L2/23-070

Proposal to encode Bima characters

Febri Muhammad Nasrullah <niomi13@proton.me>
2023-02-05

1 Proposal history

L2/22-150	Proposal to Encode Bima Characters – Febri Muhammad Nasrullah
L2/22-128	Recommendations to UTC #172 July 2022 on Script Proposals – Deborah Anderson, et al

Related documents:

L2/11-091	Indonesian and Philippine Scripts and extensions not yet encoded or proposed for encoding in Unicode — Christopher Miller
L2/16-075	Proposal to encode VIRAMA signs for Buginese – Anshuman Pandey
L2/16-119	Representing Bima in Unicode – Anshuman Pandey
L2/16-156	Recommendations to UTC #147 May 2016 on Script Proposals – Deborah Anderson, et al
L2/16-159	Preliminary proposal to encode 'Buginese Extensions' in Unicode – Anshuman Pandey
L2/16-216	Recommendations to UTC #148 August 2016 on Script Proposals – Deborah Anderson, et al

2 Introduction

The Bima script is a Brahmi-derived script used to write Bima[1] (*nggahi mbojo*), a Malayo-Polynesian language primarily spoken by 914,800 people in the eastern part of Indonesia's Sumbawa Island (BPS Provinsi Nusa Tenggara Barat, 2022).

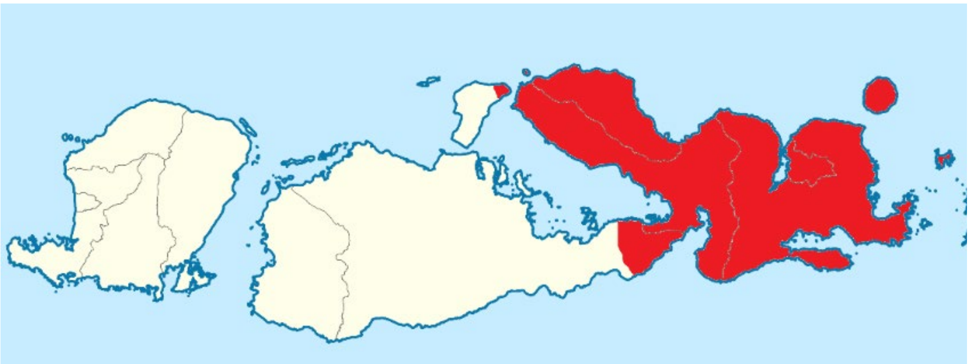


Figure 1. West Nusa Tenggara map, with red areas indicating Bimanese speaker areas.

The Bima script is closely related to the Lontara' script of South Sulawesi (encoded as "Buginese" in Unicode), sharing many basic features but with some modifications unique to Bima. At least until the 20th century, the Bima script was actively used in the courts of Bima Kingdom for various purpose, including genealogical records (fig. 6), daily registers, and religious texts (fig. 7). While the Bima script is still taught today in the local curriculum, modern Bima communities use the Latin alphabet for most everyday purpose. Contemporary use of Bima script is commonly limited to brief texts such as samples in textbooks (fig. 9), decorative glosses in street signs (fig. 11), and artworks (fig. 8), although revitalization efforts have been pursued by various parties.

[1] ISO 639-3: bhp

3 Comparison of Bima and Buginese repertoires

As previously mentioned, the repertoire of Bima script shares many similarities with the Lontara' script used for Buginese. A separate Bima script block would repeat many characters already encoded in the Buginese block, making it rather redundant. The author agrees with Miller (2011, p. 35) and Pandey's approach that Bima specific characters would be better encoded under Buginese-extension block that would also accommodate other historical and region-specific characters. But with this approach, a distinction should be made between unique characters and mere stylistic variants shared between Bima and Buginese repertoire. This is not always easy to determine, especially given the limited number of Bima and Buginese script users today, most of which are not well acquainted with naturally occurring variations in attested materials.

To clarify this, please refer to following table:

Character	**"Standard" Bugis**	**Bima**			
	glyph	glyph	EAP 988/1/4	EAP 988/1/29	EAP 988/1/145
ka					
ga[2]					
ŋa					
ca					
ja					
ɲa					
ɲca					
ta					
da[3]					
na					
nta	–				
ba[4]					
pa					
fa	–				

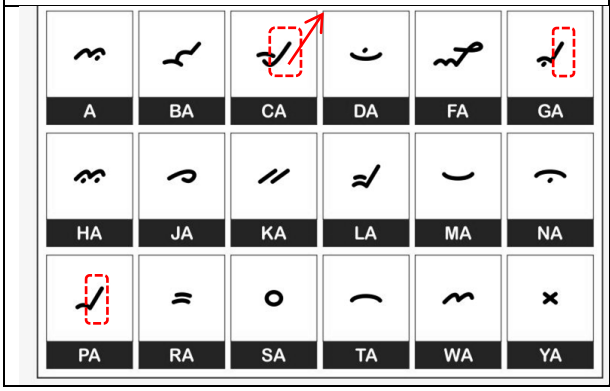
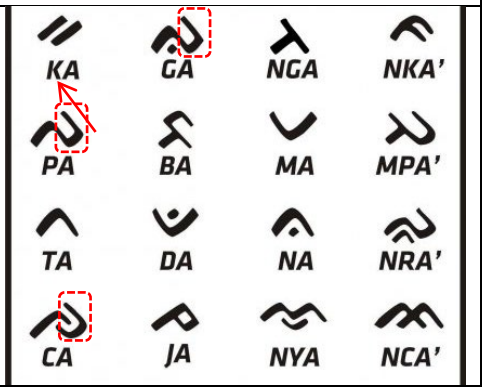
[2] Also read as /ᵑga/ in Bima
[3] Also read as /ⁿda/ and /ɗa/ in Bima
[4] Also read as /ᵐba/ and /ɓa/ in Bima

ma					
mpa					
ya					
ra					
la					
wa					
sa					
a					
ha					
-i					
-u					
-e					
-o					
reduplication	–				
gemination	–				
killer	–				
pallawa					
end of section					

Based on the above comparison, Bima characters can be divided into five groups:

1. 11 Bima characters in white-colored cells [ka, ja, ɲa, ta, da, ra, wa, sa, a, -i, -u] do not have noticeable glyph difference with their equivalence in Buginese. Separate codepoints are not necessary for these.
2. 1 Bima character in blue-colored cell [ha] has an identical glyph with another Bugis character with different sound value; [ya]. An annotation is needed.

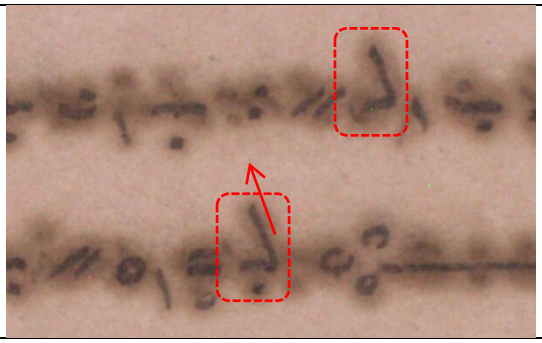
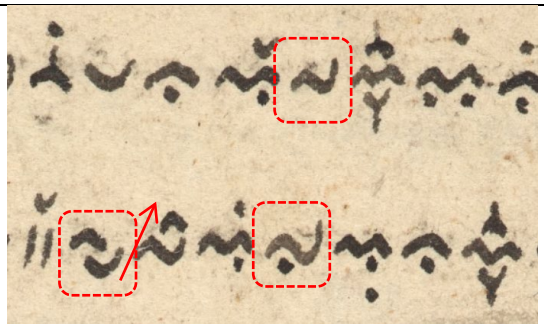
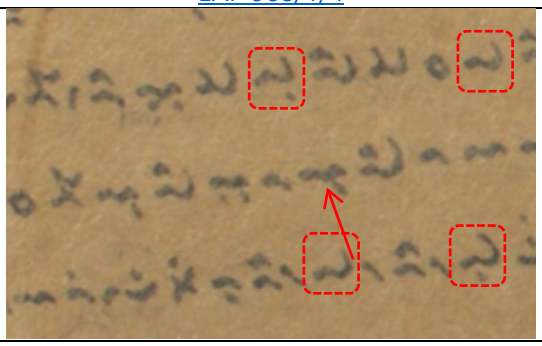
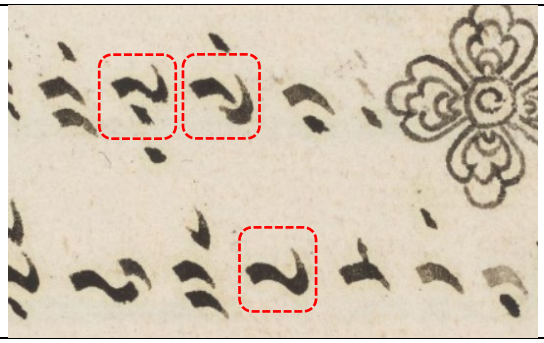
3. 5 Bima characters in green-colored cells [nta, fa, reduplication, gemination, killer] do not have Buginese equivalents and thus merit new code points.
   Contemporary Bima materials often use a circular glyph which can either be placed above ◌̊ or below the base letter ◌̥. The placement is purely stylistic and does not have any semantic significance. Historic materials also show right pointing chevron-like variant with similar behavior of placement ◌̊◌̥.[5] Based on this, some modern sources interpret Bima [killer] as a derivation of Arabic *sukun* which indeed may occur in both ◌ْ (U+FE7F) and ◌ۡ (U+06E1) forms throughout the archipelago. However, Jonker (1896, p. 12) states that the Bima [killer] mark is derived from Buginese [-ə] mark ◌ᨛ (U+1A1B).[6] Since the schwa sound does not occur in native Bimanese, the Buginese schwa mark was repurposed as a vowel silencer when writing Bimanese, comparable to how the sign is sometime repurposed to indicate nasal endings in Makassar writings. Regardless of the origin, both shapes (circular, chevron) and placements (above, below) can all co-occur in the same text without semantic significance. I decided ◌̊ as representative glyph because its shape is close to the two theories above and this shape often appears in manuscripts.
4. 11 characters in orange-colored cells [ga, ŋa, ca, ba, pa, mpa, la, -e, -o, pallawa, end of section] have similar base form but different details with Bugis norms, and Pandey proposed to encode most of them. However, it should be noted that some of them may also occur in texts originating from South Sulawesi. Take the character [ga], [pa], and [ca] for example. Contemporary tables often shows that the terminal right-hand strokes of these characters are straight in Bima but bent towards the left in Bugis:

Contemporary table of Bima characters	Contemporary table of Bugis characters
	

Based on this, Pandey (2016, p. 5) proposed that separate [ga], [pa], and [ca] should be encoded. However, it is unfortunately common for contemporary materials of Indonesian scripts to create reductive or oversimplified tables that omit once common variants. In some cases, this has led to erroneous characters that poorly represent authentic materials. Pertinent to this case, counter examples can be easily found where so called Bima-like shapes are found in historical Bugis/Makassar text and vice versa:

Bima text	Bugis Makassar text

[5] It is unclear why Pandey (2016, p. 2) used an upward pointing circumflex-like glyph when attestation he included did not resemble such shape.
[6] Which Miller (2011, p. 42) cited in his report.

	
EAP 988/1/4	Add MS 12363
	
EAP988/1/142	Add MS 12346

Unlike Pandey, I do not think that separate encoding of these characters is necessary to write Bimanese and their appearance should be handled at the font level.

For Bima [la] and [ba], while the exact same form never occurred in Bugis/Makassar texts as far as I know, in-between shapes can be found in numerous materials, including those written in Bima. I will encode both characters which often appears in the manuscript (𑻞 and 𑻟).

It should be noted however that while historic materials show a wide range of possible shape iteration, contemporary users might only be acquainted with a somewhat narrower range of shape owing to the script's limited application. Issue of cultural identity may also play a role in wedging perceived difference between Bima and Buginese script by over-representing diverging shapes and downplaying shared letterforms. While I believe that the characters shouldn't be encoded, dialog with local actors is desirable to alleviate any misunderstanding that representing Bima can only be done through encoding new characters; already encoded Buginese characters can still be modified at the font level to suit contemporary Bima community preference.

Character in yellow cells has completely different base form than the corresponding Bugis letters. I feel that stronger case could be made for encoding these rather than characters in orange cells.

4 Script details

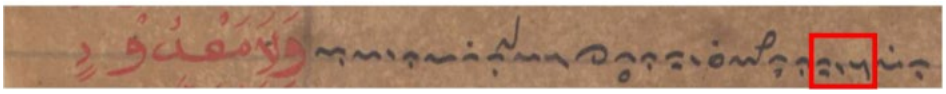
The Bima script is essentially Bugis script with a few changes or additional characters to suit the Bima language. As a result, the basic rules and structure of Bima, such as line breaking, spacing, rendering, and so on, are similar to that of Buginese. I use the standard Brahmi order because there is no special order or native collation for the Bima characters, including both newly proposed ones and already encoded characters.

4.1 Consonant reduplication

The 𑻠 REDUPLICATION SIGN is used for reduplicating the onset consonant of the previous syllable in a word. Its usage is based upon a convention opposite that of the doubling of vowel signs for the abbreviation of syllables. As there is no sign or other means for marking the inherent vowel of a consonant, it is not possible to abbreviate two contiguous syllables consisting of identical consonants by doubling their vowel signs. This applies solely to cases where the onset

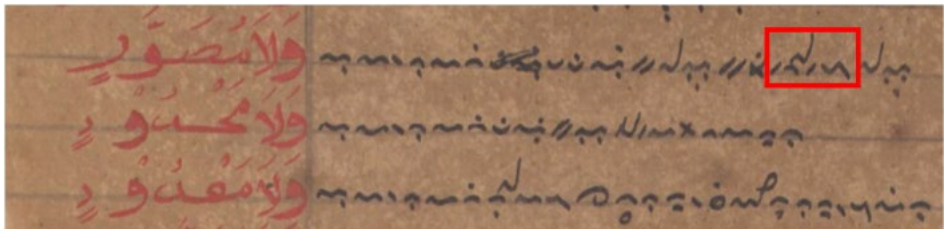
consonant and the consonant of the following syllable are identical. In such a case, the consonant following the onset is replaced with the REDUPLICATION SIGN.

The usage of REDUPLICATION SIGN is illustrated in the following examples. The boxed text in the excerpt below is ᨒ᨞ᨒ, which is the syllable ᨒ᨞ <ᨒ RA, ◌᨞ VOWEL SIGN O> followed by ᨒ REDUPLICATION SIGN:



This text is to be read as ᨒᨒ *rora*. As shown, the REDUPLICATION SIGN reduplicates the onset consonant RA of the previous syllable *ra*, but does not carry the accompanying vowel *o*; it retains the inherent vowel *a*.

The REDUPLICATION SIGN may also serve as a vowel carrier, as shown below. The boxed text shows ᨕᨊᨒ, which is the syllable ᨕᨊ BE followed by an ᨒ REDUPLICATION SIGN carrying the ᨕ◌ VOWEL SIGN E.



This text is to be read as ᨕᨊᨒ *bebe*.

The usage of REDUPLICATION SIGN is based upon the practice of using the digit '2' as a mark of repetition. The form of REDUPLICATION SIGN is derived from ᩞ U+A9CF JAVANESE PANGRANGKEP, which is itself based upon ٢ U+0662 ARABIC-INDIC DIGIT TWO. A similar system of syllable reduplication is used in Buginese. However, a separate REDUPLICATION SIGN-type character has not been encoded for Buginese and the Unicode standard states that the JAVANESE PANGRANGKEP is to be used. As pairs of base letters and combining vowel signs belonging to different script blocks may complicate rendering, syllable identification, collation, and other processing, it may not be practical to use JAVANESE PANGRANGKEP as a base letter in Bima script contexts. For this reason, the REDUPLICATION SIGN is proposed for encoding as a separate character.

5 Approach for encoding

The present 'Buginese' block is made up of 30 characters: 23 consonant letters, 5 vowel signs, and 2 punctuation marks. Unicode requires 31 characters to represent Bima: There are 23 letters, four vowel combining signs, a killer sign, a reduplication sign, a gemination sign, and a punctuation sign. Six of these letters are unique, but the other seventeen can be expressed using existing Buginese characters. Two vowel signs are identical, two are variant forms, and one does not appear in Bima. The following steps must be taken:

a. Encode the following Bima characters in 'Buginese Supplement'

U+16EA0	᩠	BUGINESE LETTER BIMA NYCA
U+16EA1	ᩡ	BUGINESE LETTER BIMA NTA
U+16EA2	ᩢ	BUGINESE LETTER BIMA BA
U+16EA3	ᩣ	BUGINESE LETTER BIMA YA
U+16EA4	ᩤ	BUGINESE LETTER BIMA LA
U+16EA5	ᩥ	BUGINESE LETTER BIMA FA
U+16EA6	ᩦ	BUGINESE SIGN REDUPLICATION
U+16EA7	◌ᩧ	BUGINESE SIGN GEMINATION
U+16EA8	◌ᩨ	BUGINESE SIGN KILLER

b. Add the following annotation to the names list for the 'Buginese' block

1A10 ᨐ BUGINESE LETTER YA

• used in Bima for /h/

5.1 Character data

UnicodeData.txt

```
16EA0;BUGINESE LETTER BIMA NYCA;Lo;0;L;;;;;N;;;;;
16EA1;BUGINESE LETTER BIMA NTA;Lo;0;L;;;;;N;;;;;
16EA2;BUGINESE LETTER BIMA BA;Lo;0;L;;;;;N;;;;;
16EA3;BUGINESE LETTER BIMA YA;Lo;0;L;;;;;N;;;;;
16EA4;BUGINESE LETTER BIMA LA;Lo;0;L;;;;;N;;;;;
16EA5;BUGINESE LETTER BIMA FA;Lo;0;L;;;;;N;;;;;
16EA6;BUGINESE SIGN REDUPLICATION;Lo;0;L;;;;;N;;;;;
16EA7;BUGINESE SIGN GEMINATION;Mn;0;NSM;;;;;N;;;;;
16EA8;BUGINESE SIGN KILLER;Mn;9;NSM;;;;;N;;;;;
```

LineBreak.txt

```
16EA0..16EA6;AL  # Lo    [7] BUGINESE LETTER BIMA NYCA..BUGINESE SIGN REDUPLICATION
16EA7..16EA8;CM  # Mn    [2] BUGINESE SIGN GEMINATION..BUGINESE SIGN KILLER
```

IndicSyllabicCategory.txt

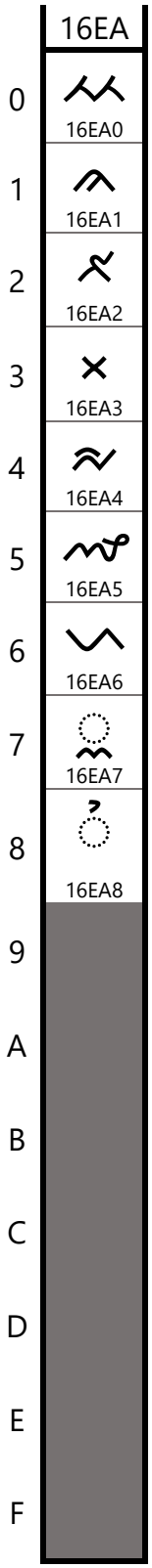
```
# Indic_Syllabic_Category=Consonant
16EA0..16EA5; Consonant  # Lo    [6] BUGINESE LETTER BIMA NYCA..BUGINESE LETTER BIMA FA
# Indic_Syllabic_Category=Consonant_Placeholder
16EA6   ; Consonant_Placeholder  # Lo    BUGINESE SIGN REDUPLICATION
# Indic_Syllabic_Category=Gemination_Mark
16EA7   ; Gemination_Mark        # Mn    BUGINESE SIGN GEMINATION
# Indic_Syllabic_Category=Pure_Killer
16EA8   ; Pure_Killer    # Mn    BUGINESE SIGN KILLER
```

IndicPositionalCategory.txt

```
# Indic_Positional_Category=Top
16EA8           ; Top           # Mn    BUGINESE SIGN KILLER
# Indic_Positional_Category=Bottom
16EA7           ; Bottom        # Mn    BUGINESE SIGN GEMINATION
```

6 Acknowledgments

The author wishes to express gratitude to Norbert Lindenberg and Aditya Bayu Perdana for their assistance and advice, particularly in the technical elements of encoding and providing excellent fonts.



Buginese letters for Bima

16EA0 BUGINESE LETTER BIMA NYCA
16EA1 BUGINESE LETTER BIMA NTA
16EA2 BUGINESE LETTER BIMA BA
16EA3 BUGINESE LETTER BIMA YA
16EA4 BUGINESE LETTER BIMA LA
16EA5 BUGINESE LETTER BIMA FA
16EA6 BUGINESE SIGN REDUPLICATION
16EA7 BUGINESE SIGN GEMINATION
16EA8 BUGINESE SIGN KILLER

- vowel killer, doesn't produce conjuncts
- always rendered visibly

Bibliography

BPS Provinsi Nusa Tenggara Barat. (2022). *Provinsi Nusa Tenggara Barat Dalam Angka 2022*. Mataram: BPS Provinsi Nusa Tenggara Barat. Retrieved from https://ntb.bps.go.id/publication/2022/02/25/81b407c481be37affd75d6f5/provinsi-nusa-tenggara-barat-dalam-angka-2022.html

Holle, K. F. (1882). *Tabel van Oud- en Nieuw- Indische Alphabetten. Bijdrage tot de palaeographie van Nederlandsch-Indië.* Batavia: W. Bruining & Co.; 's Hage: M. Nijhoff.

Jonker, J. C. (1896). *Bimaneesche spraakkunst.* Makassar.

Miller, C. (2011, March 14). *Unicode Technical Note #35: Indonesian and Philippine Scripts and Extensions.* Retrieved from Unicode Technical Note: http://www.unicode.org/notes/tn35/

Pandey, A. (2016, May 3). *Representing Bima in Unicode*. Retrieved from https://www.unicode.org/L2/L2016/16119-bima.pdf

Zollinger, H. (1851). *Verslag van eene reis naar Bima en Soembawa, en naar eenige plaatsen op Celebes, Saleijer en Floris, gedurende de maanden Mei tot December 1847.* Batavia: Bataviaasch Genootschap van Kunsten en Wetenschappen.

Source of Manuscripts

Object	Documentation/Current Location	Ref. number	Place of origin
Paper	BL EAP	EAP988/1/4	Samparaja Museum, Bima, Indonesia
Paper	BL EAP	EAP988/1/29	Samparaja Museum, Bima, Indonesia
Paper	BL EAP	EAP988/1/145	Samparaja Museum, Bima, Indonesia
Paper	BL EAP	EAP988/1/10	Samparaja Museum, Bima, Indonesia

Documentations

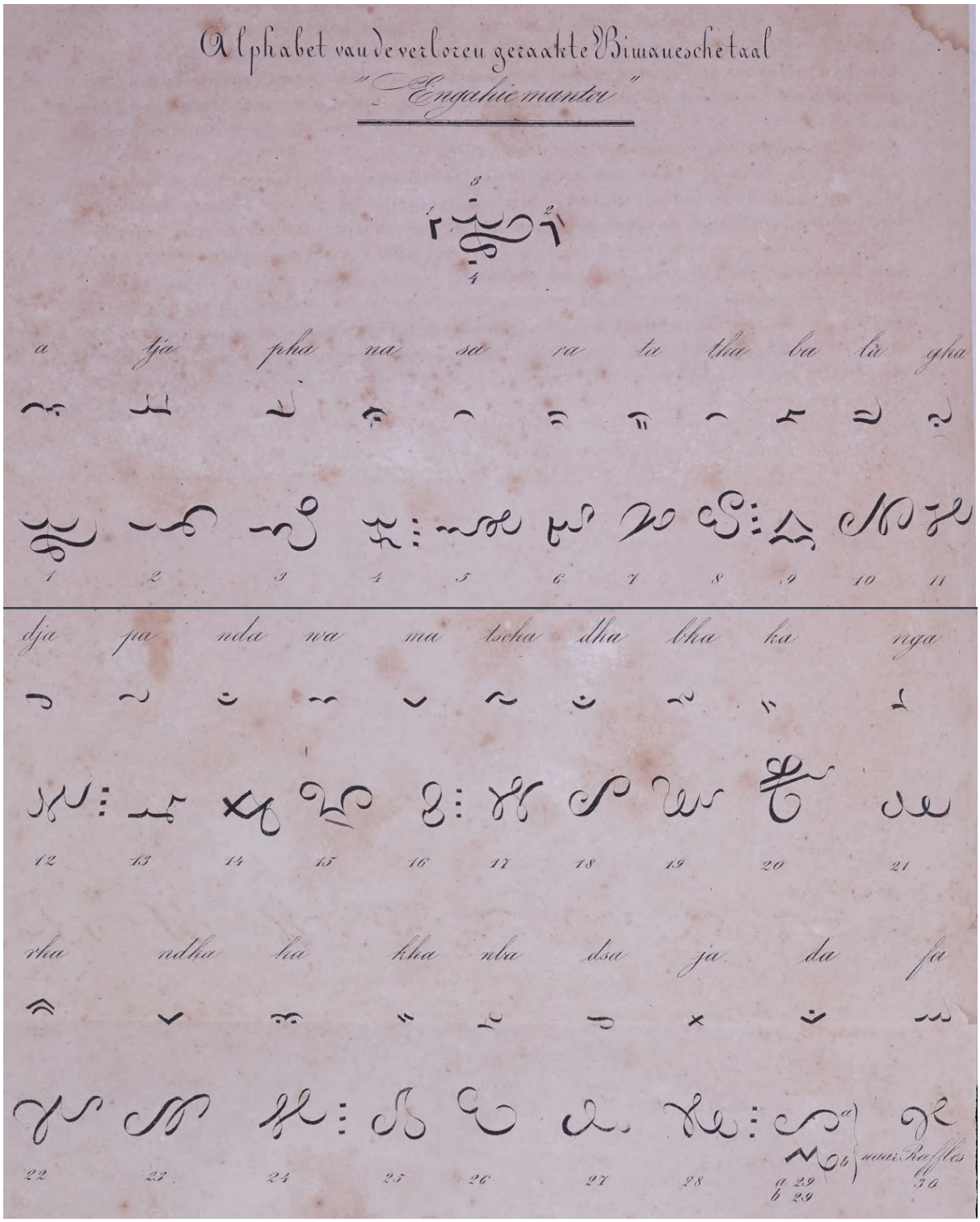


Figure 2: Bima characters (Zollinger, 1851, p. 134).
Noted that LETTER NYCA *mistakenly identified as* TJA (CA) *and* TA *as* THA.

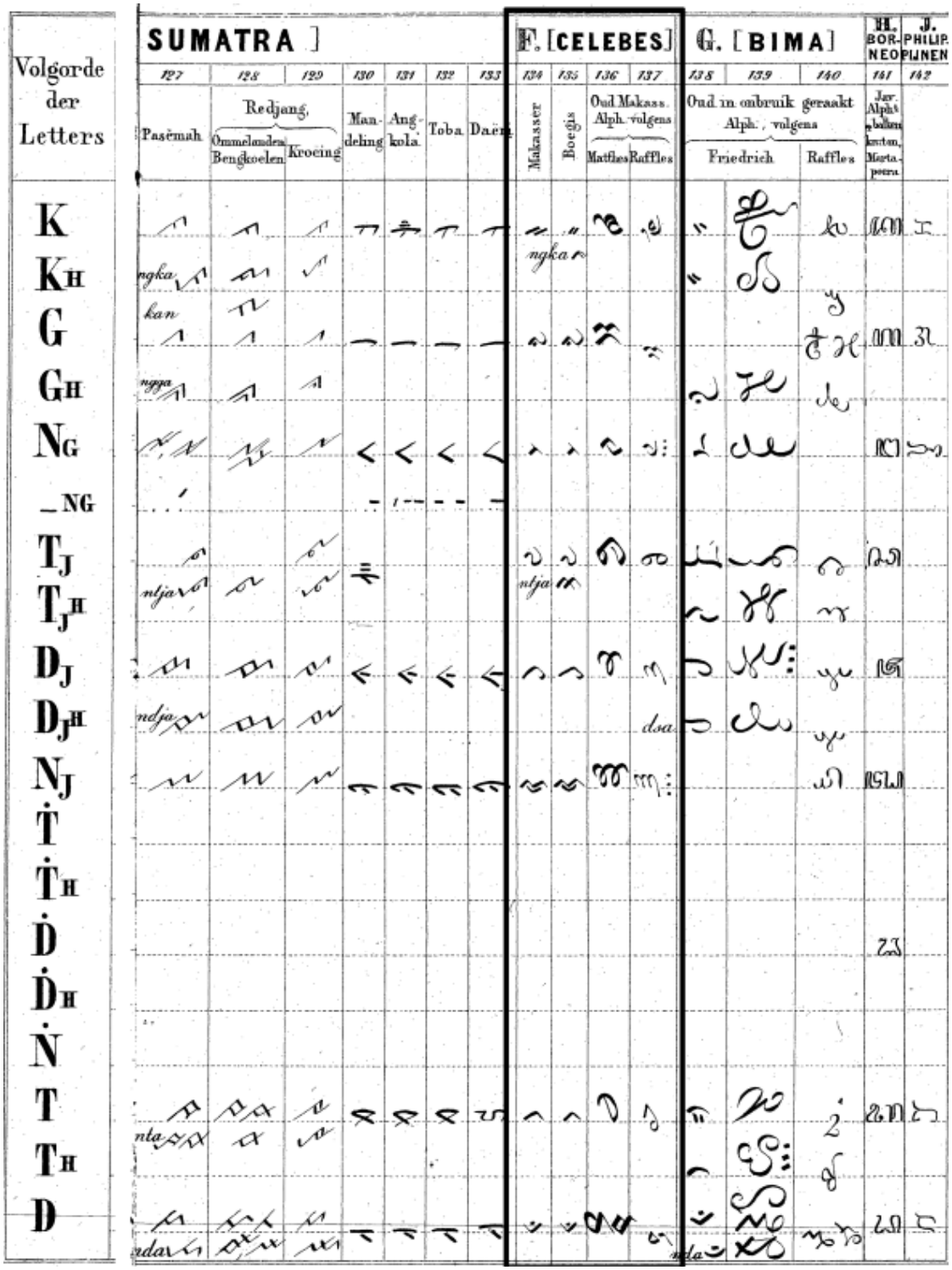


Figure 3: Chart showing scripts from "Bima" (Holle, 1882, p. 11). *Columns 138 show the Bima script. The column showing transliteration ("Volgorde der Letters") has been stitched from the previous page in Holle. Noted that* LETTER NYCA *mistakenly identified as* TJA (CA) *and* TA *as* THA.

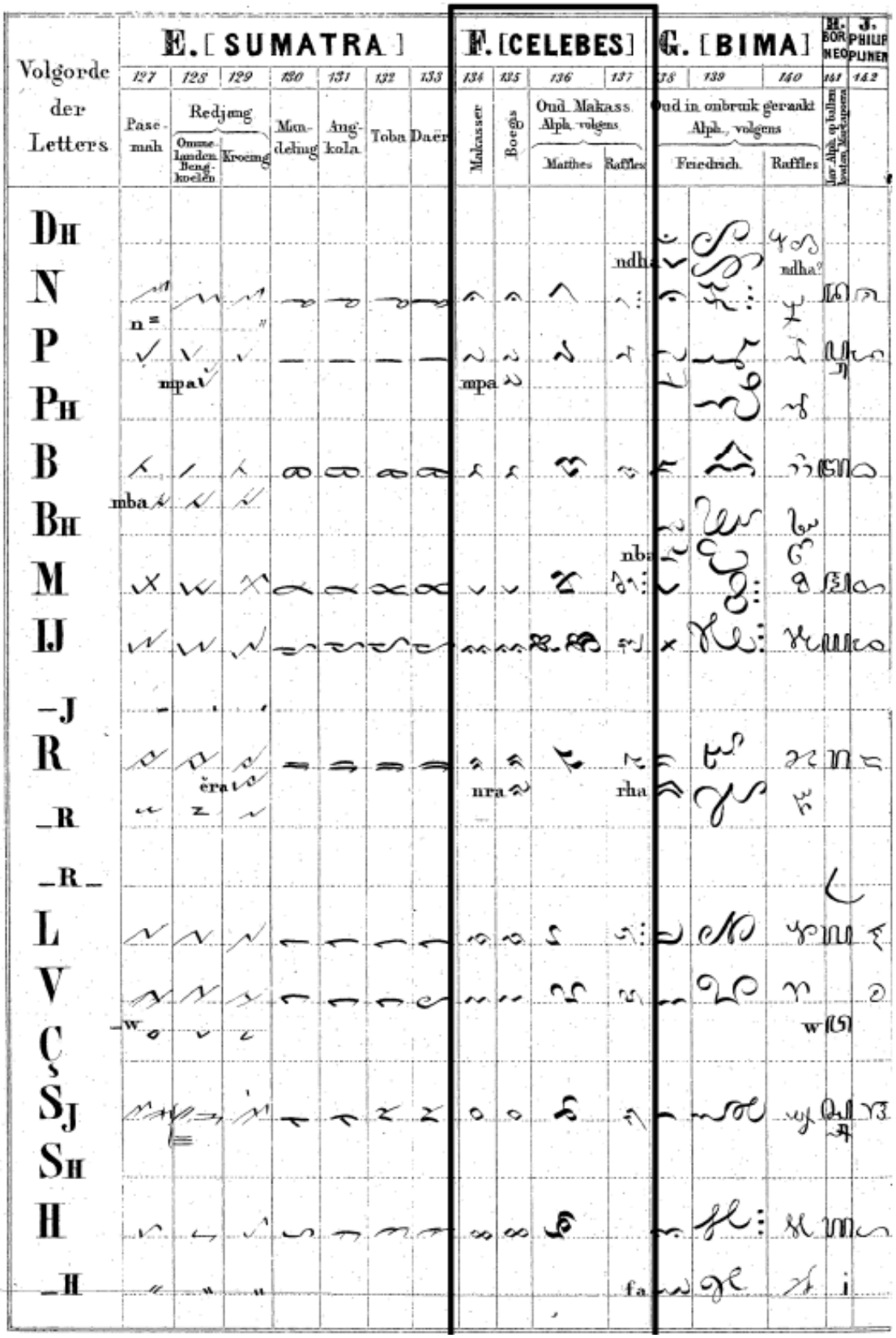


Figure 4: Chart showing scripts from "Bima" (Holle, 1882, p. 11). *Columns 138 show the Bima script.*

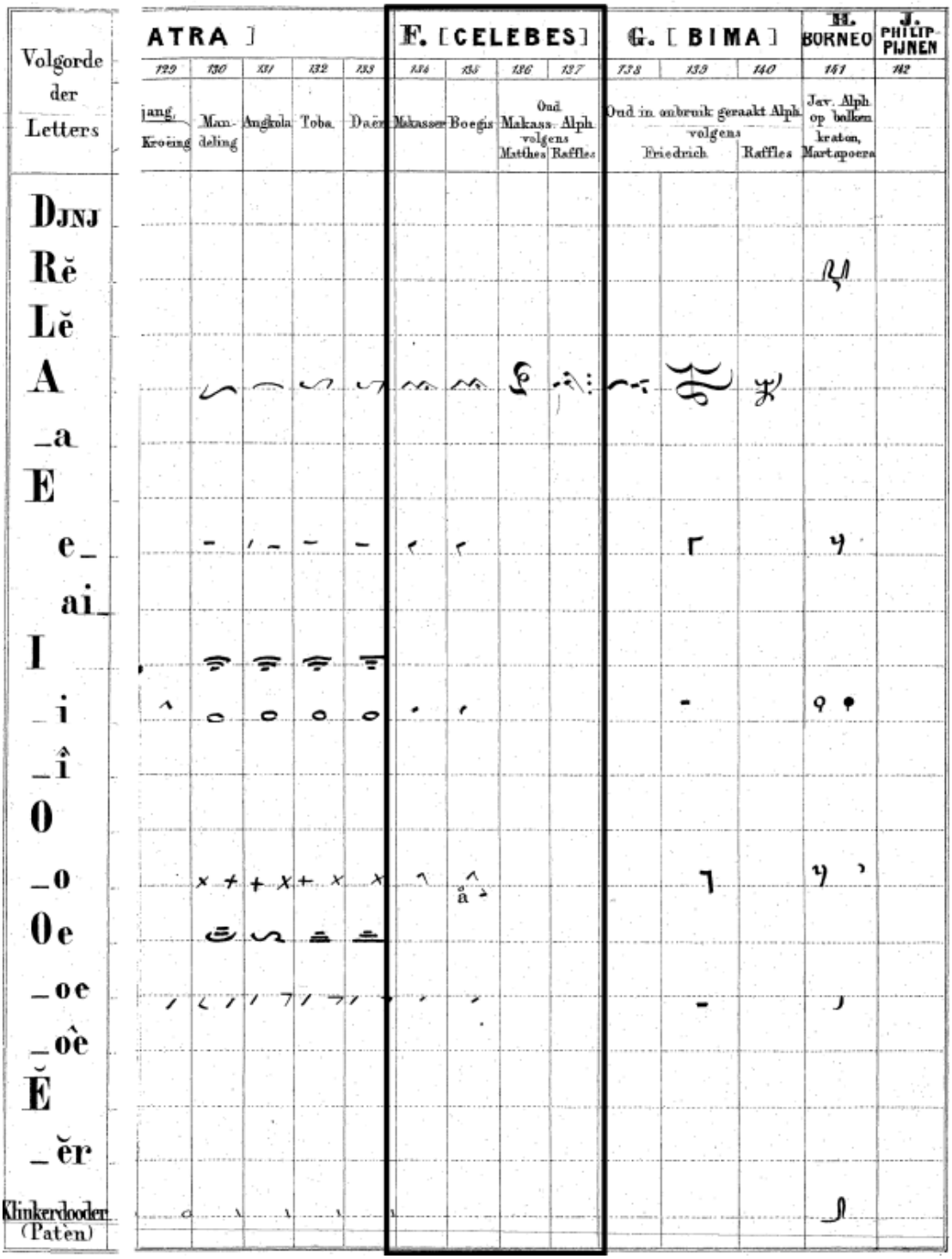


Figure 5: Chart showing scripts from "Bima" (from Holle 1882: 11). Columns 138 show the Bima script.

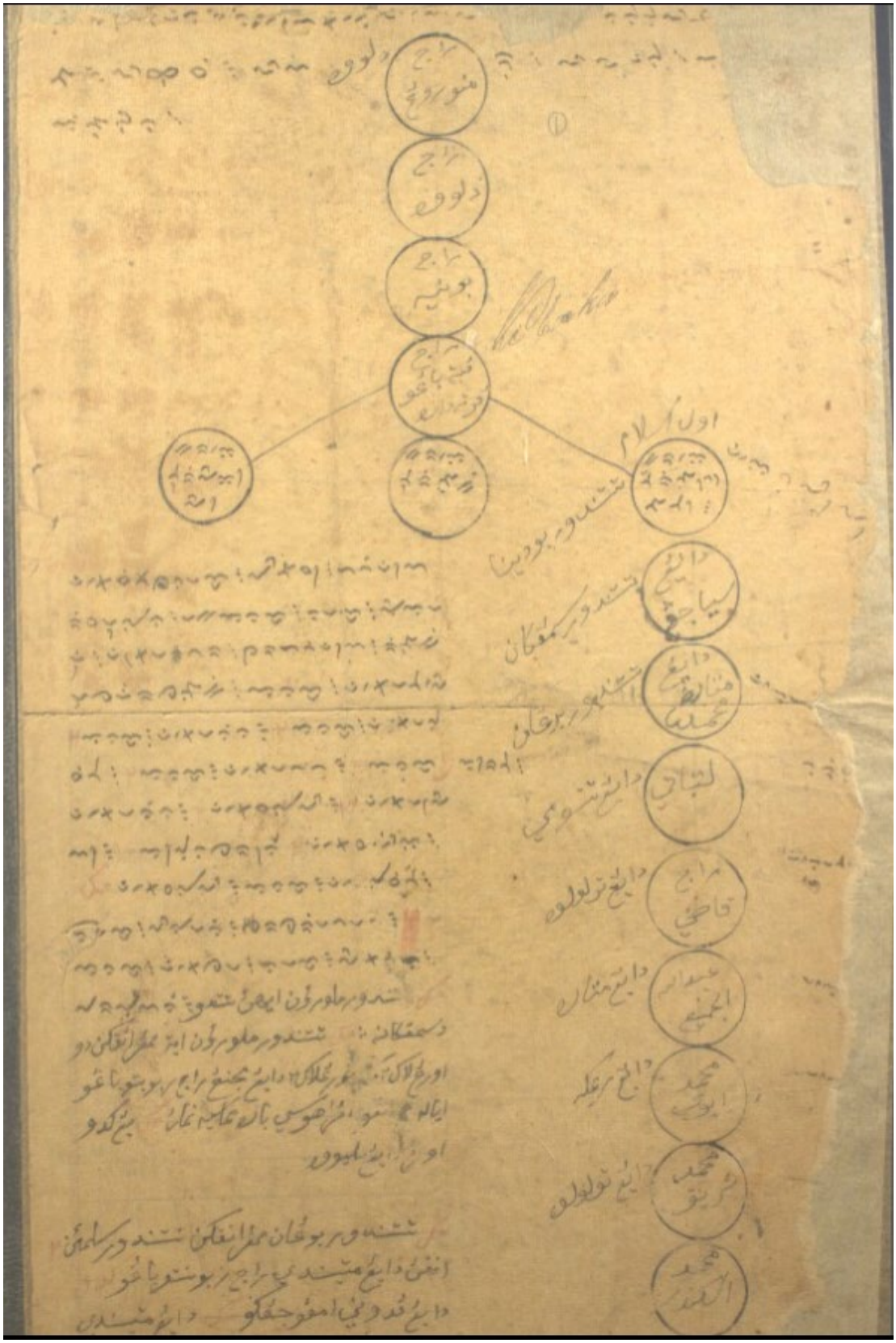


Figure 6: Excerpt from a handwritten genealogical record in the Buginese, Bima, and Jawi script (EAP988/1/145).

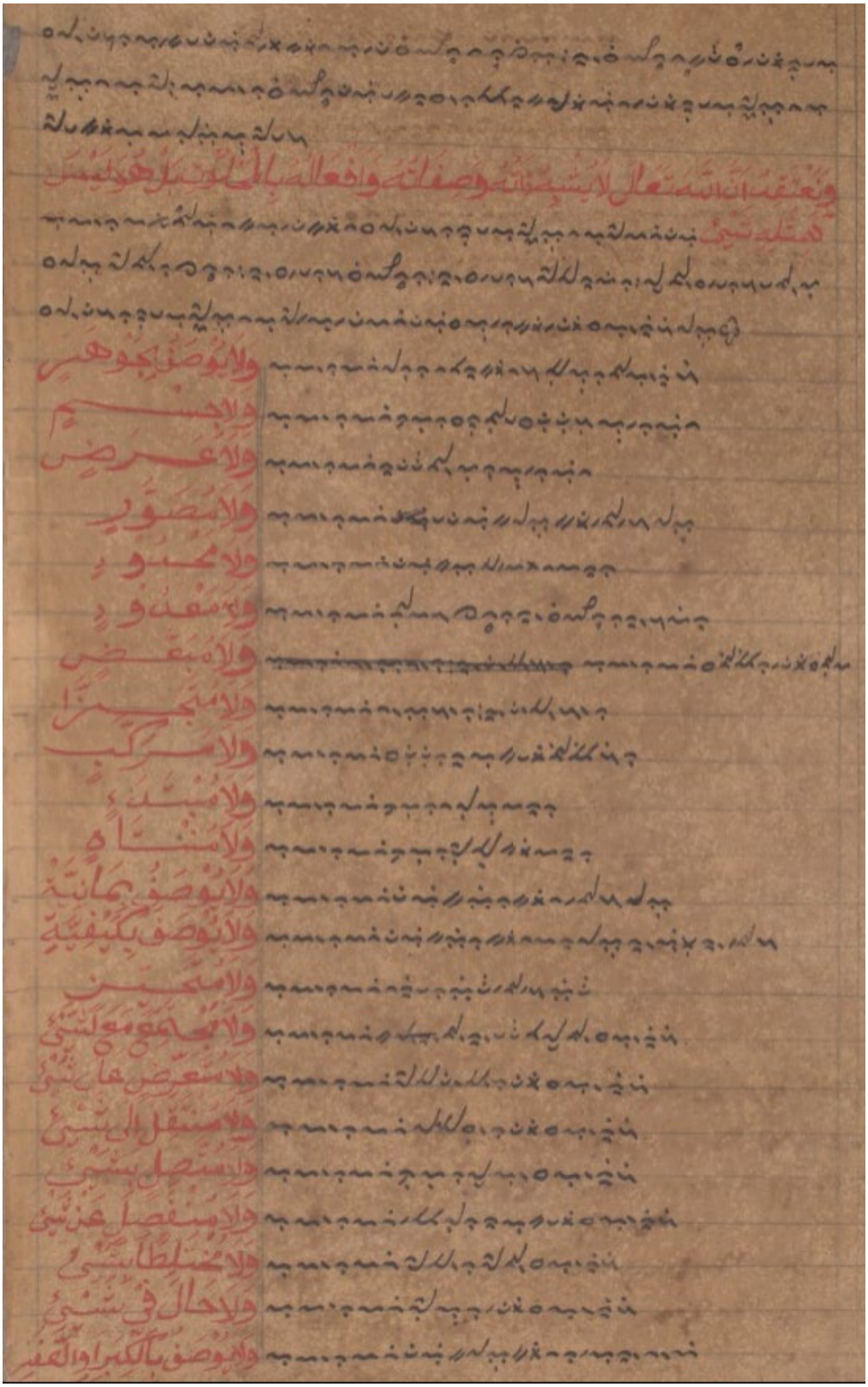


Figure 7: Excerpt from a handwritten religious book in the Bima and Arabic script (EAP988/1/29).

Figure 8: Calligraphy in Bima script. Source: https://t.co/hWIyjlvRbM

Figure 9: Bima script class at elementary school. Taken on 20 January 2022. Source: https://t.co/1MsBqGx46b

Figure 10: A Bimanese proverb in Bima and Latin script. Source: https://t.co/6MuYvIMwId

Figure 11: The tri-scriptural street signs in Bima city. Source: https://t.co/vwkg30GZ2d

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[7]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Proposal to Encode Bima Characters*
2. Requester's name: *Febri Muhammad Nasrullah*
3. Requester type (Member body/Liaison/Individual contribution): *Invidual contribution*
4. Submission date: *2023-02-05*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *Yes*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): *Yes*
      Proposed name of script: *Buginese Supplement*
   b. The proposal is for addition of character(s) to an existing block: *No*
      Name of the existing block:
2. Number of characters in proposal: *9*
3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary		B.1-Specialized (small collection)	v	B.2-Specialized (large collection)	
C-Major extinct		D-Attested extinct		E-Minor extinct	
F-Archaic Hieroglyphic or Ideographic				G-Obscure or questionable usage symbols	

4. Is a repertoire including character names provided? *Yes*
   a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? *Yes*
   b. Are the character shapes attached in a legible form suitable for review? *Yes*
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? *Aditya Bayu Perdana*
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): *Aditya Bayu Perdana, OFL*
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *Yes*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *Yes*

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

[7] Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *Yes*
   If YES explain *Proposal to Encode Bima Characters (L2/22-150)*
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? *Yes*
   If YES, with whom? *Aziz*
   If YES, available relevant documents: *This document*
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *Yes*
   Reference: *This document*
4. The context of use for the proposed characters (type of use; common or rare) *Common*
   Reference: *This document*
5. Are the proposed characters in current use by the user community? *Yes*
   If YES, where? Reference: *This document*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *No*
   If YES, is a rationale provided?
   If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? *Yes*
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? *Yes*
    If YES, is a rationale for its inclusion provided? *Yes*
    If YES, reference: *This document*
11. Does the proposal include use of combining characters and/or use of composite sequences? *No*
    If YES, is a rationale for such use provided?
    If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    If YES, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? *No*
    If YES, describe in detail (include attachment if necessary)
13. Does the proposal contain any Ideographic compatibility characters? *No*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: