cutting lexical and phonological isoglosses found throughout the family (especially Northern Athabaskan – cf. Krauss and Golla 1981) indicate a historical development involving local innovations that spread more or less widely within a vast dialect continuum. Given this scenario, it is of interest to determine how computational models perform where there are no well-established subgroups to compare the results against – and indeed, where it has been suggested that it is futile to seek such subgroups in the first place. The evolutionary models used in the present analysis are explicitly geared toward providing metrics for the degree of certainty associated with tree topologies and therefore can quantify the extent to which subgroups can be found within the family.

Another point of interest is methodological, relating to the way linguistic features (“characters” in the evolutionary idiom) are coded. Gray and Atkinson (2003) use a method for coding lexical datasets according to which characters with multiple states (i.e., meanings expressed by two or more cognates across a set of languages) are recoded in a binary scheme. Evans, Ringe, and Warnow (2006) argue that this recoding creates dependencies among characters and can potentially lead to biased results. Pagel and Meade (2006), however, maintain that dependencies among characters introduced in binary recoding will merely create scaled versions of the best topology. The present paper considers results based on both multi-state and binary codings of characters in light of this debate.

The paper is structured as follows. Section 2 provides an overview of the languages and data sources included in the study. Section 3 outlines the computational methods employed and gives some details about particular coding decisions that were made. Section 4 presents results, the main finding being that the Pacific Coast Athabaskan languages do indeed emerge as a subgroup that is well-supported under different data codings (multi-state vs. binary) and assumptions about rates of change across lineages (non-clock, strict clock, and relaxed clock evolutionary models). Another finding of interest is that differences in tree topologies are found under multi-state and binary codings of the data. Section 5 summarizes these findings and concludes.

2 Languages and Sources

The core lexical data for this research were harvested from the glottochronological study of Hoijer (1956), which contains Swadesh-100 lists for fifteen Athabaskan languages: the Northern languages Beaver, Dakelh (Carrier), Dëne Sųłiné (Chipewyan), Gwich’in (Kutchin), Hare, and Tsuut’ina (Sarcee); the Apachean

languages Chiricahua, Jicarilla, Lipan, Navajo, and San Carlos; and the Pacific Coast languages Galice, Hupa, Kato, and Mattole. Seven other Athabaskan languages were added to this set: the Pacific Coast languages Tututni (Golla 2008), Tolowa (Bommelyn 1989), and Wailaki (Li n.d.) and the Northern languages Ahtna (Kari 1990), Dena'ina (Kari 2007), Koyukon (Jetté and Jones 2000), and Minto (Krauss 1963, Kari 1994). The Athabaskan languages are widely accepted to be part of a larger Na-Dene family that also includes Eyak and Tlingit, which were included in the study using data in Krauss (1963). The PCA data were supplemented by information from various published and archival sources: Bright (1964), Collins (1989), Dorsey (1884), Golla (1996), Goddard (1902-1903, 1923), Harrington (1982-1990), Hoijer (1973), Landar (1977), Merriam (1850-1974), Tuttle (n.d.), and Waterman (n.d.).¹

More than 40 Athabaskan languages have been identified in the literature, so restricting the dataset to only slightly more than half this number exposes this study to one of Krauss' (1973) main criticisms of earlier phylogenetic research on Athabaskan: discrete, well-defined boundaries distinguishing Athabaskan languages from one another may be an artifact of the fact that documentation was collected at widely-dispersed checkpoints, thereby obscuring internal diversity and linguistic features that have diffused regionally within Krauss' Athabaskan dialect network. This criticism can be countered to some extent insofar as the dataset is surely adequate to allow reasonable inferences related to the status of PCA, the main research question addressed in this paper. Moreover, a moderate amount of dialect diversity has been accounted for by coding some characters as polymorphic. Nonetheless, Krauss' point is well-taken, and therefore for Northern Athabaskan especially the results reported in §4 must be considered provisional.

3 Methods

This study uses Bayesian phylogenetic inference using the software program MrBayes (Ronquist et al. 2012) to address questions of subgrouping related to Pacific Coast Athabaskan. The fundamental premise of this and other recent computational approaches in historical linguistics is that lexical items are heritable traits that evolve in a manner that can be understood as analogous to evolution in biology. The methodology is summarized in this section, following the discussions in Nichols and Warnow (2008), Atkinson and Gray (2006), Pagel and Meade (2006), and Bowerman and Atkinson (2012). In the interests of space, certain details about the coding procedures employed have been omitted from this paper; further discussion can be found in Spence (2013).

¹ Excluded from this study is the Oregon Athabaskan language Upper Umpqua. Preliminary analysis of the available data for this sparsely-documented language and its impact on the results presented in §4 can be found in Spence (2013).

The goal of the computational analysis is to infer the evolutionary tree that makes a set of observed comparative data – coded as a matrix of similarities and differences between languages according to the procedure outlined below – most likely given a model of evolution. Ideally the likelihood of every possible tree would be inspected and the winning candidate(s) selected, but the large number of such trees makes this computation effectively impossible. Therefore, Markov chain Monte Carlo (MCMC) sampling is employed to generate a set of trees from which the best phylogeny is inferred. An analysis proceeds by making incremental changes to a randomly-selected starting tree. At each step, if changes increase the likelihood of the data given the evolutionary model, the tree is added to the sample set. If the changes decrease the likelihood, the tree is retained in the sample some relatively low proportion of the time. At the end of the analysis – after a sufficient number of trees have been generated – the sample reflects the posterior probability distribution of the universe of possible trees. The sample can then be summarized to indicate the extent to which particular features (tree topologies and branch lengths) occur. The proportion of trees that a subgroup appears in can be taken as an approximation of its posterior probability – i.e., the degree of certainty that a subgroup is part of the true evolutionary history of the language family.

As input to MrBayes in the present study, a data matrix was constructed whose rows were the twenty-four languages and whose columns were the meanings in the Swadesh-100 list. Each cell in the matrix contained one or more integers indicating the cognate set(s) found in a given language in a given meaning, as illustrated in the table in (1) for the meanings ‘blood’, ‘two’, and ‘cloud’:

(1) Multi-state character coding

	‘blood’	‘two’	‘cloud’
Tlingit	2	2	2
Eyak	0	1	0
Gwich’in	3	0	0
Navajo	0	0	0
Hupa	1	0	1
Kato	1	0	1
Tututni	0	0	1

Taking ‘blood’ as an example, most Athabaskan languages and Eyak have a reflex of *dəl in this meaning, so these languages were assigned the integer 0 to indicate that they all have the same state for this character. The California Athabaskan languages all have a word cognate with Hupa *tse:lin* in this meaning, so these languages were assigned character state 1. Gwich’in (Kutchin) and Tlingit both have distinct words for this meaning, so they were coded with unique

integers ‘2’ and ‘3’. This procedure was repeated for all items and all languages, yielding a data matrix with 100 items for 24 languages.² One item had more than the 10 states allowed for integer-valued data in MrBayes and was discarded; an additional five items were uninformative under the multi-state coding retaining all polymorphisms.³ Most of the results reported below are therefore based on a matrix of 94 informative characters coded for 24 languages, or 2256 data points. Of these, 41 (1.8%) were missing and coded using the character ‘?’ (cf. Atkinson and Gray 2006). Polymorphic characters were permitted in the dataset, but additional analyses using alternative codings excluding them were performed as well (cf. Nakhleh et al. 2005 for discussion); some of the results obtained using these alternative codings are reported in §4.

Multi-state characters, where each character can have more than two states, can be transformed into binary-valued characters according to a method introduced by Gray and Atkinson (2003) for Indo-European. The basic procedure is to consider each state in the multi-state coding – i.e., each cognate set associated with a given meaning in a wordlist – as a binary-valued character in its own right. For example, the item ‘blood’ in (1) is a character with five states. This single multi-state character is transformed to five binary-valued characters, with each language assigned state 0 or state 1 depending on whether it has that cognate in the relevant meaning. Hupa would have state 1 (‘present’) for the cognate set *tse:lin* in the meaning ‘blood’, and would have state 0 (‘absent’) for the other four binary-valued characters.

This recoding procedure has been criticized by Evans et al. (2006:124), who point out that one of the foundational assumptions of the models invoked in computational historical linguistic studies requires that characters evolve independently of one another. Because the innovation of a new association between a meaning and a form typically involves replacement of an existing association, binary recoding of multi-state characters introduces many dependencies among characters: if a language has state 1 for a given cognate in a particular meaning, it will usually have state 0 for all other cognates associated with that meaning. For example, the innovation of *tse:lin* for the meaning ‘blood’ in California Athabaskan in all cases went hand-in-hand with the loss of the reflex of *dəl in that

² Hoijer’s (1956) original cognacy judgments were sometimes modified, especially for the PCA languages, but sometimes for other languages in the family in light of the cognate sets published in Krauss and Leer (1981) and Krauss (2005) and Leer’s (2011) unpublished comparative lexicon. Cognacy judgments for Minto, Tlingit, and Eyak were taken from Krauss (1963). For Northern languages not included in Hoijer (1956), some of the cognacy judgments were found in the original sources, whereas others were my own.

³ Uninformative characters have the same state for all languages. Some of them became informative when polymorphisms were removed, so under some codings there were as many as 96 informative lexical items in the dataset.

meaning. The fact of having *tse:lin* in this meaning thus makes it highly likely (if not inevitable) that a language will not have a reflex of *dɔ̄l. Evans et al. (2006) argue that these changes are not independent of one another, leading to “extreme violations of the independence assumption” underlying the evolutionary models, making this binary recoding “patently inappropriate.”

Atkinson and Gray (2006) respond to this critique by questioning whether a language having multiple cognates in a given meaning is really atypical, pointing to the existence of polymorphic characters that imply that the innovation of a cognate in a given meaning in some cases does not entail the loss of a cognate in that meaning. They maintain that one of the advantages of their binary coding scheme is how easily it handles such polymorphisms. Atkinson and Gray (2006) follow Pagel and Meade’s (2006) arguments that non-independence introduced by binary recoding will have two effects: it will tend to shorten the branch lengths of the tree returned by the analysis and inflate posterior probabilities at each node. However, Pagel and Meade predict that non-independent characters introduced by binary recoding will not affect the subgroups detected by the analysis. To test this prediction, the present study created two binary recodings of the original multi-state data matrix, one in which all states were retained, and another in which all unique states were discarded (Atkinson and Gray 2006:104). Results of analyses run with these alternative codings are reported in §4.

For some of the coding configurations (polymorphisms retained vs. excluded, multi-state vs. binary coding), separate analyses were run using three different evolutionary models: A non-clock model placing no constraints on the rates of evolution across lineages, a strict clock model that assumes all lineages evolve at a constant rate (analogous to the so-called “glotto-clock” of the glottochronology of yesteryear – cf. Hoijer 1956), and a relaxed clock model constraining rates of evolution across lineages but allowing for variation. Strict and relaxed clock models produce rooted trees, whereas non-clock models do not and must be rooted on an outgroup, Tlingit in the present study.

One of the main pitfalls in Bayesian analysis is that there are no criteria guaranteeing that the universe of possible trees has been adequately sampled – that is, how many generations are required to obtain a statistically reliable result, when an analysis is said to have “converged” (Nichols and Warnow 2008). One heuristic diagnostic for convergence is to run simultaneous analyses and monitor the average standard deviation of split frequencies, which is generated on the fly by MrBayes. For the present study, two simultaneous runs of between 1 million and 10 million generations were conducted until this convergence diagnostic fell below the value .01. Post hoc analysis of two additional convergence diagnostics – the Potential Scale Reduction Factor (PSRF) and Effective Sample Size (ESS) of model parameters – was conducted. PSRF values for each parameter should approach 1 as an analysis converges, and following Bower and Atkinson (2012) an ESS greater than 2000 was sought. For relaxed clock models, ESS > 1000 was

ultimately accepted due to the large number of model parameters and the number of generations required to bring each analysis above the more stringent threshold.

Results reported in §4 are given in the form of majority-rule consensus trees, which show nodes occurring in 50% or more of trees in the sample after the first 25% are discarded as burn-in. Scores at each node indicate the percentage of trees the node occurs in, which can be taken as an approximation of the posterior probability of the node (Pagel and Meade 2006). Particular nodes will be said to have, for example “90% support”, which can be interpreted to mean that they occur in 90% of the trees in the sample or, equivalently, that they have a posterior probability of .9. As Bowerman and Atkinson (2012:829) point out, there are no well-established criteria for considering a node well-supported versus merely adequately supported. Their standards are adopted here: a node with 80% support (i.e., posterior probability greater than .8) is considered well-supported.

It is important to emphasize that the main question of interest in the ensuing discussion is whether or not the Pacific Coast Athabaskan languages emerge as a well-supported subgroup of the Athabaskan family. The tree topologies and branch lengths presented here raise a host of other questions of potential interest within the family. Some of them relate to the grouping of languages within PCA itself, others to how PCA is related to other languages in the family, and still others to how non-PCA languages are related to one another. Some of these additional points of interest will be noted in passing, but in general no special effort will be made to explain them.

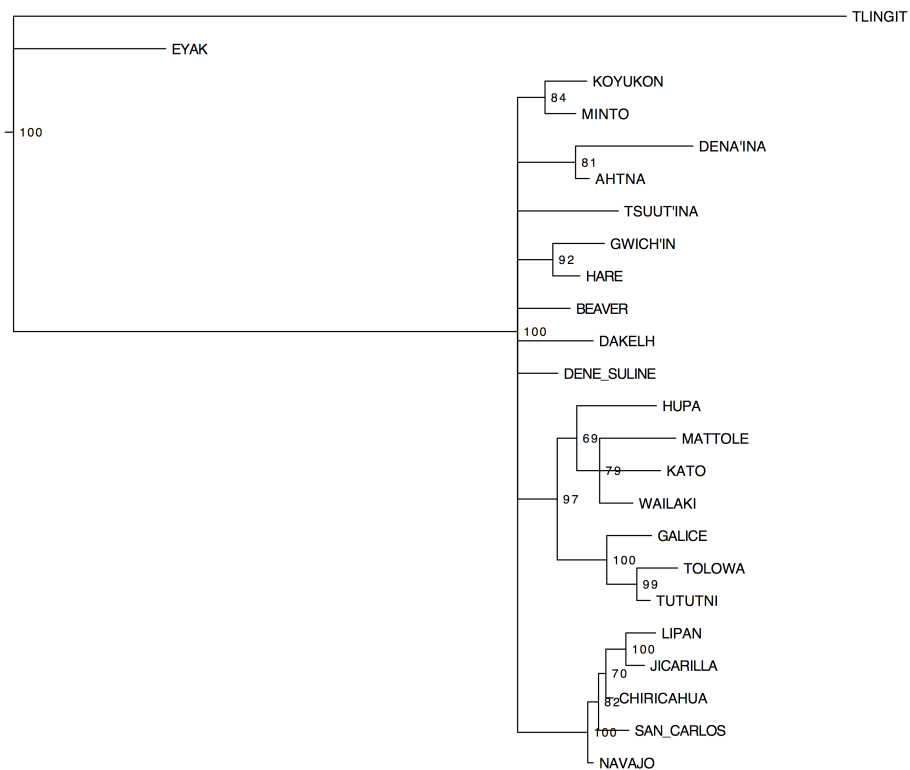
4 Results

The figure in (2) shows the majority-rule consensus tree obtained from the 94 informative lexical characters in the dataset with all polymorphisms included, run with a non-clock model of evolution. This is the least complicated analysis insofar as it has no manipulations to the underlying multi-state coding and the smallest number of model parameters in MrBayes. The main point of interest in (2) is that a Pacific Coast Athabaskan subgroup is detected with a high degree of support (97%). The Apachean languages also emerge as a subgroup with strong support; any other result would be shocking, since these languages are so similar to one another (Hojer 1956). The Northern languages show a highly unresolved, rake-like branching – Gwich'in and Hare, Koyukon and Minto, and Dena'ina and Ahtna appear as clades with over 80% support, but the rest of the Northern group does not resolve into a branching structure. This is consistent with the view of Krauss (1973) and Krauss and Golla (1981) that a highly structured family tree is an inappropriate model of Northern Athabaskan linguistic relationships.

The consensus tree in (2) can be taken as support for the view that the Pacific Coast Athabaskan languages are a well-defined subgroup within Athabaskan. Within PCA, the Oregon Athabaskan languages emerge as a subgroup with strong

support (100%). A further subgroup including Tututni and Tolowa to the exclusion of Galice also receives strong support. Interestingly, there is only moderate support for a California Athabaskan subgroup comprised of Hupa, Mattole, Kato, and Wailaki (69%); the Mattole, Kato, and Wailaki subgroup falls just below the threshold for strong support at 79%. This is consistent with Hoijer's (1962) observation that a California Athabaskan subgroup is not well-supported by lexical data.⁴

- (2) Majority-rule consensus tree produced with a multi-state coding and all polymorphisms included (non-clock evolutionary model)



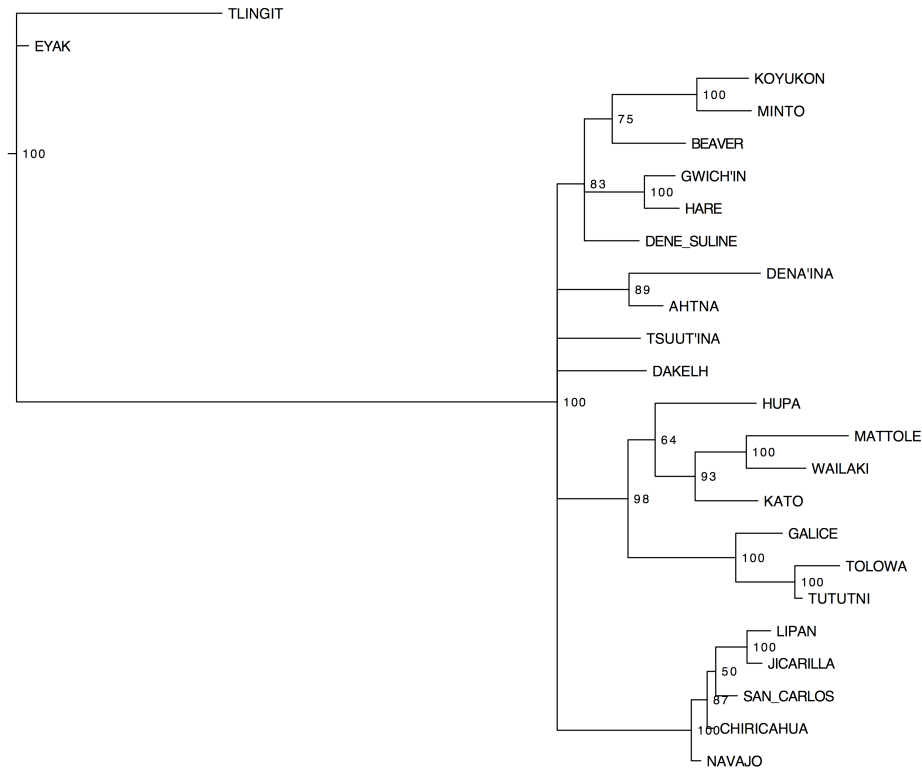
Analysis of the data was also conducted using a binary recoding. This was done in two ways: one including all character states, and another according to which all unique character states were removed, which should reduce the impact of non-independence introduced in the recoding (Atkinson and Gray 2006:104). Results from the two conditions were similar to each other; the consensus tree from the

⁴ The dataset recoded to exclude polymorphisms produced a consensus tree with minor differences from the figure in (2). The PCA subgroup had slightly lower (93%) support, but the topologies were identical except for some minor restructuring in California Athabaskan and the emergence of a subgroup including PCA and Apachean languages (with only 60-61% support).

The Status of Pacific Coast Athabaskan

latter recoding is shown in (3), where once again PCA emerges as a well-supported (98%) subgroup:⁵

- (3) Majority-rule consensus tree produced with a binary coding and unique character states removed (non-clock evolutionary model)

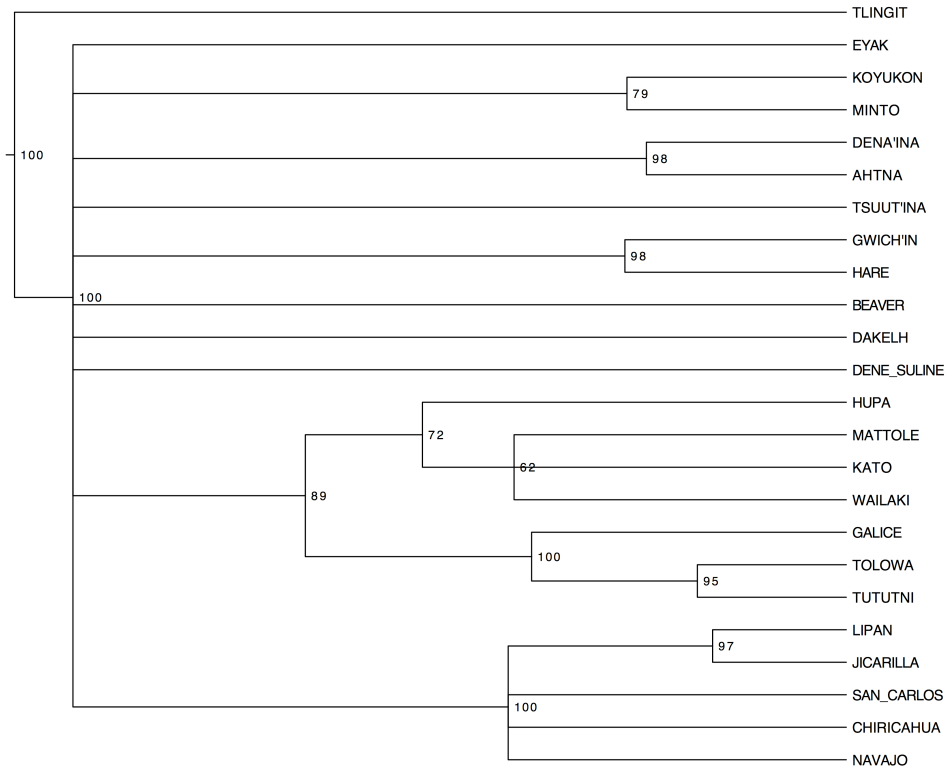


As discussed in §3, the binary recoding procedure introduces dependencies among characters. According to Pagel and Meade (2006), this will produce a scaled version of the best topology, shortening branch lengths and increasing posterior probabilities without altering the subgroups themselves. Comparison of the figures in (2) and (3) suggests that, contrary to Pagel and Meade's predictions, binary recoding had a non-trivial impact on the topology returned by the analysis. For example, within PCA, a well-supported split between Kato and Wailaki + Mattole emerges in California Athabaskan. More strikingly, many of the Northern Athabaskan languages are subsumed under a well-supported (83%) subgroup, one that also contains a subgroup with moderate support (75%) consisting of Beaver, Koyukon, and Minto. And the California Athabaskan node has slightly weaker support (64%) under the binary recoding than it had with the multi-state coding.

⁵ The shortening of branch lengths for Tlingit and Eyak in (3) relative to (2) is due to the exclusion of unique character states, many of which are found in those two languages.

Analysis of the dataset with all polymorphisms retained was also conducted using relaxed and strict clock models of evolution. The resulting consensus tree for the relaxed clock analysis is shown in (4):⁶

- (4) Majority-rule consensus tree produced with a multi-state coding and all polymorphisms included (relaxed clock evolutionary model)



The subgroups in (4) are similar to those in (2). Notably, PCA again emerges as a subgroup, one with weaker support (89%) than under the non-clock model but still comfortably above the 80% threshold to be considered well-supported. Northern Athabaskan has the same rake-like branching as found in (2), but the model does not produce the high-level Eyak-Athabaskan split, and there is much

⁶ The strict clock model produced much shorter relative branch lengths than the relaxed clock model, implying a later split within the family. There were differences in the tree topologies as well: the strict clock finds a clade with 67% support that includes the Northern and Apachean languages to the exclusion of PCA. Bayes factors analysis suggests that in general the relaxed clock model performed better than the strict clock models on each coding, $BF > 5$, in a range that is considered “substantial” but not “strong” support for one model over another (Bowerman and Atkinson 2012:830). The relaxed clock models generally performed slightly better than non-clock models but with $BF < 5$.

less resolution within Apachean. Moreover, the branch lengths within Athabaskan in (4) are much longer relative to the root, suggesting a much earlier split for Athabaskan than is implied in (2).

5 Discussion

The most pertinent result for the main research question considered here is that the Pacific Coast Athabaskan languages emerge as a well-supported subgroup of Athabaskan in all conditions: with non-clock, relaxed clock, or strict clock models; with polymorphic characters included or excluded under a multi-state coding; with unique character states included or excluded in a binary recoding. All of this can be taken as evidence in favor of a PCA subgroup within Athabaskan, as argued by Hoijer (1960). Note that the question of the subgrouping of the PCA languages is orthogonally related to the migration of Athabaskan-speaking people into California and Oregon. The results reported here do not speak to the locus of differentiation of PCA, whether it was *in situ* after a single Athabaskan-speaking group migrated into the California-Oregon region, or in some other place prior to migration. Golla (2011:69) argues that a lack of shared lexical innovations indicative of a common migration into a new biological and cultural region suggest that the latter scenario is correct. This is surprising, since *a priori* a theory of differentiation *in situ* is preferable: it requires positing only a single migration event and thus provides the simplest explanation for the concentration of a closely-related subset of the Athabaskan languages in a locale far removed from the rest of the family. Multiple migration events subsequent to the diversification of PCA are not a historical impossibility, however.⁷

Another result worth noting is that multi-state versus binary codings of the data produced different tree topologies, contrary to the expectations of Pagel and Meade (2006). This may be due to the inflationary effect of non-independence on posterior probabilities, with some nodes achieving the 50% threshold required to be included in the consensus tree in (3).⁸ Alternatively, differences in topologies in (2) and (3) might be due to a disproportionate influence of characters with large numbers of states. Binary recoding of characters with many states will introduce more dependencies than will binary recoding of characters with fewer states. Such uneven distribution of dependencies in the dataset may cause characters with more states in the multi-state coding to contribute disproportionately to likelihood calculations in a binary recoding. A similar point is made by Pagel and Meade

⁷ Golla's (2011:69, 257) suggestion of a "partially shared" history of migration that occurred in two or more "pulses" may provide the key to understanding the dynamics of Athabaskan migration from the north.

⁸ Note, however, that support for some nodes in (3) (with binary recoding) is actually lower than what is found in (2), e.g., the group subsuming Hupa and the other California Athabaskan languages.

with respect to the effect of binary recoding on branch lengths, and perhaps is implied for their argument concerning topology as well. Nonetheless, the upshot is that one cannot assume that the best topology obtained under a binary recoding will simply be a scaled version of the topology obtained under the corresponding multi-state coding.

Acknowledgments

I would like to thank audiences at BLS 39, the 2013 meeting of the Society for the Study of Indigenous Languages of the Americas, and UC Berkeley's *dhworom group for their comments on this research. Special thanks are due to Will Chang for his extraordinarily clear and patient explanations of Bayesian methodology. Figures were produced with Andrew Rambaut's FigTree software (v. 1.4.0).

References

- Atkinson, Quentin and Russell D. Gray. 2006. How Old is the Indo-European Language Family? Illumination or More Moths to the Flame? In Forster and Renfrew (2006), 91-110.
- Bommelyn, Loren, ed. 1989. *Xus we-yo: Tolowa (Tututni) Language Dictionary*. Crescent City, CA: Tolowa Language Committee.
- Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, and Quentin D. Atkinson. 2012. Mapping the Origins and Expansion of the Indo-European Language Family. *Science* 337:957-960.
- Bowern, Claire and Quentin Atkinson. 2012. Computational Phylogenetics and the Internal Structure of Pama-Nyungan. *Language* 88(4):817-845.
- Bright, Jane O. 1964. The Phonology of Smith River Athapaskan (Tolowa). *International Journal of American Linguistics* 30(2):101-107.
- Bryant, David. 2006. Radiation and Network Breaking in Polynesian Linguistics. In Forster and Renfrew (2006), 111-118.
- Collins, James. 1989. Nasalization, Lengthening, and Phonological Rhyme in Tolowa. *International Journal of American Linguistics* 55(3):326-340.
- Dorsey, J. Owen. 1884. *Tutu Vocabulary and Grammatical Notes August 20-?, 1884*. National Anthropological Archives: NAA MS 4800 (4.1.7) [380].
- Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesnik, and Angela Terrill. 2008. Structural Phylogeny in Historical Linguistics: Methodological Explorations Applied in Island Melanesia. *Language* 84(4):710-759.
- Evans, Steven N., Donald Ringe, and Tandy Warnow. 2006. Inference of Divergence Times as a Statistical Inverse Problem. In Forster and Renfrew (2006), 119-130.
- Goddard, Pliny Earle. 1902-1903. Tolowa Field Notes. American Philosophical

The Status of Pacific Coast Athabaskan

- Society MS 30 (Na20f.1).
- Goddard, Pliny Earle. 1923. Wailaki Texts. *International Journal of American Linguistics* 2(3/4):77-135.
- Golla, Victor. 1996. *Hupa Language Dictionary*. 2nd. ed. Hoopa, CA: Hoopa Valley Tribal Council.
- Golla, Victor. 2008. *Lower Rogue River (Tututni) Lexicon*. Manuscript, Humboldt State University.
- Golla, Victor. 2011. *California Indian Languages*. Berkeley, CA: University of California Press.
- Gray, Russell D. and Quentin Atkinson. 2003. Language-Tree Divergence Times Support the Anatolian Theory of Indo-European Origins. *Nature* 426:435-439.
- Forster, Peter and Colin Renfrew, eds. 2006. *Phylogenetic Methods and the Prehistory of Languages*. Cambridge: McDonald Institute for Archaeological Research.
- Harrington, John Peabody. 1982-1990. *The Papers of John Peabody Harrington in the Smithsonian Institution, 1907-1957*. Microfilm edition, part 1: Native American History, Language, and Culture of Alaska/Northwest Coast. Oregon Athabaskan, reels 25-28. National Anthropological Archives.
- Hoijer, Harry. 1956. The Chronology of the Athapaskan Languages. *International Journal of American Linguistics* 22(4):219-232.
- Hoijer, Harry. 1960. Athapaskan Languages of the Pacific Coast. In *Culture in History*, edited by S. Diamond, 960-77. New York: Columbia University Press.
- Hoijer, Harry. 1962. Linguistic Subgroupings by Glottochronology and the Comparative Method: The Athapaskan Languages. *Lingua* 11:192-198.
- Hoijer, Harry. 1973. Galice Noun and Verb Stems. *Linguistics* 104:49-73.
- Holden, Clare and Russell D. Gray. 2006. Rapid Radiation, Borrowing and Dialect Continua in the Bantu Languages. In Forster and Renfrew (2006), 19-32.
- Jetté, Jules and Eliza Jones. 2000. *Koyukon Athabaskan Dictionary*. Fairbanks, AK: Alaska Native Language Center.
- Kari, James. 1990. *Ahtna Athabaskan Dictionary*. Fairbanks, AK: Alaska Native Language Center.
- Kari, James. 1994. *Lower Tanana Athabaskan Dictionary*. Online resource: <http://www.uaf.edu/anla/collections/dictionaries>.
- Kari, James. 2007. *Dena'ina Topical Dictionary*. Fairbanks, AK: Alaska Native Language Center.
- Krauss, Michael. 1963. [Notes on Athabaskan Glottochronology.] Alaska Native Language Archive manuscript CA961K1963c.
- Krauss, Michael. 1973. Na-Dene. In *Linguistics in North America*, edited by Thomas A. Sebeok, 903-978. (Current Trends in Linguistics 10). The Hague: Mouton.
- Krauss, Michael. 2005. Athabaskan Tone. In *Athabaskan Prosody*, edited by S. Hargus and K. Rice, 55-136. Philadelphia: John Benjamins.

Justin Spence

- Krauss, Michael, and Victor Golla. 1981. Northern Athapaskan Languages. In *Handbook of North American Indians*, Vol. 6, *Subarctic*, edited by J. Helm, 67-85. Washington, D.C.: Smithsonian Institution.
- Krauss, Michael, and Jeff Leer. 1981. Athabaskan, Eyak, and Tlingit Sonorants. *Alaska Native Language Center Research Papers* 5. Fairbanks, AK: Alaska Native Language Center.
- Landar, Herbert. 1977. Three Rogue River Athapaskan Vocabularies. *International Journal of American Linguistics* 43(4): 289-301.
- Leer, Jeff. 2011. [Comparative Athabaskan Lexicon.] Manuscript, University of Alaska, Fairbanks.
- Li, Fang-Kuei. n.d. [Wailaki Noun and Verb Stem List.] Photocopy of original manuscript in the possession of Victor Golla, Humboldt State University.
- Marten, Lutz. 2006. Bantu Classification, Bantu Trees, and Phylogenetic Methods. In Forster and Renfrew (2006), 43-56.
- Merriam, C. Hart. 1850-1974. C. Hart Merriam Papers, BANC FILM 1022, Bancroft Library, University of California, Berkeley.
- Nakhleh, Luay, Tandy Warnow, Don Ringe, and Steven N. Evans. 2005. A Comparison of Phylogenetic Reconstruction Methods on an Indo-European Dataset. *Transactions of the Philological Society* 103(2):171-192.
- Nichols, Johanna and Tandy Warnow. 2008. Tutorial on Computational Linguistic Phylogeny. *Language and Linguistics Compass* 2(5):760-820.
- Pagel, Mark and Andrew Meade. 2006. Estimating Rates of Lexical Replacement on Phylogenetic Trees of Languages. In Forster and Renfrew (2006), 173-182.
- Ronquist, F., M. Teslenko, P. van der Mark, D. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M.A. Suchard, and J.P. Huelsenbeck. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space. *Systematic Biology* 61(3):539-542.
- Spence, Justin. 2013. *Language Change, Contact, and Koinéization in Pacific Coast Athabaskan*. Ph.D. thesis, University of California, Berkeley.
- Tuttle, Siri. n.d. [Notes on the Galice Creek Dialect of Oregon Athabaskan from the Melville Jacobs Papers.] Manuscript, University of Alaska, Fairbanks.
- Waterman, Thomas T. n.d. The Athapaskan Indians of Southwestern Oregon and Northwestern California. National Anthropological Archives: NAA MS 3183.

Justin Spence
Native American Studies Department
University of California, Davis
2401 Hart Hall
One Shields Avenue
Davis, CA 95616

jspence@ucdavis.edu