Chapter 7

Summary & Conclusions
Summary & Conclusion

The human genetic variation is to decipher the amount, pattern, distribution and structuring of genetic diversity across different geo-ethnic, socio-cultural and linguistic human groups. Then, assessing the information of the genetic structure or ancestry of a population to comprehend the disparity in human susceptibility to diseases, differential human response to pharmacological agents and the complex interface of genetic and environmental factors in producing diverse phenotypes. The genetic variation studies starts with deducing the genetic profile of an individual or a population and then carve out a perspective correlation of this genetic profile with the past human movements, the palenteological evidences and the socio-cultural restrictions to offer strong interpretation about the causes and consequences of the genetic variation. The causes of variation includes a glut of evolutionary forces that infuses change in the nucleotide sequence of DNA and then reject or select this change based on the viability and reproducibility of the individual carrying the changed or ‘mutated’ sequence. However, when the population under study belongs to Indian sub-continent, then the causes, amount, pattern and distribution of genetic variation acquire a highly complex view. Reasons for this complexity are many
folds; one of them is the crucial geographical location of India that lies on the postulated southern coastal route followed by the anatomically modern *H. sapiens* out of Africa. Secondly, the impending role of extensive gene flow through a series of migrations and invasions, that have shaped and distributed the contemporary genetic variation across different geographical locations of India and has created enormous amount of genetic diversity. Finally, a unique social framework fabricated with the threads of numerous religions, communities, castes and linguistic groups that have resulted into extensive population structuring and creation of numerous of endogamous groups. Overall, the genetic structure, affinity and diversity of the 1 billion Indians is often contested and postulated of holding an important key about numerous of unanswered questions concerning the evolution of modern human and a range of factors that shapes the contemporary pattern of genetic variation.

Present study was conceived with the aim of exploring the genetic composition of four tribal population of Jharkhand and there by offering an unsoiled and immaculate interpretation on the pattern, distribution and structuring of genetic variation existing in East Indian populations. The four populations selected included two are belong to Austro Asiatic linguistic family and two consanguineous sects of Dravidian and Indo-European linguistic group Oraon and Paharia belongs to this group. Among the two Austro Asiatic linguistic families, one is “Munda” and another is “Birhor”. All the four populations are ethnically Austroloid and maintain endogamy and it is consider to be the initial settler of India.

In order to dissect maximum genetic information about the population concerned. We have analysed 32 unique event polymorphisms (UEPs) and 4 STR loci located on non-recombining region of Y-chromosome (NYR) to delineate the paternal lineages.

At first, venous blood was collected from randomly selected 800 samples belonging to four populations were collected from different regions of Jharkhand state. Majority of the samples were male (*n* = 419), in order to facilitate a larger sample size for Y-chromosome analysis. At first DNA was extracted from all 419 male samples to delineate the paternal lineage. The findings of Y-chromosome analysis unravel some anonymous facts about the paternal lineages of the four
studied populations and the comparison of the Y-genealogy of the east Indian groups with that of other contemporary Indian and global populations have revealed some ravishing features about the larger degree of historical gene flow into India and a congruence of genetic and geographical/socio-cultural affinities of Indian populations. A total of 419 Y-chromosomes belonging to four Jharkhand tribal populations were analyzed for a set of 32 binary Unique Event Polymorphisms - UEPs (28 SNPs, 3 indels and one retro-transposition element) by sequencing and RFLP based analysis to designate an individual’s Y-chromosome in any of the 18 haplogroups proposed in the YCC binary tree. In addition, 4 YSTR loci were also genotyped by fragment size analysis to ascertain the level of diversity and presence of different lineages within a haplogroup in different populations.

Synthesis of Y-chromosomal haplogroup tree defined by 32 UEPs in four Jharkhand tribal populations exemplified the occurrence of 12 different haplogroups ("C", "F*", "G", "H", "J2", "K*", "L", "O", "P*", "R1a", "R1b" and "R2"), of which 9 ("C", "F*", "H", "J2", "K*", "O", "P*", "R1a", and "R2") occurs in all the four populations. The most frequent clad comprised of 5 haplogroups -"H", "O" "R1a", "R2" and "J2", that together accounts for ~84% of the total Y-chromosomes analyzed. The less frequent clad include 7 distinct haplogroups namely "C", "F*", "G", "K*", "L", "P*" and "R1b" that together constituted 16.0% of the total Y-chromosomes. The Y-genealogy demonstrated that east Indian populations carry two different Y-lineages, derived from (i) Central Asia or west Eurasia (R1a, R1b and R2 and some of the J2 lineages) and (ii) indigenous Indian Y-lineage (H, F*, O and one Brahmin individual carrying C without M217 transversion).

The analysis of 4 YSTR loci have delineated total 106 Y haplotypes, numerous of which were observed in single individuals. Maximum 53 haplotypes were observed in Paharia’s and minimum 17 haplotypes were seen in Munda tribes. It is interesting to note that Paharia tribes exhibited maximum number of unique haplotype (~69% of the total observed haplotype) and on the other hand Munda tribes exhibited maximum number of shared haplotype (~70%) signifying closer genetic relation ship of Munda tribes with other tribal population than Paharia tribes. The two tribes Oraon and Paharia shows more genetic similarity with each other.
discernible by nearly equal number of shared and unique haplotype. The analysis of
the YSTR haplotype sharing was carried out between individuals of a specific binary
haplogroup. It was noted that most of the individual sharing YSTR haplotypes also
carry same Y-binary haplogroup. This suggests that structuring of YSTR haplotype
is in accordance with the Y-binary haplogroup distribution.

Among the six Y-STR haplotype shared by the all the four population, three
were exclusively found in haplogroup “O”, two haplotype (15-22-10-12 and 16-22-
10-12) were specific to haplogroup “H” while remaining one were found exclusively
in haplogroup “R” lineage. Similar pattern of parallel structuring of YUEP-
haplogroup and YSTR-haplotype were observed in all the populations where elite Y-
UEPs lineages carry explicitly unique YSTR haplotype while frequently distributed
Y-UEPs haplogroups “O” also carried shared haplotypes. Sharing of some of the
haplotypes was also observed between different haplogroups like 15-22-10-12 was
found in both H and R2 clad. Such sharing were rare and could have been because
of haplotype motif which were based on only four YSTR loci and if more markers
would have been analyzed then these shared YSTR lineages might have also get
segregated. Some of the important haplotypes found in specific Y-UEP defined
haplogroups are as follows:

A vital outcome of the Y-chromosome analysis was the robust signals of each
of the movements in north India starting from earliest hominids that reached Indian
through southern coastal route out of Africa (haplogroup H, F*, C without M217
transversion and O), to the farmers of Neolithic agriculture expansion (haplogroup L,
R2, G and some of the J2 lineages), Indo Aryan speaking nomads from central Asian
steppes (haplogroup R1, R2 and J2) also contribute in Jharkhand tribes geneology.

At last, the geographic cline of genetic affinities of north, south, east and
west Indians with Eurasians was found more pragmatic in Y-genealogy as clearly the
native Indian haplogroups were more pronounced in south and east Indians.
Caucasian haplogroups symbolizing the Indo-Aryan dominance were more distinct
in the north Indians.

Conclusively, the combined picture represented by the empirical results of
the high-resolution analysis of paternal lineage of human genome infers that (i) The
Jharkhand tribes are genetically highly diverse people with most of the variation scattered between individuals. (ii) The genetic differentiation does exist between endogamous groups across Indian mainland but the differentiation is mainly configured geographically and linguistically, and the stringent socio-cultural norms have only a trivial contribution. (iii) The advents of migrations and invasions in northeastern corridor of India have more genetic evidences in the paternal lineages. (iv) Despite of experiencing a glut of human migrations, the paternal genealogy of east Indians still sustains the records of each of the migratory episodes and (v) The legacy of Austro-Asiatic speakers is discernible in the gene pool of each of the studied Jharkhand tribal population (vi) The Y-UEPs and Y-STR genetic diversity distribution suggests that although four Jharkhand tribal populations have high genetic similarity.

Overall, the genetic configuration of Indians is as complex the history of Indian mainland, interwoven in numerous threads of unknown facts. More genetic data from this part of the world and other critically important regions like Afghanistan, Iran and Iraq is highly necessitated in tracing the missing blocks of the causes and consequences of human genetic variation. Present study has provided some clues that might help in unwinding the complex interwoven threads of Indian tribal genetic composition.