L2/14-067

Proposal to encode five Mongolian head marks

Aaron Bell | Greg Eck | Andrew Glass | Andrew West 2014/02/06

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 exclusion of the sequences also means that their use is not conformant with Unicode since StandardizedVariants.txt is a normative contributory data file.

The authors of this document believe that the Mongolian birgas should be encoded as atomic code points, and ask that the UTC consider this proposal.

Birga

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

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

Birga types and the variation sequences

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. It may also exist for presentation purposes in horizontal layout.

	Double birga	U+1800 U+180D	This type is rare 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, but is included in the publications that established the Mongolian encoding. 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.

Tibetan head marks

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

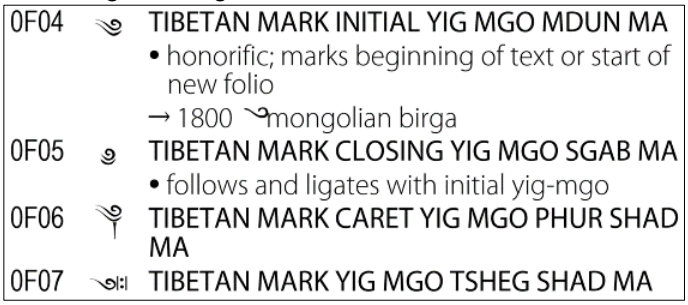
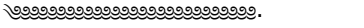


Figure 2. Excerpt from the Tibetan code chart (Unicode 6.3)

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.,  .

The extensible approach to encoding birgas is not suitable for Mongolian since the birga is written horizontally, that is, perpendicular to the writing steam, unlike Tibetan which is in line with the stream. Therefore Tibetan can accommodate an arbitrary length while Mongolian cannot. Therefore, having atomic code points for the double and triple Mongolian birgas will be suitable. The authors are not aware of a requirement for greater than a triple birga type in Mongolian.

Due to the layout differences between Mongolian and Tibetan (vertical vs. horizontal), it is not suitable to reuse the Tibetan head marks (U+0F04, U+0F05) in place of Mongolian equivalents. As such, dedicated Mongolian birga code points are required.

Concerns about the current use of undefined variation sequences

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. These sequences are not defined in StandardizedVariants.txt
4. The Tibetan head marks have been encoded atomically. For encoding consistency and because there is overlap in the user community for both Mongolian and Tibetan texts that include head marks, the Mongolian birga types should follow the same model and be encoded atomically

Proposal

Based on the concerns listed above with using variation sequences to encode these birga types, we propose they be encoded as atomic code points within the Mongolian block. The Mongolian block has sufficient space for these birgas at the end of the second column (U+181*) of the range. That space follows the Mongolian digits. Since the Mongolian digits are a complete set, there is no need to reserve this space. This proposal suggests using the following code point assignments:

 181A
 181B
 181C
 181D
 181E

Names

181A		MONGOLIAN ORNAMENTED BIRGA → 0F04 tibetan mark initial yig mgo mdun ma
181B		MONGOLIAN ROTATED BIRGA
181C		MONGOLIAN DOUBLE BIRGA
181D		MONGOLIAN TRIPLE BIRGA
181E		MONGOLIAN SWIRL BIRGA

Properties:

181A;MONGOLIAN ORNAMENTED BIRGA;Po;0;ON;;;;;N;;;;;
181B;MONGOLIAN ROTATED BIRGA;Po;0;ON;;;;;N;;;;;
181C;MONGOLIAN DOUBLE BIRGA;Po;0;ON;;;;;N;;;;;
181D;MONGOLIAN TRIPLE BIRGA;Po;0;ON;;;;;N;;;;;
181E;MONGOLIAN SWIRL BIRGA;Po;0;ON;;;;;N;;;;;

Collation

The new types of Mongolian birga should sort immediately after U+1800 MONGOLIAN BIRGA.

Appendix

Ornamented birga

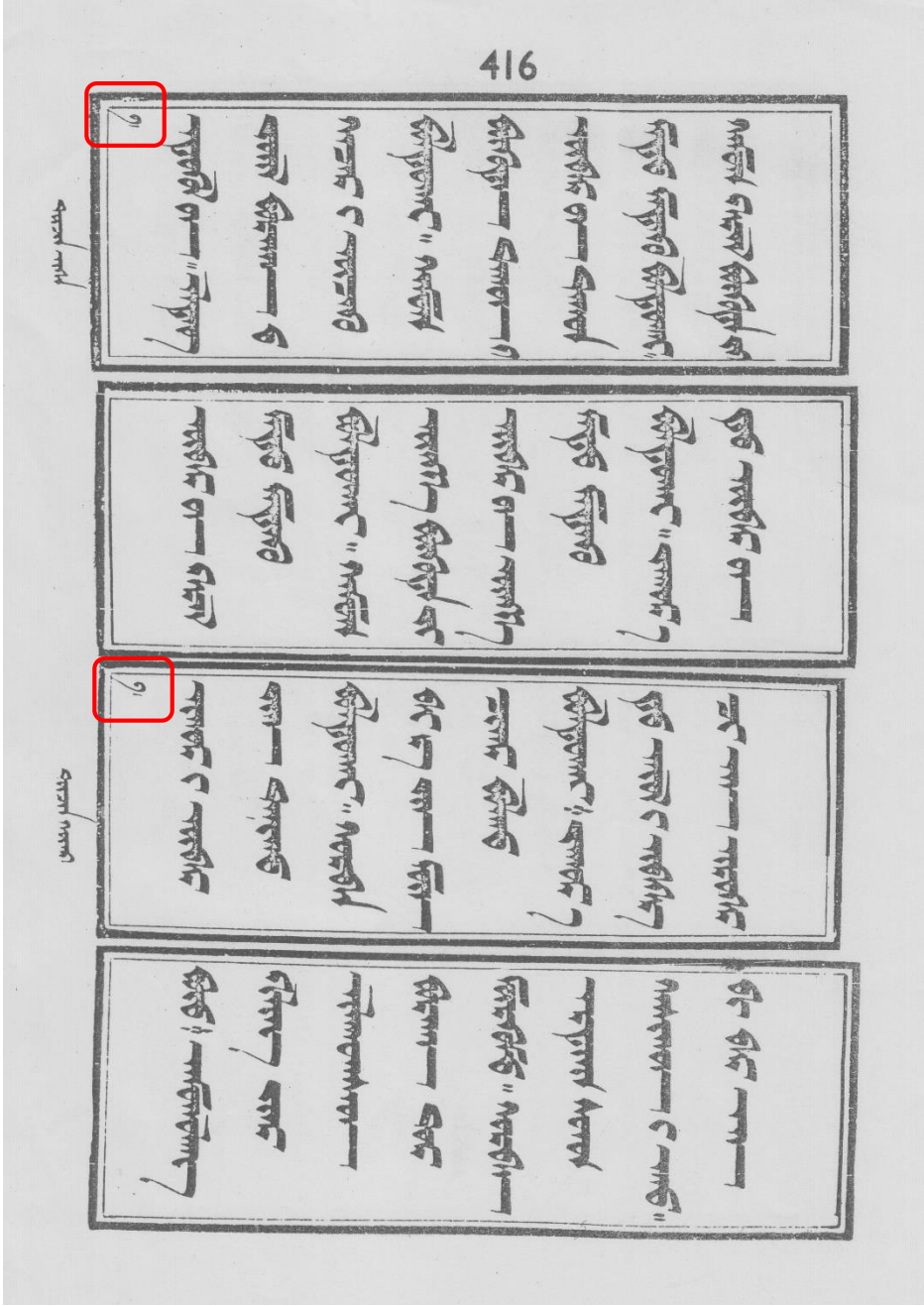


Figure 3. Evidence for the ornamental birga

Rotated birga

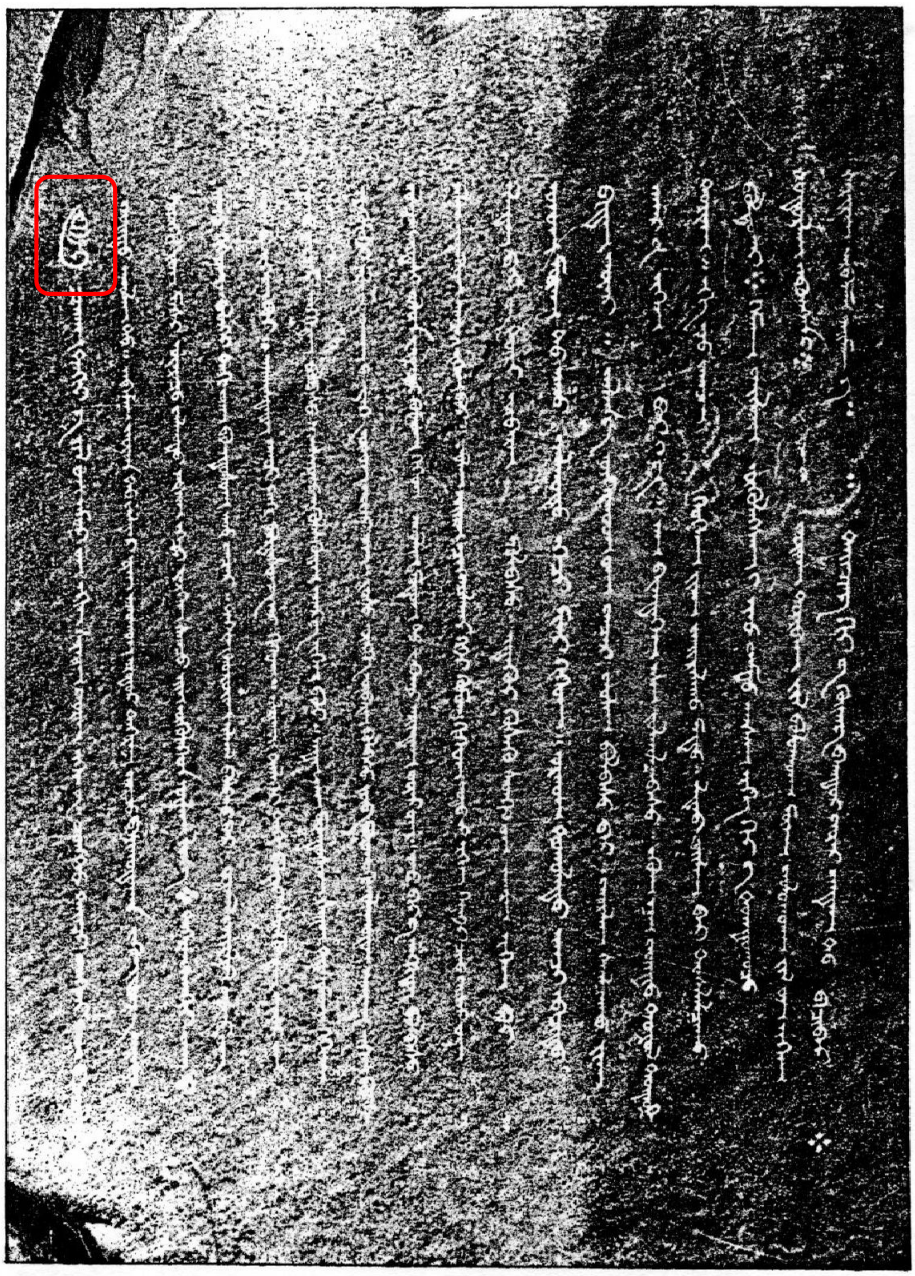


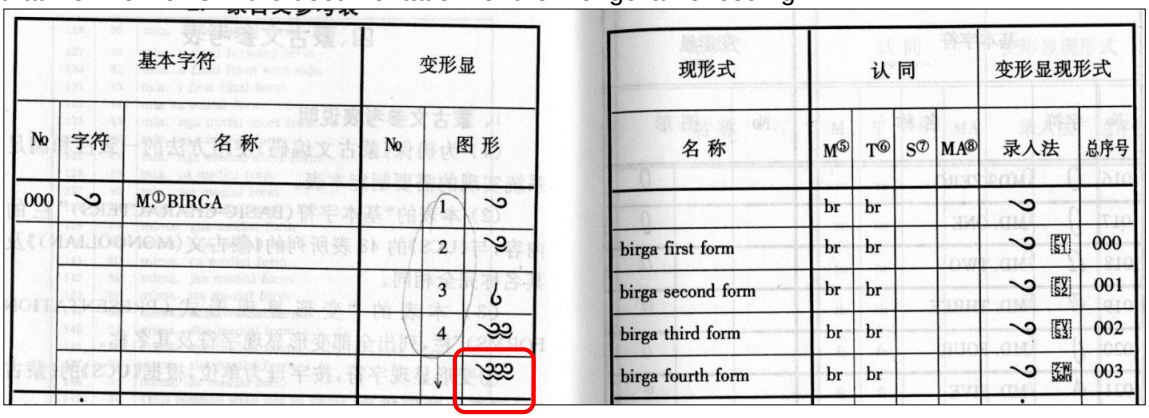
Figure 4. Evidence for the rotated birga at top of the first column from the left.

Double birga

Figure 5. Evidence for the double birga at the top left of the folio,

Triple birga

The authors have not been able to locate an example of the triple birga in a Mongolian document. The form may be very rare, and so it is a matter of locating the right document. The only example of this type that we know of is in the documentation for the Mongolian encoding.



基本字符			变形显	
№	字符	名 称	№	图 形
000	ᢀ	M.①BIRGA	1	ᢀ
			2	ᢀ
			3	ᢀ
			4	ᢀ
			↓	ᢀ

现形式	认 同				变形显现形式	
名 称	M⑤	T⑥	S⑦	MA⑧	录入法	总序号
	br	br			ᢀ	
birga first form	br	br			ᢀ EV S1	000
birga second form	br	br			ᢀ EV S2	001
birga third form	br	br			ᢀ EV S3	002
birga fourth form	br	br			ᢀ ZWJ ᠊	003

Figure 6. Evidence for triple birga in Quejingzhabu 2000.

Swirl birga

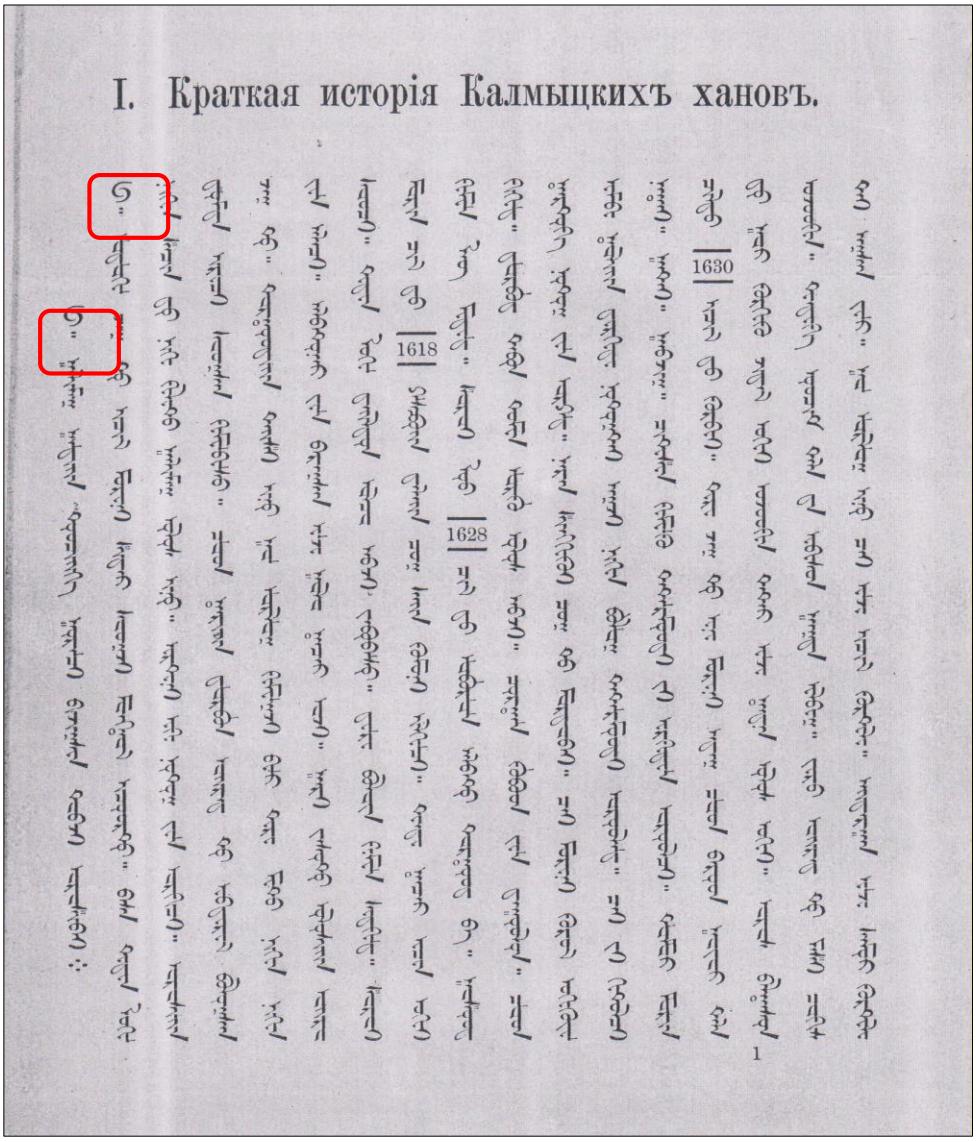


Figure 7. Page from a Kalmyk document containing evidence for the Swirl birga in the first and second columns from the left.

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.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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 five Mongolian head marks***
2. Requester's name: *Andrew Glass, Aaron Bell, Greg Eck, Andrew West*
3. Requester type (Member body/Liaison/Individual contribution): *Member*
4. Submission date: *February 6, 2014*
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):
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *Yes*
      Name of the existing block: *Mongolian*
2. Number of characters in proposal: *5*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary A | B.1-Specialized (small collection) | 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?
   b. Are the character shapes attached in a legible form suitable for review?
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
   Microsoft
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   Microsoft
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.

[1] Form number: N4102-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? *No*
   If YES explain
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? *Prof. Quejingzhabu*
   If YES, available relevant documents: *No*
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *3.3 M*
   Reference: *http://www.ethnologue.com/language/mvf*
4. The context of use for the proposed characters (type of use; common or rare) *Common & Rare*
   Reference:
5. Are the proposed characters in current use by the user community? *Yes*
   If YES, where? Reference:
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *Yes*
   If YES, is a rationale provided? *Yes*
   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?
   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? *Yes*
   If YES, is a rationale for its inclusion provided? *Yes*
   If YES, reference: *L2/14-030 - Encoding Mongolian head letters*
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:
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: