Ind. J. Phys. Anthrop. & Hum. Genet. Vol. 34. No. 2, (2015) : 193-199

EAST ASIAN ANCESTRY IN INDIA

Gyaneshwer Chaubey

ABSTRACT

It has been suggested that Indian populations constitute the second largest diversity of human populations just after Africa. Based on the mitochondrial DNA evidence it was suggested that the prehistoric South Asia (including Southeast Asia) carried half of the world populations 20-40 thousands years ago. Such scenario led some to conclude a complex and deep demographic history of the subcontinent with minor gene flow from East and the West. The relatedness of India to Central Asian, European, Middle Eastern, the Caucasus and to the East/Southeast Asians has been suggested, but none of the study has estimated the East/Southeast Asian ancestry among the extent population of India. Here the analysis of genome wide data on Indian and East/Southeast Asian demonstrated their restricted distinctive ancestry in India mainly running along the foothills of Himalaya and northeastern part. Moreover, we have also identified the consistency of East/Southeast Asian ancestry over the population history of the subcontinent leading the entry of Austroasiatic and Tibeto-Burman languages.

INTRODUCTION

Studies on Indian populations by various disciplines yielded contrasting results (Chaubey *et al.,* 2007; Petraglia *et al.*, 2007; Anthony 2009; Reich *et al.*, 2009; Xing *et al.,* 2010; Metspalu *et al.,* 2011; Moorjani *et al.,* 2013). In light of limited archaeological evidence and the limitations of linguistics, molecular anthropology is another independent line of evidence which may resolve such complexity; however, different genetic systems are also not in congruent. The mitochondrial DNA (mtDNA) overwhelmingly support an early arrival of modern humans to the subcontinent with later minor admixture from East and West (Metspalu *et al.,* 2004; Thangaraj *et al.,* 2006; Metspalu *et al.,* 2004; Sun *et al.,* 2006; Chandrasekar *et al.,* 2009). The Y chromosome studies have identified frequent autochthonous lineages (haplogroups C5, F*, H, L and R2) and also lineages with either Middle Eastern (haplogroup J2) or East/Southeast Asian (haplogroup O2a and O3a) or of unknown origins (haplogroup R1a) (Sahoo *et al.,* 2006; Sengupta *et al.,* 2006; Underhill *et al.,* 2015; Singh *et al.,* 2016; Chaubey *et al.,* 2016). On the other hand, the autosomal

Dr. Gyaneshwer Chaubey, Evolutionary Biology Group, Estonian Biocentre, Riia23, Tartu, Estonia-51010, *E-mail: gyanc@ebc.ee*

studies have identified two major components: one largely restricted to the subcontinent, whilst second is shared with Central Asia, the Caucasus, Middle East and Europe (Reich *et al.,* 2009; Metspalu *et al.,* 2011). This striking sharing was in coherence with the historic linguistics (Anthony, 2009), nevertheless the haplotype based analysis ruled it out by showing its dispersal timeline to be before Neolithic (Metspalu *et al.,* 2011).

Along with these two major components, studies on autosomes also showed minor components restricted to some populations which are however prevalent in East and West Eurasia (Chaubey *et al.,* 2011; Chaubey *et al.,* 2015). One of these components is present among Indian Austroasiatic (Munda) speakers which are thought to be associated with the arrival of Austroasiatic speakers from Southeast Asia to India (Chaubey *et al.,* 2011). The arrival of Austroasiatic speakers from Southeast Asia to India is also attested with the frequent Southeast Asian haplogroup O2a (M95) among Indian Munda speakers (Chaubey *et al.,* 2011). The virtual absence of East/Southeast Asian specific maternal lineages among Munda speakers helped researchers to conclude that the Austroasiatic (Munda) migration to India was mainly male mediated (Chaubey *et al.,* 2011). Another Austroasiatic (Khasi-Khumic) speaker of India, Khasi doesn't show sex specific admixture as Munda (Reddy *et al.,* 2007). Tibeto-Burman speakers of India also corroborate their affinity with the East/Southeast Asia (Metspalu *et al.,* 2004). The presence of distinct East/Southeast Asian specific ancestry among Indian Austroasiatic and Indian Tibeto-Burmans indicate their different population histories in the subcontinent (Chaubey *et al.,* 2014).

It is clear that most of the East/Southeast Asian ancestry to India is arrived via language dispersal (Diamond and Bellwood, 2003) and other local demographic histories (Thangaraj *et al.,* 2008; Chaubey *et al.,* 2014), but it is a question that at what extant it is diffused in to the subcontinent beyond its linguistic boundaries? To investigate this issue we have calculated the East/Southeast Asian ancestry among thirty Indian populations belong to different linguistic groups and two Negrito populations from Andaman Island.

MATERIAL AND METHODS

In this study, we have merged samples from three different studies (HUGO Pan-Asian SNP Consortium *et al.,* 2009; Reich *et al.,* 2009; International HapMap 3 Consortium *et al.,* 2010). Average IBS (identity by state) was calculated (Purcell *et al.,* 2007), and related individuals upto three generations were removed from the analysis. We used PLINK 1.07 (Purcell *et al.,* 2007) to filter the combined data set to include only SNPs on the 22 autosomal chromosomes with minor allele frequency >1% and genotyping success >99%. Our combined dataset had data for 12,622 SNPs, after excluding SNPs unique to any of the three platforms and SNPs from mtDNA and X and Y chromosomes. To calculate the East Asian specific ancestry among Indian populations we have used the *f*4 ratio estimation test (Patterson *et al.,* 2012). We used European-CEU as an out group and calculated the Han ancestry among

the Indian populations; East Asian ancestry = *f*4 (Japan, Kurumba; X, CEU)/*f*4 (Japan, Kurumba; Han, CEU), where X represent Indian populations. We plotted the % of ancestry over the geographical map of India (Fig. 1).

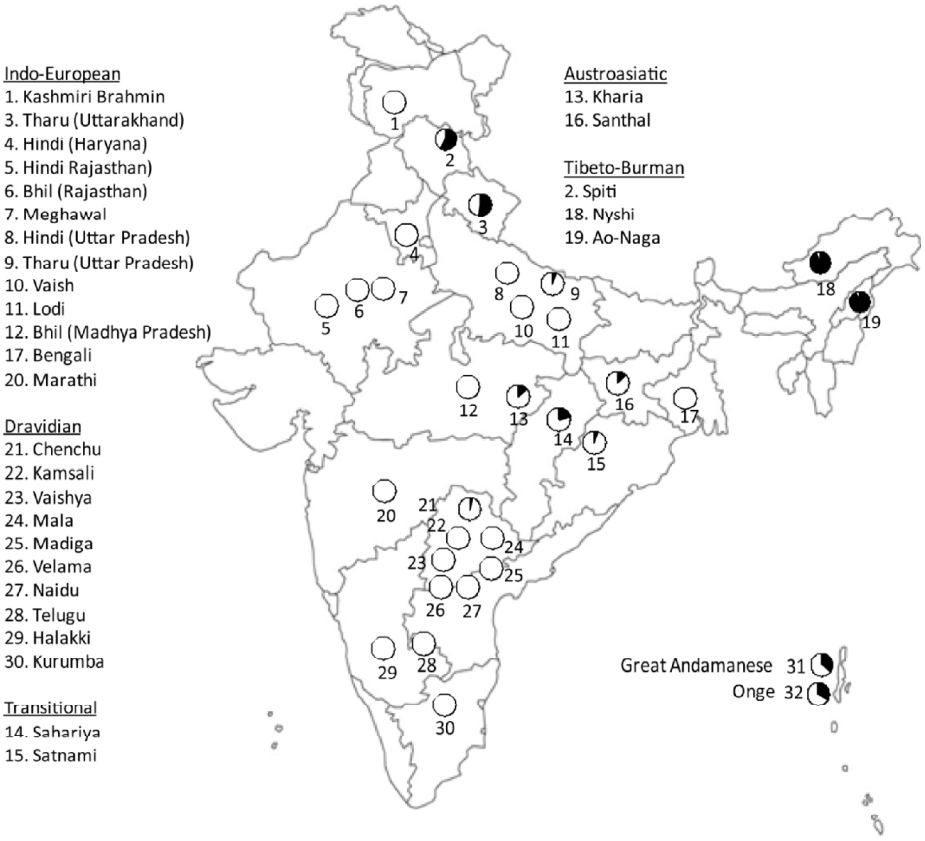

Figure 1: The pie chart, showing the geographical distribution of East/Southeast Asian ancestry among various Indian populations

RESULTS AND DISCUSSION

India is linguistically, genetically, culturally and geographically a highly heterogeneous country. The highly ripped caste and tribal groups of India are unique among all known societies in human history (Chaubey *et al.*, 2007). This uniqueness is the result of long term high effective population size and admixture to and from East and West. In the present study we have quantified the East/ Southeast Asian gene flow among several populations of India belonging to different language groups.

We observed a high frequency of East/Southeast Asian specific ancestry over the lower foothills of Himalaya and towards northeast India (Fig. 1). Nishi, a Tibeto-Burman population from Arunanchal Pradesh carried the highest amount (93%)

of East/Southeast Asian ancestry, followed by the Ao-Naga (92%) from Nagaland belongs to the same language family. The highest frequency of East/Southeast Asian ancestry among Tibeto-Burman populations support their recent arrival from East (Cordaux *et al.,* 2004; Metspalu *et al.,* 2004). Notably, the Negrito populations from Andaman Island also showed a high level of East/Southeast Asian ancestry (Fig. 1). Our previous analyses on these two populations have observed three distinct ancestry components and have suggested a deep common ancestry of Andaman Negrito with the Melanesia, Malaysian Negrito and South Asia (Chaubey and Endicott, 2013). The Han ancestry measured in Andaman Negrito is probably partially capturing both the Melanesian and Malaysian Negrito ancestry.

The Transitional and Munda groups were also harbouring a substantial amount of East/Southeast Asian ancestry (Fig. 1). Consistent with the higher genome sharing with East/Southeast Asia, South Munda (Kharia) constituted 21% of the East/Southeast Asian ancestry, in comparison with North Munda (Santhal 13%). Saharia a Transitional group, who presently speak Indo-European language, carried 15% of the East/Southeast Asian ancestry. Satnami another transitional group speaking Indo-European language carried 5% of the East/Southeast Asian ancestry (Fig. 1). It is interesting to note that the Transitional populations share the same geography as their Austroasiatic neighbours and it is highly likely that the East/Southeast Asian ancestry among them is a result of gene flow or language shift.

Among the large number of Indo-European and Dravidian speakers the East/Southeast Asian component was largely absent except in Tharu and Chenchu (Fig. 1). The case of Tharu ancestry have been already discussed elsewhere in detail (Thangaraj *et al.,* 2008; Chaubey *et al.,* 2014), whereas there is no historic information about language shift among Chenchu, who presently speak Dravidian language. The presence of 4% (though insignificant $Z < 2$) of East/Southeast Asian ancestry among Chenchu is intriguing. It is likely that they may have received minor gene flow from the Austroasiatic populations living in neighbouring state, which is evident in their maternal profile (Endicott *et al.,* 2006). Moreover, the East/Southeast Asian specific Y chromosome lineages were not observed among Chenchu in haploid DNA analysis (Kivisild *et al.,* 2003).

In conclusions, considering the overall distribution of East/Southeast Asian ancestry in India, it is clear that due to the high level of endogamy the diffusion of this ancestry component is limited with the exception of the populations who have known introgression from the population associated with the Austroasiatic or Tibeto-Burman e.g. Tharu, Saharia and Satnami. The spatial distribution of this component largely mimics the spread of Austroasiatic, Tibeto-Burman languages; Y chromosomal haplogroups O2a and O3a; and autosomal EDAR 370A variant (Chaubey *et al.,* 2011). Therefore, taking into consideration the demographic histories of the Indian populations, the distribution of East/Southeast Asian ancestry in India predominantly follow linguistic landscape model followed by 'isolation by distance' model.

ACKNOWLEDGEMENTS

This study is supported by Estonian Personal grants PUT-766 (GC). GC also acknowledges the financial support from European Union European Regional Development Fund through the Centre of Excellence in Genomics to Estonian Biocentre and University of Tartu by Tartu University grant (PBGMR06901), and Estonian Institutional Research grants IUT24-1.

REFERENCES

Anthony DW 2009. The horse, the wheel, and language: how Bronze-Age riders from the Eurasian steppes shaped the modern world. Princeton University Press.

Chandrasekar A, Kumar S, Sreenath J, Sarkar BN, Urade BP, Mallick S, Bandopadhyay SS, Barua P, Barik SS, Basu D, Kiran U, Gangopadhyay P, Sahani R, Prasad BVR, Gangopadhyay S, Lakshmi GR, Ravuri RR, Padmaja K, Venugopal PN, Sharma MB and Rao VR. 2009. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PloS one.* 4(10) : e7447.

Chaubey G and Endicott P. 2013. The Andaman Islanders in a regional genetic context: reexamining the evidence for an early peopling of the archipelago from South Asia. *Human biology.* 85(1-3): 153-172.

Chaubey G, Kadian A, Bala S and Rao VR. 2015. Genetic Affinity of the Bhil, Kol and Gond Mentioned in Epic Ramayana. *PloS one.* 10(6): e0127655.

Chaubey G, Metspalu M, Choi Y, Mägi R, Romero IG, Soares P, van Oven M, Behar DM, Rootsi S, Hudjashov G, Mallick CB, Karmin M, Nelis M, Parik J, Reddy AG, Metspalu E, van Driem G, Xue Y, Tyler-Smith C, Thangaraj K, Singh L, Remm M, Richards MB, Lahr MM, Kayser M, Villems R and Kivisild T. 2011. Population Genetic Structure in Indian Austroasiatic speakers: The Role of Landscape Barriers and Sex-specific Admixture. *Mol. Biol. Evol.* 28(2): 1013-1024.

Chaubey G, Metspalu M, Kivisild T and Villems R. 2007. Peopling of South Asia: investigating the caste-tribe continuum in India. *Bioessays.* 29(1): 91-100.

Chaubey G, Singh M, Crivellaro F, Tamang R, Nandan A, Singh K, Sharma VK, Pathak AK, Shah AM, Sharma V, Singh VK, Selvi Rani D, Rai N, Kushniarevich A, Ilumäe AM, Karmin M, Phillip A, Verma A, Prank E, Singh VK, Li B, Govindaraj P, Chaubey AK, Dubey PK, Reddy AG, Premkumar K, Vishnupriya S, Pande V, Parik J, Rootsi S, Endicott P, Metspalu M, Lahr MM, van Driem G, Villems R, Kivisild T, Singh L and Thangaraj K. Unravelling the distinct strains of Tharu ancestry. *Eur. J. Hum. Genet.* 2014; 22(12): 1404-1412.

Chaubey G, Singh M, Rai N, Kariappa M, Singh K, Singh A, Pratap Singh D, Tamang R, Selvi Rani D, Reddy AG, Kumar Singh V, Singh L and Thangaraj K. 2016. Genetic affinities of the Jewish populations of India. *Scientific reports.* 6: 19166.

Cordaux R, Weiss G, Saha N and Stoneking M. 2004. The northeast Indian passageway: a barrier or corridor for human migrations? *Mol. Biol. Evol.* 21(8): 1525-1533.

Diamond J and Bellwood P. 2003. Farmers and their languages: the first expansions. *Science.* 300 (5619): 597-603.

Endicott P, Metspalu M, Stringer C, Macaulay V, Cooper A and Sanchez JJ. Multiplexed SNP typing of ancient DNA clarifies the origin of andaman mtDNA haplogroups amongst south Asian tribal populations. *PLoS ONE.* 2006; 1:e81.

HUGO Pan-Asian SNP Consortium, Abdulla MA, Ahmed I, Assawamakin A, Bhak J, Brahmachari SK, Calacal GC, Chaurasia A, Chen CH, Chen J, Chen YT, Chu J, Cutiongco-de la Paz EMC, De Ungria MCA, Delfin FC, Edo J, Fuchareon S, Ghang H, Gojobori T, Han J, Ho SF, Hoh BP, Huang W, Inoko H, Jha P, Jinam TA, Jin L, Jung J, Kangwanpong D, Kampuansai J, Kennedy GC, Khurana P, Kim HL, Kim K, Kim S, Kim WY, Kimm K, Kimura R, Koike T, Kulawonganunchai S, Kumar V, Lai PS, Lee JY, Lee S, Liu ET, Majumder PP, Mandapati KK, Marzuki S, Mitchell W, Mukerji M, Naritomi K, Ngamphiw C, Niikawa N, Nishida N, Oh B, Oh S, Ohashi J, Oka A, Ong R, Padilla CD, Palittapongarnpim P, Perdigon HB, Phipps ME, Png E, Sakaki Y, Salvador JM, Sandraling Y, Scaria V, Seielstad M, Sidek MR, Sinha A, Srikummool M, Sudoyo H, Sugano S, Suryadi H, Suzuki Y, Tabbada KA, Tan A, Tokunaga K, Tongsima S, Villamor LP, Wang E, Wang Y, Wang H, Wu JY, Xiao H, Xu S, Yang JO, Shugart YY, Yoo HS, Yuan W, Zhao G, Zilfalil BA and Indian Genome Variation Consortium. Mapping human genetic diversity in Asia. *Science.* 2009; 326(5959): 1541-1545.

International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Bonnen PE, de Bakker PIW, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarrol SA, Nemesh J, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Gonzaga-Jauregui C, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Zhang Q, Ghori MJR, McGinnis R, McLaren W, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD and McEwen JE. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010; 467(7311): 52-58.

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Gölge M, sanga E, Papiha SS, Cinnioğlu C, King R, Cavalli-Sforza L, Underhill PA and Villems R. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* 2003; 72(2): 313-332.

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MTP, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A and Villems R. 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.* 2004; 5: 26.

Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, Hudjashov G, Nelis M, Mägi R, Metspalu E, Remm M, Pitchappan R, Singh L, Thangaraj K, Villems R and Kivisild T. Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am. J. Hum. Genet.* 2011; 89(6): 731-744.

Moorjani P, Thangaraj K, Patterson N, Lipson M, Loh PR, Govindaraj P, Berger B, Reich D and Singh L. Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* 2013; 93(3): 422-438.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T and Reich D. Ancient admixture in human history. *Genetics.* 2012; 192(3): 1065-1093.

Petraglia M, Korisettar R, Boivin N, Clarkson C, Ditchfield P, Jones S, Koshy J, Lahr MM, Oppenheimer C, Pyle D, Roberts R, Schwenninger JL, Arnold L and White K. Middle

Paleolithic assemblages from the Indian subcontinent before and after the Toba super-eruption. Science. 2007; 317(5834): 114-116.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ and Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007; 81(3): 559-575.

Reddy BM, Langstieh BT, Kumar V, Nagaraja T, Reddy ANS, Meka A, Reddy AG, Thangaraj K and Singh L. Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. *PLoS ONE.* 2007; 2(11): e1141.

Reich D, Thangaraj K, Patterson N, Price AL and Singh L. Reconstructing Indian population history. *Nature.* 2009; 461(7263): 489-494.

Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, Gaikwad S, Trivedi R, Endicott P, Kivisild T, Metspalu M, Villems R and Kashyap VK. A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc. Natl. Acad. Sci. USA.* 2006; 103(4): 843-848.

Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, Lin AA, Mitra M, Sil SK, Ramesh A, Usha Rani MV, Thakur CM, Cavalli-Sforza LL, Majumder PP and Underhill PA. 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.* 2006; 78(2): 202-221.

Singh S, Singh A, Rajkumar R, Sampath Kumar K, Kadarkarai Samy S, Nizamuddin S, Singh A, Ahmed Sheikh S, Peddada V, Khanna V, Veeraiah P, Pandit A, Chaubey G, Singh L and Thangaraj K. Dissecting the influence of Neolithic demic diffusion on Indian Y-chromosome pool through J2-M172 haplogroup. *Scientific reports.* 2016; 6: 19157.

Sun C, Kong QP, Palanichamy MG, Agrawal S, Bandelt HJ, Yao YG, Khan F, Zhu CL, Chaudhuri TK and Zhang YP. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol. Biol. Evol.* 2006; 23(3): 683-690.

Thangaraj K, Chaubey G, Kivisild T, Selvi Rani D, Singh VK, Ismail T, Carvalho-Silva D, Metspalu M, Bhaskar LV, Reddy AG, Chandra S, Pande V, Prathap Naidu B, Adarsh N, Verma A, Jyothi IA, Mallick CB, Shrivastava N, Devasena R, Kumari B, Singh AK, Dwivedi SK, Singh S, Rao G, Gupta P, Sonvane V, Kumari K, Basha A, Bhargavi KR, Lalremruata A, Gupta AK, Kaur G, Reddy KK, Rao AP, Villems R, Tyler-Smith C and Singh L. Maternal footprints of Southeast Asians in North India. *Hum. Hered.* 2008; 66(1): 1-9.

Thangaraj K, Chaubey G, Singh VK, Vanniarajan A, Thanseem I, Reddy AG and Singh L. In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. *BMC Genomics.* 2006; 7: 151.

Underhill PA, Poznik GD, Rootsi S, Järve M, Lin AA, Wang J, Passarelli B, Kanbar J, Myres NM, King RJ, Di Cristofaro J, Sahakyan H, Behar DM, Kushniarevich A, Sarac J, Saric T, Rudan P, Pathak AK, Chaubey G, Grugni V, Semino O, Yepiskoposyan L, Bahmanimehr A, Farjadian S, Balanovsky O, Khusnutdinova EK, Herrera RJ, Chiaroni J, Bustamante CD, Quake SR, Kivisild T and Villems R. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur. J. Hum. Genet.* 2015; 23(1): 124-131.

Xing J, Watkins WS, Hu Y, Huff CD, Sabo A, Muzny DM, Bamshad MJ, Gibbs RA, Jorde LB, and Yu F. Genetic diversity in India and the inference of Eurasian population expansion. *Genome Biol.* 2010; 11(11): R113.