

Poster B-18

The TIGR Rice Genome Annotation Database



Authors:

Shu Ouyang (*The Institute for Genomic Research*)
Haining Lin (*The Institute for Genomic Research*)
Kevin L. Childs (*The Institute for Genomic Research*)
Wei Zhu (*The Institute for Genomic Research*)
John Hamilton (*The Institute for Genomic Research*)
Dan Lee (*The Institute for Genomic Research*)
Mathew Campbell (*The Institute for Genomic Research*)
Rena Malek (*The Institute for Genomic Research*)
Li Zheng (*The Institute for Genomic Research*)
Joshua Orvis (*The Institute for Genomic Research*)
Brian Haas (*The Institute for Genomic Research*)
Jennifer Wortman (*The Institute for Genomic Research*)
Robin Buell (*The Institute for Genomic Research*)

Short Abstract: We have been funded to annotate the rice genome. Our Osa1 database contains pseudomolecules representing the 12 rice chromosomes. Ab initio gene finders, full length cDNA, EST and protein evidence were used to generate gene models and functional assignments for those models. The annotation is available <http://rice.tigr.org>.

Long Abstract:

We have been funded by the National Science Foundation to annotate the rice genome and to provide access to the annotated genome for the community. We have created a rice annotation database called Osa1 that houses sequence, annotation and analysis information. We released version 4 of our rice genome annotation in January 2006. The version 4 release includes a number of updates and improvements. We have re-constructed pseudomolecules of the 12 rice chromosomes that represent ~95% of the total genome. These pseudomolecules have been analyzed by gene prediction programs. Full length rice cDNAs and rice ESTs have been aligned to the pseudomolecules. Gene models were created using a combination of the ab initio gene finders and alignment evidence. The results from blastp analyses against the TIGR non-redundant protein database and from hmmpfam analyses against pfam domains have been used to provide improved functional assignments to the gene models. Alignments between the TIGR version 4 pseudomolecules and both the IRGSP build 4.0 pseudomolecules and the BGI 93-11 contigs have also been prepared. We have performed a comprehensive analysis of alternative splicing in rice and compared it with the data for Arabidopsis. Many other types of annotations are available for the rice genome including gene ontology assignments, domains and motif identification, paralogous family construction, segmental duplication analysis, and alignments with flanking sequence tags, genetic markers, and oligos from a variety of rice microarray platforms. We have also developed a series of alignments with other plant species including alignments with ESTs, genomic sequence and with genetic markers from wheat, maize, sorghum and arabidopsis. The expression of rice gene models was analysed

using full length cDNA/EST, MPSS and SAGE data. We have developed a community annotation tool to facilitate the incorporation of gene family annotations from individual rice researchers into the TIGR rice annotation. The annotation is available to the community through the TIGR Rice Genome Annotation website (<http://rice.tigr.org>) where the annotation can be viewed in the Rice Genome Browser with 51 tracks, or downloaded using our ftp site or the Data Extractor Tool.

Poster B-18

KPD: Known gene Promoter Database and tissue specific gene regulation



Authors:

Soo-Young Cho (*Hanyang university*)
Min-Young Park (*Hanyang university*)
Seong-Jin Park (*Hanyang university*)
Young-Seek Lee (*Hanyang university*)

Short Abstract: Recently, to find the target genes of transcription factor(TF)s researchers used high-throughput experiment assays. However, cells show various phenotypes in different condition although they contain same genome sequence. we construct database for tissue specific transcriptional gene regulation. Our database provides the network of tissue specific gene regulation.

Long Abstract:

One of the interesting topics of gene regulation study in the post-genome era is non-coding region of the genome. Non-coding region is a segment of DNA that does not comprise a gene and thus does not code for a protein. Promoter region, which is one of the non-coding region, contains critical information such as gene regulation, time-dependent expression and cellular conditions. Recently, to find the target genes of transcription factor(TF)s researchers used high-throughput experiment assays such as ChIP-chip. However, cells show various phenotypes in different condition although they contain same genome sequence. To study the correlations of target and TF expression, we compute probability of tissue specific expression using Microarray chip data from GEO and the gene regulation by given TF using Bayes' theorem. As a result, we construct database for tissue specific transcriptional gene regulation. Our database provides the network of tissue specific gene regulation. In comparison between normal subject and pathological cells such as cancer cells, these database also help us to understand specific gene expression and their network in disease status