Undeciphered Scripts in the Unicode Age

Challenges for encoding early writing systems of the Near East

Internationalization and Unicode Conference #42 • September 2018

Dr. Anshuman Pandey

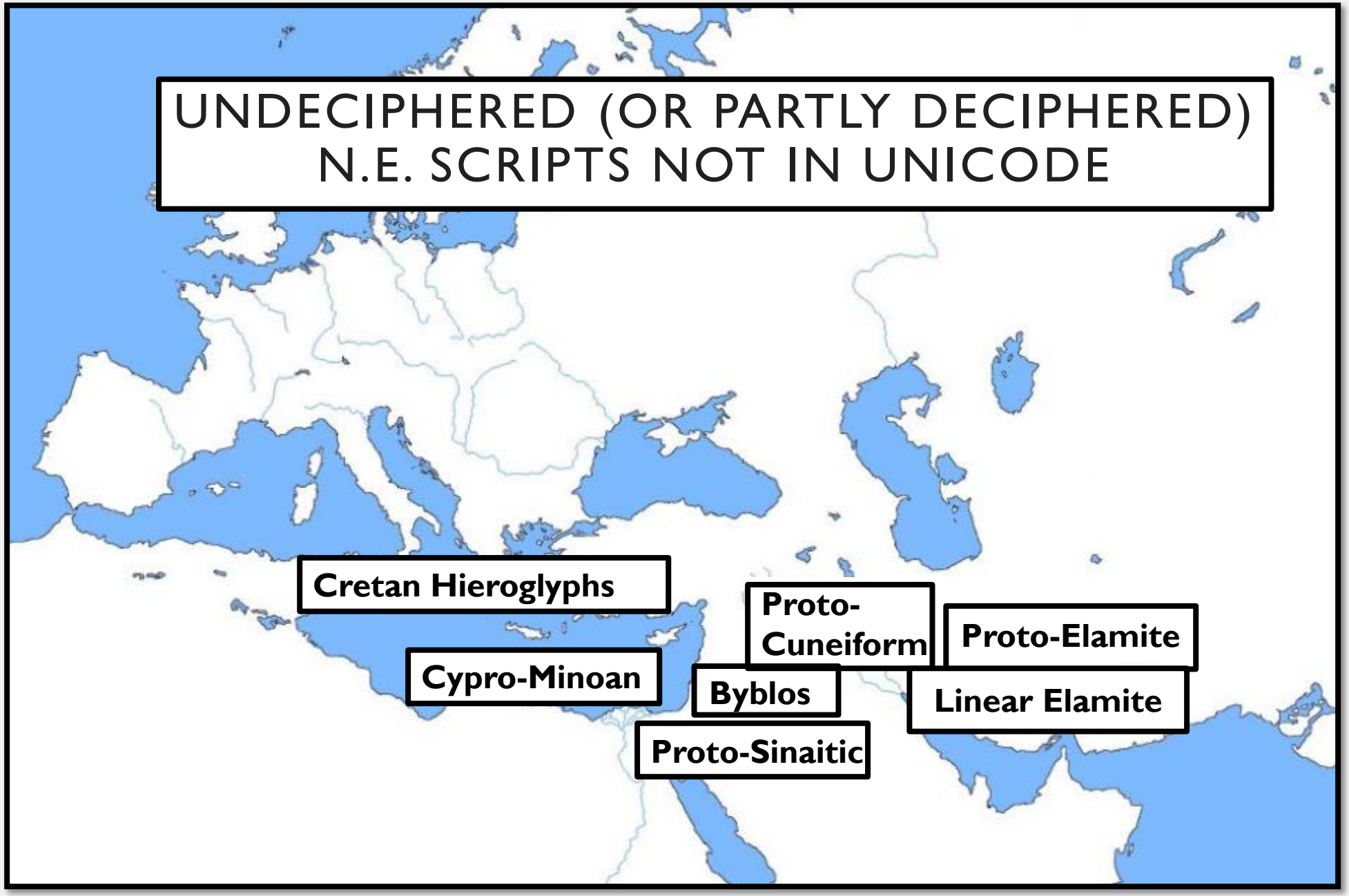
Dr. Deborah (Debbie) Anderson

Script Encoding Initiative, UC Berkeley

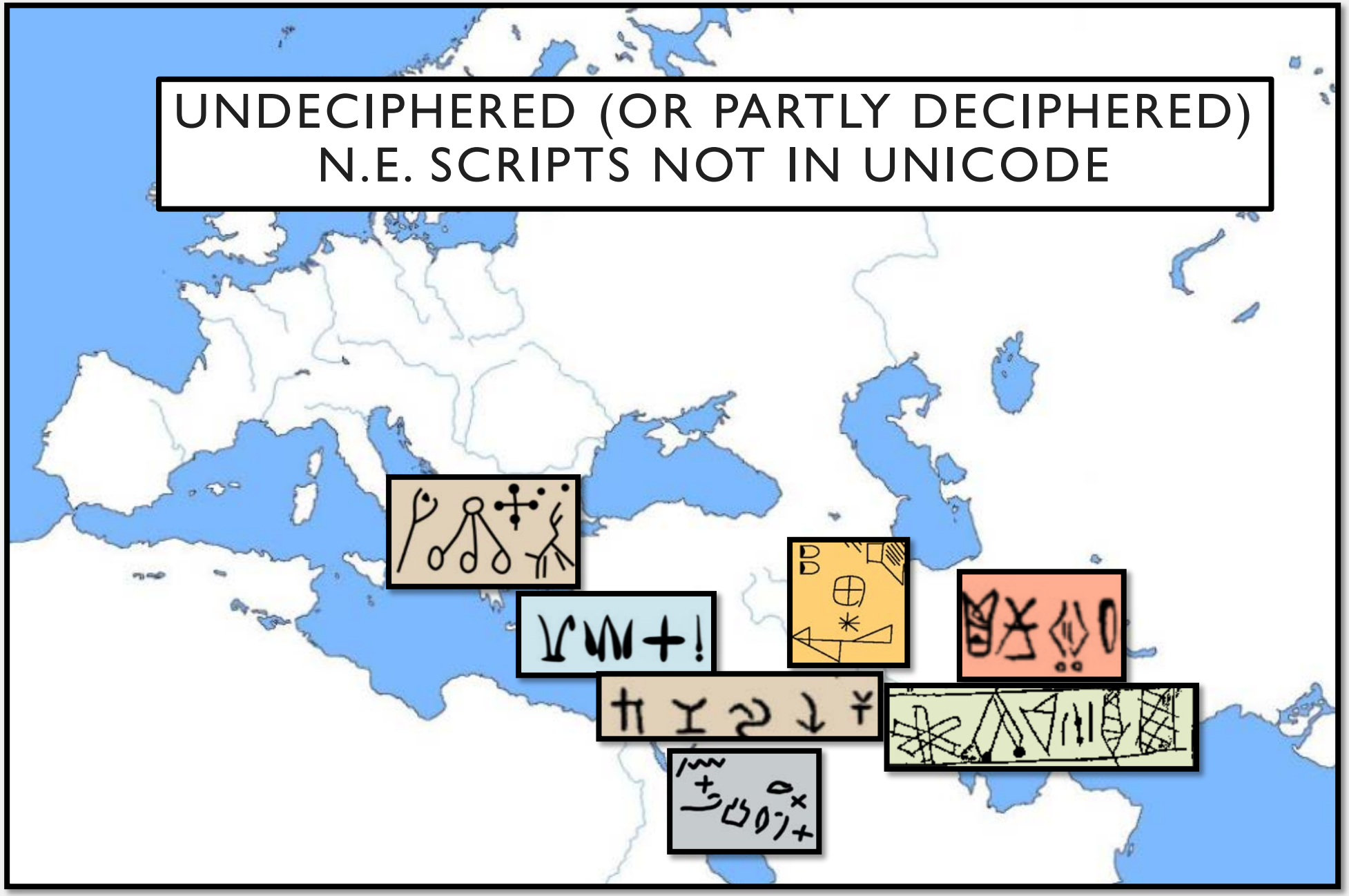
WELL-KNOWN UNDECIPHERED SCRIPTS (OUTSIDE THE NEAR EAST)

UNDECIPHERED (OR PARTLY DECIPHERED) N.E. SCRIPTS NOT IN UNICODE

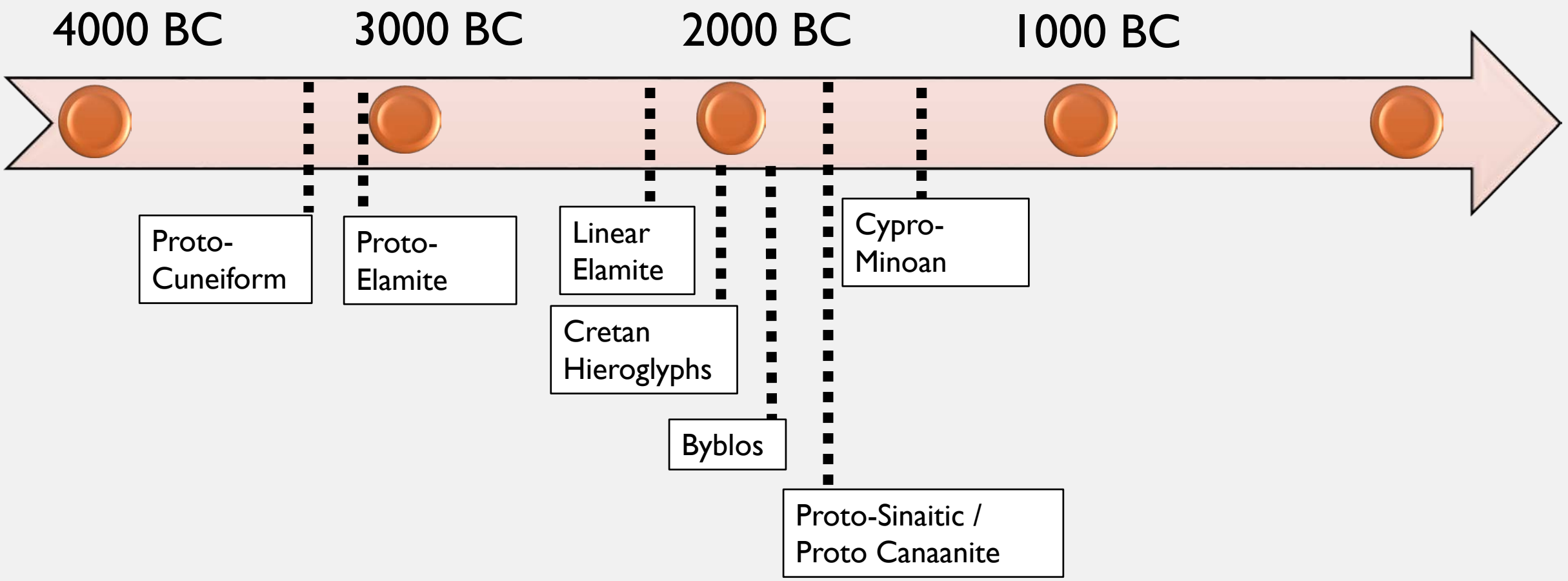
1. Proto-Cuneiform
2. Proto-Elamite
3. Linear Elamite
4. Cretan Hieroglyphs
5. Byblos
6. Proto-Sinaitic
7. Cypro-Minoan


UNDECIPHERED (OR PARTLY DECIPHERED)
N.E. SCRIPTS NOT IN UNICODE
Cretan Hieroglyphs
Cypro-Minoan
Byblos
Proto-Sinaitic
Proto-Cuneiform
Proto-Elamite
Linear Elamite

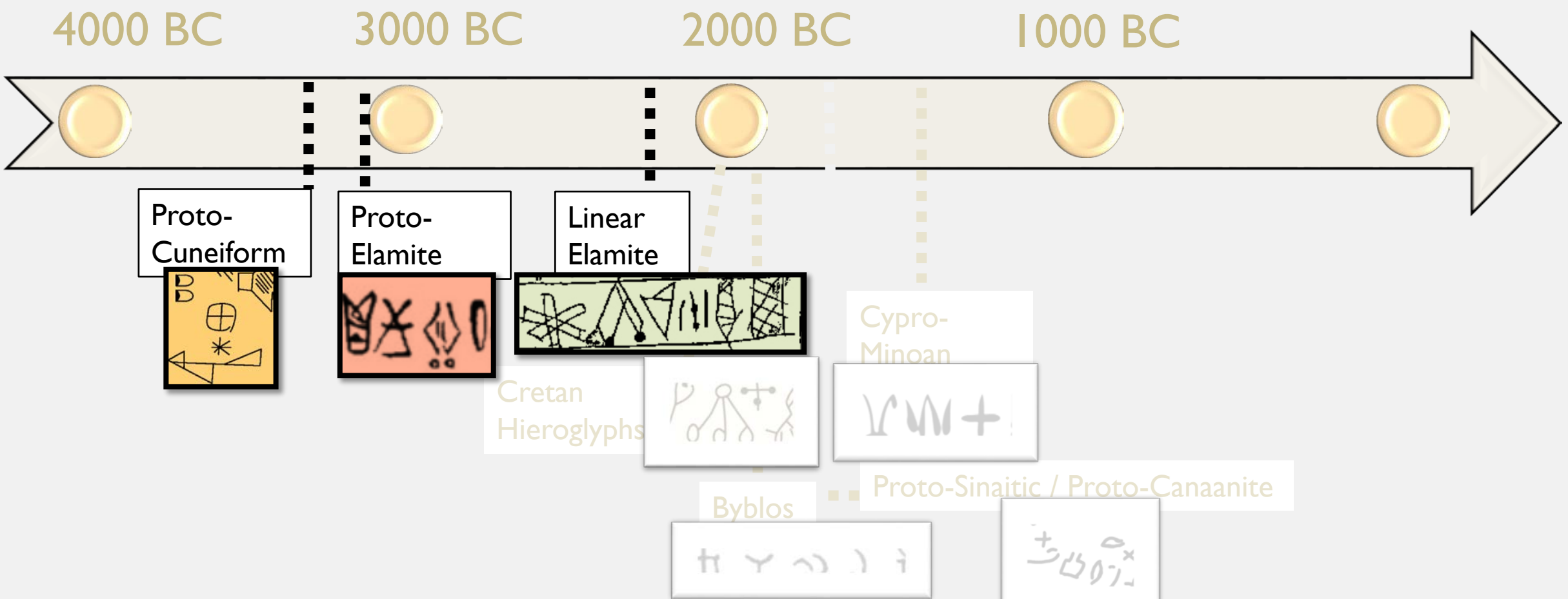
UNDECIPHERED (OR PARTLY DECIPHERED) N.E. SCRIPTS NOT IN UNICODE



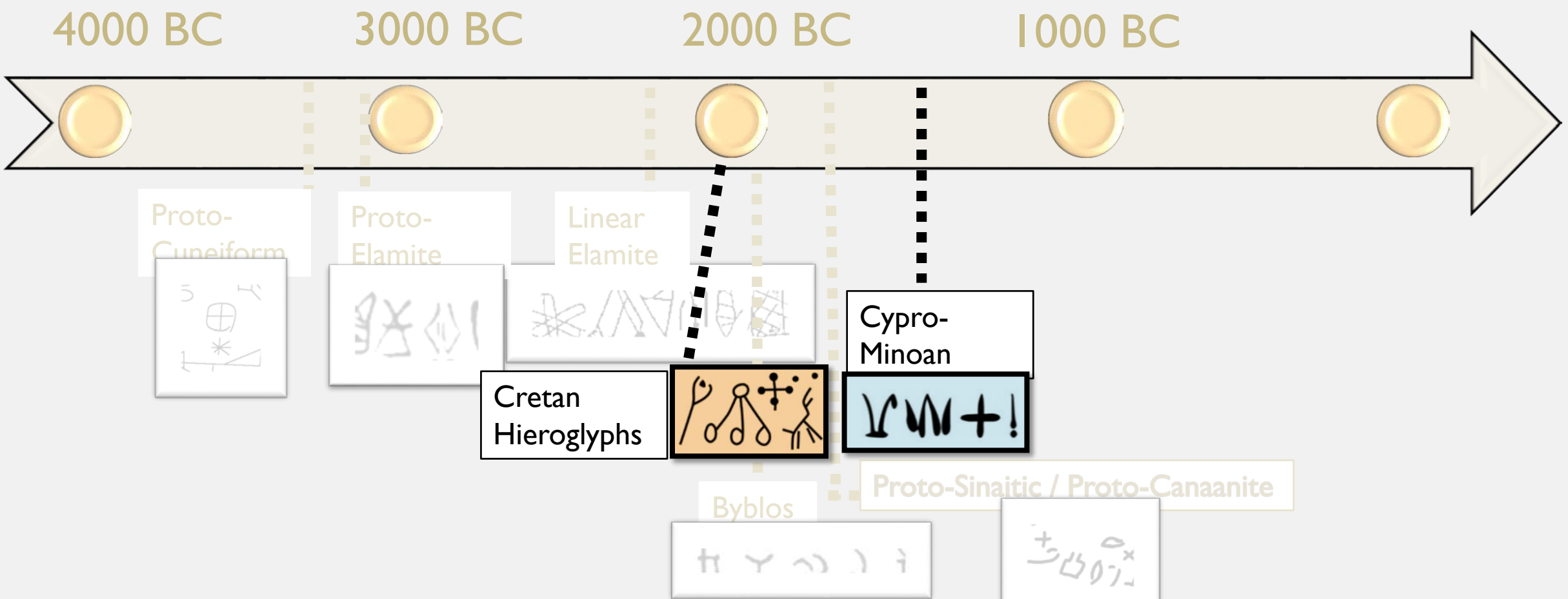
TIMELINE OF UNDECIPHERED/PARTLY DECIPHERED SCRIPTS OF THE NEAR EAST (3200-1500 BC)



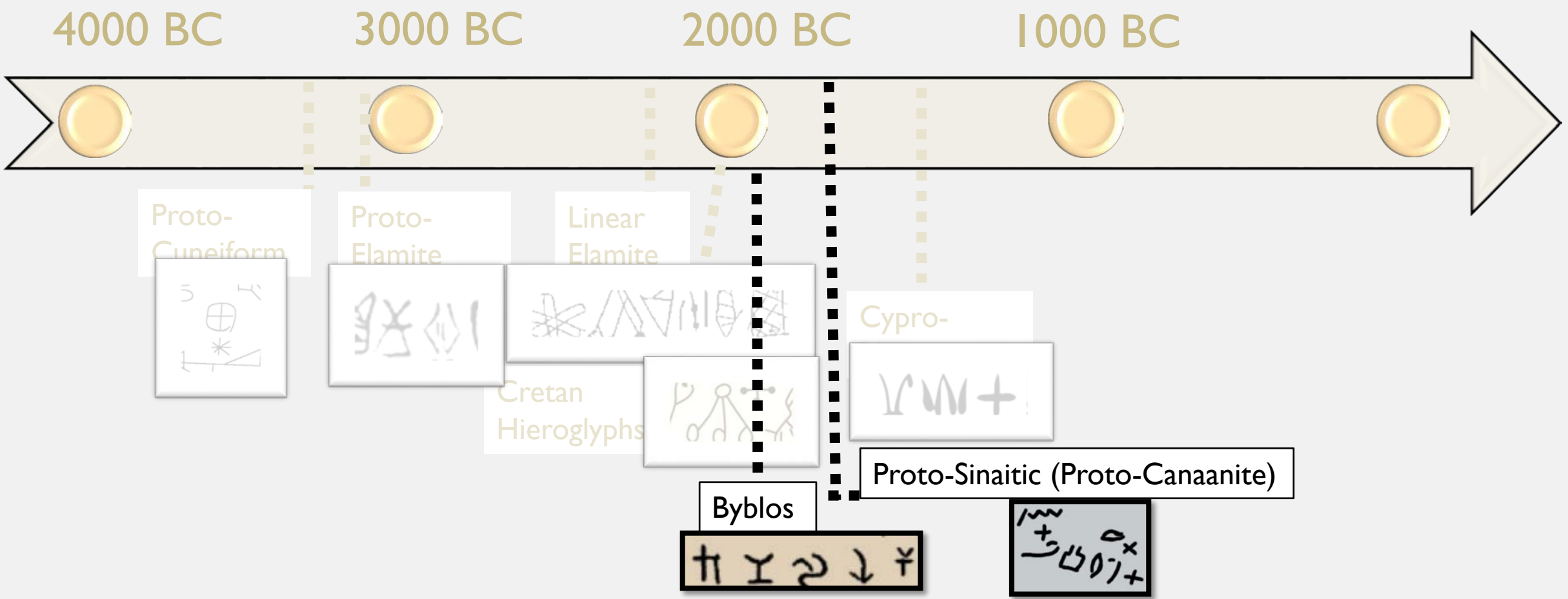
TIMELINE OF UNDECIPHERED/PARTLY DECIPHERED SCRIPTS OF THE NEAR EAST



TIMELINE OF UNDECIPHERED/PARTLY DECIPHERED SCRIPTS OF THE NEAR EAST



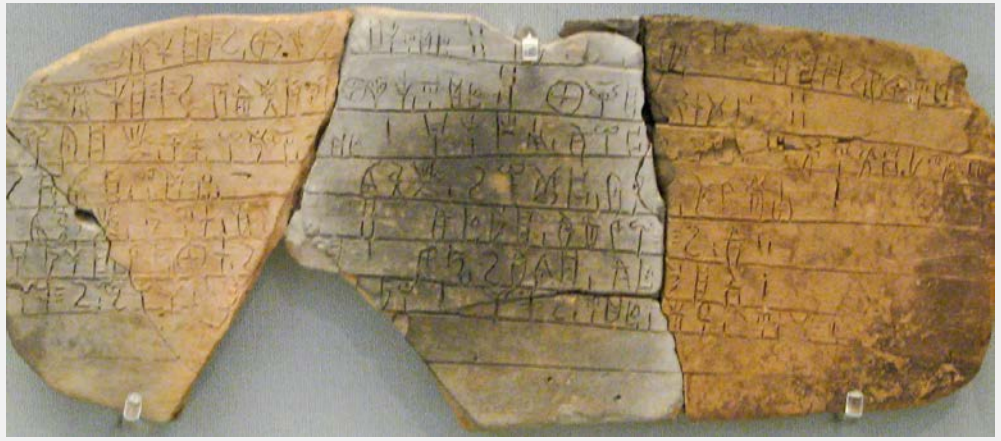
TIMELINE OF UNDECIPHERED/PARTLY DECIPHERED SCRIPTS OF THE NEAR EAST



KEYS TO DECIPHERMENT

- Corpus large enough? Can the number of signs be determined?

EXAMPLE: LINEAR B
Over 5000 tablets with inscriptions; 85-90 distinct signs



KEYS TO DECIPHERMENT

- Known relationship(s) to other scripts?

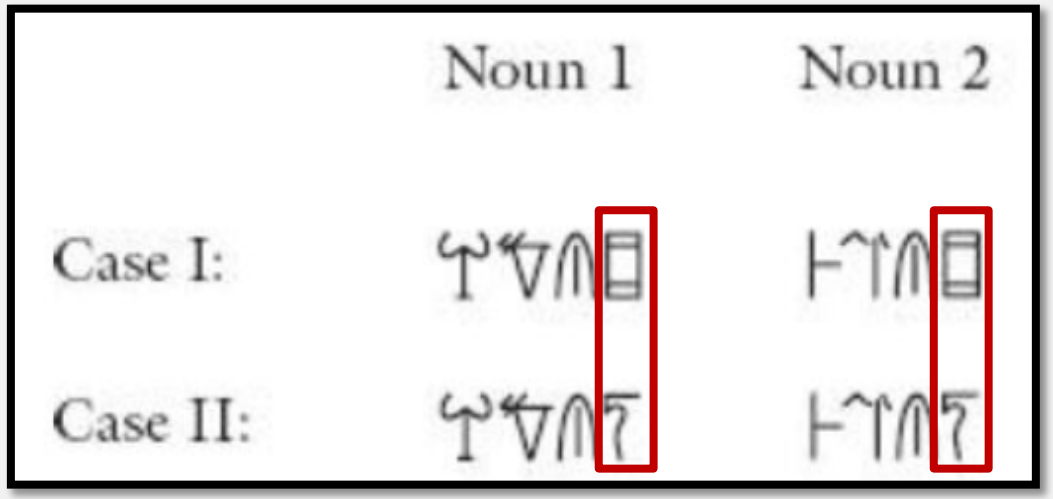
EXAMPLE: LINEAR B
Shown to have some relationship to the Cypriot Syllabary, which was deciphered earlier

Cypriot Syllabary

KEYS TO DECIPHERMENT

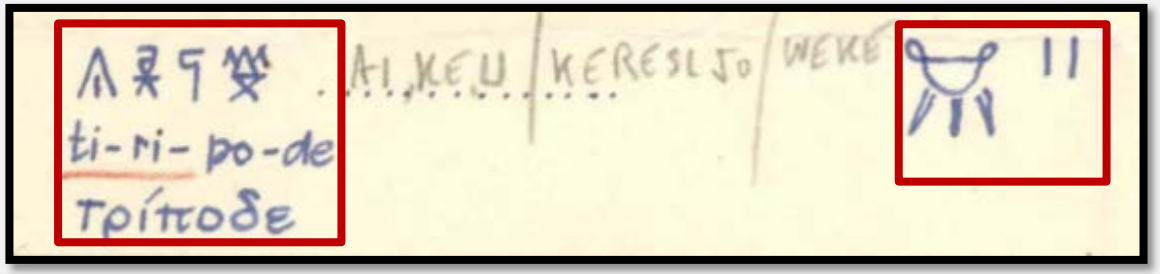
- Underlying language (or language family) known?

EXAMPLE: LINEAR B
Nouns varied only by 1 sign, suggesting the language was inflected (and led to identifying it as an Indo-European lang.)

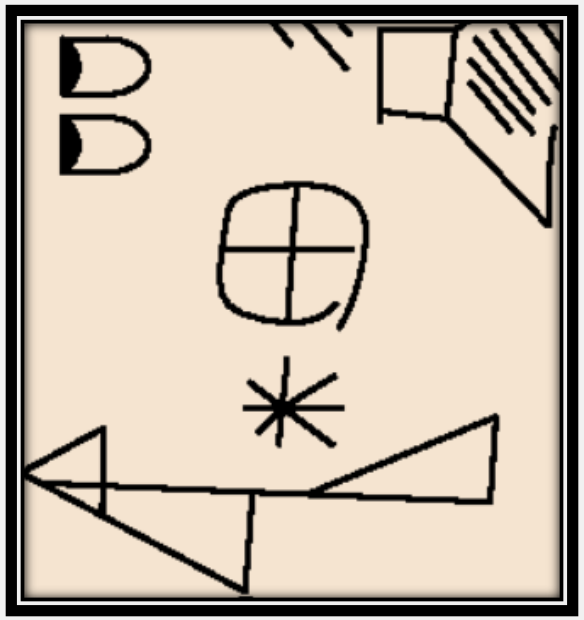


KEYS TO DECIPHERMENT

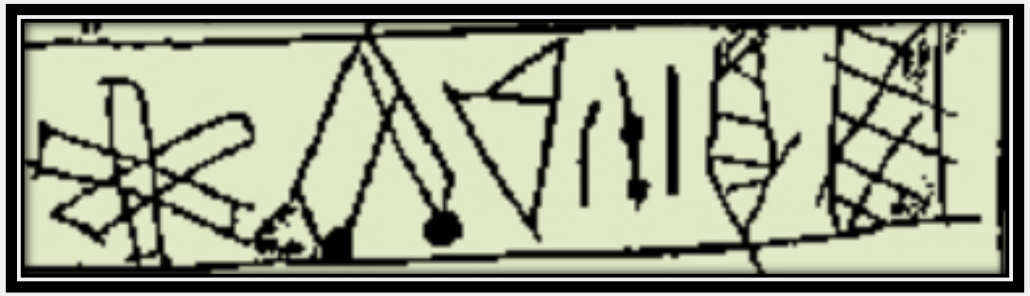
- Bilingual available? (Or are there other ways to be able to confirm a reading?)



Letter from Michael Ventris to E. Bennett, May 1953



Proto-Cuneiform



Linear Elamite

COMMENTS ON N.E. UNDECIPHERED SCRIPTS

1. Scripts in this talk reflect a spectrum, ranging from partially deciphered to completely undeciphered.

2. May attract wide-ranging, unusual theories

ENCODING PROCESS

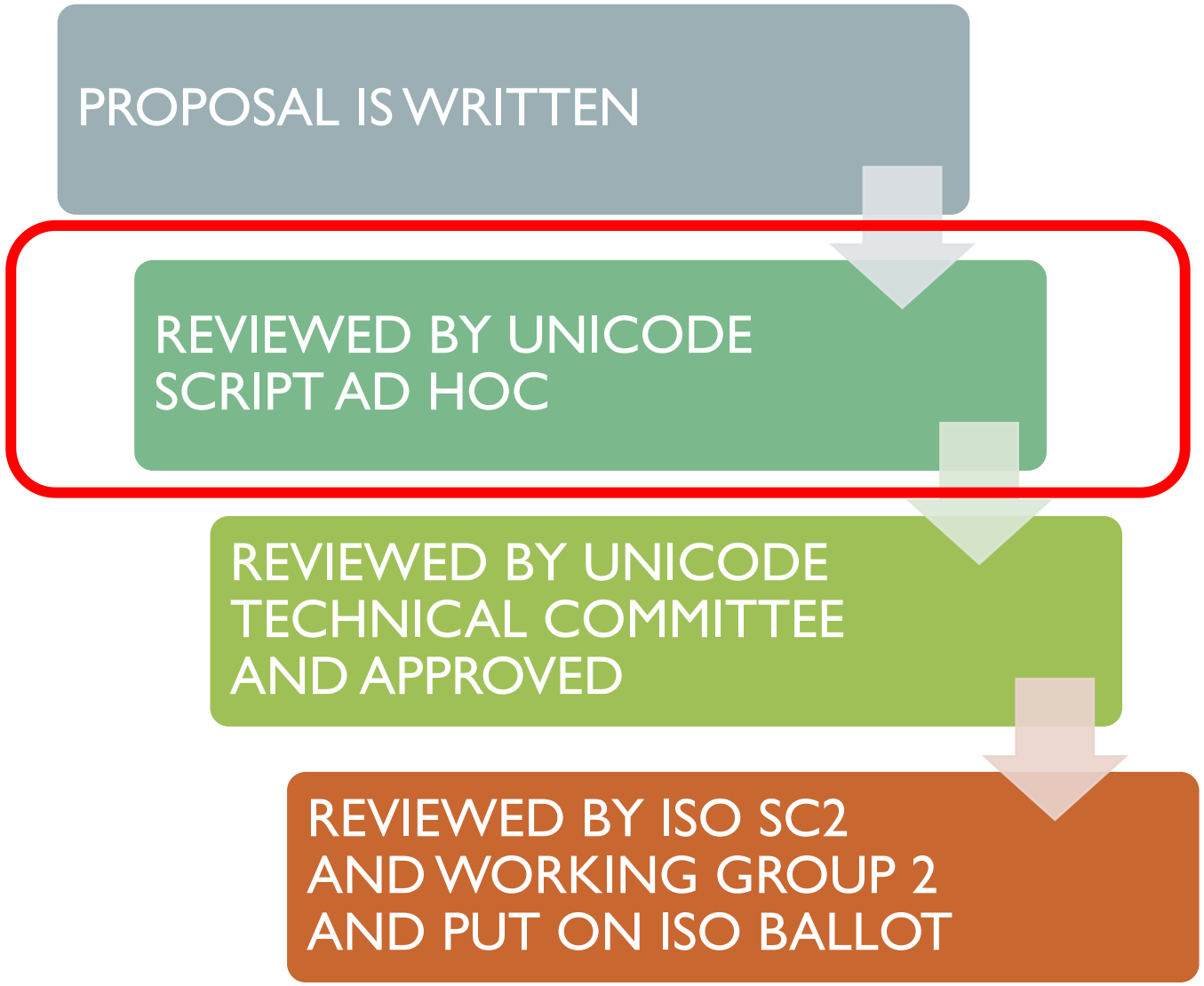
PROPOSAL IS WRITTEN

REVIEWED BY UNICODE SCRIPT AD HOC

REVIEWED BY UNICODE TECHNICAL COMMITTEE AND APPROVED

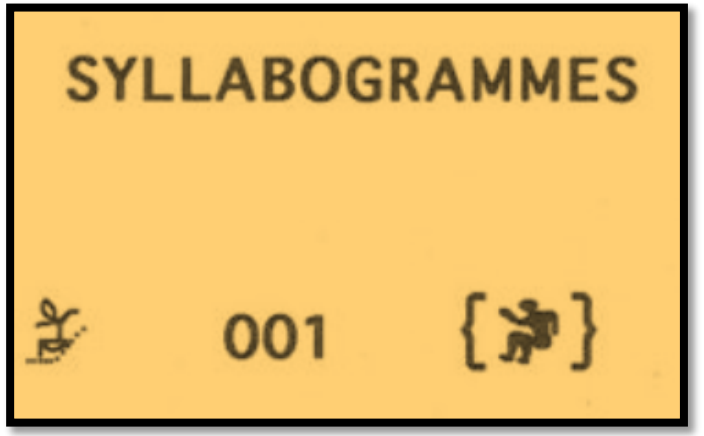
REVIEWED BY ISO SC2 AND WORKING GROUP 2 AND PUT ON ISO BALLOT

ENCODING PROCESS



FACTORS TO CONSIDER RE: UNDECIPHERED SCRIPTS (FROM THE SCRIPT AD HOC)

- Does the script have a stable list of the characters that scholars refer to?



From CHIC = Corpus Hieroglyphicarum Inscriptionum Cretae (Godart and Olivier)

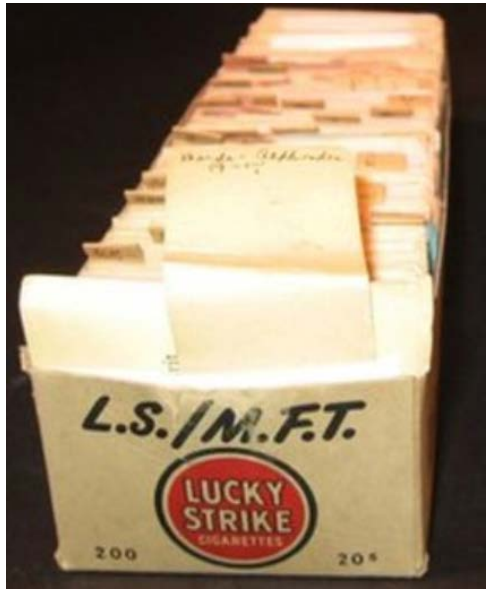
FACTORS TO CONSIDER RE: UNDECIPHERED SCRIPTS (FROM THE SCRIPT AD HOC)

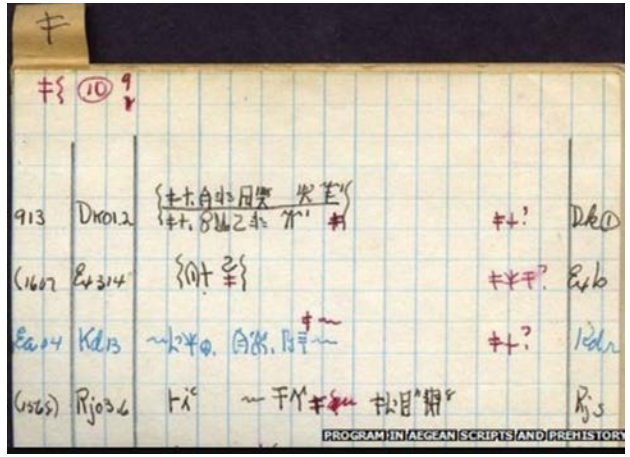
- How much material in the script exists today?

Proto-Sinaitic stele

FACTORS TO CONSIDER RE: UNDECIPHERED SCRIPTS (FROM THE SCRIPT AD HOC)

- What is the state of decipherment?
- Is the underlying language known?

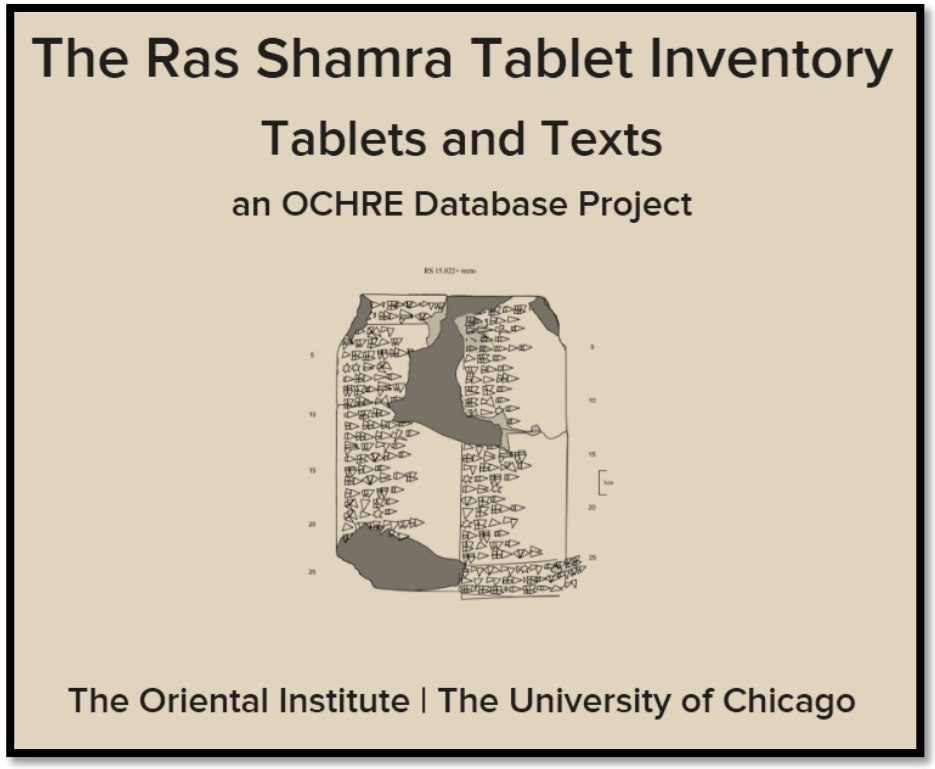




Alice Kober's files for Linear B

FACTORS TO CONSIDER RE: UNDECIPHERED SCRIPTS (FROM THE SCRIPT AD HOC)

- Can a strong case be made to encode the script? Is text in script being interchanged?



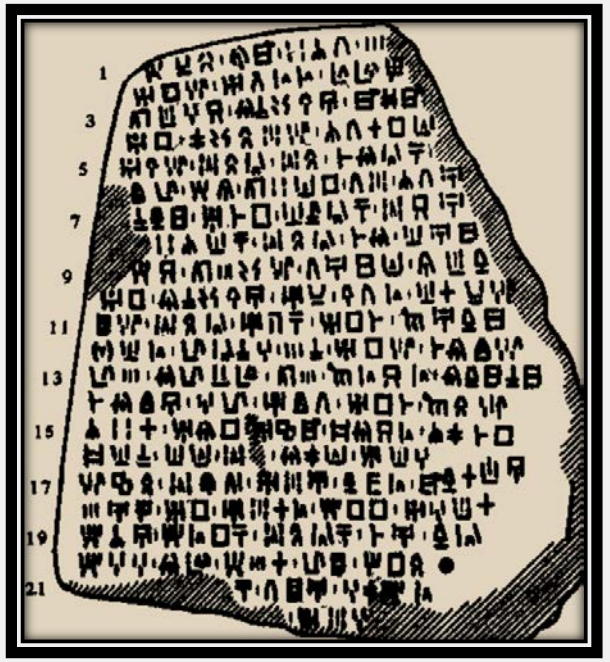
CYPRO-MINOAN

recto

recto

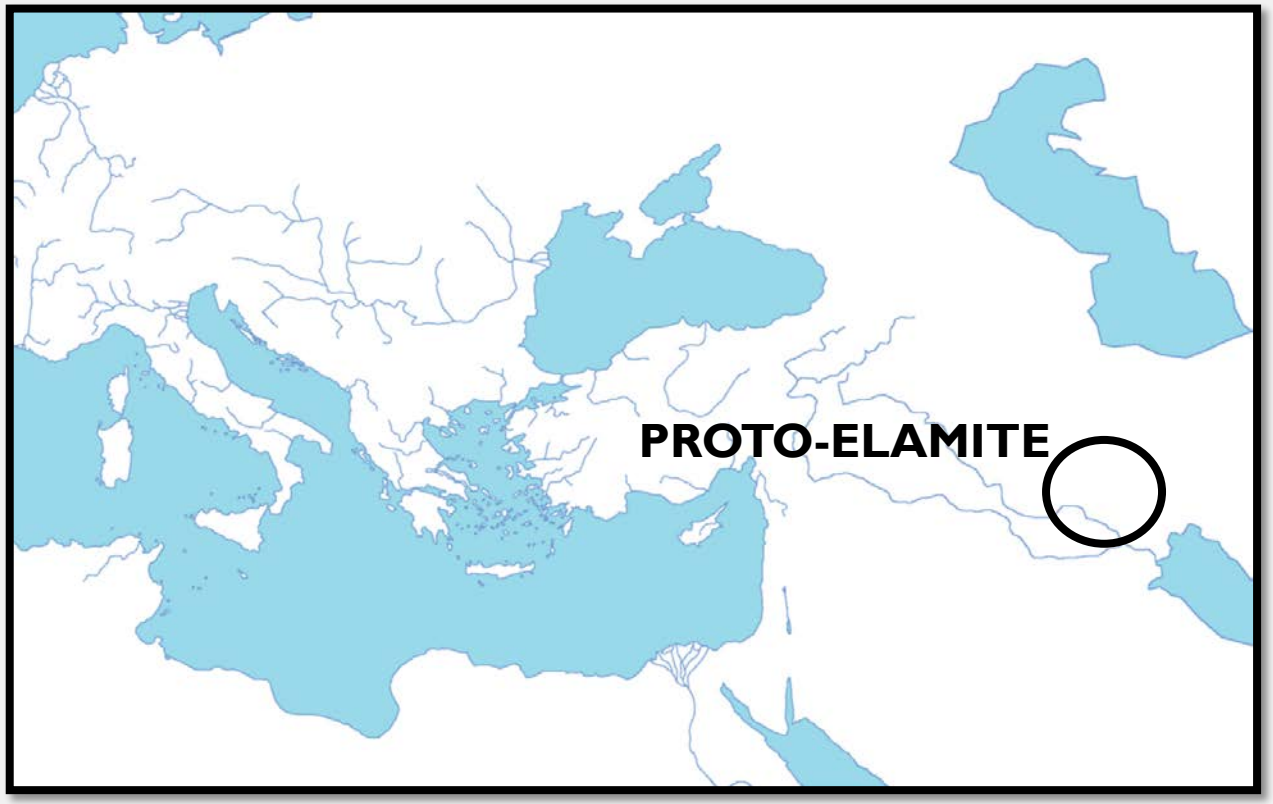
CYPRO-MINOAN

- Found on ca. 250 objects
- Current proposal is stalled:
  - Some characters in proposal are not regarded today by scholars as valid
  - Apparent duplicates in repertoire (from Enkomi tablet)

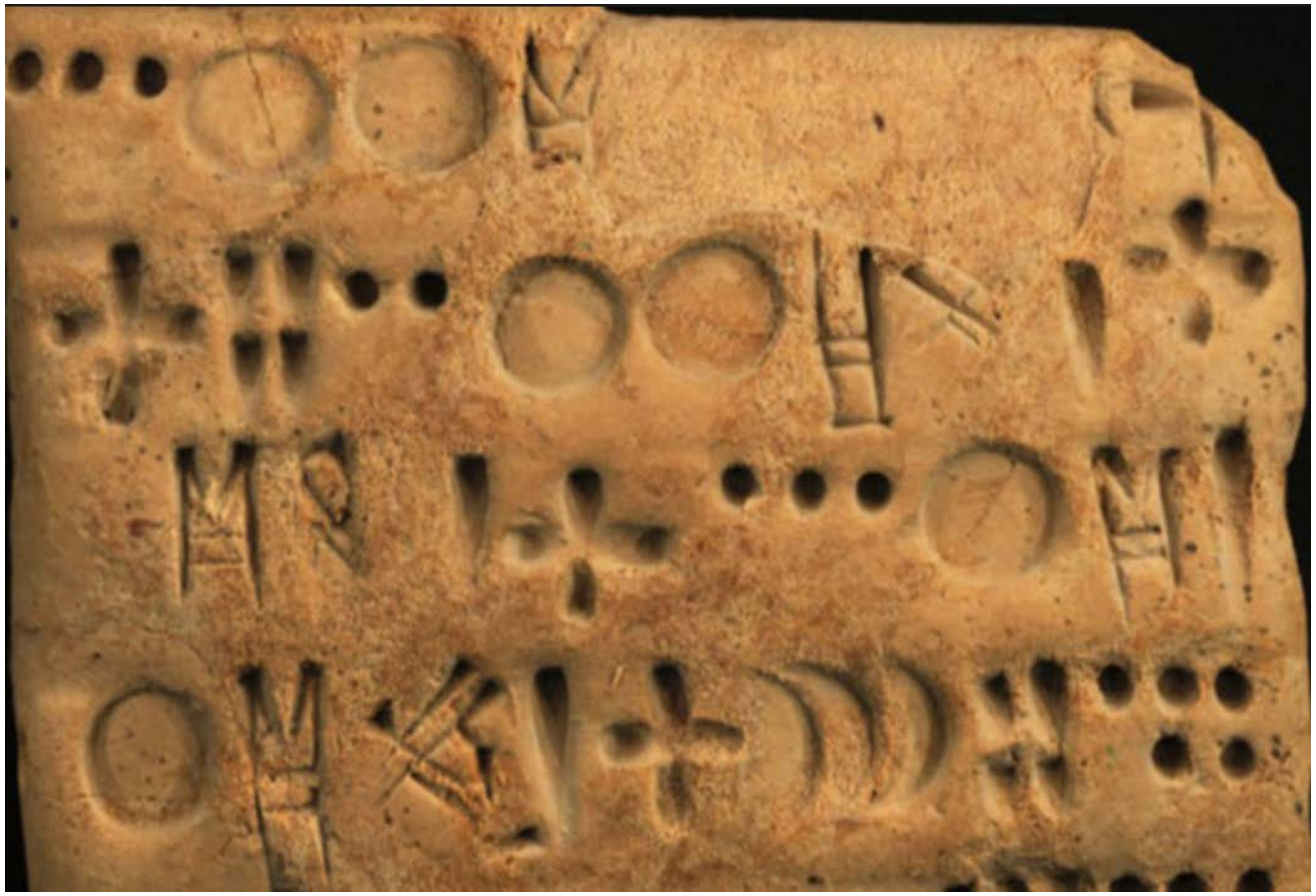


Enkomi tablet ENKO Atab 001

PROTO-ELAMITE



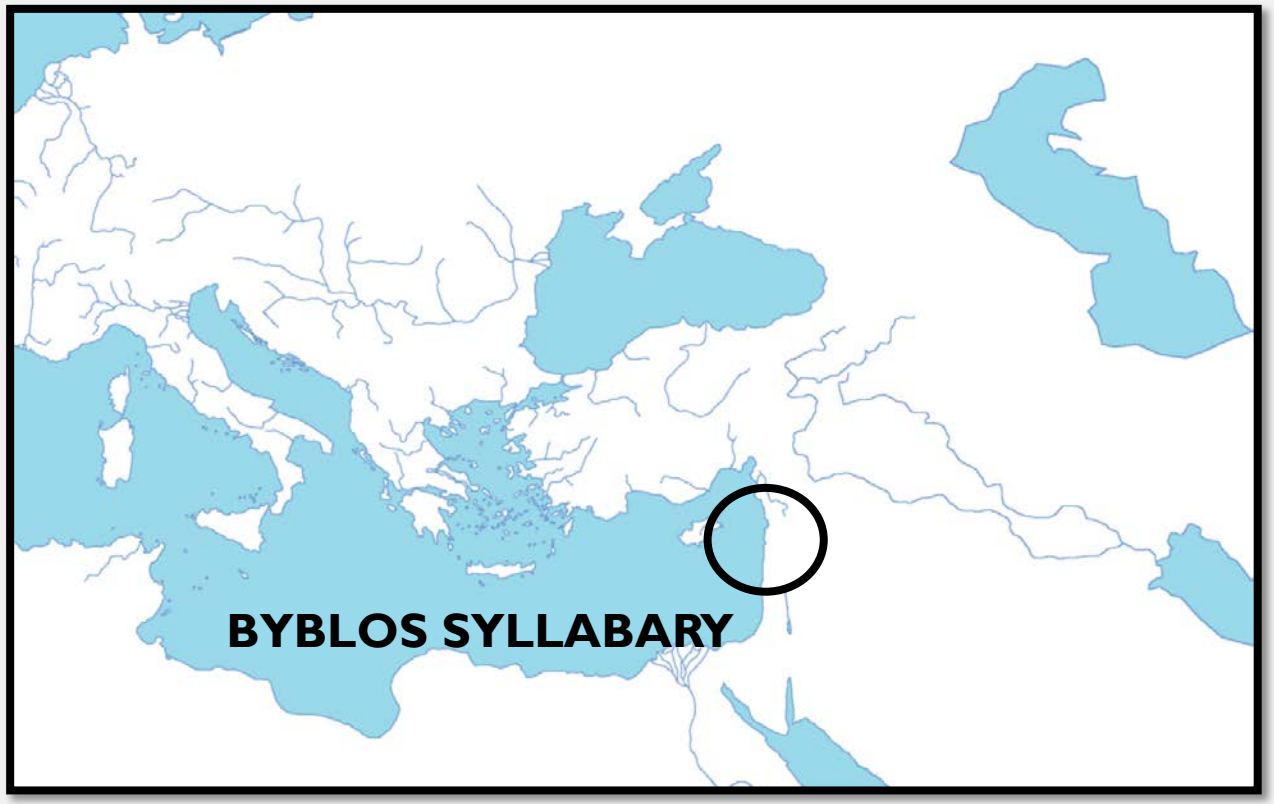
- Found on over 1600 tablets, most from Susa, in SW Iran
- A short-lived writing system (ca. 3100-2900 BC)



PROTO-ELAMITE

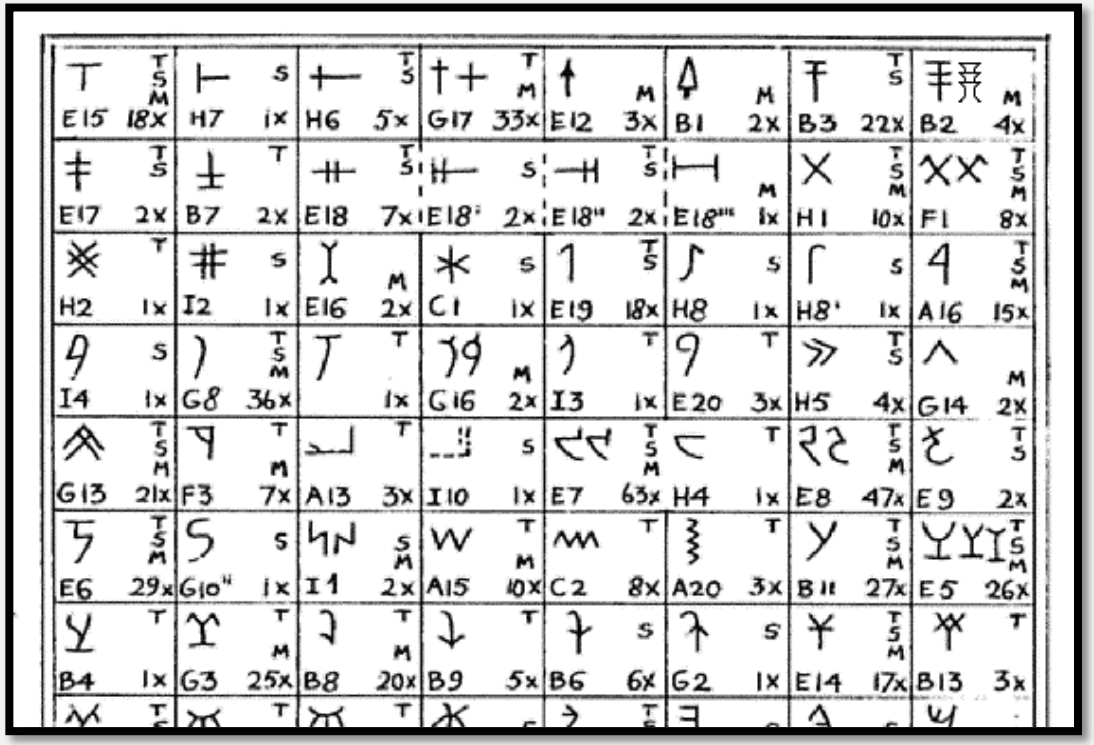
- Closed set of characters (300-400)
- Similar numerical system to Proto-Cuneiform
- New texts will be available from Tehran soon

BYBLOS SYLLABARY



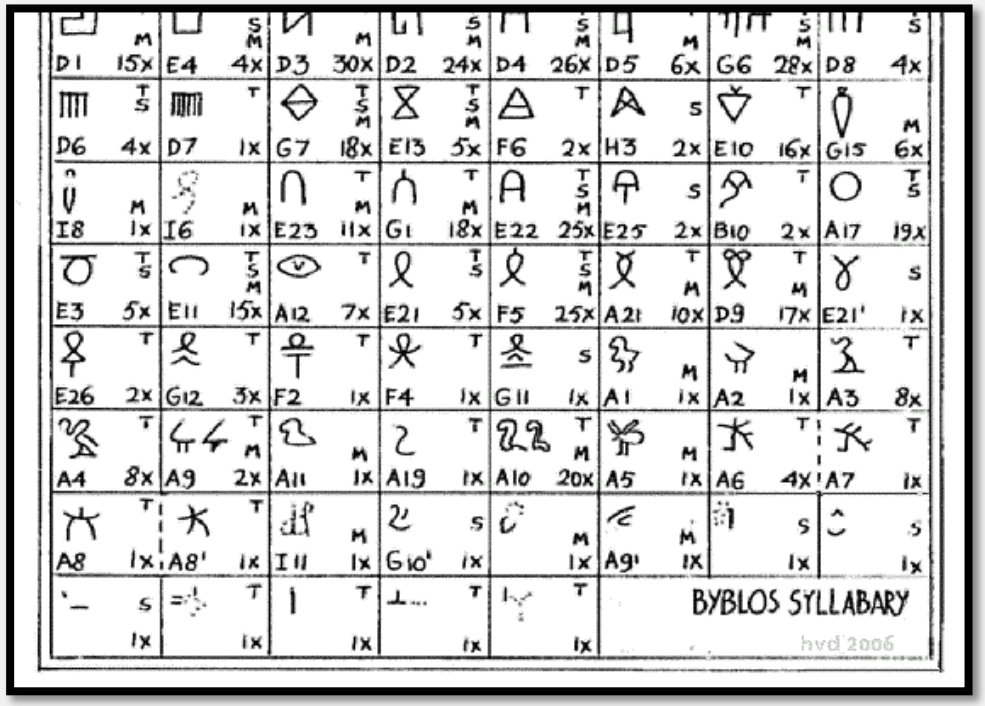
- Byblos (modern Lebanon)
  - Other parts of Mediterranean
- 18th-15th c. BCE
- 10 extant records
- Origins unknown
  - Egyptian hieratic script?
- Syllabic script

BYBLOS SYLLABARY



- Structure
  - signs represent Semitic CV syllables
- Directionality
  - believed to be left to right

BYBLOS SYLLABARY

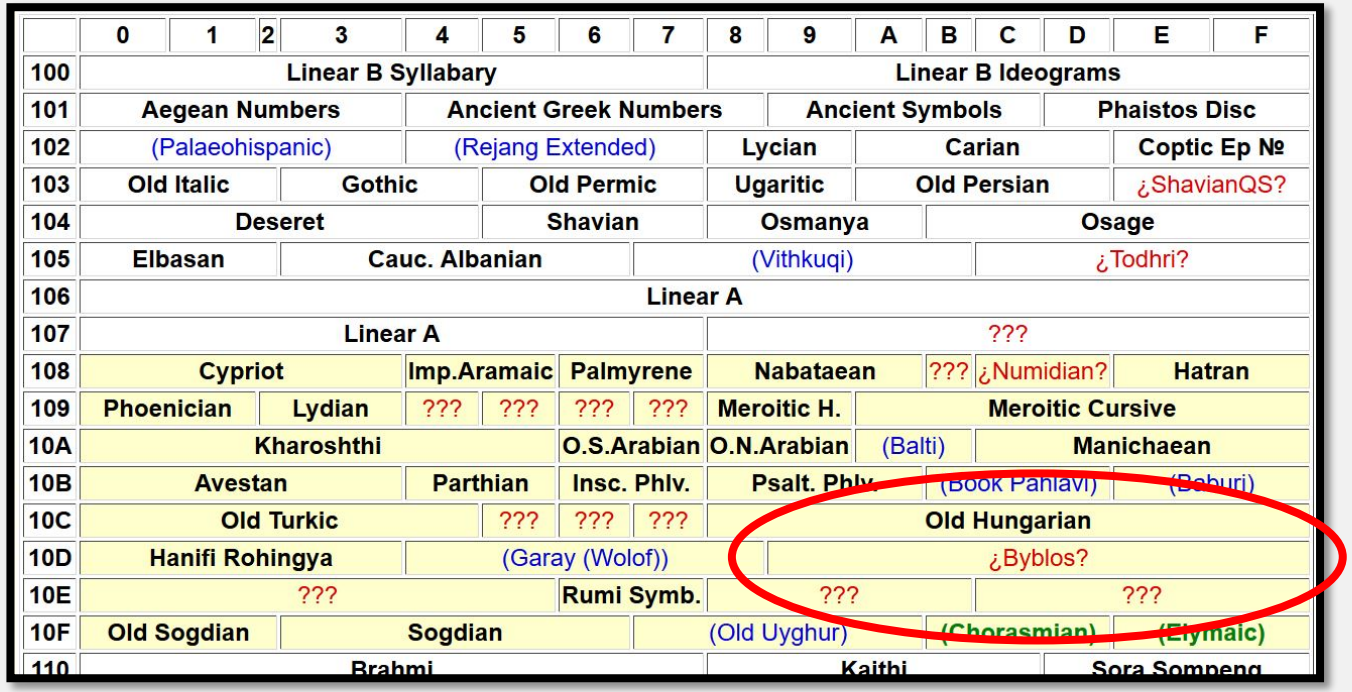


- Repertoire
  - ~1050 characters in corpus
  - ~90 to ~120 distinctive signs
  - No number signs
  - No punctuation

BYBLOS SYLLABARY

- Decipherment status
  - No consensus on repertoire
    - Variants vs. distinctive signs?

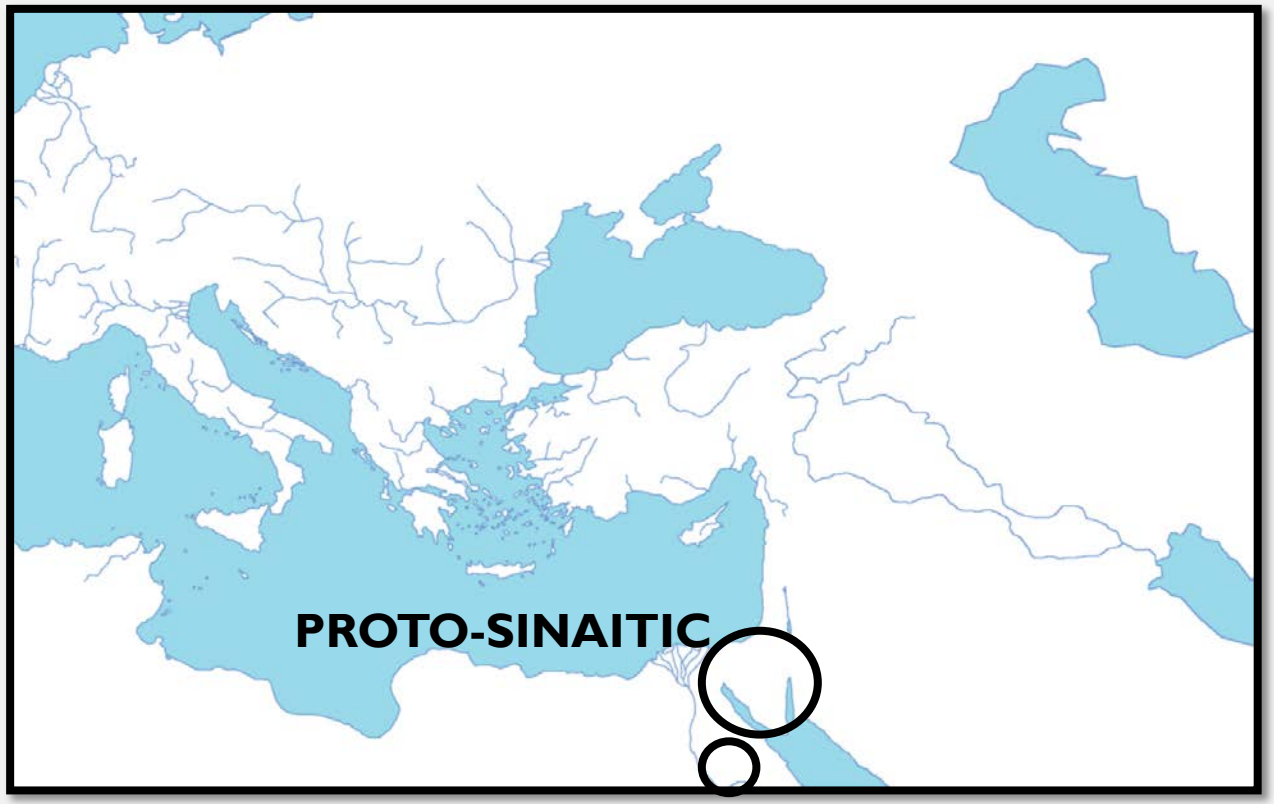
BYBLOS SYLLABARY



	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
100	Linear B Syllabary								Linear B Ideograms							
101	Aegean Numbers				Ancient Greek Numbers				Ancient Symbols				Phaistos Disc			
102	(Palaeohispanic)				(Rejang Extended)				Lycian		Carian				Coptic Ep №	
103	Old Italic			Gothic			Old Permic		Ugaritic		Old Persian				¿ShavianQS?	
104	Deseret					Shavian			Osmanya				Osage			
105	Elbasan			Cauc. Albanian					(Vithkuqi)				¿Todhri?			
106	Linear A															
107	Linear A								???							
108	Cypriot				Imp.Aramaic		Palmyrene		Nabataean		???	¿Numidian?			Hatran	
109	Phoenician		Lydian		???	???	???	???	Meroitic H.		Meroitic Cursive					
10A	Kharoshthi						O.S.Arabian		O.N.Arabian		(Balti)		Manichaean			
10B	Avestan				Parthian		Insc. Phlv.		Psalt. Phlv.			(Book Pahlavi)			(Bohuri)	
10C	Old Turkic					???	???	???	Old Hungarian							
10D	Hanifi Rohingya				(Garay (Wolof))				¿Byblos?							
10E	???						Rumi Symb.		???				???			
10F	Old Sogdian			Sogdian					(Old Uyghur)			(Chorasmian)			(Elymaic)	
110	Brahmi								Kaithi						Sora Sompeng	

- Unicode status
  - Allocated to SMP roadmap
  - No proposal
- Challenges for encoding
  - Open repertoire
  - Character-glyph distinctions
  - Unconfirmed sign values

PROTO-SINAITIC



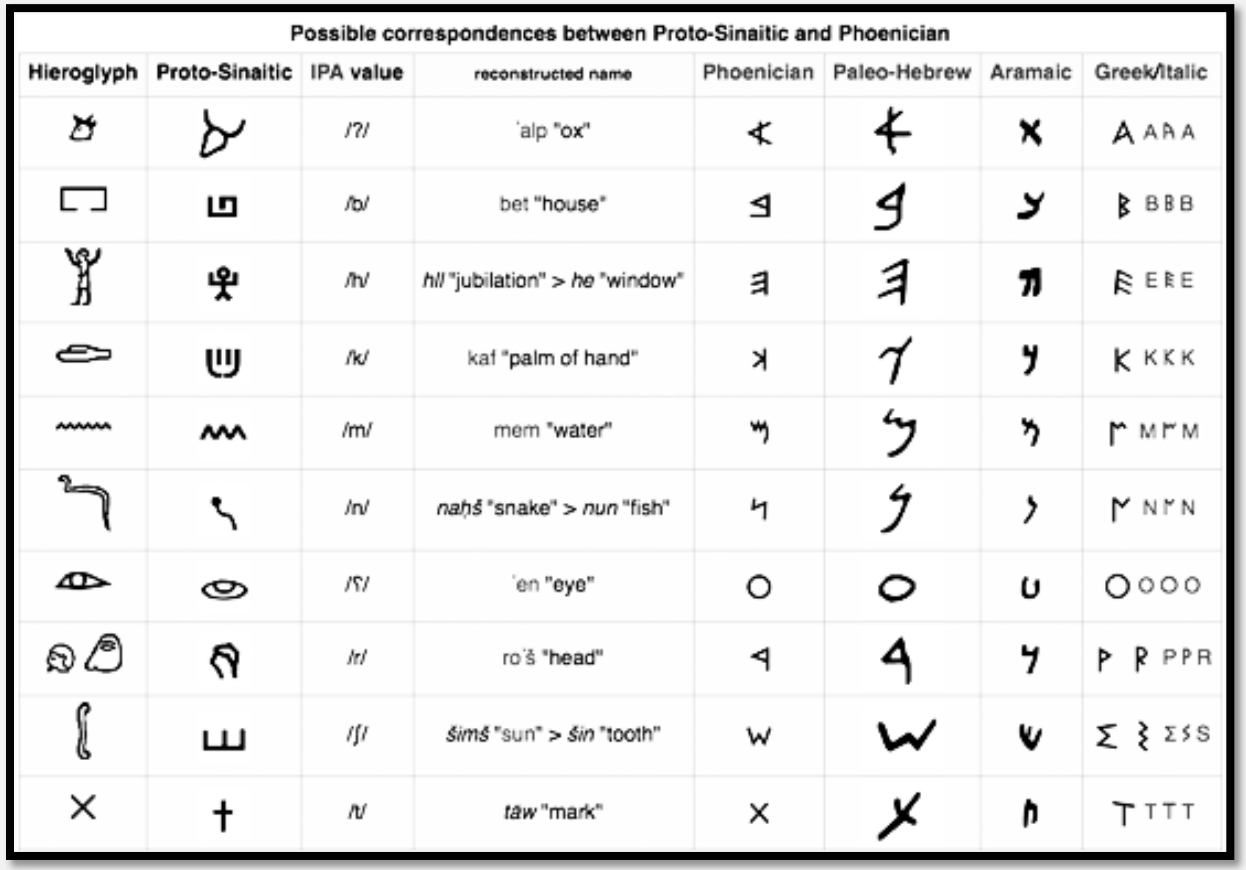
- Also known as 'Early Alphabetic'
- ~18th-17th c. BCE
- Inspired by Egyptian Hieroglyphs
- Supposed first alphabetic script

PROTO-SINAITIC

... *lb ʿlt* ('...to the Lady'), Gardiner 1916

- ~50 inscriptions
  - Serabit el-Khadim (Sinai), 17th c. BCE
  - Wadi el-Hol (Qena, Egypt)

PROTO-SINAITIC

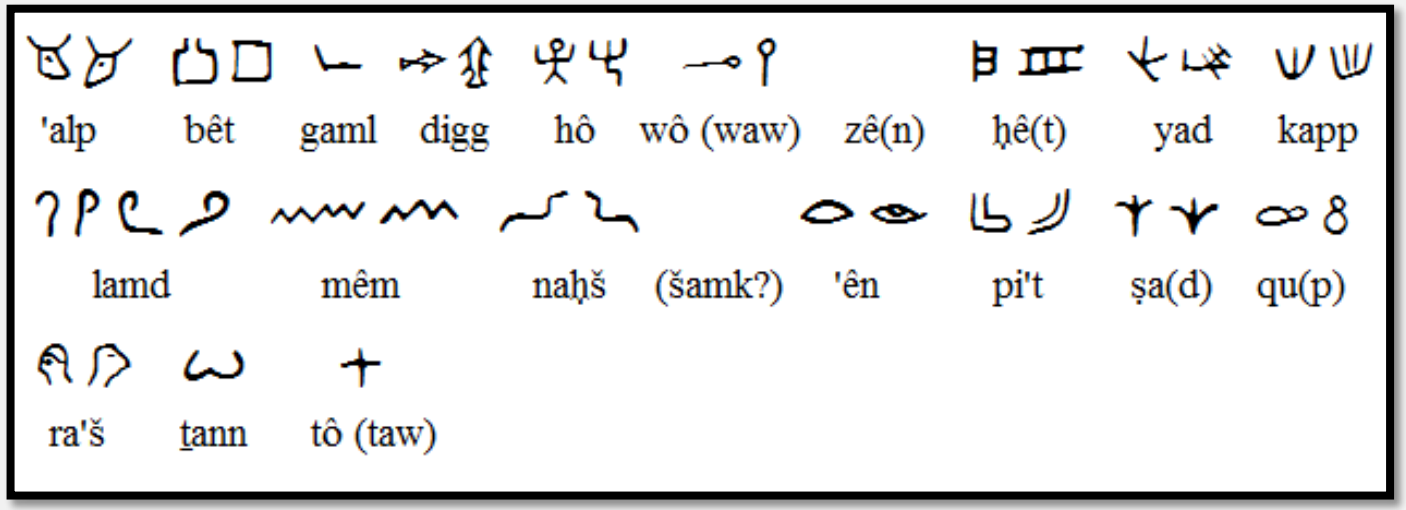


Possible correspondences between Proto-Sinaitic and Phoenician

Hieroglyph	Proto-Sinaitic	IPA value	reconstructed name	Phoenician	Paleo-Hebrew	Aramaic	Greek/Italic
		/ʔ/	ʾalp "ox"				A A A A
		/b/	bet "house"				B B B B
		/h/	*hll* "jubilation" > *he* "window"				E E E E
		/k/	kaf "palm of hand"				K K K K
		/m/	mem "water"				M M M M
		/n/	*naḥš* "snake" > *nun* "fish"				N N N N
		/ʕ/	ʿen "eye"				O O O O
		/r/	roʾš "head"				P P P P R
		/ʃ/	*šimš* "sun" > *šin* "tooth"				Σ Σ Σ S
		/t/	*taw* "mark"				T T T T

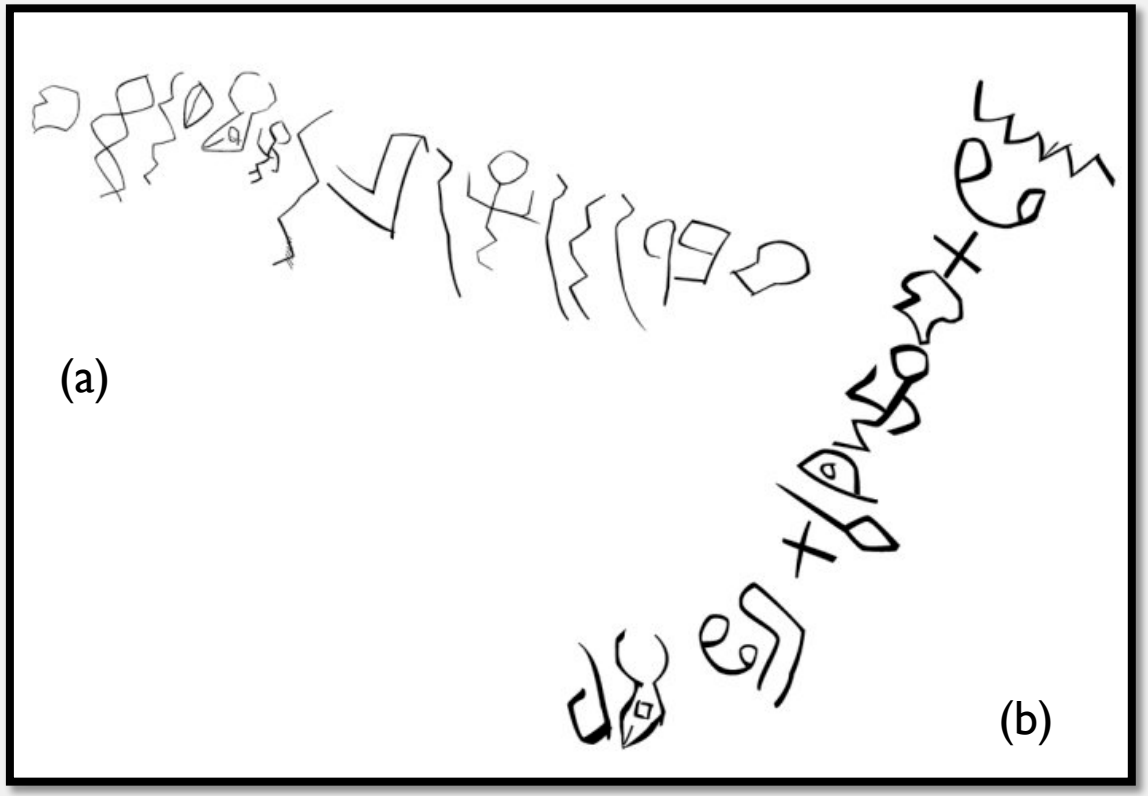
- Ancestor
  - Egyptian Hieroglyphs
- Descendants
  - Proto-Canaanite, in turn, Phoenician
  - all organically evolved alphabets, abjads, abugidas

PROTO-SINAITIC



- Repertoire
  - Closed set of characters
    - ~20 base signs
    - some variants
  - No number signs
  - No punctuation

PROTO-SINAITIC



Wadi el-Hol inscriptions

- Structure
  - Directionality:
    - Horizontal
      - left to right
      - right to left
    - Vertical: top to bottom
      - glyphs may be rotated
  - Non-joining, non-cursive

PROTO-SINAITIC

Serabit el-Khadim, Sinai

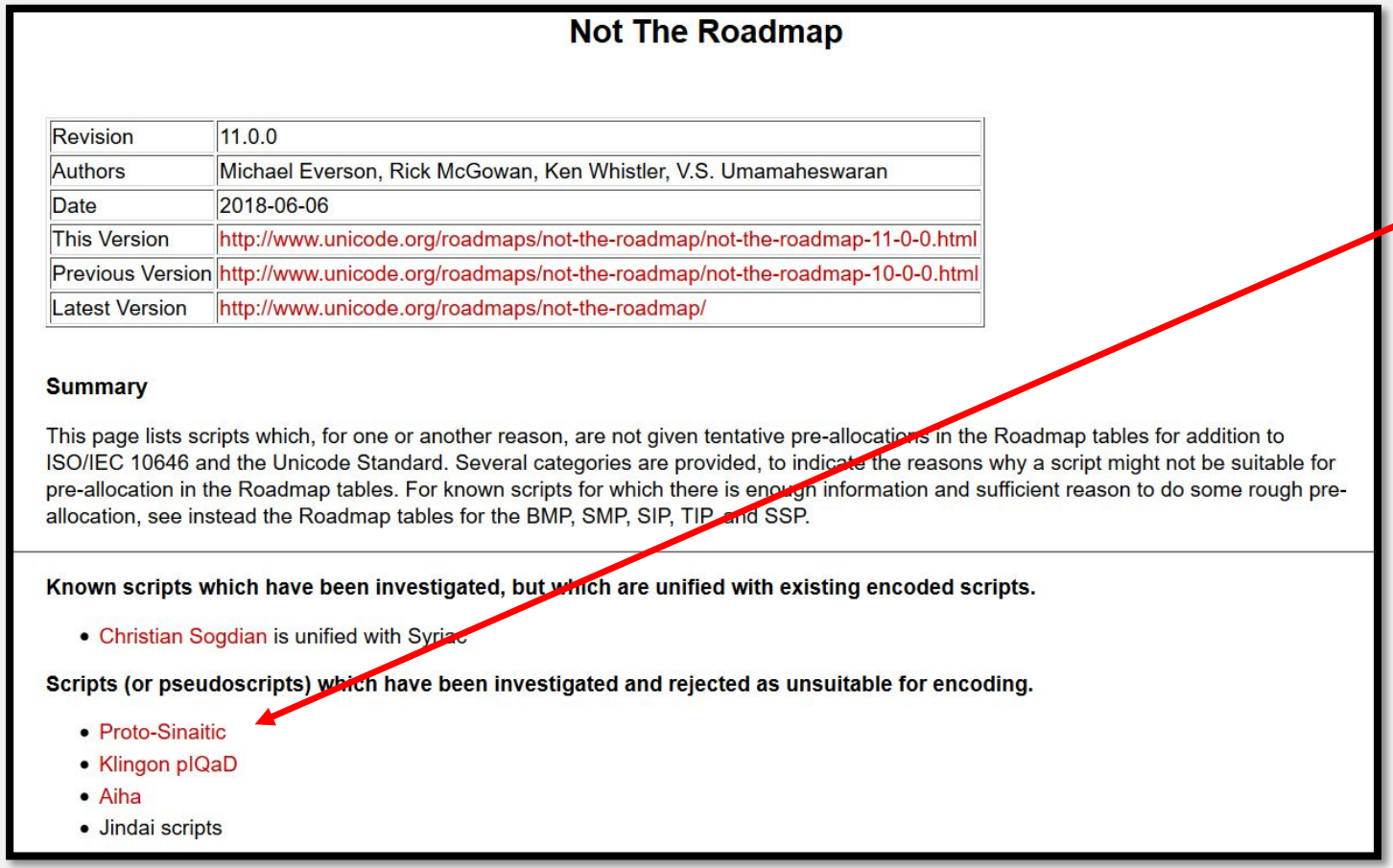
- Users / Inventors
  - Two hypotheses:
    - ‘Illiterate’ miners
    - Literate foreman

PROTO-SINAITIC

mʿhbʿl ... ('beloved of the La(dy)...'), Gardiner 1916

- Origin
  - Hieroglyphs selected by shape of the sign with a familiar object
  - No apparent semantic or phonetic connection between source and target
  - Acrophonic?, Logographic?

PROTO-SINAITIC



Not The Roadmap

Revision	11.0.0
Authors	Michael Everson, Rick McGowan, Ken Whistler, V.S. Umamaheswaran
Date	2018-06-06
This Version	http://www.unicode.org/roadmaps/not-the-roadmap/not-the-roadmap-11-0-0.html
Previous Version	http://www.unicode.org/roadmaps/not-the-roadmap/not-the-roadmap-10-0-0.html
Latest Version	http://www.unicode.org/roadmaps/not-the-roadmap/

Summary

This page lists scripts which, for one or another reason, are not given tentative pre-allocations in the Roadmap tables for addition to ISO/IEC 10646 and the Unicode Standard. Several categories are provided, to indicate the reasons why a script might not be suitable for pre-allocation in the Roadmap tables. For known scripts for which there is enough information and sufficient reason to do some rough pre-allocation, see instead the Roadmap tables for the BMP, SMP, SIP, TIP, and SSP.

Known scripts which have been investigated, but which are unified with existing encoded scripts.

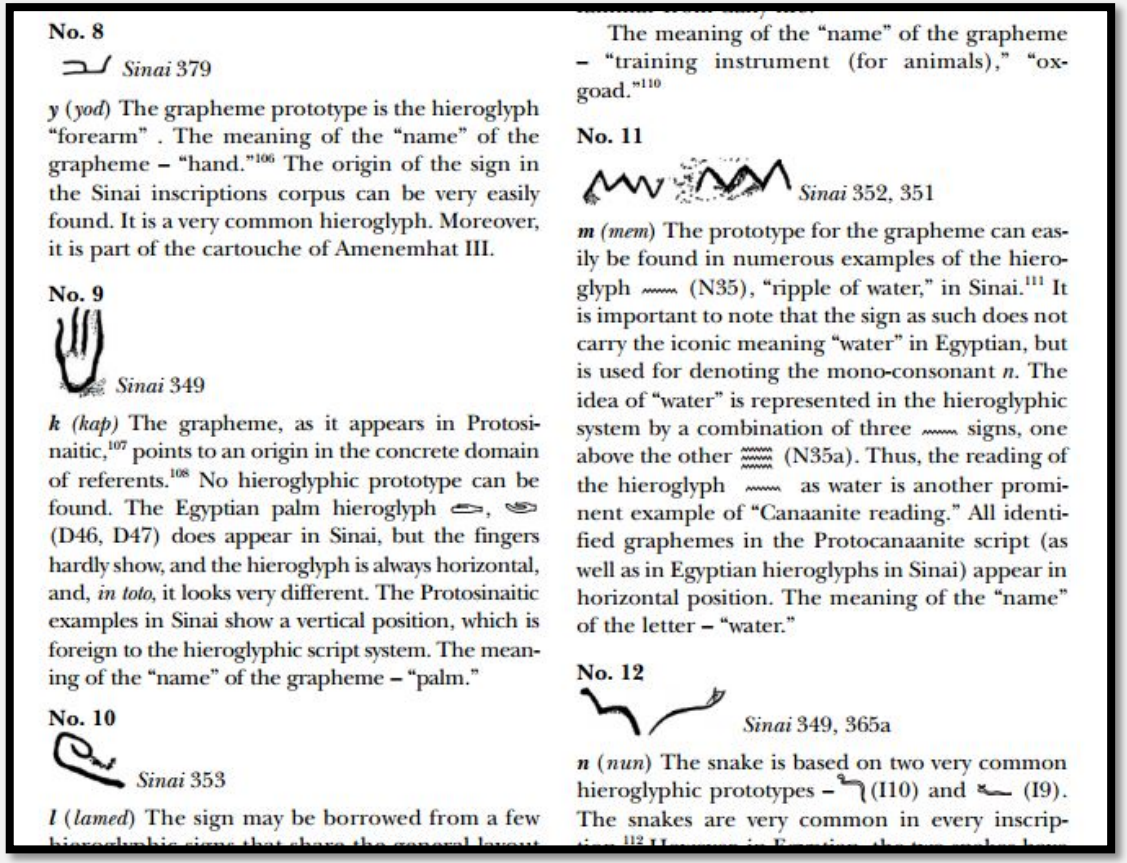
- Christian Sogdian is unified with Syriac

Scripts (or pseudoscripts) which have been investigated and rejected as unsuitable for encoding.

- Proto-Sinaitic
- Klingon pIQaD
- Aiha
- Jindai scripts

- Unicode status
- proposed, but rejected
  - Everson (N1688), 1988
- unallocated to roadmap

PROTO-SINAITIC



No. 8

Sinai 379

y *(yod)* The grapheme prototype is the hieroglyph "forearm" . The meaning of the "name" of the grapheme – "hand."[106] The origin of the sign in the Sinai inscriptions corpus can be very easily found. It is a very common hieroglyph. Moreover, it is part of the cartouche of Amenemhat III.

No. 9

Sinai 349

k *(kap)* The grapheme, as it appears in Protosinaitic,[107] points to an origin in the concrete domain of referents.[108] No hieroglyphic prototype can be found. The Egyptian palm hieroglyph , (D46, D47) does appear in Sinai, but the fingers hardly show, and the hieroglyph is always horizontal, and, *in toto*, it looks very different. The Protosinaitic examples in Sinai show a vertical position, which is foreign to the hieroglyphic script system. The meaning of the "name" of the grapheme – "palm."

No. 10

Sinai 353

l *(lamed)* The sign may be borrowed from a few

The meaning of the "name" of the grapheme – "training instrument (for animals)," "ox-goad."[110]

No. 11

Sinai 352, 351

m *(mem)* The prototype for the grapheme can easily be found in numerous examples of the hieroglyph (N35), "ripple of water," in Sinai.[111] It is important to note that the sign as such does not carry the iconic meaning "water" in Egyptian, but is used for denoting the mono-consonant *n*. The idea of "water" is represented in the hieroglyphic system by a combination of three signs, one above the other (N35a). Thus, the reading of the hieroglyph as water is another prominent example of "Canaanite reading." All identified graphemes in the Protocanaanite script (as well as in Egyptian hieroglyphs in Sinai) appear in horizontal position. The meaning of the "name" of the letter – "water."

No. 12

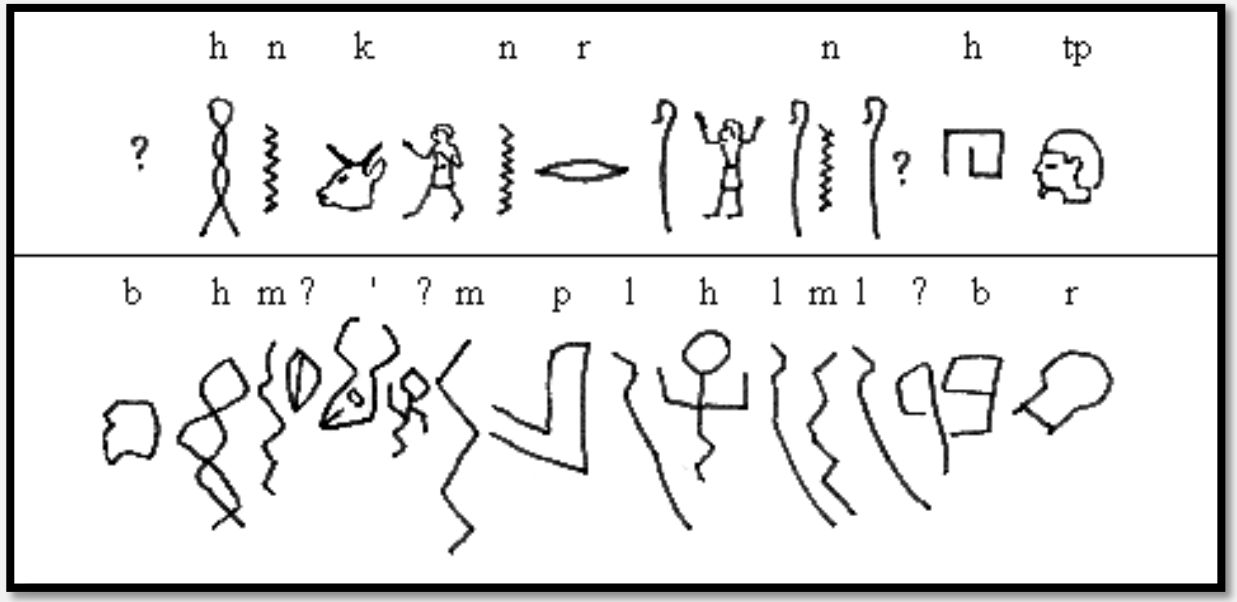
Sinai 349, 365a

n *(nun)* The snake is based on two very common hieroglyphic prototypes – (I10) and (I9). The snakes are very common in every inscrip-

Goldwasser,, "From the Iconic to the Linear", 2016

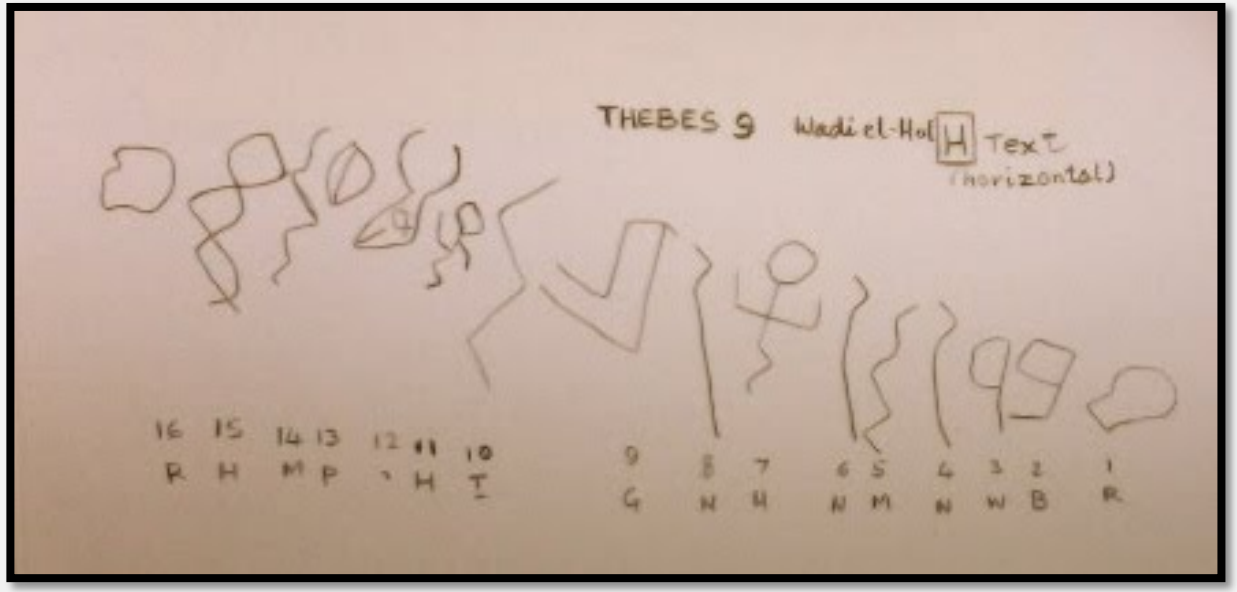
- Current usage
  - active scholarship
  - representation of signs in publications
  - exchange of documents containing signs
  - fonts

PROTO-SINAITIC



- Status of decipherment
  - value of signs not firmly deciphered
  - variance in typology (alphabetic, logographic, rebus?)

PROTO-SINAITIC

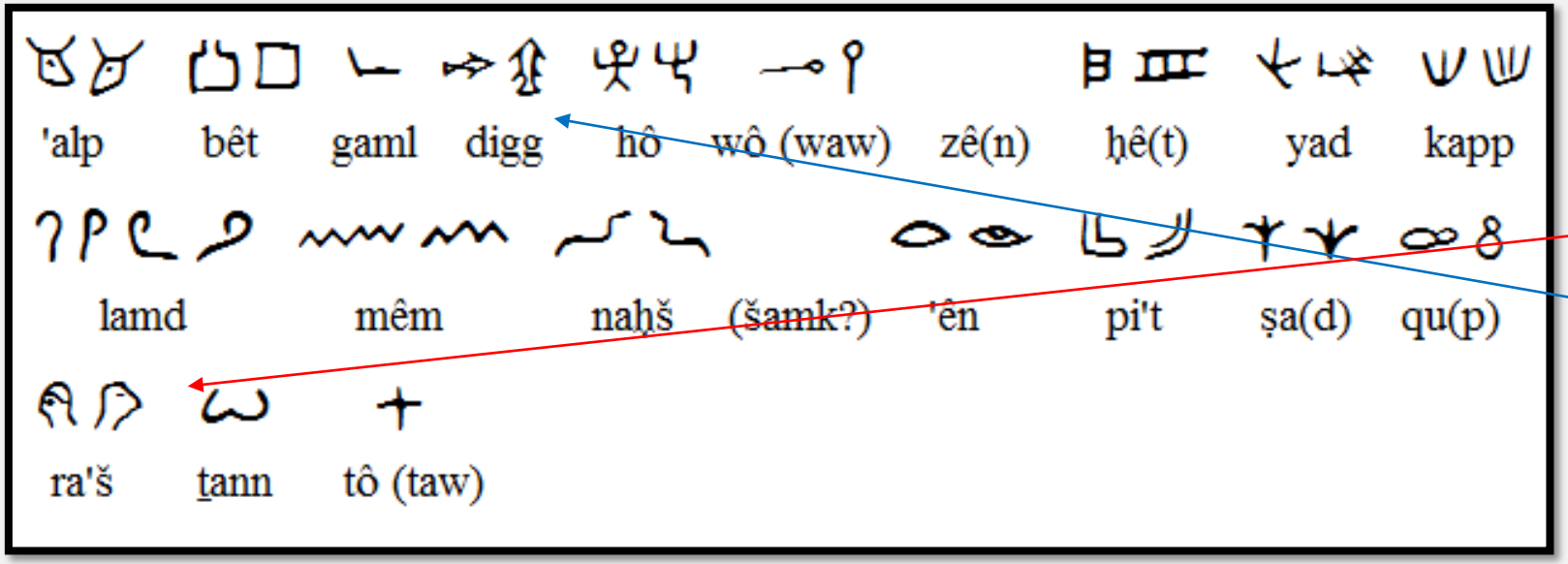


- Issues with encoding
  - Typology?
  - Sign values?

Darnell, others: *rb* ...

Colless: "Excellent (R) banquet (mšt) of the celebration (H) of `Anat (`nt). 'El ('l) will provide (ygš) plenty (rb) of wine (wn) and victuals (mn) for the celebration (H). We will sacrifice (ngt_) to her (h) an ox (') and (p) a prime (R) fatling (mX)."

PROTO-SINAITIC

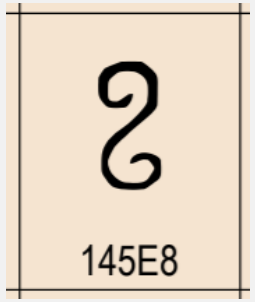


- Issues with encoding
  - Representative glyphs?
  - Directional variants?
    - Horizontal (mirror)
    - Vertical (rotated)
  - Variant vs. distinctive

N.E. SCRIPTS IN THE UNICODE STANDARD NOT FULLY UNDERSTOOD

- Scripts in Unicode containing some characters whose values are still unknown:
  - Linear B
  - Carian
  - Anatolian Hieroglyphs

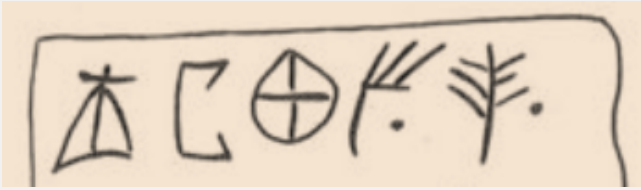
    U+145E8 ANATOLIA HIEROGLYPH A435
    = syllabic a-x?



N.E. SCRIPTS IN THE UNICODE STANDARD NOT FULLY UNDERSTOOD

- Scripts that are partly deciphered/language not fully understood:

  Linear A



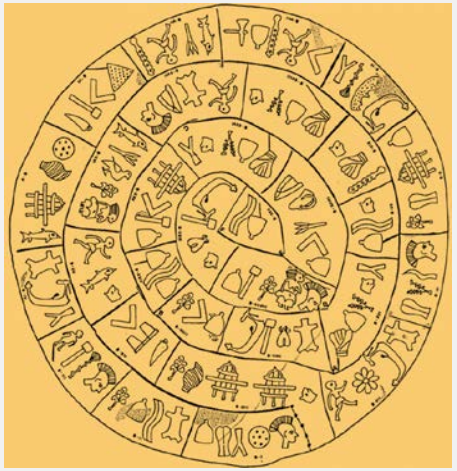
- Underlying language not fully understood:

  Etruscan, (Old Italic script)

Etruscan inscription in Old Italic script

APPROACHES TO UNDECIPHERED SCRIPTS

- Use image-based solutions/PUA until script is better understood, or a stronger case for encoding can be made
- Option: Encode characters as *symbols*, such as was done for Phaistos Disc symbols



APPROACHES TO UNDECIPHERED SCRIPTS – PROTO-SINAITIC

- Model A
  - Use Phoenician or Hebrew encoding (current practice for existing fonts)
  - prevents distinctive representation of script in plain text
- Model B
  - Encode as a separate script
  - Character repertoire
    - encode all attested signs?
    - directional variants?
  - Handle directionality using mark-up
  - Goal: interchange, not perfection

CONCLUSION

- Semantic-gap conundrum
  - Information recorded, but cannot access or decode
- Usage conundrum:
  - Represent this information digitally, but no support
- Encoding conundrum:
  - How to define semantics for the unknown?