

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document

Title: Tangut Character Additions and Glyph Corrections

Source: Andrew West, Viacheslav Zaytsev, Sun Bojun (18th Sub-project of the Project Chinese Characters Repertoire), and Jerry You

Status: Individual Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Replaces: N4850 (L2/17-313); N4851 (L2/17-314)

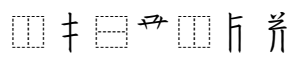
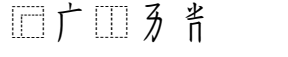
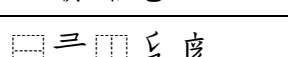
Date: 2017-09-22

This document consolidates and replaces N4850 (L2/17-313) “Proposal to encode six additional Tangut ideographs” and N4851 (L2/17-314) “Glyph Corrections for 31 Tangut ideographs and one Tangut component”.

1. Character Additions

We propose to encode six additional Tangut ideographs listed in Table 1 that have recently been identified in two Tangut sources. These characters are discussed in more detail in WG2 N4851 (L2/17-314).

Table 1: Proposed Tangut Characters

C/P	Glyph	IDS	R/S	Source	Evidence
187F2			42.15	<i>Hevajra</i> col. 223	Fig. 1 Fig. 3
187F3			185.12	<i>Hevajra</i> col. 144	Fig. 2 Fig. 4 Fig. 5
187F4			73.11	IOM Tang. 27 folio 2A col. 5	Fig. 6
187F5			383.18	IOM Tang. 27 folio 22A col. 3	Fig. 7
187F6			79.14	IOM Tang. 27 folio 29B col. 3	Fig. 8
187F7			79.19	IOM Tang. 27 folios 36A through 38A	Figs. 9–13

1.1 Hevajra's Wish-Fulfilling Gem of Abhisamaya

The two characters proposed for encoding at U+187F2 and U+187F3 are attested in a recently discovered Yuan dynasty Tangut manuscript that was sold at auction in Beijing in June 2017. This manuscript, entitled *ngiwei* ᠩᡤᡳᠪᡳ ᠨᡳᠩ ᠮᡳᠣ ᠶᡳ ᠲᡰᡢ ᠫᡵᡳ ᠯᡳᠭᡳ 祿禱席鋒辨纒 祿禱𑖦𑖧 (Chinese: *Xǐ jīngāng xiànzhèng rúyì bǎo* 喜金剛現證如意寶; English: *Hevajra's Wish-Fulfilling Gem of Abhisamaya*) [abbreviated as *Hevajra*], has been identified as a previously unknown Tangut translation of the first part of a Tibetan text relating to the Buddhist deity Hevajra which was made in 1258 by Drogön Chögyal Phagpa (1235–1280), fifth patriarch of the Sakya school of Tibetan Buddhism and Imperial Preceptor under Kublai Khan.

This text has been studied by Li Ruoyu 李若愚 in an academic article (September 2016) and in his PhD dissertation (April 2017). Li has identified two new characters (𑖀 and 𑖁) used in the *Hevajra* text as abbreviations for the Sanskrit word *vajra*. These two characters are both related to the character 𑖂 used as an abbreviation for *vajra* which was proposed for encoding at U+187F0 in WG2 N4724.

The character 𢇛 is constructed from the left side of 𢇛 (*mba* = *va-*), the top and left of 𢇛 (*ndzi* = *-j-*), and the left side of 𢇛 (*riq̣a* = *-ra*), the three parts together transcribing *vajra*. It differs from 𢇛 in the choice of component on the left side representing the *va-* sound (the left side of 𢇛 is derived from the left side of 𢇛 *mba*).

The character 𑖦 represents *vajra* in the name Hevajra, and has the same right side components as 𑖦 and 𑖦, but the expected left side component is absent. It is possible that 𑖦 is a scribal error for 𑖦, but we propose it for encoding as it is required by scholars studying this important Buddhist text.

232 231 230 229 228 227 226 225 224 223

Column 223: *om vajrakartari hevajrāya hūṃ hūṃ hūṃ phaṭ svāhā*

[illegible]

146	145	144	143	142	141	140
拜啟	報龍	新	辨	三	報	報

[illegible][illegible]

om vajra karta ri he vajrā ya hūm hūm hūm phata svāhā/ nye ba'i snying po'o

“om vajra^[2] karta ri he vajrā ya hūm hūm hūm phat svāhā”,根本心咒^[3]也。

[1] 心咒，西夏文作“𐽀𐽂𐽄”，字面意思是“心真”，对应藏文 snying po（精华；藏；要旨），在佛书中常译作“心咒”。《甘露泉》作“心咒”。

[2] vajra, 西夏文作“𐵇𐵓”, 仔细分析字形我们发现该俗体字是西夏文“𐵇𐵓𐵇𐵓”(vajra)的省写, 即是“𐵇”字的左半部加上“𐵓”字的上半部和下半部的左边, 另外加上“𐵇”字的左半边构成的俗体字, 用来对音 vajra 一词。

Fig. 4: Li Ruoyu 2017 p. 73: Transcription of Tangut text

“𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺” 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 入思，𐰇𐰺 印结，“ve tā lī pandha bam” 谓咒以自于入。次复 ghasmarī 鸣钟持𐰇𐰺，𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺，𐰇𐰺 𐰇𐰺，
“𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺” 𐰇𐰺 多发，鸣钟声一切于至思，鸣钟印结，“om bha ga vām ma hā kā ru ka he vajra sa

Fig. 5: Li Ruoyu 2017 p. 75 Note 10 for p. 74

[10] vajra, 西夏文作“𐰇𐰺”。该字不见于西夏字书，系译经人临时自造的合体俗字，实为“𐰇𐰺”（jra）二字的省写，以“𐰇”字的上半部和下半部的左边，另外加上“𐰇”字的左半边构成。

1.2 IOM Tang. 27

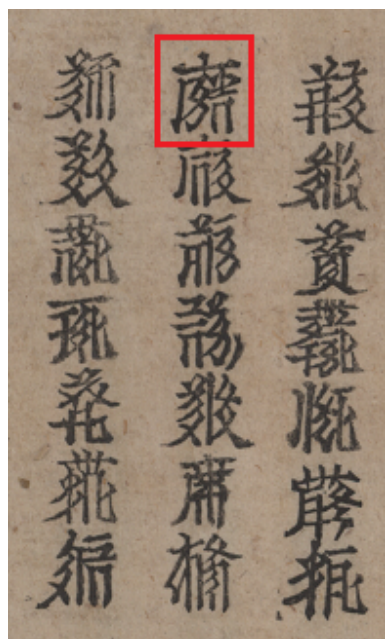
The four characters proposed for encoding at U+187F4 through U+187F7 are attested in a Tangut collection of the writings of three generations of White Cloud Sect masters entitled *sə ʃiei ndźwɪ swew ngwu ʃiow ɹiwə* 𐰇𐰺 𐰇𐰺 𐰇𐰺 𐰇𐰺 (Chinese: *Sānshìshǔ míngyán jíwén* 三世屬明言集文; English: *Collected Writings of the Shining Speech of Three Generations*) that was printed using moveable type during the Western Xia period. The only extant copy of this text is held at the Institute of Oriental Manuscripts [IOM] of the Russian Academy of Sciences (Saint Petersburg, Russia), pressmark Tang. 27.

This is an important text for the study of Buddhism under the Western Xia, and has been studied by Prof. Arakawa Shintarō (2001), Prof. Evgeny Ivanovich Kychanov (2004), Prof. Sun Bojun 孫伯君 (2011), and other scholars. Sun Bojun recently noted that there are four unencoded Tangut characters in the text, which should be encoded.

The character 𐰇𐰺 occurs twenty times on folios 36A through 38A, in a poem entitled *kai tsə tsʃia lə ndʒe* 𐰇𐰺 𐰇𐰺 𐰇𐰺 “Pain of desire and *rūpa* [material objects] hindering the way” (see Fig. 9) and in a poem entitled *mje ngje ndziwo lie ʃiwə* 𐰇𐰺 𐰇𐰺 𐰇𐰺 “Worrying that fame and fortune destroys people” (see Fig. 10). These two poems have been transcribed and translated in Arakawa 2001 (see Fig. 11 and Fig. 13) where the character 𐰇𐰺 is translated into Japanese as *indō* 引導, used in the Chinese sense of “to guide, to lead” or “guidance”.

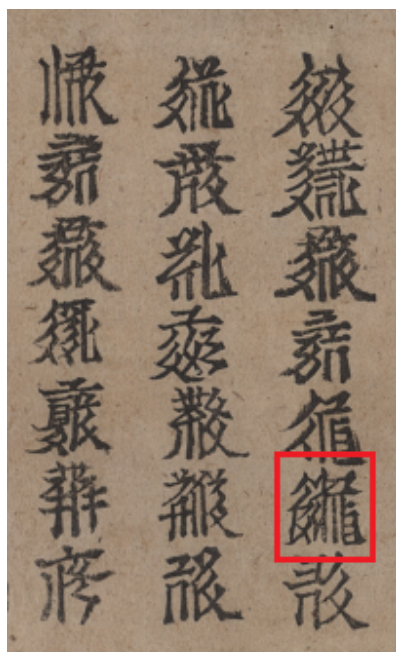
The other three characters (𐰇𐰺, 𐰇𐰺, and 𐰇𐰺) each occur only once in the text, and we do not know of any published studies that discuss them.

Fig. 6: Tang. 27 folio 2A
cols. 4-6



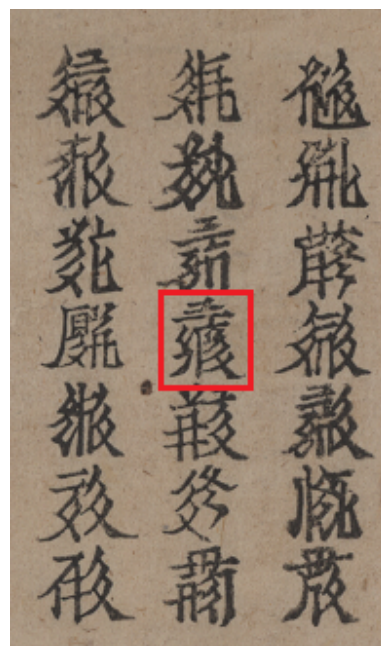
齋

Fig. 7: Tang. 27 folio 22A
cols. 3-5



齋

Fig. 8: Tang. 27 folio 29B
cols. 2-4



齋

Fig. 9: Tang. 27 folio 36A cols.
1-3: 辯

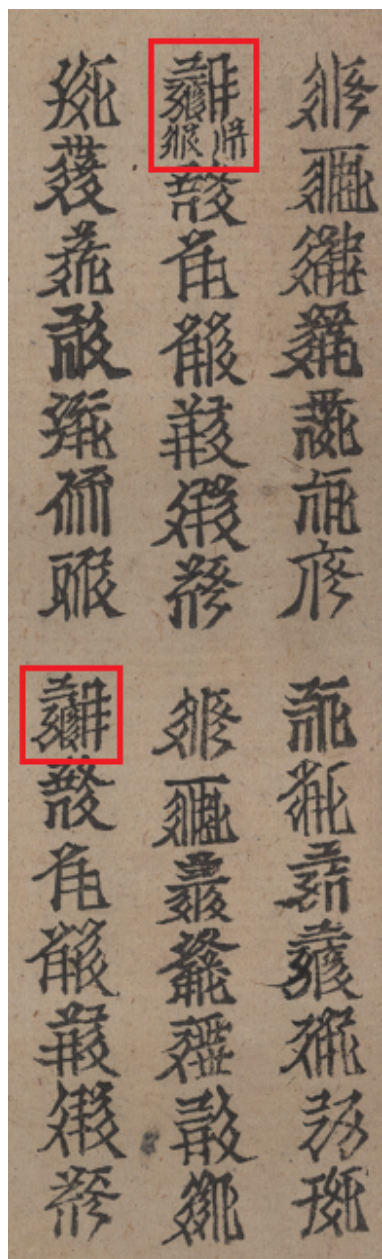


Fig. 10: Tang. 27 folio 37A
cols. 3-5: 辯

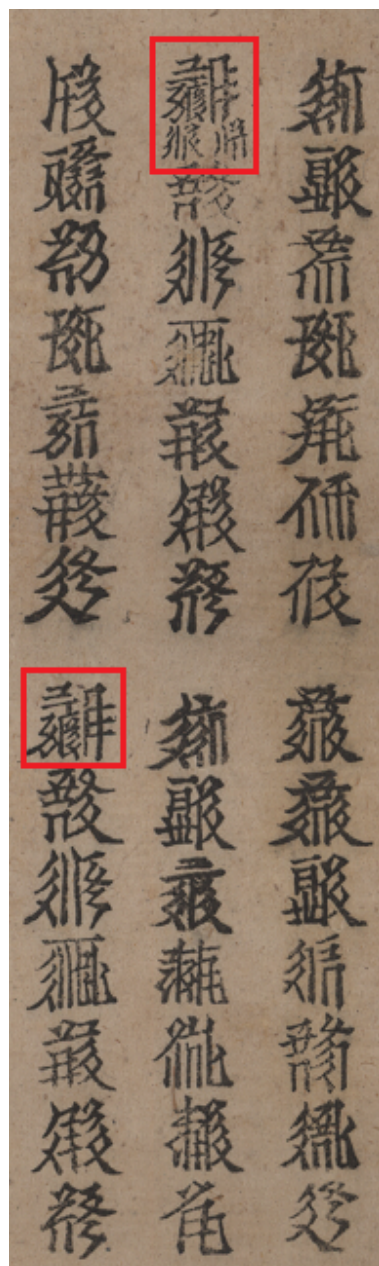


Fig. 11: Arakawa 2001 page 218: Transcription and translation of folio 36a cols. 1-2

36a-1	²ngo ²de:' ¹ni: ¹phI ¹ca: ¹jenq ²myeq'2 悲痛なるかな、出家し⁷⁰ 道を行う者 2ngo ²de:' ¹ni: ¹phI ¹ca: ¹jenq ²myeq'2	²myI' ²a ¹e: ¹chi:' ¹khwi: ²'aq ²syu 境に自らを引き出し、狂犬⁷¹ の如し 2myI' ²a ¹e: ¹chi:' ¹khwi: ²'aq ²syu
36a-2	²dze:' ²ryeq'2 ²gyu ²wa ²ryeq'2 ¹wi: 引導の⁷³ [毀れ酔う⁷⁴] 教え、その労苦は何を願うのか 2dze:' ²ryeq'2 ²gyu ²wa ²ryeq'2 ¹wi:	²ngo ²de:' ¹chwo" ¹denq ²si: ¹jyu ?lhe? 悲痛なるかな、或いは（同じ）部族 2ngo ²de:' ¹chwo" ¹denq ²si: ¹jyu ?lhe?

Fig. 12: Arakawa 2001 page 218 footnotes 70-73

- ⁷⁰ 西夏語の語順・意味では「家を一捨てる」となっている。
- ⁷¹ 西夏語の語順では「犬ー狂った」と形容詞の後置修飾である。
- ⁷² 声調・声母・韻母不明であるので推定音は ??? で表記せざるを得なかった。
- ⁷³ この字形は各種西夏語字書『文海』、『同音』などに確認できない。意味は前の行（36a-1）に現れる字形  「引き出す」から推定した。
- ⁷⁴ 字形が確認しづらいが  「酔う」（『夏漢』3547, p. 660）と判読できる。

Fig. 13: Arakawa 2001 page 218: Transcription and translation of folio 37a cols. 2-4

「名利人滅憂（名利が人を滅ぼすことを憂う）」	
37a-2	²me:' ¹genq ²dzwo: ²le:' ¹syI' 名利が人を滅ぼすことを憂う ²me:' ¹genq ²dzwo: ²le:' ¹syI'
37a-3	²me:' ¹genq ¹zyi ²syu ¹tha: ¹e: ²lyenq ¹ga ¹ga ¹genq ¹me: ²pon ²ror ²de: 名利は響く如く、その甘いこと 空空⁸⁵ であり、利はなく、自ら廻っている ²me:' ¹genq ¹zyi ²syu ¹tha: ¹e: ²lyenq ¹ga ¹ga ¹genq ¹me: ²pon ²ror ²de:
37a-4	²dze:' ²ngo ²de:' ²wa ²ryeq'2 ¹wi: ²me:' ¹genq ¹dzu ¹byu ¹nwi ¹li:q' ²ryeq'2 引導の[毀れ酔う⁸⁶] 教え、その悲痛は何を願うのか 名利は楽しみに随い、（あとで）知 2dze:' ²ngo ²de:' ²wa ²ryeq'2 ¹wi: ²me:' ¹genq ¹dzu ¹byu ¹nwi ¹li:q' ²ryeq'2

2. *Glyph Corrections*

Document WG2 N4588 (2014-08-12, last revision R2 2014-10-01) identified 76 Tangut characters with incorrectly written glyphs in the primary source for the Tangut repertoire (Li Fanwen's 2008 *Tangut-Chinese Dictionary*) whilst under ballot, and these corrections were incorporated into the published repertoire of the Tangut block (ISO/IEC 10646:2014 Amd. 2 and Unicode 10.0).

Document WG2 N4723 (2016-04-21) identified an additional three Tangut glyph corrections in the published Tangut repertoire.

Document WG2 N4850 (L2/17-313) identifies a further thirty-one Tangut ideographs that are incorrectly written in the ISO/IEC 10646 and Unicode code charts, as well as one Tangut component. Eight of these are individual corrections, and twenty-three relate to one component (U+18885 𐰇 Tangut Component-134) which has an incorrect shape. Two characters (U+17378 and U+17D36) have a revised stroke count, and one character (U+18488) has a revised radical (Radical 458 𐰇 to Radical 325 𐰇). The required glyph corrections are summarized in Table 2 below. See WG2 N4850 (L2/17-313) for detailed evidence for the correct glyph forms.

Table 2. Summary of Glyph Corrections

Character	Current Glyph	Proposed New Glyph	Old R/S	New R/S
U+17159	𐰇	𐰇	17.10	17.10
U+1732E	𐰇	𐰇	37.10	37.10
U+17336	𐰇	𐰇	41.8	41.8
U+17378	𐰇	𐰇	47.10	47.9
U+173E1	𐰇	𐰇	68.11	68.11
U+173EB	𐰇	𐰇	68.12	68.12
U+173EC	𐰇	𐰇	68.12	68.12
U+173F4	𐰇	𐰇	68.13	68.13
U+173FB	𐰇	𐰇	68.15	68.15

U+173FF	𐤏	𐤏	68.16	68.16
U+17403	𐤐	𐤐	68.20	68.20
U+1756E	𐤑	𐤑	79.14	79.14
U+176E4	𐤒	𐤒	106.13	106.13
U+1788A	𐤓	𐤓	134.9	134.9
U+1788B	𐤔	𐤔	134.11	134.11
U+1788C	𐤕	𐤕	134.13	134.13
U+1788D	𐤖	𐤖	134.14	134.14
U+17B16	𐤗	𐤗	145.12	145.12
U+17B17	𐤘	𐤘	145.15	145.15
U+17B5B	𐤙	𐤙	155.14	155.14
U+17D36	𐤚	𐤚	206.12	206.13
U+17F8B	𐤛	𐤛	267.9	267.9
U+17FBB	𐤜	𐤜	267.13	267.13
U+181B2	𐤝	𐤝	328.11	328.11
U+181C0	𐤞	𐤞	328.13	328.13
U+1820D	𐤟	𐤟	348.10	348.10
U+1823D	𐤠	𐤠	369.14	369.14
U+183B3	𐤡	𐤡	436.10	436.10

U+1841D	𐞑	𐞑	436.14	436.14
U+18488	𐞑	𐞑	458.13	325.13
U+187C6	𐞑	𐞑	729.15	729.15
U+18885	𐞑	𐞑	4	4

The following three entries in TangutSrc.txt (ISO/IEC 10646) and TangutSources.txt (Unicode) need to be modified as shown below:

Code Point	Field	Current Value	Suggested Value
U+17378	kRSTUnicode	47.10	47.9
U+17D36	kRSTUnicode	206.12	206.13
U+18488	kRSTUnicode	458.13	325.13

3. Bibliography

- Arakawa 2001.** Arakawa Shintarō 荒川慎太郎. “Seikashi no kyakuin ni mirareru inbo ni tsuite: ‘Sanze zoku meigen shū bun’ shoshū Seikagoshi” 西夏詩の脚韻に見られる韻母について: 『三世属明言集文』 所収西夏語詩 [About the rhymes in Tangut verses: Reanalysis of Tangut rhyming poetry in *San shi shu ming yan ji wen*]; *Kyōto daigaku gengogaku kenkyū* 京都大学言語学研究 = *Kyoto University Linguistic Research* vol. 20 (2001): 195–224.
- Kychanov 2004.** Кычанов, Е. И. (E. I. Kychanov). “Несколько предварительных замечаний по поводу тангутского текста «Собрание слов, передаваемых от одного к другому в трёх поколениях»” [Some Preliminary Remarks about the Tangut Text *The Collection of Words Transmitted through Three Generations from One Person to Another*]; *Письменные памятники Востока* = *Written Monuments of the Orient* no. 1 (2004, Spring–Summer): 147–159.
- Li Ruoyu 2016.** Lǐ Ruòyú 李若愚. “‘Xǐ jīngāng xiànzhèng rúyì bǎo’: Yuán dìshī Bāsībā zhùzuò de Xīxià yìběn” 《喜金刚现证如意宝》: 元帝师八思巴著作的西夏译本 [*Hevajra's Ruyibao: A Tangut Version by Yuan Imperial Preceptor Ba Siba*]; *Níngxìà Shèhuì Kēxué* 宁夏社会科学 [*Social Sciences in Ningxia*] No. 5 (September 2016): 206–212.
- Li Ruoyu 2017.** Lǐ Ruòyú 李若愚. *Xīxiàwén ‘Xǐ jīngāng xiànzhèng rúyì bǎo’ kǎoshì* 西夏文《喜金刚现证如意宝》考释 [Study of the Tangut version of Hevajra's Abhisamaya]. Ph.D. dissertation. Beijing: Zhōngguó shèhuì kēxuéyuàn yánjiūshēng-yuàn 中国社会科学院研究生院 = Graduate School of Chinese Academy of Social Sciences, 2017.
- Sun Bojun 2011.** Sūn Bójūn 孙伯君. “Yuándài báiyún zōng yìkān Xīxiàwén wénxiàn zōngkǎo” 元代白云宗译刊西夏文文献综考 [General study of Tangut documents translated and printed during the Yuan dynasty by the White Cloud sect]; *Wénxiàn Wénkuān* 文献 [Documents] No. 2 (2011): 146–157.

4. Proposal Summary Form

ISO/IEC JTC 1/SC 2/WG 2 PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹ Please fill all the sections A, B and C below. Please read Principles and Procedures Document (P & P) from http://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html . See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest Roadmaps.

A. Administrative

1. Title:	Proposal to encode six additional Tangut ideographs		
2. Requester's name:	<i>Andrew West, Viacheslav Zaytsev, Sun Bojun (18th Sub-project of the Project Chinese Characters Repertoire), and Jerry You</i>		
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>		
4. Submission date:	<i>2017-09-22</i>		
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):		NO
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:		YES
Name of the existing block:	<i>Tangut</i>	
2. Number of characters in proposal:		6
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary	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?		
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?	<i>Andrew West</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Jing Yongshi, Andrew West, Michael Everson</i>	
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)?		NO
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?	YES
If YES explain <i>Replaces WG2 N4851 (L3/17-314)</i>	
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? <i>Experts from China, Russia and USA</i>	
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	NO
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	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?	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?	NO
If YES, is a rationale for its inclusion provided?	
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:	