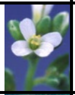
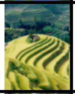
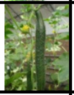
# Plant Genomics and Emerging Techniques

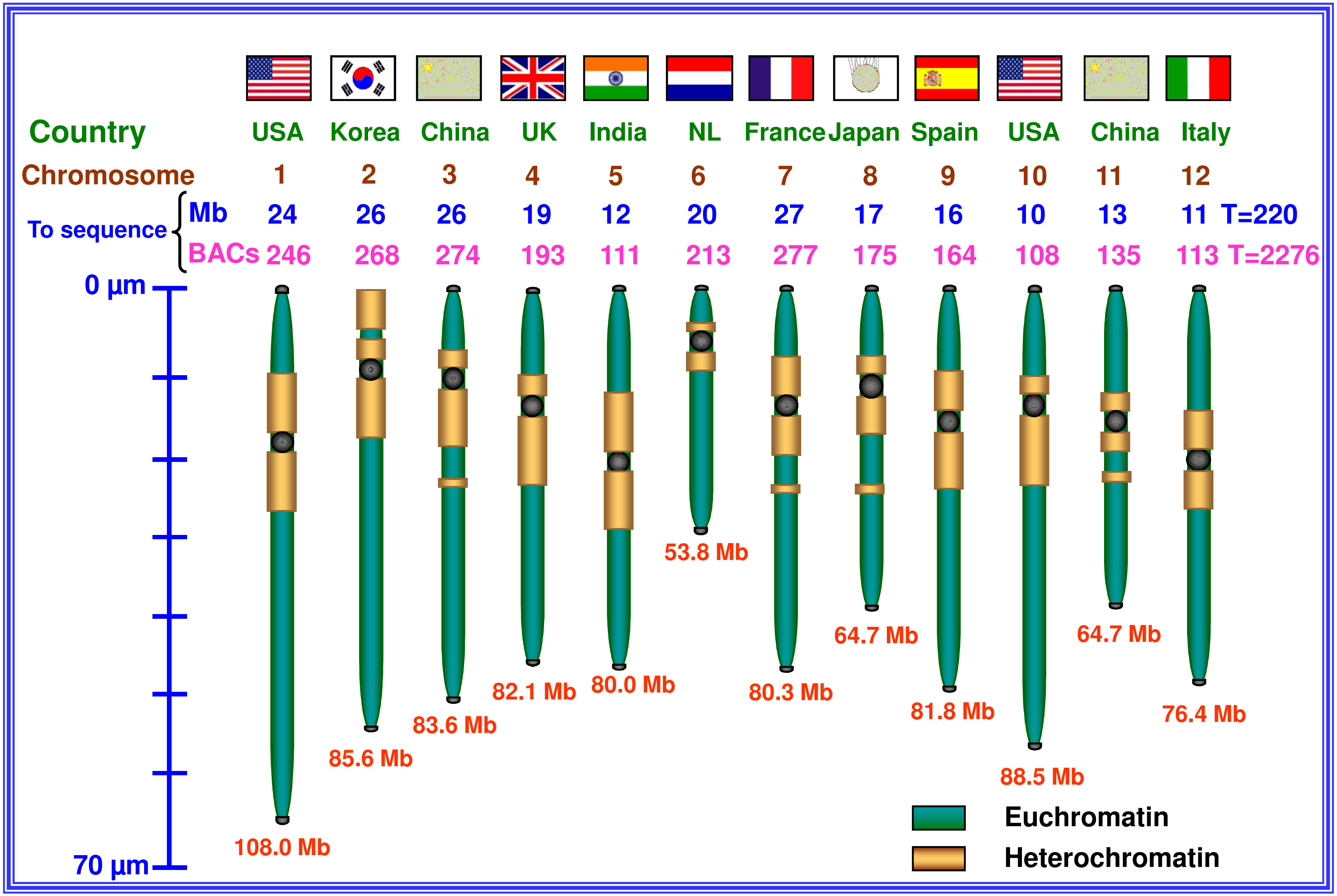
**Jitendra P. Khurana**

Inter-disciplinary Centre for Plant Genomics
Department of Plant Molecular Biology,
University of Delhi South Campus,
New Delhi-110021, India

# Sequenced Plant Genomes

| S. No. | Plant | Genome Size | Sequencing Technology | Reference |
|---|---|---|---|---|
| 1  | *Arabidopsis thaliana* | 125 Mb | Sanger | AGI 2000<br>**Nature** 408:796-815 |
| 2  | *Oryza sativa* | 389 Mb | Sanger | IRGSP 2005<br>**Nature** 436: 793-800 |
| 3  | *Populus* | 550 Mb | Sanger | Tuskan et al. 2006<br>**Science** 313:1596-1604 |
| 4  | *Vitis* | 475 Mb | Sanger | Jaillon eat al. 2007<br>**Nature** 449: 463-467 |
| 5  | *Carica papaya* | 372 Mb | Sanger | Ming et al. 2008<br>**Nature** 452:991-997 |
| 6  | Cucumber | 367 Mb | Sanger + Illumina GA | Huang et al. 2009<br>**Nat. Genet.** 41: 1275-1281 |
| 7  | *Sorghum* | 730 Mb | Sanger | Paterson et al. 2009<br>**Nature** 457: 551-556 |
| 8  | *Zea mays* | 2.3 Gb | Sanger | Schnable, et al. 2009<br>**Science** 326:1112-1115 |
| 9  | *Solanum tuberosum* | 840 Mb | Sanger + Illumina GA | Pre-publication draft release |
| 10  | *Solanum lycopersicum* | 950 Mb | Sanger+ 454 + SOLiD | Pre-publication draft release |

# International Tomato Genome Sequencing Project



| Country | USA | Korea | China | UK | India | NL | France | Japan | Spain | USA | China | Italy | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chromosome | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
| To sequence: Mb | 24 | 26 | 26 | 19 | 12 | 20 | 27 | 17 | 16 | 10 | 13 | 11 | T=220 |
| To sequence: BACs | 246 | 268 | 274 | 193 | 111 | 213 | 277 | 175 | 164 | 108 | 135 | 113 | T=2276 |

🟩 Euchromatin

🟧 Heterochromatin

# International Tomato Genome Sequencing Project

## New Initiative

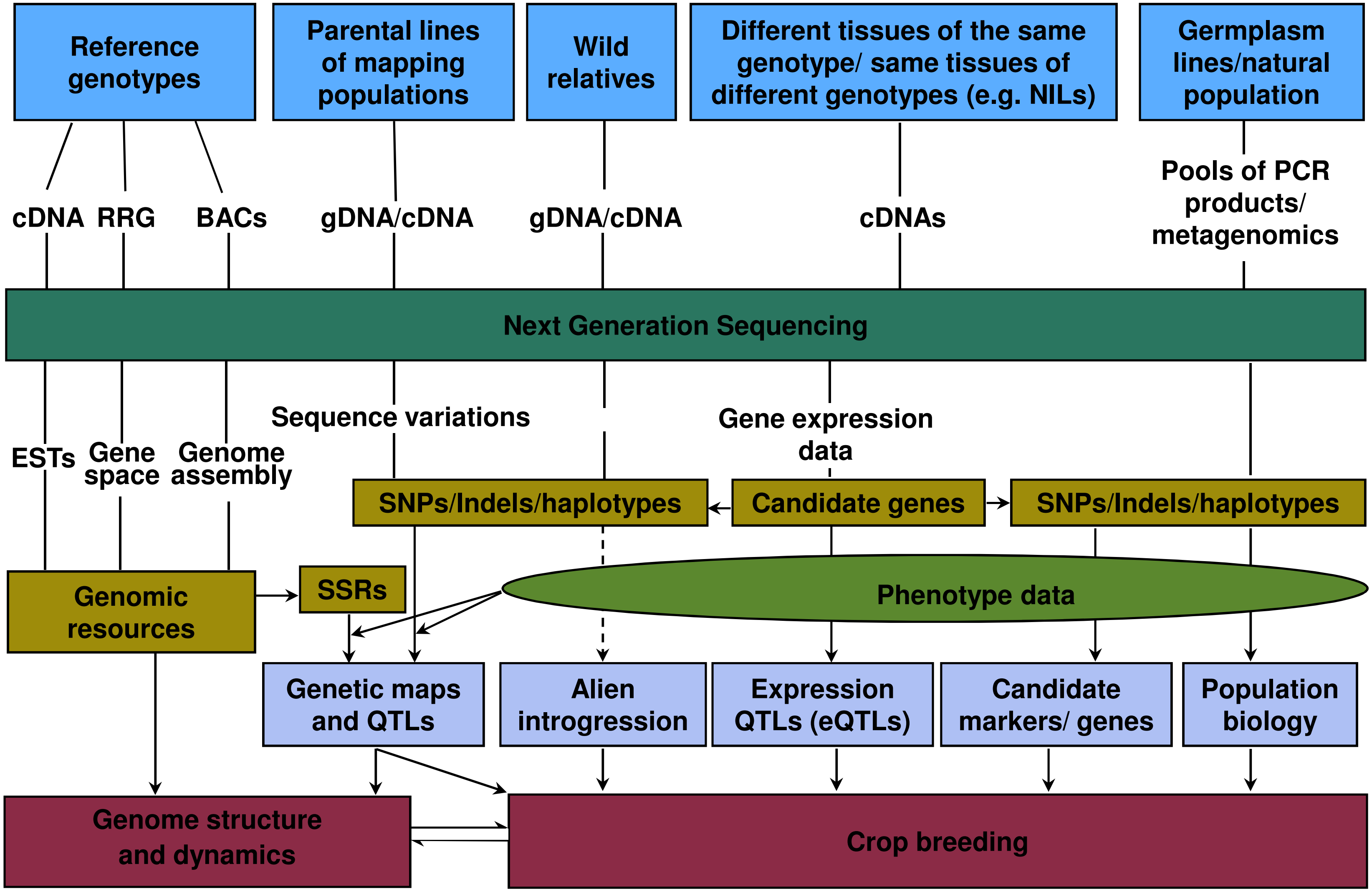
## Whole Genome Sequencing using NextGen Sequencing Technologies

| The technology | | 454 | SOLiD |
|---|---|---|---|
| The countries and their commitment | Italy | 10X | 15X |
| | Netherlands | 10X | 15X |
| | France | 10X | |
| | **India** | **05X** | |
| | Spain | | 30X |
| Resources available | Whole genome shotgun library (The Netherlands and Italy)<br>3 kb/20 kb paired-end libraries (Roche) | | |
| Data available | 15X SOLiD and 454 data produced<br>(Available at dedicated ftp site to the consortium) | | |
| The new proposed resources | New physical map using HindIII, EcoRI and MboI libraries (Keygene)<br>New randomly-sheared insert BAC library (EU-SOL and US) | | |
| Assembly | NextGen Sequences + BAC sequences (Sanger's) + Selected BAC Mixture (SBM) + BAC ends + fosmid ends | | |
| Target deadline | First draft of tomato genome by 2009 end | | |

# Sequencing Technologies: A Comparison

| | | ABI 3730xl | 454 Flx system | Solexa 1G | SOLiD |
|---|---|---|---|---|---|
| | **Technology** | Automated Sanger's di-deoxy | Emulsion PCR | Bridge PCR | Emulsion PCR |
| | | Polymerase; Fluorescent-di-deoxy terminators | Polymerase; Pyrosequencing | Polymerase; Reversible terminators | Ligase; Octamer with two-base encoding |
| 1 | **Instrument cost (approx.)** | US$ 500,000 | US$ 550,000 | US$ 500,000 | US$ 591,000 |
| 2 | **Running cost/ Mb data** | ~US$ 8,000 | ~US$ 60 | ~US$ 2 | ~US$ 1 |
| 3 | **Data generated per run** | 50 Kb | 500 Mb | 1 Gb | 6 Gb |
| 4 | **Read length** | 500– 1000 bp | 300 bp | 36 bp | 35 bp |
| 5 | **Read pair information** | Available | Available | Available | Available |

# Overview of Next Generation Sequencing applications in crop genetics and breeding



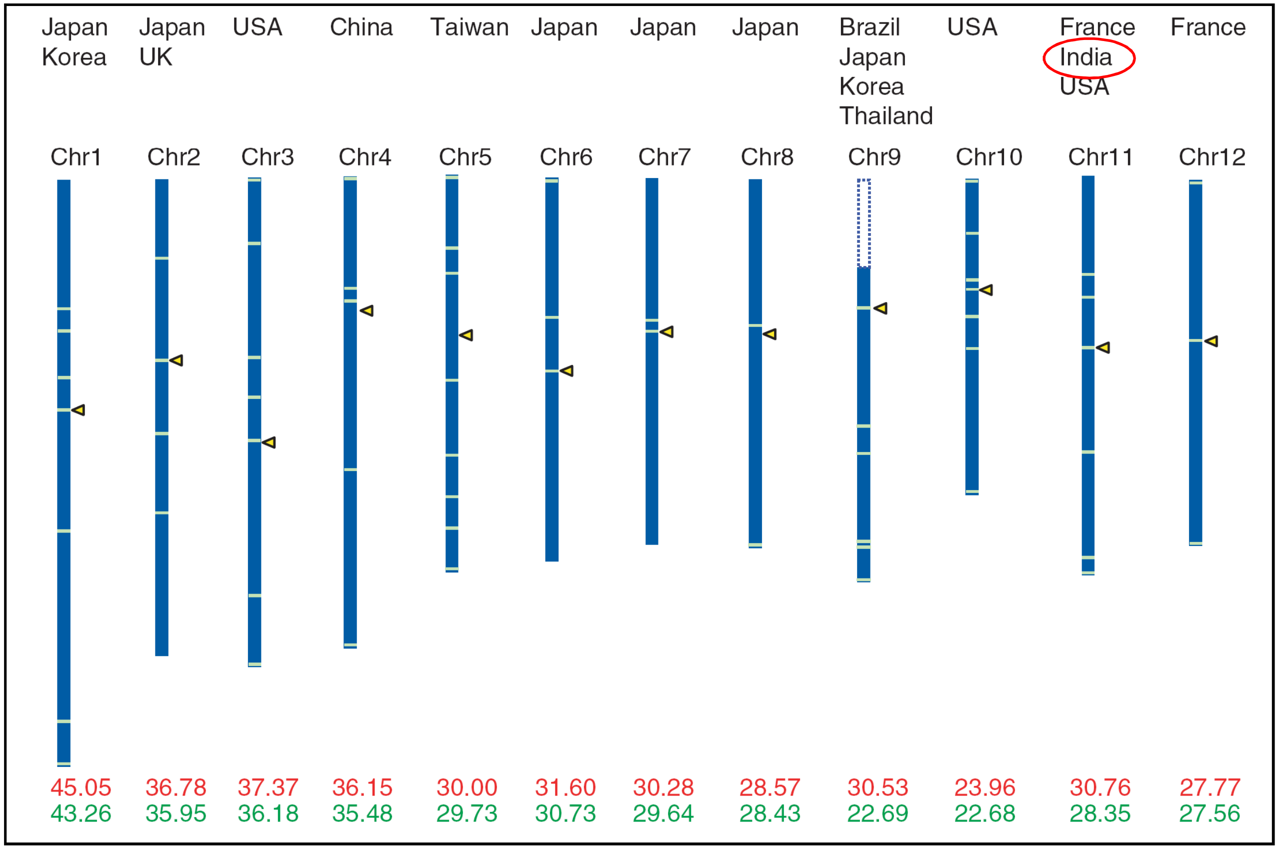
Varshney *et al.*, 2009, Trends in Biotechnology

# The map-based sequence of the rice genome

International weekly journal of science
nature

International Rice Genome Sequencing Project*

Vol 436, August 11, 2005



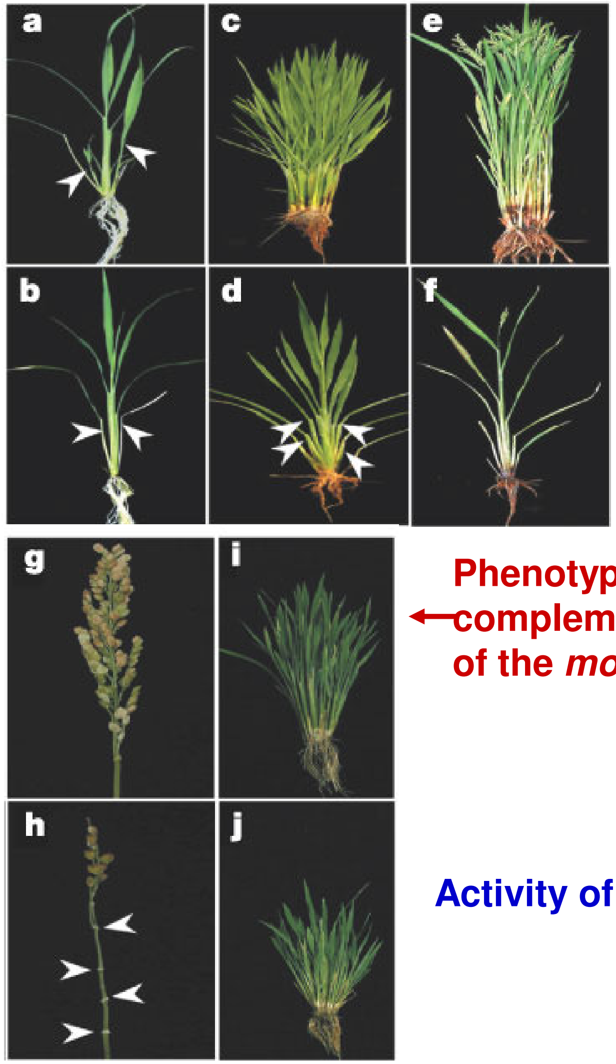
Pseudomolecules of the 12 rice chromosomes along with the participating nations responsible for sequencing each chromosome

# Examples of agriculturally important genes isolated from rice

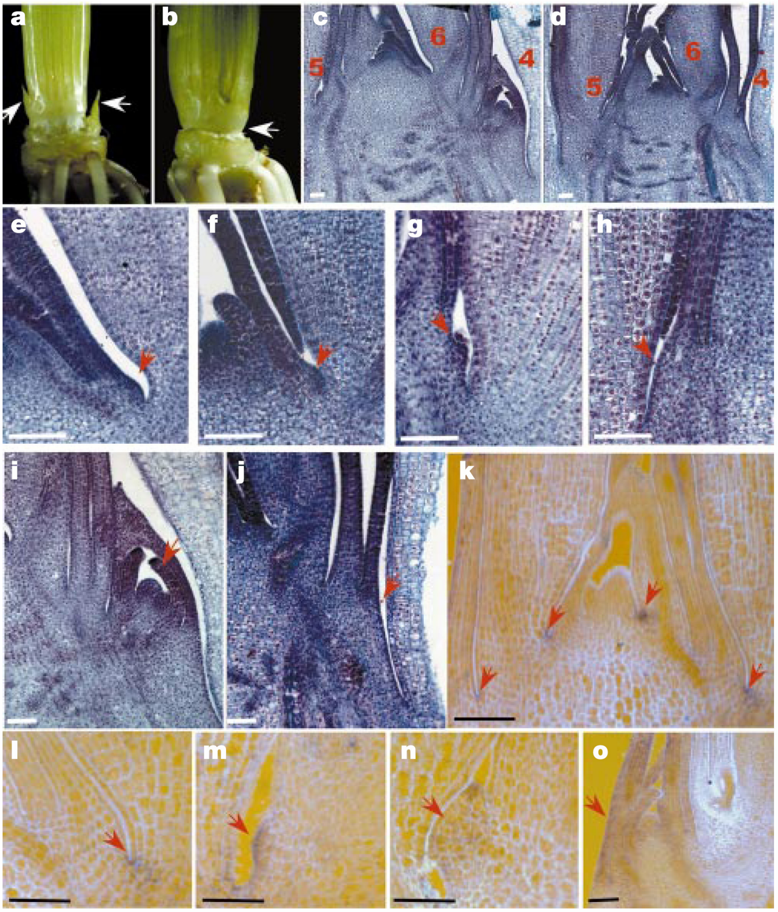
| Locus or gene | Function | Identification method |
|---|---|---|
| ***Xa21*** | Bacterial resistance | Map-based cloning |
| ***Sub1*** | Submergence tolerance | Map-based cloning |
| ***Moc1*** | Tillering number control | Map-based cloning |
| ***Pi9*** | Fungal resistance | Map-based cloning |
| ***Pi2*** | Fungal resistance | Map-based cloning |
| ***Gid1, Gid2 or Slr1*** | Gibberellin signalling pathway | Map-based cloning |
| ***Sd1*** | Gibberellin synthesis | Map-based cloning |
| ***Lsi1*** | Silicon transport | Map-based cloning |
| ***qSH1*** | Grain abscission control | Map-based cloning |
| ***Spl18*** | Fungal resistance | Activation tagging |
| ***Fon1*** | Tillering number control and the number of seeds | T-DNA |
| ***Lhs*** | Floral organ formation and seed setting | T-DNA |
| ***Udt1*** | Early anther development | T-DNA |
| ***Xb3*** | Bacterial resistance | Yeast two-hybrid |
| ***NH1*** | Bacterial resistance | Yeast two-hybrid |
| ***NRR*** | Bacterial susceptibility | Yeast two-hybrid |

**Jung *et al.*, 2008, Nature Reviews Genetics**

# Control of Tillering in Rice



← Phenotype and complementation of the *moc1* mutant



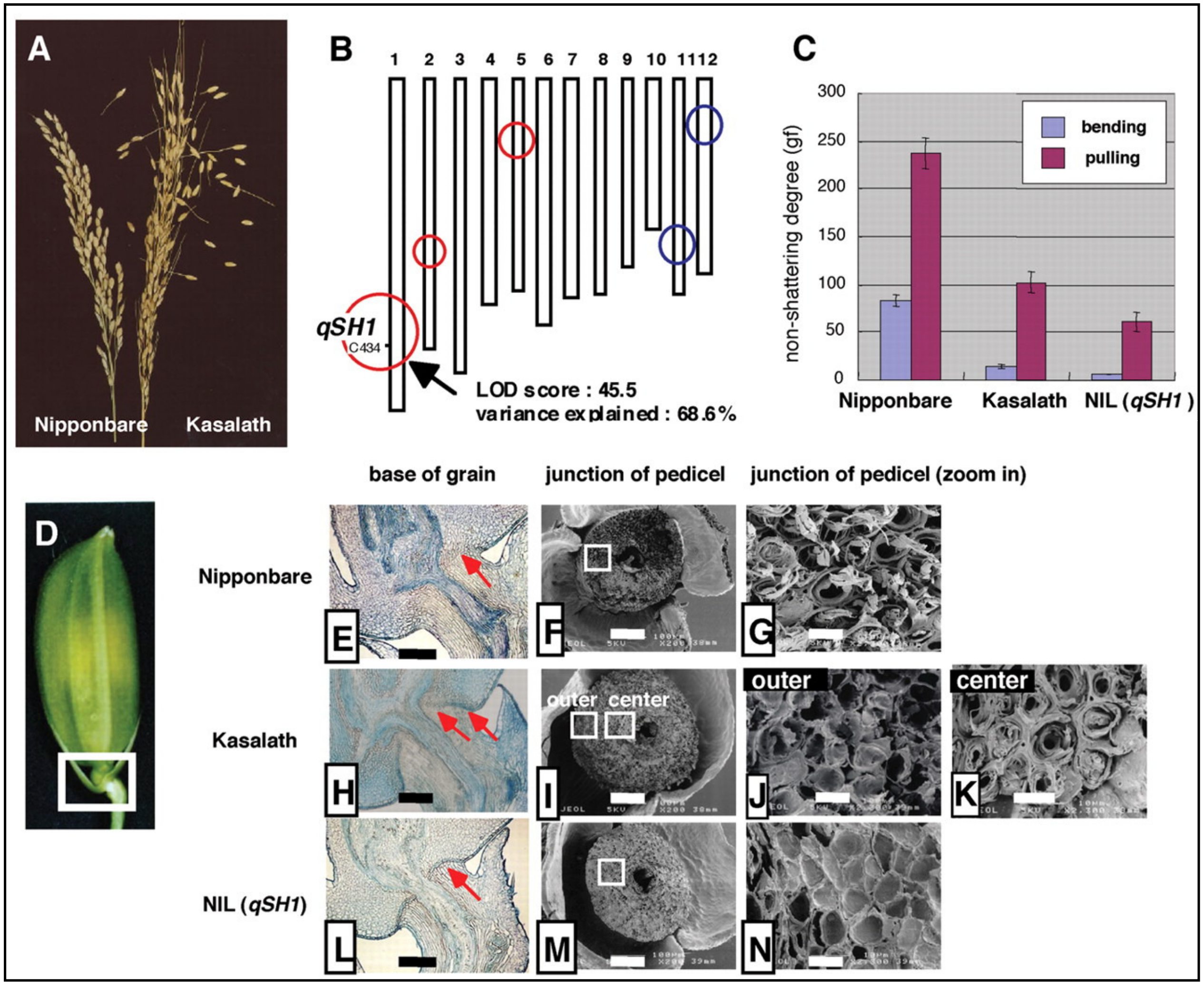
Activity of the *MOC1* →

Li et al. (2003)

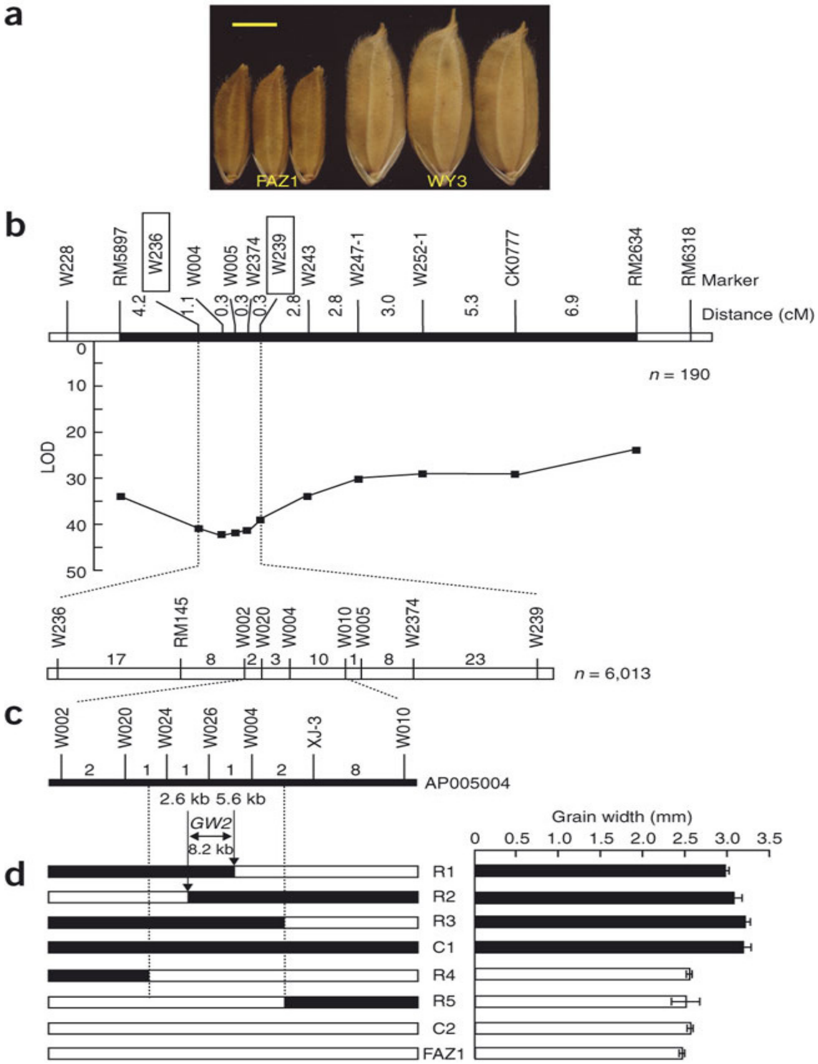
Department of Plant Molecular Biology
UNIVERSITY OF DELHI
South Campus

# qSH1 is Required for Formation of the Abscission Layer at the Base of the Rice Grain

(Konishi et al. 2006)

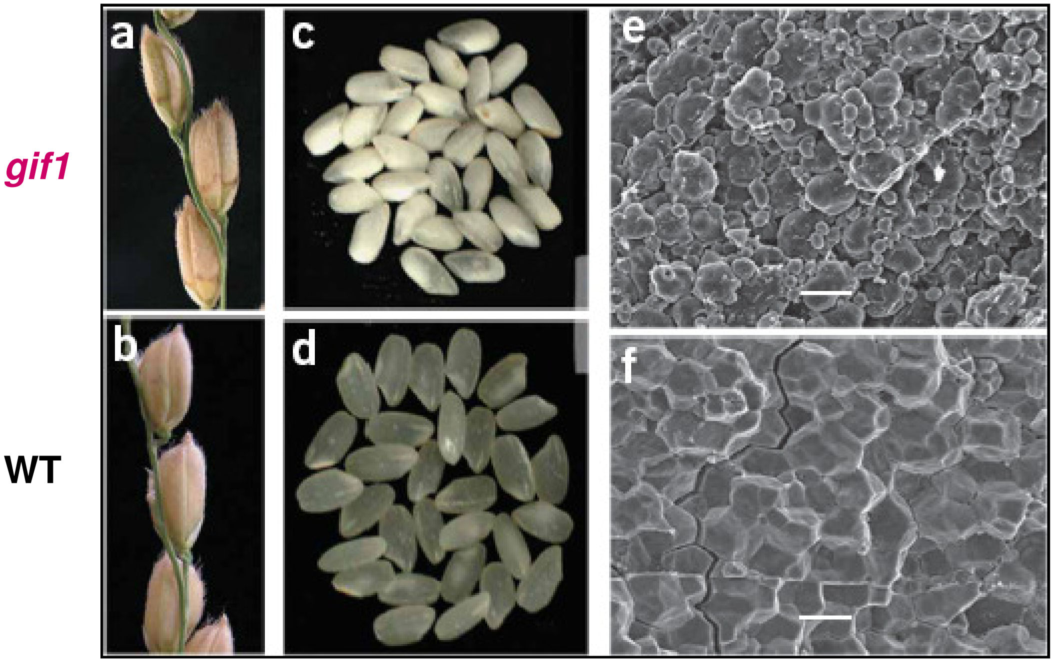


# A QTL for rice grain width and weight : RING-type E3 ubiquitin ligase

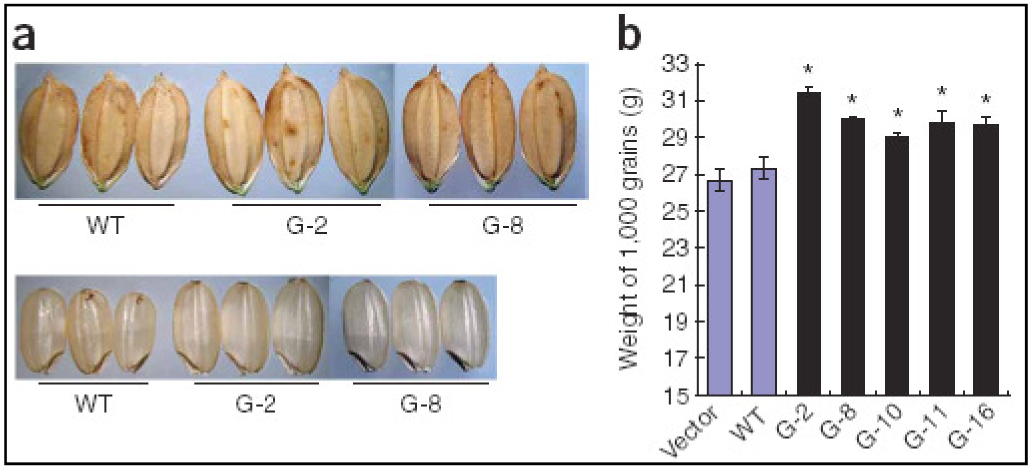


Song et al., 2007, Nat. Genet.

# Control of rice grain-filling and yield by *GRAIN INCOMPLETE FILLING 1* (*GIF1*)



Grain-filling and sugar content of *gif1* mutant and wild-type rice

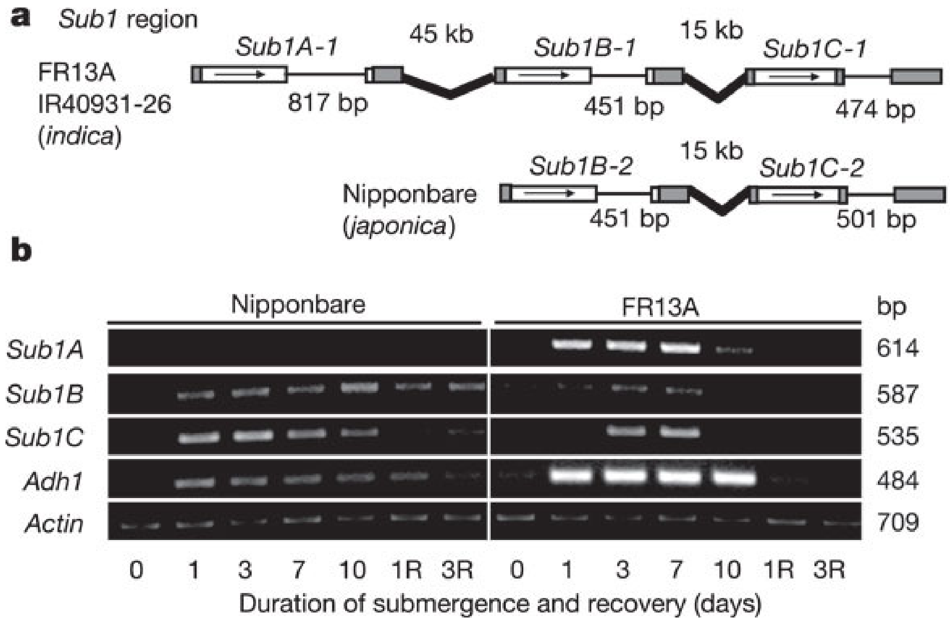


INCREASED grain size and weight in transgenic rice overexpressing *GIF1*

Wang *et al.*, 2008,
Nature Genetics

# *Sub1A,* an Ethylene-response-factor-like Gene Confers Submergence Tolerance to Rice

**Xu et al. (2006) Nature**

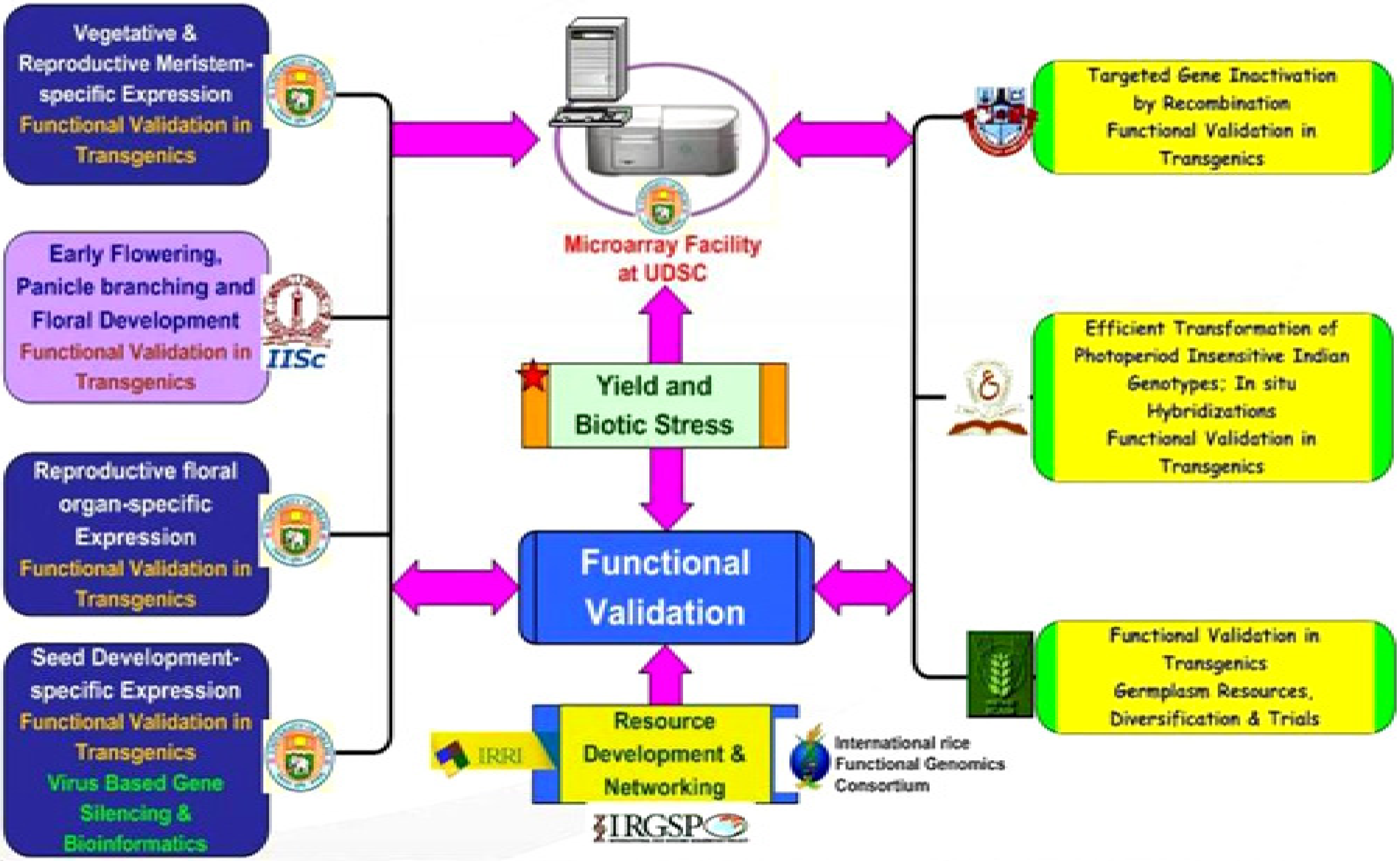


*Sub1* Region Gene Composition and Submergence induced mRNA Accumulation in Rice

Introgression of the FR13A *Sub1* Haplotype into an Intolerant Variety by MAS Confers Submergence Tolerance

# NATIONAL CONSORTIUM FOR FUNCTIONAL GENOMICS OF RICE (NCFGR)

Gene Expression Profiling During Flower and Seed Development and Functional Validation of Identified Genes



Vegetative & Reproductive Meristem-specific Expression
Functional Validation in Transgenics

Early Flowering, Panicle branching and Floral Development
Functional Validation in Transgenics

IISc

Reproductive floral organ-specific Expression
Functional Validation in Transgenics

Seed Development-specific Expression
Functional Validation in Transgenics
Virus Based Gene Silencing & Bioinformatics

Microarray Facility at UDSC

Yield and Biotic Stress

Functional Validation

Resource Development & Networking

IRRI

IRGSP

International rice Functional Genomics Consortium

Targeted Gene Inactivation by Recombination
Functional Validation in Transgenics

Efficient Transformation of Photoperiod Insensitive Indian Genotypes; In situ Hybridizations
Functional Validation in Transgenics

Functional Validation in Transgenics
Germplasm Resources, Diversification & Trials

# Developmental Time-points (19) Under Investigation

Vegetative

3

Leaf, Root, Seedlings

Floral Transition

4

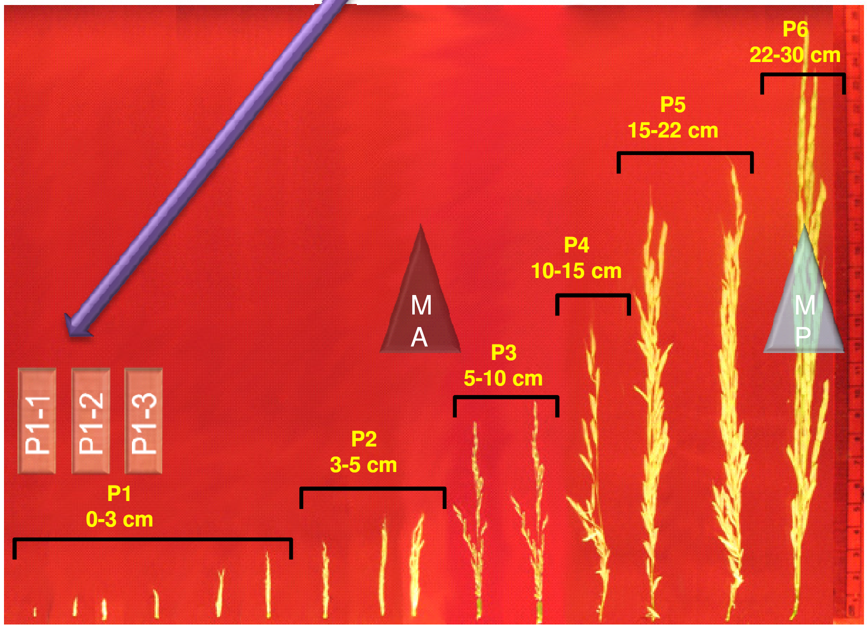
SAM, P1-1, P1-2, P1-3

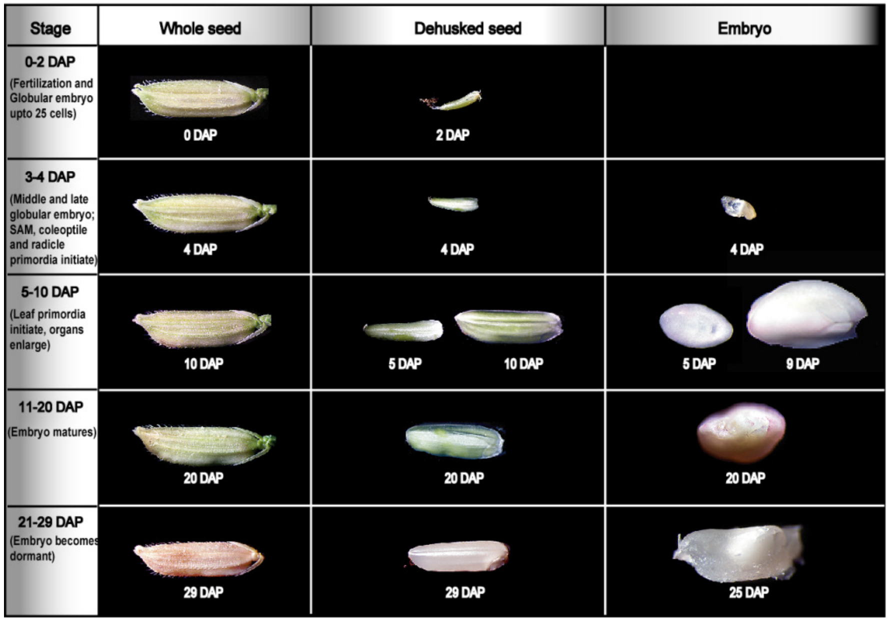
Panicle

7

Seed

5

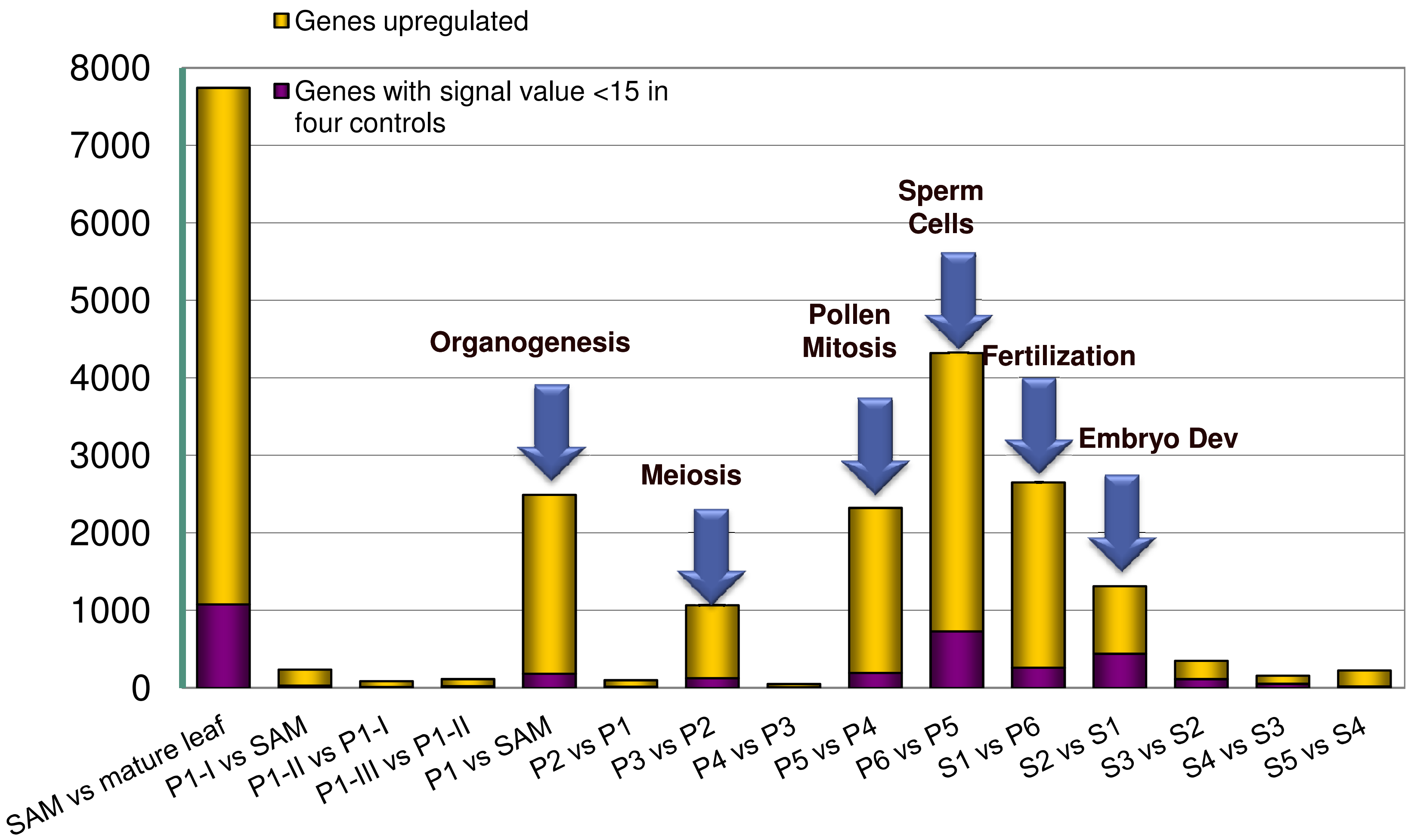




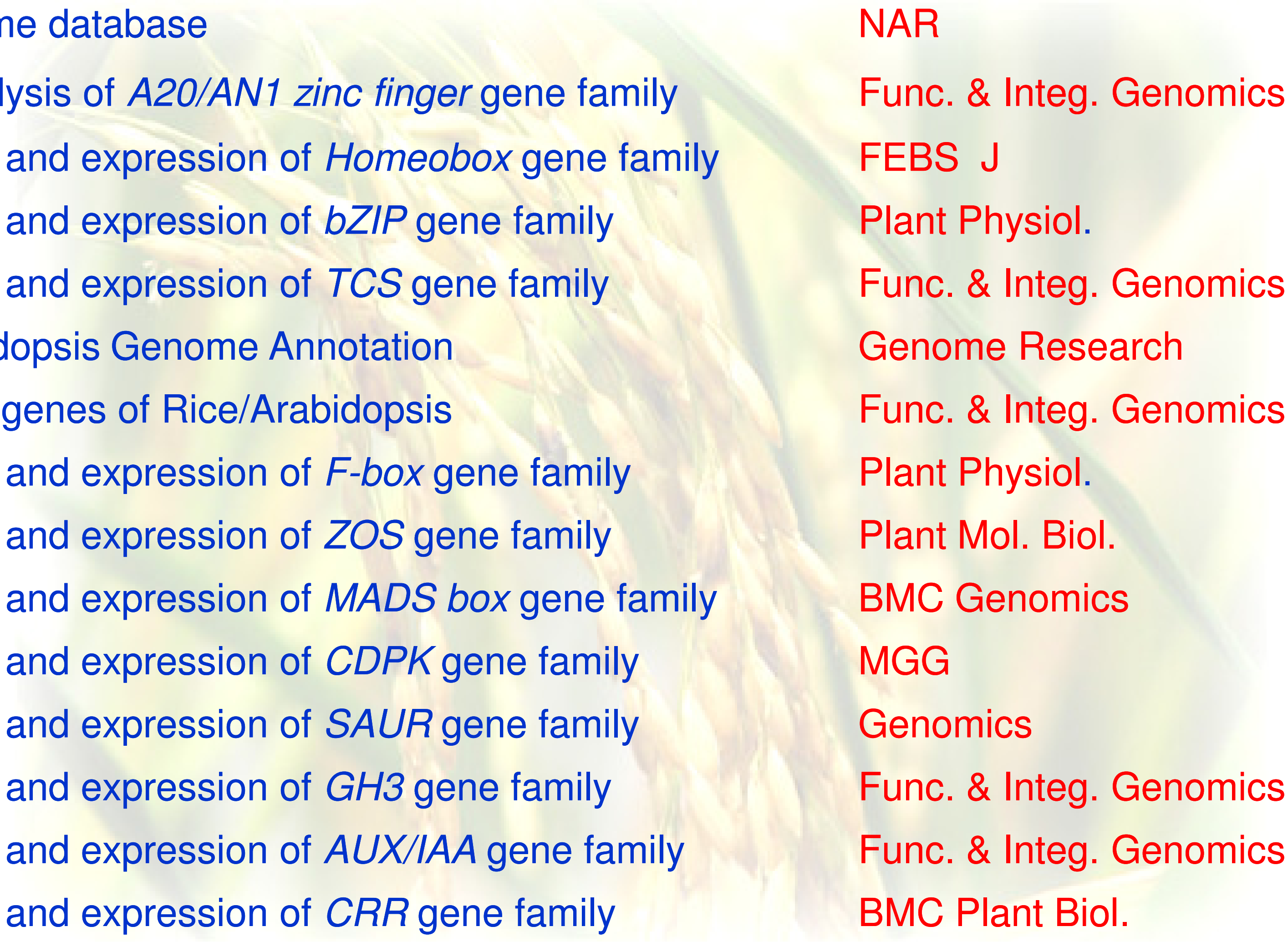
| Stage | Whole seed | Dehusked seed | Embryo |
|---|---|---|---|
| 0-2 DAP<br>(Fertilization and Globular embryo upto 25 cells) | 0 DAP | 2 DAP | |
| 3-4 DAP<br>(Middle and late globular embryo; SAM, coleoptile and radicle primordia initiate) | 4 DAP | 4 DAP | 4 DAP |
| 5-10 DAP<br>(Leaf primordia initiate, organs enlarge) | 10 DAP | 5 DAP<br>10 DAP | 5 DAP<br>9 DAP |
| 11-20 DAP<br>(Embryo matures) | 20 DAP | 20 DAP | 20 DAP |
| 21-29 DAP<br>(Embryo becomes dormant) | 29 DAP | 29 DAP | 25 DAP |

# Developmental Stages of Maximum Differential Expression Activity

(Gene Up-regulation in comparison to previous stage)



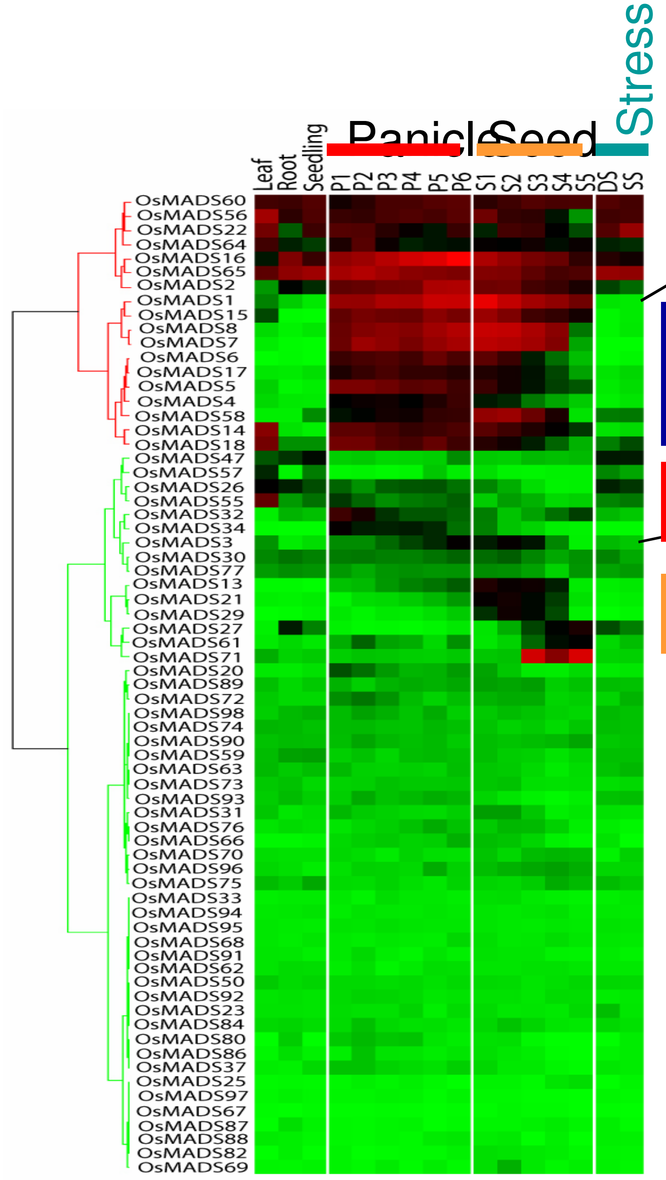
# Rice Genome Analysis



| | | |
|---|---|---|
| Annotation and expression of *RLCK* gene family | Mol Plant | 2008 |
| Rice genome database | NAR | 2008 |
| Global analysis of *A20/AN1 zinc finger* gene family | Func. & Integ. Genomics | 2008 |
| Annotation and expression of *Homeobox* gene family | FEBS J | 2008 |
| Annotation and expression of *bZIP* gene family | Plant Physiol. | 2008 |
| Annotation and expression of *TCS* gene family | Func. & Integ. Genomics | 2007 |
| Rice/Arabidopsis Genome Annotation | Genome Research | 2007 |
| Intron-less genes of Rice/Arabidopsis | Func. & Integ. Genomics | 2007 |
| Annotation and expression of *F-box* gene family | Plant Physiol. | 2007 |
| Annotation and expression of *ZOS* gene family | Plant Mol. Biol. | 2007 |
| Annotation and expression of *MADS box* gene family | BMC Genomics | 2007 |
| Annotation and expression of *CDPK* gene family | MGG | 2007 |
| Annotation and expression of *SAUR* gene family | Genomics | 2006 |
| Annotation and expression of *GH3* gene family | Func. & Integ. Genomics | 2006 |
| Annotation and expression of *AUX/IAA* gene family | Func. & Integ. Genomics | 2006 |
| Annotation and expression of *CRR* gene family | BMC Plant Biol. | 2006 |
| Genome-wide analysis of the *SAP* gene family | MGG | 2006 |

# Expression Profiles of 69 Rice MADS Box Genes

## (BMC Genomics, 2007)



**•OsMADS1**

Panicle & Seed

Panicle

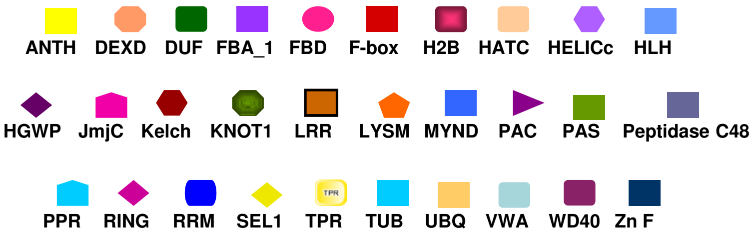
**•OsMADS3**

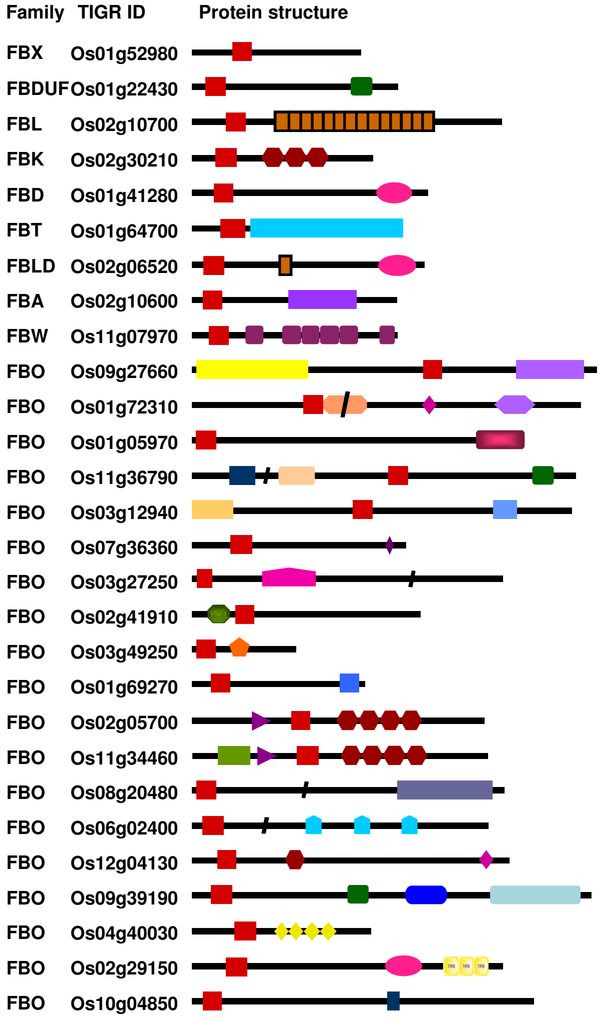
Seed

# Classification and domain organization of F-box proteins in rice

## (Plant Physiology, 2007)

| Subfamily | Domain content | No. in category |
|---|---|---|
| FBX | Unknown/none | 465 |
| FBDUF | DUF | 66 |
| FBL | LRR | 61 |
| FBK | Kelch | 25 |
| FBD | FBD | 17 |
| FBT | TUB | 14 |
| FBLD | LRR and FBD | 9 |
| FBA | FBA | 4 |
| FBW | WD40 | 2 |
| FBO | Others | 24 |

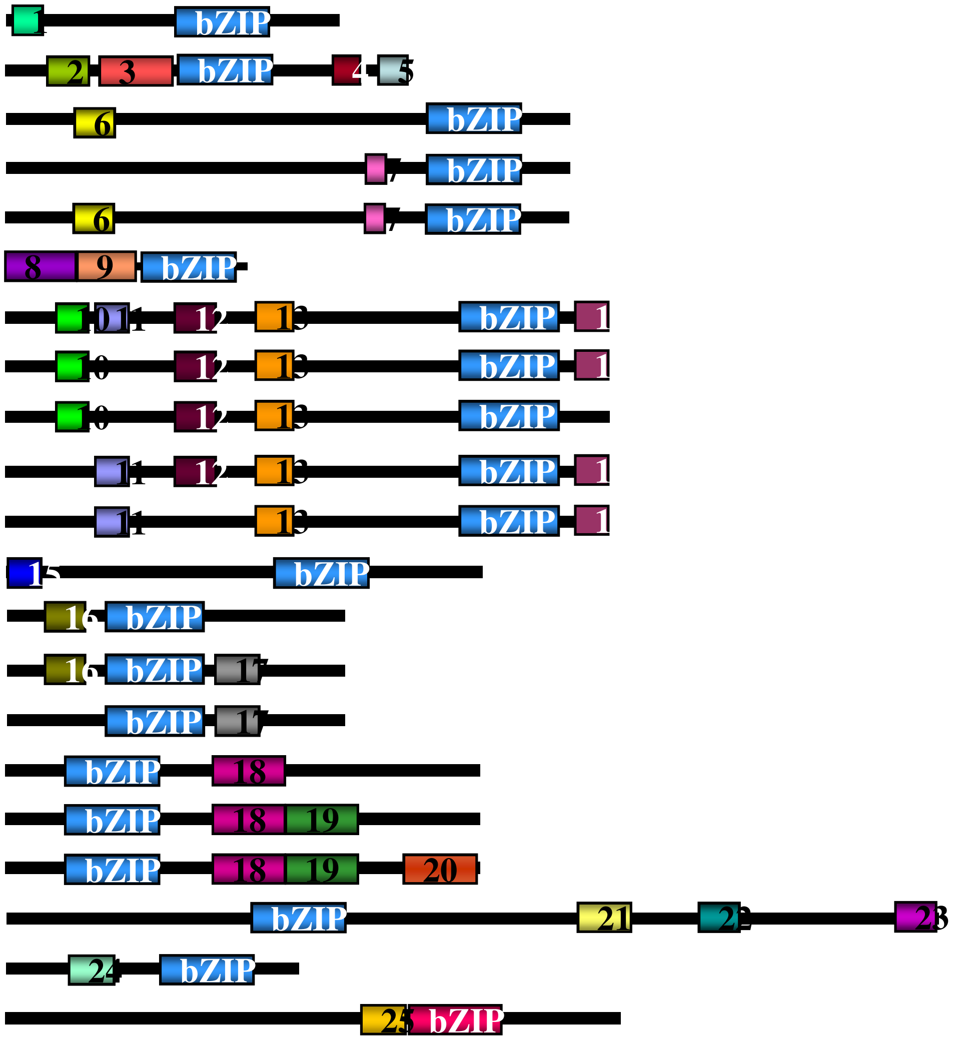




| Family | TIGR ID |
|---|---|
| FBX | Os01g52980 |
| FBDUF | Os01g22430 |
| FBL | Os02g10700 |
| FBK | Os02g30210 |
| FBD | Os01g41280 |
| FBT | Os01g64700 |
| FBLD | Os02g06520 |
| FBA | Os02g10600 |
| FBW | Os11g07970 |
| FBO | Os09g27660 |
| FBO | Os01g72310 |
| FBO | Os01g05970 |
| FBO | Os11g36790 |
| FBO | Os03g12940 |
| FBO | Os07g36360 |
| FBO | Os03g27250 |
| FBO | Os02g41910 |
| FBO | Os03g49250 |
| FBO | Os01g69270 |
| FBO | Os02g05700 |
| FBO | Os11g34460 |
| FBO | Os08g20480 |
| FBO | Os06g02400 |
| FBO | Os12g04130 |
| FBO | Os09g39190 |
| FBO | Os04g40030 |
| FBO | Os02g29150 |
| FBO | Os10g04850 |

# Structural features in OsbZIP transcription factors

## Protein Structure



## OsbZIP proteins having similar protein structure

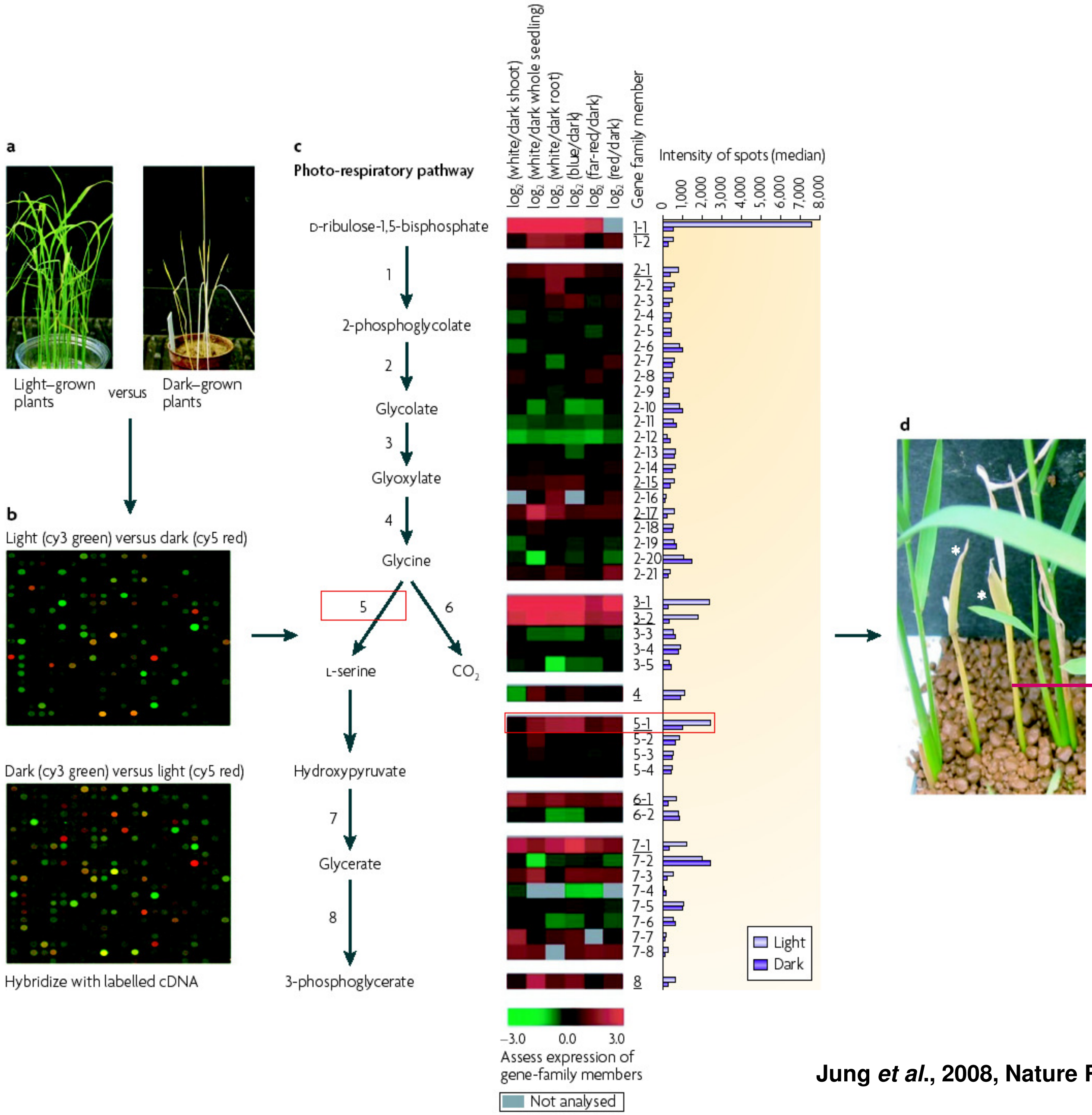
OsbZIP67, 73
OsbZIP15, 20
OsbZIP05, 45
OsbZIP34, 59
OsbZIP13, 26, 86
OsbZIP54, 55, 56
OsbZIP09, 10, 23, 42, 46, 72
OsbZIP62
OsbZIP29
OsbZIP66
OsbZIP12, 40
OsbZIP02, 06, 19
OsbZIP53
OsbZIP44
OsbZIP07
OsbZIP17
OsbZIP64
OsbZIP03, 08, 11, 28, 37, 41, 43, 47, 49,
OsbZIP39, 60
OsbZIP01, 18, 48
OsbZIP25, 30, 35, 36, 61, 68, 75, 76, 78,

**Motifs**

| | | |
|---|---|---|
| 6 Part of Pro-rich domain | 10 12 Potential | 24 COP1 interaction motif |
| 7 GCB | 13 14 phosphorylation sites | |

# Determination of gene function by a COMBINATION of whole-genome transcriptome and predicted-pathway analysis

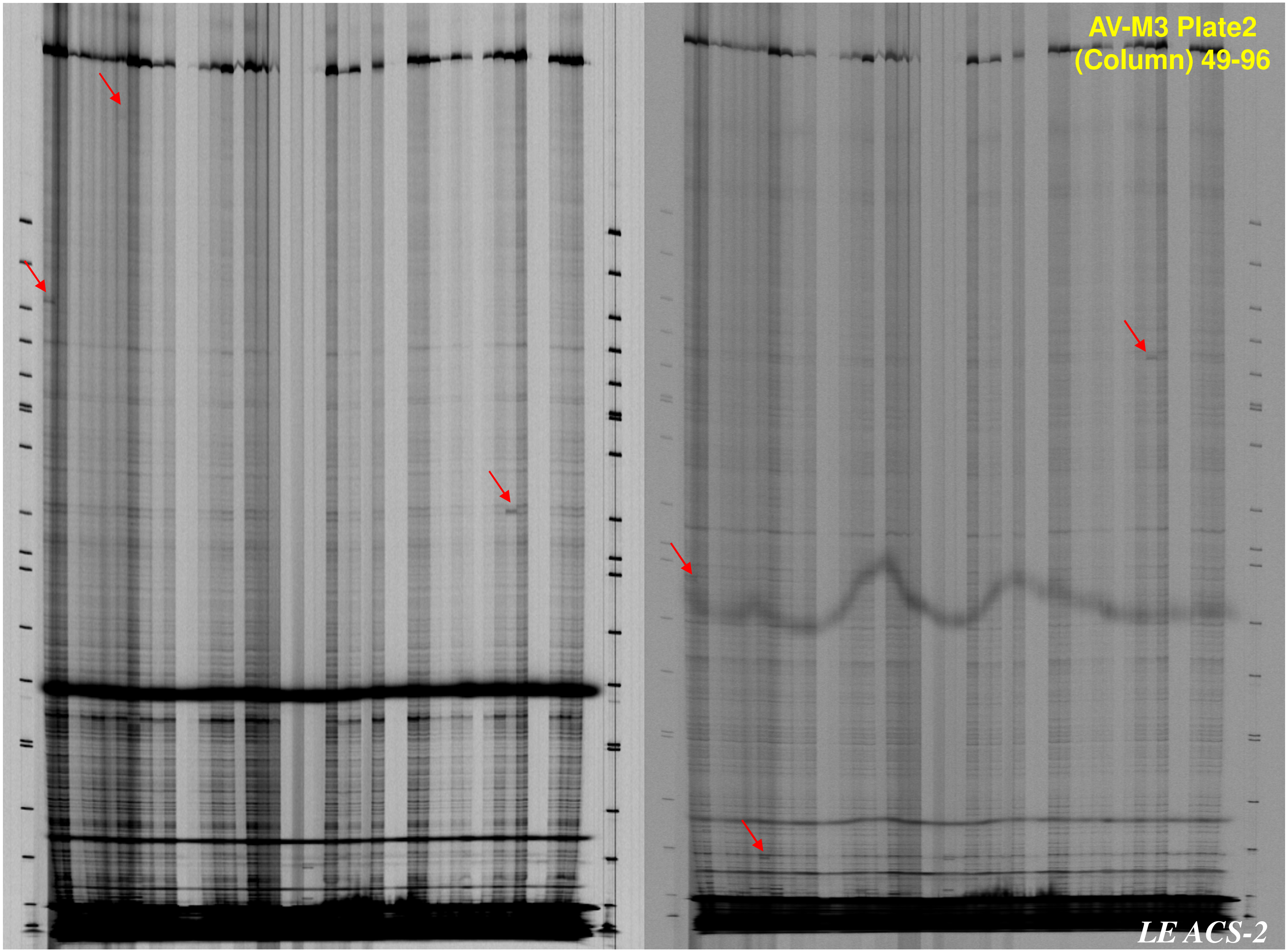
**Whole-genome expression profiling to identify light responsive genes**



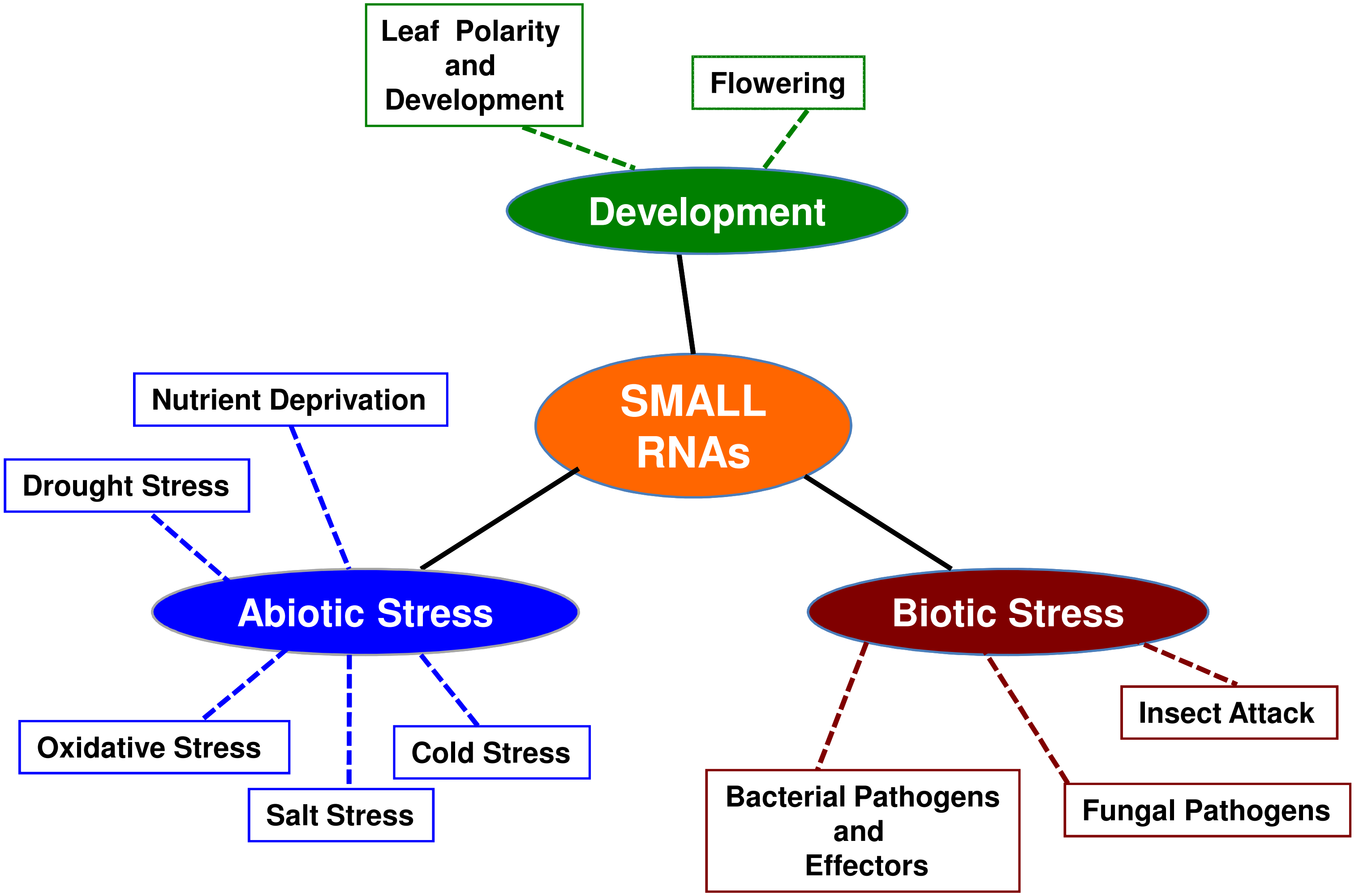
**Tos17 insertion in gene 5-1**

Jung *et al.*, 2008, Nature Reviews Genetics

# TILLING: Traditional mutagenesis meets functional genomics



# SMALL NON-CODING RNAs :
## KEY REGULATORS WITH DIVERSE ROLES



SMALL RNAs

Development

Leaf Polarity and Development

Flowering

Abiotic Stress

Nutrient Deprivation

Drought Stress

Oxidative Stress

Salt Stress

Cold Stress

Biotic Stress

Bacterial Pathogens and Effectors

Fungal Pathogens

Insect Attack

# Outlook for Future

- Sequencing/Re-sequencing of Indian crops
- Discovery of SNPs in Indian core germplasm
- Functional analysis of genes/alleles for trait-specific genetic enhancement
- Uncover relationship of genotype to phenotype in systems mode
- Genomics-assisted breeding
- Academia/Industry partnership

# THANK YOU !

*Our work is financially supported by the Department of Biotechnology*
*and*
*Department of Science and Technology, Government of India*

Department of Plant Molecular Biology
UNIVERSITY OF DELHI
# South Campus #