Bioanthropology of the Hindu Kush Borderlands: A Dental Morphology Investigation

Brian E. Hemphill

Abstract

This paper seeks to test conflicting hypotheses concerning possible migrations into or out of the Indus Valley associated, by some researchers, with the dispersal of Indo-Aryan languages. The Aryan Invasion Theory (AIT) calls for a substantial invasion of central Asian populations into the Indus Valley during the 2nd millennium B.C. This invasion is held to have led to the establishment of classical Vedic culture and its subsequent spread throughout the subcontinent by Brahminic missionaries. The Out-of-India Model (OIM) calls for the development of Indo-European languages within the Indian subcontinent, with subsequent dispersal throughout Central Asia and beyond through population movement after the 5th millennium B.C. The Early Intrusion Model (EIM) calls for two immigration events into western South Asia. The first, and most influential, entry occurred during the 5th millennium B.C. and likely signals a substantial immigration of proto-Elamo-Dravidian-speaking populations. The second entry occurred during the last two millennia B.C. and may have involved a gradual introduction of central Asian genes and perhaps Indo Aryan languages into a large resident South Asian population.

Fifty-seven non-metric variations of the permanent dentition were assessed from over 1400 living and prehistoric individuals from the Indus Valley, peninsular India, and Central Asia. Frequencies of 25 traits could be compared across all 16 samples. Eight traits were eliminated from consideration, because of few observations in living samples or insufficient heterogeneity across samples. Phenetic distances between samples based on the remaining 17 tooth-trait combinations were calculated with C.A.B. Smith's mean measure of divergence statistic and patterns of affinities were assessed with neighbor-joining cluster analysis and principal coordinates analysis. The results provide no support for either the Aryan Invasion theory or the Out-of-India model. There is no significant break in Indus Valley biological continuity during the 2nd millennium B.C. (contra AIT), nor is there any evidence of significant emigration out of the Indus Valley into Central Asia after the 5th millennium B.C. (contra OIM). Rather, the results reveal a significant discontinuity event in Indus Valley biological affinities between the 7th to 5th millennia B.C., followed by a milder discontinuity event during the last two millennia B.C. As such, the results of this study offer greatest support for the Early Intrusion model.

Introduction

Recent DNA investigations among members of the living castes and tribes of India have received great attention in the scholarly journals and these investigations have rekindled controversy over the appearance of Indo-Aryan languages and the spread of Vedic culture throughout the Indian subcontinent. Such data purportedly document greater west and central Asian admixture in high-status castes and among Indo-Aryan-speaking inhabitants of north India relative to Dravidian-speakers of south India (Bamshad 1996; Cavalli-Sforza *et al.* 1994; Mukherjee et al. 2001; Passarino *et al.* 1996). Such research has encouraged an array of non-specialists in population genetics to justify a resurrection of the Aryan Invasion theory (for a discussion see Bryant 2001).

Such efforts are unfortunate, for in their zeal to support a model of Aryan invasion these non-specialists gloss over two important caveats well-known among population geneticists. First, the relationships between biological populations and linguistic communities are not simple ones (Mair 1998). While it is true, in a general sense, that a shared language can serve to unite a community and to separate that community from speakers of other languages (Harding & Sokal 1988; Sokal *et al.* 1988), the exceptions to this pattern are so numerous as to render such general trends of little utility in specific instances (Chen *et al.* 1995). Second, the introduction of new genes into a local area is rarely, if ever, an all-or-nothing proposition (Hartl & Clark 1997). Three factors contribute to this relationship; the size of the local recipient population, the size of the intrusive population, and the proportion of overall matings that occur between members of these two populations (Falconer 1981). If the local resident population is large in number, it will take an increased number of matings over many generations for gene flow from the intrusive population to be detected (Ayala 1982; Hartl & Clark 1997). However, if the local population is small, or if resident and intrusive populations remain largely isolated, genetic changes or discrepancies may be profound (Hedrick 2000).

Advances in our understanding of patterns of population movements and interactions made possible by DNA analysis are extremely important, but are not, by themselves, definitive. This is for a number of reasons, but perhaps most importantly, the studies upon which the resurrection of the Aryan Invasion model depends are based solely on living individuals. It is difficult for such data to resolve the issue of population movements into and out of the Indian subcontinent because the convergence estimates for haplotype divergence are of insufficient resolution to distinguish between rival theories whose operationalized models encompass different expectations of the timing and patterning of biological relationships among prehistoric and living ethnic groups of South Asia. Ideally, one could resolve this dilemma by extracting DNA from the teeth and skeletons of prehistoric inhabitants of India, the Indus Valley, and Central Asia. Unfortunately, depositional conditions in these areas rarely result in the preservation of organic collagen from which ancient DNA may be extracted and amplified. So, is there any way out of this impasse?

Dental morphology provides a useful basis for a comparison of biological affinities among samples of human populations. Dental features are under moderate to strong genetic control (Nichol 1989) and, unlike cranial measurements; are not affected by sex dimorphism (Scott & Turner 1997). Hence, teeth from males, females, and individuals too fragmentary to determine sex, may be pooled to increase sample size. Finally teeth, as the only directly accessible hard tissues of the body, can be compared between living individuals and individuals of the past, thereby providing the temporal sensitivity, lacking from DNA analyses based solely on living individuals.

The current controversy over the appearance of Indo-Aryan languages and the spread of Vedic revolves around three fundamental positions that may be described as: the Aryan Invasion Model, the Out-of-India Model, and the Early Intrusion Model.

Proponents of the Aryan Invasion theory (Erdosy 1995; Parpola 1995) maintain that Indo-Aryan languages in South Asia are due to the entrance of central Asians into the Indus Valley during the 2nd millennium B.C. (Nichols 1997, 1998). They assert that this intrusive population later spread to the upper Doab of north India where Vedic culture initially became established. Proponents of this model contend that Vedic culture eventually spread to south India—through elite dominance (Bamshad *et al.* 1996; Cavalli-Sforza *et al.* 1994; Renfrew 1987).

Advocates of the Out-of-India model assert that Indo-Aryan languages arose in the Indian subcontinent then spread to their historical seats in Central Asia, West Asia, and Europe (Kazanas 2002; Misra 1992; Talageri 2000). These researchers envision a direct movement of Indo-Aryan-speakers and languages out of South Asia during the 4th millennium B.C.

Proponents of the Early Intrusion model (Hemphill & Lukacs 1993; Hemphill *et al.* 1991) assert that the population history of South Asia and the influence of language encompass more than the appearance of Indo-

Aryan languages. Rather, one must also consider the impact of Dravidian-speaking and proto-Mundic-speaking populations as well. Drawing from the work of Witzel (1999), proponents of the Early Intrusion model suggest that prior to the 5th millennium B.C., inhabitants of much of the Indian subcontinent—including the Indus Valley—were speakers of proto-Mundic languages. Following MacAlpin (1981), they suggest proto-Elamo-Dravidian-speakers entered the subcontinent from the northwest during the 5th millennium B.C. (Fairservis & Southworth 1989; Southworth 1995). Advocates of this model admit that the entrance of Indo-Aryan languages may have involved a movement of people from Central Asia into the Indus Valley and beyond, but assert that the growth of South Asian populations after the development of sophisticated agricultural technology muted the genetic impact of this later population incursion.

The current study provides a comparison of 17 permanent tooth-trait combinations among 1,451 individuals of 16 different prehistoric and living populations of Central Asia, the Indus Valley, and peninsular India. These comparisons seek to determine which model for appearance of Indo-Aryan languages and the spread of Vedic culture throughout the Indian subcontinent is best supported by the patterns of biological affinities possessed by prehistoric and living inhabitants of South Asia.

Materials and Methods

Frequencies of dental traits among 16 living and prehistoric samples from the Indus Valley, peninsular India, and Central Asia were assessed by the author and by Lukacs. In all cases, dentitions were assessed for 57 tooth-trait combinations according to the method of Turner *et al.* (1991). Intra- and interobserver variation in morphological evaluations was assessed by repeated scoring of 35 tooth-trait combinations in a random sample of 50 plaster dental casts. In the former case, observation sessions were separated by a period of 18 months. In the latter case, the author scored a random sample of 50 plaster dental casts scored by Lukacs (1976). Observer differences were assessed according to the method of Nichol and Turner (1986) and both intra- and interobserver error was found to be well within acceptable limits.

Frequencies of dental traits were calculated for each grade of expression by sex according to the individual count method of Scott and Turner (1997). This technique not only accounts for the fluctuating asymmetric effects of environmental factors, but also maximizes sample sizes in dental series obtained from archaeological contexts where remains are often fragmentary. While variation in dental trait morphology was scored along an ordinal scale of expression, scores were dichotomized into presence/absence only for comparative purposes. In most cases, any degree of trait development was considered a positive expression. The only exception is hypocone development, in which only full expressions (grade 4) were considered positive.

Frequencies of dental morphology traits among six contemporary ethnic groups of Maharashtra and Andhra Pradesh were contrasted with frequencies obtained from ten prehistoric dental series (Fig. 1). Contemporary ethnic groups from Maharashtra include Marathas (a high-status Hindu caste), Mahars (a low-status Hindu caste), and Madia Gonds (non-Hindu tribal group). Contemporary ethnic groups from Andhra Pradesh include Pakanati Reddis (a high-status Hindu caste), Gompadhompti Madigas (a low-status Hindu caste), and Chenchus (a non-Hindu tribal group).

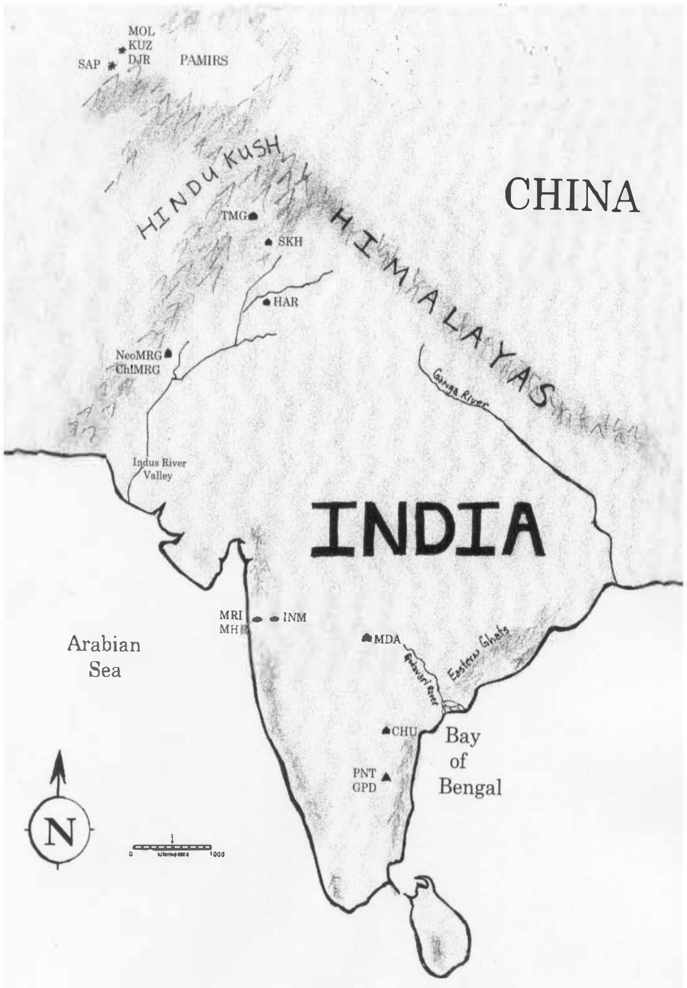

Figure 1: Geographic locations of dental samples.
Sample abbreviations from Table 1.

Five prehistoric dental series are from the Indus Valley (Neolithic Mehrgarh, Chalcolithic Mehrgarh, Harappa, Timargarha, Sarai Khola), four are from the north Bactrian oasis of southern central Asia (Sapalli Tepe, Djarkutan, Kuzali, Molali), and one (Inamgaon) is from peninsular India (Maharashtra). Together, these prehistoric dental series encompass Neolithic, Chalcolithic, Late Bronze Age, and Early Iron Age eras, spanning a temporal range from 6000 B.C. (Neolithic Mehrgarh) to 100 B.C. (Sarai Khola). Sample sizes and abbreviations are provided in Table 1.

Although 57 tooth-trait combinations were scored among all 16 dental series, only 25 tooth-trait combinations were scored in common among all samples. Of these, eight involved traits on the third molars. Since third molars were so infrequent in the six contemporary series, these traits were eliminated from further consideration. Following the procedure used in previous analyses of dental morphology data (Hemphill *et al.* 1991; Hemphill & Lukacs 1993; Lukacs & Hemphill 1991), contingency chi-square analysis was used according to the method of Rothhammer et al. (1984) to determine which of the 17 tooth-trait combinations differ significantly across all 16 samples.

Table 1. Dental Series

Sample	**Abb.**	**Date**	**N_{max}***
Mahars	MHR	Living	195
Madia Gonds	MDA	Living	169
Marathas	MRT	Living	198
Inamgaon	INM	1600-700 B.C.	41
Neolithic Mehrgarh	NeoMRG	6000 B.C.	49
Chalcolithic Mehrgarh	ChlMRG	4500 B.C.	25
Harappa	HAR	2300-1700 B.C.	33
Sarai Khola	SKH	200-100 B.C.	15
Timargarha	TMG	1400-850 B.C.	25
Djarkutan	DJR	2100-1950 B.C.	39
Sapalli Tepe	SAP	2300-2150 B.C.	43
Kuzali	KUZ	1950-1800 B.C.	24
Molali	MOL	1800-1650 B.C.	41
Chenchus	CHU	Living	194
Gompadhompti Madigas	GPD	Living	178
Pakanati Reddis	PNT	Living	182

*. N_{max} represents the greatest number of individuals scored for a non-metric trait.

Trait frequencies obtained through presence/absence dichotomization were arcsine transformed to stabilize sample variance according to the method of Green and Suchey (1976). These values were used as the basis for calculating biological distances between sample pairs with the mean measure of divergence (MMD) statistic. The variance and standard deviation of MMD values were calculated according to the method of Sjøvold (1973). Standardized distances were calculated by dividing the MMD value of a specific inter-sample comparison by its standard deviation. Such standardized distances, functionally equivalent to z-scores, are more appropriate for evaluating relative separations in cases where sample sizes differ between populations (Sofaer et al. 1986). To be significant at the 95-97% level, the MMD value must exceed twice the standard deviation of the comparison.

Since the pattern of variation among 16 samples is difficult to visualize from a 120 cell diagonal matrix, standardized MMD values were used as input for two data reduction procedures. Following standard practice, all negative standardized MMD values were set at zero prior to submission to data reduction analyses. In the first procedure, the diagonal matrix of standardized MMD values was used as input for neighbor-joining cluster analysis (Saitou and Nei 1987; Felsenstein 1989). In the second procedure, the diagonal matrix of standardized MMD values was converted to a symmetric matrix and double-centered prior to submission to principal coordinates analysis (Hair et al. 1971; Rohlf 2000). The first three principal coordinate axes were retained, group scores were calculated along these axes, and ordinated into three-dimensional space. A minimum spanning tree (Hartigan 1975) was imposed on the array of principal coordinate scores to ease interpretation of inter-sample associations.

Results

Trait frequencies and number of observations for each of the 17 tooth-trait combinations in the 16 samples is provided in Table 2. Contingency chi-square tests indicate that 15 tooth-trait combinations differ significantly in frequency across these 16 samples (Table 3). The only exceptions are entoconulid presence on first and second mandibular molars (C6M1, C6M2). MMD values and standardized MMD values are provided in Table 4.

Three fundamental groupings of samples may be recognized in the neighbor-joining tree (Fig. 2). Occupying the left side of the array, Central Asian samples from Bactria are identified as possessing closest affinities to one another. The latest prehistoric Indus Valley sample, Sarai Khola (SKH), represents the South Asian sample most closely affiliated with these central Asians.

Table 2. Frequencies of Dental Morphology Traits

			Inamgaon			Mahars			Madia Gonds			Marathas			Sarai Khola			Timargarha		
Dental Trait	**Tooth**	**Abb.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**
Shoveling	UI1	ShovI1	22	24	0.917	77	186	0.414	80	163	0.491	81	198	0.409	3	9	0.333	7	7	0.143
Shoveling	UI2	ShovI2	13	19	0.684	22	181	0.122	23	161	0.143	24	194	0.124	2	9	0.222	2	7	0.286
Med. Lingual Ridge	UI1	MlrI1	14	25	0.560	106	177	0.599	60	153	0.392	95	194	0.490	2	9	0.222	3	8	0.375
Med. Lingual Ridge	UI2	MlrI2	1	20	0.050	41	174	0.236	21	149	0.141	42	190	0.221	0	9	0.000	0	7	0.000
Hypocone Size	UM1	HypoM	27	41	0.659	163	195	0.836	155	169	0.917	170	197	0.863	11	14	0.786	17	22	0.773
Hypocone Size	UM2	HypoM	0	20	0.000	10	164	0.061	10	153	0.065	4	179	0.022	2	13	0.154	0	13	0.000
Carabelli's Trait	UM1	CaraM1	13	40	0.325	140	187	0.749	86	165	0.521	122	198	0.616	2	9	0.222	9	18	0.500
Metaconule	UM1	MtclM1	6	41	0.146	43	191	0.225	36	156	0.231	56	193	0.290	3	9	0.333	4	19	0.211
Metaconule	UM2	MtclM2	3	20	0.150	33	153	0.216	34	138	0.246	32	169	0.189	2	14	0.143	0	13	0.000
Entoconulid	LM1	C6M1	4	37	0.108	13	191	0.068	12	158	0.076	17	194	0.088	1	14	0.071	0	22	0.000
Entoconulid	LM2	C6M2	0	24	0.000	3	174	0.017	5	152	0.033	5	191	0.026	0	15	0.000	1	18	0.056
Metaconulid	LM1	C7M1	2	36	0.056	25	191	0.131	27	165	0.164	15	198	0.076	1	15	0.067	2	24	0.083
Metaconulid	LM2	C7M2	1	25	0.040	3	177	0.017	7	158	0.044	1	197	0.005	0	15	0.000	2	20	0.100
Cusp Number	LM1	CspnM1	32	39	0.821	170	192	0.885	149	161	0.925	166	195	0.851	9	15	0.600	19	25	0.760
Cusp Number	LM2	CspnM2	4	24	0.167	30	178	0.169	32	158	0.203	28	192	0.146	1	15	0.067	3	17	0.176
Y-Groove Pattern	LM1	YgrvM1	32	35	0.914	115	127	0.906	112	115	0.974	117	128	0.914	5	7	0.714	12	17	0.706
Y-Groove Pattern	LM2	YgrvM2	7	24	0.292	30	161	0.186	31	133	0.233	51	181	0.282	5	14	0.357	3	18	0.167

Table 2 Continued.

			Neolithic Mehrgarh			Chalcolithic Mehrgarh			Harappa			Sapalli Tepe			Djarkutan		
Dental Trait	**Tooth**	**Abb.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**
Shoveling	UI1	ShovI1	25	28	0.893	21	25	0.840	8	15	0.533	2	19	0.105	3	16	0.188
Shoveling	UI2	ShovI2	31	37	0.838	21	24	0.875	10	16	0.625	10	17	0.588	8	22	0.364
Med. Lingual Ridge	UI1	MlrI1	15	26	0.577	14	25	0.560	8	12	0.667	4	17	0.235	3	17	0.176
Med. Lingual Ridge	UI2	MlrI2	2	29	0.069	7	24	0.292	6	13	0.462	5	17	0.294	4	22	0.182
Hypocone Size	UM1	HypoM	35	42	0.833	22	22	1.000	16	16	1.000	36	36	1.000	30	30	1.000
Hypocone Size	UM2	HypoM	2	41	0.049	10	18	0.556	2	18	0.111	23	32	0.719	21	32	0.656
Carabelli's Trait	UM1	CaraM1	7	27	0.259	11	18	0.611	4	9	0.444	8	25	0.320	3	21	0.143
Metaconule	UM1	MtclM1	7	28	0.250	5	19	0.263	6	13	0.462	3	37	0.081	1	29	0.034
Metaconule	UM2	MtclM2	10	25	0.400	6	18	0.333	4	16	0.250	2	34	0.059	0	32	0.000
Entoconulid	LM1	C6M1	3	37	0.081	5	23	0.217	1	20	0.050	3	25	0.120	1	20	0.050
Entoconulid	LM2	C6M2	0	44	0.000	1	23	0.043	0	28	0.000	0	40	0.000	0	36	0.000
Metaconulid	LM1	C7M1	4	40	0.100	3	25	0.120	1	22	0.045	1	38	0.026	1	32	0.031
Metaconulid	LM2	C7M2	0	43	0.000	0	24	0.000	0	28	0.000	0	43	0.000	1	39	0.026
Cusp Number	LM1	CspnM1	39	43	0.907	20	23	0.870	17	20	0.850	22	28	0.786	20	21	0.952
Cusp Number	LM2	CspnM2	3	49	0.061	2	24	0.083	0	33	0.000	2	41	0.049	2	36	0.056
Y-Groove Pattern	LM1	YgrvM1	23	25	0.920	15	21	0.714	15	17	0.882	19	24	0.792	20	22	0.909
Y-Groove Pattern	LM2	YgrvM2	12	37	0.324	6	22	0.273	3	31	0.097	7	38	0.184	11	35	0.314

Table 2 Continued.

			Kuzali						Gompadhompti Madigas			Chenchus			Pakanati Reddis		
Dental Trait	**Tooth**	**Abb.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**	**P**	**N**	**Freq.**
Shoveling	UI1	ShovI1	1	13	0.077	4	25	0.160	63	175	0.360	64	194	0.330	52	176	0.295
Shoveling	UI2	ShovI2	6	14	0.429	14	27	0.519	22	174	0.126	33	191	0.173	27	177	0.153
Med. Lingual Ridge	UI1	MlrI1	2	13	0.154	9	23	0.391	85	176	0.483	88	194	0.454	110	177	0.621
Med. Lingual Ridge	UI2	MlrI2	6	14	0.429	8	25	0.320	52	177	0.294	45	191	0.236	61	177	0.345
Hypocone Size	UM1	HypoM	23	23	1.000	41	41	1.000	177	178	0.994	192	193	0.995	177	182	0.973
Hypocone Size	UM2	HypoM	11	22	0.500	23	37	0.622	55	170	0.324	80	187	0.428	40	170	0.235
Carabelli's Trait	UM1	CaraM1	2	20	0.100	13	36	0.361	108	177	0.610	105	193	0.544	136	182	0.319
Metaconule	UM1	MtclM1	2	21	0.095	3	39	0.077	49	178	0.275	50	191	0.262	58	182	0.319
Metaconule	UM2	MtclM2	1	24	0.042	3	37	0.081	36	168	0.214	32	183	0.175	34	168	0.202
Entoconulid	LM1	C6M1	0	14	0.000	3	33	0.091	21	169	0.124	13	186	0.070	22	182	0.121
Entoconulid	LM2	C6M2	0	15	0.000	0	35	0.000	5	172	0.029	1	186	0.005	5	179	0.028
Metaconulid	LM1	C7M1	0	18	0.000	2	39	0.051	23	172	0.134	48	195	0.246	31	181	0.171
Metaconulid	LM2	C7M2	0	18	0.000	1	36	0.028	19	173	0.110	18	194	0.093	11	182	0.060
Cusp Number	LM1	CspnM1	10	15	0.667	29	33	0.879	169	171	0.988	188	192	0.979	173	181	0.956
Cusp Number	LM2	CspnM2	1	14	0.071	2	35	0.057	64	172	0.372	53	191	0.277	43	180	0.239
Y-Groove Pattern	LM1	YgrvM1	10	14	0.714	25	34	0.735	149	159	0.937	151	173	0.873	142	148	0.959
Y-Groove Pattern	LM2	YgrvM2	5	15	0.333	5	33	0.152	65	166	0.392	51	182	0.280	67	165	0.406

Remaining Indus Valley and peninsular Indian samples exhibit little affinity to central Asians and are arranged into distinct two aggregates.

Table 3. Contingency Chi-Square Analysis

Trait	X^2	p	df
ShovI1	208.743	0.000	15
ShovI2	84.260	0.000	15
MlrI1	51.512	0.000	15
MlrI2	44.304	0.000	15
HypoM1	132.487	0.000	15
HypoM2	406.221	0.000	15
CaraM1	125.093	0.000	15
MtclM1	35.855	0.002	15
MtclM2	35.364	0.002	15
C6M1	16.424	0.354	15
C6M2	18.357	0.244	15
C7M1	47.186	0.000	15
C7M2	47.189	0.000	15
CspnM1	276.573	0.000	15
CspnM2	133.802	0.000	15
YgrvM1	54.962	0.000	15
YgrvM2	42.882	0.000	15

The first aggregate of samples occupies the lower center of Figure 2 and is oriented about the phenetic space occupied by the three living samples from Andhra Pradesh. These three samples are loosely affiliated with one another such that the low-status caste Gompadhomptis (GPD) represent the most divergent sample from Andhra Pradesh, the high-status Pakanatis (PNT) are the most proximate to non-Andhra Pradesh samples. The tribal Chenchu (CHU) occupies a phenetic position intermediate between low- and high-status caste samples from Andhra Pradesh. Living samples from Andhra Pradesh are linked to other South Asian and Indus Valley samples by their affinities to samples from Chalcolithic levels at Mehrgarh (ChlMRG) and mature phase Harappans (HAR). Of these, the earlier sample from Mehrgarh exhibits closer affinities to living populations from Andhra Pradesh, while the later sample from Harappa exhibits closer affinities to non-Andhra Pradesh samples.

Table 4. Mean Measure of Divergence Analysis (*MMD values not bolded, standardized MMD values in bold*)

	INM	MHR	MDA	MRT	SKH	TMG	NeoMRG	ChlMRG	HAR	SAP	DJR	KUZ	MOL	PNT	CHU
MHR	0.0278 **1.8692**														
MDA	0.0085 **0.5624**	0.0191 **4.5097**													
MRT	0.0051 **0.3483**	0.0016 **0.4192**	0.0053 **1.2910**												
SKH	-0.0130 **-0.3077**	0.0963 **3.0213**	0.0584 **1.8153**	0.0411 **1.2948**											
TMG	-0.0531 **-1.3452**	0.0100 **0.3455**	0.0058 **0.1991**	-0.0087 **-0.3041**	-0.0765 **-1.3647**										
NeoMRG	-0.0042 **-0.1835**	0.0836 **6.8389**	0.0363 **2.8986**	0.0566 **4.6765**	0.0984 **2.4522**	0.0446 **1.2155**									
ChlMRG	0.1162 **4.1075**	0.1035 **5.8849**	0.1042 **5.8276**	0.1066 **6.1004**	0.1482 **3.2800**	0.0813 **1.9600**	0.0612 **2.3775**								
HAR	0.0391 **1.1953**	0.0255 **1.1369**	0.0488 **2.1523**	0.0150 **0.6734**	0.0257 **0.5132**	-0.0014 **-0.0292**	0.0636 **2.0739**	0.0502 **1.4055**							
SAP	0.1463 **5.7116**	0.1957 **13.0860**	0.1633 **10.7207**	0.1708 **11.5077**	0.0385 **0.9003**	0.0710 **1.7811**	0.2068 **8.9408**	0.0646 **2.2909**	0.0899 **2.7023**						
DJR	0.1266 **4.8237**	0.2276 **14.4873**	0.1521 **9.5112**	0.1850 **11.8572**	0.0576 **1.3242**	0.0833 **2.0692**	0.1674 **6.9962**	0.1020 **3.5114**	0.1393 **4.0914**	-0.0437 **-1.6303**					
KUZ	0.1186 **3.5401**	0.2209 **9.6757**	0.1662 **7.1904**	0.1664 **7.3281**	-0.0149 **-0.2956**	0.0408 **0.8666**	0.1778 **5.7522**	0.1008 **2.7870**	0.0759 **1.8693**	-0.0533 **-1.5822**	-0.0442 **-1.2811**				
MOL	0.1144 **4.9052**	0.1512 **12.1040**	0.1348 **10.5505**	0.1398 **11.2992**	0.0616 **1.5295**	0.0427 **1.1471**	0.1525 **7.3644**	0.0209 **0.8035**	0.0487 **1.5845**	-0.0514 **-2.1848**	-0.0263 **-1.0856**	-0.0328 **-1.0461**			
PNT	0.0819 **5.5187**	0.0277 **7.0658**	0.0510 **12.0978**	0.0334 **8.7586**	0.1177 **3.6948**	0.0619 **2.1490**	0.1272 **10.4216**	0.0758 **4.3143**	0.0267 **1.1908**	0.1626 **10.8781**	0.1895 **12.0656**	0.1893 **8.2981**	0.1258 **10.0769**		
CHU	0.0763 **5.1917**	0.0612 **16.1941**	0.0424 **10.4131**	0.0600 **16.3699**	0.0724 **2.2824**	0.0381 **1.3288**	0.1126 **9.3363**	0.0485 **2.7840**	0.0488 **2.1935**	0.0691 **4.6692**	0.0786 **5.0470**	0.0996 **4.3929**	0.0477 **3.8650**	0.0278 **7.3844**	
GPD	0.1568 **10.5486**	0.1255 **31.7019**	0.1118 **26.3443**	0.1275 **33.1830**	0.1513 **4.7457**	0.1383 **4.7960**	0.1935 **15.8148**	0.0626 **3.5541**	0.1210 **5.3987**	0.0869 **5.8010**	0.1047 **6.6563**	0.1432 **6.2675**	0.0786 **6.2821**	0.00519 **13.1756**	0.0173 **4.5607**

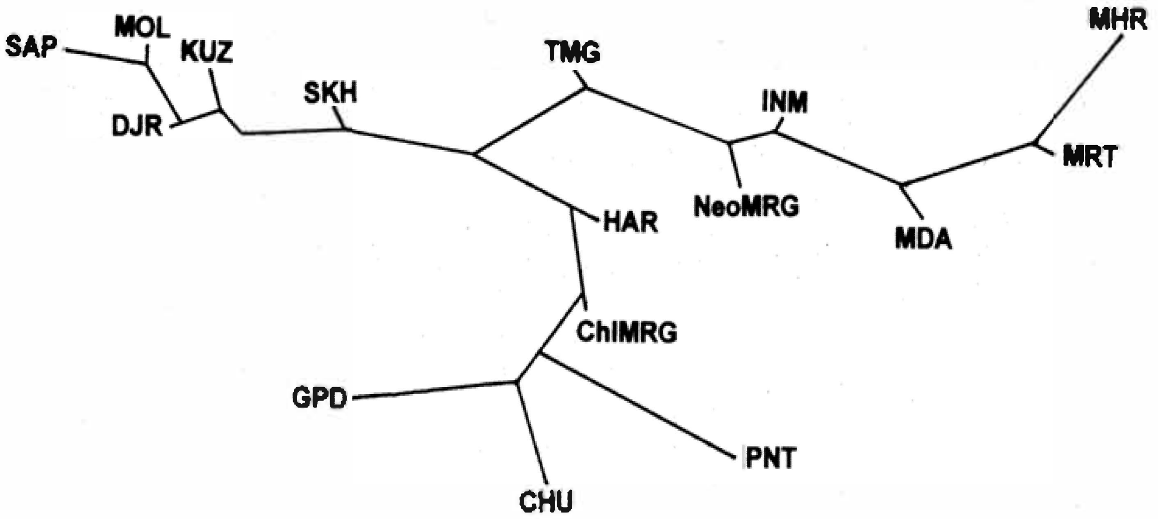

Figure 2: Neighbor-joining cluster analysis (abbreviations from Table 1)

The second aggregate occupies the right side of Figure 2 and is oriented about the phenetic space occupied by the three living samples from Maharashtra. The three samples from Maharashtra are loosely affiliated with one another such that low-status caste Mahars (MHR) represent the most divergent living sample from Maharashtra, while tribal Madia Gonds (MDA) are the most proximate to non-Maharashtrans. High-status caste Marathas (MRT) occupies a phenetic position intermediate between low-status caste Mahars and tribal Madia Gonds. Living samples from Maharashtra are linked to other south Asian and Indus Valley samples by their affinities to samples from Inamgaon (INM), Neolithic levels at Mehrgarh (NeoMRG), and post-Harappan Timargarha (TMG). Perhaps not surprisingly, it is the prehistoric sample from Maharashtra (INM) that bears closest phenetic affinities to living Maharashtrans. Together, these living and prehistoric inhabitants of western and central peninsular India share closest affinities to the early occupants of Neolithic Mehrgarh. The post-Harappan sample from Timargarha serves to link these samples to others from the Indus Valley as well as to those from central Asia and southeastern peninsular India.

Principal coordinate analysis yields three coordinate axes that account for 93% of the total variance. Three-dimensional ordination of sample scores for these axes results in three clearly defined sample aggregates (Fig. 3) that may be identified as Bactrians from central Asia, inhabitants of southeast India (Andhra Pradesh), inhabitants of western/central India (Maharashtra), and their associated Indus Valley samples, respectively.

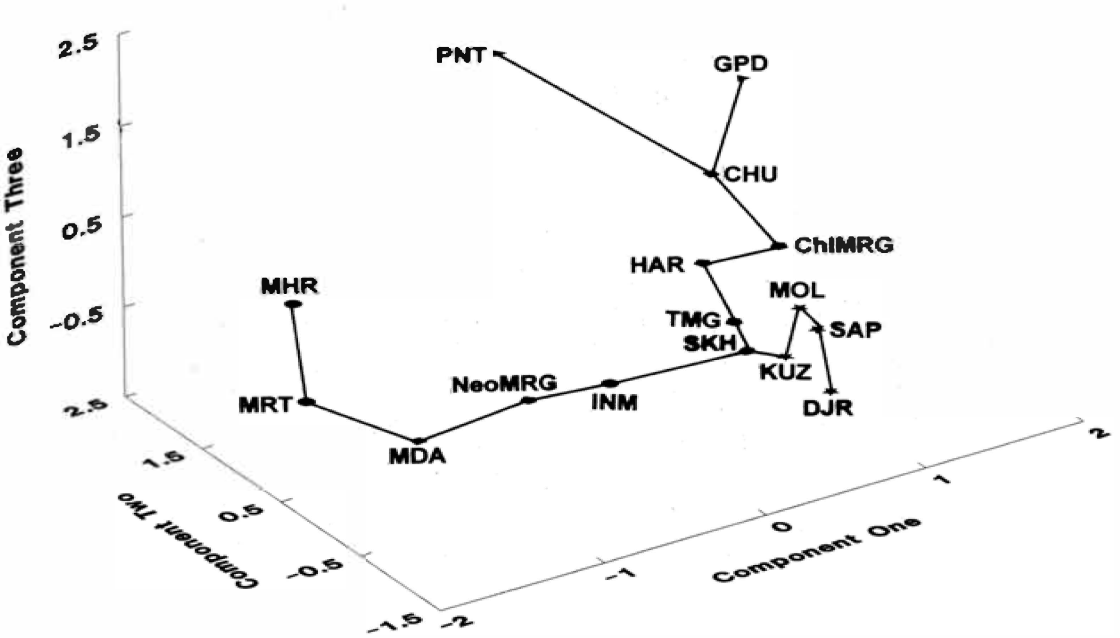

Figure 3: Principal Coordinates Analysis (abbreviations from Table 1)

Bactrian samples occupy the lower right of Figure 3 and manifest close affinities to one another. Among Bactrian samples, the Djarkutan period sample (DJR) is most divergent from non-Bactrians, while the Kuzali phase sample (KUZ) is most proximate. As identified by neighbor-joining cluster analysis, central Asian samples are linked to south Asian samples by the late prehistoric Indus Valley sample from Sarai Khola (SKH).

Samples from Andhra Pradesh occupy the upper center of Figure 3 and affinities among them are far more diffuse than those between Bactrian samples. As identified by neighbor-joining cluster analysis (Fig. 2), high-status caste Pakanatis (PNT) share only the most distant of affinities to their low-status (GPD) and tribal (CHU) counterparts from southeastern India. Connections between living inhabitants of Andhra Pradesh to other south Asians occur with prehistoric samples from the Indus Valley. These affinities are closest with the Chalcolithic inhabitants of Mehrgarh (ChlMRG), followed by mature phase Harappans (HAR), and most distantly, the post-Harappan inhabitants of Timargarha (TMG).

Maharashtran samples occupy the left side of Figure 3 and, like samples from Andhra Pradesh, affinities among them are rather diffuse. As identified by neighbor-joining cluster analysis, low-status Mahars (MHR) are the most divergent of living Maharashtrans, tribal Madia Gonds (MDA) are most proximate to non-Maharashtran samples, and high-status Marathas (MRT) occupy an intermediate position. At variance with neighbor-joining cluster analysis, principal coordinates analysis identifies living Maharashtrans as possessing closer affinities to the Neolithic inhabitants of Mehrgarh (NeoMRG) than to prehistoric Maharashtrans from Inamgaon (INM)

Discussion

If the Aryan Invasion Model is correct, the biological history of the Indus Valley should be marked by a strong discontinuity between those samples that antedate this intrusion and those that postdate it. Post-intrusion Indus Valley samples should be closely affiliated with central Asian samples, the proposed source for these intrusive

Indo-Aryan-speaking populations. Reflecting the alleged spread of Vedic culture, and hence genes through elite dominance, close affinities should exist between Late Indus Valley samples, central Asian samples, and Indo-Aryan-speaking castes. More distant affinities should occur with Dravidian-speaking castes, and there should be no biological affinities with non-Hindu tribal populations.

If the Out-of-India Model is correct—that Indo-Aryan languages arose in the Indian subcontinent and then spread to their historical seats outside South Asia through population movement—there have been no substantial population incursions into South Asia from outside. If true, the biological history of the Indus Valley should be one of long-standing continuity; in which, through the process of genetic drift, samples that differ most in antiquity and geographic propinquity should diverge most biologically. Likewise, patterns of biological affinity throughout peninsular India should largely be a product of geographic distance, regardless of caste adherence or status.

If the Early Intrusion Model is true, the introduction of significant numbers of proto-Elamo-Dravidian-speaking people into the Indus Valley during the 5th millennium B.C. should be signaled by a discontinuity in biological affinities between Indus Valley samples that antedate and postdate the 5th millennium. A second, albeit less pervasive, impact should signal the appearance of Indo-Aryan-speaking populations during the 2nd millennium B.C. Samples from peninsular India should reflect a separation of Indo-Aryan-speaking caste Hindus of west-central India from Dravidian-speaking caste Hindus of south India and caste Hindus of both regions should be markedly separated from non-caste tribals regardless of geographic propinquity.

Taken as a whole, the patterning of biological affinities reflects regional continuity in Maharashtra, southeast India, and Central Asia, but not the Indus Valley. Rather, the analyses reveal two bouts of biological discontinuity, resulting in three temporal sub-units possessing distinct affinities to different non-Indus Valley sample aggregates. The earliest Indus Valley sample (NeoMRG) has closest affinities to prehistoric and living inhabitants of Maharashtra. Chalcolithic era Indus Valley samples (ChlMRG, HAR) show closest affinities to Dravidian-speaking samples from southeast India, while the latest of the archaeologically-derived samples from the Indus Valley (SKH) is marked by some affinity to central Asians.

Conclusions

The Aryan Invasion Model is not substantiated by the results of this analysis. Contrary to expectations, post-Harappan Indus Valley samples are not marked by close affinities to central Asian samples, nor do they exhibit closer affinities to living Indo-Aryan-speaking caste members than to living Dravidian-speaking caste members. Likewise, there is no support for the Out-of-India Model. Indus Valley samples that date between 4500 and 1700 BC show no affinities to central Asian samples nor does the pattern of biological affinities attest to long-standing, uninterrupted biological continuity within the Indus Valley. Rather, the results offer best support for the Early Intrusion Model. As expected, a biological separation occurs between living Dravidian-speaking caste members of southeast India and Indo-Aryan-speaking caste members of Maharashtra. Further, these results reveal two bouts of biological discontinuity within the Indus Valley. The first, and by far the most profound, may signal the appearance of Dravidian-speaking populations during the 5th millennium B.C. The second may signal the arrival of central Asians, but contrary to expectations of the Aryan Invasion model, this introduction is far too mild in its biological impact to preserve any role for a large-scale Aryan invasion concomitant with the rise and initial spread of Vedic culture throughout south Asia.

Although the results discussed here are strongly supported, there are two caveats to be considered. First, with the sole exception of the limited Gandharan sample from Timargarha, we lack the crucial Late Bronze/Early Iron Age samples from the Hindu Kush borderlands that are most likely to preserve the remains of any central Asian invaders if they, in fact, ever existed. Second, our living samples come from peninsular India and from large inter-connected populations. As such, these are the very populations *least* likely to reflect any significant impact due to gene flow from any "Aryan invaders" of central Asia during the 2nd millennium B.C. Thus, future research efforts that seek to shed light on the Aryan invasion controversy need to address these deficiencies by including additional prehistoric and living samples from the not-so-remote environment of the Hindu Kush borderlands.

References

Ayala, F.J, 1982, *Population and Evolutionary Genetics: A Primer*, Benjamin/Cummings, Menlo Park.

Bryant, E, 2001, *The Quest for the Origins of Vedic Culture: The Indo-Aryan Migration Debate*, Oxford University Press, New Delhi.

Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A, 1994, *The History and Geography of Human Genes*, Princeton University Press, Princeton.

Chen, J., Sokal, R.R. & Ruhlen, M, 1995, Worldwide analysis of genetic and linguistic relationships of human populations, *Human Biology* 67, pp. 595-612.

Elst, K, 1999, *Update on the Indo-Aryan Debate*, Delhi, Aditya Prakashan.

Erdosy G, 1995, Language, material culture and ethnicity: theoretical perspectives. In: Erdosy G, editor, *The Indo-Aryans of Ancient South Asia*, Walter de Gruyter, Berlin, pp. 1-31.

Falconer, D.S, 1981, *Introduction to Quantitative Genetics*, 2nd Edition. Wiley, New York.

Fairservis, W.A. & Southworth, F.C, 1989, Linguistic archaeology and the Indus Valley culture, in Kenoyer, J.M. (Editor), *Old Problems and New Perspectives in the Archaeology of South Asia*, Wisconsin Archaeological Reports No 2, Madison, pp. 133-141.

Felsenstein, J, 1989, PHYLIP—phylogeny inference package (version 3.2), *Cladistics* 5, pp. 164-166.

Green, R.F. & Suchey, J.M, 1976, The use of inverse sine transformation in the analysis of non-metric cranial data, *American Journal of Physical Anthropology* 45, pp. 61-68.

Hair, J.F., Anderson, R.E., Tatham, R.L. & Grablowsky, B.J, 1971, *Multivariate Data Analysis*, PPC Books, Tulsa.

Harding, R.M. & Sokal, R.R, 1988, Classifications from the European language families by genetic distance, *Proceedings of the National Academy of Sciences of the United States of America* 85, pp. 9370-9372.

Hartigan, J.H, 1975, *Clustering Algorithms*, Wiley, New York.

Hartl, D.L. & Clark, A.G, 1997, *Principles of Population Genetics*, 3rd Edition, Sinauer Associates, Sudbury.

Hedrick, P.W, 2000, *Genetics of Populations*, 2nd Edition, Jones & Bartlett, Sudbury, 553.

Hemphill, B.E. & Lukacs, J.R, 1993, Hegelian Logic and the Harappan Civilization: An Investigation of Harappan Biological Affinities in Light of Recent Biological and Archaeological Research, in A.J. Gail and G.J.R. Mevissen (Editors), *South Asian Archaeology 1991*, Franz Steiner Verlag, Stuttgart, pp. 101-120.

Hemphill, B.E., Lukacs, J.R. & Kennedy, K.A.R, 1991, Biological Adaptations and Affinities of Bronze Age Harappans, in R. Meadow (Editor), *Harappa Excavations 1986-1990: A Multidisciplinary Approach to Third Millennium Urbanism*, Prehistory Press, Madison, pp. 137-182.

Kazanas N, 2002, Indigenous Indoaryans and the RgVeda, *Journal of Indo-European Studies 30*, pp. 275-335.

Lukacs, J.R, 1976, Anthropological Aspects of Dental Variation, Unpublished Ph.D. dissertation, Department of Anthropology, Cornell University, Ithaca.

MacAlpin, D, 1981, *Proto-Elamo-Dravidian: The Evidence and its Implications,* The American Philosophical Society, Philadelphia.

Mair, V.H, 1998, Priorities, in Mair, V.H. (Editor), *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia, Volume I,* University of Pennsylvania Museum Publications, Philadelphia, pp. 4-41.

Misra, S.S, 1992, *The Aryan Problem, a Linguistic Approach,* Munshiram Manoharlal, New Delhi.

Mukherjee, N., Nebel, A., Oppenheim, A. & Majumdar, P.P, 2001, High-resolution analysis of Y-chromosomal polymorphisms reveals signatures of population movements from Central Asia and West Asia into India, *Indian Academy of Sciences* 80, pp. 125-135.

Nichol, C.R, 1989, Complex segregation analysis of dental morphological variants, *American Journal of Physical Anthropology* 78, pp. 37-59.

Nichols, J, 1997, The epicenter of the Indo-European linguistic spread, in Blench, R. & Spriggs, M. (Editors), *Archaeology and Language I: Correlating Archaeological and Linguistic Hypotheses,* Routledge, London, pp. 220-266.

Nichols, J, 1998, The Eurasian spread zone and the Indo-European dispersal. In: Blench, R. & Spriggs, M., Editors. *Archaeology and Language II: Theoretical and Methodological Orientations,* Routledge, London, pp. 122-148.

Passarino, G., Semino, O., Bernini, L.F. & Santachiara-Benerecetti, A.S, 1996, Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms, *American Journal of Human Genetics,* 59, pp. 927-934.

Renfrew, C, 1987, *Archaeology and Language: The Puzzle of Indo-European Origins,* Cambridge University Press, New York.

Rohlf, F.J, 2000, NTSYS—PC Numerical Taxonomy and Multivariate Analysis System, Version 2.1. Exeter Software, Setauket.

Rothhammer, F., Quevado, S., Cocilovo, J.A. & Llop, E, 1984, Microevolution in prehistoric Andean populations: chronologic nonmetrical cranial variation in northern Chile, *American Journal of Physical Anthropology* 65, pp. 157-162.

Saitou, N. & Nei, M, 1987, The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Molecular Biology and Evolution* 4, pp. 406-425.

Scott, G.R. & Turner II, C.G, 1997, *The Anthropology of Modern Human Teeth,* Cambridge University Press, Cambridge.

Sjøvold, T, 1973, The occurrence of minor non-metrical traits in the skeleton and their quantitative treatment for population comparisons, *Homo* 24, pp. 204-233.

Sofaer, J.A., Smith, P. & Kaye, E, 1986, Affinities between contemporary and skeletal Jewish and non-Jewish groups based on tooth morphology, *American Journal of Physical Anthropology* 70, pp. 265-275.

Sokal, R.R., Oden, N.L. & Thomson, B.A, 1988, Genetic changes across language boundaries in Europe, *American Journal of Physical Anthropology* 76, pp. 337-361.

Southworth, F.C, 1995, Reconstructing social context from language: Indo-Aryan and Dravidian prehistory, in Erdosy G (editor), *The Indo-Aryans of Ancient South Asia,* Walter de Gruyter, Berlin, pp. 258-277.

Talageri, S, 2000, *Rigveda.A Historical Analysis,* Aditya Prakashan, New Delhi.

Turner, CG, Nichol, C.R., & Scott, G.R,. 1991, Scoring procedures for key morphological traits of the permanent dentition: The Arizona State University dental anthropology system, in Kelley, M.A. & Larsen, C.S. (Editors), *Advances in Dental Anthropology,* Wiley-Liss, New York, pp. 13-31.

Witzel, M, 1999, Substrate languages in old Indo-Aryan (Rgvedic, middle and late Vedic), *Electronic Journal of Vedic Studies* 5, pp. 1-67.