Journal of Language Evolution, 2023, **XX**, 1–31
https://doi.org/10.1093/jole/lzad004
Advance access publication 2 June 2023
Research article

OXFORD

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Subgrouping in a 'dialect continuum': A Bayesian phylogenetic analysis of the Mixtecan language family

Sandra Auderset,[*,1,2] Simon J. Greenhill,[1,3] Christian T. DiCanio,[4] and Eric W. Campbell[2]

[1]Max Planck Institute for Evolutionary Anthropology, Department of Linguistic and Cultural Evolution, Deutscher Platz 6, 04103 Leipzig, Saxony, Germany
[2]University of California Santa Barbara, Department of Linguistics, South Hall 3432, Santa Barbara, CA 93106-3100, USA
[3]The University of Auckland, Faculty of Science, School of Biological Sciences, 3A Symonds Street, Auckland 1010, New Zealand
[4]State University of New York at Buffalo, Department of Linguistics, 609 Baldy Hall, North Campus, Buffalo, NY 14260-6420, USA
[*]Corresponding author: sandra_auderset@eva.mpg.de or auderset@ucsb.edu

Subgrouping language varieties within dialect continua poses challenges for the application of the comparative method of historical linguistics, and similar claims have been made for the use of Bayesian phylogenetic methods. In this article, we present the first Bayesian phylogenetic analysis of the Mixtecan language family of southern Mexico and show that the method produces valuable results and new insights with respect to subgrouping beyond what the comparative method and dialect geography have provided. Our findings reveal potential new subgroups that should be further investigated. We show that some unexpected groupings raise important questions for phylogenetics and historical linguistics about the effects of different methods of primary data gathering and organization that should be considered when interpreting subgrouping results.

Keywords: Mixtec; phylogenetics; dialect continuum; subgrouping; Mixtecan.

1. Introduction

The comparative method of historical linguistics is still seen as the most reliable method for uncovering genealogical language relationships. It works well in many cases when languages are expanding and diversifying, but is less successful in recovering historical relationships in dialect continua. In dialect continua, speech communities do not neatly separate from each other, but rather remain in ongoing contact. This leads to differential diffusion between varieties, rendering a simple 'tree-like' model of their history inappropriate at best (Ross 1996) or invalid at worst (Kalyan and François 2018). It has thus sometimes been suggested that the comparative method of historical linguistics or Bayesian phylogenetic methods cannot be applied to dialect continua and dialect chains, given that the varieties remain in contact and there would be too much noise from internal loans (Ross 1996). This had led the critics of the simple tree model to the adoption of alternative approaches, such as the 'wave model' (proposed by Johannes Schmidt, cf. Anttila 1989), 'network models' (Willems *et al.* 2016), or 'historical glottometry' (Kalyan and François 2018). Defenders of the tree model point out that the impact and prevalence of contact on genealogical relationships are often overestimated and pose less of a problem for comparative reconstruction than is generally assumed. Apart from very specific circumstances, extensive language contact does not preclude an investigation into the languages' genealogical relationships (Bowern 2013: 426–7). However, a more fruitful way forward is to ask when and where language relationships are more like trees or more like waves and to use approaches that can quantify and incorporate both patterns. Our task is thus not to define the complete history of a language family as pertaining to one or the other, but rather to identify which parts of its history can be described as tree-like and where we find a conflicting signal that can be described as wave-like. Recently developed Bayesian phylogenetic methods provide such a way forward. Rather than providing a single tree estimate of the history of a language family, they provide a posterior

© The Author(s) 2023. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

probability distribution of many trees. Each tree in the posterior is a particular estimate of underlying history, and by aggregating over the trees we can identify which groupings are well supported by the data and which are noisier and show more conflict. These methods are also robust to even high levels of borrowing (Greenhill *et al.* 2009).

The Mixtecan language family of Mexico provides an excellent case study for the application of Bayesian phylogenetics for several reasons. First, there have been conflicting opinions about which languages belong to Mixtecan in the first place (see the debate between Swadesh 1960 and Longacre 1961, summarized in Section 2). Second, the highly diversified Mixtec subgroup within Mixtecan is often characterized as a dialect continuum consisting of chains of mutually intelligible varieties (Jiménez Moreno 1962; Josserand 1983: 457–8). Third, documentation of Mixtec varieties has significantly increased in recent years (see McKendry 2013; Becerra Roldán 2015; Mendoza Ruíz 2016; Hernández Mendoza 2017; Vázquez 2017; Rueda Chaves 2019; among others), making it possible to re-evaluate earlier proposals and apply phylogenetic methods to a large and updated Mixtecan data set. Fourth, and finally, doubts about the possibility of successful subgrouping within Mixtec proper have been explicitly expressed in seminal work by Bradley and Josserand (1982: 303). Like many other scholars, they believed that dialect areas like those characterizing the Mixtec branch are not subject to the same processes of language change as 'traditional' language families:

> "El problema es la aparente violación o infracción de las expectativas ideales del método comparativo en la linguística histórica (…). Pero este modelo presupone una efectiva separación de los grupos, después de un cambio que los divide. En contraste, en la Mixteca (…) vemos una situación más dinámica en el desarrollo de las familias lingüísticas." (Bradley and Josserand 1982: 303)[1]

After much more detailed groundwork, Josserand (1983: 461–2) expressed optimism for finer subgrouping among the Mixtec languages using the comparative method, but such a task has not yet been taken on. In this article, we show that the application of Bayesian phylogenetics to Mixtecan produces valuable results and new insights with respect to subgrouping beyond what the comparative method and dialect geography can provide. The Mixtecan language family, classified by most as belonging to the Otomanguean stock of Mesoamerica (Rensch 1976; L. Campbell 1997; E. W. Campbell 2017a; Kaufman 2006), spreads over large parts of the southern Mexican state of Oaxaca and into Puebla and Guerrero. Understanding the dispersal of these languages over time is important for the prehistory of the region and for Mesoamerica as a whole. Previous research on Mesoamerican linguistic and cultural history is heavily focused on Mayan and Nahua peoples and their languages. Mixtecan peoples have received comparatively little attention, even though they once inhabited a comparatively densely populated area in postclassic Mesoamerica (Terraciano 2001: 2) and an even larger territory than today (Joyce 2011; Pérez Rodríguez 2013). In historical linguistics, the situation is not as dire, but longstanding preconceptions about the language family have influenced research on its reconstruction and dispersal and the internal structure and development of Mixtecan are still poorly understood. In a small-scale study on the relatedness of seven Mixtec varieties spoken in the Juxtlahuaca district, Padgett (2017) found that current groupings in Ethnologue (Lewis *et al.* 2015) and INALI (INALI 2009a) do not align with speakers' reports of mutual intelligibility. This has dire consequences when community materials are created on the basis of 'established' classifications like Ethnologue and Glottolog (Hammarström *et al.* 2021), which suffer from similar issues, and then are distributed to communities who cannot understand them. Such issues can be traced back to the characterization as a dialect continuum and the comparative method and subgrouping might not be applicable here (Bradley and Josserand 1982: 303).

Bayesian phylogenetics has been applied to a wide array of language families (cf. Greenhill and Gray 2009; Lee and Hasegawa 2011; Bouckaert *et al.* 2018; Kolipakam *et al.* 2018, among others), but the only family studied with these methods (partially) located in Mesoamerica is Uto-Aztecan (Dunn *et al.* 2011; Greenhill *et al.* 2023).

2. Previous classifications and dating estimates

Longacre (1957, 1961) includes Triqui, Cuicatec, and Mixtec in the Mixtecan group of Otomanguean, with Amuzgo in a position coordinated with Mixtecan. Swadesh (1960) places Trique outside of Mixtecan and includes Amuzgo within it. In later work, Longacre (1966) reasserts that Amuzgo is not Mixtecan and expresses doubts about it even being Mixtecan's closest relation within Otomanguean. Rensch (1976) adopts Longacre's revised proposal, but more recent work by Kaufman (2006) agrees with Longacre's original proposal, classifying Amuzgo as external to, but coordinate with, Mixtecan (see also L. Campbell 1997; E. W. Campbell 2017b). Following the most recent consensus, with Amuzgo as a sister to Mixtecan, we exclude Amuzgo from our analysis.

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Proto-Mixtecan forms have been reconstructed by various scholars (Longacre 1957, 1962; Gudschinsky 1959; Arana Osnaya 1960; Swadesh 1967), and these must be considered through the lens of which classification the authors proposed or supported at the time. Throughout the body of work on Mixtecan comparative studies, the internal subgrouping of the family remains poorly understood. Further disagreements pertain to whether Mixtec and Cuicatec form a branch vis-a-vis Triqui as Belmar (1902: 4) suggests, or whether all three are coordinated with one another, cf. Figure 1. Macaulay (1996: 6) states that the Mixtec-Cuicatec group has been convincingly demonstrated by Kaufman (1983, 1988), but both of these unpublished studies lack sufficient primary data for assessing the claim. The latter proposal of a flat structure is based on negative evidence, i.e., in the absence of convincing evidence for grouping Mixtec and Cuicatec closer together (Josserand 1983: 101). In the following, we summarize the proposals of internal grouping for each of the three branches.

Triqui consists of three distinct varieties (Hollenbach 1977; DiCanio 2008; Matsukawa 2008) centered around the towns of San Andrés Chicahuaxtla, San Juan Copala, and San Martín Itunyoso, all in the state of Oaxaca (Hammarström *et al.* 2021).

Cuicatec is considered to consist of a handful of mutually intelligible varieties (Anderson and Roque 1983), but unfortunately, this group remains very sparsely documented (Campbell 2017a: 8), and we will not be able to address its internal classification further.

Mixtec is by far the largest and most diversified Mixtecan subgroup and the languages are spoken in at least 189 municipalities in Oaxaca, Guerrero, and Puebla states in Mexico (Smith Stark 1994). It is characterized as a dialect continuum for which it is difficult or impossible to know how many languages and varieties there are. Jiménez Moreno (1962) considers Mixtec a 'language' with seven 'dialect complexes', but he does not provide primary data and argumentation. Hammarström *et al.* (2021) indicate that Mixtec comprises at least 53 varieties, but this appears to be based on Ethnologue, which in turn reflects the inter-intelligibility studies by Egland (1983); INALI (2009b) lists 81 varieties, with no

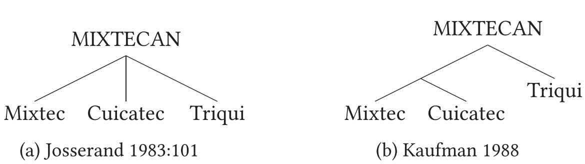

Figure 1: Two proposals for subgrouping within Mixtecan.

supporting evidence for the determination of the groups and their membership. Such confusion proliferates in much of the literature on Mixtec languages and is only exacerbated by the unfortunate practice of referring to the Mixtec group as one language.[2]

Proto-Mixtec as such has been reconstructed and refined by various scholars (Mak and Longacre 1960; Bradley and Josserand 1982; Swanton 2021; van Doesburg *et al.* 2021; among others). The most systematic Mixtec subgrouping proposal to date is Josserand's (1983) dissertation, in which she proposes 12 dialect areas (some with further subdivisions), reproduced in Figure 2. Her work is based on a word list with 188 items collected from 120 Mixtec-speaking villages and remains the most comprehensive comparative Mixtec study to date. She focuses on sound changes in vowels but mentions consonantal developments as well. The tone is excluded since at that time there was limited availability of reliable material. She established the dialect areas based on bundles of isoglosses, which reflect vowel changes. It is important to point out that Josserand's groupings do not directly translate into a family tree, nor were they intended to.

The tree provided in Glottolog (Hammarström *et al.* 2021) largely follows this outline, but also lists Amoltepec as an isolated primary daughter of proto-Mixtec. This variety is not included in Josserand's (1983) dialect areas, listed as the source of the tree in Glottolog, but it is proposed as its own group in Bradley and Josserand (1982). Lower-level groupings do not follow Josserand (1983) in most cases; we infer that they follow Ethnologue (Eberhard *et al.* 2021). The same is true for the distinction between languages and dialects, which is not made in Josserand (1983)—nor in the present study, cf. Section 3. To sum up, there are many uncertainties with regard to subgrouping among and within each of the three Mixtecan groups, especially as regards Mixtec.

Some of the earlier works mentioned above not only probed subgrouping and reconstruction but also provided dating estimates based on lexicostatistics and glottochronology (Holland 1959; Arana Osnaya 1960; Swadesh 1967). Both of these methods have severe shortcomings. Lexicostatistics is based on superficial similarity of form, without systematic cognate identification, and fails to account for situations in which lexical borrowing is extensive (Comrie 2000), and glottochronology makes the false assumption that lexical replacement occurs at a uniform, constant rate in all languages (Hoijer 1956; Nettle 1999; Holman 2010; among many others). These estimates, however, are the only ones currently available and therefore have been integrated into archaeological and anthropological studies (cf. Byers 1967; Flannery and Marcus

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

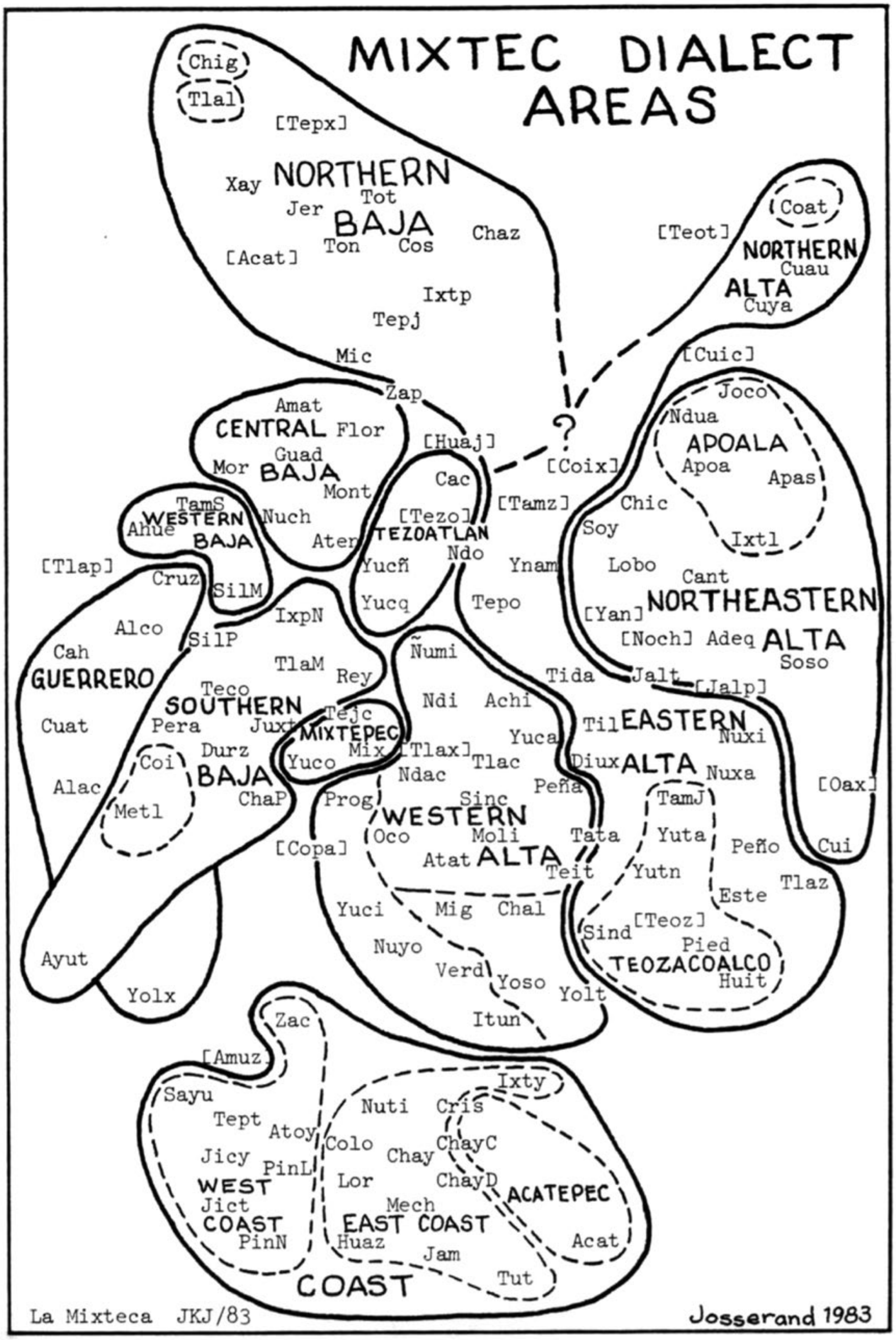

Figure 2: Mixtec dialect areas according to Josserand 1983:470.

1983; Josserand *et al.* 1984). Mixtec and Triqui are estimated to have diverged around 3900–3500 years ago, Triqui and Cuicatec around 3900 years ago, and Mixtec and Cuicatec around 3100–2500 years ago (Holland 1959: 25–6, Arana Osnaya 1960: 262–3, Swadesh 1967: 94).

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

3. Data

3.1 Languages

We included every Triqui, Cuicatec, and Mixtec variety in the sample for which there is enough documentation and the materials are available to us. We do not distinguish between languages and dialects, since there is no solid basis nor necessary or sufficient criteria to do so in Mixtecan. In this study, we use the neutral term ‘variety’ throughout, but refrain from using the term ‘dialect’ since it carries a negative connotation in Mexico and has been part of a long history of oppression of communities that speak Mixtecan languages (Cruz and Woodbury 2014).

We collected data from four Triqui varieties, three Cuicatec varieties, and 130 Mixtec varieties, i.e., 137 varieties in total. The Triqui varieties include the modern varieties mentioned in Section 2—San Martín Itunyoso, San Juan Copala, and San Andrés Chicahuaxtla—and the historical account of Chicahuaxtla Triqui spoken at the end of the 19th century (Belmar 1897). For Cuicatec, we include the closely related varieties spoken around the town of Santa María Pápalo, and the historical account of Tepeuxila Cuicatec spoken around the year 1900 (Belmar 1902). The Mixtec varieties are the 120 sampled in Josserand (1983), excluding proto-Mixtec. We added ten varieties based on our own or colleagues’ documentation work and recently published sources. Of the varieties sampled in Josserand (1983), seventeen were partially updated and provided with tone marking in Dürr (1987) and for another ten there are new materials available. An overview of all languages and sources can be found in the Appendix in Table A.1 and a geographical overview in Figure 3.

3.2 Word list and data collection

We constructed a list of 209 concepts of basic vocabulary selected through a triangulation of already existing lists (with the number of items in brackets: Josserand 1983 [188]; Dürr 1987 [110]; Padgett 2017 [98]; Swanton and Mendoza Ruíz 2021 [86]; and Campbell unpublished [340]) and considerations about ease of elicitation (Laycock 1970) and semantics (Kassian *et al.* 2010). We kept all entries that appear in three or more lists and added a few that appear in two lists, but are easy to elicit and encode important cultural concepts. We excluded verbs because they require aspect-mood inflection, which is not well understood on a comparative Mixtecan level and is not always provided or reliably identifiable in glosses in the sources. The entire list is provided in SM 1.

We collected as many list entries as possible for each variety in the orthography provided in the source. We then converted the orthographic entries to IPA using orthography profiles (one for each source), which also standardized tone notation. The details of the conversion and standardization are laid out in SM 2. After data collection, we removed all concepts for which there was only one entry. Excluding duplicates, this results in a data set with 18,060 entries. The number of entries per variety ranges from 223 to 65, with a mean at 131 and a median at 125. We can thus say that on average we were able to gather around 60 per cent of the concepts for each variety.

3.3 Cognate coding

The cognate coding was carried out based on the comparative method and mainly done by hand. We initially applied the ‘LexStat’ algorithm for automatic cognate detection (as implemented in List and Forkel 2021 and described in detail in List *et al.* 2018), but then refined and adjusted each cognate set by hand based on already established sound changes and correspondences (cf. Section 2 references), and our own knowledge of the languages in question. As Mixtecan languages exhibit multi-morphemic words in their basic vocabulary, we identified cognate morphemes within each lexical item (i.e., ‘partial cognacy’ *sensu* List *et al.* 2017). This is exemplified with the entry for ‘scorpion’ in the Mixtec and Cuicatec varieties in Table 1, where all of the languages show an animate marker and the word for ‘tail’ for this concept. Since these are both recurrent morphemes with a clear meaning assigned to them, they each get their own cognate identifier (nine for the animate marker and 635 for the ‘tail’ morpheme). In Triqui, the entry for ‘scorpion’ is not further analyzable (and not cognate with ‘tail’) and so only has one cognate identifier assigned to it (749). The entries for ‘tail’ are coded as cognate (set 635) in Cuicatec, Peñasco Mixtec, and Alacatlatzala Mixtec, because the sound correspondences found there are systematic and recur. This is not so for the Duraznos Mixtec and Triqui entries, even though they look superficially similar. These entries are, therefore, coded as separate cognate sets.

A novel feature we introduce is the annotation of tonal derivation. Mixtecan languages have highly complex systems of lexical and grammatical tone and the languages cannot be described or analyzed without making reference to tonal phenomena. With respect to cognate assignments, we treat tonal derivation the same as segmental derivation. That is, we assign the tonal derivational morpheme its own cognate ID and represent it in the appropriate spot. The only difference to segmental derivation is that the toneme cannot be visually segmented in the same way a concatenated affix may be. In Table 1, the tonal derivation is

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

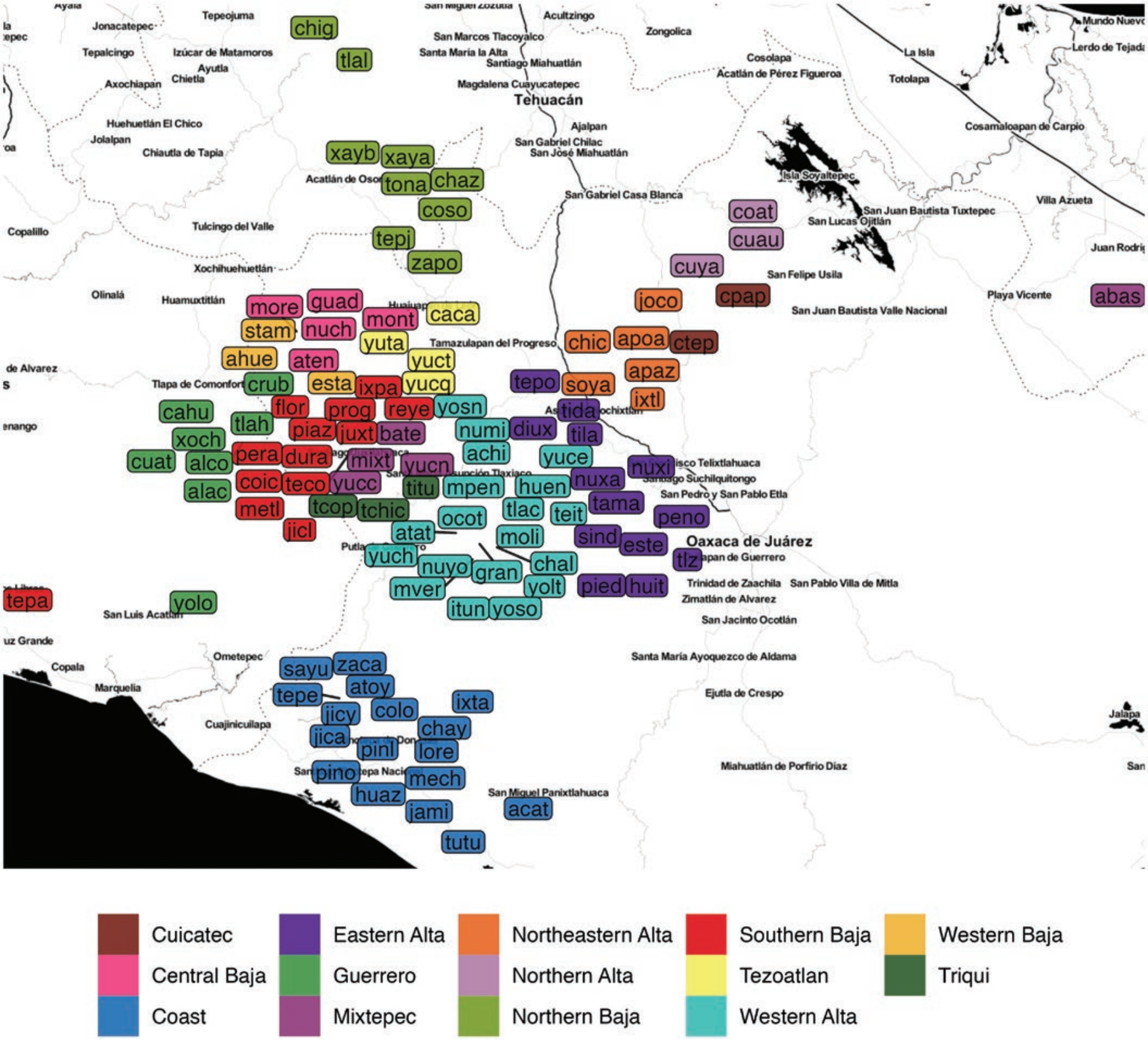

Figure 3: Sampled varieties, with colors indicating primary branches and Josserand's dialect areas of Mixtec (for label abbreviations see Supplementary Materials).

Table 1: An example of the annotation of partial cognates and tonal derivation. (Forms are given in IPA. In these examples, there is no difference between the broad and fine-grained cognate assignments.)

DOCULECT	CONCEPT	FORM	COGNATE IDs
AlacatlatzalaMixtec	TAIL	s i ^{1} ʔ m a ^{1}	635
MagdalenaPenascoMixtec	TAIL	s u ^{3} ʔ m a ^{1}	635
SanMartinDuraznosMixtec	TAIL	^{n}d o ^{3} ʔ o ^{1}	748
SantaMariaPapaloCuicatec	TAIL	^{n}d u ^{4} k u ^{1} + ð e ^{1} ʔ ẽ ^{3}	709 635
SanMartinItunyosoTriqui	TAIL	t u ^{3} n e ʔ ^{3}	749
AlacatlatzalaMixtec	SCORPION	t i ^{5} + s i ^{1} ʔ m a ^{3}	9 1044 635
MagdalenaPenascoMixtec	SCORPION	t i ^{1} + s u ^{3} ʔ m a ^{1}	9 635
SanMartinDuraznosMixtec	SCORPION	t ɕ i ^{1} + s u ^{3} ʔ m a ^{1}	9 635
SantaMariaPapaloCuicatec	SCORPION	i ^{3} t + ð e ^{3} ʔ ẽ ^{1}	9 635
SanMartinItunyosoTriqui	SCORPION	tʃ i ^{3} k ĩ ^{32}	636

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Table 2: An example of the annotation of broad vs. fine cognate IDs. (Forms are given in IPA.)

DOCULECT	CONCEPT	FORM	COG.IDs BROAD	COG.IDs FINE
SanAndresYutatioMixtec	BIRD	l a ^{3} a ^{3}	49	49L
SantaMariaPenolesMixtec	BIRD	t ɨ ^{5} + l a ^{1} a ^{5}	9 49	9 49L
SantaMariaJicaltepecMixtec	BIRD	s a ^{3} a ^{3}	49	49
SanMiguelElGrandeMixtec	BIRD	t ɨ ^{3} + s a ^{3} a ^{1}	9 49	9 49
SanAndresChicahuaxtlaTriqui	BIRD	ʃ a ^{3} + t a^{h} ^{32}	9 49	9 49
SanAndresYutatioMixtec	FROG	l a ^{5} ʔ o ^{1}	263	263L
SantaMariaPenolesMixtec	FROG	l a ^{1} ʔ β a ^{5}	263	263L
SantaMariaJicaltepecMixtec	FROG	s a ^{3} ʔ β a ^{3}	263	263
SanMiguelElGrandeMixtec	FROG	s a ^{3} ʔ β a ^{1}	263	263
SanAndresChicahuaxtlaTriqui	FROG	ʃ i ^{2} + r i ^{3} k ɨ^{h} ^{3}	9 266	9 266

exemplified by Alacatlatzala Mixtec in the word for ‘scorpion’. The second morpheme is clearly cognate with the word for ‘tail’, but in the derived word for ‘scorpion’, the second tone is raised. This tonal process is recurrent and represented with its own cognate identifier (1044). Note that in San Martín Duraznos Mixtec, we do not know whether this tonal process applied or not, since the base morpheme ‘tail’ was replaced by another (non-cognate) word. Consequently, we do not annotate it as being tonally derived.

We coded cognate sets in two ways: a broad analysis and a more fine-grained one. The fine-grained analysis takes into account potentially ‘irregular’ or ‘unexpected’ reflexes, while the broad one ignores those. The reason for these two coding schemes is that these variant reflexes could carry a phylogenetic signal that reflects shared innovations rather than parallel innovations. This could, therefore, be useful for subgrouping. To be able to address this question empirically rather than prejudging the matter, we opted for annotating the cognate sets in two ways. This allows us to run the models on both sets and assess the influence (or absence thereof) on the outcomes. The different codings are illustrated in Table 2. In both the reflexes for ‘bird’ and ‘frog’ in Table 2, some Mixtec varieties show an initial [l] while others have an [s] in this position. This correspondence surfaces in a few other items, but is not regular in the strict sense. Furthermore, it is unclear whether they represent different reflexes of the same proto-form or whether one or both of them contain remnants of fossilized prefixes. In the broad cognate coding, this difference is glossed over, while in the fine-grained one the [l]-forms are grouped into a separate class from the [s]-forms.

The assessment results in 1120 cognate sets in the broad analysis and 1197 in the fine-grained one. Details regarding cognate coding and the full database of cognate sets can be found in SM 2 and SM 3.

Table 3: Mean delta-scores and Q-residuals.

Cognate sets/Family	δ-score	Q-residual	Source
Broad	0.3911	0.02554	
Fine	0.3793	0.02472	
Polynesian	0.41	0.002	Gray *et al.* 2010
Dravidian	0.30	0.0069	Kolipakam *et al.* 2018
Chapacuran	0.262	0.016	Birchall *et al.* 2016

This cognate database was converted to a binarized cognate matrix to serve as the input for the Bayesian phylogenetic analysis. This cognate matrix has one column for each cognate set, which marks the presence (1), absence (0), or lack of information (?) of this cognate set for each variety. We excluded 27 varieties due to low coverage (marked by * in Table A.1). This coding resulted in a cognate matrix with 1183 states (columns) for the broad cognate assignment and 1254 states (columns) for the fine-grained cognate assessment. The resulting data files in Nexus format (Maddison *et al.* 1997) are provided in SM 4.

4. Methods

4.1 Tree-likeness and conflicting signal

To visualize and quantify the conflicting signal in the data set, we calculated δ-scores and Q-residuals. Both of these metrics assess how much conflicting signal there is in the data, or in other words, how tree-like the data are (Bryant and Moulton 2004). The Q-residuals and δ-scores were calculated with SplitsTree4 (Huson and Bryant 2006), which was also used to produce NeighborNets (provided in SM 5). δ-scores are computed for each tip and result in a number between 0 and 1, with higher numbers indicating more conflicting signals. Gray *et al.* (2010: 3925) argue that Q-residuals are a more direct measure of how much the tips diverge from a strict tree, because they take into account the potential effects of scaling. A lower Q-residual score reflects more adherence to a strict tree. Both scores can be averaged over all tips to give a measure of the tree-likeness of the whole network. These average scores are summarized in Table 3. The scores for each tip are provided in SM 2. The broad cognate sets result in slightly higher scores as compared to the fine-grained sets, exhibiting more conflicting signals or, in other words, a less tree-like structure. The potentially 'irregular' reflexes annotated in the fine-grained sets should thus be investigated in more detail when family-wide sound changes are worked out, since they are potentially useful for subgrouping.

Delta-scores and Q-residuals cannot be straightforwardly compared across data sets of different languages, and there are no clear cut-offs of what counts as tree-like or not (Gray *et al.* 2010: 3925–6). It is still instructive to situate our results in the context of other studies, given that the Mixtec group has been explicitly described as evolving in a non-tree-like manner (cf. Section 2). The scores of three other language families are also summarized in Table 3. Mixtecan has a δ-score almost as high as Polynesian, which is described as very reticulate, and a Q-residual score higher than all three language families. This corroborates the impressionistic views that Mixtecan languages exhibit a high degree of a conflicting signal.

The amount of reticulation, however, is not distributed evenly among Mixtecan languages. To better illustrate this, we group the languages according to Josserand's proposed dialect areas and plot the Q-residuals and δ-score against each other in Figure 4. First, we note that Cuicatec and Triqui languages show high scores—in fact higher than most of the Mixtec languages. This is not surprising with regard to Cuicatec. This branch is the least well-studied of the three, with little in-depth documentation and almost no historical work on those languages. It is thus possible that there are undetected loans from Mixtec languages in the data.

The amount of conflicting signals in Triqui languages requires another explanation. The Triqui people are a quite small group in the region and their territory is completely surrounded by and in some cases shared with Mixtec speakers. The Triqui communities have long had a high level of contact with Mixtec speakers. Historically, Triqui speakers had some degree of bilingualism with San Juan Mixtepec Mixtec (a nearby Mixtec language) and they travelled regularly to Mixtec towns (especially Cuquila and Tlaxiaco) for the purposes of commerce (see DiCanio 2022a). This could help explain their non-tree-like history.

The Mixtec groups with the lowest amount of conflicting signals are the Central and Western Baja, the overall highest scores are found in the Mixtepec group. Very high scores—or a lot of conflicting signals—are also found in a few Guerrero languages and some languages of the Western Alta.

To explore these data further, we apply a Bayesian phylogenetic model. Bayesian phylogenetic methods are well suited to explore language history based on cognate data and have several advantages over other methods. Unlike lexicostatistics, they allow for rate-of-change variation across languages, across cognates, and over time. Bayesian phylogenetic methods do not produce a single 'best' tree, but rather sample the space of possible trees to return a distribution of trees that fit the data well given the model. This posterior sample provides a natural way for calculating the support for particular groupings while allowing us to also take into account differing scenarios. This means that we can quantify uncertainty in the tree, both with respect to nodes (or splits) and with respect to model parameters. Furthermore, Bayesian phylogenetic approaches incorporate linguistic and historical knowledge into the model via priors and calibration points. It is sometimes assumed that Bayesian phylogenetic models disregard the important difference between retentions and innovations by treating them the same way, but—as opposed to lexicostatistics—these models show where traits originate versus where they remain the same (cf. Hoffmann *et al.* 2021: 12). For a more detailed introduction and explanation of Bayesian phylogenetic methods and their application in linguistics, see Greenhill and Gray 2009; Bowern and Atkinson 2012; Bowern 2018; Greenhill *et al.* 2020; Maurits *et al.* 2020; Hoffmann *et al.* 2021.

4.2 Calibration points

To infer dates, calibration points are needed. The Mixteca region, especially the Mixteca Alta, is well excavated and described in ethnohistoric and anthropological terms (cf. Bernal 1966; Flannery and Marcus 1983; Balkansky *et al.* 2000; Kowalewski *et al.* 2009; Joyce 2011; Pérez Rodríguez 2013; Spores and Balkansky 2013; among others). Despite these advances, it is still difficult to identify specific cultural groups and speech communities in the archaeological record (Pérez Rodríguez 2013: 93), especially in the early periods before people adopted a mostly sedentary lifestyle. This means that even though the archaeological record of the Mixtec people is relatively well understood, we cannot link specific artefacts or sites to specific Mixtec varieties.

The Mixtecs have chronicled their royal lineages in several codices (cf. Jansen 1990 for an overview). Those that survive to the present day are from the late

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

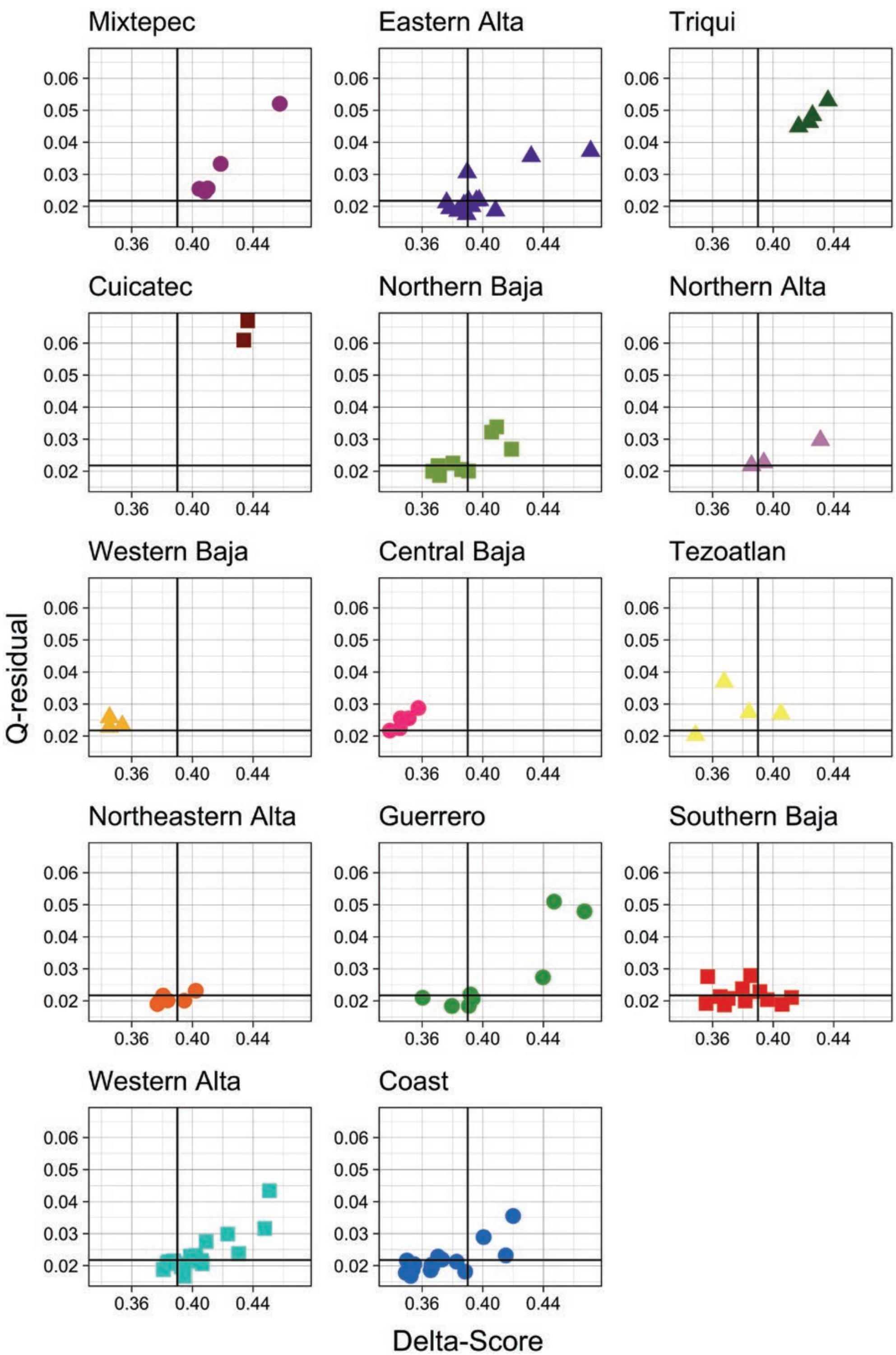

Figure 4: Delta-scores and Q-residuals plotted against each other by dialect area according to Josserand (1983).

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Table 4: Calibration points.

Language(s) affected	Type	Setting	Details
SanPedroySanPabloTeposcolula1600Mixtec	Tip date	350 BP	Age of historical document
SanJuanTepeuxila1900Cuicatec	Tip date	50 BP	Age of historical document
SanAndresChicahuaxtla1890Triqui	Tip date	60 BP	Age of historical document

post-classic and early colonial periods. These are pictorial manuscripts and, while clearly identifiable as Mixtec, we cannot tell with certainty which variety of Mixtec was spoken by their creators or by the people listed in the codices (Jansen and Pérez Jiménez 2011: 7).

This lack of granularity in the historical record leaves us with few candidates for calibration points. There are historical documents of the early and late colonial period (de los Reyes 1890 [1593]; Belmar 1897, 1902; de Alvarado 1962 [1593]) with vocabularies and grammar sketches that allow us to clearly identify the language described. The Mixtec variety of Teposcolula was used as a *lingua franca* throughout the 16th century (Jansen and Pérez Jiménez 2011: 9) and was thus documented quite extensively by the Dominicans. Given that we know when these documents were published, we can use that date as a prior setting for the Teposcolula Mixtec variety. In the late colonial period, Belmar documented a variety of Cuicatec and a variety of Triqui (Belmar 1897, 1902). These two documents also serve as calibration points, all summarized in Table 4.[3]

4.3 Model specifications and comparison

We carried out a Bayesian phylogenetic analysis with BEAST2 (Bouckaert *et al.* 2019). BEAST2 estimates the posterior distribution of trees using a Markov chain Monte Carlo. As the tree prior, we used a Birth-Death Skyline with a Serial Sample model parameterized on birth, death, and sampling rates (Stadler *et al.* 2013). This tree prior controls how the sampled trees are initially built. Under the Birth-Death Skyline model lineages are born (= 'birth') and go extinct (= 'death') at separate rates. We parameterized the birth and death rate as exponentially distributed with a mean of 0.01. This means that, on average, a new lineage is born every 100 years, while an existing lineage lives for an average of 100 years (Hoffmann *et al.* 2021).

The third parameter in this model is the sampling parameter. A language is said to be 'sampled' if it is included in the data set, but there is an unknown number of languages that existed in this family through time that fell out of use before being recorded. However, it is unlikely that Mixtec was as highly diversified as it currently is before or even at the time of the Spanish conquest, and we thus estimate that we have been able to sample more than half of the languages. We modelled this parameter following a Beta distribution with a mean of around 60 per cent (Beta [110, 80], cf. Hoffmann *et al.* 2021).

Table 5: Model comparison results sorted by Bayes factor.[4]

Cognate coding	Clock model	MlogL	logBF	SD
Broad sets	Relaxed clock	−11,325		3.72 **
Broad sets	Strict clock	−11,469	144	3.23
Fine sets	Relaxed clock	−12,253	784	3.49 **
Fine sets	Strict clock	−12,406	153	3.20

To model how the cognates change we applied a binary covarion model of cognate evolution (Penny *et al.* 2001). The covarion model estimates at which rate cognates are lost and gained over time and allow for the rates to vary, such that there can be 'slow' and 'fast' periods of changes in cognate sets. This accounts for the intuition shared by many linguists that some cognates can be relatively stable over a period of time and then change rather rapidly (e.g., due to language contact). To model rate variation over time we fitted both a strict clock and a relaxed clock model of character change (Drummond *et al.* 2006). The strict clock assumes the same rate of substitution across the whole tree, while the relaxed clock allows the rates to vary across branches.

This setup resulted in a total of four models, summarized in Table 5. The maximum clade credibility (MCC) tree was extracted with TreeAnnotator (part of the BEAST2 distribution) and the graphical representations were created with the package ggtree (Yu *et al.* 2017) in R (R Core Team 2021). The visual representation of all posterior trees was obtained with DensiTree (Bouckaert and Heled 2014). The BEAST XML files and the full MCC tree are available in SM 6 and 7.

To evaluate which model fits these data best, we used nested sampling to calculate the marginal likelihoods of each model (Maturana Russel *et al.* 2019) and Bayes factors to quantify the differing model performance (Kass and Raftery 1995). We ran each model for 100,000,000 generations, sampling every 5000th generation to avoid auto-correlation. We discarded

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Table 6: BETS (Bayesian estimation of temporal signal) results.[5]

Cognate coding	Clock model	Temp. inf.	MlogL	logBF	SD
Broad sets	Relaxed clock	+TIME	−11,325		3.72
Broad sets	Relaxed clock	−TIME	−11,311	14	3.47 **
Broad sets	Strict clock	+TIME	−11,469		3.23 **
Broad sets	Strict clock	−TIME	−11,460	9	3.08
Fine sets	Relaxed clock	+TIME	−12,253		3.49
Fine sets	Relaxed clock	−TIME	−12,248	5	3.50 * (Overlaps)
Fine sets	Strict clock	+TIME	−12,406		3.20 * (Overlaps)
Fine sets	Strict clock	−TIME	−12,409	3	3.20

the first 10 per cent as 'burn-in' where the inferred parameters were still unduly affected by the Markov chain's starting parameters. The trace files were inspected with Tracer 1.7 (Rambaut *et al.* 2018), which showed that all the models converged after burn-in and critical parameters all showed effective sample sizes above 200. Table 5 summarizes the results of the model comparison. The best-performing model, i.e., the one with the highest Bayes factor, is the one based on the broad cognate assignments with a relaxed clock. This model outperforms the strict clock model with the same data set. This is true also of the two models based on the fine-grained cognate assignments: the relaxed clock model performs better than the strict clock one. However, both models based on the fine sets have lower Bayes factors than the ones based on the broader cognate sets.

As our analysis only contained minimal calibrations we were concerned there would not be enough signal to provide a robust estimate of the age of the family. Therefore, we formally tested whether the temporal information provided sufficient signal using a Bayesian estimation of temporal signal ('BETS') analysis (Duchene *et al.* 2020)—i.e., we ran each analysis a second time without any temporal information and compared the model fit with and without temporal information. Table 6 provides the results of the BETS analysis. Overall, the temporal information has little impact on our models—as might have been expected given that all we have are three tip dates (cf. Section 4.2). In the models based on the fine sets the temporal information neither improves nor renders the analysis worse, since the Bayes factors overlap. In the model with the broad cognate sets and a strict clock, the timed analysis outperforms the untimed one, but the reverse is true for the relaxed clock model with the broad sets. However, in both cases the practical difference is marginal and the tree topology is almost the same, except that the timed analysis shows overall lower posteriors for most of the nodes. We thus base the remainder of this article on the model based on the broad cognate sets with a relaxed clock and its output.

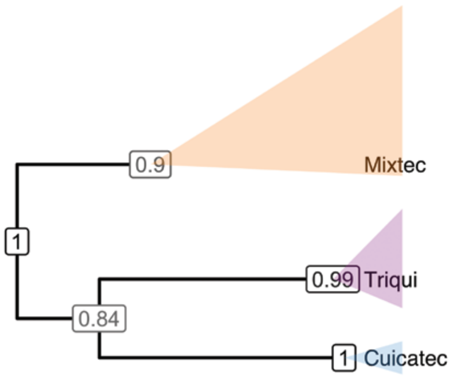

Figure 5: Primary branches of Mixtecan as recovered by our model with node posteriors.

5. Results

We discuss the results of the best-performing model (broad cognate coding with relaxed clock) with respect to subgrouping based on the MCC tree and the densitree representations. We also refer back to the NeighborNet of the broad cognates sets where this is illustrative. We consider node posteriors of the MCC tree of over 0.7 as well supported and posteriors of over 0.9 as very reliable and will primarily discuss such nodes, starting from the root of the tree moving to lower-level groupings. We refer to the groups identified in our model with numbers, using single digits for the higher-level groupings and adding digits for each lower level. The numbers are intended as neutral labels. We did not use capital letters (as is customary for example in Bantu classification) to avoid confusion with Josserand's (1983) previously established dialect areas and the geographic areas (Alta, Baja, and Costa) of the Mixteca region.

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Our model recovers the three branches of Mixtecan—Mixtec, Cuicatec, and Triqui—well, cf. Figure 5. Triqui and Cuicatec are grouped together versus Mixtec, even though we expect Mixtec and Cuicatec grouped together (cf. Section 2). This unexpected grouping was also present in the untimed model, suggesting that it is not the temporal information from the late colonial Triqui and Cuicatec documents primarily responsible for this division. As mentioned in Section 2, both Cuicatec and Triqui have been influenced by Mixtec speakers, who are and have been the majority in the area. However, there were also some shared Triqui-Cuicatec cognates which lacked Mixtec language cognate forms and are at least partially responsible for their closeness in the tree and the NeighborNet. It should be further investigated which factors led to Triqui and Cuicatec being grouped together in our model.

Since only two varieties of Cuicatec could be included, one of which is not contemporary, the internal structure of Cuicatec cannot be discussed further. Within Triqui, we do not find well-supported divisions, apart from the one that separates the colonial Chicahuaxtla variety from the contemporary ones. We thus refrain from any further interpretation of the internal groupings of Triqui.

In Mixtec, which comprises the largest number of varieties of the three branches, we discuss each well-supported group and linkage in terms of earlier proposals and implications for further research. We use the term 'linkage' to refer to low-level groups and varieties which are placed close together but do not have good support for forming a group (Ross 1988; François 2015). An overview of the higher-level groups and linkages is provided in Figure 6.

The first split in Mixtec separates the colonial era Teposcolula Mixtec from all other varieties, see Figure 7. Teposcolula was the political centre during the colonial period and the variety spoken there served as a *lingua franca* throughout the Mixteca. Josserand (1983) classified Teposcolula as belonging to the Eastern Alta group. This is not recovered in our model, but it is possible that the temporal distance influenced the placement of this variety since it is about 350 years removed from contemporary varieties. In the untimed model, Teposcolula is part of Group 1 (which corresponds to Josserand's (1983) Northern Alta), albeit not with high certainty (posterior = 0.72). In the NeighborNet, this variety is placed between varieties of Josserand's Eastern Alta and Coast groups, but closer to the Eastern Alta. All of this suggests that further research is needed to clarify the relationship of the Teposcolula variety to contemporary Mixtec varieties.

The next well-supported split sets Group 1 apart from all the other Mixtec groups and linkages, cf. Figure 8,

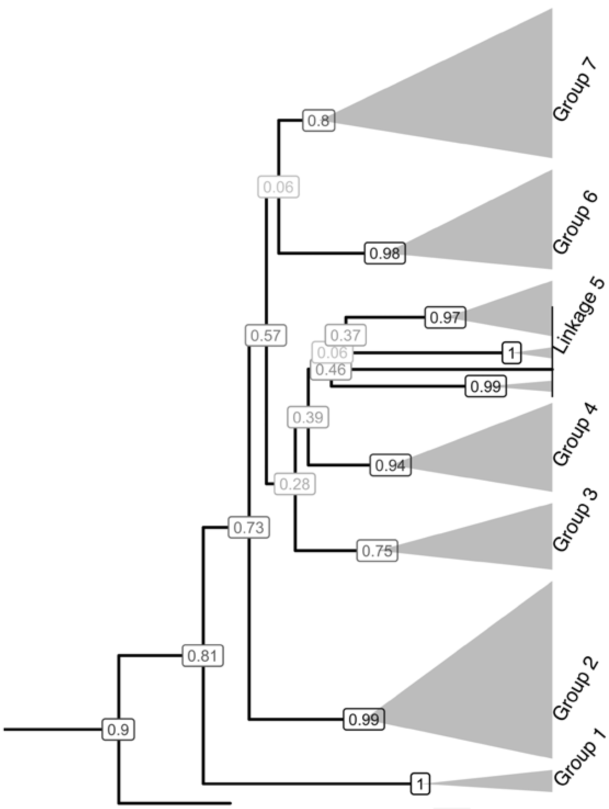

Figure 6: The Mixtec groups and linkages with node posteriors.

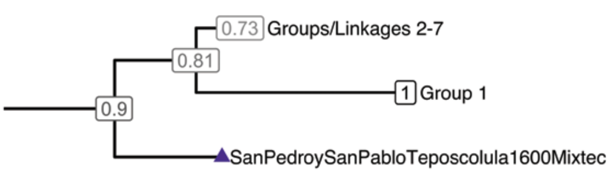

Figure 7: The first and second split within the Mixtec branch with node posteriors.

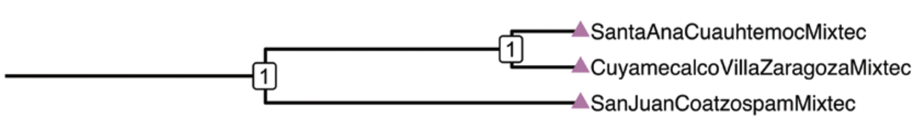

Figure 8: Mixtec Group 1 with subdivisions and node posteriors.

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

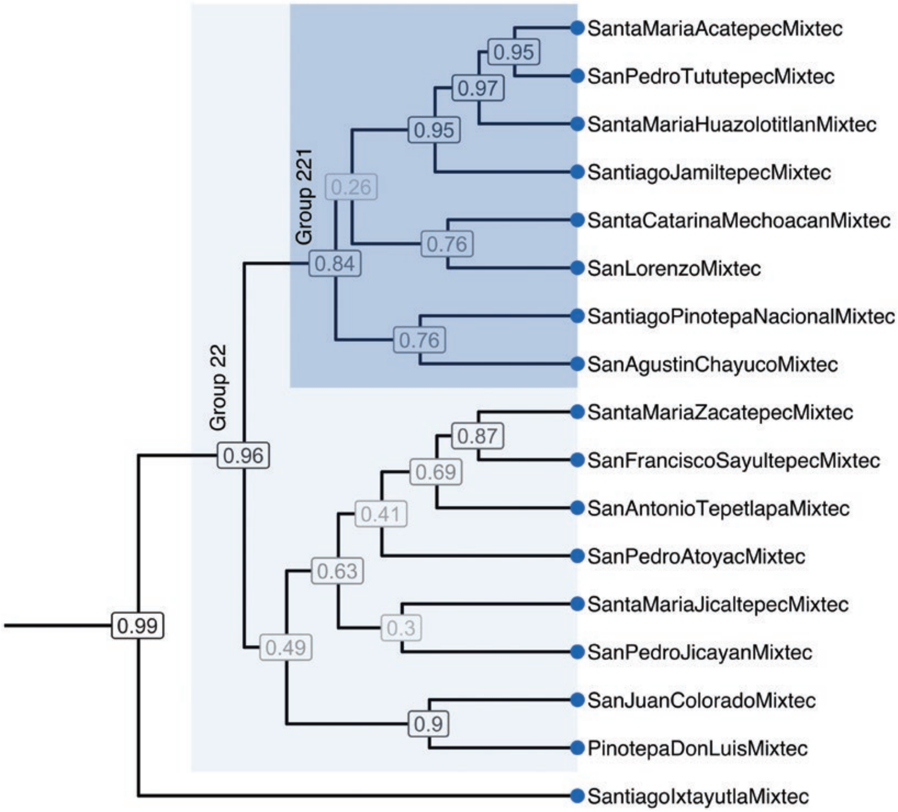

Figure 9: Mixtec Group 2 with subdivisions and node posteriors.

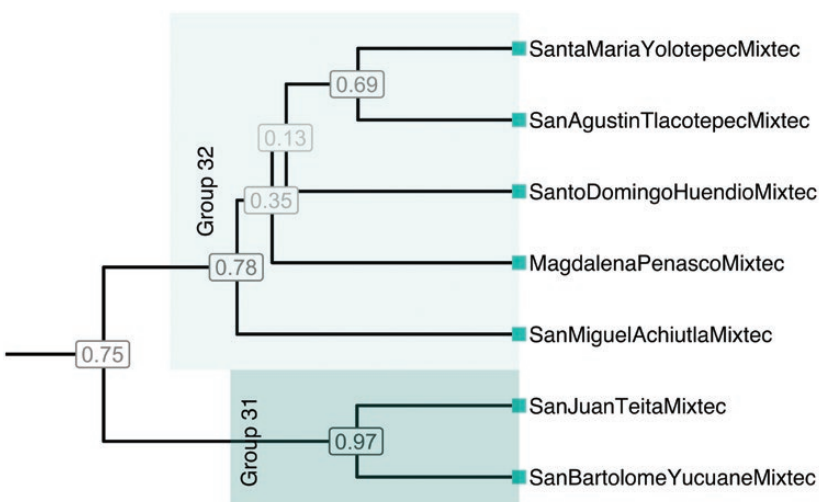

Figure 10: Mixtec Group 3 with subdivisions and node posteriors.

and corresponds to Josserand's (1983) Northern Alta group. The towns where these varieties are spoken are geographically separated from other Mixtec speakers. They are completely surrounded by Mazatec-speaking communities (Josserand 1983: 104) and Ixcatec and Chocho-speaking towns (all Otomanguean: Popolocan) and also border the Cuicatec region. The Mixtecs migrated to this area after the Mazatecan communities

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

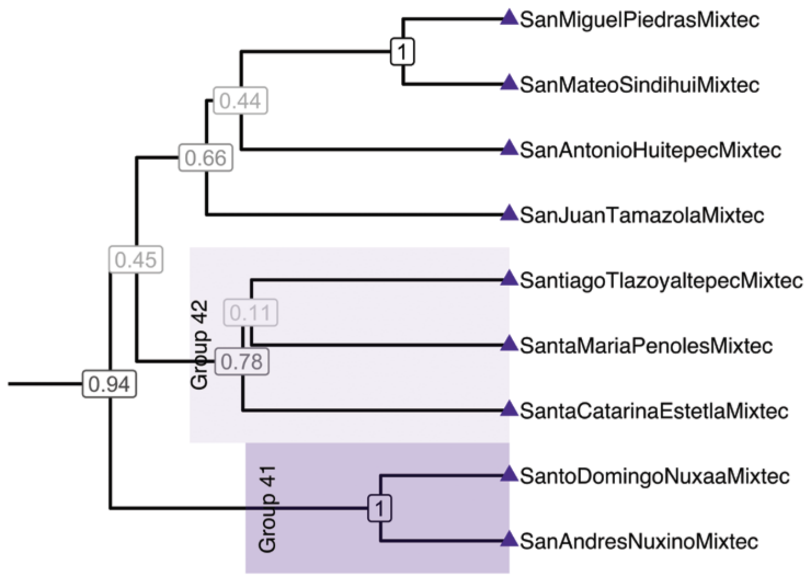

Figure 11: Mixtec Group 4 with subdivisions and node posteriors.

were already established (cf. Gudschinsky 1958 for the influence of Mixtec on Mazatec dialect history). These facts explain well why Group 1 is set apart from all others. The separation of Group 1 from the rest of the Mixtec varieties is also well reflected in the Densitree visualization and the NeighborNet (cf. SM 5 and 8).

Group 2 corresponds to the Coast group from Josserand (1983). It is clearly set apart from other groups in the Densitree visualization and also clusters together in the NeighborNet, showing less reticulation than other groups. The Mixtecs did not originally occupy territories on the coast, but rather emigrated there. Although these towns are connected to the Mixteca Alta by trade—both today and historically—they are more isolated from other Mixtec speakers than those of the other regions (Josserand 1983: 116). They are, however, not isolated from each other, which explains the high degree of cohesion of this group in terms of cognacy data. This is reflected in the Q-residuals, which are relatively low for most coastal varieties (cf. Figure 4). There is a relatively well-supported three-way partition within Group 2, cf. Figure 9. Subgroup 2.1 separates out Ixtayutla Mixtec (included in the East Coast group in Josserand 1983). Ixtayutla is relatively remote from other Mixtec towns and is described as conservative, located at the eastern boundary of the coastal Mixtec region (Josserand 1983: 116) and some villages in that region are bilingual with Zenzontepec Chatino. This explains its separate position in the MCC tree well. Subgroup 2.2 broadly reflects Josserand's (1983) West and East Coast groups. However, Subgroup 2.2.1, roughly corresponding to the East group, also contains the variety of Acatepec, singled out by Josserand as belonging to neither group. Subgroup 2.2 (roughly Josserand's West group) includes the variety of San Juan Colorado, classified as belonging to the eastern group in Josserand (1983). This opens avenues for further research. Since Josserand's classification is predominantly based on vowel correspondences, it is possible—and should be investigated—that considering sound changes of these varieties on a broader scale places them in the same groups as the lexical cognacy data.

Group 3 (see Figure 10) comprises a part of the varieties classified by Josserand (1983) as Western Alta. They are spoken in a roughly contiguous area around the Tlaxiaco district in the western part of the Mixteca Alta. Subgroup 3.1 sets the varieties of Teita and Yucuañe apart from the others, cf. 10. These varieties are also set apart from the others in the NeighborNet, where they appear closer to Linkage 5 and Group 6 varieties.

Group 4 broadly corresponds to Josserand's (1983) Eastern Alta group, although excluding the Diuxi, Tilantongo, and Tidaa varieties which are part of Linkage 5 in our model, cf. Figure 11. These varieties are also set apart from the other 'Eastern Alta' ones in the NeighborNet, where they appear closest to Linkage 5 languages. As can be seen in Figure 6, the relationship of Linkage 5 to the other groups is rather unclear, with all the intermediate nodes exhibiting very low posteriors. It is thus possible that further research could provide evidence for Josserand's (1983) original Eastern Alta group.

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

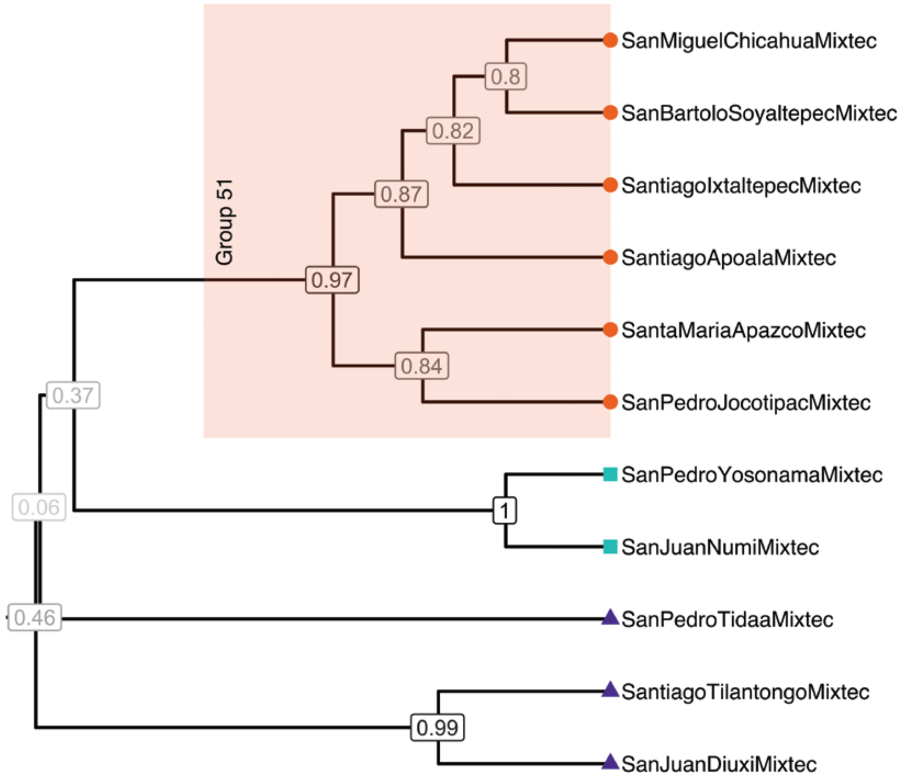

Figure 12: Mixtec Linkage 5 with subdivisions and node posteriors.

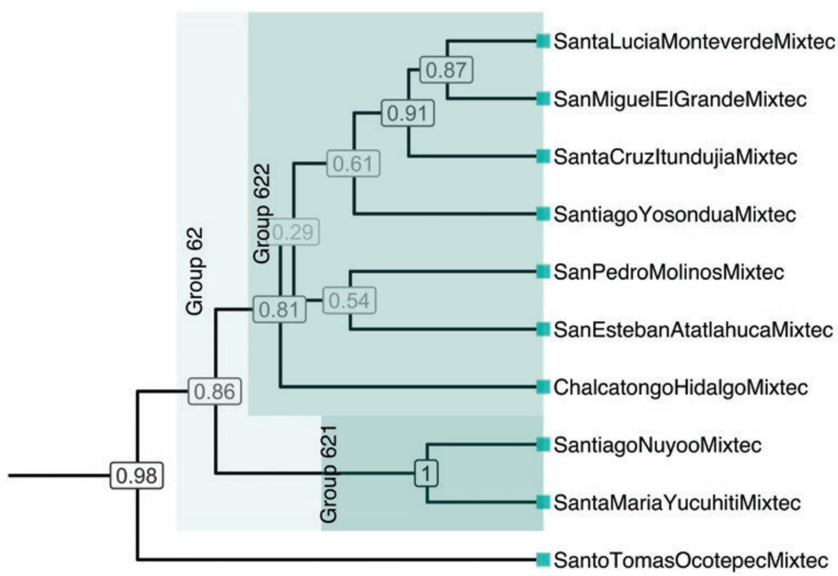

Figure 13: Mixtec Group 6 with subdivisions and node posteriors.

Linkage 5 (see Figure 12) consists of varieties and small groups that are linked together and connected to the other groups by nodes with low to very low posteriors. There is, however, one group within this linkage which has a high posterior and is well supported. This is Group 5.1, which corresponds to

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

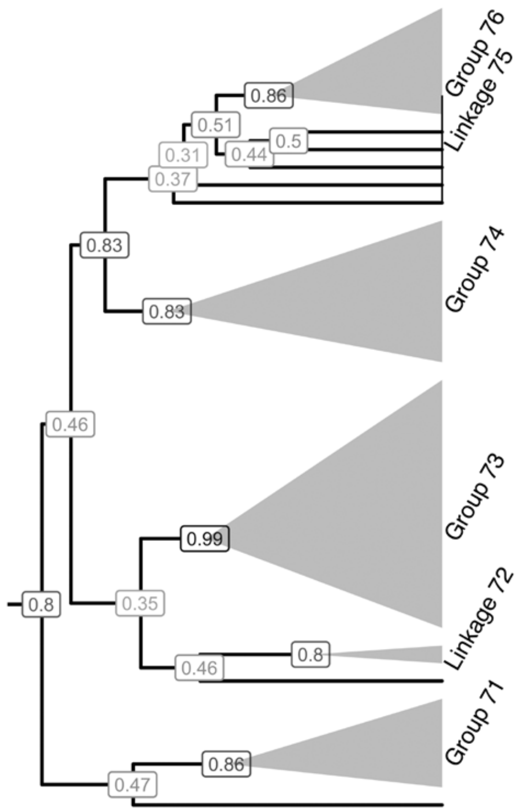

Figure 14: Subgroups of Mixtec Group 7 with node posteriors.

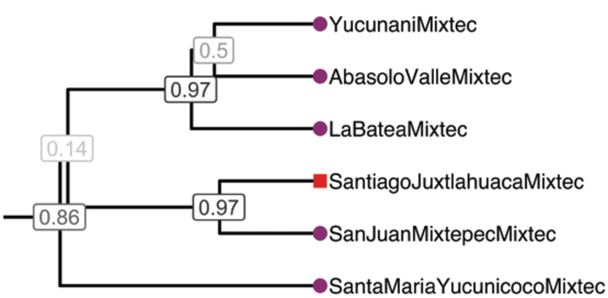

Figure 15: Mixtec Group 7.1 (purple = Josserand's (1983) Mixtepec, red = Josserand's (1983) Southern Baja).

Josserand's (1983) Northeastern Alta group. There are two well-supported subdivisions, of which one sets the varieties of Apazco and Jocotipac apart from the others. This is interesting because the varieties of Apazco and Apoala are represented as dialects of one language in Glottolog (Hammarström *et al.* 2021), while Soyaltepec is set apart as its own language. However, Apoala is a very conservative Mixtec variety and surrounding towns in the same municipality show a lot of variation. Our results suggest that Apoala and Apazco Mixtec are as distinct or even more distinct than other varieties traditionally seen as separate languages, such as Alacatlatzala and Alcozauca Mixtec.

Of the other varieties, those of Yosonama and Ñumi are closely related. In Josserand (1983) they are classified as Western Alta (our Groups 3 and 6). The varieties of Diuxi and Tilantongo are also grouped together closely. They and the not directly connected variety of Tidaa are classified in Josserand (1983) as Eastern Alta (our Group 4). As mentioned above, the posteriors connecting these varieties are very low and it thus remains an open question whether any of them form a closer relationship with each other or other groups outside of Linkage 5. These varieties also appear next to each other in the NeighborNet, but with a lot of reticulation between them.

Group 6 consists of varieties roughly situated in the western part of the Tlaxiaco district. In Josserand (1983), these varieties form part of the Western Alta group (together with our Group 3 and the varieties of Yosonama and Ñumi in Linkage 5). In the NeighborNet, they are placed close to Group 3 varieties, suggesting that further research could reveal a closer relationship. Internally, the variety of Ocotepec is set apart from the others and a larger second grouping (Subgroup 6.2) consists of the rest of the varieties, cf. Figure 13. Within Subgroup 6.2, we find a smaller division (6.2.1) consisting of the varieties of Yucuhiti and Nuyoo set apart from the rest (6.2.2). The internal composition of this subgroup invites further research, as the varieties of Chalcatongo and Molinos are said to be very similar to that of San Miguel el Grande (they are represented as dialects of it in Hammarström *et al.* 2021), but are placed relatively far from it in our results.

The next larger grouping that is relatively well supported (Group 7 with posterior = 0.8) contains varieties spreading over 7 of the groups identified by Josserand: Northern Baja, Central Baja, Western Baja, Southern Baja, Guerrero, Mixtepec, and Tezoatlan, cf. Figure 14. It is thus worth discussing its internal structure in more detail. In the NeighborNet, this group forms a large 'fan' complex apart from the other varieties with a large degree of reticulation. We identify four subgroups and two linkages, as well as two varieties not clearly placed in a group or linkage, cf. Figure 14. The higher level nodes—representing how these subgroups are connected to each other—have low posteriors, suggesting further research is needed in this area.

Subgroup 7.1 corresponds to Josserand's (1983) Mixtepec group, cf. Figure 15. It shows a further partition separating the variety of Yucunicoco from the others. This is congruent with the fact that Yucunicoco is located further away from the other villages, closer to the Triqui area. The issue in need of discussion is the classification of the variety of

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

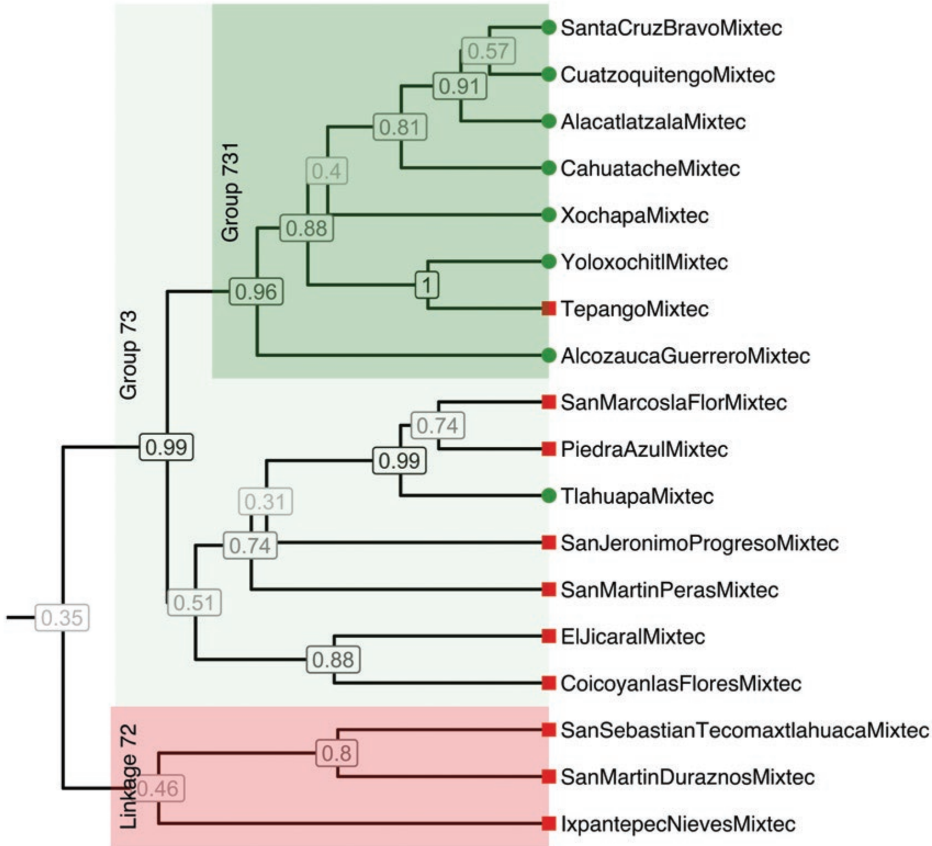

Figure 16: Mixtec Linkage 72 and Group 73 (green = Josserand’s (1983) Guerrero, red = Josserand’s (1983) Southern Baja).

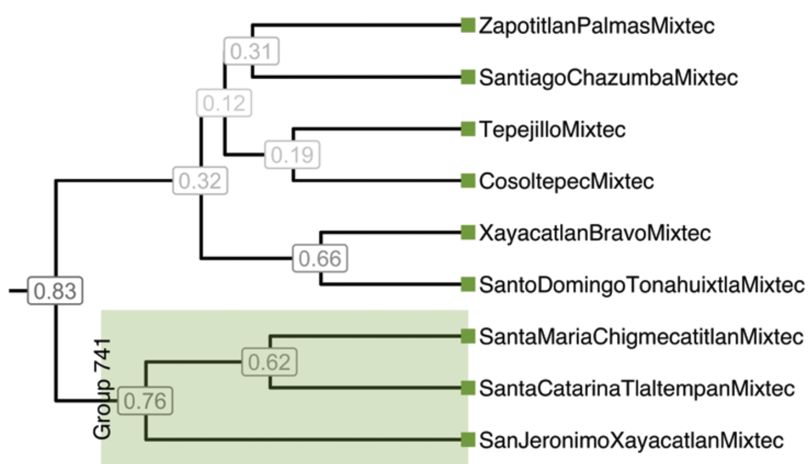

Figure 17: Mixtec Group 74 (Josserand’s (1983) Northern Baja).

Santiago Juxtlahuaca in this group rather than Linkage 7.2. This variety belongs to the Southern Baja group according to Josserand and should be most closely related to the variety of Tecomaxtlahuaca and surroundings (see below on Linkage 7.2). However, earlier proposals such as Mak and Longacre (1960) and Bradley (1968) did place Juxtlahuaca with the Mixtepec varieties. There are several reasons why

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

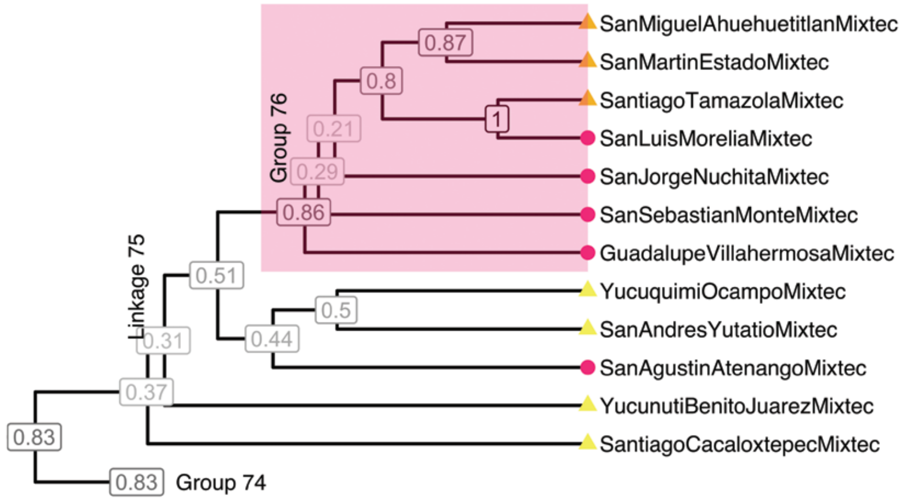

Figure 18: Mixtec Linkage 75 and Group 76 (yellow = Josserand's (1983) Tezoatlán, pink = Josserand's (1983) Central Baja, orange = Josserand's (1983) Western Baja).

difficulties might arise in the classification of this variety. Santiago Juxtlahuaca is the capital of the Juxtlahuaca district and also the seat of the municipality of Juxtlahuaca. As such, it is not only the biggest town in the area but it also sees a continuous and considerable influx of people from surrounding towns, both Mixtec and Triqui. This means that there are speakers of different varieties present (as well as many who only speak Spanish), which is likely to influence the local variety of this town, for which little data are available (only those collected in Josserand 1983).

Subgroup 7.3 covers all varieties from Josserand's (1983) Guerrero group and a portion of her Southern Baja group. Linkage 7.2, also illustrated in Figure 16, is very small and contains only varieties of Josserand's (1983) Southern Baja group. However, the two groups from Josserand are not clearly separated in our results, cf. Figure 16, and there appear to be a few 'misplaced' varieties, such as Tlahuapa Mixtec and Tepango Mixtec. The posterior of the node above Linkage 7.2 and Group 7.3 is very low, indicating uncertainty about a higher-level group combining those two—in other words, there is no strong evidence for a higher-level Southern Baja-Guerrero group.

There are, however, well-supported lower divisions. Subgroup 7.3.1 covers Josserand's (1983) Guerrero group, except for Tlahuapa and includes the Southern Baja variety of Tepango. The village of Tepango is one of the westernmost Mixtec villages, surrounded by Nahuatl and Mè'phàà speakers (Josserand 1983: 105). It is an important variety for the reconstruction of proto-Mixtec because it is one of the very few that preserves final glottal stops. It is classified by Josserand as belonging to the Southern Baja group based on vowel correspondences. A detailed study of other sound changes is outside the scope of this article, but based on the data we collected, Tepango does show a closer affinity to Yoloxóchitl Mixtec (Guerrero) than Coicoyán Mixtec (Southern Baja). We thus suggest that Tepango is in fact placed correctly in our tree. The situation is different for Tlahuapa Mixtec. Based on current knowledge, this variety should be part of the same group as the other varieties spoken in Guerrero (Subgroup 7.3.1). The reason for the 'misplacement' probably has to do with the circumstances of data collection. The data for the Tlahuapa, Piedra Azul, and San Marcos de la Flor varieties were all collected from diaspora speakers in California in a collaborative documentation project (Campbell and Reyes Basurto forthcoming). Most likely, the speakers influenced each other in selecting certain entries over others for some of the meanings, leading to a higher number of 'apparent' cognates between those three varieties—or in other words, obfuscating semantic shifts between 'cognates'. This illustrates well the importance of a detailed knowledge of the data sources and collection techniques and the need for broad-scale, rigorous language surveys.

Within Subgroup 7.3.1, there is a well-supported split that sets Alcozauca Mixtec apart from the other varieties. In Glottolog (Hammarström *et al.* 2021), the variety of Xochapa is listed as a dialect of Alcozauca Mixtec, but our study suggests that this classification should be revisited. Next, we find a well-supported division into two groups. The smaller group contains the varieties of Tepango and Yoloxóchitl, while the second, larger group (or 'northern' division) contains the rest of the varieties not already mentioned, to the exclusion of Xochapa which seems to occupy a position in between those divisions.

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

The rest of Group 7.3 covers varieties previously classified as Southern Baja. All of these varieties are spoken in the northwestern part of the state of Oaxaca very close to the border with Guerrero. These varieties are thus located in a geographically contiguous area with those of Subgroup 7.3.1, so it is not surprising that they would have a closer relationship to each other than previously assumed.

Linkage 7.2 proper contains varieties spoken further east from those of Subgroup 7.3 and are also classified as Southern Baja by Josserand (1983). However, the next node with a high posterior (0.99) does not neatly separate the varieties into these two groups, but rather into a larger group consisting of the Guerrero varieties as well as the westernmost varieties of the Southern Baja (centred around San Martín Peras and Coicoyán de las Flores). It is unclear whether the variety of Ixpantepec should be included in this group given the low posterior connecting it to the varieties of Tecomaxtlahuaca and Duraznos, but based on Josserand's classification and our own fieldwork in the area, we conclude that it does form part of this linkage.

What remains unclear and needs further investigation is how Group 7.1, Linkage 7.2, and Group 7.3 are connected to each other. The posterior of the node linking them together is low, which means that it is questionable whether a group covering all of them should be posited. The varieties of this region, called the Mixteca Baja, are not as well studied and documented as those of the Mixteca Alta or the Coast and this was even more the case at the time of Josserand's (1983) study. It is thus possible that more evidence for this new division will be found based on future research.

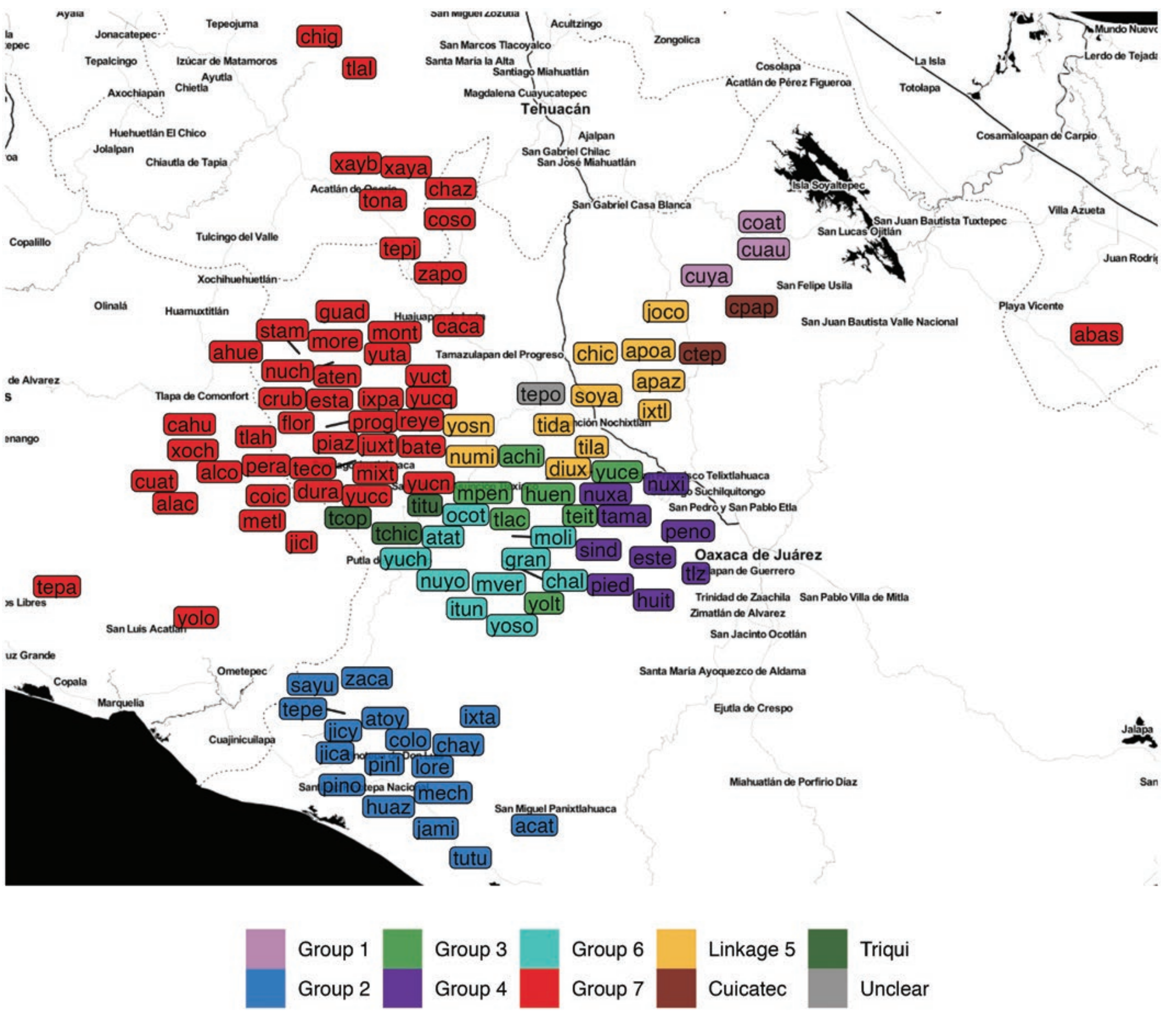

Figure 19: Map of the sample languages with colors indicating primary branches and new Mixtec subgroups proposed by our model (for label abbreviations see Supplementary Materials).

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

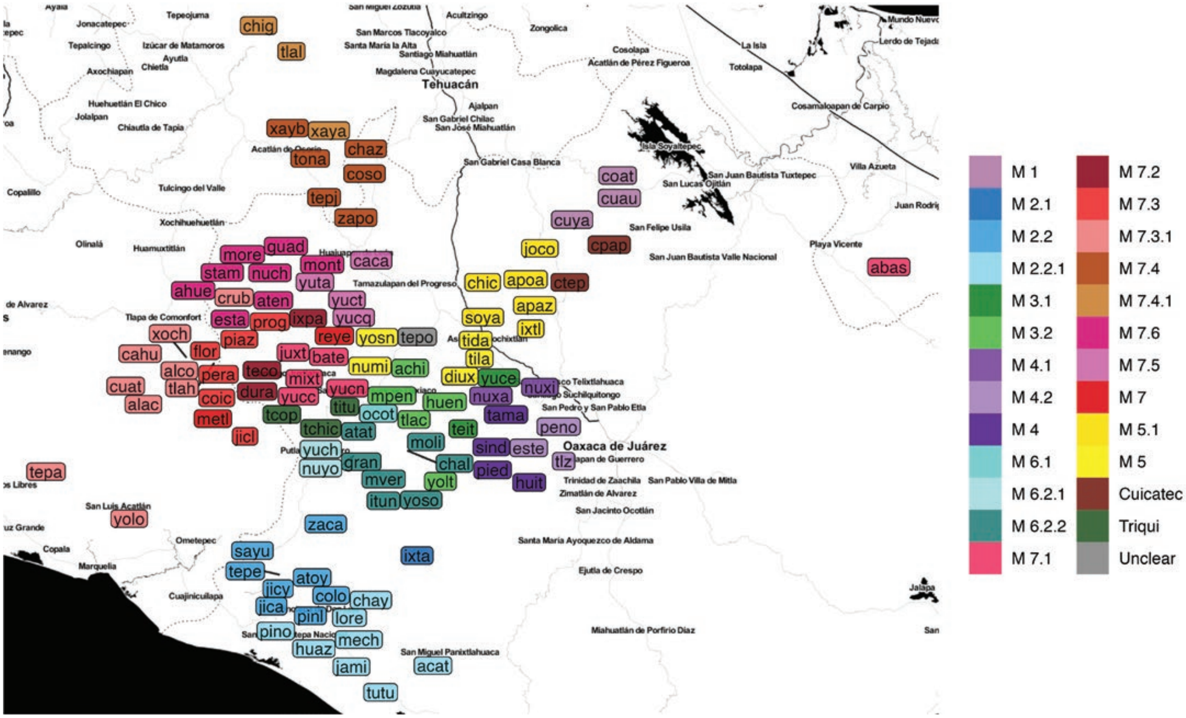

Figure 20: Map of the sample languages with colors indicating primary branches and new Mixtec subgroups (indicated with M in the legend) with lower level divisions proposed by our model (for label abbreviations see Supplementary Materials).

Subgroup 7.4 corresponds to the Northern Baja group from Josserand (1983). Internally, there is a division (Subgroup 7.4.1) that sets apart the varieties of Chigmecatitlan, Tlaltempan, and Jerónimo Xayacatlán from the others, cf. Figure 17. With respect to the first two, this corresponds well with the geographic location, being the northernmost Mixtec varieties, spoken in the state of Puebla surrounded by Nahuatl and Popoloca speakers (Josserand 1983: 105). The placement of Jerónimo Xayacatlán needs to be investigated further. This variety is spoken further south near the border to Oaxaca and in very close proximity to Tonahuixtla and Xayacatlán de Bravo, with which we would expect it to be grouped together. The relationships of the rest of the varieties of Subgroup 7.4 remain unclear with very low posteriors connecting them.

Linkage 7.5 consists of a number of varieties that cannot be clearly linked together or connected to Group 7.6. The varieties included in this agglomerate are those classified as the Tezoatlán group by Josserand (1983) plus the variety of Atenango from her Central Baja group, cf. Figure 18. Subgroup 7.6 covers Josserand's Central Baja and Western Baja varieties. The internal structure is not very well supported, but there is a smaller group with a relatively high posterior covering Josserand's (1983) Western Baja varieties and the Central Baja variety of Morelia.

We summarize our results in Table A.2 in the Appendix and Figures 19 and 20, as a new subgrouping proposal for Mixtec, which we see as a much-needed, up-to-date starting point for further research on the history of these languages. The first three splits in our MCC tree identify varieties that are either temporally removed (in the case of Teposcolula) or groups that migrated to areas not originally inhabited by Mixtec speakers (Group 1 to the far north and Group 2 on the coast). This history of relative separation from other Mixtec speakers is well reflected in our results. With respect to the dialect areas proposed by Josserand (1983), few of our groups completely overlap with her areas. Our results diverge most from hers in proposing a large high-level group (Group 7) comprising seven of the dialect areas and with respect to the internal structure of the subgroups and linkages (Groups/Linkages 71–76).

6. Conclusion

Despite claims to the contrary, our study shows that much can be learned about Mixtecan language history by applying and combining traditional historical linguistic methods (such as establishing cognate sets) and Bayesian phylogenetics. Our results indicate that at least certain sets of varieties within this language

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

family are best viewed as linkages or dialect chains, but this does not mean that we cannot investigate their genealogical relationships any further. We provide a starting point to further evaluate the linguistic relationships within these linkages as well as their relationship to other groups. In addition, we recover many well-supported, coherent groups within the Mixtec branch, suggesting that some of its history can be described as relatively tree-like.

This adds further support to the idea that the Mixtecan language family can neither be characterized completely as a tree nor accounted for solely in a wave model. In Mixtec, we identify four groups (Groups 1, 2, 4, and 6) which are very well supported by the MCC tree and two more (Groups 3 and 7) which are relatively well supported. What is less clear, as mentioned in Section 5, is the relationship between those groups. The same is true internally of some of the larger groups, i.e., there are some well-supported lower-level subgroups, while other varieties form part of loosely connected linkages (cf. the internal structure of Group 7). This could reflect a scenario in which more wave-like periods of diversification were in turn followed by more tree-like ones.

The limitations of our study, such as the exclusion of verbs and the absence of good calibration points, also leave some questions unanswered. One of these concerns is the relationship between the three primary branches of Mixtecan. Since our model grouped Triqui and Cuicatec together—probably due to Mixtec's influence on both—we cannot confirm or discard the idea that Mixtec and Cuicatec are more closely related to each other than either is to Triqui. We would also expect that some of the varieties placed outside linkages or groups can be re-evaluated for group membership once there is a better understanding of sound correspondences and sufficient analysis of verbs so that they can be integrated into the cognate database.

In Section 1, we mentioned a recent small-scale intelligibility study carried out around the town of Juxtlahuaca (Padgett 2017) and mentioned the discrepancies to classifications in reference catalogues (Lewis *et al.* 2015; Hammarström *et al.* 2021). There is no reason to believe that the result would be much different for other parts of the Mixteca, demonstrating the need for a solid basis for language classification that involves not only traditional classification methods but also studies and reports of intelligibility.

We also show that a good knowledge of data sources is crucial for the ability to correctly interpret the MCC tree, especially with respect to varieties which appear in unexpected places in the tree. This in turn means that for language families with little historical data available, the best results can be achieved by gathering data through large-scale surveys applied consistently. This could help eliminate interference from differences in data gathering and preparation. The last such survey was conducted in the Mixteca region in the late 1970s (Josserand 1983: xi), focusing on Mixtec. We hope that our work will inspire a much-needed update and expansion of this work.

Acknowledgements

First and foremost, we extend our gratitude to all the Mixtec, Triqui, and Cuicatec speakers who shared their language and knowledge with us and to all those who collaborated with the documentary linguists before us. This study would not have been possible without them. We would like to thank Konstantin Hoffmann and Denise Kühnert for their invaluable help with the Bayesian phylogenetic models and their interpretation, which much improved our study. We are also grateful to Hedvig Skirgård for assistance with the creation of the NeighborNets and the calculation of the delta-scores and Q-residuals. We thank Johann-Mattis List, Johannes Englisch, and Christoph Rzymski for the creation of the CLDF-version of the data set. Finally, we extend our thanks to two anonymous reviewers for their detailed comments.

Funding

This work was funded in part by the National Science Foundation award 1660355 to the University of California, Santa Barbara, PI: Eric W. Campbell and Mary Bucholtz; by the University of California, Santa Barbara Academic Senate Faculty Research Grant (2018-19), PI: Eric W. Campbell; and by the Endangered Languages Documentation Programme Small Grant 0566 (2019-2022) to the University of California, Santa Barbara, PI: Sandra Auderset.

Data availability statement

The data underlying this article and all supplementary materials are available on Zenodo, at https://doi.org/10.5281/zenodo.7940497. The data are also available as a CLDF data set (Forkel *et al.* 2018) on Zenodo at https://doi.org/10.5281/zenodo.7642971.

Notes

1. English version: The problem is the apparent violation or infringement of the expectations and ideals of the comparative method in historical linguistics (…). But this model assumes an actual separation of the groups, once a change has divided them. In contrast, in the Mixteca, we see a more dynamic situation concerning the development of the language families. (translation ours)
2. For example, in Macaulay (1996:1)'s grammar where the introduction states that 'Mixtec is an Otomanguean language', only to clarify later on that '"Mixtec" really

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

should be considered a group of related but distinct languages' (Macaulay 1996:6).

3. As is common in phylogenetic studies, the 'present' is set at 1950 like in archaeology, so we can cite it as 'Before Present'.
4. Abbreviations used in the tables: MlogL = Marginal log likelihood, logBF = log Bayes factor, SD = Standard deviation
5. For abbreviations see footnote 4.

References

Alexander, Ruth Mary (1980) *Gramática mixteca de Atatlahuca*. México, D.F.: Instituto Lingüístico de Verano, A.C.

——— (1988) 'A Syntactic Sketch of Ocotepec Mixtec', in C. Henry Bradley and Barbara E. Hollenbach (eds.) *Studies in the Syntax of Mixtecan Languages*, vol. 1, pp. 151–304. Dallas and Texas: Summer Institute of Linguistics and the University of Texas at Arlington.

Amith, Jonathan D. and Rey Castillo García (n.d.) 'Recursos lexicosemánticos para el mixteco de Yoloxóchitl, municipio de San Luis Acatlán, Guerrero (Glottocode yolo1241; ISO 639-3 xty)' (unpublished).

Anderson, E. Richard and Hilario Concepción Roque (1983) *Diccionario Cuicateco: Español-Cuicateco, Cuicateco-Español*. México, D.F.: Instituto Lingüístico de Verano.

Anderson, Lynn (2006) *Vocabulario de palabras que se relacionan con el maíz en mixteco de Alacatlatzala, Guerrero*, 2nd edn. México, D.F.: Instituto Lingüístico de Verano, A.C.

Anttila, Raimo (1989) *Historical and Comparative Linguistics*. Amsterdam: John Benjamins.

Arana Osnaya, Evangelina (1960) 'Relaciones internas del mixteco trique', in *Anales del Museo Nacional de México*, vol. 12, pp. 219–273. México, D.F.: Instituto Nacional de Antropología e Historia.

Balkansky, Andrew K et al. (2000) 'Archaeological Survey in the Mixteca Alta of Oaxaca, Mexico', *Journal of Field Archaeology*, 27: 365–89.

Becerra Roldán, Braulio (2015) '*Un estudio fonológico del mixteco de Santo Domingo Huendio, Oaxaca*', *MA thesis*, ENAH, México, D.F.

Belmar, Francisco (1897) *Ensayo sobre la lengua trike. Lenguas Indígenas del Estado de Oaxaca*. Oaxaca: Lorenzo San-German.

——— (1902) *El cuicateco*. Oaxaca: Imprenta del Comercio.

Bernal, Ignacio (1966) 'The Mixtecs in the Archaeology of the Valley of Oaxaca', in John Paddock (ed.), *Ancient Oaxaca: Discoveries in Mexican Archaeology and History*. Stanford University Press.

Birchall, Joshua, Michael Dunn, and Simon J. Greenhill (2016) 'A Combined Comparative and Phylogenetic Analysis of the Chapacuran Language Family', *International Journal of American Linguistics*, 82: 255–84.

Bouckaert, Remco R. et al. (2019) 'BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis', *PLoS Computational Biology*, 15: e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

Bouckaert, Remco R and Joseph Heled (2014) 'DensiTree 2: Seeing Trees through the Forest', *bioRxiv*. https://www.biorxiv.org/content/early/2014/12/08/012401.full.pdf.

Bouckaert, Remco R, Claire Bowern and Quentin D. Atkinson (2018) 'The Origin and Expansion of Pama–Nyungan Languages across Australia', *Nature Ecology and Evolution*, 2: 741–9.

Bowern, Claire (2013) 'Relatedness as a Factor in Language Contact', *Journal of Language Contact*, 6: 411–32.

——— (2018) 'Computational Phylogenetics', *Annual Review of Linguistics*, 4: 281–96.

——— and Q. Atkinson (2012) 'Computational Phylogenetics and the Internal Structure of Pama-Nyungan', *Language*, 88: 817–45.

Bradley, C. Henry (1968) 'A Method for Determining Dialectal Boundaries and Relationships', *América Indígena*, 28: 751–60.

——— and J. Kathryn Josserand (1982) 'El protomixteco y sus descendientes', *Anales de Antropología*, 19: 279–343.

Bradley, David P. (1991) 'A Preliminary Syntactic Sketch of Concepción Pápalo Cuicatec', in C. Henry Bradley and Barbara E. Hollenbach (eds.), *Studies in the Syntax of Mixtecan Languages*, vol. 3, pp. 409–506. Summer Institute of Linguistics and the University of Texas at Arlington.

Bryant, David and Vincent Moulton (2004) 'Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks', *Molecular Biology and Evolution*, 21: 255–65.

Byers, Douglas S (ed.) (1967) 'The Prehistory of the Tehuacan Valley', *Environment and Subsistence*, vol. 1. Austin, TX: University of Texas Press.

Campbell, Eric W. and Griselda Reyes Basurto (forthcoming) El Tu'un Savi (mixteco) en California: documentación y activismo lingüístico, in Marcela San Giacomo, Fidel Hernández Mendoza and Michael Swanton (eds.), *Estudios sobre lenguas mixtecanas*. México, D.F.: Seminario Permanente de Lenguas Mixtecanas, Instituto de Investigaciones Antropológicas, Universidad Nacional Autónoma de México.

Campbell, Eric W. (2017a) 'Otomanguean Historical Linguistics: Exploring the Subgroups', *Language and Linguistics Compass*, 11: e12244.

——— (2017b) 'Otomanguean Historical Linguistics: Past, Present, and Prospects for the Future', *Language and Linguistics Compass*, 11: e12240.

Campbell, Lyle (1997) *American Indian Languages: The Historical Linguistics of Native America*. Oxford: Oxford University Press.

Comrie, Bernard (2000) 'Language Contact, Lexical Borrowing, and Semantic Fields', in Dicky Gilbers, John Nerbonne and Jos Schaeken (eds.), *Languages in Contact*, pp. 73–86. Amsterdam-Atlanta, GA: Rodopi.

Cruz, Emiliana and Anthony C. Woodbury (2014) 'Finding a Way into a Family of Tone Languages: The Story and Methods of the Chatino Language Documentation Project', *Language Documentation and Conservation*, 8: 490–524.

de Alvarado, Francisco (1962 [1593]) *Vocabulario en lengua mixteca. Reproducción facsimilar con un estudio de Wigberto Jiménez Moreno*. México, D.F.: Instituto Nacional Indigenista e Instituto Nacional de Antropología e Historia.

de los Reyes, Antonio (1890 [1593]) *Arte en lengua mixteca*. Alençon: Typographie E. Renaut de Broise.

DiCanio, Christian T (2008) '*The Phonetics and Phonology of San Martín Itunyoso Trique*', *PhD dissertation*, University of California, Berkeley.

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

——— (2022a) *Itunyoso Triqui Collection of Christian DiCanio*. The Archive of the Indigenous Languages of Latin America, ailla.utexas.org. PID ailla:243667; accessed January 5, 2022.

——— (2022b) 'Vocabulario del Triqui de San Martín Itunyoso' (unpublished).

Drummond, Alexei J et al. (2006) 'Relaxed Phylogenetics and Dating with Confidence', *PLoS Biology*, 4: e88.

Duchene, Sebastian et al. (2020) 'Bayesian Evaluation of Temporal Signal in Measurably Evolving Populations', *Molecular Biology and Evolution*, 37: 3363–79.

Dunn, Michael J et al. (2011) 'Evolved Structure of Language shows Lineage-specific Trends in Word-order Universals', *Nature*, 473: 79–82.

Dürr, Michael (1987) 'A Preliminary Reconstruction of the Proto-Mixtec Tonal System', *Indiana*, 11: 19–61.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig (2021) *Ethnologue: Languages of the World*, 24th edn. Dallas, Texas: SIL International.

Egland, Steven T. (1983) *La inteligibilidad interdialectal en México: Resultados de algunos sondeos*. México, D.F.: Instituto Lingüístico de Verano, A.C.

Farris, Edwin R. (1992) 'A Syntactic Sketch of Yosondúa Mixtec', in C. Henry Bradley and Barbara E. Hollenbach (eds.), *Studies in the syntax of Mixtecan languages*, vol. 4, pp. 1–171. Summer Institute of Linguistics and the University of Texas at Arlington.

Flannery, Kent V. and Joyce Marcus (eds.) (1983) *The Cloud People: Divergent Evolution of the Zapotec and Mixtec Civilizations*. New York: Academic Press.

Forkel, Robert et al. (2018) 'Cross-Linguistic Data Formats, Advancing Data Sharing and Re-use in Comparative Linguistics', *Scientific Data*, 5: 1–10.

François, Alexandre (2015) 'Trees, Waves and Linkages: Models of Language Diversification, in Claire Bowern and Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, pp. 161–189. Routledge.

Galindo Sanchez, Bernardo (2009) *Vocabulario Basico Tu'un Savi—Castellano*, 1st edn. Xalapa, Veracruz: Academia Veracruzana de las Lenguas Indígenas.

Gittlen, Laura (2016) *Gramática popular Mixteco del norte de Tlaxiaco*. México, D.F.: Instituto Lingüístico de Verano.

Good, Claude (1978) *Diccionario triqui de Chicahuaxtla: triqui-castellano, castellano-triqui*. México, D.F.: Instituto Lingüístico de Verano, A.C.

Gray, Russell D., David Bryant and Simon J. Greenhill (2010) 'On the Shape and Fabric of Human History', *Philosophical Transactions of the Royal Society B*, 365: 3923–33.

Greenhill, Simon J and Russell D Gray (2009) 'Austronesian Language Phylogenies: Myths and Misconceptions about Bayesian Computational Methods', in Alexander Adelaar and Andrew Pawley (eds.), *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*, pp. 375–397. Canberra: Pacific Linguistics.

———, Paul Heggarty, and Russell D Gray (2020) 'Bayesian Phylolinguistics', in Richard D. Janda, Brian D. Joseph, and Barbara S. Vance (eds.), *The Handbook of Historical Linguistics*, vol. 2, chap. 11, pp. 226–253. Wiley Blackwell.

———, Thomas E. Currie and Russell D. Gray (2009) 'Does Horizontal Transmission Invalidate Cultural Phylogenies?', *Proceedings of the Royal Society of London B: Biological Sciences*, 276: 2299–306.

——— et al. (2023). A recent northern origin for the Uto-Aztecan family. *Language*. https://doi.org/10.1353/lan.0.0276.

Gudschinsky, Sarah C. (1959) *Proto-Popotecan: A Comparative Study of Popolocan and Mixtecan*. Baltimore: Waverly Press.

———, (1958) 'Mazatec Dialect History: A Study in Miniature', *Language*, 34: 469–81.

Hammarström, Harald, Robert Forkel, Martin Haspelmath, and Sebastian Bank (2021) Glottolog 4.4. Available online at http://glottolog.org.

Hernández Mendoza, Fidel (2017) '*Tono y fonología segmental en el triqui de Chicahuaxtla*', *PhD dissertation*, Universidad Nacional Autónoma De México, México, D.F.

Hernández Martínez, Carmen and Sandra Auderset (2020) *San Martín Duraznos Mixtec Vocabulary*, 1st edn (unpublished).

Hernández Mendoza, Fidel (2020) *Vocabulario Triqui de Chicahuaxtla* (unpublished).

Hills, Robert A. (1990) 'A Syntactic Sketch of Ayutla Mixtec', in C. Henry Bradley and Barbara E. Hollenbach (eds.), *Studies in the Syntax of Mixtecan Languages*, vol. 2. Summer Institute of Linguistics and the University of Texas at Arlington.

Hoffmann, Konstantin et al. (2021) 'Bayesian Phylogenetic Analysis of Linguistic Data using BEAST', *Journal of Language Evolution*, 6: 119–35.

Hoijer, Harry (1956) 'Lexicostatistics: A Critique', *Language*, 32: 49–60.

Holland, William R (1959) 'Dialect Variations of the Mixtec and Cuicatec Areas of Oaxaca, Mexico', *Anthropological Linguistics*, 1: 25–31.

Hollenbach, Barbara E. (1977) 'Phonetic vs. Phonemic Correspondence in Two Trique Dialects', in William R. Merrifield (ed.), *Studies in Otomanguean phonology*. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington.

——— (1992) 'A Syntactic Sketch of Copala Trique', in C. Henry Bradley and Barbara E. Hollenbach (eds.), *Studies in the Syntax of Mixtecan Languages*. Summer Institute of Linguistics.

Hollenbach, Barbarba E. (2017) *Diccionario mixteco de Magdalena Peñasco*. México, D.F.: Instituto Lingüístico de Verano, A.C.

Holman, Eric W. (2010) 'Do Languages Originate and Become Extinct at Constant Rates?', *Diachronica*, 27: 214–25.

Huson, Daniel H and David Bryant (2006) 'Application of Phylogenetic Networks in Evolutionary Studies', *Molecular Biology and Evolution*, 23: 254–67.

INALI (2009a) *Catálogo de las Lenguas Indígenas Nacionales*. México, D.F.: Instituto Nacional de Lenguas Indígenas INALI.

——— (2009b) 'Mixteco', in *Catálogo de las Lenguas Indígenas Nacionales*, pp. 199–218. México, D.F.: Instituto Nacional de Lenguas Indígenas INALI.

Jansen, Maarten (1990) 'The Search for History in Mixtec Codices', *Ancient Mesoamerica*, 1: 99–112.

Jansen, Maarten and Gabina Aurora Pérez Jiménez (2011) 'Introduction to Mixtec Pictography', in *The Mixtec Pictorial Manuscripts*, chap. 1, pp. 1–41. Brill.

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Jiménez Moreno, Wigberto (1962) 'Estudios mixtecos', in *Vocabulario en lengua mixteca*, pp. 9–105. Instituto Nacional Indigenista e Instituto Nacional de Antropología.

Johnson, Audrey F. (1988) 'A Syntactic Sketch of Jamiltepec Mixtec', in C. Henry Bradley and Barbara E. Hollenbach (eds.), *Studies in the Syntax of Mixtecan Languages*, vol. 1, pp. 11–150. Summer Institute of Linguistics and the University of Texas at Arlington.

Josserand, J. Kathryn (1983) '*Mixtec Dialect History*', *PhD dissertation*, Tulane University.

———, Marcus Winter and Nicholas A. Hopkins (1984) *Essays in Otomanguean Culture History*. Nashville, Tennessee: Vanderbilt University.

Joyce, Arthur A. (2011) *Mixtecs, Zapotecs, and Chatinos: Ancient Peoples of Southern Mexico*. Wiley and Sons.

Kalyan, Siva and Alexandre François (2018) 'Freeing the Comparative Method from the Tree Model', *Senri Ethnological Studies*, 98: 59–89.

Kass, Robert E and Adrian E. Raftery (1995) 'Bayes Factors', *Journal of the American Statistical Association*, 90: 773–95.

Kassian, Alexei et al. (2010) 'The Swadesh Wordlist. An Attempt at Semantic Specification', *The Journal of Language Relationship*, 4: 46–89.

Kaufman, Terrence (1983) 'New Perspectives on Comparative Otomanguean Phonology' (unpublished monograph).

——— (1988) 'Otomanguean Tense/Aspect/Mood, Voice, and Nominalization Markers' (unpublished monograph).

——— (2006) 'Oto-Manguean Languages', in Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, vol. 9, pp. 118–124. Oxford: Elsevier.

Kolipakam, Vishnupriya et al. (2018) 'A Bayesian Phylogenetic Study of the Dravidian Language Family', *Royal Society Open Science*, 5: 171504.

Kowalewski, Stephen A., Andrew K. Balkansky, Laura R. Stiver Walsh, Thomas J. Pluckhahn, John F. Chamblee, Veronica Perez, Verenice Y. Heredia Espinoza, and Charlotte A. Smith (2009) *Origins of the Ñuu: Archaeology in the Mixteca Alta, Mexico*. University Press of Colorado.

Kuiper, Albertha and Joy Oram (1991) 'A Syntactic Sketch of Diuxi-Tilantongo Mixtec', in C. Henry Bradley and Barbara E. Hollenbach (eds.), *Studies in the Syntax of Mixtecan Languages*, vol. 3, pp. 185–408. Summer Institute of Linguistics and the University of Texas at Arlington.

Laycock, Donald C. (1970) 'Eliciting Basic Vocabulary in New Guinea', in *Pacific Linguistic Studies in Honour of Arthur Capell*, 1127–76. Canberra: Australian National University.

Lee, Sean and Toshikazu Hasegawa (2011) 'Bayesian Phylogenetic Analysis Supports an Agricultural Origin of Japonic Languages', *Proceedings of the Royal Society B: Biological Sciences*, 278: 3662–9.

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (2015) *Ethnologue: Languages of the World*, 18th edn. Dallas: SIL International.

List, Johann-Mattis et al. (2018) 'Sequence Comparison in Computational Historical Linguistics', *Journal of Language Evolution*, 3: 130–44.

———, Simon J. Greenhill and Russell D. Gray (2017) 'The Potential of Automatic Word Comparison for Historical Linguistics', *PLoS One*, 12: e0170046.

———, and Robert Forkel (2021) LingPy. A Python Library for Historical Linguistics, Version 2.6.9. http://lingpy.org, DOI: https://zenodo.org/badge/latestdoi/5137/lingpy/lingpy.

Longacre, Robert E. (1961) 'Swadesh's Macro-Mixtecan Hypothesis', *International Journal of American Linguistics*, 27: 9–29.

——— (1962) 'Amplification of Gudschinsky's Proto-Popolocan-Mixtecan', *International Journal of American Linguistics*, 28: 227–42.

——— (1966) 'On Linguistic Affinities of Amuzgo', *International Journal of American Linguistics*, 32: 46–9.

Longacre, Robert E. (1957) *Proto-Mixtecan*. Indiana University.

Macaulay, Monica Ann (1996) *A Grammar of Chalcatongo Mixtec*. University of California Press.

Maddison, David R., David L. Swofford and Wayne P. Maddison (1997) 'NEXUS: An Extensible File Format for Systematic Information', *Systematic Biology*, 46: 590–621.

Mak, Cornelia and Robert E. Longacre (1960) 'Proto-Mixtec Phonology', *International Journal of American Linguistics*, 26: 23–40.

Martin, J. N. (2020) 'Diccionario del Mixteco de El Jicaral y Otros Pueblos Fronterizos de Guerrero y Oaxaca' (unpublished).

Matsukawa, Kosuke (2008) 'Reconstruction of Proto-Trique Phonemes', *University of Pennsylvania Working Papers in Linguistics*, 14: 269–81.

Maturana Russel, Patricio et al. (2019) 'Model Selection and Parameter Inference in Phylogenetics using Nested Sampling', *Systematic Biology*, 68: 219–33.

Maurits, Luke et al. (2020) 'Best Practices in Justifying Calibrations for Dating Language Families', *Journal of Language Evolution*, 5: 17–38.

McKendry, Inga (2013) '*Tonal Association, Prominence and Prosodic Structure in South-Eastern Nochixtlán Mixtec*', *PhD dissertation*, University of Edinburgh, Edinburgh.

Mendoza, Inî G. and Simon L. Peters (2020) Vocabulario del Tù'un Sàjvǐ (Mixteco) de Piedra Azul y Paredón, San Martín Peras (unpublished).

Mendoza Ruíz, Juana (2016) '*Fonología segmental y patrones tonales del Tu'un Savi de Alcozauca de Guerrero*', *MA thesis*, CIESAS, México, D.F.

Nettle, Daniel (1999) 'Is the Rate of Linguistic Change Constant?', *Lingua*, 108: 119–36.

Padgett, Erin (2017) '*Tools for Assessing Relatedness in Understudied Language Varieties: A Survey of Mixtec Varieties in Western Oaxaca, Mexico*', *MA thesis*. ND University of North Dakota, Grand Forks.

Penny, David et al. (2001) 'Mathematical Elegance with Biochemical Realism: The Covarion Model of Molecular Evolution', *Journal of Molecular Evolution*, 53: 711–23.

Pensinger, Brenda J et al. (1974). *Diccionario Mixteco: Mixteco del este de Jamiltepec, pueblo de Chayuco*. México, D.F.: Instituto Lingüístico de Verano, A.C.

Pérez Rodríguez, Verónica (2013) 'Recent Advances in Mixtec Archaeology', *Journal of Archaeological Research*, 21: 75–121.

R Core Team (2021) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Rambaut, Andrew et al. (2018) 'Posterior Summarisation in Bayesian Phylogenetics using Tracer 1.7', *Systematic Biology*, 67: 901–4.
Ramirez, Yesica (2020) Vocabulario del Mixteco de San Juan Mixtepec (unpublished).
Rensch, Calvin Ross (1976) *Comparative Otomanguean Phonology*. Bloomington: Indiana University Press.
Reyes Basurto, Griselda (2020) *Diccionario del Mixteco de Tlahuapa*, Guerrero, Mexico (unpublished).
Ross, Malcolm D. (1988) *Proto Oceanic and the Austronesian Languages of Western Melanesia*. Australian National University.
Ross, Malcolm D. (1996) 'Contact-induced Change and the Comparative Method: Cases from Papua New Guinea', in Mark Durie and Malcolm D. Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, pp. 180–217. New York: Oxford University Press.
Rueda Chaves, John Edinson (2019) '*La interacción entre el tono y el acento en el mixteco de San Jerónimo de Xayacatlán*', *PhD dissertation*, El Colegio de México, México, D.F.
Salazar, Jeremías, Saleem Alfaife, Guillem Belmar Viernes, Eric W. Campbell, Gabriel Mendoza, Olguín-Martínez Jesús, Griselda Reyes Basurto, Catherine Scanlon, Giorgia Troiani, and Alonso Vásquez-Aguilar (2021) 'Vocabulario de Yucunani Sà'án Sàvi (Mixteco)' (unpublished).
Shields, Jana (1988) 'A Syntactic Sketch of Silacayoapan Mixtec', in C. Henry Bradley and Barbara E. Hollenbach (eds.), *Studies in the Syntax of Mixtecan Languages*, vol. 1, pp. 305–449. Summer Institute of Linguistics and the University of Texas at Arlington.
Small, Priscilla C. (1990) 'A Syntactic Sketch of Coatzospan Mixtec', in *Studies in the Syntax of Mixtecan Languages*, vol. 2, pp. 261–479. Summer Institute of Linguistics and the University of Texas at Arlington.
Smith Stark, Thomas C. (1994) 'El estado actual de los estudios de las lenguas Mixtecanas y Zapotecanas', in Leonardo Manrique, Yolanda Lastra, and Doris Bartholomew (eds.), *Panorama de los estudios de las lenguas indígenas de México: Tomo II*, vol. 17, pp. 5–186. Quito: Abya-Yala.
Solano, Juvenal (2020) 'Vocabulario del Mixteco de San Sebastián del Monte' (unpublished).
Spores, Ronald and Andrew K Balkansky (2013) '*The Mixtecs of Oaxaca: Ancient Times to the Present*'. University of Oklahoma Press.
Stadler, Tanja et al. (2013) 'Birth-death Skyline Plot Reveals Temporal Changes of Epidemic spread in HIV and Hepatitis C Virus (HCV)', *Proceedings of the National Academy of Sciences*, 110: 228–33.
Stark, Sharon C., Audrey P. Johnson, and Benita González de Guzmán (2013) *Diccionario básico del mixteco de Xochapa, Guerrero*, 3rd edn. México, D.F.: Instituto Lingüístico de Verano, A.C.
———, Andrea Johnson Peterson, and Filiberto Lorenzo Cruz (1986) *Diccionario Mixteco de San Juan Colorado*. México, D.F.: Instituto Lingüístico de Verano, A.C.
Swadesh, Morris (1960) 'The Oto-Manguean Hypothesis and Macro Mixtecan', *International Journal of American Linguistics*, 26: 79–111.
Swadesh, Morris (1967) 'Lexicostatistic Classification', in Norman A. McQuown (ed.), *Handbook of Middle American Indians: Linguistics*, vol. 5, pp. 79–115. Austin: University of Texas Press.
Swanton, Michael (2021) 'Un acercamiento a la ortografía dominica del mixteco de Teposcolula: un enfoque comparativo', in Michael Swanton (ed.), *Filología Mixteca: Estudios sobre textos virreinales*, pp. 1–76. México, D.F.: Instituto de Investigaciones Filológicas UNAM.
——— and Juana Mendoza Ruíz (2021) 'Observaciones sobre la diacronía del tono en el Tu'un Savi (mixteco) de Alcozauca de Guerrero', in Francisco Arellanes and Lilián Guerrero (eds.), *Estudios lingüísticos y filológicos en lenguas indígenas mexicanas: Celebración de los 30 años del Seminario de Lenguas Indígenas*. Ciudad de México: Universidad Nacional Autónoma de México.
Terraciano, Kevin (2001) *The Mixtecs of Colonial Oaxaca: Ñudzahui History, Sixteenth through Eighteenth Centuries*. Stanford University Press.
Towne, Douglas (2011) *Gramática popular del tacuate (mixteco) de Santa María Zacatepec, Oaxaca*. México, D.F.: Instituto Lingüístico de Verano, A.C.
van Doesburg, Sebastián, Michael Swanton, Alejandro de Ávila, and Christian T. DiCanio. (2021). Flores blancas, campos quemados y quetzales: la morfología histórica mixtecana y la etimología de Chiyoyuhu (Suchixtlán), in Michael Swanton (ed.), *Filología Mixteca: Estudios sobre textos virreinales*, pp. 197–252. México, D.F.: Instituto de Investigaciones Filológicas UNAM.
Vázquez, Octavio León (2017) '*Sandhi tonal en el mixteco de Yucuquimi de Ocampo, Oaxaca*', *MA thesis*. CIESAS, México, D.F.
Willems, Matthieu et al. (2016) 'Using Hybridization Networks to Retrace the Evolution of Indo-European Languages', *BMC Evolutionary Biology*, 16(1):180. https://doi.org/10.1186/s12862-016-0745-6
Williams, Judith F., Gerardo Ojeda Morales, and Liborio Torres Benavides (2017) *Diccionario mixteco de San Andrés Yutatío, Tezoatlán, Oaxaca*. Ciudad de México: Instituto Lingüístico de Verano, A.C.
Yu, Guangchuang et al. (2017) 'ggtree: An R Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data', *Methods in Ecology and Evolution*, 8: 28–36.
Zylstra, Carol F. (1991) 'A Syntactic Sketch of Alacatlatzala Mixtec', in C. Henry Bradley and Barbara E. Hollenbach (eds.), *Studies in the Syntax of Mixtecan Languages*, vol. 3, pp. 1–177. Summer Institute of Linguistics and the University of Texas at Arlington.
——— (2012) *Gramática del Tu'un Savi (la lengua mixteca) de Alacatlatzala, Guerrero*. Instituto Lingüístico de Verano, A.C.

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Appendix

Table A.1: Varieties and sources (varieties marked with an asterisk were excluded from the final analysis due to low coverage).

No.	Variety	Branch	Sources
1	ConcepcionPapaloCuicatec*	Cuicatec	Bradley 1991
2	SanJuanTepeuxila1900Cuicatec	Cuicatec	Belmar 1902
3	SantaMariaPapaloCuicatec	Cuicatec	Anderson and Roque 1983
4	AbasoloValleMixtec	Mixtec	Galindo Sánchez 2009
5	AlacatlatzalaMixtec	Mixtec	Zylstra 2012; Anderson 2006; Zylstra 1991; Josserand 1983; Dürr 1987
6	AlcozaucaGuerreroMixtec	Mixtec	Swanton and Mendoza Ruíz 2021; Josserand 1983
7	CahuatacheMixtec	Mixtec	Dürr 1987; Josserand 1983
8	ChalcatongoHidalgoMixtec	Mixtec	Swanton and Mendoza Ruíz 2021; Macaulay 1996; Josserand 1983
9	CoicoyanlasFloresMixtec	Mixtec	Josserand 1983
10	CosoltepecMixtec	Mixtec	Josserand 1983
11	CuatzoquitengoMixtec	Mixtec	Josserand 1983
12	CuilapamGuerreroMixtec*	Mixtec	Josserand 1983
13	CuyamecalcoVillaZaragozaMixtec	Mixtec	Josserand 1983
14	ElJicaralMixtec	Mixtec	Martin 2020
15	ElRosarioMicaltepecMixtec*	Mixtec	Josserand 1983
16	GuadalupeVillahermosaMixtec	Mixtec	Josserand 1983
17	IxpantepecNievesMixtec	Mixtec	Josserand 1983
18	LaBateaMixtec	Mixtec	Ramirez 2020
19	LosTejocotesMixtec*	Mixtec	Josserand 1983
20	MagdalenaPenascoMixtec	Mixtec	Hollenbach 2017
21	MetlatonocMixtec*	Mixtec	Dürr 1987; Josserand 1983
22	PiedraAzulMixtec	Mixtec	Mendoza and Peters 2020
23	PinotepaDonLuisMixtec	Mixtec	Josserand 1983
24	SanAgustinAtenangoMixtec	Mixtec	Josserand 1983
25	SanAgustinChayucoMixtec	Mixtec	Dürr 1987; Josserand 1983; Pensinger *et al.* 1974
26	SanAgustinMonteLobosMixtec*	Mixtec	Josserand 1983
27	SanAgustinTlacotepecMixtec	Mixtec	Josserand 1983
28	SanAndresNuxinoMixtec	Mixtec	Josserand 1983
29	SanAndresYutatioMixtec	Mixtec	Williams *et al.* 2017
30	SanAntonioHuitepecMixtec	Mixtec	Josserand 1983
31	SanAntonioNduayacoMixtec*	Mixtec	Josserand 1983
32	SanAntonioTepetlapaMixtec	Mixtec	Josserand 1983
33	SanBartolomeYucuaneMixtec	Mixtec	Josserand 1983
34	SanBartoloSoyaltepecMixtec	Mixtec	Josserand 1983
35	SanCristobalJamiltepecMixtec*	Mixtec	Josserand 1983
36	SanEstebanAtatlahucaMixtec	Mixtec	Dürr 1987; Josserand 1983; Alexander 1980
37	SanFranciscoJaltepetongoMixtec*	Mixtec	Josserand 1983
38	SanFranciscolasFloresMixtec*	Mixtec	Josserand 1983
39	SanFranciscoSayultepecMixtec	Mixtec	Josserand 1983
40	SanJeronimoProgresoMixtec	Mixtec	Shields 1988; Dürr 1987; Josserand 1983
41	SanJeronimoXayacatlanMixtec	Mixtec	Dürr 1987; Josserand 1983

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Table A1. Continued

No.	Variety	Branch	Sources
42	SanJorgeNuchitaMixtec	Mixtec	Josserand 1983
43	SanJoseSinicahuaMixtec*	Mixtec	Josserand 1983
44	SanJuanCoatzospamMixtec	Mixtec	Small 1990; Dürr 1987; Josserand 1983
45	SanJuanColoradoMixtec	Mixtec	Stark *et al.* 1986; Josserand 1983
46	SanJuanDiuxiMixtec	Mixtec	Kuiper and Oram 1991; Dürr 1987; Josserand 1983
47	SanJuanMixtepecMixtec	Mixtec	Dürr 1987; Josserand 1983;
48	SanJuanNumiMixtec	Mixtec	Josserand 1983
49	SanJuanTamazolaMixtec	Mixtec	Josserand 1983
50	SanJuanTeitaMixtec	Mixtec	Josserand 1983
51	SanJuanYutaMixtec*	Mixtec	Josserand 1983
52	SanLorenzoMixtec	Mixtec	Josserand 1983
53	SanLuisMoreliaMixtec	Mixtec	Josserand 1983
54	SanMarcoslaFlorMixtec	Mixtec	Anonymous p.c.
55	SanMartinDuraznosMixtec	Mixtec	Hernández Martínez and Auderset 2020; Josserand 1983
56	SanMartinEstadoMixtec	Mixtec	Josserand 1983
57	SanMartinPerasMixtec	Mixtec	Josserand 1983
58	SanMateoPenascoMixtec*	Mixtec	Josserand 1983
59	SanMateoSindihuiMixtec	Mixtec	Josserand 1983
60	SanMiguelAchiutlaMixtec	Mixtec	Josserand 1983
61	SanMiguelAhuehuetitlanMixtec	Mixtec	Josserand 1983
62	SanMiguelAmatitlanMixtec*	Mixtec	Josserand 1983
63	SanMiguelChicahuaMixtec	Mixtec	Josserand 1983
64	SanMiguelElGrandeMixtec	Mixtec	Dürr 1987; Josserand 1983
65	SanMiguelIxtapamMixtec*	Mixtec	Josserand 1983
66	SanMiguelPiedrasMixtec	Mixtec	Josserand 1983
67	SanMiguelProgresoMixtec*	Mixtec	Josserand 1983
68	SanMiguelTlacotepecMixtec	Mixtec	Josserand 1983
69	SanPedroAtoyacMixtec	Mixtec	Josserand 1983
70	SanPedroChayucoMixtec*	Mixtec	Josserand 1983
71	SanPedroCoxcaltepecCantarosMixtec*	Mixtec	Josserand 1983
72	SanPedroJicayanMixtec	Mixtec	Josserand 1983
73	SanPedroJocotipacMixtec	Mixtec	Josserand 1983
74	SanPedroMolinosMixtec	Mixtec	Dürr 1987; Josserand 1983
75	SanPedroTidaaMixtec	Mixtec	Josserand 1983
76	SanPedroTututepecMixtec	Mixtec	Josserand 1983
77	SanPedroYosonamaMixtec	Mixtec	Gittlen 2016
78	SanPedroySanPablo Teposcolula1600Mixtec	Mixtec	Josserand 1983; de Alvarado 1962 [1593]
79	SanSebastianMonteMixtec	Mixtec	Solano 2020; Josserand 1983
80	SanSebastianTecomaxtlahuacaMixtec	Mixtec	Josserand 1983
81	SantaAnaCuauhtemocMixtec	Mixtec	Josserand 1983
82	SantaCatarinaAdequezMixtec*	Mixtec	Josserand 1983
83	SantaCatarinaEstetlaMixtec	Mixtec	Josserand 1983
84	SantaCatarinaMechoacanMixtec	Mixtec	Josserand 1983
85	SantaCatarinaTlaltempanMixtec	Mixtec	Josserand 1983
86	SantaCruzBravoMixtec	Mixtec	Josserand 1983

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Table A1. Continued

No.	Variety	Branch	Sources
87	SantaCruzItundujiaMixtec	Mixtec	Josserand 1983
88	SantaCruzNundacoMixtec*	Mixtec	Josserand 1983
89	SantaLuciaMonteverdeMixtec	Mixtec	Josserand 1983
90	SantaMariaAcatepecMixtec	Mixtec	Josserand 1983
91	SantaMariaApazcoMixtec	Mixtec	Josserand 1983
92	SantaMariaChigmecatitlanMixtec	Mixtec	Josserand 1983
93	SantaMariaHuazolotitlanMixtec	Mixtec	Josserand 1983
94	SantaMariaJicaltepecMixtec	Mixtec	Dürr 1987; Josserand 1983
95	SantaMariaNutioMixtec*	Mixtec	Josserand 1983
96	SantaMariaPenolesMixtec	Mixtec	Dürr 1987; Josserand 1983
97	SantaMariaTataltepecMixtec*	Mixtec	Josserand 1983
98	SantaMariaYolotepecMixtec	Mixtec	Josserand 1983
99	SantaMariaYucuhitiMixtec	Mixtec	Josserand 1983
100	SantaMariaYucunicocoMixtec	Mixtec	Josserand 1983
101	SantaMariaZacatepecMixtec	Mixtec	Swanton and Mendoza Ruíz 2021; Towne 2011; Josserand 1983
102	SantiagoApoalaMixtec	Mixtec	Josserand 1983
103	SantiagoCacaloxtepecMixtec	Mixtec	Dürr 1987; Josserand 1983
104	SantiagoChazumbaMixtec	Mixtec	Josserand 1983
105	SantiagoIxtaltepecMixtec	Mixtec	Josserand 1983
106	SantiagoIxtayutlaMixtec	Mixtec	Josserand 1983
107	SantiagoJamiltepecMixtec	Mixtec	Johnson 1988; Josserand 1983
108	SantiagoJuxtlahuacaMixtec	Mixtec	Josserand 1983
109	SantiagoNundicheMixtec*	Mixtec	Josserand 1983
110	SantiagoNuyooMixtec	Mixtec	Josserand 1983
111	SantiagoPinotepaNacionalMixtec	Mixtec	Josserand 1983
112	SantiagoTamazolaMixtec	Mixtec	Josserand 1983
113	SantiagoTilantongoMixtec	Mixtec	Josserand 1983
114	SantiagoTlazoyaltepecMixtec	Mixtec	Josserand 1983
115	SantiagoYosonduaMixtec	Mixtec	Farris 1992; Josserand 1983
116	SantoDomingoHuendioMixtec	Mixtec	Becerra Roldán 2015
117	SantoDomingoNundoMixtec*	Mixtec	Josserand 1983
118	SantoDomingoNuxaaMixtec	Mixtec	Josserand 1983
119	SantoDomingoTonahuixtlaMixtec	Mixtec	Josserand 1983
120	SantosReyesTepejilloMixtec	Mixtec	Josserand 1983
121	SantoTomasOcotepecMixtec	Mixtec	Alexander 1988; Dürr 1987; Josserand 1983
122	TepangoMixtec	Mixtec	Hills 1990; Dürr 1987; Josserand 1983
123	TepejilloMixtec	Mixtec	Josserand 1983
124	TlahuapaMixtec	Mixtec	Reyes Basurto 2020
125	TotoltepecGuerreroMixtec*	Mixtec	Josserand 1983
126	XayacatlanBravoMixtec	Mixtec	Josserand 1983
127	XochapaMixtec	Mixtec	Stark *et al.* 2013
128	YoloxochitlMixtec	Mixtec	Amith and Castillo García n.d.; Josserand 1983
129	YucunaniMixtec	Mixtec	Salazar *et al.* 2021
130	YucunutiBenitoJuarezMixtec	Mixtec	Josserand 1983

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Table A1. Continued

No.	Variety	Branch	Sources
131	YucuquimiOcampoMixtec	Mixtec	Swanton and Mendoza Ruíz 2021; Josserand 1983
132	YutanduchiGuerreroMixtec*	Mixtec	Josserand 1983
133	ZapotitlanPalmasMixtec	Mixtec	Josserand 1983
134	SanAndresChicahuaxtla1890Triqui	Triqui	Belmar 1897
135	SanAndresChicahuaxtlaTriqui	Triqui	Hernández Mendoza 2020; Good 1978
136	SanJuanCopalaTriqui	Triqui	Hollenbach 1992
137	SanMartinItunyosoTriqui	Triqui	DiCanio 2022b

Table A.2: Mixtec groups.

Mixtec variety	New group	Josserand area
San Pedro y San Pablo Teposcolula 1600	Unclear	Eastern Alta
Cuyamecalco Villa de Zaragoza	Group 1	Northern Alta
San Juan Coatzóspam	Group 1	Northern Alta
Santa Ana Cuauhtémoc	Group 1	Northern Alta
Santiago Ixtayutla	Group 2.1	Coast
Pinotepa de Don Luis	Group 2.2	Coast
San Antonio Tepetlapa	Group 2.2	Coast
San Francisco Sayultepec	Group 2.2	Coast
San Juan Colorado	Group 2.2	Coast
San Pedro Atoyac	Group 2.2	Coast
San Pedro Jicayán	Group 2.2	Coast
Santa María Jicaltepec	Group 2.2	Coast
Santa María Zacatepec	Group 2.2	Coast
San Agustín Chayuco	Group 2.2.1	Coast
San Lorenzo	Group 2.2.1	Coast
San Pedro Tututepec	Group 2.2.1	Coast
Santa Catarina Mechoacán	Group 2.2.1	Coast
Santa María Acatepec	Group 2.2.1	Coast
Santa María Huazolotitlán	Group 2.2.1	Coast
Santiago Jamiltepec	Group 2.2.1	Coast
Santiago Pinotepa Nacional	Group 2.2.1	Coast
San Bartolomé Yucuañe	Group 3.1	Western Alta
San Juan Teita	Group 3.1	Western Alta
Magdalena Peñasco	Group 3.2	Western Alta
San Agustín Tlacotepec	Group 3.2	Western Alta
San Miguel Achiutla	Group 3.2	Western Alta
Santa María Yolotepec	Group 3.2	Western Alta
Santo Domingo Huendio	Group 3.2	Western Alta
San Antonio Huitepec	Group 4	Eastern Alta
San Juan Tamazola	Group 4	Eastern Alta
San Mateo Sindihui	Group 4	Eastern Alta
San Miguel Piedras	Group 4	Eastern Alta
San Andrés Nuxiño	Group 4.1	Eastern Alta

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Table A2. Continued

Mixtec variety	New group	Josserand area
Santo Domingo Nuxaá	Group 4.1	Eastern Alta
Santa Catarina Estetla	Group 4.2	Eastern Alta
Santa María Peñoles	Group 4.2	Eastern Alta
Santiago Tlazoyaltepec	Group 4.2	Eastern Alta
San Juan Diuxi	Linkage 5	Eastern Alta
San Pedro Tidaá	Linkage 5	Eastern Alta
Santiago Tilantongo	Linkage 5	Eastern Alta
San Juan Ñumi	Linkage 5	Western Alta
San Pedro Yosoñama	Linkage 5	Western Alta
San Bartolo Soyaltepec	Group 5.1	Northeastern Alta
San Miguel Chicahua	Group 5.1	Northeastern Alta
San Pedro Jocotipac	Group 5.1	Northeastern Alta
Santa María Apazco	Group 5.1	Northeastern Alta
Santiago Apoala	Group 5.1	Northeastern Alta
Santiago Ixtaltepec	Group 5.1	Northeastern Alta
Santo Tomás Ocotepec	Group 6.1	Western Alta
Santa María Yucuhiti	Group 6.2.1	Western Alta
Santiago Nuyoó	Group 6.2.1	Western Alta
Chalcatongo de Hidalgo	Group 6.2.2	Western Alta
San Esteban Atatlahuca	Group 6.2.2	Western Alta
San Miguel El Grande	Group 6.2.2	Western Alta
San Pedro Molinos	Group 6.2.2	Western Alta
Santa Cruz Itundujia	Group 6.2.2	Western Alta
Santa Lucía Monteverde	Group 6.2.2	Western Alta
Santiago Yosondúa	Group 6.2.2	Western Alta
Santos Reyes Tepejillo	Group 7	Southern Baja
Abasolo del Valle	Group 7.1	Mixtepec
La Batea	Group 7.1	Mixtepec
San Juan Mixtepec	Group 7.1	Mixtepec
Santa María Yucunicoco	Group 7.1	Mixtepec
Yucunani	Group 7.1	Mixtepec
Santiago Juxtlahuaca	Group 7.1	Southern Baja
Ixpantepec Nieves	Group 7.2	Southern Baja
San Martín Duraznos	Group 7.2	Southern Baja
San Sebastián Tecomaxtlahuaca	Group 7.2	Southern Baja
Coicoyán de las Flores	Group 7.3	Southern Baja
El Jicaral	Group 7.3	Southern Baja
Piedra Azul	Group 7.3	Southern Baja
San Jerónimo Progreso	Group 7.3	Southern Baja
San Marcos de la Flor	Group 7.3	Southern Baja
San Martín Peras	Group 7.3	Southern Baja
Alacatlatzala	Group 7.3.1	Guerrero
Alcozáuca de Guerrero	Group 7.3.1	Guerrero
Cahuatache	Group 7.3.1	Guerrero
Cuatzoquitengo	Group 7.3.1	Guerrero
Santa Cruz de Bravo	Group 7.3.1	Guerrero

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023

Table A2. Continued

Mixtec variety	New group	Josserand area
Tlahuapa	Group 7.3.1	Guerrero
Xochapa	Group 7.3.1	Guerrero
Yoloxochitl	Group 7.3.1	Guerrero
Tepango	Group 7.3.1	Southern Baja
Cosoltepec	Group 7.4	Northern Baja
Santiago Chazumba	Group 7.4	Northern Baja
Santo Domingo Tonahuixtla	Group 7.4	Northern Baja
Tepejillo	Group 7.4	Northern Baja
Xayacatlán de Bravo	Group 7.4	Northern Baja
Zapotitlán Palmas	Group 7.4	Northern Baja
San Jerónimo Xayacatlán	Group 7.4.1	Northern Baja
Santa Catarina Tlaltempan	Group 7.4.1	Northern Baja
Santa María Chigmecatitlán	Group 7.4.1	Northern Baja
San Andrés Yutatío	Linkage 7.5	Tezoatlan
Santiago Cacaloxtepec	Linkage 7.5	Tezoatlan
Yucuñuti de Benito Juárez	Linkage 7.5	Tezoatlan
Yucuquimi de Ocampo	Linkage 7.5	Tezoatlan
Guadalupe Villahermosa (El Portesuelo)	Group 7.6	Central Baja
San Agustín Atenango	Group 7.6	Central Baja
San Jorge Nuchita	Group 7.6	Central Baja
San Luis Morelia	Group 7.6	Central Baja
San Sebastián del Monte	Group 7.6	Central Baja
San Martín del Estado	Group 7.6	Western Baja
San Miguel Ahuehuetitlán	Group 7.6	Western Baja
Santiago Tamazola	Group 7.6	Western Baja

Downloaded from https://academic.oup.com/jole/advance-article/doi/10.1093/jole/lzad004/7188947 by MPI Evolutionary Anthropology user on 15 August 2023