Is there a typological profile of isolates?

Marine Vuillermet*

Department of Comparative Language Science and center for the Interdisciplinary Study of Language Evolution, University of Zurich, Switzerland

E-mail: marine.vuillermet@uzh.ch

ORCID ID: https://orcid.org/0000-0002-6443-8461

David Inman*

Department of Comparative Language Science and center for the Interdisciplinary Study of Language Evolution, University of Zurich, Switzerland

E-mail: david.inman@uzh.ch

ORCID ID: https://orcid.org/0000-0002-8512-0380

Natalia Chousou-Polydouri*

Department of Comparative Language Science and center for the Interdisciplinary Study of Language Evolution, University of Zurich, Switzerland

E-mail: natalia.chousou-polydouri@uzh.ch

ORCID ID: https://orcid.org/0000-0002-5693-975X

Kellen Parker van Dam

Department of Comparative Language Science and center for the Interdisciplinary Study of Language Evolution, University of Zurich, Switzerland

Department of Languages and Cultures, La Trobe University

E-mail: kellenparker.vandam@uzh.ch

ORCID ID: https://orcid.org/0000-0002-7304-1685

Shelece Easterday

Department of Linguistics, University of Hawaiʻi at Mānoa, United States of America

E-mail: shelece@gmail.com

ORCID ID: https://orcid.org/0000-0003-1186-1044

Françoise Rose

Dynamique du Langage, CNRS / Université Lyon 2, France

E-mail: francoise.rose@cnrs.fr

ORCID ID: https://orcid.org/0000-0003-4357-1383

* These authors have contributed equally to this work.

Abstract:

Across the linguistic literature, one occasionally encounters claims of typological differences between isolates and non-isolates, but these are often vague, and tend to use isolates as proxies for small community size, hunter-gatherer societies, and/or socially/geographically isolated languages. We compared the distribution of 89 phonological and morphosyntactic typological features between isolates and non-isolates using a worldwide sample of 215 languages (68 isolates vs 147 non-isolates), in which we were unable to find a statistically significant distinction. We discuss the relevance of our results for these claims, for the suggested proxy relationships between isolates and other factors, and suggest possible avenues for future research.

Keywords: isolate; language family; typology; macroarea; statistical test.

1. Introduction

From a phylogenetic perspective, languages can be divided into two types: isolate languages, which have no (known) sister languages, and non-isolates, which are part of larger families. Most linguistic science does not regard isolate languages as qualitatively different from non-isolates, but as simply the result of different historical factors and available evidence. Still, isolates seem to hold a certain fascination for researchers: firstly, because of their undeniable typological importance as unique representatives of linguistic lineages, and secondly for the possibility that the different histories of these languages could have led to a set of common linguistic structures, i.e. a specific typological profile. This possibility requires first a careful examination of what causes a language to be termed an isolate.

The classification of a language as an isolate is dependent upon a particular period of time and a particular body of evidence available to researchers. Changes over time in the available linguistic evidence can lead to isolate formation (the new categorization of a language as an isolate) and isolate dissolution (a language no longer being classified as an isolate). One historical change leading to isolate formation is the death of related languages, a case exemplified by Ket, a language spoken in Siberia. Today, Ket has no living relatives, and could be categorized as an isolate. However, because there is documented evidence of languages in the larger Yeniseian

family which have only recently become extinct, it is usually treated as a non-isolate in typological surveys. It is possible (likely, even) that many languages we call isolates today had living relatives only a few hundred years ago (much like Ket), but for which documentation is simply lacking. Isolate formation is also caused by divergence over time. A highly divergent member of a linguistic family for which evidence of relatedness is already minimal will, in the absence of a written record, eventually have that evidence erased. The uncertain relationship between Japanese and Korean may be an example of this (see Whitman 2012). Isolate dissolution on the other hand occurs through internal diversification leading to new daughter languages. An example of this is the Nivkh language family, all members of which are closely related (Mattisen 2003: 5; Gruzdeva 2022), and which therefore could have been an isolate in the relatively recent past. Given the current state of global documentation, isolate formation through divergence and the erasure of linguistic evidence is unlikely to occur in the future, while isolate dissolution through diversification may still happen.

The formation of an isolate is thus a continuous process with fuzzy, largely epistemic boundaries that impact categorization. Small families which have recently undergone isolate dissolution (such as the families of Ainu and Nivkh) are sometimes grouped together with isolates. By the same token, highly divergent languages within established families (such as Modern Albanian, Greek, or Armenian in Indo-European) that might soon undergo

isolate formation (in the absence of a written record) could also be grouped together with isolates, although we are not aware of any examples of this in the typological literature.

Because the difference between isolates and non-isolates is largely an epistemic issue, there is no a priori reason the distinction should give rise to different typologies (Bickel 2013; Campbell 2018a). Indeed, it is rare that isolates are talked about collectively, with the exceptions of historical linguistics and language sampling in typology. For historical linguistics, diachronic research on isolates cannot benefit from known relatives, so this research has to rely only on other methods, such as internal reconstruction (Campbell 2018b). As for linguistic typology, discussion on isolates is focused around whether and how to properly represent them in language samples of various sizes and compositions (Bybee et al. 1994; Bakker 2011; Miestamo et al. 2016; Jäger & Wahle 2021).

Nevertheless, statements alluding to special properties of isolates, typically vaguely defined and lacking systematic investigation, can be found in typological literature. For example, in a discussion of their sampling procedure, Bybee and colleagues (1994: 303) write that "[t]here is a chance that language isolates are significantly different from other language types"; however, they do not elaborate on what these differences might be. Similarly, in comparing rural signing varieties to spoken languages, DeVos and Pfau (2015: 280) state that these "exhibit typologically rare and complex features

that contribute uniquely to existing typologies, as has previously been reported for spoken language isolates"; once again, the details of these features are not described. Some more specific observations about the structure of language isolates have come out of typological surveys not directed at discovering isolate profiles. For example, Jäger and Wahle (2021: 11) report that head-final languages are "quite frequent among small families and isolates."

More elaborate hypotheses that language isolates could be typologically distinct from non-isolates do not suppose that any difference is caused by the isolate or non-isolate status of a given language, but presume that a typological profile could arise as an epiphenomenon due to other characteristics that happen to be more common among isolates. In other words, in all such hypotheses language isolates are used as proxies. In the following paragraphs we give an overview of the proposed mechanisms and their corresponding predictions about isolates, some of which we investigate in the current study.

One hypothesis is that large linguistic spreads could sweep through linguistic areas leaving in their periphery (mostly) isolate languages. These languages (perhaps freshly-minted as isolates) could maintain the older regional typological profiles (Seifart & Hammarström 2017). Along these lines, Nichols (1992) offers two observations about isolates differing from their neighboring languages: many languages with class marking located

outside of noun class hotbed regions (e.g. Africa, the Caucasus) are language isolates (Nichols 1992: 138), something which could (but does not necessarily) reflect an old areal pattern that has been broken up; and certain isolates in the northeastern periphery of Eurasia (Nivkh and Ainu[1]) tend towards head marking, while the majority of Eurasia tends toward dependent marking (Nichols 1992: 204). If isolates often have older, more widespread linguistic profiles, we would expect such patterns to be *local*, as they would not necessarily involve the same features across areas, but they would contrast with the typological profiles of local non-isolate languages. To test this hypothesis, ideally one would need to compare isolates and non-isolates in predefined areas of the world where large linguistic spreads have occurred. The implied relationship between isolates and languages at the edges of linguistic spreads is tentatively supported by a study of the geographic distribution of isolates (Urban 2021), a rare example of systematic investigation of such proposed proxy relationships. Under the prediction that isolates show a retreat of a family or language's former range towards major geographical barriers in response to encroachment by other languages, and using distances to coastlines and alpine mountain areas as measures of marginal environments, Urban (2021) finds a global pattern by which language isolates are found in closer proximity to alpine mountain areas than

---

[1] Note that Nivkh and Ainu are now considered to be small families with multiple languages, a categorization difficulty that has already been discussed.

non-isolates. However, this effect does not hold up statistically when applied within macroregions.

Aside from the hypothesis of isolates being remnants of old linguistic areas, several other mechanisms have been proposed that could cause typological differences between isolates and non-isolates as an epiphenomenon, but this time at a *global* (i.e. worldwide) level:

I.   The relative size of the speaker community (Thurston 1989; Trudgill 1989)

II. Hunter-gatherer vs agricultural subsistence mode (Bickel & Nichols 2020)

III. Social isolation (Thurston 1989; Wray & Grace 2007)

IV. Geographic isolation (Nichols 1992; Urban 2021)

In fact, none of these causal mechanisms target isolates specifically. Instead, isolates are viewed as proxies: they may form a distinct typological class under these hypotheses because they are more prone than non-isolates to have the above-mentioned characteristics.

Even though many of these hypotheses sound plausible, systematic evaluation of the relevance of such proxy relationships in the literature seems rare: in fact, to our knowledge, only the aforementioned study of Urban (2021) statistically investigates the correlation between isolate language status and geographic isolation. In addition, systematic surveys exploring

correlations between these four aspects of language communities per se (i.e. without the use of isolates as proxies) and their linguistic properties are also fairly rare. Exceptions include some investigations on the effect of community size on phonological inventory size and rates of lexical change (Moran et al. 2012; Greenhill et al. 2018), as well as of hunter-gatherer vs. agriculturalist languages (Bickel & Nichols 2020).

As an additional complication, in many hypotheses, factors such as small sized communities, hunter-gatherer societies, socially and/or geographically isolated languages, and dense community networks are considered jointly, and they seem to be considered interdependent. For instance, the concept of *esoteric* (in-group oriented) and *exoteric* (out-group oriented) language use developed in Thurston (1989), and to a greater degree in Wray and Grace (2007), jointly considers community size and social isolation. Trudgill (2017: 142) also observes that polysynthesis is very frequent in "relatively small" and "traditional, non-industrialized rural tribal communities," a joint consideration of community size and subsistence mode. Bickel and Nichols (2020:67), in a statistical survey of potential distinct linguistic profiles of hunter-gatherer vs. agriculturalist languages, note that "hunter-gatherer societies are usually smaller and less complex, with lower population density," with "kinship as a main organizing factor," and live "in long-standing conditions of considerable sociolinguistic isolation," a factor that they link to "increased structural complexity," following Trudgill (2011).

This prediction of social isolation fostering grammatical complexification (or the maintenance of old complexity over time) is repeated frequently throughout the typological literature, and is often linked to low numbers of L2 speakers. The causal mechanism assumed in such cases is that the presence of many L2 speakers (and intense language contact generally) pushes a language in a certain direction (Trudgill 1989, 2004, 2017). At the extreme end of L2 influence lies the creole exceptionalism hypothesis, where pidgins and creoles are predicted to show certain types of grammatical simplification (or, equivalently, to lack certain types of grammatical complexification) caused by a recent past in which they were only spoken by L2 speakers (McWhorter 2001, 2018). The expectation under such a causal mechanism is a typological profile for languages with high numbers of L2 speakers. If non-isolates and isolates are reasonable proxies for the relative number of L2 speakers, this hypothesis would then predict a typological profile for non-isolates and diversity among isolates. It does not predict complexification of any particular linguistic feature or even set of features, but complexification of the grammar when viewed as a whole, which could manifest itself in different features from language to language. A similar consequence holds for the hypothesis that geographic isolation may be characterized by special social dynamics leading to the conservation of archaic features (Nichols 2013; DeLancey 2014; Urban 2020): again, there is no reason to believe that the same archaic features would be shared across geographically isolated languages. Therefore, a feature-by-feature

comparison is not the best way to test such hypotheses. Rather, one would need to use composite measures (e.g. of grammatical complexity or of rates of change) in order to capture such tendencies among isolated languages, if present. However, devising such measures is a complicated task beyond the scope of this study (for more info, see e.g. Miestamo et al. 2008; Mufwene et al. 2017; Ehret et al. 2021).

In this paper we take a first step towards testing some of the above mentioned hypotheses using a large dataset of 68 isolate and 147 non-isolate languages (for a total of 215 stocks [2]) and 89 phonological and morphosyntactic typological features. Because of the lack of a clear boundary between small, homogenous families and isolates, we have adopted the classification of Glottolog 4.5 (Hammarström et al. 2021). We test for isolates having a different typological profile in comparison with non-isolates at both local and global levels. Due to the very uneven distribution of isolates worldwide (see Table 1) and to the limitations of our data (many features are rare), we tested for local isolate profiles at the level of macroareas, as defined in Hammarstöm & Donohue (2014).

---

[2] We use the term stock as the top-level genetic unit, i.e. covering both isolates and language families to avoid any ambiguity of the term family (since isolates can be considered families of one member).

Table 1. Number of families, isolates, and stocks per macroarea according to Glottolog (Hammarström et al. 2021)

Macroarea	Families	Isolates	Stocks	% isolates
Africa	37	16	53	30%
Australia	24	9	33	27%
Eurasia	26	12	38	32%
North America	44	31	75	41%
Papunesia	74	50	124	40%
South America	46	64	110	58%
*World total**	*419*	*182*	*601*	*30%*

* The totals take into account that many language stocks are spanning more than one macroarea.

2. Data and methodology

The typological dataset used to test the hypotheses in this study is part of the forthcoming typological database ATLAs of potentially areal features

focused on the Americas, still under construction by the authors and collaborators.[3] Since neither the language sample nor the typological features were initially selected to test for typological hypotheses regarding isolates, we will briefly describe the characteristics of the ATLAs database below and discuss the criteria used in their selection and their appropriateness.

2.1.   Language sample

The 319 languages included in ATLAs were selected according to the following criteria: phylogenetic diversity, geographic coverage (within the restrictions of the macroarea ratio explained below), and sufficient documentation. We excluded LOL languages (Literate, Official, and Lots of users), which Dahl (2015) suggested to be fairly unusual in the sense of reflecting recent and unevenly distributed modern social changes. Finally, where possible, we maximized the overlap with samples of existing

---

[3] The ATLAs database includes 319 languages and about 200 typological features organized in 20 feature sets and is being developed as part of the Swiss National Science Foundation Sinergia project CRSII5_183578 "Out of Asia: Linguistic Diversity and Population History." It is designed to target areal signals present in the Americas. As a consequence, the language sample is skewed by a 2:1 ratio toward North and South America, and the feature sets selected are previously proposed areal features of the Americas and other features with areal potential. Each feature set was created as a questionnaire designed or adapted by the authors. Coding was then done by the authors in collaboration with research assistants. The structure and coding of these features was not altered in any way for the present study.

typological databases, like WALS (Dryer & Haspelmath 2013), SAILS (Muysken et al. 2016), AUTOTYP (Bickel et al. 2021), Easterday (2019), and personal research. As mentioned, the focus of ATLAs is the Americas, so it includes twice as many languages from the Americas in comparison with the rest of the world. To constitute the language sample used in this study we included all isolates present in ATLAs and randomly sampled one language per family for the families with more than one representative. Therefore, all the languages included in our study belong to different stocks. Our study sample's genetic and geographic distribution is shown in Table 2, while a full list of all the languages included can be found in the Appendix. Throughout this paper, we use the language names, genetic and macroarea classification of Glottolog 4.5 (Hammarström et al. 2021).

In the sample used for this study, a little more than half of the languages are from the Americas (117 of 215, or 54.4%), and there are different numbers and proportions of isolates and non-isolates per macroarea. This can be explained by the characteristics of the original sample, but also from the uneven distribution of isolates across macroareas (see Table 1). Of the 215 stocks considered in the present study, 68 represent isolates (33%) and 147 represent non-isolate (67%) languages, a distribution that comes close to the global proportion of isolates and families among the world's language stocks (30% and 70%, respectively).

Table 2. Geographical and genetic distribution of the languages in our sample

Macroareas	Families	Isolates	Total sample
South America	37	20	**57**
North America	40	20	**60**
total Americas	***77***	***40***	***117***
Papunesia	15	10	**25**
Eurasia	20	6	**26**
Australia	19	3	**22**
Africa	16	9	**25**
total non-American	***70***	***28***	***98***
TOTAL	**147**	**68**	**215**

The bias toward the Americas in the original ATLAs sample happens to be toward two of the most diverse macroareas, both in terms of number of stocks, but also in proportion of isolates, with only Papunesia coming close (see Table 1). This gives us the opportunity to have a fairly dense sample in terms of isolates, although including more languages from Papunesia would be desirable.

2.2. Features

For the present study, we selected 12 feature sets from the ATLAs database: six phonological and six morphosyntactic ones. Each feature set is focused on a general phenomenon and includes a number of individual features that represent finer-grained distinctions and more detailed typological

characteristics. As an example, for the phonological phenomenon of lateral consonants (feature set abbreviation Lat), we surveyed the presence of:

- a lateral approximant phoneme /l/ and its allophones (feature abbreviation Lat.01)

- a phonemic voiceless lateral approximant /l̥/ (Lat.02)

- a phonemic lateral fricative /ɬ/ (Lat.03)

- a phonemic alveolar lateral affricate /tɬ/ (Lat.04)

- a phonemic alveolar lateral ejective affricate /tɬ'/ (Lat.05)

- a phonemic palatal lateral approximant /ʎ/ (Lat.06)

Some features such as the last five (Lat.02-06) are binary yes-no questions, while the first feature (Lat.01) is multistate: its full version is whether the language has a lateral approximant phoneme /l/ or allophone [l] and if so, with what other allophones this /l/ or [l] alternates. We determined five possible answers (or states): <laterals or glides>; <rhotic>; <n or d>; <n or d and rhotic>; <no>.

Table 3 lists the feature sets and the 89 associated features, while a full account of each of the features and their states is available in the online Supplementary Materials (Feature List).

Table 3. List of the 12 feature sets and the associated features

Feature sets (abbreviation)	Total features
Phonology	
Coronals (Cor)	7
Laterals (Lat)	10
Glottalized consonants (Glot)	8
Velar-Uvular distinction (VelUv)	3
Tone & Prominence (ToneProm)	7
Syllable structure (Syll)	6
Morphosyntax	
Demonstratives (Dem)	8
Personal pronouns (Ppron)	8
Sociative causative (SocCaus)	2
Apprehensional morphology (Appr)	5
SG-PL verb root alternation (SgPl)	13
Word order (WordOrder)	12
Total	**89**

Many of the features in the ATLAs database have been coded so as to avoid or minimize logical dependencies between features. Therefore, some features are conditioned on the presence of a particular state in another

feature, making <Not Applicable (NA)> a possible state. In these cases, the statistical tests described in the next section were conducted on a subset of the languages (excluding all languages with <NA>). For instance, one of the features in the Personal Pronouns feature set is whether the language has a minimal/augmented system (Ppron.05), a feature that is conditioned on whether the language has an inclusive/exclusive distinction (Ppron.01). As a matter of fact, a minimal/augmented system presupposes the existence of an inclusive/exclusive distinction.[4] All languages with no inclusive/exclusive distinction were thus coded as <NA> for Ppron.05, which excluded them from the statistical tests run for that specific feature: only the languages with the states <yes> and <no> were considered in those tests.

2.3. Statistical methods

In order to test for a global or macroareal profile of isolates, we used appropriate statistical methods to assess if the prevalence of any feature among isolates is statistically significantly different than among non-isolates. Since in many cases the features investigated are rare and our sample sizes are small, we have opted for the use of exact tests rather than asymptotic ones.

---

[4] Minimal/Augmented systems are particularly well represented in Australia. Many languages with an inclusive (1+2) / exclusive (1+3) distinction do not distinguish between an *augmented* inclusive (1+2+3) and an augmented exclusive pronoun (1+3+3). See Cysouw (2009: 80-90) for a detailed overview of the phenomenon.

All tests were two-sided, since our null hypothesis is that there is no difference between isolates and non-isolates for any of the features tested and the chosen level of significance was 0.05.

In the case of binary features (the majority of our features) we used Boschloo's test (Boschloo 1970), which is an exact unconditional test that is uniformly more powerful than the more popular Fisher's exact test (Mehrotra et al. 2003; Ruxton & Neuhäuser 2010). For Boschloo's test we used the R package Exact (Calhoun 2021). The model was binomial, since the number of isolates vs non-isolates was known beforehand, and the two-sided method was set to *central*. In the case of multistate features, we used the Freeman and Halton extension of Fisher's exact test, as implemented in the R package stats (R Core Team 2019). To correct our p-values for multiple testing we controlled the false discovery rate (FDR), which is a more powerful adjustment method than controlling the family-wise error rate (FWER, e.g. using the Bonferroni correction) (Jafari & Ansari-Pour 2019). We had two composite null hypotheses:

$H_0$1. There is no macroarea-specific typological profile of isolates (vs non-isolates) for any of the features and macroareas tested.

$H_0$2. There is no global typological profile of isolates (vs non-isolates) for any of the features tested.

We performed 357 tests that could reject $H_0 1$, and 89 that could reject $H_0 2$. Therefore, we adjusted our p-values within these two groups of tests separately (Rubin 2021). We first tested for the presence of macroareal-specific profiles. As will be seen in the results section, no feature distribution was statistically significant after correction, so we pooled all our data together to subsequently test for the presence of a global profile for isolates.

3. Results

The results of the macroarea-specific profiles were evaluated first. Of the 357 tests of the features across 6 macroareas, only 11 had raw p-values smaller than 0.05, which are presented in Table 4 and described below. The full table of results for all features is given in the online Supplementary Materials.

Table 4. Features with raw p-values < 0.05 by macroarea

Macroarea	Feature	State	Isolates	Non-isolates	Raw p-value
North America	**Ppron.06** Morphological relationship between 1SG and 1PL/1EXCL	1excl = 1sg + PL	6	14	0.002
		1excl includes 1sg	6	0	

		no relation	7	22	
	Ppron.07 Morphological relationship between 1SG and 1INCL	1incl = 1sg + PL	2	1	0.008
		1incl includes 1sg	2	1	
		no relation	0	10	
South America	**WordOrder.01** Order of basic declarative transitive clause	no dominant order	3	2	0.012
		SOV	7	24	
		SVO	2	6	
		V-first	3	0	
		VSO	2	1	
		VOS	0	1	
		OVS	0	1	
		OSV	2	0	
	WordOrder.01e First element of basic transitive clause	no dominant order	3	2	0.015
		S-first	9	30	
		V-first	5	2	
		O-first	2	1	
	Syll.03 Total number of consonants in maximum syllable	1	4	4	0.034
		2	6	18	
		3	8	6	
		4	0	6	

		5	1	0	
		7	0	2	
	VelUv.01 Any velar-uvular distinction	present	0	7	0.045
		absent	19	31	
	VelUv.01a Velar-uvular distinction in plosives[5]	present	0	7	0.045
		absent	19	31	
Papunesia	**WordOrder.03a** Part of speech of nominal attributives	modifying POS	3	10	0.026
		verb	3	0	
	Ppron.09 Pronominal distinction between 3SG and 3PL	present	10	10	0.044
		absent	0	5	
Australia	**Syll.06** Syllable coda complexity	simple	2	0	0.029
		sonorant	1	2	
		complex	0	16	
Eurasia	**Syll.05** Syllable onset complexity	C	1	9	0.042
		CG	3	3	
		CC+	0	8	

[5] VelUv.01 and VelUv.01a in South America, are identical because none of the South American languages in our sample had a velar-uvular distinction only among their non-plosives phonemes.

Some of the features in Table 4 show an interdependence. Two features in South America, WordOrder.01 (basic word order) and WordOrder.01e (first element in basic word order), are essentially capturing the same pattern, since the differences in word order between isolates and non-isolates are driven mostly by the greater number of non-isolates with SOV word order, which is S-first. The two North American features with raw p-values below 0.05 have to do with the morphological relationship of the 1sg and 1pl (whether inclusive or exclusive), and non-isolates in our sample more frequently lack a relationship between 1sg and 1pl forms than isolates.

However, once FDR adjustment of p-values was applied for multiple testing, none of the above differences between isolates and non-isolates were statistically significant. Given the lack of statistically significant differences within macroareas, we pooled our entire sample to test for significant differences between isolates and non-isolates globally. From a total of 89 features, 5 had raw p-values less than 0.05, summarized in Table 5 and described below. The full results are present in the online Supplementary Materials.

Table 5. Features with raw p-value < 0.05 at the global level

Feature	State	Isolates	Non-isolates	Raw p-value

Ppron.01a Presence of an inclusive-exclusive distinction	present	21	73	0.019
	absent	45	75	
Ppron.05 Presence of a minimal-augmented distinction	present	1	17	0.041
	absent	19	49	
Ppron.06 Relationship between 1SG and 1PL or 1EXCL	1sg + PL = 1excl	13	43	0.046
	1sg includes 1excl	9	7	
	no relation	43	86	
VelUv.01 Any velar-uvular distinction	present	8	32	0.047
	absent	59	116	
SgPl.05 Number suppletion among motion verbs, if number suppletion is present	present	23	34	0.047
	absent	1	10	

Once again, none of these features remains statistically significant after FDR adjustment of p-values for multiple testing.

4. **Discussion and conclusion**

After testing for both the macroareal and global hypotheses of isolate/non-isolate typological differences, we were unable to find a statistically significant distinction for any of our features. In some cases, there were opposite tendencies at the macroareal and global level. For example, isolates in North America more frequently lacked a morphological relationship between 1sg and 1excl (or 1pl) than did non-isolates, while globally isolates were more likely than non-isolates to show such a relationship. Similarly, isolates in South America were more likely than non-isolates to lack a velar-uvular distinction in plosives, while globally the reverse was the case. However, neither of these patterns were statistically significant, and they should be interpreted as statistical noise in a large data set.

Our dataset was specifically designed to capture areal signals and thus should have picked up a macroareal difference between isolates and non-isolates, if one existed. However, there are a few ways that we could have missed an existing local profile. The first is that our definition of isolates may have been too strict to find a pattern. We followed Glottolog in categorizing languages that belong to small and young families (such as Nivkh) as non-isolates, but it may be that grouping these together with isolates would reveal a typological difference with widespread families of the area. In this case, the isolate category is a poor proxy for the difference between spreading and remnant language families, and different criteria should be used to distinguish these two groups. Another possibility is that the macroareal level is too coarse

to detect an old areal profile retained in isolates. For instance, within Eurasia, language isolates in East Asia could carry a completely different areal profile than isolates in the Middle East. However, if a difference between isolates and non-isolates exists only at a very local level, there may not be enough isolates in many parts of the world to test this hypothesis. Nevertheless, our data strongly suggest that there is no differentiation between isolates and non-isolates at the macroareal scale.

At the global level, it could be argued that we failed to find a result because our feature set did not target any hypothesized effects (such as a complexity measure or lexical differences) that could be caused by differences in community size, subsistence mode, social isolation (and relatedly the historic frequency of L2 speakers), and geographic isolation. Some of these differences might even be absent in our database because of our exclusion of LOL languages (Dahl 2015), all of which are spoken by large communities of agriculturalists with little to no social or geographic isolation (and additionally have highly literate populations with political linguistic legitimacy). As mentioned in the introduction, hypotheses of *complexification* or *conservatism* of isolates are not testable on a feature-by-feature basis. Even though several of our features plausibly correlate with a concept of complexity or irregularity (contrastive lexical tone, syllable complexity, number and type of phonemic distinctions, irregular verbal plurals, number distinctions in pronouns, and distinctions among

demonstratives), the absence of an isolate profile in these features individually should not be interpreted as weakening this hypothesis. As mentioned, testing this hypothesis would require a composite measure which would quantify complexity over many linguistic features. With these caveats in mind, our study nevertheless cannot reject the null hypothesis that globally isolates and non-isolates are typologically similar. We think this result is unlikely to change with the addition of more blindly-chosen features, especially after the appropriate corrections for multiple testing. However, more specific and targeted predictions may prove fruitful.

If future researchers would like to investigate such hypotheses further, we think our study cautions against the use of isolates as proxies for other sociolinguistic, cultural, or geographic properties. Instead, researchers would be better served by measuring these properties more directly. A few potential avenues of exploration which are not affected by our negative result are:

1. The possibility that speaker community size, subsistence mode, social isolation, or geographic isolation push a language to develop certain linguistic features, independent from the status of a language as an isolate.

2. The possibility that a much more local profile exists distinguishing locally old linguistic stocks (some of which may be isolates) from locally newer linguistic stocks (most or all of which will be non-isolates).

3. A typological difference between languages depending on phylogenetic isolation, rather than their status as an isolate.

The family of hypotheses in (1) looks at other factors besides the status of a language as an isolate. Some work along these lines has been done in the cultural domain, as in Majid and Kruspe (2018), which shows that hunter-gatherer populations are better than agriculturalists at distinguishing odors. Further differences in the linguistic and grammatical domain have yet to be established, and other causal factors await investigation. As observed in the introduction, many of these proposed causal mechanisms (community size, social isolation, etc) are interlinked and may be difficult to disentangle from one another. Definitions of what counts as sufficient social isolation or *small* community size may be difficult to come up with, and a continuous rather than binary metric might be more appropriate.

The family of hypotheses in (2) would require a repeat of the current study (perhaps with a different or expanded feature set) but with much denser sampling of a particular region or regions. Again, we think it would be fruitful to abandon the idea of a difference between non-isolates and isolates, and group languages according to their age in the area. Members of some families would be grouped together with isolates as locally old, while other families would stand apart as newcomers. Unfortunately, there may be very few regions of the world for which there is enough data to show such a difference.

The family of hypotheses in (3) also abandon the notion of *isolate* as a special category, this time in favor of a broader idea of phylogenetic

isolation. We have already discussed that the category of language isolate as commonly defined is in fact an epistemological issue. However, if the intuition is that languages with a long, *solitary* phylogenetic history could be typologically different or associated with certain sociolinguistic traits, then one should aim at languages that we *know* have this type of history, rather than isolates, which by definition have *unknown* phylogenetic histories. Such endeavors would be greatly assisted by the growing number of established phylogenies and the development of metrics of phylogenetic isolation (e.g. phylogenetic distance from the closest relatives maybe incorporating notions of duration or geographic distance).

All these directions still remain open for future researchers to investigate whether and how a language's history and ecology shape its typological properties. As it stands currently, to the best of our current knowledge, it appears that isolates and non-isolates are not distinguishable from one another on typological grounds.

5. Acknowledgements

5.1.   Author Contributions

N.C.P., K.P.v.D., D.I., S.E., and M.V. conceived of the study. N.C.P., K.P.v.D., D.I., S.E., F.R., and M.V. developed the typological variables and collected data. N.C.P. and D.I. ran statistical analyses. N.C.P., D.I., and M.V. interpreted the results. N.C.P., D.I., and M.V. wrote the paper, K.P.v.D., and F.R. reviewed the manuscript.

5.2.   Supplementary Materials

The list of features used for this study, along with short definitions, are presented in the online supplementary materials (https://osf.io/fvpz2/). Also in the supplementary materials are the full tables of statistical results at the

macroareal and global level (from which Tables 4 and 5 are derived), and a CLDF database of all features and values for the sample.

6. References

Bakker, Dik. 2011. Language sampling. In *The Oxford Handbook of Linguistic Typology*, Jae Jung Song (ed): 100–127. Oxford: Oxford University Press.

Bickel, Balthasar. 2013. Distributional biases in language families. In *Language Typology and Historical Contingency*, Balthasar Bickel, Leonore Grenoble, David Peterson & Alan Timberlake (eds): 415–444. https://doi.org/10.1075/tsl.104.19bic.

Bickel, Balthasar & Nichols, Johanna (2020). Linguistic typology and hunter-gatherer languages. In *The Language of Hunter-Gatherers*, Tom Güldemann, Patrick McConvell, Richard A. Rhodes (eds): 67–75. Cambridge: Cambridge University Press.

Bickel, Balthasar, Nichols, Johanna, Zakharko, Taras, Witzlack-Makarevich, Alena, Hildebrandt, Kristine, Rießler, Michael, Bierkandt, Lennart, Zúñiga, Fernando & Lowe, John B. 2021. *The AUTOTYP database* [Data set]. Version v0.1.2. https://doi.org/10.5281/zenodo.4574513.

Boschloo, R. D. (1970). Raised conditional level of significance for the 2× 2-table when testing the equality of two probabilities. *Statistica Neerlandica* 24(1): 1–9. https://doi.org/10.1111/j.1467-9574.1970.tb00104.x

Bybee, Joan, Perkins, Revere & Pagliuca, William. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Calhoun, Peter. 2021. *Exact: Unconditional exact test* [R package]. Version 3.1. https://CRAN.R-project.org/package=Exact

Campbell, Lyle. 2018a. Introduction. In *Language Isolates*, Lyle Campbell (ed): xi–xiv. London: Routledge.

Campbell, Lyle. 2018b. Language isolates and their history. In *Language Isolates*, Lyle Campbell (ed): 1–18. London: Routledge.

Cysouw, Michael. 2009. *The Paradigmatic Structure of Person Marking*. Oxford: Oxford University Press.

Dahl, Östen. 2015. How WEIRD are WALS languages? Paper presented at MPI-EVA, Leipzig. www.eva.mpg.de/fileadmin/content_files/linguistics/conferences/2015-diversity-linguistics/Dahl_slides.pdf

DeLancey, Scott. 2014. Creolization in the divergence of the Tibeto-Burman languages. In *Trans-Himalayan Linguistics: Historical and Descriptive Linguistics of the Himalayan Area*, Nathan W. Hill & Thomas Owen-Smith (eds): 41–70. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110310832.41

DeVos, Connie & Pfau, Roland. 2015. Sign language typology: The contribution of rural sign languages. *Annual Review of Linguistics* 2015.1: 265–88. https://doi.org/10.1146/annurev-linguist-030514-124958

Dryer, Matthew S. & Haspelmath, Martin (eds). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info

Easterday, Shelace. 2017. Highly Complex Syllable Structure: A Typological Study of its Phonological Characteristics and Diachronic Development. PhD dissertation, University of New Mexico. https://digitalrepository.unm.edu/ling_etds/51/

Easterday, Sheleace. 2019. *Highly Complex Syllable Structure: A Typological and Diachronic Study.* Berlin: Language Science Press. https://doi.org/10.5281/zenodo.3268721

Ehret, Katherina, Blumenthal-Dramé, Alice, Bentz, Christian, & Berdicevskis, Aleksandrs. 2021. Meaning and measures: interpreting and evaluating complexity metrics. *Frontiers in Communication* 61. https://doi.org/10.3389/fcomm.2021.640510

Greenhill, Simon J., Hua, Xia, Welsh, Caela F., Schneemann, Hilde, & Bromham, Lindell. 2018. Population size and the rate of language evolution: A test across Indo-European, Austronesian, and Bantu Languages. *Frontiers in Psychology* 9: 576. https://doi.org/10.3389/FPSYG.2018.00576

Gruzdeva, Ekaterina. 2022. On diversification of Nivkh Varieties. *Northern Language Studies*, Special Volume: 25–33. http://hdl.handle.net/2115/84930

Hammarström, Harald, & Donohue, Mark. 2014. Some principles on the use of macro-areas in typological comparison. *Language Dynamics and Change* 4(1): 167-187. https://doi.org/10.1163/22105832-00401001

Hammarström, Harald, Forkel, Robert, Haspelmath, Martin & Bank, Sebastian. 2021. *Glottolog 4.5* [Dataset]. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.5772642

Jafari, Mohieddin & Ansari-Pour, Naser. 2019. Why, when and how to adjust your P values? *Cell Journal (Yakhteh)* 20(4): 604–607. https://doi.org/10.22074/CELLJ.2019.5992

Jäger, Gerhard & Wahle, Johannes. 2021. Phylogenetic typology. *Frontiers in Psychology* 12: 682132. https://doi.org/10.3389/fpsyg.2021.682132

Maddieson, Ian (2005). Correlating phonological complexity: data and validation. *UC Berkeley PhonLab Annual Report 1*(1): 216-229.

Maddieson, Ian. 2013. Syllable structure. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds). Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/12

Majid, Asifa & Kruspe, Nicole. 2018. Hunter-gatherer olfaction is special. *Current Biology* 28(3): 409–413. https://doi.org/10.1016/j.cub.2017.12.014

Mattissen, Johanna. 2003. *Dependent-Head Synthesis in Nivkh. A Contribution to a Typology of Polysynthesis*. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.57

McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 10: 125–166. https://doi.org/10.1515/lity.2001.001

McWhorter, John H. 2018. *The Creole Debate*. Cambridge: Cambridge University Press.

Mehrotra, Devan V., Chan, Ivan S. F. & Berger, Roger L. 2003. A cautionary note on exact unconditional inference for a difference between two independent binomial proportions. *Biometrics* 59(2): 441–450. https://doi.org/10.1111/1541-0420.00051

Miestamo, Matti, Bakker, Dik, & Arppe, Antti. 2016. Sampling for variety. *Linguistic Typology* 20(2): 233–296. https://doi.org/10.1515/lingty-2016-0006

Miestamo, Matti, Sinnemäki, Kaius, & Karlsson, Fred (eds). 2008. *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/slcs.94

Moran, Steven, McCloy, Daniel & Wright, Richard. 2012. Revisiting population size vs. phoneme inventory size. *Language* 88(4): 877–93. https://doi.org/10.1353/lan.2012.0087.

Mufwene, Salikoko S., Coupé, Christophe & Pellegrino, François. 2017. Complexity in language: A multifaceted phenomenon. In *Complexity in Language: Developmental and Evolutionary Perspectives*, Salikoko S. Mufwene, Christophe Coupé & François Pellegrino (eds): 1–29. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107294264.001

Muysken, Pieter, Hammarström, Harald. Krasnoukhova, Olga, Müller, Neele, Birchall, Joshua, van de Kerke, Simon, O'Connor, Loretta, Danielsen, Swintha, van Gijn, Rik & Saad, George. 2016. *South American Indigenous Language Structures (SAILS) Online* [Data set]. Jena: Max Planck Institute for the Science of Human History. https://sails.clld.org/

Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.

Nichols, Johanna. 2013. The vertical archipelago: Adding the third dimension to linguistic geography. In *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives*, Peter Auer, Martin Hilpert, Anja Stukenbrock & Benedikt Szmrecsanyi (eds): 38–

60. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110312027.38

R Core Team. 2019. R: A language and environment for statistical computing [Computer software]. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Rubin, Mark. 2021. When to adjust alpha during multiple testing: a consideration of disjunction, conjunction, and individual testing. *Synthese* 199(3–4): 10969–11000. https://doi.org/10.1007/S11229-021-03276-4

Ruxton, Graeme D. & Neuhäuser, Markus. 2010. Good practice in testing for an association in contingency tables. *Behavioral Ecology and Sociobiology* 64(9): 1505–1513. https://doi.org/10.1007/s00265-010-1014-0

Seifart, Frank & Hammarström, Harold. 2017. Language isolates in South America. In *Language Isolates*, Lyle Campbell (ed): 260–287. https://doi.org/10.4324/9781315750026-10

Thurston, William R. 1989. How exoteric languages build a lexicon: esoterogeny in West New Britain. In *VICAL 1: Oceanic Languages: Papers from the Fifth International Conference on Austronesian Linguistics*, Ray Harlow & Robin Hooper (eds): 555-579. Auckland: Linguistic Society of New Zealand.

Trudgill, Peter. 1989. Contact and isolation in linguistic change. In *Language Change: Contributions to the Study of its Causes*, Leiv E. Breivik & Ernst H. Jahr (eds): 227–237. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110853063.227

Trudgill, Peter. 2004. Linguistic and social typology. In *The Handbook of Language Variation and Change*, J. K. Chambers, Peter Trudgill,

Natalie Schilling-Estes (eds): 707–728. Oxford: Blackwell. https://doi.org/10.1002/9780470756591.ch27

Trudgill, Peter. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.

Trudgill, Peter. 2017. Sociolinguistic typology: Social structure and linguistic complexity. In *The Cambridge Handbook of Linguistic Typology*, Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds): 124–150. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316135716.005

Urban, Matthias. 2020. Mountain linguistics. *Language and Linguistics Compass* 14(9): e12393. https://doi.org/10.1111/lnc3.12393

Urban, Matthias. 2021. The geography and development of language isolates. *Royal Society Open Science* 8(4): 202232. https://doi.org/10.1098/rsos.202232

Whitman, J. (2012). The relationship between Japanese and Korean. In *The languages of Japan and Korea*, Nicolas Tranter (ed): 24-38.. London and New York: Routledge.

Wray, Alison & Grace, George. W. 2007. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117(3): 543-578. https://doi.org/10.1016/j.lingua.2005.05.005

7. Appendix

The list of languages is included in Table A. The features and the associated questionnaires used are available on OSF at https://osf.io/fvpz2/.

Table A. Full list of languages included in this study with their family affiliation

Macroarea	Glottocode	Name	Family
Africa	bang1363	Bangime	
	benc1235	Bench	Ta-Ne-Omotic
	bert1248	Berta	
	cent2050	Central Kanuri	Saharan
	hadz1240	Hadza	
	kenu1243	Kenuzi	Nubian
	koyr1242	Koyraboro Senni Songhai	Songhay
	kron1241	Krongo	Kadugli-Krongo
	kuna1268	Kunama	
	lang1324	Lango (Uganda)	Nilotic
	madi1260	Ma'di	Central Sudanic
	maba1277	Maba (Chad)	Maban
	murl1244	Murle	Surmic
	nara1262	Nara	
	sand1273	Sandawe	
	shek1245	Sheko	Dizoid
	siam1242	Siamou	
	juho1239	South-Eastern Ju	Kxa
	tomm1242	Tommo So Dogon	Dogon
	yoru1245	Yoruba	Atlantic-Congo
	bamb1269	Bambara	Mande
	kamb1316	Kambaata	Afro-Asiatic
	laal1242	Laal	

	nama1264	Nama (Namibia)	Khoe-Kwadi
	shab1252	Shabo	
Australia	bard1255	Bardi	Nyulnyulan
	gaga1251	Gaagudju	
	gara1269	Garrwa	Garrwan
	goon1238	Gooniyandi	Bunaban
	djam1255	Jaminjung-Ngaliwurru	Mirndi
	kaya1319	Kayardild	Tangkic
	nucl1327	Limilngan	Limilngan-Wulna
	mang1381	Mangarrayi	Mangarrayi-Maran
	mari1424	Marithiel	Western Daly
	maun1240	Mawng	Iwaidjan Proper
	mull1237	Mullukmulluk	Northern Daly
	murr1258	Murriny Patha	Southern Daly
	ngar1284	Ngarinyin	Worrorran
	urni1239	Urningangg	Giimbiyu
	wage1238	Wageman	
	ward1246	Wardaman	Yangmanic
	nung1290	Wubuy	Gunwinyguan
	kitj1240	Kitja	Jarrakan
	naka1260	Nakara	Maningrida
	kamu1258	Kamu	Eastern Daly
	mart1255	Martuthunira	Pama-Nyungan
	tiwi1244	Tiwi	
Eurasia	abkh1244	Abkhazian	Abkhaz-Adyge
	gily1242	Amur Nivkh	Nivkh
	basq1248	Basque	

	dukh1234	Dukha	Turkic
	galo1242	Galo	Sino-Tibetan
	nucl1302	Georgian	Kartvelian
	halh1238	Halh Mongolian	Mongolic-Khitan
	hatt1246	Hattic	
	ainu1240	Hokkaido Ainu	Ainu
	ingu1240	Ingush	Nakh-Daghestanian
	iris1253	Irish	Indo-European
	iumi1238	Iu Mien	Hmong-Mien
	kett1243	Ket	Yeniseian
	kusu1250	Kusunda	
	kyon1247	Kyongsangdo	Koreanic
	negi1245	Negidal	Tungusic
	sout2750	Southern Yukaghir	Yukaghir
	nene1249	Tundra Nenets	Uralic
	akab1249	Akabea	Great Andamanese
	bulo1242	Bulo Stieng	Austroasiatic
	buru1296	Burushaski	
	chuk1273	Chukchi	Chukotko-Kamchatkan
	laoo1244	Lao	Tai-Kadai
	miya1260	Miyako-Jima	Japonic
	niha1238	Nihali	
	sume1241	Sumerian	
North America	achu1247	Achumawi	Palaihnihan
	alse1251	Alsea-Yaquina	
	ariz1237	Arizona Tewa	Kiowa-Tanoan
	chim1301	Chimariko	

	chin1286	Clatsop-Shoalwater Chinook	Chinookan
	cuit1236	Cuitlatec	
	coos1249	Hanis	Coosan
	high1242	Highland Oaxaca Chontal	Tequistlatecan
	mari1440	Maricopa	Cochimi-Yuman
	moha1258	Mohawk	Iroquoian
	nort2942	North Slavey	Athabaskan-Eyak-Tlingit
	nuuc1236	Nuu-chah-nulth	Wakashan
	quil1240	Quileute	Chimakuan
	sand1278	San Dionisio del Mar Huave	Huavean
	sout2985	Southern Sierra Miwok	Miwok-Costanoan
	sout2956	Southern Haida	Haida
	bell1243	Bella Coola	Salishan
	cadd1256	Caddo	Caddoan
	cahu1264	Cahuilla	Uto-Aztecan
	cent2127	Central Alaskan Yupik	Eskimo-Aleut
	kala1400	Central Kalapuya	Kalapuyan
	coah1252	Coahuilteco	
	cree1270	Creek	Muskogean
	hueh1236	Huehuetla Tepehua	Totonacan
	ines1240	Ineseño	Chumashan
	kich1262	K'iche'	Mayan
	karo1304	Karok	
	klam1254	Klamath-Modoc	
	kute1249	Kutenai	
	lako1247	Lakota	Siouan
	mali1285	Malinaltepec Me'phaa	Otomanguean

	misk1235	Mískito	Misumalpan
	mola1238	Molale	
	natc1249	Natchez	
	nisg1240	Nisga'a	Tsimshian
	yoku1256	Northern Yokuts	Yokutsan
	nort2951	Northwest Maidu	Maiduan
	yaki1237	Northwest Sahaptin	Sahaptian
	pure1242	Purepecha	Tarascan
	rama1270	Rama	Chibchan
	sali1253	Salinan	
	seri1257	Seri	
	shas1239	Shasta	Shastan
	sius1254	Siuslaw	
	sout2982	Southeastern Pomo	Pomoan
	swam1239	Swampy Cree	Algic
	take1257	Takelma	
	timu1245	Timucua	
	toll1241	Tol	Jicaquean
	tonk1249	Tonkawa	
	toto1305	Totontepec Mixe	Mixe-Zoque
	tuni1252	Tunica	
	wapp1239	Wappo	Yuki-Wappo
	wash1253	Washo	
	west2632	Western Keres	Keresan
	wint1259	Wintu	Wintuan
	xinc1242	Xinca-Chiquimulilla	Xincan
	yana1271	Yana	

	yuch1247	Yuchi	
	zuni1245	Zuni	
Papunesia	abun1252	Abun	
	alam1246	Alamblak	Sepik
	wara1302	Bauni	Sko
	bilu1245	Bilua	
	buki1249	Bukiyip	Nuclear Torricelli
	daga1275	Daga	Dagan
	duna1248	Duna	
	gras1249	Grass Koiari	Koiarian
	imon1245	Imonda	Border
	kuot1243	Kuot	
	lavu1241	Lavukaleve	
	nucl1622	Marind	Anim
	maib1239	Maybrat-Karon	
	meny1245	Menya	Angan
	nucl1633	Nimboran	Nimboranic
	savo1255	Savosavo	
	sulk1246	Sulka	
	taia1239	Taiap	
	tuka1248	Tukang Besi North	Austronesian
	yima1243	Yimas	Lower Sepik-Ramu
	roto1249	Rotokas	North Bougainville
	nucl1632	Sentani	Sentanic
	tabo1241	Tabo	
	tido1248	Tidore	North Halmahera
	yaum1237	Yau (Morobe Province)	Nuclear Trans New Guinea

	agua1253	Aguaruna	Chicham
	amar1274	Amarakaeri	Harakmbut
	awac1239	Awa-Cuaiquer	Barbacoan
	bora1263	Bora	Boran
	boro1282	Bororo	Bororoan
	cayu1262	Cayubaba	
	chip1262	Chipaya	Uru-Chipaya
	chiq1248	Chiquitano	
	guah1255	Guahibo	Guahiboan
	guat1253	Guató	
	cacu1241	Kakua	Kakua-Nukak
	pira1253	Pirahã	
	emer1243	Teko	Tupian
South America	yura1255	Yuracaré	
	hixk1239	Hixkaryána	Cariban
	ayor1240	Ayoreo	Zamucoan
	nant1250	Nanti	Arawakan
	chol1284	Cholón	Hibito-Cholon
	embe1260	Emberá-Catío	Chocoan
	esee1248	Ese Ejja	Pano-Tacanan
	fuln1247	Fulniô	
	hupd1244	Hup	Nadahup
	iran1263	Irántxe-Münkü	
	iton1250	Itonama	
	kain1272	Kaingang	Nuclear-Macro-Je
	kano1245	Kanoê	
	kari1254	Kariri	

kana1291	Katukína-Kanamarí	Katukinan
kwaz1243	Kwaza	
lako1248	Lakondê	Nambiquaran
leco1242	Leco	
jama1261	Madi	Arawan
mapu1245	Mapudungun	Araucanian
mose1249	Mosetén-Chimané	
movi1243	Movima	
muni1258	Muniche	
muru1274	Murui Huitoto	Huitotoan
paez1247	Páez	
pila1245	Pilagá	Guaicuruan
puin1248	Puinave	
qawa1238	Qawasqar	Kawesqar
sali1298	Sáliba	Jodi-Saliban
sana1298	Sanapaná	Lengua-Mascoy
sant1432	Santiago del Estero Quichua	Quechuan
chay1248	Shawi	Cahuapanan
sion1247	Siona-Tetete	Tucanoan
sout2996	Southern Aymara	Aymaran
tehu1242	Tehuelche	Chonan
ticu1245	Ticuna	Ticuna-Yuri
trum1247	Trumai	
urar1246	Urarina	
wara1303	Warao	
wari1268	Wari'	Chapacuran
wich1262	Wichí Lhamtés Nocten	Matacoan

	yagu1244	Yagua	Peba-Yagua
	yano1262	Yanomámi	Yanomamic
	zapa1253	Záparo	Zaparoan