J Biosci (2019)44:74
DOI: 10.1007/s12038-019-9875-5

Reconstruction of ancestral footfalls in South Asia using genomic data

Saikat Chakraborty* and Analabha Basu*

National Institute of Biomedical Genomics, Kalyani, West Bengal 741 251, India

**Corresponding author (Email, sc6@nibmg.ac.in; ab1@nibmg.ac.in)*

Due to its unique geographical position, juxtaposed in the middle of south-central Asia, east Asia and Southeast Asia, the South Asian Region (SAS) has repeatedly come into contact with people from adjacent regions throughout history and prehistory. The antiquity of the populations and the intricate history of admixture have shaped SAS as one of the most genetically diverse regions in the world. In this article we review our current understanding of the peopling and populations structure of SAS. We do not attempt to be exhaustive but summarize the salient conclusions that have been reached using genetic data and evaluate their robustness. We also identify the unanswered questions and suggest possible approaches that may lead to their answers.

Keywords. Ancestry; DNA; frequency; genome; populations

1. Introduction

South Asian Region (SAS) harbours bewildering ethnic, linguistic and cultural diversity and has been a cradle of many different populations from prehistoric times. Every one out of six people in the contemporary world calls South Asia her/his home (Worldometers Population). Today's SAS comprises India, Pakistan, Afghanistan, Nepal, Bhutan, Bangladesh, Sri Lanka and the Maldives (South Asian Regional Development Gateway 2008). For the longest span of history and prehistory, geophysical barriers have been the major impediment in people's migrations and movements, affecting the extent of admixture and intermixing of the populations. More recently, with the formation of the nation states, national borders affected population movement and isolation. The way culture affects admixture and intermixing of populations is much more intricate. Ethnic identities, language, adherence to religions and social hierarchies like castes discourage pan-mixing. To further complicate matters, the definitions of the cultural labels of ethnicity, religion and caste/tribe are not stone-carved and are temporally dynamic and often confounded. In spite of the temporal dynamism as well as the porosity of structures, it is important and often helpful to consider the broadbrush outline of the social fabric which influences the mating and admixture between individuals.

The predominant religions that SAS people subscribe to are Hinduism, Islam, Sikhism, Christianity and Buddhism. The people identified to be under the Hindu fold, by the Government of India, are further classified into castes and tribes. Brahmins, Kshatriyas, Vaisyas and Sudras are the four *varnas* which are confounded with the notions and definitions of caste. Brahmin and Kshatriyas are confounded with individuals generally perceived to belong to the 'upper' caste, the Vaishyas, 'middle' caste, and the Sudras, 'lower' caste. Castes are further divided into several sub-castes with their own intra-caste and sub-caste hierarchies. Inter-caste marriages or even inter-sub-caste marriages have been and still are quite rare. Although such ordered hierarchies have inculcated and still inculcate acute prejudice and discrimination from the people 'higher' in the hierarchy towards the 'lower', the anthropogenetics of SAS cannot be understood ignoring it (Kosambi 1965; Majumder and Basu 2015).

Another major additional layer of complexity are the tribal populations. Tribes are broadly defined as 'a community occupying a common geographic area and having a similar language and culture or beliefs and practices'. Today, besides the self-identified distinction that people of the tribal communities possess, it is also often interchangeably used as administrative category in governance. India is home to around 700 scheduled tribes (Government of India 2013). Just like the caste people, these tribes practice endogamy too.

The primary languages of almost all SAS people belong either to the Indo-European (IE), Dravidian (DR), Austro-Asiatic (AA), or Tibeto-Burman (TB) language family. People speaking AA languages are exclusively tribal. With few exceptional populations like 'Manipuri Brahmins', the same is almost true for all people speaking TB languages. Direct contact between several of the tribes and caste people

http://www.ias.ac.in/jbiosci
Published online: 08 July 2019

during the last few centuries has led to substantial influence of the language spoken by the latter on that of the former, so much so that many present-day tribes have ceased to use their ancestral language and today speak in the language of the caste people instead.

Such richness, diversity and complexity in all aspects of human existence makes this region interesting to academic historians, social scientists, linguists as well as geneticists. Modern genetics provides extremely sophisticated tools to characterize the distribution of genetic diversity over space and time and, more importantly, infer the processes that may have led to the observed pattern. Unlike other disciplines, genetics can shine light on the prehistory and history of a population just by studying the genomes of present-day populations. Such an approach is the only inferential procedure available when data from other disciplines such as archeology or physical anthropology are rare. In the context of SAS, the following questions regarding its peopling and ancestry could be argued as important:

1. When was SAS colonized by anatomically modern humans (AMH, i.e. *Homo sapiens sapiens*)?
2. Has there been subsequent migration into this region following this first colonization? If yes, when?
3. What was the source population of the initial and subsequent colonizers and what were the routes?
4. Is the variation in current SAS population continuous or can one identify discontinuities among the genomes of SAS people, i.e. into how many ancestral components can the SAS people be partitioned and what is the relation among these components?
5. What has been the degree of admixture among different populations? When did this admixture take place?
6. What signature has endogamy left on the current genetic diversity of SAS?
7. How have sizes of SAS populations changed through time and how has it impacted the genetic diversity of SAS and adjacent regions?

Geneticists use different kinds of tools to address these questions. The primary data that is used is information about genetic variation. In the absence of variation, none of the above questions can be answered. As we will see, genetic variation at different regions of the genome behaves differently, and can shine light on different aspects of the history of a population. Geneticists today primarily use three broad categories of data: (i) uniparental markers, i.e. the mitochondrial genome (MG) or the Y chromosome, (ii) autosomal markers, mainly single-nucleotide polymorphisms (SNPs) data from extant population, and (iii) ancient DNA (aSNPs).

2. Uniparental markers

The human sperm does not contain any mitochondria, but the egg does. Consequently, all the mitochondria in a human adult are copies of those initial mitochondria present in the egg. The mitochondrion has its own genome, and whenever it is copied, its genome gets copied too. However, once in a while, a 'copy error' or mutation is introduced. As each mitochondrion contains only a single circular chromosome, recombination is absent in this case. One may infer from this that two individuals with fewer number of differences in their MG share a more recent maternal ancestor than two individuals with more such differences. Thus, by comparing these differences, the maternal lineage of individuals can be reconstructed (Cann *et al.* 1987). Contrarily, only the human male carries the Y chromosome (YC), and just like the MG, recombination is absent for almost its entire length. As with MG, by comparing the variation in YC, one can glean an idea about the paternal lineage (Hammer 1995; Thomson *et al.* 2000).

Due to the lack of recombination, MG and YC behave like a single locus. Any new mutation occurs on an existing background, enabling one to trace the relationship among different MGs and YCs. The phylogenetic tree of MGs and YCs consists of major clusters which are known as haplogroups (HG). Each of the HG further consists of smaller nodes known as sub-haplogroups (sHG) (Torroni *et al.* 1993, 1994; Underhill and Kivisild 2007). These different HG or sHG can be thought of as different alleles. Each population may contain different HG and sHG in different proportions (allele frequencies). By combining the information obtained from the MG/YC phylogenetic tree and the frequencies of different HG and sHG relationship between and among different populations can be studied. Additionally, if one assumes that mutation accumulates in a clocklike manner in selectively neutral sites, one may also estimate the dates for several relevant events, such as when two lineages diverged or how long a population has been occupying a particular region.

The mitochondrial HG L1 and L2 are exclusively African and coalesce at round 190 kya (Oppenheimer 2004, 2012; Van Oven and Kayser 2009). The sHG of L1 and L2 also have coalescent times (CT) of $\sim$150 kya. The HG L1 and L2 are basal to all non-African HG. All non-African MG HG coalesce at around 80 kya. These suggest that all non-African humans are derived from African ancestors and the African/non-African split took place $\sim$80 kya (Oppenheimer 2012). All non-African MG HG are derived from the L3 HG. The two main sub-lineages of L3 are the M HG and the N HG. The M HG with a CT of $\sim$72 kya is the predominant MG HG among the Indian tribes ($\sim$70%). It also occurs at a very high frequency among the 'lower' castes (>60%) (Basu *et al.* 2003). This, along with the virtual absence of the M HG from Europe, the Middle East and Central Asia, suggests that the earliest AMH colonized India using a southern Out of Africa (OOA) route by following the Arabian coast (Kivisild *et al.* 1999; Basu *et al.* 2003). This also suggests the tribes and castes to be older than the other populations, and possibly the earliest colonizers of India (Basu *et al.* 2003). The M HG has the lowest frequency among the Muslims, followed by the 'upper' caste Hindus.

The M sHG M2, the most frequent M sHG among the Indian tribes, has the highest coalescent time among all South and Southeast Asian and the Papuan M HGs. This suggests that AMH reached India earlier than Southeast Asia or Oceania (Forster *et al.* 2001). The U HG with a CT of 45 kya is the second most frequent HG among Indians and shows a trend that is reverse to that of the M HG, i.e. the frequency of U HG is the highest among the Indian Muslims (∼35%) and lowest in the tribes (<10%) (Basu *et al.* 2003). The U sHG U2e and U2i share a common ancestor 53 kya and are specific to western Eurasia and India, respectively (Kivisild *et al.* 1999). U2e is absent among the Indian tribes and is most frequent among the 'upper' caste Hindus. U2i is the most frequent U sHG among the tribes. These suggest that (1) India was colonized by AMH probably during one of the earliest Out of Africa (OOA) events, (2) European and the indigenous tribes of India diverged very early and have been evolving independently, and (3) substantial admixture occurred between people with European ancestry and ancestors of people who today constitute the 'upper' caste in India. The overall low mitochondrial genetic diversity among Indians suggests a small number of females among the founders (Basu *et al.* 2003).

The pattern observed in case of Indian Y chromosome haplogroups suggests complex histories for different Indian populations (Sengupta *et al.* 2006). The HG with the highest frequency (∼50%) among the AA tribes (O2a), although not maximum, occurs at high frequencies in the Dravidian tribes too (27%). Interestingly, this HG is virtually absent from the caste populations. In case of the Dravidian tribes, the most frequent HG is HG1 (37%). This HG also occurs in high frequency in both DR- and IE-speaking middle and lower castes (DR: middle = 20%, lower = 14%; IE: middle = 33%, lower = 28%), but is relatively rare in upper castes (2% and 10% in DR- and IE-speaking tribes, respectively) (Sengupta *et al.* 2006). This suggests that many middle and lower caste populations, speaking both DR and IE languages today, may have been originally tribes. The most frequent HG among the TB-speaking tribes is O3e (66%), which is absent from all other South Asian populations (Sengupta *et al.* 2006). Both results from MG as well as YC HG suggest that the TB-speaking tribes have a very different origin than the other South Asian populations. The HG with the highest frequency both among the IE- and DR-speaking upper caste populations is R1a1 (DR = 29%, IE = 46%). This along with evidence from MG HG suggests a common ancestry for the upper castes irrespective of the language they speak today. HG O2a frequency is significantly correlated with longitude, but not with latitude. This is because most AA-speaking tribes, where the frequency of this HG is maximum, inhabit central India and the TB speakers (where the frequency of this HG is relatively low) reside in the northeastern part of the subcontinent. The H1 HG is negatively correlated (Spearman's rank correlation coefficient, −0.74) with latitude, suggesting a south-to-north cline in its frequency. This is expected as the DR tribes are concentrated in south and middle India. The R1a1 HG is neither correlated with latitude nor with longitude. It occurs in highest frequencies (30–40%) discontinuously in the sub-Himalayan region of north India adjacent to the western border of Nepal and southern Indus valley in the Sind area of Pakistan (Sengupta *et al.* 2006). The most parsimonious explanation for this seems to be independent increase in frequency of this HG due to random genetic drift in these two regions.

3. Autosomal markers

Although substantial knowledge about the peopling and population structure of South Asia has been obtained using uniparental markers, they suffer from considerable limitations: (1) Being uniparental, only $(½)^k$ proportion of the ancestry k generations ago is represented by the MC or HG. (2) Due to the lack of recombination, MG and YC HG behave as a single locus, and inferences drawn from differences in their frequencies among different populations is quite subjective at times. (3) As both of these markers represent only a small subset of the total genomic variation, all inferences drawn are based on this limited variation. Consequently, estimates obtained using uniparental markers have very large confidence intervals, making reliable comparisons virtually impossible. (4) Only males carry Y chromosomes and even there only one of the two sex chromosomes is Y. Therefore, Y chromosome markers have an effective population size (N_e) that is roughly ¼ of the autosomes. Therefore, YC HG are more prone to random genetic drift.

Autosomal markers, on the other hand, are very large in number and have relatively higher N_e. However, until recently it was not easy to type them. With the advent of the DNA microarray technology and more recently massively parallel sequencing technologies, popularly known as next-generation sequencing (NGS), typing a substantial proportion of the autosomal markers for a large number of individuals has become possible. Using such markers for different South Asian populations belonging to all the four major language families, four distinct genetic clusters have been identified. These are (1) the Ancestral North Indian (ANI) ancestry: This ancestry is related to the present-day Central Asians, Middle Easterns and Europeans, and in SAS consists of populations residing primarily towards northern South Asia and upper castes from both north and south India. These populations at present primarily speak in IE languages. However, south Indian Brahmins speaking DR languages also belong to this ancestry. (2) The Ancestral South Indian (ASI) ancestry: As the name suggests, these ancestries are distributed towards southern India. DR-speaking tribes and lower castes and several IE-speaking lower castes belong to this ancestry. (3) The ancestral Austro-Asiatic (AA) ancestry: Consists of AA-speaking tribes from central India. (4) The Ancestral Tibeto-Burman (ATB) ancestry: Consists primarily of TB-speaking tribes from

northeastern India (Basu *et al.* 2016). It has also been observed that the ANI ancestry decreases from northern to southern India (Reich *et al.* 2009). Major colonization from the Central Asia and the Middle East into the subcontinent is not required to explain this pattern. Continuous migration among adjacent populations under a stepping stone model of population structure is expected to lead to a cline in allele frequencies and therefore a cline in ancestries (Kimura and Weiss 1964). Among the mainland populations, the ATB populations are the most distinct. This, along with the pattern observed in the MG and YC HG, suggests that these populations could have migrated to SAS more recently than the populations of the other three ancestries, or that they have remained relatively isolated from the rest of the SAS populations. It has also been observed that the AA- and DR-speaking tribes share considerable amount of ancestral component. The Gond tribe, speaking DR language today, was found to be more closely related with the present-day AA speakers (Chaubey *et al.* 2017). Substantial admixture was observed across all the four ancestries (Basu *et al.* 2016). From the length of haplotype blocks (chromosomal segments) belonging to the four ancestries in different SAS populations, the mean time for the introduction of endogamy was estimated to be 73 generations before the present (Basu *et al.* 2016).

4. Ancient DNA

DNA recovered from ancient remains of AMH provides an invaluable glimpse into the past genetic variation, thereby illuminating the processes that have led to the contemporary pattern of genetic variation. However, obtaining uncontaminated, good quality DNA involves considerable technical challenge (Pääbo *et al.* 2004). Using 357 ancient samples from Central and South Asia, ranging from 10,000 to 50 BC, including 56 samples from the Swat valley in northern Pakistan (date range 900 BC to 50 BC), it was observed that all ancient as well as modern South Asian populations have a large proportion of ancient Iranian agriculturist (sample date $\sim$8000 BC) related ancestry (Narasimhan *et al.* 2018). This proportion was more for ancient samples (>50% in most cases) and less for modern samples (<50%). No caste-specific difference in this proportion was observed. Interestingly, almost all the ancient samples consisted of ancestry component related to Neolithic West Siberians (age $\sim$5000 BC), but among the modern-day SAS populations, only Punjabis, Sindhis and Brahmins did. This is expected as these three populations belong to the ANI ancestry (Narasimhan *et al.* 2018).

This is the only study with aDNA samples from SAS till date. However, the conclusions of this study are debatable as several populations that do not lie on a hypothetical 'Indian cline' defined based on decreasing French ancestry, or populations that the authors think to be of 'atypical' ancestry were removed (Narasimhan *et al.* 2018; Reich *et al.* 2009). The authors also define south ancient ASI populations based on Onge ancestry. It is known that modern-day Onges cluster separately from modern-day mainland South Asians (Basu *et al.* 2016). Also, coalescent modelling has revealed that Onges consistently have had a very low N_e for the last 50,000 years and, consequently, have been subjected to loss in genetic variation due to genetic drift (Mondal *et al.* 2016).

5. Discussion

The distribution and antiquity of the M HG suggest that the central and south Indian tribes separated from their African ancestors approximately 70,000 years before present (ybp) and probably took the southern exit route OOA. However, this is not adequate to draw inferences regarding the time of colonization of SAS by AMH. Historical evidence suggests that AMH has been inhabiting SAS since 55,000 ybp (Kennedy *et al.* 1987). It is possible that the present-day central and south Indian tribes are descendants of these early settlers. On the contrary, no such deep relationship with Africans is observed in the Indian YC lineages. The O2 YC HG prevalent among AA and DR tribes is a relatively new lineage (< 20,000 ybp) and does not coalesce directly with its African counterparts (A, B and E HGs). Moreover, O2 HG occurs in even higher frequencies among east Asians. On the other hand, the H1 HG (maximum frequency among Dravidian tribes) with a coalescent time >50,000 ybp is rare outside the Indian subcontinent, suggesting the antiquity of these populations. This contrasting pattern among MG and YC points towards subsequent mostly male-biased gene flow following initial colonization into the original SAS gene pool, which could have led to the loss of the original Africa-related Y chromosome signatures. YC HG frequencies also indicate complex evolutionary histories, possibly unique, for individual populations, which could be difficult to observe by grouping all the tribes speaking similar languages (AA or DB) together. The characteristic differences between the MG and YC HG also suggest that there has been at least two subsequent contacts with divergent AMH groups – one with people from Central Asia, West Asia and the Middle East and the other with people from east Asia. Using genome-wide data, four genetic components for mainland India and subsequent admixture among them was observed (Basu *et al.* 2016) and the contact between ANI and ASI populations was dated to around 4000 ybp (Moorjani *et al.* 2013). The TB speakers from northeastern India were observed to be different from the rest of India. Ancient DNA suggests that by 3000 ybp gene flow from Iran into northwest SAS had taken place and most Indian populations contain a large proportion (among possible ancient ancestries from Central, west Asia, Middle East and western Siberia) of this ancestry.

As can be observed from the preceding section, several factors confound easy conclusion from the observed pattern in genetic diversity. There are several aspects of SAS population genetics that we need to understand better: (1) Who

are the descendants of the original AMH colonizers of SAS – the present-day AA-speaking tribes or the DR-speaking tribes? (2) What is the relationship between the present day AA- and DR-speaking tribes? (3) When did the ancestors of the present-day TB populations colonize SAS? (4) When did the different populations diverge from each other? (5) How have the sizes of the different SAS populations changed over time and why?

All analyses carried with genome-wide data so far are based on allele sharing. Two populations that had diverged a long time ago may have similar allele frequencies depending on the amount and duration of subsequent admixture between them (Kimura and Weiss 1964). In such cases, although allele sharing may suggest close similarity between them, evolutionarily they could in fact be quite divergent. Therefore, one needs to look at the size of the haplotypes shared among individuals in order to obtain a better understanding of their relationship. Also, coalescent approaches could facilitate better understanding of the peopling and population structure of SAS. A coalescent approach is the only way to analyse and model complex demographic histories, likely scenarios for most SAS indigenous populations. Also aDNA, properly analysed, could give us a better glimpse into our past.

References

Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, Dey B, Roy M, Roy B and Bhattacharyya NP 2003 Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res.* **13** 2277–2290

Basu A, Sarkar-Roy N and Majumder PP 2016 Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. *Proc. Natl. Acad. Sci. USA* **113** 1594–1599

Cann RL, Stoneking M and Wilson AC 1987 Mitochondrial DNA and human evolution. *Nature* **325** 31–36

Chaubey G, Tamang R, Pennarun E, Dubey P, Rai N, Upadhyay RK, Meena RP, Patel JR, van Driem G and Thangaraj K 2017 Reconstructing the population history of the largest tribe of India: The Dravidian speaking Gond. *Eur. J. Hum. Genet.* **25** 493

Forster P, Torroni A, Renfrew C and Röhl A 2001 Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol. Biol. Evol.* **18** 1864–1881

Government of India 2013 Statistical Profile of Scheduled Tribes in India. Ministry of Tribal Affairs. Statistical Division

Hammer MF 1995 A recent common ancestry for human Y chromosomes. *Nature* **378** 376

Kennedy KA, Deraniyagala SU, Roertgen WJ, Chiment J and Disotell T 1987 Upper Pleistocene fossil hominids from Sri Lanka. *Am. J. Phys. Anthropol.* **72** 441–461

Kimura M and Weiss GH 1964 The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49** 561

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS and Dixon ME 1999 Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr. Biol.* **9** 1331–1334

Kosambi DD 1965 *The culture and civilisation of ancient India in historical outline* (London: Routledge & Kegan Paul)

Majumder PP and Basu A 2015 A genomic view of the peopling and population structure of India. *Cold Spring Harbor Perspect. Biol.* **7** a008540

Mondal M, Casals F, Xu T, Dall'Olio GM, Pybus M, Netea MG, Comas D, Laayouni H, Li Q and Majumder PP 2016 Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nat. Genet.* **48** 1066–1070

Moorjani P, Thangaraj K, Patterson N, Lipson M, Loh P-R, Govindaraj P, Berger B, Reich D and Singh L 2013 Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* **93** 422–438

Narasimhan VM, Patterson NJ, Moorjani P, Lazaridis I, Mark L, Mallick S, Rohland N, Bernardos R, Kim AM and Nakatsuka N 2018 The genomic formation of south and central Asia. *bioRxiv* 292581

Oppenheimer S 2004 *Out of Eden: The peopling of the world* second edition (London, UK: Constable)

Oppenheimer S 2012 Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **367** 770–784

Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L and Hofreiter M 2004 Genetic analyses from ancient DNA. *Annu. Rev. Genet.* **38** 645–679

Reich D, Thangaraj K, Patterson N, Price AL and Singh L 2009 Reconstructing Indian population history. *Nature* **461** 489

Sengupta S, Zhivotovsky LA, King R, Mehdi S, Edmonds CA, Chow C-ET, Lin AA, Mitra M, Sil SK and Ramesh A 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

South Asian Regional Development Gateway. South Asia Regional Overview. *https://web.archive.org/web/20081121043924/http://www.sardeg.org/marketana.asp*

Thomson R, Pritchard JK, Shen P, Oefner PJ and Feldman MW 2000 Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* **97** 7360–7365

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM and Wallace DC 1993 Asian affinities and continental radiation of the four founding native American mtDNAs. *Am. J. Hum. Genet.* **53** 563

Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T and Wallace DC 1994 Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am. J. Phys. Anthropol.* **93** 189–199

Underhill PA and Kivisild T 2007 Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu. Rev. Genet.* **41** 539–564

Van Oven M and Kayser M 2009 Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30** E386–E394

Worldometers Population. Population of southern Asia. *http://www.worldometers.info/world-population/southern-asia-population/*