

UC Santa Barbara

Himalayan Linguistics

Title

Sub-grouping Kho-Bwa based on shared core vocabulary

Permalink

<https://escholarship.org/uc/item/4t27h5fg>

Journal

Himalayan Linguistics, 16(2)

Authors

Lieberherr, Ismael
Bodt, Timotheus Adrianus

Publication Date

2017

DOI

10.5070/H916232254

Supplemental Material

<https://escholarship.org/uc/item/4t27h5fg#supplemental>

License

<https://creativecommons.org/licenses/by-nc-nd/4.0/> 4.0

Peer reviewed

Himalayan Linguistics

A free refereed web journal and archive devoted to the study of the
languages of the Himalayas

Himalayan Linguistics

Sub-grouping Kho-Bwa based on shared core vocabulary

Ismael Lieberherr

Timotheus Adrianus Bodt

Bern University/ Tezpur University

ABSTRACT

Tianshin Jackson Sun (Sun, 1992; Sun, 1993) was the first to suggest the phylogenetic relatedness of a number of highly divergent, endangered, and poorly described languages of Western Arunachal Pradesh, later named the 'Kho-Bwa cluster' by van Driem (2001). In this paper, we make use of what are predominantly new data from our own field work, covering a total of 22 linguistic varieties. In a list of 100 lexical entries, we determined cognacy manually, and computed a "cognacy percentage" for each pair of languages. The result of this analysis, and some further considerations, confirm earlier reported views of a phylogenetic relationship between these languages. The appendix contains the full data set with cognacy statements.

KEYWORDS

Kho-Bwa, Tibeto-Burman, hierarchical clusters analysis

This is a contribution from *Himalayan Linguistics*, Vol. 16(2): 26–63.

ISSN 1544-7502

© 2017. All rights reserved.

This Portable Document Format (PDF) file may not be altered in any way.

Tables of contents, abstracts, and submission guidelines are available at
escholarship.org/uc/himalayanlinguistics

Sub-grouping Kho-Bwa based on shared core vocabulary¹

Ismael Lieberherr

Timotheus Adrianus Bodt

Bern University/Tezpur University

1 Introduction

Western Arunachal Pradesh is a region of great linguistic diversity. The relatively well-known groups represented here include East Bodish (several varieties of ‘Tawang Monpa’), several varieties belonging to the unclassified Tshangla group, Central Bodish (Brokpa) and Tani (Bengni/Nyishi). Less well known, but since the 1940s presumed part of the Tibeto-Burman language inventory (Shafer 1947) is Hrusish (Miji, Bangru, Hruso Aka). Also presumed to belong to this language family as an independent language is Koro (Abraham et al. 2005; Anderson and Murmu 2010). In addition, commonly consulted handbooks (Burling 2003; Post and Burling 2017) and the online language encyclopaedias *Ethnologue*² and *Glottolog*³ add another (potential) branch of Tibeto-Burman in western Arunachal Pradesh called ‘Kho-Bwa’. The existence of the Kho-Bwa subgroup has also been suggested in other publications (e.g. Bodt 2014; Lieberherr 2015). However, as of 2017, not much, if any, linguistic evidence for the actual coherence of this group has appeared in published sources. This motivated us to write this article, in which we review the previously published material on the Kho-Bwa languages, make a count of cognate basic vocabulary of 22 linguistics varieties presumably belonging to this group, and provide data evidencing its coherence and sub-grouping.

1.1 Previous research

Linguistic and ethnic affinities among the Kho-Bwa varieties and their speakers are of course known to the people of the languages communities themselves. As early as 1952, Stonor reported that Puroik and Bugun are mutually intelligible (Stonor 1952).⁴ However, it was not until the last two decades of the previous century that the first linguistic materials on Bugun/Khowa, Puroik/Sulung, Sherdukpen and Sartang/Boot/Butpa Monpa became available: the works of the

¹ The authors wish to gratefully acknowledge all the speakers of Kho-Bwa languages who shared their wisdom, knowledge and time with us in the past five years. Furthermore, we would like to thank David Bürgin for his improvements on the computer code. We are also much indebted to the anonymous reviewers of our article and the editors of *Himalayan Linguistics* for their useful comments, suggestions and editorial work.

² <http://www.ethnologue.com/subgroups/kho-bwa> accessed in January 2017.

³ <http://glottolog.org/resource/languoid/id/khob1235> accessed in January 2017.

⁴ As of 2017 we cannot confirm this claim. Nowadays, the Puroik dialect spoken in the place Stonor visited is not even mutually intelligible with Western Puroik dialects. But origin stories linking the Buguns with the Puroiks are indeed common.

Indian research/language officers Deuri (1983), Dondrup (1988), Tayeng (1990), Dondrup (1990), and Dondrup (2004). On the Chinese side, the first Puroik data were published as part of the large-scale survey *Tibeto-Burman Phonology and Lexicon* (Sūn et al. 1991). Based on these materials and own data, Jackson Sun (Sun 1992, 1993) was the first to suggest that Puroik, Bugun, Sherdukpen and ‘Lishpa-Butpa’ are not just a random residue when all other major languages are subtracted, but that they might belong together and form a coherent linguistic group.⁵ Other researchers after him either adopted his view or independently reached the same conclusion (Rutgers 1999; Burling 2003). Van Driem (2001) named the group “Kho-Bwa cluster” in his handbook *Languages of the Himalayas*, after the reconstructions for WATER and FIRE. Blench and Post (2014) and Post and Burling (2017) are sceptical about Puroik being part of the group.

Although we do not agree with the exact phonological shape of the reconstructions **kho* WATER and **bwa* FIRE, we recommend using “Kho-Bwa” as a label for these languages. Besides the fact that it is already established to some extent, it has the advantage of not being biased toward one language like “Bugunish” (Sun 1993), or a region like “Kamengic”⁶ (Blench and Post 2014; Post and Burling 2017). Furthermore, “Kho-Bwa” offers an exhaustive definition of the group: Any language of western Arunachal Pradesh in which the word for ‘water’ starts with *k* and the word for ‘fire’ starts with *b* is a “Kho-Bwa” language.

1.2 Included “languages”

Since Sun (1992), the following languages and linguistic varieties are counted part of the Kho-Bwa group: Khispi (also known as Lishpa, Lish Monpa), Duhumbi (a.k.a. Chugpa, Chug Monpa), Sartang (a.k.a. But Monpa), Sherdukpen (a.k.a. Mey, cf. Blench 2015), Bugun (a.k.a. Khowa) and Puroik (a.k.a. Sulung). The lexical database on which this study relies consists of largely original data from 22 varieties of Kho-Bwa. Most of the data are from our own fieldwork. However, the quality of the data is not the same for each variety. On some varieties, we worked for several years together with several speakers, the data of other varieties were elicited in a single session from one single speaker. The following is a list of the Kho-Bwa varieties included in this study.

⁵ His statements are rather vague and restricted to footnotes: Sun (1992: 80 fn. 18): “The only Tibeto-Burman languages in the vicinity that show some affinity to Sulung are the obscure group consisting of Bugun, Sherdukpen and Lishpa-Butpa, but even here the relationship does not seem to be very close.”, Sun (1993: 8 fn. 14): “Sulung is a newly discovered distinct Tibeto-Burman language showing remarkable similarities to Bugun, another obscure Tibeto-Burman language spoken further to the west of the Sulung country.”, and Sun (1993: 11 fn. 18): “All of these languages have only very recently become accessible for linguistic study. From the meager published data, it seems likely that Bugun, Lishpa, and Sherdukpen may constitute a new Tibeto-Burman group yet to be recognized (Bugunish?). The peculiar Sulung language (whose autonym Puroit [pu-ɣo̯t̚ ~ pu-ro̯t̚] also seems relatable to the autonym Bugun) may also turn out to be most closely akin to this group.”

⁶ Many other languages that are not directly related to the Kho-Bwa languages are spoken in the Kameng region, such as Tshangla, Brokpa, Hruso Aka, Koro, Western Miji, Eastern Miji, and Bengni (Tani).

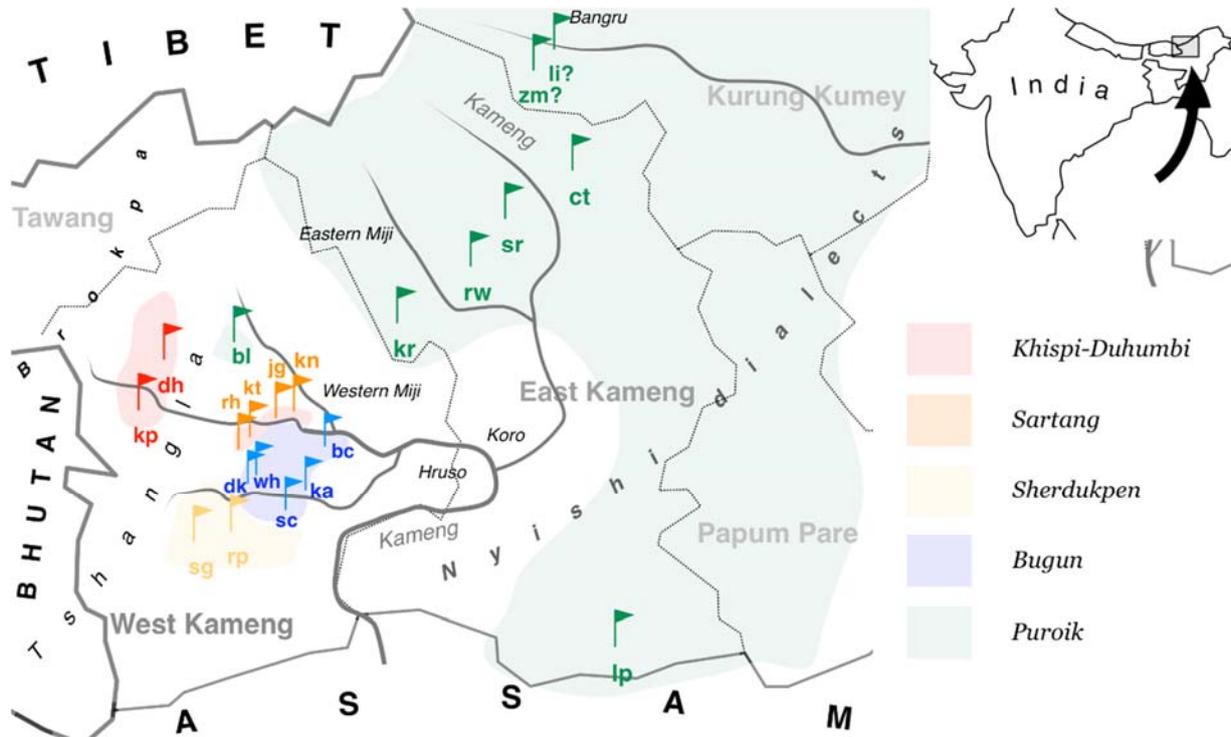


Figure 1. Linguistic map of western Arunachal Pradesh

- **Khispi-Duhumbi** Previous data of Khispi and Duhumbi have been published in Abraham et al. (2005).
 - [dh] **Duhumbi** (Chugpa) spoken by 600 people in three main villages and associated hamlets. A comprehensive grammar is in preparation by Bodt.
 - [kp] **Khispi** (Lishpa) spoken in three main villages and associated hamlets, around 1,500 speakers. A sketch grammar is in preparation by Bodt.
- **Sherdukpen** Previous data about Sherdukpen have been published in Dondrup (1988), Abraham et al. (2005) and Jacquesson (2015). The data presented here come from our own fieldwork.
 - [rp] **Rupa** spoken in three main villages and associated hamlets by perhaps 3,000 people.
 - [sg] **Shergaon** spoken in one main village by perhaps 1,500 people.
- **Sartang** is linguistically closely related to Sherdukpen. Previously, Sartang data have been published in Dondrup (2004) and Abraham et al. (2005).
 - [rh] **Rahung** spoken in one main village and associated hamlets by around 600 people.
 - [kt] **Khoitam** spoken in two main villages and associated hamlets by around 500 people.
 - [jg] **Jerigaon** spoken in one village by around 400 people.
 - [kn] **Khoina** spoken in one village and associated hamlets by around 500 people.
- **Bugun** Except for one variety, the Bugun data are from Abraham et al. (2005). Other Bugun data were published in Dondrup (1990) and Barbora (2015).

- [dk] **Dikhyang** Data from our own fieldwork, 100 speakers.
- [sc] **Singchung** 680 speakers.
- [wh] **Wangho** 220 speakers.
- [bc] **Bichom** 630 speakers.
- [ka] **Kaspi** 80 speakers.
- [np] **Namphri** 180 speakers.

• **Puroik** Previously published sources about Puroik are Deuri (1983), Tayeng (1990), Soja (2009), Remsangpuia (2008), Lǐ (2004), and Sūn et al. (1991). These sources all represent relatively closely related varieties of the Chayangtajo area. Some data of western dialects were published in Lieberherr (2015). Puroik has more dialects, and thus more speakers, in the east, which could not be included in this study.

- [bl] **Bulu** Only spoken in one village by 7-20 speakers. A comprehensive grammar is in preparation by Lieberherr.
- [kr] **Kojo Rojo** spoken in two villages Kojo and Rojo by a few hundred speakers.
- [rw] **Rawa** spoken in several villages in and around Rawa by a few hundred speakers.
- [sr] **Sario Saria** spoken in three villages by a few hundred speakers.
- [ct] **Chayangtajo** spoken in several villages in the Chayangtajo area by a few hundred speakers.
- [lp] **Lasumpatte** Puroik variety spoken in one village in Seijosa near Assam border, mainly inhabited by relatively recent migrants from the Chayangtajo area.
- [zm] Puroik variety recorded in (Sūn et al. 1991) in Tibet, possibly with speakers from Kurung Kumey.
- [li] Puroik variety recorded by Lǐ (2004) in Tibet, possibly with speakers from Kurung Kumey.

1.3 Language classification and the case of Kho-Bwa

Language classification, and historical linguistics in general, deals with “evidence from three sources: basic vocabulary, grammatical evidence (especially morphological), and sound correspondences.” (Campbell and Poser 2008: 4).⁷ These sources are not independent, and sometimes almost circularly connected. Inherited basic vocabulary is needed, in order to find sound correspondences, sound correspondences are needed in order to know what is inherited, sound correspondences are needed to find cognate morphology, and cognate morphology can help to find sound correspondences. None of the three can be studied in isolation.

However, this does not imply that the three have the same importance always and everywhere in language classification. For Indo-European, for example, the rich inflectional and derivational morphology of ancient Indo-European languages provides a huge amount of information about the character of Proto-Indo-European and about the diversification of the daughter languages. Grouping based on the lexicon is of rather subordinate importance. While in other language families a lot can

⁷ “It will not spoil any surprises to come if we disclose here at the outset that throughout the history of linguistics the criteria employed in both pronouncements about method and in actual practice for establishing language families consistently included evidence from three sources: basic vocabulary, grammatical evidence (especially morphological), and sound correspondences.”

be concluded from comparing, for example, present tense paradigms, such kind of valuable information about the past is entirely missing in Kho-Bwa. There are hardly any paradigms, no ablaut, no inflection classes. Under these circumstances, investigating shared basic vocabulary is more important than in families with a rich morphology. In this survey, we focus on core vocabulary (section 2), i.e. the words which are least likely to be borrowed, and only sketch important issues in phonology and morphology (section 3). Detailed phonological comparison has to await the analysis of the synchronic phonology of more language varieties of Kho-Bwa.

2 Shared core vocabulary

For our study of cognate core vocabulary in Kho-Bwa languages there were several questions we thought worthwhile exploring. Are there differences as to how much core vocabulary is shared between these languages? Or are the differences blurred after centuries of language contact and diffusion? Can the language-groups ‘Khispi-Duhumbi’, ‘Sherdukpen’, ‘Sartang’, ‘Bugun’ and ‘Puroik’ be confirmed based on shared core vocabulary, even if for example Bulu Puroik is geographically much closer to Sartang and Bugun than to other Puroik dialects? Are the Kho-Bwa languages as a whole lexically distinct from surrounding languages? Or have some Kho-Bwa languages become lexically so much assimilated that Kho-Bwa languages are rather substrates to those languages than languages in their own right? In order to find answers to these questions we compiled a suitable list of concepts (2.1) and translated them into 22 Kho-Bwa languages and seven other Tibeto-Burman languages (2.2), we judged for every set of words manually which are cognate and which are not (2.3), we then grouped the languages according to similarity using a hierarchical cluster algorithm (2.4), and interpreted the results (2.5 - 2.7).

2.1 *Compilation of the word list*

There is a wide range of concept lists used in comparative studies. The “concepticon”, an online resource of concept lists, contains a collection of 161 lists (List, Cysouw, and Forkel 2016). The decision which list is best suited depends on the research questions, the setup of the study, the data, and, in our view, also on the languages compared. If the research question is to determine to which extent some languages are mutually intelligible (such as Abraham et al. 2005), one would probably devise a list with the most frequently used words in discourse. Words like ‘mobile’, ‘tea’, ‘onion’ and ‘cooking oil’ are important because these are frequently used words nowadays and important to be able to understand each other. However, if the question is, *whether*, and if so, *how* these languages might have evolved from a common ancestor, the list should consist of words that are likely to be inherited from this common ancestor, i.e., it would be better to use concepts that are not easily borrowed.

Since our objective was to determine whether the languages purported to be Kho-Bwa languages derived from a common ancestor language, we started from the Leipzig-Jakarta 100 items list (Haspelmath and Tadmor 2009). This list was compiled based on a composite score with equal weight to “borrowability” (How often is the word with this meaning borrowed?), “age” (How long is the word with this meaning attested on average?), “simplicity” (Do the words with this meaning in average contain more than one morpheme?), and “representation” (Is this meaning well represented

in the languages of the world?).⁸ The resulting list is up to 62% identical with the 100-items Swadesh list (Swadesh 1971: 283).

2.2 *Translation and adjusting the list*

The way the lexicon of languages is organised differs in many ways, and the question of translation, i.e. ‘how to say x in language y?’, is not always easy to answer in an objective way. In fact, it is sometimes near to impossible.

For example, there are cases where the concept as given in the list already leaves room for interpretation. In our case, the Leipzig-Jakarta list has some items with concepts defined with a backslash such as ‘crush/grind’ and ‘hit/beat’. Perhaps some languages have a general term for ‘crush/grind’. But if a language has two words, one for ‘crush’ and one for ‘grind’, or even more than two, as in some Kho-Bwa languages that have distinct vocabulary for ‘grind (smaller grains or grain particles to flour with a hand-turned grinding stone)’, one for ‘grind (larger grains to smaller particles or flour with a water mill)’, one for ‘crush (with the hand, a stick, a hammer or rock)’ and one for ‘crush (with a pestle in a mortar or on a flat stone)’, then which one should be taken? Furthermore, there are cases where old people use an inherited word and younger speakers rather use a loan from another language. Which is then the correct translation? The “original” inherited word or the more common loan?

Geisler and List (2014) identified these problems in translation of concept lists as “concept fuzziness” (concept is not clear), “synonymous differentiation” (more than one word for one concept in English) and “linguistic diversity” (dialect forms and loans). These problems arise all due to language-specific lexicalisation patterns and cannot be foreseen by the compiler of a universal concept list.⁹ Some concepts are problematic for some languages, whereas for other languages it is relatively straightforward.

One way to deal with the translation problem would be to omit “the troublesome item when necessary” (Swadesh 1952: 457) and end up with a shorter list. However, in our case, we decided that in order to retain a list with a sufficient number of entries to be robust for a comparative study, we identified the concepts from the Leipzig-Jakarta list which we found difficult to translate in the Kho-Bwa languages, and replaced them with items from the Swadesh list which were, in our opinion, much less ambiguous.

As far as possible, we tried to replace a part of speech with a similar part of speech (i.e. replace an adjective with an adjective) and a noun from a certain semantic field with a noun from a related field (i.e. a kinship term with another kinship term).

- The body part ‘back’ by ‘fat’
- The kinship term ‘child (reciprocal of parent)’ by ‘woman’
- The verb ‘crush/grind’ by ‘die’
- The verb ‘fall’ by ‘sleep’
- The verb ‘hit/beat’ by ‘kill’
- The body part ‘thigh’ by ‘head’

⁸ Data available on <http://wold.clld.org>.

⁹ Swadesh (1952: 457): “Of course, it would be impossible to devise a list which works perfectly for all languages, and it must be expected that difficult questions will sometimes arise.”

- The noun ‘night’ by ‘moon’
- The noun ‘rope’ by ‘path’
- The noun ‘shade/shadow’ by ‘cloud’
- The noun ‘soil’ by ‘sun’
- The verb ‘suck’ by ‘sit’
- The verb ‘tie’ by ‘dry’
- The verb ‘blow’ by the noun ‘fingernail’
- The verb ‘come’ by adjective ‘white’

While it seems opportunistic to make replacements in the sample - a concern which both our anonymous reviewers expressed - it is in fact the opposite. Through the replacements we avoided having to make arbitrary decisions in the individual languages.

2.3 *Determining cognates*

As noted above (1.3), every cognacy judgment involves - at least implicitly - an analysis of the morphology and sound correspondences. What is the prefix? What is the root? Which sounds correspond?

We went through the list “manually” item by item and decided which are cognates and assigned the same number for items considered to be cognate.¹⁰ We distinguished only between “cognate” and “non-cognate”, and no “partial cognate” or “unsure”. Whenever we were in doubt, we took the careful approach and we judged items as non-cognate. The resulting list can be found in the appendix of this paper (cf. appendix A) and as a csv-spreadsheet on Github.¹¹

To count cognates in a set of languages which are yet to be proven to be related seems circular, i.e. to assume “relatedness” in order to prove “relatedness”. This working hypothesis is necessary for the use of the comparative method for identifying cognates (see, for example, Weiss 2014: 128).¹² One cannot establish regular sound correspondences, without assuming that a set of languages is related. And one cannot prove that languages are related, without having established regular sound correspondences.

There are other approaches for finding cognates or for measuring the similarity of basic vocabulary, like string comparison algorithms (e.g. Brown et al. 2008; List 2014). In this case, strings in arbitrary languages can be compared, and the “relatedness” does not have to be assumed a priori. There is no guarantee, however, that a similarity found by a string comparison algorithm, is indeed a cognate in the traditional sense and not just a lookalike. On the other hand, some cognates which have changed phonologically might be missed. For example, an algorithm will hardly judge Duhumbi *huma* and Rawa Puroik *lip* as very similar. However, Duhumbi *h* goes back to *hl* before *a* and *u*, and

¹⁰ If an item was missing in one of the two varieties compared, it is omitted from the pairwise comparison. For example, in the data of the Rawa dialect of Puroik three items are missing and in the data of the Jerigaon dialect of Sartang two items are missing. This effectively leaves 95 items for the pairwise comparison Rawa-Jerigaon.

¹¹ https://github.com/metroxylon/kho-bwa-lexicostat/blob/master/data/dataset_khobwa.csv.

¹² “The first step in applying Comparative Method is formulating a hypothesis that the given languages to be compared are in fact descended from a common source. It obviously makes little sense to apply the Comparative Method to languages that evidently aren’t related – at any reasonable time depth – and the failure of the procedure to reveal any regularity of correspondence would be a strong argument against a theory of genetic common origin.”

Rawa Puroik final stop very often compares to final nasal. Knowing that possibly Duhumbi *huma* is contracted from **blam-ma*, *-ma* being a common noun suffix, and Rawa Puroik *lip* derives from **lim*, this comparison becomes viable.

2.4 Computation

We computed percentages of shared core vocabulary and wrote it in a table coloured according to the value of the percentage (“heat map”). In order to get a different perspective on the data, we made a hierarchical cluster analysis using an algorithm known as “standard agglomerative method” or UPGMA (Unweighted Pair Group Method with Arithmetic Mean). A write-up about the cluster analysis for this paper can be found on Github¹³ along with the Python code.

One of our reviewers has pointed out the short-comings of the trees to model how languages diversify and split (citing François 2014). We feel that this is an unnecessary concern in this place. The dendrogram is a tool for finding structure in our data and as such neither true nor wrong.¹⁴ A dendrogram can even be useful for exploring data where a phylogenetic interpretation is unlikely.¹⁵

2.5 Results

The heat map and the corresponding dendrogram of the Kho-Bwa languages¹⁶ show three clearly distinct groupings of languages which share higher percentages of cognate core vocabulary: the Puroik varieties, the Bugun varieties, and the group consisting of the varieties of Khispi-Duhumbi, Sartang and Sherdukpen, what we will henceforth call “Western Kho-Bwa” (Figure 2). Further observations regarding the cognacy percentages are:

- Western Kho-Bwa and the Puroik varieties both show about the same degree of internal diversity. Bugun is somewhat more uniform. That the Puroik “dialects” are overall equally or more diverse than the Western Kho-Bwa “languages” is not surprising, given the huge extension of the Puroik language area compared to the small geographic area where we find speakers of Western Kho-Bwa (see map in Figure 1). It is remarkable, however, that in the previous literature Puroik is considered a single language, but the Western Kho-Bwa varieties as three or four languages: clearly, historical, socio-cultural and political arguments underlie this distinction between “dialect” and “language”, not linguistic considerations.
- The Western Kho-Bwa group shows a clear split between Khispi-Duhumbi on the one hand, and the Sartang and Sherdukpen varieties on the other. The Sartang and Sherdukpen varieties are all about equally close to each other.
- The Bugun varieties appear all about equally close to each other.

¹³ <https://github.com/metroxylon/kho-bwa-lexicostat/blob/master/writeup/writeup-cluster-analysis.pdf>.

¹⁴ Everitt et al. (2011: 4) about cluster analysis: “So it should be remembered that in general a classification of a set of objects is not like a scientific theory and should perhaps be judged largely on its usefulness rather than in terms of whether it is ‘true’ or ‘false’”.

¹⁵ An example for a non-phylogenetic dendrogram can be found in the analysis by Tal Galili “Votes for Republican Candidate in Presidential Elections”:

https://cran.r-project.org/web/packages/dendextend/vignettes/Cluster_Analysis.html.

¹⁶ Created with the command: `heatmap_dendrogram plot [path/to/dataset_khobwa.csv]`.

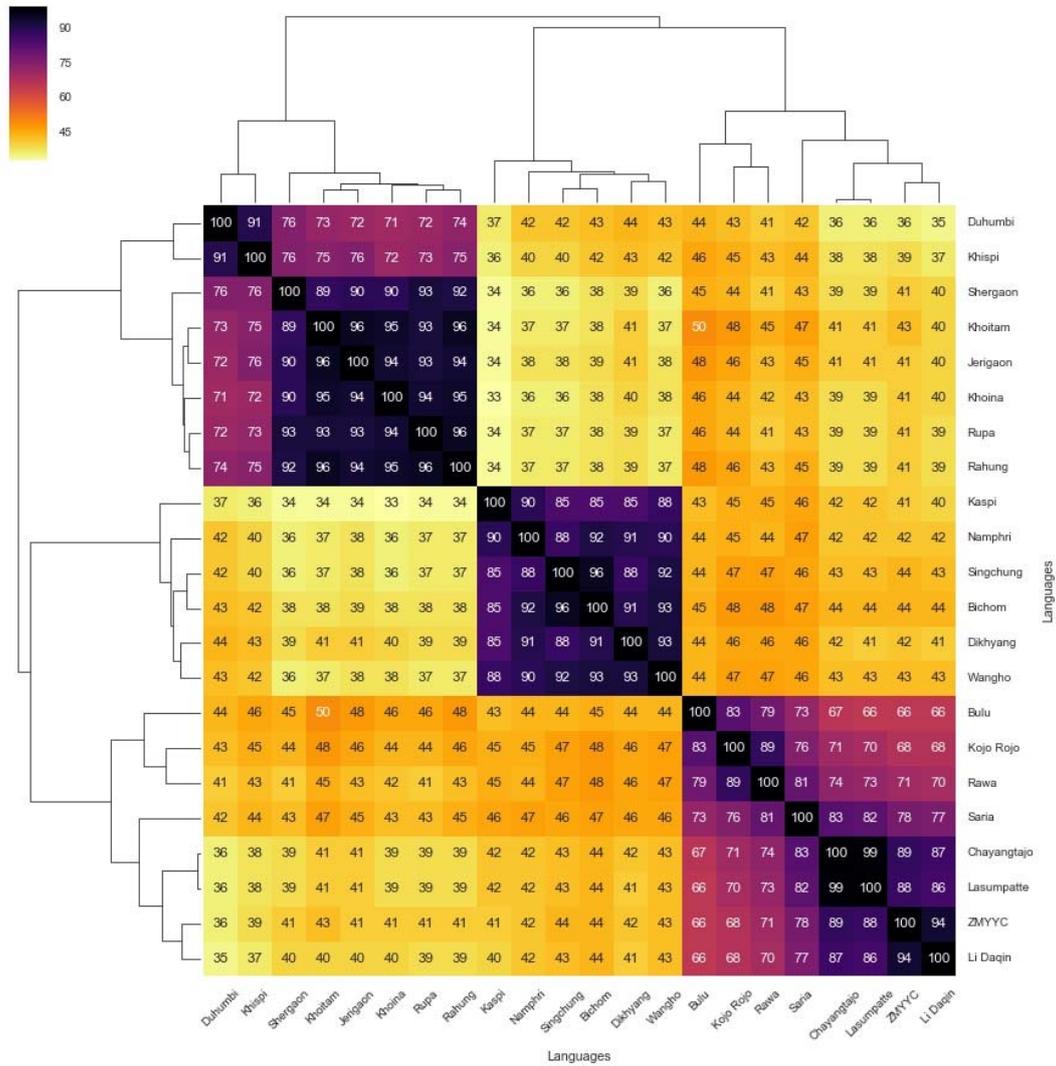
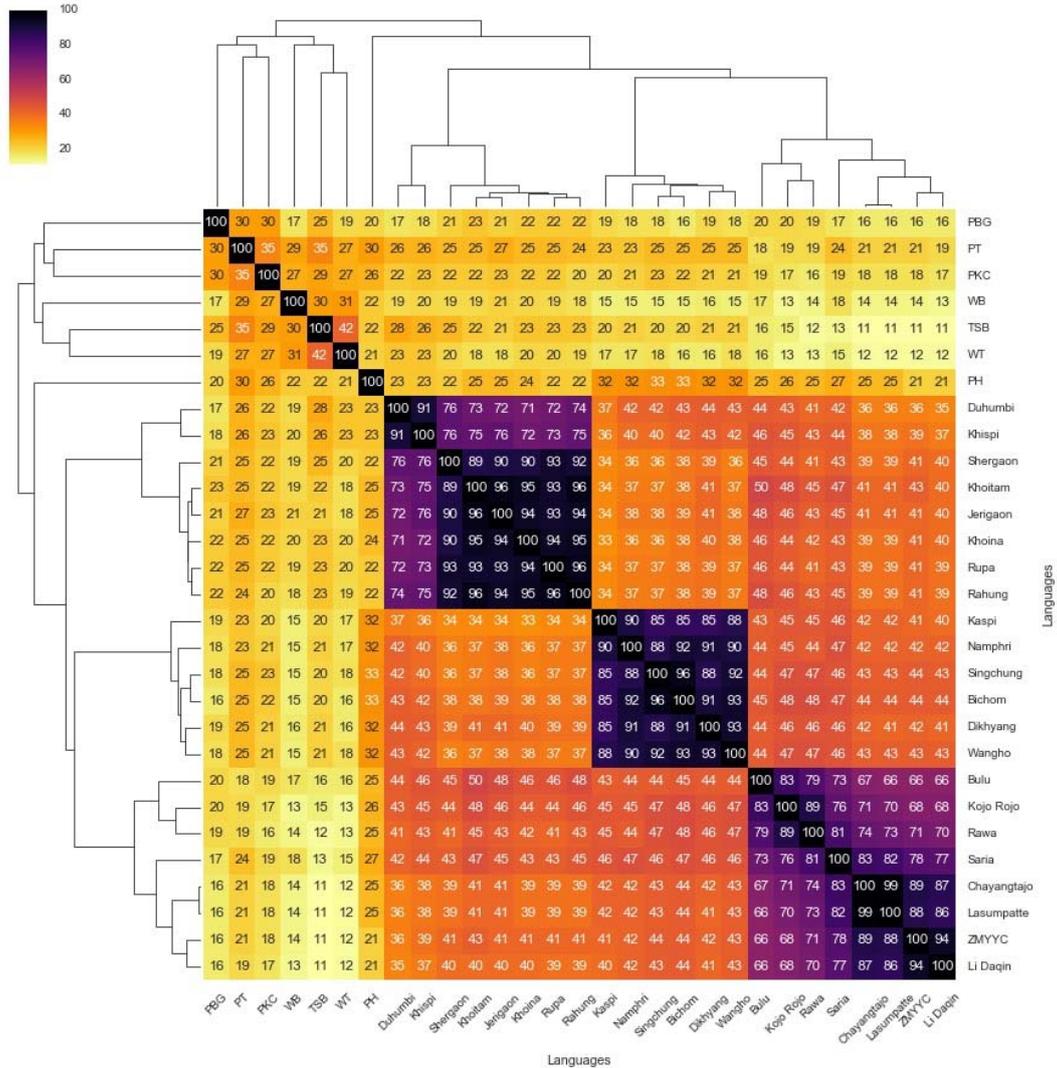


Figure 2. Heat map and dendrogram for Kho-Bwa

- The Puroik varieties show a clear split between west (Bulu, Kojo Rojo, Rawa) and east (the rest). Within the western varieties Bulu Puroik stands apart, which is given its geographic isolation understandable (see map in Figure 1).
- Bugun and the Puroik varieties are somewhat closer to each other than to the Western Kho-Bwa varieties.

When other Tibeto-Burman languages are included in the analysis, the heat map shows that none of the Kho-Bwa languages is closer to any other Tibeto-Burman language - not even intense contact languages - than to the most distant other Kho-Bwa language (Figure 3). An exception is Hrusish, which is probably a contact induced similarity.

Lieberherr and Bodt: Sub-grouping Kho-Bwa based on shared core vocabulary



There are 26 roots that are cognate in most of the Kho-Bwa varieties. These roots are: the pronouns 1SG, 2SG, 3SG (except Rawa Puroik and Kaspi Bugun); the body parts FAT (n.), FINGERNAIL, BLOOD, HAIR (ON HEAD), HEAD, ARM/HAND, LEG/FOOT, NAVEL, SKIN, NOSE and TONGUE; the verbs DIE, KILL and CRY; the cultural concepts NAME and PATH (except easternmost Puroik); the natural elements FIRE, LEAF, MOON (except Rawa Puroik), SMOKE, WATER and WOOD; and the adjective HEAVY. Figure 3. Heat map and dendrogram including other Tibeto-Burman languages

2.6 *Are the similarities due to inheritance?*

We chose a list of concepts which do not tend to be universally similar (such as onomatopoeic words, baby language) and concepts which are usually not borrowed. However, assuming that in principle “everything can be borrowed”, geography could still be a factor for explaining the picture obtained. The dendrogram does not necessarily have to be interpreted as a phylogenetic tree in the sense that languages under one node shared a common phase of development. What looks like a genealogical tree could, after all, be a tree-shaped geographic map where geographically close languages are grouped under one node. However, looking at Figures 2 and 3 again, several of the percentages indicate that language contact situations had little effect on the list of core vocabulary we investigated.

- Bulu is geographically closer to almost any Western Kho-Bwa language or any Bugun variety than to any other Puroik dialect (see map in Figure 1). 'the language has had very few speakers, never more than 20-30 in the last 100 years. All speakers of Bulu Puroik are perfectly bilingual in both Puroik and Miji, with Miji being more commonly spoken in the village at present. In addition, most senior villagers can converse fluently in Dirang Tshangla and Brokpa. This is a typical situation where one would expect a high degree of language mixture, which is indeed the case. However, language contact did not affect the core vocabulary investigated in this paper. Bulu Puroik lines up clearly with the other Puroik dialects. The dendrogram does not show in any way that Bulu Puroik is a language in West Kameng geographically close to Bugun and Western Kho-Bwa.
- The speakers of Sartang and Bugun and the speakers of Sherdukpen and Bugun are immediate neighbors that have lived in close association for a considerable time. Nonetheless, lexically Bugun is clearly separate, and both Sherdukpen and Sartang are lexically closer to Khispi-Duhumbi than to Bugun. This even though the Khispi-Duhumbi speech area forms a geographic outlier, separated from Sartang and Sherdukpen by the Tshangla speech area. Although the situation is different in other semantic fields (such as advanced implements, emotions and feelings etc.), the basic vocabulary presented by the Leipzig-Jakarta list is indicative of a genetic relation rather than borrowing.
- Puroik and Bugun are under one node, even though Bugun is geographically much closer to all Western Kho-Bwa languages than to Puroik. However, this node is less clear than others.
- Another remarkable fact is that the lowest cognacy percentage we found is for the pair Written Tibetan and the Puroik varieties recorded by the Chinese authors in Tibet. This implies that the genetic relation with and influence of Tibetan on these two Puroik varieties is more limited than on any other Kho-Bwa or Tibeto-Burman varieties discussed here, and may also support the assumption that the Puroik informants in these sources were not native to Tibet itself but had come from across the border.

2.7 *Validity of groups*

To what extent can the groupings obtained in Figures 2 and 3 be trusted? Would the picture be completely different if some cognacy judgments were wrong or other linguists came to different conclusions?

We didn't answer this question in a strict statistical sense. To get a rough impression into the effect that other decisions might have on our result, we added a matrix with random integers between -20 and + 20 to our distance matrix (see Figure 4).

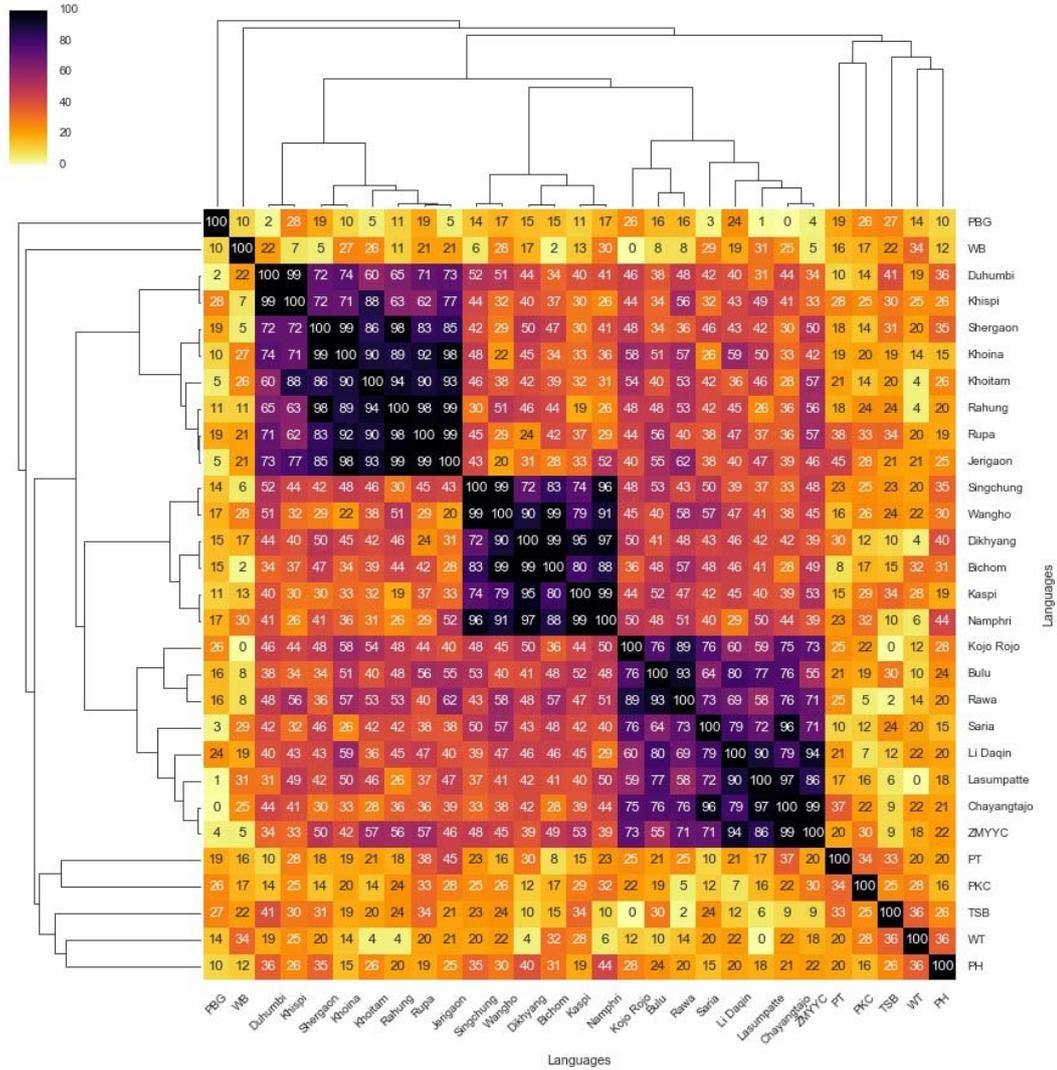


Figure 4. Dendrogram with randomly introduced variation +/- 20 percent

Introducing this random variation did not alter the main outcomes of our study, namely that the three subgroups Western Kho-Bwa, Bugun and Puroik remain intact, and that these three subgroups remain significantly distinct from other Tibeto-Burman languages and proto-languages in the study.

3 Phonology, morphology and other considerations

Investigating a small set has the advantage that a large number of languages can be surveyed and a tentative picture of the situation can be given, relatively fast and with relatively little effort. We

acknowledge, however, the limitations of this approach. Future comparative studies of the lexicon, phonology, and eventually morphology will give a more detailed picture of this group of languages. We will list here a few further considerations regarding the status of Kho-Bwa. First, it is important to note that in some semantic fields the cognacy percentage is much higher. In some cases, whole lexical paradigms are clearly cognate, e.g. the singular pronouns (cf. Table 1 for a selection of varieties).

		language	1SG	2SG	3SG	
Kho-Bwa	WKB	dh	<i>ga</i>	<i>naŋ</i>	<i>wɔj</i>	
		rp	<i>gu</i>	<i>naŋ</i>	<i>wa</i>	
		kt	<i>gu</i>	<i>naŋ</i>	<i>wa</i>	
		Bugun	dk	<i>ko:</i>	<i>nɔ̃</i>	<i>oε</i>
		Puroik	bl	<i>gu:</i>	<i>na:</i>	<i>ve:</i>
	kr		<i>go:</i>	<i>naŋ</i>	<i>wai</i>	
	ct		<i>go:</i>	<i>na:</i>	<i>wε:</i>	
TB		pt	<i>*ŋo:</i>	<i>*no:</i>	-	
		ph	<i>*na(-jaŋ)</i>	<i>*ni</i>	<i>*ʔi</i>	
		tsb	<i>ɕzaŋ</i>	<i>nan</i>	<i>dan/rɔk</i>	

Table 1. Singular pronouns

In addition to the shared the roots the Kho-Bwa languages often have cognate lexical affixes (2). For example, words for parts of the head (HEAD, HAIR, EYE, EAR) in Kho-Bwa languages contain an overt or a fused velar prefix, e.g. HAIR ON HEAD dh *ku-ɕaŋ*, rp *gɔ-zəŋ*, kt *gɔ-zəŋ*, dk *ka-zijɔŋ*, bl *kə-zəN*, kr *kə-zjaŋ*, ct *kə-zak*. Another lexical prefix characteristic for the Kho-Bwa languages is a prefix for celestial objects and weather phenomena (RAIN, SNOW, CLOUD, MOON, SUN, STAR). For example, MOON dh *nam-ba*, rp *nam-blu*, kt *nam-blu*, dk *ha-bie:*, bl *ham-bo:*, kr *ha-bu*, ct *am-bo:*. While the forms are not straightforwardly cognate, they are nevertheless similar, and it is remarkable that all languages of the group seem to have a particular prefix for this semantic domain. Finally, all the Kho-Bwa languages have an adjective prefix *a-* (or going back to **a-* following regular patterns of phonological change), distinguishing these languages from neighbouring Miji-Bangru where the adjective prefix is **mə-*. For example, HEAVY dh *u-li*, rp *a-li:*; kt *a-li:*; dk *ə-lai*, bl *a-li:*; kr *a-lei*, ct *a-lei*.

prefix	function
<i>a-</i>	adjectives (≠ Miji-Bangru adjective prefix <i>mə-</i>)
<i>k-</i>	parts of the head
<i>nam-</i> ~ <i>ham-</i>	“sky”-prefix for celestial objects and weather phenomena

Table 2. Lexical prefixes shared by Kho-Bwa languages

Of great importance for the classification of languages is the question whether a group of languages shares uncommon phonological innovations. The more uncommon shared phonological innovations are, the more likely it is that they didn't happen independently and that the languages go back to a common ancestor. All Kho-Bwa languages show some evidence for a syllable initial

denasalisation, at least for the bilabial place of articulation, which is a non-trivial change and almost unique in Tibeto-Burman.¹⁷ The prime example is the root for FIRE which has a bilabial plosive *b* as onset (other examples are NEGATION, DREAM, NAME, PERSON see Table 3). Cognates in all surrounding languages have a nasal continuant onset. Whether this development is also found for other places of articulation is a matter of debate. One candidate is the 1SG pronoun. However, the plosive onset for this pronoun is also found in other Tibeto-Burman languages.

	language	FIRE	DREAM	NEG	NAME	PERSON ¹⁸	1SG	
Kho-Bwa	WKB	dh	<i>bɛj</i>	<i>bɛn-</i> <i>kan</i> ¹⁹	<i>ba-</i>	<i>biŋ</i>	<i>bu-dun</i>	<i>ga</i>
		rp	<i>ba:</i>	<i>ban</i>	<i>ba-</i>	<i>a-zɛŋ</i>	<i>ɖə-riŋ</i>	<i>gu</i>
		kt	<i>bɛ:</i>	<i>ban</i>	<i>ba-</i>	<i>a-ɖɛŋ</i>	<i>ɖə-riŋ</i>	<i>gu</i>
	Bugun	dk	<i>bo:ɛ</i>	<i>bɔŋ</i>	-	<i>ə-bɛŋ</i>	<i>b-ran</i>	<i>ko:</i>
	Puroik	bl	<i>bɛ:</i>	<i>baN</i>	<i>ba-</i>	<i>a-bjɛN</i>	<i>p-riN</i>	<i>gu:</i>
		kr	<i>bai</i>	<i>baŋ</i>	<i>ba-</i>	<i>a-bɛŋ</i>	<i>bi:</i>	<i>go:</i>
		ct	<i>bɛ:</i>	<i>bak</i>	<i>ba-</i>	<i>a-bɛŋ</i>	<i>bi:</i>	<i>go:</i>
Other TB	ph	<i>*maj</i>	<i>*tai-mə</i>	<i>*ma-</i>	<i>*mə-mjiŋ</i>	<i>*nji</i>	<i>*na(-jaŋ)</i>	
	pt	<i>*mə</i>	<i>*juɸ-</i> <i>maŋ</i>	<i>*maŋ</i>	<i>*muun</i> <i>mruuŋ</i>	~ <i>*mi</i>	<i>*ŋo:</i>	
	tsb	<i>mi</i>	<i>moŋ-ɛi</i> ²⁰	<i>ma-</i>	<i>miŋ</i>	-	<i>ɖzaŋ</i>	
	wt	<i>me</i>	-	<i>ma-</i> , <i>mi-</i>	<i>miŋ</i>	<i>mi</i>	<i>ŋa</i>	
	wb	<i>mí</i>	<i>meʔ</i>	<i>ma-</i>	<i>nà myì</i>	-	<i>ŋà, kò</i>	
	pbg	<i>*bwar</i> ²	<i>jV³-maŋ</i>	<i>*-ya⁰</i>	<i>*muŋ</i>	-	<i>*aŋ¹</i>	
	pkc	<i>*may</i>	<i>*maŋ</i>	-	<i>*miŋ bmiŋ</i>	<i>*mii</i>	<i>*kay kay-</i> <i>maʔ</i>	

Table 3. Kho-Bwa **b* vs. *m* in other TB languages

If the etymologies in Table 3 are correct, it is a strong phonological argument for the coherence of the Kho-Bwa group. Being typologically uncommon, the change **m* > *b* is unlikely to be an independent innovation. Furthermore, the concepts FIRE, DREAM, NEG, NAME and PERSON belong to the core vocabulary and are relatively resistant to borrowing for these roots with this divergent plosive onset (see World Loan Word Database²¹). Borrowing of several of these words into several languages would be a strong assumption. Even though this could, in principle, have happened, there is still no plausible Tibeto-Burman source for these roots, other than the Kho-Bwa languages itself. Dismissing independent innovation and borrowing, the most parsimonious scenario is that the

¹⁷ It occurs in Southern Loloish Bisu, e.g. PLB FIRE **mey²* > Bisu *bi* (Matisoff (2003, p. 39)). The root for NAME has voiced plosive onset in Lepcha *ɭá-bryáŋ* (Plaisier 2007) and Nungic Trung *aŋ³¹b.uuŋ⁵³* (Sün 1991).

¹⁸ Prefix in Western Kho-Bwa, Bugun and Bulu Puroik corresponds to the root in the other Puroik dialects. The correspondence of *b* to *ɖ* in Sherdukpen (rp) and Sartang (kt) is regular.

¹⁹ root-nominaliser.

²⁰ root-suffix

²¹ <http://wold.clld.org>.

Kho-Bwa languages have a common ancestor and that the innovation happened before the split into different languages and varieties. The data in Table 3 itself are already a strong argument that the Kho-Bwa languages share a common ancestor (i.e. that they are genetically related), and that they are Tibeto-Burman.

Another shared phonological feature of the Kho-Bwa languages, which is almost certainly a shared innovation, is the absence of any reflex where other languages have *s* or *th* or (see Table 4). Like **m > b* this feature distinguishes Kho-Bwa from all neighbouring languages.

		language	DIE	KILL	THREE	
Kho-Bwa	WKB	dh	<i>i</i>	<i>at</i>	<i>ɔm</i>	
		rp	<i>iː</i>	<i>ɔ</i>	<i>uŋ</i>	
		kt	<i>iː</i>	<i>ɔ</i>	<i>uŋ</i>	
		Bugun	dk	<i>iː</i>	<i>uə</i>	<i>ũm</i>
		Puroik	bl	<i>iː</i>	<i>weʔ</i>	<i>im</i>
			rw	<i>iː</i>	<i>at</i>	<i>ɨp</i>
			ct	<i>iː</i>	<i>aiʔ</i>	<i>uk</i>
Other TB		ph	<i>*əj</i>	-	<i>*gə-ə̃m</i>	
		pt	<i>*si</i>	-	<i>*hum</i>	
		tsb	<i>ɛi</i>	<i>ɛe</i>	<i>sam</i>	
		wt	<i>si</i>	<i>bsad-pa</i>	<i>gsum</i>	
		wb	<i>ei</i>	<i>aʔ</i>	<i>thoun:</i>	
		pbg	<i>*thui¹</i>	-	<i>*tham²</i>	
		pkc	<i>*thii, *thiʔ</i>	<i>*that-I, *-thaʔ-II</i>	<i>*tbum</i>	

Table 4. Kho-Bwa \emptyset vs. reflexes of *s* in other TB languages

A morpho-syntactic feature shared by the Kho-Bwa languages is the preverbal negation. In Tshangla, Bodish, Miji, Bangru and Hruso the negation is equally preverbal. However, in Tani negation is post-verbal.

4 Conclusion

The analysis of cognate core vocabulary in section 2 shows that even the geographically most distant Kho-Bwa languages share a higher percentage of cognate core vocabulary than with any of the geographically close non-Kho-Bwa contact languages. Within Kho-Bwa, there are three groups sharing higher percentages of cognate core vocabulary (Western Kho-Bwa, Bugun, Puroik), hence, in terms of core vocabulary, Kho-Bwa is a consistent group with three sub-groups: Western Kho-Bwa, Bugun and Puroik. Important tasks for the future will be the documentation and analysis of Kho-Bwa varieties, the reconstruction of low level subgroups such as Proto-Western-Kho-Bwa, Proto-Puroik, Proto-Bugun, and ultimately the reconstruction of Proto-Kho-Bwa.

ABBREVIATIONS

bc	Bichom (Bugun)	pkc	Proto-Kuki-Chin (VanBik 2009)
bl	Bulu (Puroik)	plb	Proto-Lolo-Burmese

ct	Chayangtajo (Puroik)	pt	Proto-Tani (Sun 1993)
dh	Duhumbi	ptb	Proto-Tibeto-Burman
dh	Dikhyang (Bugun)	rh	Rahung (Sartang)
kg	Jerigaon (Sartang)	rp	Rupa (Sherdukpen)
kn	Khoina (Sartang)	rw	Rawa (Puroik)
ka	Kaspi (Bugun)	s100	Swadesh list 100 words (Swadesh 1971)
kp	Khispi	sc	Singchung (Bugun)
kr	Kojo Rojo (Puroik)	sg	Shergaon (Sherdukpen)
kt	Khoitam (Sartang)	sr	Sario Saria (Puroik)
li	Puroik China (Li 2004)	tsb	Bhutan Tshangla
lj	Leipzig-Jakarta list (Tadmor 2009)	wb	Written Burmese
lp	Lasumpatte (Puroik)	wh	Wangho (Bugun)
ng	Namphri (Bugun)	wkb	Western Kho-Bwa
pbg	Proto-Bodo-Garo (Joseph and Burling 2006)	wt	Written Tibetan
ph	Proto-Hrusish	zm	Puroik China (Sun et al. 1991)
pkb	Proto-Kho-Bwa		

REFERENCES

- Abraham, Binny et al. 2005. "A sociolinguistic research among selected groups in Western Arunachal Pradesh highlighting Monpa". Unpublished manuscript.
- Anderson, Gregory D. S.; and Murmu, Ganesh. 2010. "Preliminary notes on Koro, a 'hidden' language of Arunachal Pradesh". *Indian Linguistics* 71: 1–32.
- Barbora, Madhumita. 2015. *Bugun nyo thau: Bugun reader: A collection of Bugun folk tales, stories, proverbs, songs, rituals and lexical items*. Guwahati: EBH Publishers.
- Blench, Roger. 2015. "The Mey languages and their classification". Paper presented at the CSEHEP. University of Sydney. url: https://www.academia.edu/15108029/The_Mey_languages_and_their_classification (retrieved on 08/01/2016).
- Blench, Roger; and Post, Mark W. 2014. "Rethinking Sino-Tibetan phylogeny from the perspective of Northeast Indian languages". In: Thomas Owen-Smith and Nathan W. Hill (eds.), *Trans-Himalayan-Linguistics*, 71-104. Berlin: de Gruyter.
- Bodt, Timotheus Adrianus. 2017. *Grammar of Duhumbi (Chugpa)*. PhD thesis, Universität Bern.
- Bodt, Timotheus Adrianus. 2014. "Ethnolinguistic survey of westernmost Arunachal Pradesh: A fieldworker's impressions". *Linguistics of the Tibeto-Burman Area* 37.2: 198-239.
- Bodt, Timotheus Adrianus; and Lieberherr, Ismael. 2015. "First notes on the phonology and classification of the Bangru language of India". *Linguistics of the Tibeto-Burman Area* 38.1: 66–123.
- Brown, Cecil H. et al. 2008. "Automated classification of the world's languages: a description of the method and preliminary results". *STUF - Language Typology and Universals* 61.4: 285-308.

- Burling, Robbins. 2003. "The Tibeto-Burman languages of Northeastern India". In: Graham Thurgood and Randy J. LaPolla (eds.), *The Sino-Tibetan Languages*, 169–191. New York: Routledge [Language Family Series].
- Campbell, Lyle; and Poser, William J. 2008. *Language classification: History and method*. Cambridge: Cambridge University Press.
- Deuri, R. K. 1983. *The Sulungs*. Shillong: Research Department, Government of Arunachal Pradesh.
- Dondrup, Rinchin. 1988. *A Handbook on Sherdukpen language*. Itanagar: Government of Arunachal Pradesh.
- Dondrup, Rinchin. 1990. *Bugun language guide*. Itanagar: Government of Arunachal Pradesh.
- Dondrup, Rinchin. 2004. *An introduction to Boot Monpa language*. Itanagar: Government of Arunachal Pradesh.
- van Driem, George. 2001. *Languages of the Himalayas – An ethnolinguistic handbook of the Greater Himalayan Region*. Vol. 2. Leiden: Brill.
- Everitt, Brian S. et al. 2011. *Cluster analysis*. London: John Wiley and Sons [Wiley Series in Probability and Statistics].
- François, Alexandre. 2014. "Trees, waves and linkages: Models of language diversification". In: Claire Bowerman and Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 161-189. London and New York: Routledge.
- Geisler, Hans; and List, Johann-Mattis. 2014. "Beautiful trees on unstable ground: Notes on the data problem in lexico-statistics". In: Heinrich Hettrich (ed.), *Die Ausbreitung des Indogermanischen: Thesen aus Sprachwissenschaft und Archäologie*. Wiesbaden: Reichert. url: <https://hal.archives-ouvertes.fr/hal-01298493>.
- Haspelmath, Martin; and Tadmor, Uri (eds.). 2009. *Loanwords in the world's languages: A comparative handbook*. Berlin: Walter de Gruyter.
- Hill, Nathan W. 2010. *A Lexicon of Tibetan Verb Stems as Reported by the Grammatical Tradition*. Studia Tibetica. Munich: Bayerische Akademie der Wissenschaften.
- Jacquesson, François. 2015. *An introduction to Sherdukpen*. Vol. 39. Diversitas Linguarum. Bochum: Universitätsverlag Dr. N. Brockmeyer.
- Jäschke, Heinrich August. 1881. *A Tibetan-English dictionary: With special reference to the prevailing dialects, to which is added an English-Tibetan vocabulary*. London: Routledge and Kegan Paul.
- Joseph, U.V.; and Burling, Robbins. 2006. *Comparative phonology of the Boro Garo languages*. Mysore: Central Institute of Indian Languages Publication.
- Lǐ, Dàqín. 2004. *Sūlóngyǔ yánjiū* (Research on Puroik). Běijīng: Mǐnzú chūbǎn shè (National Minorities Publisher).
- Lieberherr, Ismael. 2017. *A grammar of Bulu Puroik*. PhD thesis, Universität Bern.
- Lieberherr, Ismael. 2015. "A progress report on the historical phonology and affiliation of Puroik". In: Linda Konnerth et al. (eds.), *North East Indian linguistics* Vol. 7, 235–286. Canberra: Asia-Pacific Linguistics, Australian National University.
- List, Johann-Mattis. 2014. *Sequence comparison in historical linguistics*. Düsseldorf: Düsseldorf University Press [Dissertations in Language and Cognition].
- List, Johann-Mattis; Cysouw, Michael; and Forkel, Robert (eds.). 2016. *Concepticon*. Jena: Max Planck Institute for the Science of Human History. url:

- <http://concepticon.clld.org/> (retrieved on 01/07/2017).
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibeto reconstruction*. Berkeley: University of California Press.
- Plaisier, Heleen. 2007. *A grammar of Lepcha*. Leiden and Boston: Brill.
- Post, Mark W.; and Burling, Robbins. 2017. "The Tibeto-Burman languages of Northeastern India". In: Graham Thurgood and Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 213-233. London: Routledge.
- Remsangpuia. 2008. *Puroik phonology*. Shillong: Don Bosco Center for Indigenous Cultures.
- Rutgers, Leopold Roland. 1999. Puroik or Sulung of Arunachal Pradesh. Paper presented at the 5th Himalayan Languages Symposium. Kathmandu Guest House: Kathmandu.
- Shafer, Robert. 1947. "Hruso". *Bulletin of the School of Oriental and African Studies* 12: 184-196.
- Soja, Rai. 2009. *English-Puroik dictionary*. Shillong: Living Word Communicators.
- Stonor, Charles Robert. 1952. "The Sulung tribe of the Assam Himalayas". *Anthropos* 47: 947-962.
- Sūn, Hóngkǎi et al. 1991. *Zàng-Miǎn-yǔ yǎyīn hé cíhuì* (Tibeto-Burman Phonology and Lexicon). Běijīng: Zhōngguó shèhuì kēxué chūbǎn shè (Chinese Social Sciences Press).
- Sun, T. -S. Jackson. 1992. "Review of Zangmianyu Yuyin He Cihui (Tibeto-Burman Phonology and Lexicon)". *Linguistics of Tibeto-Burman Area* 15: 73-113.
- Sun, T. -S. Jackson. 1993. *A historical-comparative study of the Tani (Mirish) branch in Tibeto-Burman*. PhD thesis, University of California, Berkeley.
- Swadesh, Morris. 1952. "Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos". *Proceedings of the American philosophical society* 96.4, 452-463.
- Swadesh, Morris. 1971. *The origin and diversification of language*. New Jersey: Transaction Publishers.
- Tadmor, Uri. 2009. "Loanwords in the world's languages: Findings and results". In: Martin Haspelmath and Uri Tadmor (Eds.), *Loanwords in the world's languages. A comparative handbook*, 55-75. Berlin: Mouton de Gruyter.
- Tayeng, Aduk. 1990. *Sulung language guide*. Itanagar: Directorate of Research, Government of Arunachal Pradesh.
- VanBik, Kenneth. 2009. *Proto-Kuki-Chin: A reconstructed ancestor of the Kuki-Chin languages*. Berkeley: University of California [STEDT Monograph 8].
- Weiss, Michael. 2014. "The comparative method". In: Claire Bowerman and Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 127-145. London and New York: Routledge.

Appendix A

The comparative word list contains the data of the Kho-Bwa language varieties discussed in this paper. Entries with the same number are considered to be cognate. In addition to the Kho-Bwa varieties there is a tentative comparison with other Tibeto-Burman languages. These are: Bhutan Tshangla (personal database of Bhutan Tshangla), Written Burmese (SEAlang),²² Written Tibetan (Jäschke, 1881; Hill 2010; and Nitartha Online Tibetan-English dictionary),²³ Proto-Tani (Sun, 1993), Proto-Bodo-Garo (Joseph and Burling 2006), Proto-Hrusish (Bodt and Lieberherr 2015), and Proto-Kuki-Chin (VanBik 2009). Besides the standard characters of the International Phonetic Alphabet the character -N is used for a “placeless nasal coda” which is realised as a nasalisation of the preceding vowel or a nasal segment [m, n, ŋ] depending on the environment. A reference to the concept list is in square brackets after the concept (LJ=Leipzig-Jakarta, S100=Swadesh 100).

1SG	[LJ, S100]	2SG	[LJ, S100]	3SG	[LJ]			
dh	ga	(1)	dh	naŋ	(1)	dh	wəj	(1)
kp	ga	(1)	kp	naŋ	(1)	kp	wuj	(1)
rp	gu	(1)	rp	naŋ	(1)	rp	wa	(1)
sg	gu	(1)	sg	naŋ	(1)	sg	wa	(1)
rh	gu	(1)	rh	naŋ	(1)	rh	he ~ ?ε	(1)
kt	gu	(1)	kt	naŋ	(1)	kt	wa-dzi	(1)
jg	gu	(1)	jg	naŋ	(1)	jg	wa	(1)
kn	gu	(1)	kn	naŋ	(1)	kn	laŋ-wa	(1)
df	ko:	(1)	df	nō	(1)	df	oε	(1)
sc	go	(1)	sc	no	(1)	sc	oe	(1)
wh	ko	(1)	wh	no	(1)	wh	we	(1)
bc	go ~ gu	(1)	bc	no	(1)	bc	uwe	(1)
ka	gu	(1)	ka	no	(1)	ka	t ^h e	(2)
np	go	(1)	np	no	(1)	np	oe	(1)
bl	gu:	(1)	bl	na:	(1)	bl	ve:	(1)
kr	go	(1)	kr	naŋ	(1)	kr	wai	(1)
rw	go	(1)	rw	na	(1)	rw	ho:	(3)
sr	go	(1)	sr	na	(1)	sr	wa:	(1)
ct	go	(1)	ct	na	(1)	ct	wε:	(1)
lp	go	(1)	lp	na	(1)	lp	wε:	(1)
zm	goh ⁵³	(1)	zm	nah ⁵³	(1)	zm	væ ³³	(1)
li	goh ⁵⁵	(1)	li	na ⁵⁵	(1)	li	va ⁵⁵	(1)
tsb	dzaŋ	(2)	tsb	nan	(1)	tsb	rək	(4)
wb	ŋà/kò	(2)	wb	ní, nī, nī	(2)	wb	θù	(2)
wt	ña	(2)	wt	khyed	(3)	wt	kho	(5)
pt	*ŋo:	(2)	pt	no:	(1)	pt	NA	
pbg	*aŋ ¹	(2)	pbg	*na ² /*naŋ ²	(1)	pbg	*u ¹ -	(6)
ph	*na(-jaŋ)	(2)	ph	*ni	(2)	ph	*i	(7)
pkc	*kay kay-ma?	(1)	pkc	*naŋ	(1)	pkc	NA	

²² <http://sealang.net/burmese/dictionary.htm>.

²³ <http://www.nitartha.org>.

Lieberherr and Bodt: Sub-grouping Kho-Bwa based on shared core vocabulary

ANT		[LJ]	tsb	ga-danj	(5)	bl	a-ra:	(3)
dh	k ^h in-tɕ ^h ək	(1)	wb	leʔ	(6)	kr	a-rai	(3)
kp	k ^h e-tɕ ^h ək	(1)	wt	lag-pa	(6)	rw	a-rai	(3)
rp	san-dək	(1)	pt	*lak	(6)	sr	toʔ	(1)
sg	san-dək	(1)	pbg	*yak	(6)	ct	kəla:	(4)
rh	san-dək	(1)	ph	*mə-guC	(7)	lp	kəla:	(4)
kt	san-dək	(1)	pkc	*kut khut	(1)	zm	a ³³ -bua ³³	(2)
jg	san-dək	(1)				li	a ³¹ -bua ⁵⁵	(2)
kn	san-dək	(1)	ASH		[LJ, S100]	tsb	ka-tanj	(5)
df	ci-pinj	(2)	dh	tɕ ^h u-p ^h ak	(1)	wb	tʃi (kji:)	(6)
sc	ʃinj	(2)	kp	tɕ ^h u-p ^h a	(1)	wt	chen-po	(7)
wh	ʃinj	(2)	rp	tɕ ^h a-p ^h lu:	(1)	pt	*tə- ta	(1)
bc	siinj	(2)	sg	tɕ ^h a-p ^h u:	(1)	pbg	-ma ⁰	(8)
ka	ʃinj	(2)	rh	tɕ ^h a-p ^h lu:	(1)	ph	*mə-doʔ	(1)
np	ʃinj	(2)	kt	tɕ ^h a-p ^h lu:	(1)	pkc	*lian-II, lianʔ-II	(9)
bl	dʒam-dʒuʔ	(1)	jg	tɕ ^h a-p ^h lu:	(1)	►		
kr	gam-gɔ	(1)	kn	dʒe-p ^h lu:	(1)	BIRD		[LJ, S100]
rw	ga-mə-gɔ:	(1)	df	sə-p ^h iə	(2)	dh	hwa	(1)
sr	gɔɛŋ-gɔ	(1)	sc	s-pije	(2)	kp	hwa	(1)
ct	gɔɛŋ-gɔ	(1)	wh	s-pije	(2)	rp	hu	(1)
lp	gɔɛŋ-gɔ	(1)	bc	s-pije	(2)	sg	hu	(1)
zm	gɔɛn ³³ -gɔ ⁵³	(1)	ka	s-pije	(2)	rh	hu	(1)
li	gran ³¹ -ko ⁵³	(1)	np	s-pije	(2)	kt	hu	(1)
tsb	k ^h an-dzi-la	(3)	bl	p ^h a-rəp	(3)	jg	hu	(1)
wb	pəyweʔ s ^h eiʔ	(4)	kr	bə-rəm	(3)	kn	fu	(1)
wt	grog-ma	(1)	rw	bai-rəp	(3)	df	fua	(1)
pt	*ruk ~ rup	(1)	sr	bai-rəʔ	(3)	sc	fua	(1)
pbg	NA		ct	bɛɛ-rak	(3)	wh	fua	(1)
ph	*si-ni	(2)	lp	bɛɛ-rak	(3)	bc	fua	(1)
pkc	NA		zm	bæ ³³ -ɣək ³³	(3)	ka	fo	(1)
			li	bæ ³³ -ɣəh ⁵³	(3)	np	fua	(1)
ARM/HAND		[LJ, S100]	tsb	t ^h u-lu	(4)	bl	pə-du:	(2)
dh	hut	(1)	wb	pyà	(2)	kr	pə-do:	(2)
kp	hut	(1)	wt	thal-ba	(4)	rw	pə-do:	(2)
rp	ik	(2)	pt	NA		sr	pə-do:	(2)
sg	ik	(2)	pbg	*thap ⁴ -la	(4)	ct	pə-dou	(2)
rh	ik	(2)	ph	NA		lp	pə-dou	(2)
kt	ik	(2)	pkc	*wut	(5)	zm	pə ³³ -tu ³³	(2)
jg	ik	(2)				li	pu ³¹ -tu ⁵⁵	(2)
kn	ik	(2)	BIG		[LJ, S100]	tsb	k ^h a	(3)
df	wat	(3)	dh	ə-dək	(1)	wb	ŋeʔ	(4)
sc	wat	(3)	kp	ə-dək	(1)	wt	bya	(5)
wh	wat	(3)	rp	a-dək	(1)	pt	*tanj	(6)
bc	wat	(3)	sg	a-dək	(1)	pbg	*tao ²	(2)
ka	wat	(3)	rh	a-dək	(1)	ph	*bə-dow	(2)
np	wat	(3)	kt	a-dək	(1)	pkc	*waa	(1)
bl	geʔ	(4)	jg	a-dək	(1)			
kr	geiʔ	(4)	kn	a-dək	(1)	BITE		[LJ, S100]
rw	get	(4)	df	ə-pō:	(2)	dh	zək	(1)
sr	geiʔ	(4)	sc	e-bo ^u	(2)	kp	zək	(1)
ct	geiʔ	(4)	wh	e-bo ^u	(2)	rp	zək	(1)
lp	geiʔ	(4)	bc	bo ^u	(2)	sg	zək	(1)
zm	geʔ ³³	(4)	ka	bo ^u	(2)	rh	zək	(1)
li	gi ⁵⁵	(4)	np	bo ^u	(2)	kt	zək	(1)

Lieberherr and Bodt: Sub-grouping Kho-Bwa based on shared core vocabulary

ph	NA		lp	mə-ri:	(5)	wh	ruwanj	(1)
pkc	*naa hnaa	(3)	zm	mə ³³ -ri ³³	(5)	bc	r ^h anj	(1)
EAT	[LJ, S100]		li	mur ³¹ -ti ⁵⁵	(5)	ka	r ^h anj	(1)
dh	tɕ ^h a	(1)	tsb	goi-t ^h am	(6)	np	r ^h anj	(1)
kp	tɕ ^h a	(1)	wb	u	(3)	bl	a-tʃoi	(2)
rp	tɕ ^h u	(1)	wt	sgo-ŋa	(2)	kr	a-tʃai	(2)
sg	tɕ ^h u	(1)	pt	*puu	(7)	rw	a-tʃai	(2)
rh	tɕ ^h u	(1)	pbg	*pi ¹ -tui ²	(8)	sr	kje:	(2)
kt	tɕ ^h u	(1)	ph	*do-rinj	(5)	ct	a-tʃje:	(2)
jg	tɕ ^h u	(1)	pkc	*ɖu(u)y tu(u)y	(8)	lp	a-tʃje:	(2)
kn	tɕ ^h u	(1)	EYE	[LJ, S100]		zm	a ³³ -ciɛ ³³	(2)
df	tsiə	(2)	dh	k ^h u-m	(3)	li	a ³¹ -cie ⁵⁵	(2)
sc	tʃə-mut	(2)	kp	k ^h u-mə:	(3)	tsb	t ^h a-rinj-bu	(1)
wh	tʃə-mut	(2)	rp	k ^h a-by	(1)	wb	wéi	(3)
bc	tʃə-mut	(2)	sg	k ^h i-bi:	(1)	wt	thag rin-po	(1)
ka	tʃə-mut	(2)	rh	k ^h a-by	(1)	pt	*do	(4)
np	tʃə-mut	(2)	kt	k ^h a-by	(1)	pbg	*jal ² /cal ²	(2)
bl	tʃi:	(2)	jg	k ^h a-by	(1)	ph	*mə-rəŋ	(1)
kr	tʃi:	(2)	kn	ga-by	(1)	pkc	*hlaa-I, hlaa-II	(5)
rw	tʃi:	(2)	df	miək ~ mjek	(2)	FAT	[S100]	
sr	tʃi:	(2)	sc	mi-jak	(2)	dh	ɔ-jəw	(1)
ct	tʃi:	(2)	wh	mi-jak	(2)	kp	ɔ-ju	(1)
lp	tʃi:	(2)	bc	mi-jak	(2)	rp	a-jɔ:	(1)
zm	tɕih ⁵³	(2)	ka	mi-je	(2)	sg	a-jɔ:	(1)
li	tɕih ⁵³	(2)	np	mi-jak	(2)	rh	a-jəw	(1)
tsb	za	(3)	bl	a-kəm	(4)	kt	a-jɔ:	(1)
wb	sá	(3)	kr	a-kəm	(4)	jg	a-jɔ:	(1)
wt	bzah-ba	(3)	rw	kəp	(4)	kn	a-jɔ:	(1)
pt	*do	(4)	sr	kəʔ	(4)	df	ə-jou	(1)
pbg	*ca ²	(1)	ct	(a)-kək	(4)	sc	e-ju	(1)
ph	*tca	(1)	lp	(a)-kək	(4)	wh	e-ju	(1)
pkc	*ʔay	(5)	zm	kək ³³	(4)	bc	e-ju	(1)
EGG	[LJ, S100]		li	a ³¹ -gũ ^{k55}	(4)	ka	e-jo	(1)
dh	hɔj-ɖəŋ	(1)	tsb	minj	(2)	np	e-jo	(1)
kp	ɔ-gɔŋ	(2)	wb	seʔ k ^h u	(3)	bl	a-ʒɔ:	(1)
rp	a-juŋ	(3)	wt	mig	(2)	kr	a-zja:	(1)
sg	k ^h ɔm	(4)	pt	*mik	(2)	rw	a-zja:	(1)
rh	a-juŋ	(3)	pbg	*muuk-kon	(2)	sr	a-zua	(1)
kt	a-juŋ	(3)	ph	*mə-jaʔ	(2)	ct	a-zua	(1)
jg	a-juŋ	(3)	pkc	*mik	(2)	lp	a-ɖɔua	(1)
kn	a-juŋ	(3)	FAR	[LJ]		zm	a ³³ -zua ¹¹	(1)
df	ə-ri:	(5)	dh	t ^h a-u-rinj	(1)	li	mur ³¹ -zua ⁵⁵	(1)
sc	k ^h e-re	(5)	kp	t ^h a-u-linj	(1)	tsb	jan	(2)
wh	k ^h e-re	(5)	rp	a-tɕ ^h ɔ:	(2)	wb	əs ^h i	(3)
bc	e-ri	(5)	sg	a-tɕ ^h ɔ:	(2)	wt	tshil	(3)
ka	k ^h e-re	(5)	rh	a-tɕ ^h ɔ:	(2)	pt	*fu	(4)
np	k ^h e-re	(5)	kt	a-tɕ ^h ɔ:	(2)	pbg	*mV ⁴ -tVm	(5)
bl	a-wuN	(3)	jg	a-tɕ ^h ɔ:	(2)	ph	*mə-baC	(6)
kr	a-li:	(5)	kn	a-tɕ ^h ɔ:	(2)	pkc	*thaaw	(7)
rw	tju-li	(5)	df	rɛʔɛ ruanj	(1)	FINGERNAIL	[S100]	
sr	mə-ri:	(5)	sc	r ^h anj	(1)	dh	hut-tsun	(1)
ct	mə-ri:	(5)				kp	hut-tsin	(1)

rp	ik-tsiŋ	(1)	pt	*mə	(1)	sr	mə-reʔ	(4)
sg	ik-tciŋ	(1)	pbg	*bwar ²	(2)	ct	mə-rjeʔ	(4)
rh	ik-tciŋ	(1)	ph	*maj	(1)	lp	mə-rjeʔ	(4)
kt	ik-tsiŋ	(1)	pkc	*may	(1)	zm	mə ³³ -ri ³³	(4)
jg	ik-tc ^h iŋ	(1)				li	mur ³¹ -li ^{k55}	(4)
kn	ik-tsiŋ	(1)	FISH		[LJ, S100]	tsb	ca	(1)
df	bə-sen	(1)	dh	ŋa	(1)	wb	ə-θá	(1)
sc	bə-sen	(1)	kp	ŋa	(1)	wt	ša	(1)
wh	pə-ŋi-jat	(1)	rp	ŋu	(1)	pt	*duun	(5)
bc	bə-sen	(1)	sg	ŋu	(1)	pbg	NA	
ka	pə-sen	(1)	rh	nu	(1)	ph	*su	(1)
np	pə-fen	(1)	kt	nu	(1)	pkc	*shaa	(1)
bl	a-geʔ gə-sin	(1)	jg	nu	(1)			
kr	gei-ŋin	(1)	kn	nu	(1)	TO FLY		[LJ, S100]
rw	gə-sin	(1)	df	ūā	(1)	dh	ɕɛɪ	(1)
sr	gei-siŋ	(1)	sc	ūā	(1)	kp	ɕɛɪ	(1)
ct	gei-siʔ	(1)	wh	ūā	(1)	rp	tɕ ^h an	(1)
lp	gei-siʔ	(1)	bc	u-ŋa	(1)	sg	tɕ ^h an	(1)
zm	geʔ ³³ -ci ³³	(1)	ka	ūā	(1)	rh	tɕ ^h ɛm	(1)
li	gi ¹³³ -ci ¹⁵³	(1)	np	u-ŋā	(1)	kt	tɕ ^h an	(1)
tsb	tsi-nan	(1)	bl	ʃui	(2)	jg	tɕ ^h ɛn	(1)
wb	leʔ θé	(1)	kr	ʃui	(2)	kn	tɕ ^h ɛn	(1)
wt	sen-mo	(1)	rw	ʃuei	(2)	df	mə-tso	(2)
pt	*lak-zin	(1)	sr	ŋo:	(1)	sc	k ^h e-ma-ge	(3)
pbg	*yak-si-kur	(1)	ct	kahuan	(3)	wh	k ^h e ⁱ	(3)
ph	NA		lp	kahuan	(3)	bc	k ^h e ⁱ	(3)
pkc	*tin	(1)	zm	ka ³³ fuan ⁵³	(3)	ka	k ^h ɛŋ-p ^h a	(3)
			li	ka ³³ fuan ⁵³	(3)	np	k ^h e ⁱ -p ^h a	(3)
FIRE		[LJ, S100]	tsb	ŋa	(1)	bl	ki:	(3)
dh	bej	(1)	wb	ŋá	(1)	kr	kɛi	(3)
kp	bej	(1)	wt	ŋa	(1)	rw	kei	(3)
rp	ba	(1)	pt	*ŋo	(1)	sr	piʔ	(1)
sg	ba	(1)	pbg	*na ²	(1)	ct	pjeʔ	(1)
rh	bɛ	(1)	ph	trV	(2)	lp	pjiʔ	(1)
kt	bɛ	(1)	pkc	*ŋaa hŋaa	(1)	zm	pie ³³	(1)
jg	bɛ	(1)				li	pie ⁵⁵	(1)
kn	ba	(1)	FLESH/MEAT		[LJ, S100]	tsb	p ^h ir	(5)
df	bo:ɛ	(1)	dh	ca	(1)	wb	pyã (pjan)	(1)
sc	bo ^ɛ	(1)	kp	ca	(1)	wt	hphur-ba	(5)
wh	bo ^ɛ	(1)	rp	su	(1)	pt	*bjar	(1)
bc	bo ^ɛ	(1)	sg	su	(1)	pbg	*pir ¹	(5)
ka	ba ⁱ	(1)	rh	su	(1)	ph	NA	
np	bo ^ɛ	(1)	kt	su	(1)	pkc	NA	
bl	bɛ:	(1)	jg	su	(1)			
kr	bai	(1)	kn	ʃu	(1)	FULL		[LJ, S100]
rw	bai	(1)	df	e:-cau ~ ə-cau	(1)	dh	pɕiŋ-ba	(1)
sr	bɛ:	(1)	sc	e-pije	(2)	kp	pɕiŋ-ba	(1)
ct	bae	(1)	wh	e-pije	(2)	rp	p ^h iŋ-ba	(1)
lp	bɛ:	(1)	bc	e-bije	(2)	sg	p ^h iŋ-ba	(1)
zm	bæ ³³	(1)	ka	mə-bi	(2)	rh	a-p ^h iŋ	(1)
li	bæ ⁵⁵	(1)	np	e-bei	(2)	kt	a-p ^h iŋ	(1)
tsb	mi	(1)	bl	fi:	(1)	jg	a-p ^h iŋ	(1)
wb	mí	(1)	kr	mai	(3)	kn	a-p ^h iŋ	(1)
wt	me	(1)	rw	mai	(3)			

Lieberherr and Bodt: Sub-grouping Kho-Bwa based on shared core vocabulary

df	ə-li-je:	(2)	TO GO	[LJ]	tsb	lek-pu	(7)	
sc	li-ja	(2)	dh	wa	(1)	wb	káō	(8)
wh	e-li-ja	(2)	kp	wa	(1)	wt	legs-po	(7)
bc	le ¹	(2)	rp	rɔ	(2)	pt	*-pro?	(9)
ka	le ¹	(2)	sg	rɔ	(2)	pbg	*nam ¹	(5)
np	ele ¹	(2)	rh	rɔ	(2)	ph	NA	
bl	lje:	(2)	kt	rɔ	(2)	pkc	*phraa-I, phrat-II	(9)
kr	jei	(2)	jg	rɔ	(2)			
rw	ljei	(2)	kn	rɔ	(2)	HAIR ON HEAD	[LJ, S100]	
sr	le:	(2)	df	wuə	(1)	dh	k ^h u-çan	(1)
ct	lje:	(2)	sc	lua	(3)	kp	k ^h u-çan	(1)
lp	lje:	(2)	wh	və-wa	(1)	rp	gə-zan	(1)
zm	liɛ ³³	(2)	bc	t ^h e-vəə	(1)	sg	gə-zan	(1)
li	lian ³³	(2)	ka	dun	(4)	rh	gə-zan	(1)
tsb	p ^h uŋ-ma	(3)	np	vəo-p ^h a	(1)	kt	gə-zan	(1)
wb	p ^h yó	(3)	bl	vu:	(1)	jg	gə-zan	(1)
wt	gañ-ba	(4)	kr	wu:	(1)	kn	gə-zan	(1)
pt	*bruŋ	(3)	rw	wɔ:	(1)	df	k ^h a-zijən	(1)
pbg	*p ^h uŋ ¹	(3)	sr	u:	(1)	sc	k ^h a-zijan	(1)
ph	*liŋ	(1)	ct	u:	(1)	wh	k ^h a-zija ^u	(1)
pkc	*pliŋ	(1)	lp	u:	(1)	bc	k ^h a-zijan	(1)
			zm	u ³³	(1)	ka	k ^h a-zijan	(1)
TO GIVE	[LJ, S100]		li	wu ⁵⁵	(1)	np	k ^h a-zijan	(1)
dh	ci	(1)	tsb	di	(5)	bl	kə-zaN	(1)
kp	ci	(1)	wb	tó	(2)	kr	kə-zjan	(1)
rp	t ^h i	(1)	wt	hgro	(2)	rw	kə-zja?	(1)
sg	k ^h ji	(1)	pt	*in	(6)	sr	kə-za?	(1)
rh	t ^h i	(1)	pbg	NA		ct	kə-zak	(1)
kt	t ^h i	(1)	ph	NA		lp	kə-dzak	(1)
jg	t ^h i	(1)	pkc	*kal	(7)	zm	kə ³³ -zak ³³	(1)
kn	t ^h i	(1)				li	kau ³¹ -zia ^{k35}	(1)
df	p ^h i	(1)	GOOD	[LJ, S100]	tsb	ts ^h am	(2)	
sc	p ^h i-mage	(1)	dh	ə-dzəp	(1)	wb	s ^h ã	(2)
wh	p ^h i	(1)	kp	nak-pa	(2)	wt	skra	(3)
bc	p ^h i-p ^h a	(1)	rp	a-dzap	(1)	pt	NA	
ka	p ^h i-p ^h a	(1)	sg	a-dzap	(1)	pbg	*khi ³ -ni	(4)
np	p ^h i	(1)	rh	a-dzap	(1)	ph	*go-pu	(5)
bl	taN	(2)	kt	a-dzap	(1)	pkc	*sham	(2)
kr	taŋ	(2)	jg	a-dzɛ	(1)			
rw	ta?	(2)	kn	a-dza	(1)	HARD	[LJ]	
sr	taŋ	(2)	df	ə-viə	(3)	dh	kaktar	(1)
ct	taŋ	(2)	sc	a-vije	(3)	kp	kaktal	(1)
lp	taŋ	(2)	wh	a-vije	(3)	rp	a-tɕu	(2)
zm	taŋ ³³	(2)	bc	a-vije	(3)	sg	a-tɕu	(2)
li	taŋ ⁵⁵	(2)	ka	a-vije	(3)	rh	a-tɕu	(2)
tsb	bi	(1)	np	a-vije	(3)	kt	a-tɕu	(2)
wb	kã	(3)	bl	a-mje:	(4)	jg	a-tɕu	(2)
wt	sbyin-pa	(1)	kr	a-njam	(5)	kn	a-xãŋ	(3)
pt	*bi	(1)	rw	a-njam	(5)	df	ə-kɔŋ	(3)
pbg	*ron ²	(4)	sr	a-ji:	(6)	sc	NA	
ph	*baj	(1)	ct	a-njan	(5)	wh	NA	
pkc	*pia-I, piak-II	(1)	lp	a-njan	(5)	bc	NA	
			zm	a ³³ -nian ¹¹	(5)	ka	NA	
			li	a ³¹ -nian ⁵⁵	(5)	np	NA	

Lieberherr and Bodt: Sub-grouping Kho-Bwa based on shared core vocabulary

zm	wa ^{t33}	(1)	ka	hō	(2)	rh	la-p ^h ik k ^h om	(2)
li	wa ^{rt55}	(1)	np	hō	(2)	kt	lə-p ^h ik k ^h om	(2)
tsb	ce	(1)	bl	ham	(2)	jg	la-p ^h yk kri:	(2)
wb	θa?	(1)	kr	ham	(2)	kn	ja-p ^h yk kra:	(2)
wt	bsad-pa	(1)	rw	ham	(2)	df	la:-gjuŋ	(3)
pt	*man	(3)	sr	ha?	(2)	sc	lə-spau	(4)
pbg	NA	(4)	ct	fuk	(3)	wh	la-kijuŋ	(3)
ph	*gajC	(1)	lp	fuk	(3)	bc	lə-spau	(4)
pkc	*that-I, -tha?-II	(1)	zm	tcok ³³	(3)	ka	la-gijon	(3)
			li	tcau ^{k53}	(3)	np	la-gijon	(3)
HORN		[LJ, S100]	tsb	p ^h ai	(4)	bl	a-le: kə-peɪ?	(5)
dh	k ^h ɛŋ	(1)	wb	ɛi (ein)	(5)	kr	lai ka-pɛi	(5)
kp	k ^h ɛŋ	(1)	wt	khyim	(6)	rw	lai kə-rə?	(6)
rp	k ^h ɛŋ	(1)	pt	*nam	(7)	sr	le:-kuŋ	(7)
sg	k ^h ɛŋ	(1)	pbg	*nok	(8)	ct	ka-pɛi	(5)
rh	t ^h ɛŋ	(1)	ph	*nam	(7)	lp	ka-peɪ	(5)
kt	t ^h ɛŋ	(1)	pkc	*ʔim	(6)	zm	NA	(12)
jg	t ^h ɛŋ	(1)				li	mə ³¹ -lai ⁵⁵ kə ³¹ -peɪ ⁵³	(5)
kn	tʂ ^h ajŋ	(1)	IN		[LJ]	tsb	təŋziŋ	(8)
df	ə-riəŋ	(2)	dh	naŋ	(1)	wb	dú	(9)
sc	e-rijaŋ	(2)	kp	naŋ	(1)	wt	pus-mo	(10)
wh	e-rijok	(2)	rp	a-lyŋ	(2)	pt	*lə-buŋ	(11)
bc	e-rijaŋ	(2)	sg	a-luŋ	(2)	pbg	NA	(12)
ka	e-rijaŋ	(2)	rh	a-lyŋ	(2)	ph	NA	(12)
np	e-rijaŋ	(2)	kt	a-lyŋ	(2)	pkc	*kuup khuup khuuk	(12)
bl	a-kjeN	(1)	jg	a-lyŋ	(2)			
kr	a-kɛian	(1)	kn	a-lyŋ	(2)	TO KNOW		[LJ, S100]
rw	a-kɛiet	(1)	df	ptɕ ^h au	(3)	dh	dɛn	(1)
sr	a-kuaŋ	(3)	sc	NA	(3)	kp	dɛn	(1)
ct	ko-waŋ	(3)	wh	NA	(3)	rp	dan	(1)
lp	ko-waŋ	(3)	bc	NA	(3)	sg	dan	(1)
zm	a ³³ -kə-vuaŋ ³³	(3)	ka	NA	(3)	rh	dɛn	(1)
li	a ³¹ -vaŋ ⁵³	(3)	np	NA	(3)	kt	dɛn	(1)
tsb	wa-rəŋ	(2)	bl	a-liŋ	(2)	jg	dɛn	(1)
wb	dʒò (gjou)	(6)	kr	a-loŋ	(2)	kn	dɛn	(1)
wt	rwa	(2)	rw	NA	(2)	df	it	(2)
pt	*rəŋ	(2)	sr	a-raŋ	(4)	sc	t ^h ek-ma-ge	(3)
pbg	*kron ¹	(1)	ct	a-ɽaŋ	(4)	wh	it	(2)
ph	*mə-su	(4)	lp	aŋ-ɽaŋ	(4)	bc	t ^h e-jak-pa	(3)
pkc	*kii	(5)	zm	a ³³ -ɽaŋ ⁵³	(4)	ka	it-p ^h at	(2)
			li	a ³¹ -raŋ ⁵³	(4)	np	it-p ^h a	(2)
HOUSE		[LJ]	tsb	naŋ-ka	(1)	bl	dɛN	(1)
dh	wam	(1)	wb	ətwi ⁷ (atwin:)	(5)	kr	dan	(1)
kp	wam	(1)	wt	naŋ-la	(1)	rw	dat	(1)
rp	jam	(1)	pt	NA	(1)	sr	dai?	(1)
sg	jam	(1)	pbg	*niŋ ² /*nuŋ ²	(2)	ct	dai?	(1)
rh	jam	(1)	ph	NA	(2)	lp	dai?	(1)
kt	jam	(1)	pkc	*tshuŋ	(6)	zm	dat ³³	(1)
jg	jan ~ jɛn	(1)				li	da ^{t55}	(1)
kn	jan	(1)	KNEE		[LJ, S100]	tsb	se	(2)
df	hō	(2)	dh	kə:lək-pa	(1)	wb	ta?	(1)
sc	hō	(2)	kp	kə:lək-pa	(1)	wt	šes-pa	(2)
wh	hō	(2)	rp	lə-p ^h yk	(2)	pt	*ken	(4)
bc	hō	(2)	sg	k ^h lək-pu:	(1)	pbg	*si ⁴	(2)

ph	*ni	(5)	ct	a-ləʔ	(1)	sc	NA	
pkc	*thay-I, thay?-II	(3)	lp	a-ləʔ	(1)	wh	NA	
			zm	he'n ³³ -luuk ³³	(1)	bc	NA	
			li	a ³¹ -lur ^{k53}	(1)	ka	NA	
			tsb	ɕaba	(2)	np	NA	
TO LAUGH [LJ]			wb	əyweʔ	(5)	bl	a-p ^{hi} N	(2)
dh	zej	(1)	wt	lo-ma	(3)	kr	a-pin	(2)
kp	zej	(1)	pt	*nə	(4)	rw	a-pjet	(2)
rp	za:	(1)	pbg	*lai ²	(1)	sr	a-piʔ	(2)
sg	za:	(1)	ph	*mə-rajC	(1)	ct	a-pjiʔ	(2)
rh	zɛ:	(1)	pkc	*hnaʔ	(4)	lp	a-pjiʔ	(2)
kt	zɛ:	(1)	LEG/FOOT [LJ, S100]			zm	a ³³ -pie ⁵³	(2)
jg	zɛ:	(1)	dh	ləj	(1)	li	a ³¹ -pie ⁵³	(2)
kn	zɑ:	(1)	kp	ləj	(1)	tsb	tɕ ^{hi} m-pa	(3)
df	sɔk	(2)	rp	la	(1)	wb	ə-θé	(2)
sc	ʈʂjo	(2)	sg	la	(1)	wt	mchin-pa	(3)
wh	ʈʂjo	(2)	rh	lə	(1)	pt	*zin	(2)
bc	ʃjaʔ	(2)	kt	lə	(1)	pbg	*pi ³ -kha	(2)
ka	seʔ	(2)	jg	lə	(1)	ph	*mə-θin	(2)
np	ʃeʔ-p ^h a	(2)	kn	la	(1)	pkc	*thin	(2)
bl	zɛ:	(1)	df	loɛ	(1)	LONG [LJ, S100]		
kr	zjei	(1)	sc	lo ^e	(1)	dh	u-rin	(1)
rw	zei	(1)	wh	lu ^e	(1)	kp	u-lin	(1)
sr	zɛ:	(1)	bc	lu ^e	(1)	rp	a-rɛŋ	(1)
ct	ruɛ	(3)	ka	la ¹	(1)	sg	a-rɛŋ	(1)
lp	ruɛ	(3)	np	la ¹	(1)	rh	a-rɛŋ	(1)
zm	yue ³³	(3)	bl	a-lɛ:	(1)	kt	a-rɛŋ	(1)
li	yue ⁵⁵	(3)	kr	lai	(1)	jg	a-rɛŋ	(1)
tsb	ŋar	(4)	rw	lai	(1)	kn	a-rɛŋ	(1)
wb	yè (yi)	(5)	sr	lai	(1)	df	ə-p ^h jan	(2)
wt	bzad	(1)	ct	(a)-lɛ:	(1)	sc	p ^h ijan	(2)
pt	*ŋil	(6)	lp	(a)-lɛ:	(1)	wh	e-p ^h ijam	(2)
pbg	*mi ⁴ -ni	(6)	zm	læ ³³	(1)	bc	p ^h ijan	(2)
ph	*toC	(7)	li	a ³¹ -tsæ ³³	(2)	ka	p ^h ijan	(2)
pkc	*nu(u)y-I, -nuy?-II	(6)	tsb	bi	(3)	np	e-p ^h ijam	(2)
			wb	ʈʰèi (ʈʰi)	(4)	bl	a-pjaN	(2)
			wt	rkaŋ-pa	(7)	kr	a-pɾaŋ	(2)
			pt	*lə	(1)	rw	a-pɾaŋ	(2)
			pbg	*ya ² -then	(5)	sr	pɾaŋ	(2)
			ph	*ləj	(1)	ct	a-pɾaŋ	(2)
			pkc	*kee kee	(6)	lp	a-pɾaŋ	(2)
			►[li] a ³¹ -tsæ ⁵³ probably means 'thigh'			zm	a ³³ -pɾaŋ ⁵³	(2)
			LIVER [LJ, S100]			li	a ³¹ -pɾaŋ ⁵⁵	(2)
LEAF [LJ, S100]			dh	ca tanku	(1)	tsb	rin-bu	(1)
dh	u-lap	(1)	kp	cap-ɕin	(2)	wb	myó	(3)
kp	u-lap	(1)	rp	a-ts ^h ɛ:	(3)	wt	rin-po	(1)
rp	a-lap	(1)	sg	a-tɕ ^h i:	(3)	pt	NA	
sg	a-lap	(1)	rh	a-tɕ ^h i:	(3)	pbg	*lao ²	(4)
rh	a-rap	(1)	kt	sə-si:	(2)	ph	*mə-pjan	(2)
kt	a-rap	(1)	jg	sɛj ~ sɛi	(2)	pkc	*s ^h ay	(5)
jg	a-lɛ	(1)	kn	də-tɕɛ:	(3)	LOUSE [LJ, S100]		
kn	a-laʔ	(1)	df	ə-p ^h in	(2)	dh	hɛk	(1)
df	ə-rap	(1)						
sc	e-rap	(1)						
wh	rap	(1)						
bc	rap	(1)						
ka	rap	(1)						
np	rap	(1)						
bl	a-ləp	(1)						
kr	hi-jəm	(1)						
rw	hiŋ-ləp	(1)						
sr	a-ləʔ	(1)						

df	raguŋ ju-ŋã	(3)	NOSE	[LJ, S100]	tsb	ma	(1)
sc	e-jua	(3)	dh	hɛm-p ^h ɔŋ	wb	NA	(1)
wh	ja-pu	(3)	kp	hɛm-p ^h ɔŋ	wt	ma	(1)
bc	e-jua	(3)	rp	nə-p ^h uŋ	pt	*man	(1)
ka	jua	(3)	sg	nə-p ^h uŋ	pbg	*-ya ^o	(2)
np	jua	(3)	rh	ma-p ^h uŋ	ph	NA	(1)
bl	kə-tuN-rin	(1)	kt	nə-p ^h uŋ	pkc	*law	(3)
kr	tʉŋ-rin	(1)	jg	nə-p ^h uŋ			
rw	tō-rin	(1)	kn	nə-p ^h uŋ	OLD OF THINGS	[LJ]	
sr	kə-tjuŋ	(1)	df	ə-p ^h uŋ	dh	ɔ-mɛn	(1)
ct	kə-tuŋ	(1)	sc	e-p ^h uŋ	kp	ɔ-mɛn	(1)
lp	kə-tjuŋ	(1)	wh	e-p ^h uŋ	rp	a-man	(1)
zm	kə ³³ -tuŋ ³³	(1)	bc	e-p ^h uŋ	sg	a-man	(1)
li	kur ³¹ -tor ⁵⁵	(1)	ka	e-p ^h uŋ	rh	a-mɛn	(1)
tsb	ŋaŋ	(4)	np	e-p ^h uŋ	kt	a-man	(1)
wb	lè	(5)	bl	a-p ^h uN	jg	a-mɛn	(1)
wt	mjiŋ-pa	(6)	kr	a-puŋ	kn	a-mɛn	(1)
pt	*luŋ	(7)	rw	poŋ	df	ə-hek	(2)
pbg	*Go ^o -tok	(1)	sr	poŋ	sc	e-he	(2)
ph	NA		ct	a-pok	wh	e-hek	(2)
pkc	*hŋooŋ	(4)	lp	a-pok	bc	e-hek	(2)
			zm	pək ³³	ka	e-hek	(2)
NEW	[LJ, S100]		li	a ³¹ pau ^{k55}	np	e-hek	(2)
dh	ɔ-k ^h ɔŋ	(1)	tsb	na-huŋ	bl	a-tɛN	(3)
kp	ɔ-han	(1)	wb	ŋà	kr	a-tʃɛn	(3)
rp	a-fan	(1)	wt	sna	rw	a-tʃet	(3)
sg	u-fan	(1)	pt	*nã-buŋ	sr	a-tʃei?	(3)
rh	a-hɛn	(1)	pbg	*kuŋ ¹ -thuŋ	ct	a-tʃai?	(3)
kt	a-fan	(1)	ph	NA	lp	a-tʃai?	(3)
jg	ə-hɛn	(1)	pkc	*hnaar	zm	a ³³ -tɛt ³³	(3)
kn	a-fɛn	(1)			li	a ³¹ -tɛ ⁵⁵	(3)
df	ə-vɔ̄	(2)	NOT	[LJ]	tsb	man-ma	(1)
sc	o-vo ^u	(2)	dh	ba-	wb	háɔ̄ (haun:)	(4)
wh	e-voŋ	(2)	kp	ba-	wt	rɪiŋ-pa	(5)
bc	e-wã	(2)	rp	bə-	pt	*ku ~ kju?	(6)
ka	m-vəo	(2)	sg	bə-	pbg	*Gu ^o -cam ¹	(7)
np	e-vo	(2)	rh	bə-	ph	*mə-ɔʔ	(8)
bl	a-fɛN	(1)	kt	bə-	pkc	*hlun	(4)
kr	a-fan	(1)	jg	bə-			
rw	a-fa ^t	(1)	kn	bə-	ONE	[LJ]	
sr	a-fai?	(1)	df	NA	dh	hin	(1)
ct	a-fai?	(1)	sc	NA	kp	hin	(1)
lp	a-fai?	(1)	wh	NA	rp	han	(1)
zm	a ³³ -fa ^t ³³	(1)	bc	NA	sg	han	(1)
li	a ³¹ -fua ⁵⁵	(1)	ka	NA	rh	hɛn	(1)
tsb	siŋ-ma	(3)	np	NA	kt	han	(1)
wb	θi?	(3)	bl	ba-	jg	hɛn	(1)
wt	gsar-pa	(1)	kr	ba-	kn	hɛn	(1)
pt	NA		rw	ba-	df	tsjɔ̄	(2)
pbg	*Gu ^o -dal ⁴	(5)	sr	ba-	sc	ɕjiɔ	(2)
ph	*mə-gə-nu	(6)	ct	ba-	wh	tʃo	(2)
pkc	*thar	(1)	lp	ba-	bc	ɕjiɔ	(2)
			zm	ba ³³ -	ka	ɕɛ	(3)
			li	ba ³¹ -	np	ɕɛ	(3)

Lieberherr and Bodt: Sub-grouping Kho-Bwa based on shared core vocabulary

bl	tyi	(4)	jg	nə-tc ^h u	(2)	ROOT	[LJ, S100]	
kr	kju:	(2)	kn	na:-ts ^h u	(2)	dh	ra-ba	(1)
rw	tjo:	(2)	df	ha:-p ^h iε	(3)	kp	hiŋŋa lej	(2)
sr	ʃue	(3)	sc	ha-p ^h ije	(3)	rp	ə-hi:	(3)
ct	hui	(5)	wh	ha-p ^h ije	(3)	sg	ra-bə:	(1)
lp	hui	(5)	bc	ha-p ^h ije	(3)	rh	i-hi:	(3)
zm	ɦui ³³	(5)	ka	ha-p ^h ije	(3)	kt	a-hi:	(3)
li	ʃun ⁵⁵	(6)	np	ha-p ^h ije	(3)	jg	i-hi:	(3)
tsb	t ^h u:	(7)	bl	ham-pi:	(3)	kn	a-xi:	(3)
wb	ti?	(4)	kr	ham-piə:	(3)	df	ə-rəŋ	(4)
wt	gčig	(4)	rw	ham-piə:	(3)	sc	e-raŋ	(4)
pt	kon	(8)	sr	ha?-piə	(3)	wh	e-raŋ	(4)
pbg	-sha ⁴	(9)	ct	am-pi:	(3)	bc	e-raŋ	(4)
ph	a-ken	(10)	lp	am-pju:	(3)	ka	e-raŋ	(4)
pkc	*khat ʔat hat	(11)	zm	aŋ ³³ -pru ³³	(3)	np	e-raŋ	(4)
			li	aŋ ³³ -pru ³³	(3)	bl	ʃiN-rin	(4)
			tsb	ŋam-su	(4)	kr	hi-rin	(4)
PATH	[S100]		wb	mó (mou:)	(1)	rw	a-riŋ	(4)
dh	huma	(1)	wt	char-pa	(5)	sr	a-riŋ	(4)
kp	huma	(1)	pt	*mV-dəŋ	(6)	ct	a-reŋ	(4)
rp	ləmu	(1)	pbg	*no ⁴ -kha	(7)	lp	a-reŋ	(4)
sg	ləmu	(1)	ph	*pro	(3)	zm	he ¹ n ³³ -kuŋ ³³	(5)
rh	lym	(1)	pkc	*rua?	(8)	li	a ³¹ -koŋ ⁵⁵	(5)
kt	lym	(1)				tsb	ra-tc ^h i	(1)
jg	hamu	(1)	RED	[LJ, S100]		wb	əmyi?	(6)
kn	ham	(1)	dh	ə-k ^h ek	(1)	wt	rtsa-ba	(7)
df	ləmuə duə	(1)	kp	ə-hək	(1)	pt	*puur?	(8)
sc	ləmua	(1)	rp	ə-hək	(1)	pbg	NA	
wh	ləmua	(1)	sg	ə-hək	(1)	ph	*mə-kriŋ	(4)
bc	ləmua	(1)	rh	ə-hək	(1)	pkc	*yur	(9)
ka	ləmua	(1)	kt	ə-hək	(1)			
np	ləmua	(1)	jg	ə-hək	(1)	TO RUN	[LJ]	
bl	lim	(1)	kn	a-xək	(1)	dh	dur	(1)
kr	lim	(1)	df	ə-cək	(1)	kp	dul	(1)
rw	lip	(1)	sc	e-faik	(1)	rp	tc ^h a:	(2)
sr	li?	(1)	wh	e-fək	(1)	sg	tc ^h a:	(2)
ct	puzue	(2)	bc	e-faik	(1)	rh	tcɛ:	(2)
lp	puzue	(2)	ka	fe ¹	(1)	kt	riŋ	(3)
zm	pə ³³ zue ³³	(2)	np	e-faik	(1)	jg	le ʔik ^h en	(4)
li	pu ³¹ zui ⁵⁵	(2)	bl	a-ʃi?	(1)	kn	ts ^h ə:	(2)
tsb	lam	(1)	kr	a-hi:	(2)	df	ran	(3)
wb	lá (lan:)	(1)	rw	a-rju?	(3)	sc	NA	
wt	lam	(1)	sr	a-rju?	(3)	wh	NA	
pt	*lam	(1)	ct	a-rjuk	(3)	bc	NA	
pbg	*la ¹ -ma	(1)	lp	a-rjuk	(3)	ka	NA	
ph	*lam-bar	(1)	zm	a ³³ -ji ³³	(4)	np	NA	
pkc	*lam	(1)	li	a ³³ -ji ⁵⁵	(4)	bl	rin	(3)
			tsb	tša-lu	(5)	kr	ren	(3)
RAIN	[LJ, S100]		wb	máō	(6)	rw	rit	(3)
dh	na-mu	(1)	wt	dmar-po	(6)	sr	ri?	(3)
kp	na-mu	(1)	pt	*luŋ	(7)	ct	ri?	(3)
rp	nə-tc ^h u:	(2)	pbg	*Gu ⁰ -cak	(5)	lp	ri?	(3)
sg	ni-mi	(1)	ph	*tca?	(5)	zm	ri ³³	(3)
rh	p ^h -tc ^h u	(2)	pkc	*s ^h an s ^h en	(8)			
kt	nə-tc ^h u	(2)						

li	zei ⁵⁵	(5)	np	bali	(4)	kt	nū: (-t ^{hi})	(2)
tsb	dzuk	(6)	bl	səm-pjɛ:	(6)	jg	NA	
wb	pyéi	(7)	kr	pan-tʃi:-hɿɛ:	(6)	kn	nō:-t ^{hi}	(2)
wt	rgyug-pa	(6)	rw	pā-tʃiʔ	(6)	df	gə-rak	(3)
pt	NA		sr	bɛɛ-tʃiʔ	(6)	sc	rək-ma-ge	(3)
pbg	*khat	(2)	ct	pɿɛ:-tʃeiʔ	(6)	wh	rək	(3)
ph	NA		lp	pjɛ:-tʃeʔ	(6)	bc	rəg	(3)
pkc	*pliɿ play	(7)	zm	pɿɛ ^{r33} -tsət ³³	(6)	ka	rek-p ^h a	(3)
			li	pa ^{r55} -tse ^{r33}	(6)	np	rək	(3)
SALT			[LJ] tsb	betsa	(1)	bl	niŋ	(4)
dh	ha	(1)	wb	θé	(7)	kr	njəŋ	(4)
kp	ha	(1)	wt	bye-ma	(8)	rw	niŋ	(4)
rp	hū:	(2)	pt	NA		sr	ɲi:	(4)
sg	huŋ	(2)	pbg	*han ² -ceŋ	(9)	ct	ɕɛi	(5)
rh	hūŋ	(2)	ph	*sə-gə-raj	(10)	lp	ɕɛi	(5)
kt	huŋ	(2)	pkc	*neel	(11)	zm	ŋua ³³ /dat ³³	(6)
jg	huŋ	(2)				li	ŋwa ⁵⁵ da ¹³³	(6)
kn	hū:	(2)	TO SAY		[LJ, S100]	tsb	t ^h əŋ	(7)
df	sə-ruə	(3)	dh	in	(1)	wb	myi ^ː	(4)
sc	su-ruwa	(3)	kp	in	(1)	wt	mthoŋ-ba	(7)
wh	su-ruwa	(3)	rp	am	(1)	pt	*kaŋ	(8)
bc	su-rəja	(3)	sg	am	(1)	pbg	*nuk	(2)
ka	su-ɿɛ	(3)	rh	an	(1)	ph	*gaŋ	(8)
np	su-ruwa	(3)	kt	an	(1)	pkc	*hmuu-I, hmuʔ-II	(9)
bl	fəu	(4)	jg	ɛn	(1)			
kr	fou	(4)	kn	ɛn	(1)	TO SRT		[S100]
rw	fə:	(4)	df	ja:	(2)	dh	dzu	(1)
sr	hoŋ	(2)	sc	je-ma-ge	(2)	kp	dzu	(1)
ct	hoŋ	(2)	wh	je	(2)	rp	dzau	(1)
lp	hoŋ	(2)	bc	je-p ^h a	(2)	sg	dzau	(1)
zm	həŋ ³³	(2)	ka	si-p ^h a	(3)	rh	dzo	(1)
li	hoŋ ⁵⁵	(2)	np	jo-p ^h a	(4)	kt	dzo:	(1)
tsb	in-t ^h a	(5)	bl	ri:	(5)	jg	dzo:	(1)
wb	s ^h á	(5)	kr	rjo:	(6)	kn	dzo:	(1)
wt	tshwa	(5)	rw	rjo:	(6)	df	duk	(2)
pt	*lo	(8)	sr	jɔ:	(4)	sc	duk-ma	(2)
pbg	*stum ³	(6)	ct	luiʔ	(7)	wh	tuk	(2)
ph	*lu	(8)	lp	luiʔ	(7)	bc	duk	(2)
pkc	*tsii	(7)	zm	lut ³³	(7)	ka	dək-mu	(2)
			li	lui ¹³³	(7)	np	duk-ma-ge	(2)
SAND			[LJ, S100] tsb	jik	(8)	bl	ri:	(3)
dh	betsa	(1)	wb	s ^h ə	(9)	kr	ɕgao	(1)
kp	betsa	(1)	wt	bśad	(10)	rw	tun	(2)
rp	mə-k ^h ɔ	(2)	pt	*ban- man	(11)	sr	tun	(2)
sg	betsə	(1)	pbg	*huun ¹	(1)	ct	tun	(2)
rh	k ^h ə-bɔ	(3)	ph	NA		lp	tun	(2)
kt	k ^h ə-p ^h ɔ	(3)	pkc	NA		zm	toŋ ³³	(2)
jg	k ^h -p ^h a	(3)				li	toŋ ⁵⁵	(2)
kn	gə-p ^h a	(3)	TO SEE		[LJ, S100]	tsb	lan	(4)
df	ba:li	(4)	dh	aj	(1)	wb	t ^h ai (htain)	(5)
sc	bali	(4)	kp	aj	(1)	wt	sdod-pa	(6)
wh	bali	(4)	rp	ɔ:	(1)	pt	*dun	(2)
bc	bali	(4)	sg	ɔ:	(1)	pbg	*a ⁴ -shoŋ	(7)
ka	bali	(4)	rh	ɔ:	(1)	ph	*dzuC	(1)

Lieberherr and Bodt: Sub-grouping Kho-Bwa based on shared core vocabulary

pkc	*taw	(5)	zm	yəŋ ³³	(3)	ka	baiŋ-k ^h am	(1)
			li	yəŋ ⁵⁵	(3)	np	bo ^o -mo-k ^h əm	(1)
SKIN/HIDE		[LJ, S100]	tsb	jip	(4)	bl	bɛ-k ^h i:	(1)
dh	ɕa-k ^h ək	(1)	wb	seʔ	(5)	kr	bai-kə:	(1)
kp	ɕa-k ^h ək	(1)	wt	ŋal-ba	(6)	rw	bai-ki:	(1)
rp	s-kuk	(1)	pt	*jup	(4)	sr	bai-ki:	(1)
sg	s-kuk	(1)	pbg	NA		ct	bɛ:-ki:	(1)
rh	s-kuk	(1)	ph	*dzV	(7)	lp	bɛ:-ki:	(1)
kt	s-kuk	(1)	pkc	*ʔip-I, ʔiʔ-II	(4)	zm	bæ ³³ -ku ³³	(1)
jg	s-kuk	(1)				li	bæ ³³ -ku ⁵⁵	(1)
kn	ʂ-k ^h uk	(1)	SMALL		[LJ, S100]	tsb	mu-gu	(1)
df	e-k ^h uk	(1)	dh	ka-ti	(1)	wb	ə-k ^h ó	(1)
sc	e-k ^h uk	(1)	kp	ɕa-da	(10)	wt	du-ba	(2)
wh	e-k ^h uk	(1)	rp	a-tɕan-du	(3)	pt	*mə-kuu	(1)
bc	e-k ^h uk	(1)	sg	a-tɕan-du	(3)	pbg	*muk	(3)
ka	e-k ^h uk	(1)	rh	tɕɔ:-du	(3)	ph	*maj-kən	(1)
np	e-k ^h uk	(1)	kt	a-dʒan-du	(3)	pkc	*may-khuu	(1)
bl	a-kuʔ	(1)	jg	a-du-du	(2)			
kr	a-ki:	(1)	kn	a-dʒan-du	(3)	STAND		[LJ, S100]
rw	a-kə:	(1)	df	ə-ti-dua	(1)	dh	k ^h iŋ	(1)
sr	a-kə:	(1)	sc	ēe ^u	(4)	kp	k ^h iŋ	(1)
ct	a-kə:	(1)	wh	e-juŋ	(4)	rp	k ^h i:	(1)
lp	a-kə:	(1)	bc	e-juŋ	(4)	sg	k ^h i:	(1)
zm	a ³³ -kə ⁵³	(1)	ka	e-juŋ	(4)	rh	tɕ ^h i:	(1)
li	a ³¹ -kə ⁵³	(1)	np	iu	(4)	kt	tɕ ^h i:	(1)
tsb	k ^h əp-taŋ	(1)	bl	a-zi:	(5)	jg	tɕ ^h i:	(1)
wb	ək ^h ũ (akhun)	(1)	kr	a-zi:	(5)	kn	tɕ ^h i:	(1)
wt	pags-pa	(2)	rw	a-jiŋ	(4)	df	tsai	(2)
pt	*rjo/*puŋ	(3)	sr	ɲe:-do:	(6)	sc	ri ^o -ma	(3)
pbg	*Gur ¹	(4)	ct	kə-siʔ	(7)	wh	ri ^o	(3)
ph	*mə-prja	(5)	lp	kə-siʔ	(7)	bc	ri ^o	(3)
pkc	*phoo, *wun	(6)	zm	ciaŋ ³³	(3)	ka	dʒoi	(2)
			li	dʒiaŋ ⁵⁵	(3)	np	dʒai-ma-ge	(2)
TO SLEEP		[LJ, S100]	tsb	ze-mu	(5)	bl	ʃiŋ	(1)
dh	min	(2)	wb	peʔ	(6)	kr	ʃiŋ	(1)
kp	min	(2)	wt	čuŋ-čuŋ	(8)	rw	ʃiŋ	(1)
rp	dʒiŋ	(1)	pt	NA		sr	ʃiŋ	(1)
sg	dʒiŋ	(1)	pbg	NA		ct	ʃiŋ	(1)
rh	dʒiŋ	(1)	ph	NA		lp	ʃiŋ	(1)
kt	dʒiŋ	(1)	pkc	*ʈaa-I, ʈaat-II	(9)	zm	cin ³³ -pua ¹¹	(1)
jg	dʒiŋ	(1)				li	cin ⁵⁵	(1)
kn	dʒiŋ	(1)	SMOKE		[LJ, S100]	tsb	t ^h iŋ	(4)
df	mjen	(2)	dh	bu-k ^h u	(1)	wb	t ^h a	(5)
sc	mijen-ma-ge	(2)	kp	bu-k ^h u	(1)	wt	hgyiŋ-ba	(1)
wh	mijen	(2)	rp	ba-k ^h y	(1)	pt	*dak	(5)
bc	mijan-p ^h a	(2)	sg	ba-k ^h y	(1)	pbg	*diŋ ¹	(4)
ka	men	(2)	rh	ba-k ^h y	(1)	ph	*gu	(6)
np	men-p ^h a	(2)	kt	ba-k ^h y	(1)	pkc	*diŋ-I, diŋ-II	(4)
bl	rəm	(3)	jg	ba-k ^h y	(1)			
kr	rəm	(3)	kn	ba-k ^h y	(1)	STAR		[LJ, S100]
rw	rəp	(3)	df	ba-k ^h au	(1)	dh	ka:ma	(1)
sr	rəʔ	(3)	sc	bo ^o -k ^h a ^u	(1)	kp	kalma	(1)
ct	rəm	(3)	wh	bu-k ^h u	(1)	rp	zyk	(2)
lp	rəm	(3)	bc	bam-k ^h am	(1)	sg	tɕ ^h u-zuk	(2)

rh	t ^h ɣŋ-zyk	(2)
kt	t ^h ɣŋ-zyk	(2)
jg	də-zyk	(2)
kn	dy-zyk	(2)
df	tɕa:-trəŋ	(3)
sc	sa-dijon	(3)
wh	sa-tijon	(3)
bc	ta-dijaŋ	(3)
ka	la-dijaŋ	(3)
np	ta-dijaŋ	(3)
bl	haN-wai?	(4)
kr	ha-daŋ	(3)
rw	sat	(5)
sr	sai?	(5)
ct	pə-ɕai?	(6)
lp	pə-ɕai?	(6)
zm	ha ³³ -ɣat ⁵³	(7)
li	ha ³¹ -ɣai ⁵³	(7)
tsb	karma	(1)
wb	fɛ̃ (kje)	(8)
wt	skar-ma	(1)
pt	*kar	(1)
pbg	*[ha ³ -tV-khi]	(9)
ph	*lə-tɕuŋ	(11)
pkc	*ʔaar-θii -sii	(10)
STONE/ROCK [LJ, S100]		
dh	luŋ	(1)
kp	luŋ	(1)
rp	lɣŋ	(1)
sg	luŋ	(1)
rh	rɣŋ	(1)
kt	rɣŋ	(1)
jg	lɣŋ	(1)
kn	lɣŋ	(1)
df	ləbau	(2)
sc	laba ^u	(2)
wh	laba ^u	(2)
bc	laba ^u	(2)
ka	laba ^u	(2)
np	lubu	(2)
bl	ka-liŋ	(1)
kr	ka-hoŋ	(1)
rw	kə-bɿa:	(3)
sr	kə-bɿa:	(3)
ct	kə-bɿa:	(3)
lp	kə-bɿa:	(3)
zm	kə ³³ -pə ⁵³	(3)
li	kə ³¹ -pa ⁵³	(3)
tsb	luŋ	(1)
wb	fʃaʊ? (kjau')	(4)
wt	rdo	(5)
pt	*luŋ	(1)
pbg	*loŋ ²	(1)
ph	NA	(1)
pkc	*luŋ	(1)
SUN [S100]		
dh	na-mi	(1)
kp	na-mi	(1)
rp	nə-ny:	(2)
sg	nə-ni:	(2)
rh	nə-mi:	(1)
kt	nə-mi:	(1)
jg	nə-ny:	(2)
kn	na:-ny:	(2)
df	ha:-neŋ	(3)
sc	ha-na ^l	(3)
wh	ha-ne'ŋ	(3)
bc	ha-na ^l	(3)
ka	ha-na ^l	(3)
np	ha-na ^l	(3)
bl	ha-mi	(1)
kr	ha-mi	(1)
rw	ha-mi	(1)
sr	ha-mi	(1)
ct	kri:	(2)
lp	kri:	(2)
zm	kə ³³ ri ³³	(2)
li	kə ³¹ ri ⁵⁵	(2)
tsb	ŋam	(4)
wb	nèi	(2)
wt	ñi-ma	(2)
pt	*ñi	(2)
pbg	NA	(1)
ph	*ɕu?	(5)
pkc	*nii	(2)
SWEET [LJ]		
dh	jəŋ-jəŋ	(1)
kp	ɔ-bej-da	(2)
rp	a-tɕi:	(7)
sg	a-tɕi:	(7)
rh	tɕi:	(7)
kt	tɕi:	(7)
jg	a-blä	(2)
kn	a-tɕi:	(7)
df	p ^h -jɛŋ	(1)
sc	p ^h -i-jam	(1)
wh	p ^h -i-jan	(1)
bc	p ^h -i-jan	(1)
ka	p ^h -i-jan	(1)
np	p ^h -i-jan	(1)
bl	a-pin	(3)
kr	a-pin	(3)
rw	a-pin	(3)
sr	a-piŋ	(3)
ct	a-piŋ	(3)
lp	a-piŋ	(3)
zm	a ³³ -piŋ ³³	(3)
li	a ³¹ -pin ⁵⁵	(3)
tsb	cim-pu	(7)
wb	ɕò (chou)	(4)
wt	zim-po	(7)
pt	*ti:	(5)
pbg	*tuɿ ²	(5)
ph	*mə-jaŋ	(1)
pkc	*khlum	(6)
TAIL [LJ, S100]		
dh	nəj	(1)
kp	nuj	(1)
rp	a-ɲa:	(1)
sg	a-ɲa:	(1)
rh	a-ne ^h	(1)
kt	a-ne ^h	(1)
jg	a-ne ^h	(1)
kn	a-na: ^h	(1)
df	ə-ɲui	(1)
sc	e-ɲo ^l	(1)
wh	e-ɲo ^l	(1)
bc	e-ɲo ^l	(1)
ka	e-ɲo ^l	(1)
np	e-ɲo ^l	(1)
bl	a-nyi	(1)
kr	a-sərjam	(2)
rw	a-kə-doŋ	(3)
sr	ɲi-mɛ:	(4)
ct	ni-mɛ:	(4)
lp	ni-mɛ:	(4)
zm	a ³³ -ɲuk ³³ kuaŋ ³³	(1)
li	a ³¹ -ɲiu ⁵⁵ kuaŋ ⁵⁵	(1)
tsb	campi	(5)
wb	myi	(4)
wt	mjug-ma	(6)
pt	*me~ mjo	(4)
pbg	NA	(1)
ph	*mə-lə-mrjaŋ	(4)
pkc	*may	(4)
TO TAKE [LJ]		
dh	t ^h ɔŋ	(1)
kp	t ^h ɔŋ	(1)
rp	t ^h ɔ:	(1)
sg	t ^h ɛ:	(1)
rh	t ^h ɛŋ	(1)
kt	t ^h i:	(1)
jg	t ^h ɔ:(n)	(1)
kn	t ^h ɔ:	(1)
df	rija	(2)
sc	t ^h a-rija-ma-ge	(2)

sr	kua	(1)	sc	e-mo ^u	(3)	kp	nam-suŋ	(1)
ct	kua	(1)	wh	e-moŋ	(3)	rp	nə-səm	(1)
lp	kua	(1)	bc	e-mo ^u	(3)	sg	nə-səm	(1)
zm	kua ³³	(1)	ka	mō ^u	(3)	rh	p-səm	(1)
li	kua ⁵⁵	(1)	np	e-moŋ	(3)	kt	ny-səm	(1)
tsb	ri	(2)	bl	a-rjuN	(4)	jg	nə-sən	(1)
wb	yèi	(2)	kr	a-rjuŋ	(4)	kn	na-sən	(1)
wt	chu	(3)	rw	a-rjuŋ	(4)	df	ha-sa:m	(1)
pt	*si	(4)	sr	a-rjuŋ	(4)	sc	ha-sam	(1)
pbg	*tui ¹	(5)	ct	a-rjuŋ	(4)	wh	ha-sam	(1)
ph	*wi	(6)	lp	a-rjuŋ	(4)	bc	a-fam	(1)
pkc	*tuy	(5)	zm	a ³³ -rioŋ ³³	(4)	ka	ha-sam	(1)
			li	a ³¹ -tioŋ ⁵⁵	(4)	np	ha-sam	(1)
WHAT?		[LJ]	tsb	balinmin	(5)	bl	haN-fin	(2)
dh	a-be	(1)	wb	p ^h yü	(6)	kr	ha-hin	(2)
kp	a-be	(1)	wt	dkar-po	(1)	rw	ha-hin	(2)
rp	musi a:	(2)	pt	*pun ~ puŋ	(6)	sr	a-hui?	(3)
sg	mə-ra:	(2)	pbg	*bok	(7)	ct	aŋ-hui?	(3)
rh	p ^h -te ^h õ: wa:	(3)	ph	*mə-gə-raN	(8)	lp	aŋ-hui?	(3)
kt	a-ti ja:	(4)	pkc	*(k)(r)aaŋ-I, (k)(r)aan-II	(9)	zm	aŋ ³³ -huŋ ⁵³	(3)
jg	mə-ti: ja	(4)				li	aŋ ³³ -xu ¹⁵⁵	(3)
kn	mə-ti: ja	(4)	WHO?		[LJ, S100]	tsb	ridi	(4)
df	ɲuja:	(5)	dh	mi	(1)	wb	lèi	(5)
sc	niwka	(5)	kp	mi	(1)	wt	rluŋ-ma	(6)
wh	niw	(5)	rp	ɲi:	(1)	pt	*rji	(4)
bc	niw	(5)	sg	ɲi:	(1)	pbg	*par ¹	(7)
ka	nije	(5)	rh	mi:	(1)	ph	*low	(8)
np	ɲijaga	(5)	kt	mi:	(1)	pkc	*hu(u)y	(3)
bl	he:	(6)	jg	ni:	(1)			
kr	hai	(6)	kn	ni:	(1)	WING		[LJ]
rw	hai	(6)	df	ha-ɲu	(2)	dh	pempu	(1)
sr	hai	(6)	sc	ha-niju	(2)	kp	ɲuj-p ^h a	(2)
ct	hi:	(7)	wh	ha-ɲuŋ	(2)	rp	nə-p ^h aŋ	(2)
lp	hi:	(7)	bc	ha-niju	(2)	sg	lə-p ^h aŋ	(2)
zm	hi ³³ je ^h 53	(7)	ka	ha-niju	(2)	rh	mə-p ^h aŋ	(2)
li	hi ⁵⁵	(7)	np	ha-niju	(2)	kt	nə-p ^h aŋ	(2)
tsb	haŋ	(8)	bl	mii	(1)	jg	ny-p ^h aŋ	(2)
wb	əbè	(1)	kr	ho:	(2)	kn	ny-p ^h aŋ	(2)
wt	ci	(9)	rw	ho:	(2)	df	ə-jik	(3)
pt	NA		sr	ha:	(2)	sc	NA	
pbg	*ma ⁴	(10)	ct	ha:	(2)	wh	NA	
ph	*tiŋ	(4)	lp	ha:	(2)	bc	NA	
pkc	NA		zm	ha ³³	(2)	ka	NA	
			li	ha ⁵⁵	(2)	np	NA	
WHITE		[S100]	tsb	ibi	(3)	bl	a-zuiN	(3)
dh	jaŋ-kai	(1)	wb	myi	(1)	kr	a-zun	(3)
kp	jaŋ-kal	(1)	wt	su	(4)	rw	a-zuət	(3)
rp	a-zõ:	(2)	pt	NA		sr	mə-jui?	(3)
sg	a-zõ:	(2)	pbg	*shar ¹	(5)	ct	a-jui?	(3)
rh	a-zõ:	(2)	ph	*θu	(4)	lp	a-wi?	(4)
kt	a-zõ:	(2)	pkc	*tuu	(4)	zm	a ³³ -juŋ ³³	(3)
jg	a-zā:	(2)				li	a ³¹ -jəu ¹⁵⁵	(3)
kn	a-zā:	(2)	WIND		[LJ]	tsb	ɕək-pa	(5)
df	ə-mau	(3)	dh	nam-suŋ	(1)	wb	taõ bā ⁵ (taun ban)	(6)

