

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

Doc Type: Working Group Document

Title: Second revised proposal to encode four historic Latin letters for Sakha (Yakut)

Authors: Ilya Yevlampiey, Nurlan Jumagueldinov, Karl Pentzlin

Status: Expert Contribution

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

Date: 2012-04-26

Replaces: L2/12-044, WG2 N4213

*Some remarks regarding distantly similar Cyrillic letters are added at the end of the introduction, addressing a question raised during UTC #130.
 In consequence, two characters were renamed.*

1. Introduction

Sakha (ISO 639: sah, also known as Yakut) is a language spoken by 456.288 people according to the 2002 census [2], mainly in the area which today is the Sakha Republic, a federal subject of Russia located in the north east of Siberia.

Some Cyrillic orthographies devised in the 19th century did not yield a larger success in alphabetization. Based on this fact and the feeling that the former orthographies did not reflect the phonemes of the language sufficiently, the linguist Semyon Andreyevich Novgorodov (semen nəfʊrʊdap, Səməŋ Noğoruodap, 1892–1924) devised a Latin orthography using a subset of IPA, augmented by four letters for specific diphthongs which are proposed here. This orthography did use lowercase letters only (like IPA).

This orthography was introduced 1917 in the wake of the Russian revolution, and it was the official orthography of the Sakha language until 1927, when the alphabet was changed again.

Details can be found in [3].

Regarding the letter **Ǝ** proposed as U+AB60 LATIN SMALL LETTER IOTIFIED E, it is noted that this letter somewhat resembles the Cyrillic letter **Ѧ** U+0465 CYRILLIC SMALL LETTER IOTIFIED E. In fact, it is possible that Novgorodov was inspired by the "iotified" Cyrillic letters when devising the alphabet. However, the Latin IOTIFIED E resembles a "iotified Latin e", while U+0465 resembled a "iotified Cyrillic e" despite of its name based on its use rather than on its glyphic appearance. The following lines show the Cyrillic letters U+0435, U+0454, U+0465 using some common fonts:

е **Ѧ** **Ѧ** — Arial

е **Ѧ** **Ѧ** — Times New Roman

е **Ѧ** **Ѧ** — Roman Cyrillic (<http://kodeks.uni-bamberg.de/AKSL/Schrift/RomanCyrillicStd.htm>)

This shows that the IOTIFIED E is, if any, a "iotified Latin e", not a "iotified Cyrillic e". Thus, independent of the question whether unification of Latin and Cyrillic letters is appropriate at all, it cannot be unified with **Ѧ**, as **е** and **Ѧ** are distinct Cyrillic letters. Thus, unifying the iotified forms of these letters while the base letters are different in form and function would cause confusion.

2. Proposed Characters

Block: Latin Extended-E

Historic letters for Sakha (Yakut)

These letters were used 1917–1927 in the official IPA-based Latin orthography of that era.

Ъ U+AB60 LATIN SMALL LETTER SAKHA IOTIFIED A
→ A657 cyrillic small letter iotified a

Ё U+AB61 LATIN SMALL LETTER IOTIFIED E
→ 0465 cyrillic small letter iotified e

œ U+AB62 LATIN SMALL LETTER OPEN OE
→ 0254 latin small letter open o

Ƶ U+AB63 LATIN SMALL LETTER UO

Notes on the character names

The names of the characters proposed here as Latin small letters "IOTIFIED E", "OPEN OE", and "UO" are derived from their shape. It is not unlikely that such letters are found when still unencoded repertoires of other historical orthographies or of dialectological transcriptions are re-researched (see e.g. some examples of similar letters in recent proposals for German dialectology). Thus, names which describe their obvious composition are proposed here.

As the "LATIN SMALL LETTER SAKHA IOTIFIED A" does not resemble glyphically an "a" (neither a Latin or modern Cyrillic nor a Church Slavonic one) and was devised for Sakha, it is named according to its use in the Sakha orthography which it is part of.

Properties:

AB60;LATIN SMALL LETTER SAKHA IOTIFIED A;Ll;0;L;;;;N;;;;;
AB61;LATIN SMALL LETTER IOTIFIED E;Ll;0;L;;;;N;;;;;
AB62;LATIN SMALL LETTER OPEN OE;Ll;0;L;;;;N;;;;;
AB63;LATIN SMALL LETTER UO;Ll;0;L;;;;N;;;;;

3. References

- [1] Юшманов, Николай Владимирович: Определитель Языков. Москва/Ленинград 1941
- [2] http://ru.wikipedia.org/wiki/%D0%AF%D0%BA%D1%83%D1%82%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA (retrieved 2011-09-10)
- [3] <http://ilin-yakutsk.narod.ru/1999-12/50-2.htm> (retrieved 2011-09-05; in Russian)
- [4] Новгородов С.А. Первые шаги якутской письменности. Статьи и письма. - М.: Наука, 1977.
- [5] saqalw: suruk bicik, Yakutsk 1917
- [6] nəfurɯdap, semen (Novgorodov, S. A.); et al., suruk bicik. mɔskuba (Moscow) 1923
- [7] Будищев, И. П.; Русско-якутский словарь. Якутск 1926

4. Examples and Figures

Fig. 1: [1], p. 28/29. This table shows a compilation of special Latin letters used in orthographies of languages of the Soviet Union before the publication of [1] (1941). The Yakut letters proposed here are marked. (The table shows several other letters not contained in Unicode 6.0, which are not subject of this proposal.)

28

Таблица II
Латиницы языков СССР

В алфавитный список не включены первичные латиницы, замененные впоследствии другими: *абхазская* (особенности — ə, ħ, ɪ, j, φ, ш, ʒ, ʋ; буквы с точками, как ȳ, ȳ̇, кружками, как ꞑ, ꞑ̇, зубчиками, как ꞑ̇, ꞑ̇̇, черточкой, как ꞑ̇̇̇); *кабардинская* (особенности — ȳ, ȳ̇, ȳ̇̇, ȳ̇̇̇, ȳ̇̇̇̇; сочетания типа fh, Xu, qhu); *якутская* (особенности — ȳ, ȳ̇, ȳ̇̇, ȳ̇̇̇, ȳ̇̇̇̇, ȳ̇̇̇̇̇, ȳ̇̇̇̇̇̇, ȳ̇̇̇̇̇̇̇; двоеточие как знак долготы, напр. a:, s:; прописных букв и знаков препинания нет).

Ингушская, осетинская и чеченская латиницы употребляли много сочетаний согласных с -h взамен особых букв кавказских языков. Языки народов Севера пользовались значками: — над гласными (*удэйский, эвенкийский, эвенский*), — под согласными (*ителменский, ненецкий, нивхский, саамский, селькупский*) и ' после согласных (*ителменский, нивхский, ныманский, эскимосский*).

ä вепс., ижор., ингуш., карел.	ı хакасский	ï чеченский
а адыгейский	ɹ тувинский	ï см. ï, ï̇
æ ингуш., кетский, осет., селькупский, удэйский	ķ кавказские, курдский	ï казах., каракалпакский
č ингуш., осет., чеченский	ķ, в адыгейский	ï таджикский
ç курдский, саамский	ġ адыг., кабардинский	ı ингуш., чеченский
ç [большинство ¹]	ı авар., ительм., хант., эским.	ı дагестанские
đ абхаз., адыг., башк., шуги.	ı абаз., авар., адыг., кабард.	ı дагест., шуганский
đ адыг., белуджский	ı хантыйский	ı кабард.
ə [большинство ²]	ң ингуш., тюркские, ⁴ чеченский	ı ингуш., осет., чеченский
ə курдский	ң языки народов Севера	ı кавказские
ı абаз., абхаз., адыг., кабард.	ö вепс., ижор., карел., чеченский	ı дунганский
ı абазинский, абхазский	о [большинство ⁵]	ı [большинство ⁷]
ı адыгейский	р адыгейский	ı даргинский, табасаранский
ı [большинство ³]	р абаз., абхаз., кабард.	ı кавказские, нанайский, саам., удэйск., эвенк., эвенск.
ħ абаз., адыг., дагест., курд., татск.	р дагест. (кроме авар.), курд.	ı коми
ı дагест., татск., шуганский	р адыгейский	ı саамский, шуганский
ı абаз., кабардинский	q абаз., абхаз., дагест., кабард.	ı абазинский, абхазский
ı абаз., адыг., кабардинский	ғ нивхский	ı [большинство ⁸]
ı ингуш., кабардинский, кетский, маньс., ненецкий, нивх., осет.	š адыг., ингуш., чеченский	ı адыгейский
ı адыгейский	š [большинство ⁶]	ı абаз., абхаз., кабард.
ı абазинский, абхазский	s дагест., маньс., саам., хант.	ı абаз., абхаз., адыгейский
ı таджикский	ı абазинский, абхазский	ı адыгейский
	ı белудж., кавк., курд., сир.	ı абаз., абхаз., кабард.
	ı адыг., башк., шуганский	ı вепс., таджикский, тюркские (реже другие).
	ı, ı̇ абхазский	

¹ Вепсский, ижорский, иранские, кавказские, карельский, сирийский, тюркские (кроме башк., ногайск., тувин.).
² Иранские (кроме таджик.), кавказские (кроме абаз., авар., кабард.), северные (кроме ительм., маньс., нымл.), сирийский, тюркские (кроме карач.-балк., киргиз, крым.-тат., кумык, ногайск., ойрот, тувин., узбек., хакас., шор., якут.).
³ Иранские, кавказские (кроме адыг.), тюркские (кроме ойрот, туркмен, хакас., шор.).
⁴ Кроме азербайджанского.
⁵ Тюркские (кроме узбек.) и еще: бурятский, калмыцкий, лакский, лезгинский, селькупский, удэйский, хантыйский, шуганский.
⁶ Вепсский, ижорский, иранские, кавказские (кроме адыг.), карельский, сирийский, тюркские (кроме казах.).
⁷ Вепсский, ижорский, иранские (кроме татск.), кавказские (кроме адыг.), карельский, кетский, саамский, сирийский, тюркские (кроме казах., каракалп.).
⁸ Ижорский, курдский, северные (кроме маньс., нанайск., нивх., удэйск., эвенк.), сирийский, тюркские (кроме узбек., уйгур.).

29

Fig. 2: Title page and alphabet page from [5], pictures taken from {4}.

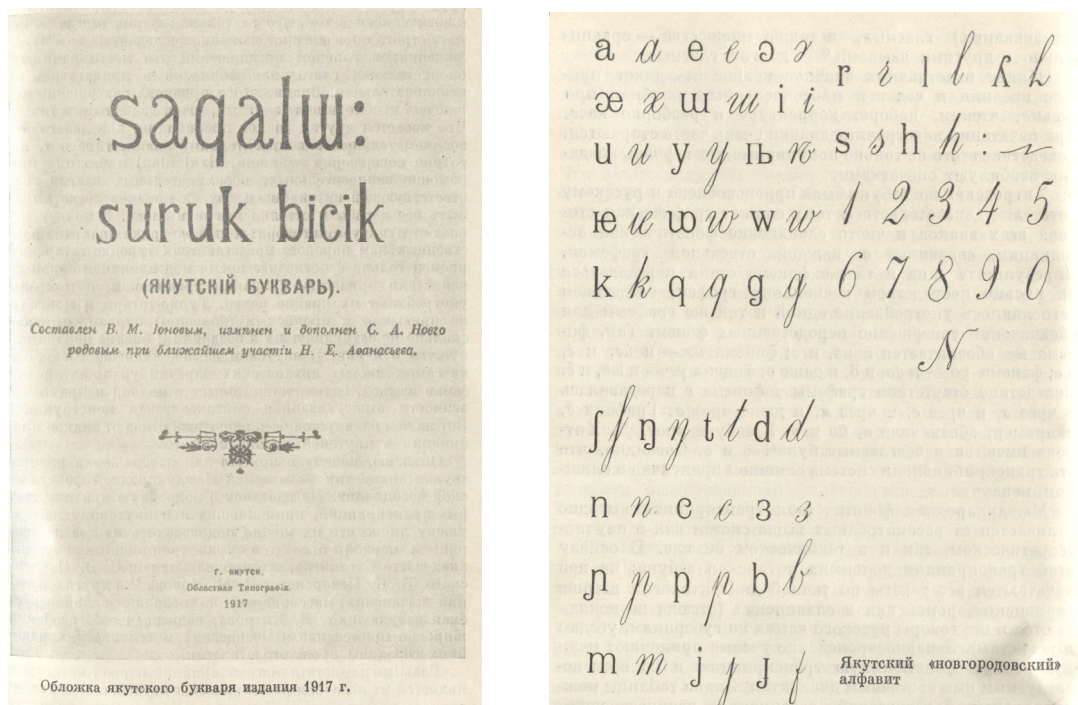
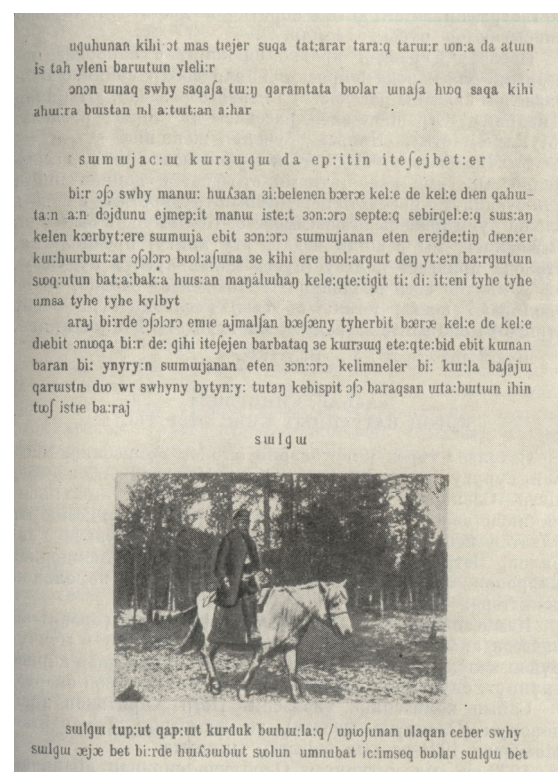


Fig. 3: Title page from {6}.



Fig. 5: Sample pages from several books using the Sakha orthography containing the proposed letters..



ije u:hun atučluur artıla

J

ofo ijetin y:tyn ispetitten xler

f

tır 300 bulbut astarun, bulbut bultarun tıj
cugahuna:tu atučhuttarga atučlu:llar. atučhūt
tus tuspa 3000 3000 qaja bafarar albunna:n
twkejden ulan bejetin bassun qagatar bular.
atučhūt bassun qagatıq keriete, beje bassun
qagatar 3000 buluq ete. 3000 qaja ije u:ha ije
u:hunan qalbahen artıl 3000, bular talar asta-
rum willerin, bultarun altarun keppereti:pke 30-
30 du:, kəmyske tıjeın 30 du:, 30 30 natu-
gar atučlu:n tuhanıq etiler. tır kihite bular
talar aha wle elbeq—arun: y:t, qajaq, swıhtyn
ete tırıte, baha tıjafa. tır kihite bultu:r bulta
alta da: elbeq—tır bultuttan tabata tıjafa, kə-
tər kunnatta:qtan kuha, qa:ha, u: bultuttan ejin
ejin tıjnaru qajula:ıun ara:ha... tıj da: kınma
suq... artıl bu manı barutun tımen qalbahen
tupsaran 3000, barar sıriger keppereti:pterge
du: kemısterge du: atučlu: turdajuna, bu ba-
rıtı ulaqan kəzyske turu ete. bu mantan ku-
ra 300 ulaqannuk tuhanı etiler.

qamu:nafa qalbahen 3000 putun 3000 but
3000 3000 suq buluqta:q.

bu:kubalar.

a _a	e _e	ə _ə	æ _æ
ı _ı	ı _ı	w _w	w _w
u _u	i _i	u _u	y _y
k _k	q _q	g _g	ı _ı
ı _ı	t _t	d _d	n _n
c _c	3 ₃	ı _ı	p _p
b _b	j _j	j _j	m _m
r _r	l _l	ı _ı	s _s
h _h	:		

Fig. 6: Two sample pages from [7].

A.	
Автономия	ap̄tan̄m̄ūja
Август	at̄ur̄z̄aq̄ ūja
Агент	eḡient
Алмаз	al̄amā:s̄ tā:s̄
Амбар	am̄pā:r̄
Амбулатория	am̄bul̄at̄īr̄ūja
Америка	em̄er̄ike
Амнистия	em̄nī:st̄ije
Анастасия	nāstā:es̄ūja
Андрей	ǣnd̄ȳrej̄, ǣnd̄rȳ:s̄
Анна	ā:nā, ā:nūs̄, ā:nūska
Англия	ēnḡil̄ije
Алексей	æl̄eks̄ǣj̄
Александр	sā:nd̄rā, ǣl̄eks̄ǣnd̄ȳre
Александра	æl̄eks̄ǣnn̄ǣr̄ǣ
Аркан	ə̄fūr̄
Апрель	mū:s̄ ūstar
Апелляция	ēppēlē:ss̄ije
Арестант	q̄ā:j̄ūt̄lā:q̄
Арестовать	q̄ā:j̄ūr̄:gā ūḡūq̄q̄ā, q̄ā:j̄ūq̄q̄ā
Аршин	ass̄ū:n̄
Армия	ā:r̄m̄ūja
Аэродром	k̄æt̄ǣr̄ ā:l̄ tȳher̄ t̄ə̄qt̄ur̄ sīre
Аэроплан	k̄æt̄ǣr̄ ā:l̄
Б.	
Бабушка	ēbe
Баба	zāq̄tar̄

84	
Большие шали есть?	ulaqan sa:l bula:t bar du?
Покажите одеяла?	suorjanna kærdær?
Шапка есть?	bergehe bar du?
Картуз есть?	qottus bar du?
Покажите шарфы?	sarta kærdær ere?
Пояс есть?	kur bar du?
Машинные нитки есть?	massu:na saba bar du?
Серые нитки есть?	dwn sap bar du?
Полушубки есть?	kulgas baran son bar du?
Что стоит?	synata tæ:hænyj? (tuguj?)
Дайте кирпич чаю?	bir kippi:ce cajda ajal?
Дайте три фунта чаю байхового	ys munta ot cajda ajal?
Дайте два фунта сахара?	ikki munta sa:garda ajal?
Свесьте карамели	su:la:q qampi:ta uja:
Дайте мюшасе	lampasnj qampi:ta ajal
Рубашечные пуговицы есть?	arba:jui timefe bar du?
Пуговицы для пальто есть?	son timefe bar du?
Какие пуговицы пужны?	qajdaq timeq na:da?
Что стоит дюжина?	zuchunata qaska turar?
Сапоги есть?	sappiuku bar du?
Катанки есть?	kattunka bar du?
Войлок есть?	bu:zæq bar du?
Почем цена?	synata tæ:hænyj?
Покажите швейные машины	istener massu:nata kærdær ere?
Хорошая машина?	yeygej massu:na du?
Что стоит?	synata tæ:hænyj?
Машинные иголки есть?	massu:na ignete bar du?
Отпустите мне один фунт пороку и пять фунтов дробн	mege bir munta bu:rafu kuutta bies munta dara-mu:nata uja: ere
Все это что стоит?	bu baruta tæ:hæfæ turar?

**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:	Second revised proposal to encode four historic Latin letters for Sakha (Yakut)		
2. Requester's name:	<i>Ilya Yevlampiev, Nurlan Jumagueldinov, Karl Pentzlin</i>		
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>		
4. Submission date:	<i>2012-04-26</i>		
5. Requester's reference (if applicable):			
6. Choose one of the following:			
This is a complete proposal:			<i>Yes</i>
(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):			<i>No</i>
Proposed name of script:			
b. The proposal is for addition of character(s) to an existing block:			<i>Yes</i>
Name of the existing block:	<i>Latin Extended-E</i>		
2. Number of characters in proposal:			<i>4</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):			
A-Contemporary	B.1-Specialized (small collection)	<i>X</i>	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?			<i>Yes</i>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?			<i>Yes</i>
b. Are the character shapes attached in a legible form suitable for review?			<i>Yes</i>
5. Fonts related:			
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Karl Pentzlin; font available at http://www.pentzlin.com/sakhaletters1.ttf</i>		
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Released into the Public Domain by the author</i>		
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?			<i>Yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?			<i>Yes</i>
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)?			<i>n/a</i>

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: N3902-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)

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? <i>Linguists, Librarians</i>	
If YES, available relevant documents: <i>see text</i>	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	Yes
Reference: <i>see text</i>	
4. The context of use for the proposed characters (type of use; common or rare)	<i>Historically common</i>
Reference: <i>see text</i>	
5. Are the proposed characters in current use by the user community?	<i>Historical and scientific</i>
If YES, where? Reference: <i>see text</i>	
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?	
If YES, reference: <i>To keep them in line with similar characters</i>	
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 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:	