CONCLUSIONS
6. Conclusions

Present study deals with the identification and characterization of the mRNA transcripts tagged with the simple repeats of the GACA, GATA and 33.15 repeats in water buffalo as a model system, which unveiled the differential organization and expression of these transcripts among different tissues and spermatozoa. Moreover, the detailed isolation and characterization of the full length Smoc-1 and \textit{c-kit} genes were also performed to gain insight into their structural and functional organization, expressional status and chromosomal localization. The following points summarize the outcome of this work highlighting the novel potential implications of the MASA uncovered mRNA transcripts in the spermatogenesis, fertilization and several other regulatory pathways:

1. The \textit{in-silico} distributional analyses of the GACA and GATA repeats in the coding and non-coding genomes of Archeas and 17 eukaryotes revealed total absence of these repeats in the prokaryotes, and their accumulation in the higher eukaryotes with an increase of their genetic complexities during evolution. This highlights the significance of these simple repeats in the genome evolution.

2. The analysis of chromosome-wise distribution for the GACA/GATA repeats highlights their preferential accumulation on the mammalian sex chromosomes which suggests their involvements in functional regulation of the sex determination.

3. MASA using the GACA, GATA and 33.15 repeats uncovered a total of 616 fragments, encompassing 148 with 33.15 repeat, 332 with GACA, and 136 with GATA, from somatic tissues, gonads and spermatozoa. This is a novel attempt revealing the existence of the 33.15, GACA and GATA-tagged transcripts in the buffalo spermatozoa highlighting their involvements during the pre- and post-fertilization events.

4. Characterization of the MASA uncovered fragments led to the identification of a total of 63 different mRNA transcripts (34, GACA-tagged; 10, GATA-tagged; and 19, 33.15-tagged) in water buffalo. This association of repeats with the mRNA transcripts, which can be study acquires considerable significance since it establishes the extrapolated
Conclusions

to establish the conclusive significance of other repeats within and adjacent to the coding regions.

5. Exclusive presence of several GACA-/33.15-tagged transcripts in a tissue or spermatozoa, and absence of the GATA-tagged transcripts in lung/heart highlights their differential transcript profiles. Several tissuespecific transcripts demonstrate their exclusive requirement in that tissue and their absence showed the transcriptional quiescence in other tissues.

6. The homology search established the novel status of about 50% of the GACA-/33.15-tagged and all the GATA-tagged transcripts corroborating their species-specific distribution. However, the transcripts showing homology to characterized genes were either involved in signal transduction or cell-cell interaction pathways indicating their crucial roles in various cellular functions essential for the life cycle of a cell.

7. Present study also established the GACA richness of the buffalo transcriptome while other species including human were observed to be GATA rich. The GC-richness seems to be unique for the buffalo genome organization and thus for replication timings, methylation and gene expression.

8. All these uncovered mRNA transcripts showed faithful evolutionary conservation across thirteen different species, suggesting broader significance of the 33.15 and GACA/GATA repeats and their tagged transcripts in eukaryotes.

9. Of all the transcripts, approximately 35% demonstrated inter-tissue and/or tissue-spermatozoal sequence polymorphisms which were confirmed in 5-10 additional animals. This may be explained either towards their various functions in different tissues and spermatozoa, or differential functions at various stages of development.

10. The quantitative expressionional studies demonstrated the uniform expression of about 30% GACA-tagged in all the sources, whereas 10% GACA-tagged and 15% 33.15 tagged transcripts with highest expression in the liver or spleen indicated their putative involvement in the hepatocellular and immunological activities, respectively.
11. Most interestingly, the exclusive or highest expression of 60% GACA-tagged, 85% 33.15-tagged, and 100% GATA-tagged transcripts in the testis and/or spermatozoa substantiating their deep involutions in various testicular functions like spermatogenesis and male gonad development.

12. The full length CDS of proto-oncogene c-kit (2973 bp) from different tissues, and Secreted modular calcium binding protein-1 (3474 bp) from liver, of water buffalo were isolated.

13. Upon comparison, the c-kit sequences showed tissue-specific nucleotide insertions, deletions and changes resulting in novel truncated peptides. These peptides lacked intracellular and/or transmembrane domains in all other tissues. However, only testis was found to encode full length c-kit protein which highlighted its tissue and stage specific functions.

14. C-kit, implicated with spermatogenesis, melanogenesis and hematopoeisis, was found to undergo tissue-specific alternate splicing. These alternately spliced transcripts were the integral parts of the open reading frame and have been reported in other mammals.

15. Multiple sequence alignment of c-kit sequences across the mammals revealed a unique tyrosine kinase domain in buffalo compared to that in other species, suggesting the species-specific organization and function of c-kit.

16. The expressional analysis of c-kit unveiled its highest expression in testis, and 10 times lesser in spermatozoa compared to that in testis which substantiates its predominant role in spermatogenesis. This study establishes unequivocal involvement of an autosomal gene c-kit receptor in testicular functions.

17. Present study demonstrates the association of the consensus sequence of minisatellite 33.15 with the Smoc-1 which was found to encode a secreted matricellular glycoprotein containing two EF-hand calcium binding motifs homologous to that of BM-40/SPARC family which suggest their property of calcium binding.

18. This gene consisting of 12 exons was mapped onto the acrocentric chromosome 11 in buffalo. Though this gene was found to be
Conclusions

evolutionarily conserved, the buffalo Smoc-1 showed conspicuous nucleotide/amino acid changes altering its secondary structure compared to that in other mammals. This suggests their species-specific organizational and functional uniqueness.

19. Two EF-hand motifs in the ECD conformed well to its calcium binding affinity and N-glycosylation site at Asn-214 suggesting its glycoprotein nature with a calcium dependent conformation.

20. For the first time, we unveiled two transcript variants of this gene, varying in their 3'UTR lengths but both coding for identical protein(s). Buffalo Smoc-1 is a single copy number gene, and thus, the presence of two variants of this gene may signify either for a backup of the transcripts if one is degraded/mutated or for the enhanced protein expression.

21. Buffalo Smoc-1 evidenced highest expression of both the variants in liver and modest to negligible in other tissues contrary to the earlier reports in mouse substantiating species and tissue-specific functions of this gene. The relative expression of variant-02 was markedly higher compared to that of variant-01 in all the tissues examined, highlighting the variant-02 as major transcript and variant-01 as minor one.

22. Moreover, the expression of Smoc-1, though moderate during the early ages, was conspicuously enhanced after 1 year age and remained consistently higher during the entire life span of buffalo, with the gradual increment in expression of variant-02, intimating its role in postnatal development besides embryonic development. Since Smoc-1 is thought to be involved in cell-matrix interaction and bone mineralization, its fulminant expression at 10 month age and beyond which signifies its requirement for growth, development and possible sustenance of the animal.