

West Eurasian mtDNA lineages in India: an insight into the spread of the Dravidian language and the origins of the caste system

Malliya Gounder Palanichamy¹ · Bikash Mitra^{1,3} · Cai-Ling Zhang¹ ·
Monojit Debnath^{2,5} · Gui-Mei Li¹ · Hua-Wei Wang¹ · Suraksha Agrawal⁴ ·
Tapas Kumar Chaudhuri³ · Ya-Ping Zhang^{1,2}

Received: 8 January 2015 / Accepted: 25 March 2015 / Published online: 2 April 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract There is no indication from the previous mtDNA studies that west Eurasian-specific subclades have evolved within India and played a role in the spread of languages and the origins of the caste system. To address these issues, we have screened 14,198 individuals (4208 from this study) and analyzed 112 mitogenomes (41 new sequences) to trace west Eurasian maternal ancestry. This has led to the identification of two autochthonous sub-haplogroups—HV14a1 and U1a1a4, which are likely to have originated in the Dravidian-speaking populations approximately 10.5–17.9 thousand years ago (kya). The carriers of these maternal lineages might have settled in

South India during the time of the spread of the Dravidian language. In addition to this, we have identified several subsets of autochthonous U7 lineages, including U7a1, U7a2b, U7a3, U7a6, U7a7, and U7c, which seem to have originated particularly in the higher-ranked caste populations in relatively recent times (2.6–8.0 kya with an average of 5.7 kya). These lineages have provided crucial clues to the differentiation of the caste system that has occurred during the recent past and possibly, this might have been influenced by the Indo-Aryan migration. The remaining west Eurasian lineages observed in the higher-ranked caste groups, like the Brahmins, were found to cluster with populations who possibly arrived from west Asia during more recent times.

Electronic supplementary material The online version of this article (doi:[10.1007/s00439-015-1547-4](https://doi.org/10.1007/s00439-015-1547-4)) contains supplementary material, which is available to authorized users.

✉ Malliya Gounder Palanichamy
empalani@yahoo.com

✉ Ya-Ping Zhang
zhangyp1@263.net.cn

¹ Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming 650 091, Yunnan, China

² State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

³ Cellular Immunology Laboratory, Department of Zoology, University of North Bengal, Siliguri 734013, West Bengal, India

⁴ Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow 226014, India

⁵ Present Address: Department of Human Genetics, National Institute of Mental Health and Neurosciences, Bangalore, India

Introduction

In the last two decades, the distribution and extent of variation of west Eurasian mitochondrial DNA (mtDNA) haplogroups in Indian populations have been evaluated quite extensively (Bamshad et al. 1998; Kivisild et al. 1999a, 2003a, b; Basu et al. 2003; Metspalu et al. 2004; Palanichamy et al. 2004; Quintana-Murci et al. 2004). The commonly found west Eurasian-specific mtDNA lineages in India come from haplogroups H, V, HV, R0, J, T, N1, W, X, and subclusters U1, U2e, U3, U4, U5, U7, U9, and K of haplogroup U. Interestingly, the overall proportion of west Eurasian haplogroups in India is found to be eightfold lower than in Europe (10 versus 85 %) (Kivisild et al. 1999a, 2003b; Quintana-Murci et al. 2004). Within India, the frequencies also vary widely in the various regions, for example, 40 % in the Punjab, 16 % in southern India and 7 % in eastern India (Metspalu et al. 2004).

The spread of west Eurasian haplogroups to the Indian mtDNA pool is discussed in only a few studies and it is believed to be linked to the spread of agriculture to India (Kivisild et al. 1999a, b, 2000). West Eurasian haplogroups are found across India and the data obtained by Kivisild et al. (1999a) suggested that these haplogroups may have been spread by the early Neolithic migrations of proto-Dravidian farmers spreading from the eastern horn of the Fertile Crescent into India. However, other studies argue in favor of an Indo-Aryan invasion and associated with the introduction of the caste system to India (Bamshad et al. 2001). Furthermore, the frequency of these haplogroups was found to be specifically higher among the upper caste groups of India (Bamshad et al. 1998, 2001; Basu et al. 2003). Contradictorily, it has also been suggested by many that the process of social stratification was established long before the Indo-Aryan invasion (Chaubey et al. 2007; Tamang and Thangaraj 2012 and references therein). Overall, there is no consensus on the spread of west Eurasian lineages being linked to the spread of agriculture, the proto-Elamo-Dravidian language, and the Indo-Aryan migration. Moreover, the interpretation of prehistoric (Indo-Aryan migration) events as well as phylogenetic network-based time estimation (calculated from the mtDNA hypervariable segment I region sequence variation) related to the origin and spread of west Eurasian haplogroups in India is still controversial.

To get a deeper insight into the spread of the west Eurasian haplogroups to India and to resolve the conflicting views, a comprehensive phylogenetic understanding of west Eurasian haplogroups including the subclades that show autochthonous development in India is essential. Although a previous study examined the complete mtDNA variation of west Eurasian haplogroups phylogeny in India (Palanichamy et al. 2004), the putative source population of the west Eurasian lineages in the extant Indian mtDNA pool was not suggested.

In this study, we have assessed the phylogeographic distribution of the west Eurasian lineages in Indian populations and classified them as region-specific subhaplogroups through complete mtDNA sequencing. We generated a phylogenetic tree based on maximum parsimony of 112 complete mitogenomes (including 41 new sequences) belonging to west Eurasian haplogroups from India. Our results were compared with the published data from geographically and ethnically targeted groups from the Middle East and southwest Asia. Furthermore, we have assessed variations and divergence times of the Indian-specific west Eurasian subclades. Finally, we reevaluated various hypotheses that have been proposed on the spread of the west Eurasian haplogroups in India.

Materials and methods

Sampling and preliminary screening of west Eurasian haplogroups

To identify the west Eurasian maternal ancestry, we examined 4208 unrelated individuals from the Indian subcontinent. All the DNA samples analyzed in this study were derived from the blood samples collected during 2001–2004. The authors obtained written informed consent from the volunteers or community leaders after explaining the collection procedures and purpose of the study in local languages. The institutional Ethical Committees of North Bengal University, Sanjay Gandhi Institute of Medical Sciences and the Yunnan University approved the protocol and ethical clearance of the study. The hypervariable segment of the mtDNA control region (HVS-I and HVS-II) was amplified and sequenced from nucleotide position 16,001 to 16,497 (HVS-I) and from 29 to 408 (HVS-II) of all the studied individuals. We identified 422 mtDNAs (42 sequences reproduced from our previous study) representing west Eurasian haplogroups based on HVS-I and HVS-II variable sites and some diagnostic coding region variants. In addition, 757 mtDNAs belonging to different west Eurasian haplogroups were identified from published Indian mtDNA dataset comprising 9990 sequences (see Supporting Information Table S1 and S2).

Complete mtDNA sequencing

To deconstruct the clustering of the west Eurasian mtDNA lineage into Indian-specific subfounders, 41 west Eurasian lineages were completely sequenced. We obtained the complete mtDNA sequences by following the methods described previously (Palanichamy et al. 2004). Sequencing was performed using 3730 Genetic Analyzer (Applied Biosystems), and the sequences were checked using SeqScape (v. 2.5-Applied Biosystems) and DNASTAR software (DNASTAR, Inc., Madison, USA). Mutations were scored by comparing the sequences with the revised Cambridge reference sequence (rCRS) (Andrews et al. 1999). The complete mtDNA sequences reported in this paper have been deposited in GenBank under accession numbers GU213243–GU213254, GU327373, KP763831–KP763851 (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>).

Phylogenetic analysis and molecular dating

The phylogenetic relationships of the complete mtDNA sequences were established using the reduced median network algorithm (<http://www.fluxus-engineering.com>) and the tree was checked manually to resolve homoplasies. To

differentiate the Indian-specific subclades, we assembled the published mitochondrial genomes from west Eurasia and compared them to the Indian mtDNA sequences. The coalescent times for the Indian-specific west Eurasian subclades were estimated with rho-statistics (ρ) (mean divergence inferred from ancestral haplotype) (Forster et al. 2001), and standard errors (σ) were calculated following the method of Saillard et al. (2000). The calculator provided by Soares et al. (2009) was used to convert the ρ -statistics and its error ranges to age estimates with 95 % confidence intervals.

Results

The distributions of the west Eurasian haplogroups in the Indian populations are shown in supplementary table S3. About 16 % of west Eurasian mtDNAs in India belong to haplogroup H and its subclades of H1, H2, H3, H5, H6, H7, H9, H13, H14, H15, and H103, of which haplogroups H2, H9, and H13 together with unclassified H* account for 12.6 % of H haplogroup variation in India. The H haplogroup is comparatively more frequent in the north (5.2 %) and the south (7.6 %) than in the west (1.7 %) and east (1.5 %) Indian populations. The haplogroups HV, JT, U (xU2a, U2b, and U2c), and W lineages have frequencies about 12.6, 12.7, 38.6, and 8.7 %, respectively, in the west Eurasian control region database of Indian populations (Supplementary table S3).

The phylogeny of west Eurasian mitogenomes in India

We have sequenced two H2 haplogroup complete mtDNA genomes, one each from the Tamil Nadu and Uttar Pradesh populations, and compared them with the published complete sequences. It was found that these two lineages belong to subclades H2a1 and H2b. A previous study has reported H2a1 lineages only in south Indian populations, whereas its derived single H2a1a haplogroup has been reported in north Indian populations (Metspalu et al. 2004). About half of the H2b haplogroups were found in higher-caste Brahmins and Indian Muslims. Our whole-genome H2a1 and H2b sequence analyses suggest that these haplogroups possibly share common ancestry with Pakistani, Iraqi, Armenian, and Siberian populations (Achilli et al. 2004; Hartmann et al. 2009; Derenko et al. 2013, 2014). We also observed four H3g sequences—two from south Indian (Tamil Nadu) and two from north Indian (Uttar Pradesh) populations. One H3g mitogenome from a north Indian individual was sequenced completely and it was found that this sequence clusters phylogenetically with the European Caucasian samples (Coble et al. 2004) (supplementary fig. 1).

The H5a1 lineages of H5 subclade have been observed mostly in the higher-caste Brahmin populations of northern India (Punjab and Uttar Pradesh). However, a single mtDNA variation that belongs to H5a1 haplogroup was reported previously in a Brahmin population of Maharashtra (Gaikwad and Kashyap 2005). Complete mtDNA genome sequence of the H5a1 lineage of north Indian Brahmins shows clustering with the Pakistan, Caucasus, Iran, Italy, and central Asian groups (Mishmar et al. 2003; Derenko et al. 2013). In addition to this, three H6a1a haplotypes were reported from south Indian (Karnataka) Muslims and Lingayat populations (Rajkumar and Kashyap 2003). In this study, two H7b lineages were detected in Brahmin population from north India (Uttar Pradesh). Interestingly, comparison of these complete genome sequences with the published data shows that Indian H7b sequences share an ancestry with Pakistani and Siberian populations (Derenko et al. 2014). The lineages of the H9 haplogroup was found to be more common in Brahmin individuals from the north (Uttar Pradesh) and the west (Maharashtra) India (Gaikwad and Kashyap 2005); they were also reported in other caste groups from West Bengal and Gujarat (Metspalu et al. 2004). Complete genome sequencing of H9 lineages has led to the identification of a distinct subclade, named herein as H9b (supplementary fig. 1).

The H13 haplogroup carrying individuals from India (Uttar Pradesh, Tamil Nadu, and Andhra Pradesh) represent mainly the H13a1a and H13a2a subclades. Amongst these, the H13a1a lineage has predominantly been found only in the Uttar Pradesh Brahmin (Bhargava and Chaturvedi) populations, which is likely to be the result of founder effects. The Indian H13a1a lineage is shown to share an ancestry with the populations from Europe, the Caucasus and the Near East (Achilli et al. 2004; Coble et al. 2004), whereas the ancestry of H13a2a is shared exclusively by the Pakistani and the Near Eastern (Iran and Iraq) populations (Derenko et al. 2013) (Fig. 1a). The haplogroup H14 is uncommon in Europe but found mostly in the Near East and the Caucasus (Richards et al. 2000; Nasidze and Stoneking 2001; Roostalu et al. 2007; Al-Zahery et al. 2011) and so far only two subclades—H14a and H14b have been identified. However, based on new complete sequence analysis we identified an additional subclade, H14c, which has not been reported earlier (Fig. 1a). H14a and H14c lineages were found in Punjab, Uttar Pradesh, West Bengal, and Andhra Pradesh, whereas a single H14b2 haplotype was reported from Pakistan (GenBank KJ446337). H15 is a very rare subclade, distributed from Western Europe to the Middle East and from Central Asia to India (Derenko et al. 2013). H15a1a1 haplotype was observed only in a Chaturvedi Brahmin population from north India (Uttar Pradesh), while H15a1b haplotype was found in a Parsi population

Armenia, and Turkey (Derenko et al. 2013). The subclades of HV12—HV12a and HV12b have been reported in the Near Eastern populations (Nasidze et al. 2004; Al-Zahery et al. 2011; Terreros et al. 2011). Whereas, in India only the lineages of subclade HV12b such as subgroups HV12b1 and HV12b1a were observed. Of these, subhaplogroup HV12b1a shares a deep ancestry with the Near Eastern (Iran and Armenia) populations (Schönberg et al. 2011; Derenko et al. 2013, 2014). The moderately high frequency (5 %) of HV14 lineages was found only in samples from south India (Tamil Nadu, Karnataka, and Andhra Pradesh) and adjoining areas of Sri Lanka (Ranaweera et al. 2014), however, these are either absent or negligibly present in other places like Maharashtra, Odisha, West Bengal and Uttar Pradesh (supplementary table S1). We completely sequenced a single Indian HV14 haplotype and compared it with our previously published south Indian sequences, and it helped us to define a new Indian-specific subclade, HV14a1 (Palanichamy et al. 2004). Nevertheless, a member of haplogroup HV14 was recently reported in Persian population from Iran (Derenko et al. 2013) (Fig. 1a).

The near eastern origin R0 haplogroup-derived R0a and R0a2 lineages were proposed based on their presence in the Arabian Peninsula, Iran, Iraq, Kuwait and Pakistan (Quintana-Murci et al. 2004; Abu-Amero et al. 2008; Rakha et al. 2011; Al-Zahery et al. 2011; Scheible et al. 2011), whereas in India, only R0a2 has been found mostly in Muslim communities from Gujarat and Tamil Nadu, and generally absent or rare in non-Muslim Indian populations. Surprisingly, we have detected the R0a2 lineages in Brahmin populations from northern India—Uttar Pradesh (supplementary table S1) which clustered phylogenetically with Iranian and Somalian samples (Cerný et al. 2011; Derenko et al. 2013). We detected eight lineages of haplogroup V in north Indian (Uttar Pradesh) Brahmin population and all were found to carry similar haplotypes, suggesting a possible founder event in recent times. The Indian V lineage shared an ancestry with the Caucasian individuals (Coble et al. 2004; Palanichamy et al. 2004) (supplementary fig. 1).

Haplogroup J makes up around 5.3 % of west Eurasian mtDNA in India. Subhaplogroup J1b and its derived J1b1a1, J1b1b, and J1b3 subclades encompass ~65 % of the total J lineages. In India, however, the J haplogroup is widely distributed and found predominantly in southern (Andhra Pradesh) and northern (Punjab and Uttar Pradesh) populations. The much rarer J2 lineages were found only once in the Punjab and Ladakh regions, respectively (supplementary table S1). Indian J1b1a1, J1c1b1a, J1c5, and J1c8 lineages are phylogenetically clustered with Pakistanis, Europeans and central Asians, whereas J1b1b and J1d4 lineages were found to cluster with the Near Easterners (Palanichamy et al. 2004; Pala et al. 2012). In addition,

Indian J1b3- and J1d3b-related haplotypes were found in the Pakistan and Iran populations (Derenko et al. 2013) (supplementary fig. 2).

Haplogroup T makes up around 7.4 % of west Eurasian mtDNAs in India. Approximately, 75 % of Indian samples having T haplogroup represent T1, T1a1 and T2 subhaplogroups. Similar to the J haplogroup distribution in India, T lineages were also found more often in the populations of Andhra Pradesh, Tamil Nadu, Uttar Pradesh, Punjab, and Maharashtra, and about two-third of the lineages (45/88) were found in Indian Muslims and Brahmins (supplementary table S1). Indian T1- and T2-derived lineages—T1a, T1a1, T1a5, T1a1b1, T1b, T2, T2a1a, T2b2, T2c2, T2d1a, and T2e2 tend to cluster mainly with the Near Eastern populations particularly from Iran, Iraq, and Azerbaijan, and some even extend to Europeans (Palanichamy et al. 2004; Pala et al. 2012) (supplementary fig. 2).

Almost 40 % of the mtDNAs types found in India fall into west Eurasian haplogroup U, predominantly subhaplogroups U1a, U2e, U3–U5, U7, and U9. Of these, U7 type attained the highest frequency (20 %), followed by U1a (8 %), U2e (4.2 %), and U5 (3.7 %). In India, U1a-derived—U1a1a4, U1a1c1, U1a1c4, U1a2, and U1a3 lineages have been reported. Amongst these, U1a1a4 and U1a1c4 encompass 80 % of U1a lineages found in southern India (supplementary table S1). U1a lineage was observed to have a relatively high frequency in the Koragas population (22/28) with low HVS-I diversity (only two haplotypes), indicating the result of genetic drift that might have occurred in this population (Cordaux et al. 2003). The fully sequenced Koragas U1a mtDNAs (Ingman and Gyllensten 2003) together with our new sequence defined a new subclade—U1a1a4. This subclade appears to be unique, being found only in the Indian populations. Indian U1a1c4 lineages phylogenetically cluster with the Pakistani and Near Eastern (Iran) populations (Palanichamy et al. 2004; Derenko et al. 2013). U1a2 has been noted in the Jews from Cochin and it shares an ancestry with the Azerbaijan population (Schönberg et al. 2011) (Fig. 1b).

In the literature, there is relatively a little information about the U2e mtDNA haplogroup. This haplogroup has a low frequency in Central Asia, Europe, and the Near Eastern region (Macaulay et al. 1999; Richards et al. 2000; Quintana-Murci et al. 2004; Nasidze et al. 2004; Comas et al. 2004; Al-Zahery et al. 2003; Abu-Amero et al. 2008; Derenko et al. 2014). In India, U2e and its derived subhaplogroups—U2e1, U2e1a1, U2e1b, and U2e3 have been found to have relatively higher frequencies in the populations of Andhra Pradesh and Uttar Pradesh. We observed that some of the Indian U2e mitogenomes (mostly from south India) differ from both the European and Near East U2e counterpart, and may represent a paraphyletic lineage within U2e. Subclade U2e1 lineage though was

detected in north India (Uttar Pradesh and Punjab), its related haplotypes were also found in Pakistan and the Caucasus (Macaulay et al. 1999; Quintana-Murci et al. 2004). Indian U2e1a1 and U2e1b mitogenomes cluster with central Asian and Eastern Europeans (Palanichamy et al. 2004; Derenko et al. 2013, 2014) (supplementary fig. 3). We identified U2e3 lineages in north Indian (Uttar Pradesh) Brahmins. It shares mutations (nucleotide position 16,400 and a retro mutation 217) with the Italian lineage, and this has helped to define a new U2e3b subclade. Besides Uttar Pradesh Brahmins, U2e3 was also reported in Punjab Brahmin and Ladakh Muslim populations (supplementary table S1).

The frequencies of lineages of haplogroups U3, U4 and U9 were found to be low in Indian populations (Supplementary Table S1). The rare U9 lineages were noticed, three in Tamil Nadu and one each from Andhra Pradesh, Uttar Pradesh and Maharashtra. We sequenced two U9 complete mtDNA genomes—one from Tamil Nadu and the other one from Uttar Pradesh. The sample obtained from Tamil Nadu shares several mutations with the previously reported U9a1 lineage from Andhra Pradesh, and this subclade lineage has so far only been reported in south Indian populations. The U9 sequence obtained from north India (Uttar Pradesh) clustered with the Pakistan-specific subclade U9b1 (Fig. 2). Haplogroup U5 comprises two major phylogenetic clusters, U5a and U5b (Malyarchuk et al. 2010). In India, only the eastern European-specific U5a1-derived lineages (Richards et al. 2000) such as U5a1a1, U5a1b, U5a1b1, and U5a1b1f were observed mostly in populations from Andhra Pradesh, Tamil Nadu, Uttar Pradesh and Punjab (Supplementary Table S1).

Haplogroup U7 is a typical Near Eastern and Indian haplogroup. Derenko et al. (2013) recently built a maximum parsimony tree and revealed three subclusters—U7a, U7b, and U7c. Surprisingly, when Indian U7 sequences were examined and compared with the published Eurasian mtDNAs, we observed seven India-specific subclades, viz., U7a2b, U7a3a2, U7a3a3, U7a3b1, U7a6, U7a7, and U7c (Fig. 1b), of which four (U7a2b, U7a3a3, U7a6 and U7a7) are newly identified subhaplogroups in the present study. We have designated the previously defined U7a3a2 as U7a3a2a and U7a3b1 as U7a3b1a. The majority of U7 clades present in Indian populations are U7a, U7a1a, U7a3 lineages and they have a widespread distribution in India (supplementary table S1). Interestingly, the U7a1a, U7a3a2, U7a3a3 and U7a3b1a sequences observed in Brahmin and Muslim populations of India were found to cluster exclusively with the Pakistani and Iran-Persian samples (Derenko et al. 2013) (Fig. 1b). The remaining U7 subhaplogroups—U7a2, U7a4, U7a6, and U7a7 are less frequent and apparently have a scattered distribution across India (supplementary table S1).

A single R1a1a lineage found in the Uttar Pradesh Brahmin population clustered with the Near Eastern and central Asian populations (Malyarchuk et al. 2008; Fedorova et al. 2013). A relatively rare subclade R1b, identified recently from the Russian population, was detected in a Tamil Nadu sample. The complete sequence analysis indicates that this sample shares ancestry with a West Bengal sample and clustered with Armenia and Finland R1b1 mtDNAs samples (Fedorova et al. 2013) (supplementary fig. S4). The R2 haplogroup, which originated in the Near East, was observed in India too, mainly in north Indian populations. Based on the R2 haplogroup phylogeny (Al-Abri et al. 2012; Derenko et al. 2013), the Indian lineages were assigned to different subclades—viz., R2a, R2a4 and R2c1. In addition, we noticed that Indian R2a4 and R2c1 lineages specifically clustered with the Pakistan and Iran sequences (Derenko et al. 2013) (supplementary fig. S4).

In India, though W haplogroup has a wide distribution in the populations of northern and southern India, it is found rarely in east Indians. Subhaplogroups of W such as W1c, W1g, W3, W3a1, W3a1b, W3a2, W4, W4a, and W6 were observed in Indian populations (supplementary table S1). W1c lineages were observed in Brahmin populations of Uttar Pradesh and Tamil Nadu Muslims and this lineage clustered with the Near East-Iranian samples (Olivieri et al. 2013; Derenko et al. 2013) (supplementary fig. S5). W3a1 and W3a1b lineages are sparsely distributed in India. Notably, W3a1b lineages are found specifically in India and Pakistan (Quintana-Murci et al. 2004; Rakha et al. 2011). Subclade W3a2 is very rare and only three complete sequences have been reported so far (Olivieri et al. 2013). In India, we found a single W3a2 lineage from the Uttar Pradesh Brahmin population, and complete sequence analysis confirmed the phylogenetic status of this lineage (supplementary fig. S5). Subclades W4 and W6 have been reported from Europe, the Near East, and the Caucasus (Richards et al. 2000; Nasidze and Stoneking 2001; Nasidze et al. 2004, 2006; Quintana-Murci et al. 2004; Abu-Amero et al. 2008; Al-Zahery et al. 2011; Terremos et al. 2011; Olivieri et al. 2013). The W4 lineages have been detected in western (Gujarat), northern (Punjab, Delhi, Uttar Pradesh, and Himachal Pradesh), and eastern (Odisha) parts of India, but are absent in southern India (supplementary table S1). A total of eight W6 sequences were observed in India (five subjects from Andhra Pradesh and one each from Tamil Nadu, Gujarat and Punjab). We sequenced two W6 mitogenomes, one from Tamil Nadu and the other one from Andhra Pradesh, and found that these lineages differed from the already known W6 subclades (i.e., W6a–W6d) (Derenko et al. 2014), hence we assigned a new subclade, W6e (supplementary fig. S5).

The N1a1a haplotypes were detected in the Havik (Karnataka) and Iranian (Maharashtra) populations. It is

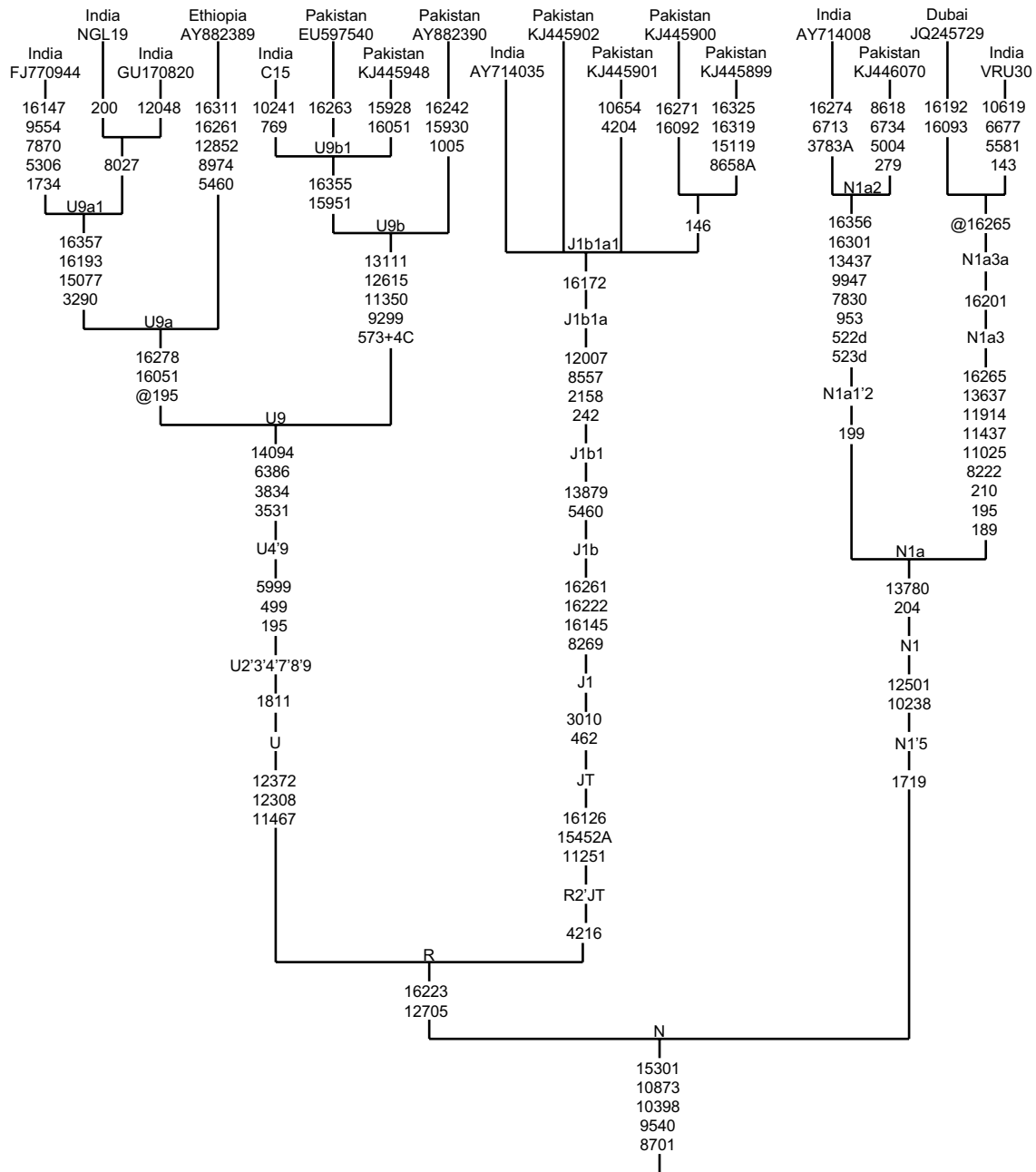


Fig. 2 Phylogenetic tree of haplogroups U9, J1b1a, N1a2, and N1a3. For additional information, see the Fig. 1 legend

noteworthy that closely related haplotypes were found in the populations from the Caucasus, Turkey and Iran (Palanichamy et al. 2010). We observed one N1a1a1 sequence in Tamil Nadu and two in West Bengal and these clustered within the Near East/European subclade N1a1a1a1 (Palanichamy et al. 2010). N1a1b1 sister clade-I haplogroup and its subhaplogroups II, I1b, I4a1, and I4b have been observed in our studied samples, but with less frequencies (~1 %). We observed the extremely rare N1a1b1a lineage in one Uttar Pradesh sample and two in Tamil Nadu, and

one N1a1b1a1 lineage from Uttar Pradesh sample. Our complete sequence analysis indicates that these samples clustered with Dubai, Iran, and Pakistan samples (Fernandes et al. 2012; Olivieri et al. 2013) (supplementary fig. S6). In addition, we noticed that the Siberian N1a1b1 lineage shares six variants with the Pakistan sample and forms a new subclade N1a1b1b (Derenko et al. 2014; GenBank KJ446053). Haplogroup N1a2 is sparsely distributed in south (one in Tamil Nadu, Two in Kerala, and six in Andhra Pradesh) and north Indian (one each in Delhi

Table 1 Estimated ages (years) for different subclades of West Eurasian haplogroups in India

Clade	No of mtDNAs	Coding region-based substitutions		Only synonymous transitions			
		Age ^a	Age ^b	Age ^b	Age ^c	Age ^d	Age ^e
H13a2a	6	7700 ± 3900	6900 ± 3500	7650 ± 5400	6800 ± 4800	7900 ± 5600	8000 ± 5600
H14c	2	10,300 ± 5100	9200 ± 4600	NA	NA	NA	NA
HV14a1	3	11,900 ± 5100	10,800 ± 4600	10,200 ± 6200	9000 ± 5500	10,500 ± 6400	10,650 ± 6500
J1b1a1	5	4100 ± 2100	3700 ± 1800	3100 ± 2200	2700 ± 1900	3100 ± 2200	3200 ± 2300
U1a1a4	2	17,900 ± 6800	16,100 ± 6,100	15,300 ± 7600	13,500 ± 6700	15,700 ± 7800	16,000 ± 8000
U7a1a	2	2600 ± 2500	2300 ± 2200	3800 ± 3700	3400 ± 3300	3900 ± 3800	4000 ± 3900
U7a3a3	2	2600 ± 2500	2300 ± 2200	3800 ± 3700	3400 ± 3300	3900 ± 3800	4000 ± 3900
U7a6	2	10,300 ± 5100	9200 ± 4600	7650 ± 5400	6800 ± 4800	7900 ± 5600	8000 ± 5600
U7a7	2	7700 ± 4400	6900 ± 4000	7650 ± 5400	6800 ± 4800	7900 ± 5600	8000 ± 5600
U7c	3	6900 ± 3400	6100 ± 3100	7650 ± 4400	6800 ± 3900	7900 ± 4500	8000 ± 4600
U9a1	3	11,900 ± 5100	10,800 ± 4600	7650 ± 4400	6800 ± 3900	7900 ± 4500	8000 ± 4600
N1a2	2	15,400 ± 6300	13,800 ± 5600	15,300 ± 7600	13,500 ± 6700	15,700 ± 7800	16,000 ± 8000

^a Mutation rate of Mishmar et al. (2003)

^b Mutation rate of Perego et al. (2009)

^c Mutation rate of Kivisild et al. (2006)

^d Mutation rate of Soares et al. (2009)

^e Mutation rate of Loogväli et al. (2009)

and Punjab) populations (Supplementary Table S1). It has also been reported in Pakistan (GenBank KJ446070) (Fig. 2), Iran and Turkey (Richards et al. 2000; Nasidze et al. 2006). N1a2 is primarily confined to India and Pakistan. Members of haplogroup N1a3 were observed in three individuals from south India (Tamil Nadu and Karnataka). We completely sequenced one N1a3 mtDNA sample originating from Tamil Nadu, which shares reverse substitution in HVS-I (16265) with the Dubai mitogenome (Fernandes et al. 2012) (Fig. 2).

Apart from complete sequence phylogeny, we noticed that the matches of Indian H2a1, H2a1a, H2b, H5a1, H6a1a, H7b, H9, H14a, H15a1a1, HV6, HV12b1, HV12b1a, U3, U4, and U5 mtDNA HVS-I haplotypes were found in the populations of the Caucasus, Armenia, Anatolia, Turkey, Iran, Iraq, Georgia, Jordan, Syria, Arabia Peninsula, Kuwait, Pakistan, central Asia, Tajikistan, Uzbekistan, Siberia, Slovaks, and Russia (Comas et al. 1998, 2000, 2004; Macaulay et al. 1999; Richards et al. 2000; Nasidze and Stoneking 2001; Loogväli et al. 2004; Nasidze et al. 2004; Quintana-Murci et al. 2004; Roostalu et al. 2007; Abu-Amro et al. 2008; Malyarchuk et al. 2008; Nasidze et al. 2008; Rakha et al. 2011; Al-Zahery et al. 2011; Terrenos et al. 2011; Derenko et al. 2013).

Dating the Indian-specific west Eurasian subclades

Coalescence times for the Indian-specific west Eurasian subclades—H13a2a, H14c, HV14a1, J1b1a1, U1a1a4,

U7a1a, U7a3a3, U7a6, U7a7, U7c, U9a1 and N1a2 were estimated (Table 1). The subclades—H14c, U1a1a4, U7a1a, U7a3a3, U7a6, U7a7 and N1a2 showed a broad range of standard errors, possibly due to their calculation from the data of only a few complete sequences. The addition of new complete mtDNA sequences to these subclades might provide more reliable time estimates. In this study, very few H14c, U1a1a4 and N1a2 haplotypes were detected and therefore, we were unable to add new complete mtDNA sequences to these subclades. However, the date of 5.7 kya determined by considering all the Indian-specific U7 subclades can be accepted as a reliable time estimate.

Discussion

The Elamite and Dravidic linguistic connection and its spread into south India

The autochthonous subhaplogroups—HV14a1 and U1a1a4 uniquely found in contemporary Dravidian speakers share their ancestry primarily with the Near East-Iran populations (Derenko et al. 2013). The coalescence times of HV14a1 and U1a1a4 were estimated to be ~10.5–17.9 kya. The shared ancestry of the Dravidian of South India and Iranian of Near East populations has been shown in the HV14 and U1a1 phylogeny (Fig. 1a) and their time estimates are consistent with the proto-Elamo-Dravidian language diffusion

hypothesis which emphasized that the proto-Dravidian language evolved over 15 kya, specifically in western Asia before the beginning of agricultural development ~11 kya. This language was introduced by Neolithic pastoralists, and was thought to be associated with the spread of these west Eurasian-specific mtDNAs to peninsular India (Pagel et al. 2013). The Y-chromosome haplogroup L1a has added a further dimension to this hypothesis. The subclades of haplogroup L such as L1a, L1b, and L1c were found predominantly in Iranian populations of western Asia (Grugni et al. 2012). In India, only the L1a lineage was observed and was largely restricted to the Dravidian-speaking populations of south India (Sahoo et al. 2006; Sengupta et al. 2006). The coalescence time (~9.1 kya) (Sengupta et al. 2006) and the virtual absence in Indo-Aryan speakers in north indicate that the L1a lineage arrived from western Asia during the Neolithic period and perhaps was associated with the spread of the Dravidian language to India. These results emphasized that the Dravidian language originated outside India and may have been introduced by pastoralists coming from western Asia (Iran).

The origins of the Indian caste system

The most frequent west Eurasian haplogroup in India is U7 (~20 %). This haplogroup is thought to have originated about 16–22 kya in west Asia–Iran (Derenko et al. 2013). Haplogroup U7 comprises several sublineages, which are found frequently in India, Iran and the Near East (Kivisild et al. 1999a; Metspalu et al. 2004; Palanichamy et al. 2004; Derenko et al. 2013). Nevertheless, subsets of U7 lineages (i.e., U7a1, U7a2b, U7a3a2, U7a3a3, U7a3b1, U7a6, U7a7, and U7c) have been identified in Indian individuals. The age of these autochthonous U7 subhaplogroups coalesce at a recent time, ranging from 2.6 to 8.0 kya with an average of 5.7 kya (95 % CI 2.9–8.5), suggesting that the spread of these subhaplogroups might be associated with the arrival of Indo-Aryan speakers from west Asia. Moreover, the high frequencies of U7 lineages in the higher-ranked Brahmin and religious Muslims and their rarity in lower-ranked caste and tribal groups are consistent with the hypothesis that Indo-Aryan speakers were perhaps responsible for the social stratification and the attaining of a higher-ranked caste status. Further support for this notion has been derived from the Y-chromosome haplogroup J2 distribution, which closely matches the distribution pattern of mtDNA haplogroup U7 in India. Interestingly, J2 lineages were found to be frequent in the higher-ranked caste/Muslims and absent or negligibly present in the lower caste and tribal groups. Taken together, our findings support the hypothesis that the social stratification was generated by the Indo-Aryan speaking immigrants.

The subclades of other west Eurasian H (H2, H3, H5, H6, H7, H9, H13, H14 and H15), HV (HV2 and HV6), R0 (R0a2), V (V2a), N1 (N1a1a), W (W1, W3 and W4), and R (R2) lineages were observed frequently among the higher-ranked caste, such as in Brahmin populations, irrespective of their geographic origins. As inferred from the phylogeny, these lineages were more diverse in west Asia, and these west Eurasian lineages in Indian populations were found to cluster with the populations from the Near East, the Caucasus, and central Asia (Macaulay et al. 1999; Richards et al. 2000; Comas et al. 2000, 2004; Nasidze and Stoneking 2001; Nasidze et al. 2004, 2006; Quintana-Murci et al. 2004; Abu-Amero et al. 2008; Al-Zahery et al. 2011; Terreros et al. 2011; Scheible et al. 2011). Unfortunately, the arrival dates for these lineages could not be detected due to the absence of Indian-specific subclades. However, the presence of these lineages lends further support toward the contribution of Aryan speakers in establishing culture and stratifying the caste rank in the pre-existing Indian populations. The presence of other west Eurasian subclades—JT, K, and I in tribal, caste, and religious Muslim populations suggest that their spread in India took place in more recent times, perhaps through the invasions by the Moghuls, Muslims, English, and other groups.

Conclusion

The present study deconstructs the clustering of west Eurasian mtDNA lineages into Indian-specific subfounders and resolves the competing hypotheses that describe the Indo-Aryan migration, the spread of the Dravidian language, and the origins of the caste system. The presence of mtDNA haplogroups (HV14 and U1a) and Y-chromosome haplogroup (L1) in Dravidian populations indicates the spread of the Dravidian language into India from west Asia (Sengupta et al. 2006; Sahoo et al. 2006; Derenko et al. 2014). However, this claim should not be considered exclusively to establish the fact that Dravidian populations of India originated in west Asia and then migrated to south India (Sengupta et al. 2006). The co-presence of these west Eurasian mtDNA and Y-chromosome haplogroups in Dravidian-speaking tribal and caste groups in South India suggest that their spread is not associated with the Indo-Aryan migration (Arunkumar et al. 2012). A large proportion of the west Eurasian mtDNA haplogroups observed among the higher-ranked caste groups, their phylogenetic affinity and age estimate indicate recent Indo-Aryan migration to India from west Asia. The Indian–west Eurasian lineage diversity and frequency are highest in higher-ranked caste, implying that the west Eurasian admixture was restricted to caste rank. It is likely that Indo-Aryan migration has influenced the social stratification in the pre-existing populations and

helped in building the Hindu caste system, but it should not be inferred that the contemporary Indian caste groups have directly descended from Indo-Aryan immigrants (Cordaux et al. 2004). Thus, our comprehensive analyses of the west Eurasian mtDNA in Indian populations suggest that the Dravidian language was introduced to India from west Asia, and perhaps the caste system was influenced by the Indo-Aryan immigrants.

Acknowledgments This work was supported by Yunnan University, the Yunnan provincial Science and Technology Department and from Grants from the National Natural Science Foundation of China (31060156 and 31240018).

Conflict of interest The authors declare that they have no competing interests.

References

- Abu-Amero KK, Larruga JM, Cabrera VM, González AM (2008) Mitochondrial DNA structure in the Arabian Peninsula. *BMC Evol Biol* 8:45
- Achilli A, Rengo C, Magri C et al (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75:910–918
- Al-Abri A, Podgorná E, Rose JI et al (2012) Pleistocene-Holocene boundary in Southern Arabia from the perspective of human mtDNA variation. *Am J Phys Anthropol* 149:291–298
- Al-Zahery N, Semino O, Benuzzi G et al (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol Phylogenet Evol* 28:458–472
- Al-Zahery N, Pala M, Battaglia V et al (2011) In search of the genetic footprints of Sumerians: a survey of Y-chromosome and mtDNA variation in the Marsh Arabs of Iraq. *BMC Evol Biol* 11:288
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147
- Arunkumar G, Soria-Hernanz DF, Kavitha VJ et al (2012) Population differentiation of southern Indian male lineages correlates with agricultural expansions predating the caste system. *PLoS One* 7:e50269
- Bamshad MJ, Watkins WS, Dixon ME et al (1998) Female gene flow stratifies Hindu castes. *Nature* 395:651–652
- Bamshad M, Kivisild T, Watkins WS et al (2001) Genetic evidence on the origins of Indian caste populations. *Genome Res* 11:994–1004
- Basu A, Mukherjee N, Roy S et al (2003) Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res* 13:2277–2290
- Behar DM, van Oven M, Rosset S et al (2012) A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet* 90:675–684
- Cerný V, Mulligan CJ, Fernandes V et al (2011) Internal diversification of mitochondrial haplogroup R0a reveals post-last glacial maximum demographic expansions in South Arabia. *Mol Biol Evol* 28:71–78
- Chaubey G, Metspalu M, Kivisild T, Villems R (2007) Peopling of South Asia: investigating the caste-tribe continuum in India. *Bioessays* 29:91–100
- Coble MD, Just RS, O’Callaghan JE et al (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. *Int J Legal Med* 118:137–146
- Comas D, Calafell F, Mateu E et al (1998) Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet* 63:1824–1838
- Comas D, Calafell F, Bendukidze N, Fañanás L, Bertranpetit J (2000) Georgian and kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. *Am J Phys Anthropol* 112:5–16
- Comas D, Plaza S, Wells RS et al (2004) Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur J Hum Genet* 12:495–504
- Cordaux R, Saha N, Bentley GR, Aunger R, Sirajuddin SM, Stoneking M (2003) Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur J Hum Genet* 11:253–264
- Cordaux R, Aunger R, Bentley G, Nasidze I, Sirajuddin SM, Stoneking M (2004) Independent origins of Indian caste and tribal paternal lineages. *Curr Biol* 14:231–235
- Derenko M, Malyarchuk B, Bahmanimehr A et al (2013) Complete mitochondrial DNA diversity in Iranians. *PLoS One* 8:e80673
- Derenko M, Malyarchuk B, Denisova G et al (2014) Western Eurasian ancestry in modern Siberians based on mitogenomic data. *BMC Evol Biol* 14:217
- Fedorova SA, Reidla M, Metspalu E et al (2013) Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol Biol* 13:127
- Fernandes V, Alshamali F, Alves M et al (2012) The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa. *Am J Hum Genet* 90:347–355
- Forster P, Torroni A, Renfrew C, Röhl A (2001) Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol Biol Evol* 18:1864–1881
- Gaikwad S, Kashyap VK (2005) Molecular insight into the genesis of ranked caste populations of western India based upon polymorphisms across non-recombinant and recombinant regions in genome. *Genome Biol* 6:P10
- Grugni V, Battaglia V, Hooshiar Kashani B et al (2012) Ancient migratory events in the Middle East: new clues from the Y-chromosome variation of modern Iranians. *PLoS One* 7:e41252
- Hartmann A, Thieme M, Nanduri LK et al (2009) Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum Mutat* 30:115–122
- Ingman M, Gyllenstein U (2003) Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res* 13:1600–1606
- Kivisild T, Bamshad MJ, Kaldma K et al (1999a) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9:1331–1334
- Kivisild T, Kaldma K, Metspalu M, Parik J, Papiha S, Villems R (1999b) The place of the Indian mtDNA variants in the global network of maternal lineages and the peopling of the Old World. In: Deka R, Papiha SS (eds) *Genomic diversity*. Kluwer Academic/Plenum Publishers, New York
- Kivisild T, Papiha SS, Rootsi S et al (2000) An Indian ancestry: a key for understanding human diversity in Europe and beyond. In: Renfrew C, Boyle K (eds) *Archaeogenetics: DNA and the population history of Europe*. McDonald Institute for Archaeological Research, Cambridge
- Kivisild T, Rootsi S, Metspalu M et al (2003a) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet* 72:313–332

- Kivisild T, Rootsi S, Metspalu M et al (2003b) The genetics of language and farming spread in India. In: Bellwood P, Renfrew C (eds) Examining the farming/language dispersal hypothesis. McDonald Institute Monographs, Cambridge
- Kivisild T, Shen P, Wall DP et al (2006) The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172:373–387
- Loogväli EL, Roostalu U, Malyarchuk BA et al (2004) Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol* 21:2012–2021
- Loogväli EL, Kivisild T, Margus T, Villems R (2009) Explaining the imperfection of the molecular clock of Hominid mitochondria. *PLoS One* 4:e8260
- Macaulay V, Richards M, Hickey E et al (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 64:232–249
- Malyarchuk B, Grzybowski T, Derenko M et al (2008) Mitochondrial DNA phylogeny in Eastern and Western Slavs. *Mol Biol Evol* 25:1651–1658
- Malyarchuk B, Derenko M, Grzybowski T et al (2010) The peopling of Europe from the mitochondrial haplogroup U5 perspective. *PLoS One* 5:e10285
- Metspalu M, Kivisild T, Metspalu E et al (2004) Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5:26
- Mishmar D, Ruiz-Pesini E, Golik P et al (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci* 100:171–176
- Nasidze I, Stoneking M (2001) Mitochondrial DNA variation and language replacements in the Caucasus. *Proc Biol Sci* 268:1197–1206
- Nasidze I, Ling EY, Quinque D et al (2004) Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Ann Hum Genet* 68:205–221
- Nasidze I, Quinque D, Rahmani M, Alemohamad SA, Stoneking M (2006) Concomitant replacement of language and mtDNA in South Caspian populations of Iran. *Curr Biol* 16:668–673
- Nasidze I, Quinque D, Rahmani M, Alemohamad SA, Stoneking M (2008) Close genetic relationship between Semitic-speaking and Indo-European-speaking groups in Iran. *Ann Hum Genet* 72:241–252
- Olivieri A, Pala M, Gandini F et al (2013) Mitogenomes from two uncommon haplogroups mark late glacial/postglacial expansions from the near east and Neolithic dispersals within Europe. *PLoS One* 8:e70492
- Pagel M, Atkinson QD, Calude AS, Meade A (2013) Ultra conserved words point to deep language ancestry across Eurasia. *Proc Natl Acad Sci* 110:8471–8476
- Pala M, Olivieri A, Achilli A et al (2012) Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am J Hum Genet* 90:915–924
- Palanichamy MG, Sun C, Agrawal S et al (2004) Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am J Hum Genet* 75:966–978
- Palanichamy MG, Zhang CL, Mitra B et al (2010) Mitochondrial haplogroup N1a phylogeography, with implication to the origin of European farmers. *BMC Evol Biol* 10:304
- Perego UA, Achilli A, Angerhofer N et al (2009) Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol* 19:1–8
- Quintana-Murci L, Chaix R, Wells RS et al (2004) Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet* 74:827–845
- Rajkumar R, Kashyap VK (2003) Haplotype diversity in mitochondrial DNA hypervariable regions I and II in three communities of Southern India. *Forensic Sci Int* 136:79–82
- Rakha A, Shin KJ, Yoon JA et al (2011) Forensic and genetic characterization of mtDNA from Pathans of Pakistan. *Int J Legal Med* 125:841–848
- Ranaweera L, Kaewsutthi S, Win Tun A, Boonyarit H, Poolsuwan S, Lertrit P (2014) Mitochondrial DNA history of Sri Lankan ethnic people: their relations within the island and with the Indian sub-continental populations. *J Hum Genet* 59:28–36
- Richards M, Macaulay V, Hickey E et al (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67:1251–1276
- Roostalu U, Kutuev I, Loogväli EL et al (2007) Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Mol Biol Evol* 24:436–448
- Sahoo S, Singh A, Himabindu G et al (2006) A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc Natl Acad Sci* 103:843–848
- Saillard J, Forster P, Lynnerup N, Bandelt HJ, Nørby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67:718–726
- Scheible M, Alenizi M, Sturk-Andreaggi K, Coble MD, Ismael S, Irwin JA (2011) Mitochondrial DNA control region variation in a Kuwaiti population sample. *Forensic Sci Int Genet* 5:e112–e113
- Schönberg A, Theunert C, Li M, Stoneking M, Nasidze I (2011) High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *Eur J Hum Genet* 19:988–994
- Sengupta S, Zhivotovsky LA, King R et al (2006) Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet* 78:202–221
- Soares P, Ermini L, Thomson N et al (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84:740–759
- Tamang R, Thangaraj K (2012) Genomic view on the peopling of India. *Investig Genet* 3:20
- Terreros MC, Rowold DJ, Mirabal S, Herrera RJ (2011) Mitochondrial DNA and Y-chromosomal stratification in Iran: relationship between Iran and the Arabian Peninsula. *J Hum Genet* 56:235–246
- van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30:e386–e394