LANGUAGE AND LINGUISTICS IN MELANESIA

Journal of the Linguistic Society of Papua New Guinea
ISSN: 0023-1959

Special Issue 2012

Harald Hammarström & Wilco van den Heuvel (eds.)
History, contact and classification of Papuan languages

Part Two

A classification of Papuan languages

Søren Wichmann
Max Planck Institute for Evolutionary Anthropology
wichmann@eva.mpg.de

ABSTRACT
This paper provides a basic classification of 737 Papuan doculects pertaining to 513 different ISO 639-3 codes, in addition to 9 doculects that have not been assigned ISO 639-3 codes. *Ethnologue* (Lewis 2009) catalogues 848 non-Austronesian languages of New Guinea. Thus, this paper covers 60% of these languages. The point of the paper is to provide a solid benchmark for the classification of languages in a region which is clearly the most poorly understood in the world. The classification combines two different proposals, one of which is the classification by Harald Hammarström (2010), augmented by personal correspondence (2012), and the other is a classification based on methods of the Automated Similarity Judgment Program (ASJP). The former represents a conservative sifting of published evidence for language family affiliations and the latter provides an automated classification based on similarity among 40 lexical items selected for maximal stability. An ASJP tree annotated for Hammarström's families allows for identifying cases where the latter apparently fail to be coherent and should therefore possibly be broken up into smaller units, as well as cases where families should possibly be merged. The resulting classification will be even more conservative than Hammarström's in many cases, but it will also contain proposals for wider relationships not considered supported by Hammarström, including several proposals that have not been made before in the literature.

KEYWORDS: Language Classification, Papuan languages, Lexicostatistics, Levenshtein Distance, Neighbor-Joining

0. Introduction

The dominating trend in the historical linguistics of Papuan languages has been to cast the net widely and quickly gather languages into sometimes vast families based on loose counts of cognates, similarities in pronouns, typological similarity or simply geographical proximity (Foley 1986). While much progress has been made, it is certainly not an exaggeration to claim that non-Austronesian New Guinea still is the most poorly understood larger world area in terms of historical linguistics. In such a situation it behooves the comparative linguist to be conservative when assessing genealogical relations. Once the minimal family units are established the search for wider relations can begin.

In this paper two approaches are combined. One is the classification into families of Hammarström (2010), updated through personal communication (2012) from the author (henceforth the HH classification). The reason why the HH classification is chosen as the basic reference is that it is (1) complete (includes all languages), (2) conservative, and (3) accompanied by references to literature where the corresponding groupings are argued for. Lewis (2009) is also complete but less conservative and fails as regards the third criterion. Other available classifications are incomplete, dealing only with subsets of Papuan languages, or are not explicit about each individual language. Each of the languages under consideration in this paper is tagged for the HH classification in the metadata contained in the database of Wichmann et al. (2012), where it is given in the first line introducing each word list, after the @ sign. The database is available for full download (see the References for the URL).

The second approach used in this paper is an automated classification of 60% of the Papuan languages based on 40 lexical items using ASJP methodology (to be explained shortly). While Papuan languages have already been subjected to lexically based classifications, the present approach differs from previous work by being more systematic and less biased by areally informed or other intuitions. A systematic, pairwise comparison of 737 word lists involves the inspection of 271,216 pairs of word lists. Such an amount of work is easily achieved by a computer but cannot be carried out manually, which is why lexicostatistic studies have been limited to subsets of Papuan languages. Usually the groups have been selected on a geographical basis, which has introduced an areal bias in the available classifications.

Each approach—the HH classification and ASJP—provides a check on the other, and where they concur in the sense that a HH family is represented by a single cluster in the ASJP tree I assume that the family is valid. When a HH family is scattered over more than one cluster I take this as an indication that the family is possibly problematic and present these cases in order to highlight a potential need for further research. In several cases a family is merely interrupted by one or a few languages that are not supposed to belong to the family in question in HH's scheme. Such cases are highlighted but not commented on further. Finally, in some cases visual inspection of the ASJP tree shows families or isolates (henceforth both will be referred to as 'families') to cluster together under a node, suggesting that the pair of families could be genealogically related (henceforth simply 'related'). For larger groups of languages the possibility of relatedness is evaluated by checking how highly the pair in question is ranked in terms of similarity among the 57,630 pairs of HH families from the entire world that are attested in the ASJP database. For single pairs of languages word lists will in several cases be inspected, and care will be taken to distinguish similarities possibly due to contact, i.e., loanwords, from cognates.

1. Introducing the ASJP tree

The ASJP tree of Papuan languages (henceforth 'the Papuan tree' or simply 'the tree') is found as Appendix 1 to this paper.[1] The language names (in capitals) are the ones assigned for the purpose of the database. They are usually taken over from the sources of the data. Following the names are three-letter ISO 639-3 codes, when available. When a code is not available this is indicated by 'XXX'. The tree is annotated for HH families. The way that the tree was produced is described in the following.

The three basic components in ASJP are: (1) some lexical input; (2) a measure of distances between words which are subsequently averaged across words; and (3) an algorithm for deriving a phylogeny from the distance matrix. There is no particular input, distance measure or phylogenetic algorithm which is hard-wired in the approach, but the following specifications of the components are used here.

(1) The lexical input is lists of words corresponding to a 40-item subset of the 100-item Swadesh list. The 40 items in question were found to be particularly stable and sufficient for optimizing classification results in Holman et al. (2008). All word lists used in the present study are contained in Wichmann et al. (2012), where doculects are uniquely identified by their names and ISO 639-3 codes are also provided (when available) for help with the identification.

(2) The distance measure is the twice-modified Levenshtein distance called LDND (Levenshtein Distance Normalized & Divided). It is based on the Levenshtein distance, a distance metric which counts the minimal number of operations (deletions, insertions, and substitutions) required to transform one word into another. The LDN distance between a pair of words is the Levenshtein distance divided by the length of the longer of the two words. Next, the LDND distance between two languages is defined as the average LDN distance between each pair of words with the same meaning, divided by the average LDN distance between each pair of words with a different meaning. The latter division is intended to control for similarity owing simply to similar phonemic inventories of the two languages (cf. Oswalt 1970 for a related approach).

(3) The algorithm used to turn the resulting distance matrix of doculects into a tree is Neighbor-Joining (Saitou and Nei 1987), which is probably the currently most widely used distance-based phylogenetic algorithm.

Large, cross-linguistic tests of the performance of this set of components (Pompei et al. 2011, Wichmann et al. 2010a, and Huff and Lonsdale 2011) have shown varying performance

[1] In addition to languages normally considered Papuan the tree also includes Kenaboi, an extinct language which has variously been depicted as some kind of Austronesian and Austro-Asiatic mix, as a regular Mon-Khmer language, or as a taboo jargon, according to Hajek (1998). The language is included because it branches with other Papuan languages in the ASJP world tree of Müller et al. (2010), but I do not wish to imply that Kenaboi should be regarded as Papuan. Nevertheless, I also do not wish to exclude the possibility that at least some of its lexical items could have a Papuan origin.

with respect to classification results across language families, from perfect to far from perfect matches with the classifications of Lewis (2009) and Dryer (2005). At least some of the variability in fit with expert classifications must be attributed to variability in the quality of these expert classifications, since ASJP should in principle work equally well everywhere. Evaluating these evaluations is therefore not straightforward.

Other tests of a more fine-grained and qualitative nature have been carried out for some individual families. Hill (2011) compares an ASJP classification of Uto-Aztecan to one exclusively based on shared phonological innovations, and finds only minor differences. Hill's paper was originally presented at The First Conference on ASJP and Language Prehistory, a conference devoted to the evaluation of ASJP classifications for different families. Other papers from this conference have not been fully published but some are available as online working papers (Mailhammer 2010 on Indo-European, Donohue 2010 on Skou, Brown and Holman 2010 on Mayan; cf. also Urban 2009 on Pomoan and Urban 2009 on Iroquoian, not presented at the conference, but similar in nature). More recently, Walker et al. (2011) compare an ASJP classification of the Tupi language family with the literature on Tupi classification, finding that ASJP replicates the overall subgrouping scheme that is standardly assumed; within the large Tupi-Guarani clade subgrouping is more controversial, and the differences between ASJP and the various published proposals are on the same order as the difference among the opinions of experts, but the ASJP scheme is most similar to the two most recent proposals.

A final test is the irregularly updated ASJP World Language Tree of Lexical Similarity, the last published version of which was uploaded as Müller et al. (2010). It shows a clear tendency for younger families to be better replicated in the sense that all languages supposed to belong to a family are gathered under a single node, uninterrupted by unrelated languages. The oldest families (using definitions from Dryer 2005 in this case[2]) that are replicated in this sense include the following, where age estimates in years before present from Holman et al. (2011) are given in parentheses after the family names: Hmong-Mien (4243 BP), Uto-Aztecan (4018 BP), Nakh-Daghestanian (3907 BP), Salishan (3827 BP), Tor-Orya (3693 BP), Northwest Caucasian (3649 BP), Austro-Asiatic (3635 BP), East Bird's Head (3590 BP), Border (3453 BP), Kiowa-Tanoan (3434 BP), Chukotko-Kamchatkan (3368 BP), Tai-Kadai (3252 BP), Uralic (3178 BP), and Barbacoan (3080 BP). The only exception, where a family having an estimated age younger than that of Hmong-Mien is not completely replicated in the world tree is Tupi (3585 BP), which has a single outlying member, Karitiâna [ktn].[3]

---

[2] The dates were produced using the *Ethnologue* classification. It is possible to simultaneously use family definitions from Dryer 2005 and the *Ethnologue* classification because Dryer's families in the cases listed here are isomorphic with either families or subgroups of families in *Ethnologue*. The following are cases where a Dryer family is an *Ethnologue* subgroup: Tor-Orya = the Orya-Tor subgroup of Tor-Kwerba; Northwest Caucasian = the West Caucasian subgroup of North Caucasian; East Bird's Head = the East Bird's Head subgroup of East Bird's Head-Sentani.

[3] In addition, a single Austronesian (Oceanic) language, Kayupulau [kzu], sits in a cluster of Papuan languages next to Austronesian, but in the latest (still unpublished) version of the ASJP world tree Kayupulau has joined Austronesian. It appears that there was earlier some error in the data which has now been corrected. Pauwasi (4102

Thus, for correctly replicating groups of related languages the method shows a high degree of reliability down to the time level of about 4000 BP. For older families the problems increase with time depth. Thus, for instance, Indo-European (4348 BP) constitutes a large, coherent segment, but the isolates Modern Greek and Albanian are attracted by accidental similarities to other regions in the world tree. Sino-Tibetan languages (5261 BP) also generally cluster, except for five languages that are found elsewhere in the world tree. Exceeding a time depth of around 5000 the method is of questionable utility, since families which are that old tend to be split into many different clusters in the world tree, even if the bulk of the languages may still cluster. Only a few of these old families, however, are uncontroversial (Afro-Asiatic, Na-Dene, Otomanguean); others tend to be controversial, at least as regards some of its supposed member groups: Australian, Macro-Ge, Niger-Congo, Nilo-Saharan, Penutian, Trans-New Guinea. Some relations picked up by the ASJP world tree have only been solidly established recently or relatively recently. For instance, Austro-Asiatic remained controversial well into the mid-20th century (Sidwell 2010: 46), and Totozoquean has only very recently received extensive confirmation (Brown et al. 2011). Thus, it is to be expected that for the Papuan languages, which are generally understudied (Hammarström and Nordhoff 2012), there are still relationships to discover within the time range where the method works well. Indeed, several such possible cases will be presented in this paper. When the method fails to replicate a family claimed to exist—such cases will also be presented—, it is no proof against the given family proposal, but cases like that do potentially point to problems with hypotheses of genealogical relationship.

2. Results comparing ASJP tree and HH classification

104 HH families (and isolates) are represented in the Papuan tree. Of these, only 18 fail to cluster under a single node in the tree. All these cases are listed in Table 1, which provides the family names, the minimal number of nodes in the tree under which the languages cluster, i.e. the number of segments that the family is split up into, and some comments. The comments distinguish different types of cases, which are now described.

(a) There are the cases of Angan, Eleman, and Lakes Plain, where all languages do cluster except one or two outliers. Such families can be regarded as supported, and it can be supposed that the outliers either have been misclassified or just are highly divergent members without close relatives within the family.

(b) Border is a family whose failure to cluster completely is due to a supposedly unrelated intruder behaving as a member of the family. I consider this supported, but in this case it should be investigated whether the intruder is really a family member or whether its behavior is due to accidental similarities or loanwords.

BP) is also split up, but it appears that Dryer (2005) actually does not operate with a Pauwasi family. The only language from the family in WALS is Karkar-Yuri, which is assigned to the Karkar-Yuri family.

(c) Then there is one case where the above two circumstances combine: Nuclear Torricelli has one intruder and two outliers. I also consider this family supported, but it should be investigated further in order to verify whether it should be expanded and/or reduced.

(d) Next, there are some cases of small families having or being represented by only 2-5 members which are in two different segments: Biksi, Dibiyaso-Doso-Turumsa, Kwalean, Morehead-Wasur, Pauwasi, and Sentanic. To be cautious I do not consider these to be supported, but since in all cases one of the two segments consists of only 1-2 languages it is possible, for instance, that data circumstances relating to these single languages are responsible for the failure to cluster. For a word list to be considered in the present study it is required that at least 70% among the 40 item be attested, i.e., a minimum of 28 items. For a single pair of languages this means that there can theoretically be as few as 16 words to compare if the number of missing items is maximal for both languages and if all those items are different. Of course this extreme situation rarely occurs, but it is also relatively rare to have full 40-item lists available for both members of a pair. Holman et al. (2008) found classification performance to increase rapidly with the addition of items up to around 40, and Wichmann et al. (2011) found evidence that missing data introduce conflicting phylogenetic signals (i.e., reticulation) into classifications. This means that even a few missing items are expected to diminish the performance substantially.

(e) There are two cases where one family is intertwined in another, raising the question whether the two families should be considered a single entity. These are the cases of East Timor-Bunaq, which (except for one outlier) is embedded into West Timor-Alor-Pantar, and Greater Kwerba, which (again except for one outlier) is mixed with Tor-Orya.

(f) Finally there are cases where a larger putative family is split into two or more different larger segments occurring in separate regions of the larger tree: Sko (2 segments), Lower Sepik-Ramu (5 segments), Nuclear Trans New Guinea (16 segments). These exhibit the sort of behavior of very old and/or controversial families like Afro-Asiatic, Altaic or Australian in the ASJP World Tree.

Table 1. Summary of behavior of HH families with aberrant behavior in the ASJP tree

Family	Nodes	Comments	Type
Angan	2	one outlier	a
Biksi	2	two languages, in different regions of tree	d
Border	2	one intruder	b
Dibiyaso-Doso-Turumsa	2	two languages, in different regions of tree	d
East Timor-Bunaq	2	three languages embedded as a cluster in West Timor-Alor-Pantar and one outlier	e
Eleman	2	one outlier	a
Greater Kwerba	3	three languages interspersed with Tor-Orya and one outlier	e
Kwalean	2	two languages in different regions of the	d

		tree	
Lakes Plain	2	one sister-pair of outlier languages	a
Lower Sepik-Ramu	5	spread over five different regions of the tree	f
Morehead-Wasur	2	three languages in two different regions of the tree	d
Nuclear Torricelli	7	one intruder and two outliers	c
Nuclear Trans New Guinea	16	sixteen different clusters and single languages spread over the entire tree	f
Pauwasi	2	two small clusters, in different regions of the tree	d
Sentanic	2	three languages, two in a cluster, the third elsewhere	d
Sko	2	two clusters in different regions	f
Tor-Orya	2	Interspersed with Greater Kwerba	e
West Timor-Alor-Pantar	3	one single language intruder and one cluster intruding, one outlier	e

In summary, among the 18 families which to a greater or lesser degree show aberrant behavior in the ASJP tree, 5 (cases a-c) can nevertheless be considered supported barring details of some individual languages. I will not be further concerned with these cases in this paper. I will also not be concerned about final decisions with regard to whether smaller (case d) or larger (case f) families that are severely fragmented in the ASJP tree nevertheless do constitute families or whether they should be split, although I assume the latter to be the case for the purposes of this paper. If ASJP fails to support these entities it means that the data immediately available cannot be tweaked into saying something different, and going beyond these data would require entire investigations and papers for each case. The cases where something *can* be done with the ASJP database to shed further light on the Papuan classification are the cases classified in the (e) group, where ASJP apparently delivers false positives. In the following paragraph I will be concerned about how to interpret the mixture of families found in these two cases.

A phylogenetic algorithm such as Neighbor-Joining has the advantage that it takes the entire distance matrix into account when placing the languages relatively to one another, something which cannot be done by hand. But there is also a disadvantage to this and other phylogenetic algorithms. False positives—the branching together of languages that are really unrelated—can occur even if the unrelated languages in question have low similarities, provided that they also have low similarities to all the other languages in the tree. Languages that do not fit into any of the well supported clusters may end up clustering with unrelated languages simply because they do not fit in anywhere else. How can such cases of 'false friends' be distinguished from true relatives? One immediate clue is the length of the (horizontal) branch connecting the node that unites the languages to the remainder of the network which, when short, should cause one to be cautious. But visual inspection of a Neighbor-Joining tree has to be impressionistic and can be inconclusive—there is not some absolute cut-off point with regard to how long a branch should be to be significant. So it is often worthwhile going beyond the tree and directly study the raw distances which the tree is based on—and one can obviously go further to inspect the actual word lists that are the basis for the distances or additional data. Indeed, this last step is recommended, but for a large groups of languages this additional step would constitute an investigation worth a whole separate paper. When judging the East Timor-Bunaq/West Timor-Alor-Pantar and Greater Kwerba/Tor-Orya connections I therefore focus directly on what the distances between the members of each pair of families say.

In and of themselves distances are not very telling, but they become so in a comparative perspective. Judgments on the evidence for respectively East Timor-Bunaq/West Timor-Alor-Pantar and Greater Kwerba/Tor-Orya will therefore be made with reference to distances between all 57,630 pairs of HH families throughout the world. Since HH's classification is so conservative, family pairs that are top-ranking in terms of average similarity between member languages are good candidates for actually being relatives, even if chance similarities could and probably do account for some of these high-ranking pairs. For the purpose of these comparisons Table 2 is offered, which includes the following information in the different columns:

- Family designations according to HH.
- *Ethnologue* family designations (only one per HH family pair if both are in the same *Ethnologue* family, otherwise separated by a slash).
- The number *N* of pairs of lists from each family. If one family is represented by *m* lists and the other by *n* lists then *N* = *m***n*.
- *CN* is a correction of *N* taking into account how similar the lists are within each family. If there are many lists representing very close speech varieties then *N* should be penalized, and this has being taken care of by *CN*, which was suggested to me by Eric Holman (p.c., 2012), whose description of the procedure is reproduced in Appendix 2.
- *SIM*, which is the average similarity expressed as percentages (100% – LDND) for pairs of doculects where the members belong to each family.
- Finally, *NSIM* is a correction of *SIM* that puts greater weight on comparisons involving many and/or divergent doculects than on a small number of comparisons involving fewer and/or more divergent doculects. *NSIM* is found by multiplying *SIM* by the square root of *CN*.
- Numbers representing the rank by *SIM* and *NSIM*.

Table 2 is ordered by the *NSIM* rank, which gives a better idea of plausible genealogical relations than *SIM*. To check which of the two works best as an indicator of relationships, the family pairs were successively ranked by *SIM* and *NSIM* and tagged as being 'possible' or 'impossible', 'possible' being defined as 'not impossible', and 'impossible' being defined in a loose sense as not spoken in the same world area—areas being Eurasia, Africa, New Guinea, Australia, North America, South America—and/or not considered related even by very enthusiastic long-range comparativists. Since no real claims are based on these judgments I will not account for them in more detail. The point of the exercise was simply to see which of the two measures, *SIM* or *NSIM*, turned up the fewest cases of 'impossible' relations along the lists of family pairs ranked for each of the two measures. The 500 top-ranking pairs for each measure were inspected, and in 4 of 5 100-pair bins *SIM* produced more 'impossible' pairs than *NSIM*. For instance, Furan (Nilo-Saharan in *Ethnologue*) and Konda-Yahadian (Trans New Guinea in *Ethnologue*) are ranked as #31 by *SIM*, but #182 by *NSIM*. Each of these two HH families is represented by just one language in the database, allowing for a greater influence of accidental similarities, and *NSIM* efficiently corrects for this. The exercise also showed that the number of 'impossible' pairs continues to grow quickly as ones moves down from the top of the list ranked by *NSIM* roughly until reaching pair #200. Within the #201-#250 bin about one half of the pairs are 'impossible', and the same holds for successive bins within the 500 pairs investigated. Thus, within the c. 200 highest-ranking pairs, but not beyond that segment, *NSIM* should be a potentially valuable indicator of possible genealogical relations. It needs to be stressed, though, that the presence of impossible pairs even among the 200 highest-ranking pairs clearly indicates that chance similarity can be at work. I will not attempt to offer a probability estimate that languages entering into pairs in the top-200 segment really are related. What I am offering is simply a list of the best candidates in the world for being related as far as the ASJP lexical evidence goes. The 200 top-ranking pairs are provided in Table 2.

Table 2. A listing of the 200 HH family pairs ranking highest with respect to *NSIM*

HH family 1	HH family 2	*Ethnologue*	*N*	*CN*	*SIM*	rank	*NSIM*	rank
West Timor-Alor-Pantar	East Timor-Bunaq	Trans-New Guinea	205	11.23	8.72	34	29.22	1
Lepki	Murkim	Both Unclassified	2	1.12	26.64	1	28.19	2
North Omotic	Mao	Afro-Asiatic	72	4.92	11.06	14	24.53	3
Garrwan	Limilngan	Australian	1	1	22.91	2	22.91	4
Amto-Musan	Left May	Amto-Musan / Arai-Kwomtari	16	3.81	11.19	12	21.84	5
Bunaban	Jarrakan	Australian	4	2.19	13.42	6	19.86	6
Eastern Daly	Northern Daly	Australian	6	1.5	16.04	3	19.64	7
Anson Bay	Northern Daly	Australian	6	1.38	15.98	4	18.77	8
Mongolic	Tungusic	Altaic	176	5.5	7.61	65	17.85	9
Central Sudanic	Birri	Nilo-Saharan	45	4.95	7.88	59	17.53	10
Kiwaian	Waia	Trans-New Guinea / South-Central Papuan	28	1.94	12.54	9	17.47	11
Bosavi	Turama-Kikori	Trans-New Guinea	52	5.25	7.44	74	17.05	12
Nyulnyulan	Pama-Nyungan	Australian	218	11.62	4.98	576	16.98	13

Quechuan	Aymara	Quechuan / Aymaran	360	1.77	12.39	10	16.48	14
Panoan	Tacanan	Panoan / Tacanan	115	3.83	8.32	41	16.28	15
Central Sudanic	Kresh-Aja	Nilo-Saharan	90	7.74	5.74	281	15.97	16
Kamula	Awin-Pa	Trans-New Guinea	1	1	15.88	5	15.88	17
Jarrakan	Worrorran	Australian	6	3.33	8.55	36	15.60	18
Mirndi	Pama-Nyungan	Australian	436	18.97	3.53	1994	15.37	19
Iwaidjan Proper	Marrku-Wurrugu	Australian	3	1.97	10.92	15	15.33	20
Gunwinyguan	Pama-Nyungan	Australian	1417	29.82	2.80	3801	15.29	21
Atlantic-Congo	Dogon	Niger-Congo	7876	37.87	2.48	5059	15.26	22
Fasu	East Kutubu	Trans-New Guinea	2	1.44	12.66	8	15.19	23
Southern Daly	Western Daly	Australian	36	2.97	8.69	35	14.98	24
Garrwan	Pama-Nyungan	Australian	109	7.74	5.34	411	14.86	25
Bunaban	Pama-Nyungan	Australian	218	10.64	4.55	806	14.84	26
Jarrakan	Pama-Nyungan	Australian	218	12.34	4.22	1085	14.82	27
Murkim	Biksi	Unclassified / Sepik	4	2.18	9.98	22	14.74	28
Maningrida	Pama-Nyungan	Australian	327	16.95	3.53	1995	14.53	29
Pama-Nyungan	Worrorran	Australian	327	16.17	3.61	1867	14.52	30
Gunwinyguan	Giimbiyu	Australian	39	4.69	6.64	133	14.38	31
Giimbiyu	Iwaidjan Proper	Australian	9	2.4	9.24	28	14.31	32
Bosavi	Dibiyaso-Doso-Turumsa	Trans-New Guinea	26	4.78	6.53	144	14.28	33
Greater Kwerba	Tor-Orya	Tor-Kwerba	25	7.9	5.01	552	14.08	34
Suki-Gogodala	Waia	Trans-New Guinea / South-Central Papuan	14	1.77	10.58	16	14.08	35
Puinave	Kakua-Nukak	Language isolate / Maku	8	1.87	10.24	19	14.00	36
Birri	Kresh-Aja	Nilo-Saharan	2	1.56	11.17	13	13.95	37
Gunwinyguan	Yangmanic	Australian	26	5.45	5.92	240	13.82	38
Bosavi	East Strickland	Trans-New Guinea	91	4.78	6.21	196	13.58	39
Atlantic-Congo	Mande	Niger-Congo	48688	47.87	1.94	7943	13.42	40
Nuclear Trans New Guinea	Pauwasi	Trans-New Guinea / Pauwasi	1350	59.24	1.74	9414	13.39	41
Bosavi	Fasu	Trans-New Guinea	26	3.67	6.99	100	13.39	42
Northern Daly	Western Daly	Australian	18	2.12	9.18	30	13.37	43
Anson Bay	Western Daly	Australian	27	1.92	9.60	24	13.30	44
Chitimacha	Huavean	Gulf / Huavean	3	1.6	10.47	17	13.24	45

South Bird's Head Family	Inanwatan	Trans-New Guinea	7	1.91	9.52	26	13.16	46
Ndu	Sepik	Sepik	126	7.93	4.67	736	13.15	47
Border	Nimboran	Border / Nimboran	35	5.93	5.40	392	13.15	48
Bosavi	Nuclear Trans New Guinea	Trans-New Guinea	2925	45.11	1.95	7864	13.10	49
Aikana	Kwaza	Both Unclassified	1	1	13.06	7	13.06	50
Mailuan	Yareban	Trans-New Guinea	3	1.62	10.17	21	12.94	51
Kresh-Aja	Maban	Nilo-Saharan	8	3.31	7.11	91	12.94	52
Bilua	Savosavo	Central Solomons	2	1.18	11.9	11	12.93	53
Border	Elseng	Border / Language isolate	7	3.25	7.13	89	12.85	54
Maningrida	Iwaidjan Proper	Australian	9	4.31	6.19	197	12.85	55
Marindic	Duna-Bogaya	Trans-New Guinea	22	4.09	6.30	184	12.74	56
Lakes Plain	Sko	Lakes Plain / Sko	364	22.05	2.70	4177	12.68	57
Gunwinyguan	Gaagudju	Australian	13	3.85	6.44	164	12.64	58
East Timor-Bunaq	West Bomberai	Trans-New Guinea	15	5.46	5.35	410	12.50	59
Matacoan	Guaicuruan	Mataco-Guaicuru	40	8.2	4.35	980	12.46	60
Kolopom	Nuclear Trans New Guinea	Trans-New Guinea	900	43.68	1.88	8362	12.43	61
Anson Bay	Pama-Nyungan	Australian	327	8.66	4.22	1086	12.42	62
Inland Gulf	Marindic	Trans-New Guinea	33	4.93	5.59	326	12.41	63
Maningrida	Gunwinyguan	Australian	39	8.42	4.27	1053	12.39	64
Nyulnyulan	Worrorran	Australian	6	3.13	6.85	114	12.12	65
Saliban	Jodi	Salivan / Unclassified	4	2.81	7.20	85	12.07	66
Anson Bay	Eastern Daly	Australian	9	1.36	10.32	18	12.04	67
Nuclear Trans New Guinea	Koiarian	Trans-New Guinea	1350	40.41	1.89	8278	12.01	68
Inland Gulf	Kiwaian	Trans-New Guinea	42	3.34	6.56	137	11.99	69
Nuclear Trans New Guinea	East Timor-Buna	Trans-New Guinea	1125	47.62	1.73	9502	11.94	70
Songhay	Amto-Musan	Nilo-Saharan / Amto-Musan	16	2.17	8.10	50	11.93	71
Nuclear Torricelli	Urim	Torricelli	42	9.86	3.77	1608	11.84	72
Bunaban	Worrorran	Australian	6	2.87	6.93	106	11.74	73
Misumalpan	Chibchan	Misumalpan / Chibchan	60	10.64	3.59	1893	11.71	74
Gaagudju	Pama-Nyungan	Australian	109	7.74	4.19	1110	11.66	75
Songhay	Left May	Nilo-Saharan / Arai-Kwomtari	64	4.16	5.69	294	11.61	76
North Omotic	Nuclear Trans New Guinea	Afro-Asiatic / Trans-New Guinea	5400	51.61	1.61	10482	11.57	77
Pauwasi	Biksi	Pauwasi / Sepik	12	6.5	4.53	824	11.55	78
Mangarrayi-Maran	Pama-Nyungan	Australian	327	17.42	2.76	3938	11.52	79
Pomoan	Cochimi-	Hokan	84	4.79	5.26	443	11.51	80

Gaagudju	Yuman Iwaidjan Proper	Australian	3	1.97	8.19	45	11.50	81
Daju	Nara	Nilo-Saharan	5	1.88	8.38	40	11.49	82
Pahoturi	Eastern Trans-Fly	South-Central Papuan / Eastern Trans-Fly	152	5.33	4.97	581	11.47	83
Mirndi	Yangmanic	Australian	8	3.47	6.15	204	11.46	84
Wagiman	Yangmanic	Australian	2	1.42	9.59	25	11.43	85
Kiwaian	Nuclear Trans New Guinea	Trans New-Guinea	3150	31.11	2.04	7332	11.38	86
Bunaban	Umbugarla-Ngurm	Australian	2	1.37	9.71	23	11.37	87
South Omotic	Ongota	Afro-Asiatic	6	2.63	6.97	102	11.30	88
Pauwasi	Koiarian	Pauwasi / Trans-New Guinea	36	7.57	4.08	1231	11.23	89
Pama-Nyungan	Umbugarla-Ngurmbur	Australian	109	7.74	4.03	1283	11.21	90
Bunaban	Mirndi	Australian	8	3.37	6.10	211	11.20	91
Nara	Nubian	Nilo-Saharan	5	2.98	6.45	162	11.13	92
West Timor-Alor-Pantar	West Bomberai	Trans-New Guinea	123	8.56	3.80	1568	11.12	93
Katla-Tima	Narrow Talodi	Niger-Congo	10	2.86	6.56	138	11.09	94
Pama-Nyungan	Yangmanic	Australian	218	10.96	3.32	2425	10.99	95
Biksi	Sko	Sepik / Sko	28	7.66	3.97	1357	10.99	96
Left May	Sko	Arai-Kwomtari / Sko	112	10.61	3.36	2332	10.94	97
Atlantic-Congo	Afro-Asiatic	Niger-Congo / Afro-Asiatic	201912	201.89	0.77	20501	10.94	98
Totonacan	Mixe-Zoque	Totonacan / Mixe-Zoque	130	2.77	6.56	139	10.92	99
Dibiyaso-Doso-Turumsa	East Strickland	Trans-New Guinea	14	3.55	5.75	275	10.83	100
Southern Daly	Pama-Nyungan	Australian	436	13.37	2.95	3337	10.79	101
Bosavi	Pauwasi	Trans-New Guinea / Pauwasi	78	8.45	3.69	1725	10.73	102
Lepki	Pauwasi	Unclassified / Pauwasi	6	3.33	5.87	251	10.71	103
Afro-Asiatic	North Omotic	Afro-Asiatic	6768	41.43	1.66	10072	10.68	104
Giimbiyu	Pama-Nyungan	Australian	327	9.44	3.45	2154	10.60	105
Suki-Gogodala	Nuclear Trans New Guinea	Trans-New Guinea	1575	28.35	1.99	7638	10.60	106
Nuclear Trans New Guinea	Goilalan	Trans-New Guinea	450	29.99	1.93	8005	10.57	107
Inland Gulf	Kolopom	Trans-New Guinea	12	4.69	4.86	634	10.53	108
Pauwasi	Namla-Tofanma	Pauwasi / Unclassified	6	3.33	5.76	272	10.51	109
Dravidian	Sepik	Dravidian / Sepik	476	15.37	2.68	4244	10.51	110

Miwok-Costanoan	Maiduan	Penutian	36	4.23	5.09	509	10.47	111
Maningrida	Nyulnyulan	Australian	6	3.28	5.78	267	10.47	112
Nara	Surmic	Nilo-Saharan	9	2.74	6.30	185	10.43	113
Kartvelian	Arawakan	Kartvelian / Arawakan	212	16.33	2.58	4652	10.43	114
Pama-Nyungan	Minkin-Tangkic	Australian	218	10.71	3.18	2750	10.41	115
Larrakiyan	Limilngan	Australian	2	1.55	8.31	42	10.35	116
Kwalean	Waia	South-Central Papuan	12	2.63	6.36	178	10.31	117
Murkim	Pauwasi	Unclassified / Pauwasi	12	3.73	5.33	416	10.29	118
Greater Kwerba	Savosavo	Tor-Kwerba / Central Solomons	5	2.99	5.95	234	10.29	119
Iwaidjan Proper	Pama-Nyungan	Australian	327	15.27	2.62	4486	10.24	120
Awin-Pa	East Strickland	Trans-New Guinea	7	1.88	7.46	73	10.23	121
Mpur	West Bird’s Head	Language isolate / West Papuan	7	2.43	6.56	140	10.23	122
Mangarrayi-Maran	Worrorran	Australian	9	4.69	4.72	706	10.22	123
Walio	Sepik	Sepik	42	12.16	2.93	3398	10.22	124
Limilngan	Northern Daly	Australian	2	1.23	9.21	29	10.21	125
Sino-Tibetan	Pama-Nyungan	Sino-Tibetan / Australian	22890	72.45	1.20	14711	10.21	126
Nuclear Trans New Guinea	Sko	Trans-New Guinea / Sko	3150	69.87	1.22	14471	10.20	127
South Bougainville	Suki-Gogodala	South Bougainville / Trans-New Guinea	21	3.46	5.48	364	10.19	128
Dem	Yareban	Trans-New Guinea	1	1	10.18	20	10.18	129
Gunwinyguan	Southern Daly	Australian	52	6.64	3.95	1376	10.18	130
Sentanic	Left May	East Bird's Head-Sentani / Arai-Kwomtari	72	7.38	3.74	1649	10.16	131
Hmong-Mien	Austroasiatic	Hmong-Mien / Austro-Asiatic	4305	34.33	1.73	9503	10.14	132
Maiduan	Yokutsan	Penutian	8	1.95	7.25	83	10.12	133
Kujarge	Afro-Asiatic	Unclassified / Afro-Asiatic	282	14.27	2.67	4275	10.09	134
Maiduan	Wintuan	Penutian	4	1.63	7.89	57	10.07	135
Kosare	Sko	Kaure / Sko	14	3.93	5.07	519	10.05	136
Dravidian	Savosavo	Dravidian / Central Solomons	34	3.28	5.54	336	10.03	137
Jarrakan	Mirndi	Australian	8	3.9	5.08	515	10.03	138
Southern Daly	Northern Daly	Australian	8	2.13	6.87	110	10.03	139
Jarrakan	Nyulnyulan	Australian	4	2.39	6.46	160	9.99	140
Taiap	Nuclear Torricelli	Language isolate / Torricelli	42	9.86	3.18	2751	9.99	141
Garrwan	Nyulnyulan	Australian	2	1.50	8.14	49	9.97	142
Dibiyaso-	Fasu	Trans-New Guinea	4	2.72	6.02	230	9.93	143

Doso-Turumsa								
Worrorran	Yangmanic	Australian	6	2.96	5.77	270	9.93	144
Pauwasi	Sepik	Pauwasi / Sepik	84	15.6	2.51	4949	9.91	145
Dibiyaso-Doso-Turumsa	Nuclear Trans New Guinea	Trans-New Guinea	450	33.48	1.71	9664	9.89	146
Kayagar	Klamath-Modoc	Trans-New Guinea / Penutian	3	1.51	8.03	52	9.87	147
Nuclear Trans New Guinea	Duna-Bogaya	Trans-New Guinea	450	28.14	1.86	8509	9.87	148
Hmong-Mien	Mailuan	Hmong-Mien / Trans-New Guinea	105	10.45	3.02	3162	9.76	149
Waia	Pomoan	South-Central Papuan / Hokan	14	1.87	7.13	90	9.75	150
Manubaran	Misumalpan	Trans-New Guinea / Misumalpan	18	2.99	5.61	322	9.70	151
Bosavi	Arawakan	Trans-New Guinea / Arawakan	689	19.14	2.20	6385	9.62	152
Molala	Sahaptian	Penutian	2	1.69	7.40	77	9.62	153
Kapauri	Nimboran	Kaure / Nimboran	5	1.83	7.11	92	9.62	154
Left May	Busa	Arai-Kwomtari / Language isolate	8	2.70	5.83	256	9.58	155
North Omotic	South Omotic	Afro-Asiatic	72	5.99	3.91	1420	9.57	156
Kolopom	Mombum	Trans-New Guinea	12	3.66	5.00	557	9.57	157
Austronesian	Touo	Austronesian / Central Solomons	1129	5.67	3.98	1337	9.48	158
Manubaran	Yareban	Trans-New Guinea	6	1.38	8.05	51	9.46	159
Kamula	Bosavi	Trans-New Guinea	13	2.54	5.93	237	9.45	160
Gaagudju	Northern Daly	Australian	2	1.23	8.52	37	9.45	161
Marrku-Wurrugu	Northern Daly	Australian	2	1.23	8.52	38	9.45	162
Kadugli-Krongo	Birri	Nilo-Saharan	11	1.88	6.88	109	9.43	163
Pomoan	Bororoan	Hokan / Macro-Ge	14	2.41	6.06	220	9.41	164
Border	Barbacoan	Border / Barbacoan	35	8.01	3.31	2452	9.37	165
Minkin-Tangkic	Worrorran	Australian	6	2.89	5.51	350	9.37	166
Savosavo	Touo	Central Solomons	1	1.00	9.36	27	9.36	167
Greater Kwerba	Mawes	Tor-Kwerba	10	3.53	4.98	577	9.36	168
Kolopom	Koiarian	Trans-New Guinea	24	5.58	3.95	1377	9.33	169
Greater Kwerba	Eastern Trans-Fly	Tor-Kwerba / Eastern Trans-Fly	190	8.17	3.26	2574	9.32	170
Greater Kwerba	Nimboran	Tor-Kwerba / Nimboran	25	5.47	3.98	1338	9.31	171
Kaure-Narau	West Timor-Alor-Pantar	Kaure / Trans-New Guinea	41	4.20	4.53	825	9.28	172
Kosare	Nuclear Trans New Guinea	Kaure / Trans-New Guinea	225	17.78	2.20	6386	9.28	173
Indo-European	Uto-Aztecan	Indo-European / Uto-Aztecan	16434	20.61	2.04	7333	9.26	174
Gunwinyguan	Iwaidjan	Australian	39	7.59	3.36	2333	9.26	175

	Proper							
Inland Gulf	Mombum	Trans-New Guinea	9	2.84	5.49	360	9.25	176
Waia	Sko	South-Central Papuan / Sko	28	4.36	4.42	917	9.23	177
Maybrat	West Bird's Head	Maybrat / West Papuan	7	2.43	5.92	241	9.23	178
Bilua	Nuclear Trans New Guinea	Trans-New Guinea / Central Solomons	450	20.91	2.01	7515	9.19	179
Songhay	Fasu	Nilo-Saharan / Trans-New Guinea	16	2.23	6.15	205	9.18	180
Great Andamanese	Jarawa-Onge	Andamanese	16	3.73	4.75	688	9.17	181
Furan	Konda-Yahadian	Nilo-Saharan / Trans-New Guinea	1	1.00	9.16	31	9.16	182
Nuclear Trans New Guinea	Sepik	Trans-New Guinea / Sepik	3150	83.23	1.00	17190	9.12	183
Ndu	West Timor-Alor-Pantar	Sepik / Trans-New Guinea	369	7.10	3.42	2214	9.11	184
Pyu	Atakapa	Arai-Kwomtari / Gulf	1	1.00	9.10	32	9.10	185
Siuslaw	Barbacoan	Penutian / Barbacoan	5	2.47	5.78	268	9.08	186
Austronesian	Pama-Nyungan	Austronesian / Australian	12306 1	43.92	1.37	12793	9.08	187
Limilngan	Southern Daly	Australian	4	1.73	6.90	108	9.08	188
Cariban	Bororoan	Carib / Macro-Ge	56	4.78	4.12	1188	9.01	189
Awin-Pa	Bosavi	Trans-New Guinea	13	2.54	5.65	314	9.00	190
Furan	West Timor-Alor-Pantar	Nilo-Saharan / Trans-New Guinea	41	4.20	4.39	946	9.00	191
Limilngan	Umbugarla-Ngurmbur	Australian	1	1.00	8.99	33	8.99	192
Kolopom	Moraori	Trans-New Guinea	4	2.46	5.73	282	8.99	193
Nuclear Trans New Guinea	West Bomberai	Trans-New Guinea	675	36.28	1.49	11570	8.97	194
Lepki	Biksi	Unclassified / Biksi	2	1.95	6.42	167	8.97	195
Mor	Matacoan	Trans-New Guinean / Mataco-Guaicuru	8	2.79	5.36	406	8.95	196
Nuclear Trans New Guinea	Angan	Trans-New Guinea	3150	66.10	1.10	15926	8.94	197
Heiban	Nubian	Niger-Congo / Nilo-Saharan	55	7.55	3.25	2595	8.93	198
Miwok-Costanoan	Yokutsan	Penutian	18	3.10	5.07	520	8.93	199
Maningrida	Kungarakany	Australian	3	2.19	6.03	228	8.92	200

Armed with Table 2 we can better approach the cases of East Timor-Bunaq/West Timor-Alor-Pantar and Greater Kwerba/Tor-Orya.The first of these is the highest-ranking of all HH family pairs in the world in terms of *NSIM*, and in terms of raw similarities (*SIM*) it ranks as #34. Thus, there is strong support for these HH families as a single genealogical unit, as the Papuan tree also suggests. Greater Kwerba/Tor-Orya rank as #34 in terms of *NSIM* and #552 in terms of

SIM. This makes the pair a very good candidate for also constituting a single genealogical unit, as suggested by the tree, and as also suggested by the *Ethnologue* classification.

The reader will no doubt have been struck by the large number of HH families whose relatedness according to *Ethnologue* seem to receive support from Table 2, but there are also many high-ranking cases that suggest a different picture than *Ethnologue*, even towards the top end of the ranked list, and, importantly, one should not forget the fact that the HH families themselves should be supported before we can trust any wider relations among them. In particular, there are many cases where different HH families considered as belonging to *Ethnologue*'s Trans-New Guinea indeed seem to be related. But 11 of the pairs involve HH's Nuclear Trans New Guinea, which is in itself not supported. When the Nuclear Trans New Guinea hub is taken out, most of the network—or 'mesh' in the sense of Swadesh (1954)—falls apart. In the next section we return to Table 2 and what it may suggest about deeper Papuan connections.

The closer look at East Timor-Bunaq/West Timor-Alor-Pantar and Greater Kwerba/Tor-Orya concludes the first part of this paper, which was intended to test the HH classification in order to arrive at a conservative set of lexically solid genealogical units as a basic classification which can be used as a framework for going back in the opposite direction to find some more distant relations using ASJP. The HH classification, represented by 104 families in the database, was largely supported (ignoring here some details of the affiliations of single languages), but 5 families with few representatives (Biksi, Dibiyaso-Doso-Turumsa, Kwalean, Morehead-Wasur, Pauwasi, and Sentanic) should possibly each be split in two, and the larger families Sko, Lower Sepik-Ramu, and Nuclear Trans New Guinea should possibly be split into respectively 2, maximally 5, and maximally 16 segments respectively. For the moment this leaves us with a classification which is even more conservative than the HH one. The merging of some of these units, however, seems to be supported, albeit not necessarily in ways envisaged by scholars who have contributed to the scheme represented by *Ethnologue*. This is the topic of the next section.

3. Possible relations among HH families

In this section I will consider cases where ASJP has constructive contributions to make to the classification of Papuan languages in the sense that it suggests relatedness among groups considered unrelated in the HH classification. For a genealogical link among HH families to be considered sufficiently interesting I will require support both from the Papuan tree and from the similarity scores in Table 2, and in cases where only a few languages are involved I will also inspect the actual word lists for likely cognates. This section is organized by Table 2, going top down but excluding the investigation of any further relations of the 8 HH families that were considered to not be supported. The task of investigating whether their (ex-)members are related to other families is postponed to future work.

When only a few languages are involved it is a simple matter to inspect the word lists for possible cognates. When doing so, I cite words directly from the ASJP database in the phonemically reduced transcriptions. These word lists come with no warranty. They are from sources that vary in quality and they are produced by different transcribers and rarely rechecked by experts. So anyone interested in pursuing work on phonological correspondences should refer to the original sources.[4] The transcription system (ASJPcode) is originally described in Brown et al. (2008) and may also be consulted in Wichmann et al (2010b) (an open access publication). Here I will just mention the function of the most non-obvious symbols, * and ~. The former represents nasalization and the latter indicates that the symbols preceding it are to be regarded as a unit (according to the transcriber who produced a given list, who may not always have had sufficient evidence for distinguishing between unit phonemes and sequences).

3.1. Lepki/Murkim

Fig. 1. Locations of Lepki (red) Murkim (blue)

This pair of languages ranks first in the world in terms of *SIM* and second in terms of *NSIM*. The languages, not surprisingly, are also sisters in the Papuan tree. Since there are just two languages we will inspect the word lists. 20 comparisons (bold-faced) out of 33 have the appearance of cognates. A few are identical across the three doculects. If it were not for the fact that there are very many similar words we might be suspicious of identical forms as representing loan words. But in this case they simply look like evidence for a close relationship.

Table 3. Lepki/Murkim lexical comparison

Meaning	LEPKI [lpe]	MILKI MURKIM [rmh]	MOT MURKIM [rmh]
one	kutuowap	hel	hel
two	**kaisi**	**kais**	**kais**
person	**ra**	**ra**	**pra**
fish	**yakEn**	**kan**	**kan**
dog	nan	sai	sai
louse	**nim, nimdEl**	**om**	**im**

[4] See http://lingweb.eva.mpg.de/asjp/index.php/ASJP, where sources are listed by the *Ethnologue* names of each language..

tree	**ya**	**yamul**	**yamul**
leaf	**nabai**	**bw~aik**	**bw~aik**
skin	**yit**	**yaith~**	**yaith~**
blood	yiri	mal	mal
bone	**kow, yiow**	**kok**	**kok**
ear	**bw~i**	**bw~i**	**bw~i**
eye	**yEmon**	**amol**	**amol**
nose	**mogw~an**	**mo*a**	**mw~a**
tooth	**kal**	**kal**	**kal**
tongue	**braw**	**prouk**	**porouk**
knee	kolbw~i	balka	balka
breast	**nom**	**mom**	**mom**
liver	b3oak	miEm	miEm
drink	yis	ksewo	kel5ilo
hear	**ofao**	**pao**	**ha**
die	di	knewo	ko
come	guyo	haro	kw~i
sun	mom	kaya7kalo	kayakalo
star	Endi	ili	ile
water	**kEl**	**kel**	**kel**
stone	saup	on	o*n
fire	**yaoala**	**yo**	**yo**
path	**masin**	**msan**	**mesain**
night	tioa**, tiTa**	**disla**	**tisla**
new	**nowal**	**brel**	**prel**
name	**gy~e**	ibe	**ka**

3.2. Amto-Musan/Left May/Busa

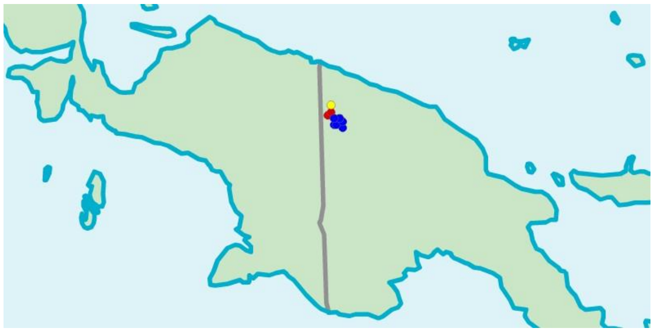


Fig. 2. Locations of Amto-Musan (red), Left May (blue), and Busa (yellow)

Amto-Musan and Left May cluster in the tree and appear in Table 2 (rank *NSIM*: #5; rank *SIM*: #12). The tree suggests a further connection to Busa and four dialects of Demta [dmy], which are supposed to belong to the problematic Sentanic group. Busa also appears with Left May in Table 2 (rank *NSIM*: #155; rank *SIM*: #256), whereas the Busa/Amto-Musan pair has a *NSIM* rank of #5789 and *SIM* rank of #3150. Thus there is indirect, chained evidence tying Busa to Amto-

Musan via Left May. The direct evidence for Busa/Amto-Musan is less clear, but the #5789 rank is still just below the top 10% of pairs of HH families in the world in terms of *NSIM*. This sort of ranking does not go strongly against a relationship although it does not strongly support it either. Thus, I tentatively regard Busa as an outlier connected with the better supported Amto-Musan/Left May family. Further connections to Demta will not be considered here since they would first require a detailed look at the evidence for Sentanic.

3.3. Kamula/Awin-Pa/Bosavi/East Strickland

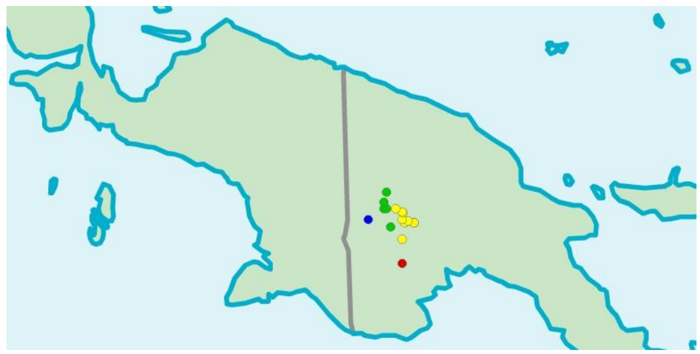


Fig. 3. Locations of Kamula (red), Awin-Pa (blue), Bosavi (yellow), and East Strickland (green)

These four HH families cluster in the tree, and 5 out of the 6 pairs among them figure in Table 2. Kamula/East Strickland is the one pair that does not rank high for *NSIM*, having a rank of #3080, and a *SIM* rank of #2414. Nevertheless, because all other pairs have high similarity ranks and because an *NSIM* rank of #3080 after all lies well within the highest 10% in the world, this looks like a strong cluster. It is interrupted by Dibiyaso [dby], whose supposed relative Doso [dol] sits elsewhere in the tree. The issue of the splintered Dibiyaso-Doso-Turumsa HH family is not considered here.

Since Kamula and Awin-Pa each consists of a single language (respectively Kamula and Pare) it is easy to inspect the word lists. This is done in Table 4. Kamula and Pare show so many similarities (boldfaced) that it would seem immediately viable to establish their relatedness with more extensive work. Inspection of some Bosavi and East Strickland word lists show a few promising possible cognates with Kamula, with Pare or with one another. To pursue the possibility of the relatedness of the entire group it would clearly be necessary to first reconstruct ancestral languages for respectively Bosavi and East Strickland and then compare proto-Kamula-AwinPa, proto-Bosavi and proto-East Strickland to each other in a pairwise fashion.

Table 4. Kamula/Awin-Pa lexical comparisons

meaning	KAMULA [xla]	PARE [ppt]
I	**nE***	**no***
you	wE*	go*

we (incl.)	diE	nigi
one	**hoto**lop	**ote**so
two	**dep**iomEtE	**di**yabo
person	**opo**loimi	**kobo**
fish	omolo	mune
dog	esemolo	Ti
louse	**iyo**	**o**
tree	toli	i*
leaf	upo	use
skin	kopolo	sia
blood	umoli	sowo
bone	ELu	ko
ear	**molo**	**mogo**
eye	**inoma**	**kinemo**
nose	mu*	kine
tooth	Epe	male
tongue	**tE**	**tE**
knee	**oLuma**	**oumu**
hand	**to**	**atowe** ‘arm’
breast	mEmE	bu
liver		
drink		
see	ele	ded
hear	**tolo**	wo**dala**
die		
come	pu	hadan
sun	soLi	gine
star	**tome**Li	pe**teme**
water	yu	ume
stone	**ewo**Lo	i**ebo**
fire	**de**Lopo	**nE**
path	opi	otigi
mountain	tomoLi	giso
night	utoLElo	hwiga
full		towate
new	omoko	kw~ane
name	**hi**	**hi**

Kamula and Awin-Pa lexical comparisons are also provided in Reesink (1976:16) and the Bosavi-East Strickland connection is suggested in Shaw (1986) based on cognate counts. The latter author also considers it “very reasonable” (p. 56) to connect Awin-Pa to Bosavi-East Strickland, but nevertheless does not follow through with this suggestion. (Apparently following McElhanon and Voorhoeve 1970, *Ethnologue* considers all four of the HH families compared here to belong to Trans-New Guinea).

3.4. Fasu-East Kubutu

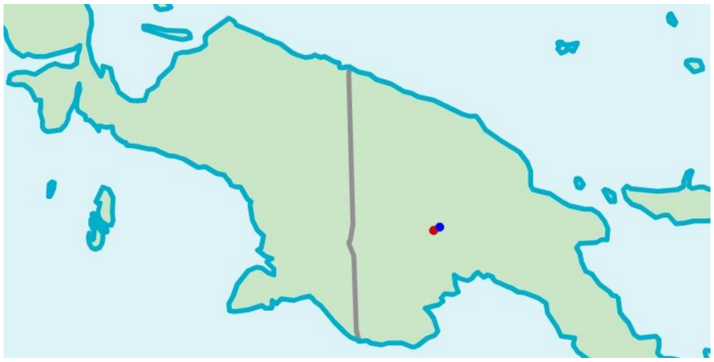


Fig. 4. Locations of Fasu (red) and East Kutubu (blue)

This pair of families is represented by just two languages by *Ethnologue*'s count, Foe [foi] (East Kutubu) and Fasu & Namumi [faa] (Fasu). The pair is a cluster in the tree and is supported by Table 2. Given that just three doculects are involved we can easily inspect the word lists for possible cognates. There are a total of 17 Foe words which are similar to forms in either Fasu or Namumi, with 15 Foe-Fasu matches out of 62 comparisons and 12 Foe-Namumi matches, out of 60 comparisons.[5] These are highlighted in bold in Table 5. Borrowing cannot be excluded, but at least for recent borrowings we would expect forms to overall be more similar, and we would also not expect as many as 20-24% borrowings on these short lists of basic vocabulary. So in my opinion, there is little doubt that Fasu and East Kutubu are related.[6] Franklin (1973b) also assumes that they are related, but the only evidence given is a list of 10 compared words—a list which is intended not to show that these two languages in particular are related, but that there is a large group of languages in the area which are all related to one another. Better evidence is given

[5] In counting the number of comparisons I regard each form as being involved in a separate comparison even if it is clearly a phonological variant, e.g., Foe *iya* and *yiya* 'we (incl.)', so the numbers are intended to err on the conservative side.

[6] One of the referees of this paper is more skeptical, arguing that Fiwaga, another East Kutubu language, which is not included in the ASJP database, lacks many of the matches with Fasu exhibited by Foe, something which could be construed as an argument that the Foe matches not present in Fiwaga are borrowings. Another possibility is to interpret this as meaning that Fiwaga is less lexically conservative. Finally, it may also be the case that some Foe matches are borrowings while others, the best candidates being the ones also shared with Fiwaga, are inherited. Indeed, this last scenario is probably the most likely. Franklin and Voorhoeve (1973:154) show the relevant cognate percentages. On a 231 item list Fasu has 18% cognacy with Foe and 10% with Fiwaga. I do not see the difference between 38 and 23 shared items as a cause for any special interpretation. Moreover, the cognate percentages between Foe and Fiwaga and the two other Fasu languages, Some and Namumi (Some is not in the ASJP database) are on the same order as Foe-Fasu: Foe-Some 18%, Foe-Namumi: 16%, Fiwaga-Some: 15%, Fiwaga-Namumi: 15%. In other words, the percentages for all six pairs of East Kutubu-Fasu pairs range between 10% and 18%, with Foe-Fiwaga being the single outlier within this range. A borrowing scenario needed to explain all these lexical similarities would have to assume that much of the borrowing took place already at an early stage between proto-Fasu and proto-East Kutubu. But since so much basic vocabulary is involved (the percentages are similar for the Swadesh list and the full 231 item list in the matrices of Franklin and Voorhoeve 1973:154), and since Franklin (2001) additionally provides grammatical evidence, I doubt that this is a viable explanation.

by Franklin (2001) in the form of shared kinship and counting terms and some regular sound correspondences and grammatical markers (but again the emphasis is on wider relations).

Table 5. Fasu/East Kutubu lexical comparison

meaning	FASU [faa]	NAMUMI [faa]	FOE [foi]
I	**ano**	**anuni**	**nano**
you	re, **ne**	**ni**	ha7a, **na7a**
we (incl.)	isu	su	iya, yiya
one	hakasa, **meno**	nakasa	**mana*xa**
two	teta	tita	ha*xa
person		**abano**	**amena**
fish	pu, pokoa	poka	zagi
dog	**kasa**	**kasa**	**gesa***, xaso
tree	**ira**	**ira**	**iro**
leaf	**ira** ku*	gu	**iroso*i**, sa*e
skin	rorofa	**kau**	**kh~a7o**
blood	yapi	kakusa	w3lia, hamage
bone	**kiki**	**kiki**	**kh~igi, kh~ikh~i**
ear	senaki	sinEki	ho xh~iyo, kh~ia
eye	**hi***	**hi***	**i*, i*y**
tooth	mere	akai	gi, ti
tongue	**aru**	**airu**	**aru, auru**
knee	**kakuna**	**kukunai**	**ga7anua**, xixi
hand	hokono	nokanu ('arm')	ya
breast	**hoko**	hotu	**o*xo*, o7o**
liver	**kasoko**		**kh~asia7o**, ku*7u*nu*
drink	**nena**		**ni**, No
see	**asera**	**asia**	ariy3y, sebe, **ere**
hear	kaira	kai a	nisi, nisibuba7ai
come	pera	piE	w3y, wa
sun	ma* i*ya*	maya	iriyabo, iriyapo
star	ti*makata	putini iya putini, (iya means 'rain')	irinibu, orowa*pa
water	he*	hi*	ibu, ipu
stone	eke	iki	kh~a*no, kh~ana
fire		**iro**kupi	**ira, iro**
mountain	akai	uri	duma, tuma
night	ereamo	idi iya idi, iya	genemo

full	komarususua	komurusai	kh~ona, kh~onoba7ai
new	kawe	kawi	isa, isa*
name	ano	iyanu	yapo

3.5. Suki-Gogodala/Waia/Kiwaian

Fig. 5. Locations of Suki-Gogodala (red), Waia (blue), and Kiwaian (yellow)

This group of three HH families is a single cluster in the tree. The ranks for each of the three pairs are as follows: Suki-Gogodala/Waia: #35 (*NSIM*) and #16 (*SIM*); Suki-Gogodala/Kiwaian: #3402 (*NSIM*) and #4404 (*SIM*); Waia/Kiwaian: #11 (*NSIM*) and #9 (*SIM*). The ASJP support for the relatedness of each of the pairs Suki-Gogodala/Waia and Waia/Kiwaian is strong and Suki-Gogodala/Kiwaian, although not very highly ranking, is still towards the top of pairs in the world. Thus, I hypothesize that all three HH families are related.

According to Franklin (1973a:17) "Waia shows generally a 10-12% lexical relationship with languages of the Kiwaian family, but over 15% with Gogodara." This proposal for a link between Suki-Gogodala, Waia, and Kiwaian is discussed in more detail by Reesink (1976:22-25). He presents a list of 39 probable cognate sets involving Waia and some Kiwaian languages, but, mainly based on dissimilarities in Waia and Kiwaian pronouns, expresses skepticism about the relationship. He also shows lexical similarities between Waia and Gogodara, but—without any specific arguments—assumes that they are borrowings. Also rejecting an earlier proposal by Wurm (1975:325) that Waia belongs with the Pahoturi languages, he concludes that "[a] genuine genetic relationship could not be found for Waia" (Reesink 1976:26). Whether the similarities between Waia and Suki-Gogodala on the one hand and Waia and Kiwaian on the other are ultimately due to borrowing or inheritance, they are in obvious need of further investigation. It would seem somewhat odd for Waia to borrow basic vocabulary from two different sources, and I expect that the three groups can be shown to be related once reconstructed proto-Suki-Gogodala and reconstructed proto-Kiwaian are drawn upon for comparisons. Unlike Reesink, I would certainly not regard differences between pronouns as evidence against a relationship if other basic vocabulary with similar degrees of stability as pronouns (Holman et al. 2008) supports a genealogical relationship.

3.6. South Bird's Head Family/ Inanwatan

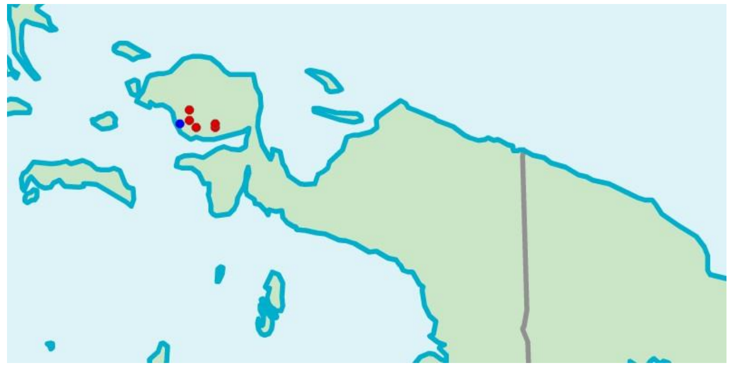


Fig. 6. Locations of South Bird's Head Family (red) and Inanwatan (blue)

This pair ranks #46 (*NSIM*) and #26 (*SIM*) and the languages are sisters in the tree. The tree suggests that Konda-Yahadian is a more distant outlier. The ranks for pairs involving Konda-Yahadian (which is represented by a single doculect) are: Konda-Yahadian/Inanwatan: #627 (*NSIM*), #105 (*SIM*); Konda-Yahadian-South Bird's Head Family: #510 (*NSIM*), #444 (*SIM*). Thus, the further connection to Konda-Yahadian is far from as well supported as the South Bird's Head Family/Inanwatan connections, but it would be worthwhile investigating further.

Following Voorhoeve (1975a), Berry and Berry (1987a) treat South Bird's Head Family, Inanwatan and Konda-Yahadian as three 'families' within the South Bird's Head 'stock'. Cognate counts show the same relations between the three groups as the ASJP Papuan tree, with Konda-Yahadian as a remote relative of the South Bird's Head Family-Inanwatan sisters. The authors also note a number of structural similarities, where the most striking is a pair of nominal gender suffixes that are identical in Inanwatan and at least one South Bird's Head Family language. The structural similarities also involving Konda-Yahadian are more run-of-the-mill.

3.7. Sepik/Ndu/Walio

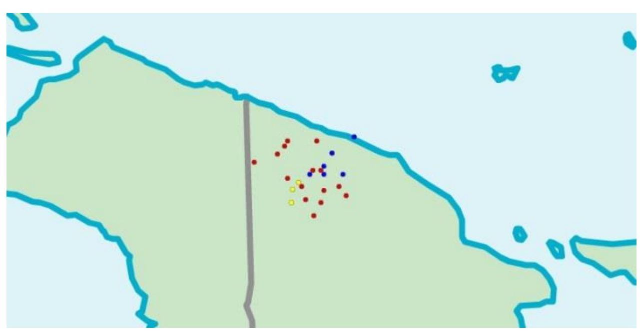


Fig. 7. Locations of Sepik (red), Ndu (blue), and Walio (yellow)

The pairs in these HH families, which form a single cluster in the tree, rank as follows. Sepik/Ndu: #47 (*NSIM*), #736 (*SIM*); Sepik/Walio: #124 (*NSIM*), #3398 (*SIM*); Ndu/Walio: #1072 (*NSIM*), #3503 (*SIM*). While the last pair is not among the top 200 in the world it is quite highly ranking. Thus, this cluster has support.

The relatedness of Sepik and Ndu is substantiated by Foley (2005:126-138), who adduces evidence from pronominals (where 6 out of 10 proto-Ndu pronominals look very similar to Sepik pronominals); basic vocabulary; and some grammatical patterns, where the strongest piece of evidence is an applicative construction involving a grammaticalized form of a proto-Sepik verb *kwV 'to give'. Foley (2005:130) raises the possibility that Kwoma [kmo], otherwise regarded as Sepik, groups with Ndu, but Aikhenvald (2008:597-605) shows that Kwoma has borrowed from the Ndu language Manambu. Her discussion, however, does not affect the larger argument by Foley of Sepik-Ndu relatedness, only the placement of Kwoma within Sepik-Ndu. Laycock and Z'Graggen (1975:753) included Walio in their Sepik-Ramu Phylum along with many other families in a big lumping attempt, but do not present substantial data in support of this possibility.

3.8. Nimboran/Kapauri/Border(/Elseng)

Fig. 8. Locations of Nimboran (red), Kapauri (blue), Border (yellow), and Elseng (Green)

These four HH families belong to the same cluster in the tree, a cluster which also involves Saberi, supposed to be a Greater Kwerba language. I will ignore the status of Saberi since it would involve a closer look at its relation to Greater Kwerba to determine whether it really belongs with that family. Instead I concentrate on the four HH families appearing in the title of this subsection. The ranks for each entailed pair are as follows. Nimboran/Kapauri: #154 (*NSIM*), #92 (*SIM*); Nimboran/Border: #48 (*NSIM*), #392 (*SIM*); Kapauri/Border: #1512 (*NSIM*), #2965 (*SIM*); Nimboran/Elseng: #4689 (*NSIM*), #3538 (*SIM*); Kapauri/Elseng: #14814 (*NSIM*), #7627 (*SIM*); Border/Elseng: #54 (*NSIM*), #89 (*SIM*). These numbers show strong support for Nimboran/Border, with Kapauri mainly being supported as a member of the cluster through its relationship with Nimboran, although Kapauri/Border is still towards the top of the world list. Elseng is potentially a spurious member given that it only scores among the top 200 world pairs for its relationship with Border. Two facts suggest that the relation is one of diffusion involving just these two languages. First, Sawa (the representative of Elseng) intrudes into Border in the tree, having Awji as a sister. Secondly, Sawa and Awji are also direct neighbors geographically. To check this possibility, the actual data are listed in Table 6. Surprisingly, it turns out that none of the word pairs looks like borrowing has been involved. Frankly there are also not any obvious cognates. Nevertheless, there are similarities throughout the list, including 10 cases of identical initial ASJPcode symbols and 3 cases where both the initial consonant and following vowel symbols are identical (marked by underscore). Thus I regard the relatedness of Nimboran, Kapauri, and Border as a sound hypothesis, whereas Elseng's membership in this group is possible but much more dubious. Its wider relation to Border needs further investigation. In the mean time I will regard it as an isolate.

Table 6. Sawa/Awji lexical comparisons

meaning	SAWA [mrf]	AWJI [auw]
I	ka	ko
you	sEm ("you pl.")	kebe ("you pl.")
we	kam	yebe
person	sisEu ("man")	kir ("man")
fish	oNgles	o
dog	w3s	w3l
louse	ku	tu
tree	s3k	ti
leaf	f3k3n	ti fiye
skin	son	f3ker
blood	w3tw3n	keane
bone	ok	sak3r
ear	uskNs	keato
eye	naf	nayo
nose	s3npok3p	nubru
tooth	an	ka
tongue	mos3n, mos	marie
knee	ambl3s	tumtkur
hand	s3k3s, s3ksan ("hand, arm")	kenie ("arm")
breast	pan	m3*
see	naf o*ni	nayo tai
hear	sko	keatik3rk3ri
come	laf	manam
sun	ninaf	mentao
star	waf	mase

water	w3t3l	wobio
stone	s3pat	ser
fire	bot	tao
path	mul	m3Ngir
mountain	Nubikin	yunu
night	yaNga	yaburoa
new	somb3n	no*mo*

Although Nimboran, Kapauri, and Border, together with many other families, have been lumped in a Central and Western Trans-New Guinea Phylum (Voorhoeve 1975b), there have been no suggestions in the literature that these three families in particular have a closer relationship.

3.9. Pahoturi/Eastern Trans-Fly

Fig. 9. Locations of Pahoturi (red) and Eastern Trans-Fly (blue)

These HH families are sisters in the tree and rank #83 (*NSIM*) and #581 (*SIM*), so their relatedness has support. They were lumped together, along with many other families, in a Trans-Fly stock by Wurm (1975:331). This author, however, did not see any particularly close relationship between Pahoturi and Eastern Trans-Fly, but actually assumed that they belong to separate divisions within his far-flung stock.

3.10. Abun/Maybrat/West Bird's Head

Fig. 10. Locations of Abun (red), Maybrat (blue), and West Bird's Head (yellow)

The HH families Abun, Maybrat, Mpur, and West Bird's Head form a cluster. The ranks among the 6 pairs are as follows: Abun/Maybrat: #483 (*NSIM*), #79 (*SIM*); Mpur/West Bird's Head: #122 (*NSIM*), #140 (*SIM*); Abun/West Bird's Head: #1020 (*NSIM*), #1379 (*SIM*); Maybrat /West Bird's Head: #178 (*NSIM*), #241 (*SIM*); Abun/Mpur: #8604 (*NSIM*), #3304 (*SIM*); Maybrat/Mpur: #21,962 (*NSIM*), #15,086 (*SIM*). If Maybrat is related to both West Bird's Head and to Abun, as suggested by the high *NSIM* ranks, then Abun, by transitivity, should also be related to West Bird's Head, and the *NSIM* rank is, indeed, relatively high for Abun/West Bird's Head. It is harder to fit Mpur into the equation. Its relation to West Bird's Head is high-ranking, its relation to Abun is relatively high ranking, but that to Maybrat ranks low. In order to decide how to interpret this case we can compare the Abun, Maybrat, and Mpur data so as to develop a better sense of how solid these relations are, cf. Table 7.

Table 7. Maybrat/Abun lexical comparisons.

meaning	MAI BRAT	ABUN	MPUR
I	tuo, tuwo	ji	in
you	nuo, n	**nan**	**nen**
we	amu, p	men	yek
one	sau, s	dik	tu
two	eok, ewok	we	dokir
person	rae	ye	man, mamir
fish	sa	boge	mw~an
dog	**mtax, mtah**	**nd~ar**	per
tree	ara	kw~e	ni
blood	mes	de	far
ear	imara	git	kw~aip
nose	naif, nayif	gwembo	minsan, wanken
drink	**ata**	**da**	kobet
see	he, xe	me	wot
hear	ari	jam	minsem
die	hai, hayi	kw~op	ut
come	**ama**	**ma**	na
sun	isie, ayo	kam	put
water	aya	Sur	war
path	**iso**	**os**	nj~an, bw~ak
mountain	atu, wiam	banbo	sor
night	mti	noru	dim
full	atot	sEs	bit, berem
name	asom	gum	muk

Among the 24 Mai Brat/Abun lexical comparisons in Table 7 there are four that look to be solidly cognate: ‘dog’,[7] ‘to drink’, ‘to come’, and ‘path’. Some other weaker candidates also appear: ‘you’, ‘person’, and ‘night’. While the evidence is not overwhelming it nevertheless looks promising. Abun and Mpur, however, do not have the appearance of relatedness. Only the words for ‘you’ look like true cognates. Even by long stretches of the imagination only a few more, such as ‘come’ and ‘water’, could be added as candidates for cognacy. Given that Mpur only seems to show relatedness to West Bird’s Head but not to West Bird’s Head’s likely relatives Maybrat and Abun, I prefer to not include Mpur in the hypothetical Abun/Maybrat/West Bird’s Head group.

The Abun/Maybrat/West Bird’s Head group is isomorphic with the West Papuan phylum of Berry and Berry (1987b), who suggest the grouping mainly based on cognate counts. Reesink (2005:187) briefly mentions pronouns, gender distinctions, and some verbal prepositions as kinds of evidence that might link Maybrat (but not Abun or Mpur) to West Bird’s Head, but still regards Maybrat (as well as Abun and Mpur) as isolates.

3.11. Yareban/Mailuan

Fig. 11. Locations of Yareban (red) and Mailuan (blue)

Investigating the possible relationship between Yareban and Mailuan also involves looking into possible connections with Dem. This language is not shown in the map in Figure 11, but it should be noted from the outset that it is located in a completely different region, namely in the western highlands of the Indonesian part of New Guinea. All three HH families form a cluster, with Dem and Yareban as sisters and Mailuan as a more distant relative. The ranks are as

[7] In the next section (3.11) I discuss another case where words for ‘dog’ are similar, this time arguing that borrowing explains the similarity. The evidence supporting the latter assertion is the widespread occurrence of similarly-shaped words in Papuan languages as well as in Oceanic. In contrast, forms similar to Mai Brat *mtax*/*mtah* and Abun *nd~ar* are not widespread. In fact, in the Papuan dataset the only words for ‘dog’ that have an initial nasal + alveolar stop sequence are Taiap [gpn] *nc~ar* and Angoram/Kambrindo [aog] *ndanda*. The former language is an isolate, the latter a member of Lower Sepik-Ramu in the HH classification. The two languages are spoken close to one another but far away from Mai Brat and Abun, on the northeastern coast of New Guinea; thus they are unlikely to be involved in diffusion of the words for ‘dog’ in Mai Brat and Abun. It is likely that the word for ‘dog’ is shared between Taiap and Angoram/Kambrindo, but this is another story.

follows: Dem/Yareban: #129 (*NSIM*), #20 (*SIM*); Yareban/Mailuan #51 (*NSIM*), #21 (*SIM*); Dem/Mailuan: #9154 (*NSIM*), # 6159 (*SIM*). It is somewhat surprising that Dem and Mailuan are apparently so different when both are very similar to Yareban. Dem and Yareban are represented by single doculects, while Mailuan is represented by three very close doculects carrying different ISO 639-3 codes. This enables us to quickly inspect the data. In order not to clutter Table 8, where lexical comparisons are made, I have arbitrarily chosen just one representative of Mailuan, namely Laua (luf).

Possible cognates between Dem and Laua, the problematical pair, are marked in bold. There are 3 such pairs which, by a stretch, may be conceived of as possible cognates in the list of 22 items, with one, 'breast', possibly to be discounted as sound symbolic. Words for 'breast' throughout the world's languages have an average of four segments and the most frequently occurring ASJPcode symbols in the four positions are *m*, *u*, *m*, and *a* (Wichmann et al. 2010b). Thus, forms like Dem *ami* and Laua *hama* are similar in shape to words for 'breast' in many languages throughout the world. This does not look like a promising relationship. That raises the question about whether one of the pairs Dem/Yareban and Yareba/Mailuan are possibly not genealogically related after all. Both pairs cannot be valid genealogical units if Dem and Mailuan are not related.

Dem/Yareba have similarities throughout the set of pronominal forms 'I', 'you', 'we'. But Dem interestingly has synonyms for two of these, increasing the probability of spurious matches. There is an identical word for 'dog' in both languages. Because of the great differences in the rest of the items, I suspect that this is a loanword. In fact, it seems to be a Wanderwort, because when inspecting words meaning 'dog' one finds similarly shaped words throughout the Papuan languages, e.g. (in ASJPcode), *3p3na, 5amp, 5imboaN, 5ombwi, 5umbakal, 5umb~ua, aga, age, agoa, agoa, amb~aipu, gwala, gwara, gw~ai, ka, kp~oro, kp~oto, kui, kw~3r, kw~a, kw~a*, oa, owa, oana, obe, obe, ofun, okw~a, pwat, u*ku*lo, ubri, ubui, uwaNku, uwi, uwura*, etc. I take it that the origin is in Oceanic, since words for the dog is similarly shaped in some Oceanic languages, e.g., Kilivila *kaukw~a*, Kove *kauwa*, Lengo, Lusi, Mbirao, Nggela, Tolo *kau*, Torau *kaukau*, Tungak *kauvek*, Vitu *kaua*. Finally, there are similarities in the words for 'fire' and 'night', respectively, but these could be accidental. Thus Dem and Yareban similarities are not convincing of a genealogical relation. Their similarity score placing them as #129 in terms of *NSIM* seems to come from a mixture of chance and borrowing.

The Yareban/Mailuan pair looks more convincing, with cognate-looking forms for 'fish', 'louse', 'tree', 'ear', 'see', 'hear' (and 'breast'). In addition, a-vowels are found in all three pronominal forms, suggesting similar systems of indicating distinction between pronouns by other means than vowel qualities. On the basis of these various considerations I hypothesize that Yareban/Mailuan to be related while Dem does not feed into the equation.

Table 8. Dem/Yareban/Mailuan lexical comparisons

meaning	DEM	YAREBA	LAUA
I	nau, no	na	ya7a
you	aN, yu	a	ga7a
we	Yu	ya	gea
fish		**erio egi, ogo egi**	**orabe**
dog	kw~a	kw~a, kw~asiri	dahari
louse	ndu, nduse	reiba**, ua**	**tuma**
tree	niye	**ana** oma	**hana**
blood	amiyep, miet	iwa, onono	lala
bone	awak	tai	gisa
ear	nado, nadoN	**ome**	**ope**
eye	**ain**gewu, eNgip	diti, natei	**ini**
tooth	naNkasa, yaNkasa	nio	ma7a
knee		yajigo	turuna
breast	**ami**	**ama**	**hama**
drink		ogo it	hihilma7a
see	aige kotak, korak	**er**	**hel**bau
hear	aindemo, nadunoye	**naut**	**nanba7a**
come	me, menaNot	ar, far	hai
star		muina, kodara	nigoru
water	da, yat	ogo	ne7ama
stone	daNat, Nga	gebiro, oma	baga
fire	**kanu**, kuna	ina	**heu**
road	dundak, mbo	daba, darei	vagorodi
mountain	dum, Na	maidani	horo
night	damuk	dumuro	garuru
full		beda, wate farinu	ma7apulaha
new		reka	gadara
name	agatiene, aluN	ifu	nim

Yareban and Mailuan are included in the far-flung South-Eastern Trans New Guinea phylum of Dutton (1975), but they are not singled out as particularly closely related. Interestingly, however, of all the different pairs of groups in the dubious phylum, Yareban-Mailuan shows the highest percentage of cognates (26%) in the count of Dutton (1975:628).

3.12. A residual case

Bilua and Savosavo is the final case where both Table 2 and the tree support a genealogical relationship. However, Dunn and Terrill (2012) argue that the lexical evidence for the relatedness between these languages (as well as the two other 'Central Solomons Papuan' languages) vanishes when Oceanic (Austronesian) loanwords are excluded. I will follow Dunn and Terrill in this assessment and not group Bilua and Savosavo together.

4. Conclusion

In conclusion, below I present the hypothetical, basic classification of Papuan languages arrived at through the above considerations. New nomenclature is not introduced. Families that are considered to not be supported are split up into the fragments suggested by the ASJP Papuan tree, and these fragments are labeled "Ex-Fam-#", where "Fam" is the HH name of the unsupported family and "#" is a number. If one of these groups is isomorphic with some subgroup in *Ethnologue*, this subgroup's name is supplied in a parenthesis. The list is given in the order in which the groups appear in the ASJP tree, from top to bottom. No attempt is made to also offer subgrouping schemes, but suggestions can be retrieved from the ASJP Papuan tree. The languages that belong to each group are indicated using ISO 639-3 codes or language names when codes are not available. Languages that are supposed to belong to a HH family considered supported, but which do not occur under the same node as the bulk of the languages in the HH family, are listed as if they nevertheless did belong to the family in question, but their potentially problematical status is indicated by a question mark.

1. *West Timor-Alor-Pantar/East Timor-Bunaq*
   abz/abz?, adn, beu, bfn?, ddg, hmu, klz, kpu, kvd, kvw, kyo, lev, mkz, nec, oia, swt, twe, woi, Kaera, Kawa, Sar Indonesia
2. *South Bougainville*
   buo, nas, siw
3. *Wiru*
   wiu
4. *Namla-Tofanma*
   tlg
5. *Ex-Pauwasi-1 (Western Pauwasi)*
   dmu, ttn
6. *Ex-Nuclear Trans New Guinea-1 (Asmat-Kamoro)*
   asc, asi, asy, cns, irx, kgq, nks, txt, xse,
7. *Mombum*
   kdw, mso
8. *Marindic*
   bgv, jaq, kvg, mrz, zik
9. *Ex-Nuclear Trans New Guinea-2 (Awyu-Dumut)*
   aax, ahh, aws, awy, bwp, khe, psa, saw, wms
10. *Inland Gulf*
   ipo, mcv, tsx
11. *Ex-Nuclear Trans New Guinea-3(Oksapmin)*
   opm
12. *Ex-Nuclear Trans New Guinea-4 (Ok)*
   bhl, fai, kti, kts, mpt, nxr, sug, tif, tlf, yon

13. *Ex-Nuclear Trans New Guinea-5 (Finisterre-Huon)*
    awx, bmu, ded, kgf, klt, kmg, kpf, ksr, mci, mlh, mpp, naf, nif, nnk, ons, spl, tbv, tim, wnc, yut
14. *Goilalan*
    fuy, ttd
15. *Ex-Nuclear Trans New Guinea-6 (Chimbu-Wahgi)*
    doa, gam, gvf, kue, med, nac, sst, wgi
16. *Kamula/Awin-Pa/Bosavi/East Strickland*
    agl, ail, bco, beo, etr, goi, jko, khs, kkc, onn, ppt, siq, smq, xla
17. *Ex-Dibiyaso-Doso-Turumsa-1*
    dby
18. *Angan*
    aak, agm?, ago, apz, byr, hmt, kcb, klp, mcr, miw, smb, ygw
19. *Duna-Bogaya*
    duc, boq
20. *Ex-Nuclear Trans New Guinea-7 (Engan)*
    bir, enq, hui, kew, kjs, kjy, kyc, leq, ssx
21. *Sepik/Ndu/Walio*
    abt, amp, bjh, bye, bzf, dju, gbe, ham, ian, iwm, kmn, kmo, mle, nnm, nud, sim, sny, tww, wla, ybx, ylg, yss
22. *Greater Kwerba/Tor-Orya*
    bkl, kwe, srl?, tmj, ury, xau
23. *Nimboran/Kapauri/Border*
    amn, auw, dnd, jet, khp, msf, nir, snu, sow, wrs
24. *Elseng*
    mrf
25. *North Halmahera*
    gbi, loa, mqo, mqs, pgu, saj, tby, tlb, tvo
26. *Yale*
    nce
27. *Ex-Dibiyaso-Doso-Turumsa-2*
    dol
28. *Kwomtari*
    kwo
29. *Ex-Nuclear Trans New Guinea-8 (Mek)*
    eip, kkl, mtg, xte
30. *Ex-Morehead-Wasur-1*
    jei, ncm
31. *Unclassified (Kenaboi)*
    xbn

32. Hatam-Mansim
had
33. Mor
moq
34. Pahoturi/Eastern Trans-Fly
bon, gdr, idi, kit, tof, ulk
35. Ex-Nuclear Trans New Guinea-9 (Kainantu-Goroka)
agd, aso, auy, awb, bef, bjr, for, gaf, gah, gaj, gim, ino, isa, kbq, snp, tbg, waj, yby, ygr
36. Yareban/Mailuan
dof, luf, mgu, yrb
37. Dem
dem
38. Ex-Nuclear Trans New Guinea-10
anh, ate, ena, faj, imi, kqa, mmq, msx, omo, pda, pmr, sbq, wdg
39. Ex-Nuclear Trans New Guinea-11 (Dani)
dni, dnt, dnw, wlw, wno, wul, yli
40. West Bomberai
bdw, ihp, kgv
41. Ex-Nuclear Trans New Guinea-12 (Wissel Lakes)
ekg, mnz
42. Koiarian
aom, bbb, kbk, kqi, mcq
43. Kaki Ae
tbd
44. Moraori
mok
45. Mawes
mgk
46. Kolopom
kig, nqm, ran
47. Bulaka River
jel, mgf
48. Molof
msl
49. Yuat-Maramba
kql
50. Kaure-Narau
bpp
51. Tirio

aup

52. Kayagar

aqm, kyt, tcg

53. Suki-Gogodala/Waia/Kiwaian

aac, bcf, ggw, kiw, kjd, kmx, knv, kxz, mdb

54. Ex-Nuclear Trans New Guinea-13

bhg, bjz, koz, kpr, sue, wsk, zia

55. Fasu-East Kubutu

faa, foi

56. Pawaia-Teberan

mps, ppo, pwa

57. Turama-Kikori

klq, meb, mgx

58. North Bougainville

kyx, roo

59. Eleman

iar?, opo, oro, tqo, uar, xeu

60. Mairasi

etz, zrs

61. Touo

tqu

62. Ex-Kwalean-1

huf, ksj

63. Tanahmera

tcm

64. Savosavo

svs

65. Bilua

blb

66. Manubaran

kqc, mds,

67. Kuot

kto

68. Burmeso

bzu

69. Amto-Musan/Left May/Busa

amm, amt, bhf, bpw, itr, mmp, nax, niw, owi

70. Ex-Sentanic-1

dmy

71. Ex-Lower Sepik-Ramu-1

kbx

72. *Taiap*

gpn

73. *Ex-Sko-1*

ksi, skv, vam, wut, Dusur, Leitre

74. *Ex-Lower Sepik-Ramu-2*

aog, can, mtf, xop, yee

75. *Geelvink Bay*

trt

76. *Konda-Yahadian*

knd

77. *South Bird's Head Family/Inanwatan*

bzp, jbj, kzm, pru, szp, xod

78. *Nuclear Torricelli*

aif, aof, aon, aun?, ape, avt, but, bvn, eit, ele, kms, lsr, mkc, mty, mwb, niz, ong, rhp, siu, tei, tua, urt, urx, van, xbi, wmo?, yev, ymb, ymo

79. *Urim*

uri

80. *Ata*

ata

81. *Monumbo*

lll, mxk

82. *Ex-Sentanic-2 (Sentani Proper)*

set, tnm

83. *Ex- Lower Sepik-Ramu-3*

byz

84. *Yawa*

yva

85. *Ex-Kwalean-2*

mfw

86. *Lavukaleve*

lvk

87. *Anem*

anz

88. *Ex-Morehead-Wasur-2*

pep

89. *Papi*

ppe

90. *Mpur*

akc

91. *Abun/Maybrat/West Bird's Head*
ayz, kgr, kzz, msg, mxn, sbg
92. *Lakes Plain*
afz, awr, bqq, dbf, ert, fau, kiy, pas, rac, spi, tad, tds, tmu, tty, wbe
93. *Pyu*
pby
94. *Ex-Biksi-1*
sbt
95. *Ex-Sko-2*
rwa, wra, Poo, Ramo, Sumararo, Womo
96. *Ex-Biksi-2*
yet
97. *Yeli Dnye*
yle
98. *Lepki/Murkim*
lpe, rmh
99. *Ex-Pauwasi-2 (Eastern Pauwasi)*
enr, wfg, yuj
100. *East Bird's Head*
mej, mnx, mtj
101. *Kosare*
kiq
102. *Usku*
ulf
103. *Ex-Nuclear Trans New Guinea-14*
abw, ali, bie, bql, buq, dmc, hih, kgu, mhl, mjj, mkr, mmi, mvq, ped, pla, prw, sks, ukg, wnb, wnu, xow, ybm, yrw
104. *Ex-Nuclear Trans New Guinea-15*
kpw
105. *Senagi*
kbv
106. *Piawi*
pnn, tmd
107. *Ex-Lower Sepik-Ramu-4*
rao
108. *Ex-Lower Sepik-Ramu-5*
geb, kct, msy
109. *Ex-Nuclear Trans New Guinea-16*
aey, asd, awm, bbd, bbr, bmh, bmx, boj, bpi, bpm, bpu, dnr, duk, eri, fad, gap, gaw, ggl, gmu, gyb, igo, jil, klm, kmf, kop, lei, mcz, mdc, mlp, mqe, mqv, mqw, mtc, nbk, pnr, pup,

rea, rmp, rpt, six, snr, snx, snz, spd, sra, ssd, ssj, swm, tya, urg, urw, usu, utu, wmc, wtf, xes, xsp, ybo, ydk, ynl

Starting from 104 families in the HH classification of the 60% of the Papuan languages under consideration here we have ended up with 109. Some of the HH families which have been split up can probably, at least in part, be reunited with more work on the data and inspection of the evidence that experts have put forward for the different proposals, but roughly the same number of families as in the HH classification for languages included in this paper may be a realistic number for a conservative classification immediately within reach. The methodology adopted is not exhausted with this study. The above new proposals for genealogical relationships should be investigated in more detail, drawing upon all data available. Some of the proposed relations may be due to chance or loanwords, so this further step is needed to establish the relations with a greater degree of confidence.

Once the proposed new relationship have been studied in more detail the exercise can be iterated using something like the above units in producing similarity measures for entire groups. The rubble left from breaking up weakly supported families has not been reused for new construction work, but there is no doubt that some larger groupings can be established. Just looking at the tree and observing branch lengths leading to nodes uniting some members of some ex-families with members of other families induces hope in this regard. For instance, promising groupings to investigate would be Mombum/Ex-Nuclear Trans New Guinea-1 (Asmat-Kamoro) or Piawi/Ex-Lower Sepik-Ramu 4/ Ex-Lower Sepik-Ramu 5. Of course, all relevant information should be extracted from the literature and used. For instance, in the case of Mombum and Asmat-Kamoro there are about a dozen Mombun words listed along with the 418 Asmat cognate sets of Voorhoeve (1980). It would obviously also improve the classification of Papuan languages to increase the current 60% coverage in the ASJP database.

The impressive genealogical diversity represented by the non-Austronesian languages of the New Guinea region represents a great challenge to comparative linguistics, and I hope to have shown that computational methods can be an aid in this enterprise. The main contribution of this paper has been to identify genealogical relations which are good candidates for becoming firmly established once more detailed work is undertaken, applying the comparative method.

Appendix 1: The ASJP tree of Papuan languages

See pages 357-386 below, for the ASJP tree of Papuan languages.

Appendix 2: Description of *CN*

N is the number of pairs with one list from each family, so if one family has *m* lists and the other family has *n* lists, then $N = m * n$. Other things being equal, the bigger *N* is, the more reliable the average similarity between the families is. One thing that is not equal is the correlation between the lists. As an extreme example, if all the lists in a family are copies of the same list, then all the copies are no better than the one original list no matter how many copies there are. In general, the more highly correlated the lists are, the less helpful additional lists are. *CN* is *N* corrected for the correlations between lists in the same family: *CN* = *m'* * *n'*, where *m'* and *n'* are *m* and *n* corrected for correlations.

To derive the correction, let a family with *n* lists be given, and let a list from a language outside the family also be given. Let s_i be the similarity between the *i*th list in the family and the list outside the family. The possibility of a relationship between the given family and the outside language can be tested by observing the mean similarity and the variability of the mean: a high mean with low variability provides evidence for a relationship. The mean similarity $\bar{s}$ is defined as:

$$\bar{s} = \Sigma_i s_i/n.$$

The variability of $\bar{s}$ can be expressed by its variance $V(\bar{s})$, which is:

(1) $V(\bar{s}) = V(\Sigma_i s_i/n) = V(\Sigma_i s_i)/n^2.$

The standard expansion for variance of a sum is:

(2) $V(\Sigma_i s_i) = \Sigma_\mathrm{i} V(s_i) + \Sigma_{i\neq j} r_{ij}\sqrt{}[V(s_i)V(s_j)],$

where r_{ij} is the Pearson correlation between s_i and s_j across all the lists outside the given family. Under the null hypothesis that languages in different families are unrelated, it is reasonable to assume that $V(s_i) = V(s_j)$ for all *i* and *j*; let $V(s)$ denote the common variance. Substituting $V(s)$ for $V(s_i)$ and $V(s_j)$ in (2) produces:

(3) $V(\Sigma_i s_i) = nV(s) + V(s)\Sigma_{i\neq j} r_{ij}.$

Now let *r* denote the mean of the r_{ij}, which is:

$$r = \Sigma_{i\neq j} r_{ij}/[n(n\text{-}1)].$$

Substituting this in (3) leads to:

$$V(\Sigma_i s_i) = nV(s)[1 + (n\text{-}1)r],$$

and substituting this back in (1) produces:

(4) $V(\bar{s}) = V(s)[1 + (n\text{-}1)r]/n.$

Finally, let *n'* be defined as $n/[1 + (n\text{-}1)r]$. With this substitution, (4) simplifies to:

$$V(\bar{s}) = V(s)/n'.$$

If $r = 0$, then $n' = n$; thus, *n'* can be interpreted as the number of independent lists that would produce the same $V(\bar{s})$ as do the given *n* correlated lists. If $r = 1$, then $n' = 1$, because *n* lists are no better than one if they are all perfectly correlated.

For comparing two families rather than a single family and a single list, *CN* is the product of the two corrected family sizes. *CN* is usually much lower than *N*, because lists in the same family tend to be highly correlated. A conventional test statistic for the relationship between two families, analogous to the *t* statistic, is the mean similarity divided by the standard deviation (SD) of the mean. Since the SD is the square root of the variance, the SD of the mean is inversely proportional to $\sqrt{CN}$. The test statistic is therefore directly proportional to the mean similarity multiplied by $\sqrt{CN}$, which is *NSIM* in Table 2. The constant of proportionality is not estimated, which precludes formal significance tests but does not affect the ranking in Table 2.

Acknowledgments

This paper has benefitted enormously from collaboration with members of the ASJP consortium over the years. Special thanks go to Eric W. Holman, who contributed directly by producing and describing the different versions of the similarity scores for HH families reported in Table 2. Four large sets of comments from the editors—Harald Hammarström and Wilco van den Heuvel—and two referees, as well as comments from Malcolm Ross and Bernard Comrie, certainly helped to improve this paper, even if it could surely have been improved even more.

References

Adelaar, Alexander K. 1995. Borneo as a cross-road for comparative Austronesian linguistics. In: Peter Bellwood, James J. Fox, and Darrell T. Tryon (eds.), *The Austronesian: Historical and Comparative Perspectives*, 75-95. Canberra: Department of Anthropology in association with the Comparative Austronesian Project, Research School of Pacific Studies, The Australian National University.

Aikhenvald, Alexandra Y. 2008. *The Manambu Language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.

Berry, Keith and Christine Berry. 1987a. A survey of the South Bird's Head Stock. *Workpapers in Indonesian Languages and Cultures* 4: 81-117.

Berry, Keith and Christine Berry. 1987b. A survey of some West Papuan phylum languages. *Workpapers in Indonesian Languages and Cultures* 4: 25-80.

Brown, Cecil H., David Beck, Grzegorz Kondrak, James K. Watters, and Søren Wichmann. 2011. Totozoquean. *International Journal of American Linguistics* 77.3: 323-372.

Brown, Cecil H. and Eric W. Holman. 2010. Comparing ASJP approaches to automated classification: correspondence-based and lexical-based trees for Mayan. http://email.eva.mpg.de/~wichmann/papers.htm.

Brown, Cecil H., Eric W. Holman, Søren Wichmann, and Viveka Velupillai. 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.

Donohue, Mark. 2010. Skou. http://email.eva.mpg.de/~wichmann/papers.htm.

Dryer, Matthew S. 2005. Genealogical language list. In: Haspelmath, Martin, Matthew Dryer, David Gil, and Bernard Comrie (eds.), *The World Atlas of Language Structures*, 584-644. Oxford: Oxford University Press.

Dunn, Michael and Angela Terrill. 2012. Assessing the lexical evidence for a Central Solomons Papuan family using the Oswalt Monte Carlo test. *Diachronica* 29: 1-27.

Dutton, Tom E. 1975. South-Eastern Trans-New Guinea Phylum languages. In Wurm, Stephen A. (ed.), *New Guinea Area Languages and Language Study*, Vol 1: *Papuan Languages and the New Guinea Linguistic Scene* (Pacific Linguistics, Series C 38), 613- 64. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Felsenstein, Joseph. 2004. *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates.

Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge University Press.

Foley, William A. 2005. Linguistic prehistory in the Sepik-Ramu Basin. In: Pawley, Andrew, Robert Attenborough, Jack Golson, and Robin Hide (eds.), *Papuan Pasts: Studies in the Cultural, Linguistic and Biological History of the Papuan-speaking Peoples* (Pacific Linguistics 572), 109-144. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Franklin, Karl J. 1973a. Introduction. In: Franklin, Karl J. (ed.), *The Linguistic Situation in the Gulf District and Adjacent Areas, Papua New Guinea*. Pacific Linguistics, Series C 26, 1-30. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Franklin, Karl J. 1973b. Other language groups in the Gulf District and adjacent areas. In: Franklin, Karl J. (ed.), *The Linguistic Situation in the Gulf District and Adjacent Areas, Papua New Guinea*. Pacific Linguistics, Series C 26, 263-277. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Franklin, Karl J. 2001. Kutubuan (Foe and Fasu) and proto Engan. In: Pawley, Andrew, Malcolm Ross, and Darrell Tryon (eds.), *The Boy from Bundaberg: Studies in Melanesian Linguistics in Honour of Tom Dutton*. Pacific Linguistics 514, 143-154. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Franklin, Karl J. and Clemens L. Voorhoeve. 1973. Languages near the intersection of the Gulf, Southern Highlands, and Western District. In: Franklin, Karl J. (ed.), *The Linguistic Situation in the Gulf District and Adjacent Areas, Papua New Guinea*. Pacific Linguistics, Series C 26, 151-167. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Hajek, John. 1998. Kenaboi: An extinct unclassified language of the Malay Peninsula. *Mon-Khmer Studies* 28: 137-162.

Hammarström, Harald. 2010. A full-scale test of the language farming dispersal hypothesis. *Diachronica* 27: 197-213.

Hammarström, Harald & Sebastian Nordhoff. 2012. The languages of Melanesia: Quantifying the level of coverage. In Evans, Nicholas and Marian Klamer (eds.), Melanesian Languages on the Edge of Asia: Challenges for the 21st Century (Language Documentation & Conservation Special Publication 5), 13-34. Honolulu: University of Hawaii Press. [Including a 946 pp appendix].

Hill, Jane. 2011. Subgrouping in Uto-Aztecan. *Language Dynamics and Change* 1: 241-278.

Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, and Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica* 42: 331-354.

Huff, Paul and Deryle Lonsdale. 2011. Positing language relations using ALINE. *Language Dynamics and Change* 1: 128-162.

Laycock, Donald C. and John A. Z'Graggen. 1975. The Sepik-Ramu Phylum. In: Wurm, Stephen A. (ed.), *New Guinea Area Languages and Language Study*, Vol 1: *Papuan Languages and the New Guinea Linguistic Scene* (Pacific Linguistics, Series C 38), 731-764. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Lewis M. Paul. 2009. *Ethnologue: Languages of the World*, 16th Edition. Dallas, Tex.: SIL International. http://www.ethnologue.com/.

Mailhammer, Robert. 2010. Subgrouping Indo-European: A fresh perspective. http://email.eva.mpg.de/~wichmann/papers.htm.

McElhanon, Kenneth A. and C. Voorhoeve. 1970. *The Trans-New Guinea Phylum: Explorations in Deep-Level Genetic Relationships* (Pacific Linguistics, Series B 16). Canberra: Research School of Pacific and Asian Studies, Australian National University.

Müller, André, Søren Wichmann, Viveka Velupillai, Cecil H. Brown, Pamela Brown, Sebastian Sauppe, Eric W. Holman, Dik Bakker, Johann-Mattis List, Dmitri Egorov, Oleg Belyaev, Robert Mailhammer, Matthias Urban, Helen Geyer, and Anthony Grant. 2010. ASJP World Language Tree of Lexical Similarity: Version 3 (July 2010). http://email.eva.mpg.de/~wichmann/language_tree.htm.

Oswalt, Robert L. 1970. The detection of remote linguistic relationships. *Computer Studies in the Humanities and Verbal Behavior* 3: 117–129.

Pompei, Simone, Vittorio Loreto, and Francesca Tria. 2011. On the accuracy of language trees. *PLoS One* 6.6, e20109.

Reesink, Ger P. 1976. Languages of the Aramia River area. In: *Papers in New Guinea Linguistics* 19 (Pacific Linguistics, Series A 45), 1-37. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Reesink, Ger. 2005. West Papuan languages: Roots and development. In Pawley, Andrew, Robert Attenborough, Jack Golson, and Robin Hide (eds.), *Papuan Pasts: Studies in the Cultural, Linguistic and Biological History of the Papuan-speaking Peoples* (Pacific Linguistics 572), 185-220. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Saitou, Naruya and Masatoshi Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425.

Shaw, Daniel R. 1986. The Bosavi language family. In: *Papers in New Guinea Linguistics* 24 (Pacific Linguistics, Series A 70), 45-76. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Sidwell, Paul. 2010. Comparative Mon-Khmer linguistics in the 20th century: Where from, where to? In: K. S. Nagaraja and Kashyap Mankodi (eds.), *Austroasiatic Linguistics: Proceedings of Third International Conference of Austroasiatic Linguistics, 26-28 November 2007* (Central Institute of Indian Languages: Publication 573), 36-104. Mysore: Central Institute of Indian Languages.

Swadesh, Morris. 1954. Perspectives and problems of Amerindian comparative linguistics. *Word* 10: 306-332.

Urban, Matthias. 2009. Pomoan (with notes on Hokan). http://email.eva.mpg.de/~wichmann/papers.htm.

Urban, Matthias. 2009. Iroquoian. http://email.eva.mpg.de/~wichmann/papers.htm.

Voorhoeve, C. L. 1975a. *Languages of Irian Jaya: Checklist, Preliminary Classification, Language Maps, Word Lists*. Pacific Linguistics, Series B 31. Canberra: Research School of Pacific Studies, The Australian National University.

Voorhoeve, C. L. 1975b. Central and Western Trans-New Guinea Phylum Languages. In Wurm, Stephen A. (ed.), *New Guinea Area Languages and Language Study*, Vol 1: *Papuan Languages and the New Guinea Linguistic Scene* (Pacific Linguistics, Series C 38), 345-460. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Voorhoeve, C. L. 1980. *The Asmat Languages of Irian Jaya*. Pacific Linguistics, Series B – No. 64. Canberra: Research School of Pacific Studies, The Australian National University.

Wichmann, Søren, Eric W. Holman, Dik Bakker, and Cecil H. Brown. 2010a. Evaluating linguistic distance measures. *Physica* A. 389: 3632-3639 (doi:10.1016/j.physa.2010.05.011).

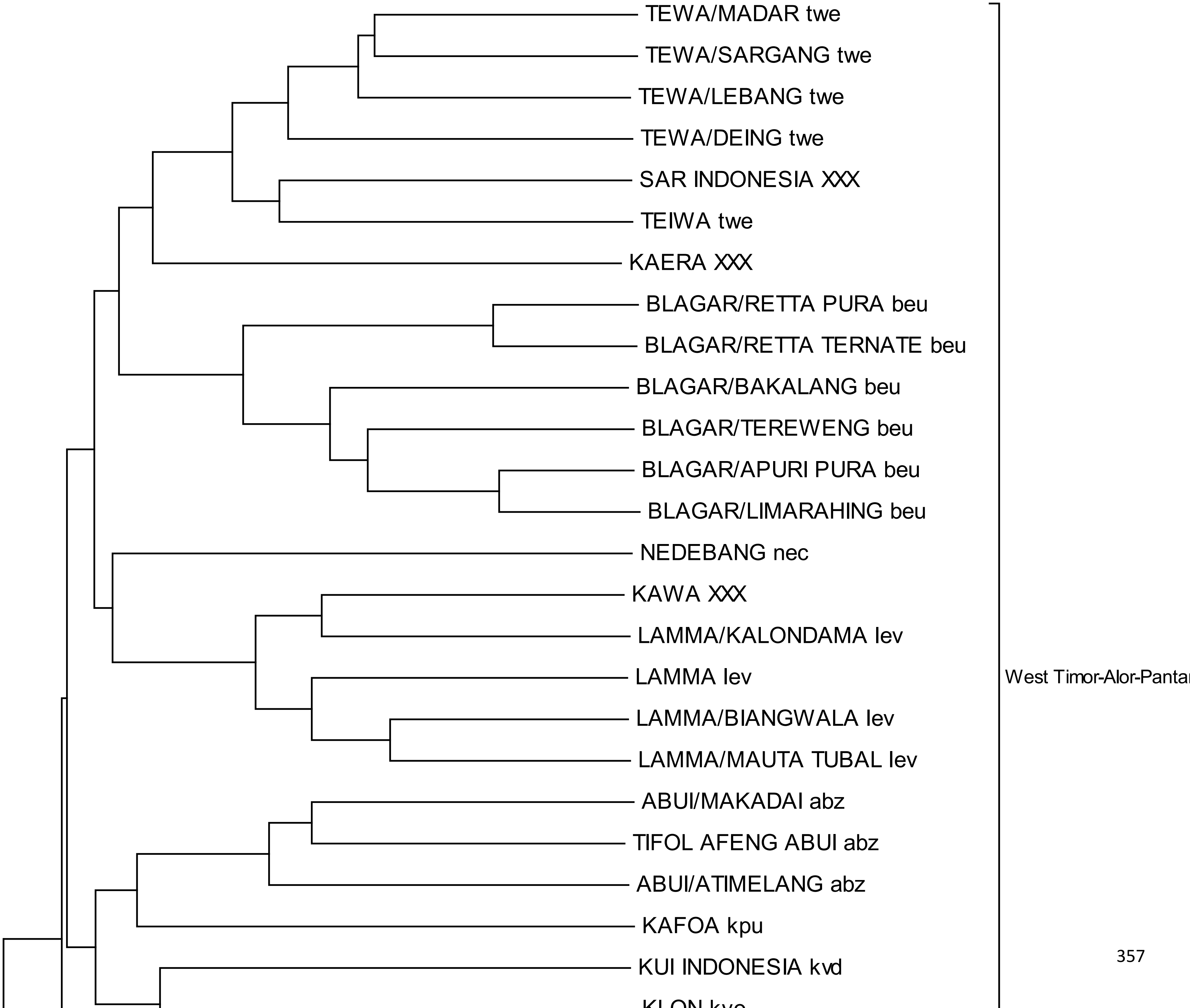
Wichmann, Søren, Eric W. Holman, and Cecil H. Brown. 2010b. Sound symbolism in basic vocabulary. *Entropy* 12(4), 844-858; doi:10.3390/e12040844.Wichmann, Søren, Eric W.

Holman, Taraka Rama, and Robert S. Walker. 2011. Correlates of reticulation in linguistic phylogenies. *Language Dynamics and Change* 1: 205-240.

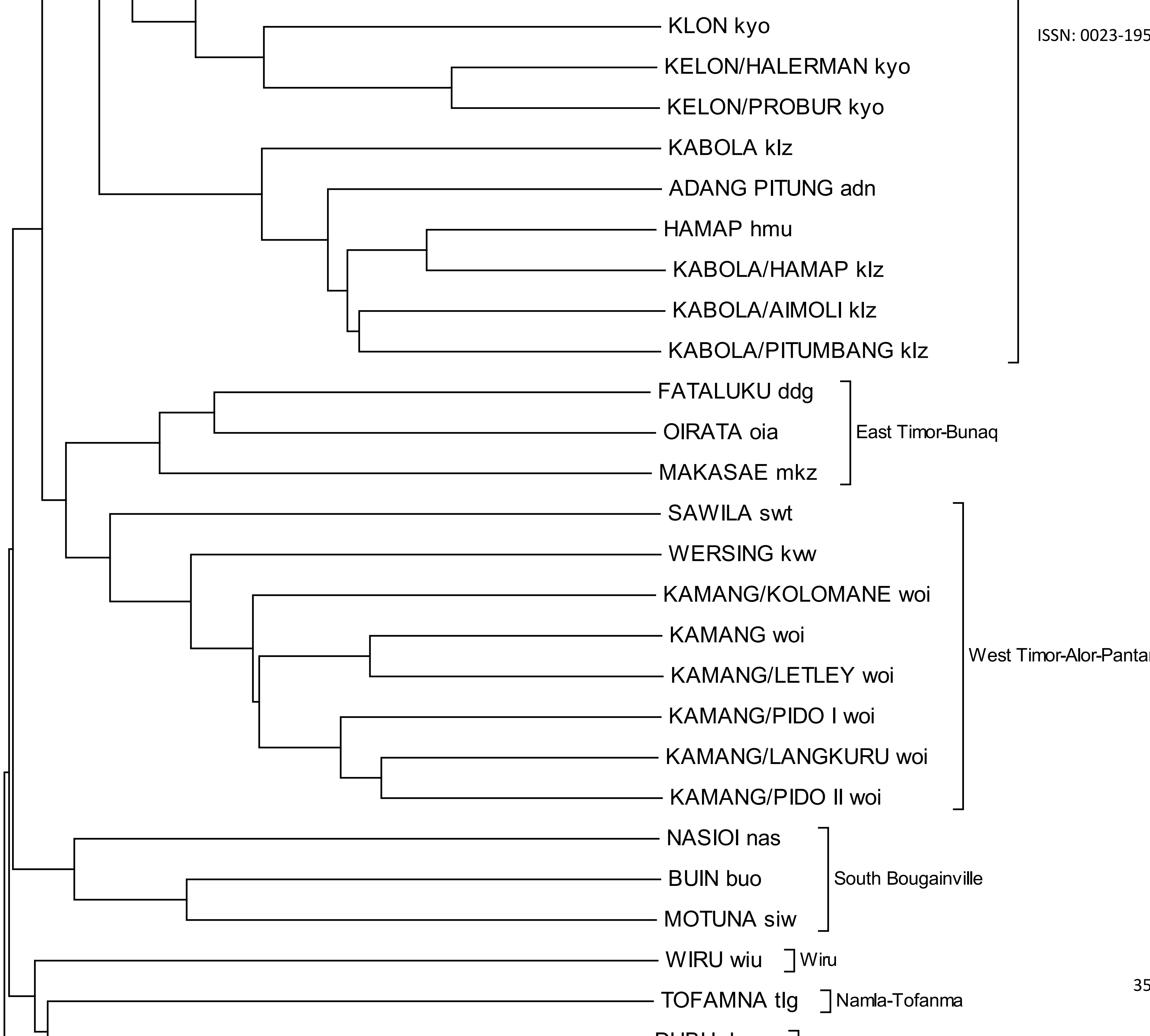
Wichmann, Søren, André Müller, Viveka Velupillai, Annkathrin Wett, Cecil H. Brown, Zarina Molochieva, Julia Bishoffberger, Eric W. Holman, Sebastian Sauppe, Pamela Brown, Dik Bakker, Johann-Mattis List, Dmitry Egorov, Oleg Belyaev, Matthias Urban, Harald Hammarström, Agustina Carrizo, Robert Mailhammer, Helen Geyer, David Beck, Evgenia Korovina, Pattie Epps, Pilar Valenzuela, and Anthony Grant. 2012. The ASJP Database (version 15). http://email.eva.mpg.de/~wichmann/listss15.zip.

Wurm, Stephen A. 1975. The central and western areas of the Trans-New Guinea phylum: The Trans-Fly (sub-phylum-level) stock. In: Wurm, Stephen A. (ed.), *New Guinea Area Languages and Language Study*, Vol 1: *Papuan Languages and the New Guinea Linguistic Sene* (Pacific Linguistics, Series C 38), 323-344. Canberra: Research School of Pacific and Asian Studies, Australian National University.

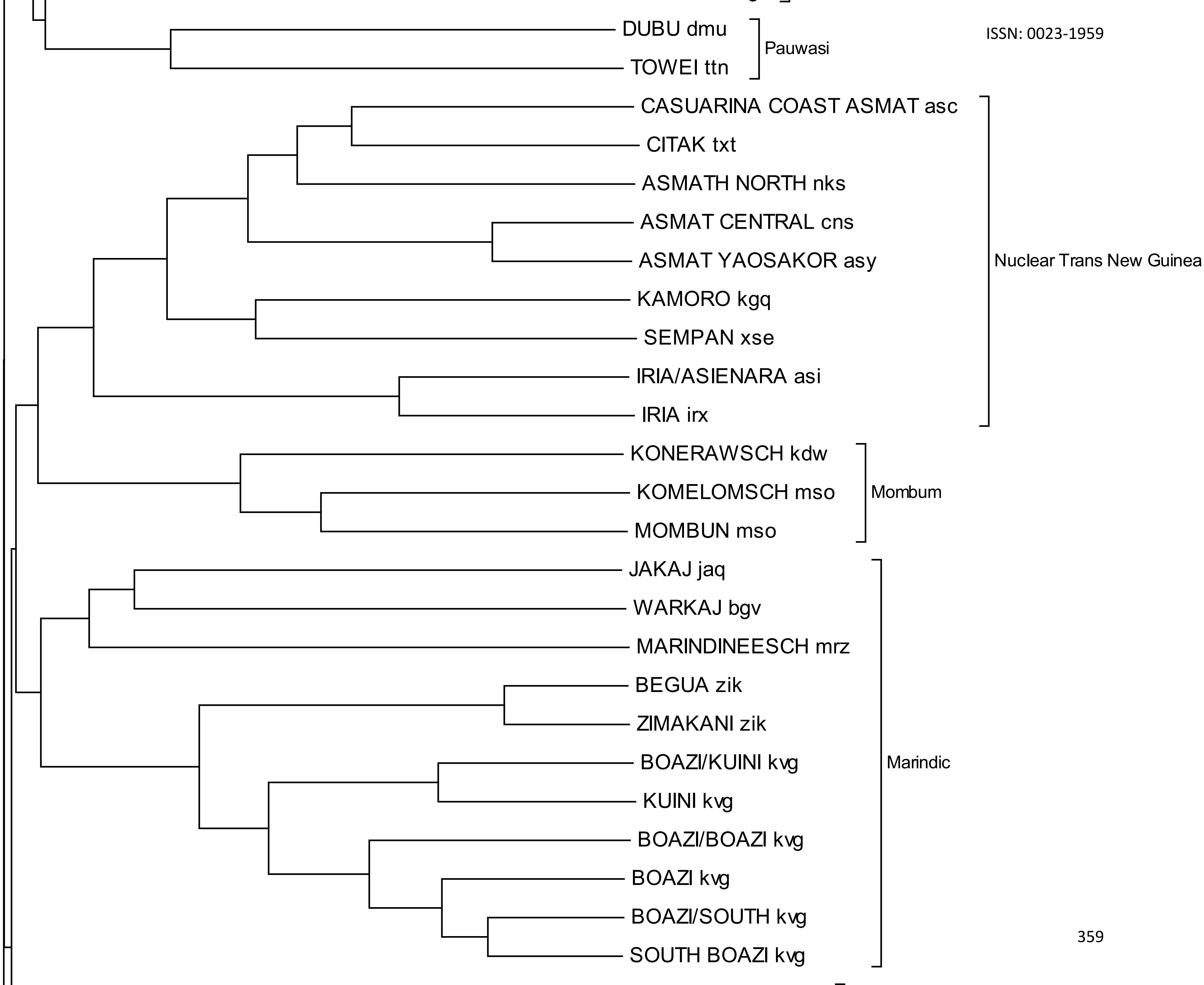
ISSN: 0023-1959



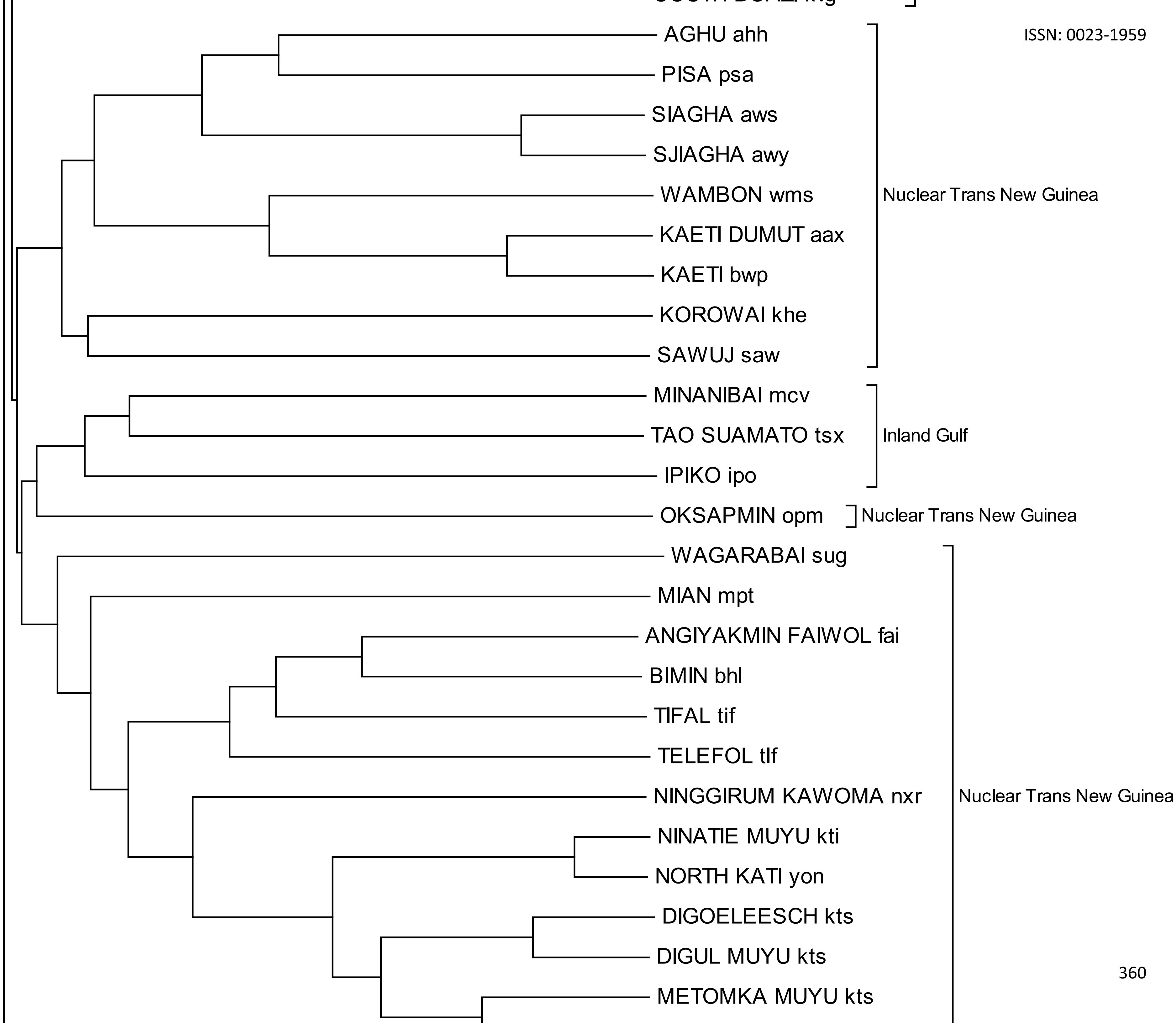
ISSN: 0023-1959



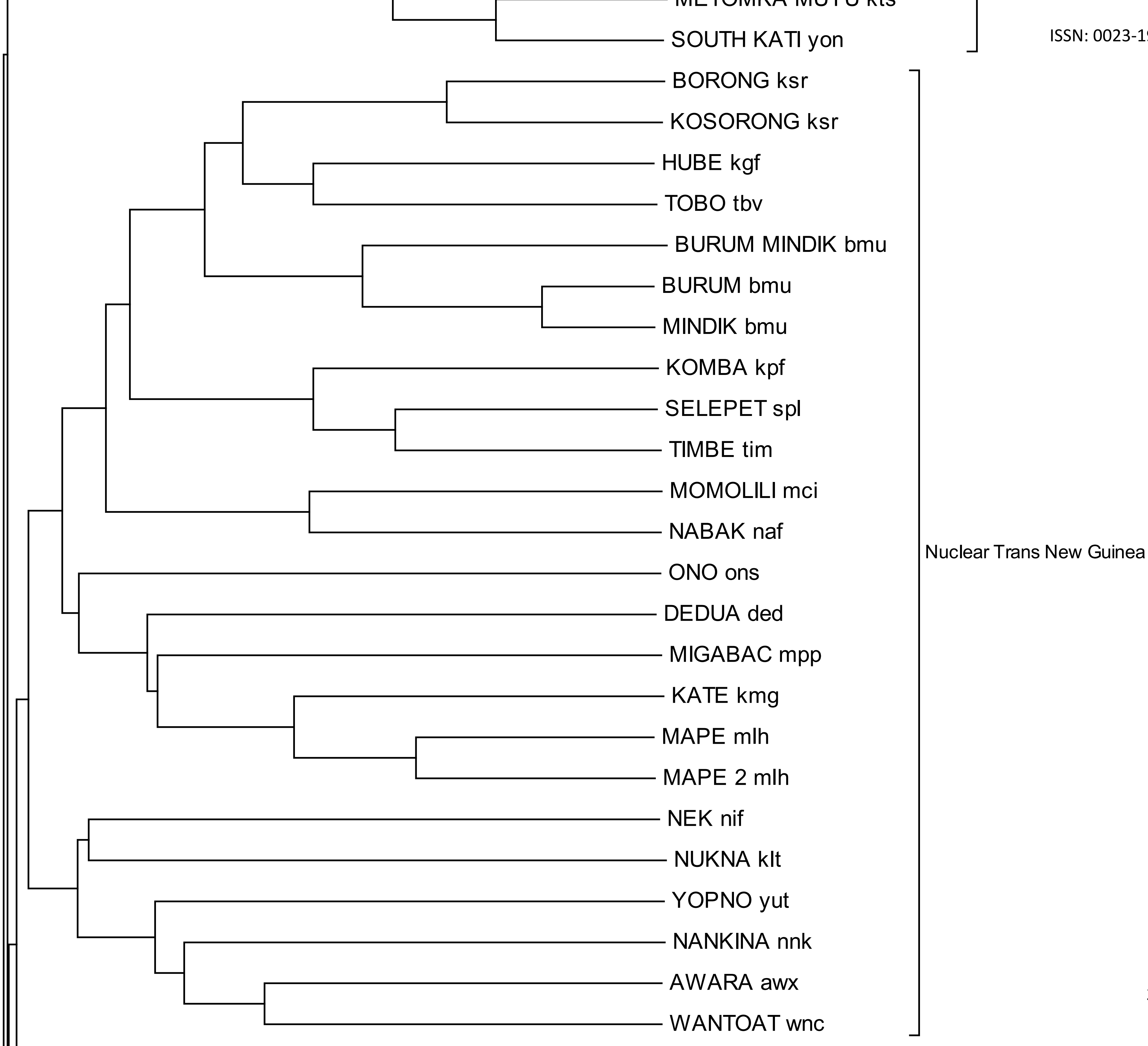
ISSN: 0023-1959



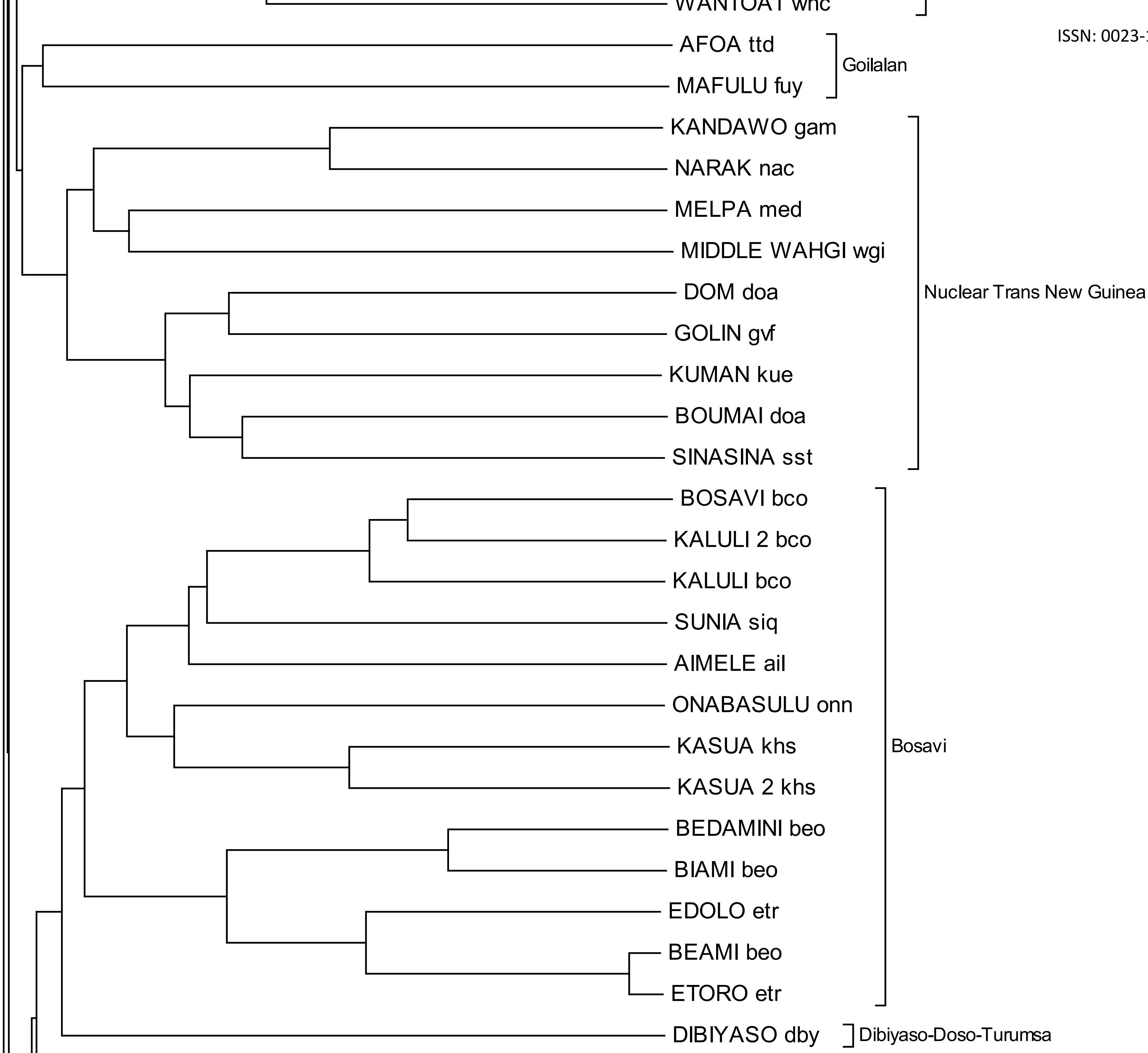
ISSN: 0023-1959



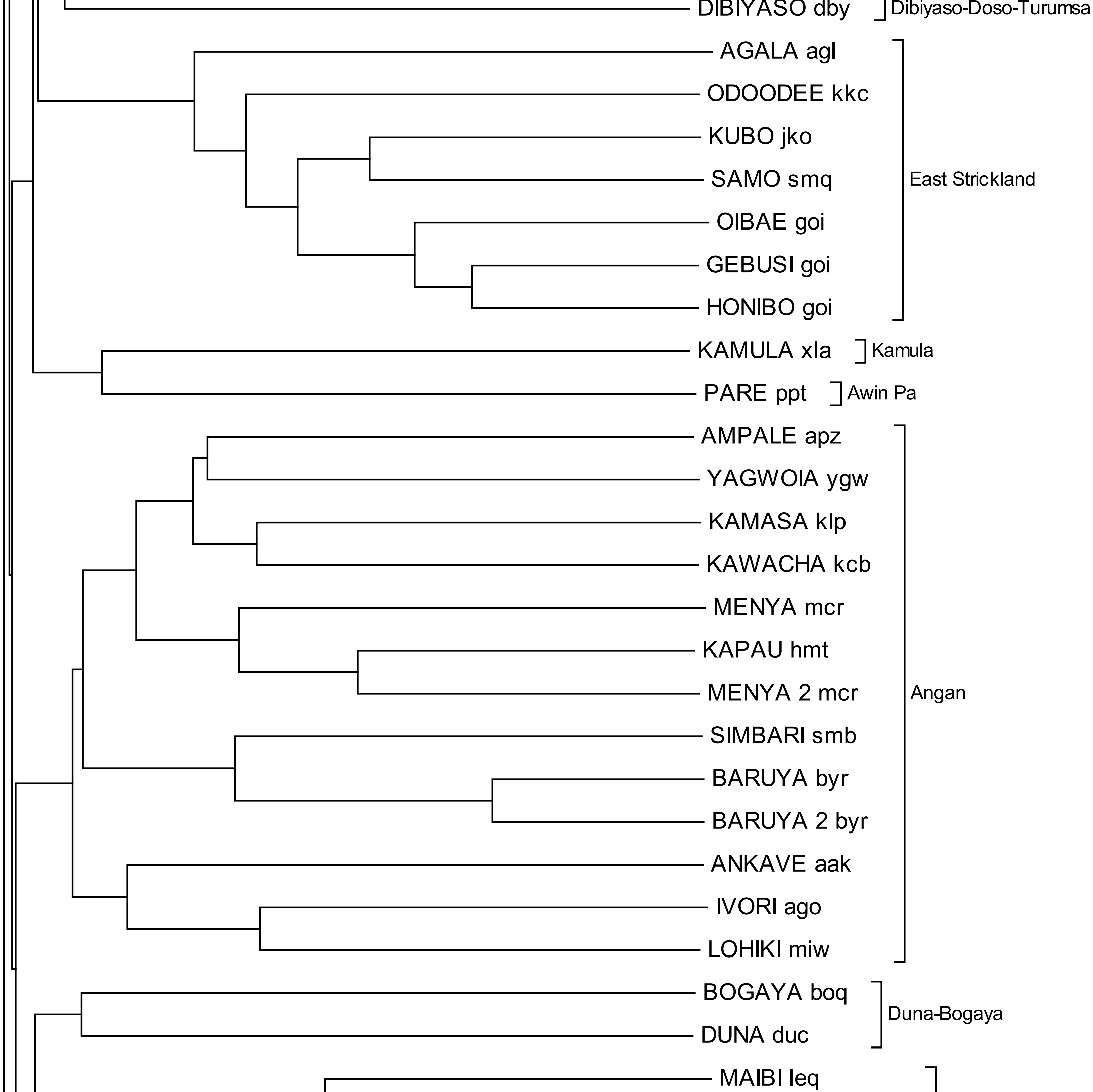
ISSN: 0023-1959



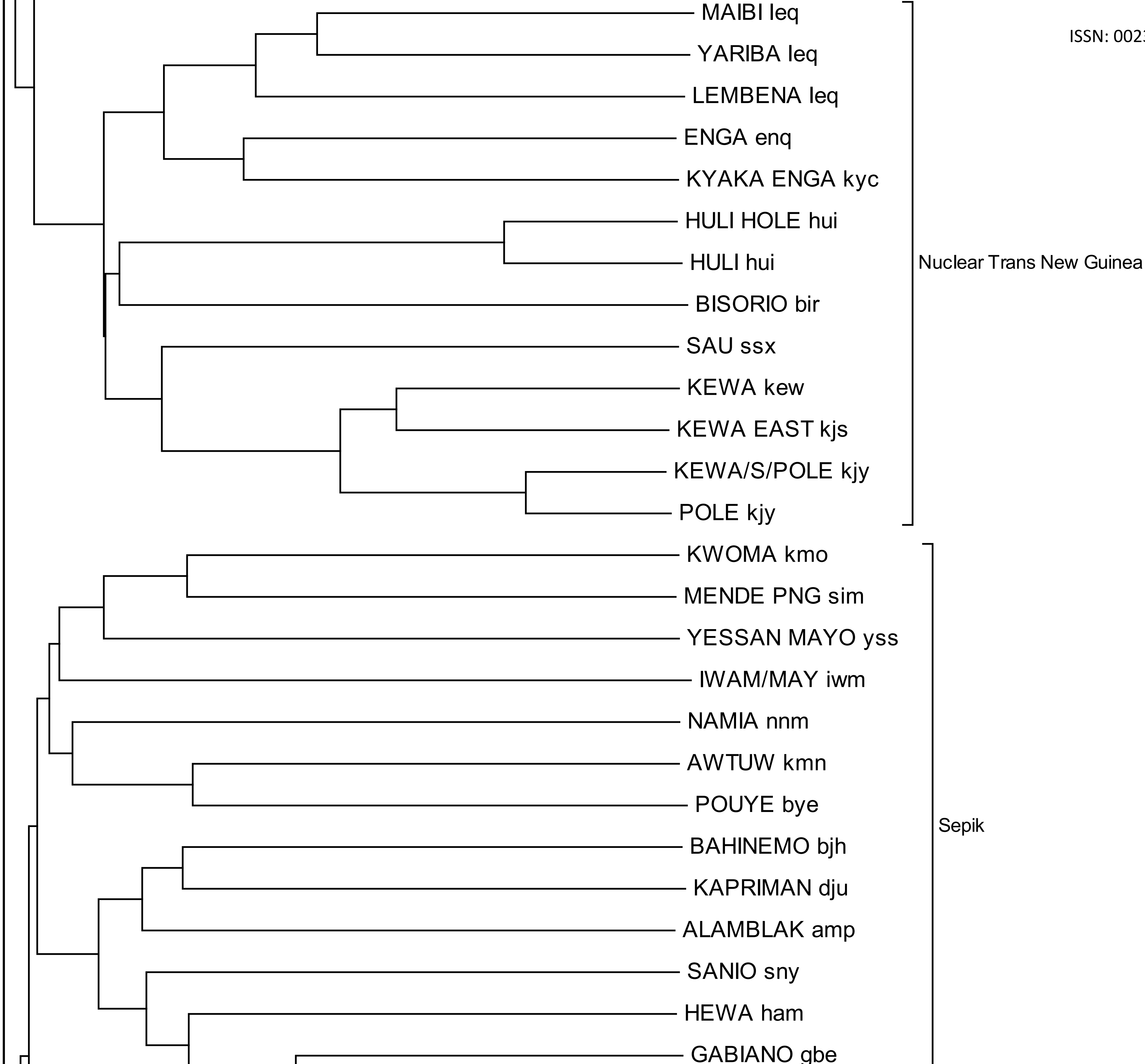
ISSN: 0023-1959



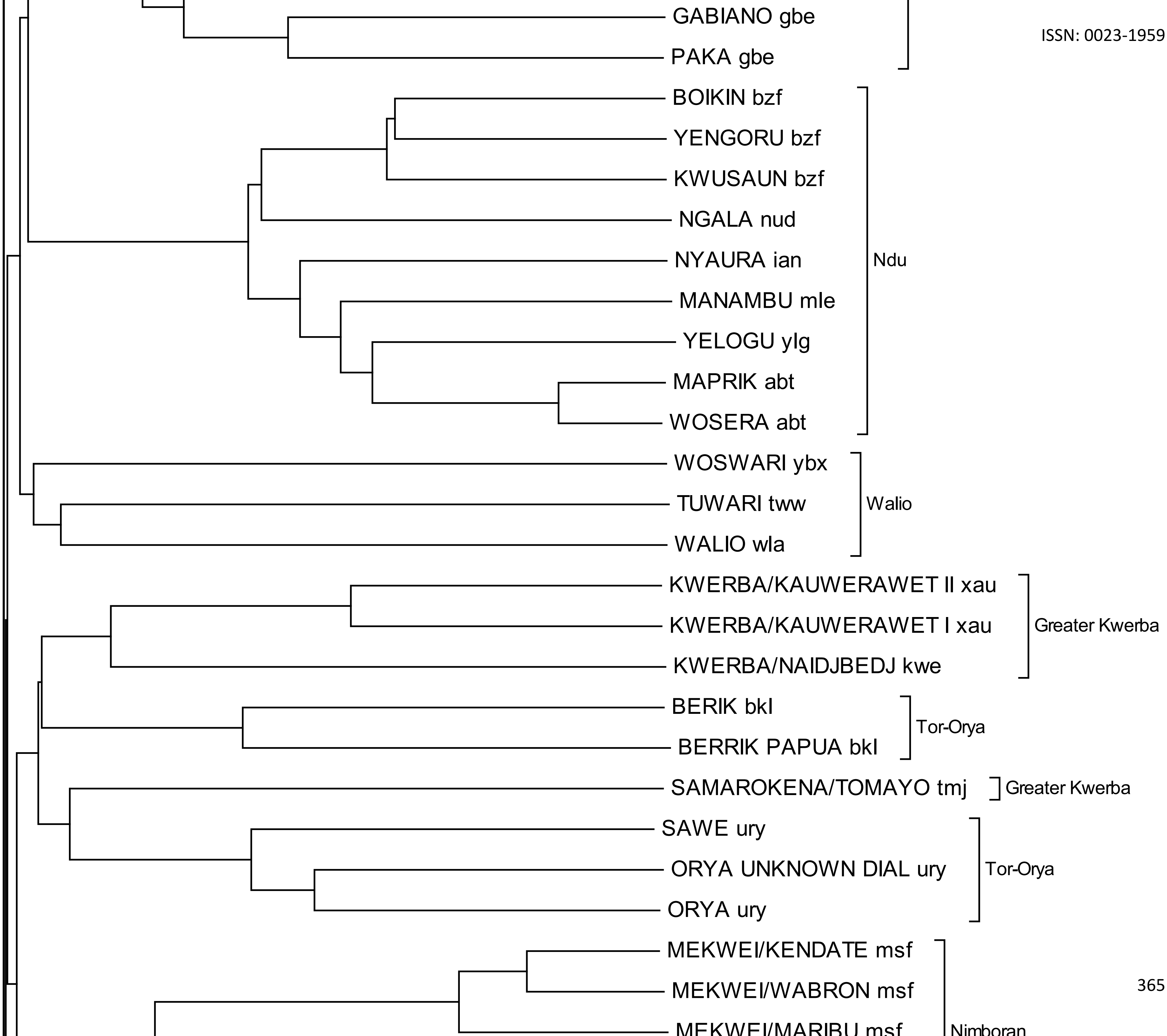
ISSN: 0023-1959



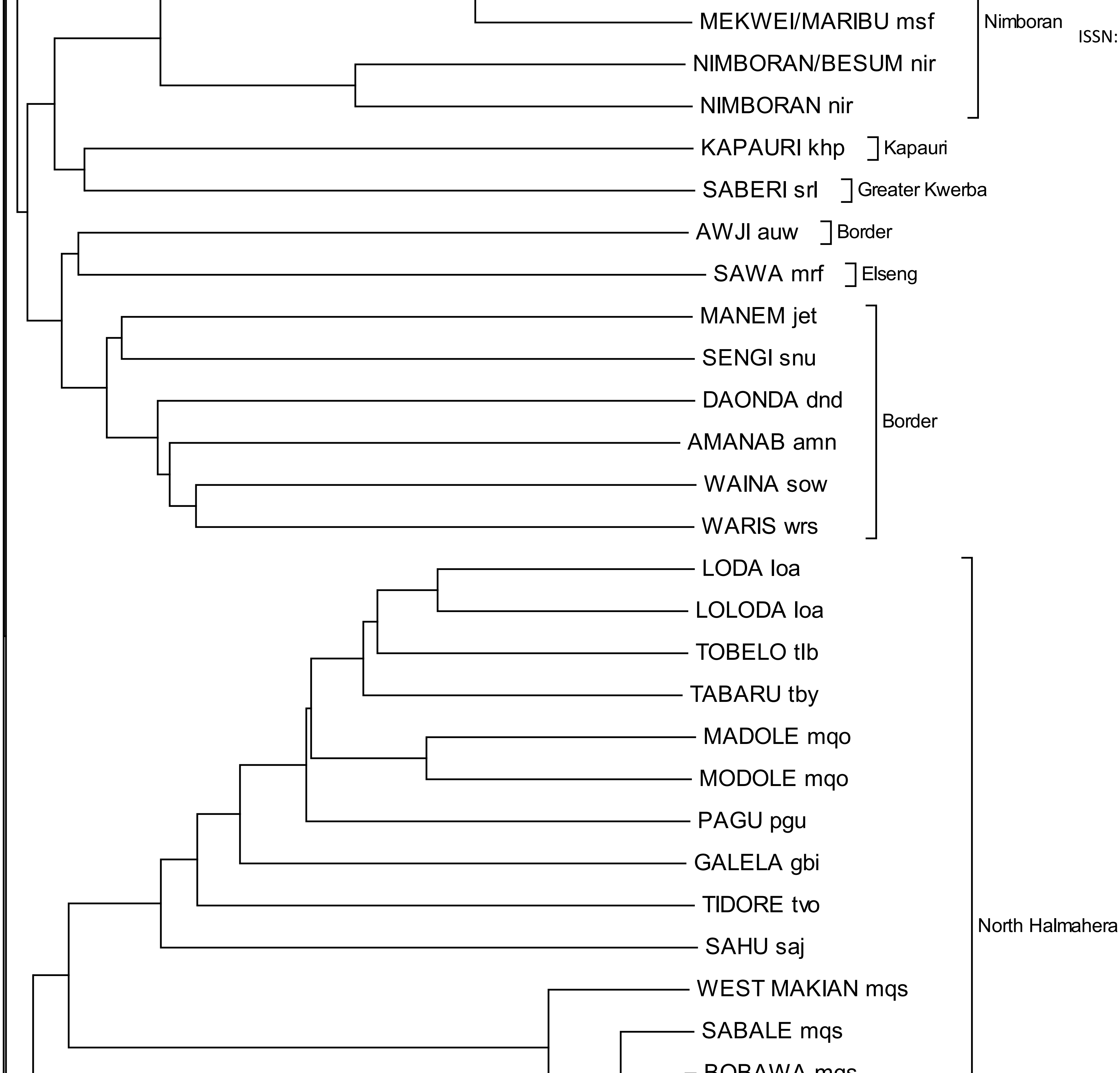
ISSN: 0023-1959



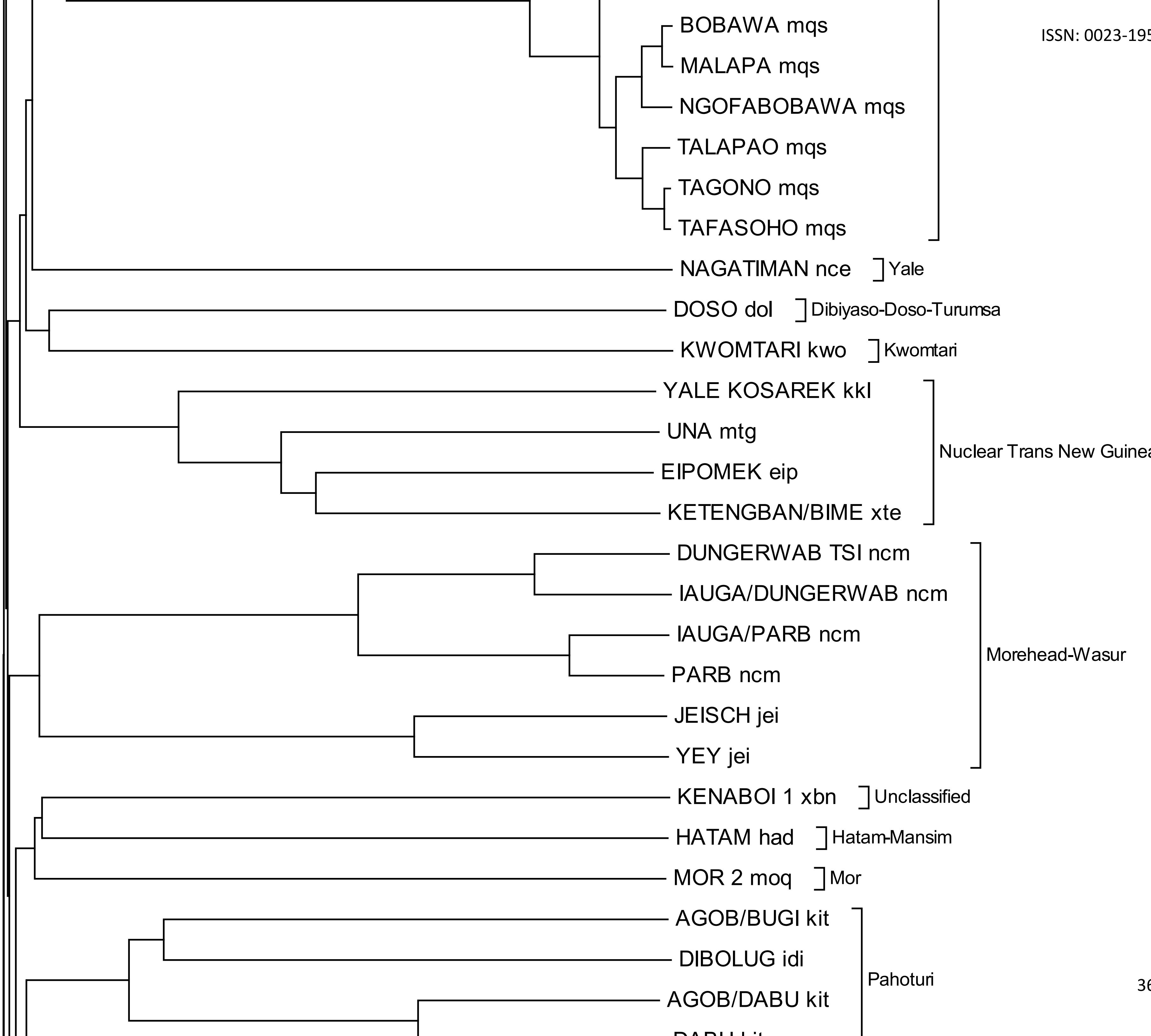
ISSN: 0023-1959



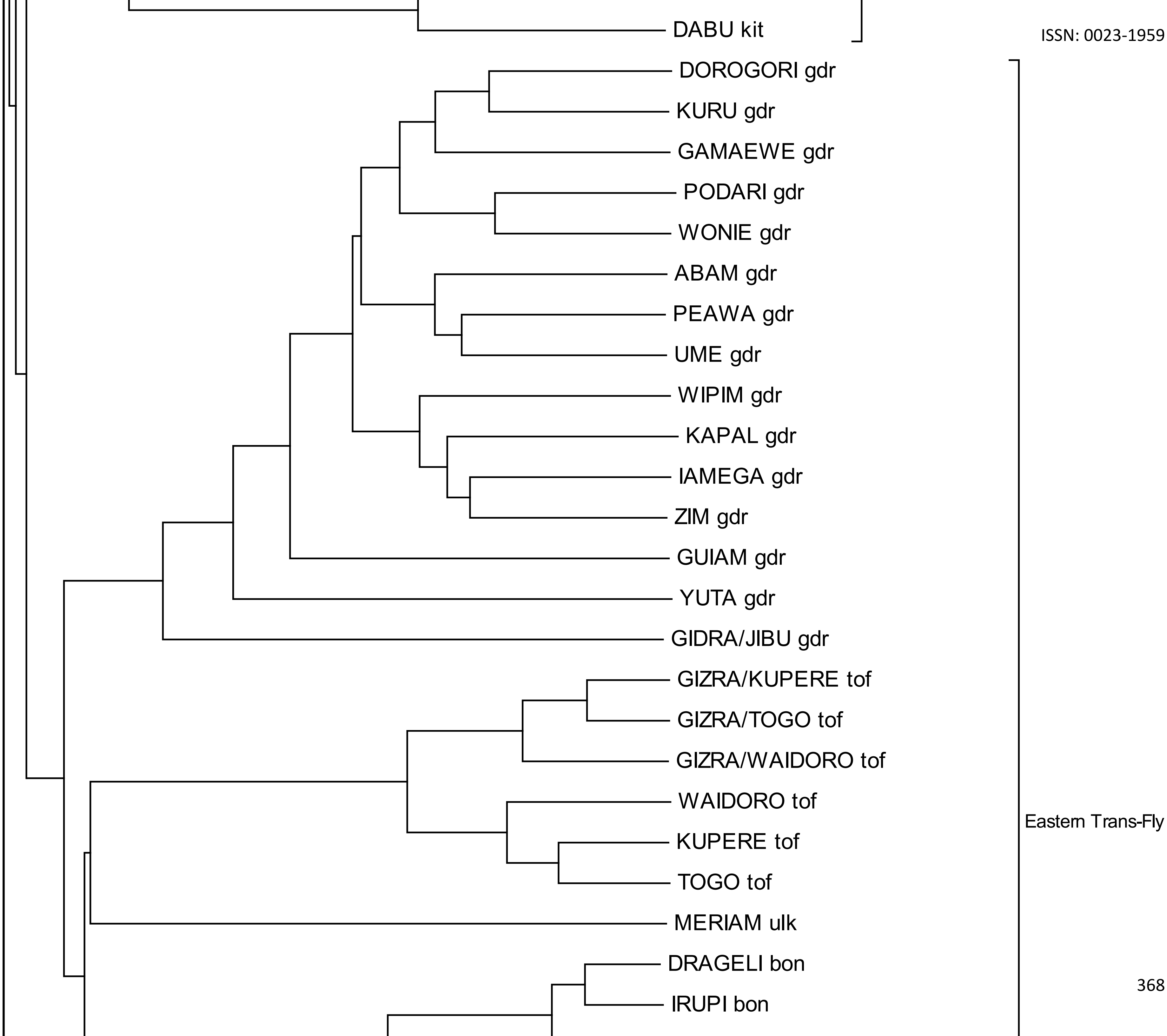
ISSN: 0023-1959



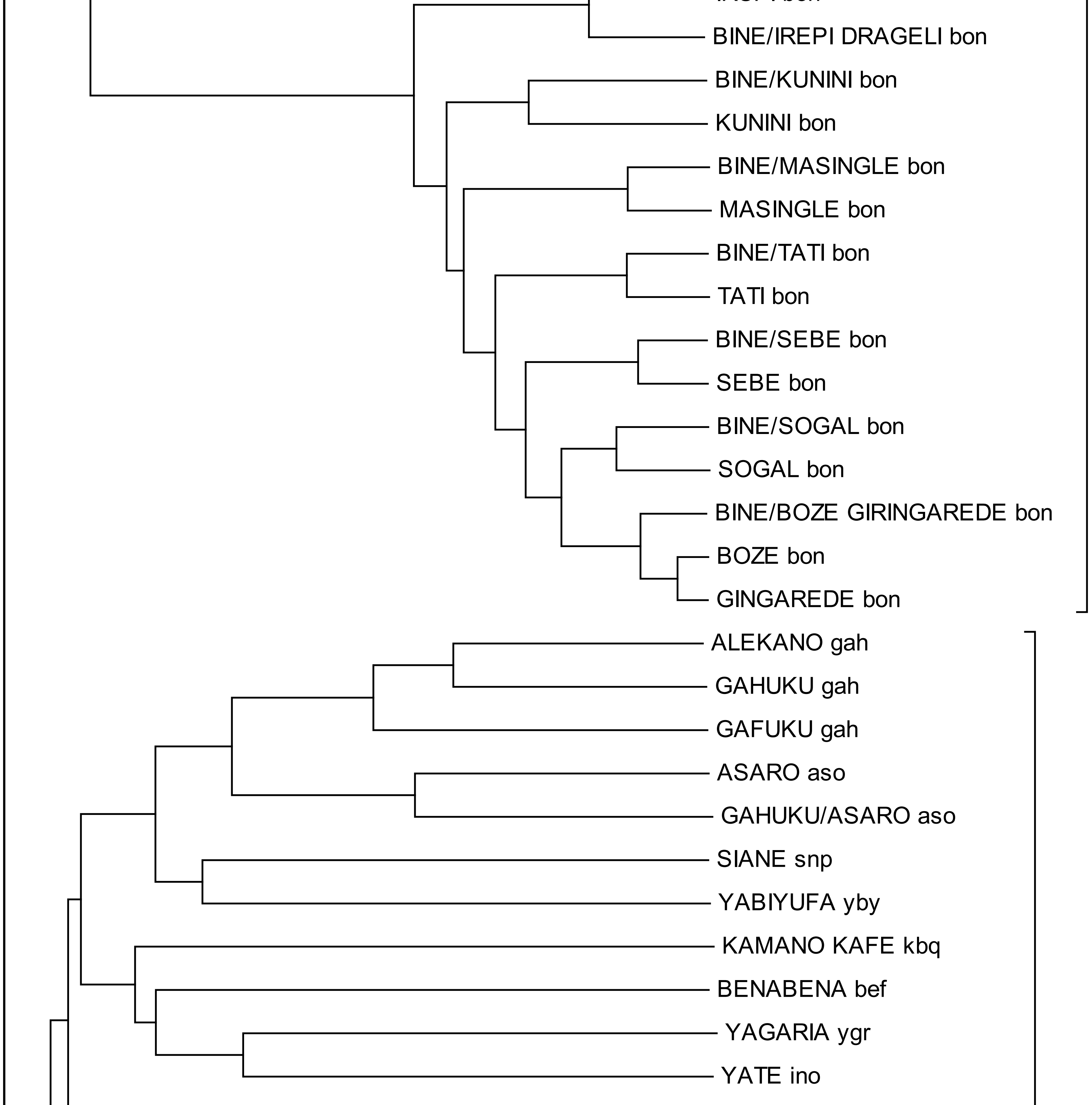
ISSN: 0023-1959



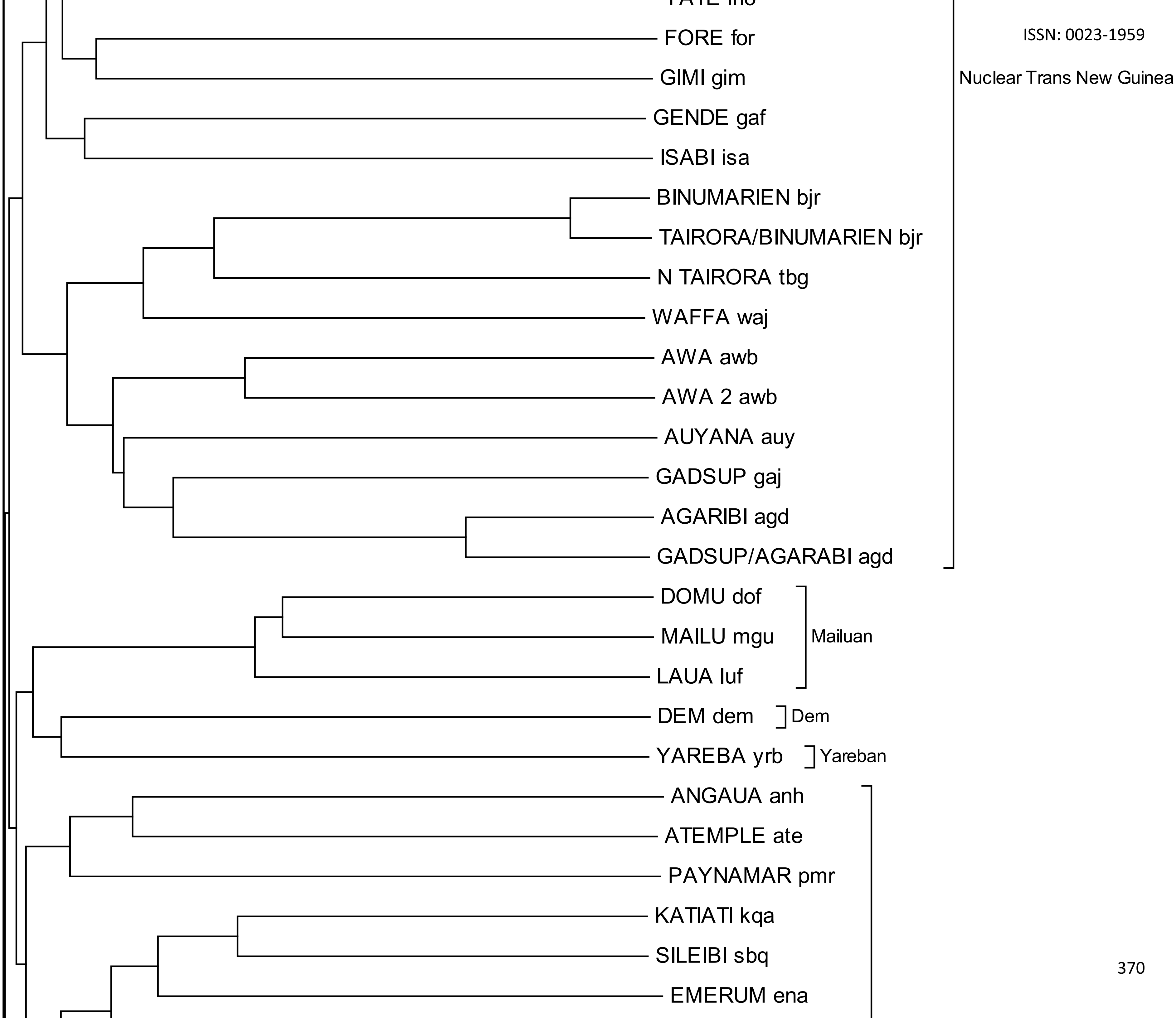
ISSN: 0023-1959



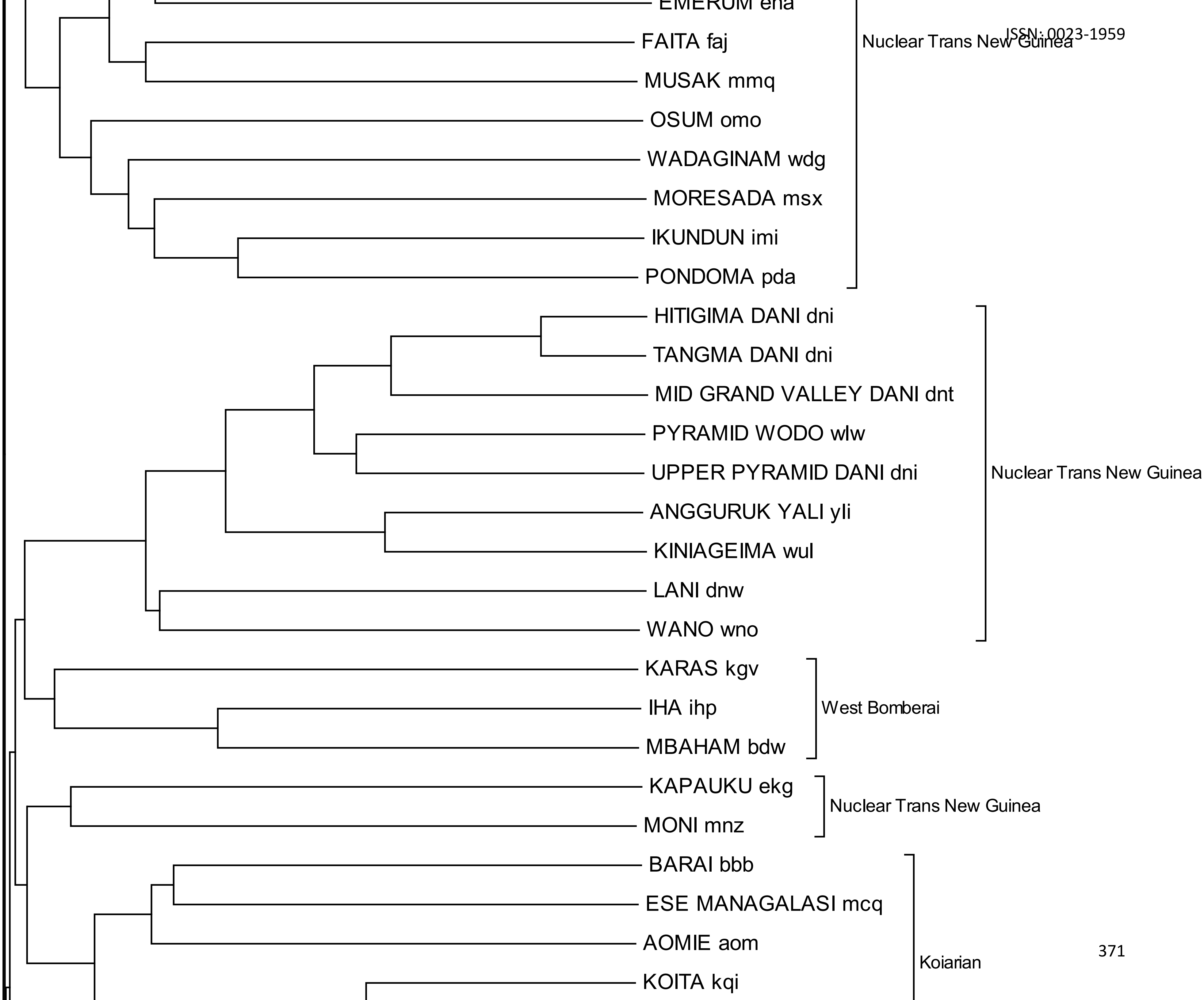
ISSN: 0023-1959



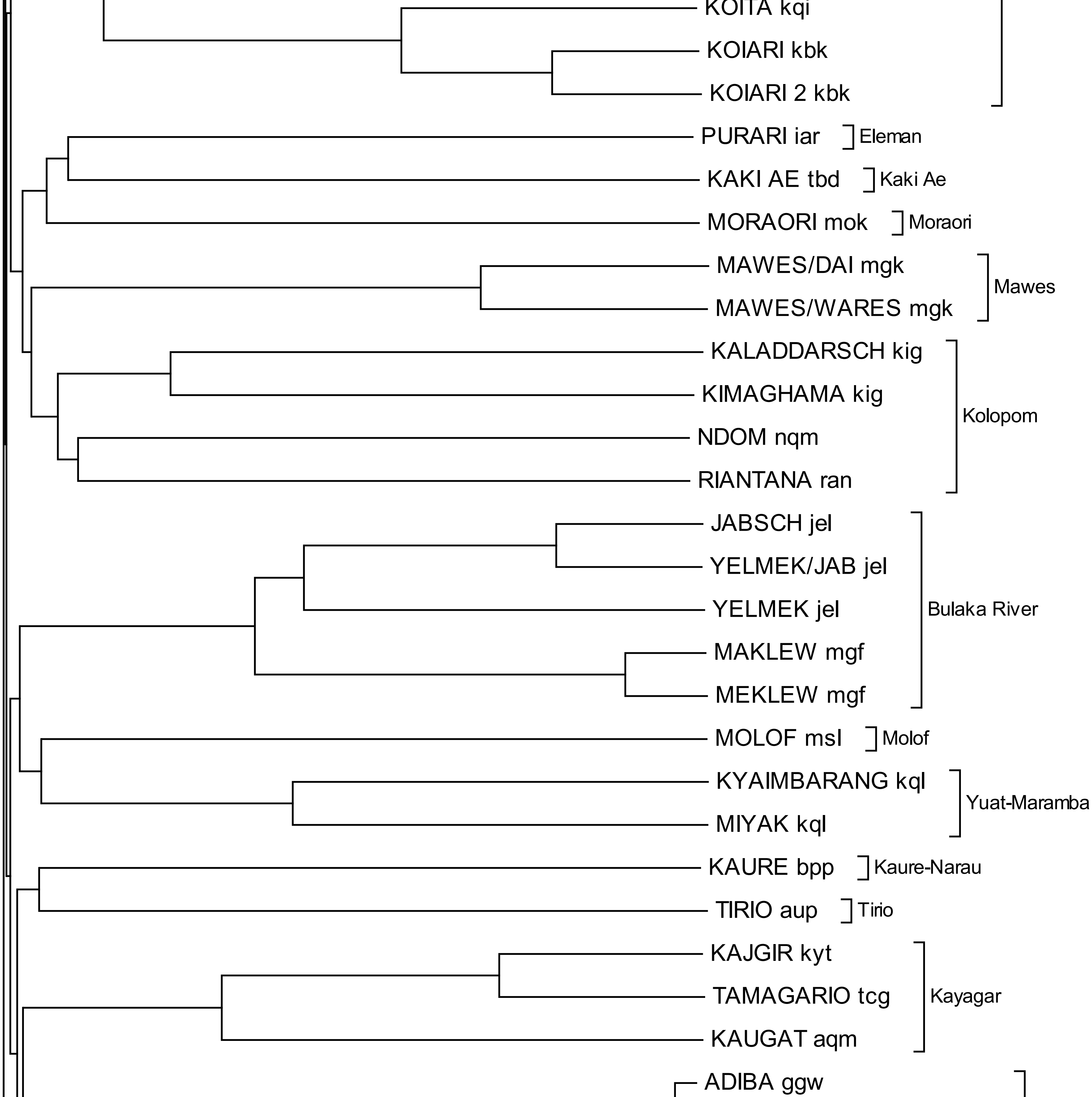
ISSN: 0023-1959



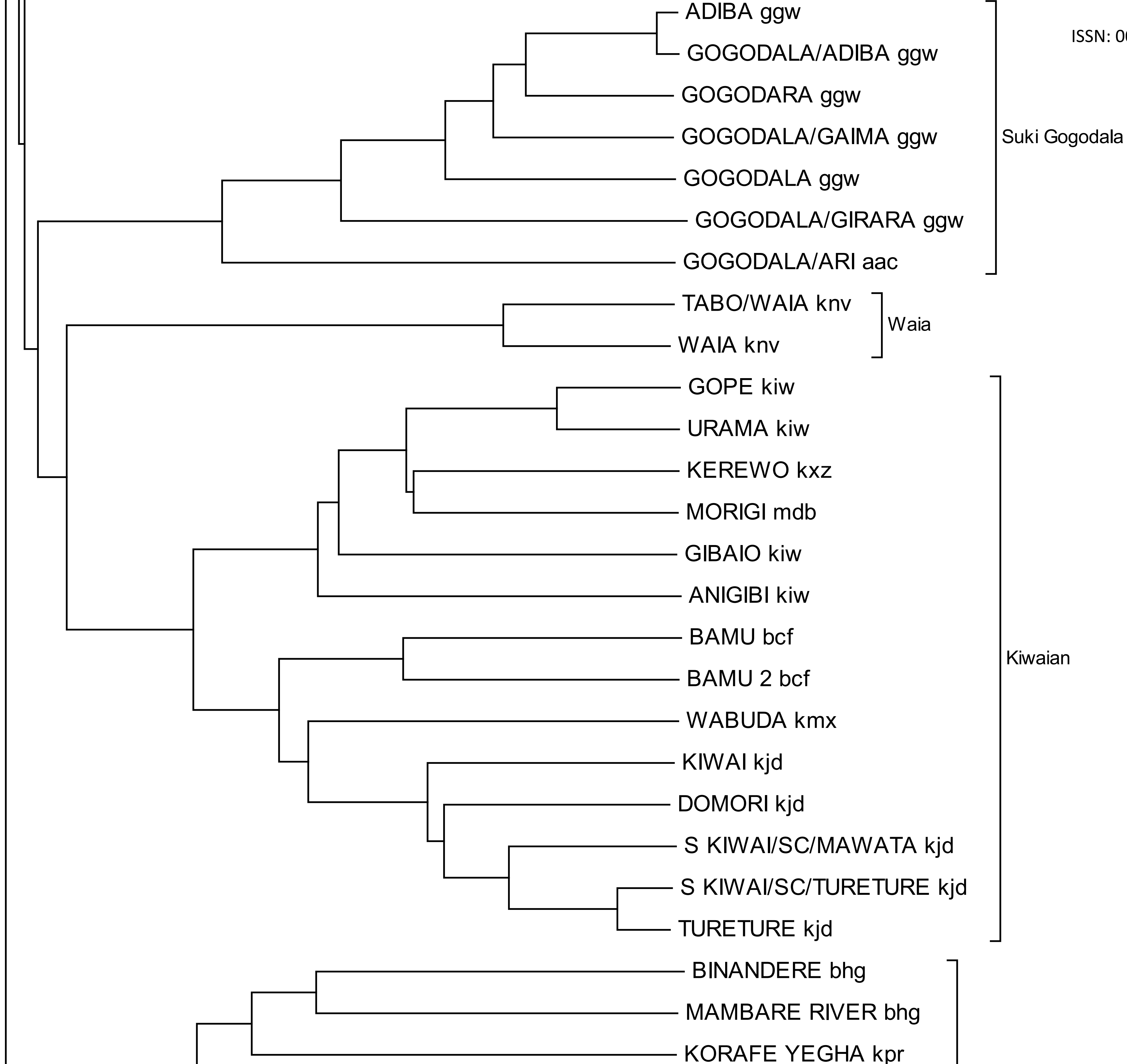
ISSN: 0023-1959



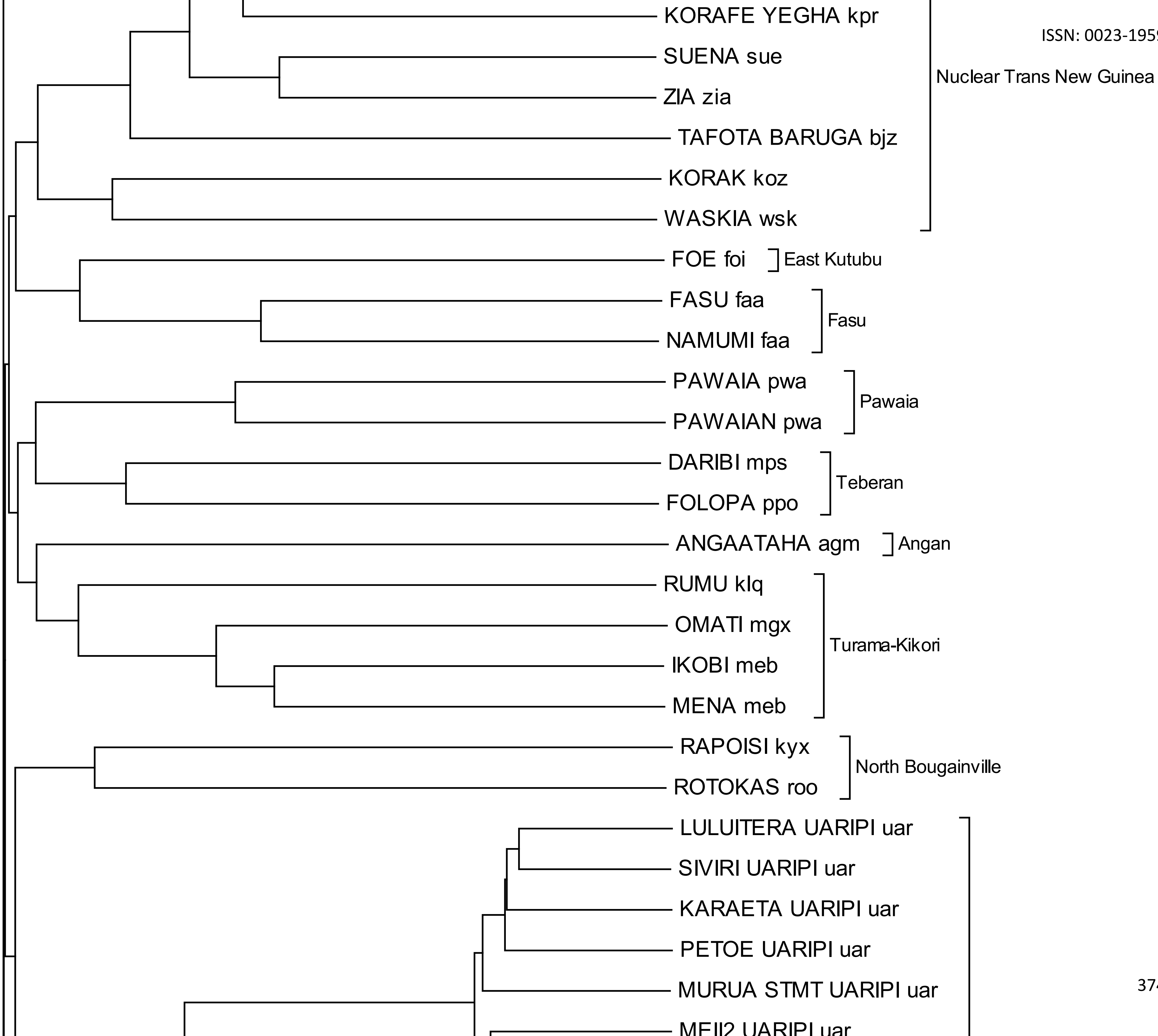
ISSN: 0023-1959



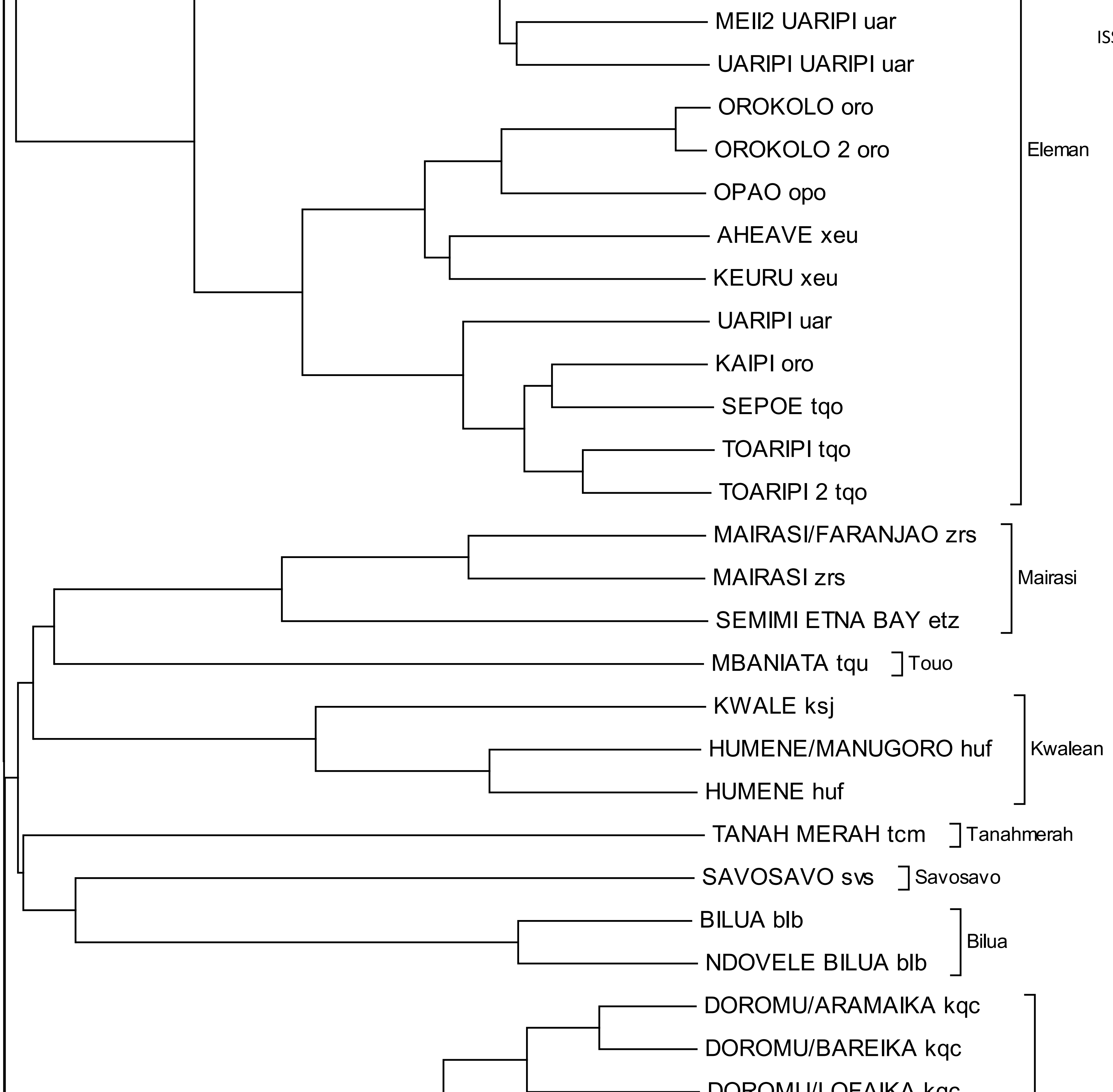
ISSN: 0023-1959



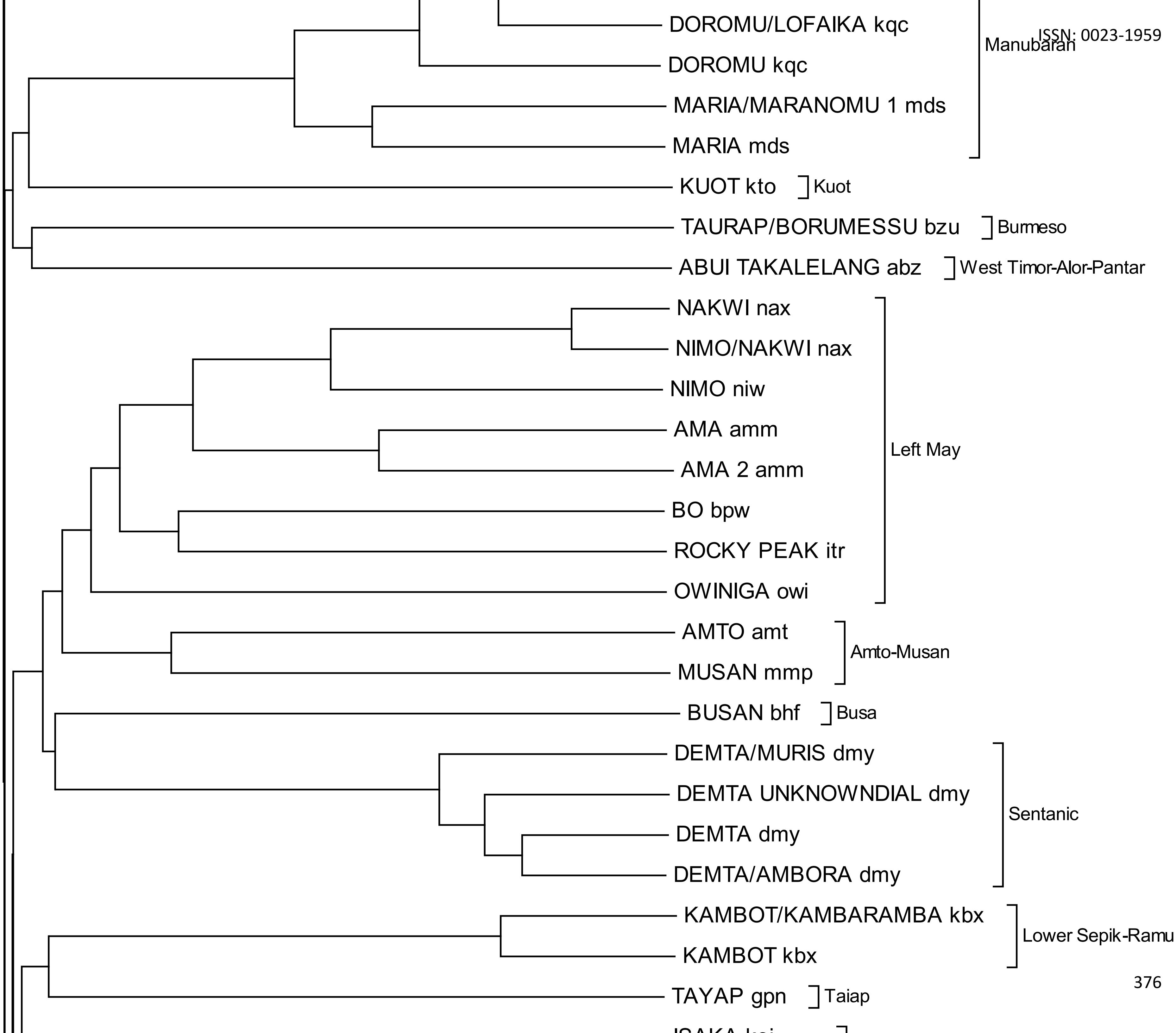
ISSN: 0023-1959



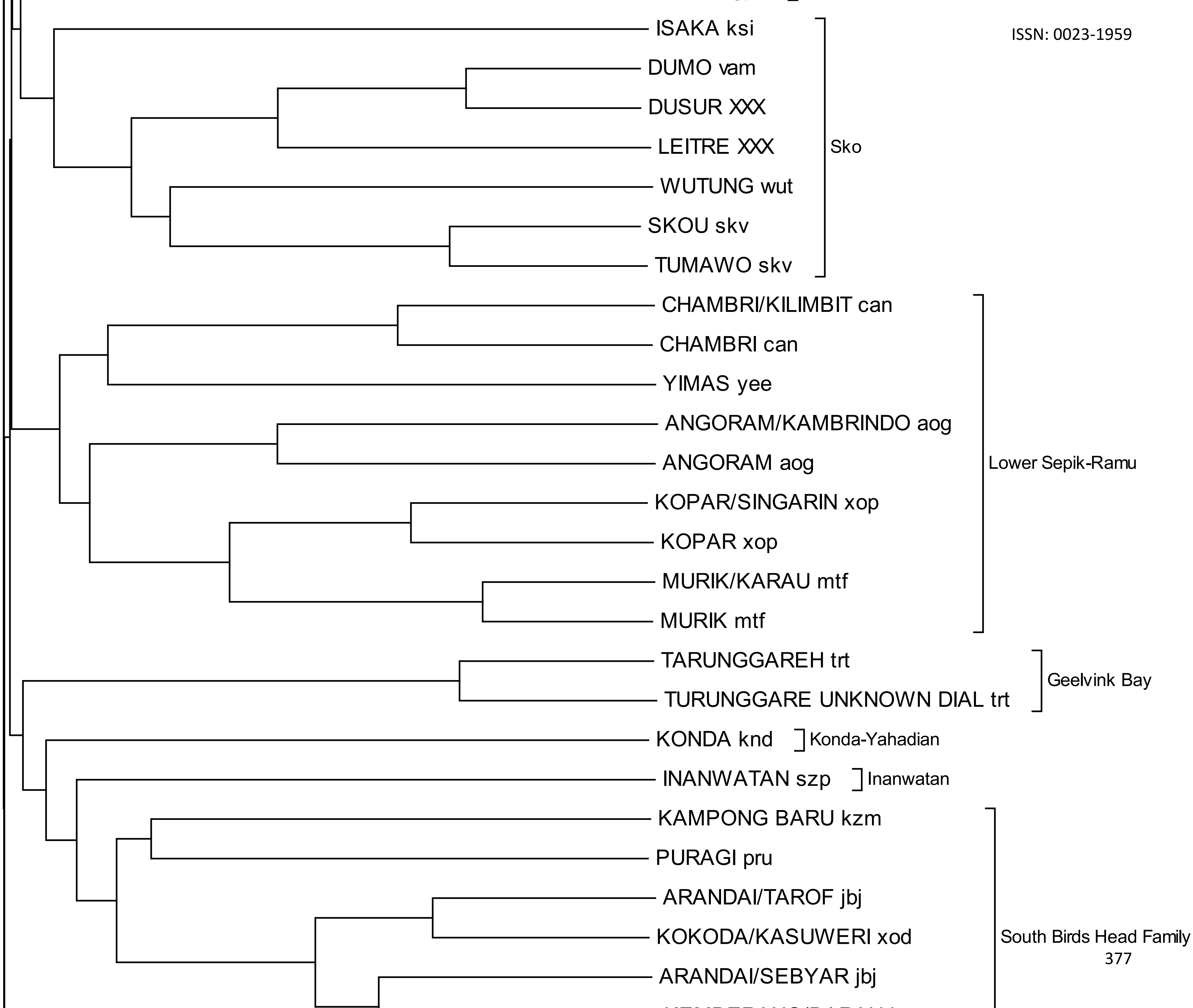
ISSN: 0023-1959



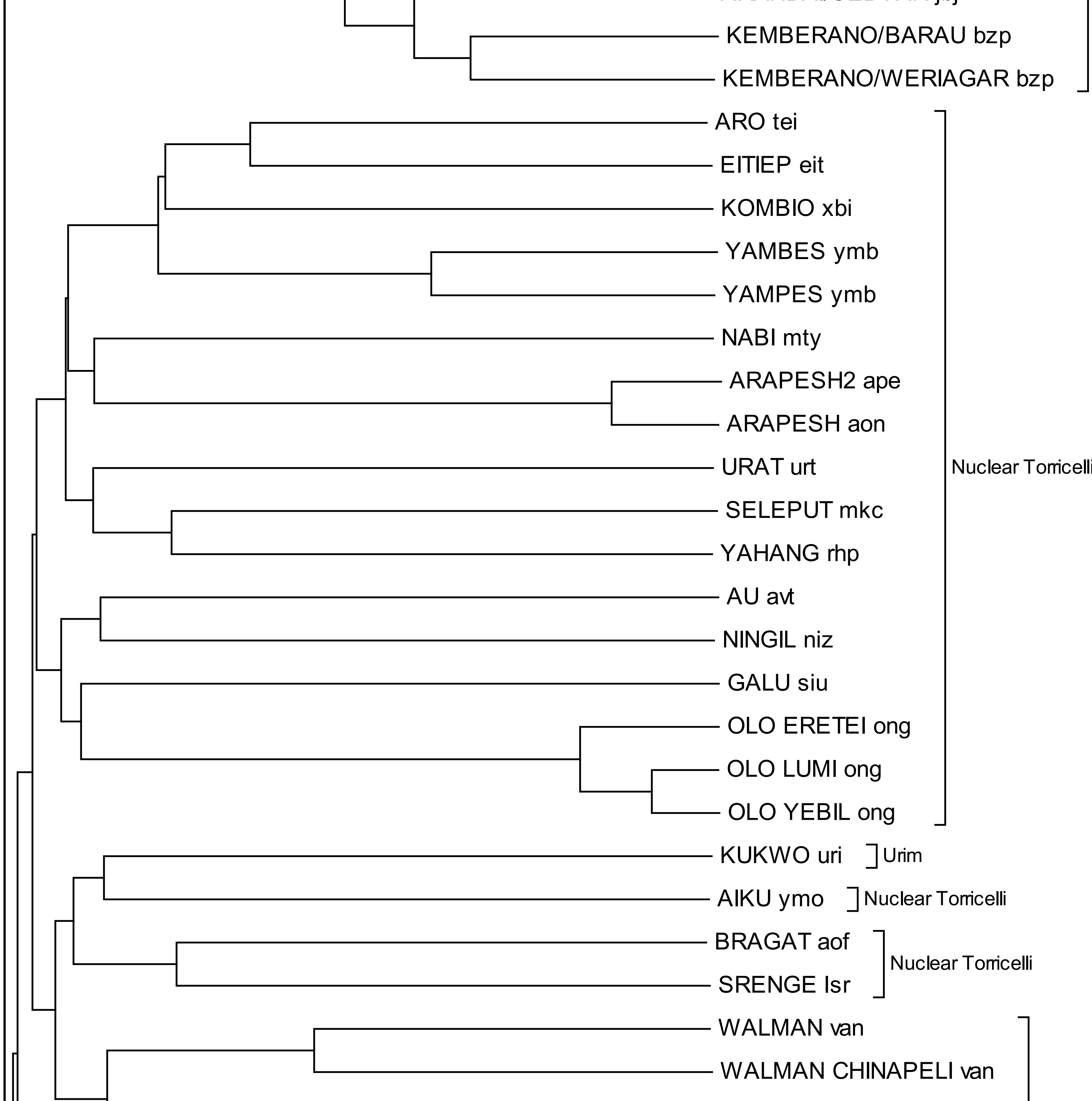
ISSN: 0023-1959



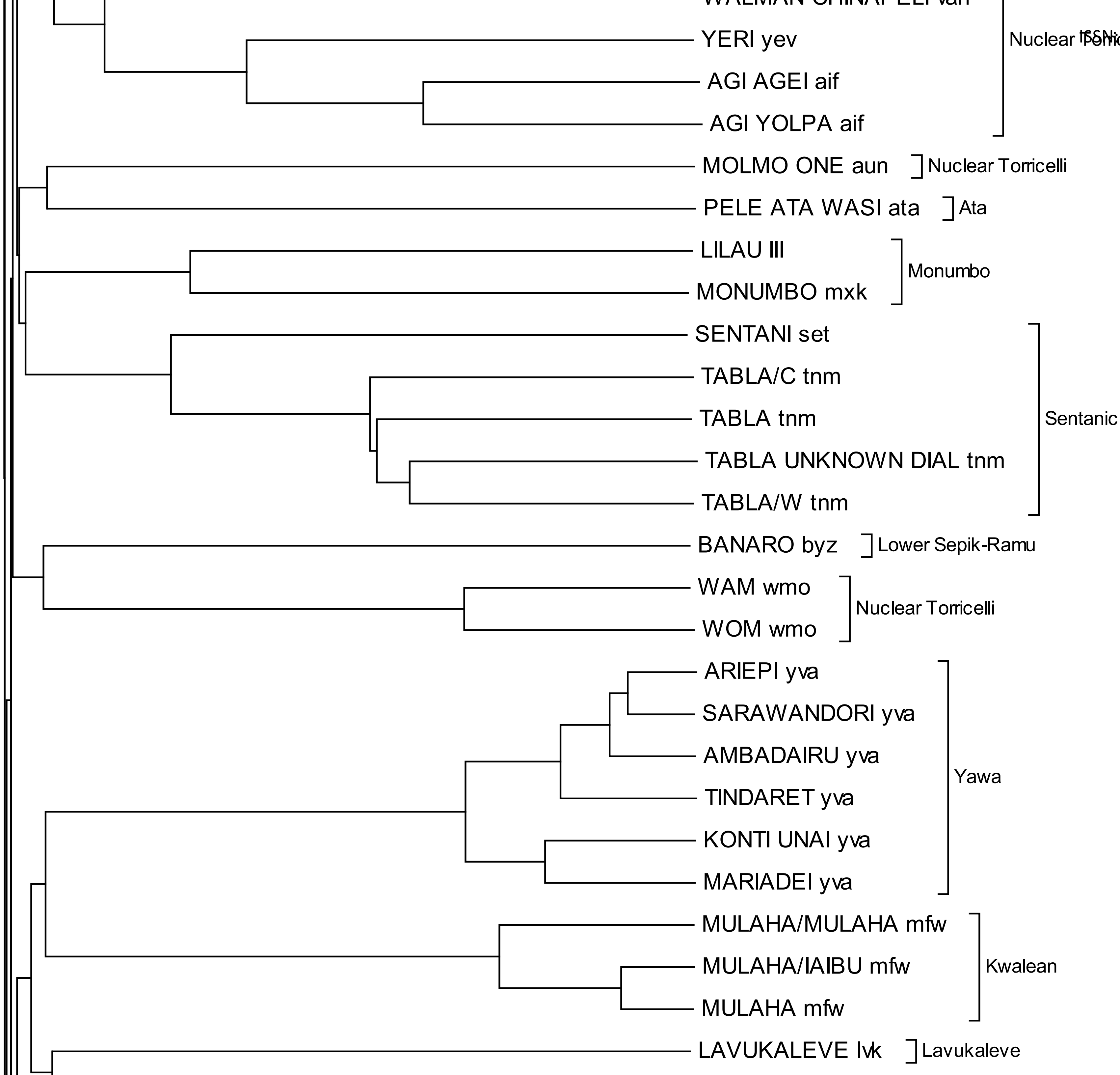
ISSN: 0023-1959



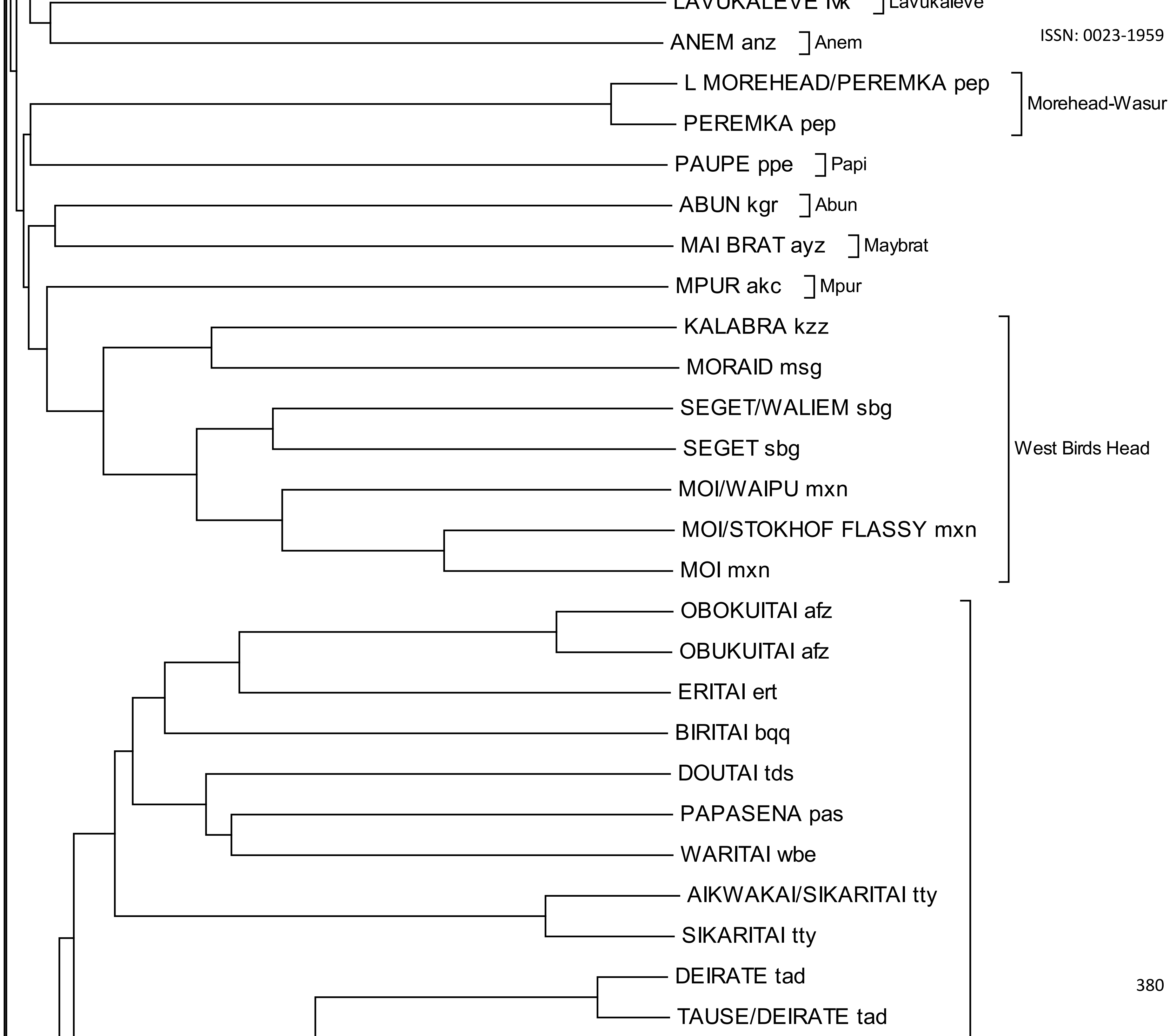
ISSN: 0023-1959



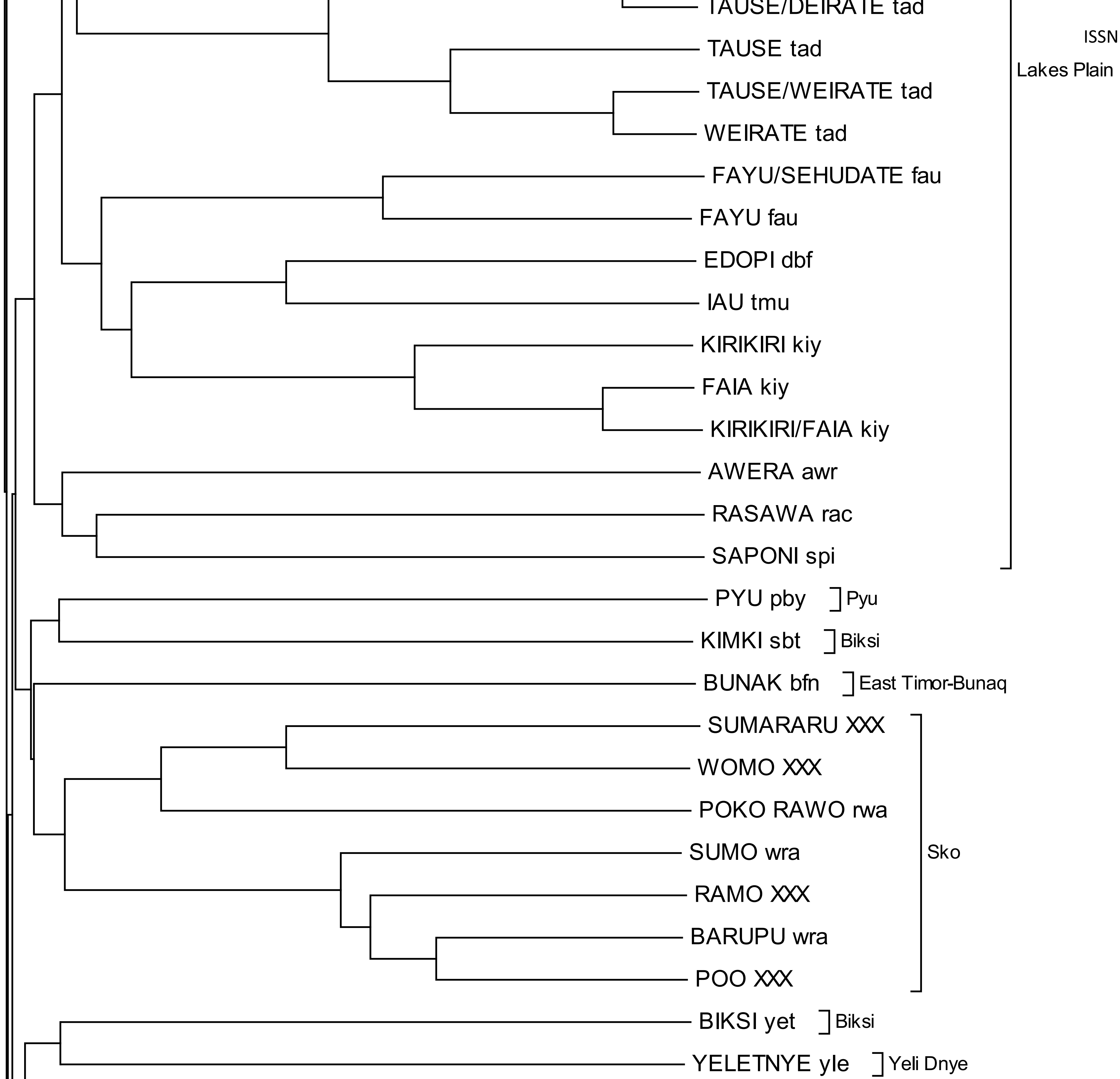
ISSN: 0023-1959



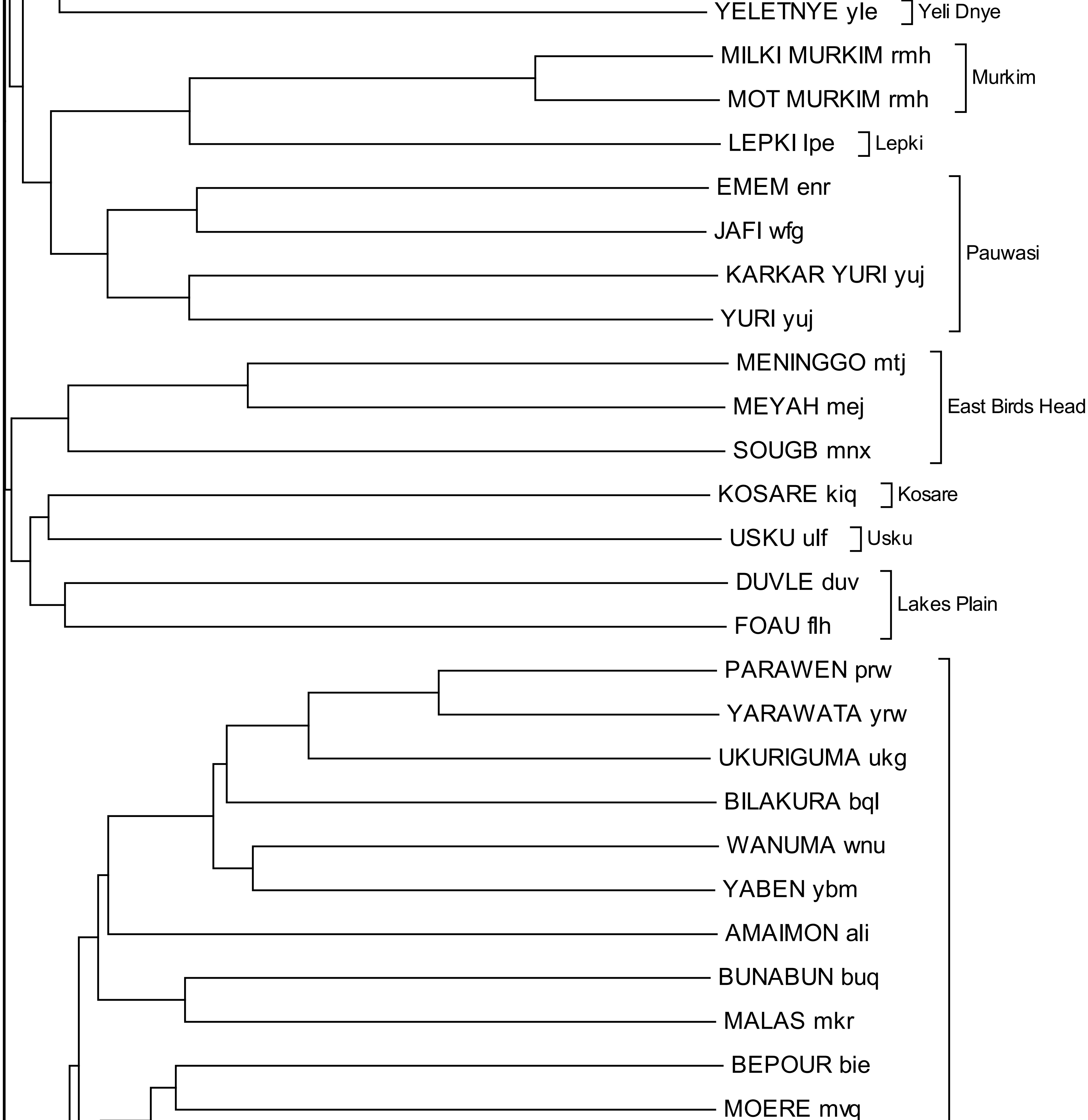
ISSN: 0023-1959



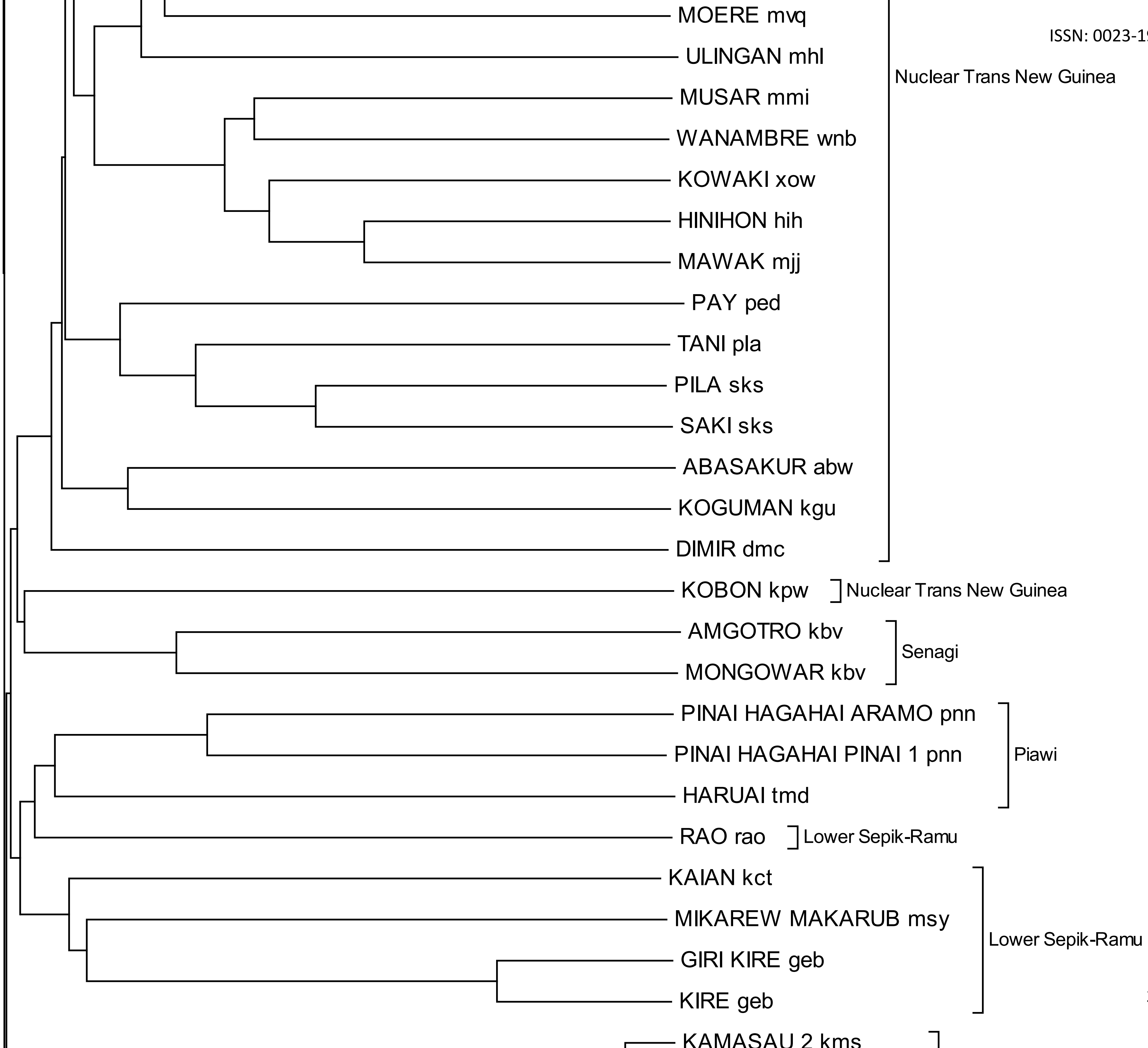
ISSN: 0023-1959



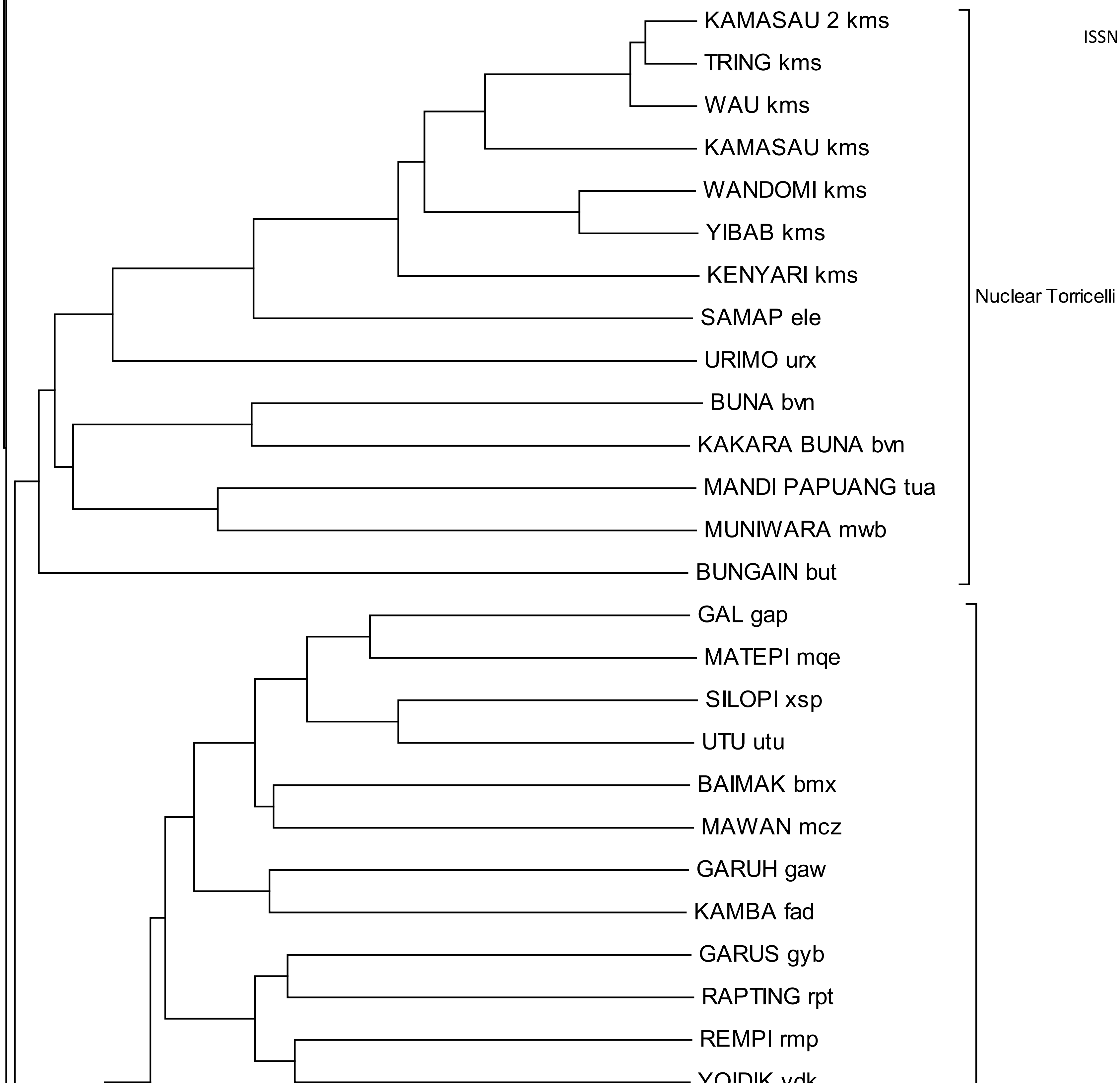
ISSN: 0023-1959



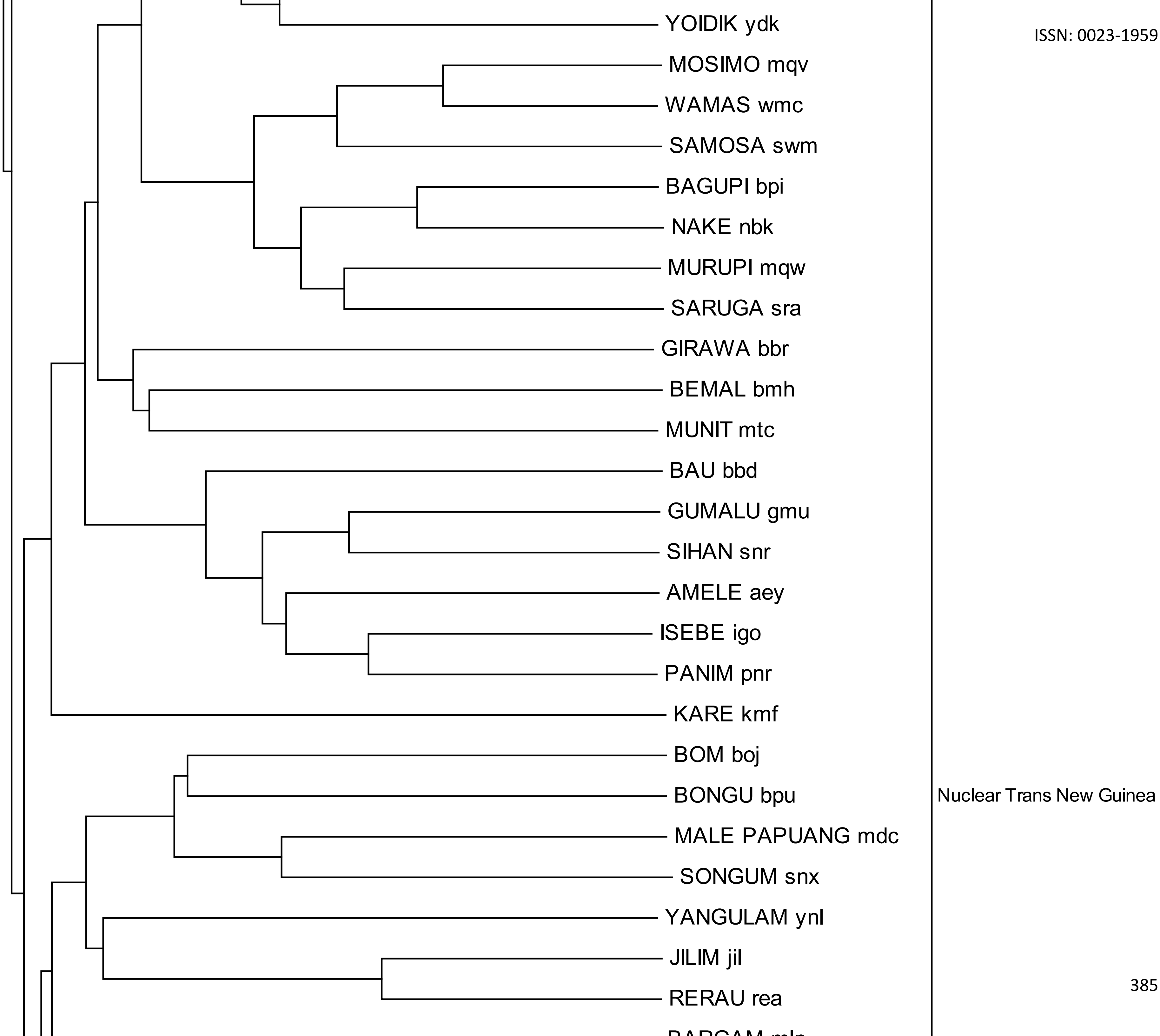
ISSN: 0023-1959



ISSN: 0023-1959



ISSN: 0023-1959



ISSN: 0023-1959