Conservation, management and documentation of livestock genetic resources and biodiversity: A biotechnology perspective

*Sher Ali and Seyed Ehtesham Hasnain**

For conservation and management of livestock genetic resources, in-depth molecular characterization of a species is required. This may encompass genome organization, sex-determination, regulation of gene expression, clad identification, breed delineation, assessment of QTL loci and their synteny in related species. Molecular methods are available that may be employed for genome analysis of farm animals as well as highly endangered species. We discuss various methodological approaches together with molecular basis of the development of genetic markers and marker systems in the context of genetic conservation.

THE conservation and management of India's immense genetic resources require great deal of effort and sustained multipronged approaches. Extensive survey, cataloguing, physical evaluation, in-depth characterization, preservation, maintenance, timely restoration and sustainable utilization of livestock genetic resources without impinging upon their delicate ecological balances and inter-clad interactions are important in addressing these issues. In the entire spectrum of cognate activity, in-depth characterization will involve many scientific approaches for which genetics and biotechnology will continue to play dominant roles.

The gigantic spectrum of biodiversity and the fast-changing world economic scenario coupled with politically ever-unfolding newer global order makes it imperative to take a serious view of our current bio-assets and unforeseen future liabilities. A marked number of breeds of cattle, buffalo, sheep, goat, pig, horse, camel, mithun, yak, dog, cat, poultry duck, geese, turkey, guinea fowl, pheasant and many more have been reported but most of them remain genetically uncharacterized. A battery of molecular markers and different experimental approaches are required not only for the genetic characterization of the species that are related to our economy but also to those that are part of our eco-system[1–4].

Startling as it may appear, but the fact remains that commensurate to the abundance of biodiversity, genetical data at every level is lacking. Much of genetical or biological information may be available on experimental animals such as rat, mouse and *Drosophila* but the same with respect to farm animals such as cattle, goat, sheep or mithun are rare. While this does not mean that studies on experimental animals in any way are less important, it does reflect lack of effort towards the understanding of the genetics of farm animals and other species that are part of critical mass of our biodiversity. The generation of detailed genetical information in turn would involve genome analysis with respect to sex determination, gene expression and its regulation, and the genomic organization and expression of the sequences involved in signal transduction. Similarly, studies on protooncogenes and their detailed characterization, molecular events leading to controlled cell proliferation and apoptosis would prove to be useful for undertaking advance studies on the desired species. Keeping in view the physical and phenotypic attributes, clad identification and their proper phylogenetic positioning[5,6], breed delineation[7], assessment of QTL loci and their synteny in related species, mutational studies[8–10] and comparative analysis of molecular events leading to modulation of signal transduction are other much-needed areas of foci in the animal species. Similarly, sex identification, analysis of the genes and molecular events involved in sex determination will be very important. Information on these lines would provide the base for subsequent manipulation of genomes related to transgenic experiments, even 'tailor-made' animals and eventually for the propagation of desired germplasm.

In order to achieve the above-mentioned goals, several experimental approaches can be adopted keeping in view the fact that a typical higher vertebrate genome is endowed with 3–4 billion haploid sequences representing at least 80,000 functional genes including a sizable part of non-coding repetitive DNA. Identification of gene(s) associated with genetic superiority of the species may be a challenging task but the same is not insurmountable owing

Sher Ali and Seyed Ehtesham Hasnain are in the National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110 067, India; Seyed Ehtesham Hasnain is also in the Centre for DNA Fingerprinting and Diagnostics, CCMB Campus, Uppal Road, Hyderabad 500 007, India.
*For correspondence. (e-mail: seyedhasnain@hotmail.com)

to the availability of many powerful techniques of molecular biology.

Recent studies have shown that a number of repetitive DNA belonging to different families are transcriptionally active, thereby, refuting the earlier contention that repetitive DNA is 'junk'. Repetitive DNA may in fact prove to be useful markers or may even assist in the development of marker systems[4,11–16]. The fact that an individual chromosome is a uninemic entity (made up of a single DNA molecule running all along the length of the chromosome), it becomes important to understand the organizational and expressional relationship of a functional gene in the context of organization of repetitive DNA. It is however, not clear if the intergenic distances with respect to different types of satellite DNA are always constant and that the similar types of spacer sequences are involved in the intergenic organization, at least in related species, if not in all the species. Information on the presence of different possible types of satellites in a species, their possible association/linkage within and between other related species will be desirable. Finally, a comprehensive synteny amongst all the desired species may be established. Such endeavour would in the long run provide working flexibilities to use marker(s) originating from one species in another. This may also help to provide an insight into the genetic susceptibility or possible resistance of a species to a disease.

In this context, the following technological approaches, though not necessarily in the same order, may be adopted and molecular genetical studies be directed towards some important species such as cow, buffalo, goat and sheep which can be subsequently, extended to other species.

To establish and ascertain different breeds of animals

A number of breeds based on physical attributes have been reported. This aspect may be substantiated scientifically for which a variety of standard experimental protocols are available. Monolocus and multilocus DNA profiling[1,3,7,17–22], random amplification of polymorphic DNA (RAPD), also known as arbitrarily primed polymerase chain reaction (AP-PCR) or more robust approach like minisatellite associated sequence amplification (MASA) may be used[23–25]. For the primers needed for RAPD or MASA, individual hypervariable repeat loci may be cloned, sequenced and screened for the presence of short tandem repeat motifs (STRs) endowed therein. These STRs in the range of 6–18 base residues may be used for RAPD or MASA studies. Some of the clones generated in the process may uncover high level of genetic polymorphism and may prove to be equally useful for routine DNA fingerprinting. A STR-based probe useful for DNA fingerprinting will also be able to score all the associated beneficial points such as pedigree analysis[5,15,26–28], allele drop out, assessing rate of mutation[8,9], studying population structure, monitoring successful breeding partners, ascertaining the propagation of desired germplasm, assessing the gene pool for its overall heterogeneity[2–4,29] and dilution of the same owing to infusion of sudden newer genetic materials in a dynamic population[2,3]. This would also enable to uncover the effect of genetic drift and migration, since allele length frequencies are likely to be affected. Similarly, some of the clones so generated may prove to be useful for breed delineation following simple DNA typing system involving Southern blot hybridization. Different recombinant clones originating from a single satellite fraction comprising consensus sequences of about 15–30 base pair may be of particular biological significance. Oligonucleotides based on these consensus sequences may be used directly as probes to uncover species or breed-specific pattern following dot blot or Southern blot hybridization. Similarly, complete sequencing of about 15–20 independent recombinant clones representing a particular satellite fraction followed by their alignment would uncover polar or non-polar mutations. Analysis of such mutations may also form the basis of breed delineation and may even be linked for their uniqueness with respect to a population. Usually, such non-polar mutations are extremely polymorphic and may reveal micro-level changes in the population[5]. This would also provide a model system to study the environmental effects on the animal population originating from two diverse environments[30,31].

Using the above-generated recombinant clone(s), expression studies may be conducted with the total RNA isolated from different somatic and germline tissues within and between individual animals employing northern-blot analysis. Subsequently, depending upon the detection of signals, the data may be substantiated by RT-PCR experiments as detailed in the following paragraph.

To explore possible transcriptional status of the satellite DNA or satellite-linked sequences

Sizable body of literature is available on different types of satellite fractions from different species including human. However, a systematic search towards the association of these sequences with functional gene or towards their own transcriptional status has never been attempted in the farm animals. This is corroborated by the fact that while a total of 1745 homologous loci in humans and 240 in cattle have been reported, only two have been characterized in the bubaline genome.

An extensive enzymatic restriction survey of the DNA (of the desired species) and identification of satellite fraction(s), their cloning and sequencing will be the first step. The sequencing of the contig will allow primers to be designed for other studies such as monitoring copy number variation, their expression by Northern analysis and RT-PCR-based expression in both the sexes during the course of development in different somatic and germ-

line tissues. It is envisaged that a number of satellite fractions may be found to be transcribing. Analysis of these sequences using *Blast Search* and their homology with entries in the GenBank would help identifying if the cloned satellite were associated with a known functional gene or the same represents still uncharacterized coding sequences. Exclusion studies using these clones may help narrowing the search of much-desired QTL loci that have *hitherto* been an elusive proposition for most of the physical and physiological attributes. Molecular characterization of a large number of satellites following the above-mentioned approaches will maximize the chances of saturating the genome with appropriate markers and marker systems.

Studies on genes and repetitive sequences implicated with sex determination and mutations known to cause infertility associated with heteromorphic sex chromosome(s)

These studies have three apparently unrelated but actually linked components. A typical vertebrate species usually has heteromorphic sex chromosomes with sizable portion of heterochromatin comprising various types of repetitive DNA. In the context of animal biotechnology, for sex identification, use of a marker that uncovers differential organization of the heterochromatic sequences (even without its involvement in actual sex determination) is possible[32,33]. Earlier work has shown that several types of repetitive DNA are involved in the phenomenon of heterochromatinization. However, it is not known if the heterochromatin in all the species is qualitatively and quantitatively similar and the same is linked with the evolutionary status of the species as has been suggested by some workers. Sex identification in a species can be conducted either by using: (i) repetitive DNA probe for dot blot or southern blot hybridization or (ii) primer(s) that amplify the sequences exclusively from the heteromorphic sex chromosome or (iii) primer(s) that amplify sex specific haploid gene(s). The genes responsible for sex determination and associated sequences differ in different species. This warrants an in-depth analysis of this phenomenon in different animal systems providing ample scope for undertaking research on these lines. The envisaged outcome would eventually assist in deciphering fertility status of animals and identifying most prevalent causative factors leading to infertility in a given population. Sex identification in itself may be a small exercise conducted routinely, but it does not tell the actual fertility status of a species. Thus, a detailed genetical study needs to be carried out to be able to establish fertility status involving all the possible genes. The rationale is based on the fact that mutation(s) known to affect fertility status of a species may involve different genes. Once most commonly occurring mutations become known, the same may be studied in a given animal population which may then be used to ascertain if the fertility/infertility situation is similar in different animal populations with respect to particular loci. Deviation, if any, may be attributed to non-genetic factors and this will open up newer vistas for additional research.

Studies on the autosomal gene(s) in the context of animal fertility

It is well established that in many vertebrates, genes located on the sex chromosome are neither sufficient nor responsible for sex determination. A number of autosomal genes have been implicated in the cascade of events leading to organic sex determination. One such autosomal pleiotropic protooncogene *c-kit* receptor has been reported to be implicated in haematopoiesis, melanogenesis and gametogenesis. Multiple mRNA transcripts of this gene and alternate splicing have been reported[34]. In recent studies, intron/exon reshuffling has been observed. In rat model system, the mRNA transcripts analysed from normal and infertile ones have shown one amino acid deletion in testes mRNA transcript compared to somatic tissues including brain. This observation is substantiated by histological studies on testes. Identification of mRNA transcript involved in gametogenesis in normal animal and its mutant forms leading to infertility would be of practical implications for the farm animals. Similarly, studies on other genes known to be involved in sex determination such as SRY, ZFY or Sox family genes would be fruitful. A collective approach towards the understanding of the phenomenon of sex determination and expression of involved genes and their mutant alleles in a given animal population will help towards conservation of biodiversity and management of genetic resources. Further, this information may be useful for reshuffling the two populations to upgrade the relatively confined gene pool of genetically less heterogeneous populations[2–4].

Studies on the rare genetic resources

There are a number of animal species endemic to India such as pigmy hog, dwarf cattle, or mithun which require genetic as well as population-based study in the light of the ever-growing menace of bio-piracy. While local research laboratories may not be fully equipped to undertake this daunting task, effective collaborative research project(s) may be undertaken involving the better equipped (nodal) and regional laboratories. This, while generating the much-needed vital information will also augment the regional infrastructures and at the same time equip us to face the challenge in view of the global patenting regime.

Studies on the highly endangered species

Similarly, genetic studies on highly endangered species such as swamp deer; Indian rhino, elephant, lion and tiger, etc. are of immediate importance. Overall genetic diversities amongst these animals may be ensured only when

we understand the extent of prevalent genetic heterogeneity. A careful and concerted approach without disturbing them or their ecological niche will go a long way to secure a safe place for these natural *beauties* that have been poached in the name of *beast*.

Gene and genome mapping

As mentioned earlier, gene and genome mapping efforts should be undertaken without any further delay. This is crucial not only for developing synteny but also for precise localization of the genes on the chromosome. For *in situ* hybridization, conventional approach using radioisotopes or fluorescence *in situ* hybridization (FISH) may be used. The FISH approach has many other advantages over conventional *in situ* hybridization techniques.

One of the most reliable approaches that have gained momentum for genome mapping is the hybridization of contig(s) with monochromosome hybrids. Use of monochromosome hybrid assigns unequivocally the correct chromosomal position to a gene though the exact location would still involve FISH approach. Thus, a research center should be equipped to deal with not only established conventional *in situ* approach but also FISH and monochromosome hybridization facilities. Monochromosome hybridization involves generation of heterokaryons and systematic propagation of retained chromosome(s) in the background of other much frequently 'knocked-out' genome in the cell culture. The technology is well established and is routinely used for human system. However, for animal species, this has remained relatively unexplored. Reports are available on the formation of human/mouse heterokaryons and selective elimination of human chromosome in the background of mouse but no effort has been made, for example, for the formation of cell hybrid of buffalo and mouse or cow and rat. It is likely that we learn the very mechanism of elimination of a selective genome in the process of generating heterokaryons using cell lines from the farm animals. Thus, animal cell fusion (along the line of human system) and generation of heterokaryons would be yet another important experimental approach not only toward gene mapping but also for understanding the molecular events leading to the selective elimination of one genome in the background of another.

Gene regulation and expression

All the genes that express do not always result in translation of the peptides. There are regulatory mRNA transcripts that are equally important for the maintenance and upkeep of cell systems. Expression studies are conducted at the total RNA level employing northern blot analysis, by RT-PCR approach, RNase protection assay or nuclear run-on assay. Expression studies during the course of development in a species both in somatic and germline cells reveal the differential expression of a gene. If the level of expression varies consistently, from the animals of one population to that of others, it would reflect difference in the population structure. Thus, the population structures may also be uncovered with respect to coding and non-coding sequences. Finally, expression studies may also help in identifying the loci linked with genetic superiority of the animals such as resistance to diseases, tolerance to drought condition or less fodder requirement, etc.

Development of DNA-based genetic markers for species identification and conservation biology: The MASA strategy

An offshoot of the above studies would be the development of DNA-based genetic markers useful for addressing issues related to diversified aspects of animal biotechnology. Screening of the libraries would generate numerous clones representing expressed sequence (from cDNA library) or regulatory elements (from genomic library). Identification and characterization of such clones and their heterologous use will provide wealth of information in the context of comparative genome analysis. Similarly, identification and characterization of genes and their mutant alleles linked with satellite sequences, by *minisatellite associated sequence amplification (MASA)* would be equally rewarding. The animal genome like any other higher eukaryotes, is also endowed with different kinds of short tandem repeat motifs (STRs) in the range of 2–15 base pairs. Such STRs are represented several hundred to thousand times depending upon the overall rate of mutation[8,9] and the natural selection in favour of these satellites[35,36]. Minisatellite associated sequence amplification (MASA) studies using three STRs representing 16 base long consensus sequence of locus 33.15 and 15 and 18 base long 5'TGTC3' repeat motifs revealed clad specific amplicons. MASA works on the basic principle of restriction fragment length polymorphism (RFLP) but the use of PCR, unlike RFLP, facilitates the quick detection of the signals obliterating the requirement of a large quantity of DNA and several days of autoradiography. In case of actual experimental condition, at the outset, about 100 STRs from different satellite fractions in the range of 2–18 base residues can be used for conducting MASA with about 200–400 animal DNA samples representing well-identified populations. The MASA amplicons represented a number of functional genes associated with STR motifs. With a single primer, this approach does not amplify all the functional genes. However, by using different STRs derived from various satellite regions encompassing larger parts of the genome, a large number of functional genes may be amplified. In the event of alteration of a gene leading to its mutant allele, this can be identified by hybridizing MASA

amplicons and assessing its allele length variation. It is envisaged that normal gene will show no change in the position of the amplicon(s) whereas the mutant alleles will show allele length variation. With this approach, a number of STRs may be screened and comprehensive gene-grids representing satellite sequences on one axis and genes linked with them on the other may be generated. It may be noted that with a single copy gene, it takes 10–15 days for obtaining hybridization signal. However, owing to a billion-fold amplification of the target substrate, the amplicons generated by MASA will produce signals in less than one hour. This approach will demonstrate most, if not all, of the coding sequences (functional genes) and their possible association with the satellite sequences. Thus, MASA approach provides ample working flexibility with respect to: (i) identification of a satellite-linked functional gene and its mutant alleles enabling us to uncover the population structure, and (ii) ascertaining the diseased loci to gain information about the susceptibility status of a population (population at risk). Subsequently, more refined gene grid, as mentioned earlier may be developed. A comparison of the data on allele length variation obtained from different populations would provide wealth of information on the population structure.

Man-power development

Simultaneously, serious attempts should also be made to train manpower (with infectious scientific enthusiasm) not only for delivering 'today' but also during the fast approaching next millennium. Despite many progresses made in the area of Science & Technology and availability of highly trained manpower, the potentials of our research scientists have remained underutilized. Thus, it is not only important to build *state of art* infrastructures that would be needed for conducting diversified experiments but also to expose researchers, particularly the new crop, to a global science culture. Similarly, strategies need to be evolved to initiate the process of developing an original thinker in a scientist than a mere research technician. In this context, an early identification of talent would prove to be as critical for success of science in the long run as asking a right kind of scientific question itself.

1. Ali, S., Verma, G. and Bala, S., *Anim. Genet.*, 1993, **24**, 199–202.
2. Ali, S., Ansari, S., Ehtesham, N. Z., Azfer, M. A., Homkar, U., Gopal, R. and Hasnain, S. E., *Gene*, 1998, **223**, 361–367.
3. Sulaiman, I. M. and Hasnain, S. E., *Theor. Appl. Genet.*, 1996, **93**, 91.
4. Sulaiman, I. M. and Hasnain, S. E., *Electrophoresis*, 1995, **16**, 1746.
5. Deka, R., Shriver, M. D., Yu, L. M., Ferrell, R. E. and Chakraborty, R., *Electrophoresis*, 1995, **16**, 1659–1664.
6. Mattapallil, M. J. and Ali, S., *DNA Cell Biol.*, (in press).
7. John, M. V. and Ali, S., *DNA Cell Biol.*, 1997, **16**, 369–378.
8. Weber, J. L. and Wong, C., *Hum. Mol. Genet.*, 1993, **2**, 1123–1128.
9. Shriver, M. D., Jin, L., Chakraborty, R. and Boerwinkle, E., *Genetics*, 1993, **134**, 983–993.
10. Di rienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M. and Freimer, N. B., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 3166–3170.
11. Ali, S. and Wallace, R. B., *Nucleic Acids Res.*, 1988, **16**, 8487–8496.
12. Raina, A., Sulaiman, I. M., Das, P., Ehtesham, N. Z., Ali, S., Dogra, T. D. and Hasnain, S. E., *Gene*, 1996, **173**, 247–250.
13. Azfer, M. A., Bashamboo, A., Ahmed, N. and Ali, S., *J. Biosci.*, 1999, **24**, 101–107.
14. Rao, B. K., Sil, S. B. and Majumder, P. P., *J. Genet.*, 1997, **76**, 181–188.
15. Shriver, M. D., Jin, L., Ferrell, R. E. and Deka, R., *Genome Res.*, 1997, **7**, 586–591.
16. Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R. and Cavalli-Sforza, L. L., *Nature*, 1994, **368**, 455–457.
17. Ali, S. and Epplen, J. T., *Indian J. Biochem. Biophys.*, 1991, **28**, 1–9.
18. Ehtesham, N. S., Das, A. K. and Hasnain, S. E., *Gene*, 1992, **111**, 261.
19. Ehtesham, N. Z. and Hasnain, S. E., *Adv. Forensic Haemogenet.*, 1992, **4**, 137.
20. Raina, A., Sulaiman, I. M., Ehtesham, N. Z., Das, P., Ali, S., Dogra, T. D. and Hasnain, S. E., *Gene*, 1996, **173**, 247.
21. Ehtesham, N. Z., Talwar, G. P., Ali, A. and Hasnain, S. E., *Indian J. Biochem. Biophys.*, 1990, **27**, 275.
22. Ehtesham, N. Z., Ma, D. P. and Hasnain, S. E., *Gene*, 1991, **98**, 301.
23. Ali, S., Azfer, M. A., Bashamboo, A., Mathur, P. K., Malik, P. K., Mathur, V. B., Raha, A. K. and Ansari, S., *Gene*, 1999, **228**, 33–42.
24. Dil-Afroze, Misra, A., Sulaiman, I. M., Sinha, S., Sarkar, C., Mahapatra, A. K. and Hasnain, S. E., *Gene*, 1998, **206**, 45.
25. Sinha, S., Dil-Afroze, Misra, A., Sulaiman, I. M., Sarkar, C., Mahapatra, A. K. and Hasnain, S. E., *FASEB J.*, 1997, **11**, 2313.
26. Mattapallil, M. J. and Ali, S., *Gene*, 1997, **206**, 209–214.
27. Takezaki, N. and Nei, M., *Genetics*, 1996, **144**, 389–399.
28. Chakraborty, R., Kimmel, M., Stivers, D. N., Davison, L. J. and Deka, R., *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 1041–1046.
29. Sulaiman, I. M., Ehtesham, N. Z. and Hasnain, S. E., *Gene*, 1995, **156**, 223.
30. John, M. V., Parwez, I., Sivaram, M. V. S., Mehta, S., Marwah, N. and Ali, S., *Gene*, 1996, **172**, 191–197.
31. Kimmel, M., Chakraborty, R., King, J. P., Bamshad, M., Watkins, W. S. and Jorde, L. B., *Genetics*, 1998, **148**, 1921–1930.
32. Ali, S., Verma, G. and Bala, S., *Mol. Cell. Probes*, 1992, **6**, 521–526.
33. Ali, S., Appa Rao, K. B. C. and Bala, S., *Assisted Reprod. Rev.*, 1993, **3**, 37–43.
34. Ali, S., Ravindranath, N., Jia, M. C., Musto, N. A., Tsujimura, T., Kitamura, Y. and Dym, M., *Biochem. Biophys. Res. Commun.*, 1996, **218**, 104–112.
35. Deka, R., Shriver, M. D., Yu, L. M., Jin, L., Aston, C. E., Chakraborty, R. and Ferrell, R. E., *Genomics*, 1994, **22**, 226–230.
36. Ali, S., Müller, C. R. and Epplen, J. T., *Hum. Genet.*, 1986, **74**, 239–243.

Received 4 June 1999; accepted 21 July 1999