

# **ASJP World Language Tree of Lexical Similarity: Version 2 (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 ASJP tree are lexically more similar than those branched less closely together. While most lexical resemblance charted in the tree almost certainly is related to genetic affiliation, closely branched languages cannot routinely be assumed to be closely genetically associated since lexical resemblance among languages can be due to factors other than genetic relatedness (see below).

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),<sup>1</sup> 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, see 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 a midpoint, i.e., the point which is equidistant between the two most lexically dissimilar 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),<sup>2</sup> with some updates from Dryer (personal communication). This annotation is presented for ease of orientation, not necessarily because ASJP agrees with it. The language names used are normally simply those

---

<sup>1</sup> <http://www.megasoftware.net/>

<sup>2</sup> <http://wals.info/>

of the sources consulted. The sources, as well as corresponding language names of *Ethnologue*, are provided in a continuously updated wiki.<sup>3</sup>

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 co-members of its family that it is found branched more closely with some language or languages with which it is not genetically related at all, 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

---

<sup>3</sup> <http://lingweb.eva.mpg.de/asjp/index.php/ASJP>

more likely explained by chance than by either inheritance or diffusion. Also, language isolates may join one another on a terminal branch because they have nowhere else to go in the tree, creating the illusion that exciting, new far-flung relations may be in evidence. One should be cautious in the interpretation of these cases.

## References

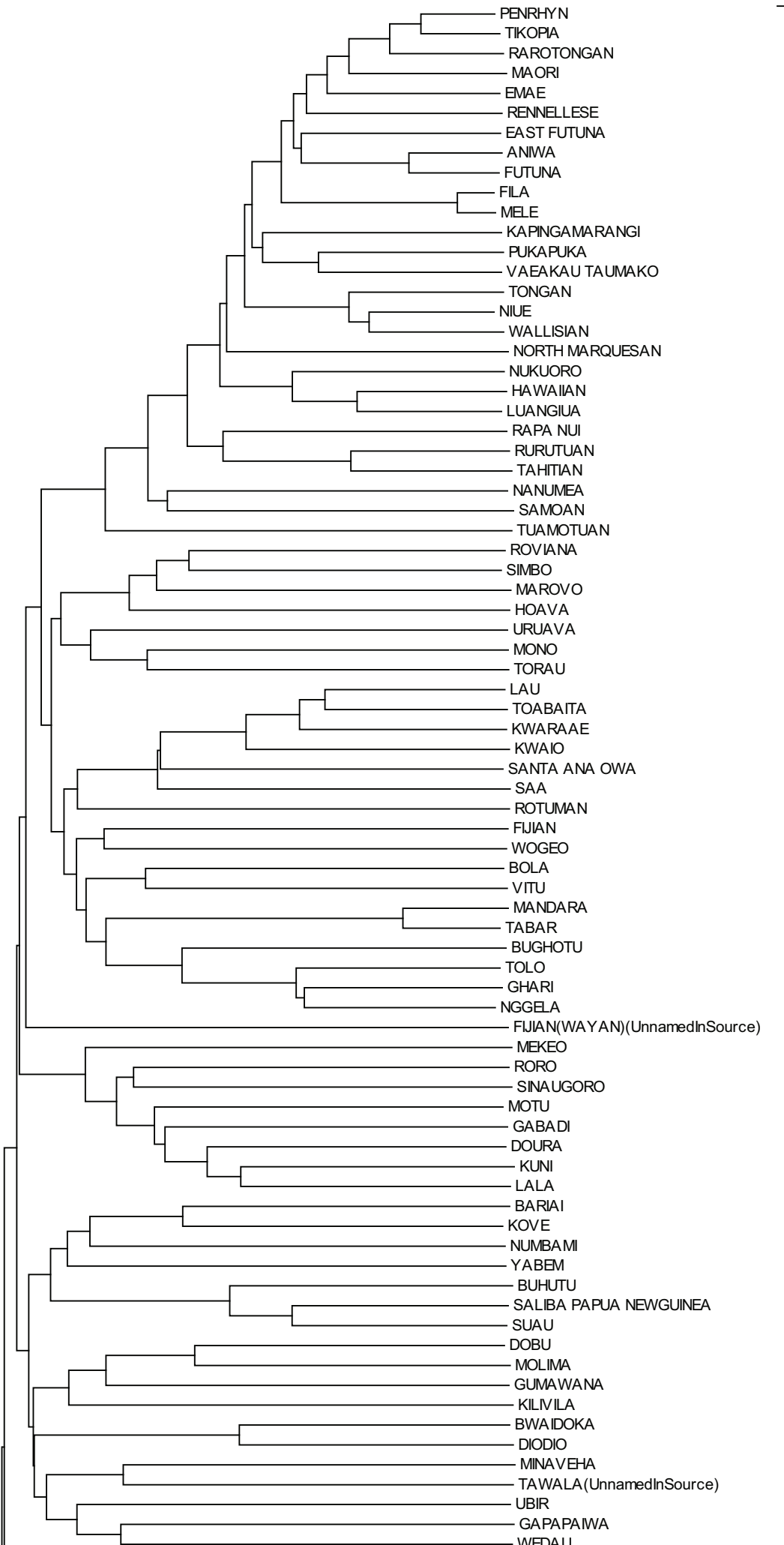
- Bakker, Dik, André Müller, Viveka Velupillai, Søren Wichmann, Cecil H. Brown, Pamela Brown, Dmitry Egorov, Robert Mailhammer, Anthony Grant, and Eric W. Holman. 2009. Adding typology to lexicostatistics: a combined approach to language classification. *Linguistic Typology* 13: 167-179.
- Holman, Eric 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

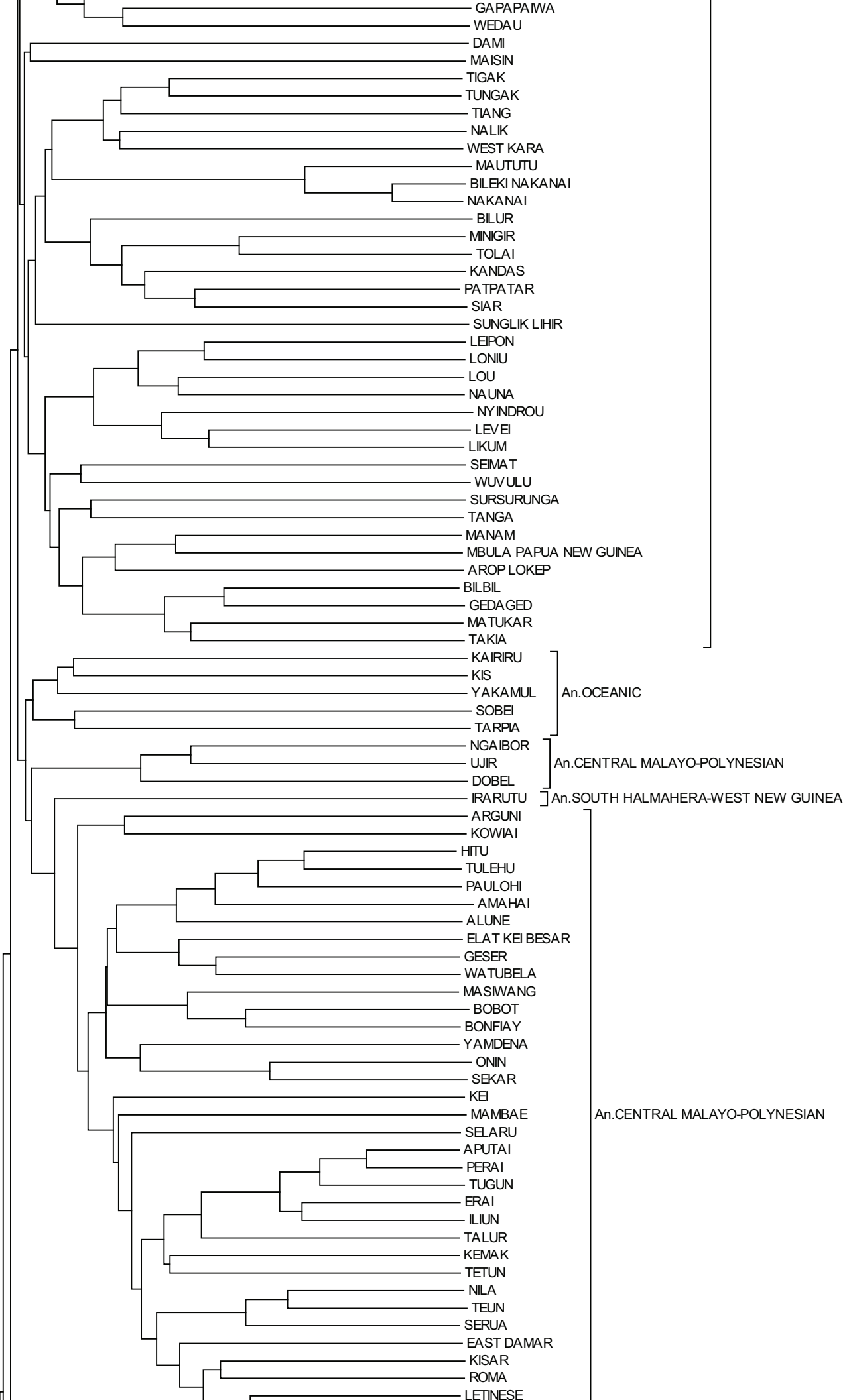
AA	Afro-Asiatic	Chu	Chumash
Aik	Aikana	CK	Chukotko-Kamchatkan
Ain	Ainu	Cmu	Chimúan
Ala	Alacalufan	CN	Cacua-Nukak
Alg	Algic	Cnd	Candoshi
Alt	Altaic	Cof	Cofán
AM	Amto-Musan	Com	Comecrudan
An	Austronesian	Cre	Creoles&Pidgins
AP	Awin-Pare	Cui	Cuitlatec
Arc	Araucanian	CW	Chapacura-Wanhan
Art	Arutani	Dos	Doso
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

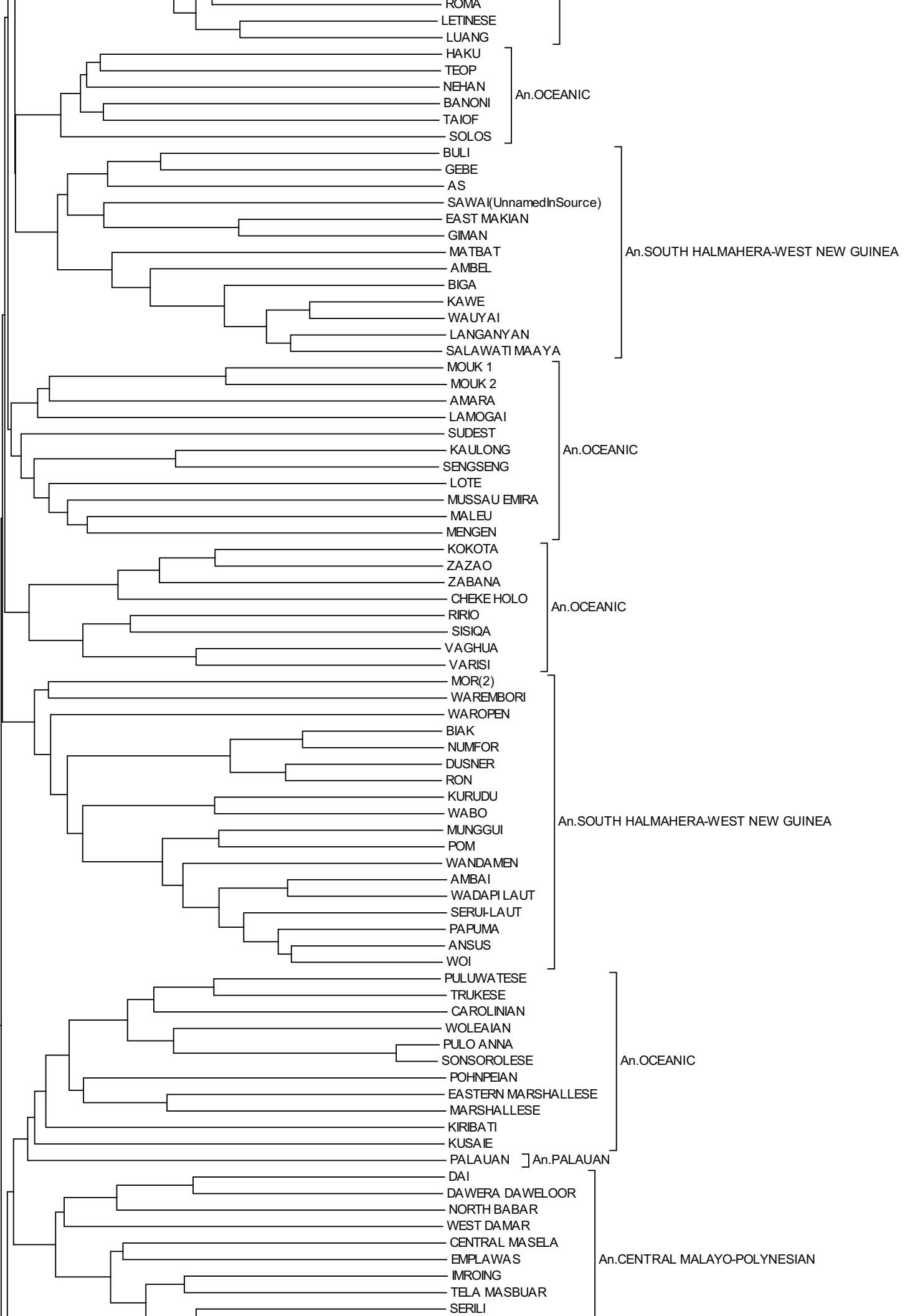
ST	Sino-Tibetan	Yal	Yale
Sum	Sumerian	Yam	Yamana
Tac	Tacanan	Yan	Yanomam
Tar	Tarascan	Yaw	Yawa
Tau	Taushiro	Yel	Yele
Teb	Teberan-Pawaiian	Yen	Yeniseian
Teq	Tequistlatecan	Yka	Yukaghir
Tic	Ticuna	Yrb	Yareban
Tim	Timucua	Yrr	Yaruro
TK	Tai-Kadai	Yuc	Yuchi
TNG	Trans-New Guinea	Yuw	Yuwana
TO	Tor-Orya	Zam	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		

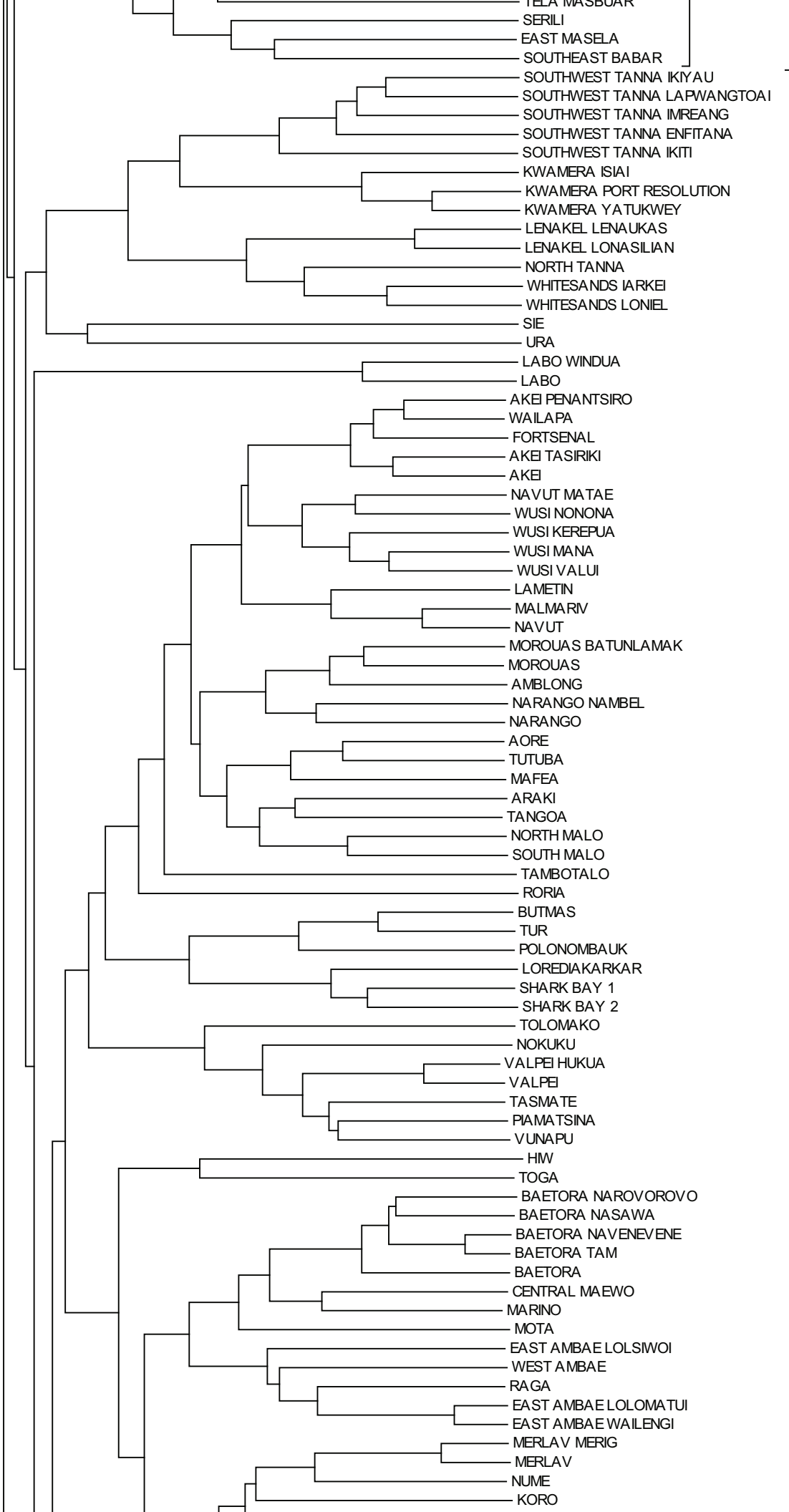


An.OCEANIC









TELA MASBUAR

SERILI

EAST MA SELA

SOUTHWEST BABAR

SOUTHWEST TAINNA IKYAU

SOUTHWEST TAINNA LAPWANGTOAI

SOUTHWEST TAINNA IMREANG

SOUTHWEST TAINNA ENFITANA

SOUTHWEST TAINNA IKITI

KWAMERA ISIAI

KWAMERA PORT RESOLUTION

KWAMERA YATUKWEY

LENAKEL LENAUKAS

LENAKEL LONASILIAN

NORTH TAINNA

WHITESANDS IARKEI

WHITESANDS LONIEL

SIE

URA

LABO WINDUA

LABO

AKAI PENANTSIRO

WAILAPA

FORTSENAL

AKAI TASIRIKI

AKAI

NAVUT MATAE

WUSI NONONA

WUSI KEREPUA

WUSI MANA

WUSI VALUI

LAMETIN

MALMARIV

NAVUT

MOROUA S BATUNLAMAK

MOROUA S

AMBLONG

NARANGO NAMBEL

NARANGO

AORE

TUTUBA

MAFEA

ARAKI

TANGOA

NORTH MALO

SOUTH MALO

TAMBOTALO

RORIA

BUTMAS

TUR

POLONOMBAUK

LOREDIAKARKAR

SHARK BAY 1

SHARK BAY 2

TOLOMAKO

NOKUKU

VALPEI HUKUA

VALPEI

TASMA TE

PIAMATSINA

VUNAPU

HIW

TOGA

BAETORA NAROVOROVO

BAETORA NASAWA

BAETORA NAVENEVENE

BAETORA TAM

BAETORA

CENTRAL MAEWO

MARINO

MOTA

EAST AMBAE LOLSIWOI

WEST AMBAE

RAGA

EAST AMBAE LOLOMATUI

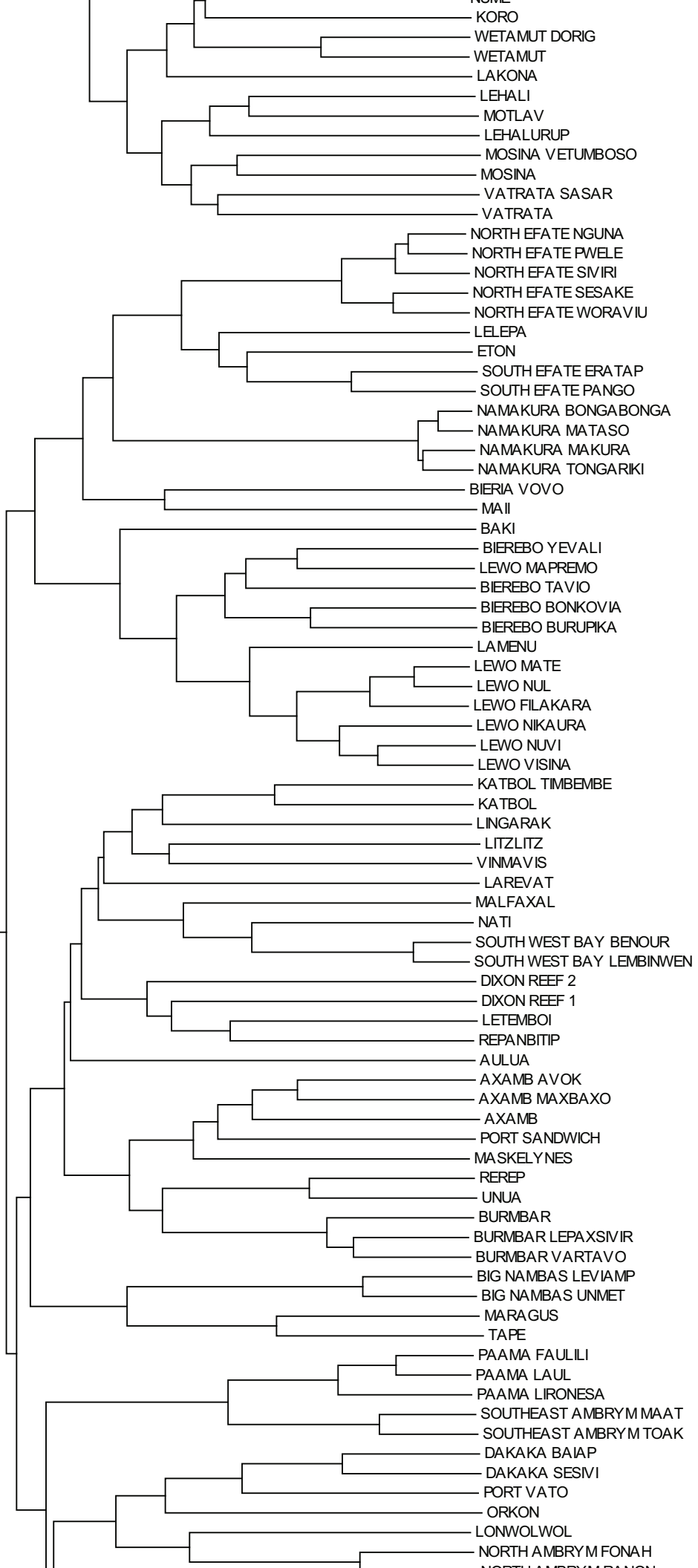
EAST AMBAE WAILENGI

MERLAV MERIG

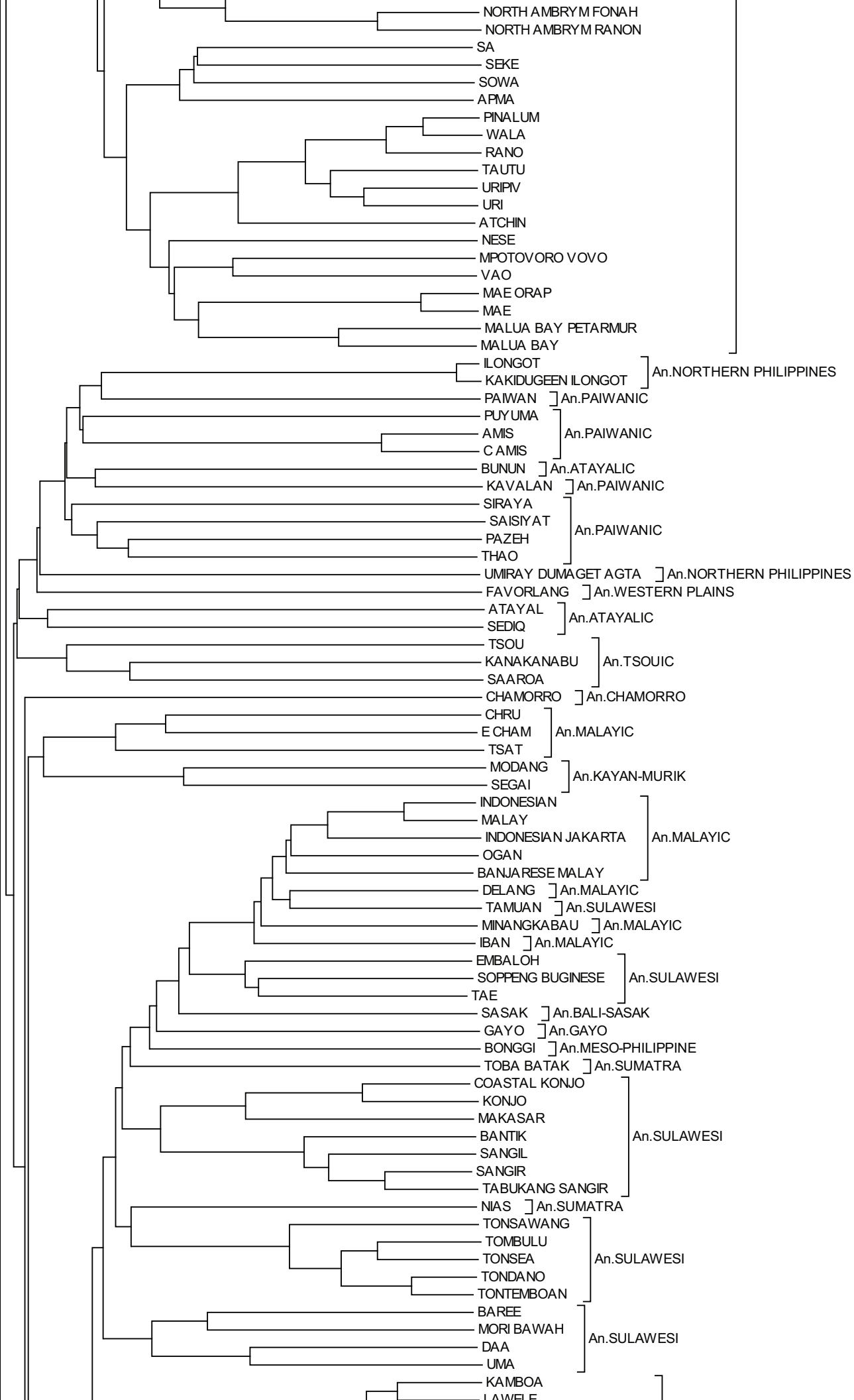
MERLAV

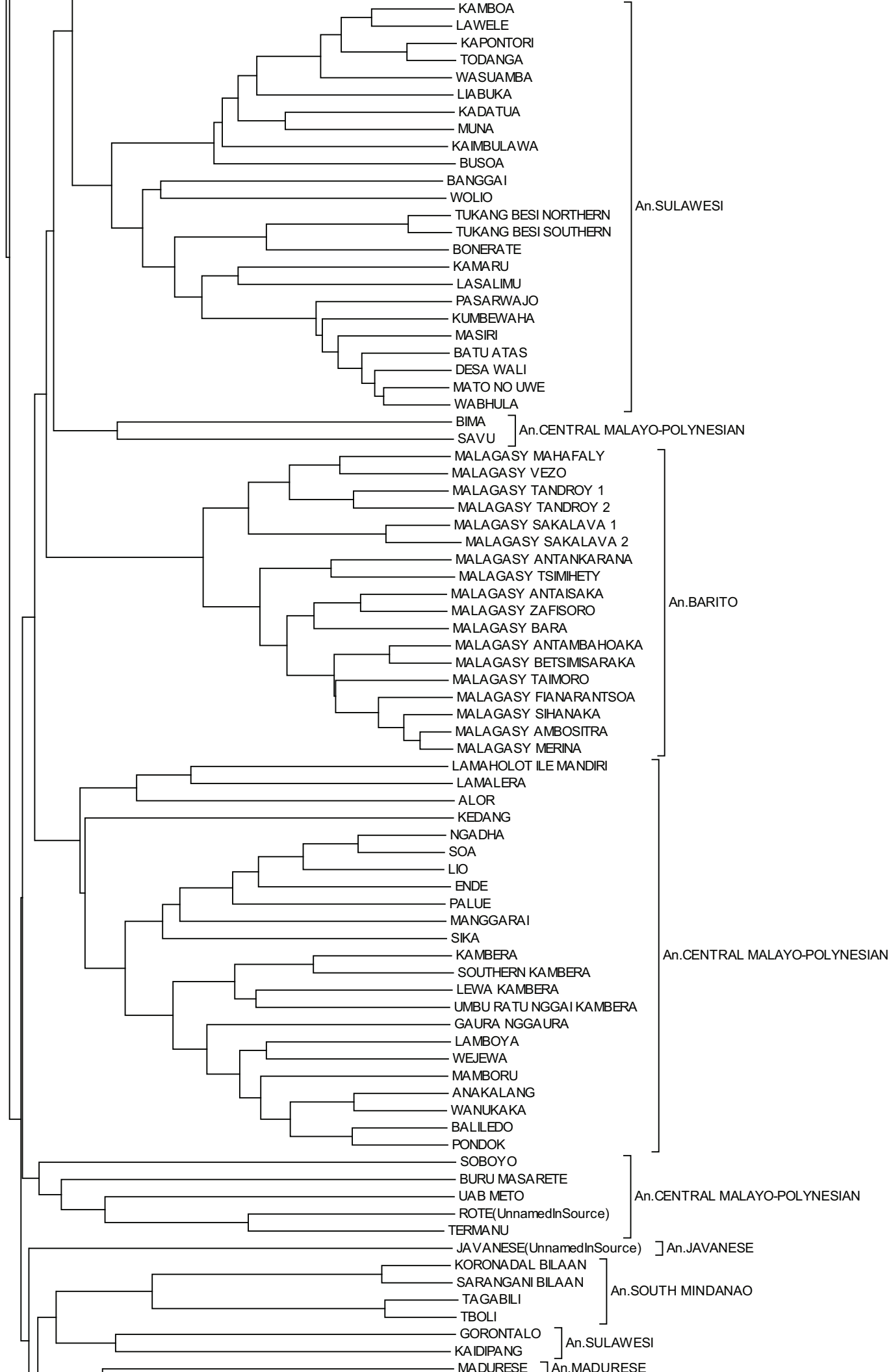
NUME

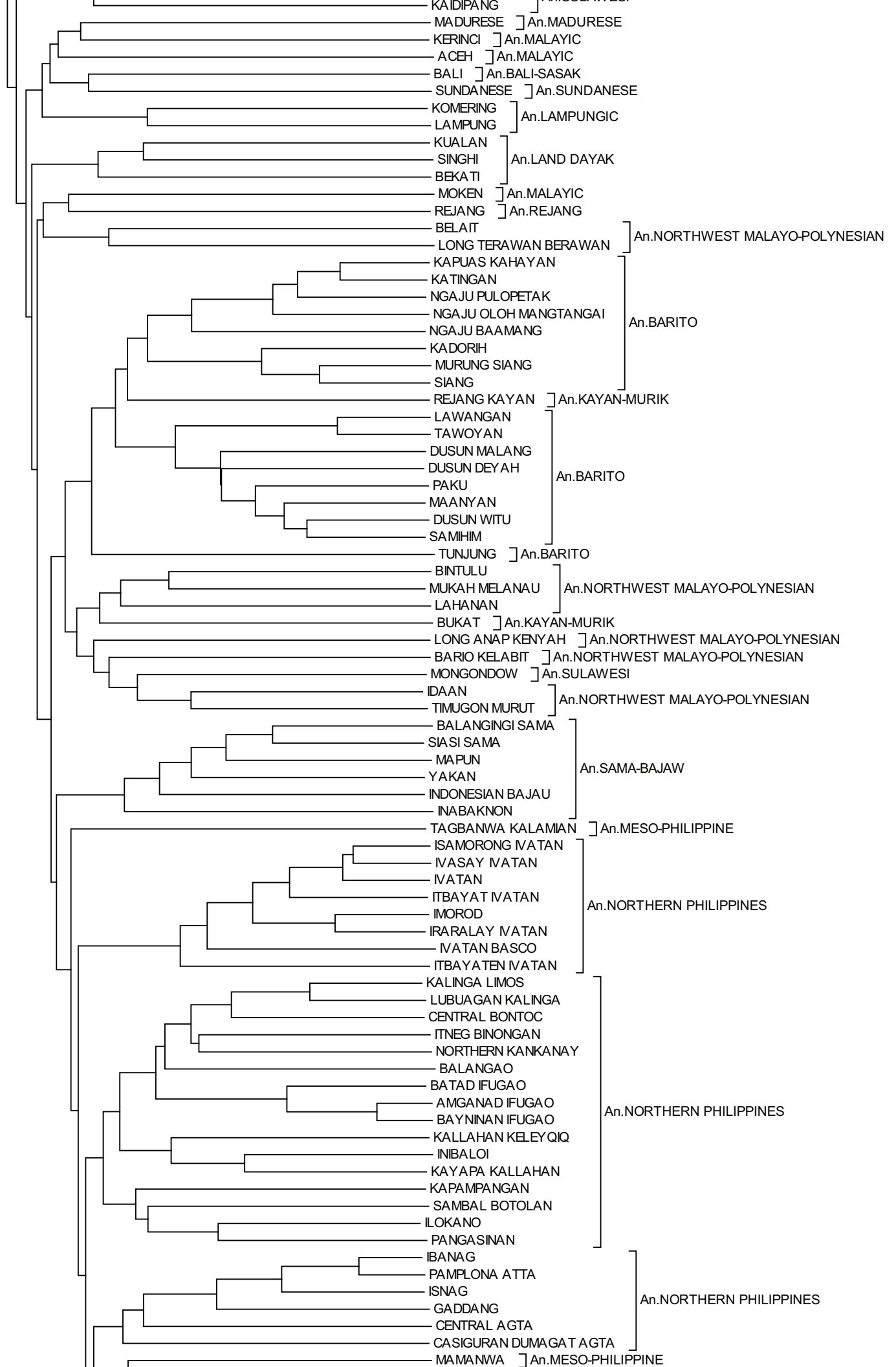
KORO

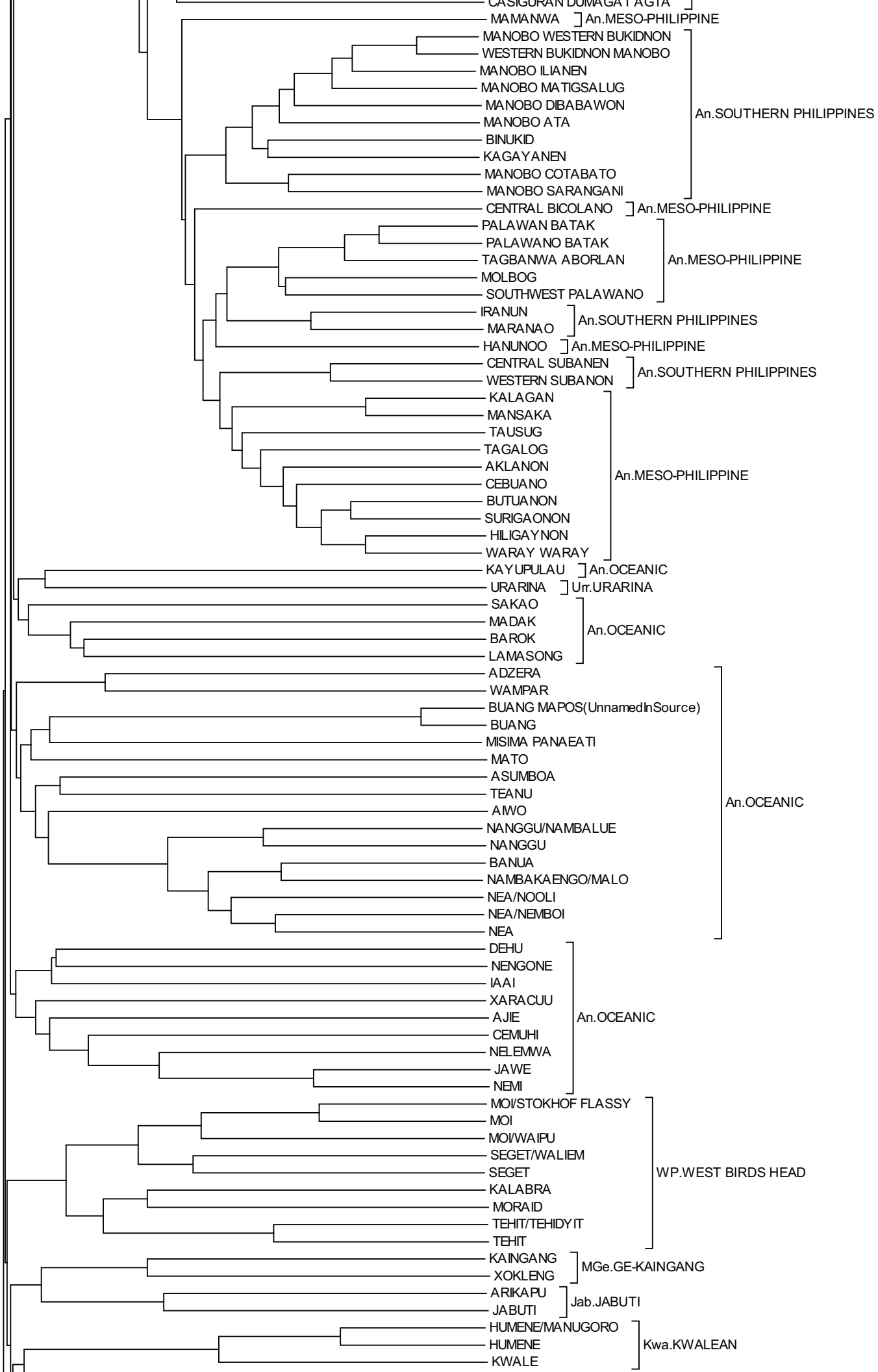


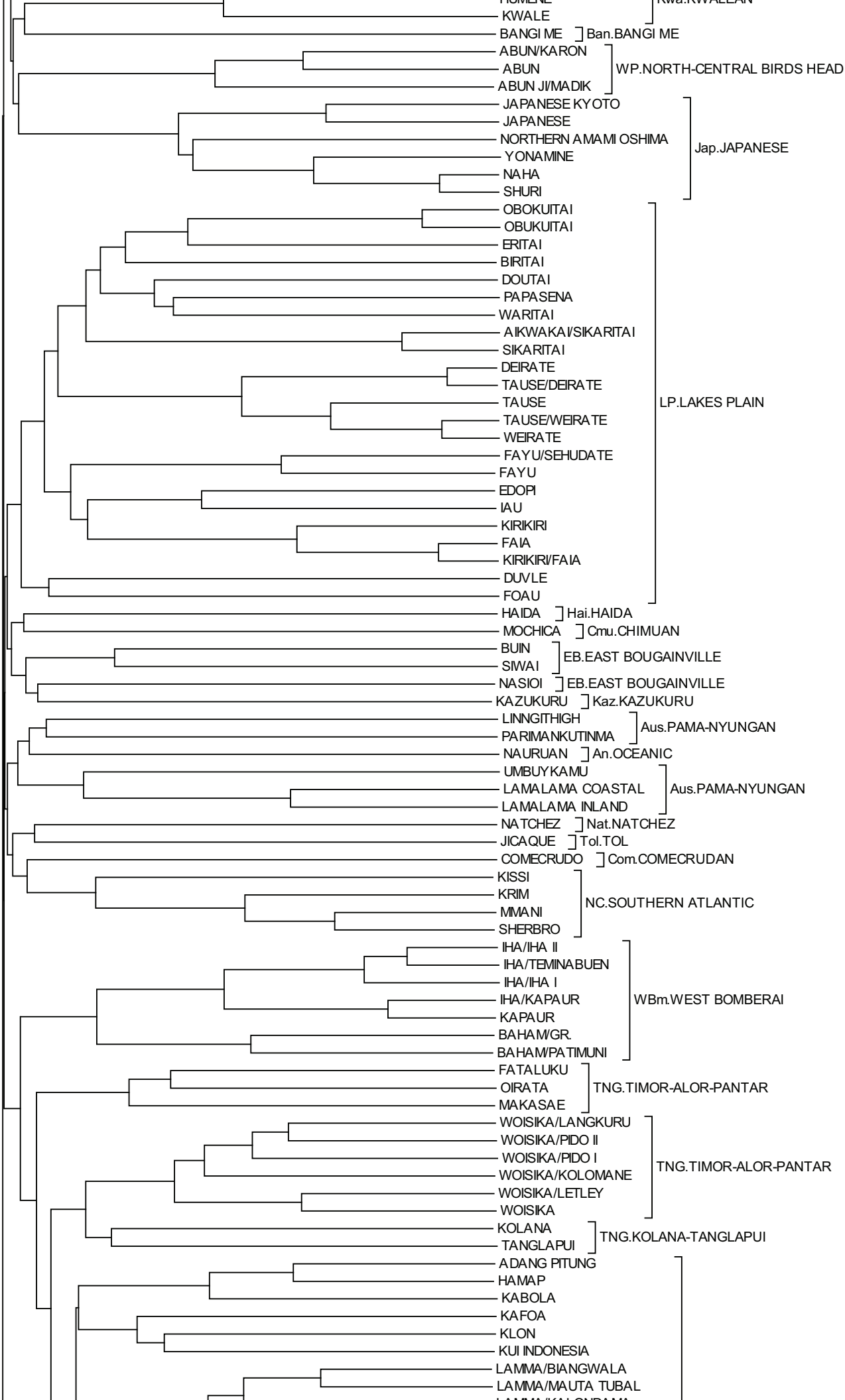
An.OCEANIC



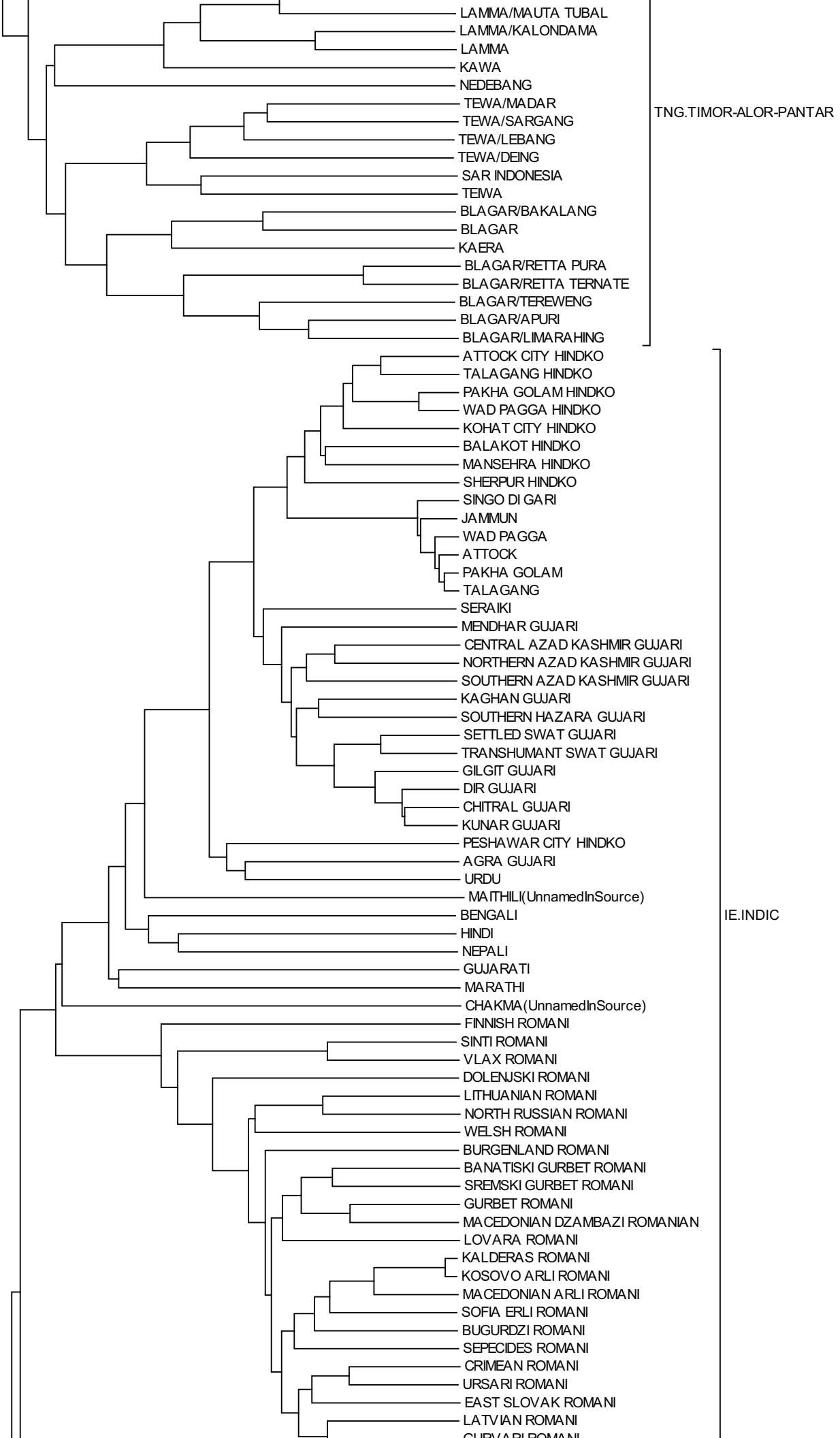


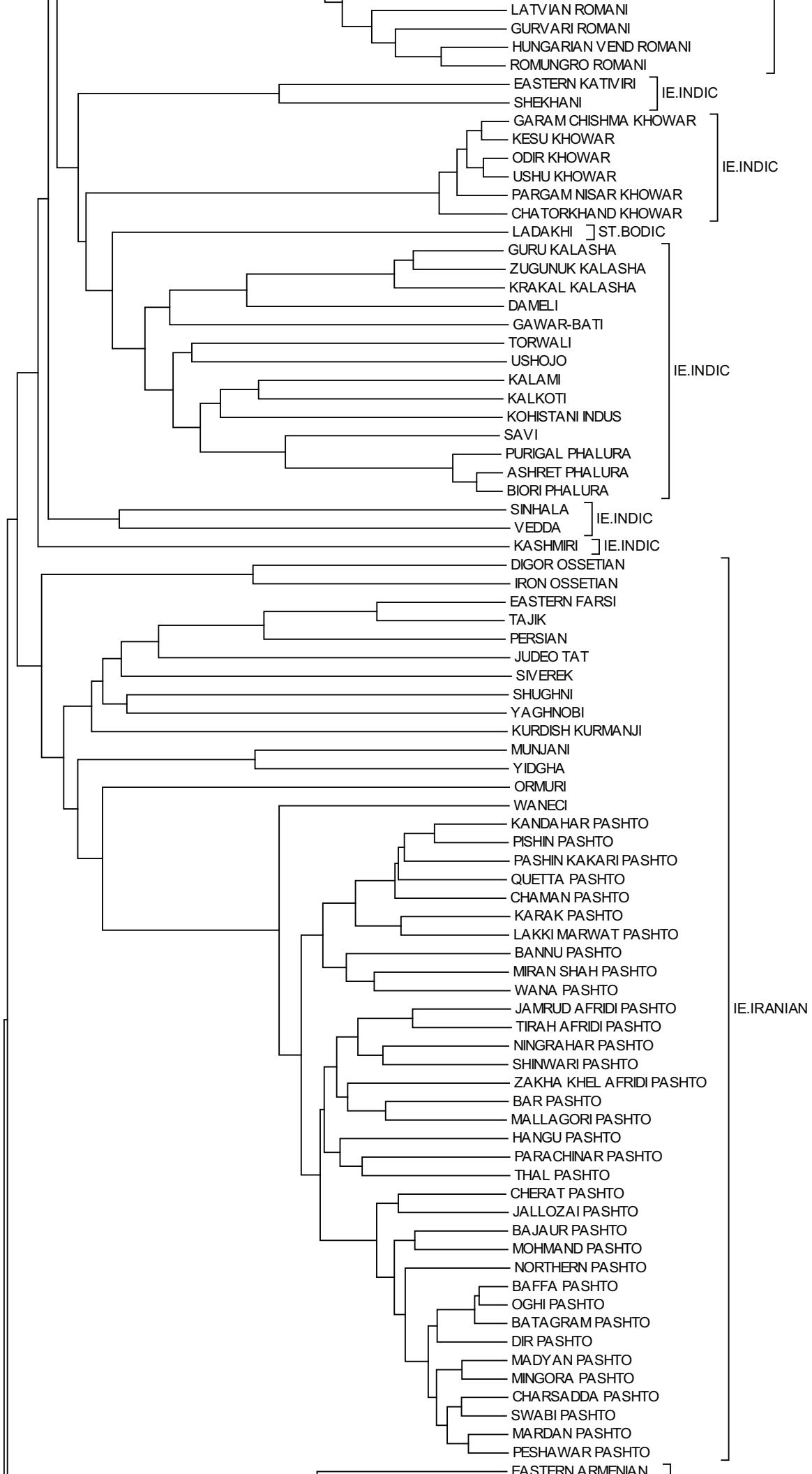


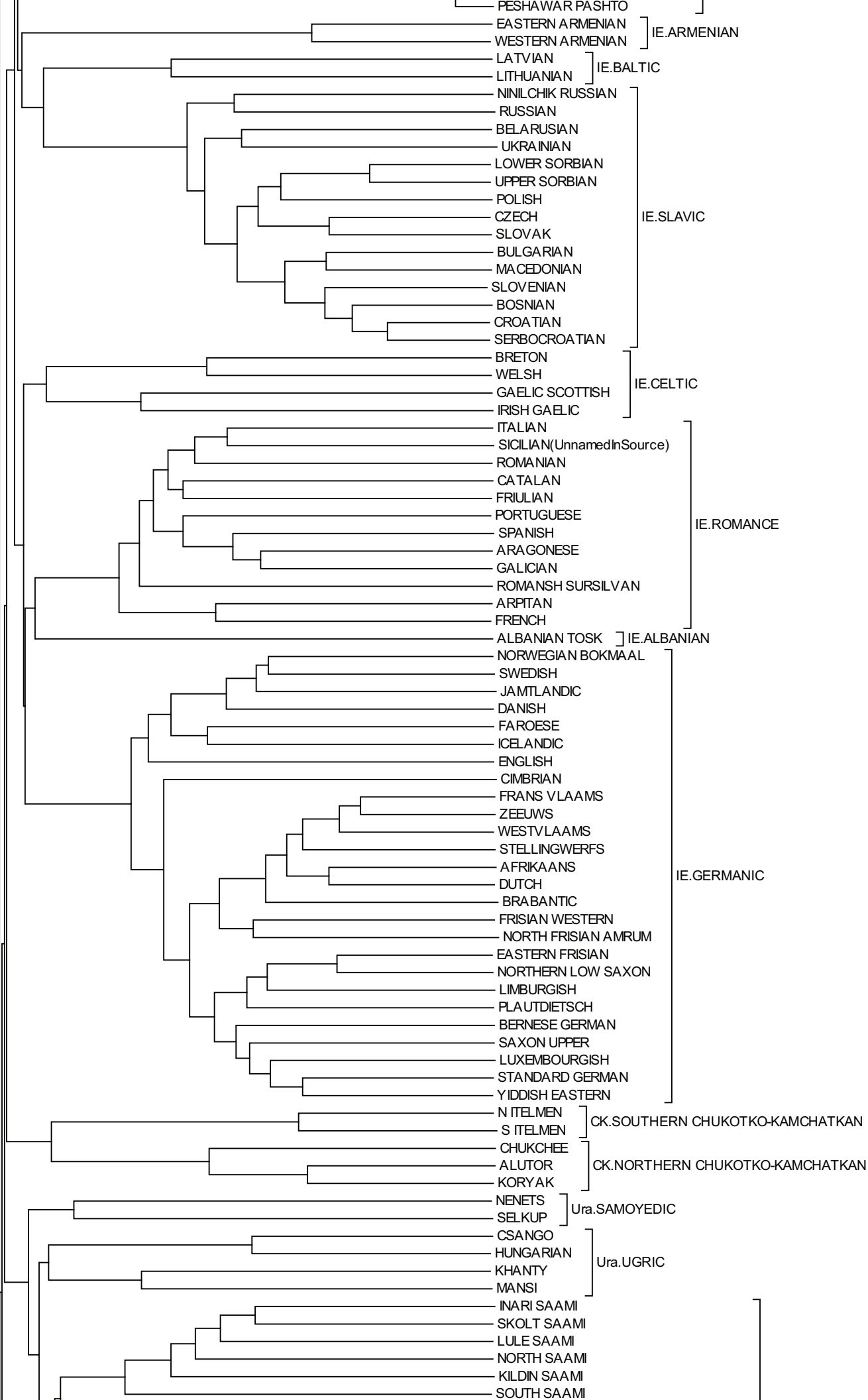


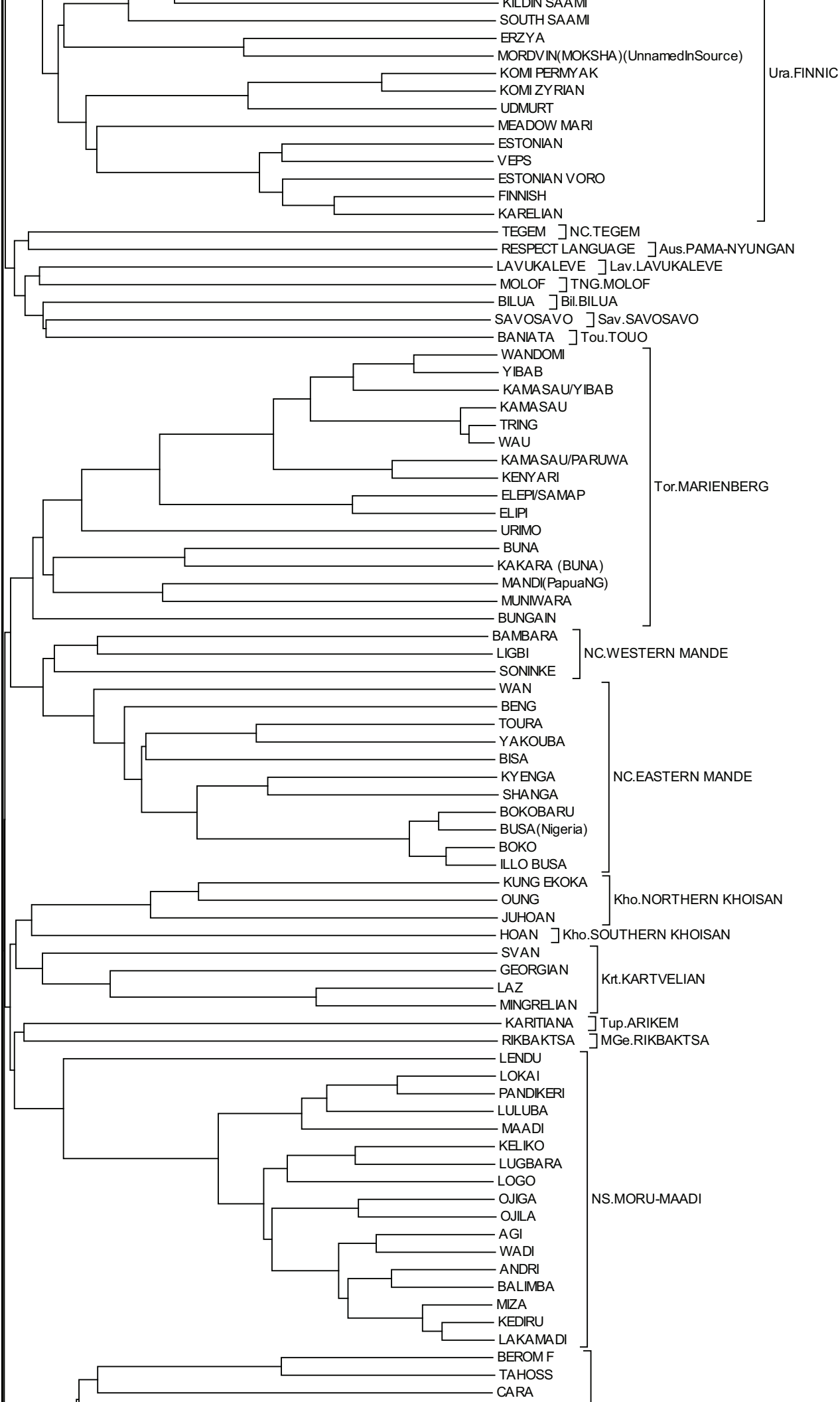


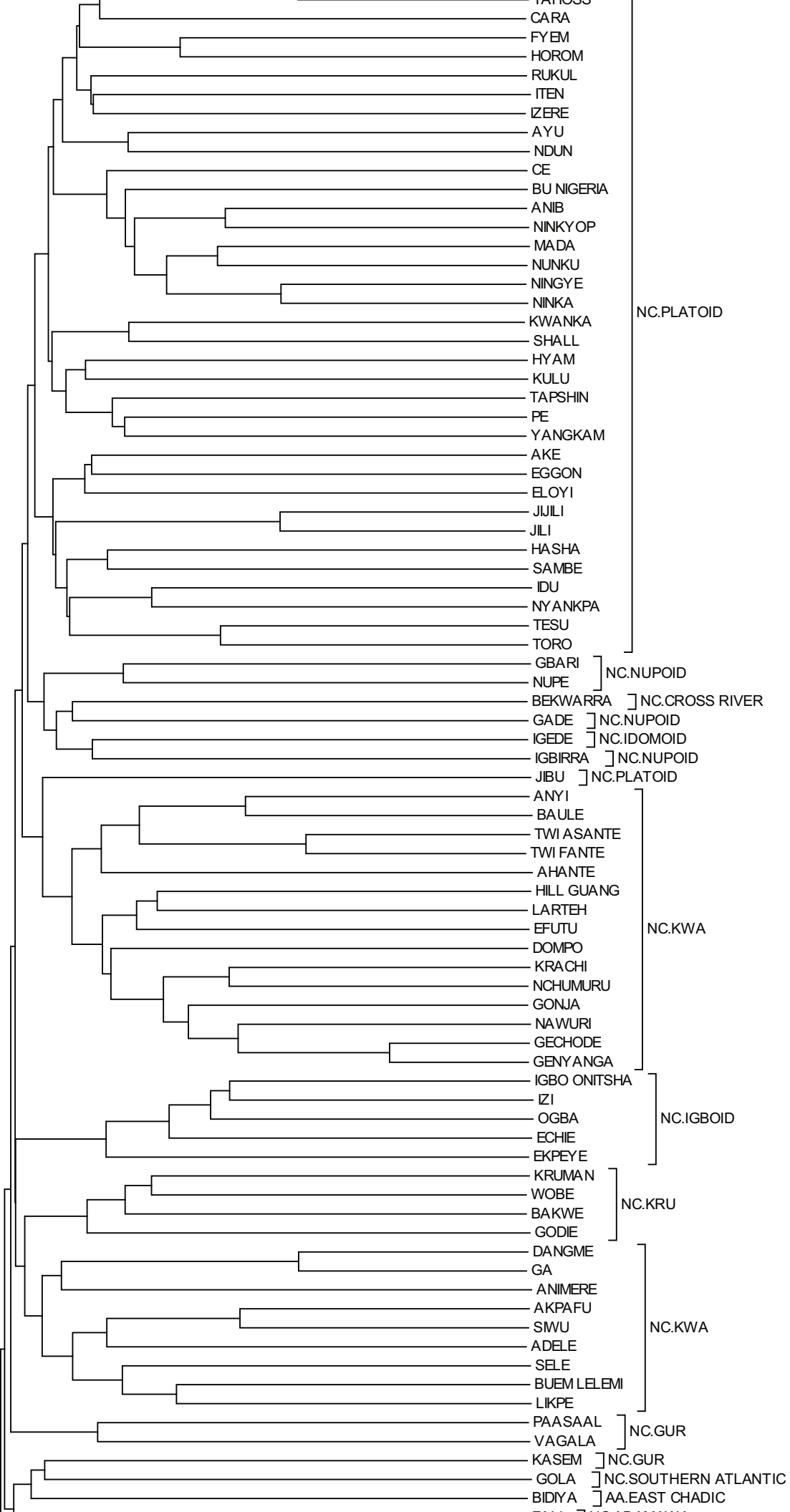


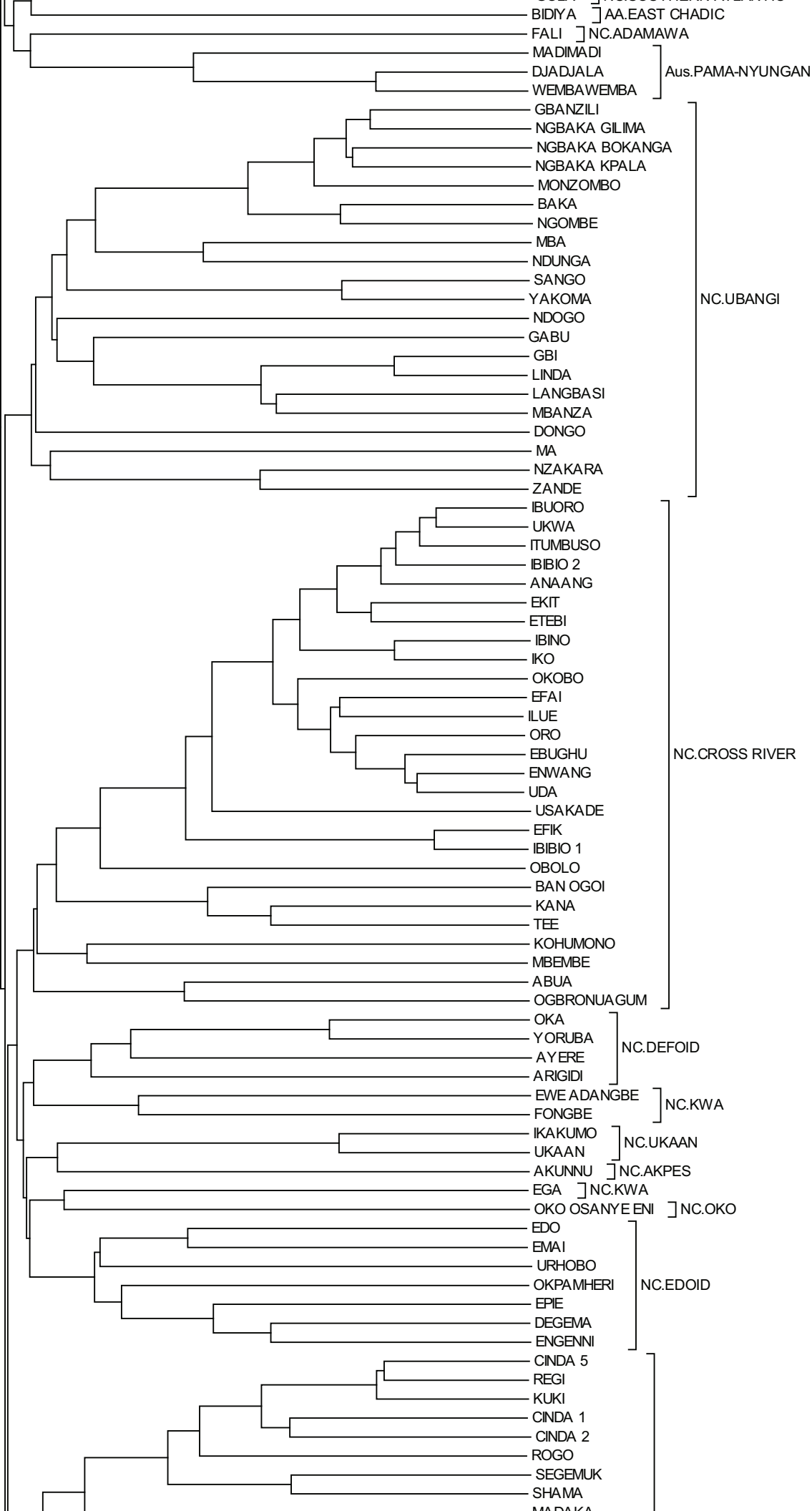


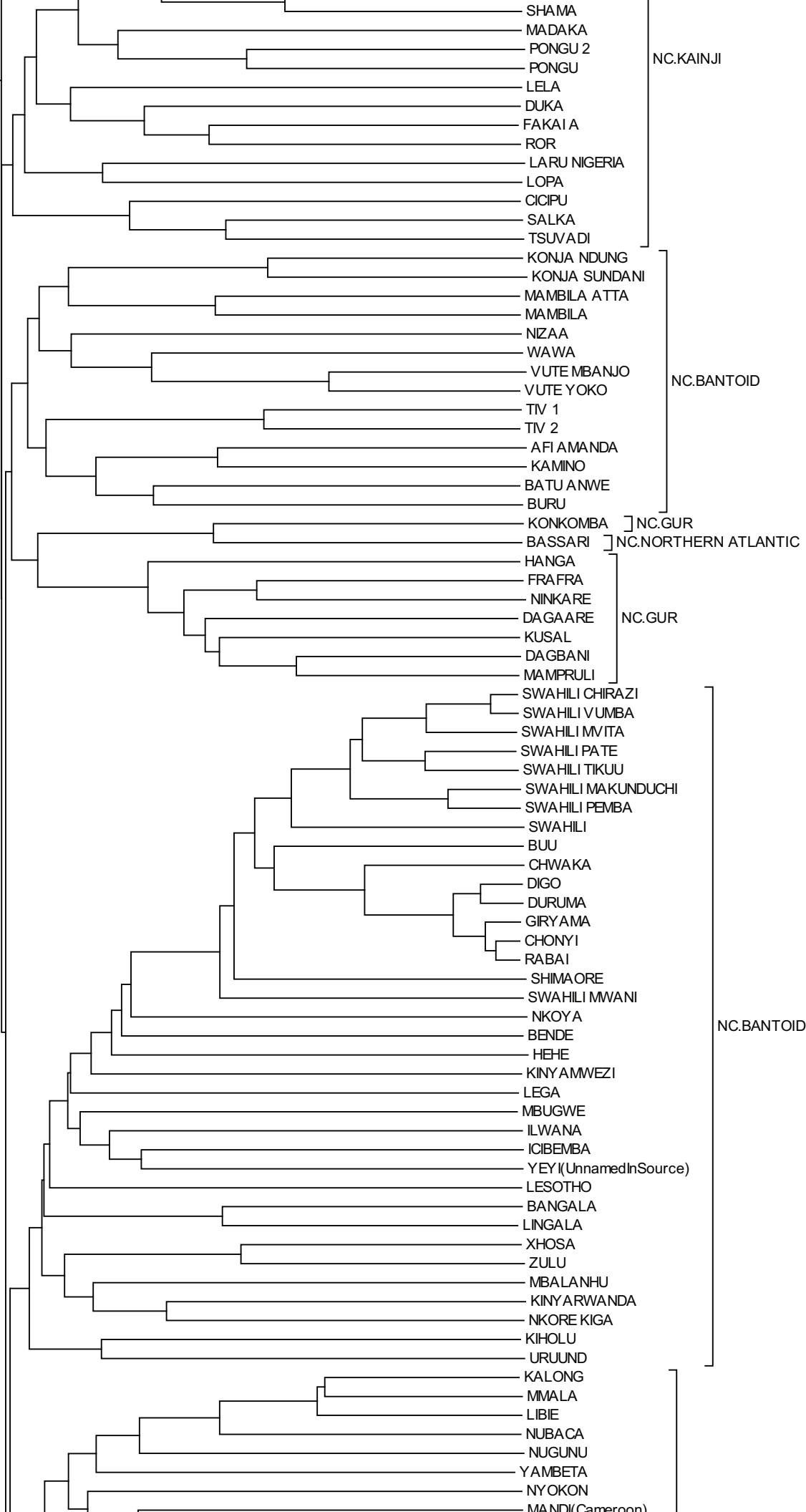


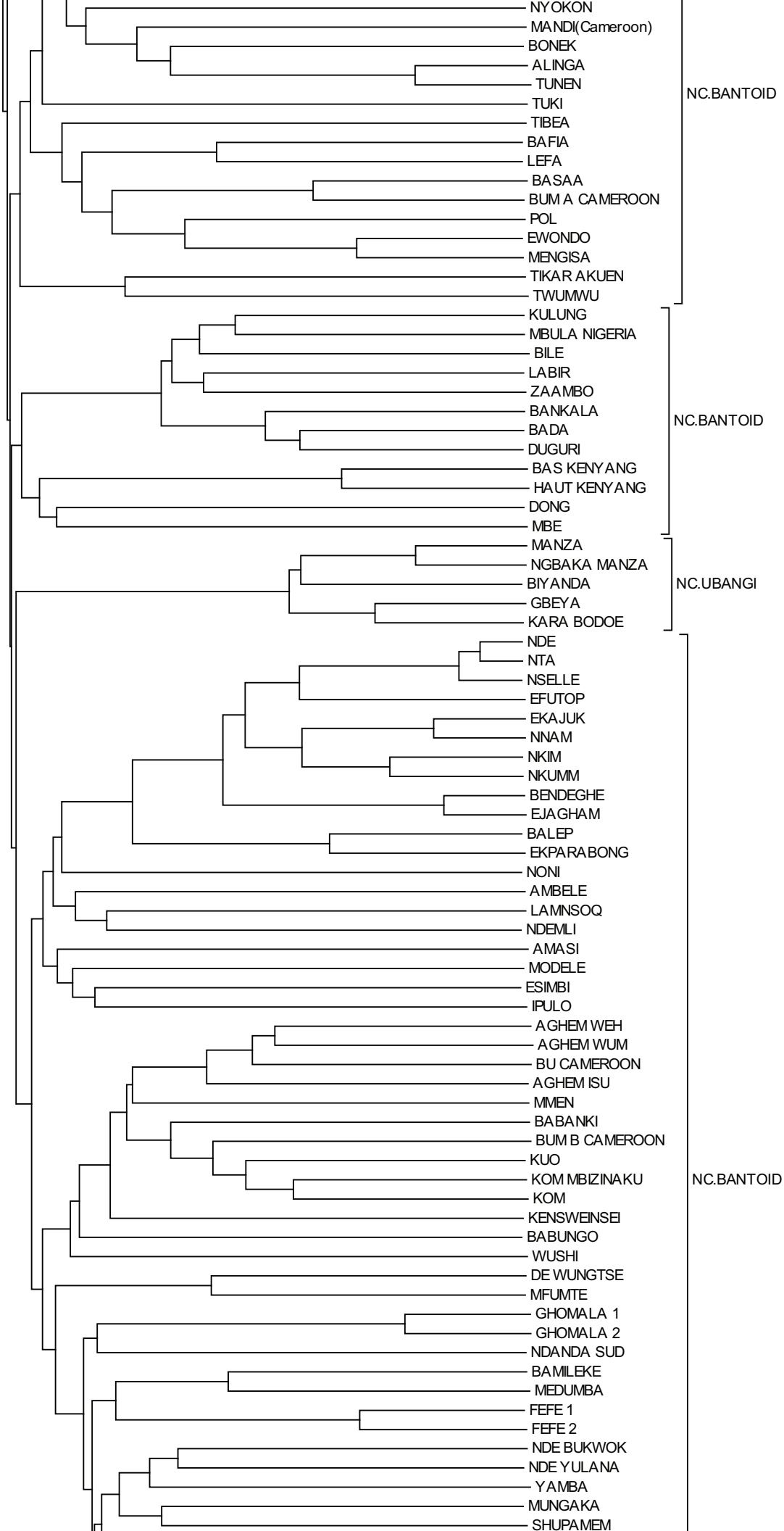












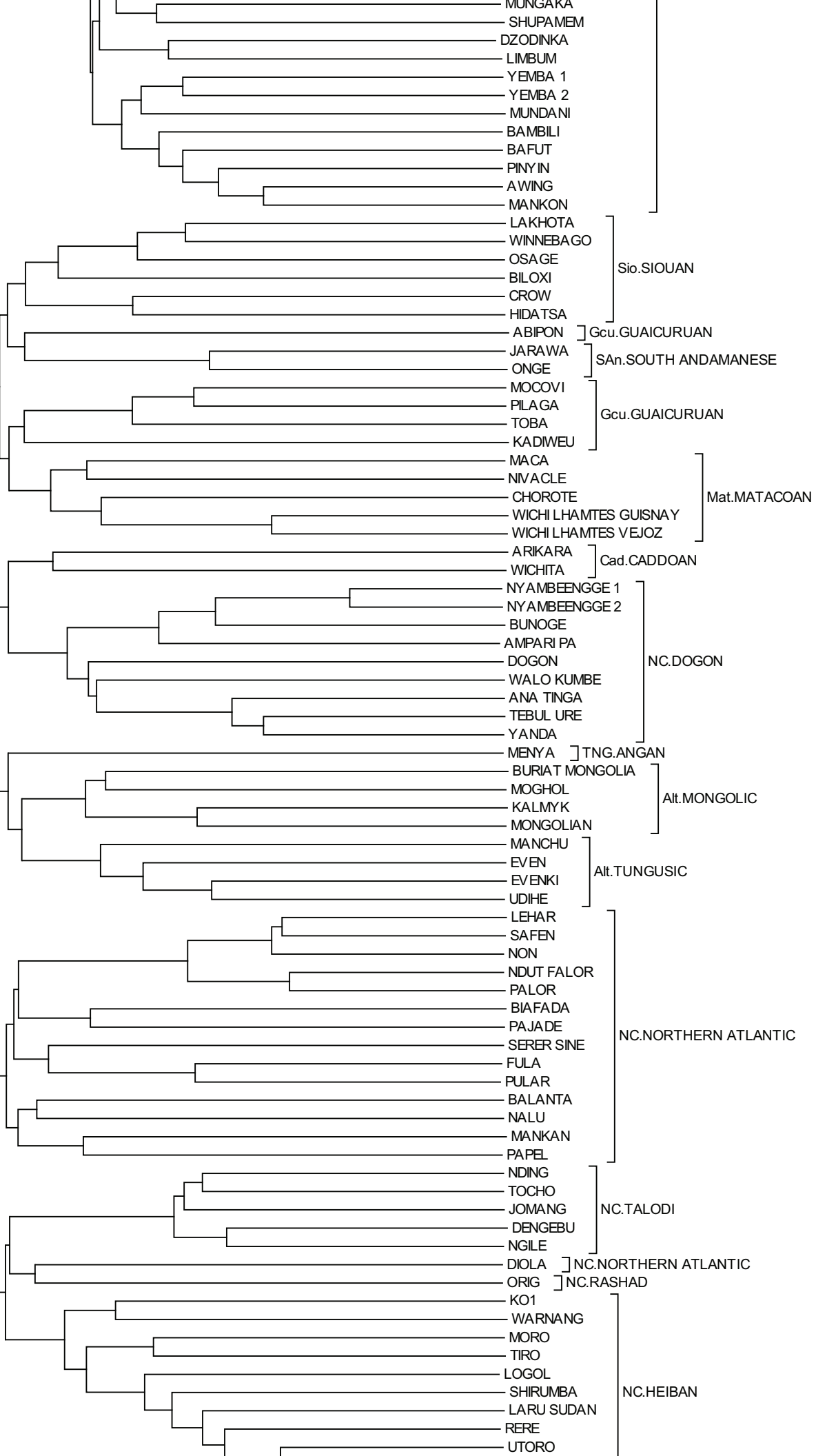
NC.BANTOID

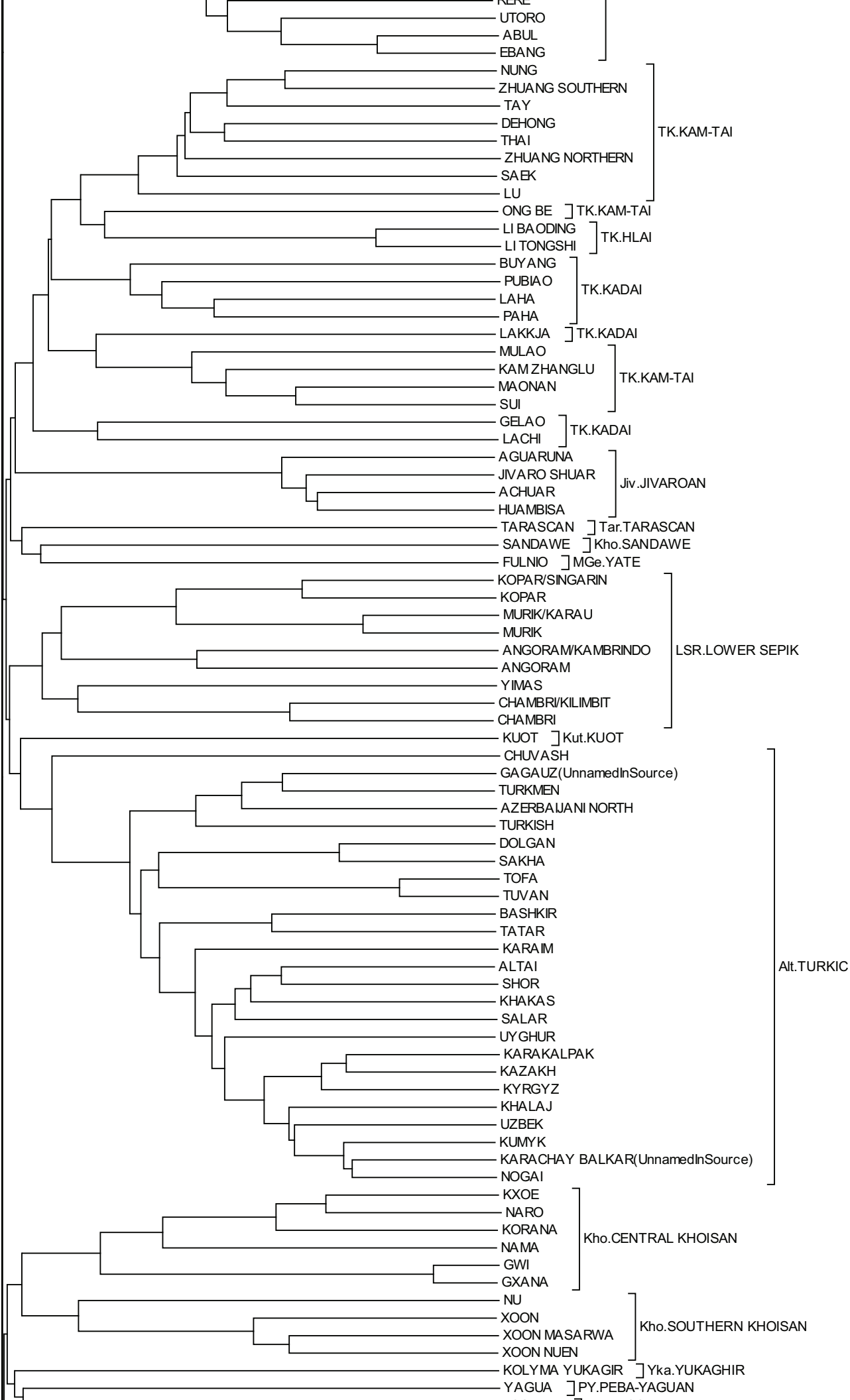
NC.BANTOID

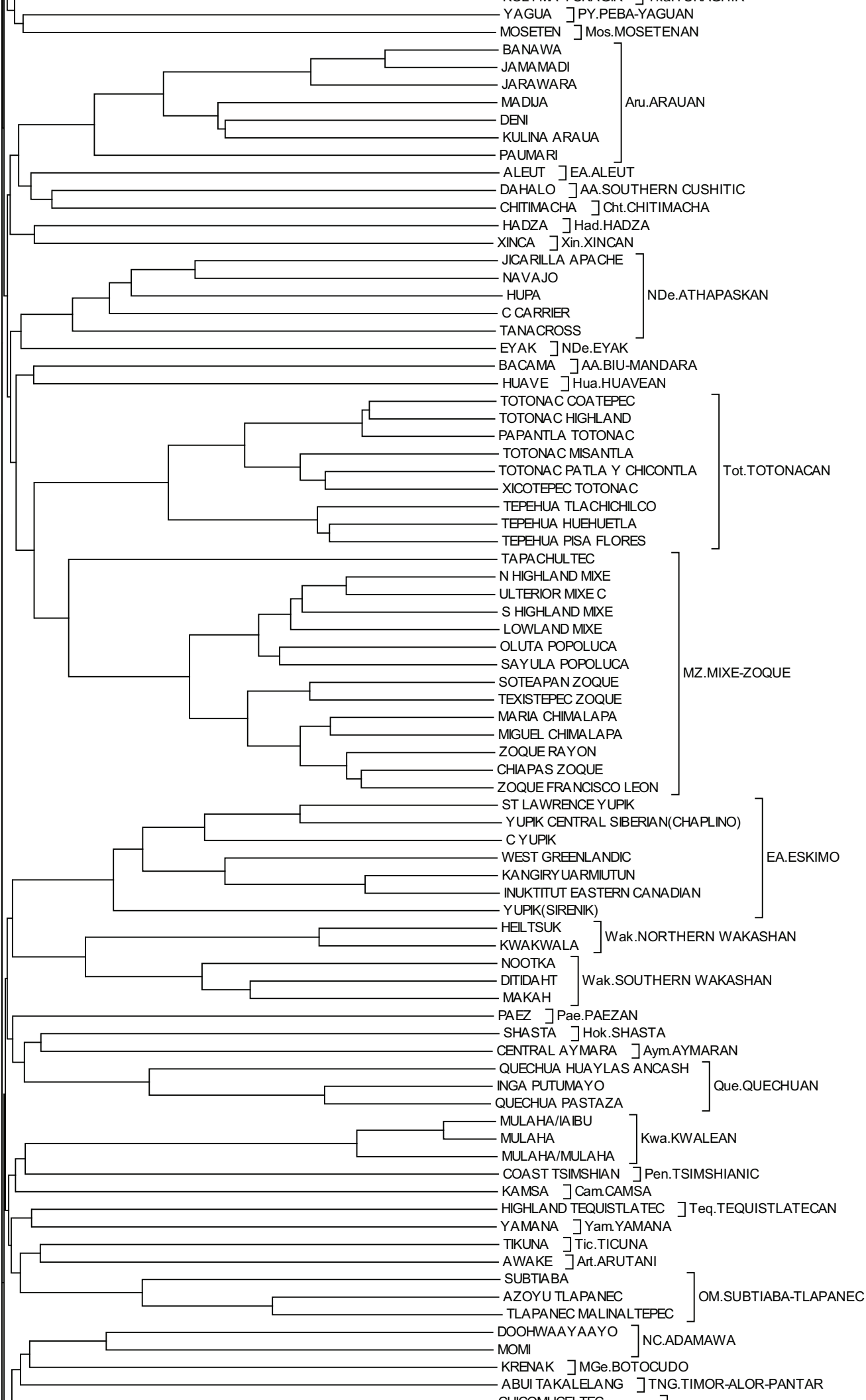
NC.UBANGI

NC.BANTOID









TNG.TIMOR-ALOR-PANTAR

ABUI TAKALELANG

CHICOMUCELTEC

HUASTE

RABINAL ACHI

SAKAPULTEKO

SIPAKAPENSE

KAQCHIKEL NORTHERN

USPANTEKO

QUICHE

TZUTUJIL

POCOMAM

POGOMCHI WESTERN

AGUACATEC

IXIL CHAJUL

TEKTIITEKO

MAM

KEKCHI

CHOL TILA

CHOL

CHONTAL TABASCO

CHORTI

MOPAN

LACANDON

ITZAJ

MAYA YUCATAN

AKATEKO

QANJOBAL EASTERN

MOCHO

JACALTEC

CHUJ

TOJOLABAL

TZELTAL OXCHUC

TZELTAL

TZOTZIL SAN ANDRES

ZINACANTAN TZOTZIL

May.MAYAN

HILDI

WAMDIU

MARGI

HUBA

KILBA

PUTAI

CIBAK

BURA

NGGWAHYI

FALI GILI

FALI KIRIYA

BAZZA

FUTU

GHYE

KAMALE

NKAFA

DABA

BACAMA MULYEN

NZANYI

GUDU

ZIZILIVAKAN

GUDE

FALI BAGIRA

FALI MUCHELLA

MAFA

TERA PIDLIMDI

TERA

HWANA

BOGA

GAANDA GABIN

GAANDA

WANDALA

DGHWEDE

GLAVDA

CINENI

GAVA

MASANA

MUSEY

MASANA POGO

PEVE PALA

MESME

PEVE LAME

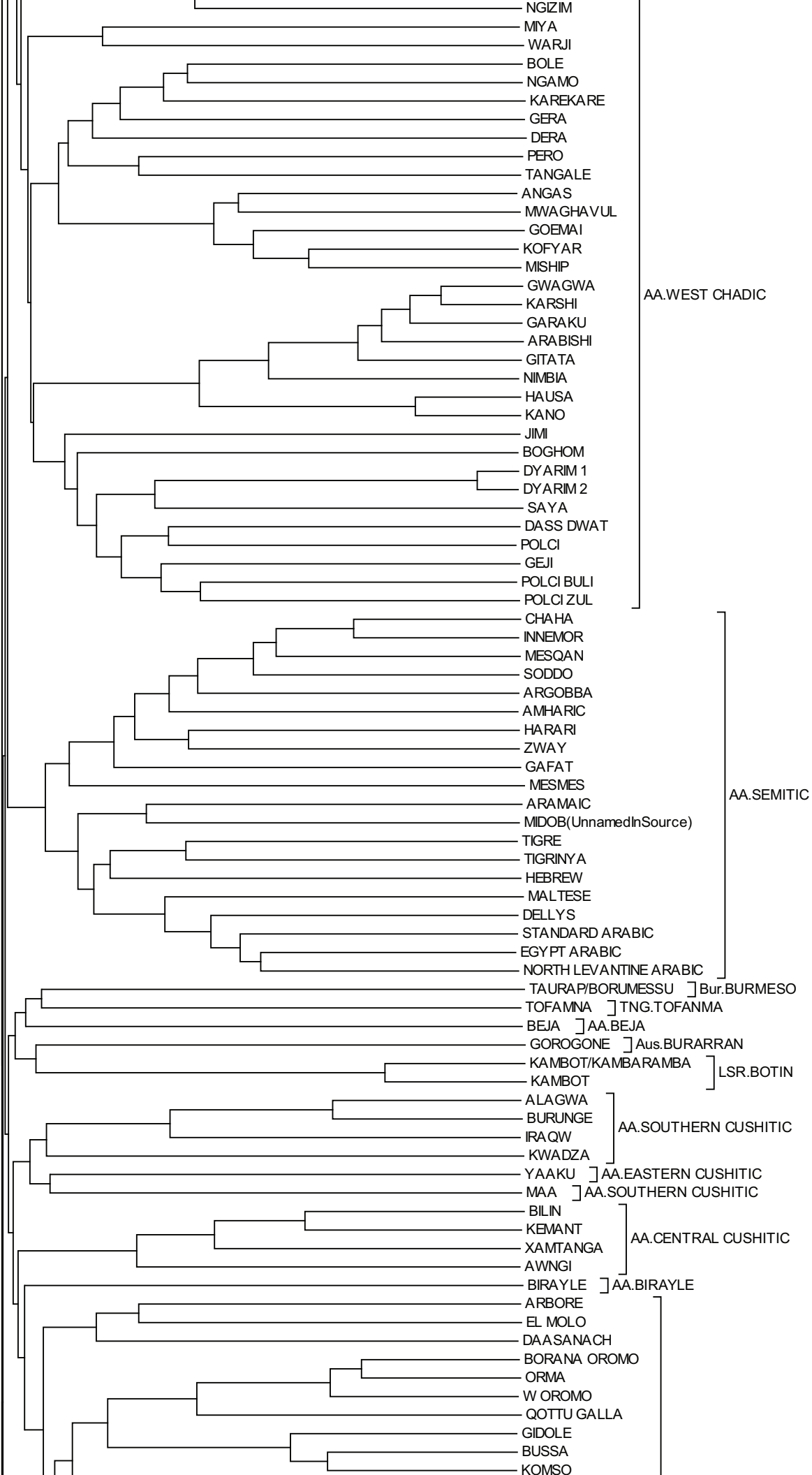
BADE

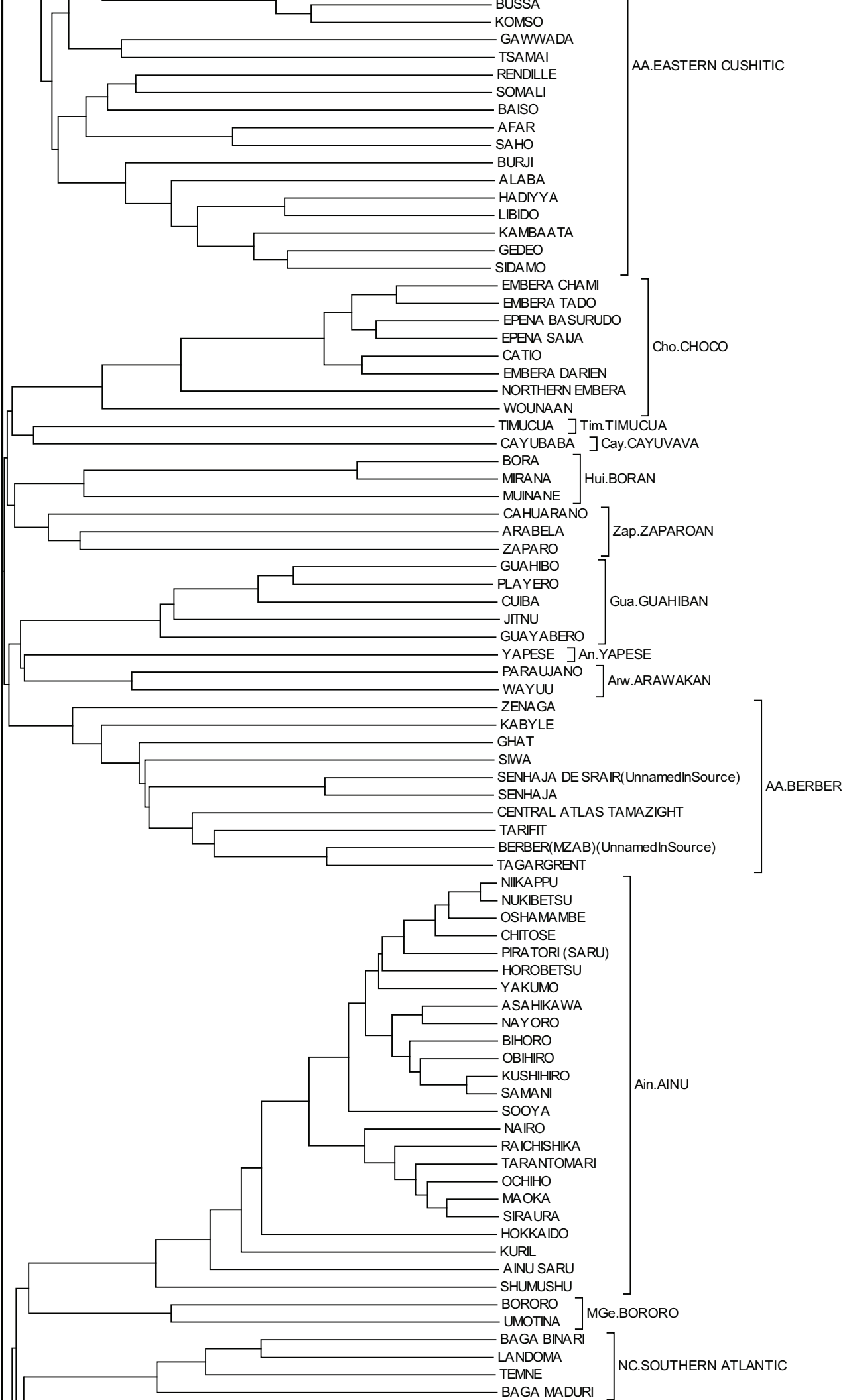
NGIZIM

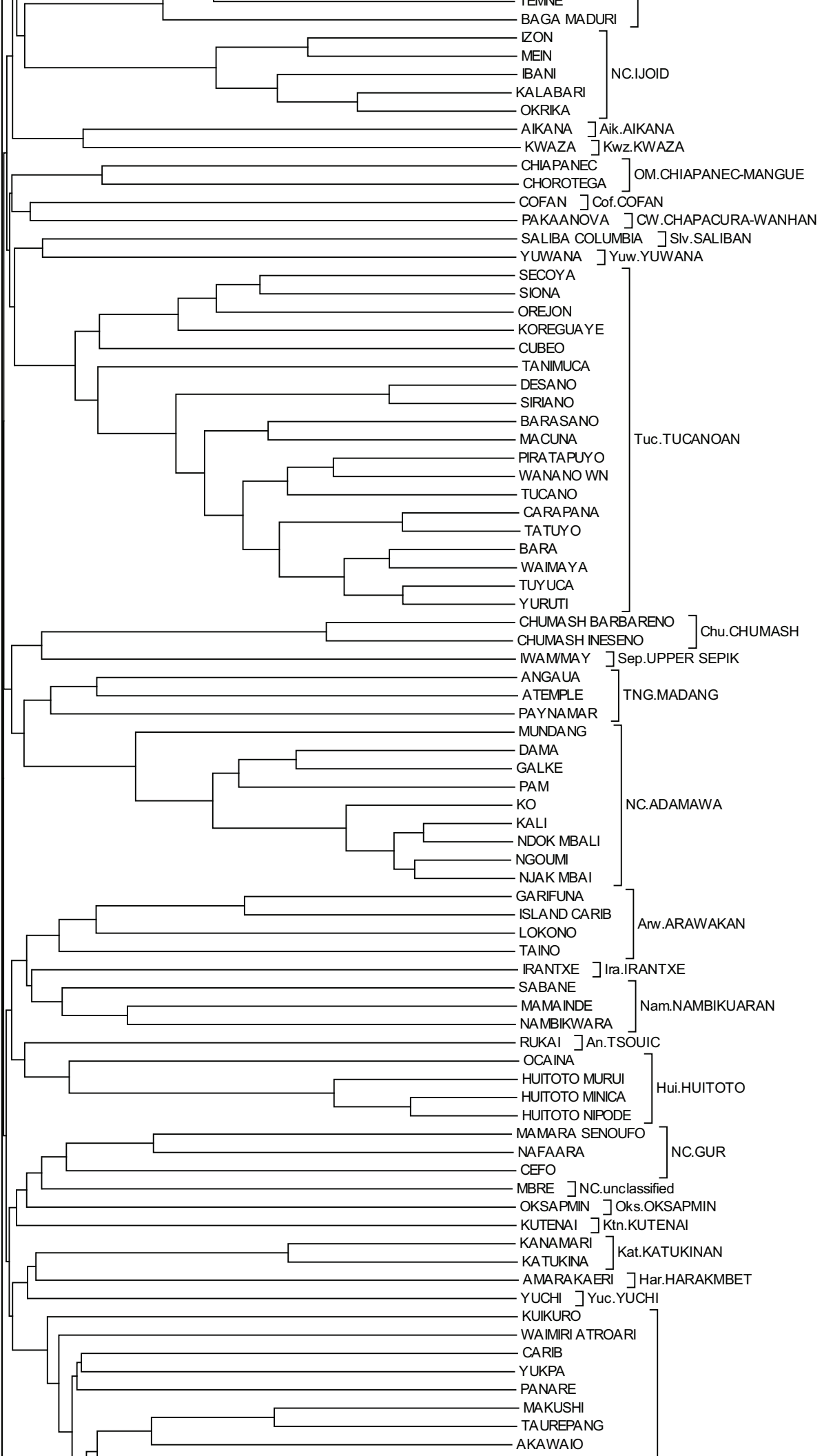
MIXA

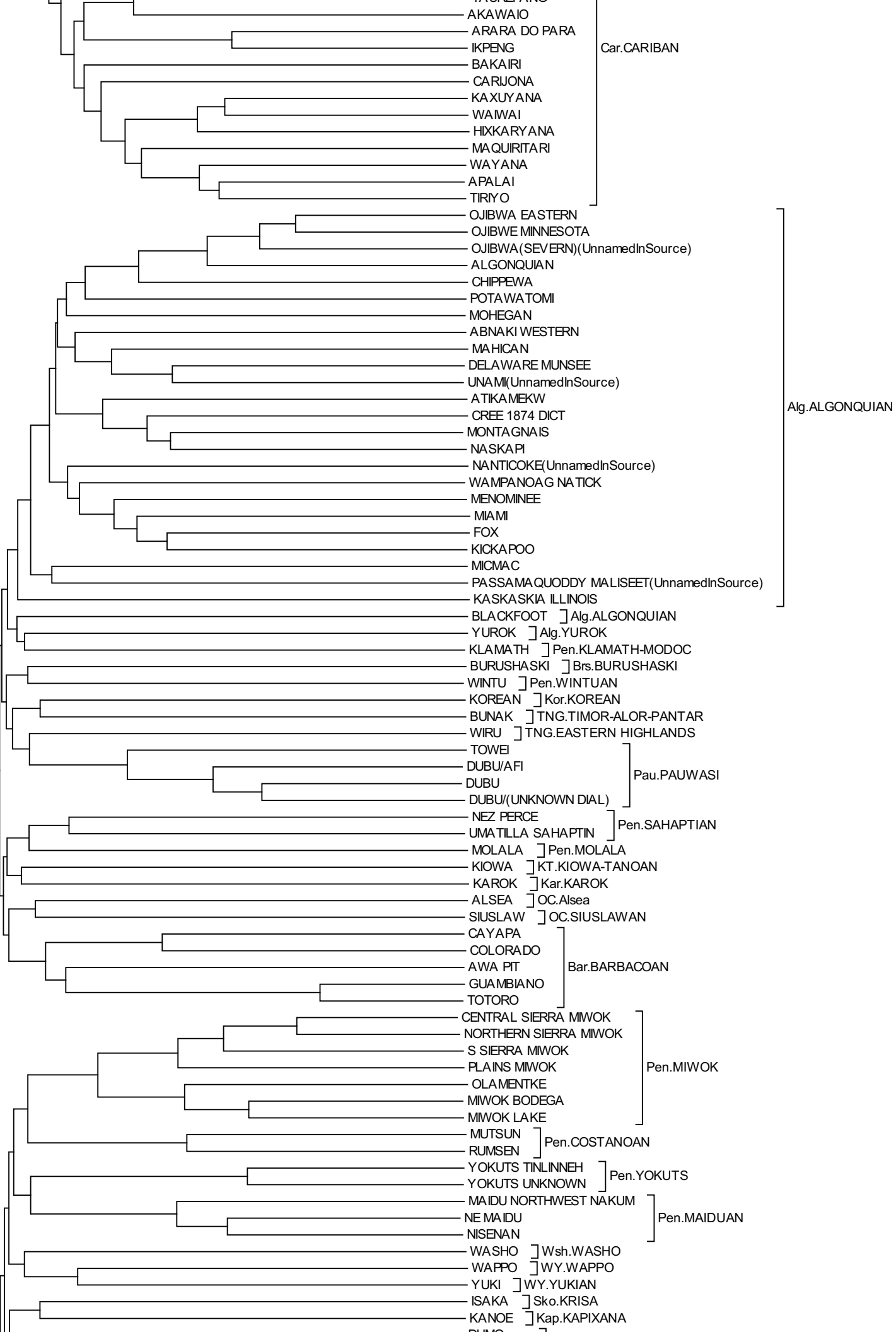
AA.BIU-MANDARA

AA.MASA

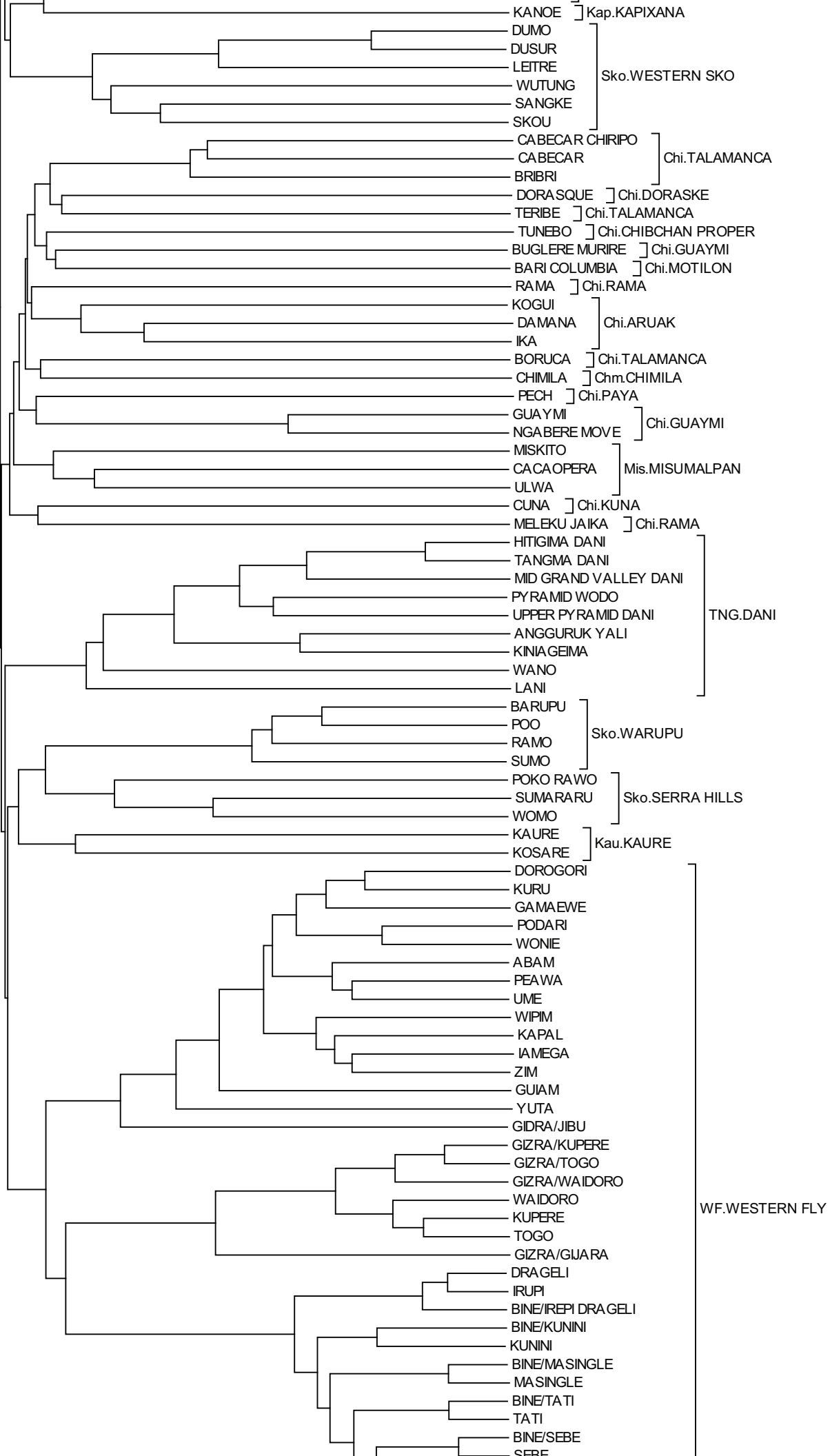


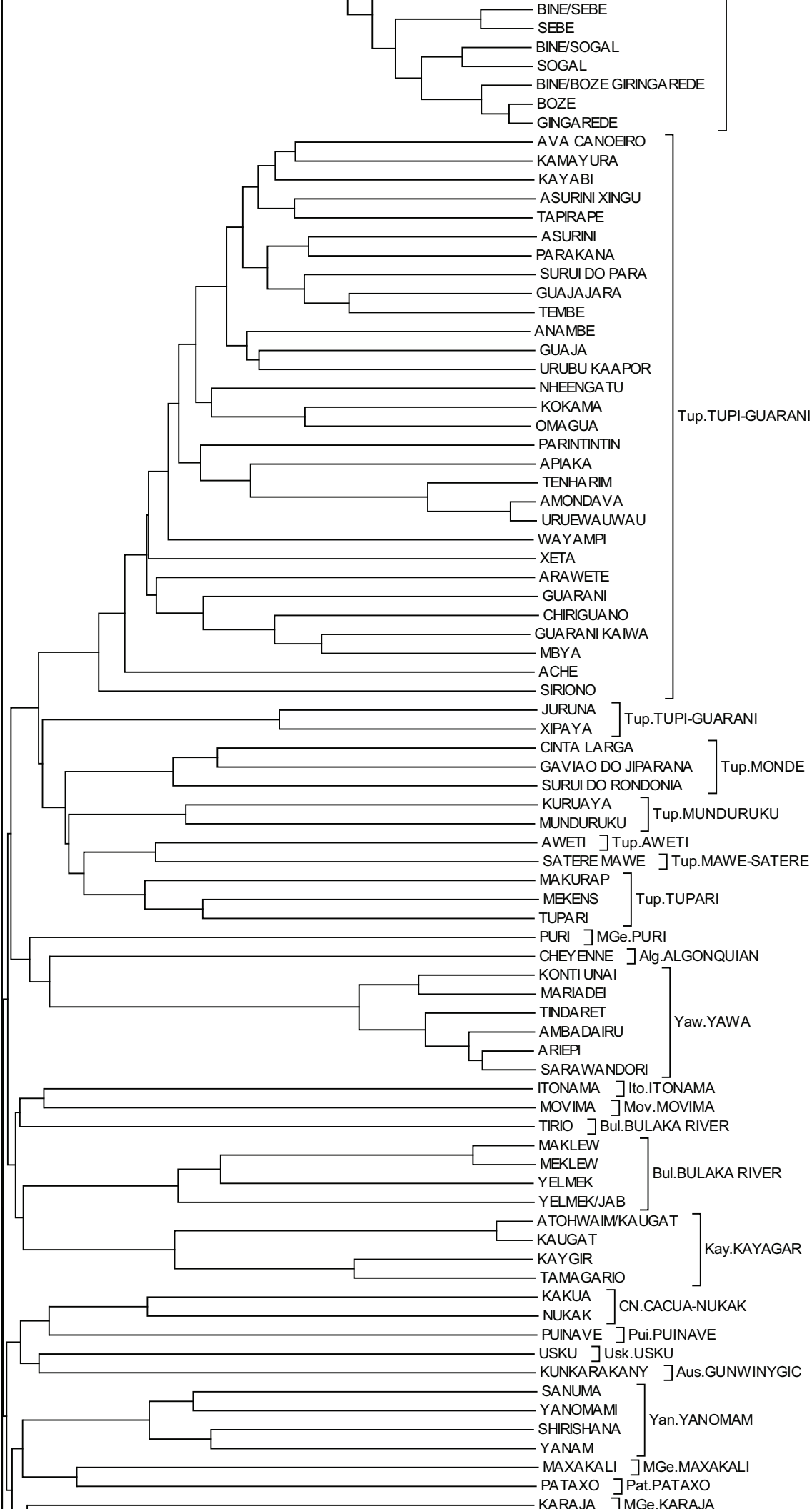


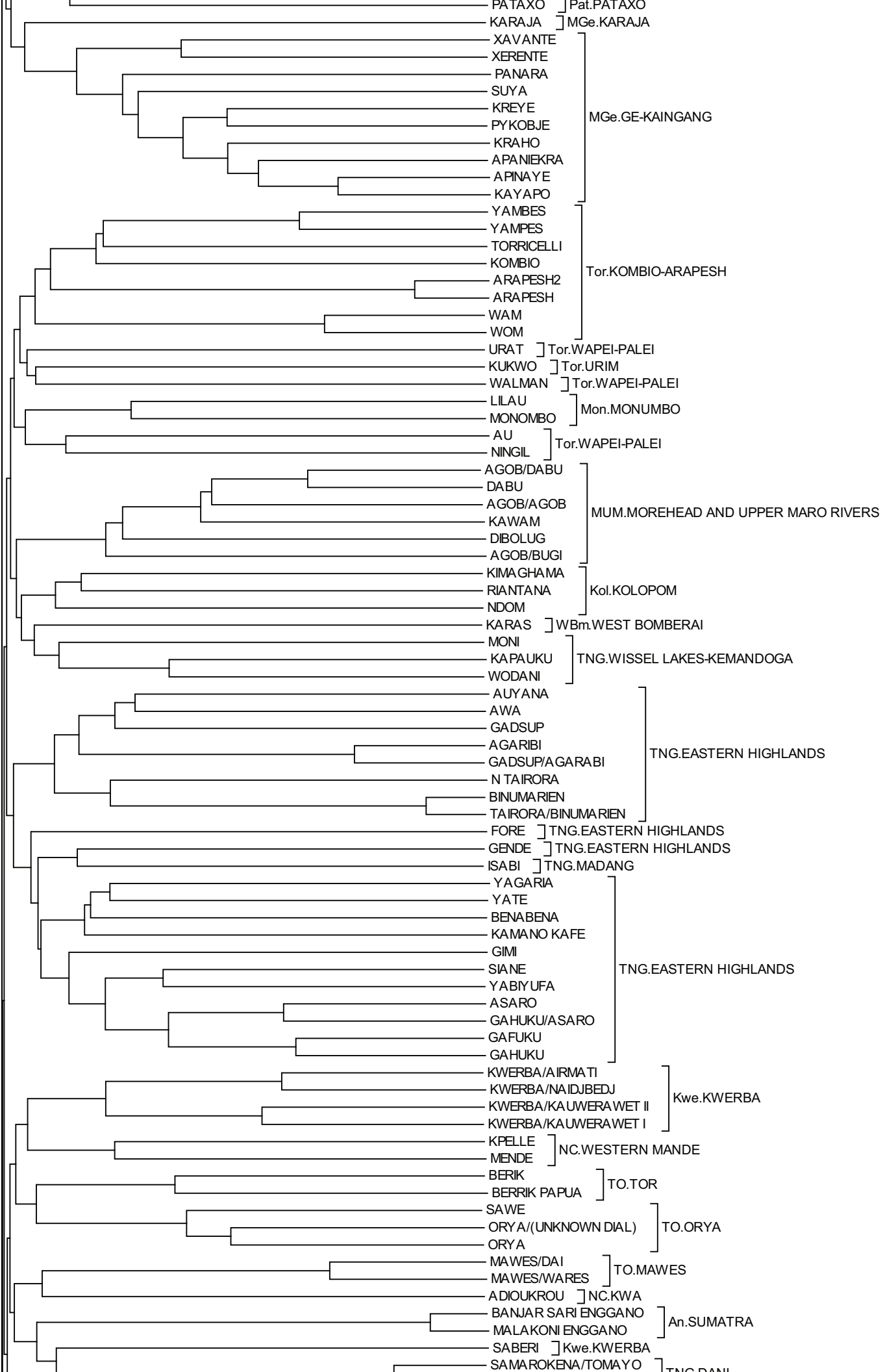


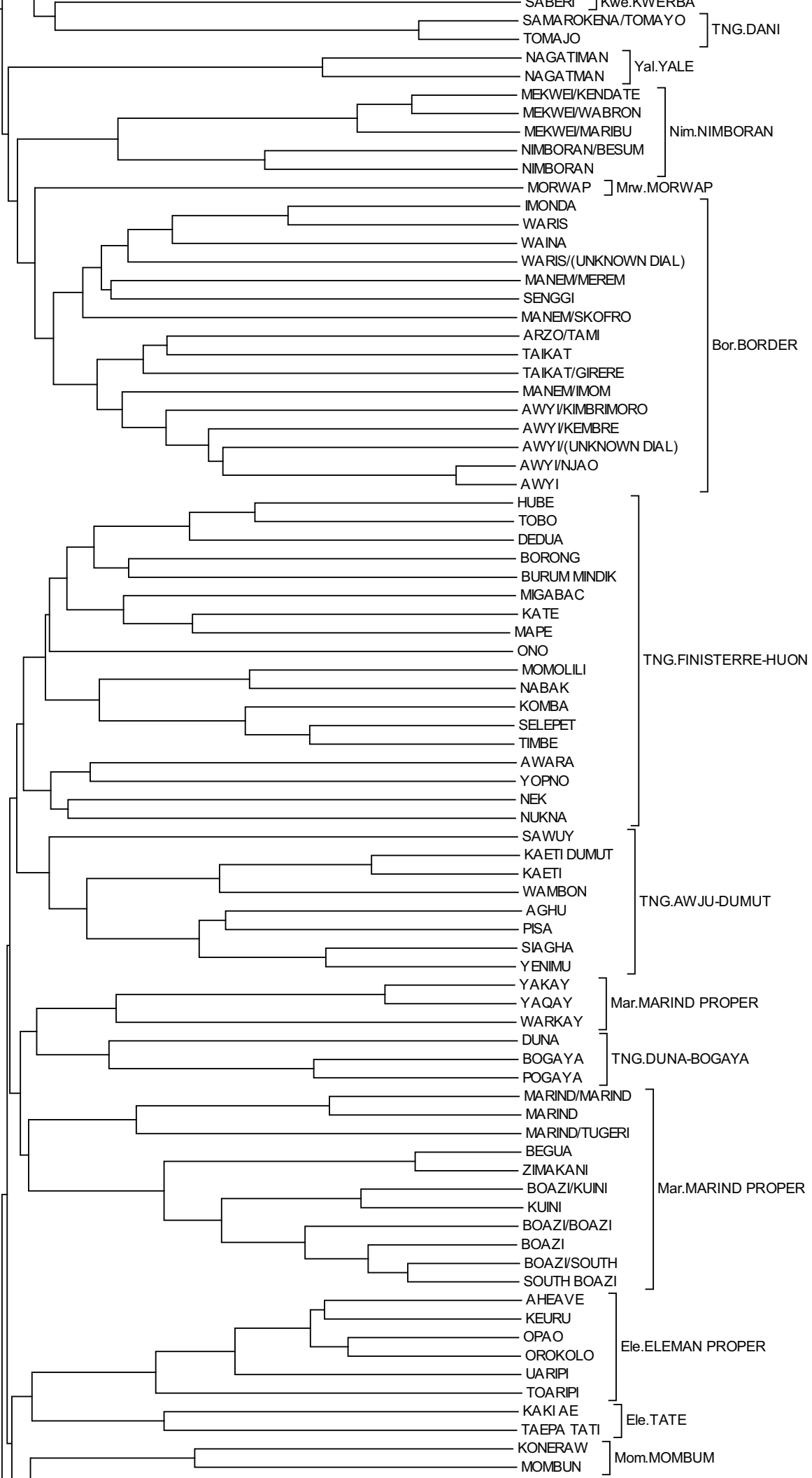












SABERI ] Kwe.KWERBA

SAMAROKENA/TOMAYO ] TNG.DANI

TOMAJO

NAGATIMAN ] Yal.YALE

NAGATMAN

MEKWEI/KENDATE

MEKWEI/WABRON

MEKWEI/MARIBU ] Nim.NIMBORAN

NIMBORAN/BESUM

NIMBORAN

MORWAP ] Mrw.MORWAP

IMONDA

WARIS

WAINA

WARIS/(UNKNOWN DIAL)

MANEW/MEREM

SENGGI

MANEW/SKOFRO

ARZO/TAMI

TAIKAT

TAIKAT/GIRERE

MANEW/IMOM

AWYI/KIMBRIMORO

AWYI/KEMBRE

AWYI/(UNKNOWN DIAL)

AWYI/NJAO

AWYI

HUBE

TOBO

DEДУА

BORONG

BURUM MINDIK

MIGABAC

KATE

MAPE

ONO

MOMOLILI

NABAK

KOMBA

SELEPET

TIMBE

AWARA

YOPNO

NEK

NUKNA

SAWUY

KAETI DUMUT

KAETI

WAMBON

AGHU

PISA

SIAGHA

YENIMU

YAKAY

YAQAY

WARKAY

DUNA

BOGAYA

POGAYA

MARIND/MARIND

MARIND

MARIND/TUGERI

BEGUA

ZIMAKANI

BOAZI/KUINI

KUINI

BOAZI/BOAZI

BOAZI

BOAZI/SOUTH

SOUTH BOAZI

AHEAVE

KEURU

OPAO

OROKOLO

UARIPI

TOARIPI

KAKIAE

TAEPATATI

KONERAW

MOMBUN

] TNG.DANI

] Yal.YALE

] Nim.NIMBORAN

] Mrw.MORWAP

] Bor.BORDER

] TNG.FINISTERRE-HUON

] TNG.AWJU-DUMUT

] Mar.MARIND PROPER

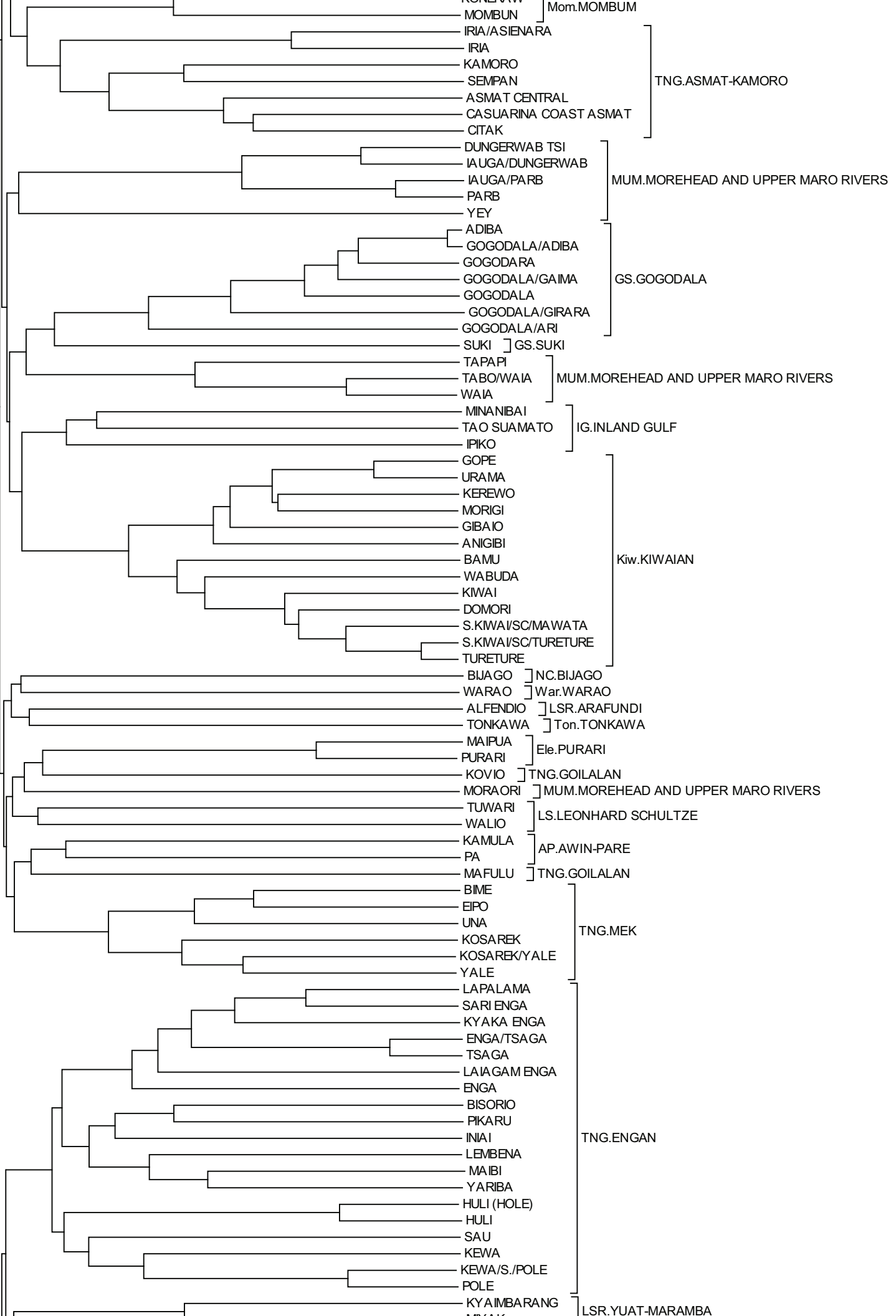
] TNG.DUNA-BOGAYA

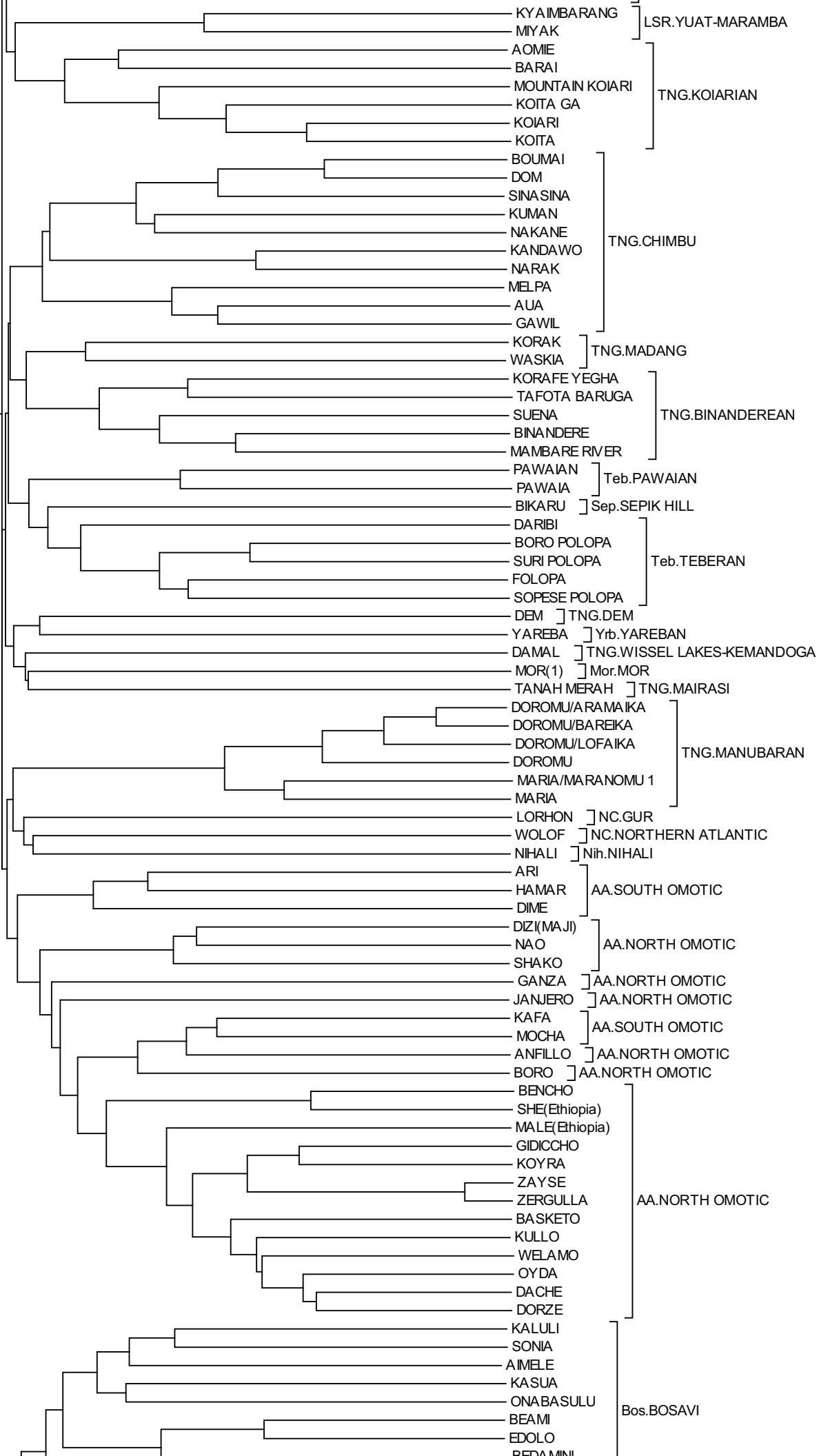
] Mar.MARIND PROPER

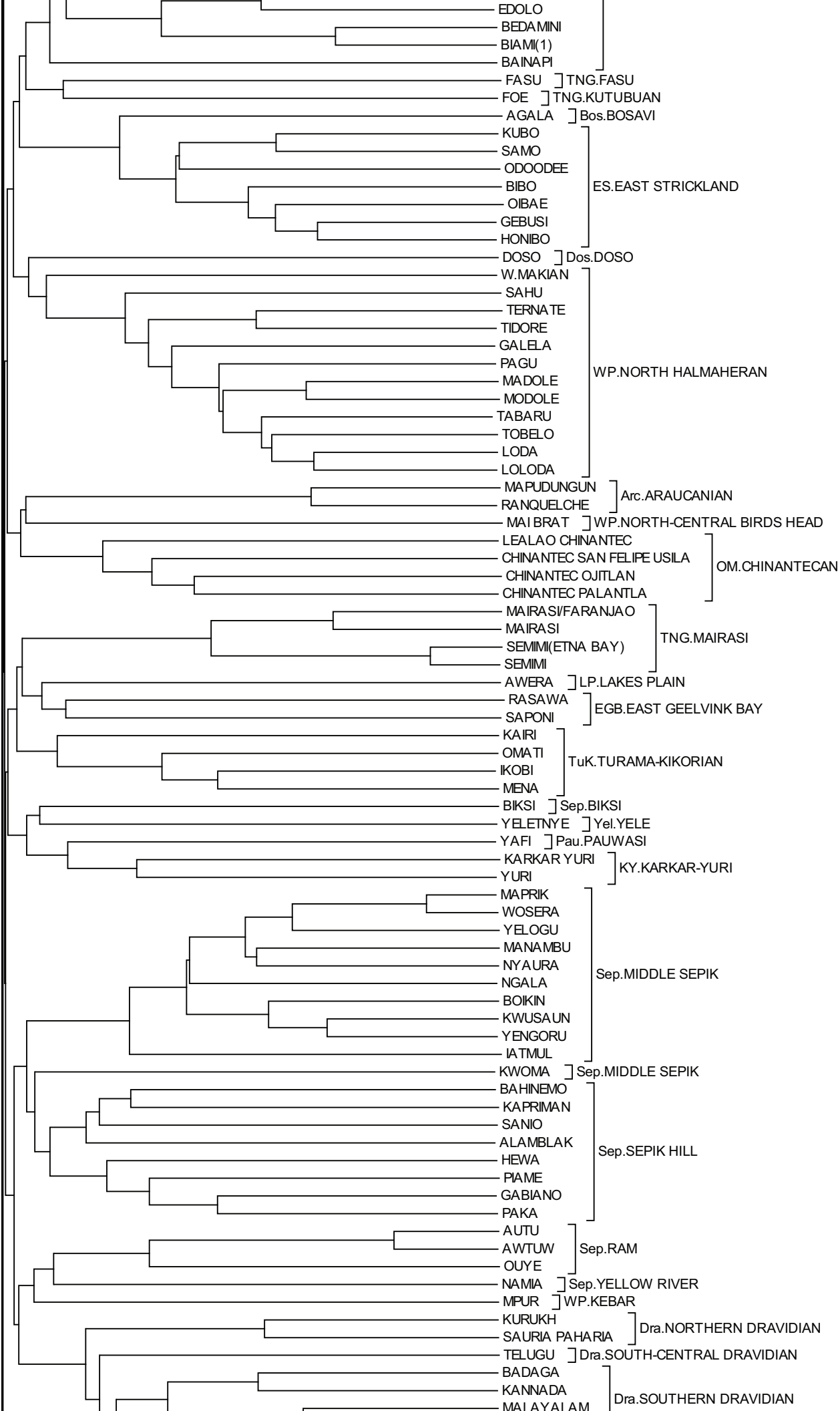
] Ele.ELEMAN PROPER

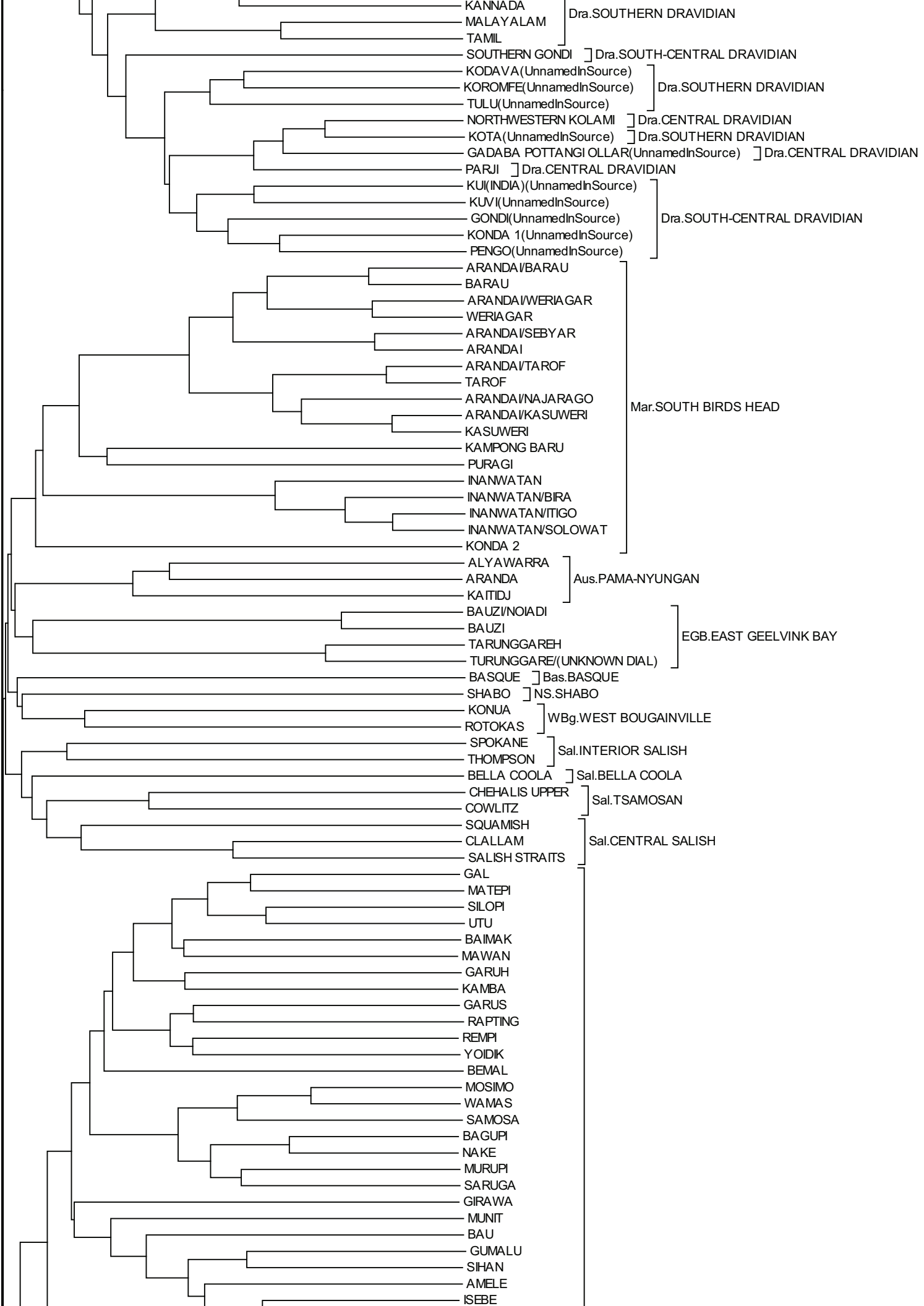
] Ele.TATE

] Mom.MOMBUM

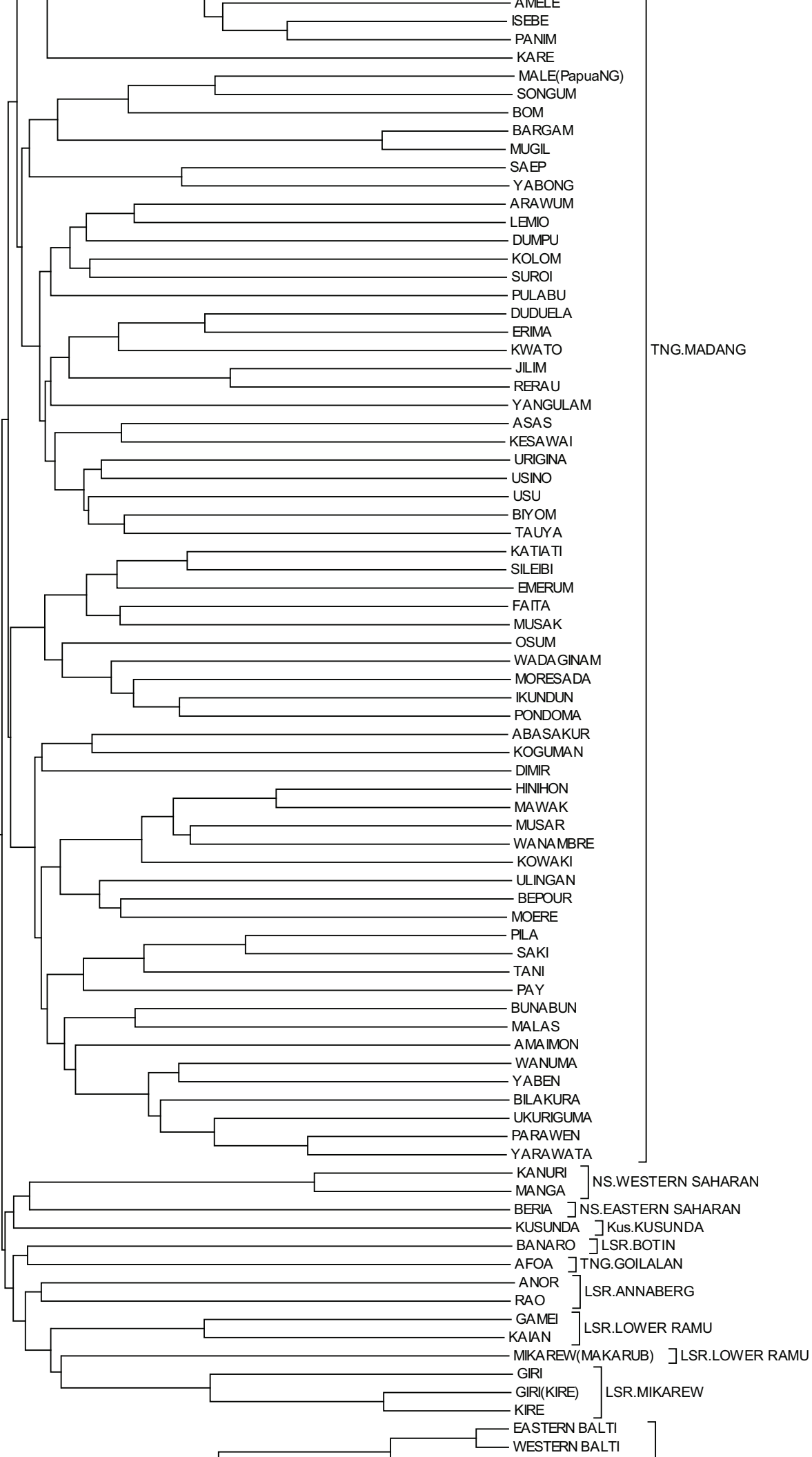


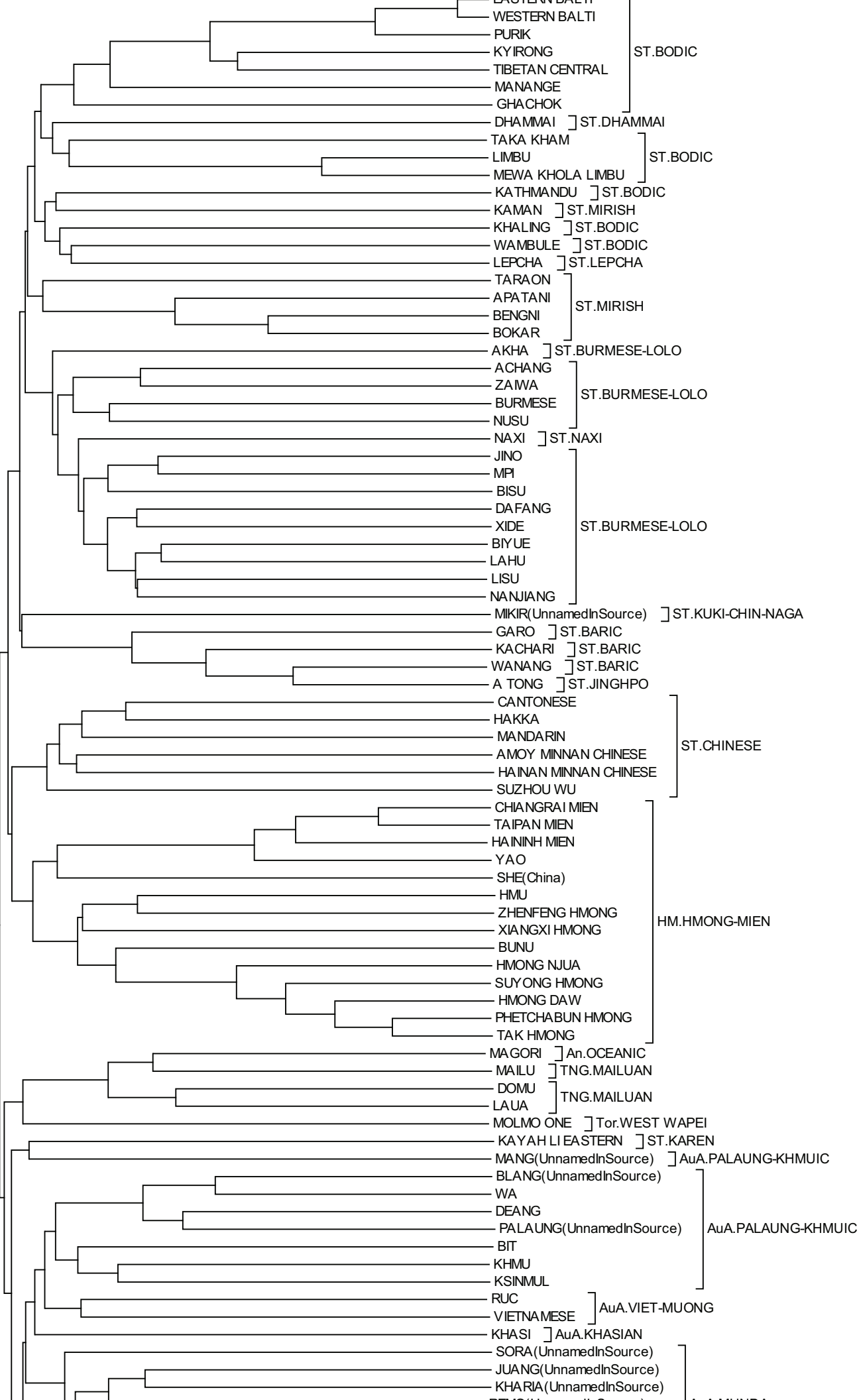


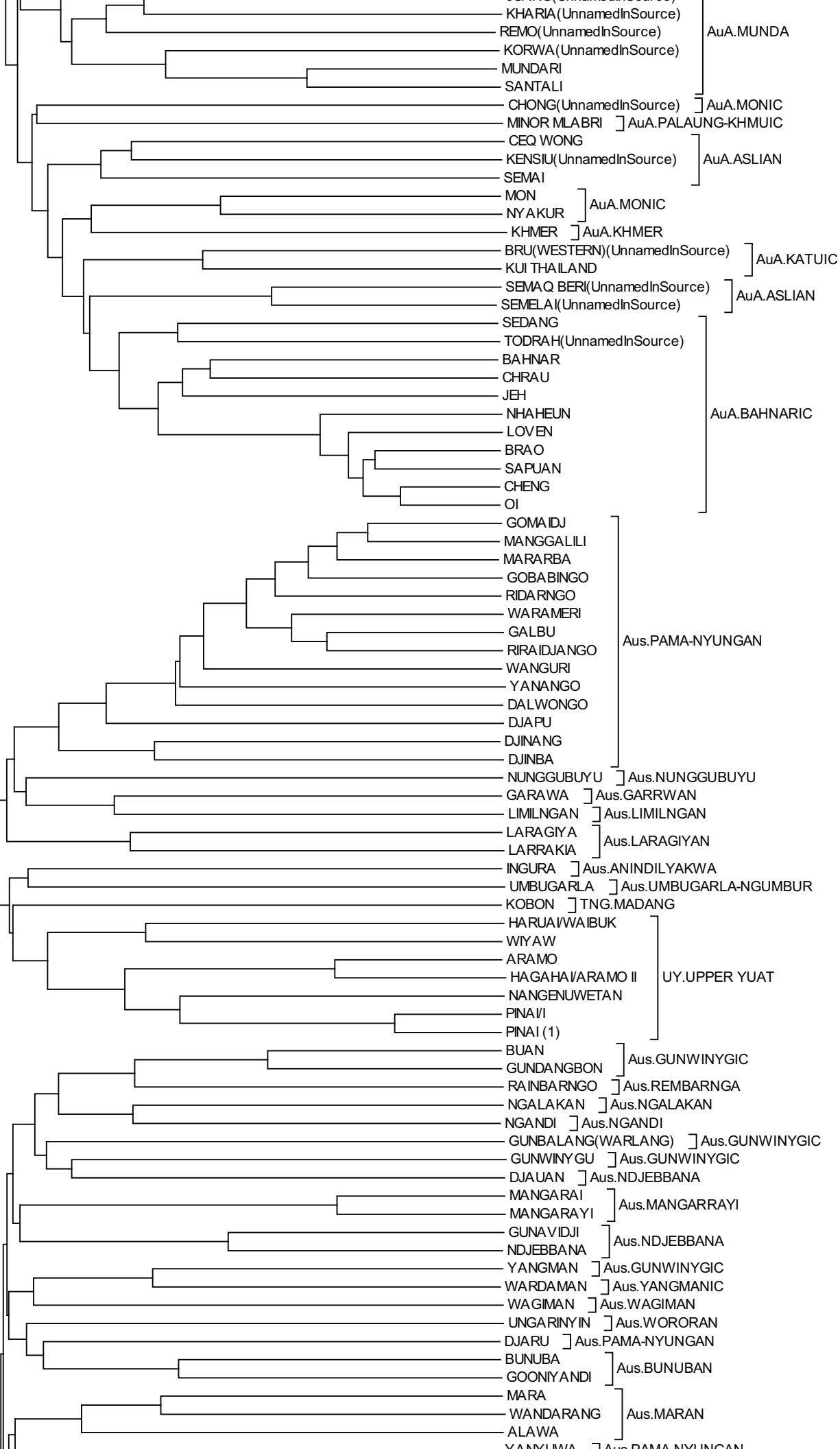


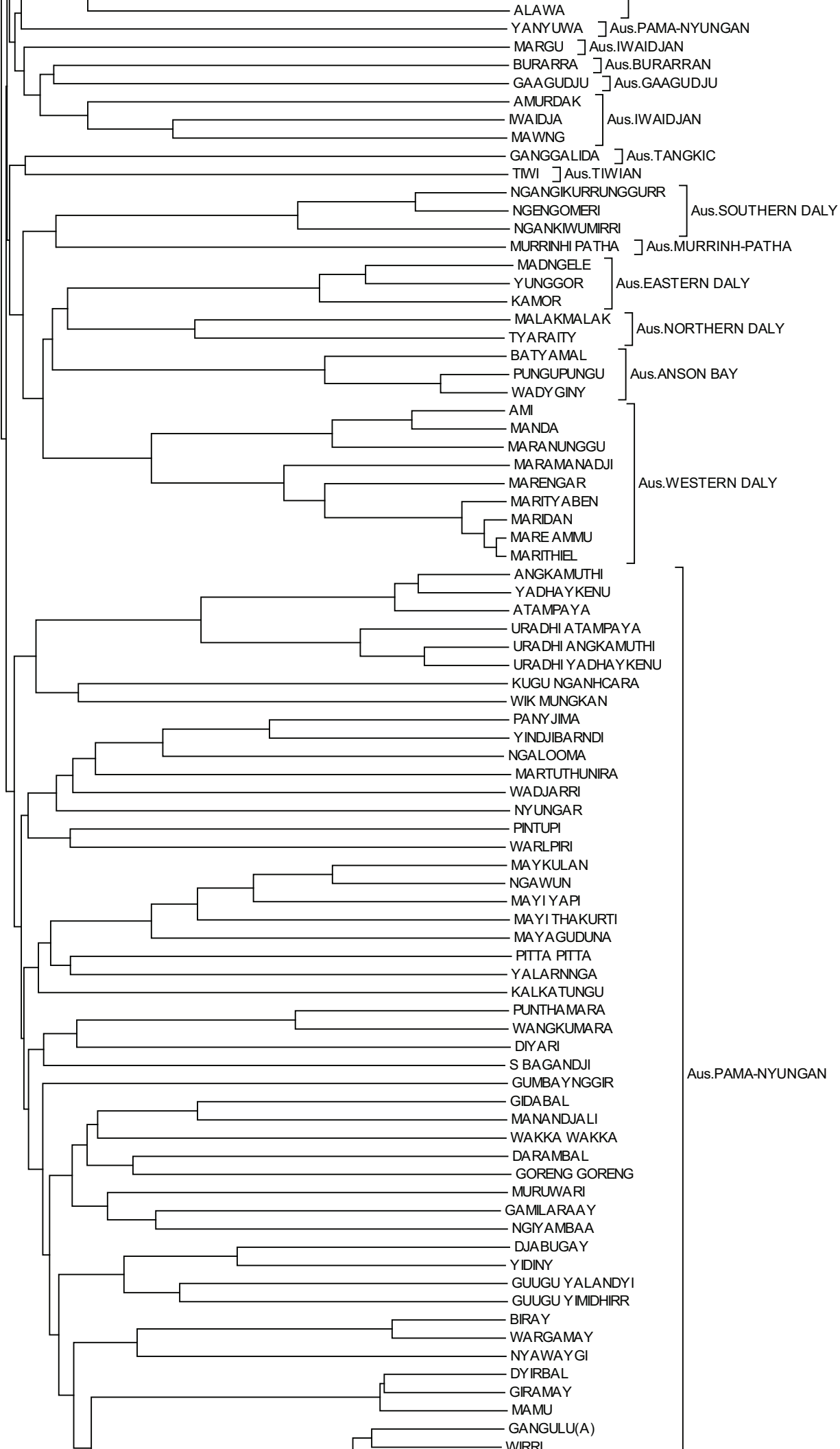












ALAWA

YANYUWA ] Aus.PAMA-NYUNGAN

MARGU ] Aus.IWAIDJAN

BURARRA ] Aus.BURARRAN

GAAGUDJU ] Aus.GAAGUDJU

AMURDAK ] Aus.IWAIDJAN

WAIDJA ] Aus.IWAIDJAN

MAWNG ] Aus.IWAIDJAN

GANGGALIDA ] Aus.TANGKIC

TWI ] Aus.TIWIAN

NGANGIKURRUNGURR ] Aus.SOUTHERN DALY

NGENGOMERI ] Aus.SOUTHERN DALY

NGANKWUMIRRI ] Aus.SOUTHERN DALY

MURRINH-PATHA ] Aus.MURRINH-PATHA

MADNGELE ] Aus.EASTERN DALY

YUNGGOR ] Aus.EASTERN DALY

KAMOR ] Aus.EASTERN DALY

MALAKMALAK ] Aus.NORTHERN DALY

TYARAITY ] Aus.NORTHERN DALY

BATYAMAL ] Aus.ANSON BAY

PUNGUPUNGU ] Aus.ANSON BAY

WADYGINY ] Aus.ANSON BAY

AMI ] Aus.WESTERN DALY

MANDA ] Aus.WESTERN DALY

MARANUNGGU ] Aus.WESTERN DALY

MARAMANADJI ] Aus.WESTERN DALY

MARENGAR ] Aus.WESTERN DALY

MARTYABEN ] Aus.WESTERN DALY

MARIDAN ] Aus.WESTERN DALY

MAREAMMU ] Aus.WESTERN DALY

MARITHEL ] Aus.WESTERN DALY

ANGKAMUTHI

YADHAYKENU

ATAMPAYA

URADHI ATAMPAYA

URADHI ANGKAMUTHI

URADHI YADHAYKENU

KUGU NGANHCARA

WIK MUNGKAN

PANYJIMA

YINDJIBARNDI

NGALOOMA

MARTUTHUNIRA

WADJARRI

NYUNGAR

PINTUPI

WARLPIRI

MAYKULAN

NGAWUN

MAYI YAPI

MAYI THAKURTI

MAYAGUDUNA

PITTA PITTA

YALARNGA

KALKA TUNGU

PUNTHAMARA

WANGKUMARA

DIYARI

SBAGANDJI

GUMBAYNGGIR

GIDABAL

MANANDJALI

WAKKA WAKKA

DARAMBAL

GORENG GORENG

MURUWARI

GAMILARAAY

NGIYAMBAA

DJABUGAY

YIDINY

GUUGU YALANDYI

GUUGU YIMIDHIRR

BIRAY

WARGAMAY

NYAWAYGI

DYIRBAL

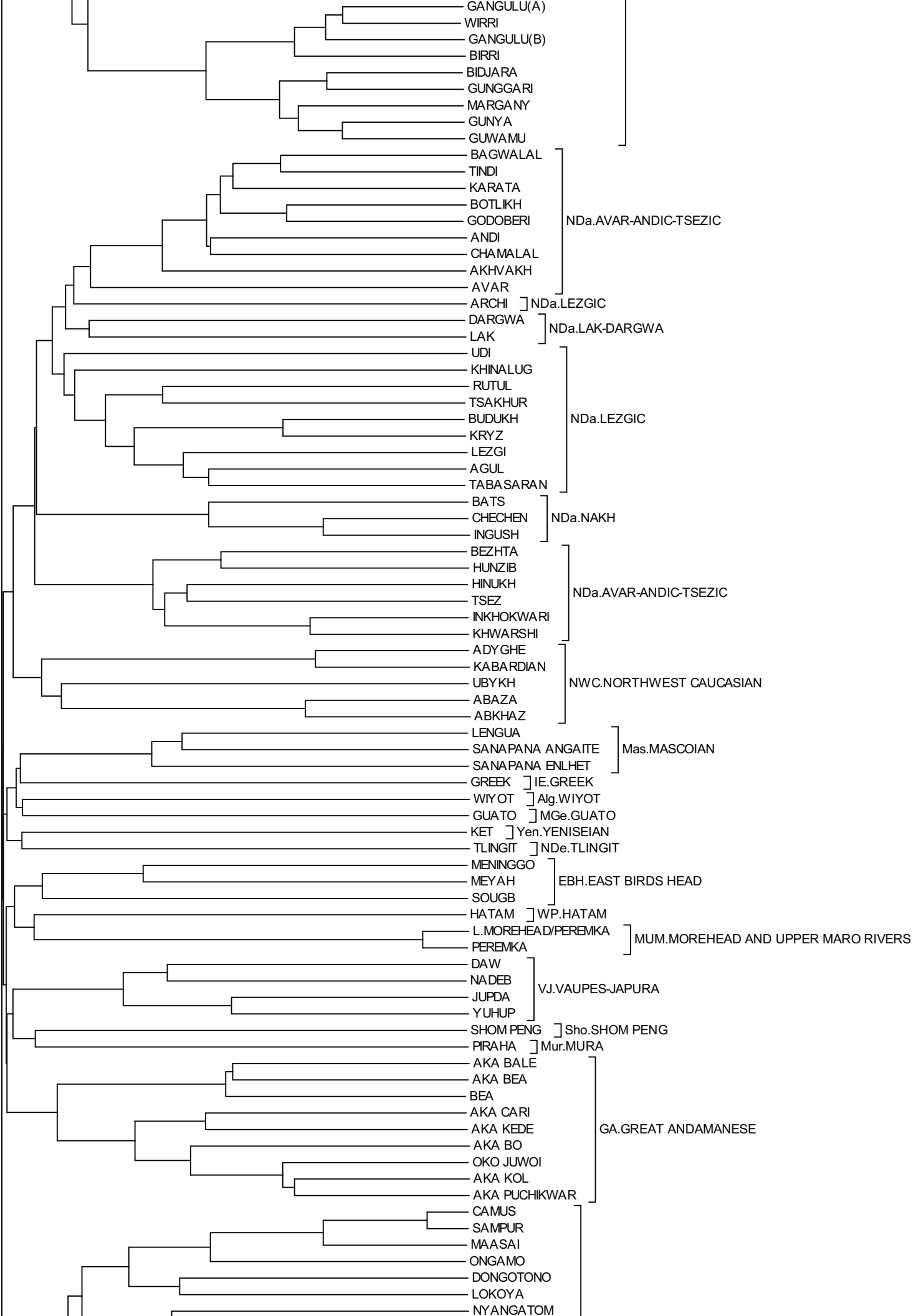
GIRAMAY

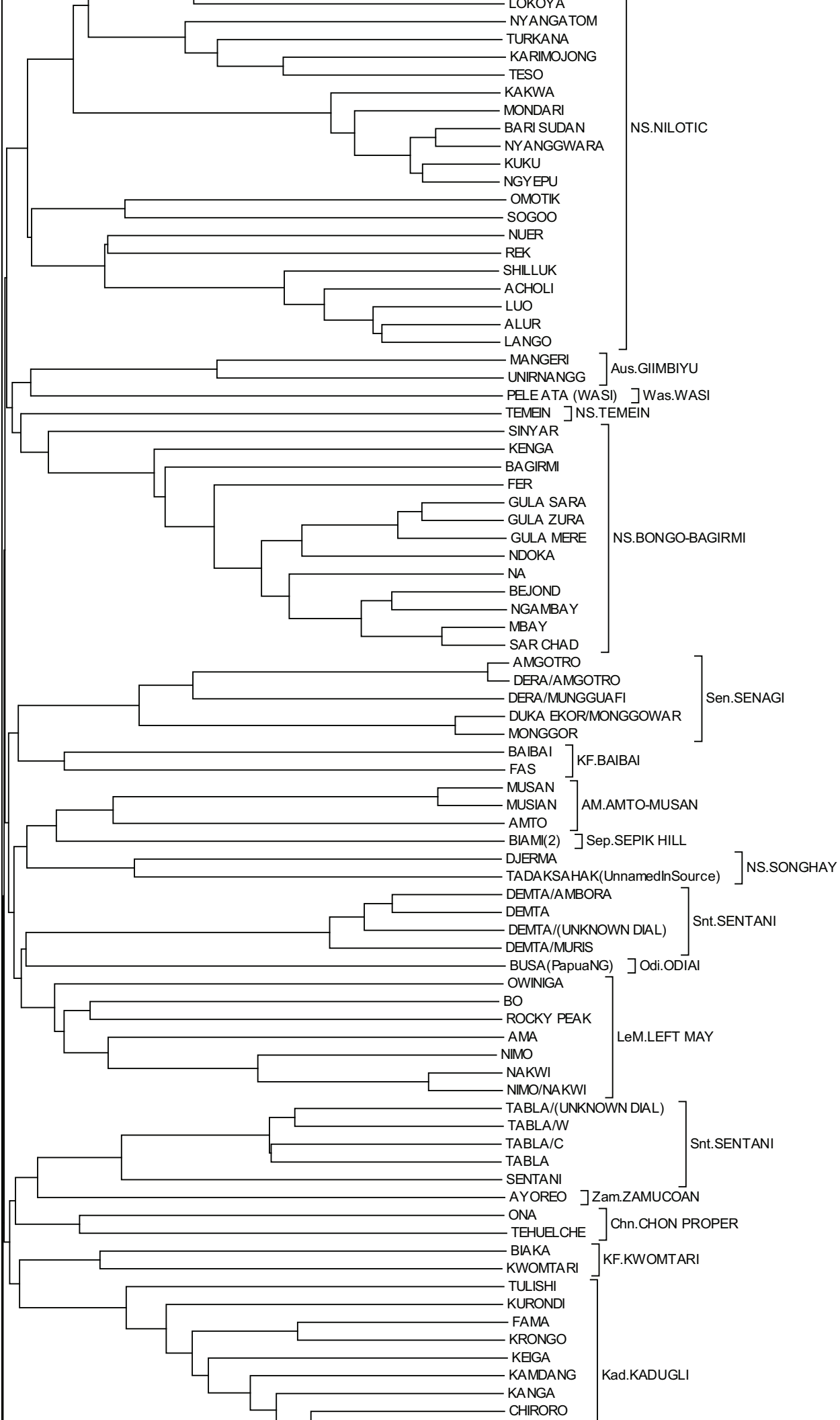
MAMU

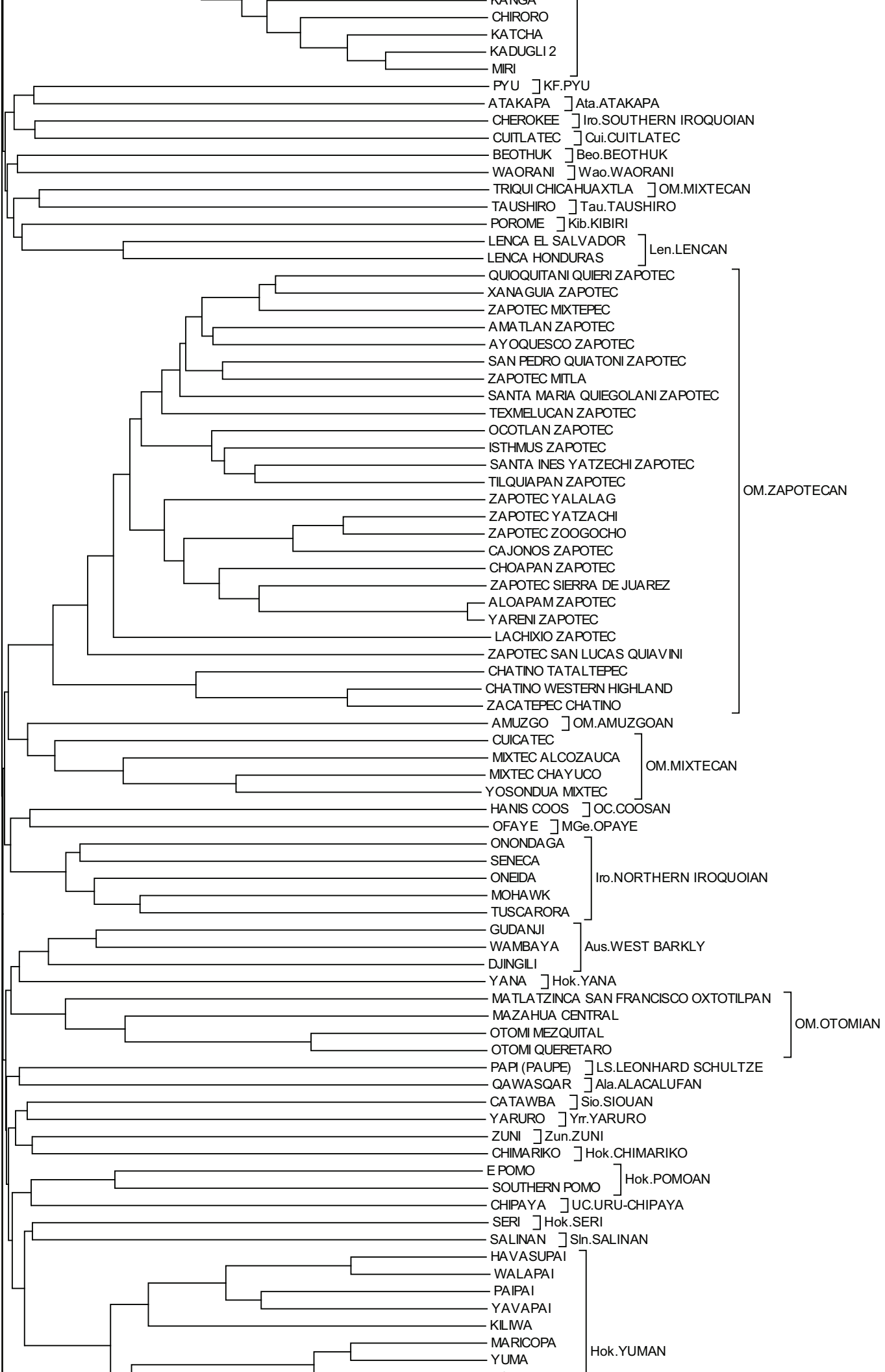
GANGULU(A)

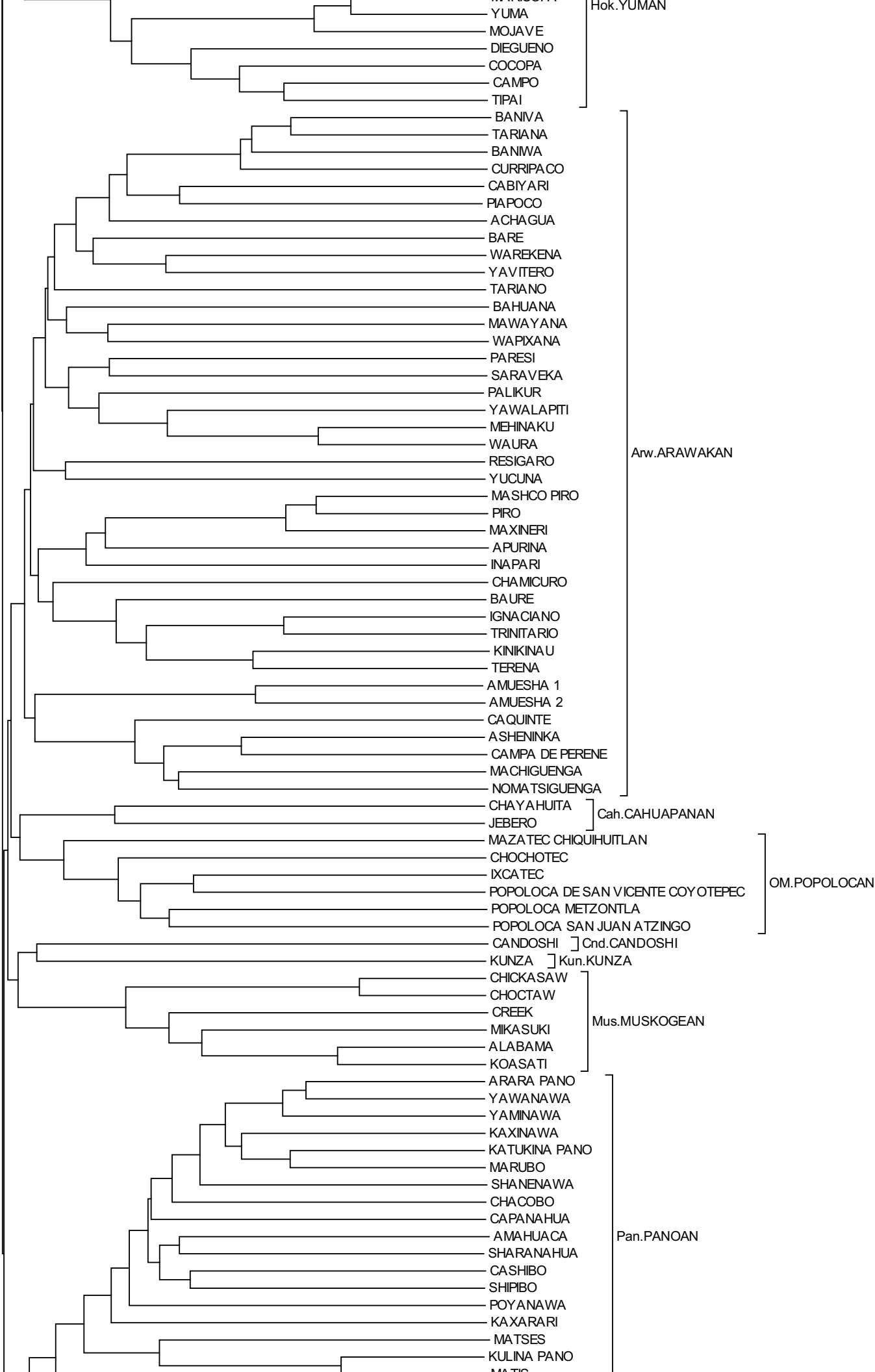
WIRRI

Aus.PAMA-NYUNGAN

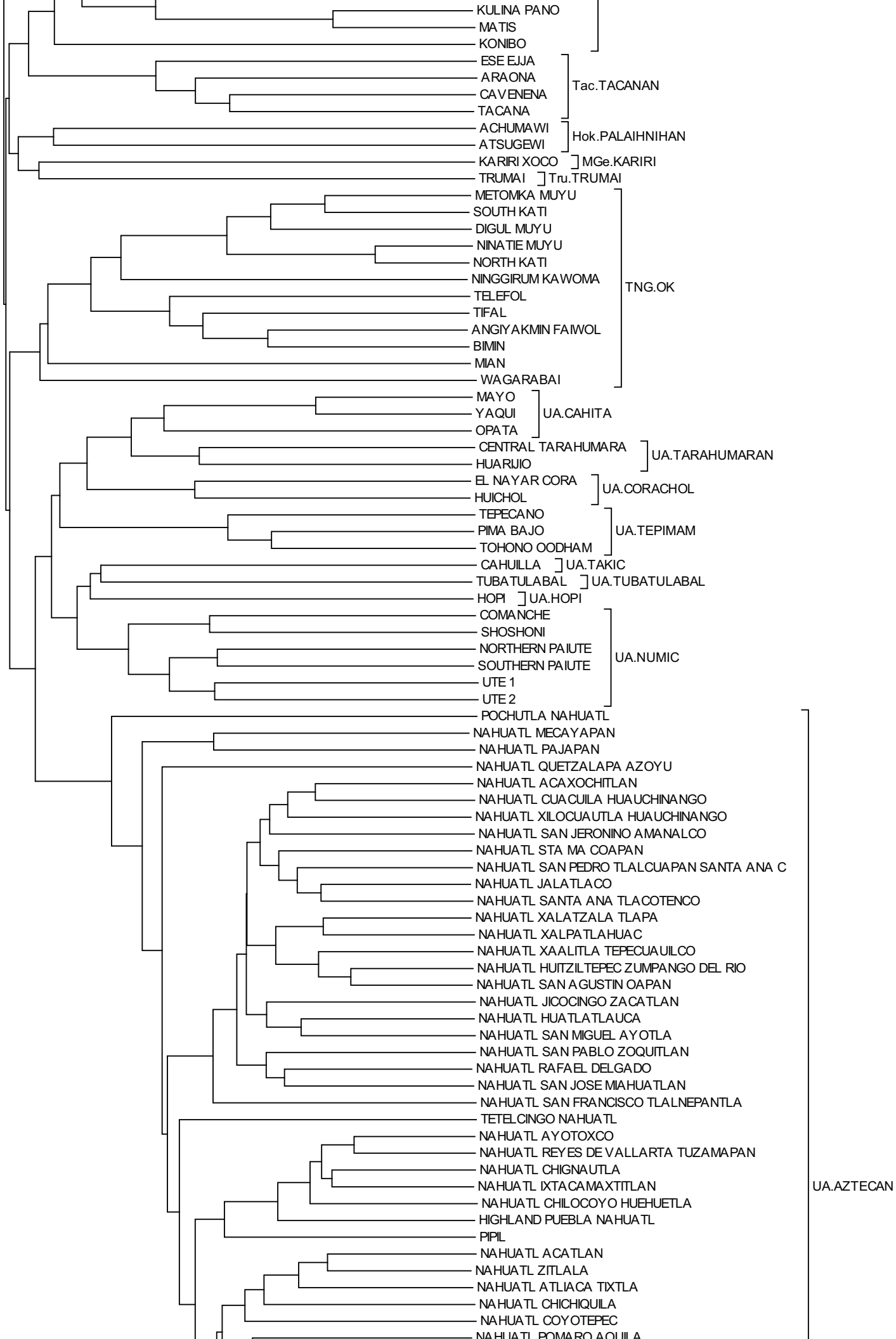












Tac.TACANAN

Hok.PALAIHNIHAN

MGe.KARIRI

Tru.TRUMAI

TNG.OK

UA.CAHITA

UA.TARAHUMARAN

UA.CORACHOL

UA.TEPIMAM

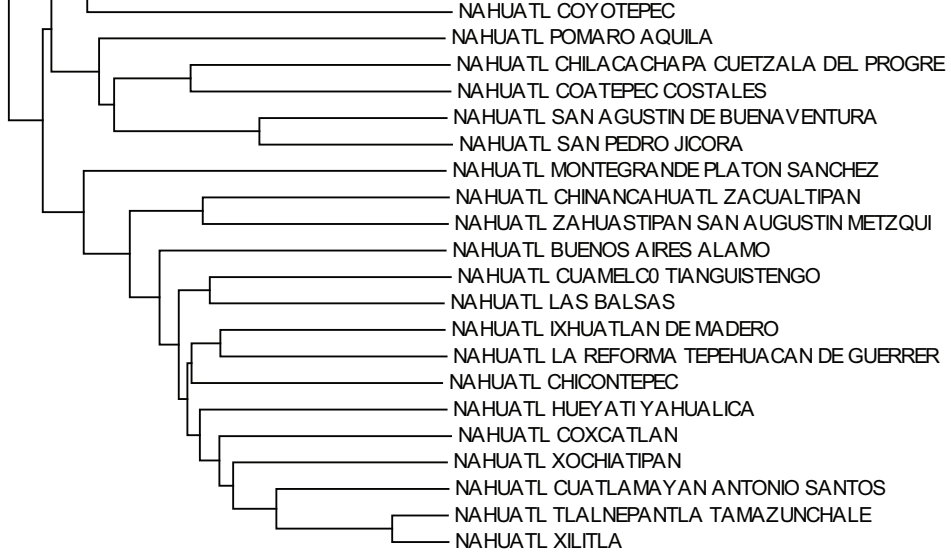
UA.TAKIC

UA.TUBATULABAL

UA.HOPI

UA.NUMIC

UA.AZTECAN



───────────  
 10