A Phylolinguistic Classification of the Quechua Language Family

Clasificación filolingüística de la familia lingüística quechua

Frederic Blum

Max-Planck Institute for Evolutionary Anthropology, Leipzig, Germany
https://orcid.org/0000-0002-2936-0275
frederic_blum@eva.mpg.de

Carlos Barrientos

Max-Planck Institute for Evolutionary Anthropology, Leipzig / Universität Leipzig, Germany
https://orcid.org/0000-0002-2847-2611
carlos_barrientos_ugarte@eva.mpg.de

Adriano Ingunza

Pontificia Universidad Católica del Perú (PUCP)
https://orcid.org/0009-0008-4074-4764
aingunza@pucp.edu.pe

Zoe Poirier

Max-Planck Institute for Evolutionary Anthropology, Leipzig / Universität Leipzig, Germany
https://orcid.org/0009-0007-2785-6414
zoe_poirier_maruenda@eva.mpg.de

Resumen: Presentamos una filogenia computacional para la clasificación interna de la familia lingüística quechua. Basándonos en una lista de conceptos de 150 elementos léxicos, analizamos manualmente los cognados de 39 variedades quechuas contemporáneas para calcular un árbol genealógico utilizando métodos filogenéticos bayesianos. Los resultados proporcionan más pruebas para la clasificación de las variedades individuales. También comparamos los resultados con las hipótesis existentes sobre la evolución de la familia lingüística quechua.

Palabras clave: quechua; clasificación lingüística; léxico básico; filogenética lingüística.

Abstract: We present a computational phylogeny for the internal classification of the Quechua language family. Based on a concept list of 150 lexical items, we manually analyzed data from 39 contemporaneous Quechua varieties for cognacy and computed a family tree using Bayesian phylogenetic methods. The results provide further evidence for the classification of individual varieties. We also compare the results to the existing hypotheses for the evolution of the Quechua language family.

Keywords: Quechua; language classification; basic vocabulary; linguistic phylogeny.

1. The history of the Quechua language family

1.1 Introduction

Despite the long history of classification proposals for the Quechua language family, much of its history remains disputed. While early classifications argued for deep links of Quechua with ancestral empires like Chavín (Torero 1974), other studies suggest a more recent expansion of Quechua (Adelaar 2012; Beresford-Jones and Heggarty 2012). Another point of dispute is whether the family is based on a primary branching between two groups (Adelaar 2013), which would suggest two separate expansion events, or whether the family is based on a dialect continuum with regional innovations as key process of diversification (Heggarty 2005). However, from both perspectives, the history of Quechua is closely related to our understanding of the Andean past. Recent documentation projects have substantially contributed to our understanding of individual varieties of Quechua (Heggarty 2005; Shimelman 2017; Juanatey 2020), which makes it worthwhile to reanalyze the internal classification of Quechua and potentially draw new conclusions about the history of this language family.

We will first discuss previous classifications that scholars have proposed. This entails both the heuristic classifications and studies on specific varieties that face different classifications or pose problems to existing models. Afterwards, we will present our database as well as the methodology we applied both for the annotation of cognacy and for the phylogenetic analysis. Finally, we report our results and discuss their possible interpretations. A central goal of our study is to reopen the debate about the historical correlations of Andean precolonial history and the Quechua language family. We hope to answer some of the open questions by the combination of quantitative and qualitative approaches to the evolution of language history.

1.2 Previous internal classifications for the language family

The first heuristic attempts to classify the varieties of Quechua are the works of Parker (1963) and Torero (1964), which later have been reanalyzed extensively by Cerrón-Palomino (2003). A central component common to all three of them is a split that separates the family into two main branches. However, both their arguments and the varieties grouped to each branch vary between the authors. Further, the binary split has later been put into doubt, as several varieties do not seem to fit this system of classification. We will summarize all the classifications in Table 1 at the end of Section 1.

Parker's "Genetic classification of Quechua dialects" (1963) proposes a central split between two groups at the root of the family tree, named Quechua A (QA) and Quechua B (QB) respectively. According to the author, the former clade further splits into two main groups: QA-Cuzco (QA-Cuz.) and QA-Ecuador (QA-Ec.). QA-Cuzco includes the variety from Cuzco and other varieties spoken in the Peruvian southern regions of Arequipa, Puno and Apurímac. It also incorporates the varieties from Bolivia, as well as

the Ayacucho variety related to the towns of Ayacucho, Arma, and Huancavelica. The QA-Ecuador group includes the varieties from regions in Ecuador, as well as Ucayali and Loreto in Perú. On the other side, the QB branch includes varieties from Ancash (Callejón de Huaylas; Huánuco and Huari) and Junín (Huancayo and Paucartambo). According to Parker (1963, 243), some important morphosyntactic differences that motivate the fissure between QA and QB are the following: the narrative past marker -sqa in QA, -naq in QB; the ablative marker -manta in QA and -piq in QB; amid others. As it will be shown in the following pages, this case distinction is also taken into consideration by Cerrón-Palomino (2003), but other scholars disagree with the consistency of those isoglosses.

Parallel to Parker (1963), Torero (1964) makes a thorough revision of several phonological, semantic, and grammatical features. The results lead the scholar to also propose two main branches: Quechua I (QI) and Quechua II (QII). In a rough comparison, QI is identical to Parker's QB, and QII is identical to QA. The varieties that are said to be part of the QI group are those spoken in Ancash, Huánuco, Pasco and Junín. The grouping also includes the varieties from Yauyos, Huaylas, Conchucos, Western Huayhuash (Bolognesi and Chancay regions in Lima), Middle Huayhuash (southeastern Cajatambo and northern Pasco), Eastern Huayhuash (Junin, Yauli and Tarma), Valle del Mantaro and Huánuco-Marañón in the Amazonian region, most of which were not part of Parker's study. Due to the proposed internal tripartite subgrouping into Quechua IIA, IIB and IIC, Torero's QII is a more complex branch than QA in Parker (1963). QIIA is considered by Torero (1964) as a linking step between both main branches because of its similarities with both QI and the rest of QII varieties. The author includes in this branch the varieties of Pacaraos, Lincha, and Cajamarca. QIIB varieties are the ones spoken in Lamas, Ucayali, Ecuador, and the extinct variety described by Domingo de Santo Tomás. The varieties spoken in Ayacucho, Cuzco, Bolivia, and Santiago del Estero are grouped as QIIC. Clearly the distinction between the Amazonian variety from the ones spoken in Cajamarca is an important distinction between Torero's work and the one developed by Parker (1963).

Among the most important classification criteria for Torero (1964) are the facts that the first-person possessor marker and the first-person subject marker are homophonous in QI varieties, while in QII varieties this is not the case. Regarding the internal classification of QII, Torero (1964, 473) claims that QIIA varieties do not merge *$t̠s$ and *$t\!f$, contrary to QIIB and QIIC varieties. Additionally, QIIA maintains the distinction between /s/ and /ʃ/, while QIIB and QIIC do not. QIIC also maintains /k/ and /q/ as distinct phonemes, while QIIB merges into the former.

Cerrón-Palomino (2003) analyzes a slightly different set of varieties than Torero (1964), but the main groups QI and QII remain the same. He further divides the QI varieties into three subgroups: Huaylas and Conchucos as the first group; Alto Pativilca, Alto Marañon, and Alto Huallaga as second group; and Huanca-Jauja, Yaru and Huangáscar-Topará as last group. Cerrón-Palomino (2003, 230) notes that the last

group cannot be subdivided further due to the internal variation and cross-cutting isoglosses. QII consists of the varieties spoken in the north of the QI territory up until the southeast of Colombia. In the south, it covers some parts of the province of Yauyos (Laraos, Lincha, Apurí, Choco, and Madean), and all the southern Peruvian territory including Huancavelica, Ayacucho, and Cuzco, reaching Bolivia and the north-eastern part of Argentina. Extending the subgroup QIIA, Cerrón-Palomino (2003) includes the Quechua from Ferreñafe. The complete grouping proposed by Cerrón-Palomino (2003) is visually presented in Figure 1.
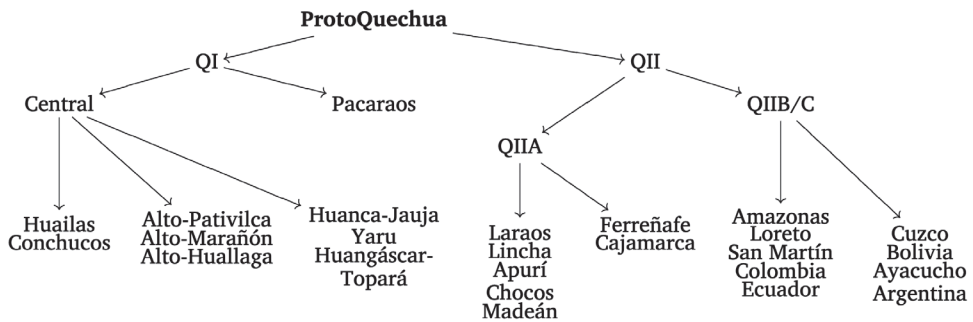

Figure 1. Classification of Quechua varieties as presented in Cerrón-Palomino (2003).

In addition to the aforementioned characteristics, the most important classification criteria for Cerrón-Palomino (2003) are the following: the contraction of *aja into *a:* (QI), the first-person object morpheme (*-ma* in QI, *-wa* in QII), the subordinating morpheme (*-r* in QI, *-ʃpa:* in QII), the locative *-tʃaw* in QI, *-pi* in QII), and the ablative (*-piqta* in QI, *-manta* in QII). An early partition inside QI between the variety of Pacaraos and the central Quechua group is proposed based on the differences in the first-person subject morpheme. The Quechua from Pacaraos would use *-y* in opposition to *-:*, the corresponding form from QI. Pacaraos Quechua has also a different subordinating morpheme, *-ʃpa:* instead of *-r*.

1.3 Reanalyzing the language family

A central reanalysis of the internal classification of Quechua has been provided by Landerman (1991). Contrary to previous studies, Landerman (1991, 238) argues that the first-person marker is not a central criterion for the internal classification of Quechua, and that many reconstructions based on this morpheme have been misleading. He also does not use the QA/II and QB/I terminology, in order to avoid the assumptions about the binary split for which the first-person maker was a central criterion. Instead, the author proposes regional clade names.

For the constitution of Central Quechua (old 'QI') as its own group, he proposes a morphological isogloss bundle which includes the ablative forms *-paq* and *-pita*, the verbal plural *-pa:ku*, grammatical vowel length, and the durative *-ja* (1991, 254). Part of this central group are different varieties from Yauyos, such as Cacra, Huangáscar, and Chocos, and the varieties of Huaylas, Conchucos, and Huallaga. He further provides evidence for the constitution of Southern Quechua (Ayacucho, Cuzco, Bolivia) and Northern Quechua (Ecuador, Colombia) as their own group. An analysis of Chachapoyas-Lamas and Cajamarca-Ferreñafe remained inconclusive with respect to their affiliation, but, contrary to some previous studies (Parker 1963), they would not form a clade together (Landerman 1991, 252).

Based on the evidence he provides, Landerman (1991) claims that the relation between the three major clades (Central Quechua, Northern Quechua, Southern Quechua) is unclear. Even though the Northern Quechua group shares many cognates with Southern Quechua (1991, 246), it is unclear whether those are actually innovations or only shared retentions. However, he does not argue in principle against the binary branching (1991, 249). This lack of clear shared innovations leads the author to be cautious with grouping Northern and Southern Quechua together, despite their superficial resemblance (1991, 268), leading to the emergence of multiple higher-level groups that are not structured hierarchically. This non-binary tree is shown in Figure 2.
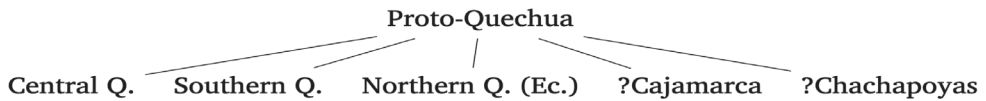
Proto-Quechua

Central Q.   Southern Q.   Northern Q. (Ec.)   ?Cajamarca   ?Chachapoyas

Figure 2. Classification of Quechua varieties following Landerman (1991).

The important morphological differences between Ecuadorian varieties and other Quechua languages led researchers to discuss the possibility of a pre-Inca Quechua ancestor in Ecuador (Torero 1974). However, this was put in doubt by other researchers who focused on the similarities between Northern and Southern varieties (Hartmann 1979). Additionally, the pre-Inca spread of Quechua by coastal traders from Chincha, hypothesized by Torero (1974, 74), was explicitly rejected by Hocquenghem (2012, 365), who instead argues that the Inca administration spread Quechua into Ecuador. More recent studies show that these differences between Ecuadorian Quechua and the other Quechua varieties are probably due to a Barbacoan substrate influence (Floyd 2022, 41).

The methodology applied by Landerman (1991) was heavily criticized by Heggarty (2005), who himself contributed an analysis based on NeighborNet (Bryant and Moulton 2004; Huson and Bryant 2005) for the classification of Quechua. In this study, he argues in favor of a dialect continuum instead of binary ramifications (Heggarty 2005, 44).

This result was established based on the distance-based lexical similarity of 150 concepts. Furthermore, Pearce and Heggarty (2011, 91) argue that there was never any binary split at the early stages of the Quechua language family. At the core of this argument are the varieties of Yauyos, which are argued to be sharing traits of both classical QI and QII. This has also led other researchers to reject a binary classification of those varieties (Shimelman 2017), or to analyze the area of Yauyos as a "dialectal microcosm" (Taylor 1984). We will discuss this special role of Yauyos in depth in Section 4.1.

Methodologically, however, it is not the case that NeighborNet can always recover the phylogenetic signal from data. In contrast to Bayesian Phylogenetics, it is only a visual, not a quantitative inference (Bryant and Moulton 2004), and thus remains highly subjective (Holman *et al*. 2011, 207). A main problem for analyzing the Quechua language family as a dialect continuum with only one expansion is that it leaves the presence of Quechua in Ecuador unexplained. Implicitly, this model includes a second expansion, namely that of some Quechua variety to Ecuador, and from there to Colombia and the north-eastern Peruvian Amazon. This would directly contradict the "single horizon expansion" (Beresford-Jones and Heggarty 2012, 74) proposed as the main driver of Quechua dispersion. However, we do hold on to the idea that expansions can originate from a certain tip within a continuum, a concept which was mentioned, but not laid out in detail by the authors.

In any case, a non-tree-like signal in the data does not always mean we are confronted with a dialect continuum. Contact situations are known to obscure the genealogical relation between languages, and taking incomplete lineage sorting into account often makes it possible to unravel the genealogical differences of superficially similar languages (Jacques and List 2019). It is up to the linguists to disentangle and interpret both the history of contact and the history of the genealogical relation in order to explain the origin of the non-tree-like signal. From our perspective, this debate is still unanswered for Quechua. The available evidence in favor of at least two expansions makes it clear that the history and diversity of Quechua cannot be answered merely by proposing a dialect continuum. Rather, the historical perspective needs to be considered. Combining tree- and wave-like expansions is likely to be a fruitful endeavor to explain such patterns.

1.4 Affiliation of individual varieties

1.4.1 Santiago del Estero

The Quechua of Santiago del Estero is commonly grouped in the subgroup QIIC, along with the varieties of Ayacucho, Cuzco, and Bolivia (Cerrón-Palomino 2003). However, the origins of the Quechua of Santiago del Estero are still a matter of investigation. Adelaar (1994) suggests that this variety might be the result of an expansion of the Quechua language caused, on one hand, by military and migration policies of the Inca Empire and, on the other hand, by the colonisation and evangelization of South American

societies driven by the Spanish administration mid 15th century. That might explain why the Quechua of Santiago del Estero shows certain similarities with some varieties from Ecuador (IIB), the south of Colombia (IIB), and Bolivia (IIC), where Quechua also arrived as a result of this second expansion. A main characteristic of QIIC is the set of glottalized and aspirated occlusive consonants. However, this feature is not shared with the variety of either Ayacucho or Santiago del Estero. Additionally, the Quechua of Santiago del Estero maintains the Proto-Quechua opposition between alveolar and alveopalatal sibilants, lost in all QIIC varieties. It does share with some varieties of QIIB the loss of the glottal fricative /h/ at the beginning of a word, a property also shared with Cajamarca and Ferreñafe varieties (Adelaar 1994, 11). The lenition of /j/ is another feature shared with northern varieties in Ecuador. The conservative treatment of the alveolar fricative /s/ is also similar to some varieties of QIIB. A unique phonological characteristic of Santiago del Estero is the loss of the semivowel */w/ between vowels. Summarizing this discussion, we can conclude that even though all scholars agree to relate Santiago del Estero Quechua with QII, the exact relations within this group remain an open debate.

1.4.2 Mixed varieties I: Yauyos and Laraos

The varieties that are spoken in the Yauyos province received dissimilar grouping by the experts throughout the decades. Torero (1964) argued for a disjunction of Alis, Vitis, Tomas, Cacra, Hongos, and Huangáscar varieties, grouping them with QI, from Lincha and Laraos, grouped as QII. As discussed in Subsection 1.2, one important argument in favor of this division is the presence of contrastive vowel length. According to this author, the homophonous expression of first-person subject and possessor and the use of the subordinator -r in the varieties of QI are the most important criteria to separate both groups of Yauyos varieties.

All in all, the most notable fact about the varieties from Yauyos is the huge diversity of cross cutting isoglosses between neighboring villages (Taylor 1984; Shimelman 2017). This leads Taylor (1984, 141) to propose the terminology "mixed dialects", but nonetheless proposes an affiliation to either QI or QII for all varieties, which is rejected for example by Shimelman (2017) in her grammar of Southern Yauyos varieties. For Heggarty (2005), the varieties of Yauyos are the main reason to abandon the binary split and to propose a dialect continuum instead.

1.4.3 Mixed varieties II: Cajamarca and Ferreñafe

The classification of the Quechua of Cajamarca has also been a matter of discussion throughout the literature. As we mentioned above, Parker (1963) grouped this variety with the Amazonian varieties, but Torero (1964) and Cerrón-Palomino (2003) considered it as part of QIIA, noting the similarity between Cajamarca and QI. Adelaar brings to the discussion the fact that the expansion of the Wari state may have spread some Quechua variety

all along the Callejón de Huaylas until Cajamarca, and relates the expansion of Wari to the expansion of Quechua (Adelaar 2012, 197). As an additional support, Adelaar (2012) mentions that there has been registered Wari ceramics all along the Callejón de Huaylas area.

According to this hypothesis, several features shared by Cajamarca Quechua with QI and QII would be explained by contact between varieties of both groups. For example, Cajamarca shares with the QI varieties the retention of the retroflex /ʈʂ/, and with QII it shares the difference between the first-person singular for the non-future tense and the first-person possessor marker. However, the picture becomes more complex when also taking into consideration the variety of Ferreñafe. According to Adelaar (2012, 207-209), the variety of Ferreñafe has several features shared with QI like the marker for first-person object *-ma* and verbal subordination *-r*. On the other hand, Cajamarca Quechua is more similar to QII varieties due to its corresponding usage of *-wa* and *-ʃpa*. In contrast to all these observations, both varieties are much closer to each other than to any other branch, sharing several morphemes unattested in other varieties.

1.4.4 Mixed varieties III: Chachapoyas-Lamas

Parker (1963, 251-252) indicates that the Lamas variety corresponds to the QA (QII) group, appending the variety to the Ecuador-Ucayali subgroup. Torero (1964) classifies the Chachapoyas variety as part of QIIA along with Cajamarca, which he states is a similar variety, while Lamas Quechua is grouped as part of QIIB (1964, 474). The former classification relies on the retention of the distinction of **tʃ* and **ʈʂ* in these varieties. The latter relies on the fusion of those two proto-phonemes in one affricate consonant. Regarding morphology, the absence of *-ku* and *-tʃik* pluralizers is proper for QIIA varieties, as well as Chachapoyas. The voicing of plosives and affricates after /n/ inside a morpheme, a process also found in the Lamas variety, is distinctive for QIIB varieties. It is interesting to notice that the fusion of **q* and **k* and the voicing of plosives after a nasal, also present in Torero (1964) but with different outcome for Chachapoyas, are the criteria used for this categorization. Contrary to this, Cerrón-Palomino (2003) classifies the Chachapoyas and Lamas varieties together as part of the QIIB subgroup, separated from Cajamarca and Ferreñafe (QIIA). Taking a completely different turn of direction, Taylor (1984, 123-124) proposes to classify both Chachapoyas and Lamas, as well as Cajamarca and Ferreñafe, as 'mixed varieties', since the classification QI and QII is argued to not fully grasp the distribution of morphemes across those varieties.

1.5 Summarising previous classifications

We summarise the proposed classifications of the different Quechua varieties in Table 1. Notable exclusions are Heggarty (2005) in general, as well as Shimelman (2017) for the varieties of Yauyos, who are in explicit disagreement with classifying the respective varieties to a binary tree of QI and QII.

Variety	Parker (1963)	Torero (1964)	Cerrón-Palomino (2003)	Others
Áncash, Huánuco, Junín, Jauja, Huanca	QB	QI	QI	
Alis, Tomas, Cacra-Hongos, Huangáscar-Topará		QI	QI	QI (Taylor 1984)
Pacaraos		QIIA	QI	QI (Taylor 1984; Adelaar 2013)
Laraos, Lincha		QIIA	QIIA	QII (Taylor 1984)
Apurí, Viñac, Madeán			QIIA	
Cajamarca, Ferreñafe	QA (EC.)	QIIA	QIIA	Mixed varieties (Taylor 1984)
Chachapoyas (Amazonas)	QA (EC.)	QIIA	QIIB	Mixed varieties (Taylor 1984)
Lamas (San Martín)	QA (EC.)	QIIB	QIIB	Mixed varieties (Taylor 1984)
Ecuador, Colombia	QA (EC.)	QIIB	QIIB	
Cuzco, Ayacucho, Bolivia, Santiago del Estero	QA (CUZ.)	QIIC	QIIC	

Table 1. Summary of proposed classifications for present-day Quechua varieties.

2. Methods

2.1 Data

We created a database of 7519 lexical items based on a 150-item concept list that was previously used in Heggarty (2005), who also published the raw data for 15 Quechua varieties.[1] Thanks to the openly available data of varieties for which we could not have gathered material otherwise, we managed to include data from 42 different varieties of Quechua in total, three of those from colonial documentations. The dataset is accessible on Github and includes all cognate sets that we analyzed.[2] The analysis is based on Version 1.0 of the dataset.[3] For our phylogenetic analysis, we included only present-day varieties. The locations of the 39 varieties are presented on a map in Figure 3. The overarching variety for each location as well as the main source is given in Table 2 in the Appendix.

---

1   https://concepticon.clld.org/contributions/Heggarty-2005-150 (21.04.2023).
2   https://github.com/lexibank/crossandean (21.04.2023).
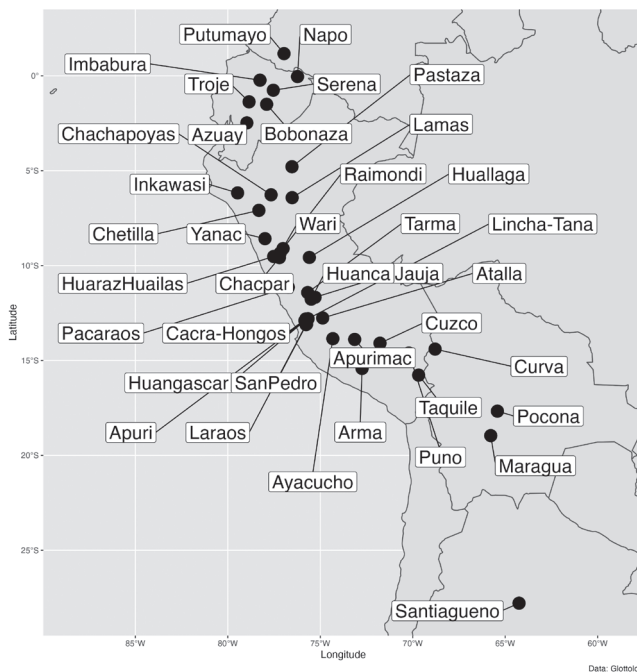3   https://doi.org/10.5281/zenodo.7491751.

Figure 3. Location of Quechua varieties included in our sample.

The database was created with the idea to be included in the Lexibank-collection in CLDF-format (List *et al.* 2022). This ensures reproducibility of our study as well as the linking of the data with other tools in linguistics, such as Glottolog (Hammarström *et al.* 2021) and Concepticon (List *et al.* 2021). After compiling the raw data, we converted the data into CLDF using CLDFBench (Forkel and List 2020). In the next step, we created automatically tagged cognates and alignments using LingPy (List and Forkel 2021) that were then manually corrected using the EDICTOR tool (List 2021). In this process, we also manually segmented morphemes that were then excluded from the annotation of cognate sets. We annotated the data for known borrowings from Aymara, Spanish, and, in a few cases, also from Uru-Chipaya. Many times, we relied on the existing literature that provides an analysis of borrowings from Aymara and Uru-Chipaya, while we relied on our own analysis for the annotation of Spanish borrowings. All those cases were excluded from the phylolinguistic analysis. For the annotation of cognate sets, we strictly followed the comparative method. Only such cases with systematic sound correspondences were treated as cognates. Even though we annotated cognacy cross-semantically, the computation is based on cognates within the same semantic concept. The cognate sets are used as input for the computation of the phylogeny and

are made publicly available as part of the Lexibank-dataset. An example of this annotation is given in Figure 4.

ID	DOCULECT	CONCEPT	FORM	TOKENS	COGIDS	SOURCE	BORROWING
6468	Chetilla Que	cloud	pu.kutay	p u k u t ai	8¹⁵	heggarty2005	
6474	Huanca Que	cloud	pukutay	p u k u t ai	8¹⁵	cerron2018	
6485	Lincha-Tana Que	cloud	pukutay	p u k u t ai	8¹⁵	shimelman2017	
6463	Ayacucho Que	cloud	puyu	p u j u	766²⁷	soto1976	
6464	Bobonaza Que	cloud	puyu	p u j u	766²⁷	orrwrisley1981	
6476	Huangascar Que	cloud	puyu	p u j u	766²⁷	shimelman2017	
6496	Santiagueno Que	cloud	puyu	p u j u	766²⁷	juanatey2021	

Figure 4. Example for the annotation of cognates for the concept 'CLOUD'.

2.2 Bayesian phylogeny

The main concept of Bayesian phylogenetics is to infer the most likely trees, given a set of alignments (i. e. cognate sets) and a phylogenetic model (Greenhill *et al.* 2020). The phylogenetic analysis has its origin in computational biology, but the parallelism between forms of direct descent in both sciences opens up the possibility of a methodological transfer (List 2016). Despite the computational approach, the main analysis is left to the linguists in the form of the annotation of cognates. Further, phylogenetic models are designed to quantitatively distinguish between shared innovations and retentions (Greenhill and Gray 2012, 525). This is achieved by computationally rooting the tree and inferring the original state of any concept set given all other concept sets. The result of this computation is not a single tree, but rather a Bayesian posterior distribution of many thousands of trees. This makes it possible to directly quantify any uncertainty involved in the inference of a summary tree. The posterior distribution of any branch will reflect the amount of evidence that it has in the model. Bayesian phylogenetics thus does not force the data onto a binary tree, but will always reflect non-tree-like signals in its posterior distribution. This can be used for evaluating claims of certain subgroupings within a language family (Greenhill *et al.* 2020, 230).

This methodology has been used to study the history of various language families such as Austronesian (Gray *et al.* 2009), Indo-European (Bouckaert *et al.* 2012; Chang *et al.* 2015), and Sino-Tibetan (Sagart *et al.* 2019). Since the rise of the phylolinguistic method it has also been widely applied to South American language families (Michael and Chousou-Polydouri 2019), for example to Chapacuran (Birchall *et al.* 2016), Panoan (Zariquiey *et al.* 2017), Tukanoan (Chacon and List 2016), and Tupí-Guaraní (Michael *et al.* 2015; Gerardi and Reichert 2021).

3. Results

3.1 Constraints on Generality

Before interpreting our results, it is important to constrain their scope of generalizability (Simons *et al.* 2017). While we do think that our results provide evidence for some hypotheses, the endeavor of re-classifying the family can only be successful when combined with qualitative results from the comparative method as well as both lexical and morphological reconstructions. In order to not reproduce any implicit assumptions, we avoid the terms 'QI' and 'QII' and use the regional names of the clades. We do not see our analysis as the single true interpretation of the data. Rather, we contribute a quantitative perspective on possibly contradictory lexical evidence, which can arise, for example, through unidentified shared retentions. We want to reopen the debate about the internal classification of the Quechua language family, rather than to put an end to it. We explicitly welcome a revision of the publicly available cognate judgments and call for further collaborations on both qualitative and quantitative studies in order to improve our knowledge on the history of the Quechua language family.

3.2 Phylogenetic tree

We implemented the model using the BEAST2 software (Bouckaert *et al.* 2014). We ran the model for 10,000,000 iterations using a BDSKY Contemporary BDS tree prior and stored every 10,000 trees to create a Maximal Clade Credibility Tree. For the visualization of the plots, we used the ggtree package (Yu 2020) and a script template of Sagart *et al.* (2019) that we modified for our purposes. We uploaded the model and all scripts that have been used and are not already part of the CLDF dataset to OSF.[4]

The phylogeny resulting from the analysis of our data is presented in Figure 5. The numbers represent the posterior value of the clade, that is, the relative number of trees sampled from the posterior distribution that include the respective clade. This translates directly to the trust we have in the corresponding branching. We analyzed all branches with a support of at least 75 % to form a higher-level group. However, all posterior values are reported explicitly to enable further discussion about the groupings.

The Quechua varieties in our sample split into two stable higher-level groups: Southern and Northern Quechua, and the Central Quechua varieties. The first group, Southern and Northern Quechua, is further split into three main clades. The first clade is formed by the Southern Quechua varieties, such as the varieties from Ayacucho, Cuzco, Bolivia, and Santiago del Estero. The second clade is formed by the varieties from Ecuador as well as the Peruvian Pastaza variety. The third clade is formed by Chachapoyas and Lamas. There is no strong evidence for a hierarchical branching between those three clades, but a tendency of Chachapoyas-Lamas to group with the Northern group. The

---

4   https://doi.org/10.17605/OSF.IO/U7JZ5.

second higher-level group consists of the varieties of Central Peru, including Jauja-Huanca, Cajamarca and Ferreñafe, Pacaraos, Tarma, Huallaga, and the varieties of Ancash. While there is evidence for some hierarchic subgroupings, it is mostly individual varieties branching off from the rest without giving rise to any larger groups. Examples for those hierarchical structures are the branches of Laraos, Jauja-Huanca, Cajamarca and Ferreñafe as well as Tarma and Pacaraos, that split off in this subsequent order.
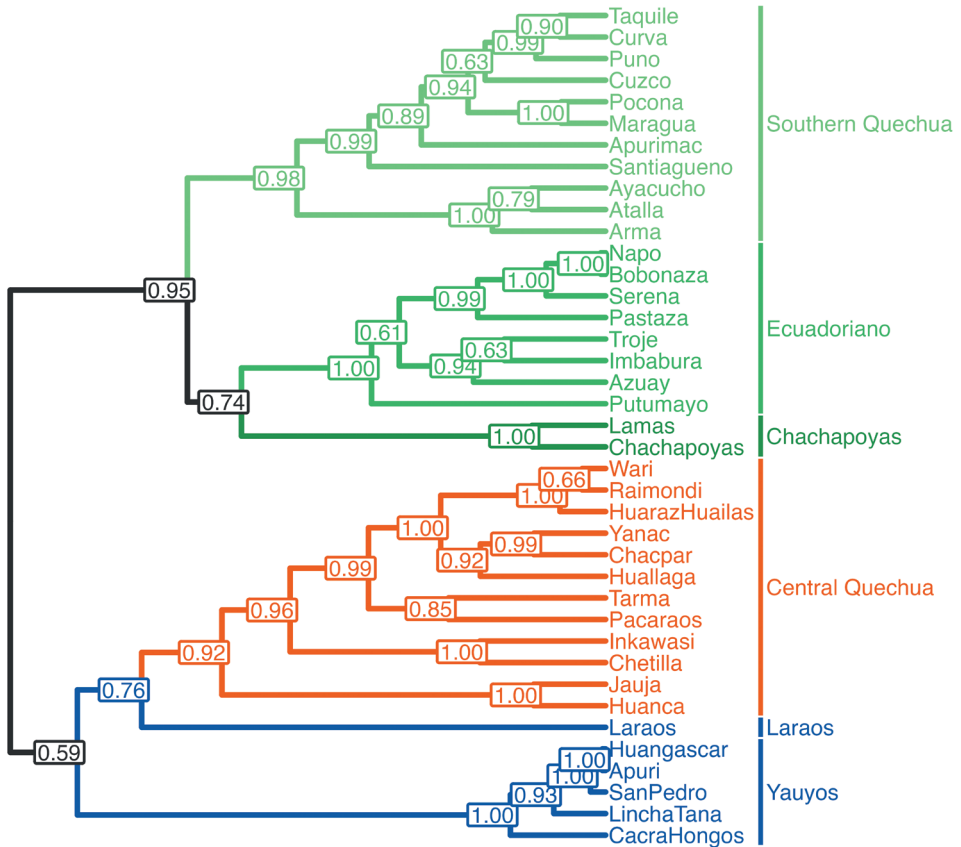
Figure 5. Phylogenetic tree for the classification with posterior values for all branches.

There is also a third group, which consists of the varieties of Central Yauyos. However, the low posterior value (59%) corresponding to their branch shows that they cannot be clearly assigned to either group, and possibly form a group of their own. However, most studies based on more traditional analysis did not claim that Yauyos forms its own clade within the family, but grouped them explicitly as either QI or QII depending on

the variety h despite cross-cutting isoglosses and regional innovations, which are seen as traits of contact rather than inheritance (Taylor 1984, 141). We will build upon those discussions in the next section.

3.3 The neighbors of Yauyos

For further investigation of these non-tree-like signals we created a NeighborNet (Bryant and Moulton 2004; Huson and Bryant 2005). Figure 6 shows that the mixed signal arises from the similarity between the varieties from Yauyos and varieties from Atalla (Huancavelica), Ayacucho, and Arma. Further, the network confirms a rather strong split between the varieties of Southern and Northern Quechua varieties, and Central Quechua. This split does not show the typical clustering of a dialect continuum across the whole language family. Instead, the network shows two closely related clusters to the left (Northern and Southern Quechua), as well as one cluster (Central Quechua) that internally looks like a dialect continuum in steplike, regional arrangement. The closest point between both higher-level groups is at the Ayacucho/Huancavelica and Yauyos varieties. However, it is important to highlight again that a NeighborNet does not represent a phylogenetic classification. It can, however, aid to find the origin of non-tree-like signals in data like the present. This is the case with the varieties from Yauyos. The closest neighbor between those groups are the varieties of Ayacucho/Huancavelica, which makes it very probable that this proximity is what causes the low posterior value for relating the Yauyos varieties to Southern Quechua.
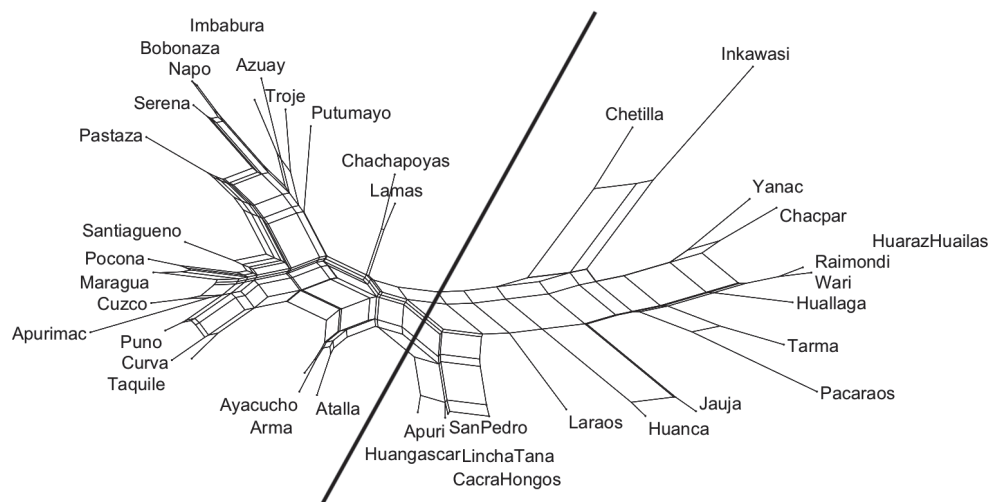

Figure 6. NeighborNet for the present-day Quechua varieties with the proposed split between Central Quechua on one side, and Southern and Northern Quechua on the other side.

It has been argued for both Ayacucho and Huancavelica that there is evidence for the historic presence of a different Quechua variety closely related to those of Central Quechua in these areas (Itier 2016; Pearce and Heggarty 2011). In the case of Ayacucho, for example, this is made evident by Quechua toponymy which is agued to be related to an earlier expansion of Central Quechua. For the Huancavelica data, we have attested variation of lexical terms for many concepts, including the terms from both Central and Southern Quechua. There are two possible explanations for this. The first possibility is that given the earlier presence of Central Quechua in this area, Huancavelica was part of the first wave of lexical innovations. However, a second expansion to this area led to the co-presence of terms related to either of the major Quechua groups. The second explanation would posit variation in the proto-stage of Quechua, of which Huancavelica would have retained the varying terms. The similarity to Yauyos would then be facilitated through a certain geographic proximity and shared retentions. Given the toponymic and anthropological evidence of the late arrival of Southern Quechua to the area of Huancavelica and Ayacucho, we think that the first explanation offers a better case for the linguistic data available. This opens up the possibility that a similar process could have taken place in the Yauyos area. Given the different social and geographic circumstances, this second wave would then have had quite a different outcome, which is why many Yauyos varieties still resemble its Central Quechua origin.

The position of Yauyos as its own clade is even more doubtful if we turn to morphological evidence. Part of Landerman's (1991) analysis was the definition of an isogloss bundle that defined Central Quechua. In our morphological database that is also annotated for cognacy, we found that all varieties of Yauyos, except Laraos, share an ablative that is related to the forms found in other varieties of Central Quechua (Blum *et al*. 2021). Further, at least Cacra-Hongos and Laraos also share the locative form /-tʃaw/ with Huallaga, Ancash, Pacaraos, and Tarma. Tarma, Ferreñafe, and the Yauyos varieties also share the durative form /-ya/. Additionally, the posterior value could be interpreted as weak evidence in favour of grouping the Yauyos varieties as part of the Central Quechua clade. This fits well with the analysis of various scholars of grouping some varieties of Yauyos along those Central Quechua varieties (Torero 1964; Taylor 1984; Cerrón-Palomino 2003). All in all, we believe that this constitutes evidence that Yauyos does not form its own higher-level group. Rather, the position of Yauyos in the tree seems like an exemplary case of incomplete lineage sorting (Jacques and List 2019). This opens up two hypotheses with various possible building stones, which do not necessarily depend on each other or are mutually exclusive:

Hypothesis 1. From the perspective of lexical phylogeny, the Yauyos varieties are part of Central Quechua.
   a) The similarity between the varieties of Yauyos and Huancavelica/Ayacucho is partially due to a substrate influence of an earlier Central Quechua variety in the area of Huancavelica/Ayacucho.
   b) Similar to the process in Huancavelica/Ayacucho, the arrival of varieties close to Southern Quechua to Yauyos is a recent process. Before, Central Quechua varieties were spoken in this area.

Hypothesis 2. Yauyos is part of an early dialect continuum from which different expansions occurred.
   a) Central and Southern Quechua are the opposite poles of a dialect continuum based upon a first expansion of Quechua.
   b) Northern Quechua arose from a second Quechua expansion on the basis of a variety of Southern Quechua, hence the similarity between those two.

In principle, our phylogeny is compatible with both hypotheses. A close investigation of the linguistic history of Yauyos and Huancavelica could possibly shed further light on the exact processes that led to the proximity of those varieties. Another likely problem is that through the regional isolation in Yauyos, it is very plausible that the varieties in question developed a set of shared innovative lexical items that are not inherited. Despite possibly belonging to different branches of the family in the first place, regional innovations that are present in the data could have led to a certain amount of convergence that the model interprets as defining for this clade. A detailed historical analysis that includes the most recent documentation efforts in the Yauyos area could clarify the affiliation of those varieties. Our quantitative model cannot resolve these questions with the limited amount of data we gathered for our database

Does this show that the phylolinguistic model does not work when complex linguistic scenarios are present, and that our tree is based on wrong assumptions? We argue that it does not. On the one hand, the low posterior value provides a first hint that the model is not confident about the positioning of this group. On the other hand, incomplete lineage sorting and horizontal transfer are a common challenge to tree models, but one that does not invalidate the whole methodology when applied correctly and with caution (Evans *et al.* 2021; Jacques and List 2019). The advantage of Bayesian phylogenies is that those problems are made explicit by the low posterior values attached to the branches. Following this discussion, our simplified tree for the Quechua language family is presented in Figure 7.
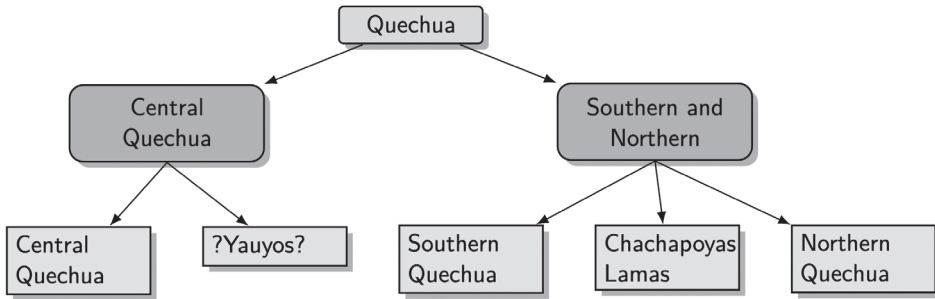
Figure 7. Higher-level groups in our Quechua phylogeny.

Our tree differs from Landerman's (Figure 2) in two main points. First, our model provides evidence for a grouping of Northern Quechua, Chachapoyas-Lamas, and Southern Quechua, within a single higher-level group. This evidence could not be established on the basis of phonological and morphological innovations, but our lexical database does provide such evidence. Second, Cajamarca and Ferreñafe are placed firmly within the Central Peruvian Quechua group in our model. This adds to the morphological evidence that at least the Ferreñafe variety is closely related to Central Quechua. We will discuss the interpretation of those results in the next section. The results provide some evidence of grouping Yauyos along the Central Quechua group, but the evidence is not fully conclusive. We will discuss some possible interpretations of the results in Section 4.1.

4. Discussion

4.1 Is Quechua a dialect continuum or tree-like?

We cannot answer this question definitively with the results of our model. However, we do think that our model provides quantitative support for the existence of two major clades that are roughly equal to the groups that have previously been proposed as QI/QB and QII/QA (Parker 1963; Torero 1964; Cerrón-Palomino 2003; Adelaar 2013). If our preliminary analysis is correct and the varieties from Yauyos can be separated by careful historical analysis and grouped with either clade, then our model provides support for this division. In order to interpret this as evidence in favour of the QI/QII distinction, it is important to be explicit about the assumptions that are made about Yauyos. This hypothesis could only be accepted if the proximity between Yauyos and Ayacucho/Huancavelica can be explained.

Considering the historic and linguistic evidence for an Inca-expansion to Ecuador (Hocquenghem 2012), the replacement of previously attested varieties in Ayacucho and Huancavelica (Itier 2016; Pearce and Heggarty 2011), and our results from the phylogenetic analysis, we argue that there were at least two major expansions. The first expansion

would then be related to Central Quechua, and a second, posterior expansion, related to Southern and Northern Quechua. In agreement with Heggarty (2005) and Adelaar (2012), the Central Quechua expansion would be driven by the Wari Empire, while the expansion of Northern and Southern Quechua can undoubtedly be related to the Inca reign and successive expansions for evangelization purposes during the colonization. The presence of two separate expansion events shows that it is not sufficient to explain the non-tree-like patterns in Quechua with a single-expansion dialect continuum. The tree model captures the relationship of each major clade with the two historic expansions of Quechua. The background of the diversity of Quechua varieties in Yauyos is secondary to this question.

We fully agree with Pearce and Heggarty (2011, 94) that internally, both groups are dialect continua. This is directly reflected in the branchings for Central Quechua, which follow a strict regional pattern. The fact that Central Quechua represents a continuum was already noted by Torero (1974, 24). Instead of a slowly expanding movement, we agree that the most reasonable model for the Central Quechua group is an initial phase of linguistic homogeneity, followed by a breakup of this status, giving rise to regional innovations. However, we do not agree with another conclusion by Pearce and Heggarty (2011, 93), who argue for a single expansion of the Quechua language family.
Arguing for a binary branching does not necessarily imply that there ever existed a Proto-Quechua I and a Proto-Quechua II simultaneously (cf. Pearce and Heggarty 2011, 93). Our tree model is certainly compatible with the hypothesis of an initial, homogeneous Quechua zone that is often analyzed as Quechua I (our Central Quechua) during the Wari Empire, and a second expansion out of that dialect continuum that gave rise to Northern and Southern Quechua. Considering the evidence for the historic presence of other Quechua varieties in Huancavelica and Ayacucho, the areal distribution of this first-expansion dialect continuum remains unclear. A reasonable hypothesis would be to posit its origin in the Huancavelica and Ayacucho region (Adelaar 2012, 214), which possibly marked the southern pole of the initial dialect continuum. This would later have been obscured by comparatively recent migrations to the areas of Huancavelica and Ayacucho (Itier 2016; Pearce and Heggarty 2011), replacing the local varieties from the Central Quechua continuous zone. By the combination of both the tree representation and the dialect continuum we can now capture both the diachronic perspective, which includes two separate expansion events, and the pattern of apparently mixed varieties.

4.2 Internal relations of Northern and Southern Quechua

A central point of critique for the grouping of Southern Quechua, Northern Quechua, and Chachapoyas-Lamas provided by Landerman (1991) was the lack of shared morphological and phonological innovations. The author claimed that based on this result, it was not possible to group those clades together, as any similarity might only be based on shared retentions. However, we think that our model actually combines these groups due

to some few shared lexical innovations. Even though it is difficult to judge in many cases which form is actually innovative, we think there is at least one convincing case, namely *puju* 'CLOUD'. The contrastive form in Central Quechua is *pukutay*, which is a nominalization of *pukuta-* 'TO BECOME CLOUDY' with the abstract nominalizer *-i*. A careful qualitative analysis of additional lexical data is necessary in order to confirm this judgment.

Notably, (Southern) Pastaza Quechua is part of the Northern Quechua clade, which confirms previous findings that this variety was likely introduced through the Ecuadorian *selva* by missionaries, and is unrelated to the varieties from Chachapoyas-Lamas. Adelaar (2012) proposes a coastal origin for Chachapoyas via *mitmaes* and observes several similarities with the lexicon of Santo Tomas (DST). However, the vocabulary of Santo Tomás is commonly analyzed as a collection of data from a multitude of Quechua varieties (Ezcurra Rivero and Bendezú-Araujo 2017), which is why we excluded this data from the phylogeny. The qualitative analysis suggests several shared lexical cognates with maritime context. For example, this includes *wayta* 'SWIM' as well as *tiyu* 'SAND', which are shared only with a limited number of other varieties in our database, such as Maragua and Pocona. The presence of maritime vocabulary in DST has been noted before and even led researchers to suggest a coastal origin of Quechua (Torero 1974, 154). However, careful analyses show that despite making a Quechua presence on the coast very probable, Quechua did most likely not originate on the coast (Urban 2021).

4.3 The position of mixed 'QIIA' varieties reconsidered

Our study further confirms previous arguments that a group 'QIIA' does not exist (Landerman 1991; Heggarty 2005; Adelaar 2012). Further, it provides strong evidence on the grouping of each of those varieties based on lexical cognates. We disagree with Adelaar (2012, 203) about the position of individual varieties such as Cajamarca and Ferreñafe, which in our analysis are firmly within the group of Central Quechua. This is in line with the analysis as Cajamarca being the northernmost outpost of the first wave of Quechua expansions (Heggarty 2005, 75). The position of Cajamarca and Ferreñafe Quechua as part of Central Quechua would mean that those are not Southern Quechua (old 'QII') varieties that came into contact with Central Quechua (old 'QI'), but rather the other way round. Strengthening this argument, Hocquenghem (2012, 363) cites evidence that the highland trading road from Cuzco to Quito led through Cajamarca, which could be a possible source of contact between the local variety and the Inca-Quechua varieties. This would provide the historical context in which the local Quechua varieties borrowed some of the lexical and morphological elements that are usually found in Southern and Northern Quechua, and seems like the more plausible scenario than the one postulated previously.

Our proposed group of Central Quechua includes many other varieties that were previously classified as 'QI' and 'QIIA'. For example, Laraos is assigned a peripheral status to this group. The next branch consists of Jauja-Huanca, followed by Cajamarca and

Ferreñafe. Tarma and Pacaraos Quechua were previously classified as a first branch of Quechua I (Adelaar 1984, 45), but are deep inside the Central Quechua branch of our phylogeny. Apart from Laraos, the posterior values indicate very strong evidence in favour of this group. Laraos, on the other hand, is repeatedly showing some weak signals related to the other Yauyos varieties such as Lincha-Tana, San Pedro, and Cacra-Hongos.

The varieties of Chachapoyas-Lamas emerge as one of three higher-level subgroups of Southern and Northern Quechua. Even though it does have some similarities with QIIB, the more likely scenario in contrast to a combined clade is to have two separated, but chronologically close expansions towards Chachapoyas and Ecuador. This is in line with the fast conquest by the Inca empire. Further, there remains no doubt about the distinctiveness of Chachapoyas-Lamas and Cajamarca-Ferreñafe. Despite their geographical proximity, they emerge on opposite sides of the tree and do not share significant lexical cognates.

5. Conclusion

We presented a quantitative study on 39 Quechua varieties using a dataset of 150 lexical concepts. The phylogeny based on the analysis of the cognate sets shows a strong division between Central Quechua on one hand, and Northern Quechua, Chachapoyas-Lamas, and Southern Quechua on the other hand. A third group is composed of the varieties of Yauyos, which cannot be assigned to either group on the basis of the phylogenetic analysis. Some varieties that previously were subject to different classifications are put strongly in certain branches, such as Cajamarca-Ferreñafe and Laraos in Central Quechua, as well as a non-tree-like structure between Northern Quechua, Southern Quechua, and Chachapoyas-Lamas. We further analyzed the situation of Yauyos with reference to the historical language situation in Huancavelica and Ayacucho, and provide different possible interpretations for the lexical similarity between those varieties. From our viewpoint, the most likely scenario is an early dialect continuum which lived through two different expansions. Given the independent evidence for at least two historic phases of expansions for the Quechua language family, we argue that the tree model successfully recovers this history, as seen from the strong split between the two central groups.

References cited

Adelaar, Willem F. H.
  1977   *Tarma Quechua: grammar, texts, dictionary*. Lisse: The Peter de Ridder Press.
  1982   *Léxico del quechua de Pacaraos*. Lima: Universidad Nacional Mayor de San Marcos/Centro de Investigación de lingüística aplicada.
  1984   "Grammatical vowel length and the classification of Quechua dialects." *International Journal of American Linguistics* 50, no. 1: 25-47. https://doi.org/10.1086/465814
  1994   "Raíces lingüísticas del quichua de Santiago del Estero." In *Actas de las segundas jornadas de lingüística aborigen*, 1-37. Buenos Aires: Universidad de Buenos Aires.
  2012   "Cajamarca Quechua and the expansion of the Huari state." In *Archaeology and language in the Andes: A cross-disciplinary exploration of prehistory*, edited by David Beresford-Jones and Paul Heggarty, 197-217. Oxford/New York: Oxford University Press. https://doi.org/10.5871/bacad/9780197265031.003.0008
  2013   "Quechua I y Quechua II: en defensa de una distinción establecida." *Revista Brasileira de Linguística Antropológica* 5, no. 1: 45-65. https://doi.org/10.26512/rbla.v5i1.16542

Beresford-Jones, David, and Paul Heggarty
  2012   "Broadening our horizons: Towards an interdisciplinary prehistory of the Andes." In *Archaeology and Language in the Andes*, edited by Paul Heggarty and David Beresford-Jones, 57-84. Oxford/New York: Oxford University Press. https://doi.org/10.5871/bacad/9780197265031.003.0003

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, no. 3: 255-284. https://doi.org/10.1086/687383

Blum, Frederic, Carlos B. Ugarte, Adriano Ingunza, and Zoe Poirier
  2021   CLDF dataset on Quechua morphology derived from Blum *et al.*'s "Una aproximación filolingüística a la clasificación interna del Quechua". V. 1.0. https://doi.org/10.5281/zenodo.5498292

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, no. 6097: 957-960. https://doi.org/10.1126/science.1219669

Bouckaert, Remco, Joseph Heled, Denise Kühnert, Tim Vaughan, Chieh-Hsi Wu, Dong Xie, Marc A. Suchard, Andrew Rambaut, and Alexei J. Drummond
  2014   "BEAST 2: A software platform for bayesian evolutionary analysis." *PLoS Computational Biology* 10, no. 4: e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Bryant, David, and Vincent Moulton
  2004   "Neighbor-Net: An agglomerative method for the construction of phylogenetic networks." *Molecular Biology and Evolution* 21, no. 2: 255-265. https://doi.org/10.1093/molbev/msh018

Cerrón-Palomino, Rodolfo
  1976   *Diccionario quechua: junin-huanca*. Lima: Ministerio de Educación.
  2003   *Lingüística quechua*. Cuzco: Centro de Estudios Rurales Andinos Bartolomé de Las Casas.

Cerrón-Palomino, Roldofo, José Francisco Ráez, and Sergio Cangahuala Castro
    2018   *Diccionario huanca quechua-castellano castellano-quechua*. Lima: Pontificia Universidad Católica del Perú (PUCP)/Instituto Riva-Agüero.

Chacon, Thiago Costa, and Johann-Mattis List
    2016   "Improved computational models of sound change shed light on the history of the Tukanoan languages." *Journal of Language Relationship* 13, no. 3-4: 177-204. https://doi.org/10.31826/jlr-2016-133-404

Chang, Will, Chundra Cathcart, David Hall, and Andrew Garrett
    2015   "Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis." *Language* 91, no. 1: 194-244. https://doi.org/10.1353/lan.2015.0005

Cordero, Luis
    1955   *Diccionario quichua-español, español-quichua*. Quito: Casa de la Cultura Ecuatoriana.

Evans, Cara L., Simon J. Greenhill, Joseph Watts, Johann-Mattis List, Carlos A. Botero, Russell D. Gray, and Kathryn R. Kirby
    2021   "The uses and abuses of tree thinking in cultural evolution." *Philosophical Transactions of the Royal Society B: Biological Sciences* 376, no. 1828: n.p. https://doi.org/10.1098/rstb.2020.0056

Ezcurra Rivero, Álvaro, and Raúl Bendezú-Araujo
    2017   "Gramáticas y vocabularios coloniales del quechua y del aimara (1560-1619)." In *Literaturas orales y primeros textos coloniales*, edited by Juan C. Godenzzi and Carlos Garatea, 123-164. Historia de las literaturas en el Perú. Lima: Pontificia Universidad Católica del Perú (PUCP)/Fondo Editorial.

Floyd, Simeon
    2022   "Ecuadorian Highland Quichua and the lost languages of the Northern Andes." *International Journal of American Linguistics* 88, no. 1: 1-52. https://doi.org/10.1086/717056

Forkel, Robert, and Johann-Mattis List
    2020   "CLDFBench: Give your cross-linguistic data a lift." In *12th Conference on Language Resources and Evaluation*, 6995-7002. Paris: European Language Resources Association (ELRA). https://doi.org/10.17613/8t0e-w639.

Gálvez Astorayme, Isabel
    2003   *Léxico del quechua de Arma*. Lima: Hipocampo.

Gerardi, Fabrício Ferraz, and Stanislav Reichert
    2021   "The Tupí-Guaraní language family." *Diachronica* 38, no. 2: 151-188. https://doi.org/10.1075/dia.18032.fer

Gray, Russel D., Alexei J. Drummond, and Simon J. Greenhill
    2009   "Language phylogenies reveal expansion pulses and pauses in Pacific settlement." *Science* 323, no. 5913: 479-483. https://doi.org/10.1126/science.1166858

Greenhill, Simon J., and Russell D. Gray
    2012   "Basic vocabulary and Bayesian phylolinguistics." *Diachronica* 29, no. 4: 523-537. https://doi.org/10.1075/dia.29.4.05gre

Greenhill, Simon J., Paul Heggarty, and Russell D. Gray
    2020   "Bayesian phylolinguistics." In *The Handbook of historical linguistics*, edited by Richard D. Janda, Brian D. Joseph, and Barbara S. Vance, 226-253. Hoboken: Wiley Blackwell. https://doi.org/10.1002/9781118732168.ch11

Gutiérrez Camacho, Leoncio, Leonidas Mantilla Gutierrez, and Shara Huaman Jullunila
    2007   *Apurimaqpaq runasimi taqe diccionario de quechua apurimeño*. Abancay: Academia Mayor de la Lengua Quechua filial Apurímac.

Hammarström, Harald, Martin Haspelmath, Robert Forkel, and Sebastian Bank
    2021   "Glottolog. Version 4.5." Leipzig: Max Planck Institute for Evolutionary Anthropology. https://glottolog.org (21.04.2023)

Hartmann, Roswith
    1979   "¿'Quechuismo preincaico' en el Ecuador?" *Ibero-Amerikanisches Archiv* 5, no. 3: 267-299.

Heggarty, Paul
    2005   "Enigmas en el origen de las lenguas andinas: aplicando nuevas técnicas a las incógnitas por resolver." *Revista Andina* 40: 9-57.

Hocquenghem, Anne Marie
    2012   "How did Quechua reach Ecuador?" In *Archaeology and language in the Andes*, edited by Paul Heggarty and David Beresford-Jones, 344-371. Oxford/New York: Oxford University Press. https://doi.org/10.5871/bacad/9780197265031.003.0014

Holman, Eric W., Robert Walker, Taraka Rama, and Søren Wichmann
    2011   "Correlates of reticulation in linguistic phylogenies." *Language Dynamics and Change* 1, no. 2: 205-240. https://doi.org/10.1163/221058212X648072

Huson, Daniel H., and David Bryant
    2005   "Application of phylogenetic networks in evolutionary studies." *Molecular Biology and Evolution* 23, no. 2: 254-267. https://doi.org/10.1093/molbev/msj030

Itier, César
    2016   "La formación del quechua ayacuchano, un proceso inca y colonial." *Bulletin de l'Institut français d'études andines* 45, no. 2, 307-326. https://doi.org/10.4000/bifea.8003

Jacques, Guillaume, and Johann-Mattis List
    2019   "Save the trees: Why we need tree models in linguistic reconstruction (and when we should apply them)." *Journal of Historical Linguistics* 9, no. 1: 128-166. https://doi.org/10.1075/jhl.17008.mat

Juanatey, Mayra
    2020   "Relaciones entre eventos y referencialidad en quichua santiagueño: de la gramática al discurso." PhD diss., Universidad de Buenos Aires. https://doi.org/10.13140/RG.2.2.30896.79365
    2021   Fieldwork notes. Unpublished manuscript.

Landerman, Peter
    1973   *Vocabulario quechua del Pastaza*. Yarinacocha: Instituto Lingüístico de Verano (ILV).
    1991   "Quechua dialects and their classification." PhD diss., University of California at Los Angeles (UCLA).

List, Johann-Mattis
 2016   "Beyond cognacy: historical relations between words and their implication for phylogenetic reconstruction." *Journal of Language Evolution* 1, no. 2: 119-136. https://doi.org/10.1093/jole/lzw006
 2021   "EDICTOR. A web-based tool for creating, editing, and publishing etymological datasets." Leipzig: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.4685130. https://digling.org/edicto (21.04.2023)

List, Johann-Mattis, and Robert Forkel
 2021   "LingPy. A Python library for quantitative tasks in historical linguistics." Leipzig: Max Planck Institute for Evolutionary Anthropology. https://lingpy.org (21.04.2023)

List, Johann Mattis, Christoph Rzymski, Simon Greenhill, Nathanael Schweikhard, Kristina Pianykh, Annika Tjuka, Carolin Hundt, and Robert Forkel
 2021   "Concepticon. A resource for the linking of concept lists. Version 2.5.0." Jena: Max Planck Institute for the Science of Human History. https: //doi.org/10.5281/zenodo.4911605. https://concepticon.clld.org/ (21.04.2023)

List, Johann-Mattis, Robert Forkel, Simon J. Greenhill, Christoph Rzymski, Johannes Englisch, and Russell Gray
 2022   "Lexibank, a public repository of standardized wordlists with computed phonological and lexical features." *Scientific Data* 9, no. 1: 316. https://doi.org/10.1038/s41597-022-01432-0

Michael, Lev, and Natalia Chousou-Polydouri
 2019   "Computational phylogenetics and the classification of South American languages." *Language and Linguistics Compass* 13, no. 12: e12358. https://doi.org/10.1111/lnc3.12358

Michael, Lev, Natalia Chousou-Polydouri, Keith Bartolomei, Erin Donnelly, Sérgio Meira, Vivian Wauters, and Zachary O'Hagan
 2015   "A bayesian phylogenetic classification of Tupí-Guaraní." *LIAMES: Línguas Indígenas Americanas* 15, no. 2: 193-221. https://doi.org/10.20396/liames.v15i2.8642301

Orr, Betsy, and Betsy Wrisley
 1981   Diccionario quichua del Oriente. Quito: Instituto Lingüístico de Verano (ILV).

Parker, Gary
 1963   "La clasificación genética de los dialectos quechuas." *Revista del Museo Nacional* 32: 241-252. https://repositorio.cultura.gob.pe/handle/CULTURA/779 (21.04.2023)

Parker, Gary, and Amancio Chavez
 1976   *Diccionario quechua ancash-huailas*. Lima: Ministerio de Educación.

Pearce, Adrian J., and Paul Heggarty
 2011   "'Mining the data' on the Huancayo-Huancavelica Quechua frontier." In *History and Language in the Andes*, edited by Paul Heggarty and Adrian J. Pearce, 87-109. New York: Palgrave Macmillan. https://doi.org/10.1057/9780230370579_5

Rendón, Jorge A. Gómez
 2009   "Imbabura Quechua." In *World loanword database*, edited by Martin Haspelmath and Uri Tadmor, n. p. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wold.clld.org/vocabulary/37 (21.04.2023)

Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin J. Ryder, Valentin Thouzeau, Simon J. Greenhill, and Johann-Mattis List
    2019   "Dated language phylogenies shed light on the ancestry of Sino-Tibetan." *Proceedings of the National Academy of Sciences* 116, no. 21: 10317-10322. https://doi.org/10.1073/pnas.1817972116

Shimelman, Aviva
    2017   *A grammar of Yauyos Quechua*. Studies in Diversity Linguistics 9. Berlin: Language Science Press.

Simons, Daniel J., Yuichi Shoda, and D. Stephen Lindsay
    2017   "Constraints on Generality (COG): A proposed addition to all empirical papers." *Perspectives on Psychological Science* 12, no. 6: 1123-1128. https://doi.org/10.1177/1745691617708630

Soto Ruiz, Clodoaldo
    1976   *Diccionario quechua Aayacucho-chanca*. Lima: Ministerio de Educación/Instituto de Estudios Peruanos.

Tandioy, Domingo, Stephen Levinsohn, and Alonso Maffla
    1978   *Diccionario inga del Valle de Sibundoy Intendencia del Putumayo*. S.l.: Townsend.

Taylor, Gerald
    1984   "Yauyos, un microcosmo dialectal quechua." *Revista Andina* 2, no. 1: 121-146.
    2006   *Diccionario quechua chachapoyas-lamas*. Lima: Instituto Francés de Estudios Andinos (IFEA)/Instituto de Estudios Peruanos (IEP)/Editorial Commentarios SAC.

Torero, Alfredo
    1964   "Los dialectos quechua." *Anales Científicos de la Universidad Agraria* 2, no. 4: 446-478.
    1974   *El quechua y la historia social andina*. Lima: Fondo Editorial del Pedagócico San Marcos.

Urban, Matthias
    2021   "Terminología marítima en el Lexicon, o Vocabulario de la lengua general del Perú de Domingo de Santo Tomás (1560) y posibles implicaciones para la historia de la familia lingüística quechua." *Boletín de la Academia Peruana de la Lengua*, 13-61. https://doi.org/10.46744/bapl.202102.001

Yu, Guangchuang
    2020   "Using ggtree to Visualize Data on Tree-Like Structures." *Current Protocols in Bioinformatics* 69, no. 1: e96. https://doi.org/10.1002/cpbi.96

Zariquiey, Roberto, Alonso Vásquez, and Gabriela Tello
    2017   "Lenguas y dialectos pano del Purús: una aproximación filogenética." *Lexis* 41, no. 1: 83-120. https://doi.org/10.18800/lexis.201701.003

Appendix

Location	Variety	Source
Apurí	Apurí-Madean-Viñac	Shimelman 2017
Apurímac	Eastern-Apurímac	Gutiérrez Camacho *et al.* 2007
Arma	Arma	Gálvez Astorayme 2003
Atalla	Huancavelica	Heggarty 2005
Ayacucho	Ayacucho	Soto Ruiz 1976
Azuay	Azuay	Cordero 1955
Bobonaza	Bobonaza	Orr and Wrisley 1981
Cacra-Hongos	Cacra-Hongos	Shimelman 2017
Chachapoyas	Chachapoyas	Taylor 2006
Chacpar	Chavín	Heggarty 2005
Chetilla	Cajamarca	Heggarty 2005
Curva	Apolobamba	Heggarty 2005
Cuzco	Cuzco	Heggarty 2005
Huallaga	Huallaga-Huánuco	Heggarty 2005
Huanca	Huaylla-Huanca	Cerrón-Palomino *et al.* 2018
Huangáscar	Chocos-Huangáscar	Shimelman 2017
Huaraz Huaylas	Huaraz-Huaylas	Parker and Chavez 1976
Imbabura	Imbabura	Rendón 2009
Inkawasi	Lambayeque	Heggarty 2005
Jauja	Jauja-Huanca	Cerrón-Palomino 1976
Lamas	San Martín	Taylor 2006
Laraos	Laraos	Heggarty 2005
Lincha-Tana	Lincha-Tana	Shimelman 2017
Maragua	Chuquisaca	Heggarty 2005
Napo	Napo	Orr and Wrisley 1981
Pacaraos	Pacaraos	Adelaar 1982
Pastaza	Pastaza	Landerman 1973
Pocona	Cochabamba	Heggarty 2005
Puno	Puno	Heggarty 2005
Putumayo	Colombian Inga	Tandioy *et al.* 1978
Raimondi	Antonio Raimondi	Parker and Chavez 1976
SanPedro	Liscay-San Pedro	Shimelman 2017
Santiagueno	Santiago del Estero	Juanatey 2021
Serena	Tena	Heggarty 2005
Taquile	Taquile	Heggarty 2005
Tarma	Yaru	Adelaar 1977
Troje	Chimborazo	Heggarty 2005
Wari	Wari	Parker and Chavez 1976
Yanac	Corongo	Heggarty 2005

Table 2. All varieties in the analysis and their main source.