

## Encoding Mongolian head letters

Aaron Bell | Greg Eck | Andrew Glass | Andrew West

2014/01/17

### Summary

The Mongolian block starts with U+1800 MONGOLIAN BIRGA which is a kind of ornament that usually marks the beginning of a text or folio. Like Tibetan, which has a related character (U+0F04), there are multiple different types of the birga symbol. Five types of birga have been identified in publications that pioneered the Mongolian encoding (Erdenechimeg et al. 1999 and Quejingzhabu 2000). These publications include guidelines that encode the birga variants using sequences based on the standard MONGOLIAN BIRGA U+1800 with one of the MONGOLIAN FREE VARIATION SELECTORS (U+180B–180D). Because there are just three MONGOLIAN FREE VARIATION SELECTORS, ZWJ is used as the fourth variation marker (U+1800 U+200D).

These sequences for the birga variants have not been accepted by the Unicode Consortium and are not included in the current version of StandardizedVariants.txt (The Unicode Consortium 2013c). All other variation sequences specified in Erdenechimeg et al. 1999 and Quejingzhabu 2000 are included in StandardizedVariants.txt. The absence of these sequences from StandardizedVariants.txt, or a recommendation on how to access them has caused confusion among users and implementers of the standard.

The authors of this document would like to discuss options for the correct encoding of these characters prior to updating the Mongolian code charts (see doc L2/14-031) so that the work on the code charts can include or exclude these variation sequences as appropriate. Three options are presented in this document and a recommendation is made for the separate encoding of these characters.

### Birga

The Mongolian block currently (6.3) includes a single character assignment for Mongolian birga:

1800	Mongolian
<b>Punctuation</b>	1825
1800 ∞ MONGOLIAN BIRGA	
	→ 0F04 ∞ tibetan mark initial yig mgo mdun ma
1801 ∴ MONGOLIAN ELLIPSIS	

Figure 1. Excerpt from the Mongolian code chart (Unicode 6.3)

In his publication *Měnggǔ wén biānmǎ* 蒙古文编码 (Quejingzhabu 2000), Prof. Quejingzhabu specifies four additional types of birga which he encodes using sequences with either a MONGOLIAN FREE VARIATION SELECTOR (U+180B–180D) or the ZERO WIDTH JOINER (U+200D).

In addition to these types, we have noticed another type attested in one manuscript using the Todo variant of Mongolian script (see appendix). The full set of known types is as follows:

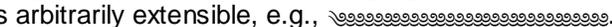
Symbol (rotated 90°)	Suggested name	Quejingzhabu 2000	Comments
?	MONGOLIAN BIRGA	U+1800	This is the usual type and is already encoded in Unicode.
?	Ornamented birga	U+1800 U+180B	This type is frequently seen in Mongolian documents.
∞	Rotated birga	U+1800 U+180C	This type is attested in archaic texts (see appendix). It may also exist for presentation purposes.

	Double birga	U+1800 U+180D	This type is unknown to the authors in Mongolian texts. It is well attested in Tibetan sources.
	Triple birga	U+1800 ZWJ	This type is unknown to the authors in Mongolian texts. It is well attested in Tibetan sources.
	Swirl birga	Not defined	This type occurs in a Kalmyk text in the Todo variant of Mongolian script. See appendix.

The Mongolian birga is related to a set of Tibetan head letters which function in the same way:

0F04		TIBETAN MARK INITIAL YIG MGO MDUN MA
		<ul style="list-style-type: none"> <li>• honorific; marks beginning of text or start of new folio</li> <li>→ 1800  mongolian birga</li> </ul>
0F05		TIBETAN MARK CLOSING YIG MGO SGAB MA
		<ul style="list-style-type: none"> <li>• follows and ligates with initial yig-mgo</li> </ul>
0F06		TIBETAN MARK CARET YIG MGO PHUR SHAD MA
0F07		TIBETAN MARK YIG MGO TSHEG SHAD MA

Figure 2. Excerpt from the Tibetan code chart (Unicode 6.3)

The approach to encoding in the Tibetan block has been to encode multiple head marks separately rather than using variation sequences. The Tibetan encoding also makes use of a closing sign (U+0F05) that ligates with U+0F04. This means that the symbol is arbitrarily extensible, e.g., .

## Concerns

The authors share the following concerns regarding the current practice of encoding variants of the Mongolian birga using sequences defined in Quejingzhabu 2000.

1. The use of the three Mongolian Free Variation Selectors is not extensible to new types since the all three Mongolian Free Variation Selectors have been used. For example, the Swirl birga doesn't fit within this encoding system
2. ZWJ should not be used as a substitute for a variation selector
3. The double and triple birga forms should be encoded in the same way as Tibetan, so that the number of loops is arbitrarily extensible
4. The authors are not certain the double and triple loop forms actually occur with Mongolian text

## Options

The following options should be considered for representing the Mongolian birga types.

1. Use the existing sequences  
Using the existing sequences defined in Quejingzhabu 2000 (p.17) would maximize compatibility with existing fonts and documents.

基本字符			变形显	
No	字符	名称	No	图形
000		M <sup>᠐</sup> BIRGA	1	
			2	
			3	
			4	
			↓	

现形式	认同				变形显现形式	
	M <sup>᠐</sup>	T <sup>᠐</sup>	S <sup>᠐</sup>	MA <sup>᠐</sup>	录入法	总序号
名称						
birga first form	br	br				000
birga second form	br	br				001
birga third form	br	br				002
birga fourth form	br	br				003

Figure 3. Excerpt from Quejingzhabu 2000: 17, showing the variation sequences in the penultimate column.

A new sequence or alternate solution would be needed for the Todo birga and possibly for the triple birga. These could use the Variations selectors from the range U+FE00–FE0F:

Symbol (rotated 90°)	Suggested name	Sequence
	Triple birga	U+1800 U+FE00
	Swirl birga	U+1800 U+FE01

The disadvantage of this approach are that the concerns above are not addressed.

2. Use sequences with variation selectors

Rather than using the limited set of Mongolian Free Variation Selectors, we could endorse using the Variation Selectors from the range U+FE00–FE0F:

Symbol (rotated 90°)	Suggested name	Sequence
	MONGOLIAN BIRGA	U+1800
	Ornamented birga	U+1800 U+FE00
	Rotated birga	U+1800 U+FE01
	Double birga	U+1800 U+FE02
	Triple birga	U+1800 U+FE03
	Swirl birga	U+1800 U+FE04

3. Encode atomic code points

It may be preferable to use atomic code points to represent these. If so, the encoding should be done following the model of the Tibetan head letters:

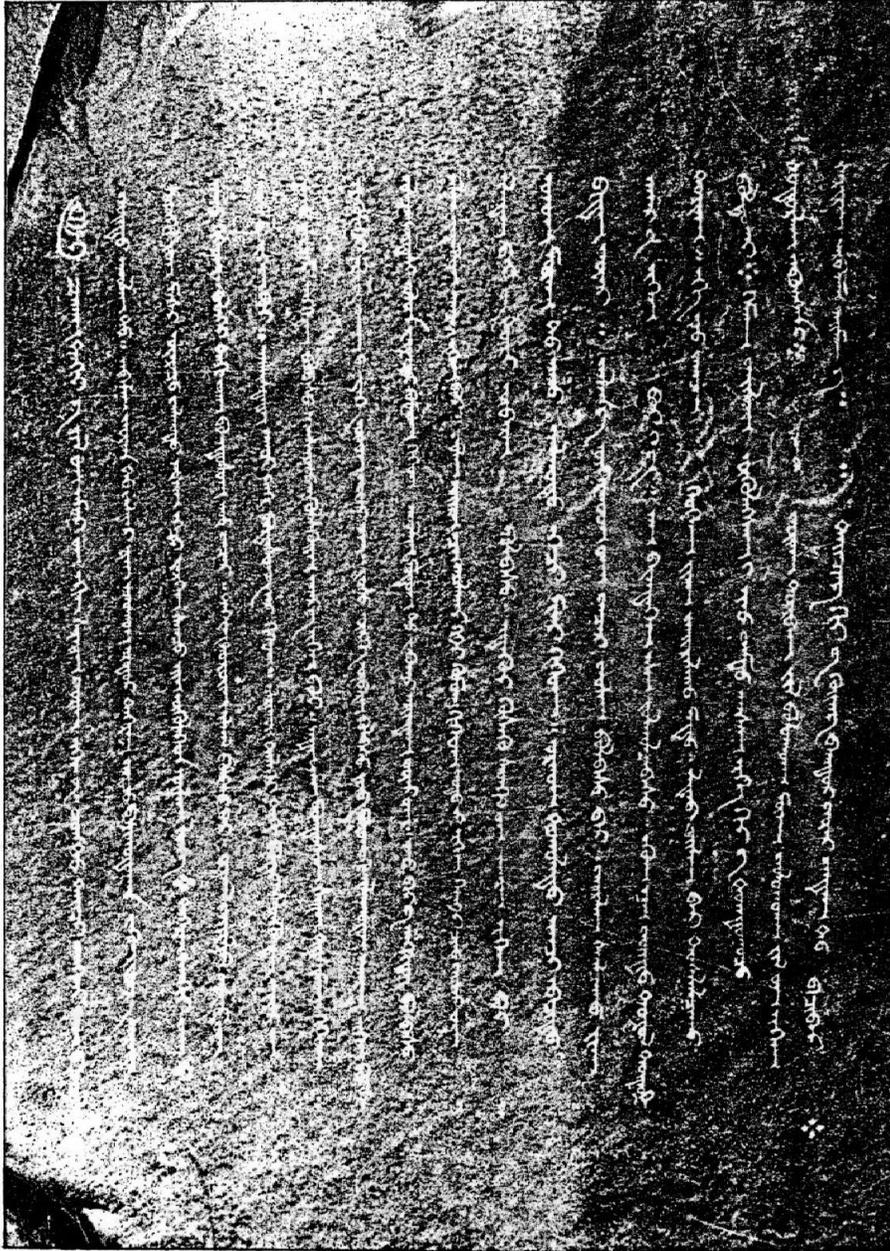
Symbol (rotated 90°)	Suggested name	Code point
	MONGOLIAN BIRGA	U+1800
	Rotated birga	U+181C
	Initial ornamental birga	U+181B
	Closing ornamental birga	U+181D
	Swirl birga	U+181E

Due to the layout differences between Mongolian and Tibetan (vertical vs. horizontal), it is not suitable to reuse the Tibetan initial and closing head letters (U+0F04, U+0F05) for Mongolian text.

## Request

The authors would like feedback from the UTC as to which approach to encoding the additional birga types should be pursued. Chiefly, does the existence of a document that recommends a particular approach to encoding these signs represent sufficient grounds to maintaining these assignments, or should an alternative solution be sought?

Appendix  
Rotated birga



Цогтын хадны "Их бичээс"-ийн байдал (1624 он)

Figure 4. Evidence for the rotated birga at top of the first column from the left.

## Swirl birga

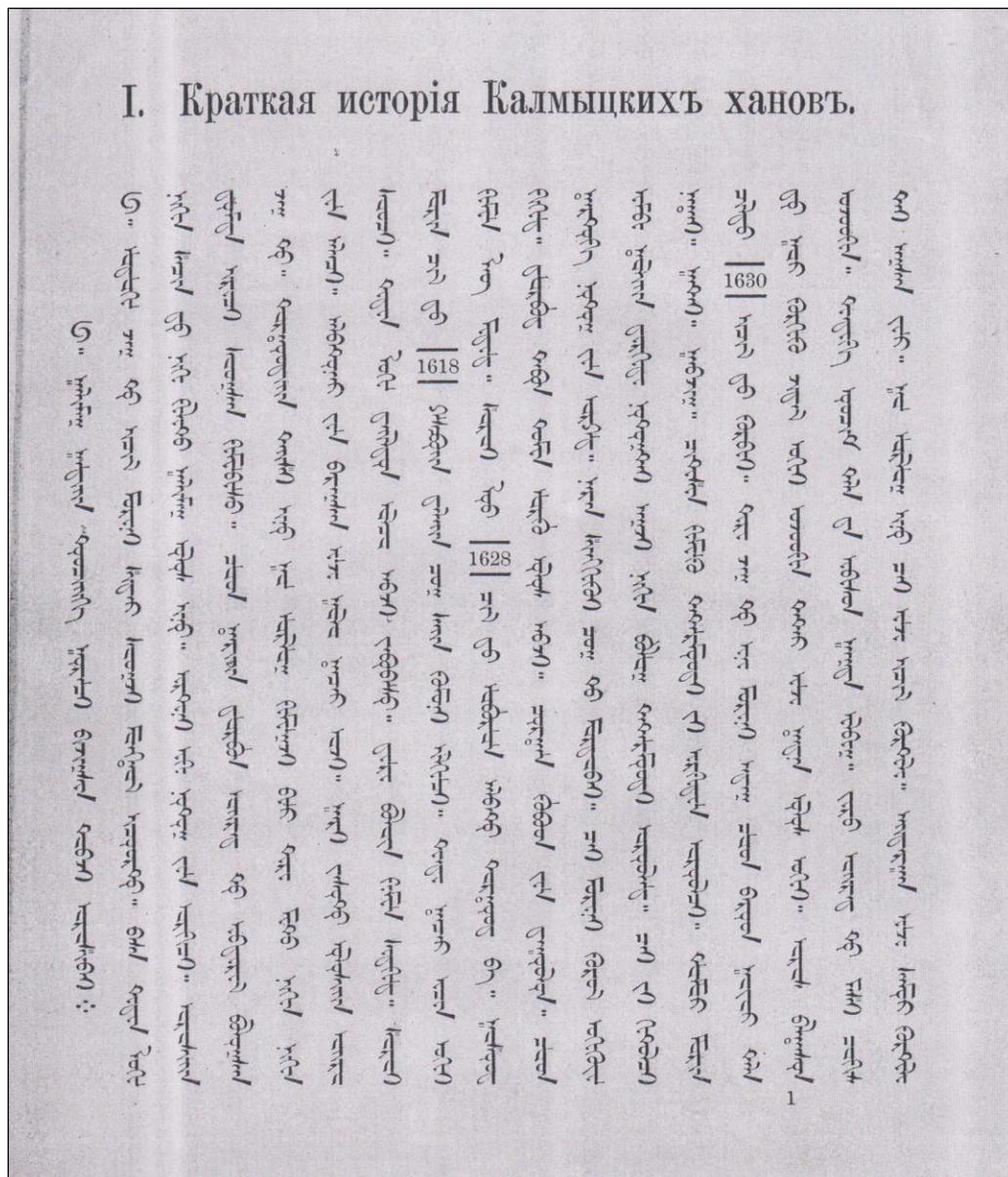


Figure 5. Page from a Kalmyk document containing evidence for the Swirl birga in the first and second columns from the left.

## References

- Erdenechimeg, Myatav, Richard Moore and Yumbayar Namsrai. 1999 "Traditional Mongolian Script in the ISO/IEC 10646 and Unicode Standards" *UNU/IIST Report No. 170*. August 1999. Accessed from: <http://www.unicode.org/~asmus/mongolian/MD001-unu-tr170.html> on 2014/01/17.
- Quejingzhabu (确精扎布). 2000. *Měnggǔ wén biānmǎ* 蒙古文编码. Hohhot: Nèi Měnggǔ dàxué chūbǎnshè 内蒙古大学出版社.
- The Unicode Consortium 2013a. "Chapter 13. Additional Modern Scripts." The Unicode Standard Version 6.3. Accessed from: <http://www.unicode.org/versions/Unicode6.2.0/ch13.pdf> on 2014/01/17.
- . 2013b. "Code Charts." The Unicode Standard Version 6.3. Accessed from: <http://www.unicode.org/Public/6.3.0/charts/CodeCharts.pdf> on 2014/01/17.
- . 2013c. "Standardized Variants." The Unicode Standard Version 6.3. Accessed from: <http://www.unicode.org/Public/UCD/latest/ucd/StandardizedVariants.txt> on 2014/01/17.