Ancient Pakistan, Vol. XXXIV, 2023: 29–79
ISSN: 0066-1600(print)/ 2708-4590(online)

Population Dynamics among Ethnic Groups Residing in Hazarewal and Chitral-Gilgit-Baltistan:

A Dental Morphology Investigation of Biological Interactions across the Northwestern Borderlands of South Asia since the Chalcolithic Era

Brian E. Hemphill*

Professor, Department of Anthropology, University of Alaska, USA.

Abstract: The ethnic groups inhabiting the lowland and highland terrain in far northern Pakistan have recently been suggested to have served as middlemen in an interregional exchange network between Turkestan and the Indus Valley — an exchange network that may have involved populations at the northern end of the Inner Asian Mountain Corridor or populations of the Bactrian Margianan Archaeological Complex (BMAC) of southern Central Asia. Alternatively, it may be that waves of Pathan immigration during the mediaeval period into northern Pakistan introduced such extensive gene flow from the west that all trace of contacts north has been lost. Patterning of morphological trait frequencies of the permanent tooth crown are known to correlate closely with DNA patterns and are largely free of natural selection making them ideal for tracing patterns of past gene flow and genetic drift. Dental morphometric trait frequencies from living members of 11 ethnic groups of the lowlands of Hazarewal and the highlands of Chitral-Gilgit-Baltistan, along with nine archaeologically-derived samples from the Khyber Pakhtunkhwa, Indus Valley, and southern Central Asia (n= 2576 individuals), were analysed. Varying operational taxonomic units (OTUs) and data reduction techniques reveal a striking absence of evidence for interactions with populations to the north, meagre evidence of Pathan-related gene flow from the west or northwest, and little support for a genetic contribution from Chalcolithic era populations from the Indus Valley, signaling long-term local biological continuity in this region of northern Pakistan.

Keywords: Pathans, Non-Pathans, Highland, Lowland, Mediaeval, Biodistance.

Introduction

Historical records and oral traditions maintain that there was an eastward movement of a whole series of Pathan ethnic groups from southern Afghanistan to the Vale of Peshawar and beyond during the first half of the second millennium CE. According to Caroe (1958), these militant invaders vied with the local population to wrest the fertile bottomlands west of the Indus River from members of the indigenous ethnic groups they encountered there. Recent molecular genetic studies have not only confirmed close affinities between extant Pathan populations of Afghanistan and Pakistan (Achakasi *et al*. 2012; Di Cristofaro *et al*. 2013; Haber *et al*. 2012, Lacau *et al*. 2012; but see Qamar *et al*. 2002) but have also hinted at much earlier incursions, either of Chalcolithic era (c. 3500 – 3000 BCE) Aeneolithic farmers from the urban centres of the Kopet Dagh foothill plain of southern Central Asia (Anthony 2007; Bhatti *et al*. 2017, 2018; Constantini 1984, 2008; de Barros Damgaard *et al*. 2018; Fairservis 1971; Gadgil *et al*. 1997; Gangal, Sarson & Shukurov 2014; Harris 1997a, 1997b; Jarrige 2006; Jarrige & Hassan 1989; Jarrige & Lechevallier 1980) or from later Bronze Age era (c. 2500 – 1500 BCE) incursions of horse-mounted Aryan-speaking warriors of the Russo-Kazakh steppe (Erdosy 1995; Kuzmina 2001; Narasimhan *et al*. 2019; Parpola 1988). Archaeological evidence corroborates these Chalcolithic and Bronze Age contacts between populations of the Indus Valley and populations located further to the north in Khyber Pakhtunkhwa and Chitral-Gilgit-Baltistan. While it has long been assumed that the directionality of these contacts was from north to south, perhaps facilitating the trade in tin via the Inner Asian Mountain Corridor

---

* ✉ bhemphill@alaska.edu

(Frachetti 2012; Kohl & Lyonnet 2008; Lyonnet 2005; Vinogradova 1993, 1994), discovery of the Harappan outpost at Shortughai in northeastern Afghanistan (Francfort 1981, 1984, 2016) and objects of Harappan manufacture at Namazga IV levels at Altyn Depe in southern Central Asia (Gupta 1979; Masson 1988) suggest the direction of exchange was either bidirectional or largely from south to north.

The current research seeks to address four important questions:

- What is the most appropriate operational taxonomic unit (OTU) for patterns of biodistance among members of these ethnic groups from Chitral-Gilgit-Baltistan and Hazarewal? Is it the ethnic group regardless of sex (i.e., sexes-pooled), ethnic groups segregated by sex (i.e., females only, males only), or is it ethnic groups in which a single model identified both ethnic group and sex (*i.e.,* sexes-stipulated)?
- Is there evidence for a Pathan presence in Hazarewal and Chitral-Gilgit-Baltistan? If so, one ought to find a kind of background scatter of biodistances among an array of long-standing indigenous populations of this region punctuated by an intrusion of a foreign element from the west and/or northwest.
- Were ethnic groups of this region differentiated spatially by adaptation to different biomic niches as suggested by Barth (1956)? If so, one ought to expect Pathan populations to dominate the rich lowland plains, while weaker indigenous ('tribal') populations were peripheralized to the adjacent highlands.
- Is there evidence of temporally distinct population incursions and settlements in Hazarewal and Chitral-Gilgit-Baltistan that reflect early intrusion from the Indus Valley to the south, from later prehistoric influence from southern Central Asia to the north and an historic intrusion of Pathans from the west?

Dynamics of Khyber Pakhtunkhwa Populations

A wide array of scholars claim that the dawn of the Christian era witnessed many invasions of Khyber Pakhtunkhwa by swarm after swarm from the north (Caroe 1958; Ghirshman 1954; McGovern 1939; Smith 1924 [2022]). However, by the mid-first millennium CE influences from South Asia via the Mauryan Empire extended through the Vale of Peshawar and perhaps as far west as Kabul. Some, such as Bellew (1872 [1998]), suggest that the ethnic groups of this region were bisected into two factions: Gar and Samil (see Caroe 1958: 61). It is held that this division reflects a pre-Islamic distinction with the Gar adhering to the older Central Asian Zoroastrian faith, while the Samil converted to the newly-introduced South Asian Buddhist faith. It is at the opening of the second millennium CE that Mahmud of Ghazni led an expanding military empire that extended from southeastern Afghanistan to encompass the Vale of Peshawar. Additional raids were conducted in the Punjab and Kashmir over the next several decades, but these raids were either unsuccessful or were undertaken under alliances with local populations. Over the next several centuries Islamic influence in South Asia was extended leading to the establishment of the Delhi Sultanate in 1206 CE (Avari 2013). However, early Islamic influence appears to have been greater in peninsular India than in Pakistan with an array of communities in the former identifying as 'Afghan' (Digby 2004), perhaps due to the impact of maritime trade during the late first millennium CE (Islam 2017).

The first Islamic entrants into Khyber Pakhtunkhwa were non-Pathan Dilazak tribals whose origins may be traced to southeastern Afghanistan (Caroe 1958). These populations were displaced by a later eastward movement by Swati Pathans who, in turn, were displaced during the 16th century by Yusufzai Pathans. Both groups of Pathans are held to have come from the Kabul Valley, though there is some debate as to whether their ultimate origins are to be found in the Kandahar Valley of south-central Afghanistan or to the north in Badakhshan. The duality of soft (Pashtun) and hard dialects (Pakhtun) suggests that multiple entries of Afghan Pathan populations were involved. The diversity of castes in lower Khyber Pakhtunkhwa has been further augmented by infiltration through local lineages claiming descent from the Prophet Muhammad, His collateral family, or from prominent saints. Barth (1969) maintains that the majority of these 'saints' likely had origins in Turkestan in southern Central Asia, for rival 'saint' groups from Persia, representing Shi'ism, have been largely unsuccessful in colonisng the nearly exclusively Sunni Khyber Pakhtunkhwa[1].

Pathan Origins

While it is commonly believed that Pashtun origins may be found in the Kandahar region of south-central Afghanistan dating to some point between the first and second millennia BCE, the specific ancestral population remains elusive. This may reflect the fact that Pashtun origins are to be found in a confederacy of ethnic groups (Lal 1846) or it may be due to multiple origins (Vogelsang 2002). Alternatively, linguistic analyses suggest that the Pashto language may have originated in the Badakhshan region of northeastern Afghanistan (Rahman & Kamran 2001). Indeed, Morgenstiern (1982) suggested that Pashto may be a Saka dialect related to the extinct Sogdian language of Bactria in southern Central Asia (Beča 1969).

If the Urheimat of the Pathans is to be found to the west of Khyber Pakhtunkhwa, the plains encompassed by the Vale of Peshawar is marked by three primary points of entry. Immediately to the west is the Khyber Pass, which provides direct access from Kabul in the west to Peshawar to the east via Jalalabad. The second is the Kohat or Darrah Pass located south of Peshawar that provides access to the southern districts of Khyber Pakhtunkhwa (i.e., Bannu, Dera Ismail Khan, Hangu, Karak, Kohat, Lakki Marwat, Tank). The third is the Malakand Pass located to the north of Peshawar, which provides access to the Bajaur and Mohmand agencies of the former FATA (Federally Administrated Tribal Area; now merged districts) as well as to the central districts of Khyber Pakhtunkhwa (i.e., Buner, Charsadda, Lower Dir, Lower Swat, Mardan, Nowshera, Peshawar, Swabi).

It is possible that the entry of Pathans into northern Pakistan did not occur once and did not involve a single point of entry. Indeed, the classical literature makes clear that as an ethnic group, Pathans are not to be conceptualized as a monolithic whole, but are marked by divisions. According to the classical literature, while Qais is acclaimed the hero of the Pathans, this literature also states he had three sons, Sarbanr, Bitan, and Ghurghusht. Sarbanr, the eldest son, is held to have sired two sons, Sharkhbun and Kharshbun. Sharkhbun is maintained to be the ancestor of the Western Afghans, namely the Durranis and kindred tribes, while Kharshbun is claimed to be the ancestor of the Eastern Afghan tribes as well as those residing within the Vale of Peshawar and the adjacent mountains to the north, namely the Yusufzai, the Mohmands, the Khalils, the Daudzais, and the Muhammadzais (Caroe 1958).

The only 'tribe' descended in the male line from Bitan, the second son of Qais, is the Bhitannis, a well-known but small group located on the eastern flank of Wazir and Mahsud country where the hills drop to the plains of Bannu. All the rest, including the largest Pashtun-speaking group of all, the Ghajlis, as well as the progenitors of two Pathan dynasties in Delhi (Lodhis and Surs) are said to be descended through the female line.

The family tree of Ghurghusht, the third and youngest son of Qais, is shorter and their descendants are maintained to be the Kakars of Zhob, Lorallai, and parts of Sibi. As such, they represent a very large tribe in extreme southern Khyber Pakhtunkhwa and northern Baluchistan. Being neighbours of the Baluch tribes, they have adopted some of the Baluch exaltation of their chiefs and are less anarchic than most Pathans (Caroe 1958: 19). The Gaduns—those that remember their Pashto—are a small tribe located on the southern slopes of the Mahaban Mountain, close to Sitana and Topi, near the western (right) bank of the Indus. In Buner and Swabi, on the right bank of the Indus, and in Hazarewal, on the other (left) bank of the Indus, is a larger group of the same tribe, with the name altered to Jadoon. The latter, living in the Rash plain adjacent to Abbottabad, are held by Caroe (1958) to have forgotten both their language (Pashto) and the ways of their ancestors and today speak the form of Panjabi current in Hazarewal.

Thus ends the genealogies of the three sons of Qais. But, queries Caroe (1958), how to account for the ancestries of such famous ethnic groups as the Afridis, the Khataks, and the Orakzais? Where are the Bangash, the Mahsuds, and the Wazirs? According to Caroe (1958), the answer is that the genealogies of all of these groups appear to have been slipped into the classical literature, almost as an after-thought. Given the stated genealogy of Qais, there ought to be three main lines of his descendants representing the progeny of his three sons Sarbanr, Bitan, and Ghurghusht, but in fact there are claimed to be four such lines. According to McMahon (1901), the ancestor of the fourth was one Karlanr, or Karlanri, who is claimed to have been adopted by Burhan, another son of Ghurghusht and was subsequently married to an Urmar (non-Pashtun) woman[2]. As Karlanri is the putative ancestor of most of the hill-tribes, he is of particular importance. It is maintained that Karlanri and his Urmar wife had two sons, Koday and Kakay. Koday had two wives. The first produced three sons, Utman, Dilazak, and Orak. Their descendants gave rise to the Utman Khel, Dilazaks (extinct), and Orakzais. The second wife had four sons: Mani, Luqman, Mangal, and Khugi. These four sons give rise to the Afridis, Khataks, Mangals, and Khugianis, respectively.

It has been asserted that the River Indus represented the easternmost border of actual population movement of Pathans into Khyber Pakhtunkhwa, for Pathan influence further east was either undertaken by local proxies or involved a process of elite dominance (Dyson 1993; Friedel 1986; Renfrew 1987; Tilley 1984). Caroe (1958) states that these immigrant Pathans always traded with the cities and towns toward the Indus, and not with the populations of Kabul or Ghazni to the west. Consequently, their links with the Eastern Afghans of the Peshawar Valley have been much closer than with the Durranis or other tribes inhabiting the regions further to the west. Conversely, the Eastern Afghans feel an undoubted sense of identity with the hill-tribes, a sense that has hardly as yet amounted to any concept of unity but transcends tribal particularism. These are the Pakhtuns or Pathans—tribes who never fell under the effective sway of any recorded imperial authority and now form the backbone of the so-called tribal belt.

Pathan Segmentation and Interrelationships

Barth (1956) asserts that while the culture area concept has been used effectively in North America (Kroeber 1939; Wissler 1927) it has proven problematic in describing Asian cultures (Bacon 1946; Kroeber 1947; Miller 1953; but see Blackburn 2009). Instead, Coon (1951) likens the arrangements of Middle Eastern societies as forming a mosaic wherein ethnic groups with radically different lifeways co-reside in the same area sharing symbiotic relationships of varying intensity. In a similar vein, Furnivall (1944) described societies of the Dutch West Indies as forming a plural society in which one finds ethnic segmentation coupled with economic interdependence. As a result, the environment occupied by any one ethnic group is defined not only by external natural or ecological conditions, but also by the presence and activities of the other ethnic groups upon which it depends. Such segmentation and co-habitation is only possible if each ethnic group utilises only a portion of the total environment thereby leaving large parts of that natural environment open for other ethnic groups to exploit.

Barth (1969) states that it is crucial to appreciate that the populations residing in the lowlands of Khyber Pakhtunkhwa—and the population of Lower Swat is no exception—have a firmly Hindu foundation. This foundation dates back to first millennium CE with the expansion of the Mauryan Empire that overlaid the pre-existing Gandharan Buddhist tradition west of the River Indus extending to the Kabul Valley, encompassing the Vale of Peshawar and much of the low-lying regions of southern Khyber Pakhtunkhwa (Dietz 2007; Falk 2003; Nasim Khan 2001; A. Samad 2011; R. Samad 2011; Scott 1985; Vogelsang 1988). According to Barth, the Muslim invasions of Pathans beginning in 1000 CE caused no break in local tradition. Instead, it appears that conversion to Islam from a Buddhist-Hindu foundation was likely due to the imposition of this religion by a small group of warrior lords through the process of elite dominance (Renfrew 1987) while the bulk of the population maintained its secular Indian traditions (but see Raza 2010-2011).

The Hindu foundation of local traditions in lowland Khyber Pakhtunkhwa is reflected by occupational specialization (Barth 1969). Here, occupations are rigidly segregated and cannot be combined, except in cases where the occupations are considered compatible. Even personal versatility is unusual. Nevertheless, the products or services of specialists in each of the occupations are all considered equally essential. Therefore, all of the occupations must be represented in a self-sufficient Pathan community. Yet, occupational status does not equate with caste status, for caste status is ascribed to individuals by virtue of their paternity, whereas occupations are the subject of individual choice in the face of pragmatic exigency. The reinforcing factors behind occupational caste are two-fold. First, all occupations are required for self-sufficiency but there is little personal versatility. Second, traditionally, exchange largely occurs through a non-monetary system of *jajmani*-like reciprocal services (Caldwell 1991).

Being Islamic, but grafted on to a local Hindu traditional system, there is much greater latitude for inter-caste mobility. In Khyber Pakhtunkhwa, this is manifested three ways. First, are claims of Pathan descent and discouragement of intermarriage with non-Pathans or even Pathans of lower social standing (Barth 1965; Caroe 1958). Second, is the preference for parallel cousin marriage. Third, reflecting the fundamental equality of all espoused by Islam, hypergamous unions especially in the case of economically successful families is not uncommon (Gilmartin 1994; Wright 1994). Such social group mobility is greater in Islamic Pakistan – accounting for some 40% of all marriages in Swat – than Hindu India, because social status is not bound up with notions of inborn purity (Ahmad 1962; Barth 1969; Mines 1972; Momin 1977).

However, the determination of caste status (*qoum*) occurs strictly through the male line as a consequence of paternity, whereas occupational status (*kasb*) permits a degree of pragmatic exigency. In this way, castes in Khyber Pakhtunkhwa are arranged into a series of clusters in which individual caste status has a degree of quasi-equality and porosity that not only offer the opportunity for social dynamism (both upward and downward) but also encourages hypergamy as a means for effecting a rise in social status and integration that is far, far rarer among the Hindu-based castes of India (Gough 1959; Kapadia 1954). Nevertheless, upon marriage a woman abandons her former status as a member of her father's household and is fully integrated into the household and kin group of her husband. In this way, women are viewed as a form of tribute from the weak to the strong. In this way, hypergamous marriages are firmly grounded within the ethos of the Pathan caste system (Barth 1969).

Although the relationship between immigrant Pathans and members of non-Pathan groups is often portrayed, both in the classical literature and in more contemporary accounts (cf. Caroe 1958) as one of conquest and conflict, Barth (1956, 1969) notes that such relationships are actually more complex and involve four factors. The first is that the distribution of ethnic groups is not controlled by fixed 'natural areas' but by the distribution of specific ecological niches to which a group, through economic and political means is able to exploit. In Swat, what appeared as a single natural area to Kohistanis has been subdivided into habitable and non-habitable regions for Pathans. We can designate this modality as *niche perception*[3]. The second aspect highlighted by Barth is that different ethnic groups will establish themselves in stable co-residence in an area if their subsistence strategy exploits different ecological niches, especially if they can establish symbiotic economic relations as described by Barth (1956) for Pathans and Gujjars. We can designate this modality as *niche segregation*. The third aspect occurs if different ethnic groups are able to exploit the same niches fully. In such cases, the militarily more powerful will displace the weaker, just as the more militaristic plough- agriculturalist Pathans displaced their weaker plough-agriculturalist Kohistani counterparts in the highly productive lowlands of Swat. This modality may be termed *competitive niche exclusion*. The fourth and final aspect occurs when different ethnic groups exploit the same ecological niches but the weaker is better able to utilise marginal environments, which can be observed among Gujjars and Kohistanis in eastern Kohistan. This modality may be termed *niche commensalism*.

Pathans live in a complex multi-caste society organized into localized, segmentary unilineal descent groups (Barth 1956, 1965, 1969). Other castes and occupational groups are tied to them as political clients and as economic serfs in a *jajmani*-like organic system of exchange. Subsistence is based on diversified and well-developed plough agriculture, but only part of the Pathan population is actively engaged in agriculture. Various occupational groups, still accorded Pathan status, perform specialized services in return for payment-in-kind, thereby requiring the agricultural segment to generate a considerable surplus. Consequently, Pathan territory is limited to areas in which two crops per year can be grown, which in this region is largely a function of altitude and is limited to the lower parts of Khyber Pakhtunkhwa.

As noted above, Pathans live within a multi-caste society into which Gujjars are assimilated as an occupational caste of herders. Thus, it is not uncommon for Pathan villages to encompass a small number of Gujjars. They may speak their native Gujri as their mother tongue and retain their separate culture, or they may be assimilated to the extent of speaking only Pashto. Their role is to care for the animals, either as servants of a landowner or as independent herdsmen. They contribute to the village economy with milk products, meat, and manure, which is important and used in the fields. In addition to their agricultural land most Pathan villages also control the neighbouring hills or mountainsides. Transhumant Gujjars, who rely primarily on water buffalo, shift their flocks to these higher areas for summer pasturage for which they pay a fixed rate in kind per animal. Transhumant Gujjars tend to be associated with specific Pathan villages to which they pay pasturage rent in animal products (especially clarified butter) and provide agricultural labour during the seasons of peak activity for which they receive agricultural produce and other Pathan assets. This is a clear example of long-standing niche segregation as well as short-term niche commensalism.

Finally, the different economic strategies and differences in organizational complexity has yielded a clear distinction which can be drawn between those who inhabit plains and open plateau on the one hand and highlanders on the other (Caroe 1958). The former have always been regarded as the senior branch of the western immigrants from Afghanistan and hence are especially entitled to the name, 'Afghan.' Additionally, these plain and plateau-dwellers can be sub-divided geographically into the Western Afghans, of whom the most important are the Abdalis (now known as the Durranis) and the Ghaljis, and the Eastern Afghans, namely the Yusufzais and other kindred Pathan ethnic groups of the Peshawar plain and the valleys to the north of it (Bellew 1864 [2013]; Ullah *et al*. 2017). Interposed between the Western and Eastern Afghans and bordering the latter to the north and east to some extent keeping them apart are the highlanders. These include most of the famous ethnic groups of Khyber Pakhtunkhwa—Afridi, Khatak, Orakzai, Bangash, Wazir, Masud and Turi. All are presented in genealogical legend as descended from a founding common ancestor named Karlanri, but are not regarded as being in the true Afghan line. Others include the Abbasis, Awans, Gujjars, Karlaars, Khowars, Swatis and Tanolis. These are likely indigenous ethnic groups, some of whom may have mixed intrusive and indigenous origins (i.e., Swatis, Tanolis), but who never fell under the effective sway of any recorded imperial authority, and who now form the constellation of ethnic groups that constitute the so-called 'tribal belt' (Watson 1908).

Materials and Methods

Materials

This research is based on the examination of dental casts made of the permanent teeth of members of living ethnic groups of Hazarewal and Chitral-Gilgit-Baltistan as well as prehistoric permanent teeth recovered from archaeological contexts in the Indus Valley, Khyber Pakhtunkhwa, and southern Central Asia. Viewed as a whole, the sample consists of 2297 living individuals from 11 ethnic groups, four of which (Awans, Gujjars, Syeds, Wakhis) were multiply sampled from geographically distinct localities. Prehistoric remains of 279 individuals were examined from seven sites, one of which—Djarkutan—is represented by three-time successive samples. These sites are located in southern Central Asia, the Indus Valley, and Khyber Pakhtunkhwa. They range in antiquity from the Chalcolithic (c. 3500 – 3000 BCE) to the Iron Age (c. 200 – 100 BCE) (Fig. 1). Taken as a whole the total number of individuals considered in the current study numbers some 2576 individuals (Table 1).

Methods

The methodology employed may be divided into two aspects: trait selection and biodistance analysis. Data was collected for a battery of 215 tooth-trait combinations encompassed by the Arizona State Dental Anthropology System (ASUDAS: Scott, Maier, & Heim 2016; Scott & Turner 1997; Turner, Nichol, & Scott 1991). Since these traits were recorded from dental casts among living individuals the entire battery of ASUDAS variables could not be assessed (*i.e.*, root traits). Where possible, the dentition of each individual was assessed for 26 dental traits scored as 53 tooth-trait combinations. Frequencies of dental traits were calculated for each grade of expression following the procedure described in Pojar, Langstieh and Hemphill (2022) in which observations were made on both right and left antimeres and the greatest expression regardless of size was recorded (Green, Suchey and Gokhale 1979; Harris and Sjøvøld 2004; Scott 1977, 1980; Scott & Turner 1997). All casts were scored by BEH whose scoring concordance was found to be highly reliable (Pojar et al. 2022). Several of the archaeologically derived samples (Sarai Khola, Timargarha) were assessed by Lukacs (1983, 1986, 1987). An unpublished assessment of inter-assessor concordance in trait assessment between Lukacs and Hemphill found minimal differences.

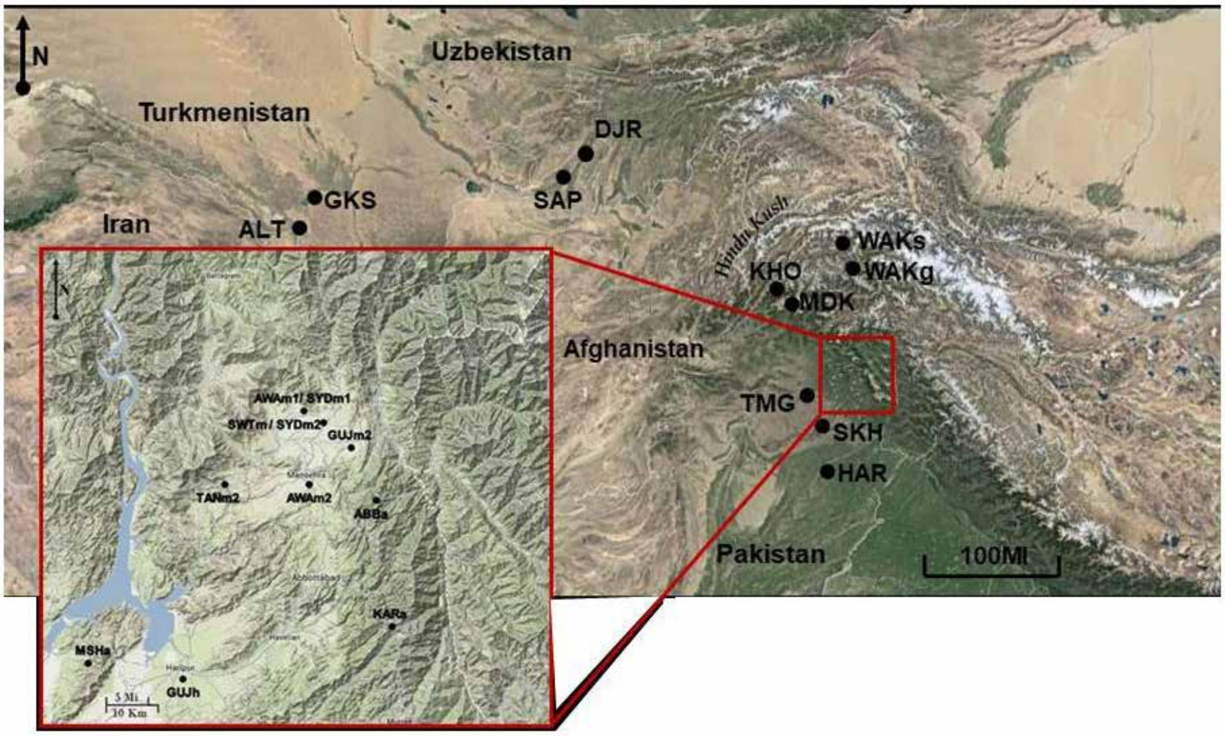

Figure 1. Map of sample locations.

Table 1. Samples used in the study.

Living Individuals

Sample	Abb.	Location	Source	N_{max}[1]
Abbassis	ABBa	Hazarewal; Haripur District	Akbar et al. (2023)	98
Awans	AWAm1	Hazarewal; Mansehra District	Hemphill (2012)	165
Awans	AWAm2	Hazarewal; Mansehra District	Akbar et al. (2023)	135
Gujjars	GUJh	Hazarewal; Haripur District	Akbar et al. (2023)	126
Gujjars	GUJm2	Hazarewal; Mansehra District	Akbar et al. (2023)	133
Karlaars	KARa	Hazarewal; Haripur District	Akbar et al. (2023)	112
Khowars	KHO	Hindu Kush; Chitral District	Blaylock & Hemphill (2007)	263
Madak Lasht	MDK	Hindu Kush; Chitral District	Hemphill (2010)	179
Mashwanis	MSHa	Hazarewal; Haripur District	Akbar et al. (2023)	135
Swatis	SWTm	Hazarewal; Mansehra District	Hemphill (2013)	189
Syeds	SYDm1	Hazarewal; Mansehra District	Akbar et al. (2023)	151
Syeds	SYDm2	Hazarewal; Mansehra District	Akbar et al. (2023)	154
Tanolis	TANm2	Hazarewal; Mansehra District	Akbar et al. (2023)	150
Wakhis	WAKg	Hindu Kush; Gilgit-Baltistan (Gulmit)	This Report	145
Wakhis	WAKs	Hindu Kush; Gilgit-Baltistan (Sost)	This Report	162
TOTAL LIVING INDIVIDUALS				2297

Prehistoric Individuals

Sample	Abb.	Antiquity	Location	Source	N_{max}
Altyn Depe	ALT	2500 – 2200 BC	Central Asia	Akbar et al. (2023)	18
Djarkutan	DJR	2100 – 1950 BC	Central Asia	Hemphill (2011)	39
Geoksyur	GKS	3500 – 3000 BC	Central Asia	Akbar et al. (2023)	41
Harappa	HAR	2600 – 1900 BC	Indus Valley	Hemphill et al. (1991)	33
Kuzali	KUZ	1950 – 1800 BC	Central Asia	Hemphill (2011)	24
Molali	MOL	1800 – 1650 BC	Central Asia	Hemphill (2011)	41
Sarai Khola	SKH	200 – 100 BC	Indus Valley	Lukacs (1983)	15
Sapalli tepe	SAP	2300 – 2100 BC	Central Asia	Hemphill (2011)	43
Timargarha	TMG	1400 – 850 BC	Indus Valley	Lukacs (1983)	25
TOTAL PREHISTORIC					279
TOTAL (ALL INDIVIDUALS)					2576

The selection of traits is of great importance in biodistance analysis (Harris 2008; Harris & Sjovøld 2004; Hemphill 2013; Irish 2010; Sjovøld 1977). As noted by Akbar, Inamullah, Ahmad, Ali and Hemphill (2023), two considerations must be taken into account when deciding which traits to include in such studies as this, which incorporates dental samples of both living and archaeologically-derived individuals. The first is that archaeologically-derived samples are often relatively few in number and biased with regard to representativeness. The potential for non-representativeness stems from two factors. First, there is an increased likelihood of loss of single-rooted teeth that leads to an under-representation of the anterior teeth, in addition there are often heightened levels of tooth wear, especially among the posterior teeth, of prehistoric individuals that obscures traits of the occlusal surface of the crown. Second, individual traits, let alone the expression of individual traits on multiple teeth within morphogenetic fields[4], are unlikely to be controlled by separate genes or even batteries of genes (see Keene 1989; Jernvall & Jung 2000; Jernvall & Thesleff 2000; Kangas, Evans, Thesleff, & Jernvall 2004; Osborn 1978; Salazar-Ciudad, Jernvall, & Newman 2003; Townsend, Harris, Lesot, Clauss, & Brook et al. 2009; Tucker & Sharpe 2004; Weiss 1990).

In a recent study, Rathmann and Reyes-Centano

(2020) identified which tooth-trait combinations yield patterns of biodistance that adhere most closely to those obtained from neutral genetic markers in a worldwide sample. Such an approach suffers from a pair of weaknesses. First, the traits most useful for distinguishing between samples are not universal across humanity but depends upon the array of samples being compared (Akbar *et al.* 2023; Harris 2008; Pojar et al. 2022; Sjovøld 1977). Therefore, the battery of variables considered in a biodistance analysis must first be tested for their discernment in distinguishing among the included samples. Second, Rathmann and Reyes-Centano only considered trait prevalence on the so-called 'key' teeth identified by Scott and Turner (1997). By doing so they failed to assess whether, by ignoring potentially unique information provided by non-key teeth within morphogenetic dental fields, key aspects of morphological differentiation between samples were not taken into account. Consequently, with regard to this second consideration, it is important to determine whether specific tooth-trait combinations are inter-correlated (Harris 2008; Sjovøld 1977), and if so whether these traits vary in the same fashion across the samples considered (Hemphill, Inamullah, Tariq, Zubair, Saadiq & Ahmad *et al*. 2018, Hemphill, Pojar & Langstieh 2019; Inamullah, Ahmad & Hemphill 2017; Tariq, Ahmad & Hemphill 2017).

The same three-step editing procedure, based on those developed by Irish (2010) and Harris and Sjovøld (2004), and elaborated upon by Akbar et al. (2023) and Pojar et al. (2023), was employed here. First, determination of inter-trait correlation of the original data (53 tooth-trait combinations: 23 mandibular, 30 maxillary in raw form (*i.e.,* by ordinal grade) was undertaken with Kendall's *tau*-b correlation coefficient among all samples of living individuals[5]. All tooth-trait combinations found to have more than 10% of their variation accounted for by another variable ($\tau_B > 0.217$) were eliminated from further consideration.

In step 2, the remaining traits were dichotomized into presence-absence categories. In most cases, any expression of a trait was considered a positive expression. However, positive scores for shoveling required a score of grade 2 or higher, hypocone development was considered fully expressed if exceeding grade 3.5, and a positive value for cusp number among the mandibular molars required retention of the hypoconulid (grade 5 or above). Those tooth-trait combinations demonstrating low inter-trait correlations with other traits were submitted to Kruskal-Wallis *H* tests (the nonparametric equivalent to a one-way ANOVA) to determine whether these traits contribute significantly to inter-sample variance in tooth-trait combination frequencies. These univariate Kruskal-Wallis *H* tests were followed with a Sheirer-Ray-Hare (1976) test, the nonparametric equivalent of a two-factor ANOVA, which permits testing the relative influence of the two main effects (ethnic group, sex) and their interaction.

In step 3, correspondence analysis was used to determine whether a trait scored on different members of the same morphogenetic field yielded redundant information. A plot of trait centroids for the first two dimensions was visually inspected. If the trait occupied the same, or very similar, positions with regard to directionality (orientation) and distance (magnitude) from the origin, the trait was considered redundant. Tooth-trait combination centroids occupying distinctly different vectors of divergence or distances from the origin were considered non-redundant and retained.

Contrasts in trait frequencies among living individuals were undertaken with sexes pooled, with sexes separated, and with sexes specified. With sexes pooled, samples by ethnic group include all individuals, both males and females without any adjustment in sample frequencies for inter-sex differences in trait prevalence. In this case, the operational taxonomic unit (OTU) is the ethnic group with all individuals pooled together. When sexes were considered separately, the OTU is the ethnic group composed of only female members or only male members, respectively. This was undertaken to control for the possibility of differential migration of dependent children due to long-term family separation (usually for wage earning prospects) or because of deceased or remarried spouses. With sexes specified, a single analysis considers biodistances among males and females of each ethnic group. In this instance, the operational taxonomic unit is the sex segregated members of an ethnic group, such as Abbasis males and Karlaar females, in a single analysis that takes both sex and ethnic affiliation into consideration.

These contrasts involved two procedures. In the first, correspondence analysis and principal components analysis were used to quantify relative differences in trait prevalence profiles directly from the patterning of dichotomized trait occurrence by individual. With correspondence analysis, obtained scores for the first two dimensions were plotted and the patterning of inter-sample affinities was identified by visual inspection. With principal components analysis, all components with eigenvalues in excess of 1.0 were retained and the most influential variable loadings described. Sample scores for the first three components were plotted in three-dimensional space and a minimum spanning tree (Hartigan 1975) was imposed on the array of points to ease interpretation of the patterning of intersample associations (for a similar approach see Arita 1997; Gillison 1978; Smit, de Geus, Boersma, Boomsma, & Stam 2016). Correspondence analysis was performed in SPSS Version 26 (IBM Corporation 2016). Principal components analysis, three-dimensional plots, and minimum spanning trees were accomplished with SYSTAT Version 11 (SYSTAT Software Inc. 2004).

In the second procedure, dichotomized trait frequencies were compared using Smith's mean measure of divergence (MMD) statistic with Freeman and Tukey's (1950) angular adjustment and Green and Suchey's (1976) correction for high- and low-frequency traits. Once negative distances were reset to zero, the patterning of pairwise inter-sample differences reflected by the triangular matrix of Smith's MMD values was simplified with neighbor-joining cluster analysis (Saitou and Nei 1987) with PHYLIP (Felsenstein 1989), while nonmetric multidimensional scaling with Kruskal's (1964a, 1964b) stress formula 1 was undertaken with SYSTAT. Multidimensional scaling was accomplished in the first three dimensions and goodness of fit was assessed through the amount of stress incurred in fitting the model. Since individual level data was either not available or was marked by high amounts of missing data for the archaeologically-derived samples, contrasts of living groups with these latter samples was limited to the second analytical procedure described above (*i.e.,* neighbor-joining cluster analysis, multidimensional scaling).

Results

Trait Selection

While 215 tooth-trait combinations are encompassed by the ASUDAS, there are factors that limit how many of these can be included in the research design. Some 98 tooth-trait combinations were either invariable or unscorable in the data set. These were due to trait fixity, trait absence, or the use of plaster casts that rendered root traits unscorable. Another 42 tooth-trait combinations had to be eliminated due to the demographic profile of the sampled living individuals, for in an effort to maximize the number of healthy tooth crowns by individual, volunteers were recruited between 16 and 19 years of age. Consequently, traits affecting the third molars could not be assessed as they had not yet erupted. Finally, those tooth-trait combinations identified as being excessively correlated with other such combinations also had to be eliminated. These steps yielded a variable list of 17 tooth-trait combinations. However, Kruskal-Wallis *H* tests (Table 2) identify four tooth-trait combinations among females that fail to differ significantly across samples: full development of the hypocone on UM1, presence of Cusp 5 on UM2, and presence of the entoconulid (Cusp 6) and metaconulid (Cusp 7) on LM2. For males, six tooth-trait combinations were identified as not differing significantly across samples. These included presence of Cusp 5 on UM2, development of the hypoconulid (CSPN) on LM1 and LM2, presence of the entoconulid on these same teeth, and presence of the metaconulid on LM2. When sexes were pooled, Kruskal-Wallis *H* tests failed to identify significant differences across samples for five tooth-trait combinations: full development of the hypocone on UM1, presence of Cusp 5 on UM1, presence of the entoconulid on

Table 2. Kruskal-Wallis tests for Significant Differences in Trait Frequencies across Samples of Living Individuals by Ethnic Group[1]

		Females		Males		Sexes Pooled	
Trait	Tooth	H	p	H	p	H	p
SHOV	UI1	103.438	>0.0001	106.267	>0.0001	205.063	>0.0001
SHOV	UI2	95.192	>0.0001	112.504	>0.0001	197.083	>0.0001
MLR	UI1	89.858	>0.0001	76.978	>0.0001	154.137	>0.0001
MLR	UI2	28.504	0.0120	35.535	0.0012	54.632	>0.0001
HYPO	UM1	16.622	0.2771*	27.582	0.0164	20.448	0.1170*
HYPO	UM2	47.564	>0.0001	27.095	0.0190	39.378	>0.0001
CARA	UM1	92.557	>0.0001	100.344	>0.0001	818.573	>0.0001
C5	UM1	388.042	>0.0001	427.457	>0.0001	22.345	0.0721*
C5	UM2	12.564	0.5615*	21.585	0.0886*	185.441	>0.0001
YGRV	LM1	32.524	0.0033	32.284	0.0044	55.913	>0.0001
YGRV	LM2	44.610	>0.0001	36.017	0.0011	64.069	>0.0001
CSPN	LM1	37.632	0.0031	16.930	0.2606*	43.239	>0.0001
CSPN	LM2	32.604	0.0033	18.106	0.2023*	36.178	0.0010
C6	LM1	26.423	0.0232	12.864	0.5379*	21.101	0.0993*
C6	LM2	17.091	0.2516*	17.328	0.2393*	23.080	0.0590*
C7	LM1	26.923	0.0200	31.781	0.0042	34.685	>0.0001
C7	LM2	17.738	0.2194*	13.311	0.5025*	14.811	0.3917*

1. Nonsignificant contrasts across the 15 samples of living individuals at $\alpha < 0.05$ marked by an asterisk.

LM1 and LM2, and presence of the metaconulid on LM2. With sexes specified, seven tooth trait combinations did not differ significantly across samples. These included full development of the hypocone and presence of Cusp 5 on UM1, development of the hypoconulid and presence of the entoconulid on LM1 and LM2, as well as the presence of the metaconulid on LM2. These tooth-trait combinations were eliminated from further consideration. Consequently, correspondence analysis by sample, principal components analyses, mean measure of divergence values, neighbor-joining trees, and multidimensional scaling plots were based on 12 combinations with sexes pooled, 13 tooth-trait combinations among females, 11 combinations among males, and 10 combinations with sexes specified.

As noted above, an additional concern with regard to trait selection is redundancy in variability of a morphological trait within a morphogenetic field. Correspondence analysis by tooth-trait combination permits assessment of trait variability by vector and by magnitude. Some 12 traits may be assessed with correspondence analysis when sexes are pooled among living individuals. Redundancy in variability can be assessed for four traits. These include shoveling and presence of the median lingual ridge on the maxillary incisors as well as the presence of the Y-groove and development of the hypoconulid (CSPN) on the first two mandibular incisors. None of these traits exhibit redundancy in variation across samples with regard to the vector of variation or magnitude (Fig. 2). Correspondence analysis among the significantly variable 13 tooth-trait combinations among living females is presented in Figure 3. Four traits, shoveling and medial lingual ridge on the maxillary incisors as well as presence of the Y-groove and cusp number on the first two mandibular molars are all identified as presenting non-redundant vectors and magnitudes of variance across samples. Correspondence analysis among the 11 remaining tooth-trait combinations among living males is presented in Figure 4. As with females, four traits could be assessed for redundancy within morphogenetic fields. These include shoveling and the median lingual ridge among the maxillary incisors, full development of the hypocone on the first two maxillary molars, and presence of the Y-groove on the first two mandibular molars. While at first glance, presence of the median lingual ridge appears to be quite similar on UI1 and on UI2, the magnitude of differentiation across samples is greater for the former than the latter. Thus, as observed among females, none of these tooth-trait combinations proved to yield redundant patterns of variation across samples. When sexes are specified among living individuals, three traits could be assessed for variability redundancy within morphogenetic field: the presence of shoveling and medial lingual ridge development on the maxillary incisors and presence of the Y-groove on the first two mandibular molars. None exhibit redundancy (Fig. 5).

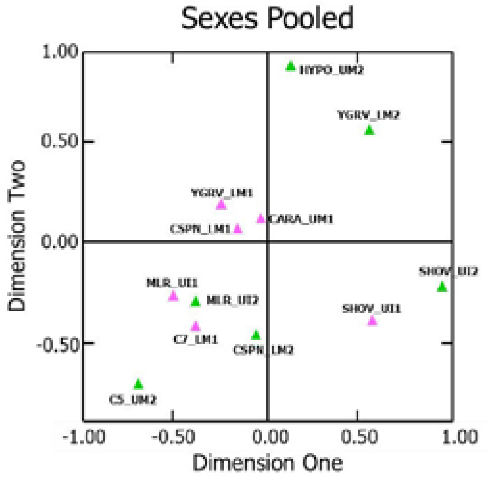

Figure 2. Correspondence Analysis of Tooth-trait Variability within Morphogenetic fields among Living Individuals with Sexes Pooled. Key Teeth within Morphogenetic Fields in Purple, Distal Teeth in Green.

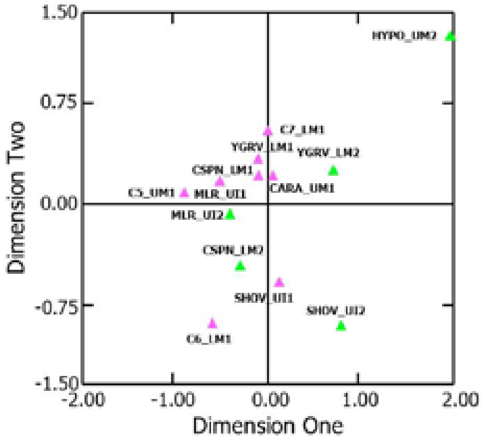

Figure 3. Correspondence Analysis of Tooth-trait Variability within Morphogenetic Fields among Living Females. Key Teeth within Morphogenetic Fields in Purple, Distal Teeth in Green.

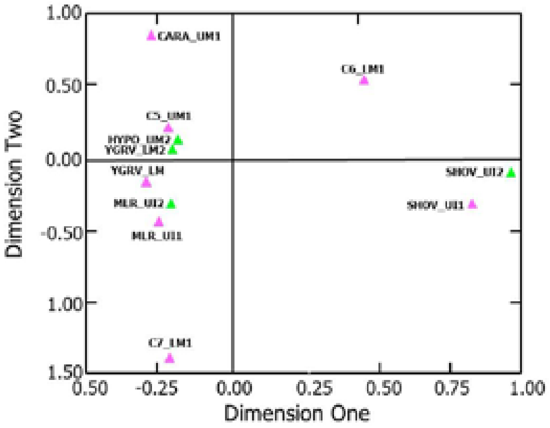

Figure 4. Correspondence Analysis of Tooth-trait Variability within Morphogenetic Fields among Living Males. Key Teeth within Morphogenetic Fields in Purple, Distal teeth in Green.

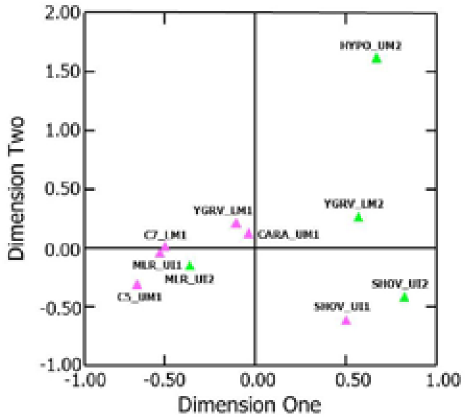

Figure 5. Correspondence Analysis of Tooth-trait Variability within Morphogenetic Fields among Living Individuals with Sexes Specified. Key Teeth within Morphogenetic Fields in Purple, Distal Teeth in Green.

Biodistance Analysis

Sexes Pooled

A plot of the first two dimensions obtained by correspondence analysis among members of living samples with sexes pooled yields three aggregates and four outliers (Fig. 6a). The first aggregate is found in the middle-right of the array and includes four samples: the two samples of Awans, Khowars, and Tanolis. The second aggregate is found in the lower left and includes three samples: the two samples of Wakhis and Swatis. The third aggregate is present in the center-right. It encompasses four samples: the two samples of Syeds and the two samples of Gujjars. Outliers are found in the upper right (ABB_a), the extreme centre-right (KAR_a), and the lower left (MSH_a, MDK). Intriguingly all but one of these (MDK) were sampled in Abbottabad District. When considered by ecological setting (Fig. 6b) lowland groups tend to occupy positions in the upper right, while highland groups occupy the lower left. Again, however, there are exceptions, for Swatis and Mashwanis occupy positions adjacent to highland samples, while the Khowars are surrounds by lowland samples. When ethnicity is taken into account (Fig. 6c) there is no separation between groups claiming Pathan origin and those believed to be of non-Pathan descent.

Principal component analysis of tooth-trait frequencies across members of living samples with sexes pooled yields four components with eigenvalues greater than 1.0 (Table 3). The first accounts for 29.8% of the total variance and draws a distinction between shoveling on UI1 and cusp number on the first two mandibular molars. The second component accounts for 23.1% of the variance. It draws a distinction between samples with relatively high prevalence of the median lingual ridge on the maxillary incisors versus samples that have relatively high prevalence of shoveling on UI2, Carabelli's trait on UM1 and the Y-groove on LM2. Component three accounts for 15.7% of the variance. It draws a distinction between samples with relatively high prevalence of the Y-groove on the mandibular molars and the metaconulid on LM1 versus samples with relatively high prevalence of the median lingual ridge on UI2 and retention of a well-developed hypoconulid on LM1. Component 4 accounts for 11.8% of the variance and it draws a contrast between samples with relatively high prevalence of shoveling on UI1 and presence of the metaconulid on LM! versus samples with retention of a fully developed hypocone and presence of Cusps 5 on UM2.

A plot of scores for the first three components (Fig. 7a) yields three sample aggregates and five isolates. These three components combine to

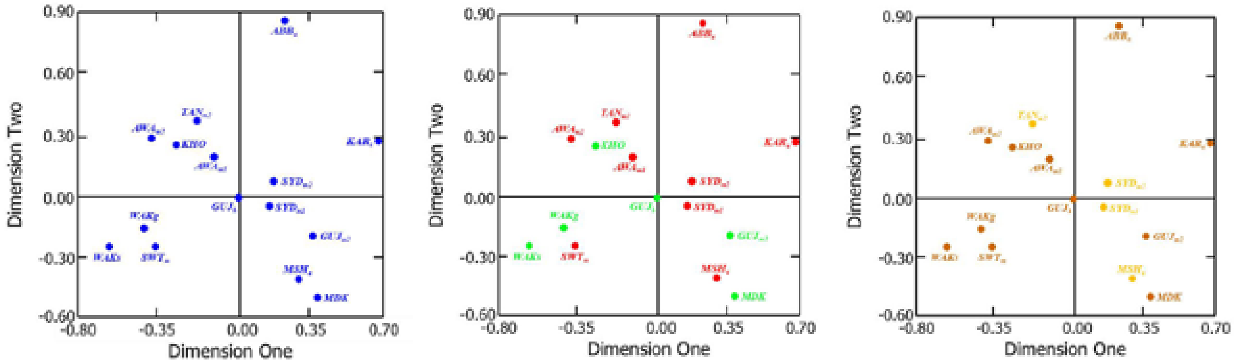

Figure 6. Correspondence Analysis of Tooth-trait Variability across Samples of Living Individuals with Sexes Pooled: a) All Samples, b) Highland Samples (in green) versus Lowland Samples (in red), c) Pathan (in light brown) versus Non-Pathan Samples (in purple).

Table 3. Principal Components Analysis among Samples of Living Sex-pooled Individuals[1]

		Loadings			
Trait	Tooth	1	2	3	4
SHOV	UI1	***0.707***	-0.473	0.091	***0.405***
SHOV	UI2	0.566	***-0.667***	0.142	0.387
MLR	UI1	0.412	***0.690***	0.199	0.018
MLR	UI2	0.602	***0.593***	***-0.324***	-0.103
HYPO	UM2	-0.084	-0.379	***0.485***	***-0.716***
CARA	UM1	0.672	***-0.485***	-0.191	-0.059
C5	UM2	0.672	0.379	0.352	***-0.460***
YGRV	LM1	0.003	0.343	***0.771***	0.174
YGRV	LM2	0.209	***-0.590***	***0.598***	-0.137
CSPN	LM1	***0.750***	-0.005	***-0.308***	-0.355
CSPN	LM2	***0.795***	0.116	-0.161	0.017
C7	LM1	0.279	0.527	***0.485***	***0.433***
Eigenvalue		3.575	2.776	1.878	1.416
% Var. Expl.		29.790	23.130	15.650	11.800

1. Most influential variables by component are in bold and italicized.

account for 68.6% of the total variance. The first aggregate is found on the left side and is a tight association of the two Wakhi samples, the two Awan samples and Khowars. The second occupies the left foreground and encompasses a loose association between Tanolis, Abbasis and Karlaars. The third aggregate occupies the background center of the array. It includes the two samples of the Gujjars and one of the Syed samples (SYD_{m1}). The four outliers include the second Syed sample (SYD_{m2}) in the lower foreground, Mashwanis and residents of Madak Lasht on the right side and Swatis in the upper right background. When considered by ecological setting (Fig. 7b) highland groups are split on the left, right and center background of the array, while lowland groups tend to occupy positions in the center and left foreground. Two glaring exceptions to this pattern are Mashwanis and Swatis who occupy positions on the extreme right and upper background, respectively. When sex pooled samples are considered by ethnicity (Fig. 7c), groups claiming Pathan ancestry tend to occupy positions at the bottom of the array, while groups believed to be of non-Pathan ancestry occupy positions higher in the array. There are, however, several exceptions with the Gujjar sample from Haripur District (GUJ_h) and the residents of Madak Lasht tending to occupy positions more like those of Pathan-affiliated groups.

The zero-corrected pairwise matrix of mean measure of divergence distances across samples of living individuals with the sexes pooled together is presented in Table 4. Some 78 of these pairwise contrasts (74.3%) are significant at $\alpha < 0.05$ level. When considered by ecological setting 37 of 53 pairwise contrasts (69.8%) differ between highland and lowland groups differ significantly, while 37 of 44 of pairwise contrasts (84.1%) differ between those who claim Pathan ancestry and those believed to be of non-Pathan origin. Such results suggest that in strong contrast to when consideration is limited to females, sex-pooled samples of living ethnic groups from Hazarewal and Chitral-Gilgit-Baltistan clearly differ from one another with greater discernment (+14.3%) when attributed ancestry is taken into account than when consideration is limited to ecological setting (-14.3%).

A neighbour-joining cluster analysis based on the matrix of pairwise mean measure of divergence values among sex-pooled samples of living individuals is provided in Figure 8a. Two aggregates and three isolates may be identified. The first aggregate is composed of six samples and is found on the left side of the array. It includes the two Awan samples, who share close affinities to one another, the two Wakhi samples, who likewise show close affinities to one another, Khowars and Swatis. The second aggregate occurs in the center of the array and encompasses five samples. These include one of the Gujjar samples (GUJ$_h$), Tanolis, the two Syed samples, who are identified as very similar to one another, and Abbasis. The other sample of Gujjars (GUJ$_{m2}$) stands as an intermediate between the second aggregate and three highly isolated samples. These isolated samples include Mashwanis, Karlaars, and the residents of Madak Lasht. When considered by ecological setting (Fig. 8b) there appears to be no consistent pattern. Highland groups, while largely concentrated on the left side of the array, also occupy positions of the right side with the inhabitants of Madak Lasht being the most extreme. The same is true of lowland groups, for such samples are widely dispersed across the array and separated from one another by highland samples. A similar lack of patterning is evident when samples are considered by ethnicity (Fig. 8c). While Pathan groups tend to occupy the center of the array, their affinities to one another are interrupted by one of the Gujjar samples (GUJ$_{m2}$) and the sample of Abbasis.

Multidimensional scaling into three dimensions with Kruskal's stress formula number 1 was accomplished in three iterations and accounted for 99.0% of the variance with a stress level of 0.039. This is considered an excellent fit of the data (Clarke 1993; De Leeuw & Stoop 1984). Two aggregates may be identified among the sex pooled samples of living individuals (Fig. 9a). In the upper center and left side there

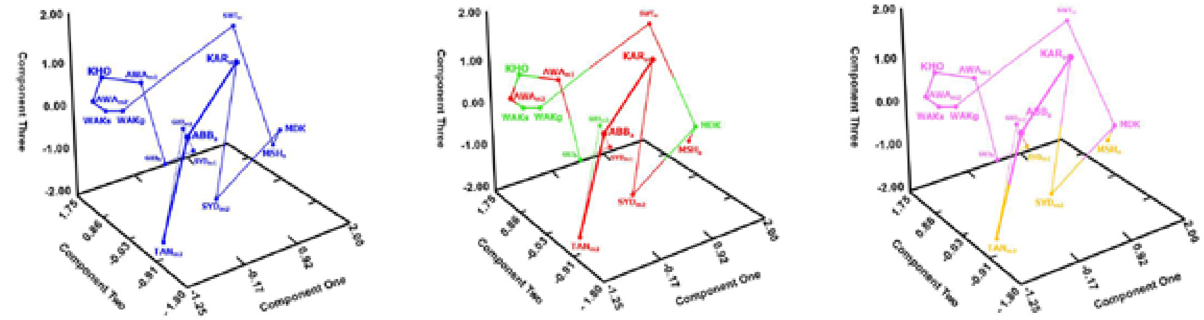

Figure 7. Three-dimensional Plot of Sample Centroids among Samples of Living Individuals yielded by Principal Components Analysis with Sexes Pooled: a) All Samples, b) Highland Samples (in green) versus Lowland Samples (in red), c) Pathan (in light brown) versus Non-Pathan Samples (in purple).

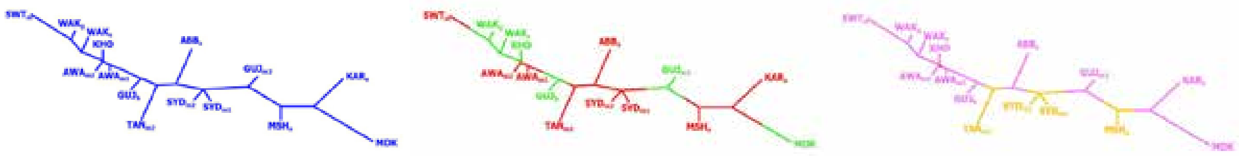

Figure 8. Neighbor-joining Cluster Analysis among Samples of Living Individuals with Sexes Pooled: a) All Samples, b) Highland Samples (in green) versus Lowland Samples (in red), c) Pathan (in light brown) versus Non-Pathan Samples (in purple).

Table 4. Zero-corrected Mean Measure of Divergence Analysis among Samples of Living Individuals Only with Sexes Pooled[1]

	ABBa	**AWAm1**	**AWAm2**	**GUJh**	**GUJm2**	**KARa**	**KHO**	**MDK**	**MSHa**	**SWTm**	**SYDm1**	**SYDm2**	**TANm2**	**WAKg**	**WAKs**
ABBa	---	0.020	0.010	0.015	0.011	0.018	0.009	0.022	0.014	0.026	0.011	0.011	0.011	0.019	0.010
AWAm1	0.037	---	0.007	0.012	0.008	0.015	0.007	0.020	0.012	0.024	0.008	0.008	0.008	0.016	0.007
AWAm2	**0.028**	0.004	---	0.013	0.009	0.016	0.008	0.020	0.013	0.025	0.009	0.009	0.009	0.017	0.008
GUJh	**0.041**	0.001	0.019	---	0.008	0.017	0.009	0.021	0.014	0.026	0.010	0.010	0.010	0.018	0.009
GUJm2	**0.047**	**0.016**	**0.050**	0.008	---	0.016	0.008	0.020	0.013	0.025	0.009	0.009	0.009	0.017	0.008
KARa	**0.063**	**0.057**	**0.116**	**0.061**	0.012	---	0.006	0.021	0.014	0.026	0.010	0.010	0.010	0.018	0.009
KHO	**0.030**	0.004	0.000	**0.023**	**0.043**	**0.096**	---	0.019	0.011	0.024	0.007	0.007	0.006	0.015	0.006
MDK	**0.175**	**0.079**	**0.157**	**0.075**	0.033	**0.046**	**0.138**	---	0.011	0.024	0.008	0.007	0.007	0.016	0.007
MSHa	**0.074**	**0.043**	**0.081**	**0.030**	0.006	**0.037**	**0.079**	0.021	---	0.025	0.010	0.009	0.009	0.018	0.009
SWTm	**0.130**	0.032	**0.068**	**0.058**	**0.081**	**0.122**	**0.063**	**0.083**	**0.072**	---	0.007	0.008	0.008	0.016	0.007
SYDm1	**0.024**	**0.026**	**0.028**	0.008	0.001	**0.048**	**0.035**	**0.079**	0.012	**0.097**	---	0.008	0.010	0.018	0.009
SYDm2	**0.026**	**0.023**	**0.027**	0.009	0.005	**0.049**	**0.033**	**0.077**	0.017	**0.092**	0.000	---	0.007	0.017	0.008
TANm2	**0.040**	**0.041**	**0.032**	0.017	**0.059**	**0.126**	**0.036**	**0.159**	**0.084**	**0.100**	**0.028**	**0.026**	---	0.017	0.008
WAKg	**0.090**	0.005	0.012	0.014	**0.047**	**0.117**	0.012	**0.094**	**0.064**	0.029	**0.045**	**0.040**	**0.039**	---	0.006
WAKs	**0.098**	**0.017**	0.014	**0.026**	**0.064**	**0.151**	**0.012**	**0.135**	**0.094**	**0.059**	**0.061**	**0.059**	**0.046**	0.000	---

1. MMD values below the diagonal, standard deviations above the diagonal, significant differences at $\alpha < 0.05$ in bold. Abbreviations from Table 1.

is loose aggregate of five samples that includes the two samples of Syeds, Mashwanis, and one of the samples of Gujjars (GUJ$_{m2}$). Peripherally associated with them are Tanolis in the upper center, Karlaars in the lower left, and residents of Madak Lasht in the left foreground. In the lower center is a second aggregate of three samples. These include the two samples of Awans and Khowars. The other sample of Gujjars (GUJh) stands as an intermediate between the Syeds of aggregate one and the two samples of Wakhis, which occupy a highly isolated position on the right side. The remaining two samples, Abbasis and Swatis occupy isolated positions in the center background and right foreground, respectively. No distinct pattern is present when inter-sample affinities are considered between highland and lowland samples (Fig. 9b). In dramatic contrast, a distinct pattern may be observed when samples are considered by ethnicity (Fig. 9c), with a single exception (MSHa),: groups claiming Pathan ancestry exhibit closest affinities to one another in the upper centre, while groups believed to be of non-Pathan origin are widely scattered throughout the left, right and foreground of the array.

Females Only

A plot of the first two dimensions obtained by correspondence analysis among all living females by sample yields four sample aggregates (Fig. 10a). The first is a fairly tight cluster in the upper left that includes one of the samples of Awans (AWA$_{m2}$), the sample of Khowars (KHO), Swatis (SWT$_{m}$) and one of the samples of Wakhis (WAK$_{g}$). The second is a loose aggregate in the upper right that includes Abbasis (ABB$_{a}$), Tanolis (TAN$_{m2}$) and Karlaars (KAR$_{a}$). The third is an aggregate found in the center of the array that is divided into two sub-clusters, one to the left the other in the center. On the left are three samples that include the second sample of Awans (AWA$_{m1}$), the second sample of Wakhis (WAK$_{s}$) and one of the samples of Gujjars (GUJ$_{h}$). In the center with close affinities to one another are the two samples of Syeds (SYD$_{m1}$, SYD$_{m2}$). The final aggregate is located in the lower center of the array and includes the second sample of Gujjars (GUJ$_{m2}$), Mashwanis (MSH$_{a}$), and the sample from Madak Laskt (MDK). When considered by ecological setting there appears to be no discernable pattern distinguishing between females of highland groups from their lowland counterparts (Fig. 10b). The same is largely true when females are divided between those claiming Pathan ancestry and those believed to be of non-Pathan origin. However, Pathan groups tend to occupy the upper right of the array, whereas local groups tend to occupy the left and lower center (Fig. 10c). Exceptions include the two samples of Awans, Swatis and Mashwanis on the one hand and one of the Gujjar samples (GUJ$_{m2}$) on the other.

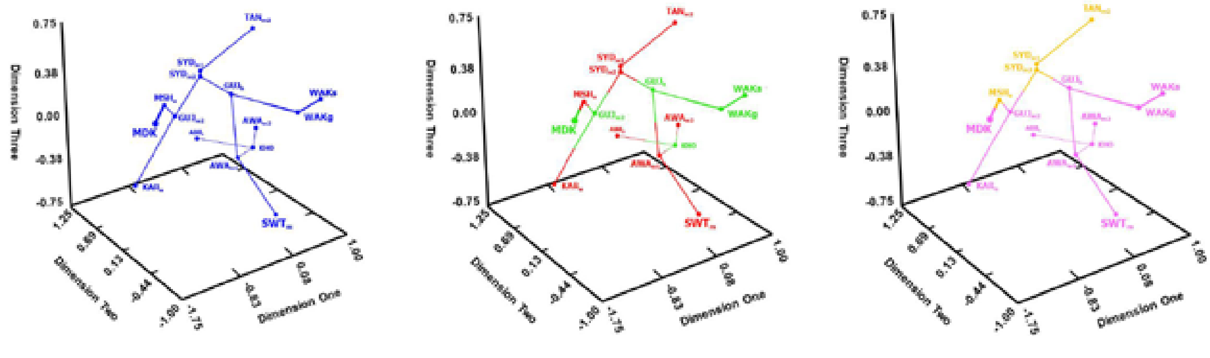

Figure 9. Multidimensional Scaling among Samples of Living Individuals with Sexes Pooled: a) All Samples, b) Highland Samples (in green) versus Lowland Samples (in red), c) Pathan (in light brown) versus Non-Pathan Samples (in purple).

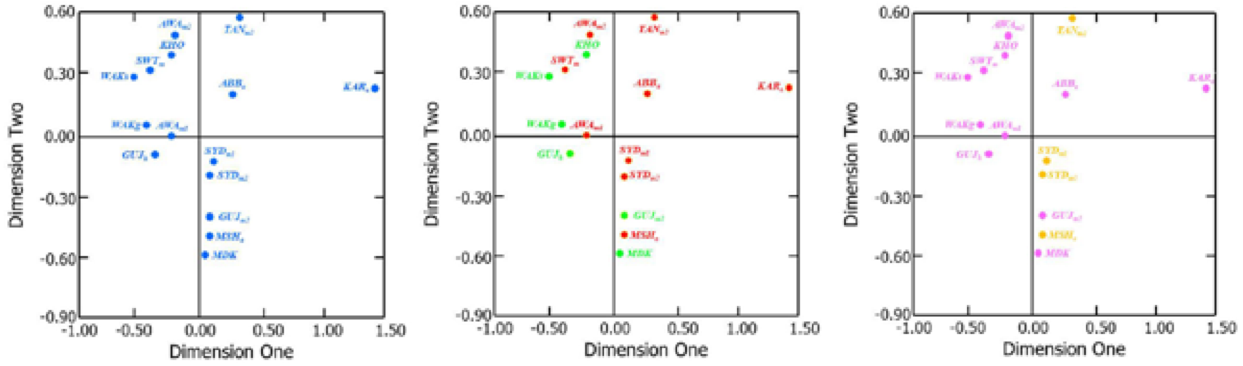

Figure 10. Correspondence Analysis of Sample Centroids among Samples of Living Females: a) All Females, b) Highland Females (in green) versus Lowland Females (in red), c) Pathans (in light brown) versus Non-Pathans (in purple).

Principal component analysis of tooth-trait frequencies among living females yields four components with eigenvalues greater than 1.0 (Table 5). The first accounts for 26.7% of the total variance and draws a distinction between shoveling on UI2 versus and presence of the Y-groove on LM2 versus and the presence of Cusp 5 on UM1. The second component accounts for 21.0% of the variance. It emphasizes shoveling on UI1, medial lingual ridge development on UI2, presence of a well-developed hypoconulid on LM2 and the presence of the entoconulid on LM1. Component three accounts for 12.8% of the variance. It emphasizes the presence of the Y-groove and the metaconulid on LM1. Component 4 draws a contrast between medial lingual ridge development on UI1 and a fully developed hypocone on UM2 versus presence of Carabelli's trait on UM1 and a well-developed hypoconulid on LM1. It accounts for 11.5% of the variance.

A plot of scores for the first three components (Fig. 11a) yields two sample aggregates and four isolates. These three components combine to account for 60.6% of the total variance. The first aggregate is found in the foreground and includes the two samples of Syeds, Tanolis, and one of the samples of Gujjars (GUJ_h). The second aggregate occupies the center of the array. It includes the two samples of Wakhis, the two samples of Awans and Khowars. Abbasis and Swatis represent distant peripheral members of this aggregate. The residents of Madak Lasht, Mashwanis, Karlaars and the second sample of Gujjars (GUJ_{m2}) found in the background of the array are isolates with little affinity to one another or to any of the other samples included in this analysis. When considered by ecological setting highland groups tend to occupy the center of the array while lowland grounds occupy more peripheral locations (Fig. 11b). Again, there are several exceptions. These include the sample of Wakhis from Gulmit (WAK_g) on one hand and the two samples of Awans on the other. A much clearer distinction across samples of living females emerges when ethnicity is the basis of comparison (Fig. 11c). With only a single exception (MSH_a), Pathan groups occupy positions in the lower foreground with especially close affinities between the two samples of Syeds. Remaining samples form a single aggregate that includes Khowars, one of the Awan samples (AWA_{m2}) and Wakhis, with two peripheral members (SWT_m, ABB_a) and three isolates (MDK, KAR_a, GUJ_{m2}).

The zero-corrected matrix of pairwise mean measure of divergence values are presented in Table 6. Some 54 of these contrasts of samples among living females (51.4%) are significant at $\alpha< 0.05$ level. When considered by ecological setting 30 of 51 pairwise contrasts (58.8%) differ between highland and lowland groups, while 28 of 44 of pairwise contrasts (63.6%) differ between those who claim Pathan ancestry and those believed to be of non-Pathan origin. Such results suggest that greater discernment among living females is achieved when attributed ancestry is taken into account (+12.2%) than a consideration by ecological setting (+7.4%).

A neighbor-joining cluster analysis based on the matrix of pairwise mean measure of divergence values among samples of living females is provided in Figure 12a. Two aggregates, three isolates and one intermediary sample may be identified. The first aggregate is composed of six samples and is found on the right side of the array. It includes the two Awan samples, the two Wakhi samples, Khowars and Swatis. The second aggregate occurs in the center of the array and encompasses five samples. These include one of the Gujjar samples (GUJ$_{m2}$), Mashwanis, the two Syed samples and Tanolis. The other sample of Gujjars (GUJ$_{h}$) stands as an intermediate between aggregates one and two. Abbasis stand as outliers to aggregate 2 while the Madak Lasht and Karlaars stand on the left as true isolates. When considered by ecological setting (Fig. 12b) there appears to be no consistent pattern with highland groups divided into two aggregates on opposite sides of the array. This is definitely not the case when the pattern of affinities is considered by ethnicity (Fig. 12c). All of the groups claiming Pathan ancestry are found in the center of the array and bear closest affinities to one another. Groups believed to be of non-Pathan origin are divided into largely highland groups on the right side and isolated groups on the left. Awans and Swatis stand out as exceptions among the highland groups on the right.

Multidimensional scaling into three dimensions with Kruskal's stress formula number 1 was accomplished in four iterations and accounted for 97.2% of the variance with a stress level of 0.066. This is considered a near-excellent

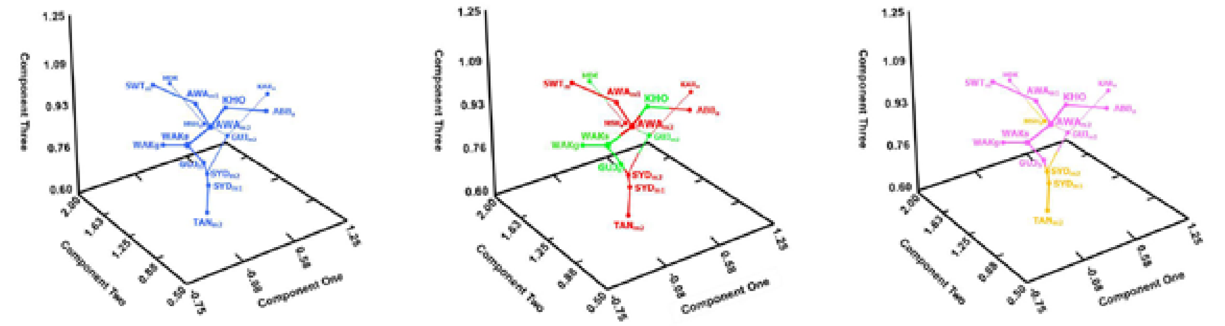

Figure 11. Three-dimensional Plot of Sample Centroids among Samples of Living Females yielded by Principal Components Analysis: a) All Females, b) Highland Females (in green) versus Lowland Females (in red), c) Pathans (in light brown) versus Non-Pathans (in purple).

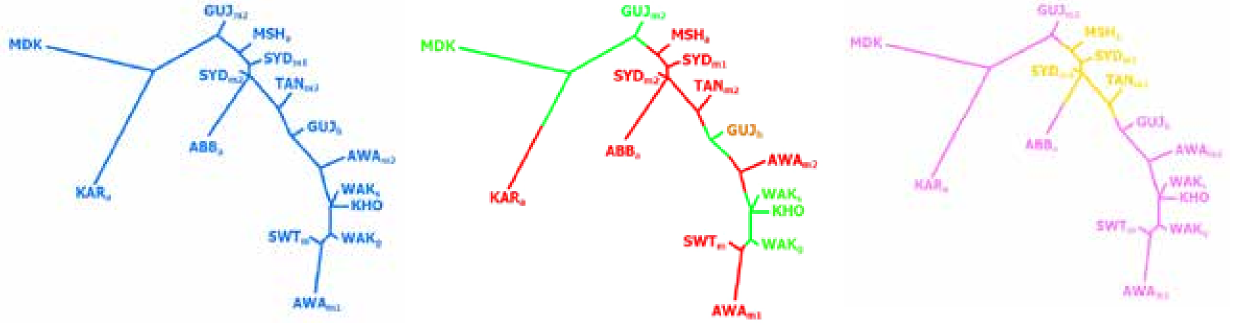

Figure 12. Neighbor-joining Cluster Analysis among Samples of Living Females: a) All Females, b) Highland Females (in green) versus Lowland Females (in red), c) Pathans (in light brown) versus Non-Pathans (in purple).

Table 5. Principal Component Loadings among Living Samples of Females[1]

		Loadings			
Trait	Tooth	1	2	3	4
SHOV	UI1	0.486	***0.747***	-0.005	0.071
SHOV	UI2	***0.792***	0.464	-0.151	0.059
MLR	UI1	-0.562	0.391	0.255	***0.564***
MLR	UI2	-0.485	***0.603***	-0.242	-0.043
HYPO	UM2	0.534	-0.083	-0.057	***0.471***
CARA	UM1	0.557	0.220	0.109	***-0.436***
C5	UM1	***-0.791***	0.156	0.045	0.045
YGRV	LM1	-0.244	-0.200	***0.870***	-0.060
YGRV	LM2	***0.813***	-0.094	0.337	0.299
CSPN	LM1	-0.222	0.417	-0.206	***-0.570***
CSPN	LM2	0.173	***0.639***	0.313	-0.327
C6	LM1	-0.167	***0.810***	0.412	0.246
C7	LM1	0.105	-0.277	***0.569***	-0.417
Eigenvalue		3.475	2.733	1.668	1.501
% Var. Expl.		26.729	21.026	12.831	11.543

1. Most influential variables by component are in bold and italicized.

Table 6. Zero-corrected Mean Measure of Divergence Analysis among Samples of Living Females[1]

	ABBa	AWAm1	AWAm2	GUJh	GUJm2	KARa	KHO	MDK	MSHa	SWTm	SYDm1	SYDm2	TANm2	WAKg	WAKs
ABBa	---	0.0236	0.0213	0.0221	0.0212	0.0295	0.0166	0.0173	0.0200	0.0218	0.0231	0.0189	0.0197	0.0190	0.0182
AWAm1	0.0372	---	0.0189	0.0195	0.0188	0.0271	0.0142	0.0150	0.0175	0.0195	0.0208	0.0166	0.0173	0.0167	0.0159
AWAm2	0.0062	0.0145	---	0.0173	0.0164	0.0248	0.0118	0.0126	0.0152	0.0171	0.0184	0.0142	0.0149	0.0143	0.0135
GUJh	0.0207	0.0000	0.0000	---	0.0173	0.0256	0.0125	0.0133	0.0161	0.0177	0.0190	0.0148	0.0158	0.0149	0.0142
GUJm2	0.0269	0.0030	**0.0598**	0.0000	---	0.0247	0.0117	0.0124	0.0151	0.0169	0.0183	0.0140	0.0149	0.0141	0.0133
KARa	**0.0719**	**0.1058**	**0.1466**	**0.1116**	0.0490	---	0.0200	0.0208	0.0234	0.0252	0.0266	0.0223	0.0232	0.0224	0.0217
KHO	0.0259	0.0084	0.0000	0.0014	**0.0452**	**0.1419**	---	0.0079	0.0104	0.0124	0.0138	0.0095	0.0102	0.0096	0.0088
MDK	**0.1551**	**0.0516**	**0.1492**	**0.0523**	0.0243	**0.0960**	**0.1435**	---	0.0112	0.0132	0.0146	0.0103	0.0110	0.0104	0.0096
MSHa	0.0319	0.0344	**0.0422**	0.0179	0.0087	**0.0810**	**0.0746**	**0.0364**	---	0.0157	0.0170	0.0128	0.0136	0.0129	0.0121
SWTm	**0.0778**	0.0000	0.0000	0.0000	**0.0517**	**0.1590**	0.0141	**0.0729**	**0.0552**	---	0.0191	0.0148	0.0155	0.0149	0.0141
SYDm1	0.0161	**0.0454**	0.0154	0.0075	0.0103	**0.0911**	**0.0561**	**0.0827**	0.0000	**0.0501**	---	0.0161	0.0168	0.0162	0.0154
SYDm2	0.0115	0.0275	0.0166	0.0040	0.0043	**0.0860**	**0.0433**	**0.0696**	0.0000	**0.0448**	0.0000	---	0.0125	0.0120	0.0111
TANm2	0.0361	0.0303	0.0122	0.0028	**0.0394**	**0.1048**	**0.0311**	**0.1300**	**0.0136**	**0.0349**	0.0098	0.0130	---	0.0126	0.0118
WAKg	**0.1090**	0.0101	0.0131	0.0000	**0.0560**	**0.1845**	**0.0204**	**0.0859**	**0.0600**	0.0000	**0.0511**	**0.0389**	**0.0283**	---	0.0113
WAKs	**0.0817**	0.0147	0.0000	0.0000	**0.0574**	**0.2023**	0.0000	**0.1211**	**0.0737**	0.0000	**0.0598**	**0.0473**	**0.0262**	0.0000	---

1. MMD values below the diagonal, standard deviations above the diagonal. Significant differences at $\alpha < 0.05$ in bold. Abbreviations from Table 1.

fit of the data (Clarke 1993; De Leeuw & Stoop 1984)[6]. Four aggregates may be identified among samples of living females (Fig. 13a). On the left is a loose aggregate of three samples that includes Karlaars, the Madak Lasht and one of the samples of Gujjars (GUJ_{m2}). The second is located in the center of the array and is composed of a loose aggregate of four samples. The Khowars occupy the most isolated position in the right foreground, followed by one of the samples of Awans (AWA_{m1}). The other sample of Gujjars (GUJ_h) and Swatis are identified as possessing rather close affinities to one another. The third aggregate occupies the right foreground and encompasses five samples arrayed into two sub-clusters. One sub-cluster in the upper center is composed of the two samples of Syeds and Mashwanis. The other sub-cluster occupies the right foreground and includes Tanolis, Abbasis and the other sample of Awans (AWA_{m2}). The fourth aggregate is composed of the two Wakhi samples who occupy an isolated position in the right background. When considered by ecological setting there is a nearly complete between lowland and highland groups (Fig. 13b). Lowland groups occupy the right and upper foreground whereas highland groups are scattered in the background although widely dispersed along the first dimension. There are, however, several exceptions. Karlaars, one of the Awan samples (AWA_{m1}) and Swatis occupy position most proximate to highland groups, while the Khowar appear isolated from all other samples. When considered by ethnicity (Fig. 13c), groups claiming Pathan ancestry exhibit closest affinities to one another, while non-Pathan groups are widely scattered throughout the array. There is a single exception, one of the Awan samples (AWA_{m2}) appears to share closest affinities with Tanolis.

Males Only

A plot of the first two dimensions obtained by correspondence analysis among samples of living males yields two aggregates and two isolates (Fig. 14a). The first occurs in the middle left and includes the two samples of Awans, Swatis, Khowars, the sample of Wahkis from Gulmit (WAK_g) and Tanolis. The second occupies the right side and includes six samples. These include the two samples of Gujjars and the two samples of Syeds along with the samples of Karlaars and the inhabitants of Madak Lasht. The two isolates are Abbasis, which occupy a position in the upper center, and the sample of Wakhis from Sost (WAKs), which occupies a position in the lower left corner of the array. When considered by ecological setting there is a nearly complete separation along dimension two between highland samples with low scores and lowland samples with higher scores (Fig. 14b). There is only one exception, Khowars, who occupy a position on the left side in between one of the Awan samples (AWA_{m2}) and Tanolis. In contrast, when samples of males are divided between those claiming Pathan ancestry and those believed to be of non-Pathan origin there is no discernable pattern by ethnicity (Fig. 14c).

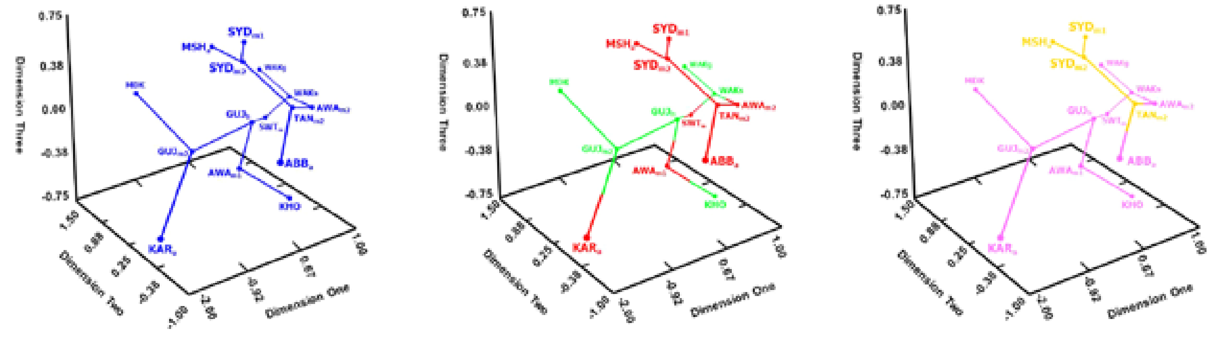

Figure 13. Multidimensional Scaling among Samples of Living Females: a) All Females, b) Highland Females (in green) versus Lowland Females (in red), c) Pathans (in light brown) versus Non-Pathans (in purple).

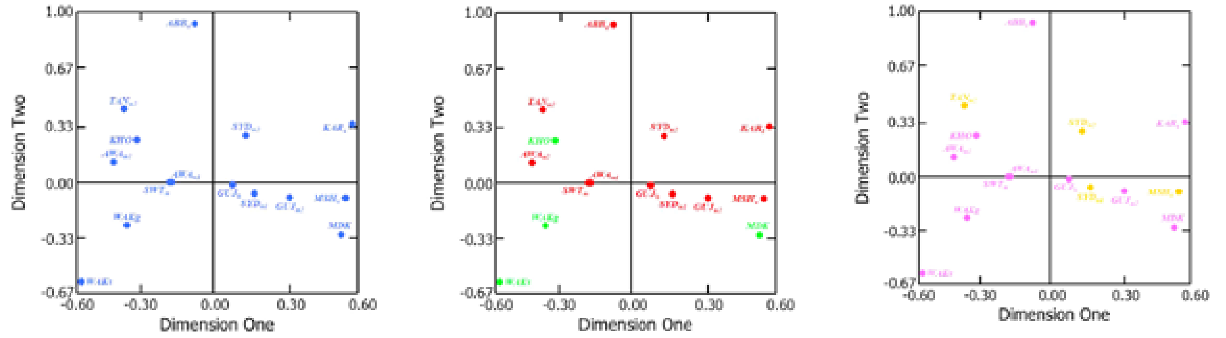

Figure 14. Correspondence Analysis of Sample Centroids among Samples of Living Males yielded by Principal Components Analysis: a) All Males, b) Highland Males (in green) versus Lowland Males (in red), c) Pathans (in light brown) versus Non-Pathans (in purple).

As with living females, principal component analysis of tooth-trait frequencies among living males yields four components with eigenvalues greater than 1.0 (Table 7). The accounts for 27.1% of the total variance and draws a distinction between samples based on the presence of shoveling and median lingual ridge development on the maxillary incisors. The second component accounts for 21.7% of the variance. It draws a contrast between samples with median lingual ridge development on the maxillary incisors and accessory cusps on UM1 (Cusp 5) and LM1 (metaconulid) versus samples with high prevalence of shoveling on the maxillary incisors and the presence of the Y-groove on LM2. Component three accounts for 13.6% of the variance. It emphasizes the prevalence of accessory cusps on UM1 (Carabelli's trait, Cusp 5). Component 4 likewise accounts for 13.6% of the total variance. In this case high scores are received by samples with relatively high prevalence of a fully developed hypocone on UM2 and retention of the Y-groove on LM1.

A plot of scores for the first three components (Fig. 15a) yields three sample aggregates and four isolates. These three components combine to account for 62.4% of the total variance. The first aggregate is found in the left and encompasses five samples. These include the two samples of Awans, the two samples of Wakhis, and Khowars. The second aggregate is a bit loose and occupies disparate positions along component one in the foreground. This aggregate encompasses four samples: Abbasis, the two Samples of Syeds, and one of the Gujjar samples (GUJ_{m2}). The third aggregate is an even looser association of three samples in the background. It encompasses three samples: Karlaars, the other sample of Gujjars (GUJ_h), and Swatis. Three samples stand as outliers: the residents of Madak Lasht and Mashwais on the left side, while Tanolis occupy a position in the extreme left foreground. When considered by ecological setting (Fig. 15b) lowland samples tend to occur in the center of the array while highland groups tend to occupy peripheral positions on the left, the right and in the background. There are, however, exceptions. These include Awans, which are found with highland groups on the left side, Tanolis, which occupy the upper background, and Mashwanis who occupy the extreme right side. Unlike females there is no clear distinction between groups who claim Pathan ancestry and those believed to be of non-Pathan origins among males. As can be seen in Figure 15c, while Pathan samples tend to possess low scores for component two they are widely scattered across components two and three sharing little affinity to one another.

The zero-corrected matrix of pairwise mean measure of divergence values among samples of living males are presented in Table 8. Some 58 of these contrasts (55.2%) are significant at $\alpha < 0.05$ level. When considered by ecological setting 28 of 51 pairwise contrasts (54.9%) differ between highland and lowland groups, while 24 of 44 of pairwise contrasts (54.5%) differ between those who claim Pathan ancestry and those believed to be of non-Pathan origin. Such results suggest that, unlike females, ecological setting is equally important as ethnicity in the patterning of affinities among living males of these social groups from Hazarewal and Chirtal-Gilgit-Baltistan.

A neighbour-joining cluster analysis based on the matrix of pairwise mean measure of divergence values among samples of living males is presented in Figure 16a. Two aggregates and a vector of increasing phenetic uniqueness may be identified. The first aggregate is composed of six samples and may be found on the lower left of the array. It includes Swatis, the two Wakhi and Awan samples as well as Khowars. The second aggregate occupies the upper left. It is composed of three samples: Abbasis, Tanolis, and one of the Gujjar samples (GUJ_h). The vector of increasing phenetic uniqueness spans the right side of the array. The two Syed samples are the least isolated and share fairly close affinities to one another. They are followed by the second sample of Gujjars (GUJ_{m2}), Mashwanis, Karlaars, and most unique of all, the residents of Madak Lasht. Just as with females there is no consistent patterning among living males of different ecological settings among these samples from Hazarewal and Chitral-Gilgit-Baltistan (Fig. 16b). Greater patterning among samples of living males is obtained on the basis of ethnicity (Fig. 16c), for with only a single exception (TAN_{m2}), all of the samples occupying the left side of the array are held to be of non-Pathan origin. The two samples of Syeds bear closest affinities to one another, but are separated from their fellow Mashwani Pathans by one of the samples of Gujjars (GUJ_{m2}). Karlaars and the residents of Madak Lasht stand apart as isolates on the extreme right side of the array.

Multidimensional scaling of mean measure of divergence scores for pairwise distances between living males with Kruskal's stress formula number 1 was accomplished in three iterations and accounted for 98.3% of the variance with a stress level of 0.050. This is considered an excellent

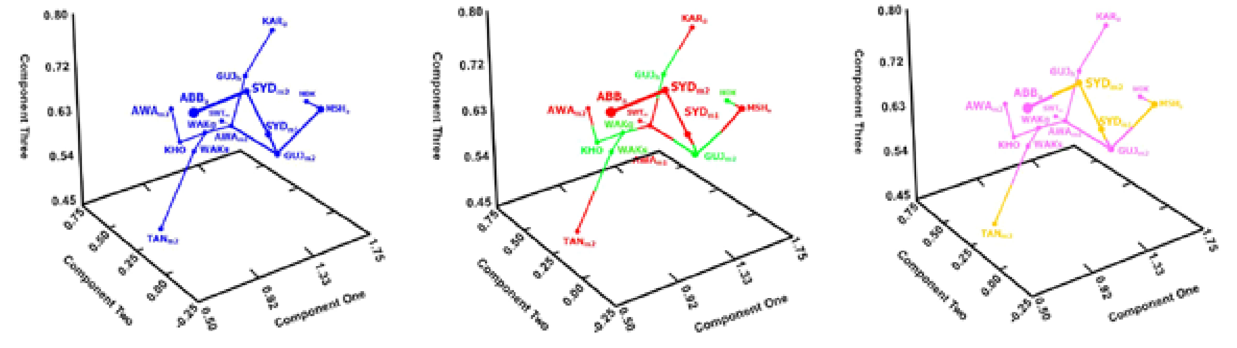

Figure 15. Three-dimensional Plot of Sample Centroids among Samples of Living Males yielded by Principal Components Analysis: a) All Males, b) Highland Males (in green) versus Lowland Males (in red), c) Pathans (in light brown) versus Non-Pathans (in purple).

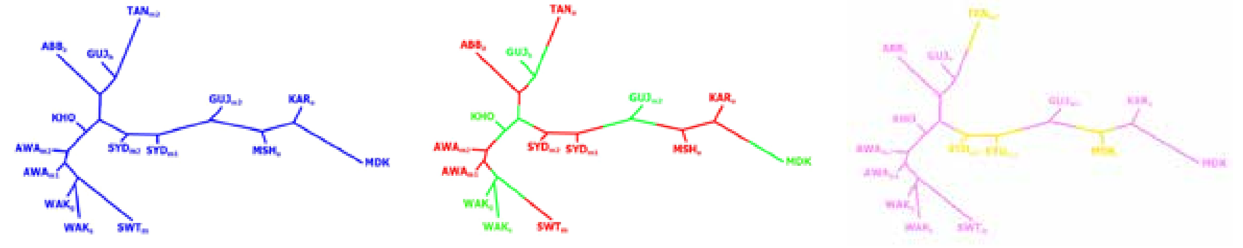

Figure 16. Neighbor-joining Cluster Analysis among Samples of Living Males: a) All Males, b) Highland Males (in green) versus Lowland Males (in red), c) Pathans (in light brown) versus Non-Pathans (in purple).

Table 7. Principal Components Analysis across Samples of Living Males[1]

		Loadings			
Trait	Tooth	1	2	3	4
SHOV	UI1	***0.801***	***-0.426***	0.132	0.049
SHOV	UI2	***0.719***	***-0.454***	0.379	0.163
MLR	UI1	***0.585***	***0.667***	-0.207	0.026
MLR	UI2	***0.571***	***0.489***	-0.156	-0.369
HYPO	UM1	-0.516	0.383	0.264	0.276
HYPO	UM2	-0.190	-0.171	0.324	***0.571***
CARA	UM1	0.456	0.153	***0.719***	-0.392
C5	UM1	-0.076	***0.543***	***0.694***	0.309
YGRV	LM1	0.296	0.275	-0.324	***0.728***
YGRV	LM2	0.506	***-0.584***	-0.146	0.324
C7	LM1	0.548	***0.637***	-0.136	0.205
Eigenvalue		2.986	2.733	1.495	1.496
% Var. Expl.		27.148	21.704	13.588	13.588

1. Most influential variables by component are in bold and italicized.

Table 8. Zero-corrected Mean Measure of Divergence Analysis across Samples of Living Males[1]

	ABBa	AWAm1	AWAm2	GUJh	GUJm2	KARa	KHO	MDK	MSHa	SWTm	SYDm1	SYDm2	TANm2	WAKg	WAKs
ABBa	---	0.0118	0.0095	0.0164	0.0146	0.0139	0.0113	0.0135	0.0163	0.0122	0.0157	0.0145	0.0148	0.0147	0.0142
AWAm1	**0.0301**	---	0.0068	0.0122	0.0104	0.0096	0.0071	0.0092	0.0121	0.0079	0.0115	0.0102	0.0106	0.0104	0.0100
AWAm2	0.0178	0.0000	---	0.0151	0.0133	0.0126	0.0100	0.0122	0.0150	0.0108	0.0144	0.0132	0.0135	0.0133	0.0129
GUJh	0.0257	0.0018	0.0242	---	0.0151	0.0144	0.0118	0.0139	0.0169	0.0126	0.0162	0.0149	0.0153	0.0151	0.0147
GUJm2	**0.0432**	0.0089	**0.0321**	0.0103	---	0.0125	0.0100	0.0121	0.0150	0.0108	0.0144	0.0132	0.0135	0.0133	0.0128
KARa	**0.0688**	**0.0536**	**0.1151**	**0.0561**	0.0032	---	0.0092	0.0114	0.0143	0.0100	0.0136	0.0124	0.0128	0.0125	0.0121
KHO	0.0167	0.0000	0.0000	**0.0256**	**0.0274**	**0.0872**	---	0.0088	0.0117	0.0075	0.0111	0.0098	0.0102	0.0100	0.0095
MDK	**0.1532**	**0.0655**	**0.1354**	**0.0714**	0.0167	0.0077	**0.1074**	---	0.0138	0.0096	0.0132	0.0120	0.0123	0.0121	0.0117
MSHa	**0.0837**	**0.0337**	**0.0951**	0.0111	0.0000	0.0000	**0.0758**	0.0000	---	0.0125	0.0161	0.0149	0.0152	0.0150	0.0146
SWTm	**0.1039**	0.0112	0.0542	**0.0382**	**0.0536**	**0.0735**	**0.0379**	**0.0475**	**0.0437**	---	0.0119	0.0106	0.0110	0.0108	0.0104
SYDm1	0.0154	0.0074	0.0232	0.0000	0.0000	0.0208	0.0106	**0.0453**	0.0072	**0.0571**	---	0.134	0.0132	0.0144	0.0139
SYDm2	0.0221	0.0117	0.0226	0.0046	0.0000	**0.0346**	0.0111	**0.0539**	0.0188	**0.0524**	0.0000	---		0.0132	0.0127
TANm2	**0.0412**	**0.0627**	**0.0467**	0.0151	**0.0867**	**0.1630**	**0.0451**	**0.1817**	**0.1195**	**0.1186**	**0.0412**	**0.0360**	---	0.0135	0.0131
WAKg	**0.0601**	0.0000	0.0000	0.0143	**0.0289**	**0.0917**	0.0000	**0.0760**	**0.0538**	0.0007	0.0205	0.0196	**0.0531**	---	0.0002
WAKs	**0.0988**	0.0119	0.0054	**0.0435**	**0.0544**	**0.1389**	0.0141	**0.1137**	**0.0967**	**0.0392**	**0.0428**	**0.0473**	**0.0638**	0.0000	---

1. MMD values below the diagonal, standard deviations above the diagonal. Significant differences at α< 0.05 in bold. Abbreviations from Table 1.

fit of the data (Clarke 1993; De Leeuw & Stoop 1984). Two aggregates, one on the right side of the array, the other on the left, and three isolates may be identified (Fig. 17a). The first aggregate encompasses five samples and one peripheral member. These include the two Wakhi samples, the two samples of Awans, Khowars, and as an outlier, Swatis. The second aggregate, which is on the left side, includes five samples and three peripheral members. The members include the two samples of Syeds, the two samples of Gujjars and Mashwanis. The three peripheral members are Karlaars in the left foreground, the inhabitants of Madak Lasht in the center-left, and Tanolis, which occupy a highly-isolated position in the upper center of the array. When considered by ecological setting there is a nearly complete separation between lowland and highland groups (Fig. 17b). Lowland groups occupy peripheral positions in the lower right and the middle left, as well as isolated positions in the lower left (KAR_a), centre-left (MSH_a), and upper center (TAN_{m2}) of the array. In contrast, although spread out along dimension one, highland samples occupy the center of the array with only a single exception, the sample of Gujjars from Haripur (GUJ_h). When considered by ethnicity (Fig. 17c) there is nearly a complete separation between those claiming Pathan ancestry and those believed to be of non-Pathan origin. With a single exception (GUJh), the former are found in the upper center of the array, while the latter are widely dispersed along dimension one in the lower portion of the array.

Sexes Specified

A plot of the first two dimensions obtained by correspondence analysis among members of all living samples with sexes specified produces two aggregates and two isolates (Fig. 18a). The first aggregate is a fairly tight scatter of samples located in the lower right of the array. This aggregate encompasses males and females of both Syed samples, both males and females of Mashwanis and the residents of Madak Lasht, as well as Karlaar males. The second aggregate is a looser scatter of samples spanning the upper and lower left side. This large aggregate encompasses males and females of both Awan and Wakhi samples, male and female Tanolis, Khowars, and Gujjars from Haripur District (GUJ_h) Males of the other Gujjar sample (GUJ_{m2}) are also a member of this aggregate, while their female counterparts (GUJm2F) are members of the first aggregate. Abbassi females represent a peripheral member of this second aggregate. The two isolates are found in the upper right and include Abbassi males and Karlaar females. When considered by ecological setting there is a fairly clean separation between lowland samples, which occupy positions in the upper right, lower right and upper left of the array, and highland samples that tend to occupy the lower left (Fig. 18b). There are, however, exceptions. Lowland samples found intermingled with highland samples include both male and female Awans and Swatis. Highland groups found intermingled with lowland samples include one of the samples of Gujjar females ($GUJ_{m2}F$) and male and female residents of Madak Lasht. A nearly

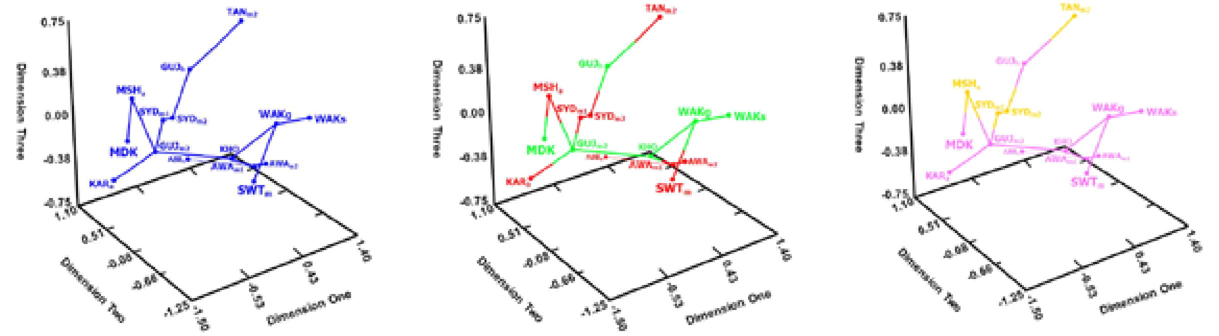

Figure 17. Multidimensional Scaling among Samples of Living Males: a) All Males, b) Highland Males (in green) versus Lowland Males (in red), c) Pathans (in light brown) versus Non-Pathans (in purple).

complete separation occurs between Pathan and non-Pathan samples when considered by ethnicity (Fig. 18c). Pathan samples tend to be found in the lower right of the array while non-Pathan samples occupy the left side and upper right. There is only one exception to this pattern for Pathans and non-Pathan samples, respectively. For the former it is the Tanolis who occupy disparate position in the upper center and for non-Pathans it is the male residents of Madak Lasht that occupy a position in between females of one of the Syed samples (SYD_{m1}F) and Mashwani males and females.

As in the previous analyses, principal component analysis of tooth-trait frequencies across samples of living males specified yields four components with eigenvalues greater than 1.0 (Table 9). The first component accounts for 27.2% of the total variance and draws a distinction between samples with high prevalence of shoveling and median lingual ridge development. The second component accounts for 21.7% of the variance. It draws a contrast between samples with relatively high prevalence of shoveling on the maxillary incisors and a cusp 5 on UM1 from samples with relatively high median lingual ridge development on these same teeth as well as retention of a well-developed hypoconulid on LM1. Component three accounts for 13.6% of the variance. It emphasizes the presence of Carabelli's trait and cusp 5 on UM1. Component 4 accounts for 13.6% of the total variance. In this case high scores are received by samples with relatively high prevalence of retention of a fully developed hypocone on UM2 and retention of the Y-groove on LM1.

A plot of scores for the first two components (Fig. 19a) yields three sample aggregates and four isolates. These components combine to account for 45.7% of the total variance[7]. The first aggregate is found in the center of the array and includes seven samples. These include male and female Mashwanis, males and females of one of the Gujjar samples (GUJ_{m2}), females from Madak Lasht, as well as Abbassi and Karlaar males. The

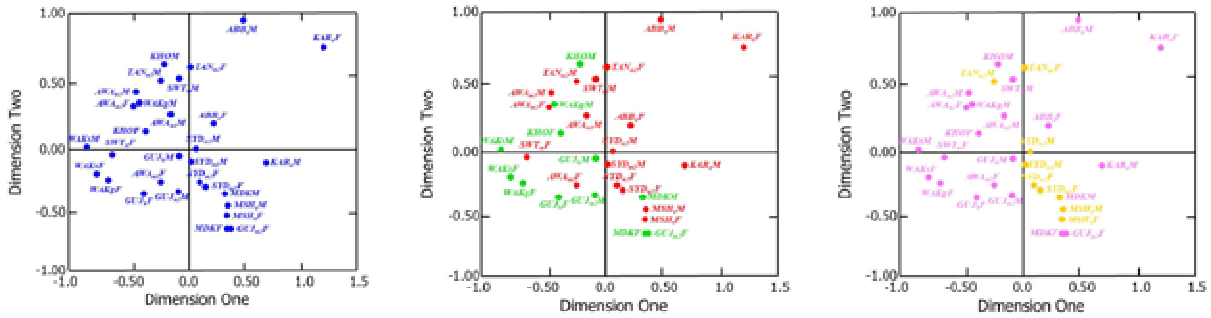

Figure 18. Correspondence Analysis of Sample Centroids among Samples of Living Individuals with Sexes Specified: a) All Individuals, b) Highland Individuals (in green) versus Lowland Individuals (in red), c) Pathans (in light brown) versus Non-Pathans (in purple).

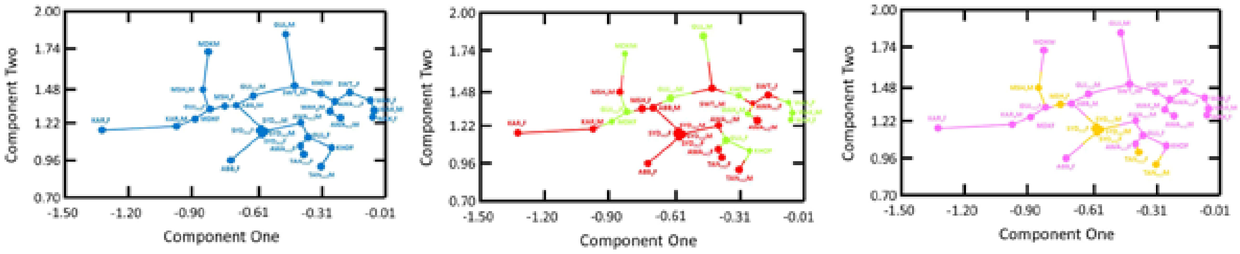

Figure 19. Three-dimensional Plot of Sample Centroids among Samples of Living Individuals with Sexes Specified yielded by Principal Components Analysis: a) All Females, b) Highland Females (in green) versus Lowland Females (in red), c) Pathans (in light brown) versus Non-Pathans (in purple).

Table 9. Principal Component Loadings among Living Samples with Sexes Specified[1]

		Loadings			
Trait	Tooth	1	2	3	4
SHOV	UI1	-0.612	***0.594***	-0.222	***0.315***
SHOV	UI2	***-0.831***	0.390	-0.132	0.094
MLR	UI1	0.438	***0.669***	0.044	0.240
MLR	UI2	0.300	***0.584***	***-0.458***	-0.198
HYPO	UM2	-0.464	***-0.237***	0.240	0.121
CARA	UM1	-0.468	0.293	-0.086	***-0.596***
C5	UM2	0.582	0.289	-0.079	***0.359***
YGRV	LM1	0.189	0.300	***0.819***	0.171
YGRV	LM2	***-0.759***	0.042	0.306	0.238
C7	LM1	0.117	0.466	***0.574***	***-0.495***
Eigenvalue		2.746	1.828	1.444	1.034
% Var. Expl.		27.455	18.283	14.437	10.343

1. Most influential variables by component are in bold and italicized.

second aggregate occupies the upper right. It includes nine samples. These include males and females of the two Wakhi samples, males and females of one of the Awan samples (AWA_{m2}) male and female Swatis as well as Khowar males. The third aggregate is found in the lower right and includes 10 samples. These include male and female Tanolis, males and females of the two Syed samples, and males and females of the other sample of Awans (AWA_{m1}) as well as Khowar females and females of the second Gujjar sample (GUJ_{h}). The four isolates include Karlaar females on the extreme left side, males from Madak Lasht in the upper center, males of one of the Gujjar samples (GUJ_{h}) in the upper right, and Abbassi females in the lower center. There is no discernable pattern when the relative position of samples along the first two components is considered by ecological setting (Fig. 19b) A clearer pattern is obtained when sample positions are considered by ethnicity (Fig. 19c). Pathan samples tend to trend from the upper left to the lower right in the center of array. However, their affinities to one another are interrupted by non-Pathan Abbassi males and females of one of the Gujjar samples (GUJ_{m2}). This is especially the case for Tanoli males and females who are segregated away from other Pathan samples in the lower right by non-Pathan Awan males and females (AWA_{m1}), and by females Khowars and from the second of the Gujjar samples (GUJ_{h}). In contrast, non-Pathan samples while largely found on the right center of the array are nevertheless widely scattered along component one.

The zero-corrected pairwise matrix of mean measure of divergence values across samples of living individuals with sexes specified is presented in Table 10. Some 252 of these pairwise contrasts across all samples (57.9%) are significant at $\alpha < 0.05$ level. When considered by ecological setting 126 of 216 pairwise contrasts (58.3%) differ between highland and lowland groups, while 109 of 176 of pairwise contrasts (61.9%) differ between those who claim Pathan ancestry and those believed to be of non-Pathan origin. Such results suggest that while ecological setting offers little improvement in discernment between groups by sex (+0.4%) a consideration by ethnicity across these ethnic groups of Hazarewal and Chitral-Gilgit-Baltistan provides ten times greater improvement in discernment (+4%) than provided by ecological setting.

Multidimensional scaling of mean measure of divergence scores for pairwise distances between samples of living individuals with sexes specified using Kruskal's stress formula number 1 was accomplished in four iterations and accounted for 96.7% of the variance with a stress level of 0.075.

Table 10. Zero-corrected Mean Measure of Divergence Analysis across Samples of Living Individuals with Sexes Specified[1]

	ABBaF	**ABBaM**	**AWAm1F**	**AWAm1M**	**AWAm2F**	**AWAm2M**	**GUJhF**	**GUJhM**	**GUJm2F**	**GUJm2M**	**KARaF**	**KARaM**	**KHOF**	**KHOM**	**MDKF**	**MDKM**	**MSHaF**	MSHaM	SWTmF	SWTmM	SYDm1F	SYDm1M	SYDm2F	SYDm2M	TANm2F	TANm2M	WAKgF	WAKgM	WAKsF	WAKsM
ABBaF	---	0.023	0.027	0.019	0.024	0.022	0.026	0.024	0.025	0.022	0.034	0.021	0.019	0.018	0.020	0.021	0.023	0.024	0.025	0.019	0.026	0.023	0.022	0.020	0.023	0.022	0.022	0.020	0.021	0.019
ABBaM	0.000	---	0.021	0.012	0.018	0.016	0.019	0.017	0.018	0.016	0.027	0.015	0.013	0.012	0.014	0.014	0.016	0.017	0.019	0.013	0.020	0.017	0.015	0.014	0.016	0.016	0.015	0.014	0.015	0.014
AWAm1F	**0.057**	**0.069**	---	0.016	0.022	0.019	0.023	0.021	0.022	0.019	0.031	0.018	0.016	0.016	0.017	0.018	0.020	0.021	0.022	0.017	0.024	0.020	0.019	0.017	0.020	0.019	0.019	0.017	0.018	0.017
AWAm1M	0.035	**0.035**	0.001	---	0.014	0.011	0.015	0.013	0.014	0.011	0.023	0.010	0.008	0.008	0.009	0.010	0.012	0.013	0.014	0.008	0.016	0.012	0.011	0.010	0.012	0.011	0.011	0.010	0.010	0.010
AWAm2F	0.022	0.023	0.021	0.000	---	0.017	0.020	0.019	0.019	0.017	0.029	0.016	0.014	0.013	0.015	0.015	0.018	0.018	0.020	0.014	0.021	0.018	0.016	0.015	0.017	0.017	0.016	0.015	0.016	0.015
AWAm2M	**0.045**	**0.033**	0.026	0.000	0.000	---	0.018	0.016	0.017	0.014	0.026	0.013	0.011	0.011	0.012	0.013	0.015	0.016	0.017	0.011	0.019	0.015	0.014	0.013	0.015	0.014	0.014	0.013	0.013	0.012
GUJhF	0.041	**0.060**	0.000	0.000	0.003	0.007	---	0.020	0.021	0.017	0.030	0.017	0.015	0.014	0.016	0.016	0.019	0.020	0.021	0.015	0.022	0.019	0.018	0.016	0.019	0.018	0.017	0.016	0.017	0.016
GUJhM	0.033	0.032	0.038	0.004	0.025	0.029	0.000	---	0.019	0.016	0.028	0.015	0.013	0.013	0.014	0.015	0.017	0.018	0.019	0.013	0.021	0.017	0.016	0.014	0.017	0.016	0.016	0.015	0.015	0.014
GUJm2F	0.048	**0.075**	0.003	**0.042**	**0.084**	**0.097**	0.006	0.013	---	0.016	0.029	0.016	0.014	0.013	0.015	0.015	0.018	0.019	0.020	0.014	0.021	0.018	0.016	0.015	0.018	0.017	0.016	0.015	0.016	0.015
GUJm2M	0.041	**0.051**	0.007	0.012	0.026	**0.038**	0.003	0.014	0.000	---	0.026	0.013	0.011	0.011	0.012	0.013	0.015	0.016	0.017	0.011	0.019	0.015	0.014	0.013	0.015	0.014	0.014	0.013	0.013	0.012
KARaF	**0.106**	**0.087**	**0.125**	**0.126**	**0.201**	**0.219**	**0.151**	**0.125**	**0.074**	**0.092**	---	0.025	0.023	0.022	0.024	0.025	0.027	0.028	0.029	0.023	0.030	0.027	0.026	0.023	0.027	0.026	0.026	0.023	0.025	0.023
KARaM	**0.066**	**0.078**	0.030	**0.061**	**0.115**	**0.129**	**0.061**	**0.064**	0.000	0.006	0.008	---	0.010	0.010	0.011	0.012	0.014	0.015	0.016	0.011	0.018	0.014	0.013	0.012	0.014	0.014	0.013	0.012	0.012	0.012
KHOF	**0.039**	**0.043**	0.000	0.000	0.000	0.000	0.001	**0.033**	**0.064**	**0.034**	**0.196**	**0.102**	---	0.008	0.009	0.010	0.012	0.013	0.014	0.008	0.016	0.012	0.011	0.010	0.012	0.011	0.011	0.010	0.010	0.010
KHOM	**0.041**	0.020	0.029	0.000	0.000	0.000	0.017	**0.030**	**0.085**	**0.032**	**0.158**	**0.098**	0.006	---	0.009	0.009	0.012	0.012	0.014	0.008	0.015	0.012	0.010	0.009	0.011	0.011	0.010	0.010	0.010	0.009
MDKF	**0.204**	**0.233**	**0.061**	**0.119**	**0.198**	**0.203**	**0.066**	**0.107**	0.022	**0.058**	**0.126**	**0.046**	**0.168**	**0.181**	---	0.011	0.013	0.014	0.015	0.009	0.017	0.013	0.012	0.011	0.013	0.012	0.012	0.011	0.011	0.011
MDKM	**0.158**	**0.171**	**0.041**	**0.074**	**0.141**	**0.151**	**0.051**	**0.081**	0.016	0.021	**0.079**	0.011	**0.127**	**0.120**	0.000	---	0.014	0.015	0.016	0.010	0.017	0.014	0.013	0.012	0.014	0.013	0.013	0.012	0.012	0.011
MSHaF	0.032	**0.057**	0.027	**0.030**	**0.051**	**0.064**	0.007	0.010	0.000	0.000	**0.092**	0.005	**0.059**	**0.056**	**0.048**	0.023	---	0.017	0.018	0.012	0.020	0.016	0.015	0.014	0.016	0.015	0.015	0.014	0.014	0.013
MSHaM	**0.062**	**0.093**	0.029	**0.035**	**0.089**	**0.103**	0.007	0.014	0.000	0.000	**0.056**	0.000	**0.085**	**0.080**	0.009	0.000	0.000	---	0.019	0.013	0.021	0.017	0.016	0.014	0.017	0.016	0.016	0.014	0.015	0.014
SWTmF	**0.111**	**0.115**	0.000	0.005	0.009	0.015	0.000	**0.038**	**0.066**	**0.048**	**0.212**	**0.122**	0.007	0.023	**0.104**	**0.087**	**0.075**	**0.069**	---	0.015	0.022	0.018	0.017	0.016	0.018	0.018	0.017	0.016	0.016	0.015
SWTmM	**0.131**	**0.116**	**0.037**	0.014	**0.062**	**0.062**	0.020	**0.044**	**0.082**	**0.061**	**0.119**	**0.083**	**0.060**	**0.043**	**0.099**	**0.054**	**0.083**	**0.046**	0.019	---	0.016	0.013	0.011	0.010	0.012	0.012	0.011	0.010	0.010	0.010
SYDm1F	0.036	0.035	**0.069**	**0.038**	0.031	**0.048**	0.019	0.001	0.018	0.000	**0.121**	**0.054**	**0.065**	**0.041**	**0.110**	**0.079**	0.000	0.020	**0.076**	**0.111**	---	0.020	0.019	0.017	0.019	0.019	0.018	0.017	0.018	0.017
SYDm1M	0.015	0.019	0.031	0.010	0.016	0.028	0.007	0.000	0.010	0.000	**0.105**	0.025	**0.037**	0.014	**0.101**	**0.052**	0.000	0.005	**0.061**	**0.065**	0.000	---	0.015	0.014	0.016	0.016	0.015	0.014	0.014	0.013
SYDm2F	0.028	0.030	**0.038**	**0.028**	0.031	**0.040**	0.011	0.005	0.010	0.000	**0.117**	**0.038**	**0.050**	**0.034**	**0.091**	**0.068**	0.000	0.015	**0.066**	**0.094**	0.000	0.000	---	0.013	0.015	0.014	0.014	0.013	0.013	0.012
SYDm2M	0.029	0.024	**0.035**	0.013	0.022	0.025	0.011	0.006	0.026	0.000	**0.110**	**0.037**	**0.040**	0.013	**0.096**	**0.056**	0.000	0.015	**0.061**	**0.054**	0.000	0.000	0.000	---	0.015	0.014	0.014	0.013	0.013	0.012
TANm2F	**0.058**	0.010	**0.048**	0.014	0.020	0.016	0.011	0.000	**0.056**	**0.040**	**0.132**	**0.095**	**0.033**	0.008	**0.158**	**0.118**	**0.048**	**0.066**	**0.045**	**0.053**	0.021	0.015	**0.021**	0.011	---	0.015	0.015	0.013	0.014	0.013
TANm2M	**0.097**	**0.048**	**0.119**	**0.071**	**0.060**	**0.054**	**0.055**	0.019	**0.110**	**0.098**	**0.253**	**0.182**	**0.080**	**0.052**	**0.245**	**0.202**	**0.096**	**0.131**	**0.107**	**0.133**	**0.054**	**0.048**	**0.058**	**0.042**	0.000	---	0.014	0.013	0.013	0.013
WAKgF	**0.153**	**0.128**	0.014	0.017	0.026	0.018	0.003	**0.036**	**0.077**	**0.042**	**0.251**	**0.126**	**0.025**	**0.032**	**0.116**	**0.090**	**0.070**	**0.079**	0.000	**0.035**	**0.074**	**0.051**	**0.056**	**0.046**	**0.038**	**0.083**	---	0.013	0.013	0.012
WAKgM	**0.075**	**0.063**	0.016	0.000	0.000	0.000	0.000	0.017	**0.071**	**0.031**	**0.170**	**0.094**	0.005	0.000	**0.129**	**0.078**	**0.054**	**0.055**	0.000	0.002	**0.059**	0.022	**0.050**	0.020	0.015	**0.055**	0.000	---	0.013	0.012
WAKsF	**0.118**	**0.107**	0.011	0.011	0.005	0.000	0.000	**0.033**	**0.083**	**0.050**	**0.276**	**0.143**	0.003	**0.022**	**0.150**	**0.117**	**0.075**	**0.092**	0.000	**0.047**	**0.080**	**0.051**	**0.065**	**0.048**	**0.033**	**0.063**	0.000	0.000	---	0.011
WAKsM	**0.105**	**0.101**	0.029	0.014	0.004	0.008	0.014	**0.045**	**0.097**	**0.056**	**0.258**	**0.141**	0.011	0.016	**0.170**	**0.116**	**0.083**	**0.094**	0.005	**0.041**	**0.088**	**0.045**	**0.080**	**0.049**	**0.045**	**0.066**	0.005	0.000	0.000	---

1. MMD values below the diagonal, standard deviations above the diagonal, significant differences at α< 0.05 in bold. Abbreviations from Table 1.

This is considered a near-excellent fit of the data (Clarke 1993; De Leeuw & Stoop 1984). Three aggregates, one on the right side of the array, one in the center left, and one in the lower left, and two isolates may be identified (Fig. 20a). The first aggregate, which is on the right side, encompasses 12 samples and two peripheral members. These include males and females of the two Gujjar samples, the two samples of Syeds, as well as male and female Mashwanis and Karlaars. The peripheral members are male and female residents of Madak Lasht, which occupy an isolated position on the right foreground. The second aggregate, which is on the centre left, includes three samples. These include male and female Abbasis and female Tanolis. Their male counterparts occupy an isolated position in the upper left. The third aggregate, which is present in the lower left, includes 11 samples. These are males and females of the two samples of Awans, Khowars, and males and females of the two Wakhi samples, as well as Swati males. Swati females occupy a highly isolated position in the center foreground. When considered by ecological setting lowland samples tend to occupy the center of the array, while highland samples form two aggregates on the left and right sides, respectively (Fig. 20b). Highland samples on the left include Khowars and both males and females of the two Wakhi samples. Highland samples on the right are limited to male and female residents of Madak Lasht. Males and females of the two Gujjar samples stand apart as an exception to this pattern and they occupy the centre and hence intermingle with lowland samples. When considered by ethnicity (Fig. 20c) there is a near complete separation between those groups claiming Pathan ancestry found in the upper centre and upper left of the array, and those believed to be of non-Pathan origin that occupy positions in the foreground and lower left of the array. The only exceptions to this pattern are males and females of the two Gujjar samples.

A Dynamic Perspective based on Inclusion of Archaeologically-derived Samples

Nine archaeologically derived samples from southern Central Asia, Khyber Pakhtunkhwa, and the Indus Valley that date from the Chalcolithic to the Iron Age were included to determine whether members of living ethnic groups of Hazarewal and Chitral-Gilgit-Baltistan evince evidence of gene flow from outside the region prior to the medieval era. Given the absence of samples of living Pathans from their alleged Urheimat (southern Afghanistan, the Kabul Valley, Nuristan) such a comparison can rule out any significant role played by immigrant Neolithic era farming populations of southern Central Asia (de Barros Damgaard *et al*. 2018; Harris 197a, 1997b; Renfrew 1987), by Bronze Age horse-mounted 'Aryan' invaders from the Russo-Kazak steppe (Erdosy 1995; Kuzmina 2001; Narasimhan *et al*. 2018; Parpola 1988) or long-standing continuity with Late Chalcolithic populations of the Indus Valley (Kennedy Chiment, Distotell & Meyers 1984; Kenoyer 2005).

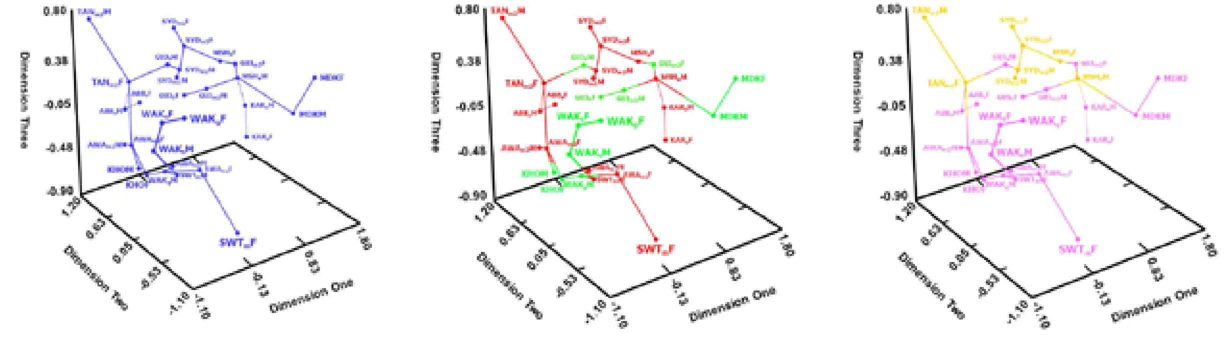

Figure 20. Multidimensional Scaling among Samples of Living Individuals with Sexes Specified: a) All Individuals, b) Highland Individuals (in green) versus Lowland Individuals (in red), c) Pathans (in light brown) versus Non-Pathans (in purple).

A plot of the first two dimensions obtained by correspondence analysis among all members of all samples both living and archaeologically-derived with sexes pooled yields three sample aggregates (Fig. 21a)[8]. The first is a fairly tight cluster in the lower right that includes Abbasis, Karlaars, both samples of Gujjars, and the residents of Madak Lasht. Also joining this aggregate of living individuals is the latest of the Bactrian-Margianan Archaeological Complex (BMAC) from southern Central Asia samples (MOL) and individuals recovered from the Late Bronze Age Gandharan Grave Culture site (TMG) located in Khyber Pakhtunkhwa. The second aggregate is found in the upper right. It includes the two samples of Awans, the two samples of Wakhis, Khowars, Swatis, and Tanolis. In addition to these living samples and occupying a somewhat peripheral position to this aggregate is the Chalcolithic era sample from Harappa (HAR) of the Indus Valley and the Iron Age sample from Sarai Khola (SKH) located at the border between southern Khyber Pakhtunkhwa and northern Punjab. The third aggregate is located on the left side and, with only one exception (Mashwanis), is limited to prehistoric samples from southern Central Asia. When living samples are considered by ecological setting there is no consistent tendency for the archaeologically-derived samples to aggregate with either highland or lowland groups (Fig. 21b). The same is true when ethnicity is taken into account (Fig. 21c).

Principal component analysis of tooth-trait frequencies among living and prehistoric individuals with sexes pooled yields four components with eigenvalues greater than 1.0 (Table 11). The first component accounts for 34.6% of the total variance and draws a distinction among samples between those with relatively high prevalence of shoveling and median lingual ridge development on UI1, Carabelli's trait on UM1 and retention of a well-developed hypoconulid on LM2 from those samples with high prevalence of hypocone reduction on UM2 and loss of the Y-groove on LM2. The second component accounts for 18.2% of the variance. It distinguishes samples with relatively high prevalence of shoveling on UI2 and retention of a well-developed hypoconulid on LM1 from samples that have relatively high prevalence of median lingual ridge development on UM1, and the metaconulid on LM1. Component three accounts for 13.5% of the variance. It emphasizes shoveling on the maxillary incisors and draws a contrast to samples with relatively high prevalence of Cusp 5 on UM2 and median lingual ridge development on UI2. Component 4 draws a contrast between retention of the Y-groove on the first two mandibular molars and the presence of the metaconulid on LM1 versus medial lingual ridge development on UI2. It accounts for 10.0% of the variance.

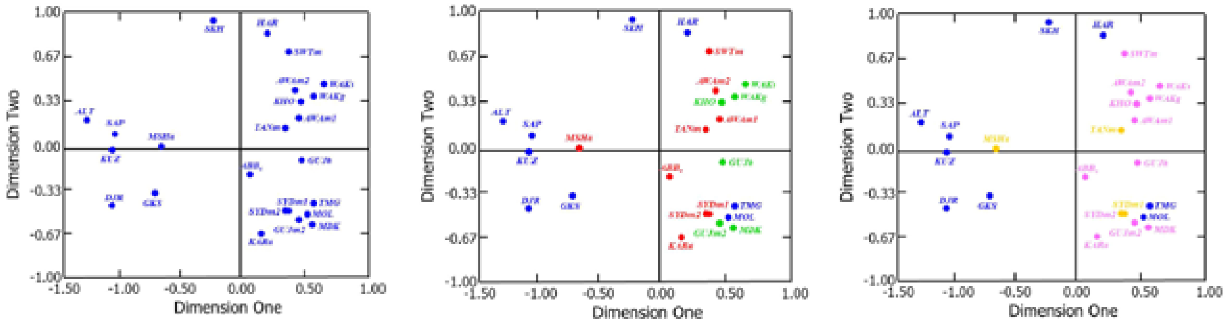

Figure 21. Correspondence Analysis of Sample Centroids among Samples of Archaeologically-derived and Living Individuals with Sexes Pooled: a) All Samples, b) Highland Samples (in green) versus Lowland Samples (in red) versus Archaeologically-derived Samples (dark blue), c) Pathans (in light brown) versus Non-Pathans (in purple), versus Archaeologically-derived Samples (dark blue).

Table 11. Principal Components Analysis among Samples of Sex-pooled Living and Prehistoric Individuals[1]

Trait	Tooth	Loadings			
		1	2	3	4
SHOV	UI1	***0.734***	0.304	***0.475***	0.003
SHOV	UI2	-0.114	***0.777***	***0.478***	0.013
MLR	UI1	***0.852***	-0.083	-0.267	-0.165
MLR	UI2	0.144	0.598	-0.467	***-0.421***
HYPO	UM2	***-0.764***	0.498	-0.080	-0.044
CARA	UM1	***0.883***	-0.038	0.234	-0.074
C5	UM2	0.386	0.251	***-0.669***	0.146
YGRV	LM1	0.289	0.423	-0.414	***0.442***
YGRV	LM2	-0.411	0.390	0.192	***0.628***
CSPN	LM1	0.410	***0.685***	0.075	-0.247
CSPN	LM2	***0.730***	-0.024	0.353	0.085
C7	LM1	0.627	-0.092	-0.157	***0.551***
Eigenvalue		4.157	2.182	1.616	1.195
% Var. Expl.		34.640	18.184	13.463	9.956

1. Most influential variables by component are in bold and italicized.

A plot of scores for the first three components (Fig. 22a) yields three sample aggregates and three isolates. These three components combine to account for 66.3% of the total variance. The first aggregate is found on the left side and is wholly composed of prehistoric samples from southern Central Asia. The second aggregate is found in the upper right and includes 11 samples. All but one of these samples is of living individuals mostly from the lowlands of Khyber Pakhtunkhwa. The only exceptions are the highland samples from Madak Lasht and the two Gujjar samples, as well as the prehistoric sample from the Gandharan Grave site of Timargarha. The third aggregate is found in the lower right. It is composed of five samples. Most are highland samples (KHO, WAKg, WAKs) but also included are the two samples of Awans. The three isolates are found in the lower center of the array. They include Swatis and two prehistoric samples—the Iron Age sample from Sarai Khola and the Chalcolithic sample from Harappa. When sample positions in the three-dimensional plot are considered by ecological setting (Fig. 22b) it is clear that the archaeologically-derived samples tend to aggregate with lowland samples and not with those obtained from the Hindu Kush and Karakoram highlands. When considered by ethnicity (Fig. 22c) archaeologically-derived samples are associated with groups claiming Pathan descent. There are two exceptions. The latest of the Central Asian samples (MOL) is associated with Abbassis, while the Chalcolithic era sample from Harappa is associated with Swatis and one of Awan samples (AWA_{m2}).

The zero-corrected matrix of mean measure of divergence values among archaeologically-derived and living samples with sexes pooled are presented in Table 12. Some 188 of these pairwise contrasts (68.1%) are significant at $\alpha < 0.05$ level. When considered by antiquity 104 of 128 pairwise contrasts (81.3%) differ between archaeologically-derived and living samples, only 5 of 36 pairwise contrasts (13.96%) differ between archaeologically-derived samples, while 79 of 105 (75.2%) pairwise contrasts between sex-pooled samples of living individuals are statistically significant. The disparity in the prevalence of significant pairwise distances between members of living and archaeologically-derived samples (+61.3%) much more likely a consequence of much greater effective sample size for the living samples than to the biological distances separating them.

A neighbour-joining cluster analysis based on the matrix of pairwise mean measure of divergence values among sex-pooled samples of

Table 12. Zero-corrected Mean Measure of Divergence Analysis among Samples of All Individuals with Sexes Pooled[1]

	ABBa	**ALT**	**AWAm1**	**AWAm2**	**DJR**	**GKS**	**GUJh**	**GUJm2**	**HAR**	**KARa**	**KHO**	**KUZ**	**MDK**	**MOL**	**MSHa**	**SAP**	**SKH**	**SWTm**	**SYDm1**	**SYDm2**	**TAN2**	**TMG**	**WAKG**	**WAKS**
ABBa	---	0.059	0.020	**0.010**	**0.023**	0.026	**0.015**	**0.011**	**0.031**	**0.018**	**0.009**	**0.031**	**0.022**	**0.019**	**0.014**	**0.023**	0.044	**0.026**	**0.011**	**0.011**	**0.011**	0.041	**0.019**	**0.010**
ALT	0.103	---	**0.057**	**0.060**	0.073	0.076	**0.063**	**0.061**	0.082	**0.064**	**0.059**	0.080	**0.065**	0.069	**0.062**	0.073	0.093	**0.068**	**0.061**	**0.060**	**0.060**	0.091	**0.064**	**0.060**
AWAm1	0.037	**0.190**	---	0.007	**0.020**	**0.023**	0.012	**0.008**	0.029	**0.015**	0.007	**0.029**	**0.020**	**0.016**	**0.012**	**0.020**	**0.041**	0.024	**0.008**	**0.008**	**0.008**	0.039	0.016	**0.007**
AWAm2	**0.028**	**0.180**	0.004	---	**0.021**	**0.024**	0.013	**0.009**	0.029	**0.016**	0.008	**0.029**	**0.020**	**0.017**	**0.013**	**0.021**	0.042	**0.025**	**0.009**	**0.009**	**0.009**	0.040	0.017	0.008
DJR	**0.133**	0.000	**0.199**	**0.183**	---	0.034	**0.028**	**0.024**	**0.044**	**0.030**	**0.023**	0.044	**0.033**	0.032	**0.027**	0.036	0.057	**0.037**	**0.024**	**0.024**	**0.024**	**0.054**	**0.030**	**0.023**
GKS	0.050	0.000	**0.100**	**0.107**	0.000	---	**0.021**	**0.023**	**0.044**	**0.029**	**0.022**	0.043	**0.032**	0.032	**0.026**	0.036	0.057	**0.036**	**0.024**	**0.023**	**0.023**	0.054	**0.029**	**0.022**
GUJh	**0.041**	**0.184**	0.001	0.019	**0.224**	**0.122**	---	0.008	0.030	**0.017**	**0.009**	**0.030**	**0.021**	**0.018**	**0.014**	**0.022**	**0.043**	**0.026**	0.010	0.010	0.010	0.040	0.018	**0.009**
GUJm2	**0.047**	**0.207**	**0.016**	**0.050**	**0.218**	**0.118**	0.008	---	0.029	0.016	**0.008**	**0.029**	0.020	**0.017**	0.013	**0.021**	**0.042**	**0.025**	0.009	0.009	**0.009**	0.040	**0.017**	**0.008**
HAR	**0.089**	0.123	0.000	0.000	**0.170**	**0.090**	0.002	0.041	---	**0.030**	0.031	**0.052**	**0.040**	0.040	**0.035**	**0.044**	0.065	0.044	0.033	0.032	0.032	0.062	0.038	0.031
KARa	**0.063**	**0.148**	**0.057**	**0.116**	**0.178**	**0.075**	**0.061**	0.012	**0.112**	---	**0.006**	**0.030**	**0.021**	**0.018**	**0.014**	**0.022**	**0.043**	**0.026**	**0.010**	**0.010**	**0.010**	0.040	**0.018**	**0.009**
KHO	**0.030**	**0.200**	0.004	0.000	**0.201**	**0.107**	**0.023**	**0.043**	0.004	**0.096**	---	**0.027**	**0.019**	**0.015**	**0.011**	**0.019**	0.040	**0.024**	**0.007**	**0.007**	**0.006**	0.038	0.015	**0.006**
KUZ	**0.138**	0.000	**0.200**	**0.173**	0.000	0.017	**0.178**	**0.206**	**0.108**	**0.196**	**0.183**	---	**0.028**	0.043	**0.038**	0.047	0.067	**0.048**	**0.035**	**0.035**	**0.035**	0.064	**0.042**	**0.034**
MDK	**0.175**	**0.344**	**0.079**	**0.157**	**0.368**	**0.227**	**0.075**	0.033	**0.114**	**0.046**	**0.138**	**0.369**	---	**0.016**	0.011	**0.020**	**0.041**	**0.024**	**0.008**	**0.007**	**0.007**	**0.039**	**0.016**	**0.007**
MOL	**0.104**	0.000	**0.118**	**0.116**	0.000	0.000	**0.126**	**0.137**	0.050	**0.124**	**0.120**	0.000	**0.239**	---	**0.017**	0.032	0.053	**0.033**	**0.020**	**0.019**	**0.019**	0.050	**0.026**	**0.019**
MSHa	**0.074**	**0.228**	**0.043**	**0.081**	**0.271**	**0.152**	**0.030**	0.006	**0.078**	**0.037**	**0.079**	**0.262**	0.021	**0.174**	---	**0.021**	**0.043**	**0.025**	0.010	0.009	**0.009**	0.040	**0.018**	**0.009**
SAP	**0.112**	0.000	**0.172**	**0.142**	0.000	0.000	**0.184**	**0.202**	**0.109**	**0.182**	**0.150**	0.000	**0.346**	0.000	**0.253**	---	0.054	**0.036**	**0.023**	**0.023**	**0.023**	0.054	**0.029**	**0.022**
SKH	0.031	0.074	**0.104**	0.055	0.074	0.088	**0.110**	**0.136**	0.068	**0.186**	0.052	0.000	**0.350**	0.085	**0.218**	0.053	---	**0.040**	**0.048**	**0.048**	0.047	0.076	**0.054**	**0.047**
SWTm	**0.130**	**0.234**	0.032	**0.068**	**0.295**	**0.168**	**0.058**	**0.081**	0.003	**0.122**	**0.063**	**0.283**	**0.083**	**0.163**	**0.072**	**0.237**	**0.214**	---	**0.007**	**0.008**	**0.008**	**0.039**	0.016	**0.007**
SYDm1	**0.024**	**0.164**	**0.026**	**0.028**	**0.188**	**0.092**	0.008	0.001	0.042	**0.048**	**0.035**	**0.173**	**0.079**	**0.105**	0.012	**0.153**	**0.116**	**0.097**	---	0.008	**0.010**	0.040	**0.018**	**0.009**
SYDm2	**0.026**	**0.178**	**0.023**	**0.027**	**0.200**	**0.093**	0.009	0.005	0.039	**0.049**	**0.033**	**0.187**	**0.077**	**0.110**	0.017	**0.159**	**0.132**	**0.092**	0.000	---	**0.007**	0.039	**0.017**	**0.008**
TANm2	**0.040**	**0.182**	**0.041**	**0.032**	**0.201**	**0.099**	0.017	**0.059**	0.030	**0.126**	**0.036**	**0.154**	**0.159**	**0.091**	**0.084**	**0.130**	0.061	**0.100**	**0.028**	**0.026**	---	0.039	**0.017**	**0.008**
TMG	0.000	0.136	0.000	0.000	**0.116**	0.042	0.000	0.000	0.000	0.035	0.000	0.073	**0.114**	0.058	0.019	0.094	0.000	**0.102**	0.000	0.000	0.000	---	0.039	0.042
WAKg	**0.090**	**0.251**	0.005	0.012	**0.258**	**0.145**	0.014	**0.047**	0.000	**0.117**	0.012	**0.226**	**0.094**	**0.132**	**0.064**	**0.196**	**0.139**	0.029	**0.045**	**0.040**	**0.039**	0.009	---	0.006
WAKs	**0.098**	**0.318**	**0.017**	0.014	**0.312**	**0.189**	**0.026**	**0.064**	0.000	**0.151**	**0.012**	**0.269**	**0.135**	**0.180**	**0.094**	**0.241**	**0.124**	**0.059**	**0.061**	**0.059**	**0.046**	0.000	0.000	---

1. MMD values below the diagonal, standard deviations above the diagonal, significant differences at α< 0.05 in bold. Abbreviations from Table 1.

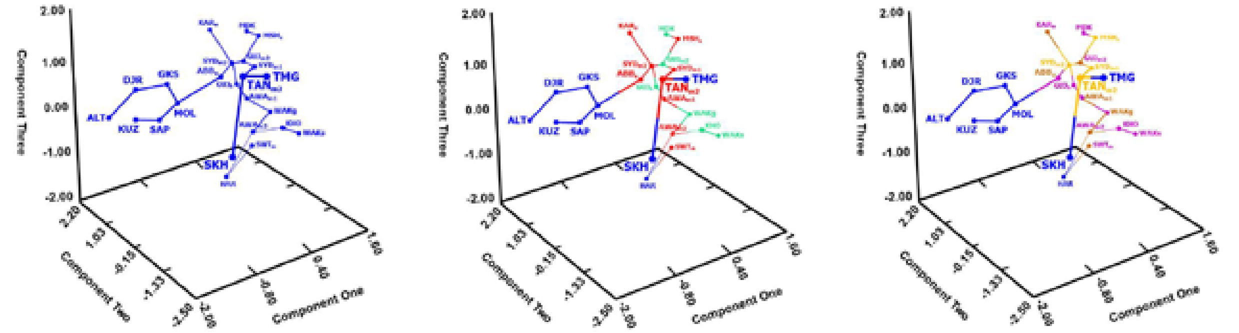

Figure 22. Three-dimensional Plot of Sample Centroids among Samples of Archaeologically-derived and Living Individuals with Sexes Pooled yielded by Principal Components Analysis: a) All Samples, b) Highland Samples (in green) versus Lowland Samples (in red) versus Archaeologically-derived Samples (dark blue), c) Pathans (in light brown) versus Non-Pathans (in purple), versus versus Archaeologically-derived Samples (dark blue).

archaeologically-derived and living samples is provided in Figure 23a. Three aggregates may be identified. The first aggregate is found on the left side of the array and is composed of seven samples. All seven are archaeologically-derived samples from southern Central Asia. The second aggregate is located in the lower center and likewise encompasses seven samples. All except the Chalcolithic era sample from Harappa are samples of living individuals. The third aggregate is a loose grouping of samples on the right side and the upper center. Some 10 samples are included in this aggregate one of which is the Late Bronze Age sample from the Gandharan Grave Culture site of Timargarha. When considered by ecological setting (Fig. 23b)[9] there appears to be no consistent pattern with highland and lowland samples scattered across aggregates two and three. This is definitely not the case when the pattern of affinities is considered by ethnicity (Fig. 23c), for all of the groups claiming Pathan ancestry are members of aggregate three, while none are included in aggregate two.

Multidimensional scaling of pairwise mean measure of divergence distances across sex-pooled archaeologically-derived and living samples with Kruskal's stress formula number 1 was accomplished in four iterations and accounted for 96.9% of the variance with a stress level of 0.074. This considered a good fit of the data (Clarke 1993; De Leeuw & Stoop 1984)[10]. Two aggregates and two isolates may be identified (Fig. 24a). The first aggregate is present on the centre left and is composed of six samples. All are prehistoric samples from southern Central Asia. The second aggregate is a loose collection of 15 samples found in the upper centre and right side of the array. This aggregate may be divided into two subclusters. The first subcluster is present on the right side and encompasses six samples. The second subcluster is present in the upper centre and eight samples. Two peripheral members of this second aggregate are the Abbasis found in the upper centre and the Iron Age sample from Sarai Khola found on the extreme left side. The Chalcolithic sample from Harappa and living Swatis may be considered isolates occupying unique positions in the lower centre. When considered by ecological setting (Fig. 24b) there is no clear patterning among the living samples. Prehistoric samples from Sarai Khola, Timargarha, and Harappa share mixed affinities to lowland and highland samples. When considered by ethnicity (Fig. 24c), groups claiming Pathan ancestry are all members of the second aggregate. Intriguingly, the earlier prehistoric samples from Harappa and southern Central Asia share affinities to non-Pathan samples while the later samples from Sarai Khola and especially Timargarha share closest affinities to groups claiming Pathan ancestry.

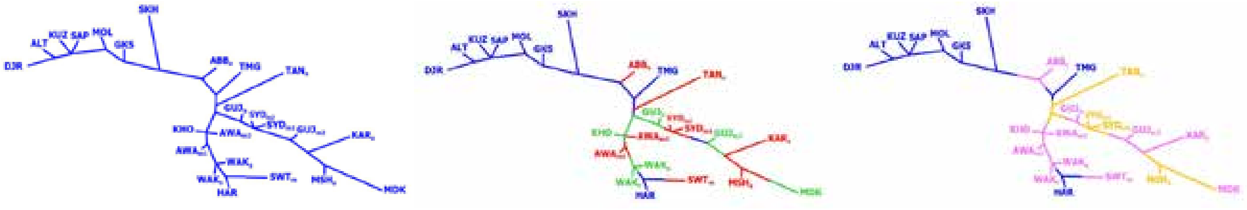

Figure 23. Neighbor-joining Cluster Analysis among Samples of Archaeologically-derived and Living Individuals with Sexes Pooled: a) All Samples, b) Highland Samples (in green) versus Samples Individuals (in red), versus Archaeologically-derived Samples (dark blue) c) Pathans (in light brown) versus Non-Pathans (in purple), versus Archaeologically-derived Samples (dark blue).

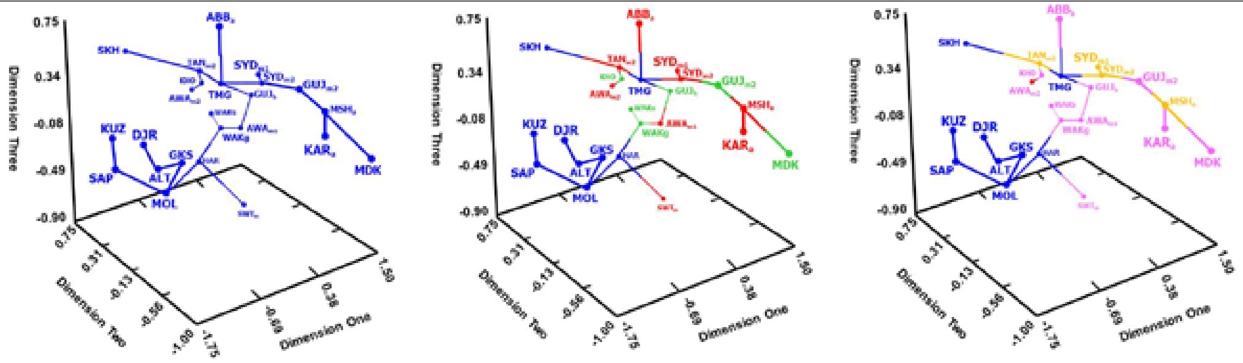

Figure 24. Multidimensional Scaling among Samples of Archaeologically-derived and Living Individuals with Sexes Pooled: a) All Samples, b) Highland Samples (in green) versus Lowland Samples (in red), versus Archaeologically-derived Samples (dark blue), c) Pathans (in light brown) versus Non-Pathans (in purple), versus Archaeologically-derived Samples (dark blue).

Discussion

Of Meaningful Biological Entities and Operational Taxonomic Units (OTUs)

The first issue to be addressed is absolutely fundamental and it is this: Are ethnic groups meaningful biological entities? To determine this multiple geographically distinct samples of members of the same ethnic group were considered. These include Awans, Gujjars, Syeds, and Wakhis. If these ethnic groups are biologically meaningful then geographically distinct samples drawn from the same biological population ought to exhibit closer affinities to one another than to the other samples considered (Bonder, Abein, Zaura & Brandt 2012; Schmidt, Rodrigues & von Mering 2015). Of course, there may be some departure from absolute parity due to subsequent gene flow (Mysara *et al*. 2017), especially in cases of pragmatic, strategic, or hypergamous marriages (Milner 1988) that often comes at considerable postmarital cost (Chaudhry 2019; de Neve 2016; but see Grover 2009).

An examination of the results obtained shows that of the 16 possible pairwise comparisons between multiple-sampled ethnic groups across the various data reduction techniques, only half (50%) of such comparisons identified samples of the same ethnic group as being most similar when sex-pooled samples served as the OTU. Results were better when sexes were considered separately, for 56.3% (9 of 16 contrasts) yielded closest similarities among multiple-sampled ethnic group samples for females and 62.5% (10 of 16 contrasts) for males. The highest proportion of matched ethnic group identifications was obtained when sexes were stipulated, for 75% of such contrasts identified multiple sampled members of the same ethnic group as most similar to one another.[11]

A second fundamental issue concerns statistical significance of pairwise differences. In actuality two items are of concern here. The first is whether specific variables differ from sample to sample and the second concerns the statistical significance of the biodistances between sample pairs. With regard to the first a Shierer-Ray-Hare test (1976) was used to assess the significance of the two main effects (sample, sex) and their interaction. The results of this test are provided in Table 13. The results show that all but one tooth-trait combination (Cusp 7 on LM1) differ significantly across samples. In contrast, none of the tooth-trait combinations differ significantly by sex, nor is the interaction between samples and sex significant. These results are important because while there is evidence for inter-sample differences, there is no basis to claim a sex dimorphism effect (*i.e.,* no significant differences by sex) and no evidence for differential marital migration as has been claimed by an array of workers (ArunKumar 2015; Bamshad 1996, 1998; but see Sengupta et al 2006; Sahoo et al. 2006). However, it should be noted that inter-sample differences only account for about 13% of the variation, while differences due to sex and the interaction between sex and sample account for only 2% and 1.6% of the variance, respectively.

The second issue concerns the statistical significance of pairwise mean measure of divergence biodistances. According to Harris and Sjovøld (2004), there are two considerations that call the testing of Smith's MMD into question. First, just what does a significant biodistance actually mean? The meaning is quite vague from a biological sense as it can be contingent upon the groups selected for analysis. After all, if samples are separated by relatively great geographic distances (Slatkin & Maddison 1990; Wright 1943), if they have different mating cycles (Feder, Chilcote, & Bush 1988; Feder & Filchak 1999; Smith 1988), different mating behaviors (Ambaryan, Voznessenskaya, & Kotenkova 2019; Giglio & Dyer 2013; Ryan & Causey 1989; Setoguchi, Takamori, Aotsuka, Sese, Ishikawa, & Matsuo 2014), or if they differ markedly with respect to culture and/or language (Cavalli-Sforza, Piazza, Menozzi & Mountain 1988; Gray, Atkinson & Greenhill 2011; Renfrew 1987; but see Skutnabb-Kangas & Harmon 2018), then they are already likely to represent distinct populations. This is especially the case in studies like this one

Table 13. Shierer-Ray-Hare tests for Significant Differences in Trait Frequencies across Samples of Living Individuals by Ethnic Group and by Sex[1]

		Sample		Sex		Sample x Sex	
Trait	Tooth	*H*	*p*	*H*	*p*	*H*	*p*
SHOV	UI1	6.813	>0.0001	2.628	0.1050*	1.149	0.3100*
SHOV	UI2	6.772	>0.0001	1.325	0.2500*	1.130	0.3270*
MLR	UI1	5.614	>0.0001	2.419	0.1200*	0.790	0.6810*
MLR	UI2	2.185	0.0070	0.001	0.9700*	0.958	0.4950*
HYPO	UM2	1.991	0.0160	1.016	0.3140*	1.834	0.0300*
CARA	UM1	6.192	>0.0001	0.257	0.6120*	0.695	0.7800*
C5	UM1	47.023	>0.0001	3.703	0.0550*	0.538	0.9110*
YGRV	LM1	4.261	>0.0001	0.000	0.9950*	0.607	0.8610*
YGRV	LM2	2.965	>0.0001	0.707	0.4010*	0.574	0.8860*
C7	LM1	1.444	0.1270*	2.164	0.1420*	0.813	0.6560*

1. Nonsignificant contrasts across the 15 samples of living individuals at α< 0.05 marked by an asterisk.

	H	*p*	Sig.	η^2
Sample	8.249	<0.001	Yes	0.128
Sex	1.571	0.111	No	0.020
Sample x Sex	0.165	0.715	No	0.016

in which, due to the inclusion of archaeologically-derived remains, samples differ in time and hence cannot be considered to constitute a single biological population of potentially interbreeding individuals (Mayr 1942, 1963).

Second, as mentioned earlier, a crucial point in any investigation of biological divergence or distance is the choice of variable(s) that form the basis of the comparison (Sjovøld 1977; Xu, Liu, An & Jia 2016). Because Smith's biodistance statistic is the *mean* measure of divergence, the size of the difference between sample pairs can be increased or reduced simply by varying the trait list used (see Sokol & Sneath 1963; Reyment 1991). Two considerations are relevant here and have been taken into consideration. First, do the traits differ significantly across the samples being compared and second, does the inclusion of the variable(s) increase or decrease intergroup differences among the samples being compared.

To these two considerations identified by Harris and Sjovøld a third may be added. When the researcher is contrasting samples, such as the living ethnic groups considered in the current study in which gene flow across such groups is possible, the OTUs are inherently porous to within species introgression (Cronquist 1978; Pigliucci 2003; Sneath 1976; Zachos 2018). Consequently, the boundaries between them are inherently "fuzzy." In such cases, statistical significance is not as important as the *patterning* of pairwise differences between samples.

Consequently, what the researcher is seeking to determine is at which OTU level the results from different data reduction techniques yield consistent results (Choi, Neiminen, Bahr, & Bahr 2002; Das, Sengupta & Bhattacharyya 2018; Dash, Liu & Motoda 2000; Fjeldså 2003). In this way, the results obtained may be considered robust and not but a mere artifact of the data reduction technique employed. An examination of the aggregates and isolates identified by the four data reduction techniques among members of living samples with sexes pooled yields many inconsistencies. While correspondence analysis identifies Tanolis as members of an aggregate that includes the two samples of Awans and Khowars this result was not yielded by any of the other data reduction techniques. Likewise, while three data reduction techniques identified the two samples of Syeds as members of the same aggregate this was not the case for principal components analysis. A similar disagreement occurs for Swatis. Two of the data reduction techniques (PCA, MDS) identify Swatis as an outlier to all other samples, whereas correspondence analysis and neighbor-joining cluster analysis identify them as possessing affinities to the two Wakhi samples. It is likely this volatility in obtained results reflects conflation of differences due to sample and differences due to sex in which these two vectors of divergences can obfuscate patterns of affinities among samples.

An obvious alternative is to limit contacts between samples to members of a single sex. In this way differences due to sex are not conflated with differences by ethnic group. When consideration is limited to females, consistency across data reduction techniques in the identification of aggregates and isolates is only improved slightly. While both principal components and neighbor-joining cluster analysis identify a large aggregate that encompasses the two samples of Awans, the two samples of Wakhis, Khowars and Swatis, this aggregate was not yielded by either correspondence analysis or multidimensional scaling. The position of one of the Gujjar samples (GUJ_{m2}) appears particularly volatile. It is identified as an isolate by CA and PCA, whereas neighbor-joining cluster analysis places it in an aggregate with Mashwanis, Tanolis and the two samples of Syeds, while MDS identifies it as part of an aggregate along with Karlaars and the residents of Madak Lasht. Results are somewhat more consistent when consideration is limited to males. Three of the four data reduction techniques yield an aggregate that includes the two samples of Awans, the two samples of Wakhis, and Khowars. The exception is correspondence analysis, which identifies the Wakhi sample from Sost as an outlier to all other samples. Less consistently, an aggregate is identified that includes the two samples of Gujjars and the two samples of Syeds. However, this is not the case for either PCA or for neighbor-joining cluster analysis. Such results *could* indicate differential post-marital migration, but this is difficult to determine from two sex-segregated analyses.

A consideration of the patterning of sample aggregates and isolates with sexes-stipulated yields the most consistent results. All three data reduction techniques identified a large aggregate that includes the two samples of Awan males and females, the two samples of Wakhi males and females and one if not both sexes of Khowars. A second aggregate that includes both samples of Syed males and females and Mashwani males and females and, in most cases, males and females of the two Gujjar samples are identified by CAs and MDS, but not by PCA. All three data reduction techniques identify Abbasis as isolates with little affinity to males and females of any of the other samples. Likewise, all three data reduction techniques identify males and females of the two Syed samples as sharing close affinities to one another. Considered as a whole, it appears that in most cases males and females of the same ethnic group exhibit closest affinities to one another. There are a few discrepancies, but such results suggest strongly that there has been no differential post-marital migration across the ethnic groups of Chitral-Gilgit-Baltistan or Hazarewal. Such results stand at odds with the mtDNA and Y-chromosome survey of Tariq, Ahmad, Hemphill, Farooq, & Schurr (2022). But given that their study was conducted on members of ethnic groups located further west in Buner and Swabi Districts, the differences observed in post-marital migration may be a consequence of a lack of Pathan incursion further east beyond the River Indus (see below).

Is there Evidence of a Pathan Presence in Hazarewal and Chitral-Gilgit-Baltistan?

The historical record suggests entry of Pathan populations into the Vale of Peshawar from the west and/or northwest soon after the beginning of the second millennium CE (Barth 1969; Caroe 1958; Ghirshman 1954; McGovern 1939; Smith 1924 [2022]). However, some maintain that the River Indus served as the eastern border of actual Pathan immigration and that any further Pathan influence eastward was accomplished via local non-Pathan intermediaries. If there was a mediaeval era intrusion of Pathans into Hazarewal and Chitral-Gilgit-Baltistan then there ought to be a biodistance discontinuity between ethnic groups whose ancestors are to be found locally, or even further to the east in the case of Gujjars, and those who claim to be descendants of emigrant Pathans.

Regardless of the OTU considered, the results obtained from a fair number of the data reduction techniques does identify a distinction between groups claiming Pathan ancestry and those believed to have non-Pathan origins. When sexes are pooled both PCA and MDS show a fairly clear separation of groups by ethnicity. Results are less clear-cut for correspondence analysis and neighbor-joining cluster analysis. Nevertheless, regardless of data reduction technique employed there are non-Pathan samples that occupy positions within or immediately adjacent to those occupied by Pathan samples. Still further, while there is a distinction between Pathan and non-Pathan ethnic groups, none show the profound separation expected if there was a substantial population intrusion into Hazarewal and Chiral-Gilgit-Baltistan from outside. Similar results were obtained when segregated sexes serve as the OTU. In fact, a fairly clear distinction by ethnicity is identified among males in three of the four data reduction techniques, yielding distinct patterns separating groups claiming Pathan origins from those believed to be of non-Pathan descent. The only exception is correspondence analysis. However, as was the case when sex-pooled samples served as the OTU, there are exceptions and the most common is for the Gujjar samples to be identified as most similar to Pathan groups. When stipulated sexes serve as the OTU the three data reduction techniques identify an ethnic distinction between Pathans and non-Pathans. Once again there are exceptions. In this case, not only are Gujjar males and females identified as proximate to Pathan groups, but Tanoli males and females are identified as divergent from all other Pathan samples considered.

Two important findings may be identified when ethnic affinities are considered as a whole across the various data reduction techniques and with the various OTUs employed. First, there is evidence of an ethnic distinction between groups claiming Pathan ancestry and those believed to be of non-Pathan origin. However, in most cases, these differences are not profound and do not indicate any large-scale influx of populations into the local region from the believed Pathan Urheimat to the west or northwest in Afghanistan. Second, exceptions to an ethnic segregation among samples are not infrequent. Intriguingly, these most often take two forms, Gujjars being identified as proximate to Pathan groups and Tanolis being identified as divergent from other Pathan groups. Turning to Barth (1956, 1969), it may be that the phenetic proximity between Gujjar and Pathan samples may be a reflection of long-standing *jajmani*-like relationships between those Gujjars living in close proximity or even within Pathan communities under a scenario of niche segregation. Alternatively, it may be that relatively few Pathan immigrants made it this far east in Khyber Pakhtunkhwa. As such, they may have found it pragmatic to engage in inter-ethnic marriages and Gujjar populations proved receptive to such strategic marriages. Resolution of these issues will have to await additional research, especially in regions of Khyber Pakhtunkhwa further to the west within the Vale of Peshawar.

Were Ethnic Groups of This Region Differentiated Spatially by Adaptation to Different Biomic Niches?

As noted in the introduction, Caroe (1958) claimed that invading Pathan ethnic groups with their greater organization and demand for surplus agricultural production favored settlement of the most productive lowland environments. In a related vein, Barth (1956) claimed that Pathan and some Gujjar populations were engaged in a scenario of niche segregation in which a small number of Gujjars were engaged as political clients and economic serfs serving settled Pathan communities. Other Gujjars, to which he refers to as 'transhumant Gujjars', occupied the surrounding highlands that provided summer pasturage for their herds and were associated with specific Pathan villages to which they paid pasturage rent. If Caroe and Barth are correct with regard to Pathan settlement patterns, one ought to see Pathan ethnic groups occupying the lowlands while groups of non-Pathan ancestry occupy the surrounding uplands to which they had been marginalized by these more highly-organized invaders. A possible exception are Gujjars which may have occupied lowland areas as clients of their Pathan overlords or as reciprocal agents residing in the uplands with their herds but engaged in recurrent economic activities with members of these lowland Pathan communities.

Patterning by biome is much less clear than by ethnicity. Many of the analyses, regardless of data reduction technique or the OTU considered, yield no discernable pattern. This was the case for two of the data reduction techniques when pooled sexes served as the OTU and when females were the basis of comparison. Results were even poorer when males served as the OTU. When sexes were stipulated two of three analyses failed to yield any discernable pattern. The only exception was correspondence analysis in which highland-residing groups tended to occupy positions in the lower left of the array while lowland-residing groups occupy the upper right. As was the case when samples were compared by ethnicity, there were numerous exceptions in which highland-residing groups were found in association with lowland groups and vice versa. Nevertheless, definite trends could be identified. Two of the data reduction techniques employed with sex-pooled OTUs identified the Awan samples as associated with highland samples. Correspondence analysis identified one of the Gujjar samples (GUJ_{m2}) as proximate to lowland samples, while PCA identified the two Gujjar samples as occupying positions intermediate between lowland and highland samples. Similar results were obtained when females served as the OTU. Awans and Swatis were associated with highland samples by correspondence analysis, while PCA identified Awans as associated with highland samples, but the Gujjar samples—especially GUJ_h—were placed with lowland samples being most proximate to Syeds. When males served as the OTU all but neighbor-joining cluster analysis associated the two Gujjar samples with lowland samples. With sexes specified, both CA and MDS associated Awan males and females with highland samples. The former also identified Swati females as associated with highland samples, while the latter identified their male counterparts as associated with highland samples. Correspondence analysis identified GUJ_{m2} females and Khowar males as associated with lowland samples, while MDS identified GUJ_h males and GUJ_{m2} females as associated with lowland samples.

Viewed as a whole it appears clear that there is far less distinction between highland and lowland groups than between groups claiming Pathan versus non-Pathan ancestry. Two important observations may be made. First, the biomic stratification identified by Barth in Swat does not appear to manifest further east in Hazarewal or further north in Chitral-Gilgit-Baltistan. As noted above, this may be a consequence of the relatively few Pathan immigrants to these regions. Second, there are consistent exceptions to lowland samples being associated with other samples from lowland contexts and highland samples being associated with other samples obtained from the highlands. These exceptions appear fairly consistent regardless of OTU or data reduction technique employed. The first involves Awans and Swatis, which were obtained from lowland contexts, but which were often associated with samples obtained from the highlands. It may be that in the past Awans and Swatis once resided in the highlands but were subsequently displaced into lowland contexts. Second, one or both of the samples of Gujjars were found to be either associated with lowland context samples—especially Syeds—or occupy positions intermediate between the phenetic spaces occupied by highland and lowland samples. It may be that the sampled Gujjars are those who reside within or adjacent to Pathan communities in the lowlands and are engaged in *jajman*i-like economic relationships with their more powerful Pathan overloads. Unfortunately, the specific contexts of the Gujjar individuals sampled by economic lifeway (economic serfs vs. transhumant herders) was not undertaken at the time these dental casts were obtained. Resolution of this issue will have to await further studies that take such relationships into account.

Is there Evidence of pre-Mediaeval Temporally Distinct Population Incursions into the Region?

It has been suggested that the populations of this region were affected by gene flow from Indus Valley populations during the Chalcolithic era due to expansion of the Indus Civilization and the quest by populations of this region for raw materials not available on the valley floor (Kenoyer 1998; Law 2011). Indeed, the presence of the Harappan outpost at Shortugai in northeastern Afghanistan (Francfort 1981, 1988) and the presence of Harappan artifacts in Namazga Period IV deposits at Altyn Depe in southern Turkmenistan (Gupta 1979; Kohl 1992; Masson 1988; Masson & Sarianidi 1972) suggest communication across Khyber Pakhtunkhwa and Chitral-Gilgit-Baltistan during the third millennium BCE. Other researchers emphasize contacts between Bronze Age era communities of southern Central Asia –either of the BMAC urban centres or of steppe Andronovo populations—across the Hindu Kush and into Khyber Pakhtunkhwa and beyond during the second millennium BC (Dani 1967, 1968; de Barros Damgaard *et al*. 2018; Erdosy 1995; Kuzmina 2001; Narasimhan *et al*. 2019; Parpola 1988).

If the first scenario is correct, then there ought to be a unique association between the Chalcolithic era human remains recovered from Cemetery R37 at Harappa (HAR) and at least some of the ethnic groups from Hazarewal and Chitral-Gilgit-Baltistan, especially those of the highland regions who are claimed to have been least affected by gene flow from intrusive Pathan populations of the mediaeval period. The same ought to be true if the second scenario of an influx of southern Central Asians crossed the Hindu Kush and swept across Chitral-Gilgit-Baltistan and Hazarewal. Such gene flow may have compounded earlier influences from the Indus Valley or may represent a single instance of gene flow into these populations.

Because of small sample sizes, fragmentary remains and uncertain identification of individuals by sex, analysis was limited to correspondence analysis, neighbor-joining cluster analysis and multidimensional scaling in which the OTU were pooled sexes. Correspondence analysis places prehistoric Central Asians on the left side of the array and samples of members of living ethnic groups of Hazarewal and Chitral-Gilgit-Baltistan on the left. There are several exceptions. Mashwanis occupy a position on the left along with prehistoric Central Asians, while the prehistoric samples from the Indus Valley (SKH, HAR) occupy rather isolated positions in the upper center. In contrast, the Gandharan Gave

Culture sample from Timargarha (TMG), which is located in Lower Dir to the west in Khyber Pakhtunkhwa, and Molali (MOL)—the latest of the BMAC samples from southern Central Asia—occupy positions in the lower right adjacent to Syeds, Gujjars, and the residents of Madak Lasht. Neighbor-joining cluster analysis segregates the prehistoric samples from southern Central Asia and from Sarai Khola (SKH) on the left. The Chalcolithic era sample from Harappa is associated with highland samples, while the Gandharan Grave Culture sample from Timargarha is positioned in between Abbasis and Tanolis and near to Gujjars from Haripur (GUJ_h) and the two samples of Syeds. Multidimensional scaling segregates the prehistoric samples from southern Central Asia on the lower left. Chalcolithic era HAR from the Indus Valley stands as intermediate between these samples and the two Wakhi samples. Iron Age Sarai Khola occupies an isolated position in the upper left, while Swatis occupy an isolated position in the lower foreground. TMG occupies a position in the upper center that stands in between the positions occupied by Tanolis, Gujjars from Haripur (GUJ_h) and the two samples of Syeds.

Considered as a whole, there is little support for a contribution from Chalcolithic era populations from the Indus Valley, as represented by the Cemetery R37 from Harappa, on living ethnic groups of Hazarewal and Chitral-Gilgit-Baltistan. Contrary to the assertions of a number of researchers on the basis of archaeology (Kuzmina 2001), linguistic references in the *Rig Veda* (Erdosy 1995; Parpola 1988) or from amplification of aDNA (de Barrros Damgaard et al 2018; Narasimhan *et al*. 2019), there is likewise little support for a lasting Bronze Age era influx from southern Central Asia. Instead, despite the claims of an external origin for the Gandharan Grave Culture (Dani 1967, 1968), the phenetic affinities of Late Bronze/Early Iron Age Timargarha firmly situate it within the constellation of ethnic groups of Hazarewal and Chitral-Gilgit-Baltistan with essentially equivalent affinities to ethnic groups residing in the highlands and those residing in the lowlands. Intriguingly, apart from Mashwanis, the remains from Timargarha exhibit closer affinities to those groups claiming Pathan descent—including Pathan-associated Gujjars (see above)—than with ethnic groups whose ancestry is believed to be non-Pathan. There are two possibilities that may have contributed to this result. First, it may be that Pathan origins— at least for those whose descendants are to be found in Hazarewal —may be in northeastern Afghanistan, rather than either the regions Kabul or Kandahar. Second, it may be that the phenetic proximity of Timargarha to the Pathan ethnic groups of Hazarewal is a secondary consequence of Timargarha's location further to the west in Lower Dir and hence incoming Pathan populations may have experienced greater gene flow with ethnic groups found in this region of Khyber Pakhtunkhwa and these affinities were subsequently transferred, albeit in a diluted fashion, to affinities with the Pathan ethnic groups of Hazarewal. Alternatively, the phenetic affinities shared between Timargarha and the ethnic groups of Hazarewal and Chitral-Gilgit-Baltistan may simply reflect long-term local biological continuity in this region of northern Pakistan—a continuity that extends at least over the course of the five millennia (Kennedy *et al*. 1984; Sahoo *et al*. 2006; Sengupta *et al*. 2006).

Notes

1. This is especially the case in lowland agriculturally productive southern Khyber Pakhtunkhwa, which is for all intents and purposes exclusively Sunni. The only enclaves of Shi'ism are to be found in the highlands of Gilgit-Baltistan where ecological conditions and topography render agricultural production a difficult and highly labour intensive endeavour.
2. An alternative genealogy has Karlanri being adopted by Sharkhbun, the elder son of Sarbanr, and being given the name Urmar.
3. A good example of this has been described by Walker and Hewlett (1990) among Central African foragers and farmers. The same physical ecological setting of the Ituri forest of the southern region of the Central African Republic and the northeastern region of Zaire. Aka, Mbuti

and Efe pygmy groups are mobile tropical forest foragers who wrest a living as net- and bow-hunters. They view the forest as a nurturing friendly place. In contrast, Bantu-speaking farmers, who occupy cleared regions of the forest and who subsist on manioc, plantains, maize, rice and peanuts, view the very same forest as dark, forbidding and dangerous.

4. As originally conceived by Butler (1939), morphogenetic fields are specific regions of the dentition that largely, but not entirely, conform to tooth types that are considered to be under genetic control that affects their differentiation, size and shape (Line 2001; Mitsidasis & Smith 2006; Sharpe 1995; Townsend & Brown 1981; Townsend, Harris, Lesot, Clauss & Brook 2009).

5. Consideration of inter-trait correlation was limited to living samples because complete data by individual is quite rare among archaeologically-derived samples, not only because of small samples, but also for the two reasons identified above that affect the representativeness of archaeologically-derived samples for both anterior (post-depositional loss) and posterior teeth (excessive wear)

6. While the stress threshold is not a concern with the large sample sizes involved with living samples, some caution against over-fitting of models can occur with the small sample sizes available for archaeologically-derived samples (Borg & Groenen 2005; Kruskal & Wish 1978; McCune & Grace 2002) and when dissimilarity measures yield many tied distances (McCune & Mefford 2016).

7. With some 30 data points it was found that the three-dimensional plot of PCA scores was overly confusing and did not add any new information over two-dimensional plot. Consequently, the two-dimensional plot is presented here.

8. Consideration was limited to sex-pooled samples for two reasons. First, considered overall, the archaeologically-derived samples are much smaller with regard to the number of individuals represented relative to the samples of living individuals. Second, while there is complete certainty, with regard to the sex of individuals included in the living samples, this is not the case for the archaeologically-derived samples. This is for two reasons: 1) diagnosis of sex may be ambiguous based on cranial and/or pelvic remains, and 2) sex could not be identified for individuals represented by fragmentary remains.

9. All of the archaeologically-derived samples may be considered lowland samples, but for consistency with previous analyses consideration by ecological setting was limited to living samples to determine whether inclusion of the archaeologically-derived samples had any effect on their patterning with regard to ecological setting.

10. As noted previously, while the stress threshold is not a concern with the large sample sizes involved with living samples, some caution against over-fitting of models can occur with the same sample sizes incurred with archaeologically-derived samples (Borg & Groenen 2005; Kruskal & Wish 1978; McCune & Grace 2002) and when dissimilarity measures yield tied distances (McCune & Mefford 2016). Indeed, an inspection of Tables 1 and 12 indicates that the archaeological samples included in this analysis suffer from both compromising factors.

11. The sex-stipulated contrast is somewhat different than those based on pooled sexes and on sex-segregated samples in three ways. First, due to the number of data points, ordination of principal component scores was limited to the first two components to ease interpretation. Second, no neighbor-joining analysis was performed. Third, if three of four data points for the geographically-distinct, sex-stipulated matched ethnic group samples were found to be most similar, then this was considered a positive result.

References

Achakzai N.M., Rahman, Z., Shahzad, M.S., Saud, S., Zar, M.S., Israr, M.,…Roewer, L.
2012. Y-chromosomal STR analysis in the Pashtun population of southern Afghanistan. *Forensic Science International: Genetics* 6(4): e103-e105. https://doi.org/10.1016/j.fsigen.2011.10.005

Ahmad, Z.
1962. Muslim caste in Uttar Pradesh. *The Economic Weekly* 14: 325-336.

Akbar, N., Inamullah, Ahmad, H., Ali, Ihsan, Hemphill, B.E.
2023. The Role of Hazarewal Populations in the Peopling of South Asia: A Dental Morphology Investigation. *American Journal of Biological Anthropology* 180(4): 673-702.

Ambaryan, A.V., Voznessenskaya, V.V., & Kotenkova, E.V.
2019. Mating Behavior Differences in Monogamous and Polygamous Sympatric Closely Related Species *Mus musculus* and *Mus spicilegus* and their Role in Behavioral Precopulatory Isolation. *Russion Journal of Theriology* 18(2): 1-13.

Anthony, D.W.
2007. *The Horse, the Wheel and Language: How Bronze Age Riders from the Eurasian Steppes shaped the Modern World*. Princeton: Princeton University Press.

Arita, H.T.
1997. Species Composition and Morphological Structure of the Bat Fauna of Yucatan, Mexico. *Journal of Animal Ecology* 66(1): 83-97.

ArunKumar, G., Tatarinova, T.V., Duty, J., Rollo, D., Syama, A., Arun, V.S…The Genographic Consortium.
2015. Genome-wide Signatures of Male-mediated Migration Shapinng the Indian Gene Pool. *Journal of Human Genetics* 60: 493-499.

Avari, B.
2013. *Islamic Civilization in South Asia*. Abingdon: Routledge.

Bacon, E.
1946. A preliminary attempt to determine the culture areas of Asia. *Southwestern Journal of Anthropology* 2: 117-132.

Bamshad, M., Fraley, A.E., Crawford, M.H., Cann, R.L., Busi, B.R., Naidu, J.M., & Jorde, L.B.
1996. mtDNA Variation in Caste Populations of Andhra Pradesh, India. *Human Biology* 68(1): 1-28.

Bamshad, M., Watkins, W.S., Dixon, M.E., Jorde, L.B., Rao, B.B., Naidu, J.M…Hammer, M.F.
1998. Female Gene Flow Stratifies Hindu Castes. *Nature* 395: 651-652.

Barth, F.
1956. Ecologic relationships of ethnic groups in Swat, North Pakistan. *American Anthropologist* 58: 1079-1089.

Barth, F.
1965. *Political Leadership among Swat Pathans*. London: Athlone Press.

Barth, F.
1969. The system of social stratification in Swat, North Pakistan. In: Barth, F. (Ed.), *Aspects of Caste in South India, Ceylon and North-West Pakistan* (pp. 113-146). Cambridge: Cambridge University Press.

Beča, J.
1969. *A Study in Pashto Stress*. Prague: Oriental Institute in Academia, Czechoslovak Academy of Sciences

Bellew, H.W.
1864 [2013]. *A General Report on the Yusufzais*. Lahore: Sang-e-Meel.

Bellew, H.W.
1872 [1998]. *From the Indus to the Tigris: A Narrative of a Journey through the Countries of Balochistan, Afghanistan, Khorassan in Iran in 1872*. Lahore: Sang-e-Meel.

Bhatti, S., Abbas, S., Aslamkhan, M., Attimonelli, M., Trinidad, M.S., Aydin, H.H.,…Gonzalez, G.R.
2018. Genetic perspective of uniparental mitochondrial DNA landscape on the Punjabi population, Pakistan. *Mitochondrial DNA Part A* 29(5): 714-726.

Bhatti, S., Aslamkhan, M., Abbas, S., Attimonelli, M., Aydin, H.H., & Silva de Souza, E.M.

2017. Genetic analysis of mitochondrial DNA control region variations in four tribes of Khyber Pakhtunkhwa, Pakistan. *Mitochondrial DVA Part A* 28(5): 687-697.

Blackburn, S.

2009. Oral stories and culture areas: from northeast India to southwest China. *Journal of South Asian Studies* 30(3): 419-437.

Blaylock, S.R., & Hemphill, B.E.

2007. Are the Koh Indigenous Inhabitants of the Hindu Kush? II. A Dental Morphology Investigation. *American Journal of Physical Anthropology* 132(S44): 76.

Bonder, M.J., Abein, A., Zaura, E., and Brandt, B.W.

2012. Comparing Clustering and Pre-processing in Taxonomy Analysis. *Bioinformatics* 28: 2891-2897.

Borg, I. & Groenen, P.

2005. *Modern Multidimensional Scaling—Theory and Applications*. Berlin: Springer.

Butler, P.T.

1939. Studies of the mammalian Dentition. Differentiation of the Post-canine Dentition. *Proceedings of the Zoological Society of London*, *Series B* 109: 1-36.

Caldwell, B.

1991. *The Jajmani System: An Investigation*. Delhi: Hindustan Publishing Corporation.

Caroe O.

1958. *The Pathans 550 B.C. – A.D.*

1957. London: Macmillan & Company Ltd.

Cavalli-Sforza, L.L., Piazza, A., Menozzi, P. & Mountain, J.

1988. Reconstruction of Human Evolution: Bringing Together Genetic, Archaeological, and Linguistic Data. *Proceedings of the National Academy of Sciences USA* 85: 6002-6006.

Choi, S.H.-J., Nieminen, T.A., Bahr, M., & Bahr, N.

2002. Improving Behaviour Classification Consistency: A Technique from Biological Taxonomy. In: *AARE—Australian Association for Research in Education Conference 2002*. Brisbane (published online at http://www.aare.edu.au/02pap/cho02101.htm).

Choudhry, S.

2019. 'For How Long can your Pīharwāle Intervene?': Assessing Natal Kin Support in Rural North India. *Modern Asian Studies* 53(5): 1613-1645.

Clarke, K.R.

1993. Non-parametric Multivariate Analyses of Changes in Community Structure. *Australian Journal of Ecology* 18: 117-143.

Constantini, L.

1984. The beginning of agriculture in the Kachi Plain: the evidence of Mehrgarh. In Allchin, B. (Ed.), *South Asian Archaeology 1981* (pp. 29-33). Cambridge: Cambridge University Press.

Constantini, L.

2008. The first farmers in western Pakistan: the evidence of the Neolithic agropastoral settlement of Mehrgarh. *Pragdhara* 18: 167-178.

Coon, C.S.

1951. *Caravan*. New York: Henry Holt.

Cronquist, A.

1978. Once Again, What is a Species? In: Ramberger, J.A. (Ed.), *Biosystematics and Agriculture* (pp. 3-20). Monclair, NJ: Allanheld & Osmun.

Dani, A.H.

1967. *Timargarha and Gandhara Grave Culture*. Peshawar: *Ancient Pakistan* 3.

1968. Ganhara Grave Complex in West Pakistan. *Asian Perspectives* 11: 99-110.

Das, A.K., Sengupta, S., & Bhattacharayya, S.

2018. A Group Incremental Feature Selection Classification using Rough Set Theory based Genetic Algorithm. *Applied Soft Computing* 65: 400-411.

Dash, M., Liu, H. & Motoda, H.

2000. Consistency based Feature Selection. *Proceedings of the Fourth Pacific Asia Conference on Knowledge Discovery and Data Mining* (pp. 98-109). London: Springer-Verlag.

de Barros Damgaard, P., Martiniano, R., Kamm, J., Moreno-Mayar, J.V., Kroonen, G.,… Willerslev, E.
2018. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* 360: eaar7711.

de Leeuw, J. & Stoop, I.
1984. Upper Bounds for Kruskal's Stress. *Psychometrika* 49: 391-402.

de Neve, G.
2016. The Economies of Love: Love Marriage, Kin Support, and Aspiration in a South Indian Garment City. *Modern Asian Studies* 50(4): 1220-1249.

Di Cristofaro J, Pennarun E, Mazières S, Myres NM, Lin AA, Temori SA,…Chiaroni, J.
2013. Afghan Hindu Kush: Where Eurasian sub-continent gene flows converge. *PLoS ONE* 8(10): e76748.

Dietz, S.
2007. Buddhism in Gandhara. In Hirman, A. & Bumbacher, S.P. (Eds.), *The Spread of Buddhism* (pp. 60-85). Leiden: Brill

Digby, S.
2004. Before Timur Came: Provincialization of the Delhi Sultanate through the Fourteenth Century. *Journal of Economic and Social History of the Orient* 47(3): 298-356.

Dyson, S.L.
1993. From New to New Age Archaeology: Archaeological Theory and Classical Archaeology—A 1990s Perspective. *American Journal of Archaeology* 97(2): 195-206.

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

Fairservis, W.A.
1971. *The Roots of Ancient India*. New York: Macmillan.

Falk, H.
2003. A Copper Plate Donation Record and some Seals from the Kashmir Smast. *Beiträge zur Allegemeinen und Vergleichenden Archälogie* 23: 1-19.

Feder, J.L., Chilcote, C.A., & Bush, G.L.
1988. Genetic Differentiation between Sympatric Host Races of the Apple Maggot *Rhagoletis pomonella Nature* 336: 61-64.

Feder, J.L., & Filchak, K.E.
1999. It's About Time: The Evidence for Host Plant-mediated Selection in the Apple Maggot Fly *Rhagoletis pomonella*, and its Implications for Fitness Trade-offs in Phytophagous Insects. In: Simpson, S.J., Mordue, A.J., & Hardie, J. (Eds.), *Proceedings of the 10th International Symposium on Insect-Plant Relationships.* Dordrecht: Springer, Series Entomologica, Volume 56.

Felsenstein, J.
1989. PHYLIP-Phylogeny Inference Package (version 3.2). *Cladistics* 5: 164-166.

Fjeldså, J.
2003. Patterns of Endemism in African Birds: How much does Taxonomy Matter? *Ostrich* 74: 30-38.

Frachetti, M.D.
2012. Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia. *Current Anthropology* 53(1): 2-38.

Francfort, H.-P.
1981. About the Shortughai sequence from Mature Harappan to Late Bactrian in Eastern Bactria (N.E. Afghanistan). *Puratattva* 10: 91-94.

Francfort, H.-P.
1984. The early periods of Shortughai (Harappan) and the western Bactrian culture of Dashly. In Allchin, B.A. (Ed.), *South Asian Archaeology 1981* (pp. 170-175). Cambridge: Cambridge University Press.

Francfort, H.-P.
2016. How the twins met: Indus and Oxus Bronze Age civilizations in Eastern Bactria. Shortughaï forty years later. In Dubova, N., Antonova, E.V., Kozhin, P.M., Kosarev, M.F., Muradov, R.G., Sataev,, R.M. & Tishkin, A.A. (Eds.), *Transactions of Margiana Expedition,*

Volume 6 (pp. 461-476). Moscow: N.N. Miklukho-Maklay Institute of Ethnology and Anthropology of Russian Academy of Sciences.

Freeman, M.F., & Tukey, J.W.
1950. Transformations Related to the Angular and Square Root. *Annals of Mathematical Statistics* 21: 607-611.

Freidel, D.A.
1986. Maya Warfare: An Example of Peer Polity Interaction. In Renfrew, C. & Cherry, J.F. (Eds.), *Peer Polity Interaction and Socio-Political Change* (pp. 25-35). Cambridge: Cambridge University Press.

Furnivall, J.S.
1944. *Netherlands India—A Study of Plural Economy*. Cambridge: Cambridge University Press.

Gadgil, M., Joshi, N.V., Prasad, U.V.S., Manharan, S., & Patil, S.
1997. Peopling of India. In Balasubramanian, D. & Rao, N.N. (Eds.), *The Indian Human Heritage* (pp. 100-129). Hyderabad: University Press.

Gangal, K., Sarson, G.R., & Shukurov, A.
2014. The Near-Eastern roots of the Neolithic in South Asia. *PLoS One* 9(5): e95714.

Ghirshman, R.
1954. *Iran: From the Earliest Times to the Islamic Conquest*. Harmondswoth: Penguin.

Giglio, E.M., & Dyer, K.A.
2013. Divergence of Premating Behaviors in the Closely Related Species *Drosophila subquinaria* and *D. recens*. *Ecology and Evolution* 3(2): 365-374.

Gillison, A.N.
1978. Minimum Spanning Ordination—A Graphic-Analytical Technique for Three-dimensional Ordination Display. *Australian Journal of Ecology* 3: 233-238.

Gilmartin, D.
1994. *Biraderi* and bureaucracy: the politics of Muslim kinship solidarity in 20th century Punjab. *International Journal of Punjab Studies* 1(1): 1-29.

Gough, E.
1959. The Nayars and the Definition of Marriage. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 89: 23-34.

Gray, R.D., Atkinson, Q.D., & Greenhill, S.J.
2011. Language Evolution and Human History: What a Difference a Date Makes. *Philosophical Transactions of the Royal Society of London, Series B Biological Sciences* 366 (1567): 1090-1100.

Green, R.F., & Suchey, J.M.
1976. The use of Inverse Sine Transformation in the Analysis of Non-metrical Data. *American Journal of Physical Anthropology* 45: 61-68.

Green, R.F., Suchey, J.M., & Gokhale, D.V.
1979. The Statistical Treatment of Correlated Bilateral Traits in the Analysis of Cranial Material. *American Journal of Physical Anthropology* 50: 629-634.

Grover, S.
2009. Lived Experiences: Marriage, Notions of Love, and Kinship Support amongst Poor Women in Delhi. *Contributions to Indian Sociology* 43(1): 1-33.

Gupta, S.P.
1979. *Archaeology of Soviet Central Asia, and the Indian Borderlands*. Delhi: B.R. Publishing Corporation.

Haber, M., Platt, D.E., Bonab, M.A., Youohanna, S.C., Soria-Hernanz, D.F., Martinez-Cruz, B.,… Zalloua PA.
2012. Afghanistan's ethnic groups share a Y-chromosomal heritage structured by historical events. *PLoS ONE* 7(3): e34288.

Harris, D.R.
1997a. First Farmers were Colonists After All. *British Archaeology* 27: 8-9

Harris, D.R.
1997b. Jeitun and the Transition to Agriculture in Central Asia. *Archaeology International* 1: 28-31.

Harris, E.F.
2008. Statistical Applications in Dental Anthropology. In J.D. Irish & G.C. Nelson (Eds.), *Technique*

and application in dental anthropology (pp. 35-67). Cambridge: Cambridge University Press.

Harris, E.F., & Sjovøld, T.
2004. Calculation of Smith's Mean Measure of Divergence for Intergroup Comparisons using Nonmetric Data. *Dental Anthropology* 17: 83-93.

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

Hemphill, B.E.
2010. Dental Anthropology of the Madaklasht I: A Description and Analysis of Variation in Morphological Features of the Permanent Tooth Crown. *Pakistan Heritage* 2: 1-36.

Hemphill, B.E.
2012. The Awans of northern Pakistan: emigrants from Central Asia, Arabs from western Afghanistan, or colonists from peninsular India? A dental morphometric investigation. *American Journal of Physical Anthropology* 147 (S54): 163.

Hemphill, B.E.
2013. Grades, gradients, and geography: a dental morphometric approach to the population history of South Asia. In Scott, G.R. & Irish, J.D. (Eds.), *Anthropological perspectives on tooth morphology: genetics, evolution, variation* (pp. 341-387). Cambridge: Cambridge University Press.

Hemphill, B.E., Inamullah, Tariq, M., Zubair, M., Saadiq, N., & Ahmad, H.
2018. Is There Evidence of a Medieval Intrusion of Pathans into the Vale of Peshawar, Pakistan? A Dental Morphology Approach. *American Journal of Physical Anthropology* 165(S66): 116.

Hemphill, B.E., Pojar, T., & Langstieh, B.T.
2019. An Initial Examination of Biological Affinities of Northeastern Indian populations Relative to other South Asians. A Dental Morphology Investigation. *American Journal of Physical Anthropology* 168(S68): 100-101.

IBM Corporation.
2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corporation.

Inamullah, Ahmad, H., & Hemphill, B.E.
2017. Historic era Immigrants to Northern Pakistan? A Dental Morphology Investigation of Pathans, Gujars, and Kohistanis. *American Journal of Physical Anthropology* 162(S64): 213-214.

Irish, J.D.
2010. The Mean Measure of Divergence: Its Utility in Model-free and Model-bound Analyses relative to the Mahalanobis D2 Distance for Nonmetric Data. *American Journal of Human Biology* 22: 378-395.

Islam, A.
2017. The Civilizational Role of Islam in the Indian Subcontinent: The Delhi Sultanate. *Intellectual Discourse* 25(1): 101-134.

Jarrige, J.-F.
2006. Mehrgarh Neolithic. *Pragdhara* 18:135-154.

Jarrige, J.-F., & Hassan, M.U.
1989. Funerary complexes in Baluchistan at the end of the third millennium in light of recent discoveries at Mehrgarh and Quetta. In K. Freifelt K & P. Sørensen P (Eds.), *South Asian archaeology 1985* (pp. 150-166). London: Curzon Press.

Jarrige, J.-F., & Lechevallier, M.
1980. Excavations at Mehrgarh, Baluchistan: their significance in the prehistorical context of the Indo-Pakistan borderlands. In: Taddei, M. (Ed.), *South Asian archaeology 1977* (pp. 463-536). Naples: Istituto Universitario Orientale.

Jernvall, J., & Jung, H.S.
2000. Genotype, Phenotype and Developmental Biology of Molar Tooth Characters. *Yearbook of Physical Anthropology* 43: 171-190.

Jernvall, J., & Thesleff, I.
2000. Reiterative Signaling and Patterning during Mammalian Tooth Morphogenesis. *Mechanisms of Development* 92: 19-29.

Kangas, A.T., Evans, A.R., Thesleff, I. & Jernvall, J.
2004. Nonindependence of Mammalian Dental Characters. *Nature* 432: 211-214.

Keene, H.J.
1989. The Morphogenetic Triangle: A New Conceptual Tool for Application to Problems

in Dental Morphogenesis. *American Journal of Physical Anthropology* 59: 281-287.

Kennedy, K.A.R., Chiment, J., Distotell, T., Meyers, D.
1984. Principal-Components Analysis of Prehistoric South Asian Crania. *American Journal of Physical Anthropology* 64: 105-118.

Kenoyer, J.M.
1998. *Ancient Cities of the Indus Valley Civilization*. Karachi: AIPS Oxford.

Kenoyer, J.M.
2005. Culture Change during the Late Harappan Period at Harappa: New Insights on Vedic Aryan Issues. In: Bryant, E.F., Patton, L.L. (Eds.), *The Indo-Aryan Controversy: Evidence and Inference in Indian History* (pp. 21-49). London: Routledge.

Kohl, P.L.
1992. Central Asia (Western Turkestan): Neolithic to the Early Iron Age. In: Erich, R.W. (Ed.), *Chronologies in Old World Archaeology, Volume 1* (pp. 179-196). Chicago: University of Chicago Press.

Kohl, P.L. & Lyonnet, B.
2008. By land and by sea. The circulation of materials and peoples, *ca* 3500-1800 B.C. In Oldham, E. & Spoor, R.H. (Eds.), *Intercultural Relations between South and Southwest Asia* (pp. 29-42). Oxford: BAR International Series No. 1826.

Kroeber, A.L.
1939. Cultural and natural areas of native North America. Berkeley: University of California Press.

Kroeber, A.L.
1947. Culture groupings in Asia. *Southwestern Journal of Anthropology* 3: 322-330.

Kruskal, J.B.
1964a. Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika* 29: 1-27.

Kruskal, J.B.
1964b. Nonmetric Multidimensional Scaling: A Numerical Method. *Psychometrika* 29: 115-129.

Kruskal, J.B. & Wish, M.
1978. *Multidimensional Scaling*. Newbury Park: SAGE

Kuzmina, E.E.
2001. The first migration wave of Indo-Iranians to the south. *Journal of Indo-European Studies* 29(1-2): 1-40.

Lacau, H., Gayden, T., Regueiro, M., Chennakrishnaiah, S., Bukhari, A., Underhill, P.A.,…Herrera, R.J.
2012. Afghanistan from a Y-chromosome perspective. *European Journal of Human Evolution*, 20, 1063-1070. http://doi:10.1038/ejhg.2012.59

Lal, M.
1846. *Life of Amir Dost Mohammad Khan of Kabul*. New York: Crabtree Publishing Company.

Law, R.W.
2011. *Inter-Regional Interactions and Urbanism in the Ancient Indus Valley*. Kyoto: Research Institute for Humanity and Nature, Occasional Paper No. 11.

Line, S.R.P.
2001. Molecular Morphogenetic Fields in the Development of the Human Dentition. *J. Theot. Biol.* 211: 67-75.

Lukacs, J.R.
1983. Dental anthropology and the origins of two Iron Age populations from northern Pakistan. *Homo* 34: 1-15.

Lyonnet, B.
2005. Another possible interpretation of the Bactro-Margiana Culture (BMAC) of Central Asia. In Jarrige, C. & Lefévre, V. (Eds.), *South Asian Archaeology 2001* (pp. 191-200). Paris: CNRS.

Masson, V.M.
1988. *Altyn-Depe*. Philadelphia: The University Museum, University of Pennsylvania.

Masson, V.M., Sarianidi, V.I.
1972. *Central Asia: Turkmenia Before the Achaemenids*. New York: Praeger Publishers.

Mayr, E.
1942. *Systematics and the Origins of Species*. New York: Columbia University Press.

Mayr, E.
1963. *Animal Species and Evolution*. Cambridge, MA: Belknap Press.

McCune, B. & Grace, J.B.
2002. *Analysis of Ecological Communities*. Gleneden Beach, OR: MJM Software Design.

McCune, B. & Mefford, M.J.
1999. *PC-ORD. Multivariate Analysis of Ecological Data*. Gleneden Beach, OR: MJM Software Design

McGovern, W.
1939. *The Early Empires of Central Asia.* Chapel Hill, NC: University of North Carolina Press.

McMahon, A.H.
1901. *Tribes of Dir, Swat and Bajaur*. London: Witherby.

Miller, R.J.
1953. Areas and institutions in Eastern Asia *Southwestern Journal of Anthropology* 9: 203-211.

Milner Jr., M.
1988. Status relations in South Asian marriage alliances: toward a general theory. *Contributions to Indian Sociology* 22: 145–169.

Mines, M.
1972. Muslim social stratification in India: the basis for variation. *Southwestern Journal of Anthropology* 28(4): 333-349.

Mitsiadis T.A. & Smith, M.M.
2006. How do Genes make Teeth to Order through Development? *Journal of Experimental Zoology (Mol Dev Evol)* 306BL: 177-182.

Momin, A.R.
1977. The Indo-Islamic Tradition. *Sociological Bulletin* 26(2): 242-258.

Morgenstiern, G.
1982. Afghanistan vi. Pasto. *Encyclopedia Iranica.* New York: Columbia University, Center for Iranian Studies.

Mysara, M., Vandamme, P., Props, R., Kerckhof, F.-M., Leys, N., Boon, N…Monsiuers, P.
2017. Reconciliation between Operational Taxonomic Units and Species Boundaries. *FEMS Microbiology* 93(4): fix029.

Narasimhan, V.M., Patterson, N., Moorjani, P., Lazaridis, I., Lipson, M., Mallick, S.,…Reich, D.
2019. The genomic formation of South and Central Asia. *Science*, 365(6457), eaat7487.

Nasim Khan, M.
2001. Exploration and Excavation of the Earliest Śivaite Monastic Establishment at Kashmir Smast (A Preliminary Report). *Ancient Pakistan* 218-309.

Osborn, J.W.
1978. Morphogenetic Gradients: Fields versus Clones. In Butler, P.M., & Joysey, K.A. (Eds.), *Development, Function and Evolution of Teeth* (pp. 171-201). London: Academic Press.

Parpola, A.
1988. The coming of the Aryans to Iran and India and the cultural and ethnic identity of the Dasas. *Studia Orientalia* 64: 195-302.

Pigliucci, M.
2003. Species as Family Resemblance Concepts: The (Dis-)solution of the Species Problem? *BioEssays* 25(6): 596-602

Pocock, D.F.
1954 [1994]. The Hypergamy of the Patidars. In: Uberoi, P. (Ed.), *Family, Kinship and Marriage in India* (pp. 330-340). Oxford: Oxford University Press.

Pojar, T., Langstieh, B.T. & Hemphill, B.E.
2022. An Initial Investigation of Dental Morphology among Three Southern Naga Ethnic Groups of Northeast India. *American Journal of Biological Anthropology* 179(2): 184-210.

Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A.,…Mehdi, S.Q.
2002. Y-chromosomal DNA variation in Pakistan. *American Journal of Human Genetics* 70: 1107-1124.

Rahman R.U. & Kamran, U.
2021. Historical developments of Pashto and Sogdian phonologies from the proto-Iranian language. *Journal of Asian Civilizations* 44(1): 111-137.

Rathmann, H. & Reyes-Centeno, H.
2020. Testing the Utility of Dental Morphological Trait Combinations for Inferring Human Neutral Genetic Variation. *Proceedings of the National Academy of Sciences USA*. 117(20): 10769-10777.

Raza S.J.
2010-2011. Hindus under the Ghaznavids. *Proceedings of the Indian History Congress* 71: 213-225.

Renfrew, C.
1987. *Archaeology and Language*. Cambridge: Cambridge University Press.

Reyment, R.A.
1991. *Multidimensional Palaeobiology*. Oxford: Pergamon Press.

Rose, H.A.
1914 [1990]. *A glossary of the tribes and castes of the Punjab and North-West Frontier Province*. 3 vols. New Delhi: Asian Educational Services.

Ryan, M.J., & Causey, B.A.
1989. "Alternative" Mating Behavior in the Swordtails *Xiphophorus nigrensis* and *Xipgophorus pygmaeus* ((Pices: Poeciliidae). *Behavioral Ecology and Sociobiology* 24: 341-348.

Sahoo, S., Singh, A., Himabindu, G., Banerjee, J., Sitalaximi, T, Gaikwad, S…Kashyap, V.K.
2006. A Prehistory of Indian Y Chromosomes: Evaluating Demic Diffusion Scenarios. *Proceedings of the National Academy of Sciences USA* 103(4): 843-848.

Saitou, N., & Nei, M.
1987. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution* 4: 406-425.

Salazar-Ciudad, I., Jernvall, J., & Newman, S.A.
2003. Mechanisms of Pattern Formation in Development and Evolution. *Development* 130: 2027-2037.

Samad, A.
2011. Emergence of Hinduism in Gandhara: An Analysis of Material Culture. Unpublished Ph.D. dissertation. Berlin: Freie Universität.

Samad, R.
2011. *The Grandeur of Gandhara: The Ancient Buddhist Civilization of the Swat, Peshawar, Kabul and Indus Valleys. New York: Algora.*

Schmidt, T.S.B., Rodrigues, J.F., and von Mering, C.
2015. Limits to Robustness and Reproducibility in the Demarcation of Operational Taxonomic Units. *Environmental Microbiology* 17(5): 1689-1706.

Scott, D.A.
1985. Ashokan Missionary Expansion of Buddhism among the Greeks (in N.W. India, Bactra and the Levant). *Religion* 15(2): 131-141.

Scott, G.R.
1977. Classification, Sex Dimorphism, Association, and Population Variation of the Canine Distal Accessory Ridge. *Human Biology* 49(3): 453-469.

Scott, G.R.
1980. Population Variation of Carabelli's Trait. *Human Biology* 52(1): 63-78.

Scott, G.R., Maier, C., & Heim, K.
2016. Identifying and Recording Key Morphological (Nonmetric) Crown and Root Traits. In Irish, J.D. & Scott, G.R. (Eds.), *A Companion to Dental Anthropology* (pp. 247-264). Malden, MA: Wiley-Blackwell.

Scott, G.R., & Turner II, C.G.
1997. *The Anthropology of Modern Human Teeth: Dental Morphology and its Variation in Recent Human Populations*. Cambridge: Cambridge University Press.

Sengupta, S., Zhivotovsky, L.A., King, R., Mehdi, S.Q., Edmonds, C.A., Chow, C.-E.T…Underhill, P.A.
2006. Polarity and Temporality of High-resolution Y-chromosome Distributions in India Identify both Indigenous and Exogenous Population Expansions and Reveal Minor Genetic Influence of Central Asian Pastoralists. *American Journal of Human Genetics* 78: 202-221.

Setoguchi, S., Takamori, H., Aotsuka, T., Sese, J., Ishikawa, Y., & Matsuo, T.
2014. Sexual Dimorphism and Courtship Behavior

in *Drosophila prolongata*. *Journal of Ethology* 32: 91-102.

Sharpe, P.T.
1995. Homeobox Genes and Orofacial Development. *Connect Tissue Research,* 32: 17-25.

Sheirer, J., Ray, W.S., & Hare, N.
1976. The Analysis of Ranked Data Derived from Completely Randomized Factorial Designs. *Biometrics* 32(2): 429-434

Sjovøld, T.
1977. Non-metrical Divergence between Skeletal Populations: The Theoretical Foundation and Biological Importance of C.A.B. Smith's Mean Measure of Divergence. *Ossa*, 4 (suppl.): 1-133.

Skutnabb-Kangas, T. & Harmon, D.
2018. Biological Diversity and Language Diversity: Parallels and Differences. In: Penz, H. & Fill, A. (Eds.), *Handbook of Ecolinguistics* (pp. 11-25). New York: Routledge.

Slatkin, M. & Maddison, W.P.
1990. Detecting Isolation by Distance using Phylogenies of Genes. *Genetics* 126: 249-260.

Smit, D.J.A., de Geus, E.J.C., Boersma, M., Boomsma, D.I., & Stam, C.J.
2016. Life-span development of brain network integration assessed with phase lag index connectivity and minimum spanning tree graphs. *Brain Connectivity* 6(4): 312-325.

Smith, D.C.
1988. Heritable Divergence of *Rhagoletis pomonella* Host Races by Seasonal Asynchrony. *Nature* 336: 66-67.

Smith, V.
1924 [2022]. *Early History of India from 600 B.C. to the Mohammadan Conquest*. Charleston, SC : Legare Street Press .

Sneath, P.H.A.
1976. Phenetic Taxonomy at the Species Level and Above. *Taxon* 25: 437-450.

Sokol, R.R. & Sneath, P.H.A.
1963. *Principles of Numerical Taxonomy*. San Francisco: W.H. Freeman and Company.

SYSTAT Software, Inc.
2004. SYSTAT 11. Richmond, CA: SYSTAT Software, Inc.

Tariq, M., Ahmad, H., & Hemphill, B.E.
2017. Are the Socially Recognized Social Groups of Northern Pakistan Meaningful Biological Entities for Reconstruction of Population Histories? A Dental Morphology Investigation. *American Journal of Physical Anthropology* 162(S64): 377.

Tariq, M., Ahmad, H., Hemphill, B.E., Farooq, U., & Schurr, T.G.
2022. Contrasting Maternal and Paternal Genetic Histories among Five Ethnic Groups from Khyber Pakhtunkhwa, Pakistan. *Scientific Reports* 12: 1027.

Tilley, C.
1984. Ideology and Legitimization of Power in the Middle Neolithic of Southern Sweden. In Miller, D. & Tilley, C. (Eds.), *Ideology, Power and Prehistory* (pp. 111-146). Cambridge: Cambridge University Press.

Townsend, G.C., & Brown, T.
1981. Morphogenetic Fields within the Dentition. *Australian Dental Journal* 7(1): 3-13.

Townsend, G.C., Harris, E.F., Lesot, H., Clauss, F., & Brook, A.H.
2009. Morphogenetic Fields within the Human Dentition: A New, Clinically Relevant Synthesis of an Old Concept. *Archives of Oral Biology* 54: S34-S44.

Tucker, A.S., & Sharpe, P.T.
2004. The Cutting-edge of Mammalian Development: How the Embryo makes Teeth. *Nature Review of Genetics* 5: 499-508.

Turner II, C.G., 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.G. (Eds.), *Recent advances in dental anthropology* (pp. 13-32). New York: Alan R. Liss.

Ullah, I., Olofsson, J.K., Margaryan, A., Ilardo, M., Ahmad, H., Sikora, M…Allentoft, M.E.
2017. High Y-chromosomal Differentiation among

Ethnic Groups of Dir and Swat Districts, Pakistan. *Annals of Human Biology* 81: 234-248.

Vinogradova, N.
1993. Interrelation between farming and 'steppe' tribes in the Bronze Age South Tadjikistan. In Gail, A.J. & Mevissen, G.J.R. (Eds.), *South Asian Archaeology 1991* (pp. 289-301). Stuttgart: Franz Steiner Verlag.

Vinogradova, N.
1994. The farming settlement of Kangurttut (South Tadjikistan) in the Late Bronze Age. *Archaeologische Mitteilungen aus Iran* 27: 29-47.

Vogelsang, W.
1988. A Period of Acculturation in Ancient Gandhara. *South Asian Studies* 4(1): 103-113.

Vogelsang, W.
2002. *The Afghans*. New York: Wiley-Blackwell.

Walker, P.L. & Hewlett, B.S.
1990. Dental Health, Diet and Social Status among Central African Foragers and Farmers. *American Anthropologist* 92(2): 383-398.

Watson, H.D.
1908. *Gazetteer of the Hazara District, 1907*. London: Chatto & Windus

Weiss, K.M.
1990. Duplication with Variation: Metameric Logic in Evolution from Genes to Morphology. *Yearbook of Physical Anthropology* 33: 1-24.

Wissler C.
1927. The culture area concept in social anthropology. *American Journal of Sociology* 32(6): 881-891.

Wright, J.P.
1994. Can there be a melting pot in Pakistan? Interprovincial marriage and national integration. *Contemporary South Asia* 3(2): 131-144.

Wright, S.
1943. Isolation by Distance. *Genetics* 28: 114-138.

Xu, Z., Liu, J., An, G. & Jia, X.
2016. The Impact of Feature Selection on Defect Prediction Performance: An Empirical Comparison. In: *27th IEEE International Symposium on Software Reliability Engineering* (pp. 309-320). ISRE: Ottawa.

Zachos, F.E.
2018. (New) Species Concepts, Species Delimitation and the Inherent Limitations of Taxonomy. *Journal of Genetics* 97: 811-815.