A COMBINED COMPARATIVE AND PHYLOGENETIC ANALYSIS OF THE CHAPACURAN LANGUAGE FAMILY[1]

Joshua Birchall
Radboud University Nijmegen
Museu Paraense Emílio Goeldi

Michael Dunn
Uppsala University
Max Planck Institute for Psycholinguistics

Simon J. Greenhill
ARC Centre of Excellence for the Dynamics of Language
Australian National University
Max Planck Institute for the Science of Human History

The Chapacuran language family, with three extant members and nine historically attested lects, has yet to be classified following modern standards in historical linguistics. This paper presents an internal classification of these languages by combining both the traditional comparative method (CM) and Bayesian phylogenetic inference (BPI). We identify multiple systematic sound correspondences and 285 cognate sets of basic vocabulary using the available documentation. These allow us to reconstruct a large portion of the Proto-Chapacuran phonemic inventory and identify tentative major subgroupings. The cognate sets form the input for the BPI analysis, which uses a stochastic Continuous-Time Markov Chain to model the change of these cognate sets over time. We test various models of lexical substitution and evolutionary clocks, and use ethnohistorical information and data collection dates to calibrate the resulting trees. The CM and BPI analyses produce largely congruent results, suggesting a division of the family into three different clades.

[Keywords: Chapacuran languages, historical linguistics, Bayesian phylogenetics, comparative method, Amazonian languages]

1. Introduction. The Chapacuran language family is composed of at least 12 attested lects spoken by different indigenous communities in the upper basin of the Madeira River in southwestern Amazonia, in modern-day

[1] We would like to thank Sérgio Meira, Henri Ramirez, Malcolm Ross, and two anonymous *IJAL* reviewers for comments on earlier versions of this paper. We also thank Willem Doelman for providing the original map vectors. Fieldwork for Joshua Birchall was supported by the Endangered Languages Documentation Programme and the Society of Endangered Languages (GBS). Simon Greenhill was supported under Australian Research Council's Discovery Projects funding scheme (project number DE120101954). We also acknowledge support from the Max Planck Society through Michael Dunn's Max Planck Research Group "Evolutionary Processes in Language and Culture."

[*IJAL*, vol. 82, no. 3, July 2016, pp. 255–84]

0020–7071/2016/8203–0001$10.00

Brazil and Bolivia.[2] Historical circumstances and demographic decline after contact have resulted in only three of these languages still being spoken today: Wari', Moré, and Oro Win. To date, there have been no proposals for a comprehensive internal classification of the language family based on an explicit data set and a clear methodology.

In this paper, we apply a number of techniques to the available materials on these languages that allow for the inference of historical relationships among the members of this family. First, following the traditional comparative method, we are able to identify a number of systematic sound correspondences across the languages that point to three distinct clades within the family. A number of proto-segments can be tentatively reconstructed based on these correspondences.

Second, the available documentation of these languages allows for the identification of 285 cognate sets of basic vocabulary. These cognate sets form the basis for a Bayesian phylogenetic analysis. Statistical phylogenetic methods are becoming increasingly important in historical linguistics and studies of language evolution as they provide a powerful framework for making inferences about language history (Greenhill and Gray 2009 and Greenhill, Drummond, and Gray 2010). Our Bayesian phylogenetic analysis infers a phylogeny that is consistent with the three subgroupings proposed using the preliminary sound correspondences, while also providing a more detailed sub-classification within these clades. The Bayesian phylogenetic inference technique also quantifies the quality of evidence for each aspect of the classification through a confidence statistic and gives an estimate of the amount of evolutionary change that the branch has undergone. Furthermore, we are able to integrate ethnohistorical information into our model to calibrate branches against time, in order to infer the chronology of the various dispersals of different language groups.

[2] It is worth noting that these 12 documented lects ("doculects"; see Cysouw and Good 2013) may not all constitute distinct languages in terms of mutual intelligibility. Since speaker communities do not exist for many of these languages, such tests are no longer possible to carry out. Fieldwork by Joshua Birchall has shown that Oro Win and Wari' are not mutually intelligible, but there is a degree of shared comprehension between speakers, partially due to extensive contact along the Pacaás Novos River. When Moré speakers were presented with speech from Wari' and Oro Win, they were unable to comprehend all but the most basic cognate phrases. Duran (2000) notes that Moré and Cojubim are generally mutually intelligible even though the latter maintains a number of archaic forms and does not have a number of phonological distinctions that are present in Moré. When discussing the language spoken by the Urupá at the Colônia Indígena Rodolfo Miranda, Nimuendajú (1925:159) notes that "the Jarú speak the same language, with minor differences in accent." In d'Orbigny (1839*a*:596), the Kitemoka and Napeka are considered members of the same "nation," while d'Orbigny (1839*b*:289) notes that "the tribe of the Quitémocas possesses many terms that are entirely different from those of the Chapacuras, which possibly comes from ancient relations with another distinct nation," suggesting that Kitemoka and Napeka are dialects of a single language distinct from Tapakura. These different doculects are referred to as different languages throughout this article.

Through the combined application of these different techniques to the available data, we propose a detailed internal classification of the Chapacuran family and are able to infer additional information about the dispersal of these groups.

2. Sources of language data. The collection of data on Chapacuran languages has taken place over the past 200 years by researchers working both in Bolivia and Brazil. The sources are listed below along with basic information about each group.

1. Tapakura (no ISO code)

   The Tapakura (also Chapakura or Huachi) were located along the upper and middle Blanco River near Lake Chitiopa in lowland Bolivia (d'Orbigny 1839*b*). Portions of the population were settled at the mission Concepción de Baures in 1708, with the remaining population settled in 1794 at the mission Nuestra Señora de Carmen (Métraux 1948:397 and Meireles 1989). The primary source of data is an extensive word list collected by d'Orbigny in the 1830s and published in Créqui-Montfort and Rivet (1913). Cardús (1886) also provides an additional 24 Tapakura words and phrases recorded from a terminal speaker, noting that the language was no longer being spoken during the time of his visit in 1883–84.

2. Kitemoka (no ISO code)

   The Kitemoka (also Quitemoca) were located along the upper Blanco River in Bolivia and later settled at Concepción de Chiquitos and Concepción de Baures (d'Orbigny 1839*b*). At least one speaker of Kitemoka was encountered in the Bolivian town of Concepción in the 1960s by the anthropologist Jürgen Riester, who recorded an interview with this speaker. Some of this material has been transcribed and published in Wienold (2012). The primary source of data on this language is a word list collected by d'Orbigny and published in Créqui-Montfort and Rivet (1913). There are no known Kitemoka speakers remaining.

3. Napeka (no ISO code)

   The Napeka were located along the upper Blanco River in Boliva and later settled at Concepción de Chiquitos with members of the unrelated Chiquitano tribe (d'Orbigny 1839*b*).[3] The primary source of data is a word list collected by d'Orbigny and published in Créqui-Montfort and Rivet (1913). Cardús (1886) provides an additional 48 words and phrases. A few remaining speakers of Napeka were encountered in the Bolivian town of Concepción in the 1960s by Jürgen Riester, who collected about 50 words in his field notebook and also recorded a number of untranscribed songs and interviews with them. Some of this material has been published in Wienold (2012).

[3] Note that the ethnonyms Napeka and Kitemoka end with the Chiquitano (Besïro) plural suffix *-ka* (see Métraux 1942:86). Hervás (1800:187) mistakenly considers the "Quimomocas" and "Tapacuracas" subgroups of the Chiquitano tribe. Certainly, these groups lived in close proximity with Chiquitanos on the Jesuit missions, and it is possible that members of these Chapacuran groups spoke the Chiquitano language.

4. Moré (ISO code: ite)

The Moré (also Itene or Itoreauhip) are located around the confluence of the Guaporé (Iténez) and Mamoré rivers in Bolivia. They were originally settled by Jesuits onto the missions of San Miguel, San Judas, and San Simón (Métraux 1948:398). After the expulsion of the Jesuits, the Moré returned to their traditional territory and were only recontacted in 1935 by the German ethnographer Emil Heinreich Snethlage (Snethlage 1937). The primary source of Moré data is a Ph.D. dissertation by Angenot-de Lima (2002). An extensive list of Moré words and phrases is also found in Leigue Castedo (1957). Angenot-de Lima (2002:40) lists 21 active speakers of Moré at the time it was written, but recent fieldwork by Birchall suggests no more than a dozen active speakers today.

5. Cojubim (no ISO code)

The Cojubim (also Kaw Tayo, Cautario, or Kuyubí) were first contacted by rubber tappers along the Cautário River in Brazil in the early twentieth century. The primary source of data on the language comes from a M.A. thesis by Duran (2000). A few descendants of this group now live on the Ricardo Franco indigenous post and in the nearby city of Guajará-Mirim. There are currently two elderly non-active speakers of Cojubim living in these locations.

6. Torá (ISO code: trz)

The Torá traditionally lived along the Maici River and occupied territory from the right banks of the lower Ji-Paraná River to the headwaters of the Marmelos River (Nimuendajú 1925). The Torá were first contacted in the early 1700s when they began to raid the boats of traders along the Solimões and Amazon rivers. The only available linguistic data were recorded by Curt Nimuendajú while he was working with the last speakers of the language living in settlements along the lower Marmelos River (Nimuendajú and do Valle Bentes 1923 and Nimuendajú 1925).

7. Rokorona (no ISO code)

The Rokorona (also Rocotona, Orocotona, Ocoróna, or Rogorona) are among the least known of the Chapacuran peoples. Historical documents suggest that their original homeland was located in the area between the Blanco and the Guaporé rivers in Bolivia (Birchall 2013). They were first settled at the missions of Santa Rosa and San Martín, and were later relocated westward to San Borja (Quintana 2005 and Hervás 1784). The only source of language data on Rokorona is a collection of three prayers published in Teza (1868).

8. Wanyam (no ISO code)

The Wanyam (also Pawumwa, Miguelenho, or Cabixi) were first contacted along the São Miguel River in the Brazilian state of Rondônia in the early twentieth century. The earliest source of data on the Wanyam language is Haseman (1912), followed by Nordenskiöld (1915). The primary source of Wanyam data is an extensive word list published in Ribeiro (1998), which resulted from work with the last known speaker of the language. Hanke (1975) documents the Cabixi dialect of Wanyam, which she states is nearly identical to the Miguelenho dialect that was later documented by Ribeiro.[4] There are no known remaining speakers of Wanyam.

9. Urupá (no ISO code)

The Urupá lived on the Urupá River, a left tributary of the Ji-Paraná River. After the invasion of their territory by Kawahiva people (Tupí-Guaraní), and later by colonists and rubber tappers, the Urupá were settled at the Colônia Indígena Rodolfo Miranda (Nimuendajú 1925). The only published source of data on Urupá is a word list in Nimuendajú (1925). The same word list was recorded by Nimuendajú with a different speaker in Manaus in 1927, which he sent to E. H. Snethlage in a personal correspondence (Gleice Mere, personal communication). There are no known remaining speakers of Urupá.

10. Jarú (no ISO code)

The Jarú lived on the Jarú River, another left tributary of the Ji-Paraná River. Suffering the same circumstances as their neighbors the Urupá, the Jarú were settled at the Colônia Indígena Rodolfo Miranda (Nimuendajú 1925). The only source of language data was collected in 1927 by members of the Rondon Commission (Rondon and de Faria 1948). There are no known remaining speakers of Jarú.

11. Wari' (ISO code: pav)

The Wari' (also Pakaa Nova or Orowari') live along the Pacaás Novos, Riberão, and Lage rivers, all right tributaries to the lower Mamoré River in Brazil. Their first permanent contact was with missionaries and government officials in the 1950s (Vilaça 2006). The primary source of data is a grammatical description published by Everett and Kern (1997). There is also an unpublished dictionary of the language (Kern 1996). There are currently over 2,000 Wari' speakers living in this area.

12. Oro Win (ISO code: orw)

The Oro Win (also Oro Towati' or Oroin) live along the headwaters of the Pacaás Novos River and the Igarapé Água Branca. They were first contacted by rubber tappers in the 1960s. De França (2002) presents a phonetic and phonological analysis of the language. The primary source of data comes from field notes collected by Joshua Birchall between 2009 and 2011. There are currently fewer than ten speakers of Oro Win.

[4] See Price (1983) for a discussion of the various uses of the term "cabixi" in the region. Haseman (1912:349) records the item *kabifi* with the gloss 'bad man, dangerous, savage, enemy' among the Wanyam group he identifies as the Pawumwa.

The names listed above are those that are attributed to specific linguistic data. A number of other groups have been considered speakers of Chapacuran languages by various researchers, but no data are currently available to verify this. These include the groups called Uomo, Urunamakan, Matawa, Herisobocona, and Aricoroni. Hervás (1800:250) mentions that the Herisobocona spoke a language similar to Rokorona. Data from languages called Abitana and Kumaná, collected by German ethnographer Emil Heinrich Snethlage during the 1930s and partially published in Loukotka (1963:18–19), show striking similarities to the languages Wanyam and Cojubim, respectively, that were more thoroughly documented at the end of the twentieth century. Since the approximate geographic locations of the earlier groups given in Loukotka (1963) match the current position of the later language groups, it is likely that the documentation of Abitana is of a variety of Wanyam and the documentation of Kumaná is of a variety of Cojubim. Approximate locations based on centroids of the known geographic distributions of the language groups around the time of contact are shown in figure 1.

Due to a lack of sufficient data, Rokorona and Napeka are not included in the lexical analysis in **5** below. Only Napeka is included in certain sound correspondence tables when the data are available. These languages clearly belong to the Chapacuran family, with obvious morphological and lexical correspondences. The internal position of Rokorona within the family still needs further investigation, but it is generally accepted that Napeka is the sister language of Kitemoka (see Birchall 2013 and Wienold 2012).

3. Previous classifications. The Chapacuran family was first identified by French naturalist Alcide d'Orbigny (1839*b*) during his travels through the Bolivian lowlands after the expulsion of the Jesuits, when he noticed a relationship between the Tapakura and Kitemoka languages. Chamberlain (1912) further included Wanyam and Moré in the family. Créqui-Montfort and Rivet (1913) expanded the Chapacuran family to include a number of other languages for which data had become available since d'Orbigny's travels and made a first attempt at comparing the different members of the family.

A number of more recent proposals have been made regarding the internal subgrouping of the Chapacuran family. Kaufman (1994) and Lewis, Simons, and Fennig (2013) both propose classifications using only a subset of the known members of the family and without presenting any evidence in support of their claims. Kaufman's classification shows no resemblance whatsoever to the one proposed here. Lewis, Simons, and Fennig's proposal recognizes that the Moré (Itene) and Torá languages are members of a subgroup distinct from the group that includes Oro Win and Wari', a view that is supported by the analysis presented in this paper. Ramirez (2010) presents a preliminary bipartite classification of the family based on a lexicostatical analysis.

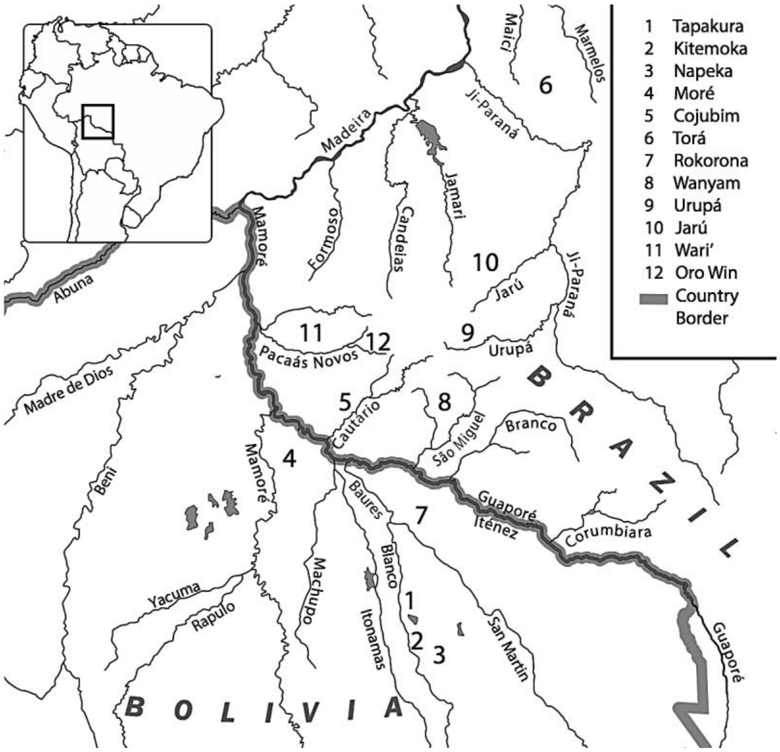

FIG. 1.—Map of the approximate location of attested Chapacuran language groups around the time of initial contact.

Unfortunately, neither the data used nor the details of the method are provided. This classificatory proposal includes all attested members of the family and shows a number of similarities to the classification presented in this paper, particularly with respect to the recognition that Wari', Urupá, Jarú, Wanyam, and Oro Win form a distinct clade within the family.

4. Sound correspondences. A number of systematic sound correspondences can be recognized based on the identifiable cognate sets in the currently available linguistic data. As an initial hypothesis, we divide the family into three branches based on the sound correspondences discussed throughout the following sections and a preliminary investigation of lexical distance (see **5** below), with the Rokorona language still unplaced within the tree. The members of each clade are as follows:

Waric:	Wanyam, Urupá, Jarú, Wari', Oro Win
Moreic:	Moré, Cojubim, Torá

Tapakuric: Tapakura, Kitemoka, Napeka
Unclassified: Rokorona

It should be noted that the majority of the data come from sources that were not compiled by professional linguists and each source varies in terms of its coverage and consistency. An attempt has been made to reconcile conflicting transcriptions of the same item across various sources and to interpret orthographic choices of earlier documents based on data from contemporary languages. When possible, reconstructed phonemes are proposed based on the principles of economy (parsimony) and naturalness of sound change. When applicable, our reconstructed phonemes are compared to the proposals given in Angenot and Angenot-de Lima (2000), the only other work to date that gives explicit proposals on the reconstruction of Proto-Chapacuran phonology.

Due to the presence of inalienably possessed nouns in some Chapacuran languages, many lexical items, especially body-part terms, occur in the original sources with a clearly identifiable possessive suffix. The citation form for body-part terms that are primarily associated with humans often occurs with the first-person plural inclusive possessive suffix *-tʃi*/*-tʃe*/*-ʃi*/*-si*, and body-part terms primarily associated with animals often occur with the third-person neuter possessive suffix *-n*/*-ɲ* or third-person masculine possessive suffix *-kon*/*-kum*. Segmentable possessive morphology rarely occurs in the citation forms for body parts in Moré and Cojubim. No attempt has been made to identify the non-possessed forms when they are not present in the original source due to a frequent occurrence of stem alternation in the non-possessed forms, such as Wari' *aran* 'its bone' (citation form) and *at* 'bone'.

4.1. On the Proto-Chapacuran phonemic inventory. A number of phonemes can be reconstructed as part of the phonemic inventory of Proto-Chapacura using the traditional comparative method. The easily reconstructible consonant segments include **p*, **t*, **k*, **ʔ*, **m*, **n*, **ɾ*, **j*, and **w*. The nature of specific fricative and affricate segments is somewhat more difficult to reconstruct. Angenot and Angenot-de Lima (2000) reconstruct an affricate series composed of $*t^s$, $*d^z$, and **tʃ*, as well as a fricative series composed of **h* and $*h^w$. The reconstruction of these segments is discussed in the context of the systematic sound correspondences presented in the following sections. To reconstruct the additional consonants identified in Angenot and Angenot-de Lima (2000), namely, $*p^w$, $*m^w$, $*m^ʔ$, $*n^ʔ$, **ɲ*, and $*w^ʔ$, is especially problematic due to inconsistencies in the available documentation pertaining to the representation of glottal stops, glides, and vowel segments. These segments tend to have a restricted distribution in the languages where they occur. A more comprehensive reconstruction of the Proto-Chapacuran consonant system remains for future research.

A typical five-vowel system with the segments **i*, **e*, **a*, **o*, and **u* can be reconstructed for Proto-Chapacura. The presence of a front rounded vowel

/ʏ/ is attested in Wari', Oro Win, and Urupá. Wari' also has a contrastive close-mid front rounded vowel /ø/ that only occurs in a few lexical items. The precise correspondences between vowels across the language family is often obscured by complex language-specific vowel harmony rules and inconsistency in transcription.[5]

4.2. Reflexes of **T*. There is a regular correspondence between /j/ in the Tapakuric languages with /z/ and /ʒ̟/ in the Moreic languages that corresponds to certain intervocalic realizations of /t/ in the Waric languages. Since all three groups of languages still present /t/ as an identity correspondence in a number of lexical items in different positions, the proto-phoneme cannot be reconstructed to **t*, even though the reflex of this proto-segment has merged with /t/ in the Waric languages.[6] Until a more comprehensive study can be carried out on the historical development of the Proto-Chapacuran consonant system, we represent this reconstructed phoneme as **T*.

Tapakuric languages have a palatal approximant /j/ as a reflex of **T* in certain intervocalic positions.[7] No cognate sets have yet been identified that show a reflex of this segment in word-initial position in this subgroup. It may be the case that the realization of /j/ in these languages results from a phonotactic restriction on vowel-vowel sequences, and thus is realized through the insertion of an approximate between vowels, similar to the process described in Wari' for the realization of diphthongs (Everett and Kern 1997:441).

The Moreic languages are unique within the Chapacuran family due to their development of voiced fricatives. There are multiple occurrences of the voiced alveolar fricative /z/ in Torá and Cojubim and the advanced voiced palatal fricative /ʒ̟/ in Moré that correspond to /t/ in the Waric languages. In Wari', the reflex of this segment is palatalized before /i/ and merges with /tʃ/ (realized as [ʃ] in certain dialects), as shown in the cognate sets SUMMER-A and FOOT-A in table 1.

[5] Cognate sets are labeled in the Supplementary Materials (see the online-only Appendix) and referred to throughout the text according to their base meaning (e.g., TOOTH) and the letter identifier for the set in which it belongs (e.g., A, B, C, etc.). All cognate sets in the tables are from the A set of cognates for that meaning unless otherwise specified.

[6] Identity correspondences for **t* can be seen intervocalically in FATHER-A and TOOTH-A, word-initially in EYE-A and WHITE-A, and in word-final coda positions for at least the Moreic and Waric languages in HEAR-A.

[7] The lexical entry for 'coati' in Tapakura is transcribed in Créqui-Montfort and Rivet (1913:150) as <kaxuel'a>. The status of <l'> in this data set is not immediately clear. Wienold (2012) compares the Reister recordings of Kitemoka and Napeka with the d'Orbigny transcriptions published in Créqui-Montfort and Rivet (1913) and finds that the use of <l'> in the d'Orbigny data corresponds to a number of different phones in the recorded data: [z] in Kitemoka and [ʎ] in Napeka for *kal'ao* 'corn' and [j] in Kitemoka for *pil'ahu* 'star'. Wienold (2012:54) considers [j] an allophone of /ʎ/, but the evidence for positing /ʎ/ as the phoneme instead of /j/, which is present in all the other Chapacuran languages, is not immediately clear. For lack of further evidence, we provisionally assume that <l'> in the d'Orbigny data is a representation of /j/.

TABLE 1
Reflexes of *T

Language	I	FOOT	SUMMER	LIE	YAM	KNEE	Reflex
Tapakura	wa**j**a	ka**j**imatʃi	?	—	?	tuk**ai**matʃi	**j/∅**
Kitemoka	wa**j**a	ka**j**imatʃe	?	—	?	tuk**ai**vatʃe	**j/∅**
Napeka	wa**j**a	kema-	?	?	?	?	**j/∅**
Moré	wa**ʒ̣**a	—	kawa**ʒ̣**i	**ʒ̣**ak	ma**ʒ̣**an	tuku**ʒ̣**im	**ʒ̣**
Cojubim	wa**z**a	—	kawa**z**i	—	ma**z**an	toko**z**im	**z**
Torá	wa**z**a	ka**z**ima	?	—	—	tuka**z**im	**z**
Wanyam	wa**t**aʔ	ka**t**imaʃiʔ	kawa**t**iʔ	—	ma**t**an	—	**t**
Urupá	wa**t**a	ka**t**imasi	?	**t**ak	ma**t**an	?	**t**
Jarú	wa**t**a	ka**t**imasi	?	—	ma**t**an	toke**t**ekilipasi	**t**
Wari'	wa**t**aʔ	ka**tʃ**imatʃiʔ	kawa**tʃ**iʔ	**t**ak	tama**t**an	—	**t/tʃ**
Oro Win	wa**t**aʔ	—	kawa**t**iʔ	—	ma**t**an	—	**t**

Angenot and Angenot-de Lima (2000:55) partially identify this correspondence and reconstruct the segment $*d^{z}$ for Proto-Chapacura. Considering that the reflex of this proto-phoneme is the voiceless plosive /t/ in the Waric languages, and the fact that this correspondence most commonly occurs in intervocalic positions, it is unlikely that the proto-phoneme was voiced, primarily due to a universal bias against devoicing in intervocalic contexts (see Garrett and Johnson 2013). Since intervocalic weakening can often lead to spirantization (as in Moreic) or elision (as in Waric), we tentatively reconstruct the proto-phoneme as a voiceless stop or affricate, most likely with an alveolar place of articulation (see Bauer 1988 and Kirchner 2001).

4.3. Reflexes of **s*. There is a regular correspondence between /s/ in the Tapakuric languages with either /ʃ/ or /tʃ/ in the Moreic languages and either /s/, /ʃ/, or /tʃ/ in the Waric languages.[8] Angenot and Angenot-de Lima (2000) reconstruct these sound correspondences as reflexes of **tʃ*.

Among the Waric languages, Urupá, Jarú, and Oro Win invariably have /s/ as a reflex of this segment. However, reflexes of this segment in Wari' and Wanyam vary depending on the dialect considered and the different sources consulted. In Wari', there is considerable variation between different idiolects and dialects, with realizations of both [ʃ] and [tʃ] for the phoneme that Everett and Kern (1997) propose as /tʃ/.

[8] The grapheme <z> occurs in a few Napeka words in an intervocalic position or immediately following <s> in Créqui-Montfort and Rivet (1913). According to the phonological analysis in Wienold (2012:58–62), [z] is not phonemic in Napeka nor is it clearly attested in modern audio recordings. However, the use of this different grapheme in the transcription could possibly reflect phonetic variation in the realization of the phoneme /s/ in Napeka, similar to the variation noted for Wanyam amd Wari'.

TABLE 2
REFLEXES OF *s

Language	FIRE	HAMMOCK	RAIN	PINEAPPLE	GOOD-B	Reflex
Tapakura	ise	?	—	—	nawasa	s
Kitemoka	ise	?	—	?	navisa	s
Napeka	ise	siat	?	?	nawiza	s
Moré	itʃeʔ	tʃat	tʃuwe	katʃin	—	tʃ
Cojubim	itʃeʔ	tʃat	—	katʃin	—	tʃ
Torá	iʃeʔ	ʃiat	—	?	—	ʃ
Wanyam	iʃeʔ	ʃijat	ʃuwiʔ	kaʃin	—	ʃ
Urupá	ise	siat	sui	?	wasap	s
Jarú	ise	siat	soi	?	asap	s
Wari'	tʃeʔ	tʃijat	tʃowiʔ	—	—	tʃ
Oro Win	se	sat	sowiʔ	kasikasin	isap	s

The most problematic case of variation is seen in Wanyam, whose last speaker passed away around the turn of the century before a detailed phonological description could be carried out. The most recently collected Wanyam data in Ribeiro (1998) show the reflex of the proto-phoneme in question realized as [ʃ], or less commonly as [s], with free variation attested in multiple examples. Hanke (1975) presents data from the Cabixi dialect of this language and also shows reflexes of this segment as either [s] or [ʃ]. However, Haseman (1912) reports this reflex to be consistently realized as [tʃ]. Loukotka (1963:19) presents data collected by E. H. Snethlage from the group referred to as Abitana, known today to be a subgroup of the Wanyam, where reflexes of this segment are transcribed as <ts>, presumably representing [ts].[9] Until a more comprehensive study of Wanyam phonology can be carried out on the existing materials, we provisionally treat the language as having the phoneme /ʃ/, while recognizing that considerable variation must have existed across speakers and/or dialects (see table 2).

Just as the language-internal variation of the reflexes of this proto-phoneme complicates the synchronic phonemic descriptions of these languages, it also complicates the reconstruction of this segment. One possibility is that there was considerable variation in the pronunciation of the proto-segment. Since all attested reflexes of this segment are either fricatives or affricates, it is most likely that the proto-segment was realized with one of these manners

[9] Gleice Mere (personal communication) points out that the original transcription of this segment in Snethlage's unpublished fieldnotes is <z>. Following standard German orthographic conventions, Loukotka interprets this segment as [ts]. There is still the possibility that this grapheme was used to represent [tʃ] in the original source, but it would be expected that the sound [ʃ] would have been transcribed as <sch> following standard German orthography.

TABLE 3
Reflexes of *tr

Language	macaw	liver	forehead	Reflex
Tapakura	**tiar**amuin	?	na**tar**atʃĩ	**t**(**V**)**r**
Kitemoka	?	?	**tiar**atʃe	**tVr**
Moré	**s**amin	**s**awan	na**t**an	**s/t**
Cojubim	**tʃ**amin	**tʃ**awan	na**tʃ**an	**tʃ**
Torá	**tr**amin	**tr**awana	na**tar**a	**tr**
Wanyam	**tr**amin	**tr**awaneʃĩʔ	na**tr**aʃi	**tr**
Urupá	**tr**amin	?	?	**tr**
Jarú	**tr**amin	**tr**awanasi	—	**tr**
Wari’	**tr**amin	**tr**awanatʃĩʔ	—	**tr**
Oro Win	**tr**amin	**tr**awanasi	na**tr**asi	**tr**

of articulation. A possible approach to this problem would be to base the reconstruction on the classification proposed in **5.2.2** below and posit the proto-form that requires the fewest sound changes. Such an analysis is appropriate here since the changes do not appear conditioned by any specific phonological contexts. A reconstruction of the phoneme **tʃ* would require a total of five changes over the tree, as would a reconstruction of the proto-phoneme as **ʃ*. The most economical reconstruction of this phoneme is **s*, with a parsimony score of four. For this reason, we tentatively propose that this set of reflexes reconstructs to **s* in Proto-Chapacura, with the possibility that there was considerable variation in its phonetic realization.

4.4. Reflexes of **tr*. Two of the Moreic languages, Moré and Cojubim, have /s/ and /tʃ/ respectively in correspondence with <tr> sequences observed in word-initial position for the rest of the languages. Different sources of language data across the family vary as to whether a vowel is represented between <t> and <r>. In sources where the vowel is represented, such as Everett and Kern (1997) for Wari’, the vowel between the two segments is identical to the vowel immediately following the segments, as in *tarawanatʃiʔ* ‘liver’. See Angenot and Angenot-de Lima (2000:61–66) for an in-depth discussion of the development of the different reflexes of **tr* across the family (see table 3).

4.5. Reflexes of **p* before V [+round]. Before the reconstructed vowels **i*, **e*, and **a* in Proto-Chapacura, the segment **p* shows a clear identity correspondence where the reflexes are /p/ in all of the daughter languages.[10]

[10] See, for example: moon-A, neck-A, kill-A, and stick-A for reflexes of **p* before **a*; star-A and flower-A for **p* before **i*; and sit-A and sleep-A for **p* before **e*. Note that in Moré, the reflex in this correspondence is /p^w/ before /e/. There is little evidence to suggest that **p^w* reconstructs across the family, as proposed in Angenot and Angenot-de Lima (2000), although

TABLE 4
REFLEXES OF *p BEFORE V[+ROUND]

Language	FAT	FISH	COATI	TAIL	LEG	SHOOT	Reflex
Tapakura	momikum	i**ʃ**uam	ka**h**ueja	—	ukutʃi	?	**h/∅**
Kitemoka	—	i**h**am	kakoja	—	ko**h**otʃi	?	**h**
Napeka	?	i**h**am	ka**h**oja	?	?	**h**urua	**h**
Moré	ma**p**om	—	ka**p**oʒaʔ	ki**p**un	**p**ok	**p**uru	**p**
Cojubim	na**p**um	—	ka**p**ozaʔ	ki**p**un	**p**ok	**p**uru	**p**
Torá	taa**h**on	**h**oam	?	?	**h**ok	-	**h**
Wanyam	wa**h**umaɲ	i**h**am	ka**h**ota	ka**h**ire	**h**okiʃi	**h**iri	**h**
Urupá	?	i**h**am	?	?	?	?	**h**
Jarú	a**h**omi	i**h**amʔ	—	?	ekisi	—	**h**
Wari'	**h**omaɲ	**hw**am	**hw**ataʔ	ka**hw**ereɲ	kokotʃiʔ	**h**YRY	**h/hw**
Oro Win	ma**f**oman	i**f**am	—	ke**f**eren	**f**okisi	**f**riʔ	**f**

However, the situation is more complex for reflexes of **p* before rounded vowels, as shown in table 4.

Within the Moreic branch, /p/ is realized as [ɸ] before the front vowels /o/ and /u/ in Moré, whereas in Cojubim, the realization is [p] in all contexts (Angenot-de Lima 2002:58 and Duran 2000). In the only available Torá data, Nimuendajú (1925) distinguishes between [h] and [p] with different graphemes, <h> and <p>. In only one instance does <pu> occur before a rounded vowel, in the possessed forms of *upik* 'head', <puye> *puje* 'my head', <puyebm> *pujem* 'your head', and so forth.[11] All instances of [h] in the Torá data occur before rounded vowels, with the exception of <hịhị> *hiʔhiʔ* 'here'. These two counterexamples in the small available corpus suggest that [h], which originally occurred only before round vowels as an allophone of /p/, has been reanalyzed as the phoneme /h/ in Torá.

In the Waric branch, Wanyam, Urupá, and Jarú have /h/ corresponding to **p* before rounded vowels. In Wari', the reflex is either /h/ or /hʷ/ depending on the dialect, with the northern groups showing the former realization and the southern groups showing the latter. In Oro Win, the reflex is /ɸ/, represented here with the grapheme <f>. In the vast majority of cases, these segments occur before rounded vowels, although they also appear before non-rounded vowels, especially in lexical items that do not appear to be reconstructible

in the cognate set SUN-A, the Tapakura form <wapuito> presents the sequence <pu>, which may represent a segment [pʷ] corresponding to Moré /pʷ/. At present, we are reluctant to reconstruct this additional proto-phoneme based on a single attested example, especially since this correspondence does not appear to be regular (cf. SIT-A, STONE-A, and HEAD-A).

[11] This is in contrast to the rest of the data and may be due to regressive vowel harmony rules triggered by the addition of the possessive suffix, similar to the process found in Oro Win (see Birchall 2014). This analysis is motivated by the fact that the non-possessed form <u̥pík> *upik* 'head' (also written as <upạ>) does not have a rounded vowel directly following the <p>.

to Proto-Chapacura (i.e., are branch-internal innovations) or those that have undergone vowel alternations as a result of vowel harmony processes and stem alternations. For example, the initial syllable of the lexeme ***po**rona* 'bow' in Oro Win historically derives from the lexeme ***pa**ro* 'peach palm', the primary wood used in the construction of bows (cf. Wanyam *paruna*, Torá *parina*, and Moré *pari* 'bow'). In a number of lexical items that have reflexes of these segments and are reconstructible to Proto-Chapacura, it is possible that there was originally a rounded vowel present that assimilated into the preceding consonant. Two likely cases for this are shown in table 4: Tapakura *ka**hu**eja* compared to Wari' ***hw**ataʔ* 'coati', and Torá ***ho**am* and Tapakura *iʃuam* compared to Oro Win *iʃam* and Wari' ***hw**am* 'fish'.

Relevant to the reconstruction of **p* before rounded vowels in the Tapakuric languages is the status of <x> in the d'Orbigny data published in Créqui-Montfort and Rivet (1913). In a brief discussion on the sounds that he perceived in Tapakura, d'Orbigny (1839*b*:289) identifies "the guttural sound of Spanish *j*" without mentioning which grapheme he used to represent this sound in his word lists. Wienold (2012) identifies two lexical items in the Riester tapes on Kitemoka that contain a <x> in the d'Orbigny data, <ixam> 'fish', which she transcribes as [ixam], and <tiaxoti> 'white', which she transcribes as [tjahutʔi]. In Reister's fieldnotes, he transcribes <iham> 'fish' for Napeka. Following Weinold's analysis for Kitemoka, we provisionally assume that the segment <x> in the d'Orbigny data represents a phonemic /h/ for both Tapakura and Kitemoka. This is reflected in the transcriptions in table 4. Note that in a number of Tapakura cognate forms, as in FAT-A and LEG-A above, the reflex of **p* has disappeared. The reflex of **p* as [ʃ] in Tapakura does not appear to be regular and only occurs in the single item in our data set, FISH-A.

4.6. Summary of sound correspondences. A summary of the sound correspondences demonstrated in the previous sections is presented in table 5.

Based on these correspondences, a number of shared innovations can be identified among different language groups. For the reflexes of **T* discussed in **4.2**, the Tapakuric languages display lenition of this segment to the point of either elision or conflation with the approximant /j/. The Moreic languages innovated a voiced fricative as a reflex of this segment, while the Waric languages merged reflexes of **T* with /t/.

In **4.3**, we posited that the Tapakuric languages retained the original phonetic value of **s*, as did the Urupá, Jarú, and Oro Win languages. Within the Waric branch, both Wari' and Wanyam changed **s* to /tʃ/ or /ʃ/, respectively. The reflexes of **s* in the Moreic languages are likely the result of two innovations. The first was the development of **ʃ* in Proto-Moreic as a reflex of Proto-Chapacuran **s*. The phonetic value of this reflex was maintained in Torá but subsequently underwent affrication in Proto-Moré-Cojubim, resulting in the /tʃ/ reflex seen in the modern languages.

TABLE 5
SUMMARY OF IDENTIFIED REGULAR SOUND CORRESPONDENCES

Language	*T	*s	*tr	*p/_____V[+round]
Tapakura	j/∅	s	t(V)r	h/∅
Kitemoka	j/∅	s	tVr	h
Napeka	j/∅	s	?	h
Moré	ʒ̹	tʃ	s	p
Cojubim	z	tʃ	tʃ	p
Torá	z	ʃ	tr	h
Wanyam	t	ʃ	tr	h
Urupá	t	s	tr	h
Jarú	t	s	tr	h
Wari'	t	tʃ	tr	h/hw
Oro Win	t	s	tr	f

As discussed in **4.4**, all languages show an identity correspondence for **tr* outside of Moré and Cojubim. In Moré, the reflex of **tr* is /s/, while in Cojubim, the reflexes of **tr* merged with /tʃ/, with both changes considered innovative.

For the reflexes of **p* before rounded vowels discussed in **4.5**, Moré and Cojubim retain the original phoneme. While Moré shows allophonic alternation with [ɸ] before rounded vowels, Cojubim always realizes this phoneme as [p]. The Tapakuric and Waric languages, as well as Torá, show parallel innovations of /h/. The data suggest that [ɸ] was initially a reflex of **p* before rounded vowels, much like in Moré, but its use was later expanded into different phonological contexts, resulting in a phonemic contrast between /p/ and /h/ or /ɸ/ in the daughter languages. Due to the parallel innovation of /h/ in members of each branch of the family, we do not consider the development of /h/ within the Waric branch as sufficient evidence to identify a subgroup within the branch that includes all languages with the exception of Oro Win, which maintained the /ɸ/ realization.

Given the identified sound correspondences and the shared innovations posited to explain their distribution, the tree in figure 2 can be proposed for the preliminary classification of the Chapacuran languages.

5. Analyses of lexical data. A data set of 285 cognate sets of basic vocabulary was compiled using the sources presented in **2** above. The list of basic vocabulary is a modified Swadesh 207-word list that combines the meanings included in Swadesh (1952;1955). Meanings that are inappropriate for lowland Amazonian societies such as 'snow' and 'ice' were excluded from consideration. A number of meanings in the original list were substituted for similar culturally appropriate concepts; 'year' was substituted with 'dry season', the local convention used to describe the passing of time, and

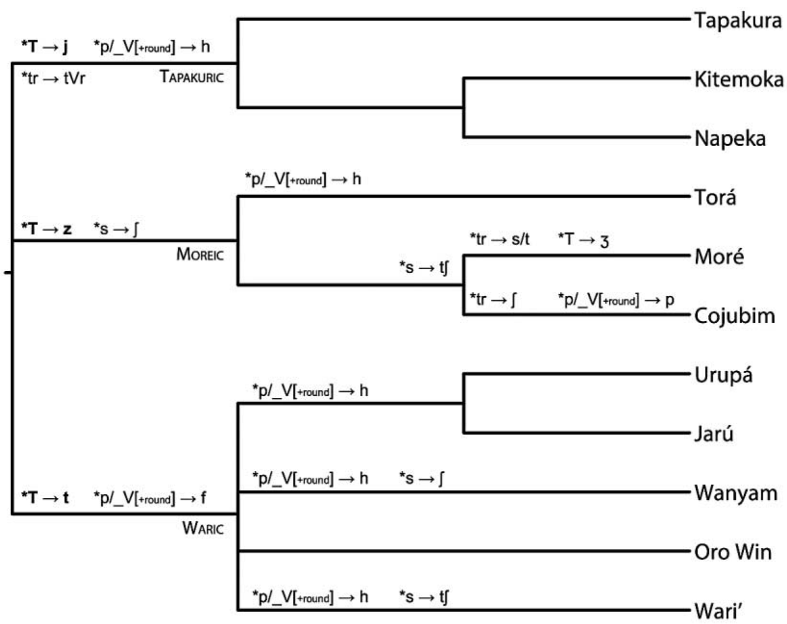

Fig. 2.—Classification of Chapacuran languages based on shared phonemic innovations and historical accounts of mutual intelligibility.

'hound' was substituted with 'peccary', a local mammal of high importance for sustenance and spiritual life. Two additional meanings were added to the list of basic vocabulary: 'afternoon' and 'brocket deer'. Meanings that were not attested in the available data, or that were attested only for members of a single subgroup of languages (most often Waric languages), were not included in the final list. The final list of basic vocabulary includes 126 meanings (for the full data set, see the Supplementary Materials presented in the online Appendix).

Cognate vocabulary items were identified based on phonetic similarity and the current state of knowledge of systematic sound correspondences in the Chapacuran language family. Variation in the orthographies adopted by the different authors of the sources were factored into the cognate judgments. A number of phonological features that were not identified consistently across sources (and sometimes even within sources) were also taken into account when identifying cognacy, such as the inclusion of the first vowel in *t*V*ɾ*V sequences (see **4.4**) and the presence of word-final glottal stops.

A degree of semantic shift was accounted for in the identification of cognate forms in the data. Take, for example, cognate set B for the meaning mother. In Urupá and Jarú, the terms *ive* and *uwe*, respectively, mean 'mother', with no additional terms documented for this meaning in the available materials. However, in Wari', the cognate form *we* refers to one's elder sister, one's father's sister, or one's father's sister's female descendants (Everett and Kern 1997:439). Oro Win, Wanyam, Cojubim, and Moré also have similar elder female relative meanings for forms cognate to Urupá and Jarú 'mother'. These forms have been included in cognate sets coded for the meaning mother since they are cognate with forms in other lects whose base meaning is indeed 'mother'. As a rule, at least one of the forms in the cognate set must present a meaning identical to the base meaning in the list of basic vocabulary, without any alternative forms. This results in some languages having cognate forms identified in multiple cognate sets for a single meaning. In the Wari' case, the term *naʔ* 'mother, mother's sister' is included as a cognate in the MOTHER-A set, while the term *we* 'elder sister, et al.' is included in the MOTHER-B set.[12]

5.1. Network analysis. To visualize the degree to which historical relationships can be recovered from our lexical data set (i.e., its phylogenetic signal), we constructed a NeighborNet network using *SplitsTree v.4.13.1* (Huson and Bryant 2006). The network is calculated from the cognate data using Hamming distances between pairs of languages. The NeighborNet visualizes the relationships in our data without enforcing the assumption of a strict family-tree-like history and reveals the conflicting signal in the data. Here, the branches are proportional to the amount of signal such that longer branches have more weight of evidence behind them. The conflicting signal, such as that caused by borrowing, is represented in figure 3 as boxes sized proportionally to the amount of conflict (Gray, Bryant, and Greenhill 2010).

The network shows the three major branches we have identified. The languages of the Tapakuric subgroup are set apart from the other languages. The languages of Waric and Moreic languages are more closely linked to each other. Lower-level groupings are found between Oro Win and Wari', between Urupá and Jarú, and between Moré and Cojubim.

To quantify the degree of conflicting signal in these data, we calculated two statistics, the δ-score (Holland et al. 2002) and the Q-residual (Gray, Bryant, and Greenhill 2010). These two statistics provide a quantitative measure of how much conflict (or reticulation) there is in the network. The mean δ-score for these languages was 0.262 (s.d. = 0.033) and the mean Q-residual score was 0.016 (s.d. = 0.004). Gray, Bryant, and Greenhill (2010) compare these two statistics across a range of linguistic and cultural data sets and find δ-scores ranging from 0.21 for very tree-like Indo-European basic vocabulary data to

[12] The semantic shifts accounted for in the data set generally follow well-attested historical pathways such as metaphor (human hair → animal feather), metonymy (animal eye → plant seed), and part–whole relationships (neck → throat). See the Supplementary Materials in the online Appendix for additional discussion of the procedure used to differentiate absent from unknown data points.

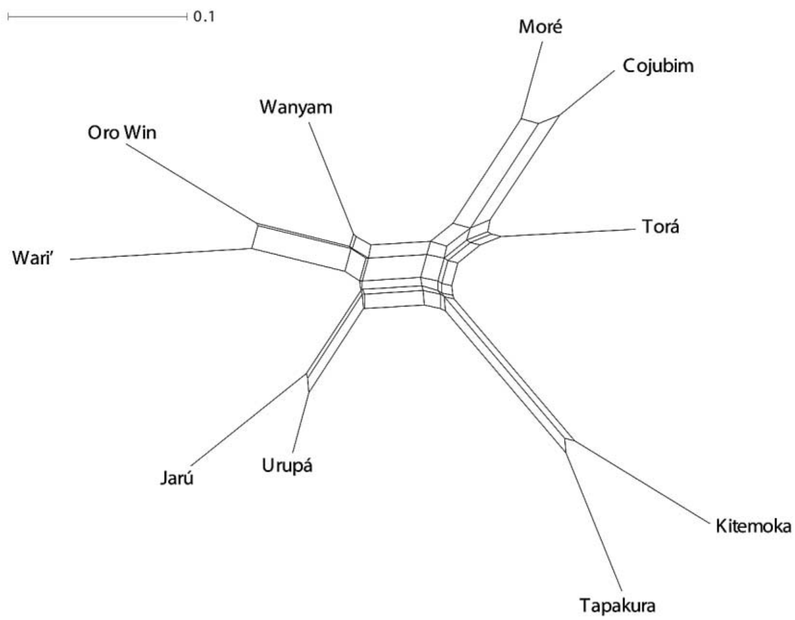

FIG. 3.—NeighborNet visualization of lexical distance.

0.41 for the highly reticulate Polynesian languages. To put the Chapacuran values in context, the combined δ-scores obtained in our analysis are very similar to those in the highly tree-like Indo-European basic vocabulary data. This finding indicates that the Chapacuran basic vocabulary is relatively stable and has undergone a minimum of borrowing between languages.

Calculating the individual δ-scores and Q-residuals for each language allows us to identify any outliers that have undergone excessive borrowing (see table 6). One language, Wanyam, shows a slightly elevated δ-score of 0.324 but has a modal Q-residual value, suggesting that Wanyam might have a more complex reticulate history than its sisters. However, all of the other individual δ-scores and Q-residuals were low, suggesting that each of the word lists—and languages—in our analyses have not undergone substantial borrowing, at least in their basic vocabulary.

5.2. Bayesian phylogenetic analyses. Next we analyzed the cognate data using Bayesian phylogenetic inference methods. Given a model of evolution (including, for instance, a tree topology and a mathematical expression of the kinds of rates of change present in the data; see below), it is possible to calculate a value expressing the likelihood that this model would have produced the observed data. In principle, it is possible to calculate the

TABLE 6
δ-Score and Q-Residual for Each Chapacuran Language

Language	δ-Score	Q-Residual
Cojubim	0.246	0.017
Jarú	0.246	0.012
Kitemoka	0.259	0.015
Moré	0.261	0.015
Oro Win	0.245	0.015
Tapakura	0.215	0.014
Torá	0.292	0.021
Urupá	0.300	0.025
Wanyam	0.324	0.014
Wari'	0.232	0.012

tree topology and other model parameter values under which the observed data are most likely, but this can be computationally intractable for large data sets. Bayesian phylogenetic inference gets around this problem using an algorithm that searches the space of possible answers and samples the tree topologies and model parameters of the proposals that produce the highest likelihood values (Gray and Atkinson 2003, Greenhill and Gray 2009, and Dunn 2014). Unlike lexicostatistics, Bayesian phylogenetic methods take into account the individual histories of cognate sets (i.e., they do not simply reduce the relationship between the cognate histories of different languages as pair-wise distance scores); they allow cognates to change at different rates in different lineages, in different words, and over time. Moreover, Bayesian phylogenetic methods allow us to quantify the support in the data for given subgrouping hypotheses in terms of a probability between 0.0 (no support) to 1.0 (complete support). Detailed descriptions of these methods in a linguistic context are provided in Dunn (2014) and Greenhill and Gray (2009), and a more technical discussion is available in Felsenstein (2004).

The first step in a Bayesian phylogenetic analysis is to model how a cognate set can change over time. A cognate set arises when it is innovated in a language and then passed down to that language's descendants. A cognate form is lost when it is no longer used in a language and is not passed down to that language's descendants. We implemented two different models of cognate change into our analysis. The first model of cognate substitution simply assumed that cognates are gained and lost at the same rate, referred to as a reversible Continuous-Time Markov Chain model (Bouckaert et al. 2012). This model does not accord very well with what is known about the historical behavior of lexical cognates, but the simplicity of the model provides a useful baseline. The second model was a Stochastic Dollo model that assumes that cognates appear once on a tree but can be lost many times (Nicholls and

Gray 2006 and Alekseyenko, Lee, and Suchard 2008). More complex models are possible; however, the relatively small set of languages and cognates means that adding additional complexity runs the risk of OVER-PARAMETERIZING the analyses and giving incorrect results (Burnham and Anderson 1998 and Greenhill and Gray 2009).

We also need a clock model to describe how rates of change vary over time. These clock models allow us to account for variation in the rates of cognate change and to estimate the age of the Chapacuran language family, given some historical calibration information. We applied two different clock models. The first clock model was a STRICT CLOCK that has a single rate of cognate replacement on all branches of the tree. The second clock model was a RELAXED CLOCK that allows rates to vary across lineages (Drummond et al. 2006). In this relaxed clock model, the rates are "autocorrelated" such that they can freely vary across the entire tree, but the rates will tend to be more similar on neighboring branches than rates on distant branches. This follows the assumption that we expect rates of cognate change to be more similar in more similar languages.

The comparison of two different models of cognate substitution and two different clock models resulted in a total of four different sets of analyses. We analyzed the lexical cognate data using *BEAST v1.8.0* (Drummond et al. 2012). The outcome of each analysis is a set of trees, sampled from the space of all possible trees. This tree sample is called the POSTERIOR PROBABILITY DISTRIBUTION. The trees in the posterior probability distribution are sampled according to their LIKELIHOOD: a measure of how well a given tree fits the data, given the model and the clock (Greenhill and Gray 2009 and Dunn 2014). The frequency of particular tree topologies and parameter values in the sample give us an estimate of how well supported the topology and parameter combinations are.

To obtain the posterior probability distribution we use a method called MARKOV-CHAIN MONTE CARLO or MCMC. MCMC starts with a random tree topology and parameter settings and randomly permutes them, retaining the new combination if it is a better fit to the observed data or, if it fits worse than the previous combination, rejecting it with a probability proportional to how much worse the fit is. The MCMC is run for a large number of these permutations—generations—and trees are sampled sporadically. Essentially, the MCMC process jumps around the space of possible trees, "climbing" into regions of high probability. The first 2,000,000 generations of the search were discarded as BURN-IN, where the trees were still heavily influenced by the initial random parameters and topologies. All four analyses were run for 20,000,000 generations, sampling every 1,000 to avoid autocorrelation between closely successive permutations. We chose 20,000,000 generations after inspection of trace plots of the parameters in the analysis showed that this was sufficient for the parameter values to move from their initial random values and stabilize.

Autocorrelation and convergence checks were carried out using *Tracer v1.5* (Rambaut and Drummond 2007), with all parameters showing high effective sample sizes (> 2,000), demonstrating that the analyses had sampled from the posterior distribution effectively.

5.2.1. Calibrations. To estimate the age of the Chapacuran language family, we incorporated known historical information about language divergence times as PRIORS in the model. These priors allow the clock models to pin down parts of the tree where we know what happened, and to extrapolate the rates and dates into parts of the tree where we have no information.

The historical information we used is of two types: tip dates based on the time of collection of the primary lexical data sources, and likely scenarios for the divergence of earlier language groups based on the appearance of distinct cultural entities in the historical record. We specified these dates in years past, rounded to the nearest decade and assuming 2010 as the calibration date for the present. The tip dates were integrated into the Bayesian analyses following Shapiro et al. (2011), while the internal calibrations were implemented as probabilistic priors on the heights of the relevant nodes. For example, Meireles (1989:123) suggests that the Moré and Cojubim once formed a contiguous population that was broken up with the arrival of the Jesuits into their territory in the 1740s. We included this information as a probability distribution of the age of the Moré and Cojubim divergence between 213 and 723 years ago, with a median of 347 years. The clock models take this information—as well as the lexical cognate data and overall rates in all the languages—into account when estimating the age of the language family.

TAPAKURA AND KITEMOKA TIP DATES

Although the primary source of data on Tapakura and Kitemoka was only published in the early twentieth century in Créqui-Montfort and Rivet (1913), the data were originally collected during d'Orbigny's trip through lowland Bolivia, which took place between 1826 and 1833.

Calibration: 180 years ago.

TAPAKURA AND KITEMOKA DIVERGENCE

The Tapakura people were first contacted in the 1580s during an expedition led by Lorenzo Suarez de Figueroa (Meireles 1989:66). By 1630, the Tapakuras were already in close contact with Jesuit missionaries and portions of the population were considered *yanacona*, a regional term meaning serf or servant of the Spaniards (Maurtua 1909 [cited in Métraux 1948]). However, the first clear mention of the Kitemoka people is not until the settlement of the mission Concepción de Chiquitos in 1707 (d'Orbigny 1839*a:*596).

Calibration: This calibration was implemented as a log-normal distribution with a mean of 250 years (in real space), a standard deviation of 0.6, and an offset of 236 years. We chose log-normal distributions for these calibrations because we have good evidence that the languages were separated by 1707, but we do not know when the initial split was. The log-normal distribution is a good match for this situation as it allows us to specify a hard upper boundary (~ 303 years ago) and assume that the actual split occurred somewhat earlier (Ho and Phillips 2009). For example, in this calibration, we place a prior distribution on the split age where 95% of the distribution is between 300 and 913 years, with the median estimate sitting around 445 years.

TORÁ AND URUPÁ TIP DATES

The data available for both Torá and Urupá were collected in 1922 by the anthropologist Curt Nimuendajú. The Urupá data were collected at Colônia Rodolfo Miranda and the Torá data were collected along settlements on the lower Marmelos River (Nimuendajú 1925).

Calibration: 90 years ago.

TORÁ DIVERGENCE

The earliest known mention of the Torá comes from a letter written by Jesuit priest Bartolomeu Rodrigues in 1714 during his travels along the lower Madeira River (Menéndez 1981). In 1716, the Torá were attacked by an expedition led by the captain general of Pará, João de Barros Guerra, in reprisal for their raids on the villages and transport ships along the lower Madeira and middle Amazonas rivers (Nimuendajú 1925).

Calibration: This calibration was implemented on the stem of the lineage leading to the Torá tip. The probability distribution was modeled using a log-normal distribution with a mean of 300 years (in real space), a standard deviation of 0.4, and an offset of 170 years. This distribution had 95% of its values between 296 years and 777 years, with a median of 447 years.

MORÉ AND COJUBIM DIVERGENCE

The Moré and the Cojubim speak similar languages but are geographically separated by the lower Guaporé River. The Moré were first contacted in the 1740s as Jesuit missions spread into the Bolivian lower Guaporé basin. No known mention is made of the Cojubim (Cautarios) as a distinct ethnic group in the historical record until an expedition led by Portuguese military engineer Ricardo Franco de Almeida Serra in 1781 (Almeida Serra 1857:422).

Calibration: This calibration was implemented as a log-normal probability distribution with a mean of 201 years (in real space), a standard deviation of 0.6, and an offset of 179 years. This gave a probability distribution where 95% of the values were between 231 years and 723 years, with a median of 347 years.

JARÚ TIP DATE

The only data available on the Jarú language were collected in 1927 at the Colônia Rodolfo Miranda by members of the Comissão Rondon (Rondon and de Faria 1948).

Calibration: 80 years ago.

5.2.2. Results. To identify the best-fitting model of cognate evolution for the Chapacuran languages, we calculated marginal likelihoods (Suchard, Weiss, and Sinsheimer 2001) and Bayes Factors (Kass and Raftery 1995). The Bayes Factors allow us to quantify the relative support in the data for a given model. The best-fitting model was the CTMC analysis with relaxed clock ($lnP(model|data) = -1209.00 \pm 0.12$), followed by the Stochastic Dollo model with a relaxed clock ($lnP(model|data) = -1212.82 \pm 0.06$). The Bayes Factor comparing the two models was 3.818, which shows substantial support for the CTMC relaxed clock analysis over the Stochastic Dollo (Kass and Raftery 1995).

These two models were then followed by their strict clock variants with the CTMC strict clock having a marginal likelihood of ($lnP(model|data) = -1215.62 \pm 0.39$), while the Stochastic Dollo model with a strict clock scored ($lnP(model|data) = -1213.62 \pm 0.02$). The Bayes Factor between these two models was 2.011.

These results show a preference for the analyses with relaxed clocks over the strict clock variants, indicating that the rates of cognate change for these languages are not clock-like but vary across branches. The relative fits of the CTMC model strongly prefer the relaxed clock: relaxed (lnP(model|data) = −1209.00) vs. strict clock (lnP(model|data) = −1215.63), Bayes Factor = 6.63. The Stochastic Dollo shows a weak preference for the relaxed clock: relaxed (lnP(model|data) = −1212.82) vs. strict clock (lnP(model|data) = −1213.62), Bayes Factor = 0.796. The Maximum Clade Credibility Tree of the relaxed clock CTMC analysis is shown in figure 4 (for the additional analyses, see the Supplementary Materials in the online Appendix). The values on the branches in figure 4 are the posterior probability of the relevant node and the branch lengths are proportional to estimated time in years.

The inferred tree topology was very stable across all four analyses with high posterior probabilities for the major subgroups. The three major branches we identified with the comparative method are all evident in these trees. First, all four analyses place the Tapakuric branch as the primary split breaking off before the rest of the Chapacuran languages (posterior probability (p) = 1.00). Second, there is strong support for the Moreic branch composed of Moré, Cojubim, and Torá (p = 0.96−1.00). Within this group, Moré and Cojubim are strongly identified as sister languages (p = 1.00).

Third, there is good support for the Waric branch with posterior probabilities ranging from 0.93−0.99. Within this clade, there is strong support for grouping Oro Win and Wari' together (p = 1.00) and for grouping Urupá and Jarú together (p = 1.00). There is uncertainty in the location of Wanyam within the Waric branch—in both the CTMC relaxed clock and strict clock analyses, Wanyam is placed as a sister to Oro Win and Wari' (p = 0.70, p = 0.88). In contrast, both of the Stochastic Dollo analyses place Wanyam as a sister to Urupá and Jarú (p = 0.78, p = 0.81).

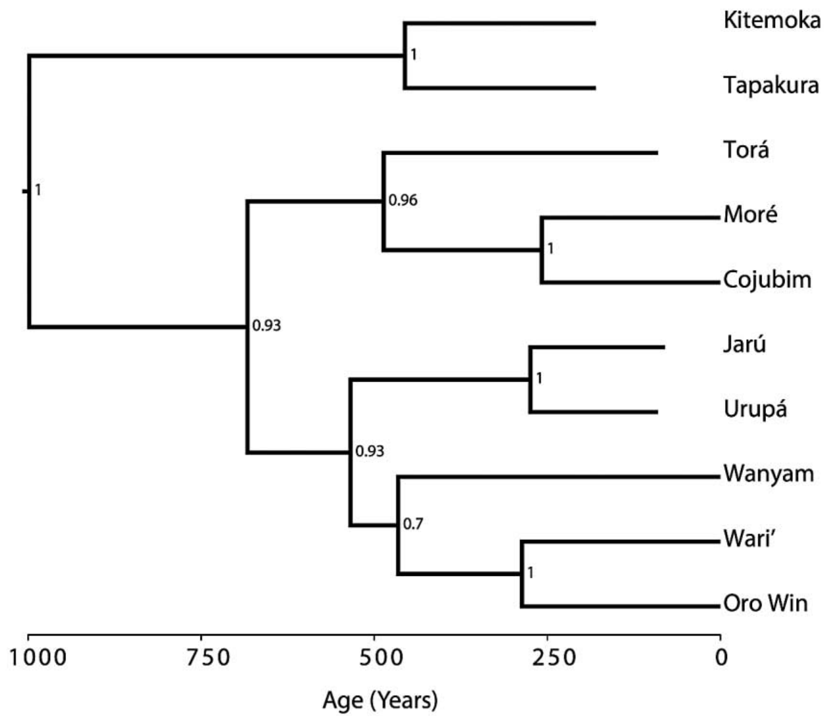

FIG. 4.—Maximum Clade Credibility Tree summary of posterior probability distribution of the relaxed clock CTMC analysis.

The dating of the Chapacuran language family varies between the models, with the CTMC relaxed clock analysis preferring a slightly younger age with a mean of 1,039 years. We are able to calculate the uncertainty in the modeled estimates of the parameters in this analysis through a credible interval (95% Highest Posterior Density (HPD) Interval = 525−1,619). The other analyses preferred slightly older root times with the Stochastic Dollo relaxed clock estimating a mean of 1,276 years (95% HPD = 793−1,831), the CTMC strict clock estimating 1,207 years (95% HPD = 783−1,692), and the Stochastic Dollo with a strict clock inferring a mean of 1,320 years (95% HPD = 899−1,812), as shown in figure 5.

To assess the impact of the calibrations on the topology, we ran the CTMC model with a relaxed clock without the Tapakura-Kitemoka and Moré-Cojubim calibrations. The posterior probabilities of these two clades were still high (both 1.0), indicating that the effect of constraining those two nodes to date the tree did not affect the subgroupings found in the tree (see the Supplementary Materials in the online Appendix).

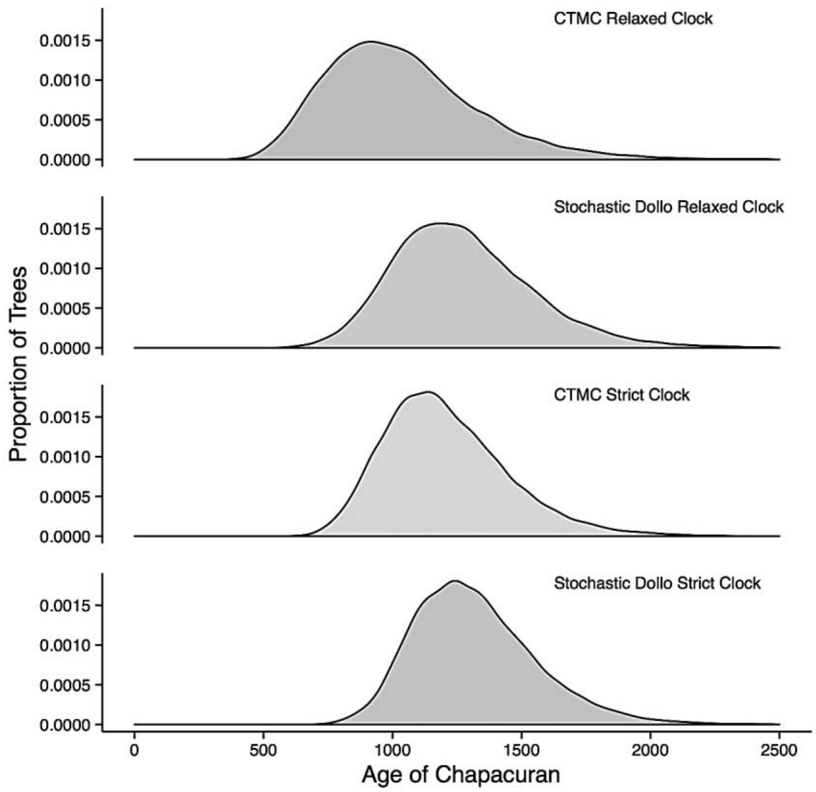

FIG. 5.—Estimated age of the Chapacuran language family based on the posterior probability distribution of all four analyses.

6. Discussion. Within the limitations of the data, the comparative method analysis has enabled us confidently to establish a number of clades on the basis of regular sound changes. These subgroups are consistent with the subgroups inferred using Bayesian phylogenetic inference. While both analyses use cognate sets, the Bayesian phylogenetic inference method models only the appearance and disappearance of reflexes of cognate forms, and does not refer to sound change at all; in contrast, the comparative method infers sound changes but ignores the aggregate history of cognate sets, including the appearance of new sets and the rates of loss of their lexical reflexes. The NeighborNet analysis of lexical distance using δ-scores and Q-residuals indicated that the lexical data contains little reticulation, which suggests that there is a low degree of unidentified borrowings in the data set. The Bayesian analysis is also consistent with the lexical distance network, which we can take as a further indication that the phylogenetic signal in the data is strong.

Bayesian phylogenetic inference produces a completely bifurcating tree and therefore makes stronger proposals than the comparative method. No regular sound changes were found that enabled us to resolve completely the branching within the Waric subgroup of languages, but the Bayesian analysis found strong evidence for grouping Jarú with Urupá and Wari' with Oro Win within this clade.

The Bayesian analysis was unable to resolve the precise placement of Wanyam within the Waric branch of languages. The CTMC relaxed clock and strict clock analyses place Wanyam as a sister to Oro Win and Wari', but the Stochastic Dollo analyses place Wanyam as a sister to Urupá and Jarú. We hypothesize that this is the result of there being unidentified loanwords in Wanyam. This hypothesis is consistent with the slightly elevated δ-score of 0.324 for Wanyam that the network analysis provided in **5.1**. One possible source for these loanwords could have been the indigenous populations speaking Tupian or Jabutian (Macro-Jê) languages within and around the basins of the São Miguel and Branco rivers in Brazil, such as the nearly extinct and poorly documented Puruborá language that was spoken by a group known to have been in contact with the Wanyam in the early twentieth century (see Galucio 2005).[13]

Phylogenetic inference using calibrated clock models gives a rooted tree as output. This gives additional evidence to allow us to determine the directionality and order of sound change. For instance, while the comparative method allows us to detect a three-way split of *j*/*z*/*t*, the branch ordering of the rooted tree shows that the change of **T* to /j/ in Tapakura-Kitemoka is independent of the change of **T* to /z/ in Proto-Moreic, since the latter group shares a more recent common ancestor with the languages showing **T* to /t/.

The resolved classification of the Chapacuran linguistic family presented in figure 4 also helps us to make inferences about the populations who speak, or once spoke, these languages. For example, the breakup of Moré and Cojubim is dated to approximately 250 years before present (CTMC Relaxed Clock estimate: mean = 266, 95% HPD = 202–345 years), consistent with the hypothesis that these groups once formed a contiguous population that split around the time of the arrival of the Jesuits into their territory (Meireles

[13] As noted by an anonymous reviewer, the unpublished fieldnotes of Erland Nordenskiöld, available at the Museum of World Culture in Gothenburg, Sweden, contain a Wanyam word list with a number of loanwords of likely Tupi-Guaranian origin, such as *jasín* 'moon'. It is hoped that through further efforts to identify lexical borrowings in the available documentation, the prehistoric contact relations between the Chapacuran peoples and their neighbors can be better understood. For a more detailed treatment of the multi-ethnic landscape of the Guaporé and Mamoré river basins, see Métraux (1948), Meireles (1989; 1991), and Ramirez (2006); for further discussion on structural convergence and lexical borrowing among the languages of this region, see Crevels and van der Voort (2008).

1989:123).[14] Additionally, the analysis suggests that Tapakura and Kitemoka once formed a contiguous speech community that diverged only around the time of the incursion of the Spaniards into lowland Bolivia in the middle sixteenth century, before their settlement into various Jesuit missions (CTMC Relaxed Clock estimate: mean = 470, 95% HPD = 300−666 years). Our analysis also suggests that the immediate ancestor of the Urupá and Jarú groups split away from the cluster of Waric languages spoken in the lower Guaporé and Mamoré basins approximately 570 years ago (CTMC Relaxed Clock estimate: mean = 571, 95% HPD = 275−915 years), presumably by crossing over the Serra dos Parecis into the Ji-Paraná and possibly Jamarí river basins. As archaeological work continues to develop in this region, it may be possible to refine the chronology of this language dispersal.

7. Conclusions. The comparative method and Bayesian phylogenetic analysis look at different aspects of linguistic data to infer the genealogy of languages. The comparative method identifies mutations in the sound system that can define splitting events in the family tree and help to establish cognate sets of vocabulary. Bayesian phylogenetic inference takes the cognates identified in the comparative method and models the innovation of new lexical items and the loss of cognate forms as a probabilistic process, producing a tree with a probabilistically defined topology and quantified branch length. Taking these two different perspectives from the data has allowed us to triangulate the most likely underlying history of the Chapacuran language family.

This analysis has established, with a fair degree of confidence, the first well-evidenced family tree for the Chapacuran languages. In addition, we have been able to propose specific and falsifiable hypotheses about the age of the root and branches of the family tree, which should inform research directions in archaeology and allied disciplines. Through this example we have shown that the comparative method and Bayesian phylogenetic inference are complementary, and that both together can and should be part of the toolbox of historical linguistics.

REFERENCES

Alekseyenko, Alexander V.; Christopher J. Lee; and Marc A. Suchard. 2008. Wagner and Dollo: A stochastic duet by composing two parsimonious solos. Systematic Biology 57:772–84.

Almeida Serra, Ricardo Franco de. 1857. Diario do Rio Madeira: Viagem que a expedição destinada à demarcação de limites fez do Rio Negro até Villa Bella, capital do governo de Matto-Grosso. Revista do Instituto Historico e Geographico Brazileiro 20:397–432.

[14] Note that the calibration discussed in **5.2.1** for the Moré and Cojubim divergence allows for either a pre-Jesuit or circum-Jesuit dispersal of these language groups, since dates between 1287 and 1779 fall within the first two standard deviations of the distribution.

Angenot, Jean-Pierre, and Geralda Angenot-de Lima. 2000. Sobre a reconstrução do Proto-chapakura. Ensaios sobre lenguas indígenas de las tierras bajas de Sudamérica: Contribuciones al 49o Congreso Internacional de Americanistas en Quito 1997, ed. Hein van der Voort and Simon van de Kerke, Lenguas Indígenas de América Latina (ILLA), vol. 1, pp. 53–70. Leiden: Research School of Asian, African and Amerindian Studies (CNWS), Universiteit Leiden.

Angenot-de Lima, Geralda. 2002. Description phonologique, grammaticale et lexicale du Moré, langue amazonienne de Bolivie et du Brésil. Ph.D. dissertation, Rijksuniversiteit te Leiden.

Bauer, Laurie. 1988. What is lenition? Journal of Linguistics 24:381–92.

Birchall, Joshua. 2013. A look at the Rokorona language. Language Typology and Universals (Sprachtypologie und Universalienforschung) 66:257–71.

———. 2014. The multi-verb benefactive construction in Wari' and Oro Win. Incremento de Valencia em Lenguas Amazónicas, ed. Francesc Queixalós, Stella Telles, and Ana Carla Bruno, pp. 115–33. Bogotá: Instituto Caro y Cuervo.

Bouckaert, Remco; Philippe Lemey; Michael Dunn; Simon J. Greenhill; Alexander V. Alekseyenko; Alexei J. Drummond; Russell D. Gray; Marc A. Suchard; and Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. Science 337:957–60.

Burnham, Kenneth P., and David R. Anderson. 1998. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. New York: Springer.

Cardús, José. 1886. Las Misiones Franciscanas entre los Infieles de Bolivia: Descripción del estado de ellas en 1883 y 1884. Barcelona: Librería de la inmaculada concepción.

Chamberlain, Alexander F. 1912. The linguistic position of the Pawumwa Indians of South America. American Anthropologist 14:632–35.

Créqui-Montfort, Georges de, and Paul Rivet. 1913. Linguistique bolivienne: La famille linguistique Čapakura. Journal de la Société des Américanistes 10:119–72.

Crevels, Mily, and Hein van der Voort. 2008. The Guaporé-Mamoré region as a linguistic area. From Linguistic Areas to Areal Linguistics, ed. Pieter Muysken, pp. 151–79. Amsterdam: John Benjamins.

Cysouw, Michael, and Jeff Good. 2013. Languoid, doculect, glossonym: Formalizing the notion 'language'. Language Documentation and Conservation 7:331–59.

de França, Maria Cristina Victorino. 2002. Aspectos da fonologia lexical da língua Oro Towati' (Oro Win). Ph.D. dissertation, Universidade Federal de Rondônia.

d'Orbigny, Alcide Dessalines. 1839*a*. L'homme américain (de l'Amérique Méridionale): Considéré sous ses rapports physiologiques et moraux. Vol. 2. Paris: Pitois-Levrault et C.

———. 1839*b*. L'homme américain (de l'Amérique Méridionale): Considéré sous ses rapports physiologiques et moraux. Vol. 4. Paris: Pitois-Levrault et C.

Drummond, Alexei J.; Simon Y. W. Ho; Matthew J. Phillips; and Andrew Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4:e88.

Drummond, Alexei J.; Marc A. Suchard; Dong Xie; and Andrew Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29:1969–73.

Dunn, Michael. 2014. Language phylogenies. Routledge Handbook of Historical Linguistics, ed. Claire Bowern and Bethwyn Evans, pp. 190–211. London and New York: Routledge.

Duran, Iris Rodrigues. 2000. Descrição fonológica e lexical do dialeto Kaw Tayo (Kujubi) da língua Moré. M.A. thesis, Universidade Federal de Rondônia.

Everett, Daniel L. 1998. Wari' (Amazonian). The Handbook of Morphology, ed. Andrew Spencer and Arnold Zwicky, pp. 690–706. Oxford: Blackwell.

Everett, Daniel L., and Barbara Kern. 1997. Wari': The Pacaás Novos Language of Western Brazil. London and New York: Routledge.

Felsenstein, Joseph. 2004. Inferring Phylogenies. Sunderland, Mass.: Sinauer Associates.

Galucio, Ana Vilacy. 2005. Puruborá: Notas etnográficas e lingüísticas recentes. Boletim do Museu Paraense Emílio Goeldi, Ciências Humanas 1:159–92.

Garrett, Andrew, and Keith Johnson. 2013. Phonetic bias in sound change. Origins of Sound Change: Approaches to Phonologization, ed. Alan C. L. Yu, pp. 51–97. Oxford: Oxford University Press.

Gray, Russell D., and Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature 426:435–39.

Gray, Russell D.; David Bryant; and Simon J. Greenhill. 2010. On the shape and fabric of human history. Philosophical Transactions of the Royal Society of London, series B, Biological Sciences 365:3923–33.

Greenhill, Simon J., and Russell D. Gray. 2009. Austronesian language phylogenies: Myths and misconceptions about Bayesian computational methods. Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust, ed. K. Alexander Adelaar and Andrew Pawley, pp. 375–97. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Greenhill, Simon J.; Alexei J. Drummond; and Russell D. Gray. 2010. How accurate and robust are the phylogenetic estimates of Austronesian language relationships? PLoS One 5, no. 3:1–6.

Hanke, Wanda. 1975. Notas lingüísticas del idioma Huanyan. Pumapunku 8:46–49.

Haseman, J. D. 1912. Some notes on the Pawumwa Indians of South America. American Anthropologist 14:333–49.

Hervás, Lorenzo. 1784. Catalogo delle Lingue Conosciute e Notizia della loro Affinità, e Diversità. Idea dell'Universo, vol. 17. Cesena: Gregorio Biasini all'Insengna di Pallade.

_______ . 1800. Lenguas y naciones Americanas. Catálogo de las lenguas de las naciones conocidas, y numeracion, division, y clases de estas, vol. 1. Madrid: Imprenta de la Administración del real arbitrio de beneficencia.

Ho, Simon Y. W., and Matthew J. Phillips. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Systematic Biology 58:367–80.

Holland, Barbara R.; K. T. Huber; A. Dress; and V. Moulton. 2002. δ plots: A tool for analyzing phylogenetic distance data. Molecular Biology and Evolution 19:2051–59.

Huson, Daniel H., and Bryant, David. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23:254–67.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. Journal of the American Statistical Association 90:773–95.

Kaufman, Terrence. 1994. The Americas. Atlas of the World's Languages, ed. Christopher Moseley and R. E. Asher, pp. 1–76. Cambridge: Cambridge University Press.

Kern, Barbara. 1996. Versão preliminar do dicionário da língua Wari' (Pacaas Novos). Ms.

Kirchner, Robert. 2001. An Effort Based Approach to Consonant Lenition. New York: Routledge.

Leigue Castedo, Luis. 1957. El Itenez Salvaje. La Paz: Ministerio de Educación.

Lewis, Paul W.; Gary F. Simons; and Charles D. Fennig, eds. 2013. Ethnologue: Languages of the World. 17th ed. Dallas: SIL International.

Loukotka, Čestmír. 1963. Documents et vocabulaires inédits de langues et de dialectes sudaméricains. Journal de la Société des Américanistes 52:7–60.

Maurtua, Victor. 1909. Juicio de límites entre el Perú y Bolivia: Prueba peruana presentada al Gobierno de la Republica Argentina. Vol. 9. Barcelona: Heinrich y Comp.

Meireles, Denise Maldi. 1989. Guardiães da Fronteira: Rio Guaporé, século XVIII. Petrópolis: Vozes.

_______ . 1991. O complexo cultural do marico: Sociedades indígenas dos rios Branco, Colorado e Mequens, afluentes do médio Guaporé. Boletim do Museu Paraense Emílio Goeldi, Série Antropologia 7:209–69.

Menéndez, Miguel. 1981. Uma contribuição para a etno-história da área Tapajós-Madeira. Revista do Museu Paulista 28:289–388.

Métraux, Alfred. 1942. The Native Tribes of Eastern Bolivia and Western Matto Grosso. Bureau of American Ethnology Bulletin 134. Washington, D.C.: Smithsonian Institution.

———. 1948. Tribes of Eastern Bolivia and the Madeira headwaters. Handbook of South American Indians, vol. 3, ed. Julian H. Steward, pp. 381–454. Washington, D.C.: Smithsonian Institution.

Nicholls, Geoff K., and Russell D. Gray. 2006. Quantifying uncertainty in a stochastic model of vocabulary evolution. Phylogenetic Methods and the Prehistory of Languages, ed. Peter Forster and Colin Renfrew, pp. 161–71. Cambridge: McDonald Institute for Archaeological Research.

Nimuendajú, Curt. 1925. As tribus do Alto Madeira. Journal de la Société des Américanistes 17:137–72.

Nimuendajú, Curt, and Euclydes Henrique do Valle Bentes. 1923. Documents sur quelques langues peu connues de l'Amazone. Journal de la Société des Américanistes 15:215–22.

Nordenskiöld, Erland. 1915. Forskningar och äventyr i Sydamerika. Stockholm: Bonnier.

Price, David. 1983. Pareci, Cabixi, Nambiquara: A case study in Western classification of native people. Journal de la Société des Américanistes 69:129–48.

Quintana, Alberto de. 2005. Descripción de las missiones de los Mojos y se da noticia del nuevo Obispo de Baltimore. Mojos, Seis Relaciones Jesuíticas: Geografía, Etnografía, Evangelización, 1670-1763, ed. Josep Barnadas and Manuel Plaza, pp. 135–59. Cochabamba: Historia Boliviana.

Rambaut, Andrew; and Alexei J. Drummond. 2007. Tracer v1.4. <http://beast.bio.ed.ac.uk/Tracer.

Ramirez, Henri. 2006. As línguas indígenas do Alto Madeira: Estatuto atual e bibliografia básica. Língua Viva 1:1–16.

———. 2010. Etnônimos e topônimos no Madeira (séculos XVI–XX): Um sem-número de equívocos. Revista Brasileira de Linguística Antropológica 2, no. 2:1–47.

Ribeiro, Michela Araújo. 1998. Léxico Português/Miguelenho. Guajará-Mirim: Universidade Federal de Rondônia.

Rondon, Cándido M. S., and João Barbosa de Faria. 1948. Glossário Geral das Tribos Silvícolas de Mato Grosso e Outras da Amazônia e do Norte do Brasil. Vol. 1. Rio de Janeiro: Comissão Rondon.

Shapiro, Beth; Simon. Y. W. Ho; Alexei J. Drummond; Marc A. Suchard; Oliver G. Pybus; and Andrew Rambaut. 2011. A Bayesian phylogenetic method to estimate unknown sequence ages. Molecular Biology and Evolution 28:879–87.

Snethlage, Emil Heinrich. 1937. Atiko y: Meine Erlebnisse bei den Indianern des Guaporé. Berlin: Klinkhardt and Biermann.

Suchard, Marc A.; Robert E. Weiss; and Janet S. Sinsheimer. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Molecular Biology and Evolution 18:1001–13.

Swadesh, Morris. 1952. Lexicostatistic dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. Proceedings of the American Philosophical Society 96:452–63.

———. 1955. Towards greater accuracy in lexicostatistic dating. IJAL 21:121–37.

Teza, Emilio. 1868. Saggi Inediti di Lingue Americane. Annali delle Università Toscane (parte prima): Scienze Neologiche 10:117–43.

Vilaça, Aparecida. 2006. Quem Somos Nós: Os Wari' Encontram os Brancos. Rio de Janeiro: Editora UFRJ.

Wienold, Christin. 2012. The Chapacuran language family: Kitemoka and Napeka. M.A. thesis, Universität Leipzig.