

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

Doc Type: Working Group Document

Title: Proposal to encode 114 Tangut components

Source: Andrew West

Status: Individual Contribution

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

Date: 2023-05-31

1. Introduction

This document proposes to encode 114 Tangut components that are used for character indexing purposes in the recently-published dictionary of Tangut words, *Xīxiàwén Cídiǎn* (*shìsú wénxiàn bùfēn*) 西夏文词典（世俗文献部分） (Beijing: Zhongguo Shehui kexue chubanshe, 2021), compiled by Hán Xiǎománg 韩小忙. This is an immense nine-volume dictionary of secular Tangut vocabulary (i.e. not covering Buddhist literature, which is anticipated to be covered in a future dictionary), with entries for 6,150 Tangut characters (including 263 variant forms) spanning some 4,872 pages. This is the most important contribution to Tangut lexicography in a generation, and there is no doubt that it will remain the standard reference work for secular Tangut vocabulary for many years to come.

The final volume of the dictionary provides indices and appendices, including two separate Tangut character indices using radical and stroke-count lookup. The first radical index (vol. 9 pp. 2–94) is a radical index based on the left side or top part of a Tangut character, as is normal for Tangut dictionaries compiled in China, and comprises 444 primary radicals numbered LR1 through LR444, and 29 supplementary radicals numbered LR+1 through LR+29. All but ten of these 473 radicals are already encoded as Tangut components, and two of the ten unencoded radicals are alternative forms of encoded components where the alternative form (given in parentheses below) is only used to classify its own corresponding ideograph:

- LR364: 𐰇 (𐰇) where the parenthetical form only indexes U+186D6 𐰇
- LR419: 𐰇 (𐰇) where the parenthetical form only indexes U+187A4 𐰇

According to the principles adopted for encoding Tangut components, radicals that are only used to classify their own corresponding ideograph are not encoded separately (see WG2 N4636 = L2/14-228 *Proposal to encode Tangut radicals in the UCS* p. 5 point D). Therefore the two alternative form radicals shown above are not proposed for encoding.

The second radical index (vol. 9 pp. 95–196) is a radical index based on the right side or bottom part of a Tangut character, as is normal for Tangut dictionaries compiled in Russia (e.g. Kychanov’s 2006 Tangut dictionary), and comprises 492 radicals numbered RR1 through RR492. 243 components are used as both left-side and right-side radicals, and 149 of the remaining right-side components are already encoded. This leaves 100 unencoded right-side components, as well as six unencoded alternative forms of encoded components. 41 of these 106 unencoded right-side components correspond to encoded ideographs, but the component is presented in different proportions (i.e. narrower) compared with the corresponding ideograph (e.g. RR479 𐞪 compared with U+17C1E 𐞪). In the original proposal to encode Tangut components (WG2 N4636 = L2/14-228), radicals which corresponded to ideographs were not excluded from separate encoding if they occurred as a radical for multiple characters, so on that basis, the 41 radicals corresponding to ideographs in Hán Xiǎománg’s dictionary are included in this proposal.

Additionally, the following six radicals in the dictionary are unifiable with existing encoded components, so are not proposed for encoding:

- [illegible]

2. Proposed characters

In total, 114 Tangut components are proposed for encoding, as listed in Table 1. Assuming that the component numbering continues after U+18AFF TANGUT COMPONENT-768, the proposed characters would be named TANGUT COMPONENT-769 through TANGUT COMPONENT-882. However, the two components proposed in WG2 N5217 could be inserted before, after, or within this sequence.

The Tangut Components block (18800..18AFF) is completely full. Therefore the 114 Tangut components proposed in this document, and the two Tangut components proposed in N5217, will need to be allocated in a new Tangut Components Supplement block, with a recommended block size of 128 code points.

In the table below, the proposed glyph form of the radical is based on the code chart form of ideographs classified under this radical, and where the actual glyph form shown in Hán Xiǎománg's dictionary not exactly the same the dictionary form is given in parentheses. An example ideograph is given for each radical, and where this is a unifiable variant of an encoded character the corresponding encoded character form is given in parentheses. The seven radicals marked with an asterisk consist of a primary and secondary radical forms, and the proposed character is an unencoded secondary radical form (e.g. RR128 is listed in the dictionary as 𠂇 (𠂇) where 𠂇 is Component 250, and 𠂇 is not encoded).

Table 1: Proposed Tangut Components

No.	Radical	Strokes	Glyph	Example Ideograph	Ideograph	Note
769	RR3	1	\	关		
770	RR4	1	、	爰		
771	RR14	2	𠂇 (𠂇)	𠂇, 𠂇 (𠂇)		1
772	RR34	3	𠂇	𠂇		
773	RR53	4	𠂇	𠂇		
774	RR73	4	𠂇 (𠂇)	𠂇		
775	RR81	4	𠂇	𠂇		
776	RR102	5	𠂇	𠂇		
777	RR128*	5	𠂇	𠂇 (𠂇)		
778	RR138	5	𠂇 (𠂇)	𠂇		
779	RR154	5	𠂇	𠂇		
780	RR160	5	𠂇	𠂇		
781	LR177	5	𠂇 (𠂇)	𠂇, 𠂇		2
782	LR190 RR185	5	𠂇	𠂇 (𠂇), 𠂇 (𠂇)		3

¹ RR14 𠂇 occurs in the two semantically-related characters U+17121 𠂇 *kji*² ‘scorched’ and U+17C51 𠂇 *tšjiw*² ‘burnt’. It looks like a cosmetic variant of Component 34 𠂇, but it is consistently written as 𠂇 in *Homophones* and *Synonyms*, and so should actually be a distinct component. The code chart glyphs for U+17121 and U+17C51 should be corrected to show this component.

² LR177 𠂇 is used both as the unencoded head component in the ideographs 𠂇 and 𠂇, and also as a left-side component corresponding to Component 323 𠂇 in the ideographs 𠂇 and 𠂇 (in this usage it is the same as RR176 which corresponds to the right side of ideographs such as 𠂇 and 𠂇).

³ LR190/RR185 𠂇 (with a dot) contrasts with LR105 𠂇 (without a dot) which is encoded as Component 198, and indexes U+17D0C 𠂇 in Hán Xiǎománg’s dictionary. In Lǐ Fànwén’s 2008 Tangut dictionary, which was the main source for the encoded Tangut repertoire, U+17D0B 𠂇 and U+180DF 𠂇 are both written with the dotless component, which are the forms inherited in the code charts. In 2014, we discussed the correct glyph forms for these and other related characters (see N4588R2 = L2/14-209R *Tangut glyph corrections* section 4.1 on pp. 82–83), but did not reach a conclusion. As the dotted forms of U+17D0B 𠂇 and U+180DF 𠂇 used in Hán Xiǎománg’s dictionary match the glyph forms shown in *Homophones*, it is reasonable to now update the code chart glyphs for these two characters to use the dotted form of the component.

No.	Radical	Strokes	Glyph	Example Ideograph	Ideograph	Note
783	RR189	5	𠂔	𠂔	U+1724C 𠂔	
784	RR191	5	𠂕	𠂕		
785	RR193	6	𠂖	𠂖		
786	RR198	6	𠂗	𠂗		
787	RR204	6	𠂘	𠂘	U+17000 𠂘	
788	LR210	6	𠂙	𠂙		
789	RR226*	6	𠂚	𠂚		
790	LR231	6	𠂛	𠂛(𠂛)		
791	RR248	6	𠂜	𠂜		
792	RR256	6	𠂝	𠂝		
793	RR261	6	𠂞	𠂞		
794	RR264	6	𠂟	𠂟		
795	RR265	6	𠂠	𠂠		
796	LR274 RR271	6	𠂡	𠂡(𠂡), 𠂡		4
797	RR274*	7	𠂢	𠂢		
798	RR275	7	𠂣(𠂣)	𠂣	U+17100 𠂣	
799	RR276	7	𠂤	𠂤	U+17684 𠂤	
800	RR284	7	𠂥	𠂥		
801	RR294	7	𠂦	𠂦		
802	RR295	7	𠂧(𠂧)	𠂧	U+171CA 𠂧	
803	RR296	7	𠂨	𠂨		
804	RR306	7	𠂩(𠂩)	𠂩		

⁴ RR187 has the same glyph as LR274/RR271 𠂡 (two dots), but is counted as five strokes. I suspect that the glyph for RR187 is a mistake for Component 343 𠂡 (only one dot), which is the expected component for the characters indexed under RR187.

No.	Radical	Strokes	Glyph	Example Ideograph	Ideograph	Note
805	RR320	7	𠂔	𠂔		
806	RR321*	7	𠂔	𠂔	U+185EA 𠂔	
807	RR323	7	𠂔 (𠂔)	𠂔		
808	RR324	7	𠂔	𠂔		
809	RR326	7	𠂔 (𠂔)	𠂔		
810	RR332	7	𠂔	𠂔		
811	RR337*	7	𠂔	𠂔		
812	RR340	7	𠂔	𠂔		
813	RR341	7	𠂔	𠂔		
814	RR347	8	𠂔	𠂔		
815	RR348	8	𠂔	𠂔		
816	RR349	8	𠂔	𠂔		
817	RR353	8	𠂔	𠂔		
818	RR354	8	𠂔	𠂔		
819	RR355	8	𠂔	𠂔 (𠂔)		
820	RR356	8	𠂔	𠂔		
821	RR362	8	𠂔	𠂔	U+1713C 𠂔	
822	RR365	8	𠂔 (𠂔)	𠂔	U+1740C 𠂔	
823	RR366	8	𠂔 (𠂔)	𠂔	U+1740E 𠂔	
824	RR368	8	𠂔	𠂔		
825	RR374	8	𠂔	𠂔		
826	RR375	8	𠂔	𠂔		
827	RR378	8	𠂔	𠂔		
828	RR379	8	𠂔 (𠂔)	𠂔		

No.	Radical	Strokes	Glyph	Example Ideograph	Ideograph	Note
829	RR382	8	𦰩	樣	U+17430 𦰩	
830	RR385	8	𦰪	𦰪		
831	RR386*	8	𦰫	𦰫	U+17BE6 𦰫	
832	RR393	8	𦰬	𦰬	U+171E1 𦰬	
833	RR395	8	𦰭 (𦰭)	𦰭		
834	RR399	8	𦰮	𦰮		
835	RR401	8	𦰯 (𦰯)	𦰯		
836	RR408	9	𦰰	𦰰		
837	RR409	9	𦰱	𦰱	U+17363 𦰱	
838	RR410	9	𦰲 (𦰲)	𦰲		
839	RR411	9	𦰳 (𦰳)	𦰳		
840	RR413	9	𦰴	𦰴	U+1768D 𦰴	
841	RR416	9	𦰵 (𦰵)	𦰵	U+17861 𦰵	
842	RR417	9	𦰶	𦰶	U+17012 𦰶	
843	RR418	9	𦰷 (𦰷)	𦰷	U+1714B 𦰷	
844	RR419	9	𦰸	𦰸	U+17160 𦰸	
845	RR420	9	𦰹	𦰹	U+17F08 𦰹	
846	RR423	9	𦰺	𦰺		
847	RR425	9	𦰻	𦰻		
848	RR426	9	𦰼	𦰼	U+1835E 𦰼	
849	RR430	9	𦰽	𦰽		
850	RR431	9	𦰾	𦰾	U+1754D 𦰾	
851	RR432	9	𦰿	𦰿	U+1754E 𦰿	
852	LR407 RR436	9	𦱀	𦱀 (𦱀), 𦱀		

No.	Radical	Strokes	Glyph	Example Ideograph	Ideograph	Note
853	RR438	9	𦍋	𦍋		
854	RR439	9	𦍌	𦍌		
855	RR441	9	𦍍	𦍍	U+1818F 𦍍	
856	RR445	9	𦍎 (𦍎)	𦍎		
857	RR446	10	𦍇	𦍇		
858	RR447	10	𦍈 (𦍈)	𦍈	U+17D32 𦍈	
859	RR448	10	𦍉	𦍉	U+17690 𦍉	
860	RR450	10	𦍊	𦍊	U+1769A 𦍊	
861	RR451	10	𦍋	𦍋		
862	RR457	10	𦍌	𦍌	U+187A4 𦍌	
863	RR458	10	𦍍	𦍍		
864	RR459	10	𦍎	𦍎	U+1744C 𦍎	
865	RR460	10	𦍇	𦍇	U+17BE7 𦍇	
866	RR461	10	𦍈	𦍈		
867	RR462	10	𦍉	𦍉		
868	RR465	10	𦍊	𦍊		
869	RR468	10	𦍋	𦍋		
870	RR469	10	𦍌	𦍌	U+18672 𦍌	
871	RR470	11	𦍍	𦍍	U+1710A 𦍍	
872	LR433 RR476	11	𦍎	𦍎 (𦍎), 𦍎		
873	RR478	11	𦍇	𦍇	U+17555 𦍇	
874	RR479	11	𦍈	𦍈	U+17C1E 𦍈	
875	RR480	12	𦍉 (𦍉)	𦍉	U+17F57 𦍉	

No.	Radical	Strokes	Glyph	Example Ideograph	Ideograph	Note
876	RR487	12 (13)	𦍋 (𦍋)	𦍋	U+18050 𦍋	5
877	RR483	12	𦍌	𦍌	U+1876A 𦍌	
878	RR485	12	𦍍	𦍍	U+187E3 𦍍	
879	LR443* RR486	13	𦍎 (𦍎)	𦍎	U+187E5 𦍎	
880	RR488	13	𦍇	𦍇	U+174B9 𦍇	
881	RR490	13	𦍈	𦍈	U+17C78 𦍈	
882	RR491	14	𦍉	𦍉	U+181D2 𦍉	

⁵ RR487 is listed under 13 strokes, and the dictionary glyph has a dot on right side (13 strokes); the three ideographs indexed under RR487 (U+18050 𦍋, U+17AC1 𦍋, and U+181CD 𦍋) also show a final dot on the right side in the dictionary entries. However, the code chart glyphs for these three ideographs are all without a final dot, which matches the glyph forms shown in primary sources such as *Homophones* and *Sea of Writing* where none of the entries show a final dot on these three ideographs. Therefore, I believe that the final dot on RR487 and three associated ideographs in Hán Xiǎománg's dictionary should be a mistake, and propose to encode the dotless form of the component (12 strokes) rather than the actual form shown in the dictionary.

3. Evidence

Xìxiàwén Cídiǎn vol. 9 p. 69: LR177 𠂔

LR177 𠂔 部

⊕ ↑ ⊕ ——— ⊕ ↑ ⊕

𠂔 (13) V 104 1. 17 khwa

⑥410

𠂔 (16) V-20 1. 22 khwaa

⑥410

𠂔 (16) V-20 1. 22 khwaa

(𠂔之异体字) ⑥411

☞ □ ⊕ ——— ☞ □ ⊕

𠂔 (11) V 21 2. 56 ka ⑥411

𠂔 (17) IX5 2. 72 rjir ⑥412

Xìxiàwén Cídiǎn vol. 9 p. 70: LR190 𠂔

LR190 𠂔 部

☞ □ ⊕ ——— ☞ □ ⊕

𠂔 (07) V-14 1. 26 kjwā ⑥491

Xìxiàwén Cídiǎn vol. 9 p. 72: LR210 𠂔

LR210 𠂔 部

☞ □ ⊕ ——— ☞ □ ⊕

𠂔 (12) VI128 1. 17 sa ⑦50

LR231 𠂔部



𠂔 (08) Ⅱ-34 1.23 xjwe

(𠂔之异体字) ⑦161

𠂔 (14) Ⅱ-14 1.10 xjwi

(𠂔之异体字) ⑦161

LR274 𠂔部



𠂔 (10) Ⅶ22 1.28 sia ⑦581

𠂔 (11) Ⅶ95 2.43 dzio ⑦582

𠂔 (11) Ⅰ-28 1.3 mju ⑦582

𠂔 (11) Ⅴ70 2.33 gjij ⑦583

𠂔 (11) Ⅲ135 2.45 noo ⑦583

LR407 𪛗 部



𪛗 (13) Ⅱ -7 2. 72 wjir
(𪛗之异体字) ⑧517

𪛘 (15) Ⅱ 13 2. 52 wju
(𪛘之异体字) ⑧517

LR433 𪛙 部

𪛙 (11) Ⅸ21 1. 49 lo
(𪛙之异体字) ⑧575



𪛚 (17) Ⅶ-18 2. 29 sjh
(𪛚之异体字) ⑧575

𪛛 (19) Ⅶ38 1. 45 tshjiw
(𪛛之异体字) ⑧575

LR443 𪛜 (𪛜) 部

𪛜 (13) Ⅰ 22 2. 25 bə ⑧603

𪛝 (18) Ⅰ 22 2. 25 bə ⑧605

RR3 \ 部



𠂔 (07) IX124 2. 37 lji ③483

RR4 , 部

※ ※ ※



𠂔 (11) IX21 1. 49 lo

(𠂔之异体字) ⑧575

𠂔 (15) VII24 (声调不详) dzj

(𠂔之异体字) ③397



𠂔 (10) IX-? 2. 11 lee

(𠂔之异体字) ①335

RR14 𠂔 部



𠂔 (08) V 81 2. 61 kj ①204

RR34 彡部



彡 (07) IX157 2. 48 lhiow

③456

彡 (08) III147 2. 10 dji ⑥174

彡 (09) VI-78 1. 33 swej ⑦324

彡 (11) VI23 1. 54 sow ⑧343

彡 (12) V149 2. 3 khju ⑥37



彡 (11) VII-37 2. 23 dзиа

(彡之异体字) ③64

彡 (11) VII-37 2. 23 dзиа ③64

RR53 彣部



彣 (09) I5 2. 68 mjijr ③302

彣 (12) I-133 1. 74 mjijr

⑧385

RR73 𠂔 部



𠂔 (08) Ⅲ52 2. 33 djij
(縫之异体字) ④26

縫 (08) Ⅲ52 2. 33 djij ④591

RR81 𠂔 部



𠂔 (08) Ⅷ29 1. 27 ywə ④595

RR102 𠂔 部



𠂔 (09) Ⅰ61 1. 20 pja ③717

𠂔 (12) Ⅰ61 1. 20 pja ②37

𠂔 (13) Ⅰ61 1. 20 pja ②42

RR128 𠂔 (𠂔) 部



𠂔 (09) Ⅲ 66 2. 20 niaa ③390



𠂔 (07) Ⅰ 35 2. 25 phə ①213

𠂔 (07) Ⅲ -16 1. 72 njə ①566

𠂔 (11) Ⅲ -169 1. 69 njwɿ
①631



𠂔 (07) Ⅲ -16 1. 72 njə
(𠂔之异体字) ①565

RR138 𠂔 部



𠂔 (12) Ⅰ 3 1. 11 bji ⑧142

RR154 𠂔 部



𠂔 (09) Ⅰ 84 1. 11 pji ③259

RR160 𠂔 部



𠂔 (10) VIII-97 1. 21 ·jaa ⑥308

RR185 𠂔 部



𠂔 (10) VIII-93 1. 26 ·jwā ⑥256

RR189 𠂔 部

𠂔 (05) I 126 2. 3 pju ①522



𠂔 (07) I 13 2. 25 mǎ ①571



𠂔 (12) V 191 2. 38 kew ②573

RR191 𡗗 部



𡗗 (10) VI124 1. 17 tswa ④613

𡗘 (11) I 83 2. 8 bie ⑤323

𡗙 (11) I 83 2. 8 bie ⑥454

𡗚 (13) I 26 2. 3 bju ⑧219



𡗛 (13) III 78 1. 63 da ③96

RR193 𡗛 部



𡗛 (10) IX57 2. 9 lji ④71

𡗜 (10) IX75 2. 30 lej ③319

𡗝 (12) IX76 1. 33 lej ⑥357

𡗞 (13) V 122 1. 30 kjɿ ⑧294

RR198 𡗟 部



𡗟 (10) I 70 1. 39 mjiiɰ ④93

𡗠 (16) V 25 1. 36 khjij ⑧556

RR204 𪛗 部

𪛗 (06) Ⅲ 55 1.49 do ①1



𪛘 (10) Ⅲ 55 1.49 do ③52



𪛙 (11) V 176 2.6 gjuu ⑥103

𪛚 (12) IX 153 (声调不详) la
⑦123

RR226 𪛛 (𪛛) 部



𪛜 (10) IX-? 1.56 ljow ④128

𪛝 (10) I -36 2.73 mar ③503

𪛞 (11) I 124 1.83 maar ③508



𪛟 (14) IX 43 2.49 lhjow ⑦501

𪛠 (15) I 124 1.83 maar
③167



𪛡 (12) V 112 1.30 khjwi
⑦475

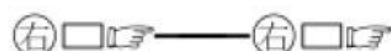
RR248 𠂔 部



𠂔 (10) V 80 2. 85 kjir ④147

𠂔 (13) V 47 2. 70 kjur ⑧234

RR256 𠂔 部



𠂔 (11) VII 130 2. 12 sjwii

⑥162

𠂔 (13) VII 76 2. 61 dzjɿ ⑧44

RR261 𠂔 部



𠂔 (12) IX-? 1. 69 lhjɿ ⑦314

𠂔 (14) VI-36 2. 7 -e ④445

RR264 𠂔 部



𠂔 (10) Ⅲ33 2. 47 thow ④151

𠂔 (15) Ⅸ61 2. 47 low ④495

RR265 𠂔 部



𠂔 (11) Ⅲ67 1. 67 tji ⑤436

RR271 𠂔 部



𠂔 (10) Ⅲ35 2. 71 ner ③395

RR274 𠂔 (𠂔) 部



𠂔 (09) Ⅲ113 1. 68 tɔ ①588



𠂔 (11) Ⅸ34 1. 30 lhji ④184

RR275 𠂔 部

𠂔 (07) (声韵不详) ①181



𠂔 (11) VIII40 1. 34 yiej ③413

RR276 𠂔 部

𠂔 (07) III 55 1. 49 do ③48



𠂔 (11) I 143 2. 6 mjuu ③2

𠂔 (12) IX105 2. 55 ljɿj ⑥314

RR284 𠂔 部



𠂔 (11) II 31 1. 64 wja ③13

𠂔 (11) III146 1. 11 dji ③241

𠂔 (11) I 1 2. 10 bji ③267

𠂔 (12) VI7 1. 42 dzjɿj ④297

𠂔 (13) III131 1. 3 djwu ⑦551

RR294 𧈧 部



𧈧 (11) VI26 1. 11 sji ③579



𧈧 (16) VI26 1. 11 sji ③179

RR295 𧈩 部

𧈩 (07) III17 2. 33 thjij ①415



𧈩 (11) III64 1. 5 duu ④220

𧈩 (13) III17 2. 33 thjij ⑦83

RR296 𧈪 部



𧈪 (12) I 49 1. 1 pu ⑤324

𧈪 (12) I 49 1. 1 pu ⑥113

RR306 𦵏 部



𦵏 (10) V 170 1.9 kiwe ②86

𦵏 (10) VIII 104 1.3 xju ②583

𦵏 (11) VII 5 1.57 síoow ②31

𦵏 (12) VII-119 1.9 síwe ⑥485

RR320 𦵏 部



𦵏 (15) VII 34 1.2 tsju ⑧237

RR321 𠂔 (𠂔) 部



𠂔 (09) IX8 2. 14 la

(𠂔之异体字) ①331

𠂔 (13) VI-18 2. 38 dzew ⑦61



𠂔 (07) IX62 1. 3 ljwu ⑧237



𠂔 (09) IX8 2. 14 la ①331

𠂔 (13) VI-18 2. 38 dzew

RR323 𠂔 部



𠂔 (09) IX115 1. 3 lhju ①600

RR324 𠂔 部



𠂔 (09) I 79 1. 61 mjij ①332

𠂔 (10) I 79 1. 61 mjij ②87

RR326 𦍋 部



𦍋 (16) VI137 2. 28 dzjwi

⑧481



𦍋 (20) VI137 2. 28 dzjwi

③206

RR332 𦍋 部



𦍋 (13) III147 2. 10 dji ④374

𦍋 (15) VI3 2. 60 dzji ⑧361

𦍋 (15) VI92 2. 86 tsjiir ⑧443

RR337 𦍋 (𦍋) 部



𦍋 (13) I145 1. 59 pju ⑦88



𦍋 (09) VII-48 2. 44 sjō ①601

RR340 𠂔 部



𠂔 (15) IX48 1.49 lwo ⑧347



𠂔 (14) V-112 1.95 khjoor
②334

𠂔 (14) V-112 1.95 khjoor
(𠂔之异体字) ②337

RR341 𠂕 部



𠂕 (09) I 101 1.69 bjɿ ①602

𠂕 (12) III 20 2.42 to
(𠂕之异体字) ④364

𠂕 (13) III 20 2.42 to ④428

𠂕 (14) VII45 1.10 tshji ⑧318

RR347 𠂔 部



𠂔 (10) VIII105 1. 66 yie ①517

𠂔 (11) IX172 2. 60 zji ②562

𠂔 (11) V14 1. 84 ɲwər ②584

RR348 𠂕 部



𠂕 (11) IX4 1. 67 lji ②503

𠂕 (13) II45 1. 27 wə ⑤206

𠂕 (15) VII162 2. 54 dzjij
(𠂕之异体字) ⑧36

RR349 𠂖 部



𠂖 (13) VIII44 2. 74 ·jar ①129

RR353 𪛗 部



𪛗 (09) IX16 2. 68 rjijr
(𪛗之异体字) ①60

𪛘 (12) VI13 1. 68 tsə ③467

𪛙 (15) I 22 2. 25 hə ⑧77

RR354 𪛚 部



𪛚 (10) IX-? 2. 11 lee ①335

RR355 𪛛 部



𪛛 (10) IX-? 2. 11 lee
(𪛛之异体字) ①335

RR356 𠂔 部

𠂔 (08) Ⅲ68 1. 67 tji ③247



𠂔 (10) Ⅲ52 2. 33 djij ①604

𠂔 (12) Ⅶ75 1. 92 dzjɿr ③450

𠂔 (14) V-52 2. 5 kuu ④446

𠂔 (15) Ⅷ-109 1. 70 ɿ ⑧76



𠂔 (18) Ⅲ16 1. 36 thjij ⑦527

RR362 𠂔 部

𠂔 (08) Ⅸ-? 1. 20 lja ①228



𠂔 (12) Ⅸ87 1. 69 ljɿ ③73

𠂔 (14) I-136 2. 33 phjij ⑦1

𠂔 (14) Ⅵ35 1. 36 tshjij ⑦501



𠂔 (14) Ⅸ88 1. 69 ljɿ ⑦431



𠂔 (17) Ⅷ-26 2. 53 yiej ②386

RR365 𩚑 部

𩚑 (08) II 35 1. 17 wa ②148



𩚒 (12) VI 5 2. 3 dzju ④308

𩚓 (12) III 21 1. 49 to ③561

𩚔 (12) I 144 2. 52 pju ①176

𩚕 (13) II 45 1. 27 wə ⑤207

𩚖 (13) VIII 78 2. 51 •wɿ ⑥464

𩚗 (14) III 115 1. 87 dewr
⑦278

RR366 𩚘 部

𩚘 (08) I 114 1. 30 pji ②148



𩚙 (12) I 61 1. 20 pja ③725

RR368 𩚚 部



𩚚 (12) I 121 1. 24 phã ③25

𩚛 (12) I 121 1. 24 phã ④317

RR374 𪗇 部



𪗇 (09) III 107 1. 3 nju ②212

𪗈 (13) III 54 2. 10 nji ④398

𪗉 (13) V -135 2. 33 kjij ⑥47

𪗊 (15) II 16 1. 77 wer ⑦299

𪗋 (18) VI 101 1. 30 sjwi ⑧538



𪗌 (16) I 19 1. 51 mjo ②376

RR375 𪗍 部



𪗍 (09) IX 163 1. 63 lwa ①62

𪗎 (12) VIII 25 1. 7 •juu ④327

𪗏 (13) III 57 1. 54 dow ⑤241

RR378 𪔐 部



𪔐 (13) VIII 10 1. 30 ·jǐ ⑥49

𪔑 (13) III 148 1. 64 njǎ ⑥423

𪔒 (13) III 159 2. 52 tju ⑤201

RR379 𪔓 部



𪔓 (14) III 60 1. 30 njǐ ⑦156

RR382 𪔔 部

𪔔 (08) I 35 2. 25 phǎ ②208



𪔕 (11) III 115 1. 87 dewr ②66

𪔖 (13) VIII 92 2. 42 ·o ⑤223

𪔗 (16) III 切韵 1. 87 tewr
(𪔗之异体字) ⑥400

𪔘 (17) III 切韵 1. 87 tewr ⑦460

RR385 𠂔 部



𠂔 (11) Ⅲ-128 2. 6 njuu ②504

𠂔 (12) Ⅱ 6 2. 9 wji ④338

𠂔 (13) Ⅲ 9 2. 3 nju ④426

RR386 𠂔 (𠂔) 部



𠂔 (14) Ⅱ 14 1. 59 wju ⑦438

𠂔 (15) V 122 1. 30 kjǐ ⑧308



𠂔 (14) Ⅱ 14 1. 59 wju
(𠂔之异体字) ⑦439

RR393 𠂔 部

𠂔 (08) IX10 2. 37 rjĭj ①437



𠂔 (12) IX16 2. 68 rjiĭr ③468

𠂔 (13) IX10 2. 37 rjĭj ⑥287



𠂔 (16) IX10 2. 37 rjĭj ②384

RR395 𠂔 部



𠂔 (13) I 24 1. 30 mĭj ⑤82

𠂔 (13) VII53 1. 2 dĭjwu ⑤208

RR399 𠂔 部



𠂔 (10) IX-? 1. 54 lwow ①345

𠂔 (12) VII134 1. 42 tshjwĭj ④350

𠂔 (13) VI111 1. 1 tshu ⑥372

RR401 𪛗 部



𪛗 (12) II 54 2. 42 wo ③227

𪛘 (12) II 54 2. 42 wo ④352

𪛙 (15) IX15 1. 11 zji ⑧309

RR408 𪛚 部



𪛚 (12) VI118 1. 3 tshjwu
②584

𪛛 (13) V194 (声调不详) khiwə
③27

𪛜 (14) VI143 1. 32 tshji
⑥423

RR409 援部

援 (09) I 60 2. 17 bja ②47



𢇛 (11) V 43 1. 87 ŋewr ①347

𢇜 (13) I 66 1. 20 phja ④368

𢇝 (19) III -2 2. 22 thwǎ ⑧538



𢇞 (13) VIII 34 1. 63 ya ②293

𢇟 (20) V 18 1. 61 gjij ⑧205

𢇠 (21) V -128 2. 64 kjo ①155

Xìxiàwén Cídiǎn vol. 9 p. 189: RR410 𦵀 (𦵀)

RR410 𦵀部



𦵁 (11) III 151 1. 3 dju ①348

𦵂 (13) VI 3 2. 60 dzji ③360

Xìxiàwén Cídiǎn vol. 9 p. 189: RR411 𦵃 (𦵃)

RR411 𦵃部



𦵄 (13) I 87 2. 42 bo ④373

𦵅 (15) I 87 2. 42 bo ⑦444

RR413 𦵑部

𦵑 (09) VII-29 1. 26 dzjwā ③51



𦵑 (17) V-119 1. 24 gā ⑧392



𦵑 (16) VI-159 1. 24 dzā ③647

Xìxiàwén Cídiǎn vol. 9 p. 190: RR416 𦵑 (𦵑)

RR416 𦵑部

𦵑 (10) II 29 1. 31 wəə ③302



𦵑 (13) VIII-104 2. 49 ·jow
②292

𦵑 (14) IX-? (声韵不详)
③124

𦵑 (15) IX-? 2. 17 lhja ⑥418



𦵑 (14) II 30 1. 31 wəə ④443

RR417 𪔐 部

𪔐 (09) VII42 2.48 sjow ①17



𪔑 (11) III70 2.55 tjij ①350

𪔒 (12) IX44 2.76 lwər ②505

𪔓 (13) VIII69 1.39 ·jiij ④375

𪔔 (14) VI91 1.93 tsjiir ⑤241

𪔕 (14) VI113 2.38 tshew
⑤332

RR418 𪔖 部

𪔖 (09) I41 1.11 mji ①267



𪔗 (13) III-69 2.73 -ar ③87



𪔘 (15) III145 2.19 naa
(𪔘之异体字) ③169

𪔙 (15) III145 2.19 naa ①676

RR419 𪔐 部

𪔐 (10) V-24 1. 75 gur ①338



𪔑 (13) V 173 1. 47 gjiiw

②294

𪔒 (14) V 173 1. 47 gjiiw

③125

𪔓 (15) I 57 1. 2 bju ⑤60

𪔔 (15) IX60 2. 66 rejr ⑥142

RR420 𪔕 部

𪔕 (09) III-158 1. 64 njə ⑤352



𪔖 (12) III 148 1. 64 njə ②271

𪔗 (13) VII63 1. 16 šji ③91

𪔘 (14) VIII50 2. 66 •wejr ⑤34



𪔙 (14) (声韵不详) ⑥289



𪔚 (22) V 117 1. 5 guu ⑧393

RR423 𪔐 部



𪔐 (13) V 80 2. 85 kj̥ir ④399

𪔐 (14) IX 158 2. 13 zē ⑤333

RR425 𪔑 部



𪔑 (11) VI 79 1. 87 tsewr ②245

𪔑 (15) VI 79 1. 87 tsewr ⑤336

𪔑 (16) VI 99 1. 42 dzj̥ij ⑥79

RR426 𪔒 部

𪔒 (09) VIII 96 1. 71 ·iə ⑦264



𪔒 (13) VIII 96 1. 71 ·iə ③98



𪔒 (14) V 87 1. 68 kwə ⑥265

RR430 𠂔 部



𠂔 (10) V 102 2. 28 khji ②425

𠂔 (14) II 59 2. 23 wiǎ ④460

𠂔 (14) IX 5 2. 72 rjir ⑤334

RR431 𠂔 部

𠂔 (09) IX 77 1. 93 ljiir ②410



𠂔 (19) III 40 2. 28 dji ⑧558

RR432 𠂔 部

𠂔 (09) III 3 2. 41 neew ②411



𠂔 (14) III 3 2. 41 neew ⑤334

RR436 𠂔 部



𠂔 (16) II 19 2. 68 wjijr
(𠂔之异体字) ①151

𠂔 (18) III 69 2. 60 tji
(𠂔之异体字) ②191

RR438 𠂔 部



𠂔 (11) II 18 2. 71 wer ①636

𠂔 (14) VIII-19 2. 22 yā ⑤153

𠂔 (15) III 51 2. 38 dew ④524

𠂔 (16) VI 48 2. 52 dzju ⑤390

RR439 𪔐 部



𪔐 (10) V 147 1. 79 kjir ①87

𪔑 (13) VI60 1. 30 tsjwi ④417

𪔒 (15) V -45 2. 77 ŋjir ⑦450



𪔓 (19) VI60 1. 30 tsjwi ③655

RR441 𪔔 部

𪔔 (09) IX-? 1. 32 ljhi ⑥412



𪔕 (13) VII126 2. 32 dzjwij ③286

𪔖 (13) V -21 2. 57 kja ③470

RR445 𪔗 部



𪔘 (11) I 145 1. 59 pju ①371

𪔙 (14) III 191 1. 32 djhi ⑥230

RR446 𠂔 部



- 𠂔 (13) VII48 2. 32 dzjij ②89
𠂔 (14) II 2 2. 7 we ④430
𠂔 (15) V -108 1. 7 kjuu ⑤223
𠂔 (15) (声韵不详) ⑤246
𠂔 (16) I 163 1. 5 puu ⑦263
𠂔 (16) I 158 2. 8 phie ⑦454
𠂔 (17) IX37 2. 71 rer ⑧50

Xìxiàwén Cídiǎn vol. 9 p. 192: RR447 𠂔 (𠂔)

RR447 𠂔 部

- 𠂔 (10) VII135 2. 32 dzjwij ⑤1



- 𠂔 (12) VII126 2. 32 dzjwij
①374

- 𠂔 (14) VII135 2. 32 dzjwij
④433

- 𠂔 (15) III24 2. 28 njwi ⑤214

- 𠂔 (15) VIII61 1. 17 ·a ⑤335

RR448 菝部

菝 (10) IX71 1. 60 liej ③52



𦵑 (12) IX123 1. 66 lie ①640

𦵒 (14) I -1 1. 66 bie ④434

𦵓 (15) IX71 1. 60 liej ⑤335

𦵔 (15) IX71 1. 60 liej ⑥469

RR450 菱部

菱 (10) IX73 2. 38 zew ③55



𦵕 (14) VII30 2. 44 tshjo ③230

𦵖 (16) VII108 1. 79 tsjir ⑦455

RR451 蒨部



蒨 (14) II 40 1. 15 wě ④437

蒨 (15) II 40 1. 15 wě ⑥68

蒨 (16) V 2 1. 17 ka ⑦455

RR457 𦍋 部

𦍋 (10) IX61 2.47 low ⑧542



𦍋 (16) I 149 1.9 pie ⑦230



𦍋 (20) I 149 1.9 pie ③207

RR458 𦍌 部



𦍌 (13) VI118 1.3 tshjwu ②585

𦍌 (14) VI66 1.77 tser ④452

𦍌 (15) VI49 2.3 sju ⑤389

𦍌 (15) VI12 1.68 tsə ⑥73



𦍌 (19) VI12 1.68 tsə ③204

RR459 薹部

薹 (10) V 22 1. 34 kiek ②232



薹 (14) IX97 2. 54 ljwij ③709

薹 (15) V -94 2. 38 khew

⑤456

RR460 薹部

薹 (10) V 67 2. 29 khjii ③484



薹 (11) V 168 1. 6 khjuu ①175

RR461 𪚩 部



𪚩 (13) II 31 1. 64 wja ②2

𪚪 (13) VI 118 1. 3 tshjwu

②585

𪚫 (14) VI 12 1. 68 tsə ④456

𪚬 (15) IX 57 2. 9 lji ⑥168

𪚭 (16) III 135 2. 45 noo ⑦587

𪚮 (19) IX 189 1. 32 lji ⑧524



𪚯 (18) VI 12 1. 68 tsə ③198

RR462 𦍋 部



𦍋 (14) I 44 1. 32 mjii ④461

𦍋 (14) IV 3 2. 48 niow ③328

𦍋 (15) I 44 1. 32 mjii ⑥77



𦍋 (18) I 45 2. 29 mjii

(𦍋之异体字) ③198

𦍋 (19) I 45 2. 29 mjii ③205

RR465 𦍍 部



𦍍 (12) IX 134 2. 2 zju ①661

𦍍 (16) I 切韵 2. 2 pju ⑦231

𦍍 (16) I 17 1. 20 mja ⑦472

RR468 𪛗部



𪛗 (14) IX64 1.22 laa ④476

𪛘 (16) I 69 2.14 ba ⑦158

𪛙 (16) IX93 1.76 rjur ⑦642



𪛚 (17) V 4 2.48 ŋjow ②390

Xìxiàwén Cídiǎn vol. 9 pp. 193–194: RR469 𪛛

RR469 𪛛部

𪛛 (10) VII-129 1.51 dzjo ⑧324



𪛜 (13) VII-23 (声韵不详) ②323

𪛝 (15) V 71 1.36 gjij ⑤462

𪛞 (16) VIII39 2.60 ·ji ⑦459



𪛟 (21) V 54 2.10 gjj ③364

RR470 𪚩 部

𪚩 (11) IX14 2. 10 zji ①188



𪚪 (14) IX133 2. 61 zjɿ ②67

𪚫 (15) IX14 2. 10 zji ④476

𪚬 (17) IX14 2. 10 zji ⑦491

RR476 𩇛部

𩇛 (11) IX21 1.49 lo
(𩇛之异体字) ⑧575



𩇛 (15) VII24 (声调不详) dzjɿ
(𩇛之异体字) ③397



𩇛 (12) IX-? 1.14 lhjii
(𩇛之异体字) ①100

𩇛 (13) VI62 1.20 tsja
(𩇛之异体字) ①522

𩇛 (14) III166 1.68 thə
(𩇛之异体字) ②525

𩇛 (15) V2 1.17 ka
(𩇛之异体字) ③289

𩇛 (15) III37 1.1 thu
(𩇛之异体字) ④483

RR478 𪚩 部

𪚩 (11) I 36 1. 27 phə ②433



𪚪 (15) I 35 2. 25 phə ③289

𪚫 (16) I 35 2. 25 phə ⑤463

RR479 𪚬 部

𪚬 (11) III 149 1. 21 njaa ③551



𪚭 (14) III 157 1. 1 nu ②410

𪚮 (14) III 2 1. 17 na ②342

𪚯 (15) III 149 1. 21 njaa ③168



𪚰 (13) III 149 1. 21 njaa ①669

𪚱 (15) V 28 1. 54 kow ④519

𪚲 (16) IX-? 1. 60 lej ⑥88

RR480 𡇗 部

𡇗 (12) IX148 2. 25 lə ⑤444



𡇗 (14) VII59 2. 81 tsior ①671

𡇗 (17) VIII-59 2. 37 ·jij ⑤307

𡇗 (17) IX-? 1. 14 zjii ⑥400



𡇗 (20) IX-? 2. 12 zjii ②394

RR481 𡇗 部

𡇗 (12) IX130 1. 86 zjir ⑧592

RR483 𪔐部

𪔐 (12) VII7 1. 56 dzjow ⑧509

⊕↓⊕——⊕↓⊕

𪔐 (14) V3 2. 14 ka

(𪔐之异体字) ①431

𪔐 (14) V3 2. 14 ka ①468

𪔐 (15) V3 2. 14 ka

(𪔐之异体字) ②359

𪔐 (16) I -129 2. 34 phiej

③183

𪔐 (17) V -51 2. 73 kar ⑤131

⊕□⊕——⊕□⊕

𪔐 (13) I 161 2. 20 phiaa

①135

𪔐 (14) VI-47 1. 8 dze ①403

𪔐 (16) VII-10 1. 56 tshjow

④540

𪔐 (16) VI-40 1. 86 tsjwɪr

③418

𪔐 (17) VII75 1. 92 dzjɪr ④576

RR485 𪔐部

𪔐 (12) III 151 1. 3 dju ③612



𪔑 (16) VIII 26 1. 11 ·ji ④552

𪔒 (17) VII 48 2. 32 dzjij ⑥88



𪔓 (20) VIII 26 1. 11 ·ji ③207

Xìxiàwén Cídiǎn vol. 9 pp. 195–196: RR486 𪔔 (𪔕)

RR486 𪔔部

𪔔 (13) I 22 2. 25 bæ ⑧603



𪔕 (17) I -105 1. 31 bæə

③472

𪔖 (18) VI 118 1. 3 tshjwu

⑥117

RR487 𦍋 部

𦍋 (13) I 26 2.3 bju ⑥141



𦍋 (17) VIII-61 1.9 yiwe ④560

𦍋 (18) VIII78 2.51 ·wu ⑥469

RR488 𦍌 部

𦍌 (13) VII71 1.64 sj̥a ②301



𦍌 (18) VII_{切韵} 1.64 ts̥ja
(𦍌之异体字) ⑤97



𦍌 (21) VII_{切韵} 1.64 ts̥ja ③211

RR490 𦍍 部

𦍍 (13) IX55 1.58 lu ③620



𦍍 (15) III66 2.20 niaa ①513

RR491 𪔐部

𪔐 (13) II -33 1. 78 wier ⑥473



𪔐 (18) IX138 1. 7 zjuu ③654



𪔐 (16) I 35 2. 25 phə ①410

𪔐 (19) I -11 2. 14 pa ⑥293

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 10646⁶

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:	<i>Proposal to encode 114 Tangut components</i>
2. Requester's name:	<i>Andrew West</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>2023-05-31</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>YES</i>
Proposed name of script:	<i>Tangut Components Supplement</i>
b. The proposal is for addition of character(s) to an existing block:	<i>NO</i>
Name of the existing block:	
2. Number of characters in proposal:	<i>114</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary	<input type="checkbox"/>
B.1-Specialized (small collection)	<input type="checkbox"/>
B.2-Specialized (large collection)	<input type="checkbox"/>
C-Major extinct	<input type="checkbox"/>
D-Attested extinct	<input type="checkbox"/>
E-Minor extinct	<input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic	<input checked="" type="checkbox"/>
G-Obscure or questionable usage symbols	<input type="checkbox"/>
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>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>Andrew West</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>NO</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.

⁶ 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.)?	NO
If YES, with whom?	
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?	YES
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:	