

# Plan for the day

---

- Objectives
  - Update each other
  - Share materials/results
  - Identify tangible results
  - Possible joint experiments
  - Context for looking at the future of GCP
- SP2 presentations
  - Organized by Outputs
  - Discussion
    - Specialized genetic stocks
    - Maps and markers
    - Validate candidate genes
- Parallel sessions
- Brainstorming on Day 3

## SP2 project presentation

8:10-8:20	<b>Output 1. Assembly of genomics and germplasm resources</b>	Wheat genetic stock assembly and utilization	<b>Hans-Joachim Braun, CIMMYT</b>
8:20-8:30		Legume mutant resource development/drought mapping in bean	<b>Mathew Blair, CIAT</b>
8:30-8:40		Tuber genetic stocks and gene function validation tools	<b>Marc Ghislain, CIP</b>
8:40-9:00		Stress response-enriched EST resources for targeted species (cowpea and pearl millet)	<b>Sarah Hearne, IITA, and Tom Hash, ICRISAT</b>
9:00-9:30		<b>Discussion on specialized genetic stocks/resources</b>	
9:30-9:50	<b>Output 2. Comparative maps and markers within and across species</b>	Validation of conserved orthologous markers	<b>Merideth Bonierable CIP and Jizeng Jia, CAAS</b>
9:50-10:00		QTL mapping for drought tolerance	<b>Nour Ahmadi, CIRAD</b>
10:10-10:30		Targeted <i>Musa</i> genome sequencing and frame map construction	<b>Takuji Sasaki, NIAS and Nicolas Roux, INIBAP</b>
10 :30		<b>Coffee break</b>	
10:45-10:55		Genome-wide SNP discovery across multiple rice varieties	<b>Ken McNally, IRRI</b>
10:55-11:15		<b>Discussion on markers and maps</b>	

## SP2 project presentation

11:15-11:35	<b>Outputs 3 and 4</b>  <b>Assign genes and pathways to phenotypes</b>  <b>Gene function validation</b>	Functional genomics of cross-species resistance to fungal diseases in rice and wheat (CEREALIMMUNITY)	<b>Pietro Piffanelli, CIRAD</b>
11:35-11:55		Determination of a common genetic basis for tissue growth rate under water-limited conditions across plant organs and genomes	<b>Mark Sawkins, CIMMYT</b>
11:55-12:10		Crop gene expression profiles and stress-gene arrays	<b>Tiegang Lu, CAAS</b>
12:10-12:30		Identifying Genes Responsible for Failure of Grain Formation in Rice and Wheat under Drought	<b>John Bennett, IRRI</b>
12 :30-13 :30		<b>Lunch</b>	
13:35		Isolation and Characterization of Aluminum Tolerance Genes in the Cereals: An Integrated Functional Genomic, Molecular Genetic and Physiological Analysis	<b>Jurandir Magalhaes, Embrapa</b>
13:55-14:45		<b>Discussion on candidate genes and their validation</b>	<b>Whole group and invite Shoshi Kikuchi, NIAS, and Richard Bruskiwich, IRRI</b>
14:45-15:15	<b>Whole subprogram discussion across all outputs</b>	Linkages and common themes Identify intermediate, useful products Items for breakout and brainstorming sessions	

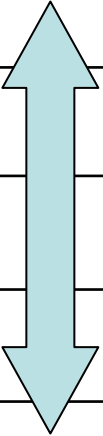
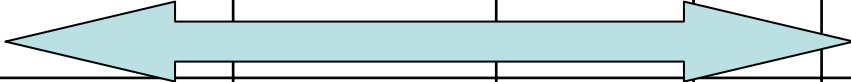
# Discussion

- Specialized genetic stocks
  - Creation and maintenance
  - Selective use
- Maps and markers
  - Orthologs
  - Polymorphisms and “mappability”
  - SNP
- QTL and expression convergence (genome-wide scan)
  - Introgression lines and transcript maps
- Validation of candidate genes (targeted genes)
  - Functional SNP
  - complementation

# Parallel discussion sessions

1. Strategies for mining diversity and how to cope with large collections
2. Strategies for allele discovery and how to cope with complex traits
3. Reaching the breeders

# Convergence of expression polymorphism, QTL, SNP, and phenotyping data

Crop-specific experiment	Trait-specific expt	Phenotype (developmental stage, conditions)	Differentially expressed gene list	Gene Ontology	QTL	SNP	Statistical test for association (Fischer Exact Test, correlation, others)
	Diseases	Different diseases					
	Drought	Vegetative vs reproductive					
	Salinity						
	Hormone						
	etc						
							

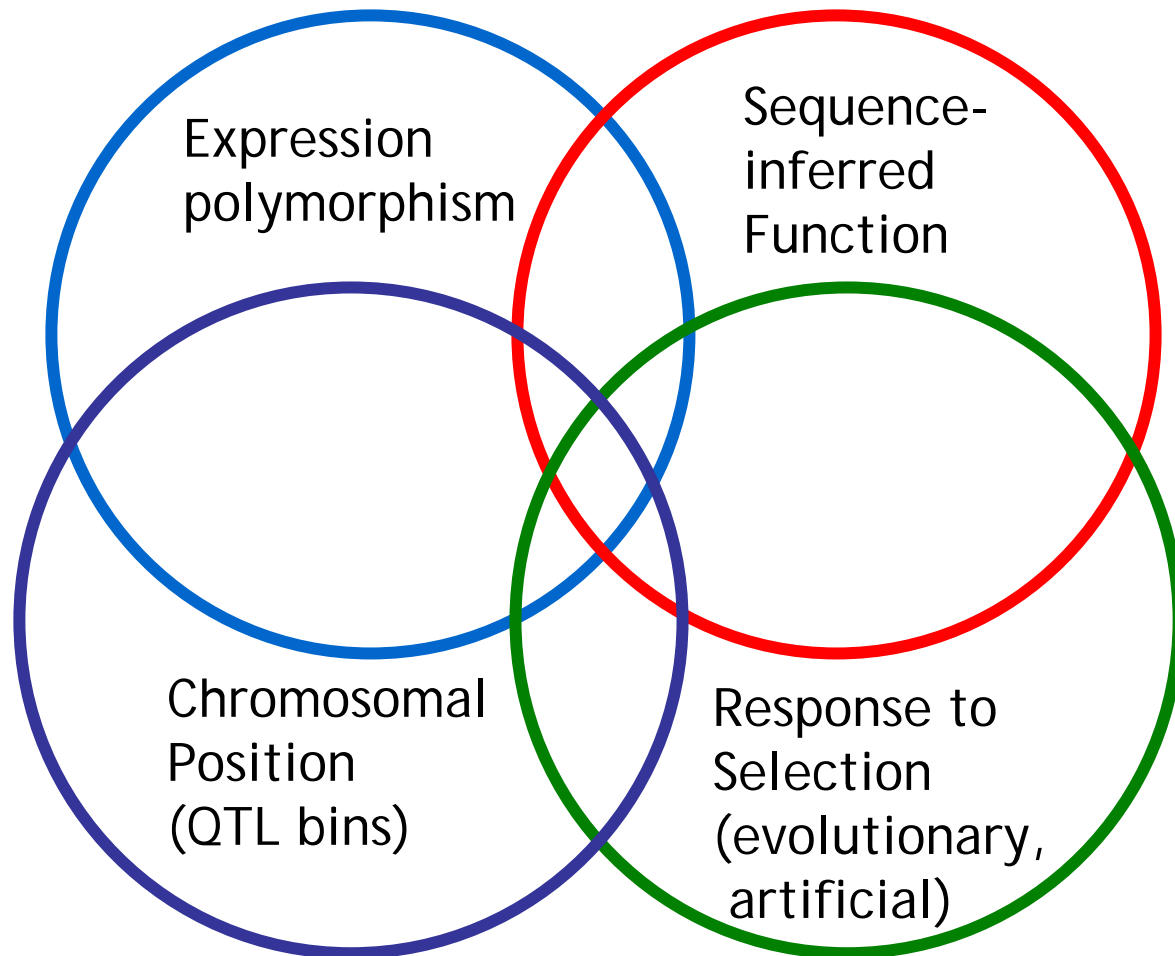


# Parallel discussion sessions

1. Strategies for mining diversity and how to cope with large collections
2. Strategies for allele discovery and how to cope with complex traits
3. Reaching the breeders

# Candidate genes: 4-Ring Circus model of gene discovery

---



Converging evidence for candidate gene validation

# “Open-search” and “short-list” candidate genes

## Broad and genome-wide search for candidates

- Convergent approach
- Reiterative process of **binning** and **cross-mapping** to narrow candidate gene list
  - Segregating populations
  - Introgression lines, CSSL
  - Heterogeneous inbred lines (pair of contrasting segregants from a heterozygous (genes or chromosomal regions) line)
  - NIL
- Orthologs as links across genomes

## Short-list candidates

- **Allelic diversity of orthologous candidate genes**
  - Allelic states (haplotypes)
  - Functional SNP
- Validation
  - Marker pyramids (multiple backgrounds)
  - Transformation
  - Over- and under-expression (available mutant stocks, Arabidopsis, rice, maize, RNAi)
- Field performance

# A database of candidate gene diversity in GCP crops

- **Convergent evidence to suggest roles in key agronomic traits.**
  - co-localisation with QTLs
  - expression patterns
  - possible physiological roles inferred by sequence annotation,
  - response to directional selection relative to the target traits.
- **Allelic polymorphism of candidate genes in a particular crop,**
  - within a reference germplasm sample, undergoes proper phenotyping,
  - testing association between observed polymorphism and trait variability.
  - Enable validation of the involvement of specific genes in a trait,
  - Mining favorable alleles within a species.
- **Shortcuts to transfer crop-specific information across species**
- **Shed light on a range of adaptation and polymorphism,**
- **Help resolve functionally important variations from non-functional ones**

**Collective outputs from a range of coordinated research investments.**



# Summary

- Highlights of SP2 results
  - What come out in terms more linkages?
  - Lessons learned
  - Problems encountered; what can be done
- 
- Across GCP
- 
- Parallel sessions
  - Brainstorming sessions
  - Distilling different views,

# Brainstorming session on Day 3

- **What outputs should the GCP have produced by the end of its halfway point, 2008?**
- ***develop a list of 8-10 specific outputs, indicate the category you believe they belong to, and add any new categories you think are suitable, taking into account:***
  - 1) what the GCP should be doing to ensure its niche as a unique program,***
  - 2) what will benefit the largest numbers/proportion of poor people, and***
  - 3) what is realistically achievable and deliverable to users in a 2-3 year framework with the available GCP funds (roughly \$13 million per year for 2006-2008)***
- **How should these 8-10 specific outputs, given the issues identified in response to question 2, be grouped into research projects (competitive and/or commissioned)?**

# Assessing the degree of concordance between expression and genetic data using genotype-specific and generic expression datasets

Expression data with different levels of specificity

Genotype-specific and trait-specific expression

Trait-specific: e.g., stress responses of different genotypes to different stresses

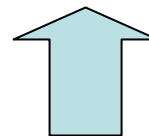
Crop-specific expression: Within single species

Datasets from multiple crops/between different species



Genetic data from specific genotypes and populations segregating for target traits (e.g. disease resistance, drought tolerance)

QTL mapping    Introgression lines

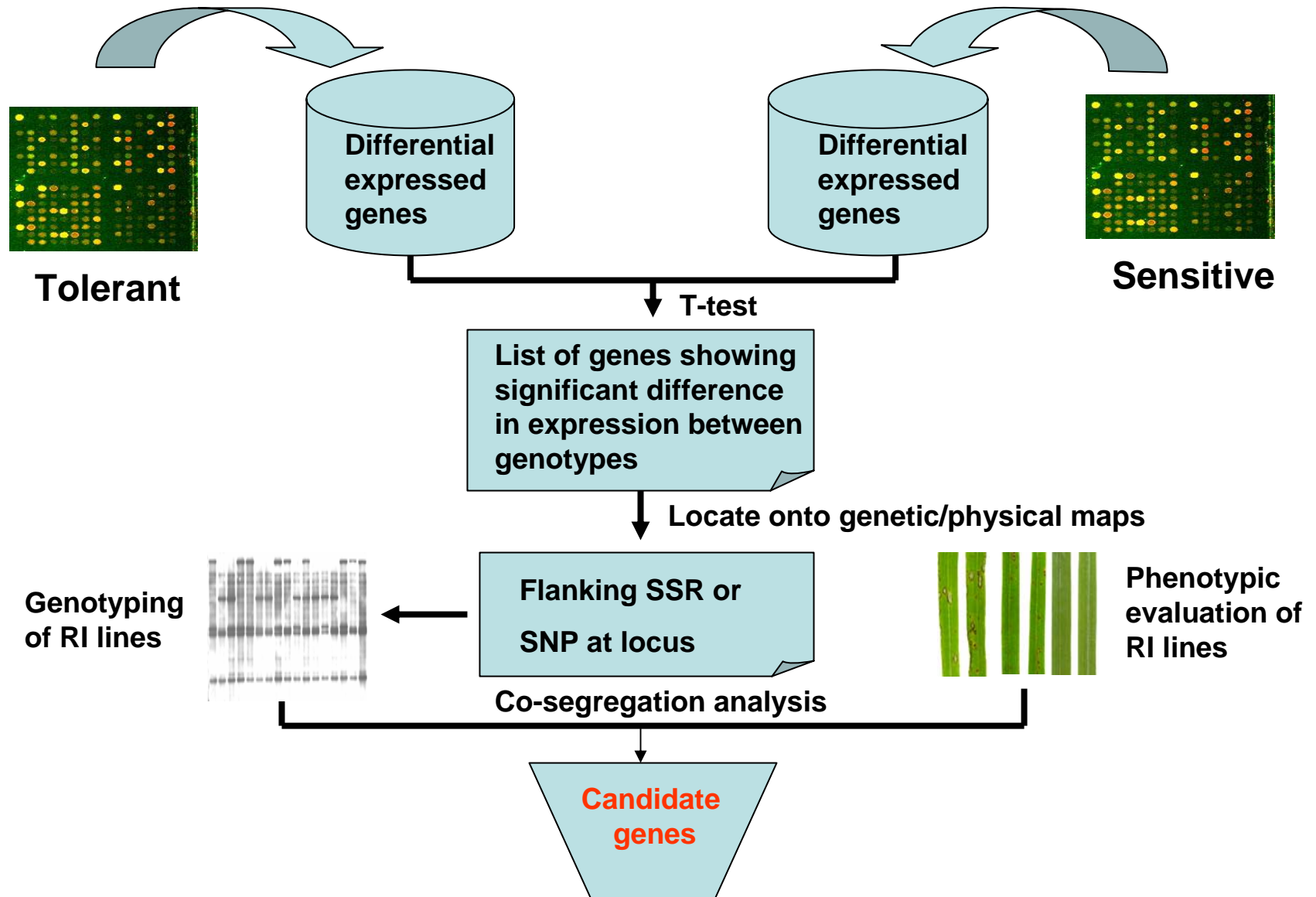


Expression data from genotypes under stress

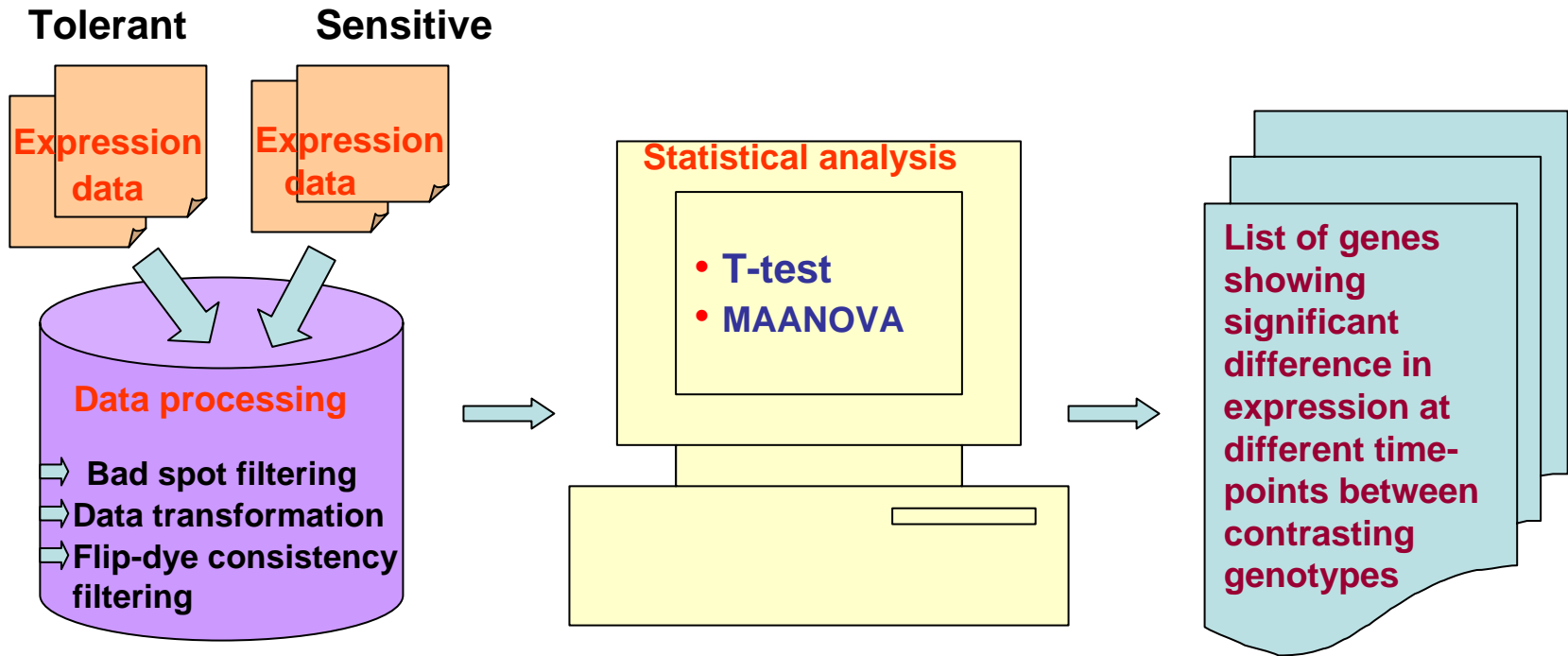
Expression data from other genotypes but not limited to stress

Expression data from related species (maize, wheat)

# Co-segregation analysis of gene expression and phenotypes to identify genes for stress tolerance



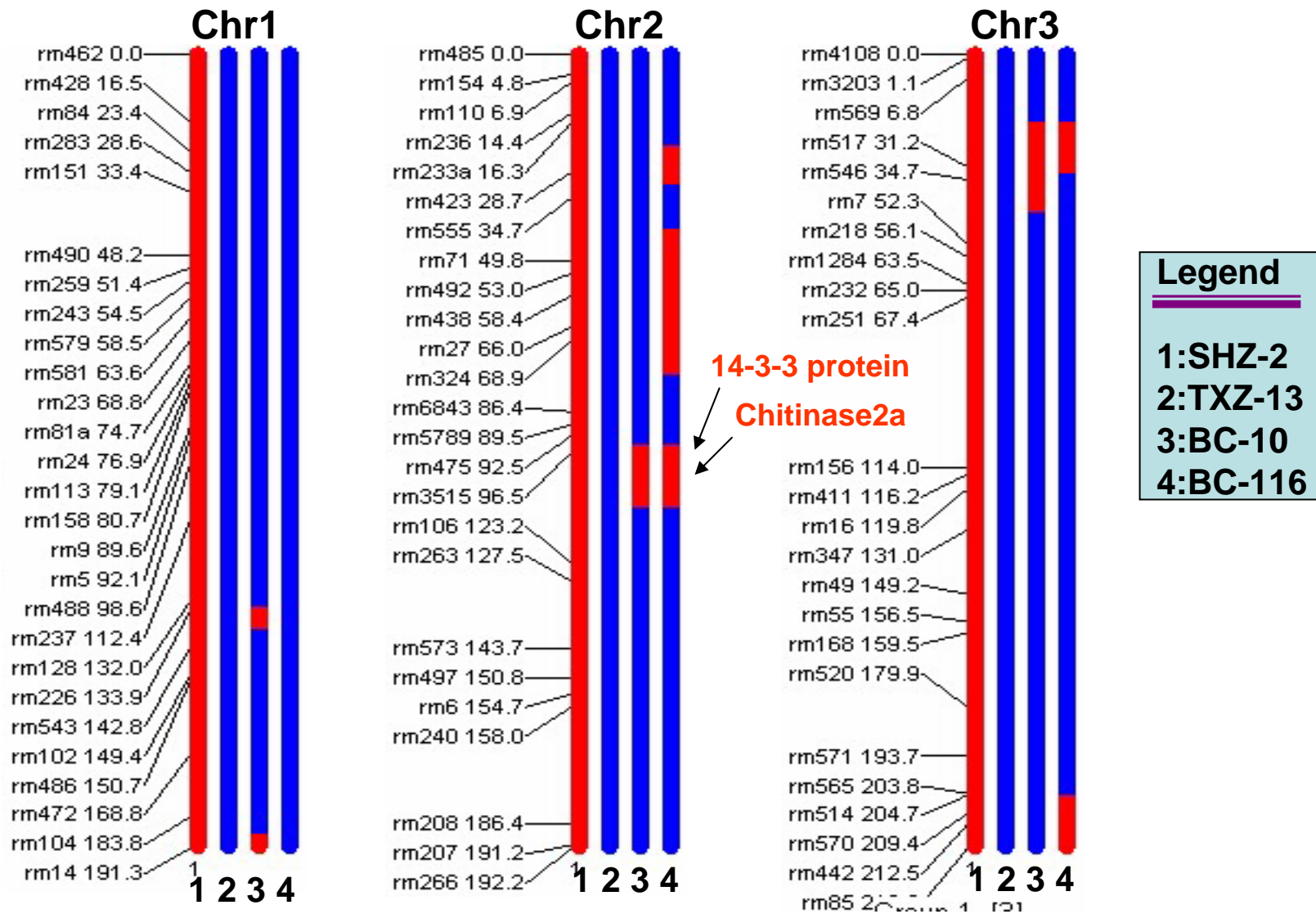
# Data processing and statistical analysis in the microarray experiment



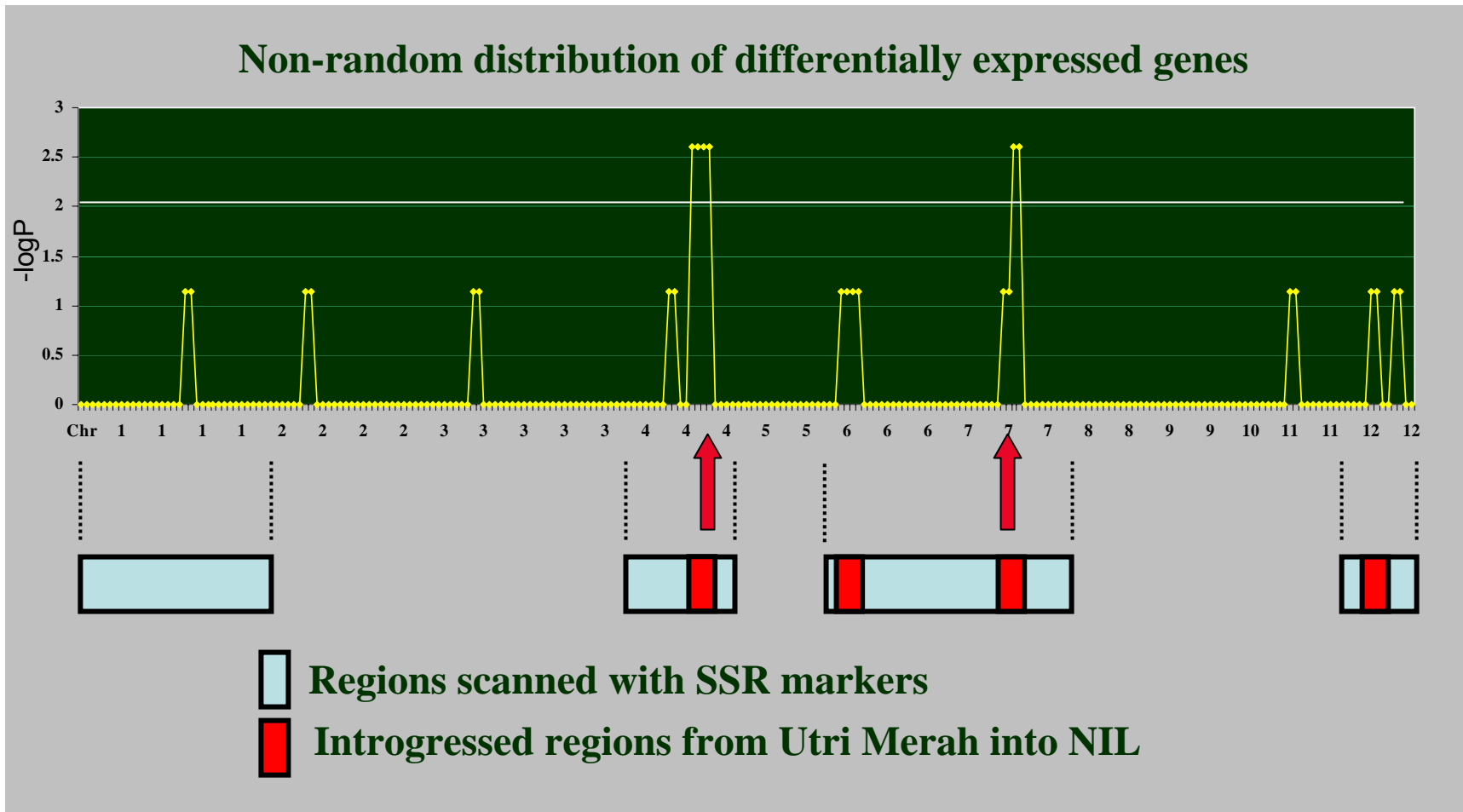
## Experimental Data: association with blast resistance and expression pattern of identified candidate genes

Gene ID	Gene name	Chr	Association with blast resistance		Expression pattern	
			R <sup>2</sup> (%)	P	Relative ratio at 48h	Relative ratio at 96h
AK062495	Subtilisin-cymotrypsin inhibitor	1	5.50	0.0215	1.59 <sup>***</sup>	0.49 <sup>**</sup>
AK111087	myb protein	2	7.26	0.0090	1.45 <sup>***</sup>	1.28 <sup>*</sup>
AK101841	Selenium-binding protein	2	6.88	0.0326	NS	2.15 <sup>***</sup>
AK109729	Unknown protein	3	3.61	0.0451	1.56 <sup>***</sup>	2.26 <sup>***</sup>
AK069447	Transketolase	4	3.33	0.0432	2.49 <sup>***</sup>	0.73 <sup>**</sup>
AK105947	Dehydration-responsive protein	5	3.15	0.0771	20.06 <sup>***</sup>	NS
AK107868	Unknown protein	8	27.2	0.0001	1.79 <sup>**</sup>	2.14 <sup>***</sup>
AK060990	Cytochrome P450	8	23.88	0.0001	10.27 <sup>***</sup>	NS
AK060251	Far-red impaired response protein	8	20.26	0.0001	0.42 <sup>***</sup>	NS
AK063193	Unknown Protein	10	3.49	0.0683	2.31 <sup>***</sup>	2.01 <sup>**</sup>
AK103834	Sumergence induced protein 2	10	2.87	0.0991	2.17 <sup>***</sup>	1.57 <sup>**</sup>

# Graphical genotyping of SHZ-2, TXZ-13, BC-10 and BC-116



# *Association between the Introgression and the Clustering of Differentially Expressed Genes*



Contrasting a pair of NILs (BC6) with R and S response to tungro virus  
Expression patterns across 1 to 12 rice chromosomes  
Considered significant differential expression if  $-\log P > 2$

# Dealing with low polymorphism in conserved orthologous candidate genes

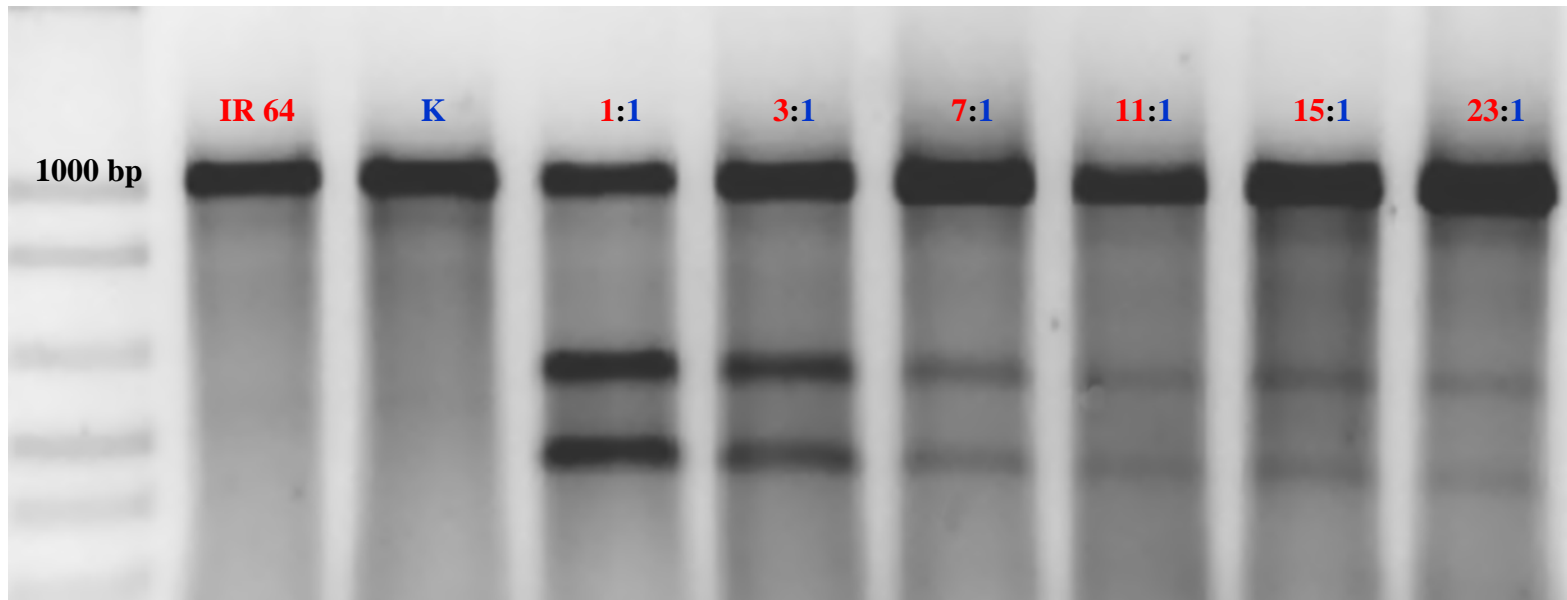
---

- Cross-species conserved markers usually lack polymorphism
- Go after SNP
  - Upfront investment
  - TILLING-type approach to identify SNP haplotypes, → reduce sequencing
  - Simple method to practice in most labs
    - Agarose-gel SNP detection

## Steps involved in TILLING

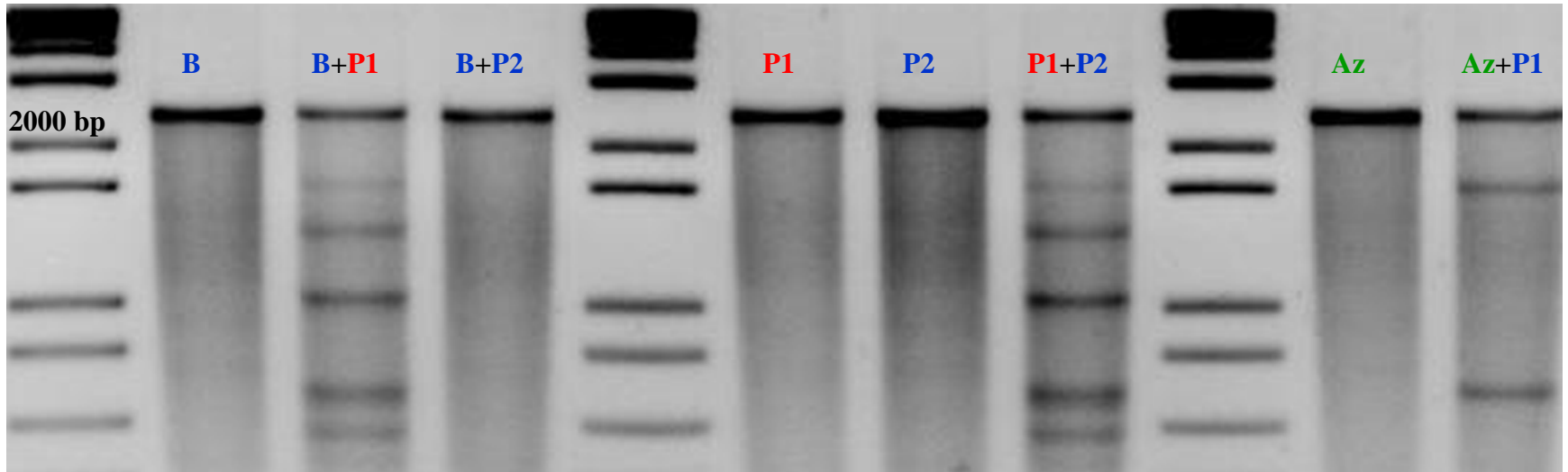
- DNA extraction
- Quality check; quantification; normalization
- Pooling DNA
- PCR amplification of genomic region under query
- Creating heteroduplexes
- Cleaving the heteroduplex at the point of mismatch
- Detection of the cleaved products
- Sequencing the full-length amplicon

# Pooling efficiency on agarose



TPP gene – SNP between IR64 (*indica*) and Kun Min Tsieh Hunan (K) (*indica*)

# Application in mapping



B – NIL (BC3F3) – BC10

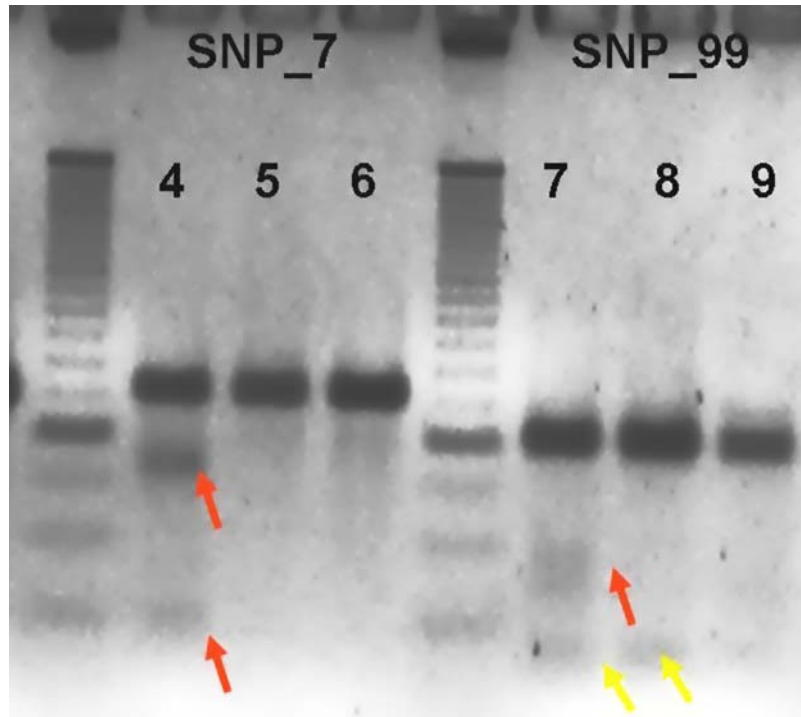
P1 – SHZ-2 (*indica*)

P2 – TXZ-13 (*indica*)

P3 – Azucena (*japonica*)

# SNP-based markers

## Moving from rice to other crop species



EcoTILLING in maize  
Red-SNP  
yellow - artifact

Currently detecting SNPs in orthologs  
of defense-related genes (rice-maize)

# DETECTION

## LiCor genotyper

- Labeled primers
- Amplicon 1 kb
- Cleaning and concentrating PCR product
- PAGE
- LiCor genotyper
- Software GIMP (Gnu Image processor) & SQUINT
- Identifies position and number of SNPs in most cases but not all
- Time consuming and expensive

## Agarose gel based detection

- Unlabeled primers
- Can span ~2-2.5 kb
- No cleaning or concentrating PCR product
- Agarose
- Regular electrophoresis
- Does not identify the position or number of SNPs just make a detection call
- Rapid and cheap

# SUMMARY

- Shorter time
- Cost effective
- Simple
- More participation
  - NARS; Small scale projects

Broader applications of SNP-based markers  
for germplasm characterization and  
mapping studies



Association analysis of 1MB  
window of rice genome with gene  
expression results in TW16-69

# Intro

- Order, linkage grouping, and physical distance between each KOME flcDNA is known
- We already determined which genes are significantly expressed for the entire oligo set
- We wanted to determine the statistical association between 1MB genome regions with significant gene expression

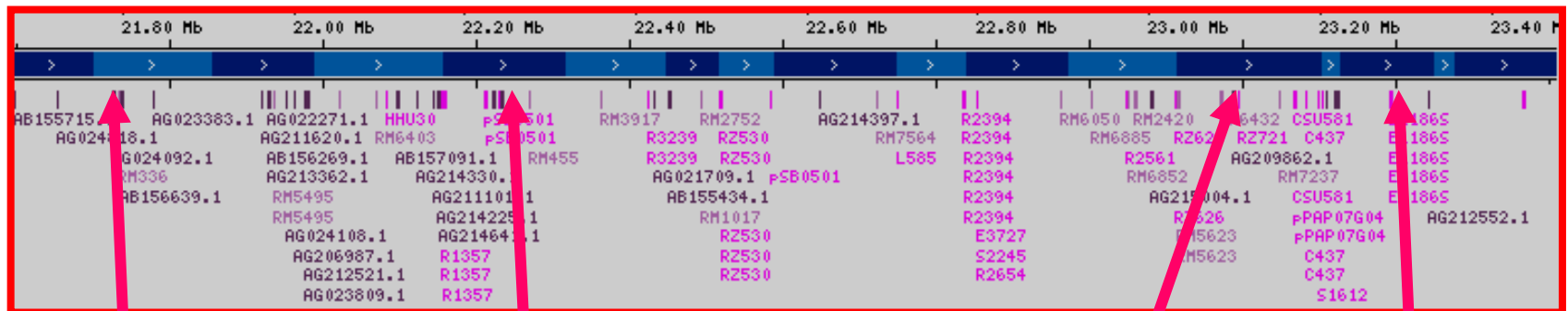
# Analysis algorithm

- Divide the genome into 1 MB windows (regions) & determine the KOME genes inside each region
  - 232 regions in all
- Determine the number of significantly expressed gene vs ns genes for 1 region
  - NS:S gene ratio determined for the region
  - NS:S gene ratio outside this current region also determined
- Test for independence (Chi-sq, Fisher's exact) between the NS:S ratio within the region vs NS:S ratio outside the region
- Repeat test for all 232 regions



# *Constitutive Genes Differentially Expressed between Susceptible and Resistant Plants – Chromosome 7*

Chromosome 7 : 29.7 Mb



**RM336**

**AK064484**

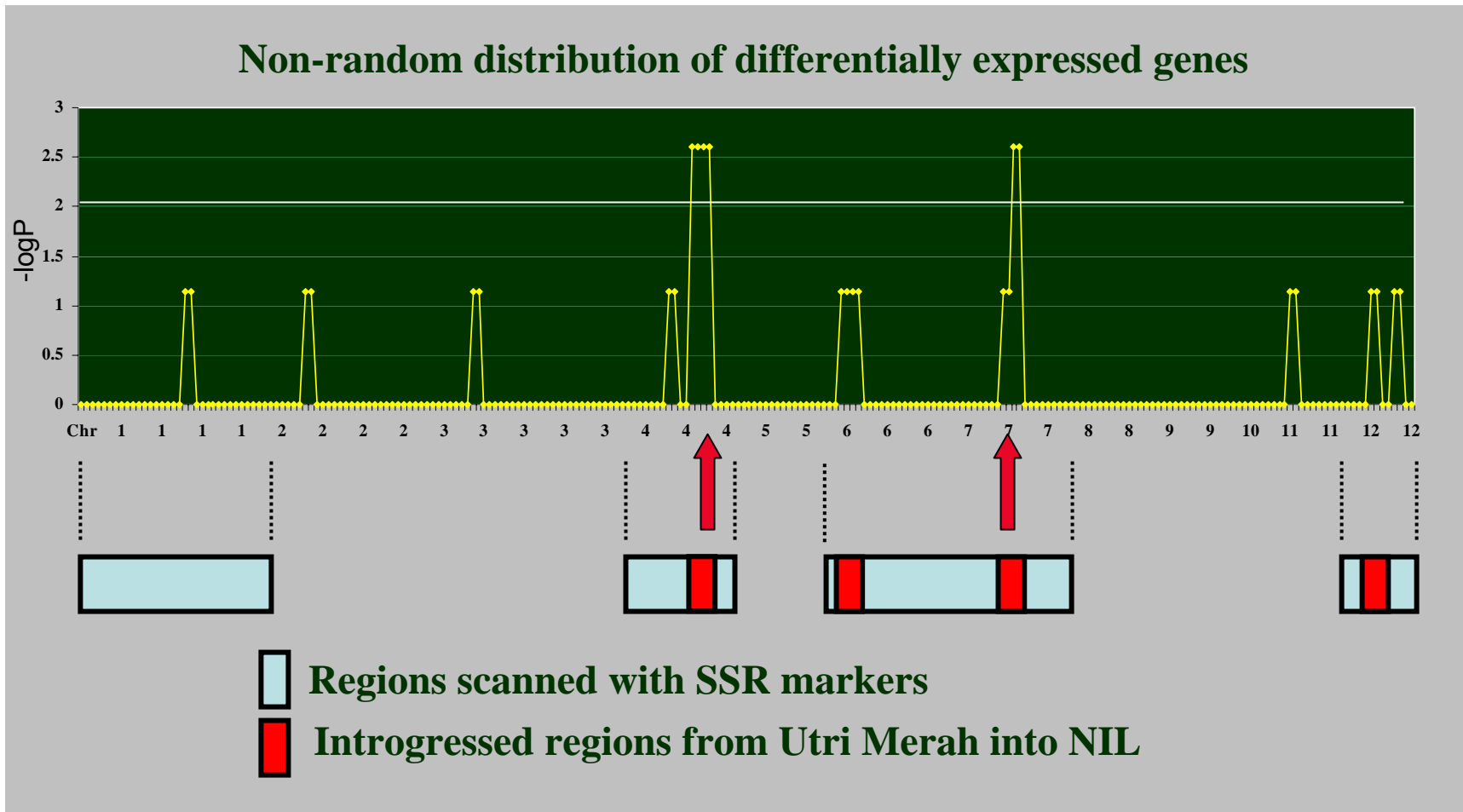
**AK072835**

**AK082833**



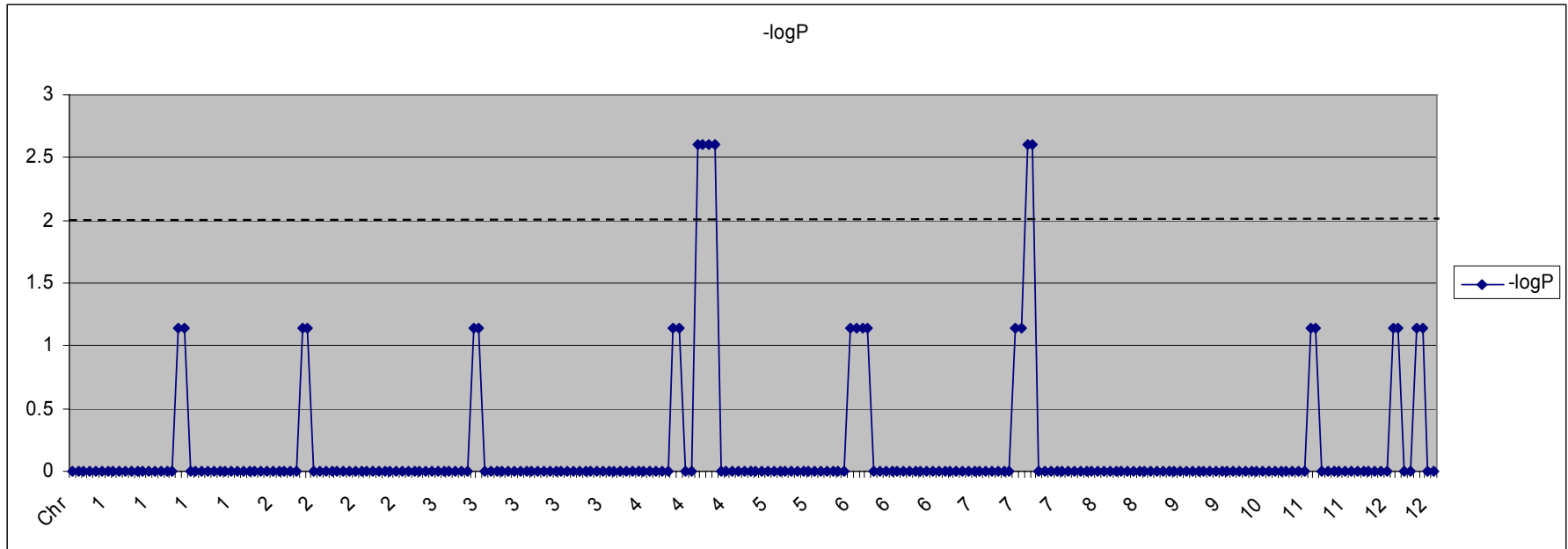


# *Association between the Introgression and the Clustering of Differentially Expressed Genes*



Contrasting a pair of NILs (BC6) with R and S response to tungro virus  
Expression patterns across 1 to 12 rice chromosomes  
Considered significant differential expression if  $-\log P > 2$

# Testing association of transcript map and introgression map (virus resistance as an example)



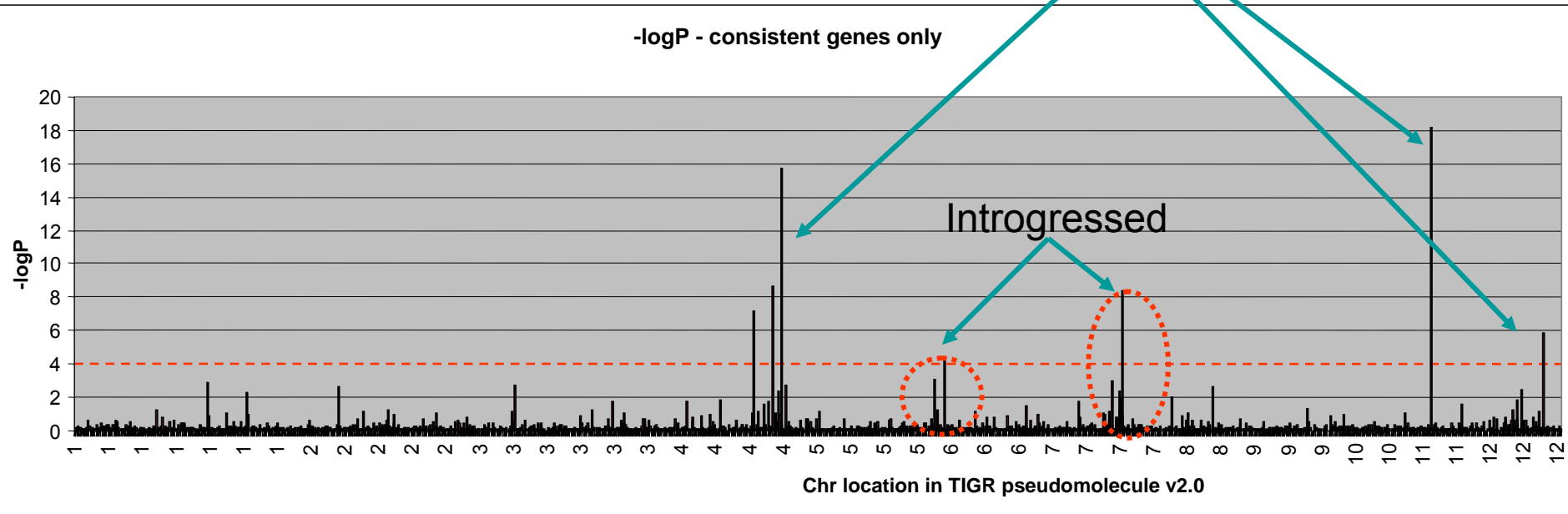
Regions in chr4 & 7 are significantly associated with differential gene expression. 2MB contiguous regions are associated in ch4 while a 1MB region is associated in ch7.

# Association analysis of constitutively expressed genes & introgressed segment

# Background review

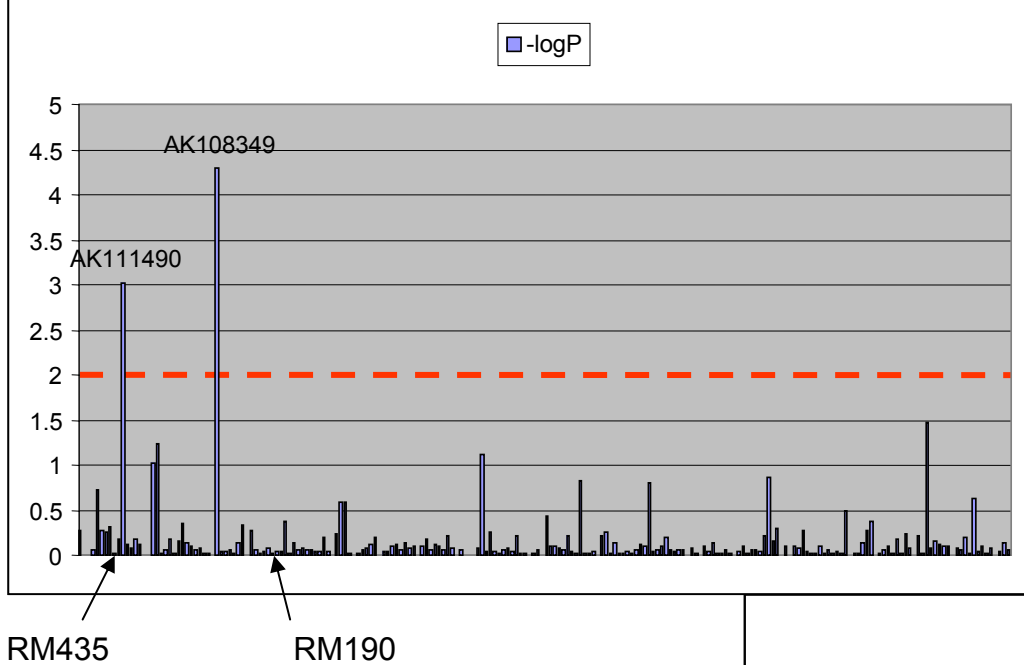
- Genetic Study (Muhsin & Choi, 2004)
  - Two introgressed regions in Chr 6 & 7 associated with resistance to tungro
- Expression study (Choi 2005)
  - 18 genes constitutively expressed (significant at  $p < 0.01$ ) in three experiments
  - 5 genes in chrs. 6 & 7

For verification of introgression



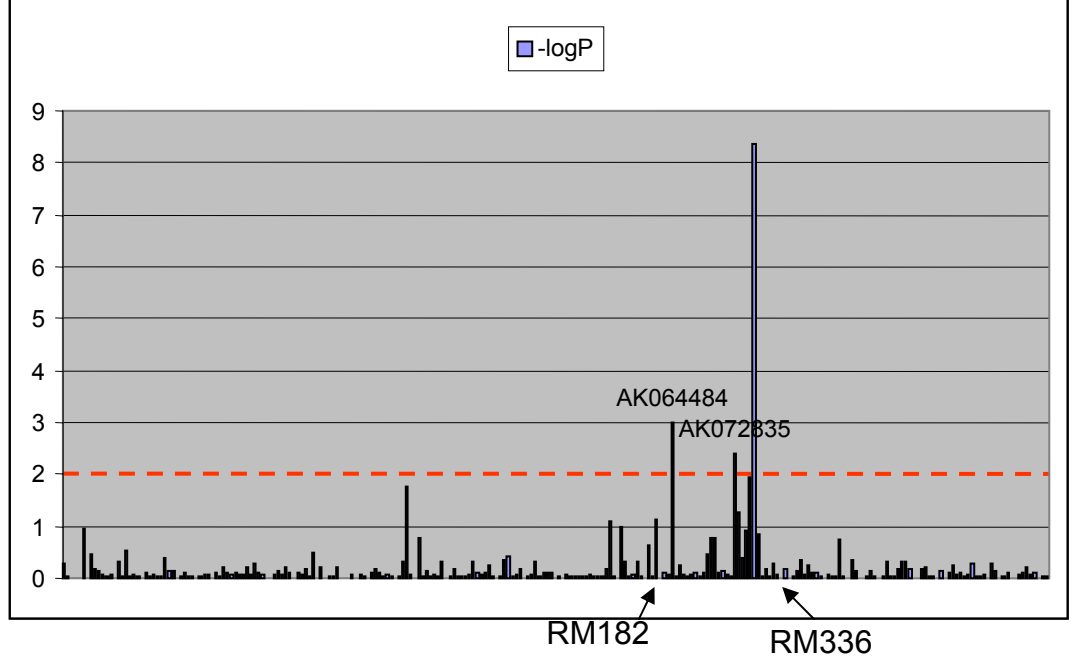
Genome-wide scan of significantly expressed genes. We can localize interesting chromosome regions according to a  $-\log P$  cutoff. In this example, a cutoff of  $\geq 4$  shows interesting regions in chrs 4, 6, 7, 11, and 12

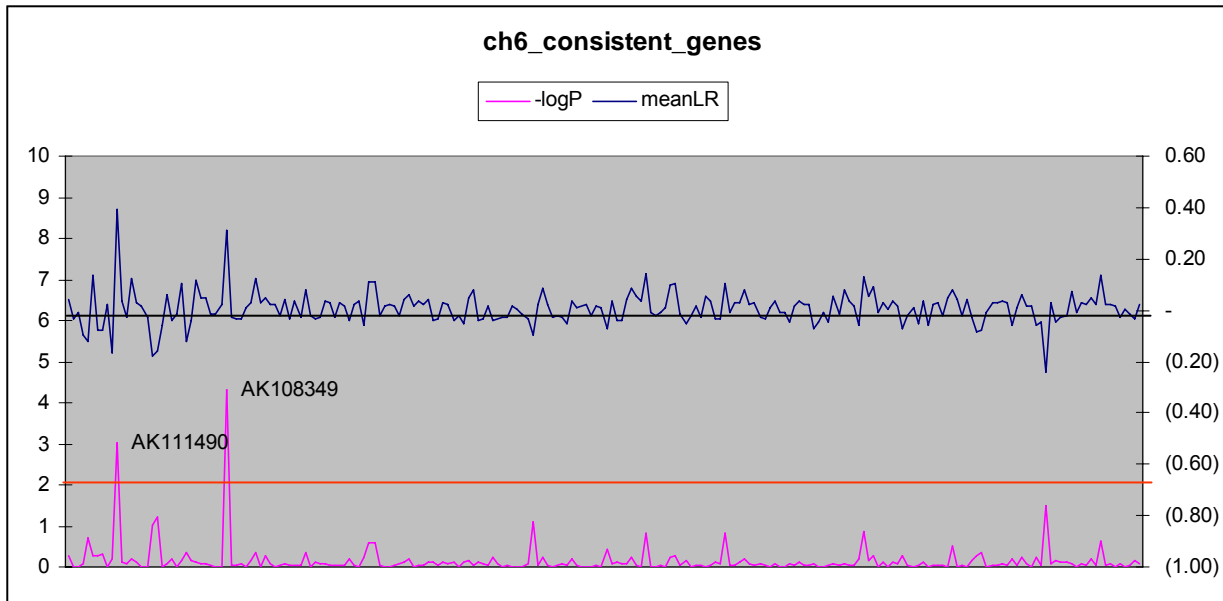
ch6\_logP>2\_introgressed



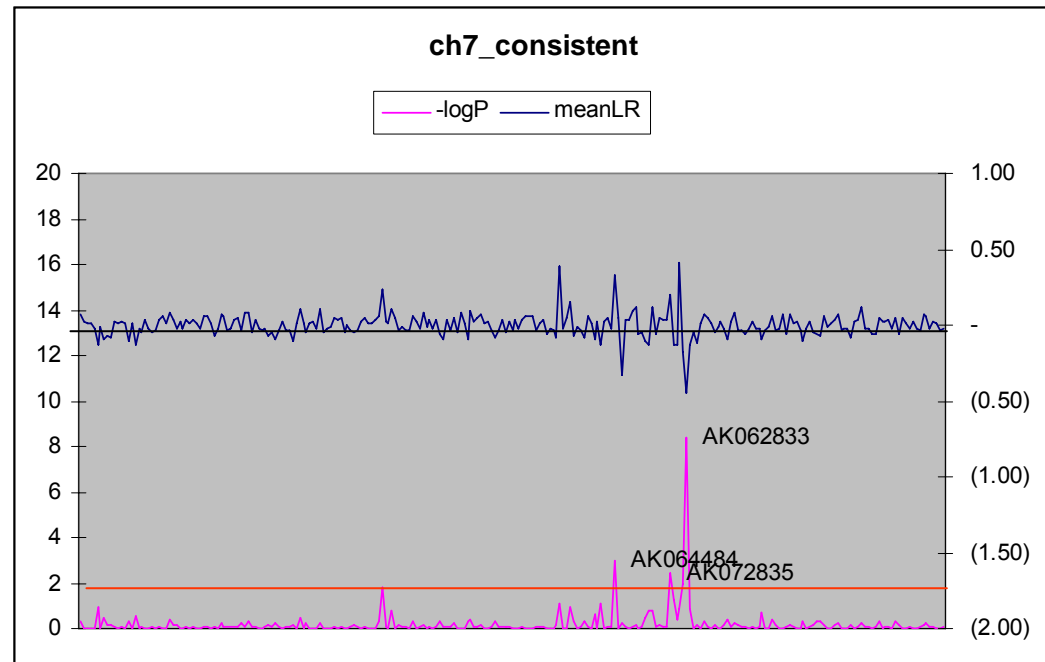
Looking at genes within the introgressed segments, at  $-\log P$  cutoff  $\geq 2$ , some genes show high significance

ch7\_-logP>=2\_introgression





Expression profile of these significantly expressed genes

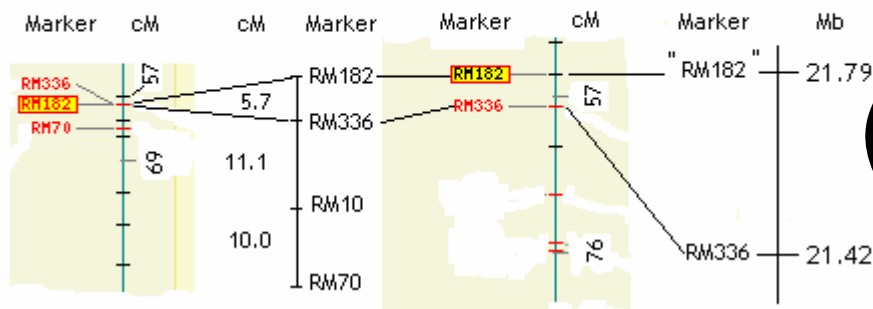


Are these genes statistically associated with the introgressed segments?

# Step 1: Categorize the genes in the microarray platform

Genetic map to pseudomolecule translation of introgression via marker location (E-PCR)

Cornell IR64/Azu DH QTL Chr7    Map of Chr7 This study    M202/IR50 QTL 2003 Chr 7    Physical Map Chr 7 Modified from TIGR gene trans.



22K Agilent genes  
(based on  
KOME  
FLcDNA)

•PM2 region bound by SSRs +/- cM size of introgressed segment

Categorize (~243kb/cM)

Genes outside introgressed segment (21,207)

Genes within introgressed segment (288)

# Is the association statistically significant?

- Step 2. Biological theme analysis via EASE (Hosack et al, 2003)
  - Set of genes under study = 18 constitutively expressed genes
    - 4 within introgression, 14 outside
  - System to compare against: Introgressed-non-introgressed gene bins in 22k chip
    - 288 introgressed: 21,207 non-introgressed

# Contingency table

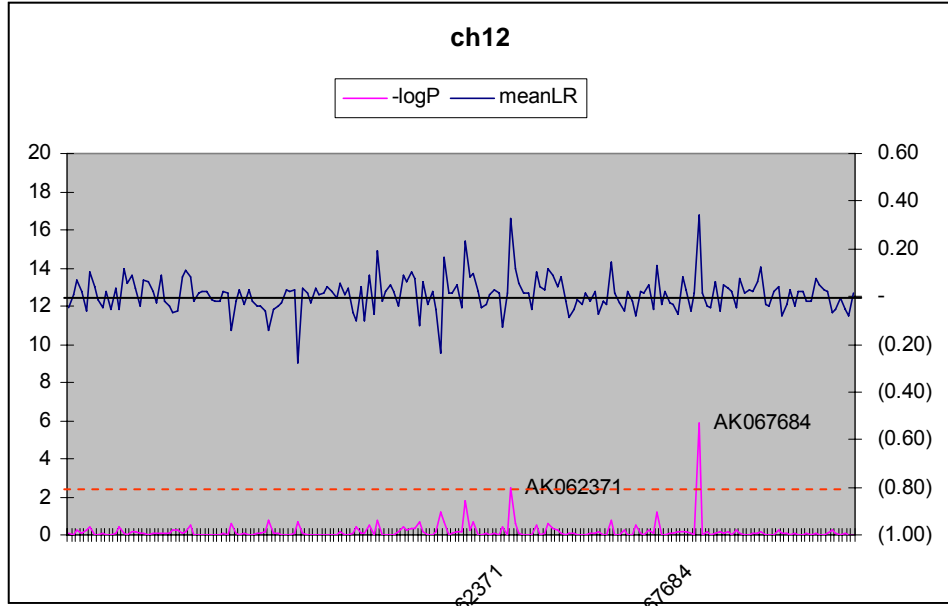
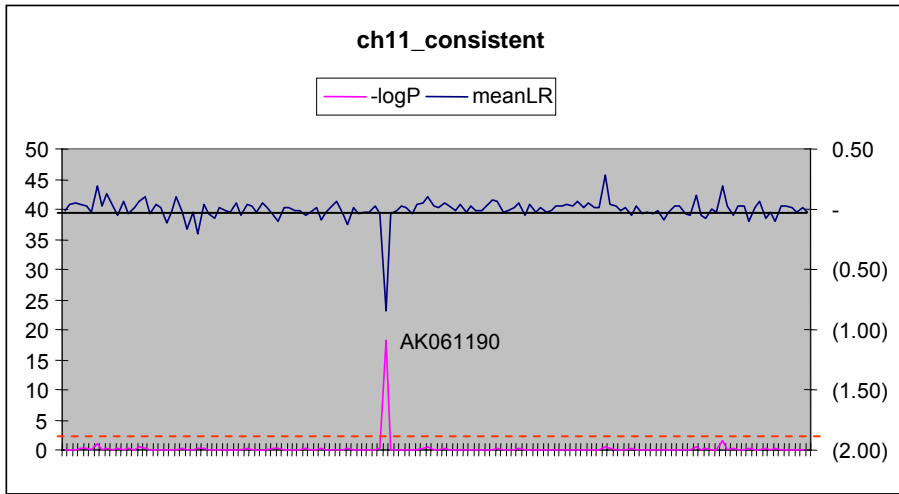
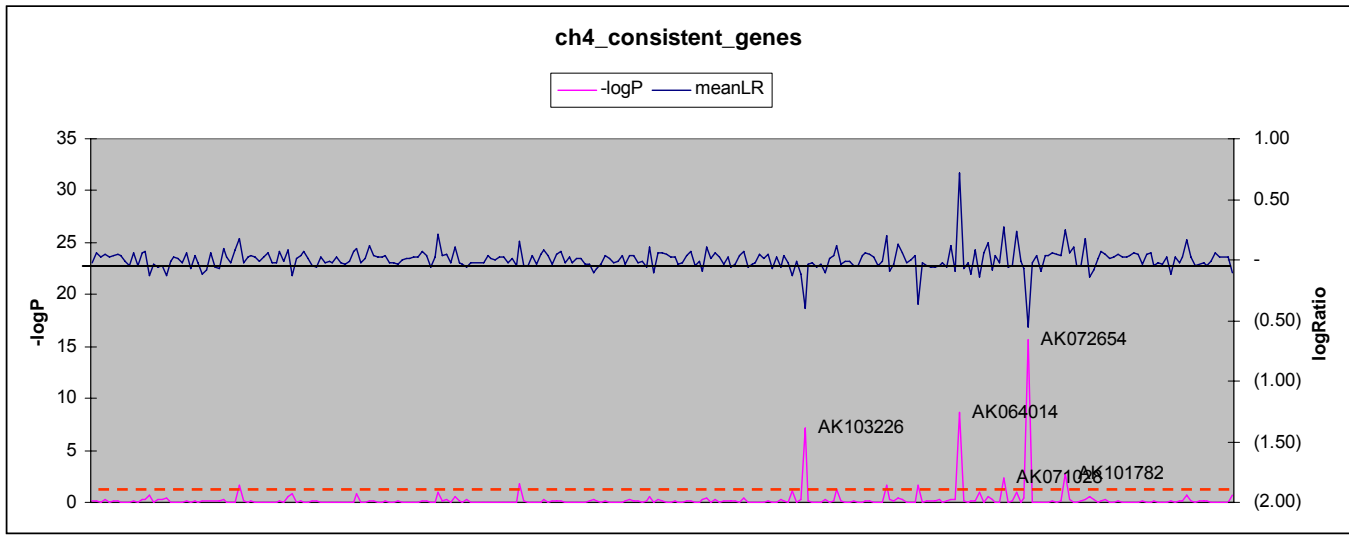
Gene set	Genes within PM2 introgressed segment	Non-associated genes
E: 22K gene chip (n=21,495)	288	21,207
O: Constitutively expressed genes (n=18)	4	14

# EASE result for 18 genes

Gene Category	List Hits	List Total	EASE score (modified Fisher exact test score)	Rice_genes_UID
ch6_7_introgressed	4	18	<b>0.001408245*</b>	AK064484; AK072835; AK108349; AK111490
non_introgressed	14	18	0.999997738	AK061190; AK061754; AK062371; AK062833; AK064014; AK065358; AK067684; AK071028; AK072654; AK100715; AK101245; AK101782; AK103226; AK104668

Of the 18 genes, association of 4 genes on ch6 & ch7(out of 5 in the same segments) with the introgressed region is significant (not by random chance).

Other interesting genome regions  
based on expression profile



Log Ratio and  $-\log P$  plots of interesting regions.  
 The labeled genes have  $-\log P$  val  $\geq 2.0$

# Graphical genotyping of introgressed lines to identify QTL for blast resistance in SHZ-2

