Introduction:

The centromere mediates normal mitotic and meiotic process including kinetochore assembly, spindle attachment, sister chromatid cohesion and subsequently proper separation of sister chromatids during mitotic anaphase or segregation of homologous chromosomes during meiosis I (Choo 1997). Centromere malfunction is the leading cause of aneuploidy, classically defined as a deviation of chromosome number from the normal chromosome complement (Hook 1985). Aneuploidy has catastrophic consequences for chromosomal disorders. The epidemiological study has shown that, in human, 45% of spontaneous abortions or 7% of all conceptions are associated with aneuploidy (Jacobs and Hassold 1995), and aneuploidy causes common birth defects (e.g. Down, Klinefelter and Turner syndromes). Aneuploidy is also tightly correlated with almost all types of cancer: there are >84,000 documented cases of abnormal karyotypes associated with human neoplastic disorders (Vig et al. 1989; Mitelman 1994). Detail studies in the structure and molecular building blocks of a centromere would be essential for understanding its crucial role in maintaining the exact chromosome complement of daughter cells or offsprings.

However, centromeric regions of eukaryotic chromosomes still present an enigma that a conserved centromere function is in the face of rapidly evolving centromeric DNA sequence (Sullivan 2001). Each species studies to date, from yeast to worms to flies to humans, the DNA sequences found in the centromere region are composed of quite diverse satellite DNAs. One still could speculate that a specific conserved DNA sequence, like the centromeric DNA of the budding yeast, exits to assemble kinetochore proteins and then to form a functional centromere in higher eukaryotypes. In order to unveil a specific conserved centromeric DNA fundamental for a functional centromere, comparative sequence analysis of the centromere regions in various species is urgently necessary. The full DNA sequences in centromere region in most mammalian species remain unknown. Only the centromeric sequence of human is almost completed (Schueler 2001) by the time the human genome project had been finished (Venter 2001). Never the less, the competent centromeric sequences analysis from human X chromosome revealed as a functional centromere DNA. This may serve as a model for the comparative sequence analysis of centromere loci in other mammalian species for the search of functional centromeric DNA in that species. Consequently, it will provide better insights into the evolution of cenromeric DNA that leads to the formation of currently active centromere.

Almost as a rule, the repetitive satellite DNAs is the main composition of centromere and centromeric DNA sequences diverse among species. Generally, most similar satellite DNAs are restricted to closely related species. In other words, the related (or congeneric) species share a "library" of similar related satellite sequences, some of which could be amplified onto a major satellite DNA in some particular species (Salser et al. 1976; Meštrović et al. 1998). The

evolutionary dynamics of satellite DNAs could be a driving force for the speciation process, thus forming a species-specific profile of satellite DNAs (Ugarković and Plohl 2002). The species-specific profile of centromeric satellite DNA is possible to be demonstracted by means of comparative genomic studies that showed a large fraction of centromeric sequences were repositioned either by independent chromosomal rearrangements or by *de novo* centromere emergence in the independent lineage to become unique to a species (Murphy et al. 2005). Therefore, mapping and sequencing of centromeric DNAs from phylogenetically closely or divergent species will provide a better understanding on the mechanism for centromeric satellite DNA evolution. Furthermore, it could also unravel the puzzle of a functional conserved centromere in the face of rapidly evolving centromeric DNA sequence.

BAC library has tremendous potential to facilitate such studies to shed light on the temporal relationships of extant species by comparative sequence analysis and cytogenetic mapping of genome rearrangement, expansion, and loss (Gomyo et al. 1999; Srinivasan et al. 2002). In addition to decipher genome evolution, BAC resources could be used to address specific questions pertaining to centromeric and pericentromeric evolution, and the mechanisms of karyotypic change between closely related species. The Indian muntjac (*Muntiacus muntjac vaginalis*) has the lowest chromosome number in mammalian species with 6 chromosomes in female and 7 in male. Recent molecular cytogenetic studies landed support to the hypothesis that the karyotype of the Indian muntjac could evolve from an ancient deer species with a karyotype 2n=70 through extensive tandem fusions and several centric fusions (Li et al. 2000). Moreover, the centromere of X+3 chromosome of the species is compound and unique. Therefore, it is an ideal species for studying chromosome rearrangement, comparative genomic and karyotype evolution. Thanks to the constructed BAC library of Indian muntjac in our lab; therefore, it is availability to explore the following studies: the centromeric evolution and function, the underlying mechanism of tandem fusion of Indian muntjac chromosomes, the chromosomal rearrangement, and the karyotype evolution.

Specific Aim:

Studying the genomic organization of Indian muntjac centromere

Systematically sequencing and mapping the genomic organization of Indian muntjac centromeric BAC clones would make an effort in identifying candidate functional centromeric DNAs of Indian muntjac.

Method:

Systematically sequencing:

The hierarchical sequencing strategy was used to sequence the centromeric BAC clone. First, the centromeric BAC DNA was digested into smaller fragments by two different restriction enzymes separately. All restriction fragments then were subcloned. All subclones were sequenced. Second, Paired-end sequencing was used to connect each sequences of subclone.

Computational analysis and comparative sequences analysis:

Sequences of centromeric DNA will be aligned with GenBank database using the BLAST programs for searching the identified satellite DNAs elements. Tandem repeats will be determined by single-base-shift self-comparison (Pluciennicza k et al. 1982). Palindromic and mirror repeats will be identified by eye. The interspersed repeats and low complexity DNA sequences will be screened using RepeatMasker program from http://www.repeatmasker.org/. The composition of sequences will be determined using nucleic acid statistics programs from Biological Workbench (http://workbench.sdsc.edu/).

Comparative sequence analysis will be performed in order to understand the evolution of centromere sequences. This will be achieved by comparing the above identified muntjac centromeric sequences with cenromeric sequences of different species that had been deposited in the GenBank database.

Physical mapping the subclones of BAC:

All subclones, with exception of those containing the known satellite DNA elements, from BAC clones of interest will be mapped onto Indian muntjac chromosomes by FISH experiment for identifying the chromosome localization. BAC-fiber FISH will be also conducted using every subclones as FISH probes for confirming the order of assemble based on the sequences. Two continuous subclones will be labeled with different fluorochromes as FISH probes to hybridize the chromatin fiber for identifying the high resolution genomic organization. .

For BAC-fiber FISH, the freshly isolated BAC DNA is fixed onto the poly-L-lysine-coated slide by heating at 65^{o}C for 10 min. For chromatin-fiber FISH, the cells are fixed onto the slide and then soak in lysis buffer to release the chromatin fiber. The FISH will be performed using digoxigenin labeled BAC vector, biotin-labeled subclones, and FITC labeled the other subclones as probes. The digoxigenin labeled probe is detected with cy5-conjugated anti-DIG antibodies, while the biotin labeled probe is detected with AF568-conjugated avidin. Finally, samples are mounted in antifade mounting medium with DAPI (vector). Fluorescent signals are captured under an Olympus BX51

fluorescence microscope equipped with appropriate filter sets and a cooled CCD camera (Photometrics, Sensys). The FISH image will be normalized and enhanced using the IPlab software. The detailed FISH protocol had been established in our laboratory (Li et al. 2002).

Results and discussions:

I. Construction of an Indian muntjac BAC library and production of the most highly density FISH map of the species

The frequency of clones with inserts was 88% and thus this library corresponds to approximately 4X coverage of Indian muntjac genome. The individual chromosomal location of 2,242 BAC clones on the Indian muntjac metaphase chromosomes were identified by fluorescence in situ hybridization (FISH). Among these clones, 2,107 BAC clones were mapped onto specific loci and 135 BAC clones were mapped onto the centromeric region. This provided the most high density FISH BAC clone map for the species. This dense ordered map could be used as a blueprint for comparative FISH mapping studies of other deer species in order to understand the mechanism of genomic rearrangement and karyotypic evolution. Moreover, the centromeric BAC clones will provide an excellent resource for studying the structure and function of mammalian centromeres. **The result had been published in Zoological Studies 47:282-292 (2008).**

II. Studying the genomic organization of Indian muntjac centromere

In our preliminary BAC mapping study, four BAC clones with the parallel signals on the kinetochore position by FISH experiment were observed (Fig 1). This special parallel signal of BAC clones probably interprets that these BAC clones contain a DNA sequence likely to be responsible for centromere function by associating with kinetochore proteins. Therefore, it is very meaningful to further characterize the genomic organization of these BAC clones by full sequencing. A hierarchical sequencing was performed. First, 17 *Eco*RI digesting fragments were subcloned into pBSK vector and then were sequenced. Second, all sequenced subclones were aligned (Fig 2). The sequence result showed that the BAC clone contained not only cervid satellite II and IV DNAs but also other sequences, such as long interspersed nucleotide elements (LINEs), short interspersed nucleotide elements (SINEs), long terminal repeats (LTRs), other DNA elements and unidentified sequence. It suggested that the centromeric DNA organization is complex than it was expected. Furthermore, EMSA and the artificial chromosome assay will be performed for identifying the critical sequence for centromeric function.

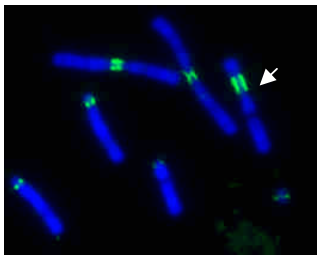

Fig.1: 1296A1 BAC clone DNA was hybridized onto Indian muntjac chromosomes. The FISH signals show parallel pattern on the lateral sides of the primary constriction where kinetochore located.

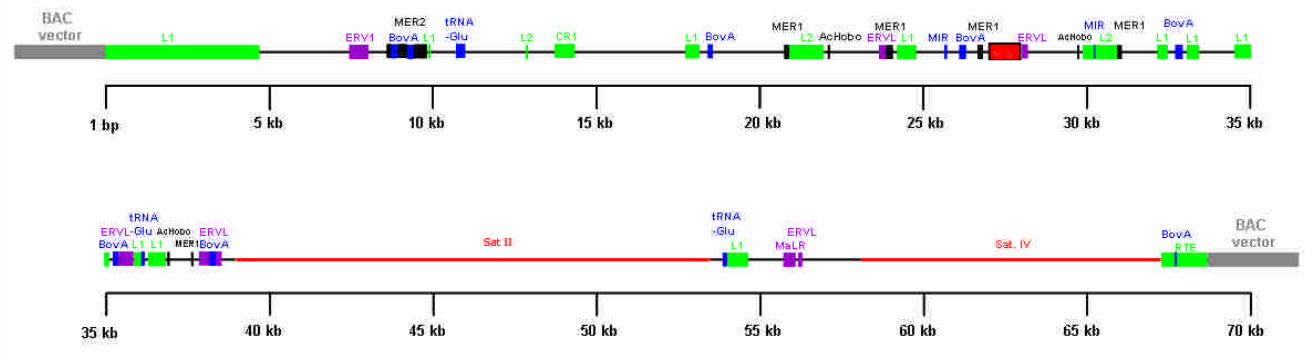

Fig 2. Sequence organization of the Indian muntjac centromeric BAC DNA clone. The sequences contains satellite tracks (red arrows), LINEs (green), SINEs (blue), LTR elements (purple), and DNA elements (black).

Self-evaluation:

In this year's project, our major aim is to studying the genomic organization of Indian muntjac centromere. We had constructed 4X coverage of Indian muntjac BAC library and mapped 2242BAC clones onto the Indian muntjac metaphase chromosomes by fluorescence in situ hybridization (FISH) at the previous project. This part of result had been published in **Zoological Studies 47:282-292 (2008).** In this study, we further characterized the centromeric BAC clone to understand genomic organization of centromere. This result would shed light on the genomic organization of centromere. It is not easy to be identified the centromeric sequence, because of highly repeatitive DNA sequence. Therefore, it was taken much time to identify the centromeric sequence. All together, we should have achieved the progress of the project.

References:

Choo, K.H.A. (1997) The centromere. Oxford University Press, Oxford.

Hook, E. (1985) The impact of aneuploidy upon public health: mortality and morbidity associated with human chromosomal abnormalities. In Aneuploidy: Etiology & Mechanisms, vol.36. Edited by Dellarco V, Voytek P, Hollaender A. New York: Plenum press;7-33.

Jacobs, P.A and Hassold, T.J. (1995) The origin of numerical chromosome abnormalities. *Adv. Genet*. **33**:101-133.

Mitelman F (1994): Catalog of chromosome aberrations in cancer, 5th Edition. New York: Wiley

Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K. and Willard, H.F. (2001) Genomic and genetic definition of a functional human centromere. *Science* **294**:109-115.

Sullivan, B.A., Blower, M.D., Karpen, G.H. (2001) Determining centromere identity: cyclical stories and forking paths. *Nat. Rev. Genet* **2**:584-596.

Venter, J.C., Adams, M.D., Myers,E.W. et al. (2001) The Sequence of the Human Genome. *Science* **291**:1304-1351.

Vig, B.K., Sternes, K.L. and Paweletz, N. (1989) Centromere structure and function in neoplasia. *Cancer Genet Cytogenet* **43**:151-178.

Li, Y.C., Lee, C., Sanoudou, D., Hseu, T.H., Li, S. Y. and Lin, C.C. (2000) Interstitial colocalization of two cervid satellites DNAs involved in the genesis of the Indian muntjac karyotype. *Chromosome Res* **8**:363-373.

Li, Y.C., Lee, C., Chang, W.S., Li, S. Y. and Lin, C.C. (2002) Isolation and identification of a novel satellite DNA family highly conserved in several Cervidae species. *Chromosoma* **111**:176-183.

Meštrović N, Plohl M, Mravinac B, Ugarković D: Evolution of satellite DNAs from the genus Palorus--experimental evidence for the "library" hypothesis. Mol Biol Evol 15:1062-1068 (1998).

Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L, Hitte C, Meyers SN, Milan D, Ostrander EA, Pape G, Parker HG, Raudsepp T, Rogatcheva MB, Schook LB, Skow LC, Welge M, Womack JE, O'Brien SJ, Pevzner PA, Lewin HA (2005) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps Science 309:613-617

Salser W, Bowen S, Browne D, el-Adli F, Fedoroff N, Fry K, Heindell H, Paddock G, Poon R, Wallace B, Whitcome P: Investigation of the organization of mammalian chromosomes at the DNA sequence level. Fed Proc Fed Am Soc Exp Biol 35:23-35 (1976).

Ugarković D, Plohl M: Variation in satellite DNA profiles-cause and effects. EMBO J 21:5955-5959 (2002).

Plucienniczak, A., Skowronski, J., Jaworski, J. (1982) Nucleotide sequence of bovine 1.715 satellite DNA and its relation to other bovine satellite sequences. *J. Mol. Biol.* **158**:293-304.