A Computational Study of the Evolution of Cretan and Related Scripts

Peter Z. Revesz

Abstract— Crete was the birthplace of several ancient writings, including the Cretan Hieroglyphs, the Linear A and the Linear B scripts. Out of these three only Linear B is deciphered. The sound values of the Cretan Hieroglyph and the Linear A symbols are unknown and attempts to reconstruct them based on Linear B have not been fruitful. In this paper, we compare the ancient Cretan scripts with four other Mediterranean and Black Sea scripts, namely Phoenician, South Arabic, Greek and Old Hungarian. We provide a computational study of the evolution of the three Cretan and four other scripts. This study encompasses a novel translation of the scripts to a DNA encoding, which enables the use of hypothetical evolutionary tree reconstruction algorithms from the area of bioinformatics.

Keywords—Cretan Hieroglyph, Linear A, Linear B, Evolution, Neighbor Joining, Phylogenetic tree.

I. INTRODUCTION

CRETE was the birthplace of several ancient writings that were first categorized by Arthur Evans, the explorer of Knossos Palace, as the Cretan Hieroglyph, the Linear A and the Linear B scripts [5]. Linear A, which dates back to about 2500 BC, was the main script used in the Minoan palaces of ancient Crete. The Cretan Hieroglyph script, which may predate Linear A, was used for centuries simultaneously with Linear A. Linear A was replaced around 1450 BC by Linear B, which was used in Mycenaean Greece and is the oldest known Greek writing [10]. In 1952 Michael Ventris gave a decipherment of Linear B as described in Chadwick [2]. However, the Cretan Hieroglyph and the Linear A scripts are still not deciphered.

In order to understand better these three ancient Cretan scripts, in this paper we study their relationship with four other scripts. The other scripts are the Phoenician, the South Arabic, the Greek and the Old Hungarian alphabets.

The Phoenician alphabet [25] was a major influence on the development of many other alphabets due to the Phoenicians' widespread commercial influence in the Mediterranean area. Both the Phoenician and the South Arabic alphabets derive from the Proto-Sinaitic alphabet, which is assumed to have originated in the Sinai Peninsula sometime between the mid-19th and mid-16th century BC [26]. Phoenician represents a northern branch while South Arabic represents a southern branch of Proto-Sinaitic.

The classical Greek alphabet from about 800 BC had a major influence for many other European alphabets. The classical Greek alphabet derives from the Phoenician alphabet except for the letters Φ, X, Ψ and Ω [24].

The Old Hungarian alphabet is the alphabet used by Hungarians before the adoption of the Latin alphabet. Parallel with the Latin, it was used sporadically until the 20th century in some Hungarian ethnic minority areas of Rumania. The origin of Old Hungarian is still debated. Hosszú [11] presents a detailed view of the development from Phoenician via Aramaic and Turkish and Proto-Rovas scripts. On the other hand, Forrai [8] and Varga [23] claim that the Old Hungarian script already existed in the Bronze Age and cite putative translations of engraved artifacts going back to 1000 BC.

In computational biology, the question of evolutionary relationships is greatly facilitated by the wide availability of genomic data and the development of a growing number of phylogenetic tree construction algorithms. Some of the best-known phylogenetic tree algorithms are Saitou and Nei's *neighbor-joining* method [19] and Sokal and Michener's UPGMA method [21]. The books by Baum and Smith [1], Hall [9] and Lerney et al. [12] review the *maximum likelihood* and several other methods. Recently, Revesz [15] also proposed the *Common Mutations Similarity Matrix* or CMSM method for phylogenetic tree construction. The CMSM method derives from a series of previous evolutionary biology studies, including [14], [16]-[18], [20] and [22].

Some of the efficient phylogenetic tree algorithms are able to reconstruct hypothetical evolutionary trees in a few minutes of computational time. Moreover, they are based on statistical techniques that are free of human bias, which sometimes prevent the objective evaluation of linguistic artifacts. Human translation attempts are inherently prone to error. For example, the Phaistos Disk, which contains some form of Cretan Hieroglyph writing, was translated in numerous contradictory ways by a large number of professional and amateur linguists. Faucounau [6] and Fisher [7] are example decipherment attempts, and Duhoux [4] is a critique of previous decipherment attempts. In this paper, we strongly advocate computerized approaches to the study of linguistic questions in order to eliminate human bias.

This paper is organized as follows. Section II presents a comparative table of the script symbols. Section III describes the DNA encoding of scripts. Section IV presents a computational reconstruction of the evolutionary tree of the scripts and a discussion of the results. Finally Section V gives some conclusions and directions for future work.

Peter Z. Revesz is with the Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA (revesz@cse.unl.edu).

Table 1 A comparison of the script symbols

Hieroglyph and Phaistos	Linear A	Linear B	Value	Phoenician	Value	South Arabic	Value	Greek	Old Hungarian	Value
					ʔ		ʔ	A		A
			SE		B		B	B		P
					G		G	Γ		G
			DA		D		D	Δ		D, T
					H		H	E		E
					W		W	Υ		US
					Z		Z	Z		U
					Ḥ		ḏ lˁ	H		GY
			KA		Tˁ		Tˁ	Θ		TY
					Y		Y	I		J
			WE		K		K	K		G^H, K
			PU		L		L	Λ		L
			TWE		M		M	M		M
					N		N	N		NT
			TE		S		S	Ξ		H
			QE		ˁ		ˁ	O		J, L
					P			Π		
			ZO		ṣ		ṣ	Μ		C
			QA		Q		Q	Ϙ		K
					R		R	P		*R
			TI		š		š	Σ		š
			RO		T		T	T		D
							F	Φ		F
								X		H
			RE				Ḥ	Ψ		ZS
			TA					Ω		O

II. A COMPARATIVE TABLE OF SCRIPT SYMBOLS

As a first step, we built a comparative table of script symbols as shown above in Table 1. In Table 1, the Phoenician alphabet and the South Arabic alphabet columns are taken from [25] with minor modifications. The Greek alphabet column is taken from [24]. The Old Hungarian alphabet column is our addition. The sound values of the Old Hungarian alphabet are from [8], [11] and [23]. The symbols marked with a star * are Proto-Rovas symbols that were used in the early phases of Old Hungarian according to Hosszú [11]. Our reconstruction assumed that the * symbols represent the more archaic form of Old Hungarian. It is possible that these archaic forms were changed to the latter forms due to Turkish or other influences.

Our reconstruction of Old Hungarian was guided by a combination of visual and sound value correspondences. The visual and the sound value correspondences almost always support each other. There are a few exceptions. For example, the Old Hungarian "US" sound value is different from the Phoenician and South Arabic semivowel "W" sound value. However, in languages where the "W" was not used, it was commonly translated as the vowel "U," including in ancient Greek, where the symbol was named "UPSILON." The Old Hungarian "US" may be a similar adaptation of "W" to "U."

Linear B and its values are from Chadwick [2] and Hooker [10]. The Cretan Hieroglyph and Linear A correspondences to Linear B are our reconstructions but are based in part on previous observations by Evans [5], Fisher [7] and Young [28]. Since the sound values of the Cretan Hieroglyph and Linear A symbols are unknown, their correspondences are based only on visual observations. Finally, the correspondence between the Linear B symbols and the four alphabets is also based primarily on visual observations. For example, the Linear B wheel symbol ⊕ has an obvious parallel in Phoenician ⊗ even though they denote different sound values. The major difficulty here is not simply that Linear B is a syllabary while Phoenician is an alphabet. A syllabary with consonant-vowel syllable combinations can have a natural evolution into an alphabet when either the consonant or the vowel is dropped. However, in this case the Linear B sound value is "RA" which cannot be reconciled with the Phoenician sound value or the corresponding Greek Θ or "THETA." When there was a conflict between the visual and the sound value correspondences, we always took the visual correspondence as having the primary significance.

III. THE DNA ENCODING OF SCRIPT SYMBOLS

After the alignment of the script symbols as shown in Table 1, we took a careful look at each row. In each row, we divided the set of symbols into groups such that in each group the symbols were closer together than they were to members of other groups. For example, in the first row both the Hieroglyph symbol and the Linear A symbol clearly denote persons. Hence they are grouped together. The Phoenician and the Greek letters are only rotations of each other. Hence they also are grouped together. The South Arabic and the Old Hungarian are much more different than the others. Hence we placed the South Arabic into the third group, and the Old Hungarian into the fourth group.

We call the first group the A group, the second group the C group, the third group the G group, and the fourth group the T group. These groups are named after the four DNA nucleotides. After the grouping of the symbols in a row of Table 1, we wrote down the group labels in column where the rows corresponded to the seven scripts. The final result is shown in Figure 1.

```
CLUSTAL 2.1 multiple sequence alignment

Linear_A        AAAACA-CC-AAAA-A-ACATA--AG
Linear_B        -A-A---CC-AAAACA-AC-TA--AG
Hieroglyph      AAAAAAAAAAAAAAAAA-AAATAG-AG
O_Hungarian     TGAACAACCAGAAAAA-ATATAGTAG
S_Arabic        GGCACAAGAAGCAAAA-ACATAC-A-
Phoenician      CCCCGCCGCCCCCACAACGCCA----
Greek           CCCCGCGTCGCCCACAACGCCAGTAG
```

Fig. 1 The DNA encoding of the seven alphabets

IV. COMPUTATIONAL RECONSTRUCTION OF AN EVOLUTIONARY TREE USING PHYLOGENETICS

A. Computational Reconstruction Using ClustalW2

We used ClustalW2's phylogenetic algorithms because they are currently among the most frequently used in bioinformatics and are available free to all users from the website http://www.ebi.ac.uk/Tools/msa/clustalw2/

For the DNA encoding in Figure 1, ClustalW2 computed a hypothetical phylogenetic tree as shown in Figure 2. This type of phylogenetic trees is also called a *cladogram*. ClustalW2 provides two phylogenetic tree generation algorithms the Neighbor Joining (NJ) method, which is the default, and the UPGMA method. We chose the NJ method.

B. Discussion of the Results

The results shown in Figure 2 suggest that the seven scripts have a common ancestor from where three branches descend.

Linear_A 0.00058
Linear_B -0.00058
Hieroglyph 0.14665
O_Hungarian 0.11021
S_Arabic 0.17348
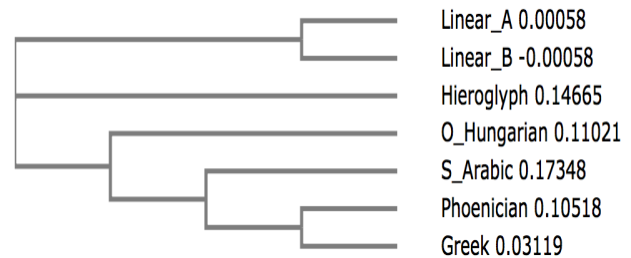Phoenician 0.10518
Greek 0.03119

Fig. 2 The ancient scripts evolutionary tree reconstructed by the Neighbor Joining algorithm in ClustalW2

These branches are as follows:

1. Hieroglyph
2. Linear A and Linear B
3. Greek, Old Hungarian, Phoenician, South Arabic

Since the Hieroglyph script is the oldest attested writing, the common ancestor was likely very close to the Hieroglyph writing. Linear A and Linear B were developed in Crete although they spread with Minoan and Mycenaean Greek cultures to other areas. Many researchers already noted the close visual relationship between the Linear A and the Linear B script symbols.

In the third branch some relationships are completely as expected. For example, the close relationship between Phoenician and Greek is due to widely recognized ancient Greek adoption of the Phoenician script. Phoenician and South Arabic are also recognized to have a common ancestor called Proto-Sinaitic [26]. Therefore the Proto-Sinaitic could be the common root of the Greek, Phoenician and South Arabic scripts. However, it seems new information in the third branch that Old Hungarian and Proto-Sinaitic are also related as sister scripts. This result seems to contradict the view presented in Hosszú [11] that Old Hungarian is a derivative of Phoenician because then Old Hungarian would be placed closer to Phoenician. The results split the third branch into an Old Hungarian sub-branch and a Proto-Sinaitic sub-branch. Therefore, the results are consistent with the view of Forrai [8] and Varga [23] in Old Hungarian being an ancient script whose archaic form existed concurrently with Proto-Sinaitic.

Previously, Naddeo [13] suggested a relationship between the Ugaritic and the Old Hungarian scripts. Colless [3] claims the Ugaritic script is derived from Proto-Sinaitic. If Colless' and Naddeo's claims are true, then they provide another link between Old Hungarian and Proto-Sinaitic. Further, Ugaritic may be a root of the third branch of the evolutionary tree in Figure 2. However, the Ugaritic script is a cuneiform abjad (consonants only alphabet), and the translations provided by Naddeo and by Colless from Ugaritic cuneiform to Old Hungarian and Proto-Sinaitic, respectively, are not very convincing. Hence more research needs to be done to decide whether the Ugaritic script also belongs to the third branch of the evolutionary tree.

V. CONCLUSIONS AND FUTURE WORK

In this study, we focused on the shape of the script symbols and the objects that they may depict because the sound values of the Cretan Hieroglyph and Linear A symbols are unknown. When the sound values are known, the correspondence is close for Phoenician, South Arabic, Greek and Old Hungarian. Only the Linear B sound values are markedly different. It has been attempted to read Cretan Hieroglyph and Linear A scripts using Linear B sound values without any fruitful result. Instead of the Linear B sound values, now it is possible to read Cretan Hieroglyph and Linear A scripts using some possible common ancestor sound values of the corresponding

Phoenician, Greek, South Arabic, and Old Hungarian alphabet symbols. The common ancestor sound values may correspond well to the initial sounds of the syllables that the Cretan Hieroglyph and the Linear A symbols stand for, in case they also denote syllables. We believe that the third branch of the evolutionary tree of Figure 2 may have preserved the original sound values better than Linear B did. We hope that this realization will open a new phase in the understanding of the ancient Cretan scripts.

REFERENCES

[1] D. Baum and S. Smith, *Tree Thinking: An Introduction to Phylogenetic Biology*, Roberts and Company Publishers. 2012.

[2] J. Chadwick, *The Decipherment of Linear B*, Cambridge University Press, 1958.

[3] B. Colless, "Cuneiform alphabet and picto-proto-alphabet," https://sites.google.com/site/collesseum/cuneiformalphabet, downloaded July 5, 2015.

[4] Y. Duhoux, "How not to decipher the Phaistos Disc," *American Journal of Archaeology*, Vol. 104, No. 3 (2000), pp. 597–600.

[5] A. J. Evans, *Scripta Minoa: The Written Documents of Minoa Crete with Special Reference to the Archives of Knossos*, Volume II, Classic Books, 1909.

[6] J. Faucounau, *Le Déchiffrement du Disque de Phaistos: Preuves et conséquences*. L'Harmattan, Paris/Montreal 1999.

[7] S. R. Fisher, *Glyph-Breaker*, Springer, 1997.

[8] S. Forrai, The Old Hungarian Writing from Ancient Times to the Present, (in Hungarian), Antológia Kiadó, 1994.

[9] B. G. Hall, *Phylogenetic Trees Made Easy: A How to Manual*, 4th edition, Sinauer Associates, 2011.

[10] J. T. Hooker, *Linear B: An Introduction*, Bristol Classical Press, 1980.

[11] G. Hosszú, *Heritage of Scribes: The Relation of Rovas Scripts to Eurasian Writing Systems*, Rovas Foundation Hungary, 2013.

[12] P. Lerney, M. Salemi, and A.-M Vandamme, editors. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd edition, Cambridge University Press, 2009.

[13] M. Naddeo, *The Ugarit Abjad ... A Rovás Alphabet*, self-published book, 2007.

[14] P. Z. Revesz, *Introduction to Databases: From Biological to Spatio-Temporal*, Springer, New York, 2010.

[15] P. Z. Revesz, "An algorithm for constructing hypothetical evolutionary trees using common mutations similarity matrices," *Proc. 4th ACM International Conference on Bioinformatics and Computational Biology*, ACM Press, Bethesda, MD, USA, September 2013, pp. 731-734.

[16] P. Z. Revesz and C. J.-L. Assi, "Data mining the functional characterizations of proteins to predict their cancer relatedness," *International Journal of Biology and Biomedical Engineering*, 7 (1), 2013, pp. 7-14.

[17] P. Z. Revesz and T. Triplet, "Classification integration and reclassification using constraint databases," *Artificial Intelligence in Medicine*, 49 (2), 2010, pp. 79-91.

[18] P. Z. Revesz and T. Triplet, "Temporal data classification using linear classifiers," *Information Systems*, 36 (1), 2011, pp. 30-41.

[19] N. Saitou and M. Nei, "The neighbor-joining method: A new method for reconstructing phylogenetic trees," *Molecular Biological Evolution*, 4, 1987, pp. 406-425.

[20] M. Shortridge, T. Triplet, P. Z. Revesz, M. Griep, and R. Powers, "Bacterial protein structures reveal phylum dependent divergence," *Computational Biology and Chemistry*, 35 (1), 2011, pp. 24-33.

[21] R. R. Sokal, and C. D. Michener, "A statistical method for evaluating systematic relationships," *University of Kansas Science Bulletin*, 38, 1958, pp. 1409-1438.

[22] T. Triplet, M. Shortridge, M. Griep, J. Stark, R. Powers, and P. Z. Revesz, "PROFESS: A protein function, evolution, structure and sequence database," *Database -- The Journal of Biological Databases and Curation*, 2010, Available: http://database.oxfordjournals.org/content/2010/baq011.full.pdf+html

[23] G. Varga, *Bronzkori Magyar Irásbeliség*, Irástörténeti Kutató Intézet publication, 1993.

[24] Wikipedia, "History of the Greek alphabet," downloaded July 6, 2015. Available: https://en.wikipedia.org/wiki/History_of_the_Greek_alphabet

[25] Wikipedia, "Phoenician alphabet," downloaded July 6, 2015. Available: https://en.wikipedia.org/wiki/Phoenician_alphabet

[26] Wikipedia, "Proto-Sinaitic script," downloaded July 6, 2015. Available: https://en.wikipedia.org/wiki/Proto-Sinaitic_script

[27] Wikipedia, "South Arabian alphabet", downloaded July 5 2015. Available: https://en.wikipedia.org/wiki/South_Arabian_alphabet

[28] J. G. Young, "The Cretan Hieroglyphic script: A review article," *Minos* 31-32 (1996-1997[1999]) 379-400.

Peter Z. Revesz holds a Ph.D. degree in Computer Science from Brown University. He was a postdoctoral fellow at the University of Toronto before joining the University of Nebraska-Lincoln, where he is a professor in the Department of Computer Science and Engineering. Dr. Revesz is an expert in databases, data mining, big data analytics and bioinformatics. He is the author of *Introduction to Databases: From Biological to Spatio-Temporal* (Springer, 2010) and *Introduction to Constraint Databases* (Springer, 2002). Dr. Revesz held visiting appointments at the IBM T. J. Watson Research Center, INRIA, the Max Planck Institute for Computer Science, the University of Athens, the University of Hasselt, the U.S. Air Force Office of Scientific Research and the U.S. Department of State. He is a recipient of an AAAS Science & Technology Policy Fellowship, a J. William Fulbright Scholarship, an Alexander von Humboldt Research Fellowship, a Jefferson Science Fellowship, a National Science Foundation CAREER award, and a "Faculty International Scholar of the Year" award by *Phi Beta Delta*, the Honor Society for International Scholars.