The Indus Valley Script Decipherment: Sanskrit vs. Dravidian Approaches and the Genetic Evidence for South Asian Origins

© Noel D. Zinn, 2025, November, ndzinn@gmail.com
Published in TNS's Journal *Vidya*, 450 | 451, Dec 2025 | Jan 2026

Preface

This is a big topic, and controversial. The Archaeological Survey of India recently hosted a conference on the decipherment of the Indus Valley Script August 20–22 this year at the Institute of Archaeology in Greater Noida (UP). I've emailed for the proceedings, but apparently they're not yet published. I'm hoping for a digital copy.

Recently I discovered Perplexity AI, a revolutionary AI product for academic (and other, less formal) papers that includes its references. The Perplexity CEO, Aravind Srinivas, is Tamil, 31 years old, born in Chennai, PhD from Berkeley. I, too, speak (and read) Tamil from my years in India in the late 1960s as a Peace Corps Volunteer, and I, too, am a Berkeley graduate (AB, 1967). So, the stars seem to have aligned! Closely following the many efforts to decipher the Indus Valley script since my years in India, I favor the thesis that its origins are Dravidian (specifically proto-Tamil), which is also the conclusion of Perplexity AI, my partner in this paper. Notice that the ends of most paragraphs have links to the end references with more information online, basically the sources of Perplexity's commentary that constitute much of the paper. This is the key feature that distinguishes Perplexity from the AI competition. But no AI can do the job alone. It requires a long and detailed prompt for a topic this complex as well as a (trial and error) tuning of Perplexity's many settings to guide the AI where to search. Turning that (re)search into coherent, scholarly text is the marvel of a Large Language Model.

By way of immediate clarification, Sanskrit belongs to the Indo-European language family, specifically the Indo-Iranian branch, and within that, the Indo-Aryan subgroup. As an ancient classical language, it serves as a common ancestor to many modern Indian languages, primarily in the north, and is known for its holy and literary significance in

Hinduism, Buddhism, and Jainism. On the other hand, the Dravidian language group is a major family of over 80 languages spoken by around 220 million people, primarily in southern India and surrounding countries, with major examples including Tamil, Telugu, Kannada, and Malayalam. These languages are distinct from the Indo-European languages found in northern India and have a rich history dating back approximately 4,500 years.

Overview

The ongoing quest to decipher the Indus Valley script represents one of archaeology's most tantalizing challenges, with profound implications for understanding the origins of Indian civilization. Recent computational advances, archaeological discoveries, and genetic studies have reinvigorated the debate between Sanskrit and Dravidian approaches to decipherment, while DNA evidence provides new insights into the population movements that shaped South Asia during the Bronze Age. The 2014 discovery of an Indus Valley site at Keeladi in Tamil Nadu has particularly intensified this academic and political contest, with the Tamil Nadu government offering a substantial reward for successful decipherment.[1][2][3][4][5][6][7]

NB: If you click on the hyperlinked graphic of Indus Valley seals at the end of this paper, you'll be redirected to a long list of Indus Valley ideograms.

The Computational Revolution in Script Analysis

The past decade has witnessed unprecedented advances in the computational analysis of the Indus script, offering new methodological approaches where traditional linguistic methods have struggled. Statistical network analysis has revealed syntactic patterns suggesting the script represents a genuine language system rather than mere pictographs. Researchers have employed n-gram Markov chain analysis to identify text beginners and enders, demonstrating clear directionality and structure consistent with written language. Deep learning algorithms have been developed to automate the

recognition and classification of Indus signs from archaeological artifacts, potentially revolutionizing epigraphical research.[8][9][10][11][12]

Machine learning approaches have also identified significant redundancies in the proposed sign lists, with statistical analysis of mirrored asymmetric signs revealing that the Indus script was multi-directional, where sign mirroring often denotes writing direction without semantic difference. This reduction of the sign inventory from over 400 to a more manageable corpus may facilitate future decipherment efforts. Recent vowel harmony analysis using Exponential Random Graph Models suggests the underlying language possessed vowel harmony characteristics, providing potential clues about its linguistic family.[13][3]

Dravidian Approaches: Parpola and the Southern Connection

The Dravidian hypothesis for Indus script decipherment, most prominently advanced by Finnish scholar Asko Parpola, represents the most sustained and academically rigorous approach to understanding the script. Parpola's methodology combines archaeological iconography with comparative Dravidian linguistics, particularly drawing on proto-Dravidian reconstructions and Tamil literary traditions. His work identifies recurring sign combinations that correspond to Dravidian morphological patterns, particularly in the realm of religious and administrative terminology.[14][15][16]

Iravatham Mahadevan, a leading Indian epigraphist, has similarly advocated for a Dravidian reading of the script, emphasizing the geographical continuity between Harappan sites and regions where Dravidian languages persist. The Dravidian approach gains support from the script's apparent agglutinative characteristics, which align with Dravidian linguistic structure rather than the inflectional patterns of Indo-Aryan languages. Archaeological evidence of cultural continuity in pottery styles, urban planning, and religious iconography between Harappan and later South Indian traditions further supports this linguistic connection.[17]

Sanskrit and Indo-Aryan Interpretations

Sanskrit-based decipherment attempts, while less systematically developed than Dravidian approaches, draw on the historical prominence of Sanskrit in Indian civilization and the early arrival of Indo-Aryan languages in the subcontinent. Some scholars have attempted to identify Sanskrit roots in Indus inscriptions, particularly in seal texts that might represent personal names or official titles. However, these approaches face significant chronological challenges, as mainstream archaeological evidence suggests Indo-Aryan languages entered the subcontinent after the decline of the Indus civilization.

The Sanskrit hypothesis confronts the problem of linguistic dating, as proto-Indo-Aryan features necessary for such early attestation are difficult to reconstruct with confidence. Additionally, the morphological characteristics of the Indus script appear more consistent with agglutinative language structures than with the synthetic inflectional patterns characteristic of early Indo-Aryan. Despite these challenges, some researchers continue to explore potential Indo-Aryan connections, particularly through analysis of religious and astronomical terminology.[18]

The Keeladi Revolution: Archaeological Evidence for Tamil Antiquity

The 2014 Archaeological Survey of India excavations at Keeladi in Tamil Nadu have dramatically altered the landscape of Indus civilization research. Located near Madurai, about 150km from where I lived in the Peace Corps, this site has yielded pottery styles, urban planning features, and material culture remarkably similar to Harappan sites, extending the geographical reach of Indus civilization far into peninsular India. Carbon dating of organic materials from Keeladi suggests continuous occupation from the 6th century BCE, with some scholars arguing for even earlier dates that would overlap with the late Harappan period.[4][5]

The Keeladi discoveries have profound implications for understanding the relationship between Harappan and Dravidian civilizations. The site's material culture demonstrates

clear technological and stylistic continuities with Harappan traditions, including sophisticated urban planning, standardized weights and measures, and similar pottery traditions. Most significantly, the discovery of Tamil-Brahmi script at later levels of the site suggests a potential evolutionary connection between Indus script traditions and early Tamil writing systems.

The political significance of these discoveries cannot be understated. The Tamil Nadu government's announcement of a one million dollar reward for Indus script decipherment reflects the high stakes involved in establishing Tamil claims to India's earliest civilization. Chief Minister M.K. Stalin's administration has explicitly framed these archaeological discoveries as evidence for Tamil cultural primacy, challenging dominant narratives that privilege Sanskrit and Indo-Aryan contributions to Indian civilization.[4]

Genetic Evidence and Population Movements

Recent advances in ancient DNA analysis have provided unprecedented insights into the population movements that shaped South Asia during the crucial period of Indus civilization. Large-scale genomic studies analyzing over 500 ancient individuals from Central and South Asia reveal a complex pattern of migration and admixture that coincides with major cultural transitions. These genetic studies identify distinct population groups that contributed to modern South Asian ancestry: Ancestral North Indians (ANI) and Ancestral South Indians (ASI).[6][19]

The genomic data reveals that the people of the Indus Valley civilization, termed "Indus Periphery" in genetic studies, possessed a distinctive mixture of ancestry related to Iranian agriculturalists and South Asian hunter-gatherers. This population shows no direct genetic contribution from the Eurasian Steppe pastoralists who spread Indo-Aryan languages, suggesting the Indus civilization predates significant Indo-Aryan migration. The genetic evidence indicates that Steppe-derived ancestry entered South Asia primarily in the second millennium BCE, after the decline of major Harappan urban centers.[6]

Mitochondrial DNA and Y-chromosome studies of modern South Asian populations provide additional insights into prehistoric migrations. These analyses suggest substantial genetic continuity from the Bronze Age, with modern populations of the Indus Valley region showing ancestry patterns consistent with ancient Harappan populations. Particularly significant is the finding that South Asian populations derive largely from Pleistocene-era settlers with limited external gene flow, supporting models of long-term population continuity rather than wholesale replacement.[7][20][21]

The genetic evidence also reveals important patterns in the distribution of ancestry components across the subcontinent. Dravidian-speaking populations show higher proportions of ASI ancestry, while Indo-Aryan speakers demonstrate greater ANI components, consistent with the historical spread of these language families. However, the genetic data also indicates substantial admixture between these populations, suggesting complex patterns of cultural and linguistic transmission that complicate simple models of population replacement.

Linguistic and Cultural Synthesis

The integration of archaeological, genetic, and linguistic evidence presents a nuanced picture of South Asian prehistory that transcends simple dichotomies between Sanskrit and Dravidian origins. The genetic data strongly suggests that the Indus Valley civilization preceded significant Indo-Aryan migration, supporting arguments for indigenous South Asian origins rather than external importation of civilization. This evidence aligns with Dravidian linguistic models while not entirely excluding the possibility of early Indo-European presence in the region.

Archaeological continuities between Harappan and later South Indian cultures, exemplified by the Keeladi discoveries, provide material support for Dravidian connections to Indus civilization. The persistence of similar urban planning principles, craft traditions, and religious iconography suggests cultural transmission that could plausibly include linguistic elements. However, the complexity of cultural change over

millennia warns against oversimplified equations between ancient and modern linguistic communities.

The computational analysis of the Indus script itself provides the most objective evidence for assessing decipherment claims. The identification of syntactic patterns, directional writing, and potential morphological structures offers testable hypotheses for linguistic approaches. While these analyses cannot definitively identify the underlying language family, they provide constraints that any successful decipherment must satisfy. The apparent agglutinative characteristics of the script align more closely with Dravidian than Indo-Aryan morphology, though other language families cannot be excluded.

Contemporary Implications and Future Directions

The contemporary political dimension of Indus script decipherment reflects broader tensions over cultural identity and historical narrative in modern India. Tamil Nadu's substantial financial incentive for decipherment represents an unprecedented politicization of archaeological research, highlighting the stakes involved in establishing cultural primacy. This political investment has positive aspects in promoting research funding and public interest, but also risks compromising scholarly objectivity and encouraging premature or politically motivated claims.

The integration of computational methods with traditional archaeological and linguistic approaches offers the most promising avenue for future progress. Machine learning algorithms capable of processing large corpora of inscriptions, combined with statistical analysis of sign patterns and distributions, may identify structures invisible to human analysis. These computational tools, when integrated with rigorous archaeological contextualisation and comparative linguistic analysis, could provide the breakthrough necessary for decipherment.

Genetic studies will continue to refine our understanding of population movements and cultural transmission in prehistoric South Asia. Higher resolution ancient DNA sampling, particularly from Harappan sites themselves, could provide direct evidence of the

biological relationships between Indus Valley populations and later linguistic communities. Such evidence would provide crucial context for evaluating competing decipherment hypotheses.

The ultimate resolution of the Indus script mystery will likely require a synthesis of multiple lines of evidence rather than any single breakthrough. The convergence of computational linguistics, ancient genomics, and archaeological analysis represents an unprecedented opportunity to solve one of archaeology's greatest puzzles. Whether the underlying language proves to be Dravidian, Indo-Aryan, or represents an entirely different linguistic tradition, the process of decipherment continues to illuminate the complex origins of South Asian civilization and the diverse streams that contributed to its development.

The stakes remain high: successful decipherment would not only unlock the written records of one of humanity's earliest urban civilizations but also provide definitive evidence for the linguistic and cultural origins of the Indian subcontinent. As computational tools become more sophisticated and archaeological discoveries continue to expand our knowledge base, the prospect of finally hearing the voices of the Indus Valley civilization grows increasingly promising.

Addendum

There’s an unfortunate political/economic twist to all this. The Central government of India is ruled by the BJP (Bharatiya Janata Party) of Narendra Modi, primarily a north Indian party. The (Indo-Aryan) north is poor, the (Dravidian) south is prosperous (consider Bangalore and Hyderabad for example). Modi is gerrymandering the legislature to favor the north. A Peace Corps colleague of mine, an economist who follows this situation, is convinced that the south will ultimately secede from the north, or worse a civil war.

Now the twist. The BJP is pressuring the Archaeological Survey of India (ASI) to change the dating of the Keeladi civilization to disassociate it from the Indus Valley civilization. The BJP is even threatening to withdraw funding. So far the ASI is not yielding.

Examples of Indus Valley Seals (click for source and more)

References

1. https://ijcaonline.org/archives/volume186/number31/ramani-2024-ijca-923836.pdf
2. https://policyjournalofms.com/index.php/6/article/view/229
3. https://www.nature.com/articles/s41599-021-00713-0
4. https://pmc.ncbi.nlm.nih.gov/articles/PMC7666134/
5. https://heritagesciencejournal.springeropen.com/track/pdf/10.1186/s40494-020-00402-2
6. http://biorxiv.org/lookup/doi/10.1101/292581
7. https://pmc.ncbi.nlm.nih.gov/articles/PMC6288199/
8. http://arxiv.org/pdf/1005.4997.pdf
9. https://arxiv.org/abs/1702.00523
10. https://arxiv.org/abs/0901.3017
11. https://pmc.ncbi.nlm.nih.gov/articles/PMC2721819/
12. https://pmc.ncbi.nlm.nih.gov/articles/PMC2841631/
13. https://dl.acm.org/doi/10.1145/3548785.3548788
14. https://www.jstor.org/stable/1580175?origin=crossref
15. https://www.semanticscholar.org/paper/bc98b843f5acfd17fa4ac05080c2fefcdbe5a124
16. https://www.semanticscholar.org/paper/811385f52811767da75f51c7e0d6d1394dd8a321

17. https://www.semanticscholar.org/paper/b57de5e51d82d0bff46016fa0b38ad1c44f54548
18. https://arxiv.org/pdf/1204.3800.pdf
19. https://www.science.org/doi/10.1126/science.aat7487
20. https://pmc.ncbi.nlm.nih.gov/articles/PMC1181978/
21. https://pmc.ncbi.nlm.nih.gov/articles/PMC1380230/
22. https://www.ijfmr.com/research-paper.php?id=11726
23. https://www.degruyter.com/document/doi/10.1515/9783110800258/html
24. https://onlinelibrary.wiley.com/doi/10.1111/j.1600-0498.1977.tb00351.x
25. https://linkinghub.elsevier.com/retrieve/pii/0024384184900378
26. https://www.semanticscholar.org/paper/ab6372074aef84016bcdb71bed7bce74969d0f49
27. https://journal.lps2h.com/ijlsh/article/view/238
28. https://sciresol.s3.us-east-2.amazonaws.com/IJST/Articles/2022/Issue-16/IJST-2022-88.pdf
29. https://aclanthology.org/2023.acl-short.71.pdf
30. https://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00354/1924241/tacl_a_00354.pdf
31. https://aclanthology.org/2023.findings-emnlp.4.pdf
32. https://arxiv.org/pdf/2308.05780.pdf
33. https://zenodo.org/record/2099918/files/article.pdf

34. https://arxiv.org/html/2406.06570

35. https://heritagesciencejournal.springeropen.com/counter/pdf/10.1186/s40494-024-01169-6

36. https://arxiv.org/pdf/1002.4513.pdf

37. http://arxiv.org/pdf/1110.1091.pdf

38. http://openarchaeologydata.metajnl.com/articles/10.5334/joad.57/galley/63/download/

39. https://uclpress.scienceopen.com/document_file/6d489c8d-8926-4c56-b818-4944f8a60f70/ScienceOpen/5c2c93a2-6b81-4f16-96db-a34a6df2f4fe.pdf

40. https://arxiv.org/pdf/1303.1426.pdf

41. https://www.mdpi.com/2072-4292/9/7/735/pdf?version=1500607918

42. https://zenodo.org/record/2233073/files/article.pdf

43. https://www.mdpi.com/2075-163X/12/7/794/pdf?version=1656907847

44. https://www.mdpi.com/2076-3263/9/1/11/pdf?version=1545735377

45. http://www.ancient-asia-journal.com/articles/10.5334/aa.12324/galley/114/download/

46. https://zenodo.org/record/2358502/files/article.pdf

47. https://www.tandfonline.com/doi/pdf/10.1080/00934690.2018.1464332?needAccess=true

48. https://pmc.ncbi.nlm.nih.gov/articles/PMC10888882/

49. https://www.semanticscholar.org/paper/ed7a0bae624ed6f72a289a0675164e3e4eaedf8e

50. https://read.dukeupress.edu/journal-of-asian-studies/article/29/3/714/328853

51. https://www.semanticscholar.org/paper/b028466f13bf26c5f72cb71dcce0c09a4d951aa1

52. https://www.cambridge.org/core/product/identifier/S0041977X00031980/type/journal_article

53. https://journals.sagepub.com/doi/10.1177/026272809601600114

54. https://www.cambridge.org/core/product/identifier/S1356186300006829/type/journal_article

55. https://arxiv.org/pdf/2107.11801.pdf

56. http://arxiv.org/pdf/1812.00049.pdf

57. http://biorxiv.org/lookup/doi/10.1101/2024.03.14.585067

58. https://linkinghub.elsevier.com/retrieve/pii/S0002929707643523

59. https://www.semanticscholar.org/paper/c604511e8560c6b216ded7189356a1cbd8228800

60. https://onlinelibrary.wiley.com/doi/10.1002/evan.20267

61. https://www.nature.com/articles/s41431-022-01057-2

62. http://link.springer.com/10.1007/s12038-019-9889-z

63. https://www.semanticscholar.org/paper/2be2a199f25094937ebd45f63413941377736c58

64. http://biorxiv.org/lookup/doi/10.1101/2020.03.25.004606

65. https://pmc.ncbi.nlm.nih.gov/articles/PMC5426639/

66. https://pmc.ncbi.nlm.nih.gov/articles/PMC4414352/

67. https://pmc.ncbi.nlm.nih.gov/articles/PMC1182158/

68. https://pmc.ncbi.nlm.nih.gov/articles/PMC379225/

69. https://dx.plos.org/10.1371/journal.pone.0271767

70. https://academic.oup.com/gbe/advance-article-pdf/doi/10.1093/gbe/evad225/54256551/evad225.pdf

71. https://pmc.ncbi.nlm.nih.gov/articles/PMC6476732/

72. https://threatenedtaxa.org/index.php/JoTT/article/download/8855/9870

73. https://www.mdpi.com/2072-4292/11/18/2089/pdf

74. https://pmc.ncbi.nlm.nih.gov/articles/PMC11064065/

75. https://zenodo.org/record/2035103/files/article.pdf

76. https://arxiv.org/pdf/1702.04149.pdf

77. https://pmc.ncbi.nlm.nih.gov/articles/PMC4000863/

78. https://www.semanticscholar.org/paper/b63574d5e209b033e0bbddefaac4cb4fc5f3ea18

79. https://zenodo.org/record/2232606/files/article.pdf

80. https://zenodo.org/record/2156982/files/article.pdf

81. https://zenodo.org/record/3524126/files/Salomon 1998.pdf

82. https://zenodo.org/record/1783431/files/article.pdf

83. https://zenodo.org/record/1925049/files/article.pdf

84. https://zenodo.org/record/2119412/files/article.pdf

85. https://zenodo.org/record/1641856/files/article.pdf

86. https://irjt.iorpress.org/index.php/irjt/article/download/1101/760

87. https://zenodo.org/record/1793549/files/article.pdf

88. https://zenodo.org/record/1606762/files/article.pdf

89. https://zenodo.org/record/1580906/files/article.pdf

90. https://storage.googleapis.com/jnl-sljo-j-ijaicterict-files/journals/1/articles/7239/649a54cc8560a.pdf

91. https://irjt.iorpress.org/index.php/irjt/article/download/606/316

92. https://zenodo.org/record/2031044/files/article.pdf

93. https://zenodo.org/record/1516944/files/article.pdf