

ASJP World Language Tree: Version 1 (April 2009)

by

André Müller, Viveka Velupillai, Søren Wichmann, Cecil H. Brown, Pamela Brown, Eric W. Holman, Dik Bakker, Oleg Belyaev, Dmitri Egorov, Robert Mailhammer, Anthony Grant, and Kofi Yakpo

The World Language Tree graphically illustrates relative degrees of lexical similarity holding among 3384 of the world's languages and dialects (henceforth, languages) currently found in the ASJP database (ASJP stands for Automated Similarity Judgment Program). Languages branched more closely together on the tree are generally lexically more similar than languages branched less closely together.

The tree is generated through use of the neighbour-joining computer algorithm originally designed to depict phylogenetic relationships in biology (Saitou & Nei 1987). This is implemented in [MEGA 4](#) (Kumar et al. 2008), the software that we use. The algorithm is applied to a matrix of lexical similarity scores based on Levenshtein (or edit) distances holding between all possible pairs of the 3384 languages (for details about this, including how we modify the Levenshtein distances for our purposes cf. Bakker et al. 2009: 169). All languages of the database are compared to one another with respect to lexical similarity relating to their words for 40 referents determined statistically in Holman et al. (2008) to be most stable among core vocabulary items commonly used in lexicostatistical analysis. The tree is unrooted, but organized around the midpoint, i.e., the point which is equidistant between the two most distant languages in the network. Finally, the tree is annotated to show how it corresponds to the classification used in the latest version of the online [World Atlas of Language Structures](#) (Haspelmath et al. 2008), with some updates from Dryer (personal communication). This classification is presented for ease of orientation, not necessarily because ASJP agrees with it. The language names we use are normally simply those of the sources that we use. The sources, as well as corresponding language names of *Ethnologue*, are provided [here](#).

Four factors influence lexical similarity registered in the tree: (1) genetic or genealogical relationship of languages, (2) diffusion (language borrowing), (3) universal tendencies for lexical similarity such as onomatopoeia, and (4) random variation (chance).

Languages branched closely together on the tree may be so because of strong lexical similarity produced by any one or a combination of the four factors. Genetic relationship would appear to be the most dominant factor accounting for close branching, followed next by diffusion. Universal tendencies and chance are less significant contributors to close branching than either genetic relationship or diffusion, but nonetheless clearly contribute to the overall structure of the tree.

Typically, all languages of non-controversial language families such as Austro-Asiatic, Uralic, or Mayan, are respectively branched together on the tree. When some languages of a non-controversial family are not found branched together, this is because they are substantially lexically different from other members of their family despite unambiguously belonging to that family. Occasionally, a language can be so lexically different from other family members that it is found branched more closely with some language or languages with which it is not genetically related at all than with co-family members, usually because of chance lexical similarity or similarity due to borrowing. (When such languages are geographically remote from one another, chance usually explains close branching.)

Typically, branching accords closely with genetic subgroups recognized by experts within non-controversial language families. When branching is not isomorphic with genealogical subgrouping, this often reflects diffusion among languages of the family promoted by language contact. Thus, when used in conjunction with expert classifications of non-controversial language families, the tree can be helpful in calling attention to historical relationships (contact) among genetically related languages that sometimes might not be otherwise apparent.

The tree may also suggest relationships heretofore not noticed among languages that may be profitably investigated. For example, if two languages not known to be related in any way are found together on a terminal branch, this may indicate a relationship between them entailing either inheritance or contact, especially if they are not geographically remote from one another. If the two languages are geographically distant, their close lexical similarity is more likely explained by chance than by either inheritance or diffusion. On the other hand, isolates often join one another because they have nowhere else to go in the tree, creating the illusion that exciting, new far-flung relations may be established. One should be cautious in the interpretation of these cases. In general, the tree is only trustworthy with respect to genetic relationship at relatively shallow time-depths corresponding to uncontroversial or only moderately controversial genealogical units.

References

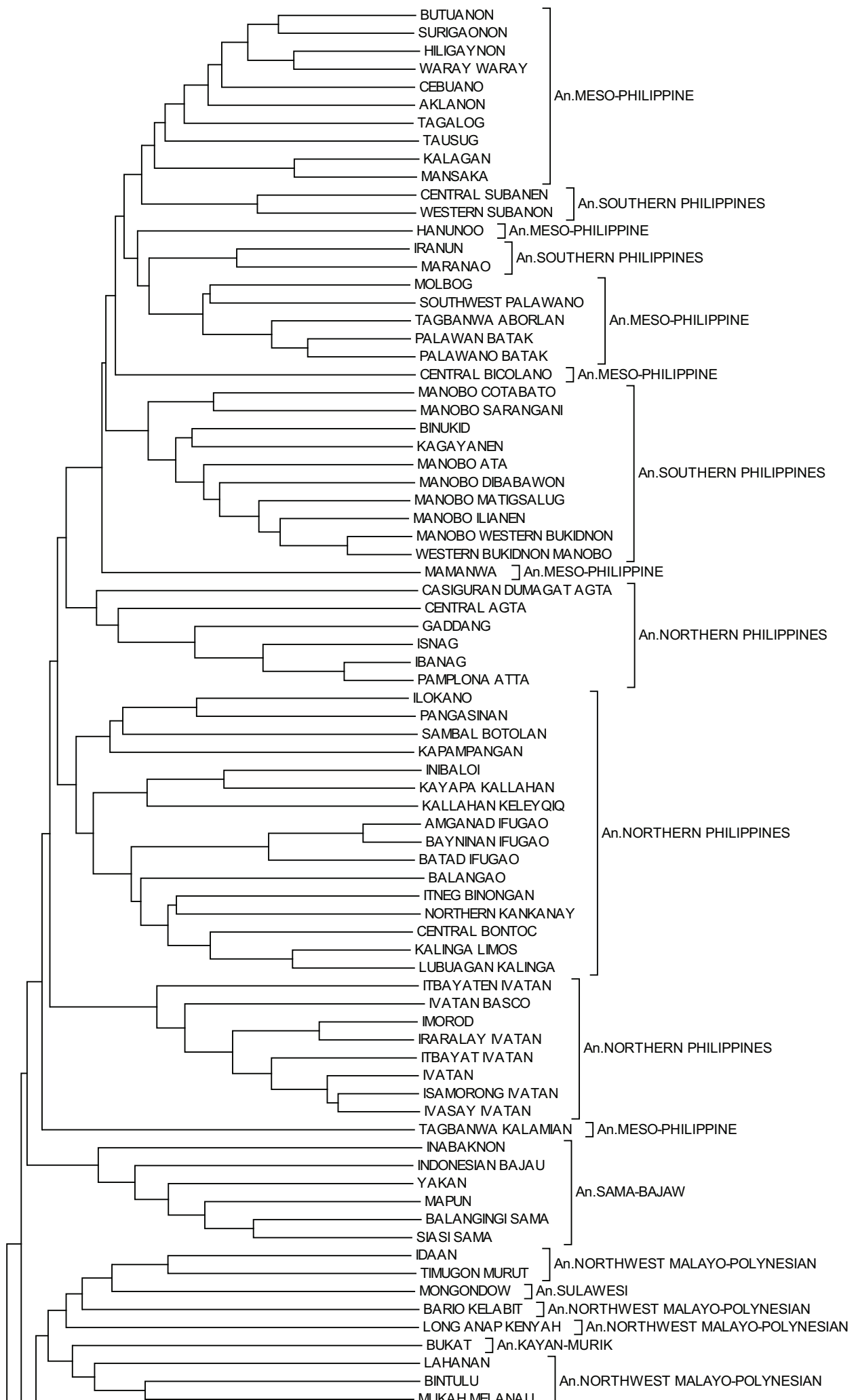
- Bakker, Dik, André Müller, Viveka Velupillai, Søren Wichmann, Cecil H. Brown, Pamela Brown, Dmitry Egorov, Robert Mailhammer, Anthony Grant, and Eric W. Holman. 2009. Adding typology to lexicostatistics: a combined approach to language classification. *Linguistic Typology* 13: 167-179.
- Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, and Dik Bakker. 2008. Explorations in automated lexicostatistics. *Folia Linguistica* 42.2: 331-354.
- Haspelmath, Martin, Matthew S. Dryer, David Gil and Bernard Comrie. 2008. *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library.
- Kumar S., J. Dudley, M. Nei, and K. Tamura K. 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* 9: 299-306.
- Saitou, Naruya and Masatoshi Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425.

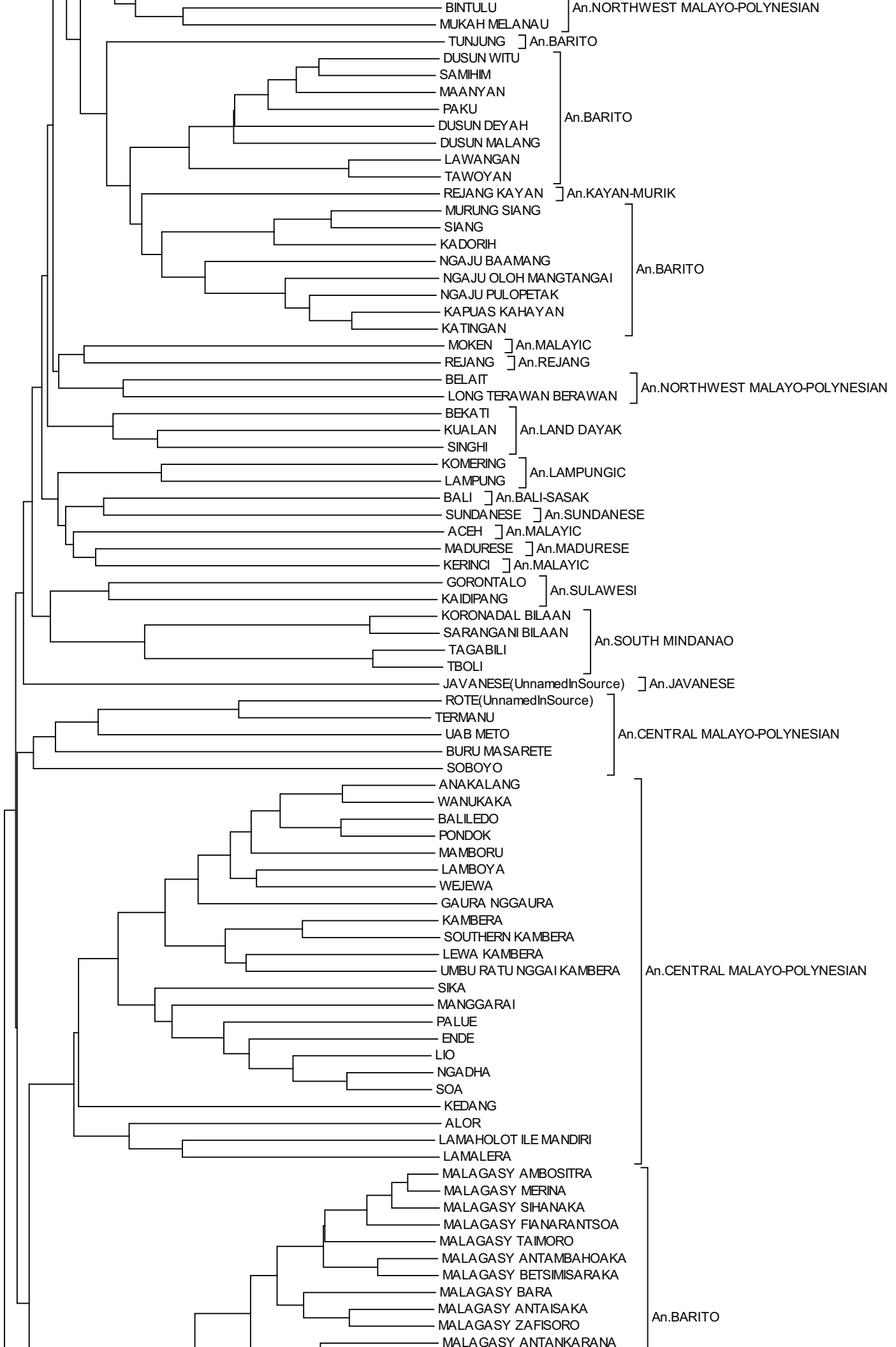
Language family abbreviations

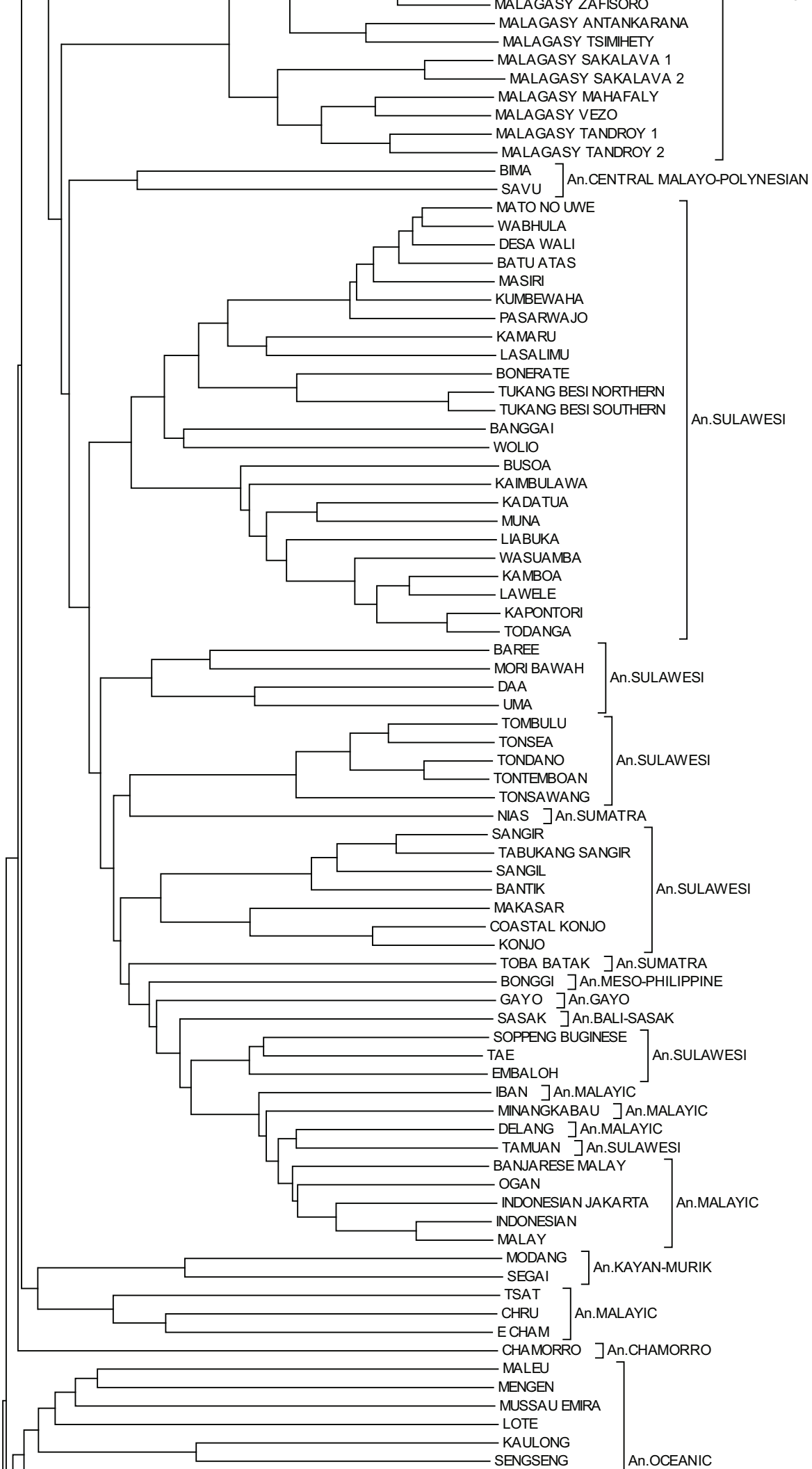
AA	Afro-Asiatic	Chu	Chumash
Aik	Aikana	CK	Chukotko-Kamchatkan
Ain	Ainu	Cmu	Chimúan
Ala	Alacalufan	CN	Cacua-Nukak
Alg	Algic	Cnd	Candoshi
Alt	Altaic	Cof	Cofán
AM	Amto-Musan	Com	Comecrudan
An	Austronesian	Cre	Creoles&Pidgins
AP	Awin-Pare	Cui	Cuitlatec
Arc	Araucanian	CW	Chapacura-Wanhan
Art	Arutani	Dos	Doso
Aru	Arauan	Dra	Dravidian
Arw	Arawakan	EA	Eskimo-Aleut
Ata	Atakapa	EB	East Bougainville
AuA	Austro-Asiatic	EBH	East Bird's Head
Aus	Australian	EGB	East Geelvink Bay
Aym	Aymaran	Ela	Elamite
Ban	Bangi Me	Ele	Eleman
Bar	Barbacoan	ES	East Strickland
Bas	Basque	GA	Great Andamanese
Beo	Beothuk	Gcu	Guaicuruan
Bil	Bilua	GS	Gogodala-Suki
Bor	Border	Gua	Guahiban
Bos	Bosavi	Had	Hadza
Brs	Burushaski	Hai	Haida
Bul	Bulaka River	Har	Harakmbet
Bur	Burmeso	HM	Hmong-Mien
Cad	Caddoan	Hok	Hokan
Cah	Cahuapanan	Hua	Huavean
Cam	Camsá	Hui	Huitotoan
Car	Cariban	IE	Indo-European
Cay	Cayuvava	IG	Inland Gulf
Chi	Chibchan	Ira	Irantxe
Chm	Chimila	Iro	Iroquoian
Chn	Chon	Ito	Itonama
Cho	Choco	Jab	Jabuti
Cht	Chitimacha	Jap	Japanese

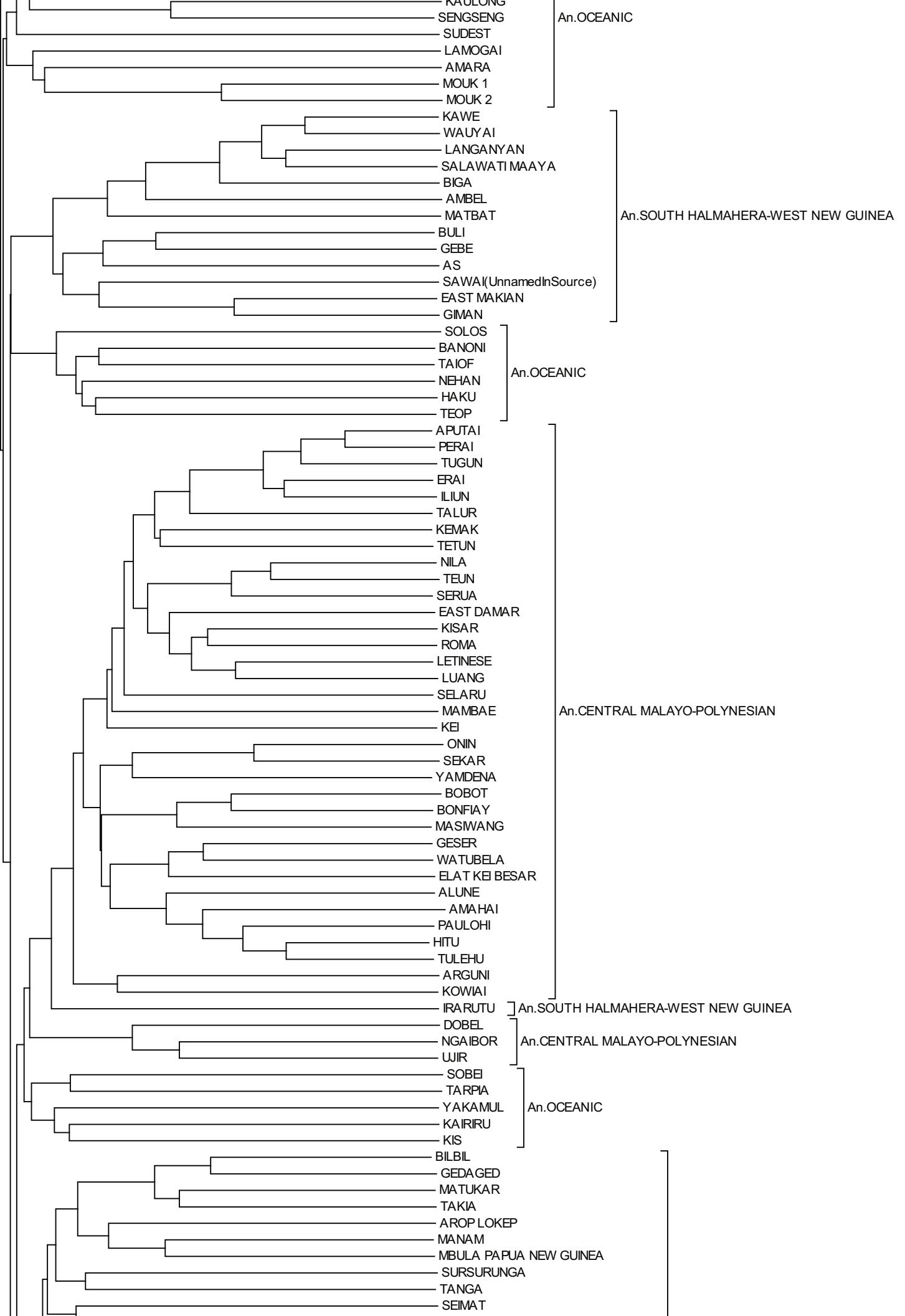
Jiv	Jivaroan	Mos	Mosetenan
Kad	Kadugli	Mov	Movima
Kap	Kapixana	Mrw	Morwap
Kar	Karok	MUM	Morehead and Upper Maro Rivers
Kat	Katukinan	Mur	Mura
Kau	Kaure	Mus	Muskogean
Kay	Kayagar	MZ	Mixe-Zoque
Kaz	Kazukuru	Nam	Nambikuaran
KF	Kwomtari-Fas	Nat	Natchez
Kho	Khoisan	NC	Niger-Congo
Kib	Kibiri	NDa	Nakh-Daghestanian
Kiw	Kiwaian	NDe	Na-Dene
Kol	Kolopom	Nih	Nihali
Kor	Korean	Nim	Nimboran
Krt	Kartvelian	NS	Nilo-Saharan
KT	Kiowa Tanoan	NWC	Northwest Caucasian
Ktn	Kutenai	OC	Oregon Coast
Kun	Kunza	Odi	Odiai
Kus	Kusunda	Oks	Oksapmin
Kut	Kuto	OM	Oto-Manguean
Kwa	Kwalean	Pae	Paezan
Kwe	Kwerba	Pan	Panoan
Kwz	Kwaza	Pat	Pataxo
KY	Karkar-Yuri	Pau	Pauwasi
Lav	Lavukaleve	Pen	Penutian
LeM	Left May	Pui	Puinave
Len	Lencan	PY	Peba-Yaguan
LP	Lakes Plain	Que	Quechuan
LS	Leonhard Schultze	Sal	Salishan
LSR	Lower Sepik-Ramu	SAn	South Andamanese
Mar	Marind	Sav	Savosavo
Mas	Mascoian	Sen	Senagi
Mat	Matacoan	Sep	Sepik
May	Mayan	Sho	Shom Peng
MGe	Macro-Ge	Sio	Siouan
Mis	Misumalpan	Sko	Sko
Mom	Mombum	Sln	Salinan
Mon	Monumbo	Slv	Sáliban
Mor	Mor	Snt	Sentani

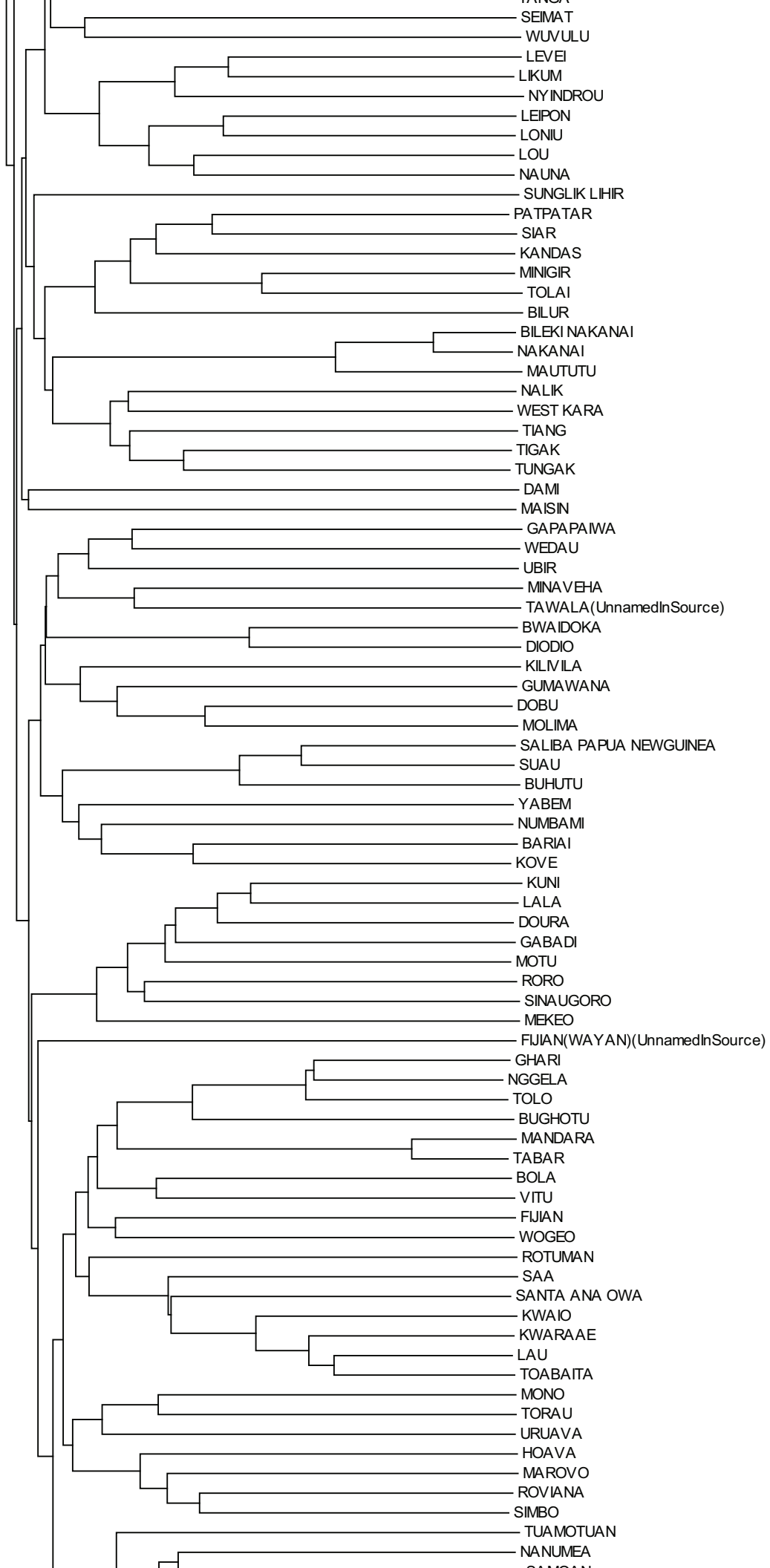
ST	Sino-Tibetan	Yal	Yale
Sum	Sumerian	Yam	Yamana
Tac	Tacanan	Yan	Yanomam
Tar	Tarascan	Yaw	Yawa
Tau	Taushiro	Yel	Yele
Teb	Teberan-Pawaiian	Yen	Yeniseian
Teq	Tequistlatecan	Yka	Yukaghir
Tic	Ticuna	Yrb	Yareban
Tim	Timucua	Yrr	Yaruro
TK	Tai-Kadai	Yuc	Yuchi
TNG	Trans-New Guinea	Yuw	Yuwana
TO	Tor-Orya	Zam	Zamucoan
Tol	Tol	Zap	Zaparoan
Ton	Tonkawa	Zun	Zuni
Tor	Torricelli		
Tot	Totonacan		
Tou	Touo		
Tru	Trumai		
Tuc	Tucanoan		
TuK	Turama-Kikorian		
Tup	Tupian		
UA	Uto-Aztecan		
UC	Uru-Chipaya		
Ura	Uralic		
Urr	Urarina		
Usk	Usku		
UY	Upper Yuat		
VJ	Vaupés-Japurá		
Wak	Wakashan		
Wao	Waorani		
War	Warao		
Was	Wasi		
WBg	West Bougainville		
WBm	West Bomberai		
WF	Western Fly		
WP	West Papuan		
Wsh	Washo		
WY	Wappo-Yukian		
Xin	Xincan		



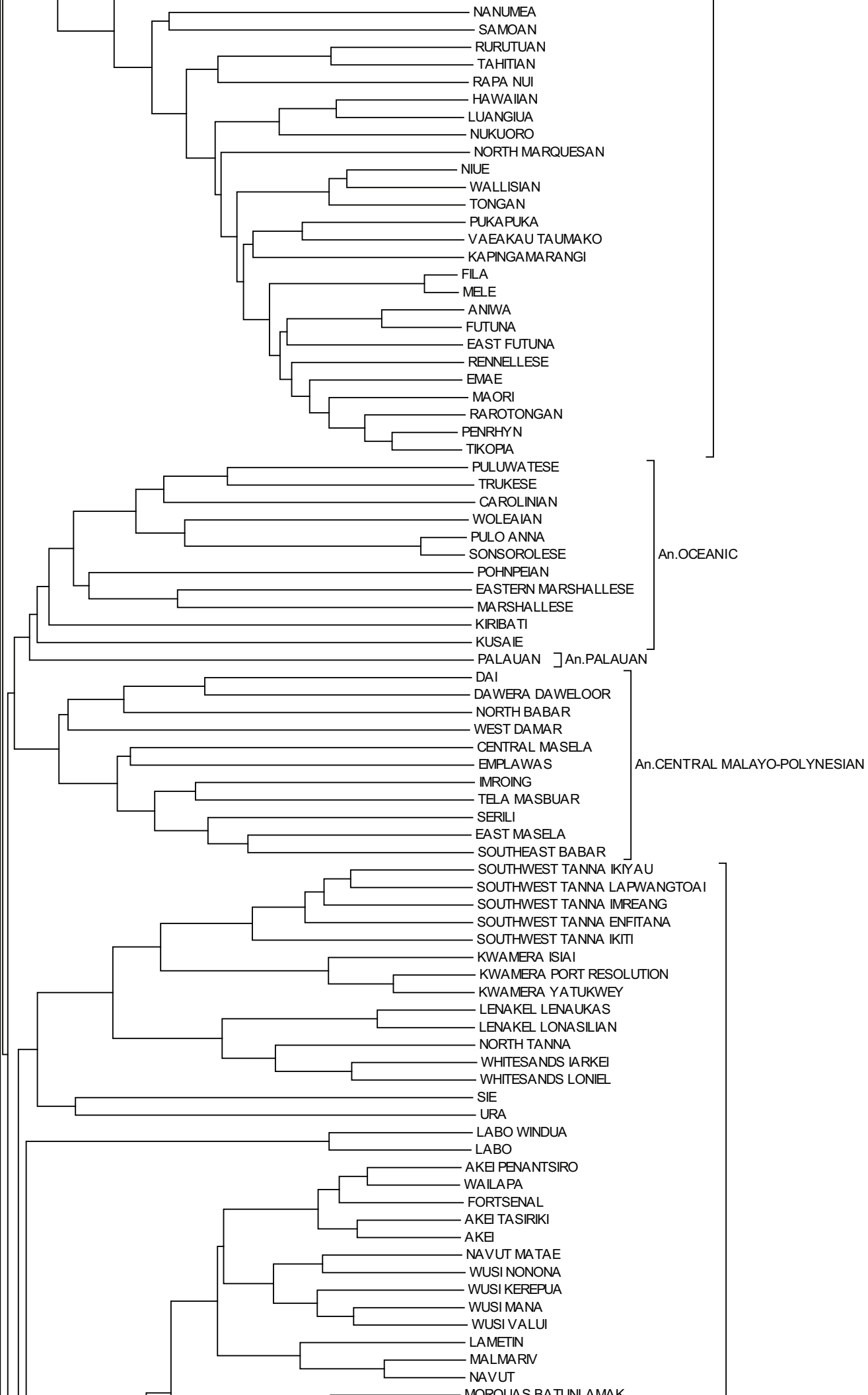


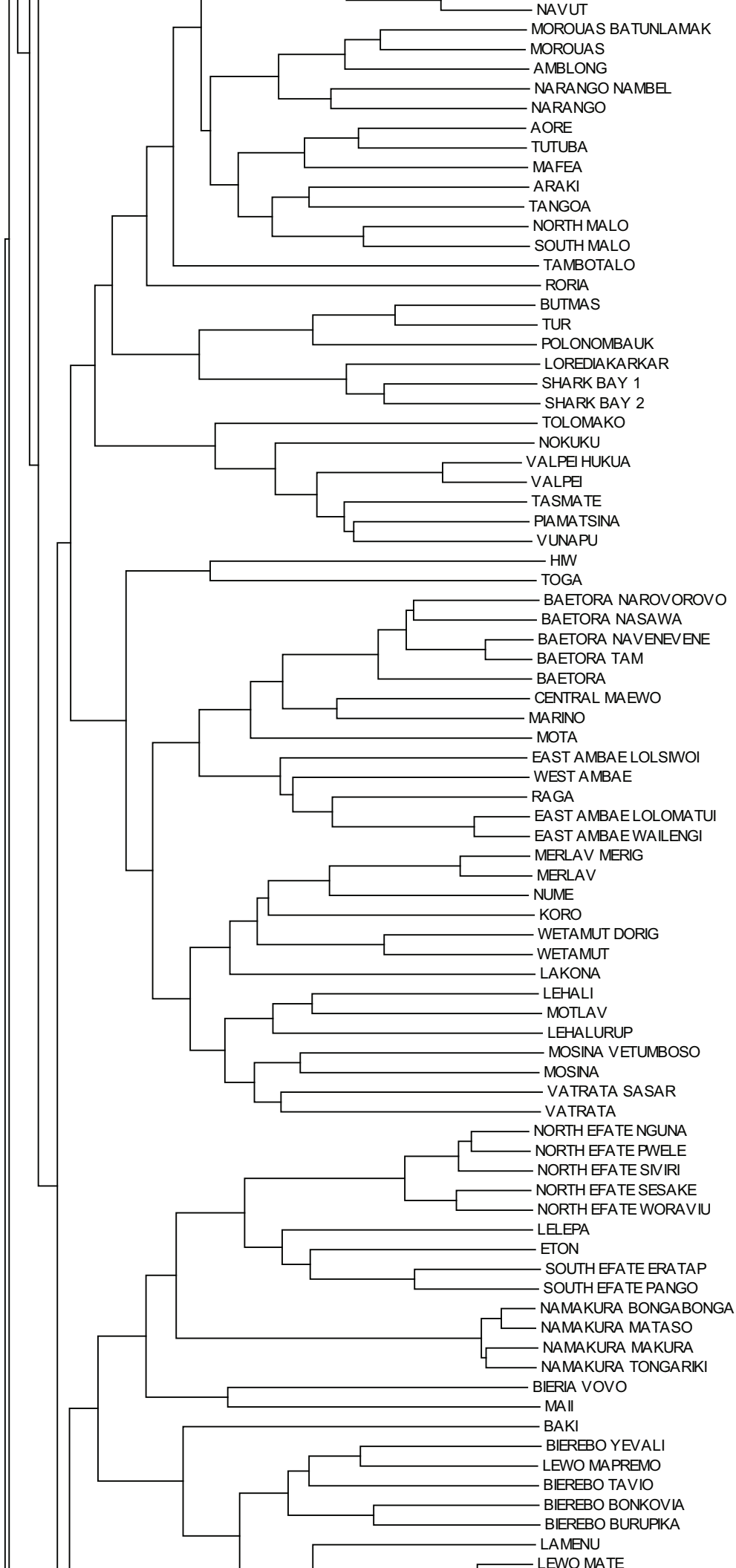




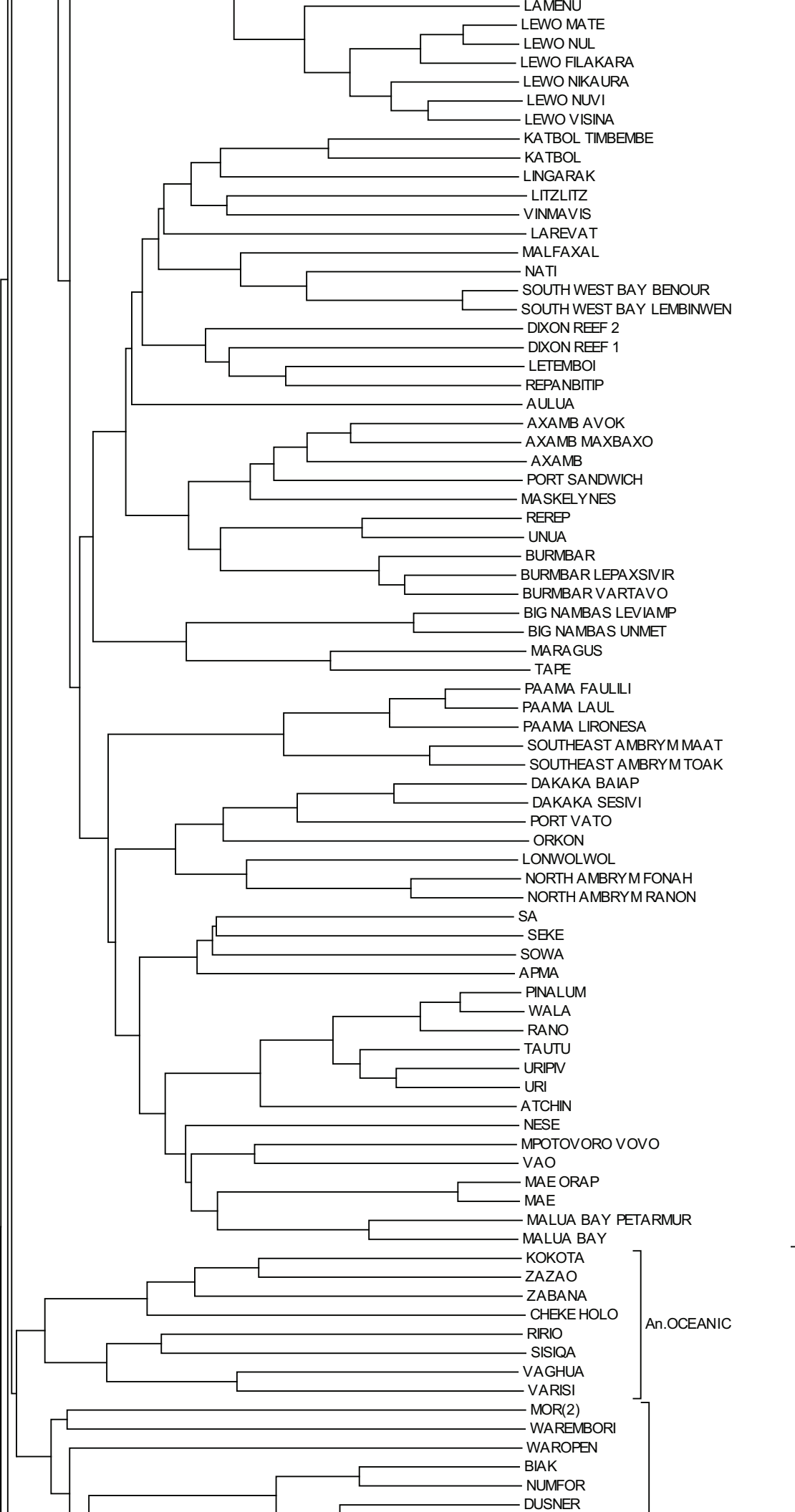


An.OCEANIC

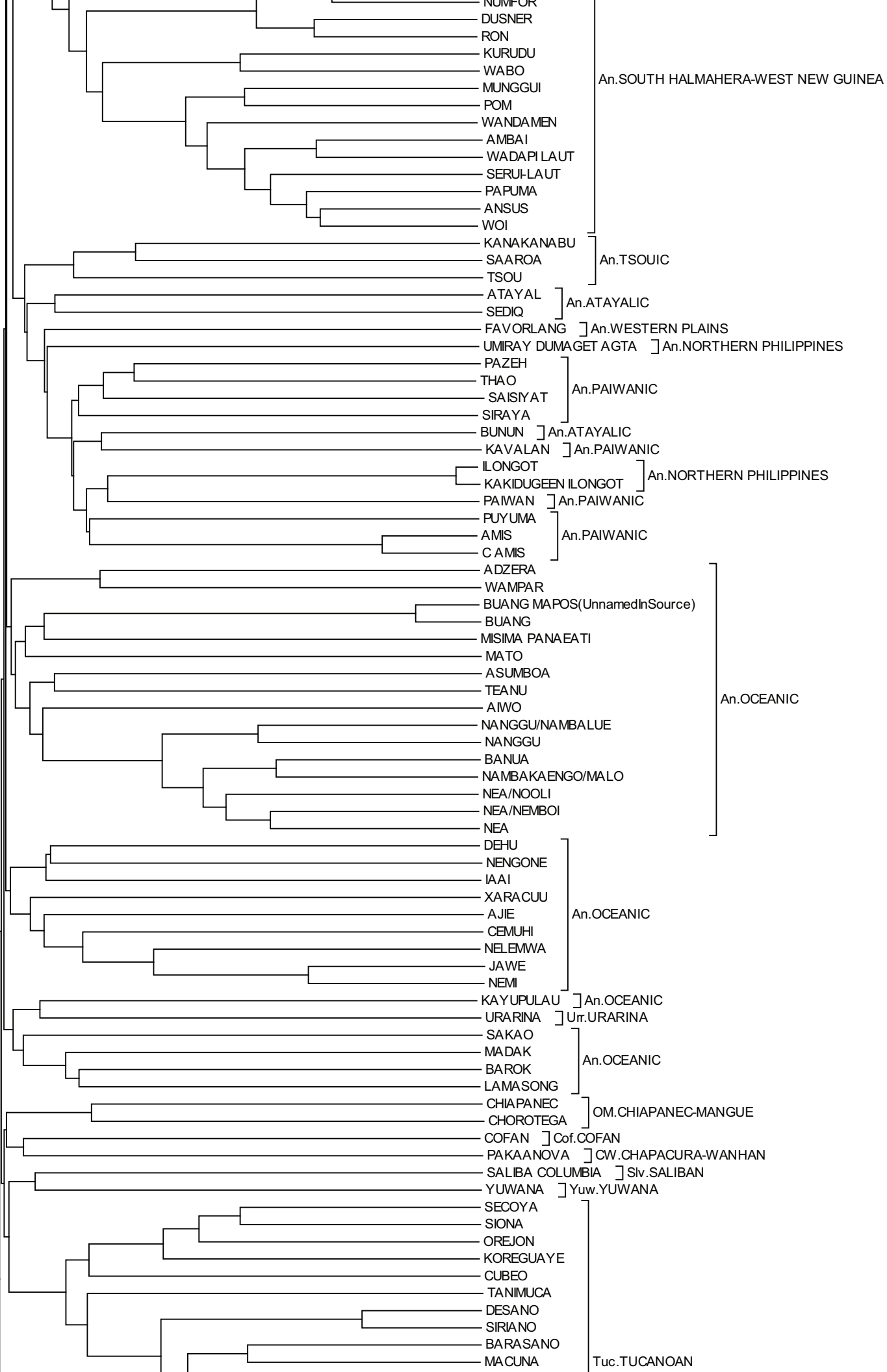




An.OCEANIC



An.OCEANIC



NUSFOR

DUSNER

RON

KURUDU

WABO

MUNGGUI

POM

WANDAMEN

AMBAI

WADAPI LAUT

SERUI-LAUT

PAPUMA

ANSUS

WOI

KANAKANABU

SAAROA

TSOU

ATAYAL

SEDIQ

FAVORLANG

UMIRAY DUMAGET AGTA

PAZEH

THAO

SAISIYAT

SIRAYA

BUNUN

KAVALAN

ILONGOT

KAKIDUGEN ILONGOT

PAIWAN

PUYUMA

AMIS

CAMIS

ADZERA

WAMPAR

BUANG MAPOS (UnnamedInSource)

BUANG

MISIMA PANAEATI

MATO

ASUMBOA

TEANU

AIWO

NANGGU/NAMBALUE

NANGGU

BANUA

NAMBAKAENGO/MALO

NEA/NOOLI

NEA/NEMBOI

NEA

DEHU

NENGONE

IAAI

XARACUU

AJIE

CEMUHI

NELEMWA

JAWE

NEVI

KAYUPULAU

URARINA

SAKAO

MADAK

BAROK

LAMASONG

CHIAPANEC

CHOROTEGA

COFAN

PAKAAANOVA

SALIBA COLUMBIA

YUWANA

SECOYA

SIONA

OREJON

KOREGUAYE

CUBEO

TANIMUCA

DESANO

SIRIANO

BARASANO

MACUNA

An.SOUTH HALMAHERA-WEST NEW GUINEA

An.TSOUIIC

An.ATAYALIC

An.WESTERN PLAINS

An.NORTHERN PHILIPPINES

An.PAIWANIC

An.ATAYALIC

An.PAIWANIC

An.NORTHERN PHILIPPINES

An.PAIWANIC

An.PAIWANIC

An.OCEANIC

An.OCEANIC

An.OCEANIC

Urr.URARINA

An.OCEANIC

OM.CHIAPANEC-MANGUE

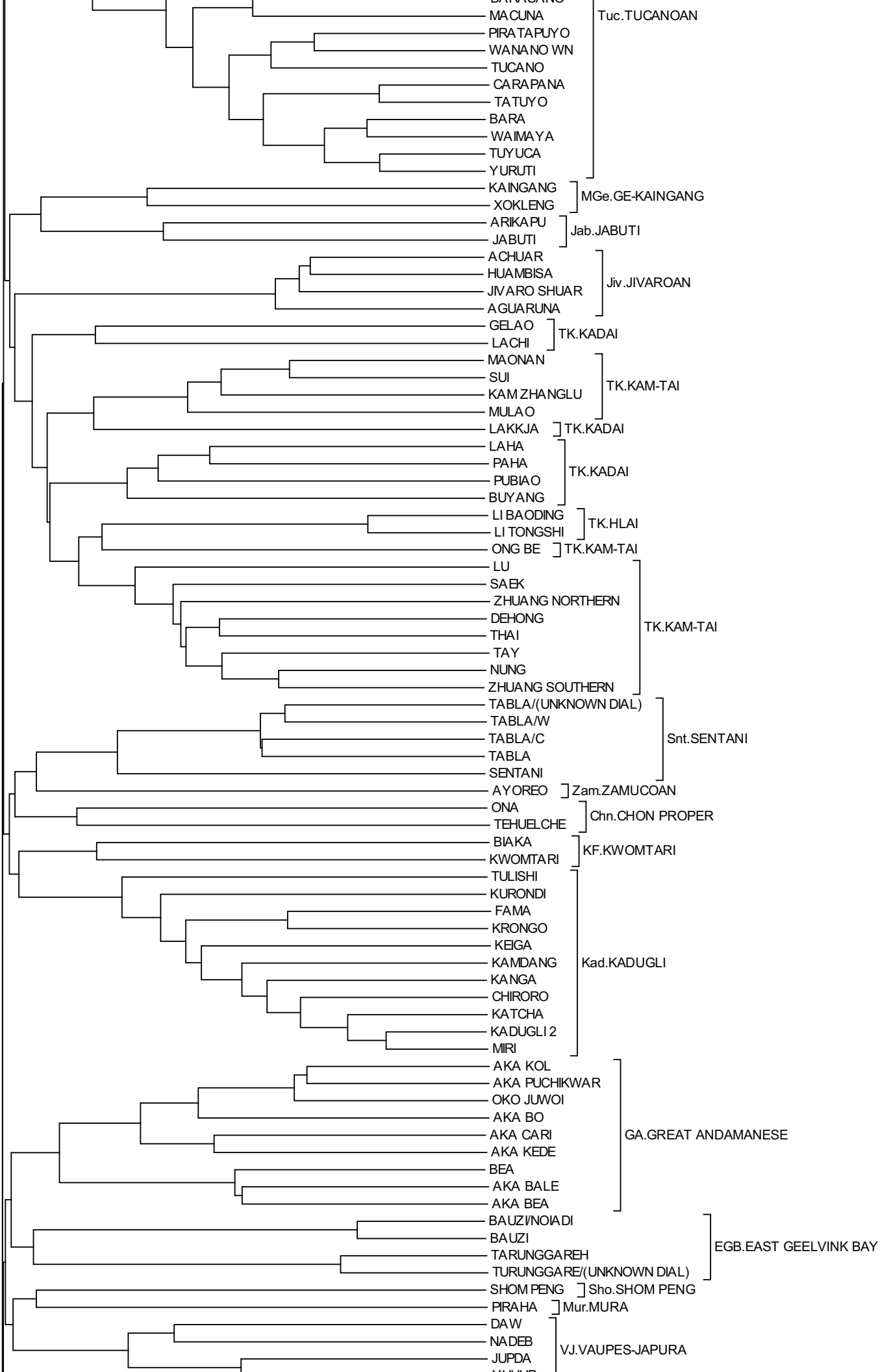
Cof.COFAN

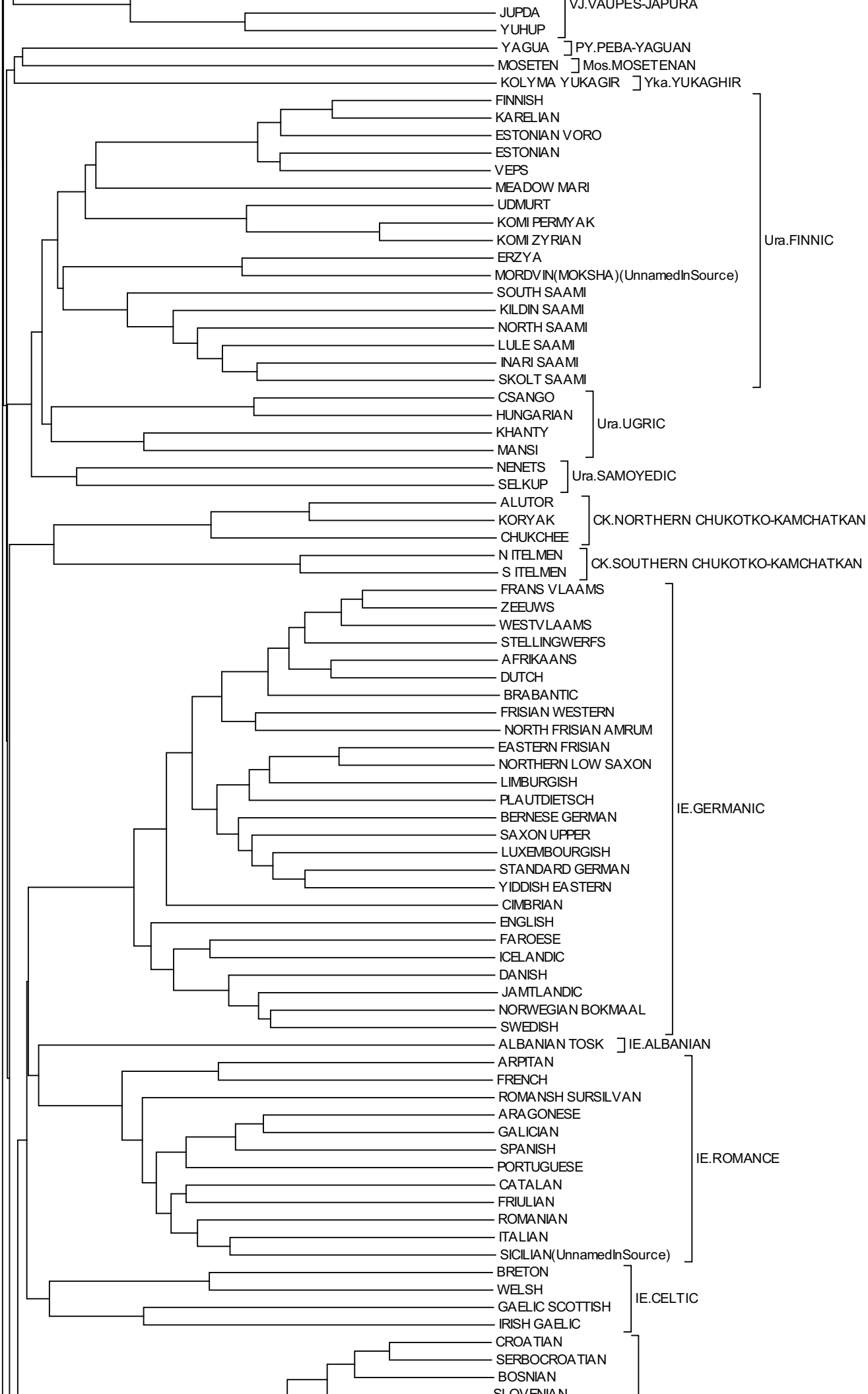
CW.CHAPACURA-WANHAN

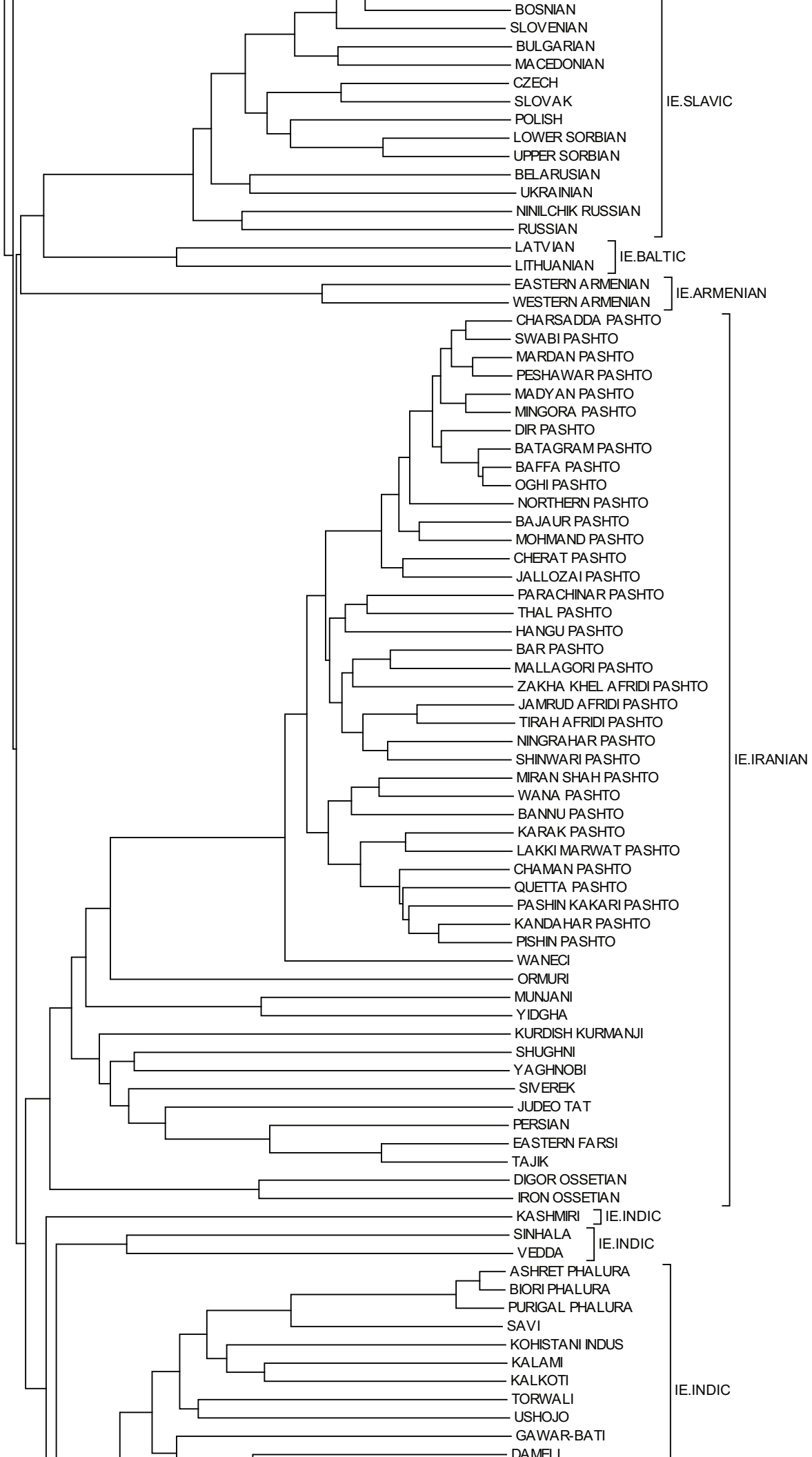
Slv.SALIBAN

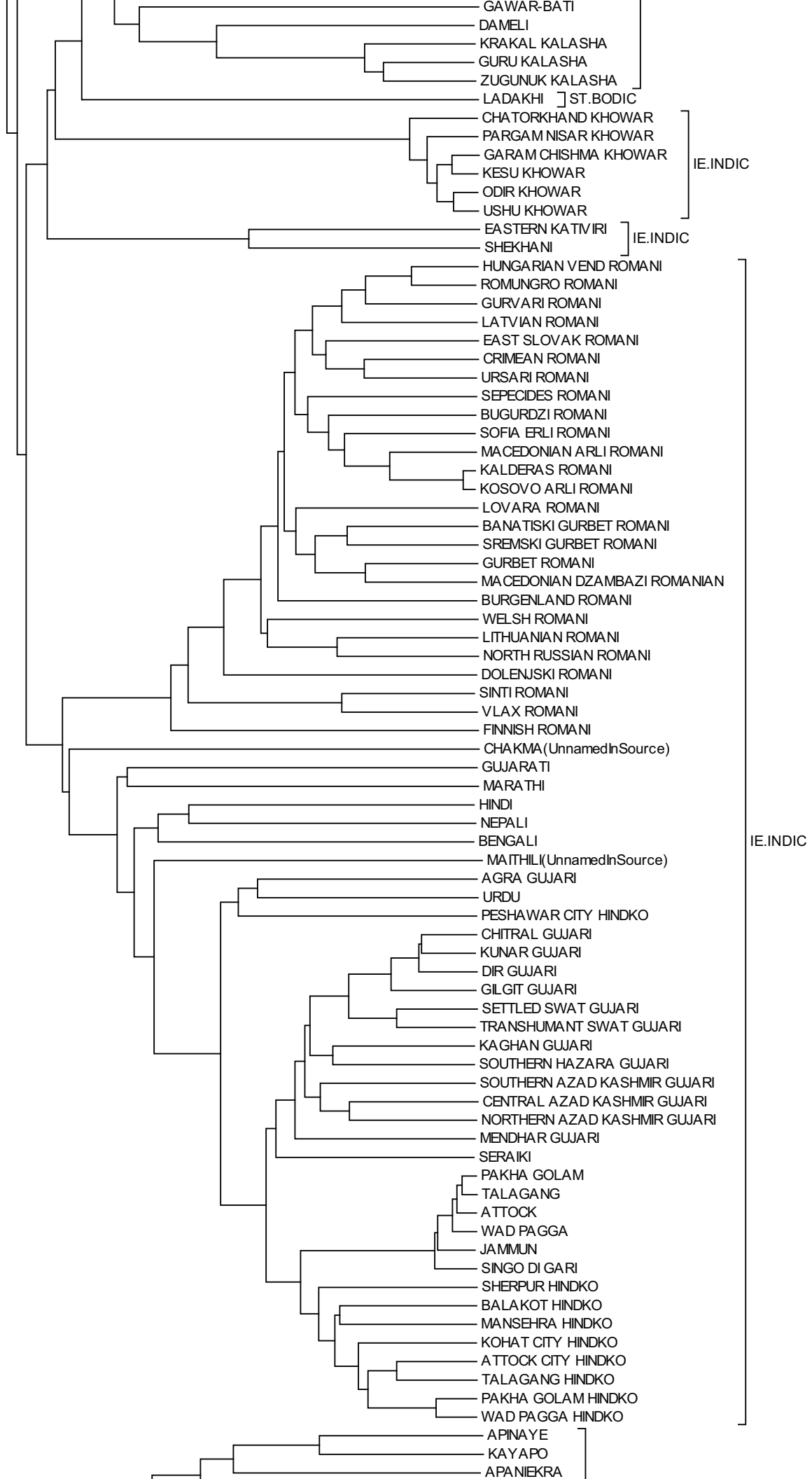
Yuw.YUWANA

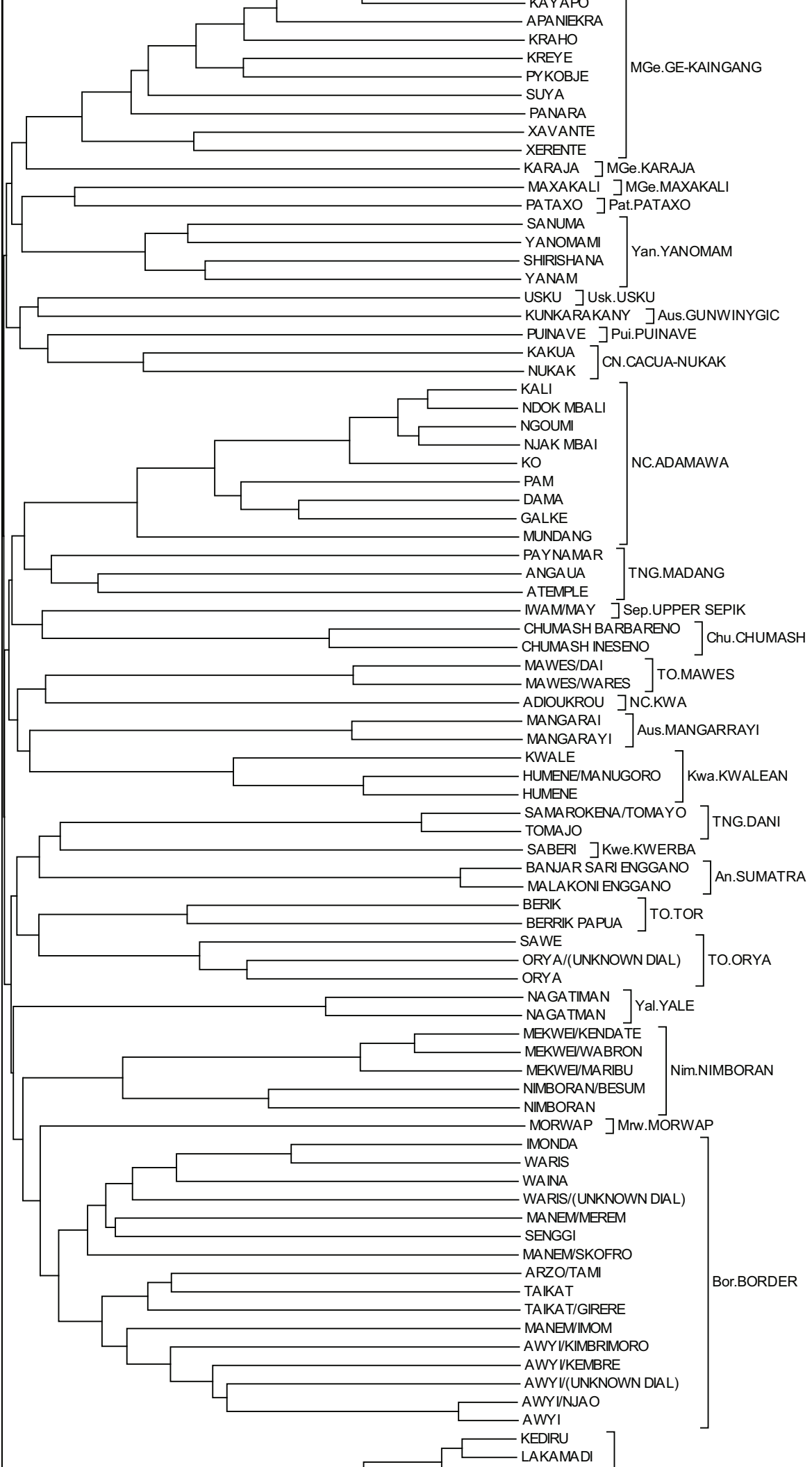
Tuc.TUCANOAN











MGe.GE-KAINGANG

MGe.KARAJA

MGe.MAXAKALI

Pat.PATA XO

Yan.YANOMAM

Usk.USKU

Aus.GUNWINYGIC

Pui.PUINAVE

CN.CACUA-NUKAK

NC.ADAMAWA

TNG.MADANG

Sep.UPPER SEPIK

Chu.CHUMASH

TO.MAWES

NC.KWA

Aus.MANGARRAYI

Kwa.KWALEAN

TNG.DANI

Kwe.KWERBA

An.SUMATRA

TO.TOR

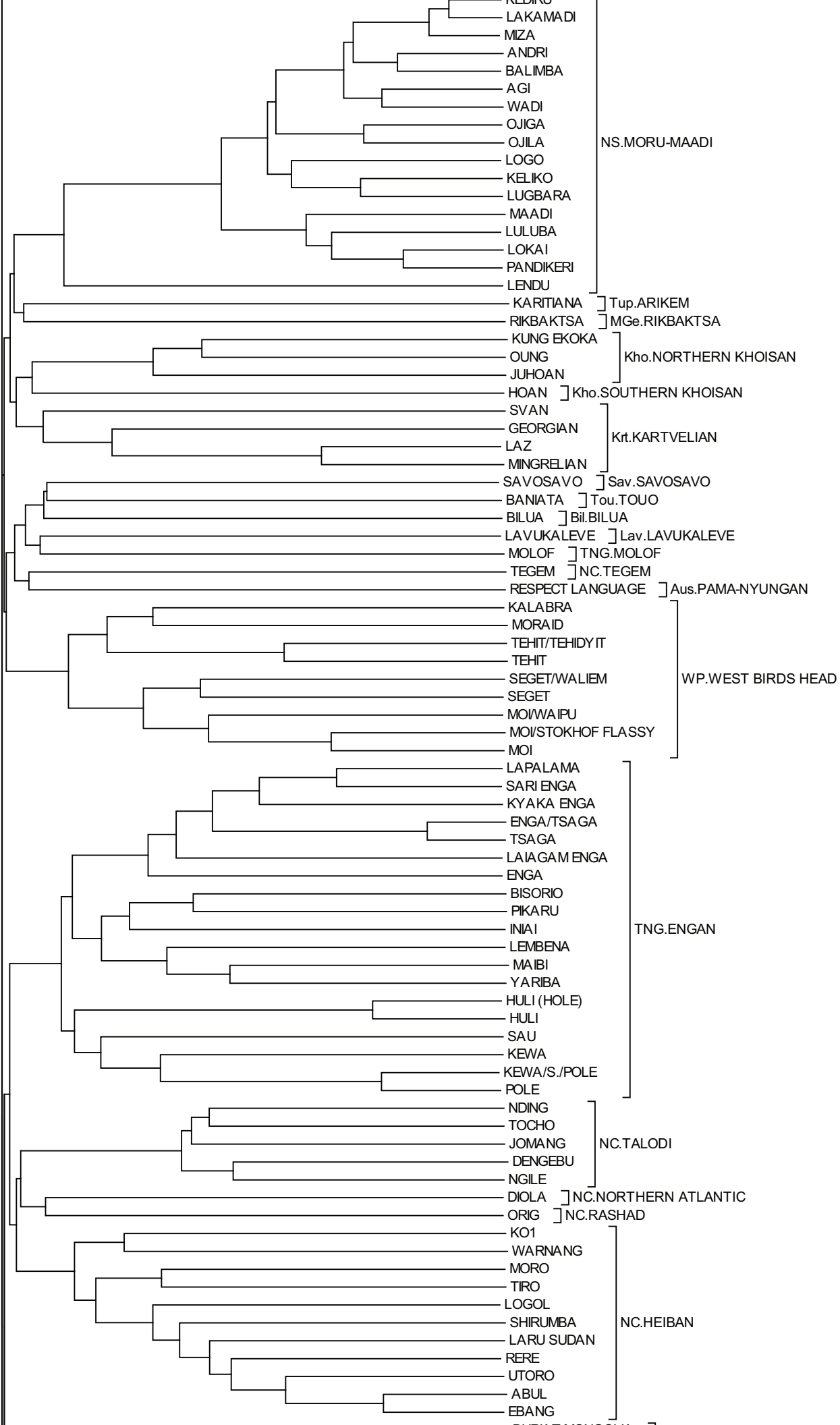
TO.ORYA

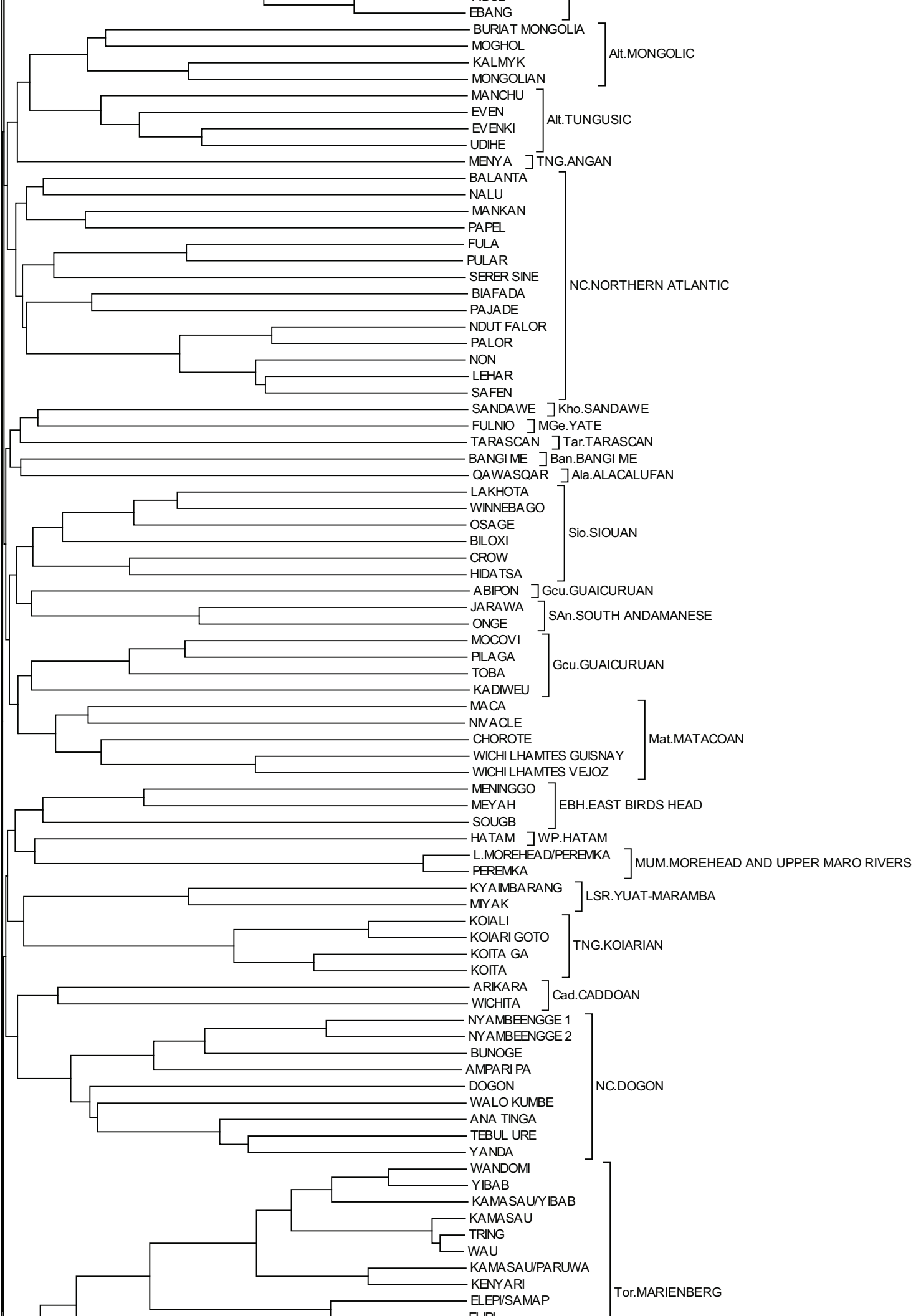
Yal.YALE

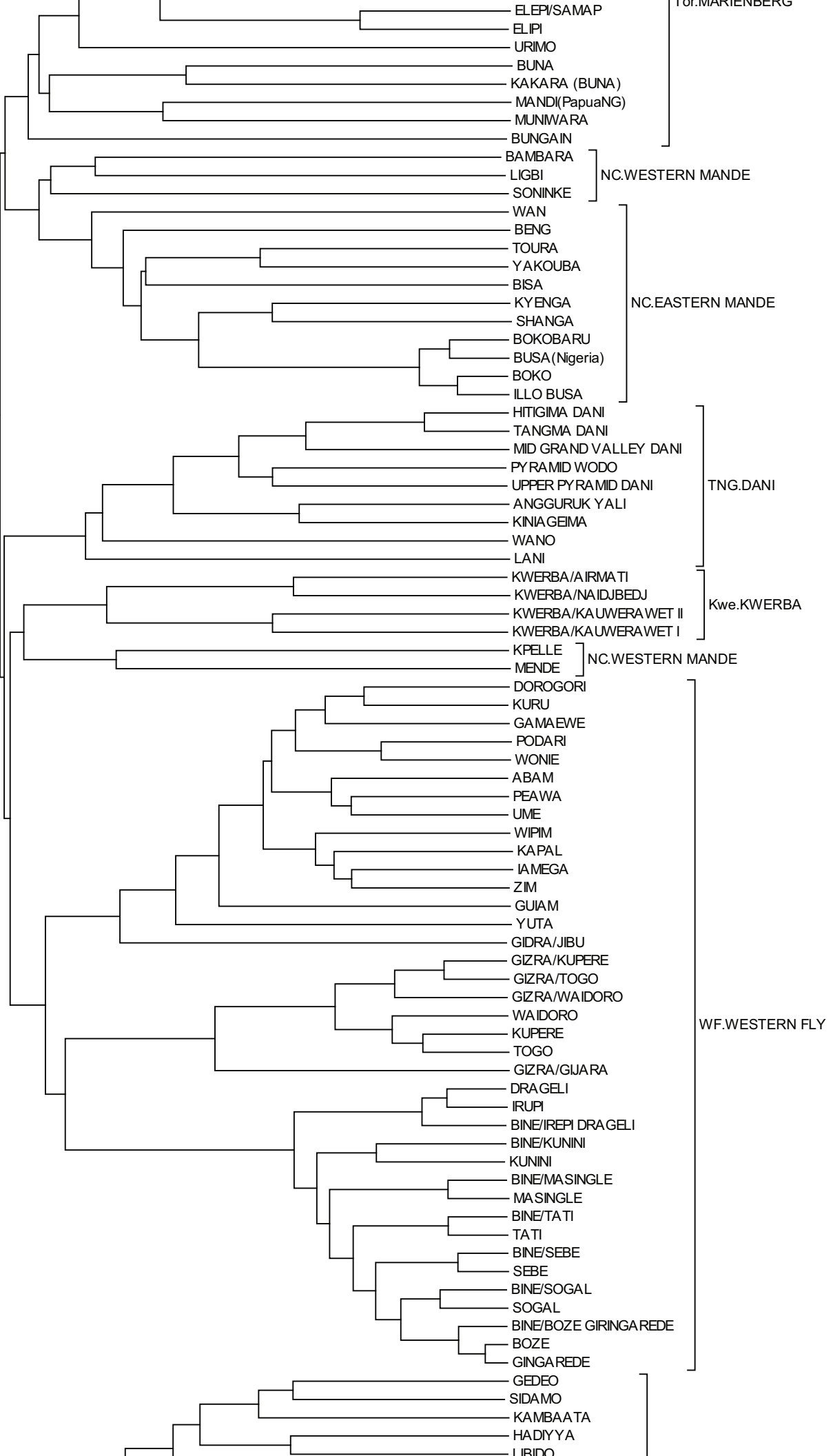
Nim.NIMBORAN

Mrw.MORWAP

Bor.BORDER







NC.WESTERN MANDE

NC.EASTERN MANDE

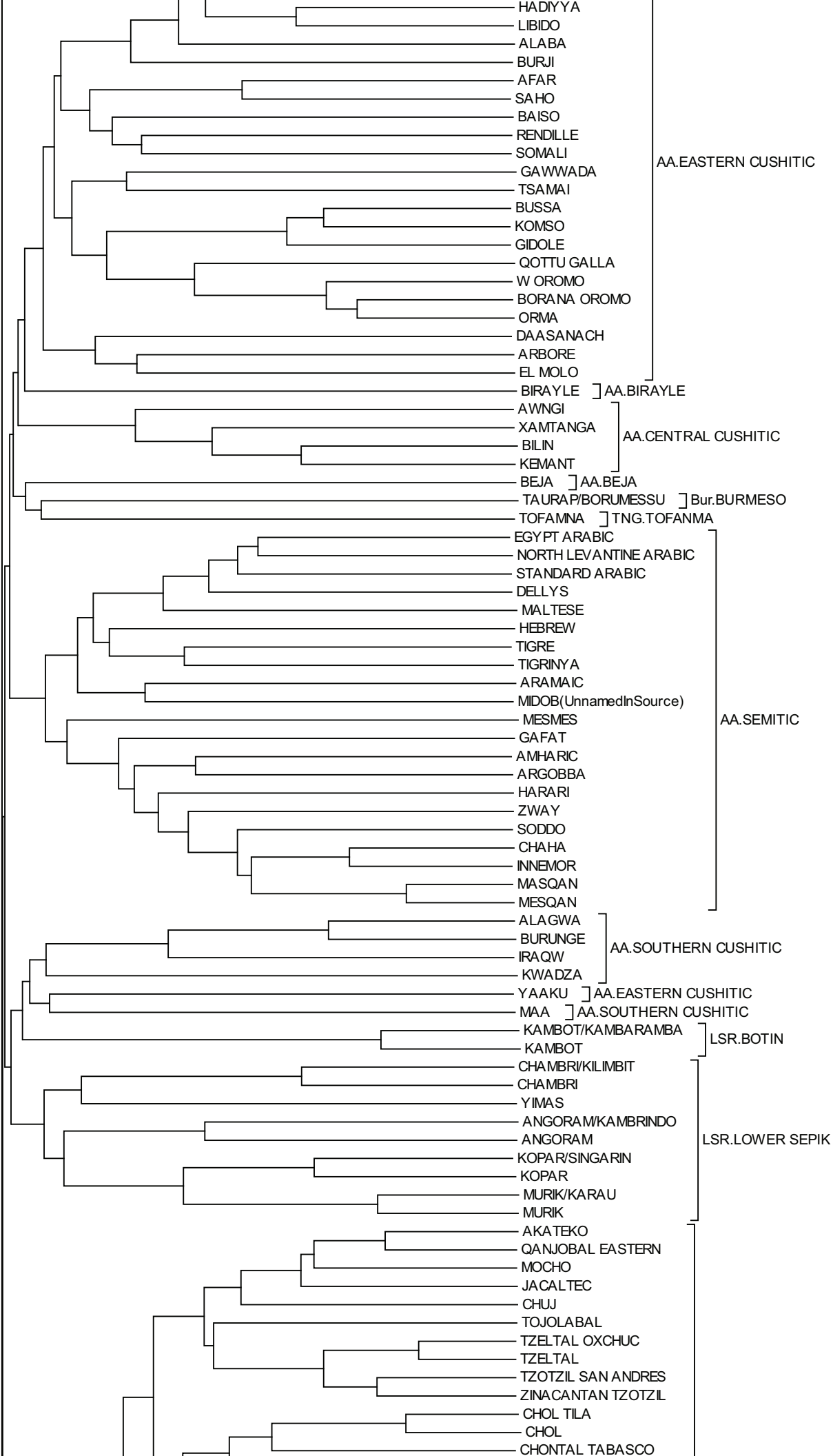
TNG.DANI

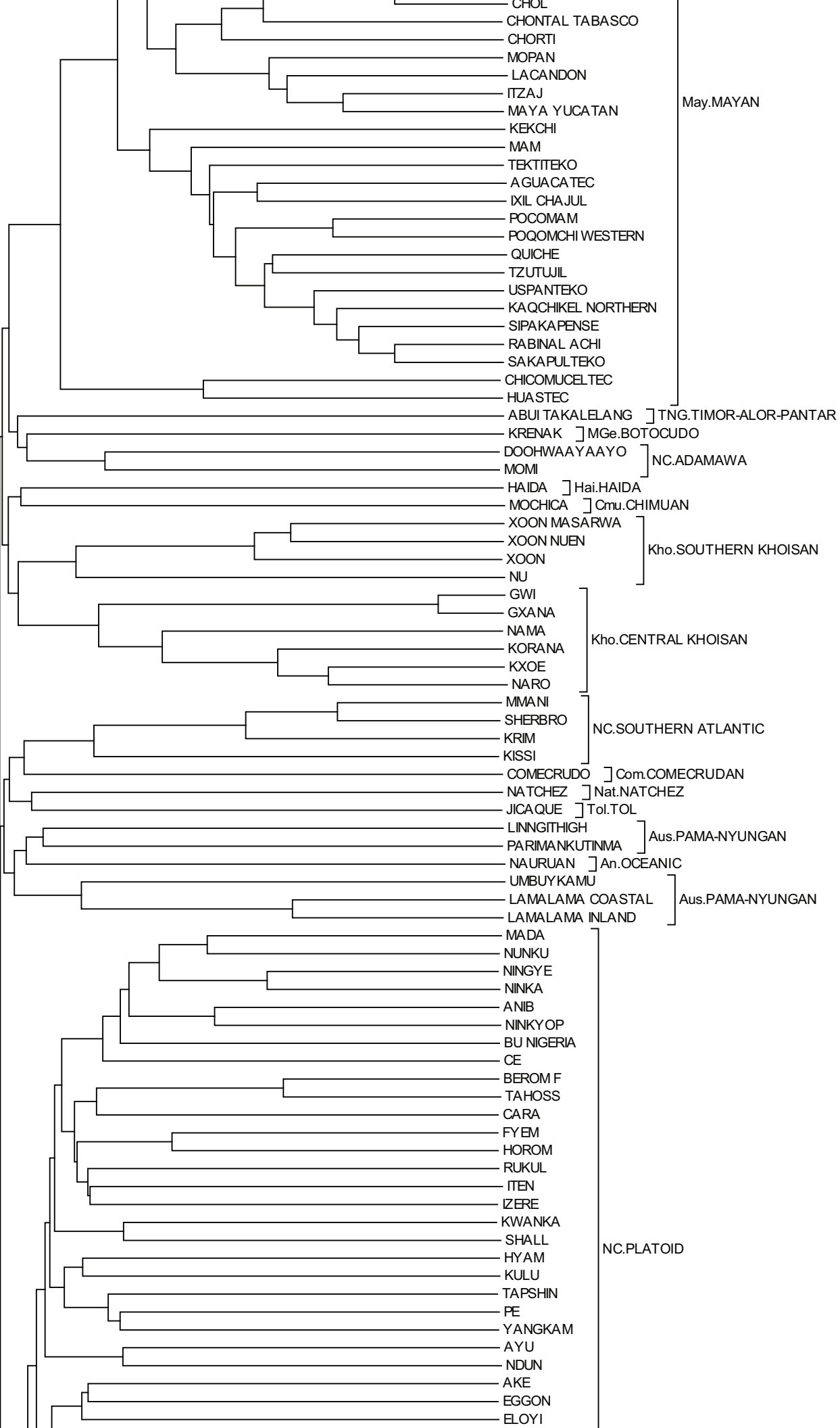
Kwe.KWERBA

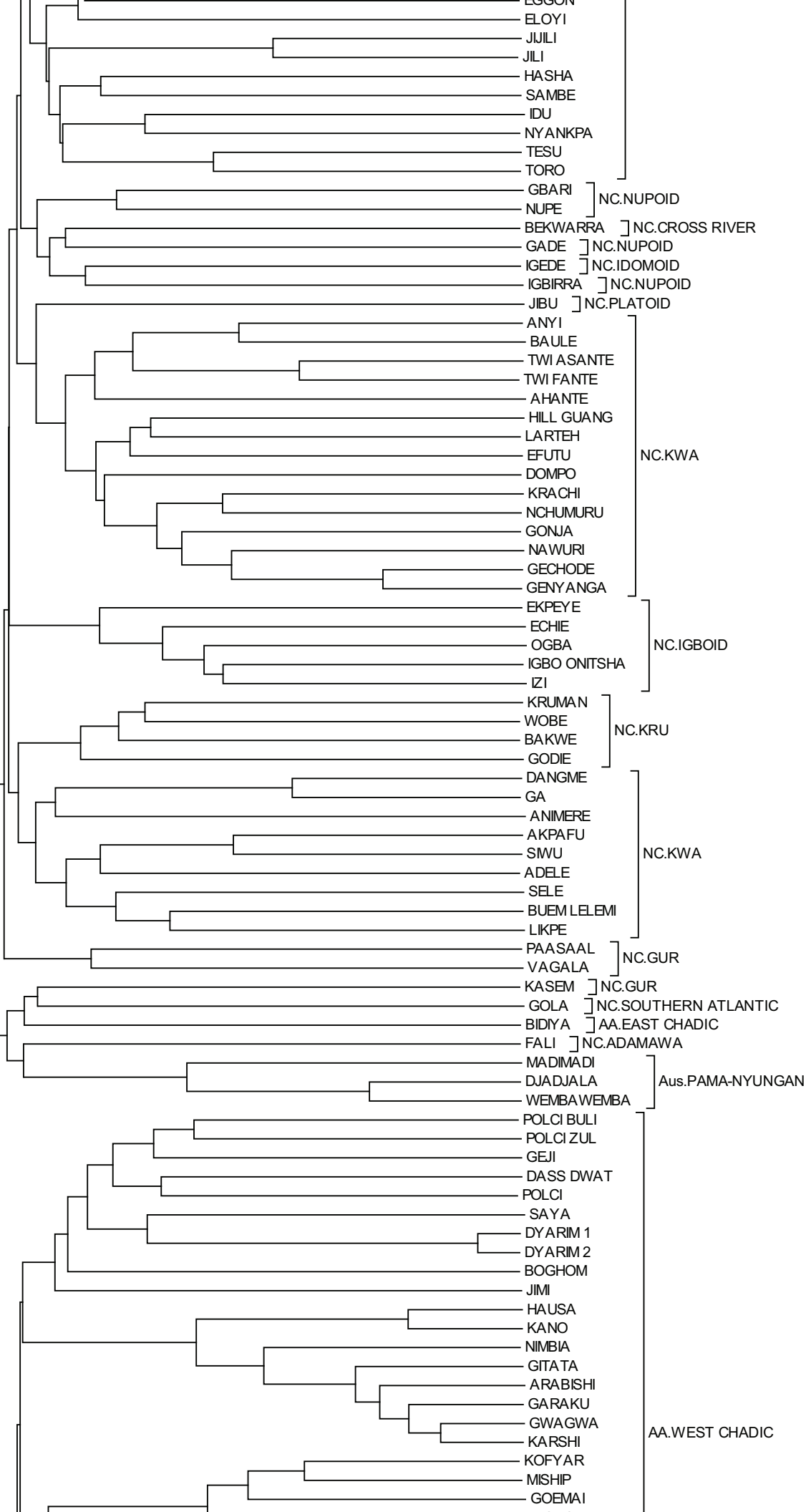
NC.WESTERN MANDE

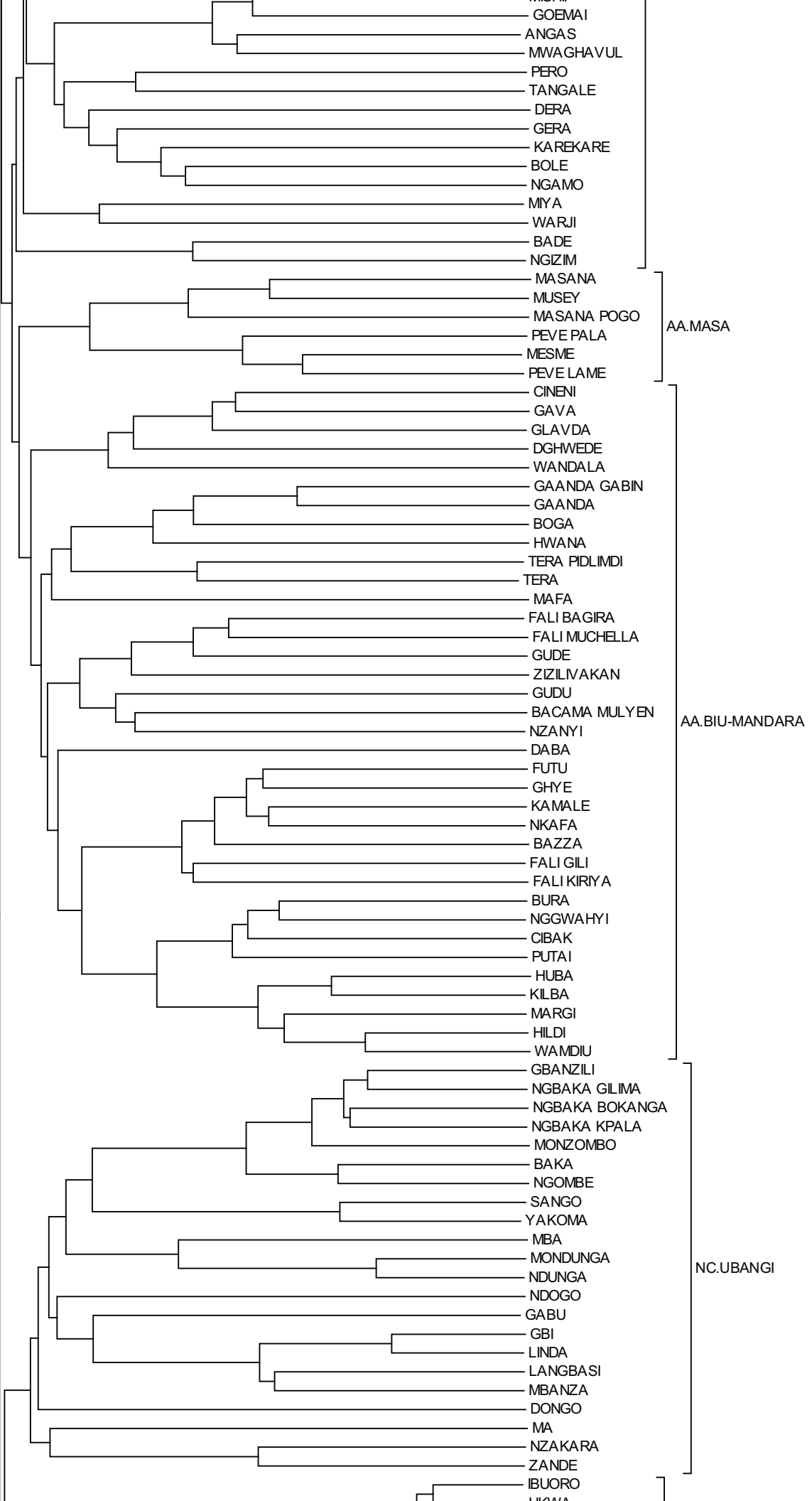
WF.WESTERN FLY

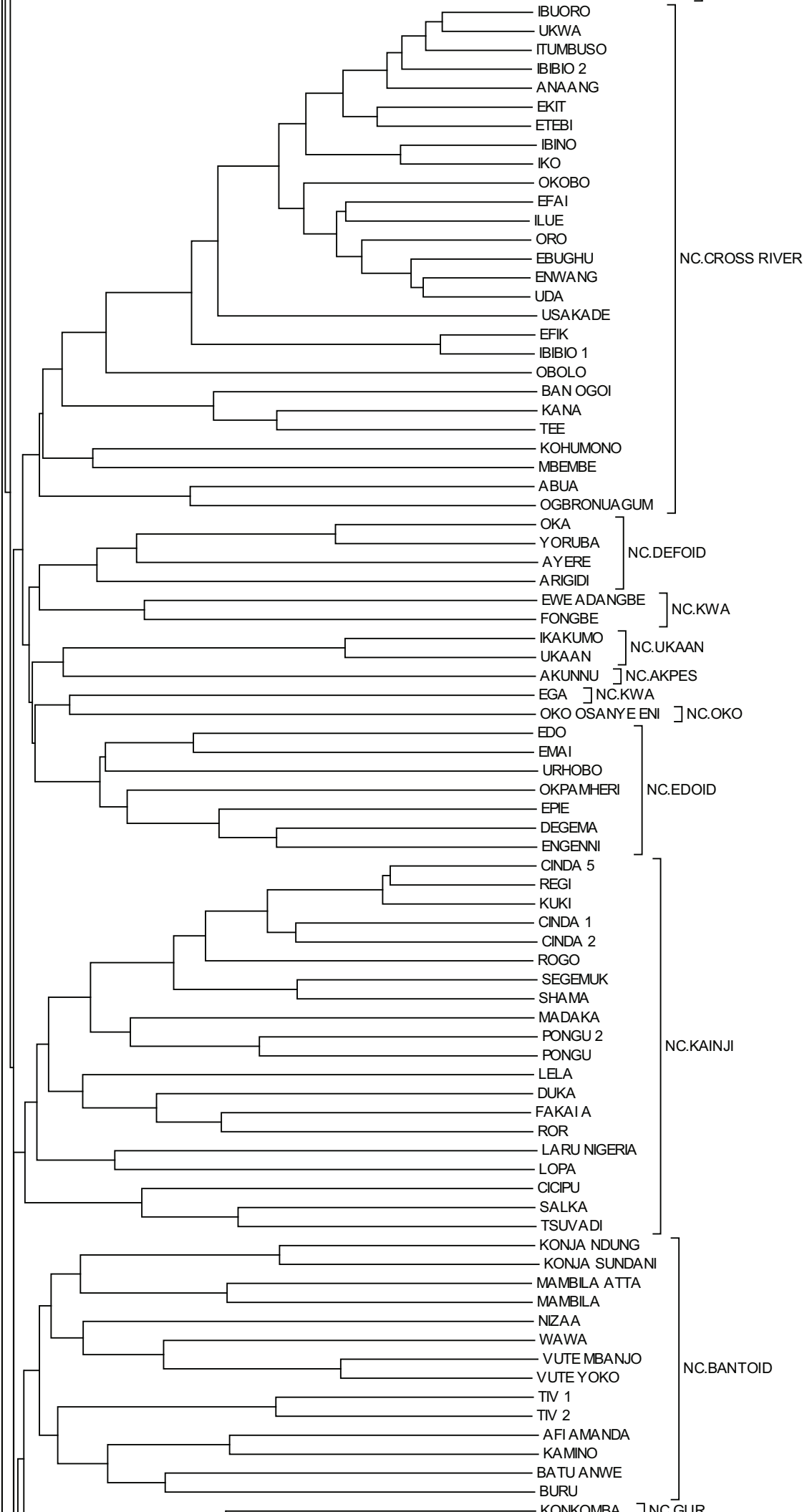
GEDEO

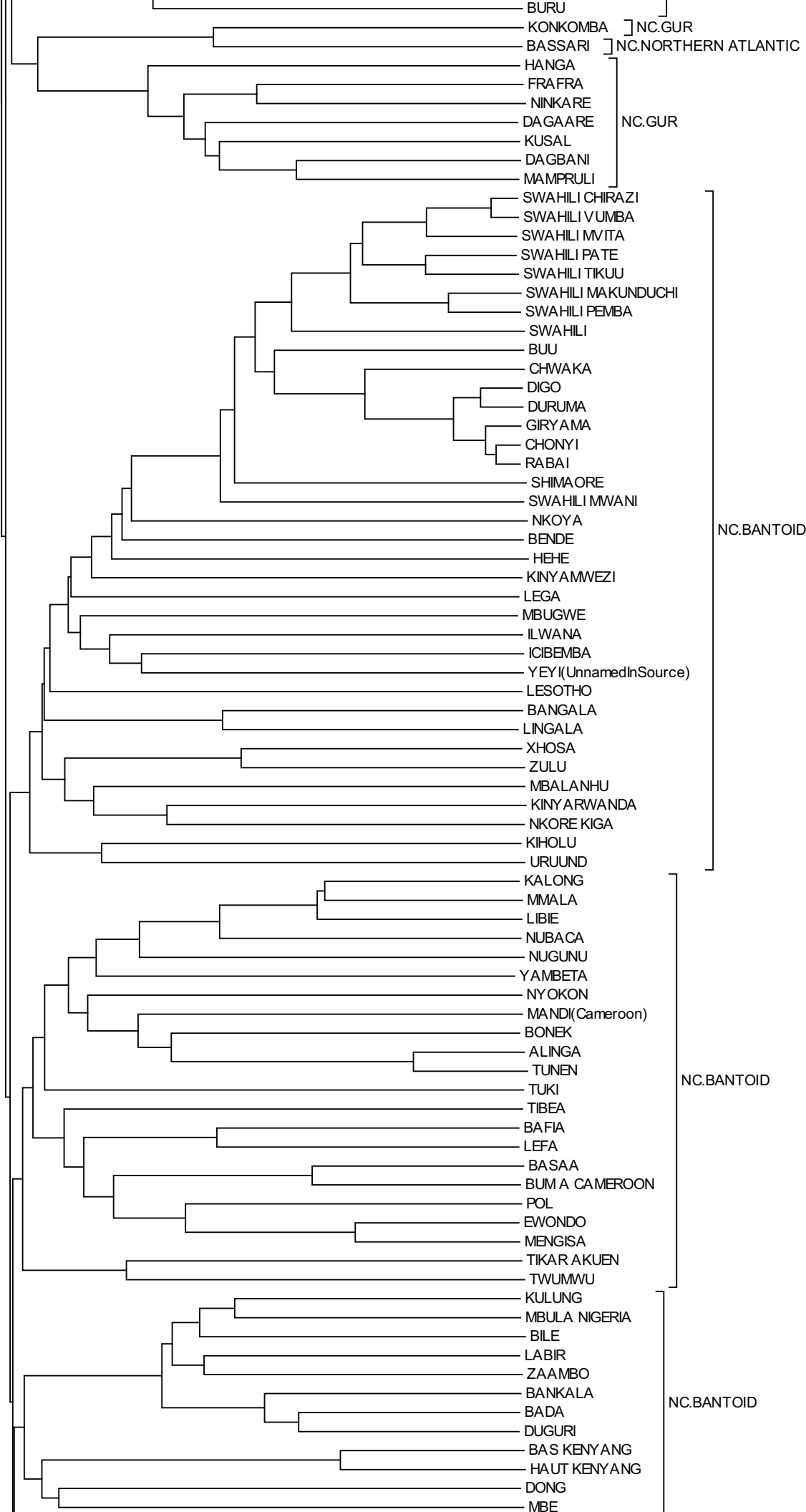


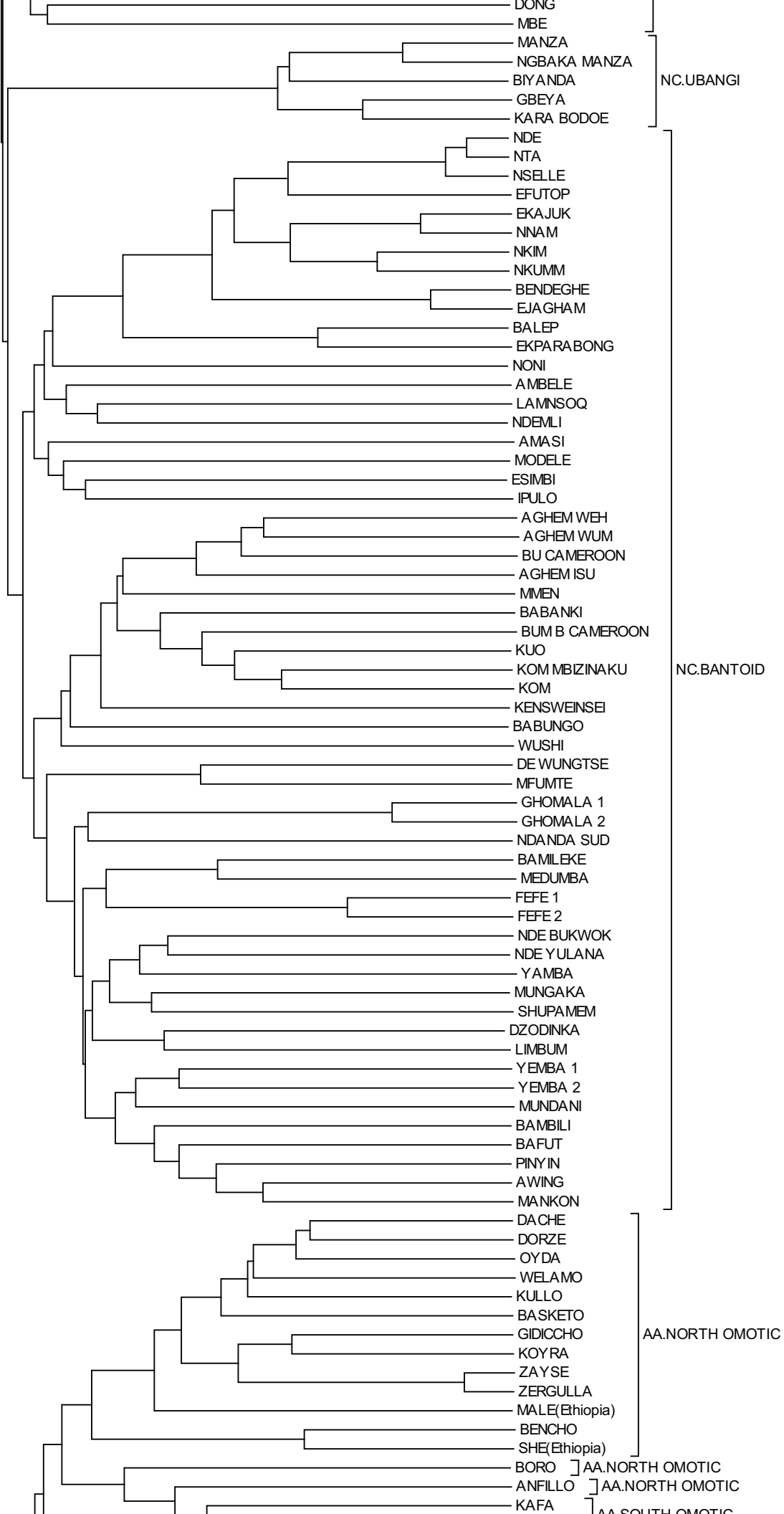








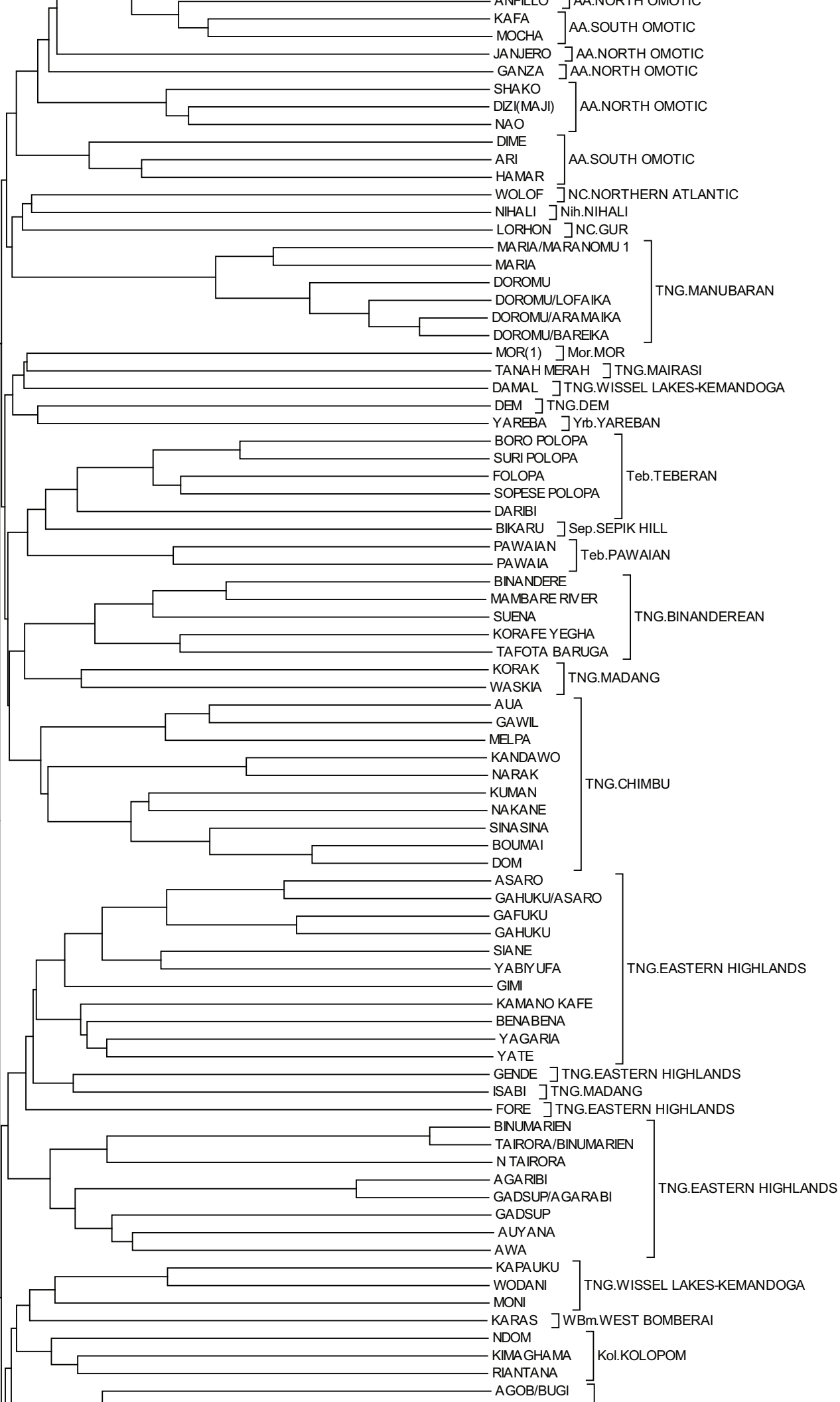


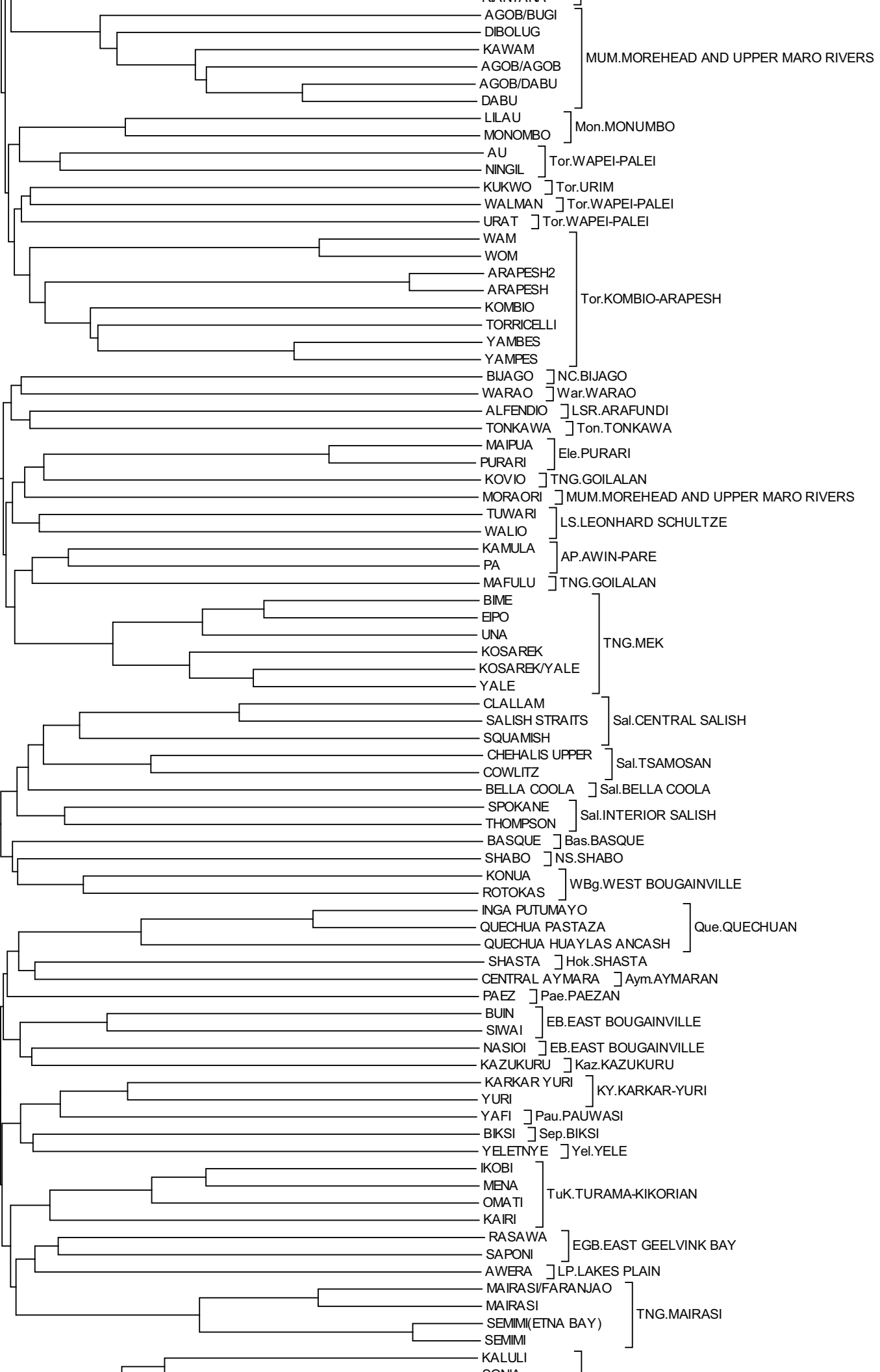


NC.UBANGI

NC.BANTOID

AA.NORTH OMOTIC





AGOB/BUGI
 DIBOLUG
 KAWAM
 AGOB/AGOB
 AGOB/DABU
 DABU
 MUM.MOREHEAD AND UPPER MARO RIVERS

LILAU
 MONOMBO
 Mon.MONUMBO

AU
 NINGIL
 Tor.WAPEI-PALEI

KUKWO
 WALMAN
 Tor.URIM
 Tor.WAPEI-PALEI

URAT
 Tor.WAPEI-PALEI

WAM
 WOM
 ARAPESH2
 ARAPESH
 KOMBIO
 TORRICELLI
 YAMBES
 YAMPES
 Tor.KOMBIO-ARAPESH

BIJAGO
 WARAO
 ALFENDIO
 TONKAWA
 NC.BIJAGO
 War.WARAO
 LSR.ARAFUNDI
 Ton.TONKAWA

MAIPUA
 PURARI
 Ele.PURARI

KOVIO
 MORAORI
 TNG.GOILALAN
 MUM.MOREHEAD AND UPPER MARO RIVERS

TUWARI
 WALIO
 LS.LEONHARD SCHULTZE

KAMULA
 PA
 AP.AWIN-PARE

MAFULU
 TNG.GOILALAN

BIME
 EIPO
 UNA
 KOSAREK
 KOSAREK/YALE
 YALE
 TNG.MEK

CLALLAM
 SALISH STRAITS
 SQUAMISH
 Sal.CENTRAL SALISH

CHEHALIS UPPER
 COWLITZ
 Sal.TSAMOSAN

BELLA COOLA
 SPOKANE
 THOMPSON
 Sal.BELLA COOLA
 Sal.INTERIOR SALISH

BASQUE
 SHABO
 Bas.BASQUE
 NS.SHABO

KONUA
 ROTOKAS
 WBg.WEST BOUGAINVILLE

INGA PUTUMAYO
 QUECHUA PASTAZA
 QUECHUA HUAYLAS ANCASH
 Que.QUECHUAN

SHASTA
 CENTRAL AYMARA
 Hok.SHASTA
 Aym.AYMARAN

PAEZ
 Pae.PAEZAN

BUIN
 SIWAI
 EB.EAST BOUGAINVILLE

NASIOI
 KAZUKURU
 EB.EAST BOUGAINVILLE
 Kaz.KAZUKURU

KARKAR YURI
 YURI
 KY.KARKAR-YURI

YAFI
 BIKSI
 Pau.PAUWASI
 Sep.BIKSI

YELETNYE
 Yel.YELE

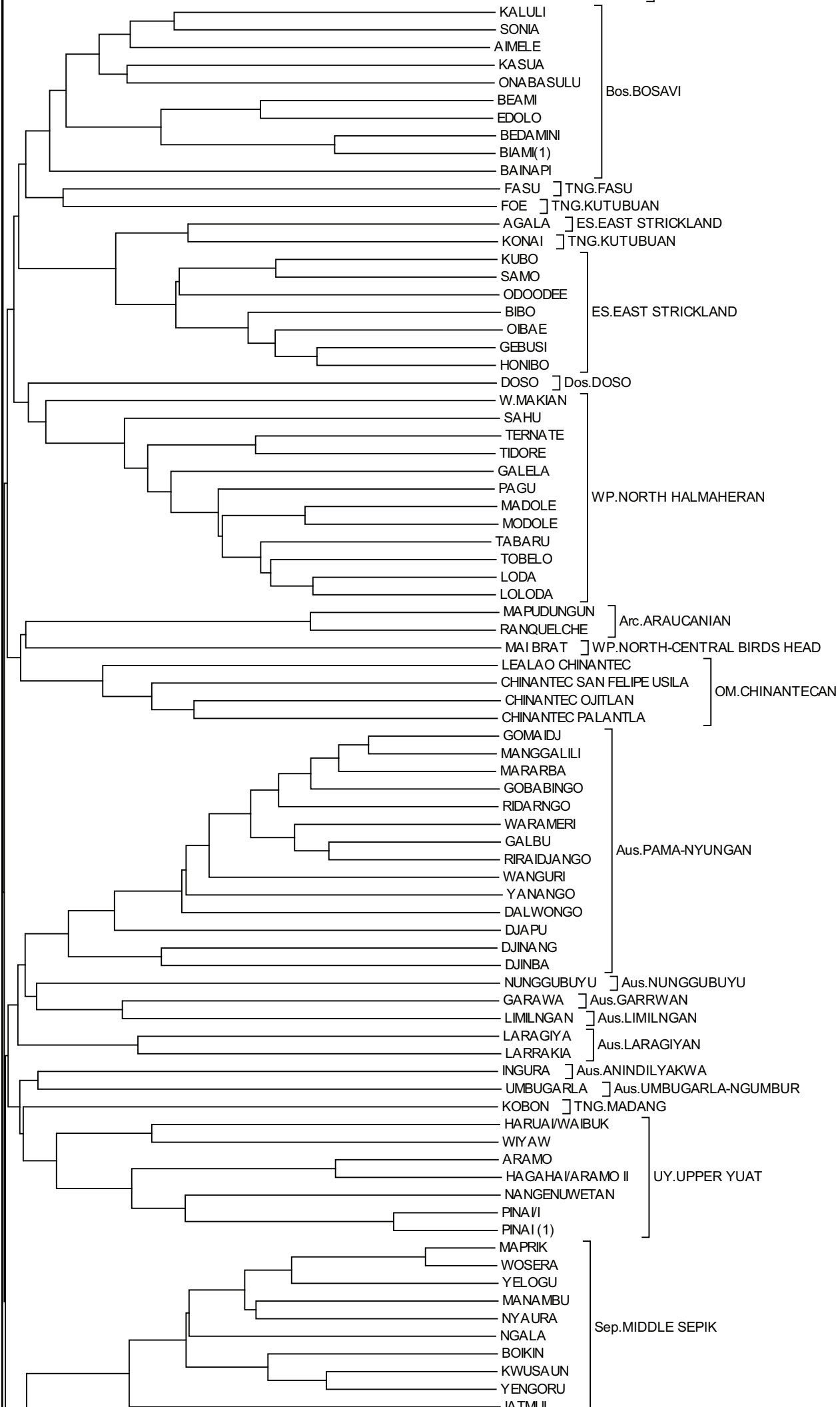
IKOBI
 MENA
 OMATI
 TuK.TURAMA-KIKORIAN

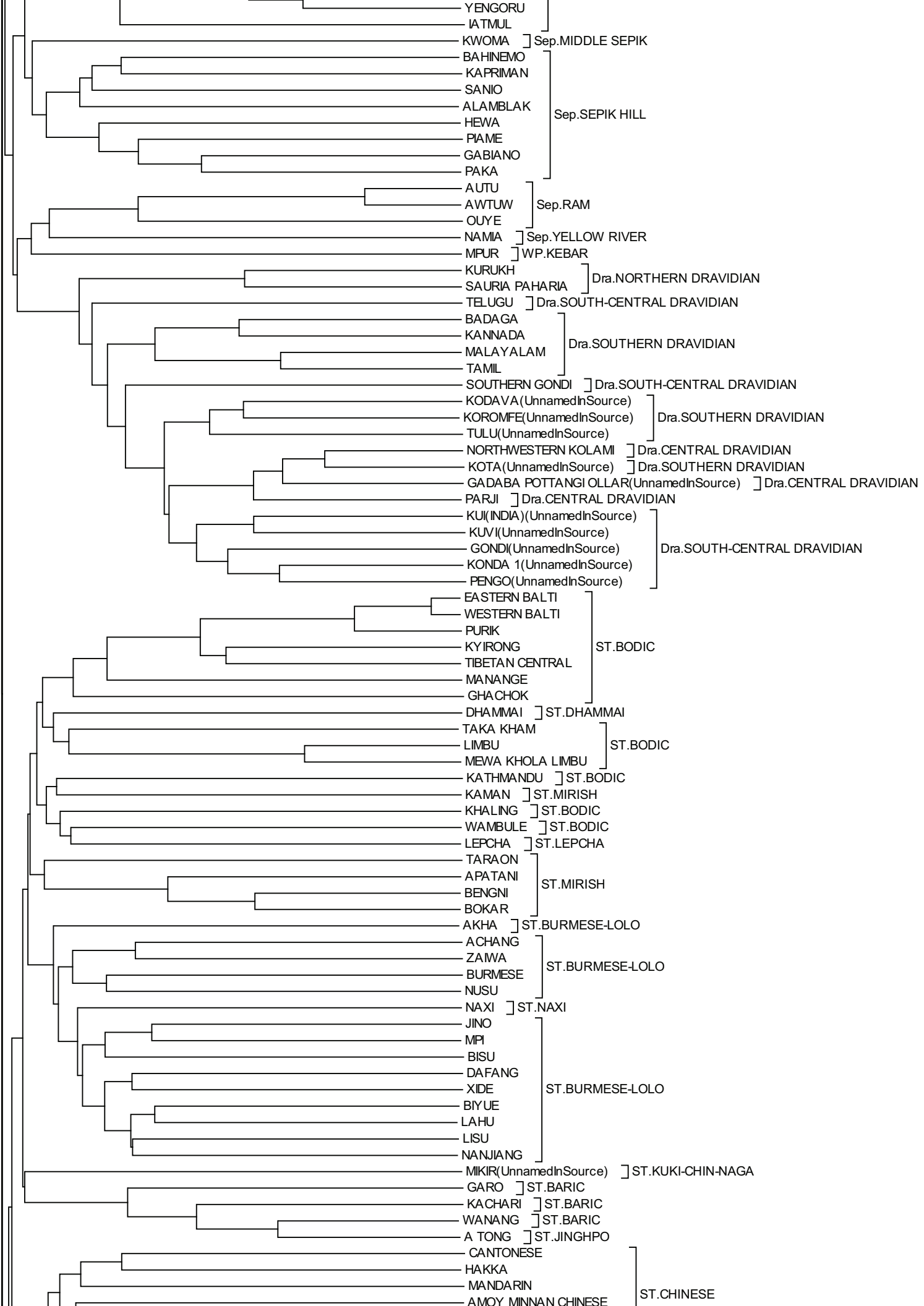
KAIRI
 RASAWA
 SAPONI
 EGB.EAST GEELVINK BAY

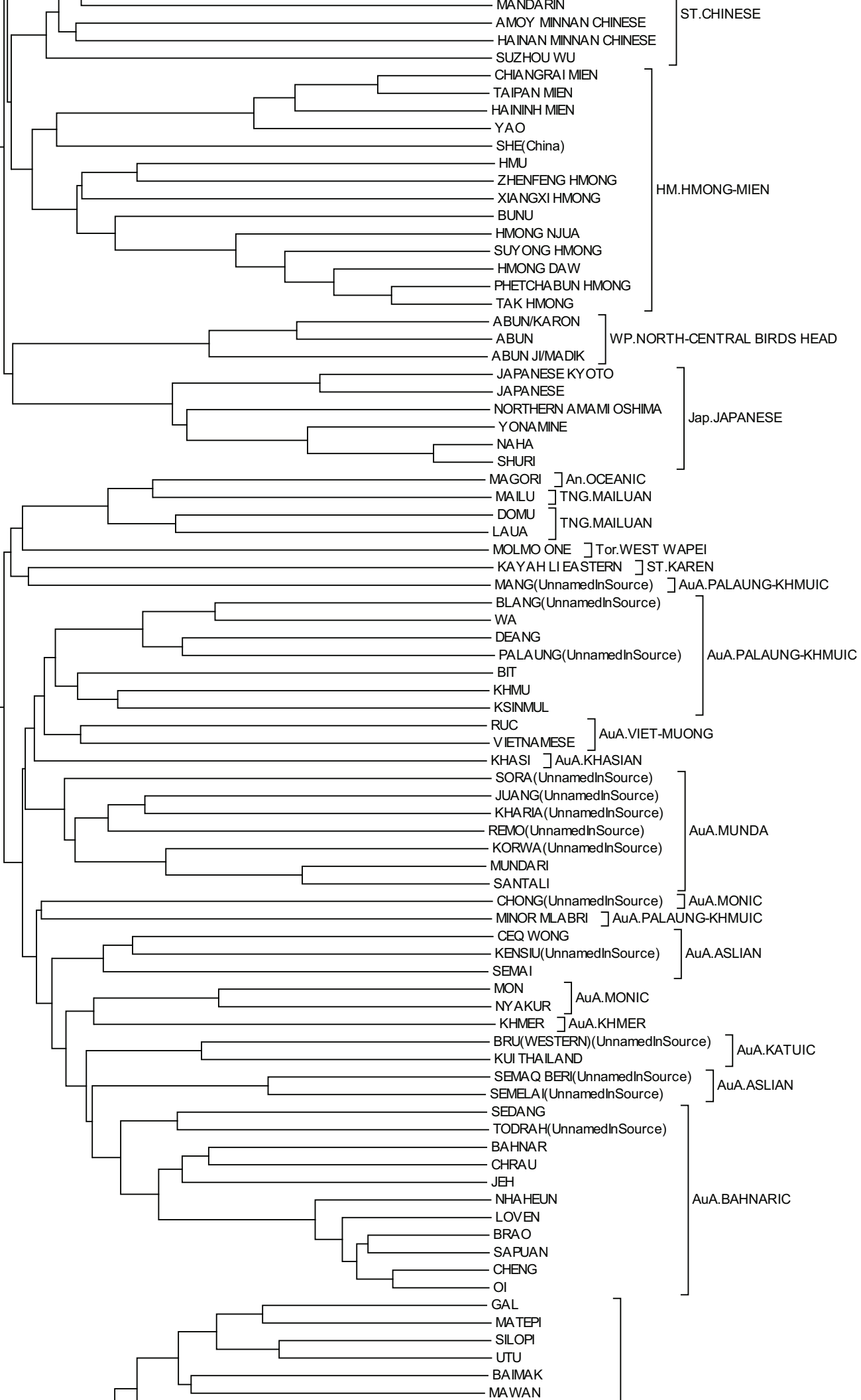
AWERA
 LP.LAKES PLAIN

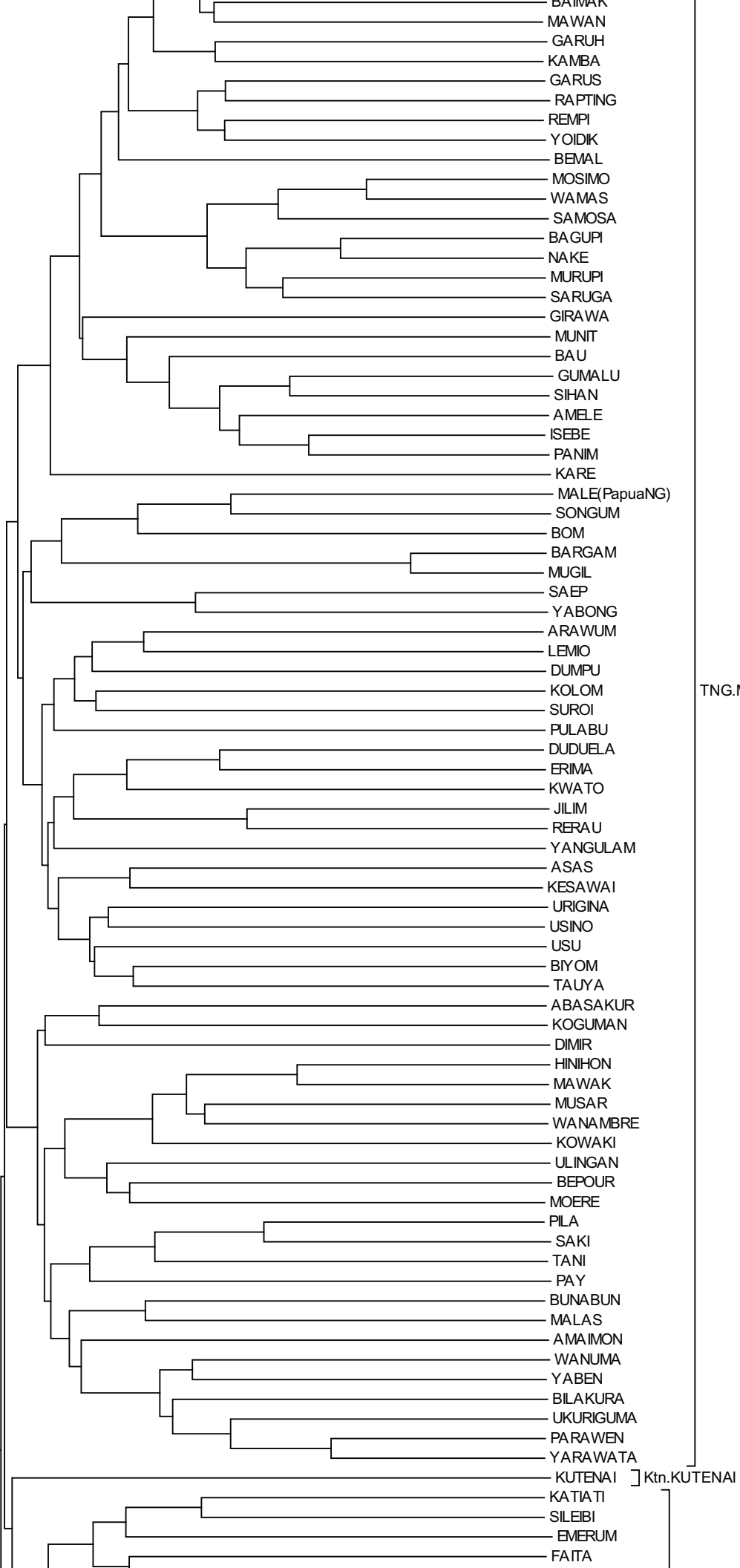
MAIRAS/FARANJAO
 MAIRASI
 SEMIMI(ETNA BAY)
 SEMIMI
 TNG.MAIRASI

KALULI







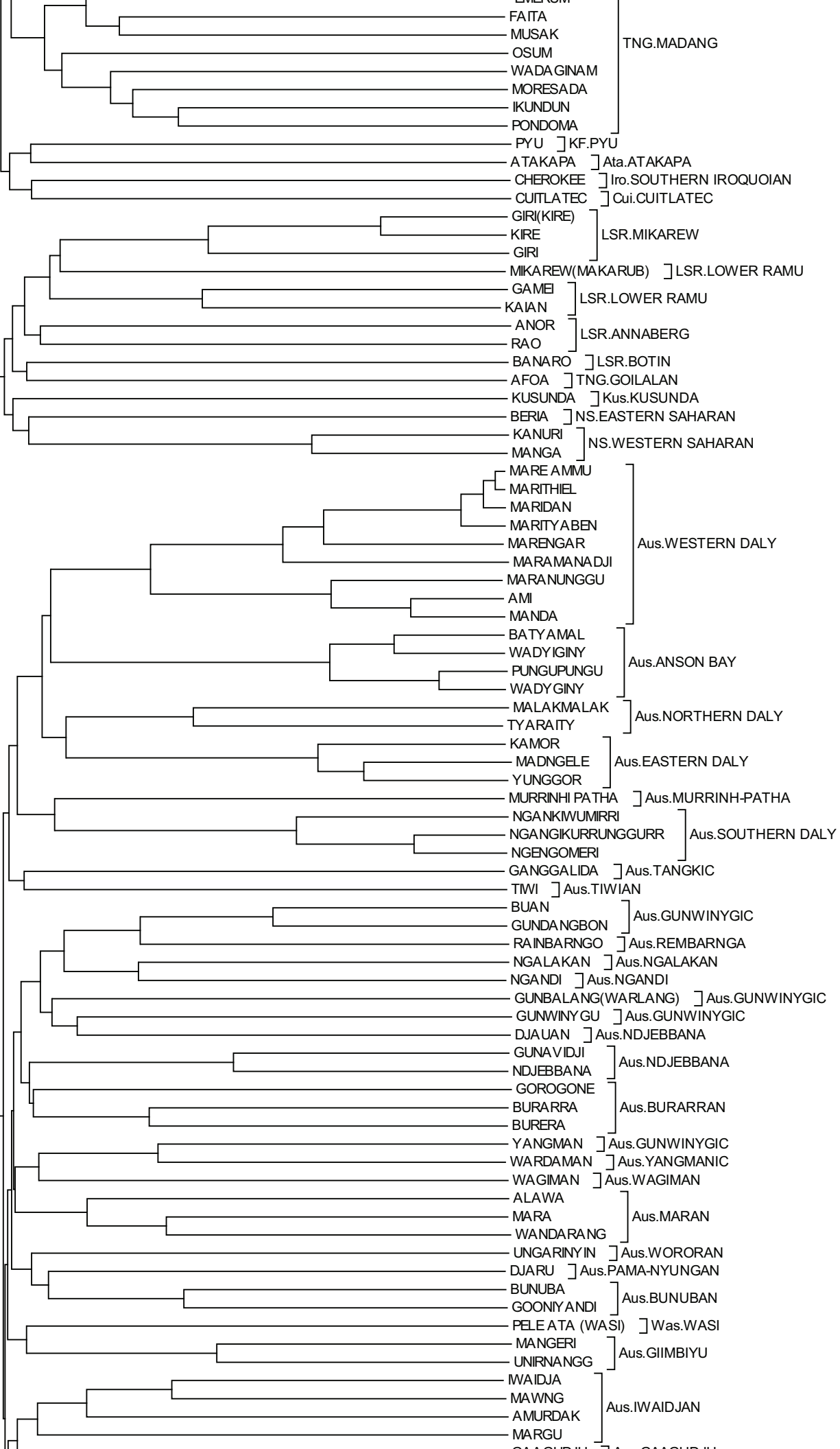


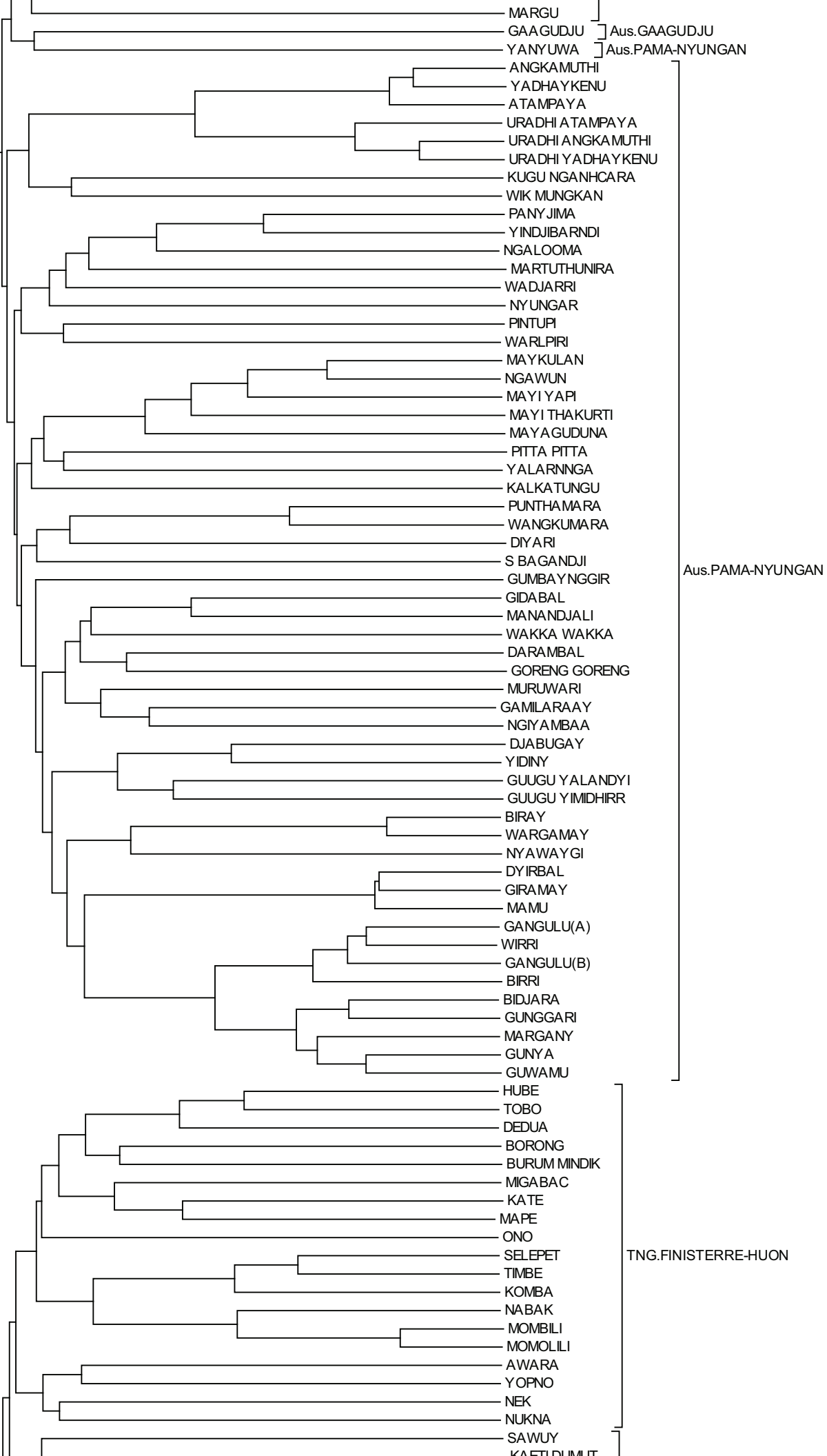
MAWAK
 GARUH
 KAMBA
 GARUS
 RAPTING
 REMPI
 YOIDIK
 BEMAL
 MOSIMO
 WAMAS
 SAMOSA
 BAGUPI
 NAKE
 MURUPI
 SARUGA
 GIRAWA
 MUNIT
 BAU
 GUMALU
 SIHAN
 AMELE
 ISEBE
 PANIM
 KARE
 MALE(PapuaNG)
 SONGUM
 BOM
 BARGAM
 MUGIL
 SAEF
 YABONG
 ARAWUM
 LEMIO
 DUMPU
 KOLOM
 SUROI
 PULABU
 DUDUELA
 ERIMA
 KWATO
 JILIM
 RERAU
 YANGULAM
 ASAS
 KESAWAI
 URIGINA
 USINO
 USU
 BIYOM
 TAUYA
 ABASAKUR
 KOGUMAN
 DIMIR
 HINIHON
 MAWAK
 MUSAR
 WANAMBRE
 KOWAKI
 ULINGAN
 BEPOUR
 MOERE
 PILA
 SAKI
 TANI
 PAY
 BUNABUN
 MALAS
 AMAIMON
 WANUMA
 YABEN
 BILAKURA
 UKURIGUMA
 PARAWEN
 YARAWATA

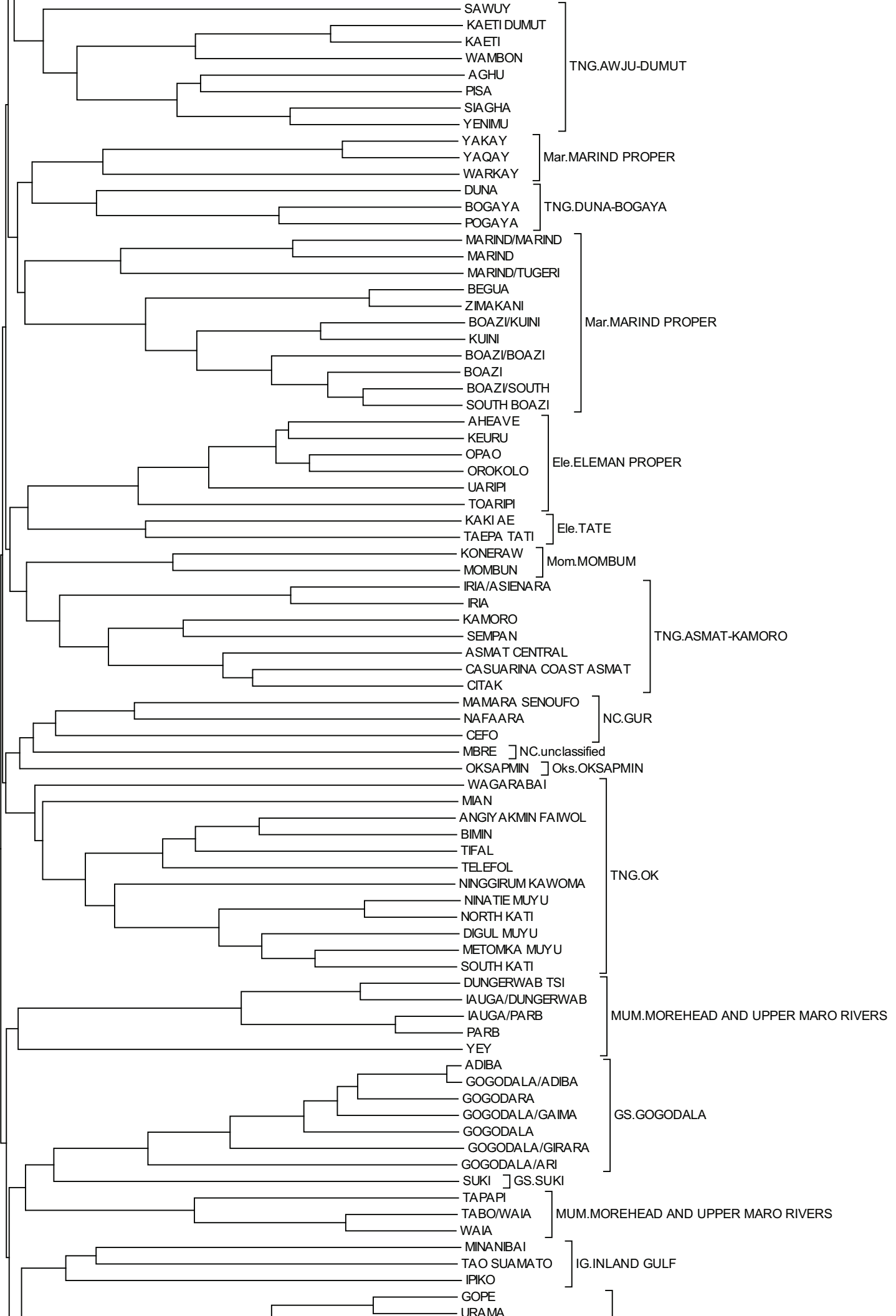
TNG.MADANG

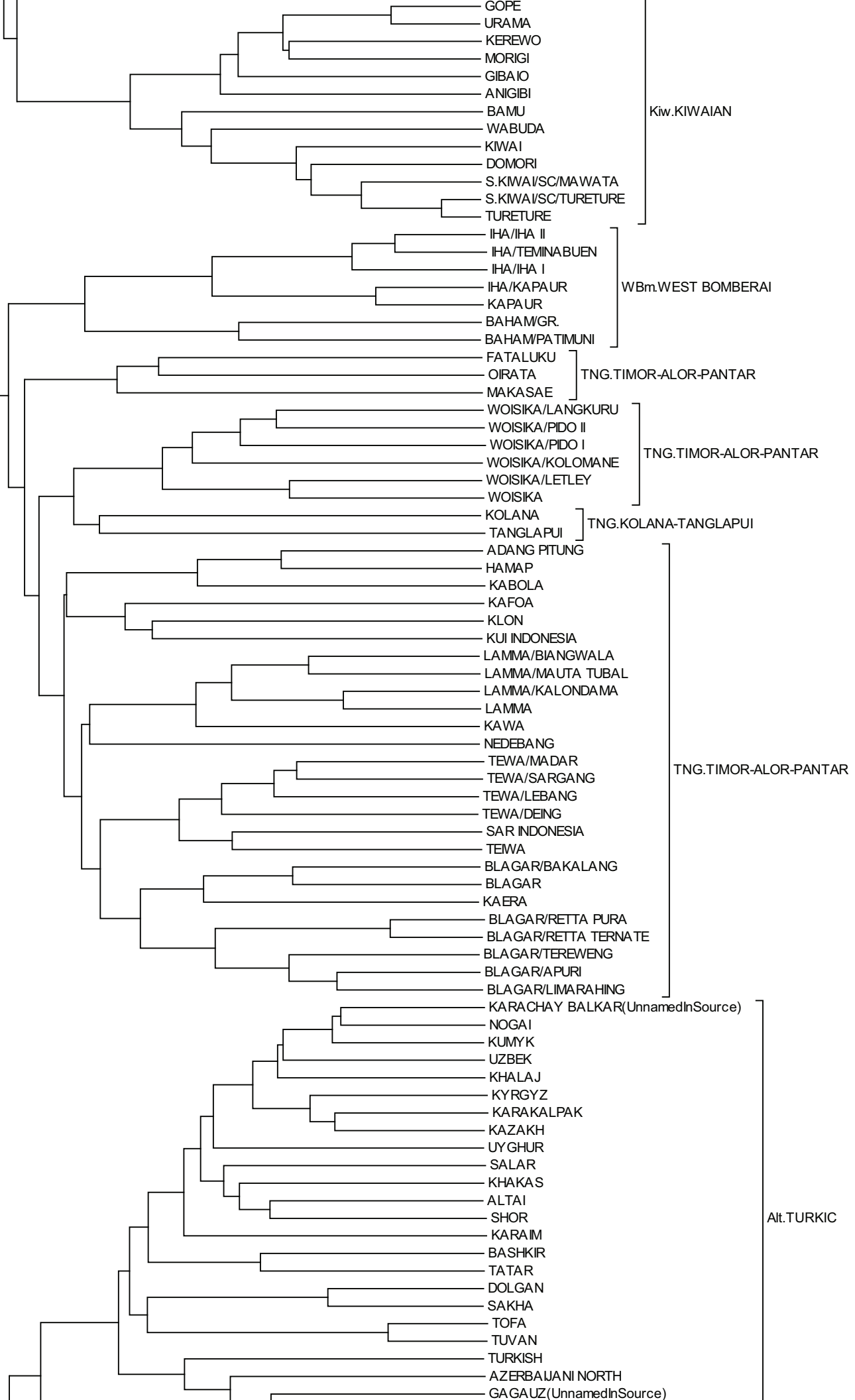
] Ktn.KUTENAI

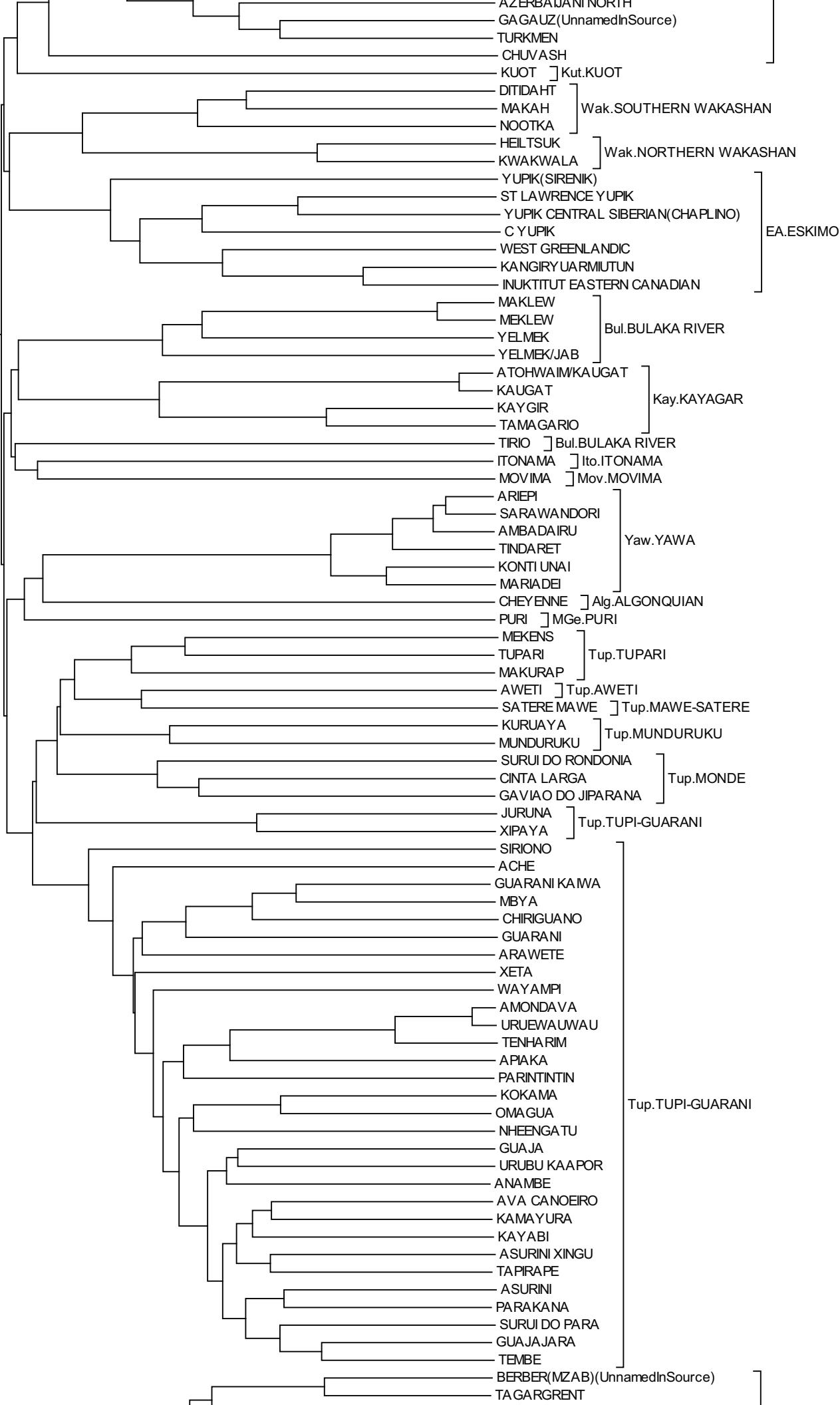
KUTENAI
 KATIATI
 SILEBI
 EMERUM
 FAITA

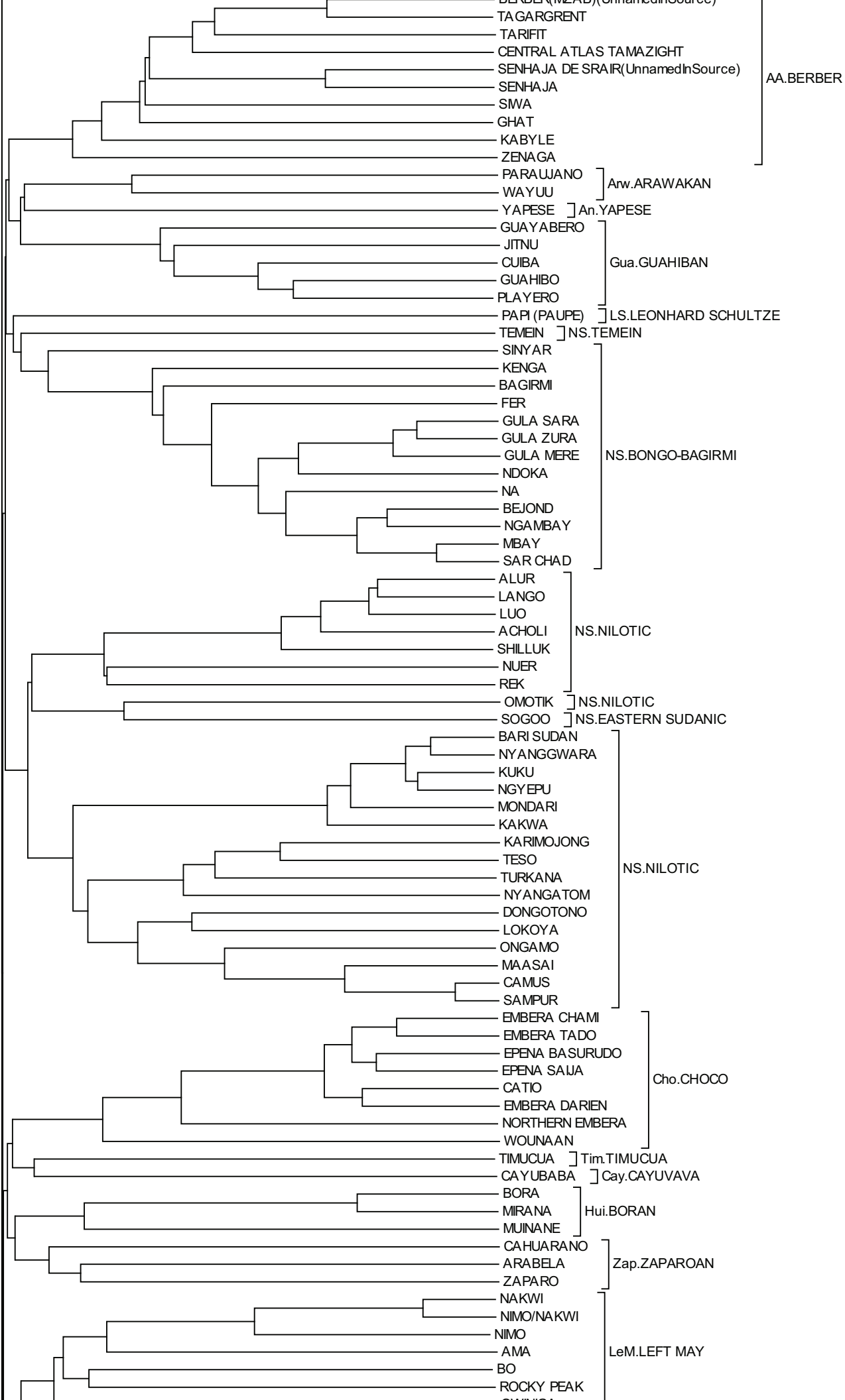


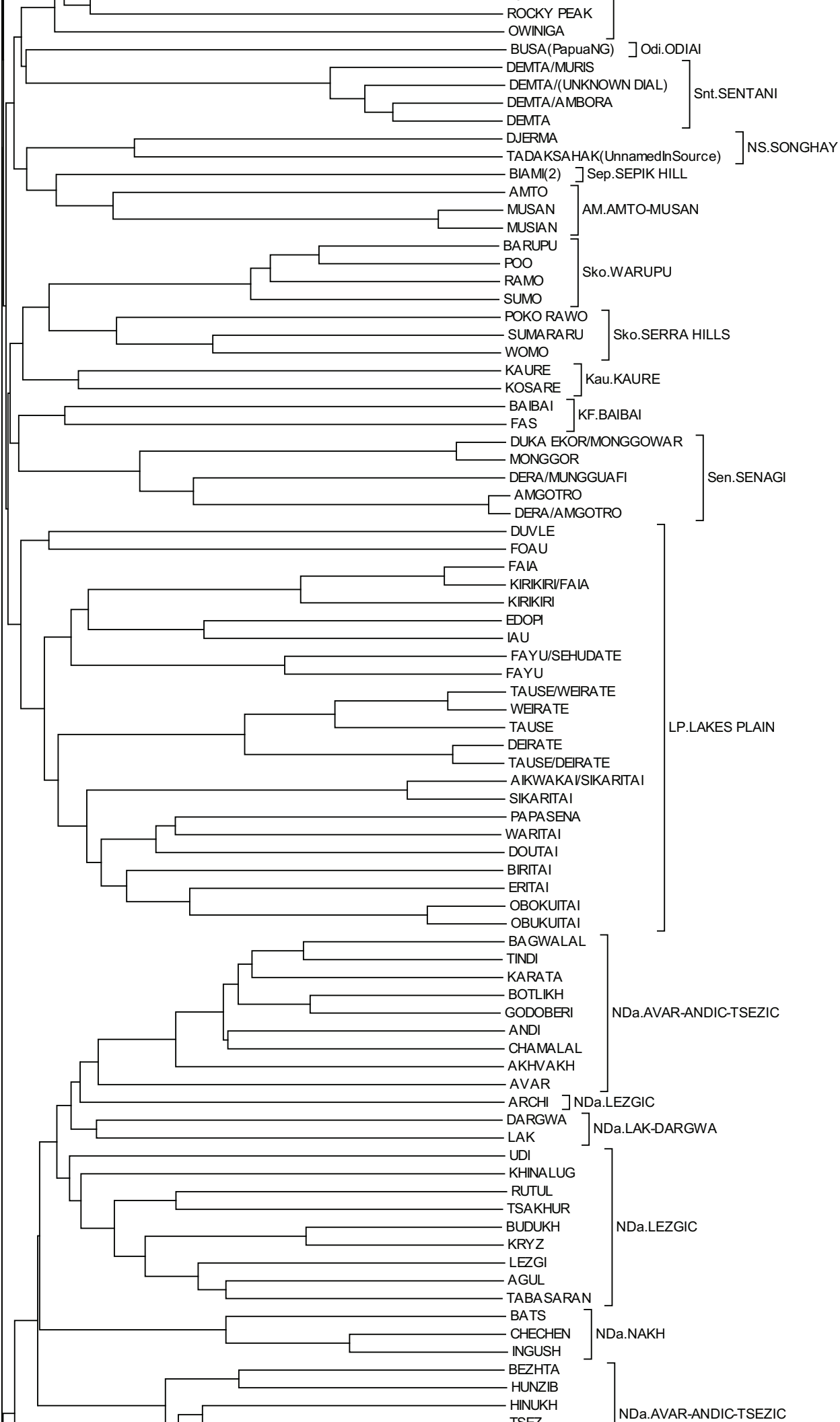


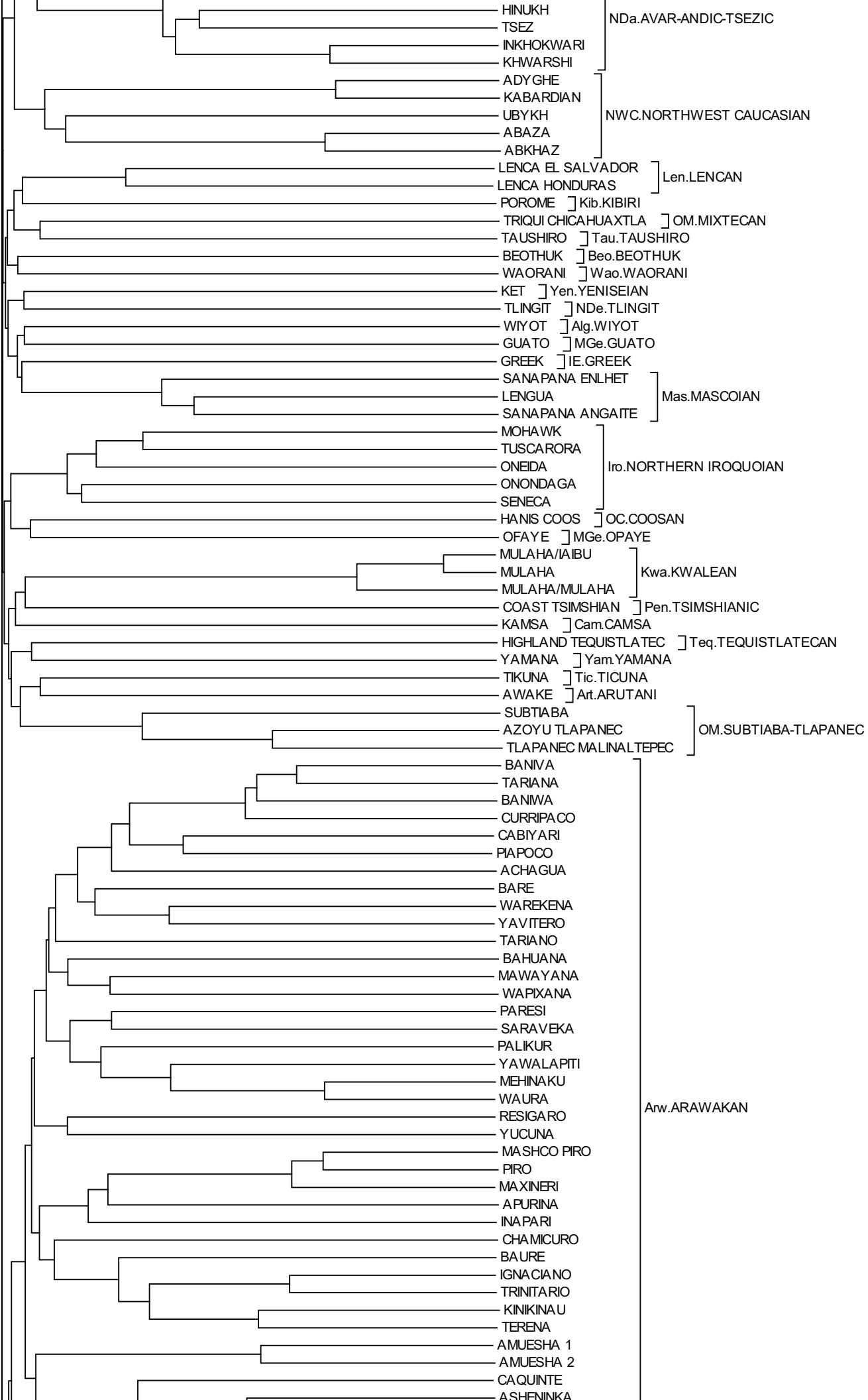


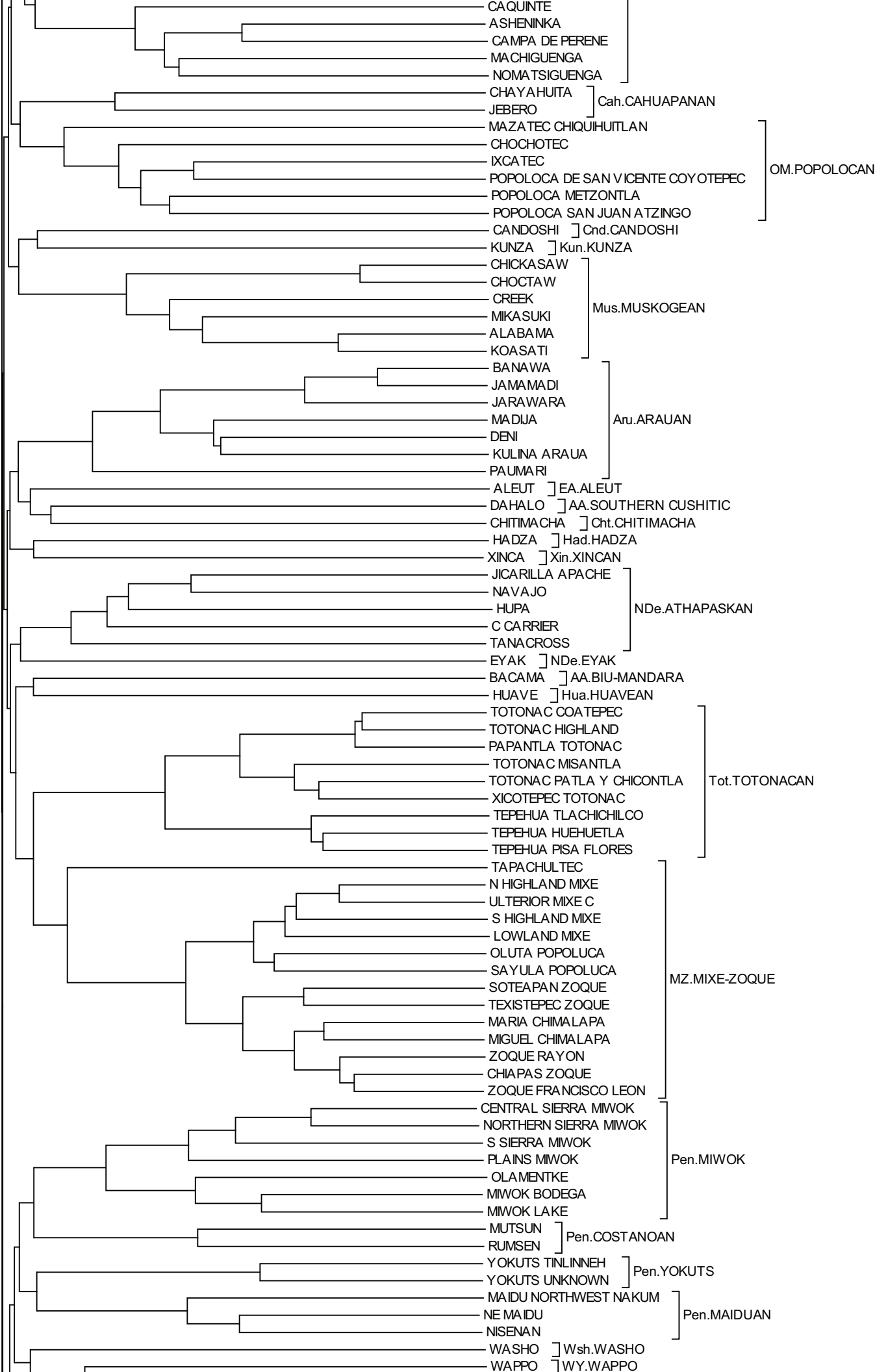


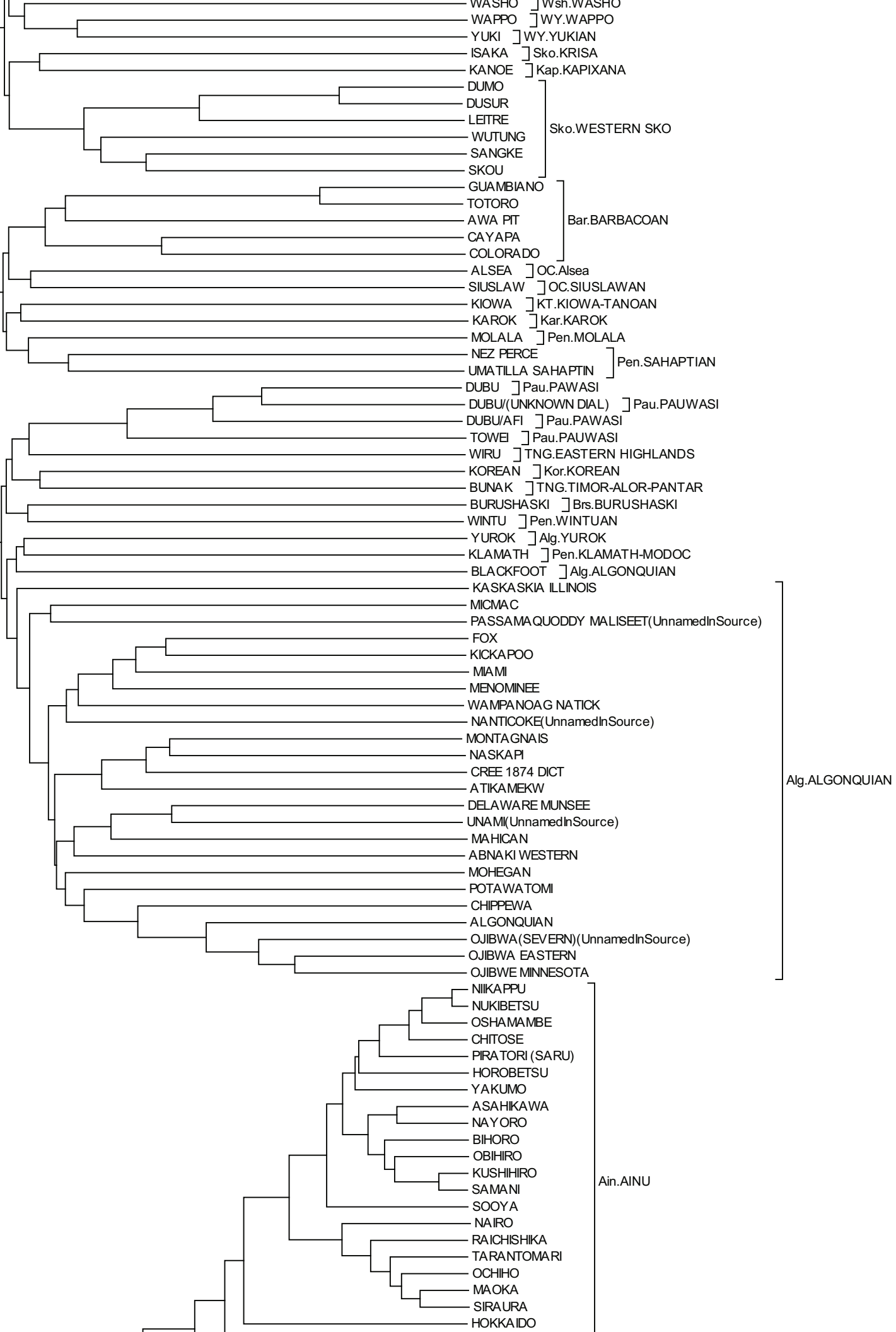


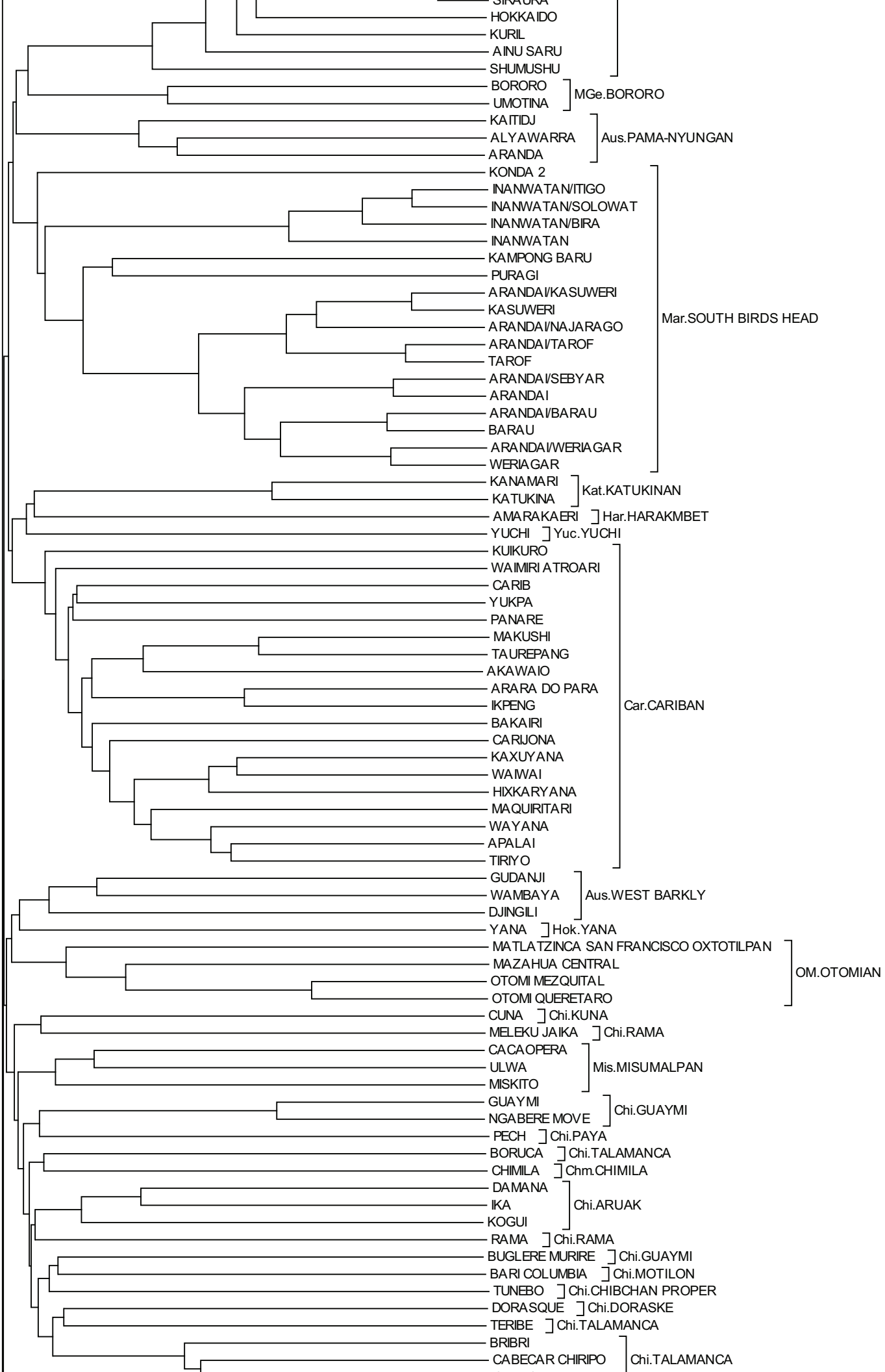


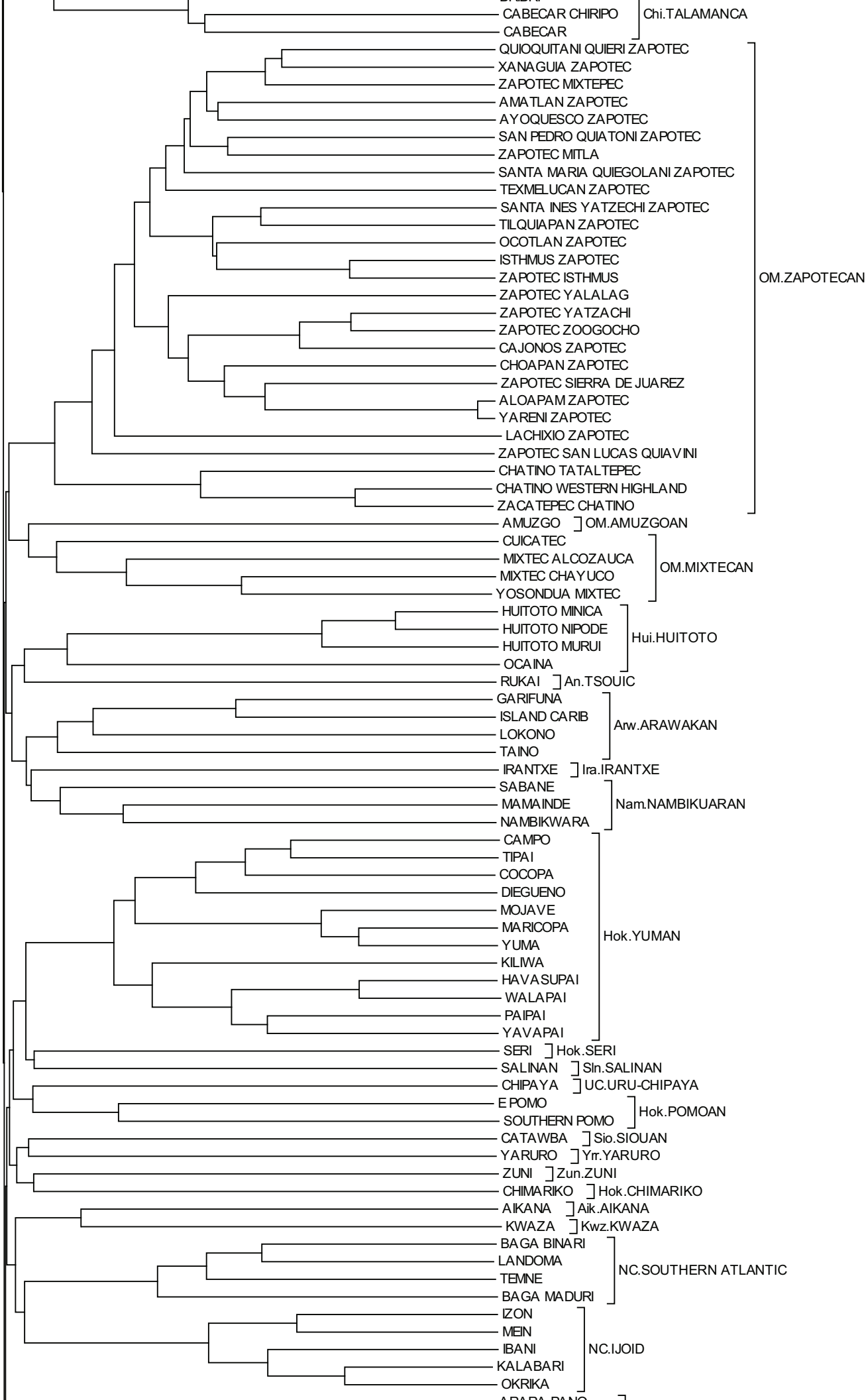


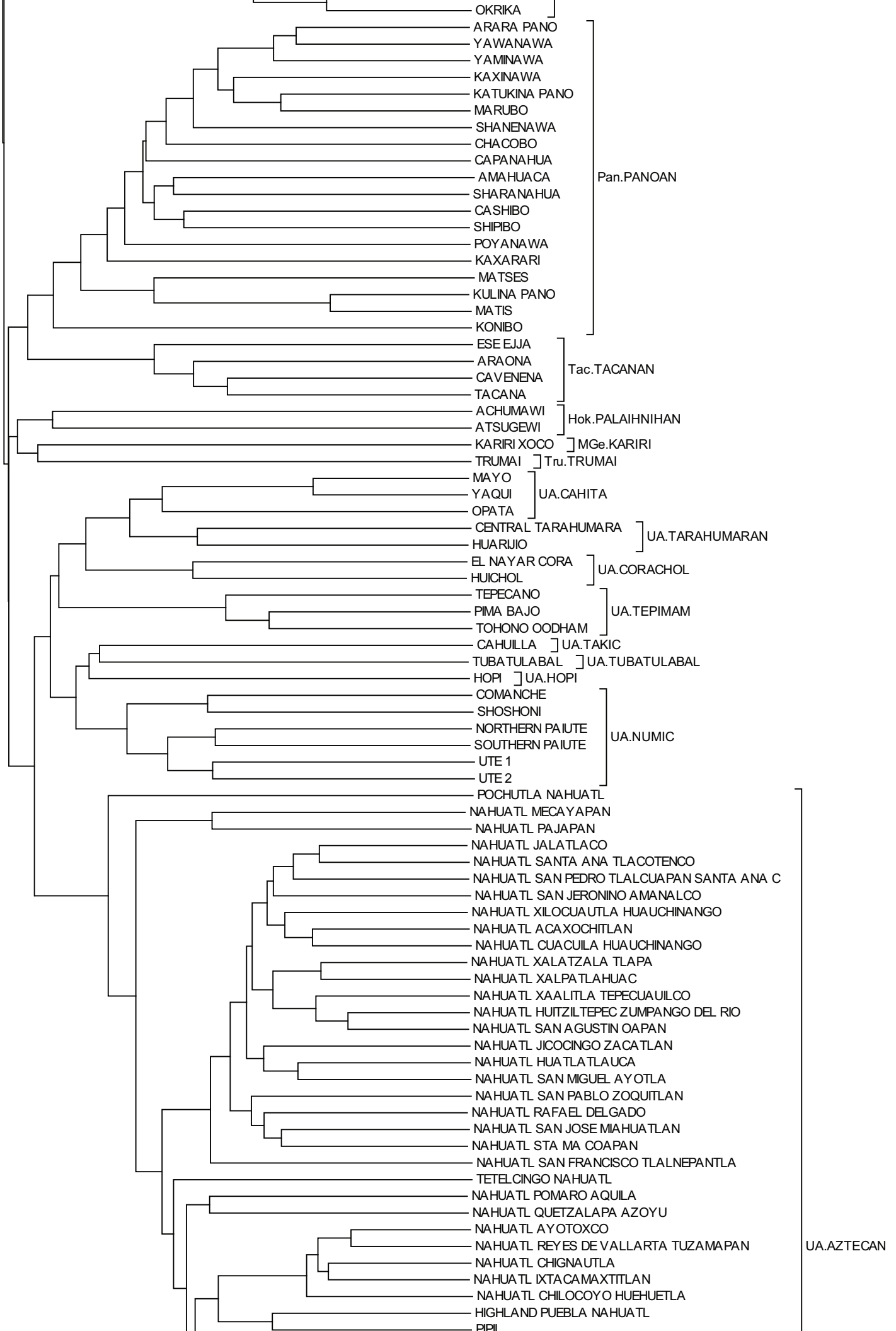


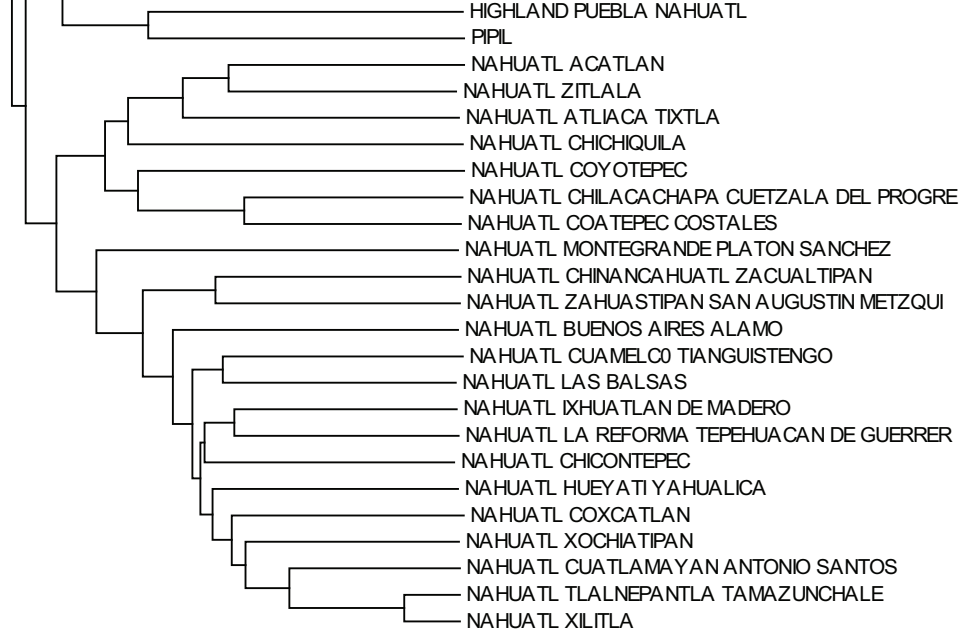












┌──────────┐
 10