

# **Evening Workshop on Comparative Expression Analysis and Orthologs Identification**

**7:30-9:30 pm, September 21, 2004**

- Compare notes from different expression analyses and results**
- Find ways to connect data from different assays**
- Discuss current approaches of finding orthologs**
- How to bring together gene expression patterns and sequence inference to help ortholog identification**

			<b>Lead</b>
<b>7:30</b>		<b>Objectives of brainstorming</b>	<b>Hei Leung</b>
	<b>Comparative expression analysis</b>	<b>Rice</b>	<b>Shoshi Kikuchi</b>
		<b>Maize</b>	<b>Mark Sawkins</b>
		<b>Arabidopsis</b>	<b>Andy Pereira</b>
		<b>Potato</b>	<b>Meredith Bonierable</b>
		<b>Making use of EST libraries Cassava, legumes, <i>Musa</i></b>	<b>Manabu Ishitani et al.</b>
		<b>Comparative approaches</b>	<b>Open discussion</b>
<b>8:30</b>	<b>Orthologs</b>	<b>COS markers</b>	<b>Theresa Fulton</b>
		<b>Gene tree construction</b>	<b>Ken McNally</b>
		<b>Phylogenomics and orthologs identification</b>	<b>Richard Bruskiwich</b>
		<b>Convergent approaches using expression data and sequence comparison</b>	<b>Open discussion</b>
		<b>Ways to move forward</b>	

# Identifying functionally conserved genes

Stuart et al. 2003 “Gene-coexpression Network....” *Science* 302:249-255

---

**Define metagenes (groups of orthologs with reciprocal best hits): 6307 metagenes from human-yeast-fly-worm**

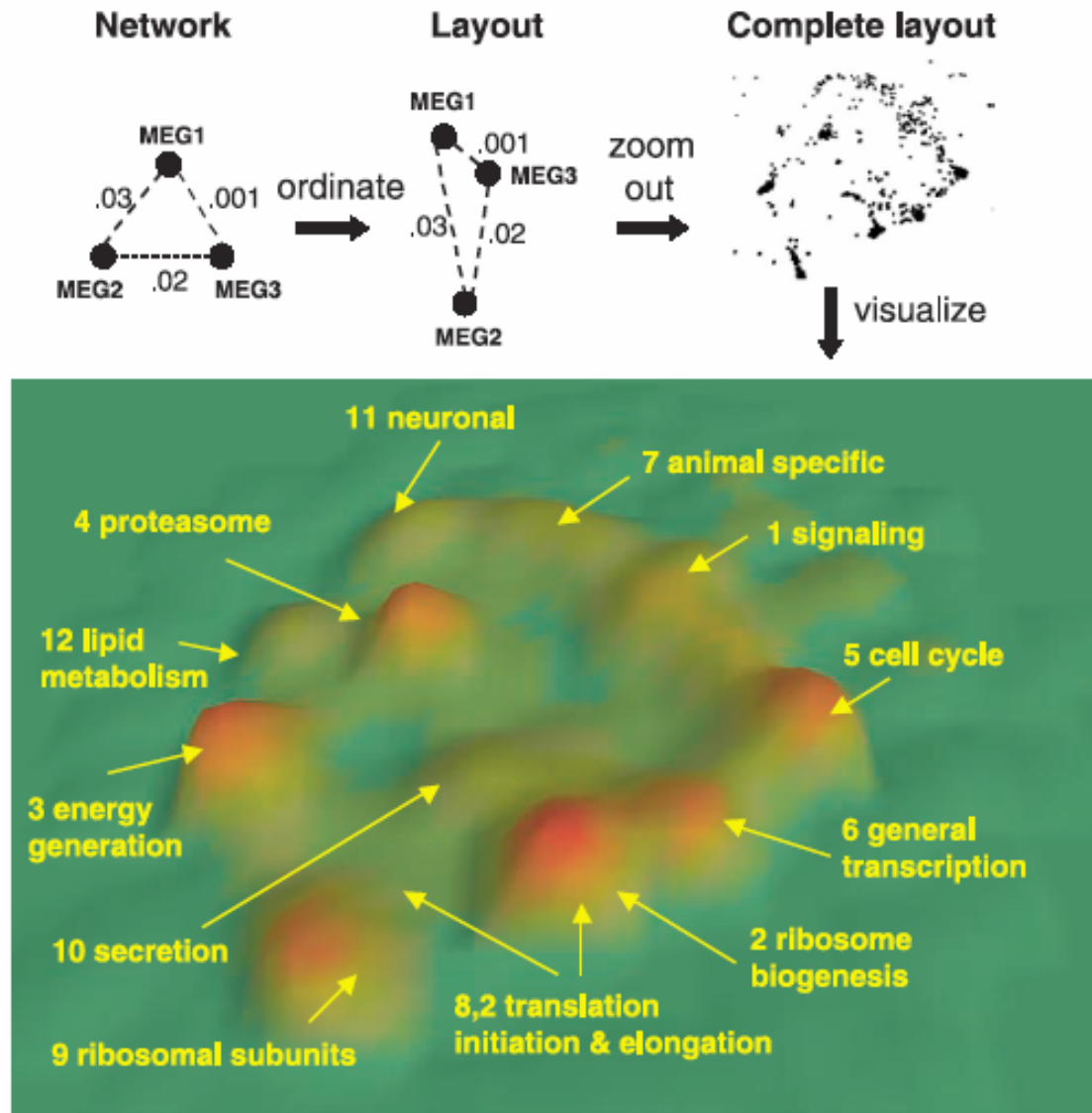
**Pairs of metagenes: co-expressed in one organism and also in different organisms**

**Identify pairs of co-expressed metagenes in multiple organisms: 3416 metagenes connected by 27,163 expression interactions ( $P < 0.05$ )**

**Terrain map: discover new genes falling in same components**

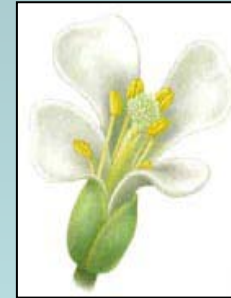
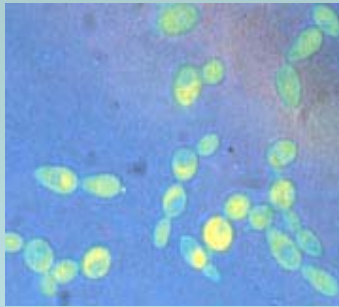
**Validate function by gene-knockouts and silencing (RNAi)**

**Fig. 3.** The negative logarithm of the  $P$  values computed for conserved coexpression links were used to position the metagenes on a 2D grid using Vx-Insight's ordination tool (20). Metagenes with smaller  $P$  values (indicating a higher significance of conserved coexpression) were placed close to each other, whereas metagenes with larger  $P$  values were placed farther apart. The altitude in the final visualization indicates the local density of genes. The bottom panel shows the 3D representation for 3416 metagenes. Twelve components of highly interconnected metagenes are shown along with the main biological functions for which they were enriched. The entire data set can be queried for individual genes using VxInsight, which can be downloaded from <http://cmgm.stanford.edu/~kimlab/multiplespecies>.





# *Decade of genomes*



<b>Organism</b>	<b>Year</b>	<b>Genome size, Mb</b>	<b>Number of predicted genes</b>
<b>Yeast</b>	<b>1996</b>	<b>13</b>	<b>6,000</b>
<b>Nematode</b>	<b>1998</b>	<b>97</b>	<b>19,000</b>
<b><i>Drosophila</i></b>	<b>2000</b>	<b>180</b>	<b>14,000</b>
<b><i>Arabidopsis</i></b>	<b>2000</b>	<b>150</b>	<b>26,000</b>
<b>Human</b>	<b>2001</b>	<b>3,000</b>	<b>35,000</b>
<b>Rice</b>	<b>2002</b>	<b>430</b>	<b>35,000-50,000</b>

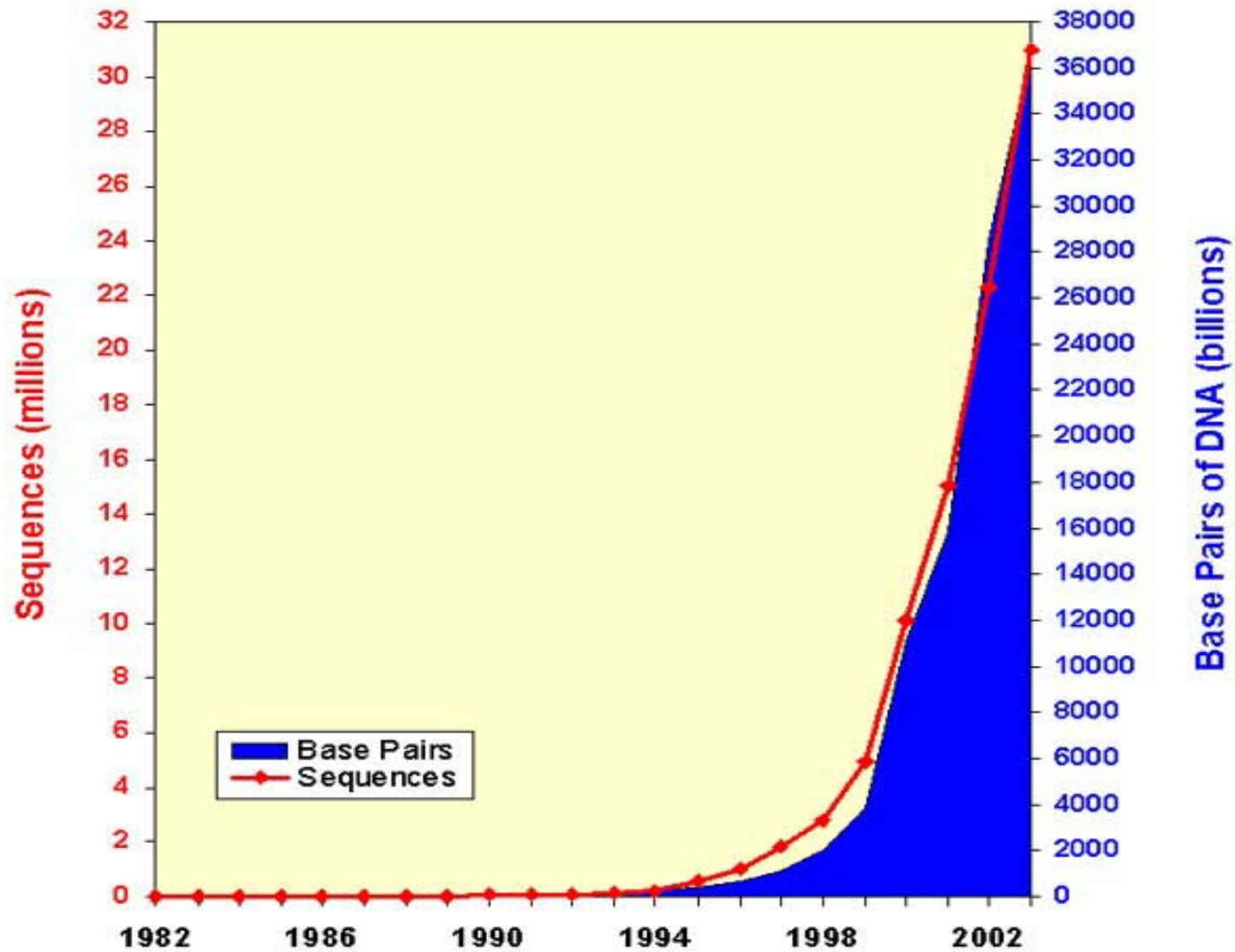


# Can we address intractable problems?

## Some encouraging signs

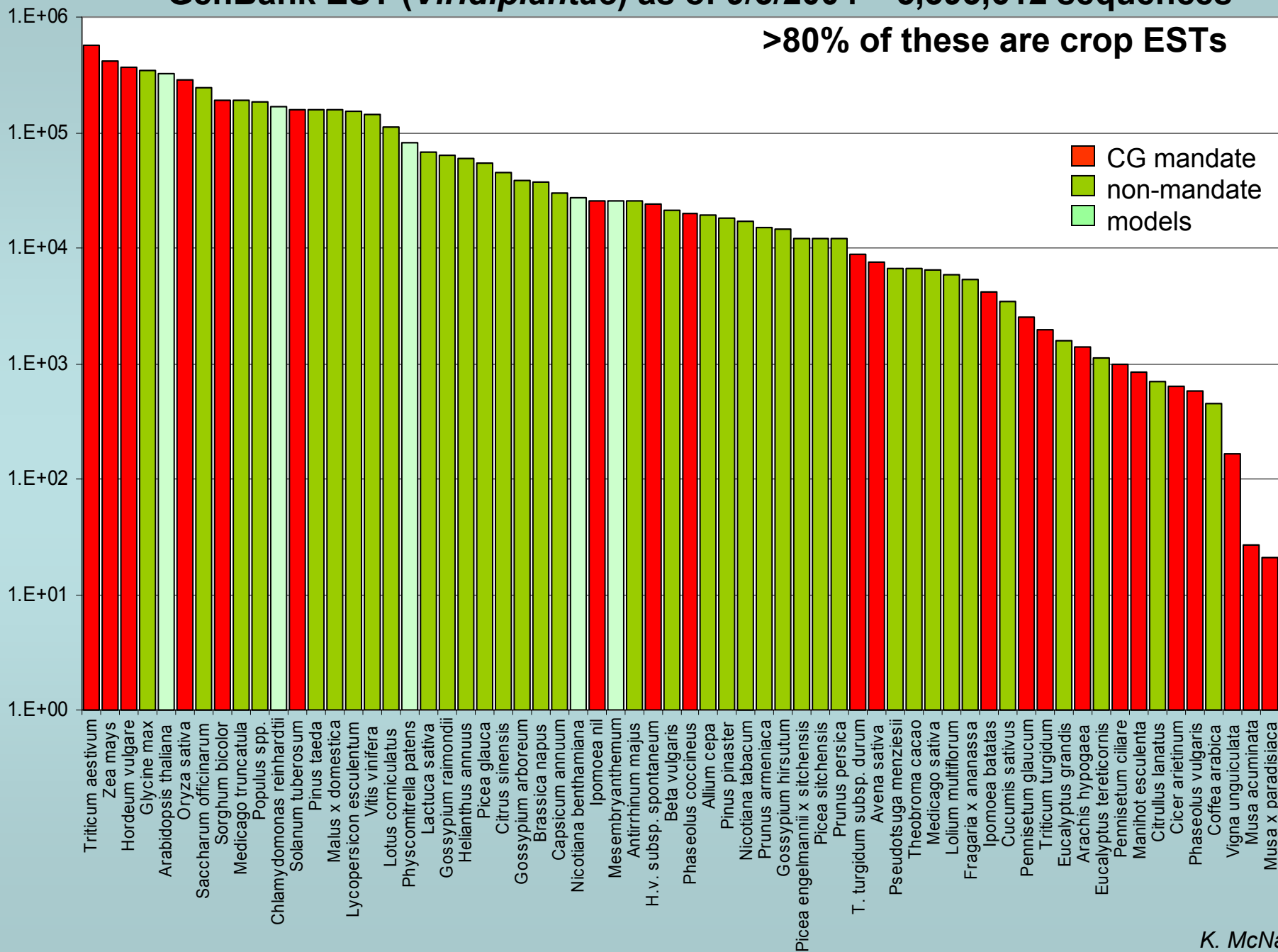
- Regulatory genes with large effects (e.g., the DREB gene for drought/cold tolerance)
  - First identified in *Arabidopsis*, now being tested in cereals and other crops
- SNP applications in medical and human genetics
  - Asthma-causing genes (human)
  - Bacterial disease immunity (*Drosophila*)

# Growth of GenBank

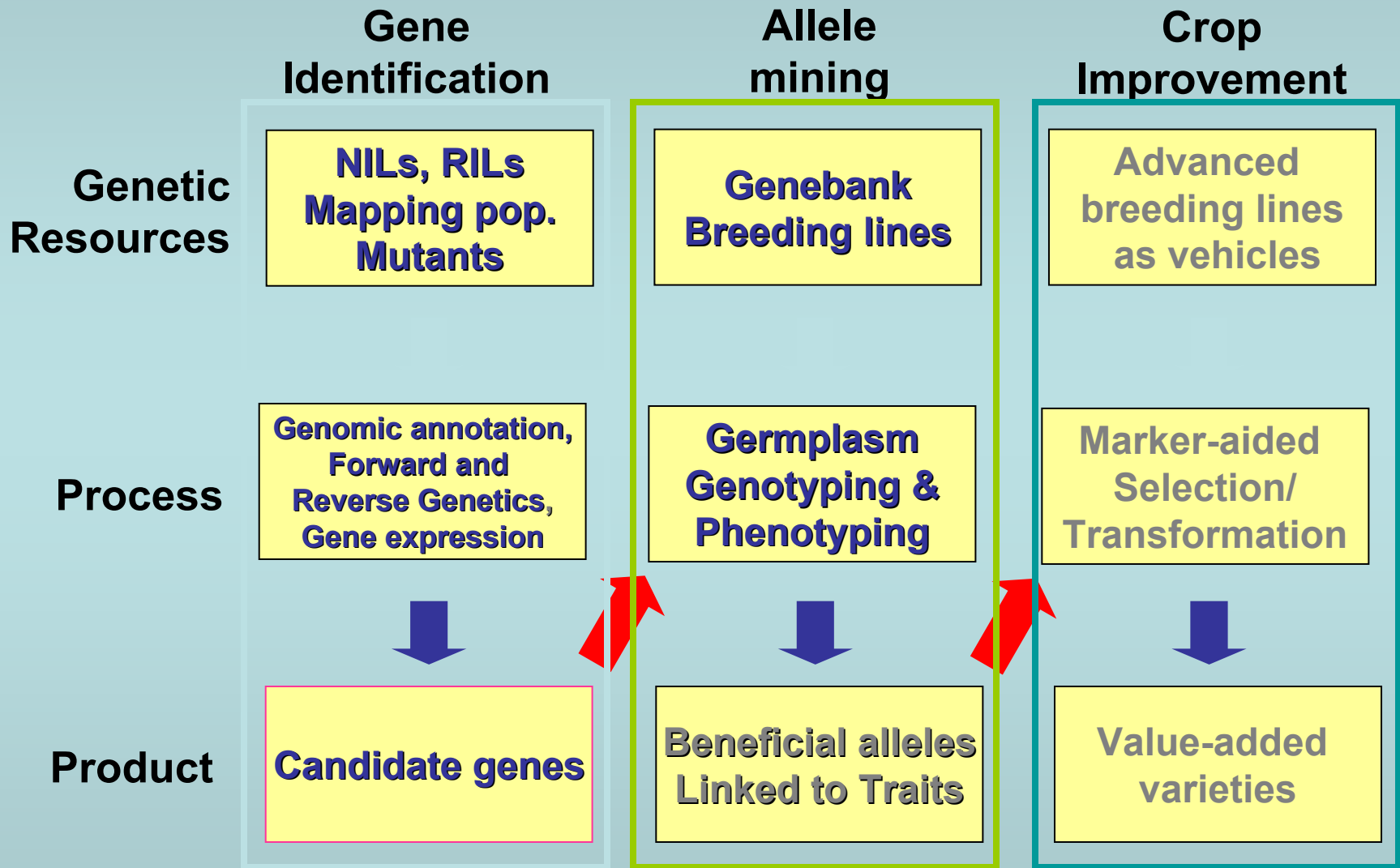


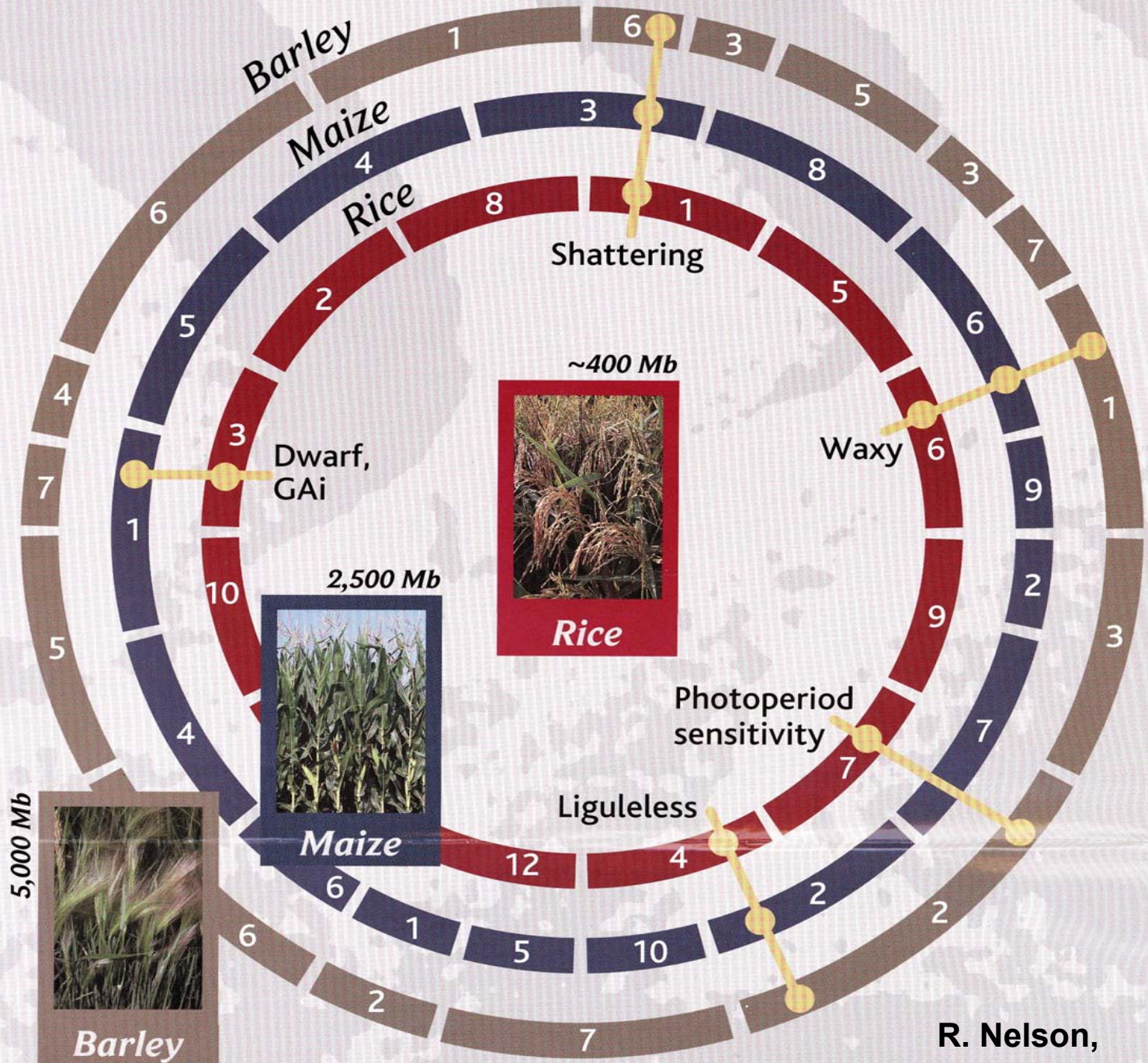
# GenBank EST (*Viridiplantae*) as of 9/3/2004 – 5,393,612 sequences

>80% of these are crop ESTs



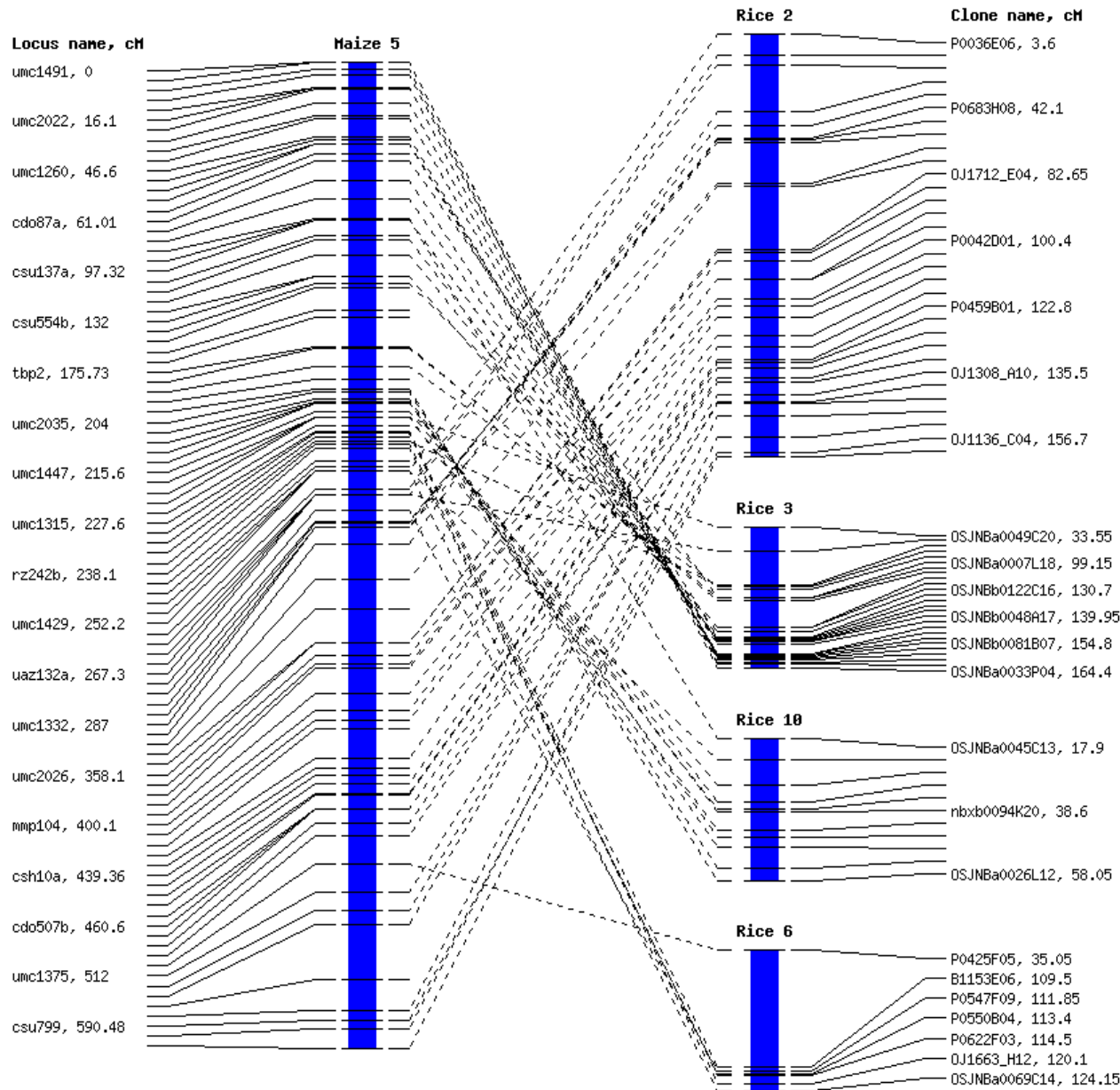
# Gene Discovery to Trait Synthesis





R. Nelson,  
McKnight Foundation

# Syntenic Mapping between Rice and Maize



•Maize 5=> Rice 2,3,10

•Maize 5: 387 unique markers

•Found 151 maize markers matched rice genome

•108/151 markers were on syntenic chromosomes

# External Environment

**Complete rice genome and growing sequence resources**

**Comparative genomics  
Generation  
Challenge Program**

**Human genetics: Power of genetic association analysis (SNP) to resolve complex traits**



**Application in crops**

# Beyond gene order

- Gene contents—actual sequence
  - orthologs and paralogs
- Evolutionary conserved functions
- Narrowing the monocot-dicot divide

# Finding functionally related genes

---

**Homologous genes:** derived from a common ancestral gene, level of sequence similarity often reflects time of divergence



**Orthologous genes** in different species derived from the same gene in the last common ancestral species, usually have similar function



**Paralogous genes**— duplicated gene pairs that have diverged and usually have different functions.

# Identify candidate genes with conserved function

---

**Multiple steps in comparative sequence analysis (Frazer et al. 2003)**

**Identify evolutionarily related genomic sequences: homologs, orthologs, paralogs**

**Annotate and align sequences**

**Identify and visualize conserved regions**

**Supporting experimental evidence**

**Syntenic relationship**

**Gene expression**

**Cross-species mapping**

**Co-expression**

**Macro and Micro-synteny**

**Functional prediction**

**Functional validation of candidate genes**

## Subprogram 2: Comparative genomics for gene discovery

- Acquisition of genetic resources and capacities
- Development of comparative maps *within and between* crop species (**increasingly based on gene contents**)
- Assignment of genes and pathways to putative phenotypes
  - Gathering functional evidence
  - Reverse genetics
  - Genetic engineering
  - Drought case study
- Validation of genes and pathways

# Tentative 3-year Medium Term Plan (MTP)

- Assemble genetic knowledge on stress tolerance (drought and others) through informatics analysis and empirical studies using **advanced genetic stocks**
- High-quality phenotyping for functional validation.
  - mutants
  - NIL and Advanced Backcross lines
- **Parallel studies** on stress tolerance as a basis to investigate drought tolerance.
- Interactive effects (synergistic or antagonistic effects) of stress responses--packaging multiple stress tolerance
- Develop genomic tools and “designer” genetic stocks in selected crops to enable application of successful approaches.
- Initiate SNP analysis in target genes (orthologs and paralogs) in multiple species.

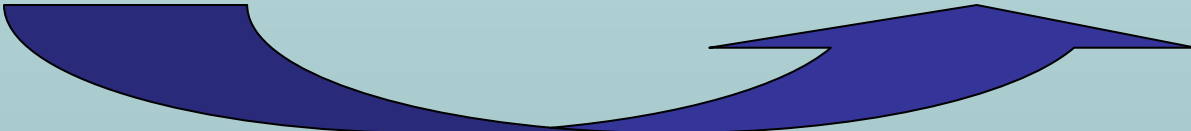
# Connecting genomic variation with phenotype and performance

---

**Genetic variation**  
**SNP and indels**

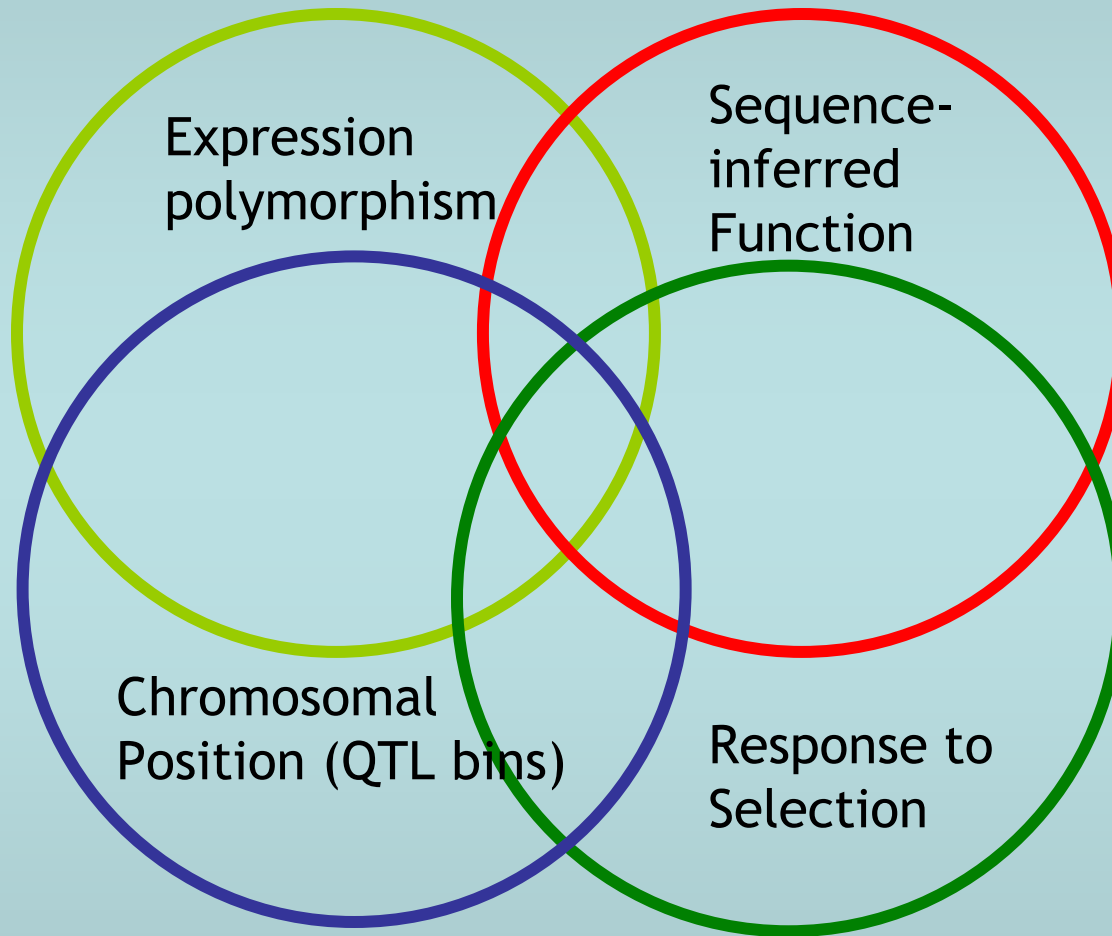
**Profiles of**  
• **Transcripts**  
• **Proteins**  
• **Metabolites**  
**Interactions**

**Phenotypes**  
**Performance**



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

---



Converging evidence for candidate gene validation

# Connecting expression and DNA polymorphism

---

**Expression  
polymorphism**

**Correlation and co-  
segregation**

**Candidate genes**

**DNA polymorphism**

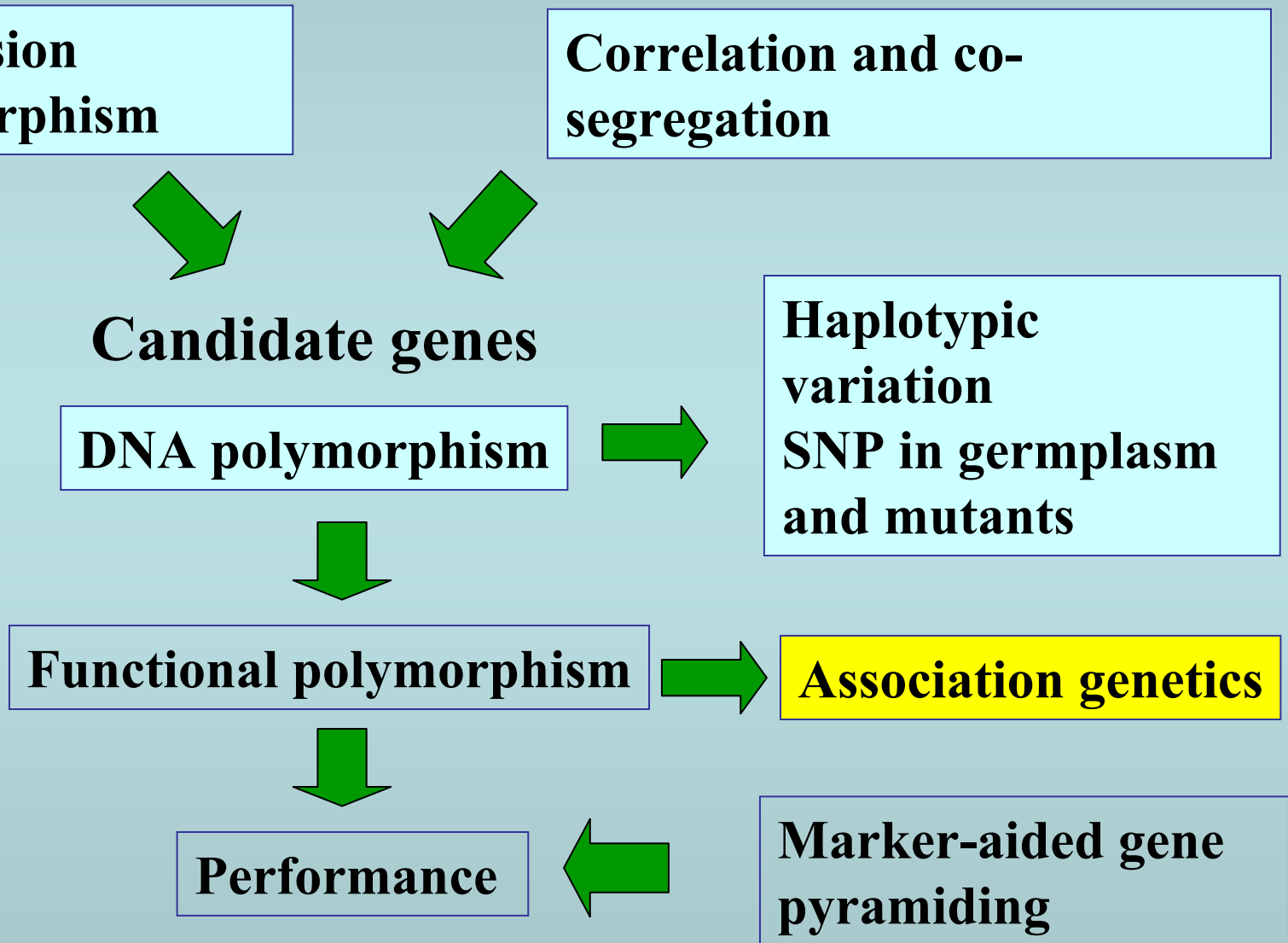
**Haplotypic  
variation  
SNP in germplasm  
and mutants**

**Functional polymorphism**

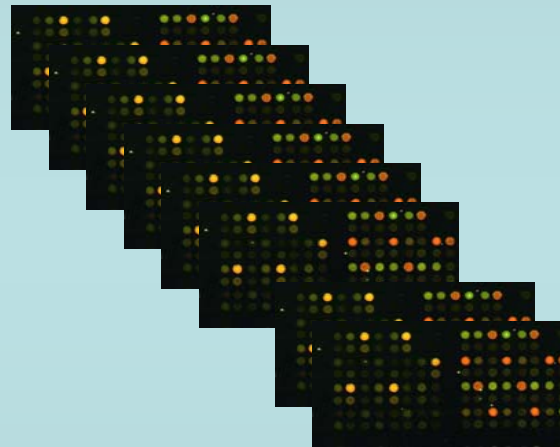
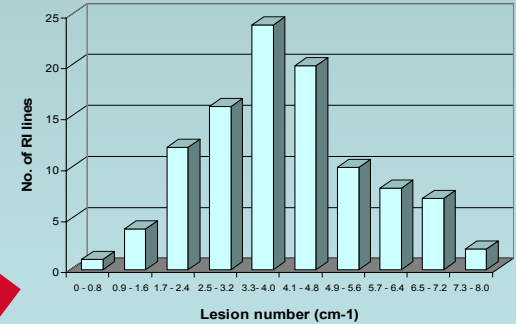
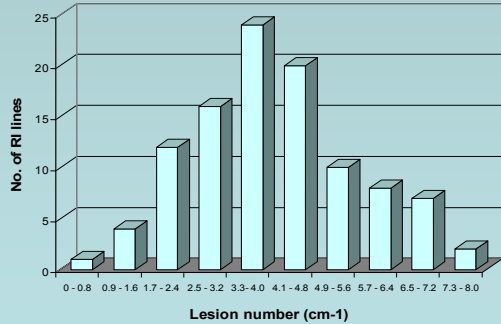
**Association genetics**

**Performance**

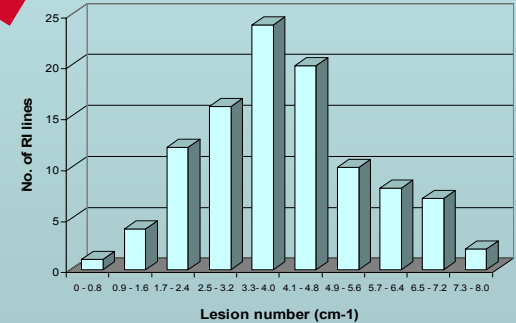
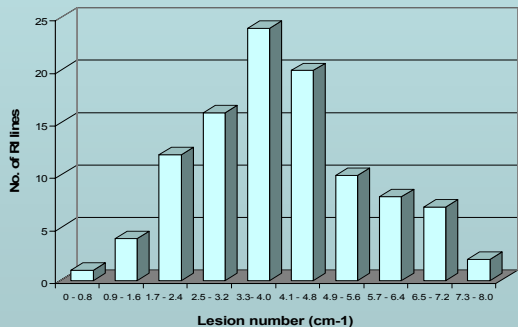
**Marker-aided gene  
pyramiding**



# Segregation genomics: make use of the power of genetic segregation to determine causative effects (not correlation)



Map expression levels of all genes on a reference population with multiple QTL and phenotypes segregating



# Co-expression analysis

---

- Co-expression of pairs of orthologs between species to predict functions
- Exploit evolutionary conservation to reveal function
- Used across human, mice, nematodes, drosophila, yeast
- Opportunities in crops
  - Data mining and experimental validation

Vera van Noort et al. 2003. TIG 19:238

Stuart et al. 2003. Science 302: 249

McCarroll et al. 2004. Nature Genetics: 36:197

# Identifying functionally conserved genes:

Lessons from Stuart et al. "Gene-Coexpression Network...." Science 302:249-255

---

Define metagenes (groups of orthologs with reciprocal best hits): 6307 metagenes from human-yeast-fly-worm

Pairs of metagenes: coexpressed in one organism and also in different organisms

Identify pairs of co-expressed metagenes in multiple organisms: 3416 metagenes, connected by 27,163 expression interactions ( $P < 0.05$ )

Functional terrains: discover new genes falling in same components

Validation of function by Gene-knockouts, gene silencing (RNAi)

# SNP in candidate genes conditioning complex traits

- Lazzaro et al. (2004)
  - Bacterial disease resistance in *Drosophila*.
  - SNP in 27 candidate defense genes
  - Variation in coding region vs upstream non-coding region.
  - Variation in regulatory genes vs variation in downstream genes
- Cohen et al. (2004).
  - SNP in 3 major genes conditioning cholesterol levels in human
  - Use selected, phenotypic extremes

# Targeting Induced Local Lesions IN Genomes (TILLING)

---

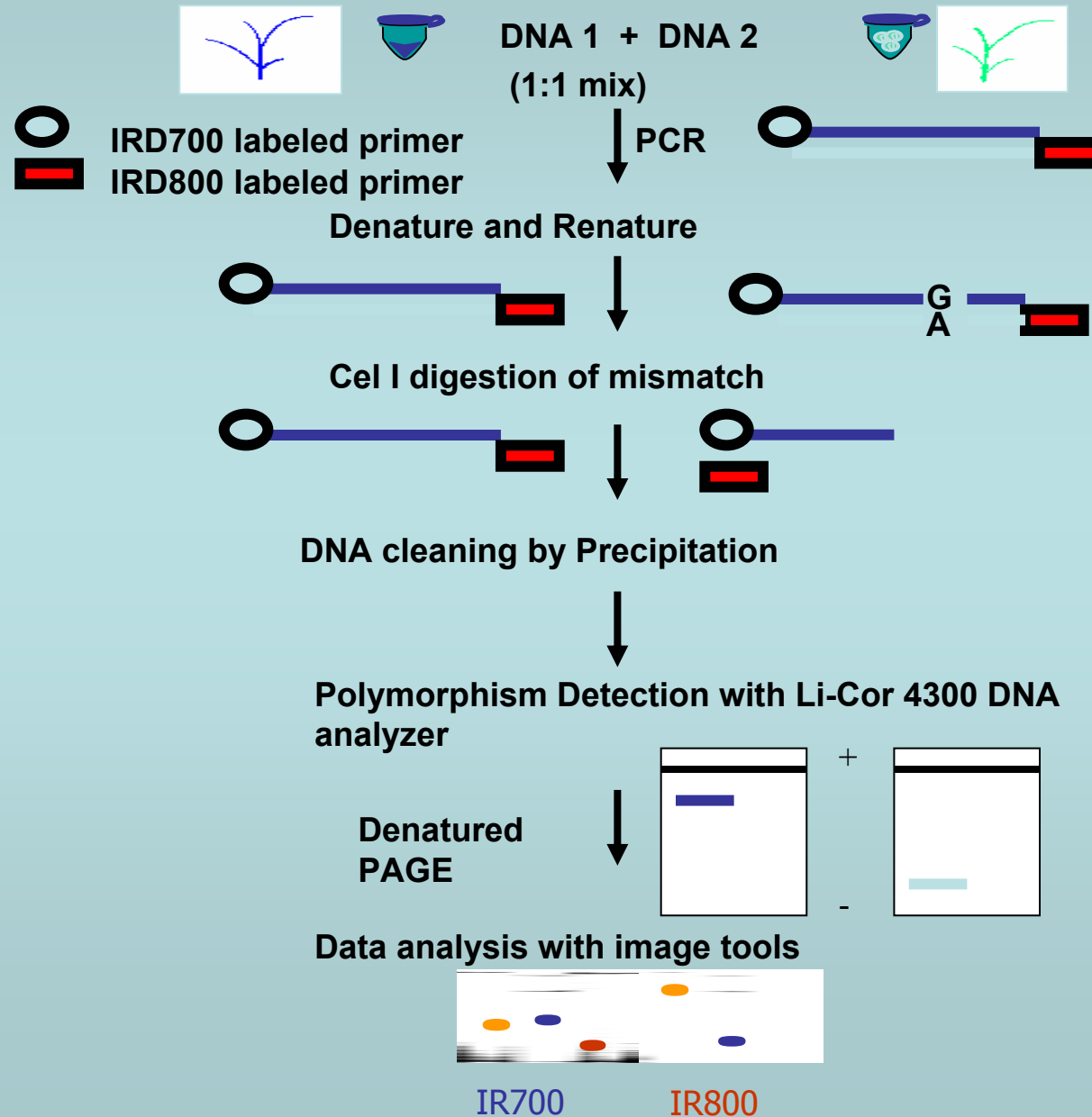
- A reverse genetics method
- The Arabidopsis TILLING project
- Applications to other organisms: rice, maize, hexaploid wheat, zebrafish, fly
- **Detection of natural sequence variation**



Steve Henikoff, Fred Hutchinson Cancer Research Center  
Luca Comai, University of Washington

*Henikoff and Comai 2003. SNP for Plant Functional Geomics. Annu Rev. Plant Biol 54:375-401*

# TILLING to detect allelic variants



**Fig. 1.** EcoTILLING of TPS on *Oryza sativa*. Trehalose 6-phosphatase (TPS) primers tested on panel of 46 *Oryza sativa* germplasm accessions contrasted to IR64. Circles indicate a clear SNP detected in the IRD700 (a) and IRD800 (b) channels for Kun Min Tsieh Hunan. This germplasm was pooled with IR64 and one or more others at 1:2, 1:4, 1:8, 1:12, and 1:16. Mismatch detected in 1:16 pool for IRD700 (c) and in 1:8 pool for IRD800 (d).

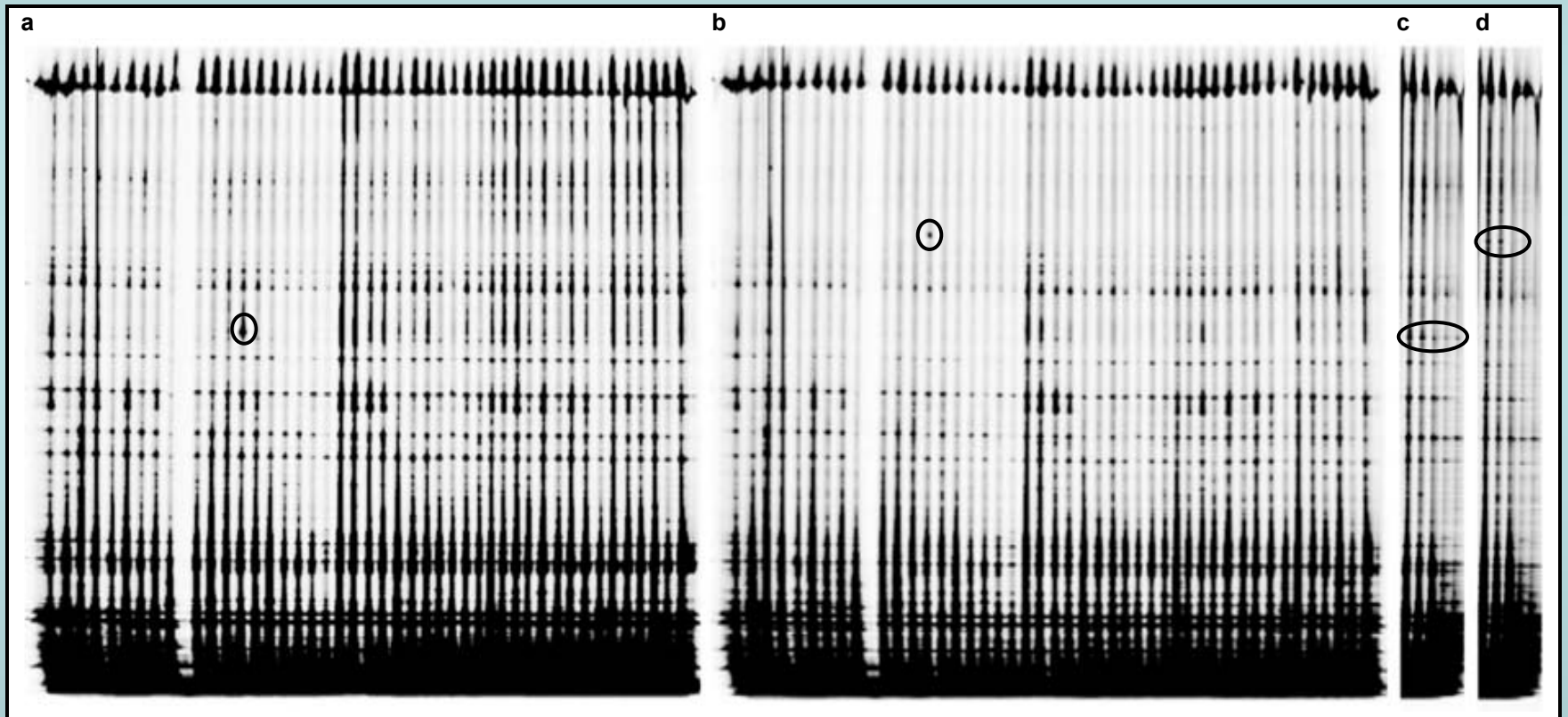
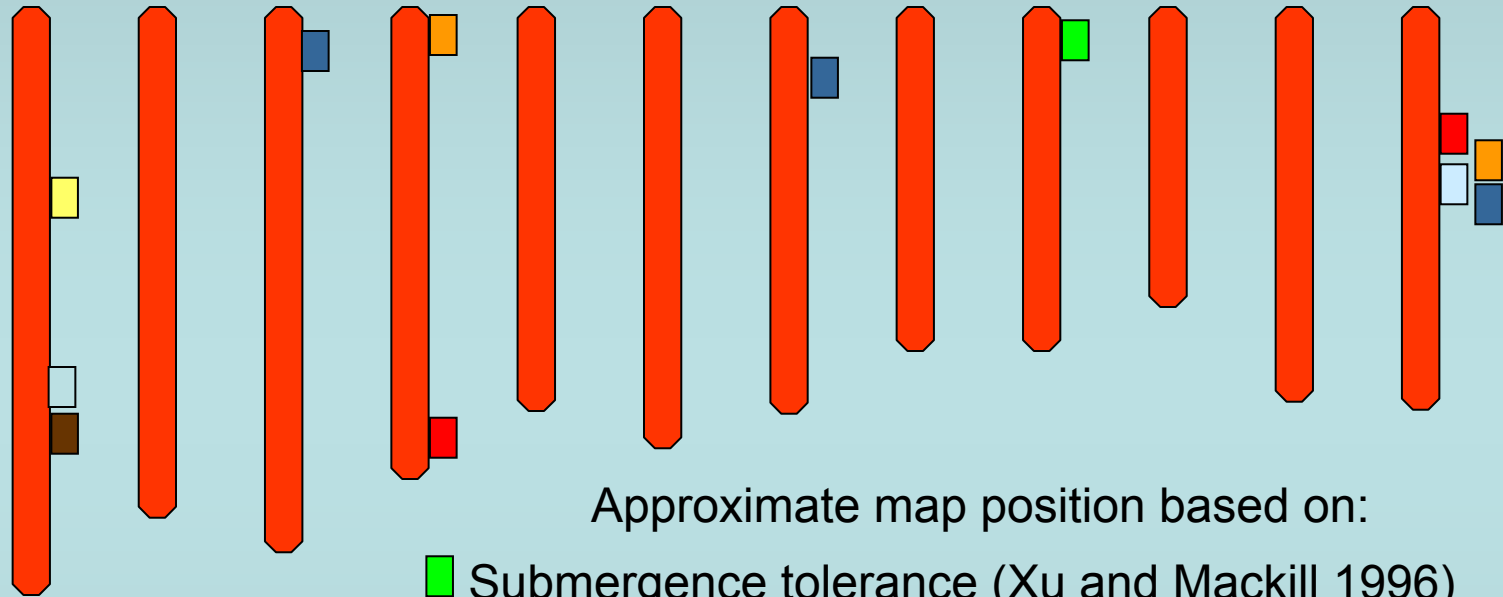


Table 4. Drought candidate gene targets for EcoTILLING implementation in *Oryza* (to be used in *Sorghum*).

Gene	Function	Primers	Sequence L (IRD700), R (IRD800)	CHR	Prod (bp)	BAC/PAC clone	Region	Amplicon
Pp2a4	protein phosphatase	pp2a4L	ggttggggcatatctcctcgtggt	CH10	928	nbxb0015J03		
		pp2a4R	Tcctaggagctggttcaaactgcaa	CH03	668	OSJNBa0091P11		
14-3-3	membrane stability	CG18-1L	ggcatgcttgattgggacatgaga	CH02	1005	B1079H01	75278-76282	-1000 - exon 1
		CG18-1R	tctcctcatacctctcggcctgctc		838		76215-77052	Exon 1 – Exon 3
		CG18-2L	Tgagctttccctgaggagaatgtg					
		CG18-2R	caagattgcaagcacggtcaggaga					
		CG18-3L	acctcgcagagttaagactggagctgaga					
		CG18-3R	agatatcgggaagtcacagcgtcaggttat					
DREB1	drought response binding protein 1	DREB1-1L	acgtcaaaacccaacccaacccat	CH01	988	P0701D05	33870-34857	-1000 - 5'-UTR
		DREB1-1R	gattgaatcagggccatcgctttc		969			
		DREB1-2L	ccgttgattgctgatagcctcctga					
		DREB1-2R	tgaaatattcctattgaccgcagca					
TPS	Trehalose 6- phosphatase	TPS-1L	ctatcttgggctcatggcgtgactg	CH02	1000	P0708H12	10934-11933	-1000 - Exon 1
		TPS-1R	cgcagcattcgcaaccaacaaata		997			
		TPS-2L	ggcacactgtgcctattgtggatg					
		TPS-2R	gtttacgagccgtgaccagtttc					
VP14	viviparous14 (9-cis-epoxy- carotenoid dioxygenase)	VP14-1L	gccattgcaaagcccagggtgtat	CH12	1065	OSJNBb0101110	15368-16432	-1000 - Exon 1
		VP14-1R	cccctcttcttctgccacacctc		1013			
		VP14-2L	tggcaagaagaaggatgggctgaac					
		VP14-2R	tccacaggtggaagcagaagcagtc					
MAP kinase	Transcriptional control of stress response	MAPK-1L	caccatctccttcagcctccgtttc	CH7	975	QJ1699_E05	36834-37808	Exon 1- Exon 4
		MAPK-1R	cacacctccacccaatcaaattcc		1006		38419-39424	Exon 6 – Exon 7
		MAPK-2L	atgggcatgaagatgagcctctgg					
		MAPK-2R	gggtgccggctatggtacgctagat					
Extensin	Cell-wall protein for growth expansion	Ext-L	aggagaagatggcgatggccaataa	CH10	977	OsjnbA0093B11	85447-86423	Exon 3
		Ext-R	gagctcgagatgctgatgatgtcc					
TRAB1	ABA responsive TF	TRAB1-L	gcaggggtcgtactactactgacg	CH8	982	P0013B04	26512-27493	Exon 1
		TRAB1-R	tcaacacagcaaggaccaatctcg					
CDPK	Signal transduction, regulatory cascade	CDPK-L	atthtgattccgatccctggcctgt	CH2	1000	QJ1717_A09	96775-97774	Exon 2 – Exon 5
		CDPK-R	caaggaaccagggtggcagacttca					
FIERY1	Signal cascade (Inositol polyP 1-Pase)	FIERY1-L	cctctctgccaccggccttatctt	CH3	997	QSJNBa0094J08	26843-27839	Exon 1 – Exon 3
		FIERY1-R	cagggaaagaaattggaagggcagtg					
Sucrose synthase	Carbon reallocation	Suc-L	cgctcagcgagtgctgagactatt	CH7	1004	B1056G08	71840-71973	Exon 2 – Exon 6
		Suc-R	ggcgattgtacactgcacatgatgg					
BZIP	Transcriptional control	Bzip-L	atggcatcggagatgagcaagaacg	CH1	991	P0489B03	42505-43495	5'UTR - Exon 1
		Bzip-R	ttcggagcaacatgtgtacctgcttt					

Extended list of >32 includes other drought genes and genes for disease, submergence, and grain quality.

# Target QTLs for Abiotic Stress Tolerance in Rice



Approximate map position based on:

- Submergence tolerance (Xu and Mackill 1996)
- Deepwater elongation (Sripongpankul et al. 2002)
- Drought (Babu et al. 2003)
- Al toxicity (Nguyen et al. 2003; Wu et al. 2000)
- P uptake (Wissuwa et al. 1998)
- Salt tolerance (Bonilla et al. 2002)
- Cold tolerance (Andaya and Mackill 2003)
- Fe toxicity tolerance (Wan et al 2003)

Genomic resources

Orthologous genes  
phylog

Gene expression

Functional validation  
mutants  
RNAi, VIGS  
Linkage disequilibrium  
NIL, gene pyramids



# SP2 Highlights

- Cluster activities
- Tangible collective results
  - Sharing materials
  - Cooperative process
  - Commons-based strategies: peer-production (Benkler 2004. *Science* 305:1110)
- Lead-in to more detailed discussion the next two days

# SP2 year 1 commitments and deliverables

## **CL1. Germplasm evaluation (drought-focus)**

- Detailed phenotypic characterization of selected lines and populations
- Adoption and evaluation of phenotyping protocols

## **CL2. Marker development**

- Test and validate consensus markers, develop comparative maps (dicot and monocots)
- Assemble sets of orthologous candidate genes for drought resistance

## **CL3 Gene expression**

- Consolidate genes/sequences for targeted stress tolerance response pathways
- Compare expression profile of selected genetic stocks (wheat, rice, maize barley, potato, Arabidopsis)-
- Resource (EST) development (cassava, bean, cowpeas)

## **CL4 QTL analysis**

- Drought-tolerance phenotypic data on segregating population and genetic mapping
- Comparative analysis of genotypic and phenotypic data to infer genet/QTL positions
- Data mining of drought-QTL on multiple crops maps

# CL1. Germplasm evaluation (drought-focus)

- A set of germplasm with drought-tolerant attributes or representative of donor gene pools has been selected from individual crops (e.g., sorghum, barley, wheat, rice, maize, chickpea, common beans, cassava, and potato) for detailed phenotypic analysis under water stress conditions.
- Phenotyping protocols for assessing on drought tolerance are being implemented in selected crops and field methodologies for drought tolerance are under testing using advanced breeding lines.

# Common bean

## Mechanisms of drought adaptation

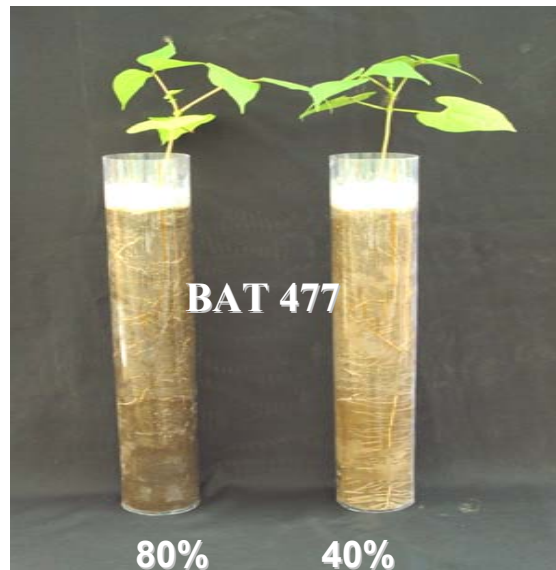
- Yield under drought
- Early maturity
- Deep rooting
- Efficient translocation of photosynthates under stress
  - Grain filling index, seed P concentration, seed and shoot TNC; seed ash content

**J. White, S. Singh, S. Beebe, I. Rao**

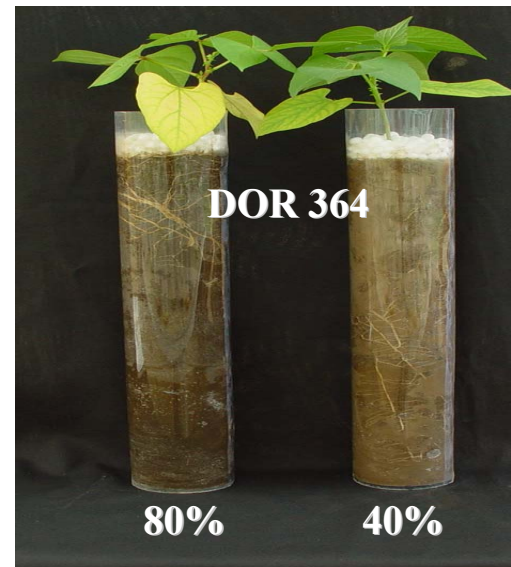
# Rooting Strategy: Greenhouse trials

**Beans grown in soil cylinders**

**Drought tolerant  
Deep branched roots**



**Drought sensitive  
Shallow roots**

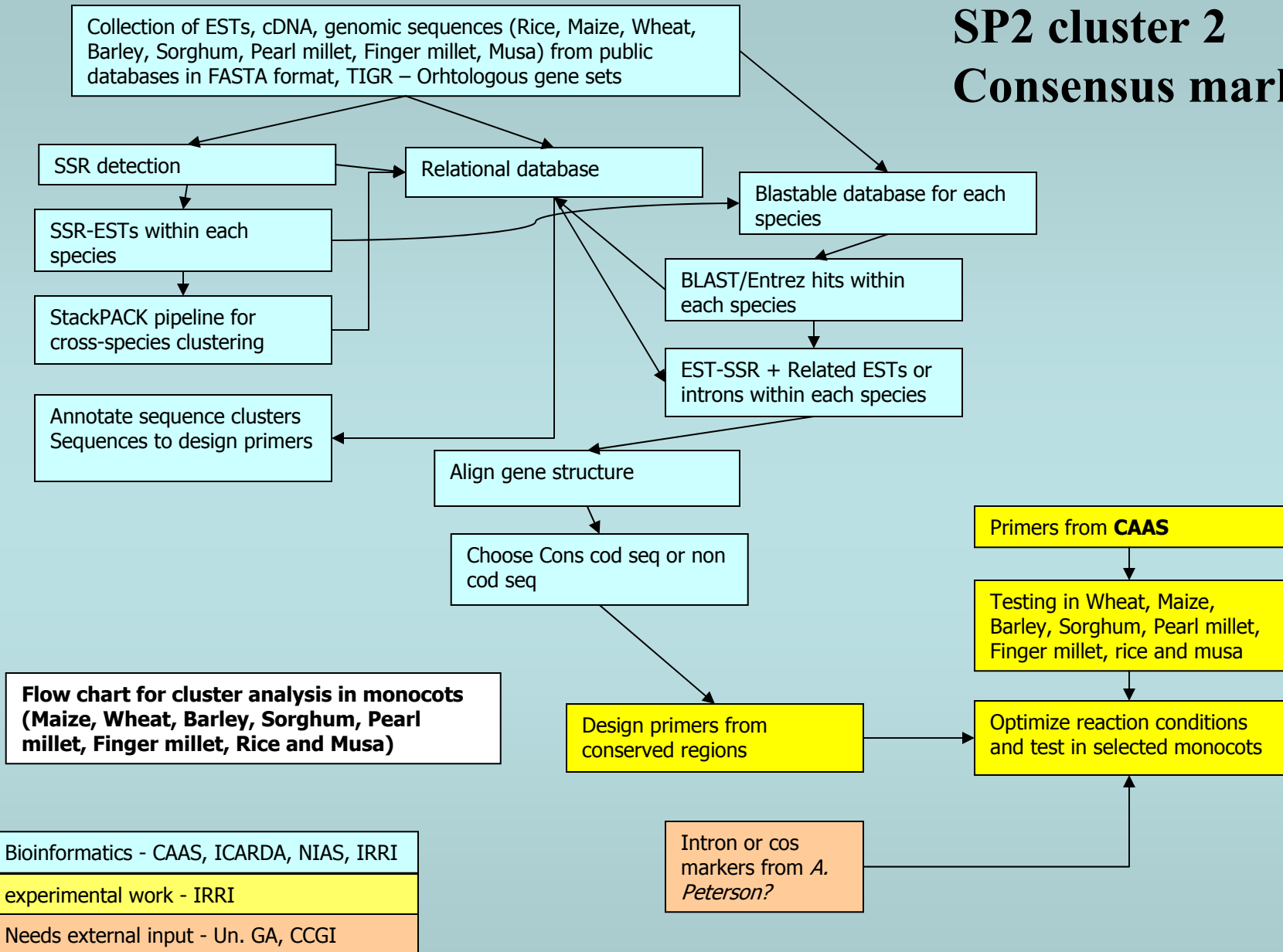


## CL2. Marker development

- **Conserved orthologous markers (COS) are under design and evaluation in monocots and dicots.**
- **Algorithms for extraction of candidate sequences for conserved genes from EST databases are being tested and shared.**

# SP2 cluster 2

## Consensus markers



## Objective

Development of conserved orthologous set of markers and genes across monocots

## Germplasm

- Rice (Nipponbare, Azucena, IR36, Co39)
- Wheat (Trigo-1, Trigo-2, Trigo-3, UPLB-14, UPLB-16)
- Maize (Phl-11310, GB-51352, GB-53961, GB-53962, GB-53963)
- Sorghum (A-1, A-2, A-3, A-244, A-2231)
- Banana (Lacatan, Philippine local)
- Onion (Philippine red variety)

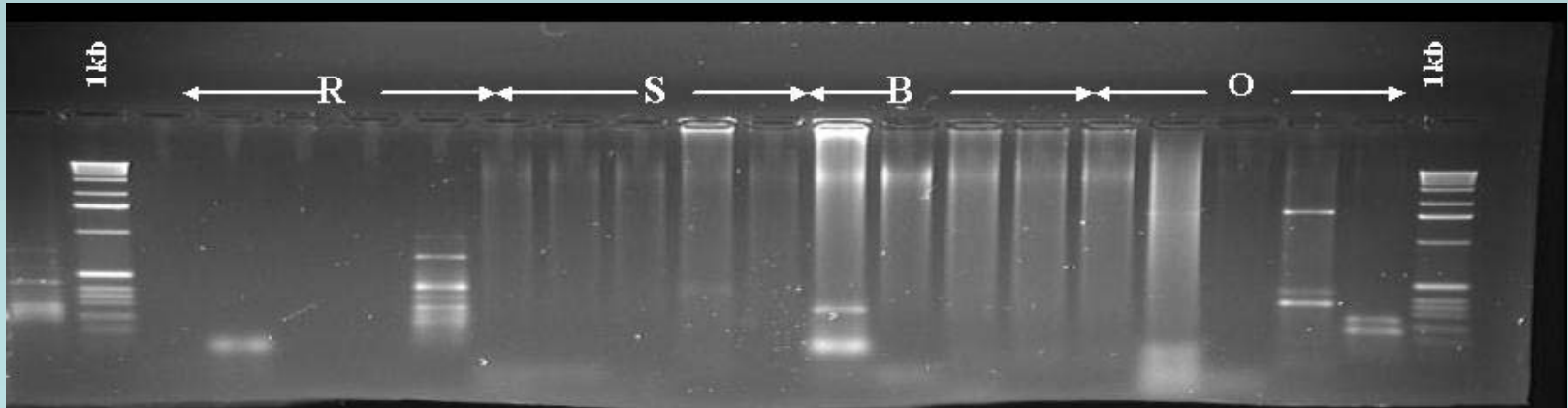
## Primers used in this study

Set of 96 primers from CAAS will be tested across monocots (Wheat, Sorghum, Maize, Banana, Onion, and Rice).

Optimizing PCR conditions for primer amplification and PCR products.

These primers already found to work in wheat germplasm.

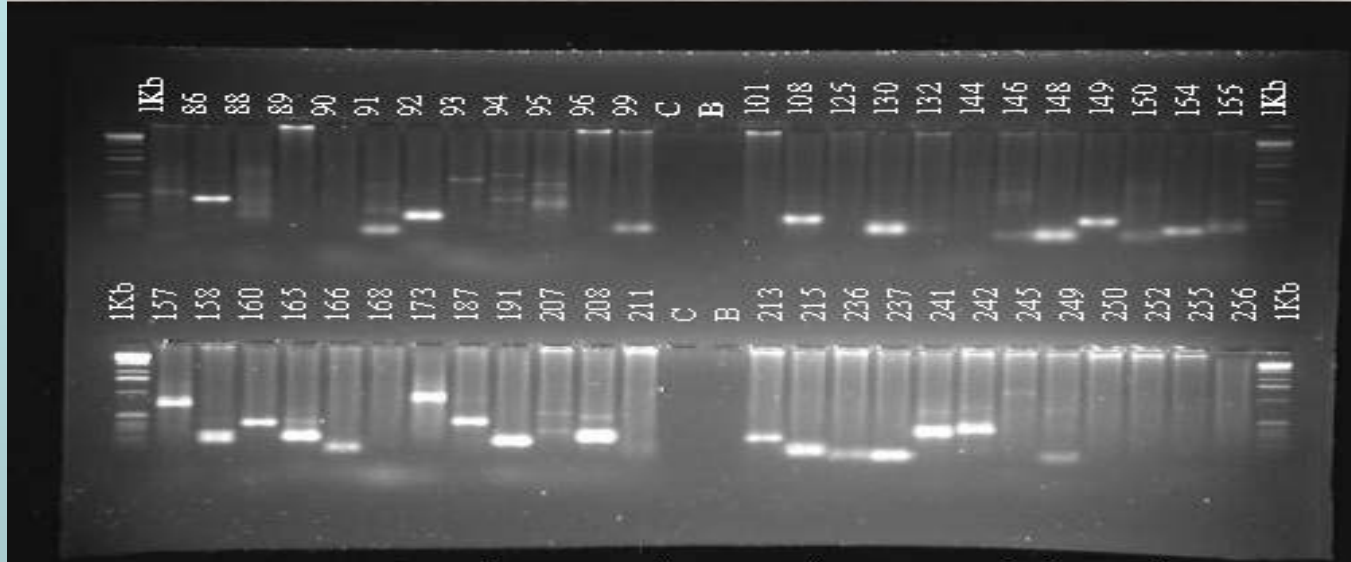
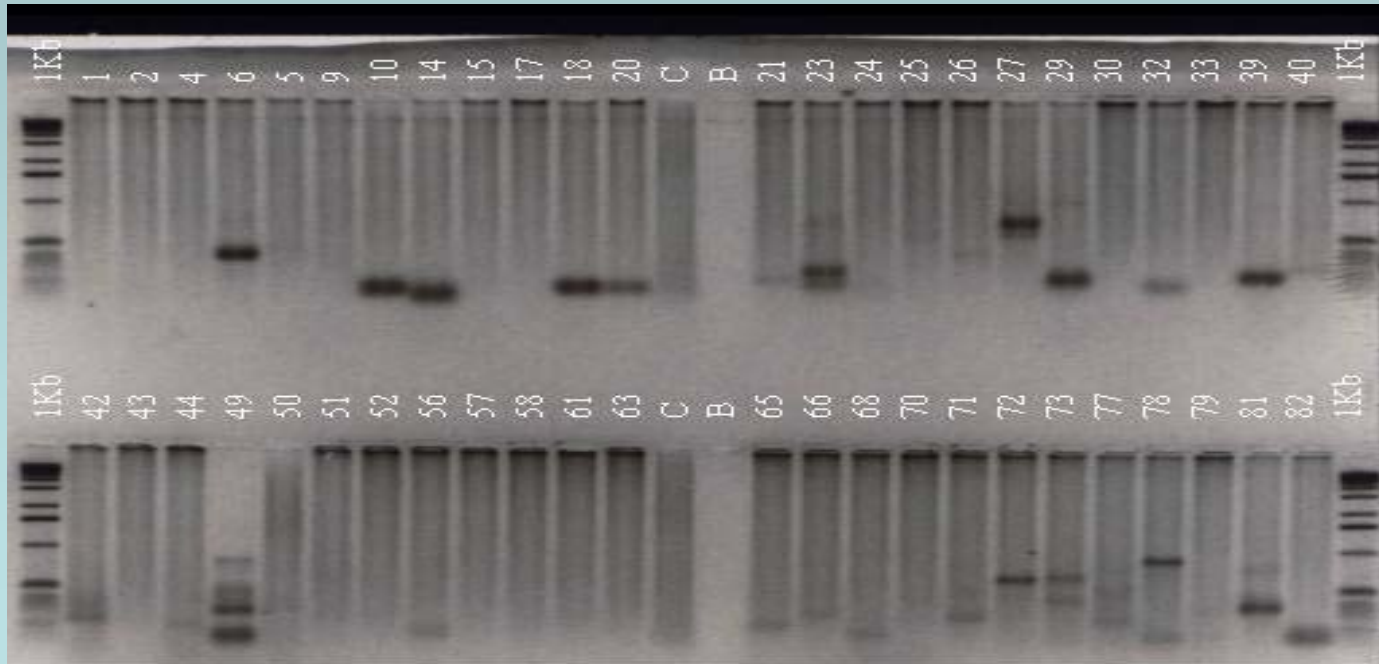
## Random sample of CAAS wheat EST derived primers



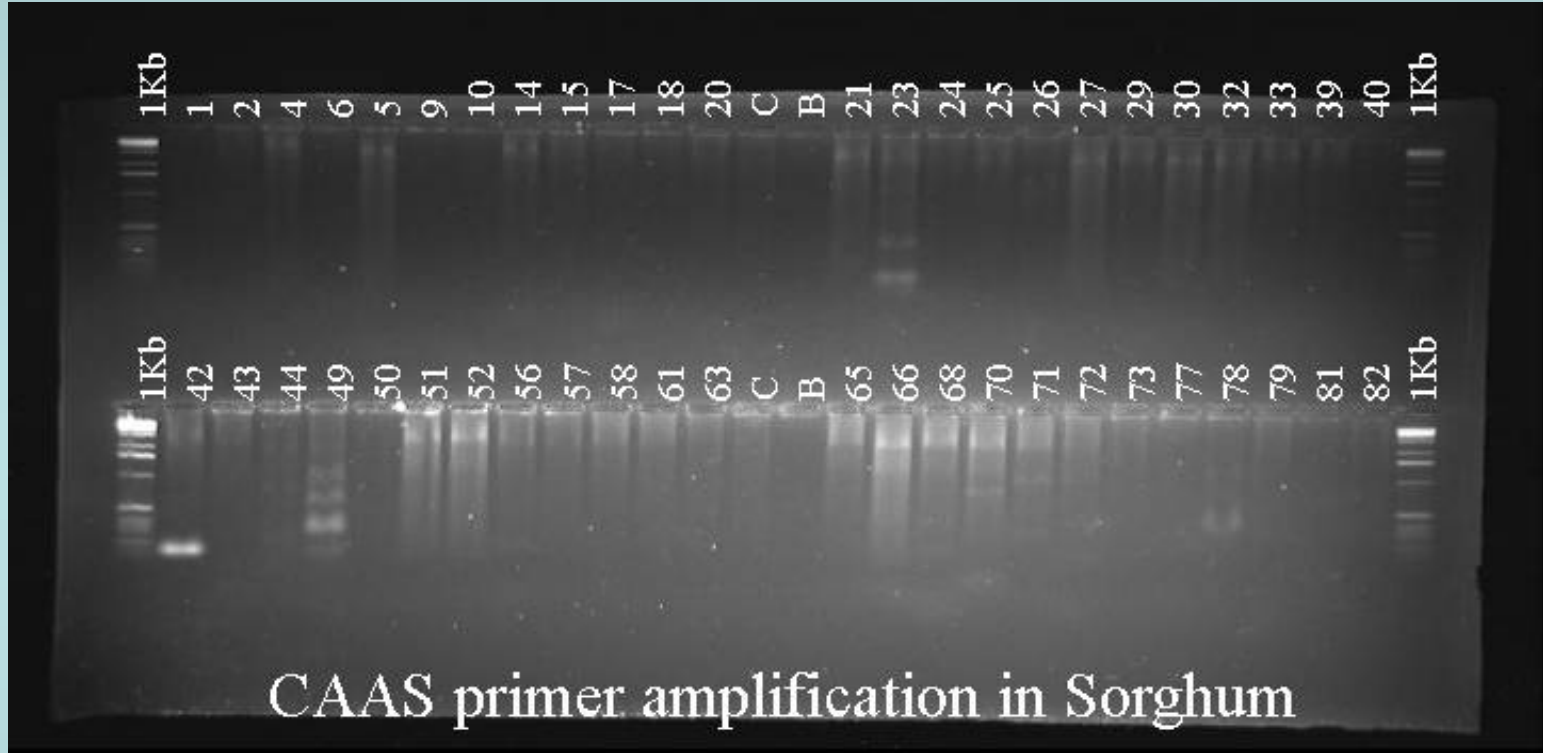
Testing of primers across monocots (rice, sorghum, banana, onion)



Randomly tested primers in rice with 2 different PCR profiles



CAAS primers (1-96) tested in wheat



# CL3 Gene expression

- **Producing and evaluating gene expression datasets for cross-species comparative analyses.**
- **EST libraries expanded for several species under stress conditions.**

# Co-expression analysis

- Co-expression of pairs of orthologs between species to predict functions
- Exploit evolutionary conservation to reveal function
- Used across human, mice, nematodes, drosophila, yeast
- Opportunities in crops
  - Data mining and experimental validation

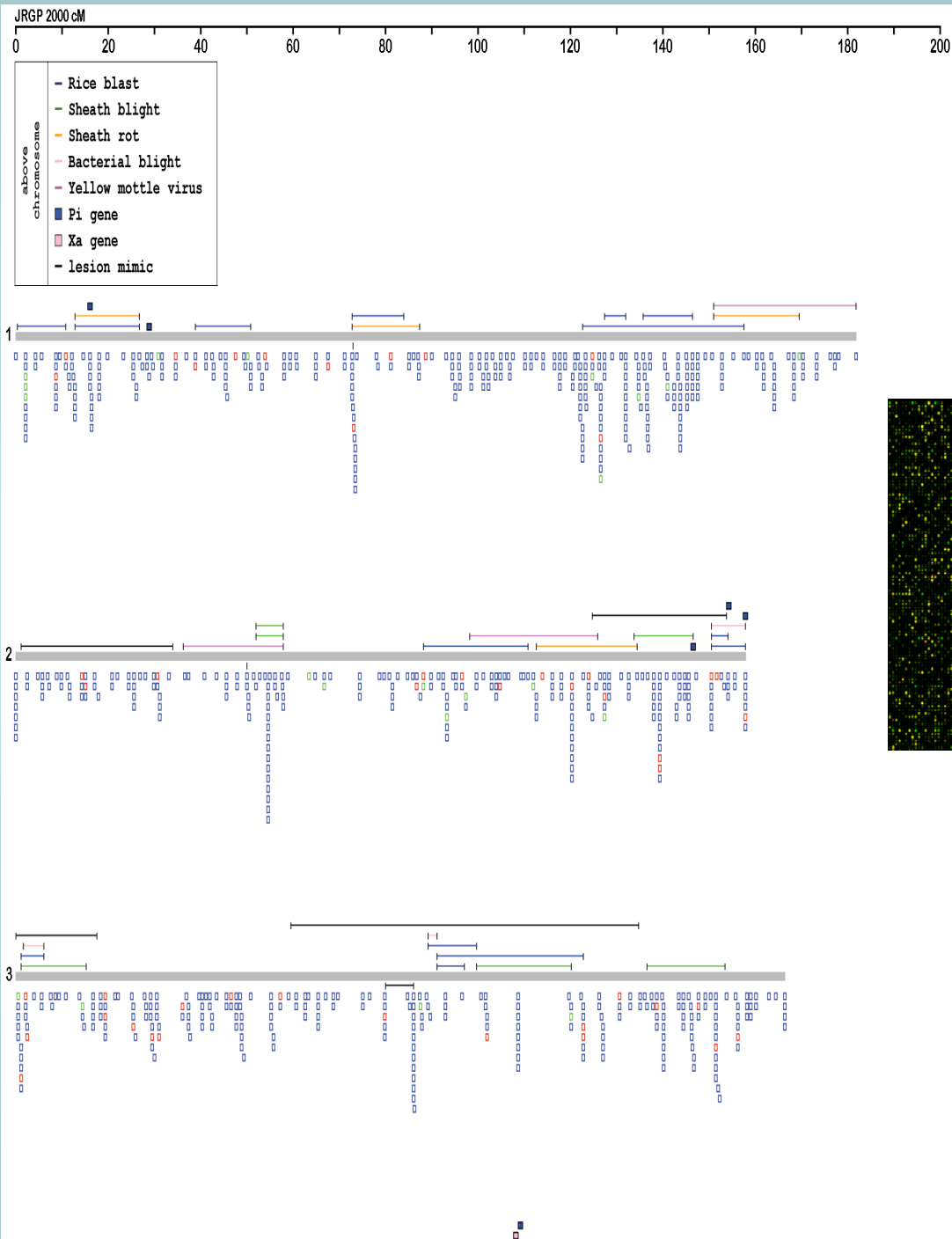
Vera van Noort et al. 2003. TIG 19:238

Stuart et al. 2003. Science 302: 249

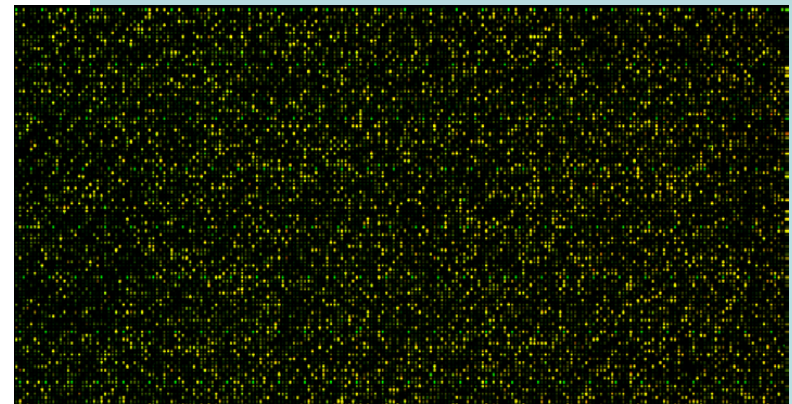
McCarroll et al. 2004. Nature Genetics: 36:197

# Comparative genomics and functional analysis

- **Gene function within species**
  - Mutants (forward and reverse genetics)
  - map-based analysis
    - Consensus map
  - segregation genomics
- **Gene function inferred by cross-species comparison**
  - Functional relationships between orthologs
  - Co-expression and gene regulation
- **Functional SNP**
  - Allelic variation associated with function
  - Association genetics at the genic level



# Binning up/down expressed genes into stress QTL consensus map



22,000 Oligo Array (NIAS-Agilent) hybridized with RNA from a disease resistant *indica* variety SHZ-2 at 48 hours after inoculation by the blast pathogen (NIAS and IRRI)

Satoh/Kikuchi, NIAS,  
Liu et al IRRI,  
Randy, Nelson Cornell

# CL4 QTL analysis

- **Mapping population characterization for drought QTL analysis.**

Table 4. Drought candidate gene targets for EcoTILLING implementation in *Oryza* (to be used in *Sorghum*).

Gene	Function	Primers	Sequence L (IRD700), R (IRD800)	CHR	Prod (bp)	BAC/PAC clone	Region	Amplicon	
Pp2a4	protein phosphatase	pp2a4L	ggttggggcatatctcctcgtggt	CH10	928	nbxb0015J03			
		pp2a4R	Tcctaggagctggttcaaactgcaa	CH03	668	OSJNBa0091P11			
14-3-3	membrane stability	CG18-1L	ggcatgcttgattgggacatgaga	CH02	1005	B1079H01	75278-76282	-1000 - exon 1	
		CG18-1R	tctcctcatacctctcggcctgctc		838		76215-77052	Exon 1 – Exon 3	
		CG18-2L	Tgagctttccctgaggagaatgtg				1140	76699-77838	Exon 2 – Exon 4
		CG18-2R	caagattgcaagcacggtcaggaga						
		CG18-3L	acctcgcagagttaagactggagctgaga						
		CG18-3R	agatatcgggaagtcacagcgtcaggttat						
DREB1	drought response binding protein 1	DREB1-1L	acgtcaaaaccaacccaacccat	CH01	988	P0701D05	33870-34857	-1000 - 5'-UTR	
		DREB1-1R	gattgaatcagggccatcgctttc		969				
		DREB1-2L	ccgttgattgctgatagcctcctga						
		DREB1-2R	tgaaatattcctattgaccgcagca						
TPS	Trehalose 6- phosphatase	TPS-1L	ctatcttgggctcatggcgtgactg	CH02	1000	P0708H12	10934-11933	-1000 - Exon 1	
		TPS-1R	cgcagcattcgcaaccacaaata		997				
		TPS-2L	ggcacactgtgcctattgtggatg						
		TPS-2R	gtttacgagccgtgaccagtttc						
VP14	viviparous14 (9-cis-epoxy- carotenoid dioxygenase)	VP14-1L	gcccattgcaaagcccagggtgtat	CH12	1065	OSJNBb0101110	15368-16432	-1000 - Exon 1	
		VP14-1R	cccctccttcttctgccacacctc		1013				
		VP14-2L	tggcaagaagaaggatgggctgaac						
		VP14-2R	tccacaggtggaagcagaagcagtc						
MAP kinase	Transcriptional control of stress response	MAPK-1L	caccatctccttcagcctccgtttc	CH7	975	QJ1699_E05	36834-37808	Exon 1- Exon 4	
		MAPK-1R	cacacctccacccaatcaaattcc		1006		38419-39424	Exon 6 – Exon 7	
		MAPK-2L	atgggcatgaagatgagcctctgg						
		MAPK-2R	gggtgccggctatggtacgctagat						
Extensin	Cell-wall protein for growth expansion	Ext-L	aggagaagatggcgatggccaataa	CH10	977	OsjnbA0093B11	85447-86423	Exon 3	
		Ext-R	gagctcgagatgcctgatgatgtcc						
TRAB1	ABA responsive TF	TRAB1-L	gcaggggtcgtactactcactgacg	CH8	982	P0013B04	26512-27493	Exon 1	
		TRAB1-R	tcaacacagcaaggaccaatctcg						
CDPK	Signal transduction, regulatory cascade	CDPK-L	atthtattccgatccctggcctgt	CH2	1000	QJ1717_A09	96775-97774	Exon 2 – Exon 5	
		CDPK-R	caaggaaccagggtggcagacttca						
FIERY1	Signal cascade (Inositol polyP 1-Pase)	FIERY1-L	cctctctgccaccggccttatctt	CH3	997	QSJNBa0094J08	26843-27839	Exon 1 – Exon 3	
		FIERY1-R	cagggaaagaaattggaagggcagtg						
Sucrose synthase	Carbon reallocation	Suc-L	cgctcagcagtgctgagactatt	CH7	1004	B1056G08	71840-71973	Exon 2 – Exon 6	
		Suc-R	ggcgattgtacactgcacatgatgg						
BZIP	Transcriptional control	Bzip-L	atggcatcggagatgagcaagaacg	CH1	991	P0489B03	42505-43495	5'UTR - Exon 1	
		Bzip-R	ttcggagcaacatgtgtacctgcttt						

Extended list of >32 includes other drought genes and genes for disease, submergence, and grain quality.

**Fig. 2. EcoTILLING of DREB1 on wild *Oryza*.** DREB1 amplicons of a similar size were produced from 47 wild *Oryza* species using primers designed from Nipponbare (a). Mixtures of these products were combined with either IR64 (left set of lanes) or Nipponbare (right set of lanes). The IRD700 channel is shown in panel b and the IRD800 in c. Circles indicate clear difference present in both channels. Arrows indicate a mismatch present in japonica but absent from most wild and indica.

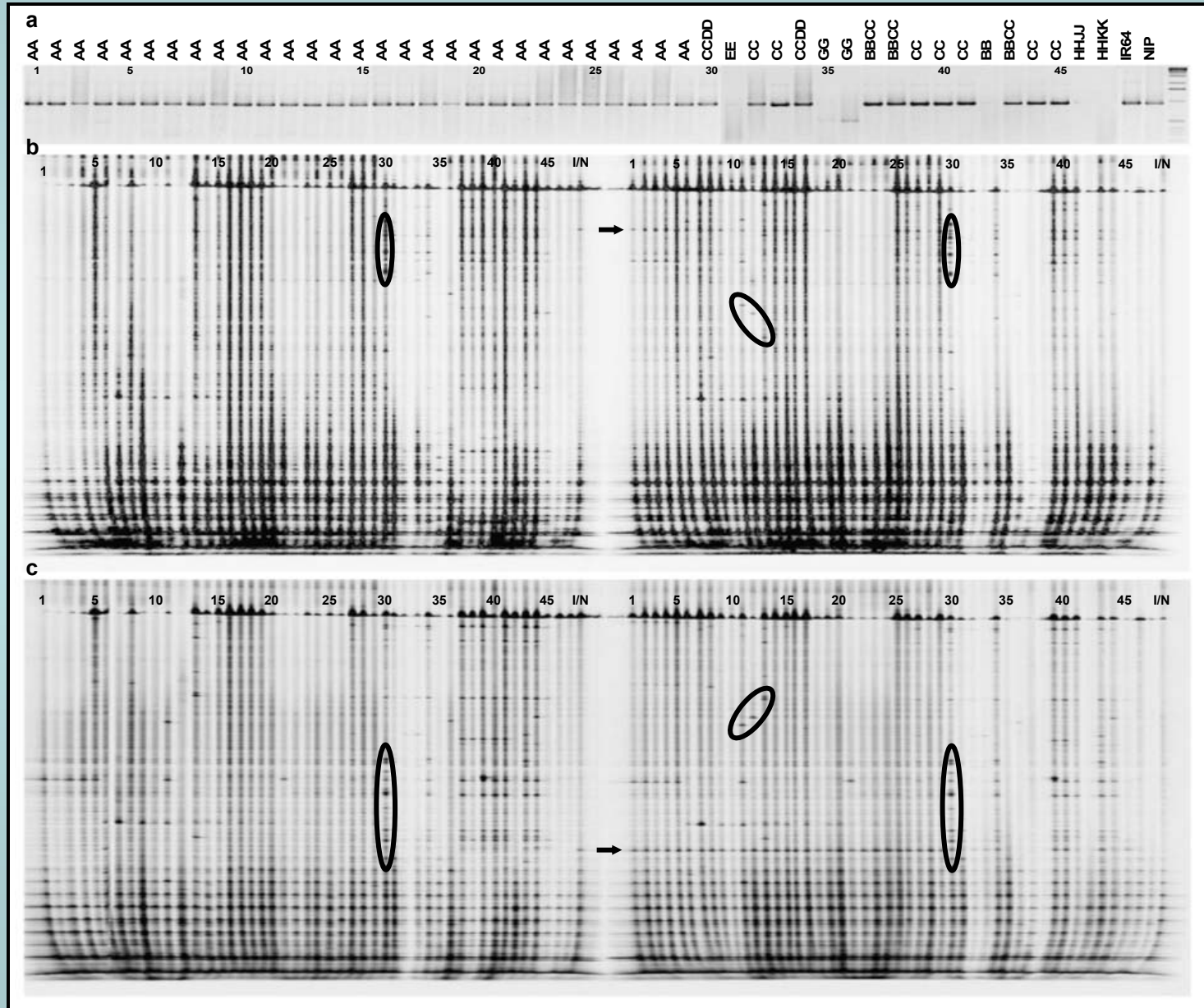


Table 1. *Oryza sativa* germplasm used for EcoTILLING implementation.

Germplasm	ACCNO	VG	Germplasm	ACCNO	VG
NIPPON-BARE	12731	temp jap	JING XI 17	67676	ind
IR64	66970	ind	TE-QING	81093	ind
CT9993-5-10-1-M		ind/jap	VANDANA		aus
IR62266-42-6-2		ind/jap	N 22	6264	aus
CAROLINA GOLD	1723		ARC 10177	12386	aus
KALUKANTHA	7755		JAGLI BORO	27516	aus
VARY VATO 462	10964		DULAR	32561	aus
FANDRAPOTSY 104	10984		AUS JOTA	66767	aus
MILYANG 23	34393	ind/jap	ASWINA	26289	dp3
BLACK GORA	40275		MATIA AMAN 53-13	37764	dp4
PACHOLINHA	50531		RAYADA	77210	dp4
APO (IR55423-01)		ind	BASMATI 370	6426	aro
DEE-GEO-WOO-GEN	123	ind	CHHOTE DHAN	58930	aro
SINTANE DIOFOR	5418	ind	TCHAMPA	32368	aro
KUN MIN TSIEH HUNAN	8195	ind	IGUAPE CATETO	4122	jap
RTS 12	8234	ind	RATHAL	31524	jap
POKKALI	8948	ind	IR60080-46A		trop jap
KHAO KAP XANG	23423	ind	AZUCENA	328	trop jap
KHAO DAWK MALI 105	27748	ind	MOROBEREKAN	12048	trop jap
LEUANG PRATEW	27762	ind	YANCAOUSSA	16071	trop jap
PETA	32571	ind	CHODONGJI	55471	temp jap
IR36	39292	ind	GERDEH	32301	temp jap
BALA	50927	ind	LEMONT	66756	temp jap
CO 39	51231	ind	PATBYEO	55607	temp jap

Extended set of 1536 accessions (including these 48) are in the pipeline.

## Summary of EcoTILLING Results

- PP2a4 locus: Tested on 22 Korean *japonica* and tongil (*indica/japonica* hybrid) varieties contrasted to Nipponbare.  
SNP detected in Milyang30, Taebaek, Samgang, Sindongjin and Milyang23.  
Second was found in Milyang30 and Milyang23.  
Third occurred in Suwon345.  
SNP detected in SHZ2 by LTH, the parental lines of a breeding population.
- TPS locus: Tested on 48 cultivated germplasm contrasted to Nipponbare and IR64.  
Distinct SNP occurred in the Kun Min Tsieh Hunan with Nipponbare and IR64.  
Tested on the wild germplasm panel.  
Amplicons were obtained only for AA genome species.  
Several haplotypes occurred.
- DREB1 locus: Tested on 96 accessions of cultivated rice.  
SNP detected IR64 and Nipponbare is common to *indica/japonica* contrasts.  
Second haplotype occurred in Mimidam by IR64.  
Third was between Taichung Native 1 and Nipponbare.  
Tested on the wild germplasm panel.  
Amplicons obtained as distant as GG but not EE, HHJJ, or HHKK genomes.  
Many putative SNPs including one absent from most wild and *indica* but present in *japonica*.
- Pooling: Pools up to 1:8 definite -- 768 germplasm per gel.  
Pools up to 1:16 possible with better optimization.

# Overall collective progress

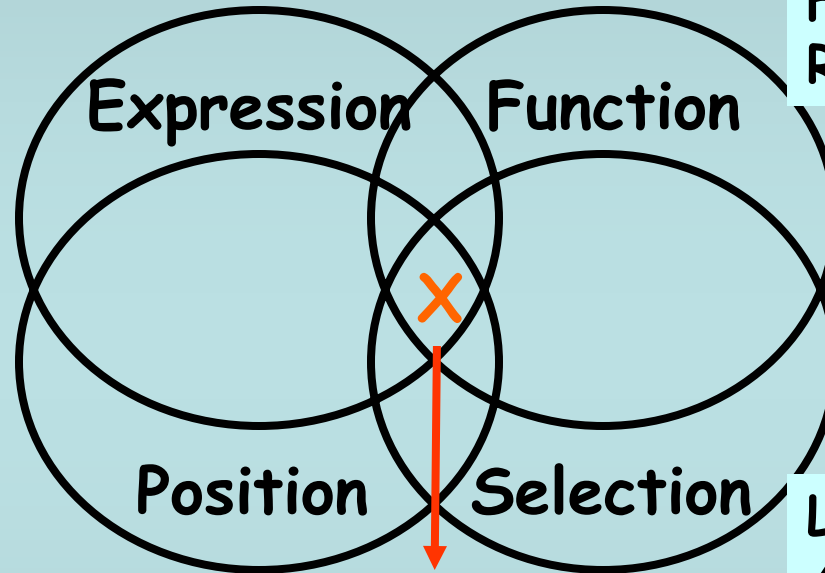
- Coordinated strategy in using diverse genetic materials yet exhibiting certain attributes of drought tolerance.
- With SP1, a common phenotyping framework (though not necessarily the same methodologies) in terms of techniques, plant developmental stages, parameters, has been developed to enable cross-species comparison.
- Selected stocks planted for detailed drought phenotyping in the field studies.
- Expanded characterization of QTL in selected species
- Initiated marker design using orthologous sequences across species.
- Genetic materials (markers, RNA) from different species shared among laboratories and being tested with different gene array platforms.
- Expanded and improved the characterization of EST libraries of *Musa* and cassava.



# Convergent evidence to identify candidate genes

RT-PCR  
Microarrays  
Proteomics  
Metabolomics  
*In situ*  
hybridization

Sequence homology  
Expression clusters  
Protein 3D prediction  
Forward genetics  
Reverse genetics



Genetic map  
Physical map  
FISH map

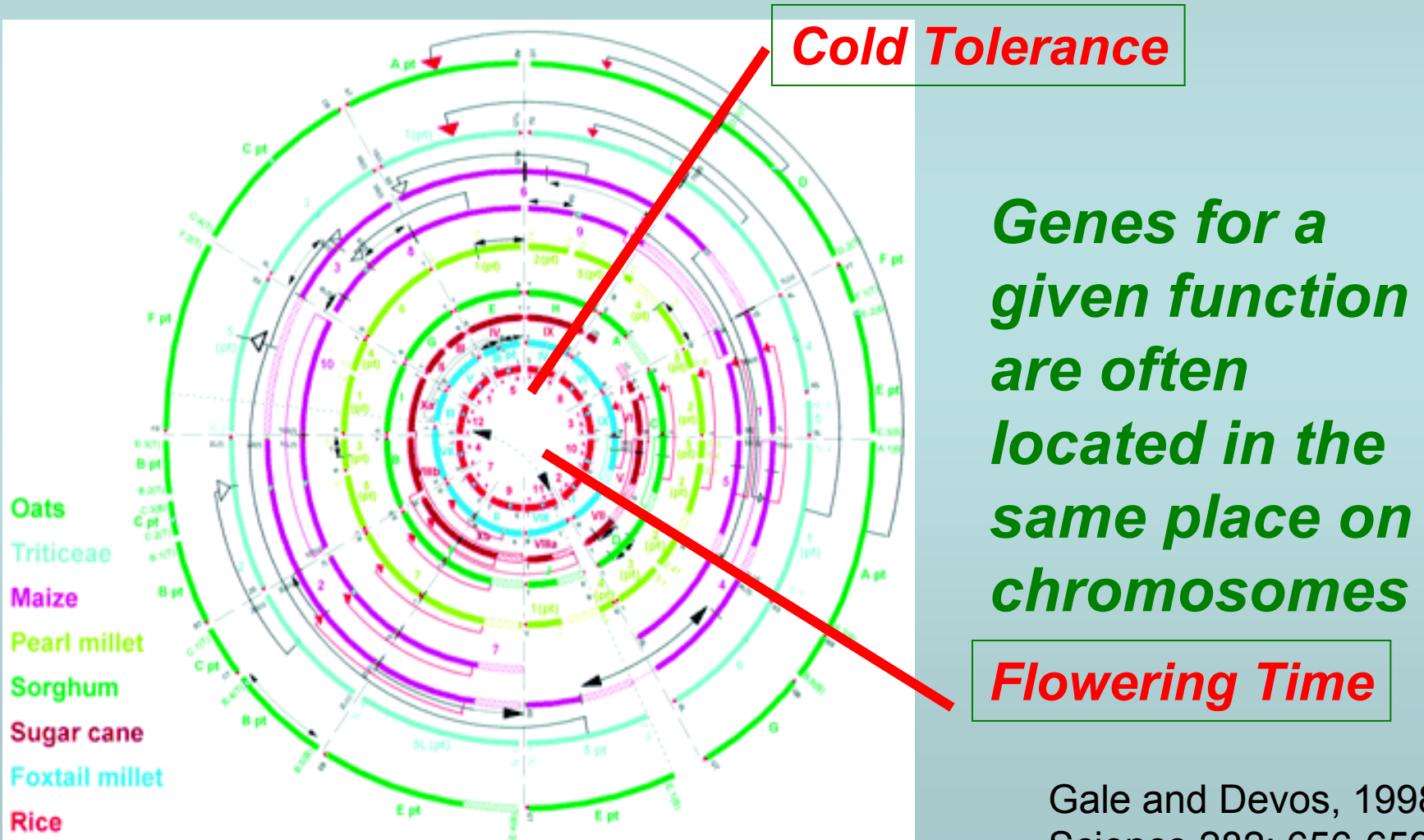
Linkage disequilibrium  
Association genetics  
Pedigree analysis

**Candidate genes**

Application to other  
crops

Validation through:  
transformation  
marker-aided selection

# Genomes of different cereals are similarly arranged



Gale and Devos, 1998  
Science 282: 656-659.

# Networks: capitalizing on investment in training and capacity building

---

- **Training and capacity building embedded in research projects.**
- **Thesis research.**
- **Mentoring:** sister-lab relationship to support young scientists to conduct work at home institutions.
- **Conduct targeted workshops and technical backstopping by CP researchers.**
- **Continuing education:** packaged activities: shuttle research, thesis work, backstopping workshops, and/or mentoring
- **Mobile or virtual resource library.**
- **Lighthouse concepts**—successful cases and models to mobilize national support