

# Characters for Classical Latin

David J. Perry  
Version 2, March 28, 2006

## Introduction

The purpose of this document is to identify *all* characters of interest to those who work with Classical Latin, no matter how rare. Epigraphers will want many of these, but I want to collect any character that is needed in any context. Those that are already available in Unicode will be so identified; those that may be available can be debated; and those that are clearly absent and should be proposed can be proposed; and those that are so rare as to be unencodable will be known.

If you have any suggestions for additional characters or reactions to the suggestions made here, please email David Perry at [hospes@scholarsfonts.net](mailto:hospes@scholarsfonts.net). No matter how rare, let's get all possible characters on this list.

Please note that abbreviations constructed with lines above or through existing letters are not considered separate characters except in unusual circumstances, nor are the space-saving ligatures found in Latin inscriptions unless they have a unique grammatical or phonemic function (which they normally don't).

Characters that I think **should be proposed** for inclusion in Unicode are highlighted in gold to make them easy to spot. Characters **that require more discussion** are highlighted in green. **Missing or incorrectly shaped glyphs** are highlighted in purple; these will be fixed in future versions of this document.

Some references are made to “smart font technology.” This refers to OpenType fonts (Windows and Linux, and occasionally on Mac OS) or AAT (Mac) fonts that can do things like position diacritics correctly over a variety of base letters or replace one sequence of characters with a ligature glyph. Such technology is available today, although not supported by many applications; there is every reason to think it will be more widely supported in the future, especially after the release of Windows Vista. If you want more information about this, email me. What users need to know, however, is that they must tell font developers what they need so that fonts can be created.

The Medieval Unicode Font Initiative is a group of medievalists working hard to get their characters into Unicode and to create a standardized solution for those that cannot be encoded. Some of their work overlaps with this document, and when appropriate cross-references are given. See the MUFI website at <http://gandalf.aksis.uib.no/mufi/>.

# List of Classical Latin Characters

## I. Letters, including letters with diacritics

### A. Long I and other tall letters

?	missing	Unicode does not normally encode letters that are variant forms of items already in the Standard. However, the fact that the long I is intended to represent the phoneme /i:/, distinct from /i/, makes it a candidate for encoding, and it <b>should be proposed</b> . Other letters are occasionally found written in an extra-tall form, but unless they can be shown to have specific phonemic or grammatical values (not the case, as far as I know), they should be treated as glyph variants.
---	---------	--

### B. Claudian letters

ⱦ	2132	Two of the Claudian letters (uppercase forms) are already in Unicode. The fact that U+2183 is named ROMAN NUMERAL REVERSED ONE HUNDRED might seem to make it unsuitable as a regular letter, but Michael Everson, author of the proposal for the Claudian letters and expert on Unicode, recommends this use. The third, the half-H character, has been proposed, along with lowercase forms for all three Claudian letters. <u>Tentative</u> Unicode values are U+214E TURNED SMALL F, U+2184 LATIN SMALL LETTER REVERSED C, and U+2C75/76 LATIN CAPITAL/SMALL LETTER HALF H.
ↀ	2183	
ⱦ	proposed	

### C. Name abbreviations

ↀ	2183	The reversed C sometimes found for Gaia should be represented by this character; see above under Claudian letters for justification of using this “numeral” as a regular letter.
ⱦ	missing	The five-legged variant of M used for Manius is missing and <b>should be proposed</b> . The MUFI spec has it in the PUA at F227 with the name MARIAN M, which I don’t understand, but cites a German work on Latin inscription. More research needed about that.

### D. Ligatures

ZWJ	200D	Unicode does not encode ligatures, except when they can be shown to have special phonemic values (not true for ancient Roman usage, where ligatures are mainly space-saving devices). Encode them by entering the first letter, then a zero width joiner, then the second letter. Smart fonts can replace such sequences with the appropriate glyph.
-----	------	--

## E. Apex and other markers of vowel length

Á	00C1	For the apex, use the acute accent. A, E, I, and O exist in precomposed form; V-apex should be encoded with the combining acute (U+0301). If a PUA codepoint for the latter is needed until combining marks are handled better, use E33A, as found in the Titus font and the MUF1 spec. Fonts may contain glyph variants if it is desired to use a different shape for the apex, such as the curved hook form. Using other diacritics such as xx for alternate apex shapes is not recommended.
É	00C9	
Í	00CD	
Ó	00D3	
Ā/ā	0100/1	If it is desired to use the modern convention of a macron to indicate length, all six vowels with macron are present in precomposed form in Unicode (given at left for easy reference). The 19 <sup>th</sup> century convention of using a circumflex for long vowels can be represented, if needed, with the precomposed circumflex vowels.  To mark a vowel specifically as short, the precomposed vowel + breve combinations are recommended. Y/y plus breve is not encoded, however, so Y/y followed by U+0306 COMBINING BREVE would be used.
Ē/ē	0112/3	
Ī/ī	012A/B	
Ō/ō	014C/D	
Ū/ū	016A/B	
Ȳ/ȳ	0232/3	

## F. Sicilicus

◌̣	0357	This rare diacritic (representing a doubled consonant) can be encoded using U+0309 COMBINING HOOK ABOVE or 0357 COMBINING HALF RING ABOVE; the latter seems the better choice to me since the former is intended as a tone mark in Vietnamese.
----	------	--

## G. Overline for general abbreviations

◌̄	0305	U+0305 COMBINING OVERLINE is specifically designed to connect visually to form a continuous line, which is usually what we find in Latin inscriptions. If adjacent letters with non-connecting overline are desired, use a macron, either the precomposed forms when they exist, or the COMBINING MACRON U+0304 after a base letter. Smart font technology is needed to position the combining marks correctly, both horizontally and vertically.
----	------	---

## H. Line through for general abbreviations

◌̅	0336	The COMBINING LONG STROKE OVERLAY is the appropriate character for such abbreviations. Smart font technology is needed to position the overlay at the right height and to supply a glyph with an overlay of the proper length. There is also a COMBINING SHORT STROKE OVERLAY at U+0335. I suggest standardizing on one overlay for abbreviations and using font technology to get the desired appearance.
----	------	--

## II. Numerals

Roman numerals should normally be represented by the standard Roman letters I, V, X, L, D, C, and M. The Roman numeral forms in Unicode that duplicate the standard Latin letters (U+2160–217F) were put there for use in East Asian text (often printed vertically) and should not be used in Latin script.

◌ <sup>¯</sup>	0305	Roman numerals are barred above to indicate multiplication by 1,000, when used as a prefix on words (IIviri = duumviri), when used as ordinal adverbs, and sometimes for other reasons. In all these cases, use U+0305 COMBINING OVERLINE; see I.G above
	007C	A bar above and a vertical line on either side of a numeral indicate multiplication by 100,000. Use U+007C VERTICAL BAR for this purpose. Smart font technology may be needed to get the corners of the overline and the vertical bar lined up right.
ϸ	missing	This form of the numeral six is missing and <b>should probably be proposed</b> , unless it is just too rare. It is found most often in Christian inscriptions.
Ɔ	glyph var.	The Romans wrote the numeral for 500 with a horizontal bar. I <b>suggest treating this as a glyph variant of the letter D</b> , since standard Latin letters are normally used for Roman numerals. This also makes it easy to search for (users don't have to enter a special codepoint). It could, however, be encoded as D followed by 0336, COMBINING LONG STROKE OVERLAY.
↯, ↰, ⊥	glyph var.	For similar reasons, I <b>suggest treating the early forms of the numeral fifty (downward pointing arrow, inverted T, anchor shape) as glyph variants of L</b> .
∞	glyph var.?	The Romans never wrote 1000 as the letter M except in some abbreviations (MP = <i>milia passuum</i> , e.g.). The numeral originally looked like a circle with a vertical line through its center, later like a modern infinity sign. <b>As suggested above in the case of 500, one might argue for treating this as a glyph variant of M. However, the shape is so totally different that a new character might be proposed. But there would then be two different ways of representing the Roman numeral 1000 in Unicode. A better way might be to handle it as a glyph variant of U+2180 (see below), from which the infinity-shape may have evolved.</b> The mathematical infinity sign in Unicode U+221E is not well suited for use as a Roman numeral.
ↀ	2180	ROMAN NUMERAL ONE THOUSAND CD
ↁ	2181	ROMAN NUMERAL FIVE THOUSAND
ↂ	2182	ROMAN NUMERAL TEN THOUSAND

↷ 2183

ROMAN NUMERAL REVERSED ONE HUNDRED

Unicode states that the letter C can be used with ↷ 2183 and the letter I to represent larger values; for example, GORDON. Alternate glyphs may be needed for good appearance in such situations.

### III. Monetary Signs

NB: currency property

⌘ missing

The denarius sign is missing and **should be proposed**. The GREEK DENARIUS SIGN is found at ????, but should not be used for Latin because of the different shape and the general practice of not mixing scripts.

Ϝ missing

The quinarius sign is missing. **Is it common enough to propose?**

HS missing

The sestertius sign is missing and **should be proposed**. Glyph variants include IS, SS, S, others?


/ missing

The Vindolanda tablets have an abbreviation for the *as* (slash with a horizontal tick). I have not seen this elsewhere; **should it be proposed? I think a simple vertical line is sometimes used for the *as*; check this.**


missing

**Are there subdivisions of the *as* that should be proposed?** Unciae are marked with a horizontal line; two unciae are shown as two lines, one above the other. How to show this in Unicode?? An em-dash might work for one uncia.

#### IV. Other characters

	missing	The centurio sign can stand for several forms (centurio, centurioni, centurione, centuriae, others?). It <b>should be proposed</b> for Unicode. Glyph variants include a reversed C, the letter O, and an angle bracket shape. Sandys also mentions a 7-shape and the letter Z (p. 228), although I have not seen these. <b>Which is most common and so should be proposed as the reference shape?</b>
Ϸ	2627	CHI RHO

#### V. Punctuation

·	00B7	MIDDLE DOT should be used for the interpunct. A horizontal wedge-shape is a common glyph variant and should be handled as such rather than proposing a new character. <b>Or, since hederæ are also used as interpuncts (esp. in later inscriptions) and is already separately encoded, we could encode a wedge-shaped interpunct.</b>
⬮	2767	ROTATED FLORAL HEART BULLET is specified in Unicode as the hederæ. U+2619 REVERSED ROTATED FLORAL HEART BULLET ⬮ is also available if needed.
	missing	The palm branch, occasionally found to mark divisions of text and as a decoration on coins, is missing and <b>should be encoded</b> . MUFI spec has it in the PUA at U+F1DB.

## VI. Modern Editorial Signs

*	002A	
[ ]	005B/005D	
{ }	007B/007D	
†	2020	
‡	2021	
◌̇	0323	Unicode contains a large number of precomposed letters with underdot in the Latin Extended Additional range, including all the Latin alphabet. Any additional characters with underdot that may be required should use combining underdot, U+0323, in which case smart font technology will be needed for good appearance.
< >	2329/2A	LEFT-POINTING ANGLE BRACKET / RIGHT-POINTING ANGLE BRACKET. Do not use the similar-looking U+3008/09, which are specific to East Asian text and are normally double-width, or U+27E8/E9, which are for mathematics.
[[ ]]	27E6/E7	
⊂ ⊃	missing	The sideways-pointing U shaped brackets used in CIL to denote Claudian letters, inverted letters, or images are missing and <b>should be proposed</b> . The MUFI spec has them in the PUA at F1DC/DD. The SUBSET OF and SUPerset OF characters (U+2282/83) are mathematical operators and so should not be used; in addition the legs of the glyphs are too short.
(( ))	missing	Some editors use double parentheses for similar purposes similar to the sideways U brackets. These also are missing and <b>should be proposed</b> . The MUFI spec has them in the PUA at F1DE/DF. <sup>†</sup>

---

<sup>†</sup> The MUFI group is preparing a proposal for medievalist and Latinist punctuation, the first draft of which contains both the sideways U brackets and the double parentheses.