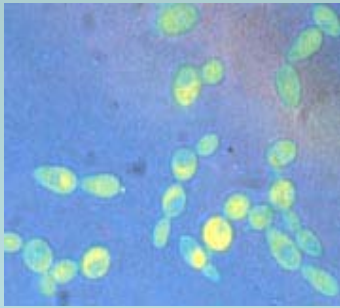


Subprogram 2: Comparative genomics for gene discovery

- Scope and contents of SP2
- Rapidly changing external environments
- Examples of technical approaches
- Peer-Production

Decade of genomes



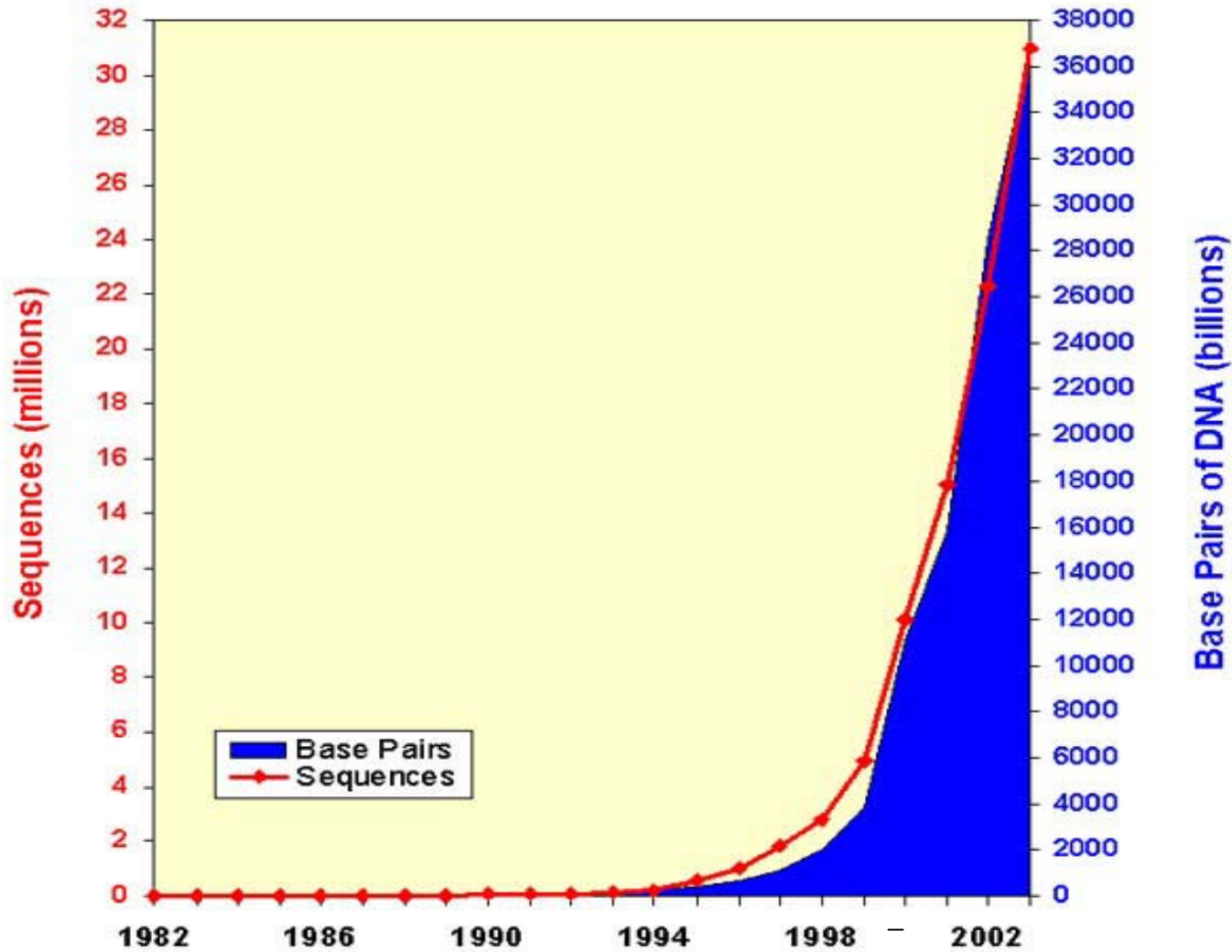
Organism	Year	Genome size, Mb	Number of predicted genes
Yeast	1996	13	6,000
Nematode	1998	97	19,000
<i>Drosophila</i>	2000	180	14,000
<i>Arabidopsis</i>	2000	150	26,000
Human	2001	3,000	35,000
Rice	2002	430	35,000-50,000



Subprogram 2: Comparative genomics for gene discovery

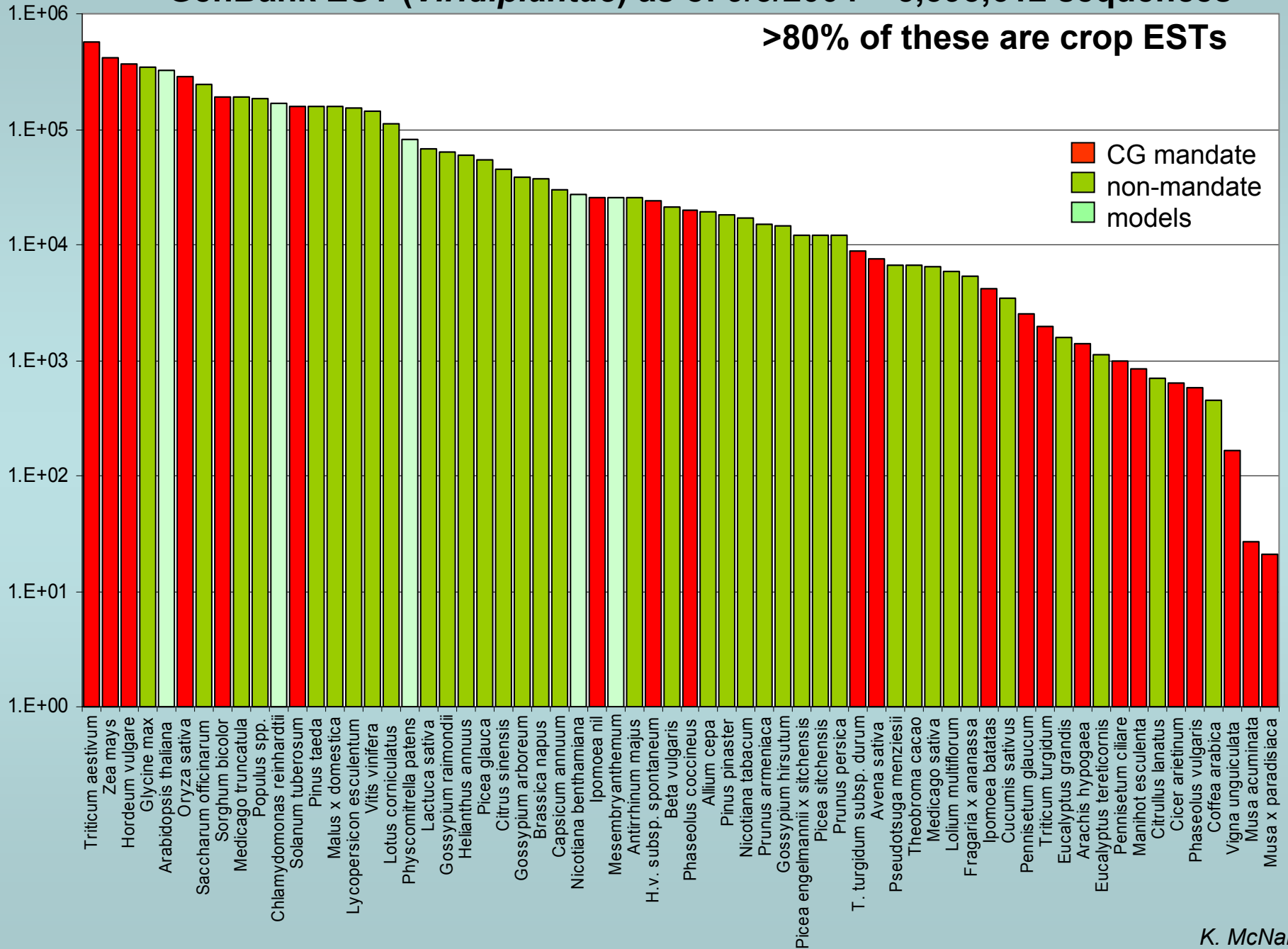
- **Acquisition of genetic resources and capacities**
- **Development of comparative maps *within and between* crop species (**increasingly based on gene contents**)**
- **Assignment of genes and pathways to putative phenotypes**
 - **Gathering functional evidence**
 - **Reverse genetics**
 - **Genetic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

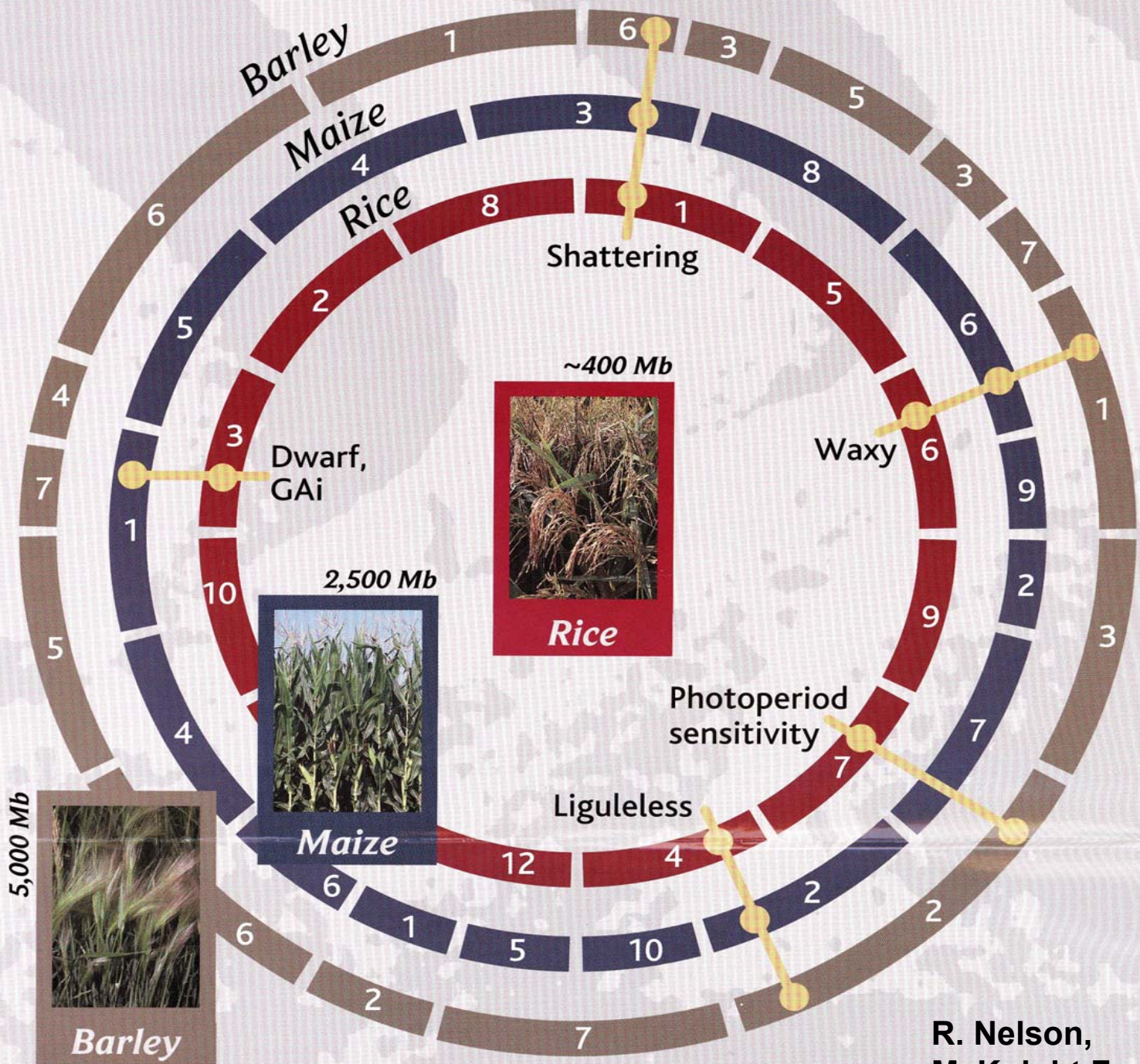
Growth of GenBank



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

>80% of these are crop ESTs





R. Nelson,
McKnight Foundation

Beyond gene order

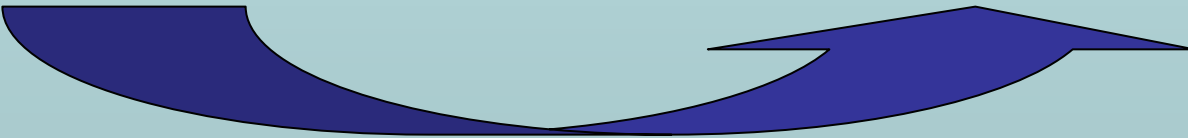
- Gene contents—actual sequence
 - orthologs and paralogs
- Evolutionary conserved functions
- Narrowing the monocot-dicot divide

Connecting genomic variation with phenotype and performance

Genetic variation
SNP and indels

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

Phenotypes
Performance



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

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 ($P < 0.05$)

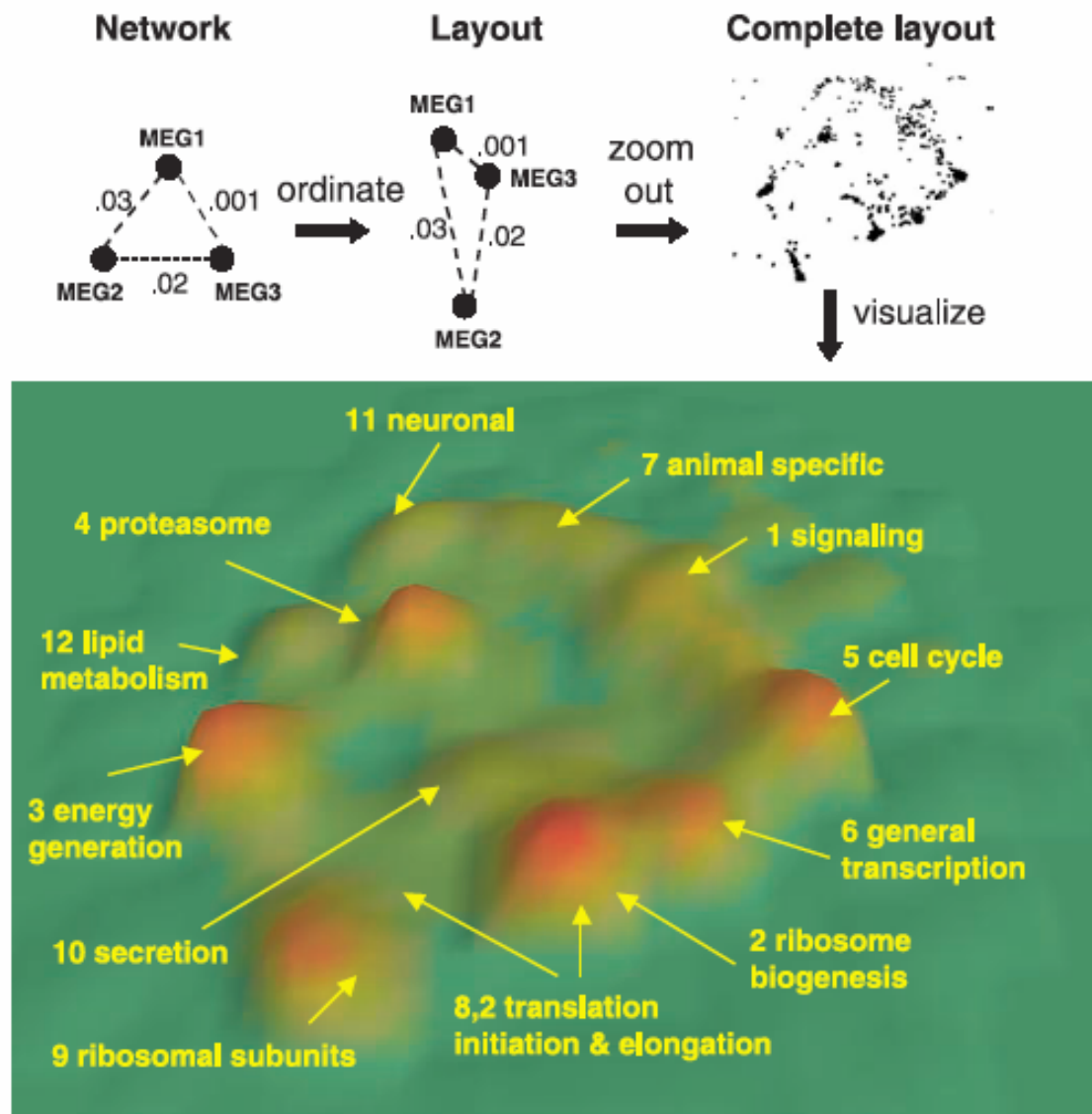


Terrain map: discover new genes falling in same components



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>.



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

External Environment

Complete rice and Arabidopsis genomes and growing sequence resources

Comparative genomics

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

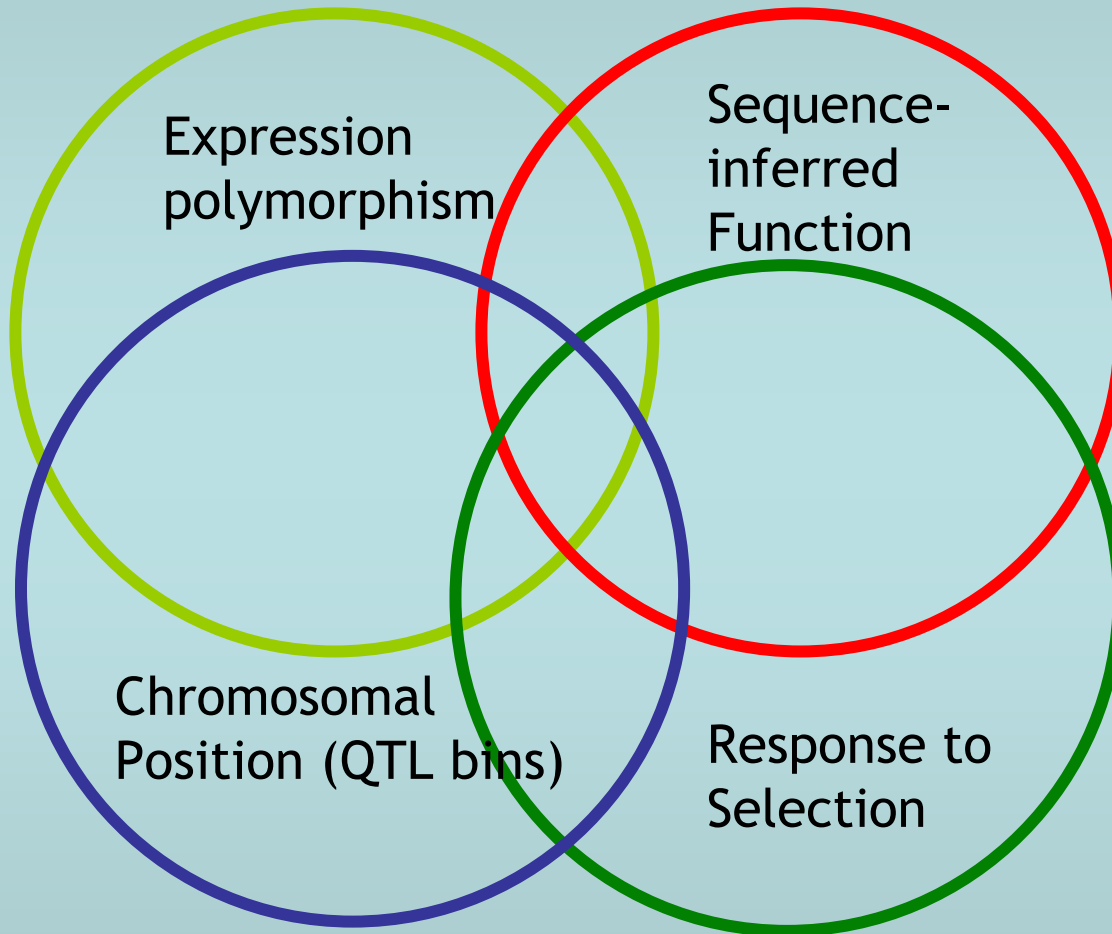


Application in crops

Three-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.
- SNP analysis in target genes (orthologs and paralogs) in multiple species.

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

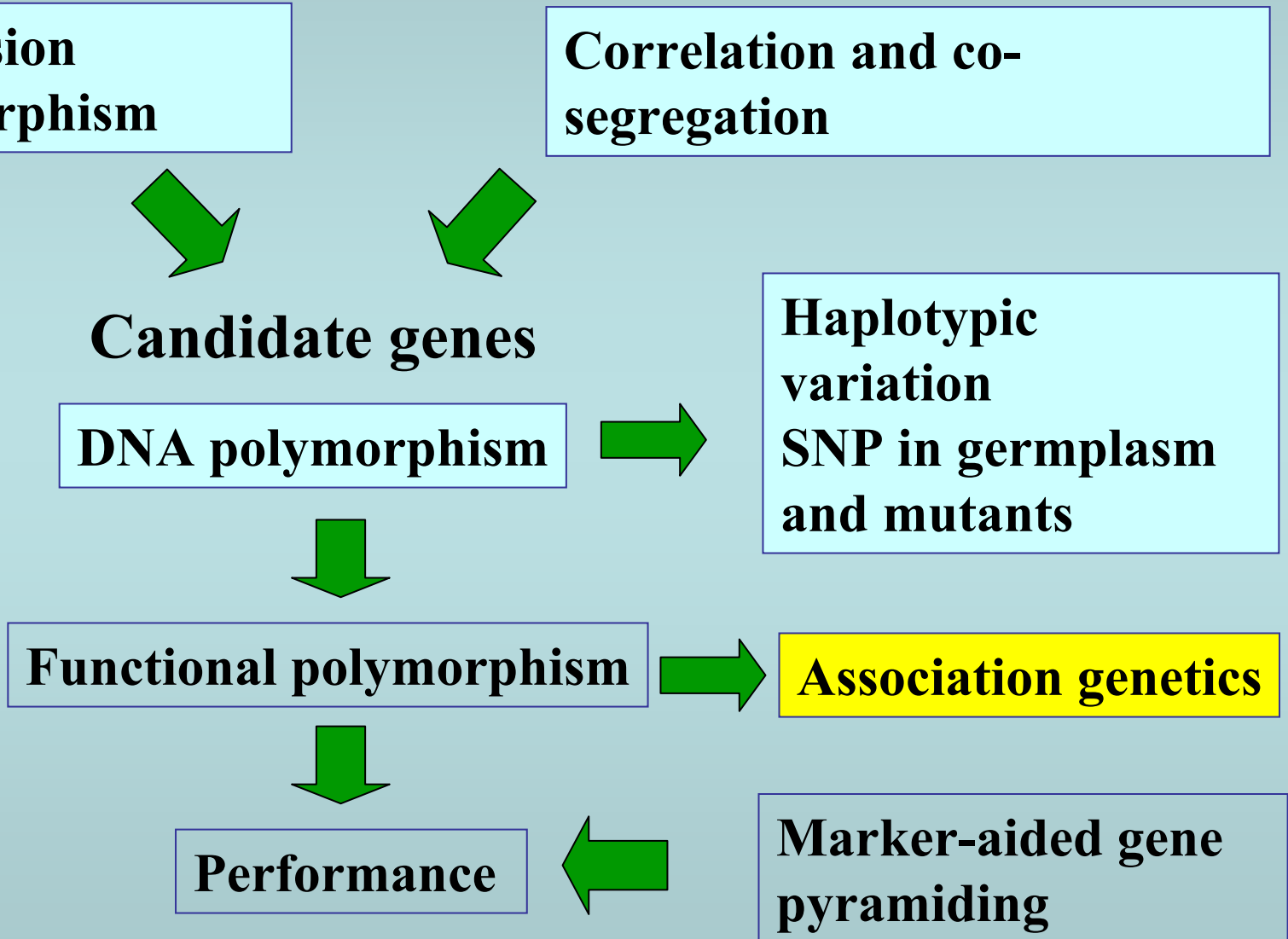
**Haplotypic
variation
SNP in germplasm
and mutants**

Functional polymorphism

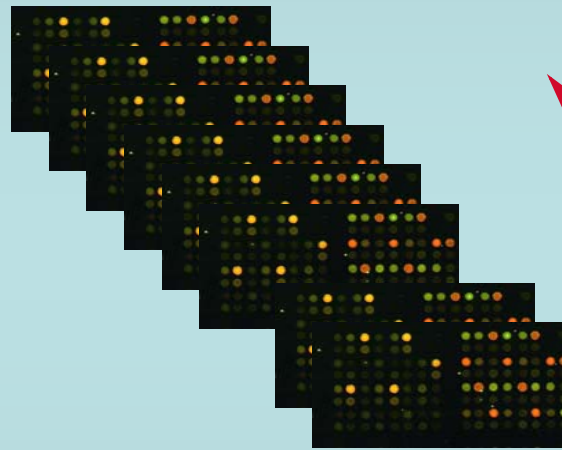
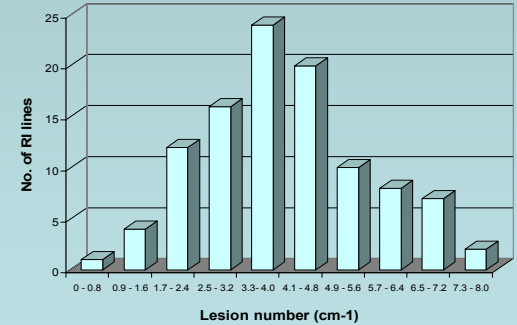
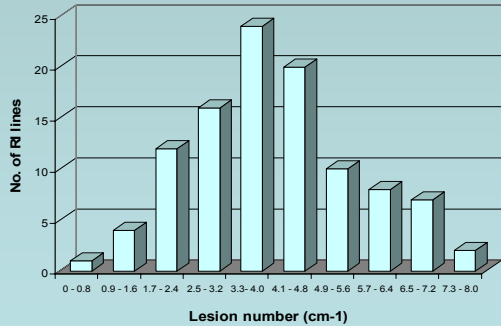
Association genetics

Performance

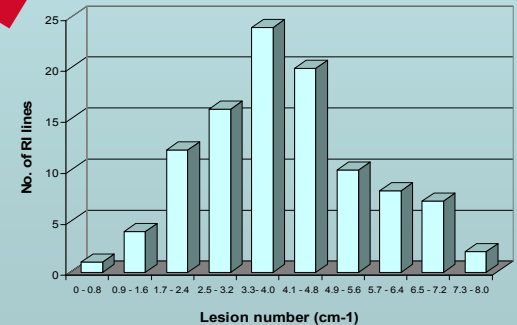
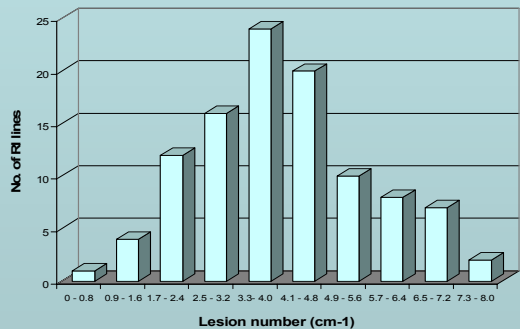
**Marker-aided gene
pyramiding**



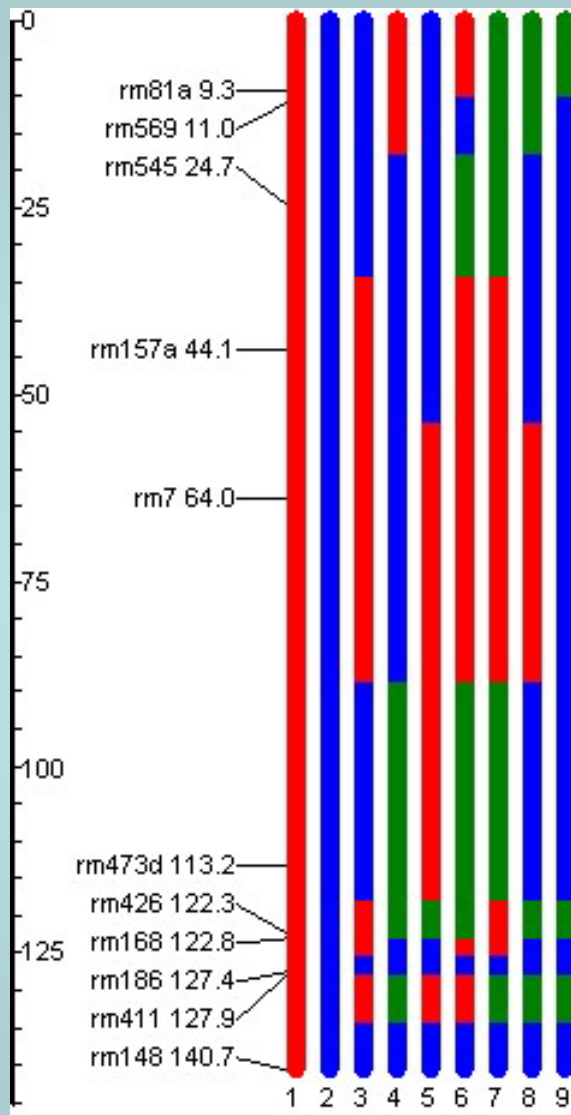
Segregation genomics: make use of the power of genetic segregation to determine causative effects (not correlation)



“Mendelize” expression levels using populations with multiple QTL and phenotypes segregating



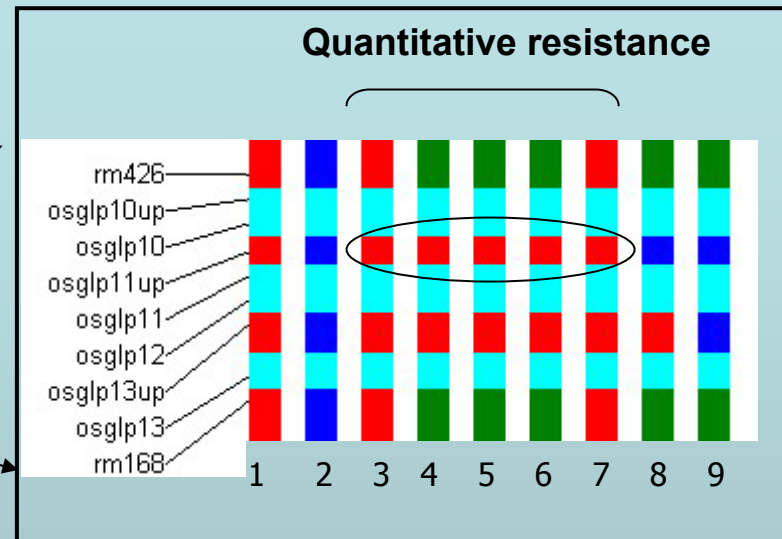
Recombinant chromosomes



Chromosome 3

Associate phenotypes with molecular variation in upstream region of members of the oxalate oxidase gene family

Cis element	OsGLP11	OsGLP13
ASF1MOTIFCAMV	1	0
WBOXATNPR1	1	1
WBOXHVIS01	3	0



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

TILLING to detect allelic variants

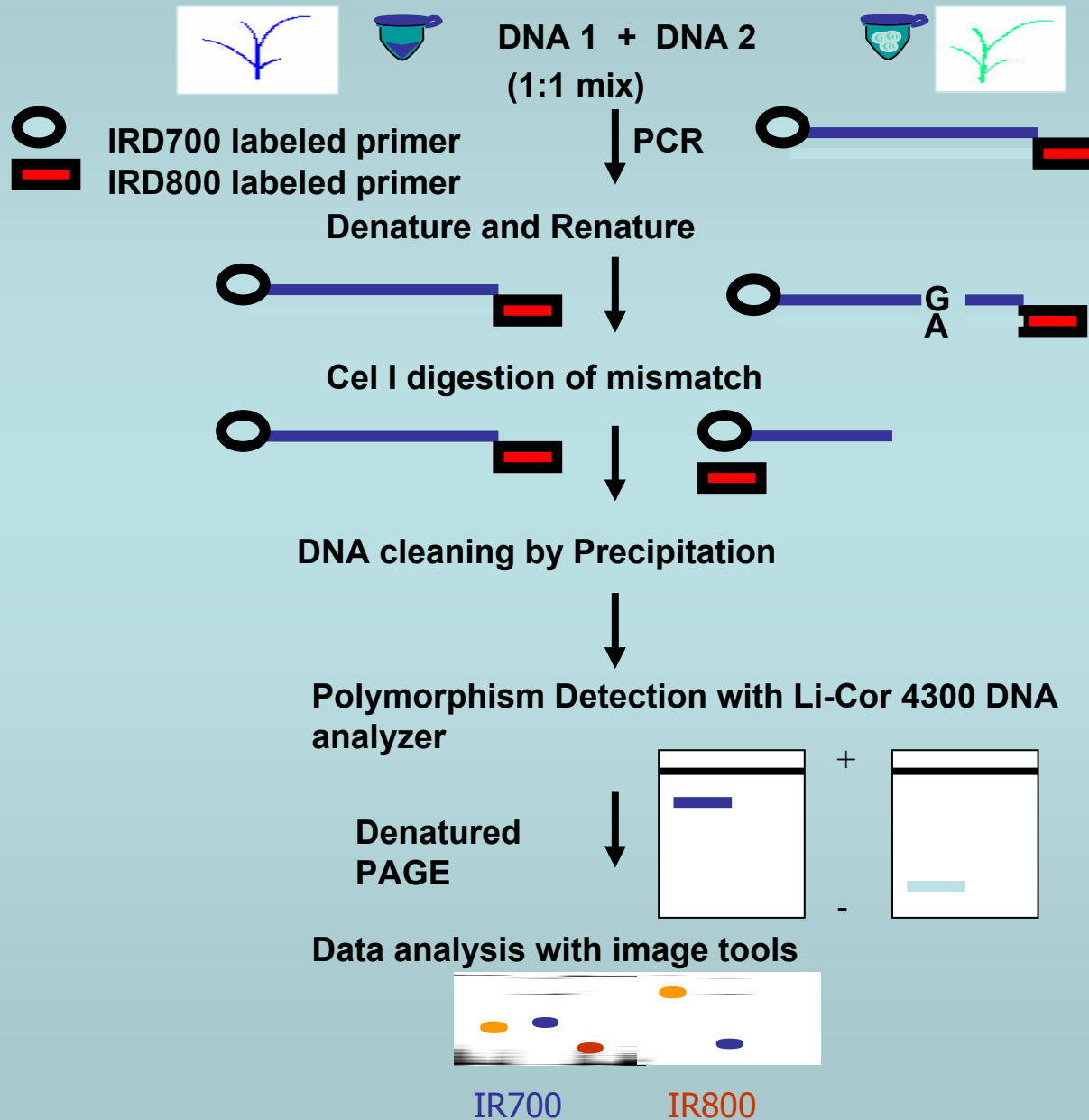
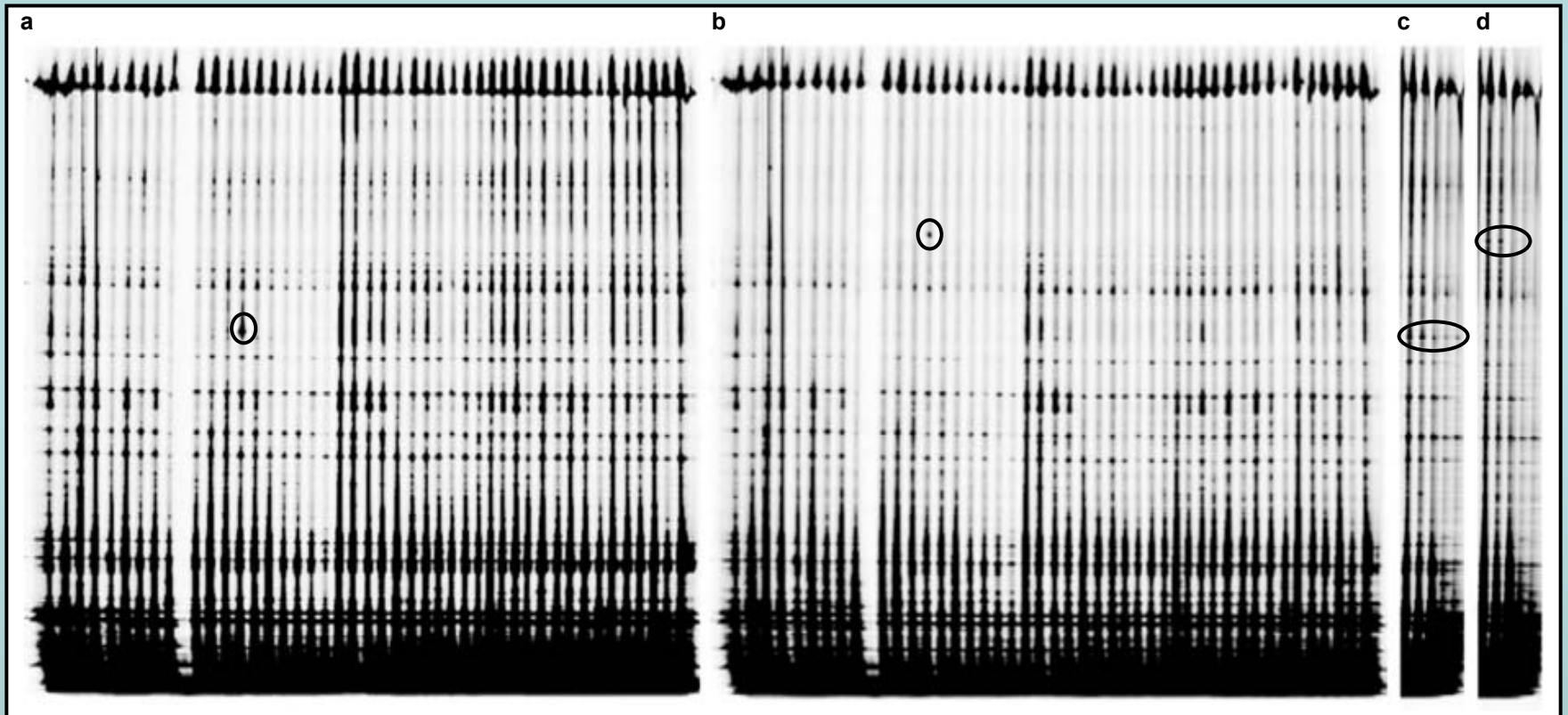
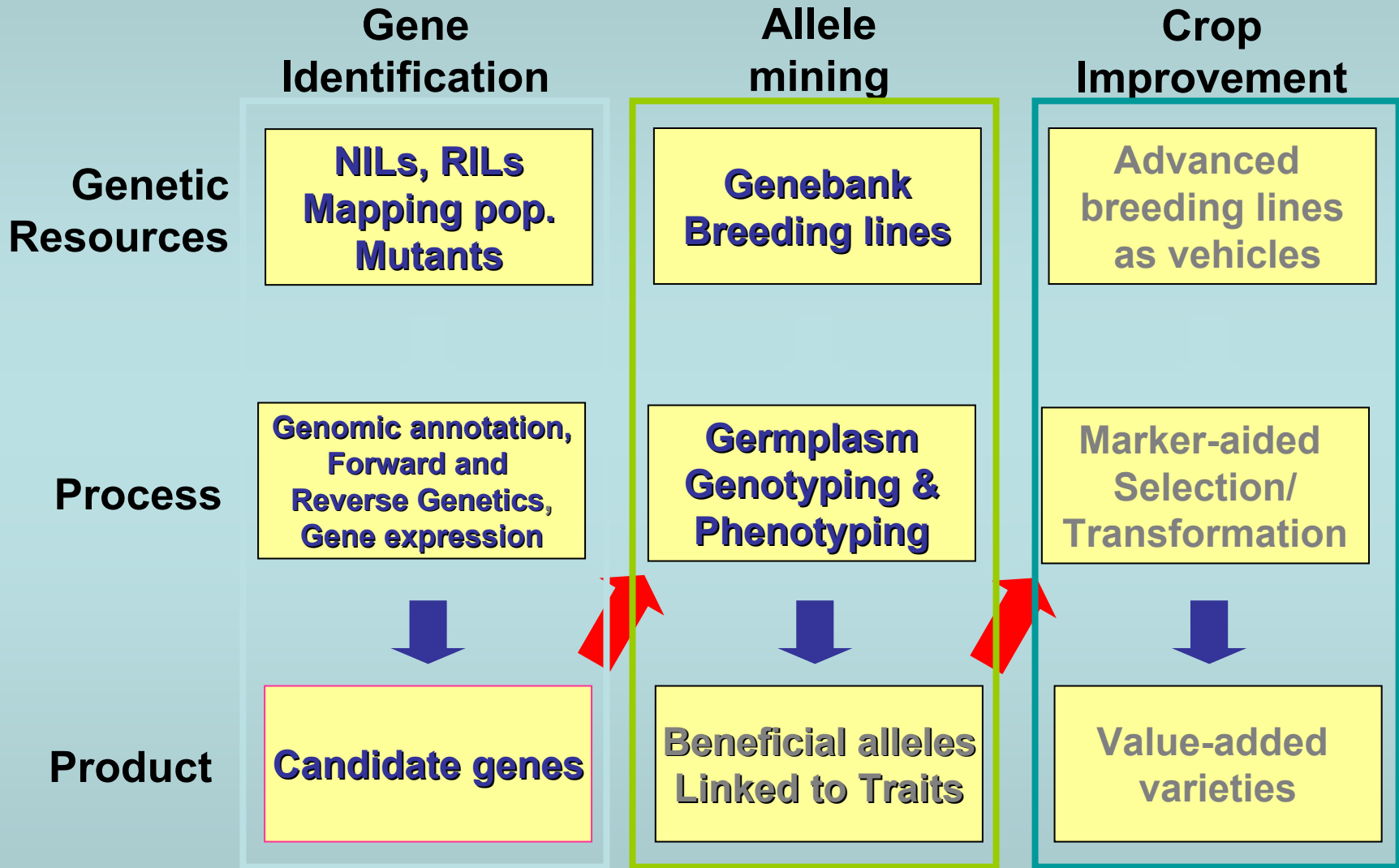


Fig. 1. EcoTILLING of TPS on *Oryza sativa*. Trehalose 6-phosphatase (TPS) primers tested on panel of 46 *Oryza sativa* germplasm accessions contrasted to IR64. Circles indicate a clear SNP detected in the IRD700 (a) and IRD800 (b) channels for Kun Min Tsieh Hunan. This germplasm was pooled with IR64 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).



Gene Discovery to Trait Synthesis



Subprogram 2: Comparative genomics for gene discovery

- Scope and contents of SP2
 - Cross-cutting resources and research platform
 - Connecting genomic variation with phenotype and performance
- The changing external environments are in our favor
- Technical approaches---current and robust
- Peer-production for the commons