Final Technical Report

Project: Structure, Function, and Evolution of Rice Centromeres
DOE Award Number: FG02-01ER15266
Principle Investigator: Jiming Jiang
Institute: University of Wisconsin-Madison

The centromere is the most characteristic landmark of eukaryotic chromosomes. The centromere appears as a distinct primary constriction on condensed metaphase chromosomes. Centromeres are responsible for sister chromatid cohesion and are the sites for kinetochore assembly and spindle fiber attachment, thereby allowing for faithful segregation of sister chromatids during cell division. In the majority of eukaryotic species, centromeres are embedded within megabases of highly repetitive DNA which cannot be precisely sequenced. Determining the precise DNA boundary of a centromere has proven to be a difficult task. Rice provides an excellent model for centromere study because rice centromeres contain limited amount of repeats. This project aimed for an in-depth analysis of the structure, function, and evolution of rice centromeres.

The functional domains of rice centromeres, which are defined by the presence of a centromere-specific histone protein CENH3, contain a 155-bp satellite repeat CentO. The amount of the CentO satellite varies significantly among different rice centromeres, ranging from 65 kilobases (kb) to approximately 2 megabases (Mb). We have successfully cloned and sequenced the centromere of rice chromosome 8 (Cen8), representing the first fully sequenced centromere from any multicellular eukaryotes. The functional domain of Cen8 spans approximately 800 kb, including 65 kb of CentO. We conducted an in-depth annotation of the 800 kb region for gene composition by integrating both DNA and protein homolog searching, computational predictions, and intensive manual inspection. A total of 29 gene models were found in this region and the expression of 16 genes was confirmed by reverse transcriptase-polymerase chain reaction assay. We have analyzed the histone modification patterns associated with these 16 genes and more than 100 genes that flank the centromere core domain. We did not find difference of histone modification patterns of genes located within or outside of the centromere core domain. These results support that rice Cen8 was originally evolved from a gene-containing genomic region. We also applied a genome-wide approach to define the functional core of all 12 rice centromeres. Using chromatin immunoprecipitation combined with next generation sequencing we mapped the precise boundaries of the functional cores of nine of the 12 rice centromeres. The sizes of the defined functional centromeres ranged from approximately 500 kb to 1500 kb.

Rice centromeres also contain a centromere-specific retrotransposon CRR. The CRR elements are highly enriched in the functional core of rice centromeres. In addition, retrotransposons homologous to the CRR elements have been discovered in distantly related plant species, including wheat and maize. These results suggest that the CRR elements may play a role in centromere function. We collected all of the CRR elements from rice chromosomes 1, 4, 8, and 10. Phylogenetic analysis revealed that the CRR elements are structurally diverged into four subfamilies, including two autonomous subfamilies (CRR1 and CRR2) and two nonautonomous subfamilies (noaCRR1 and noaCRR2). The CRR and noaCRR elements share substantial sequence similarity in regions required for DNA replication and for recognition by integrase during retrotransposition. Thus, the noaCRR elements were likely mobilized through

the retrotransposition machinery from the autonomous CRR elements. CRR elements were transcribed in root, leaf, and panicle tissues, suggesting a constitutive transcription of this retrotransposon family. However, the overall transcription level was low and the CRR transcripts appeared to be derived from relatively few loci. The majority of the CRR transcripts had chimerical structures and contained only partial CRR sequences. We detected small RNAs (smRNAs) cognate to nonautonomous CRR1 (noaCRR1) and CRR1, but not CRR2 elements. These results suggest that different CRR subfamilies may play different roles in the RNAi-mediated pathway for formation and maintenance of centromeric chromatin.

Publications resulted from the project:

Nagaki, K., Cheng, Z.K., Ouyang, S., Talbert, P.B., Kim, M., Jones, K.M., Henikoff, S., Buell, C.R., and Jiang, J. (2004) Sequencing of a rice centromere uncovers active genes. *Nature Genet.* 36: 138-145.

Nagaki, K., Neumann, P., Zhang, D., Ouyang, S., Buell, C.R., Cheng, Z., and Jiang, J. (2005) Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. *Mol. Biol. Evol.* 22: 845-855.

Lee, H.-R., Zhang, W.L., Langdon, T., Jin, W.W., Yan, H.H., Cheng, Z.K., and Jiang, J. (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. *Proc. Natl. Acad. Sci. USA* 102: 11793-11798.

Yan, H.H., Jin, W.W., Nagaki, K., Tian, S., Ouyang, S., Buell, C.R., Talbert, P.B., Henikoff, S., and Jiang, J. (2005) Transcription and histone modifications in the recombination-free region spanning a rice centromere. *Plant Cell* 17: 3227-3238.

Yan, H.H., Ito, H., Nobuta, K., Ouyang, S., Jin, W.W., Tian, S.L., Lu, C., Venu, R.C., Wang, G.-L., Green, P.J., Wing, R.A., Buell, C.R., Meyers, B.C., and Jiang, J. (2006) Genomic and genetic characterization of rice *Cen3* reveals extensive transcription and evolutionary implications of a complex centromere. *Plant Cell* 18: 2123-2133.

Lee, H.-R., Neumann, P., Macas, J., and Jiang, J. (2006) Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice. *Mol. Biol. Evol.* 23: 2505-2520.

Neumann, P., Yan, H.H., and Jiang, J. (2007) The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. *Genetics* 176: 749-761.

Yan, H.H., Talbert, P.B., Lee, H.-R., Jett, J., Henikoff, H., Chen, F., and Jiang, J. (2008) Intergenic locations of rice centromeric chromatin. *PLoS Biol.* 6: 2563-2575.