Journal of Genetics and Molecular Biology
Vol. 14, No. 4, 201-206, December 1, 2003

Rice Structural and Functional Genome Research in ASPGC, Academia Sinica

Ya-Ting Chao[1], Chin-San Chen[1], Hong-Hwa Chen[2], Teh-Yuan Chow[1], Mei-Chu Chung[1], Jaw-Shu Hsieh[3], Yue-Ie Hsing[1*], Pei-Fang Lee[1], Yann-Rong Lin[4], Jei-Fu Shaw[1], Chun-Lin Su[1], Hong-Pang Wu[1], Su-May Yu[5]

1. Institute of Botany, Academia Sinica, Taipei, Taiwan;
2. Dept. of Biotechnology, National Cheng-kung University, Tainan, Taiwan;
3. Dept. of Agronomy, National Taiwan University, Taipei, Taiwan;
4. Dept. of Biological Science, Fu-jen University, Taipei, Taiwan;
5. Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

Due to its economic importance, small genome size, and co-linear gene organization with other cereal grasses, rice has been chosen as a model organism for genome sequencing. ASPGC is currently working on the high-throughput sequencing project of rice chromosome 5. Bioinformatics study is also carried out to understand the genome and its complexity. ASPGC also works on EST, SAGE and insertional mutants of rice. The crop improvement relies on genomic technology more and more recently. Through the structural and functional genomic study, this team hopes to add more information and tools for the rice research community.

Structural Genomes

Genome sequencing

Rice (Oryza sativa) has been chosen as the first crop to be sequenced by an international sequencing consortium, the IRGSP (International Rice Genome Sequencing Project) for the following reasons: (1) Rice is an important crop in the world, feeding about one half of the world 's population; (2) Rice ' s genome size, 430 Mb, is the smallest among crops; (3) Rice linkage and physical maps have been well established, and over 200,000 ESTs available in public databases. (4) The transgenic technology for rice has been established. (5) Rice shares a co-linear gene organization with other cereal grasses (Fig. 1). ASPGC, a member of IRGSP, works on the sequencing work of chromosome 5 (Fig. 2) and adopts the map-based clone-by-clone shotgun strategy. We already submitted 308 BAC/PAC or 42 Mb sequences up to now, and sequence information as well as other links are shown in our website at http://genome.sinica.edu.tw (Fig. 3).

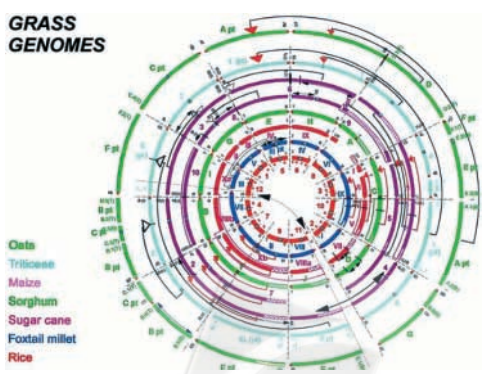

Fig 1. The synteny of grass genomes

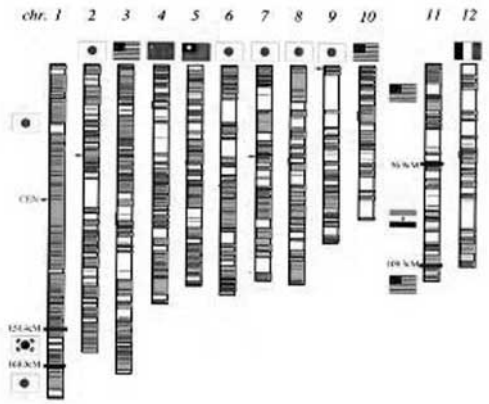

Fig 2. The International Rice Genome Sequencing Project Consortium

Fig 3. Web site of ASPGC

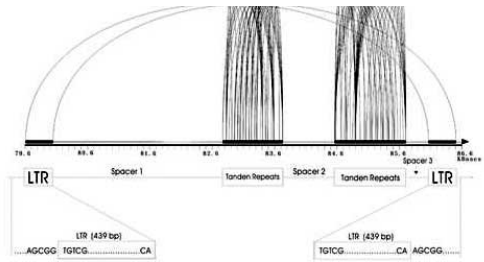

Fig 4. Structures and organization of TEOS1. Panel A. Sequence motifs are illustrated using Miropeats program, the arch lines are drawn between matching sites to indicate the tandem repeats. Panel B. Nucleotide sequences of flanking the LTRs of TEOS1. LTRs are boxed, and the 5-bp target-site duplications are indicated by arrows.

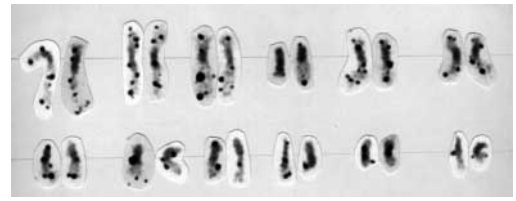

Fig 5. Localization of RIRE3 revealed by in situ hybridization to prometaphase rice chromosomes using digoxigenin-labeled RIRE3 DNA as the probe.

The DNA complexity of rice chromosome 5

The complexity of rice genomic DNA was studied by systematical computer-based repeat mining in three PACs selected from gene-rich regions of chromosome 5. Our data provided an average density of one SSR every 5 kb. The average TEs contribution was 14.5% of the genomic sequence. There were at least one or two retrotransposon present in each BAC/PAC we sequenced. A novel family of transposable elements TEOS1, (Transposable Element of Oryza sativa) were uncovered (Fig. 4).

Rice Insertional Mutants Project

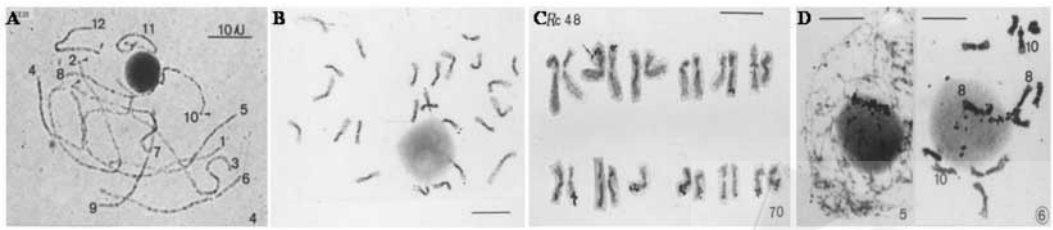

Fig 6. Cytogenetic study on rice chromosomes. Panel A: Rice pachytene
Panel B: Rice mitotic prometaphase chromosomes. Panel C: In situ localization of a repetitive Os48. Panel D: Nucleolar chromosomes are #8 and 10.

Functional Genome Research

EST project

We worked on the EST (expressed sequence tags) sequencing and analysis of young developing panicles. For the variety we used, Tainung 67, the panicles with 6 cm in length were always in meiosis stage. The panicles with 0.5 cm in length, however, were in the spikelets/branches development stage. We collected both of these tissues, constructed cDNA libraries, and ran EST sequencing. About 3000 sequences from each tissue have been obtained. Many novel genes have been discovered, e.g. DNA binding proteins, RNA binding proteins, protein kinase, etc (Table 1).

Fig 7. Rice panicles.

Table 1. Lists of EST sequences and annotations.

Query ID	Annotation	Source
5mmPR-A1A01S	GTP-binding protein	Cicer arietinum
5mmPR-A1A03S	MAP kinase homolog	
5mmPR-A1A06S	Elicitor-responsive gene 3 [imported]	Oryza sativa
5mmPR-A1A08S	Superoxide dismutase [CU-ZN]	Oryza sativa
5mmPR-A1A10S	Gibberellin-stimulated transcript 1 like protein	Arabidopsis thaliana
5mmPR-A1A11S	Putative small nuclear ribonucleoprotein G	Oryza sativa
5mmPR-A1B01S	Hypothetical protein	Arabidopsis thaliana
5mmPR-A1B02S	Sucrose synthase 2	Oryza sativa
5mmPR-A1B05S	Metallothionein-like protein type 2	Oryza sativa
5mmPR-A1B06S	Probable protease inhibitor P322 precursor	
5mmPR-A1B07S	Putative 40S Ribosomal protein	Oryza sativa
5mmPR-A1B10S	Putative protein	Arabidopsis thaliana
5mmPR-A1B12S	Histone H4	Triticum aestivum
5mmPR-A1C03S	Unknown protein	Arabidopsis thaliana
5mmPR-A1C04S	Hlycine-rich RNA-binding protein 2	Oryza sativa
5mmPR-A1C05S	Receptor-like protein kinase	Oryza sativa
5mmPR-A1C07S	CG9248 gene product	Drosophila melanogaster
5mmPR-A1C08S	Ubiquitin-conjugating enzyme E2-17 kD	Lycopersicon esculentum
5mmPR-A1C09S	Similar to pigpen protein from Mus musculus	Arabidopsis thaliana
5mmPR-A1C10S	Putative 60S ribosomal protein L10A	Arabidopsis thaliana
5mmPR-A1D01S	Galactokinase like protein	Arabidopsis thaliana
5mmPR-A1D02S	Mitochondrial phosphate transporter	Oryza sativa

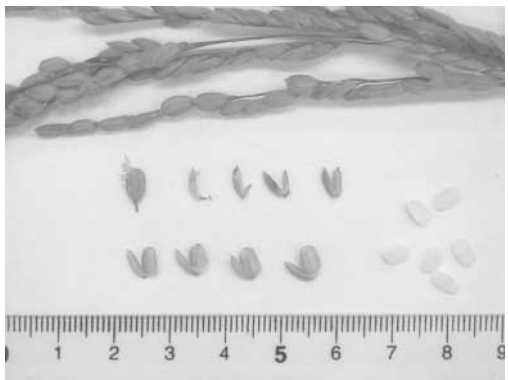

Fig 8. rice seeds

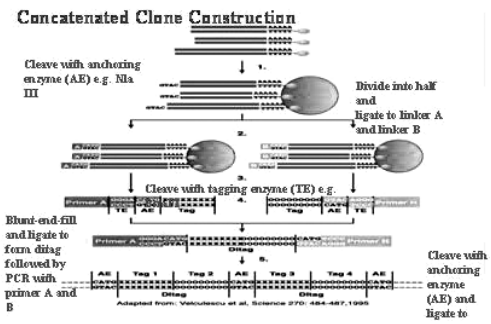

Fig 9. Overview of SAGE Technique

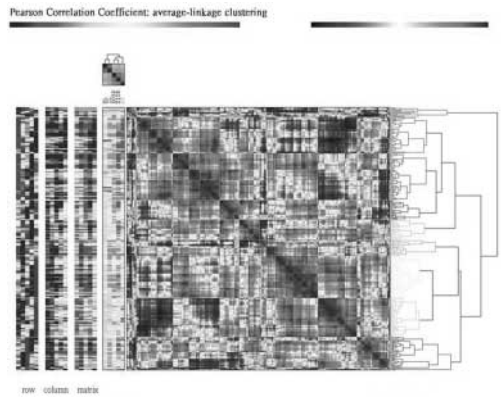

Fig 10. Frequencies of expressed gene in grain (x axis) and root (y axis) library. Genes with 0 count in one of the library were not plotted.

SAGE (serial analysis of gene expression) project

In order to understand the spatial and temporal expression of entire set of mRNA trancripts (transcriptome) during rice root, seed and flower development, we carried out serial analysis of gene expression (SAGE) profiling of transcripts derived from rice

Table 2. Protein synthesis machinery

	Root		2-4 DAP seeds		9-11 DAP seeds	
	Counts	%	Counts	%	Counts	%
Ribosomal protein	765	1.70	197	2.11	66	0.90
IF, EF, etc.	252	0.56	17	0.18	24	0.33
Total	1017	2.26	214	2.29	90	1.22

Table 3. The messages most abundantly expressed

Root		2-4 DAP seeds		9-11 DAP seeds	
RCc3	0.80	Prolamin family	4.14	Prolamin family	10.50
Profilin A	0.50	Globulin family	2.41	Globulin family	4.72
PAL	0.50	Glutelin family	1.11	Glutelin family	3.26
PR-1	0.44	Gly-rich RNA-binding	0.93	Allergen RA5	2.54
Gly-rich RNA-binding	0.39	Allergen RA5	0.92	Protease inhibitor	1.00
Glutathione S-transfer	0.36	Othophosphate dikinase	0.43	Gly-rich RNA-binding	0.72
γ-Tip	0.35	Protease inhibitor	0.41	Allergen RA16	0.62
Water channel protein	0.29	Branching enzyme III	0.30	Allergen RA14	0.53
Metallothionein	0.26	Germin-like prot.	0.27	Allergen RAG2	0.53
Glucan exohydrolase	0.26	Ribosomal prot. L36	0.25	Oleosin ole16	0.50

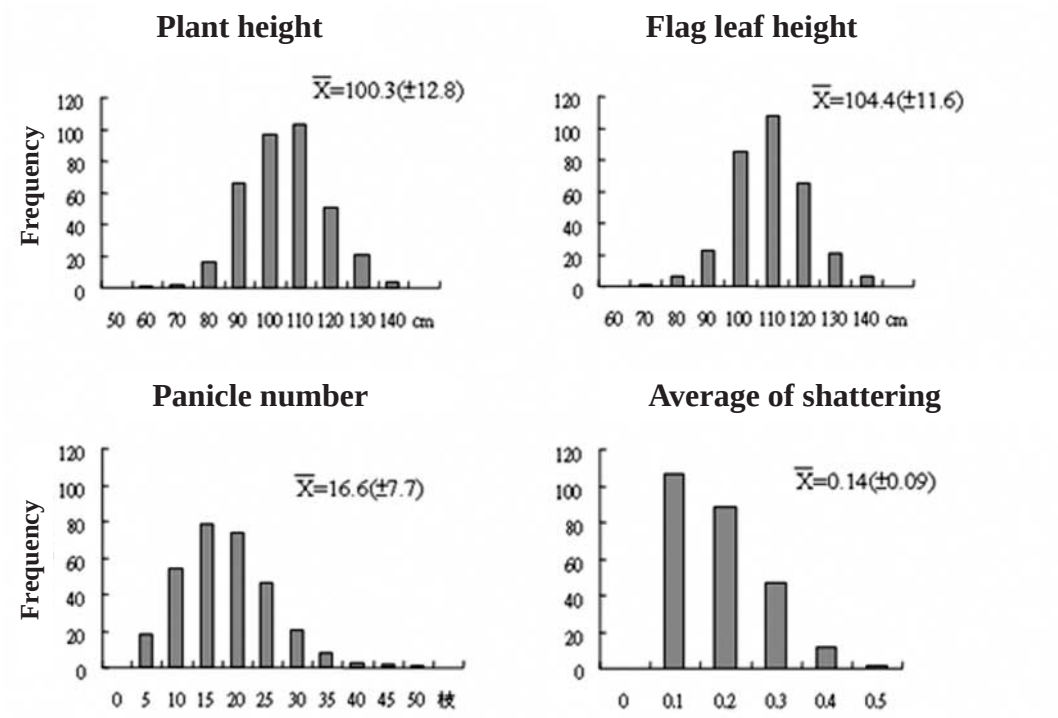

Fig 11. Phenotype distribution in the F2 population of IR1545-339 and Nipponbare. The numbers in parentheses represent the standard deviations for the F2 population.

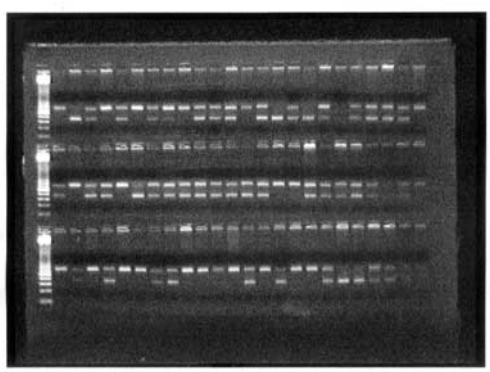

Fig 12. The genotypes of microsatellite marker of RM 207 for subpopulation of F2 progenies. The first lane of each row is 20 bp ladder. The second and third lanes of the first row are IR1545-339 and Nipponbare, respectively. The rest of lanes are the F2 progenies.

2.5% SPR 250V 10min

root, 2-4 DAP and 9-11 DAP seeds. SAGE is both a quantitative and qualitative approach in analyzing transcripts representative for cellular responses. Approximately 17,000 SAGE tags were extracted from rice seeds libraries and analyzed, representing at least 7,584 unique transcripts. The root library contained 45,050 SAGE tags, representing at least 16,266 unique transcripts.

QTL (quantitative trait loci) mapping study

Oryza sativa, Nipponbare and Oryza sativa, IR1545-339 (ssp. indica) were used as the parents in this study. The F2 population was used for phenotypic measurements of plant height, flag leaf height, panicle number average of panicle length, seed-set, panicle weight and total grain weight (Fig. 11). The genotypes of the 320 F2 individuals for six polymorphic microsatellite markers (Fig. 12) have been used to construct a partial

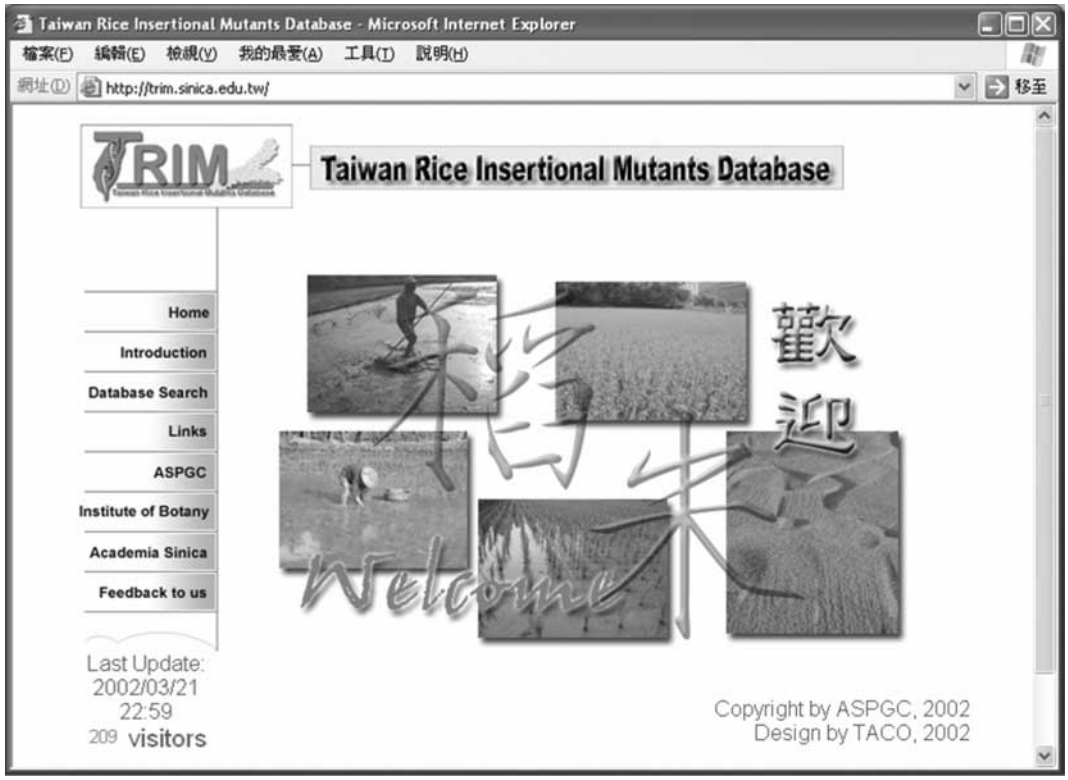

Fig 12. Web site of TRIM database.

molecular linkage map of rice chromosome 2. An interval mapping approach will be used to estimate the effects and locations of QTLs .

Rice Insertional Mutants Project

Knock-out mutants (naturally occurring alleles, deletion mutants or insertion mutants) are the central dogma of functional genomics study for all organisms. We also join a rice activation-tagging/ knockout mutants project, and focus on the high throughput sequencing of the region flanking to the insertion site. We'll also keep these flanking sequences and annotated information in the TRIM database (Taiwan rice insertional mutants database) (Fig. 12). Researchers may search with any keywords or their query sequence of interest in this FST (flanking sequence tags) database. The URL address of TRIM is http://trim.sinica.edu.tw/. This should provide very helpful tools for the rice functional genomic research.