Gramene's Genetic Diversity Module:
a case study in flowering time variation across three plant species

Charles Chen[1*], Genevieve DeClerck[1], Terry Casstevens[2], Peter Bradbury[3], Ken Youens-Clark[4], Immanuel Yap[1], Chih-Wei Tung[1], Doreen Ware[3,4], Pankaj Jaiswal[5], Susan McCouch[1] and Edward Buckler[1,2,3]

[1] Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York 14853, USA [1*]Email: cc859@cornell.edu
[2] Institute For Genomic Diversity, Cornell University, Ithaca, New York 14853 USA
[3] United States Department of Agriculture- Agriculture Research Service (USDA-ARS)
[4] Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 11724, USA
[5] Dept. of Botany and Plant Pathology, 3082 Cordley Hall, Oregon State University, Corvallis, Oregon

Abstract Flowering time is an important adaptive character which greatly impacts yield and grain quality in crop plants. Although the genetic basis of flowering time variation has been most intensively studied in the temperate, model dicot system, Arabidopsis, the availability of full genome sequences and genome-wide association studies in both maize and rice provide new insights into how genes and gene family members have been recruited to perform similar functions in the flowering time pathways of these tropical grass species. Taking advantage of Gramene's Genetic Diversity module with its expert-curated, web-accessible data resources and analysis packages, we demonstrate the power of this database environment for elucidating both conserved and divergent genetic mechanisms underlying complex trait variation in plant species.

All data and statistics programs will be available via www.gramene.org
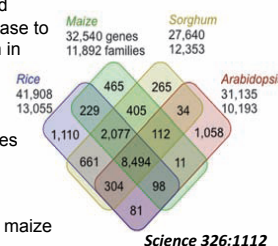
Objectives

Other Gramene posters: P822, P835, P843, P859

Gramene Diversity Module will exploit and integrate the complementary strength brought by our database to study genetic mechanism underlying complex trait variation in crop species

Science 326:1112

Orthology predictions increasingly useful- *Arabidopsis* shares 90% of gene families with grasses

Large, mutagenized populations available in *Arabidopsis*

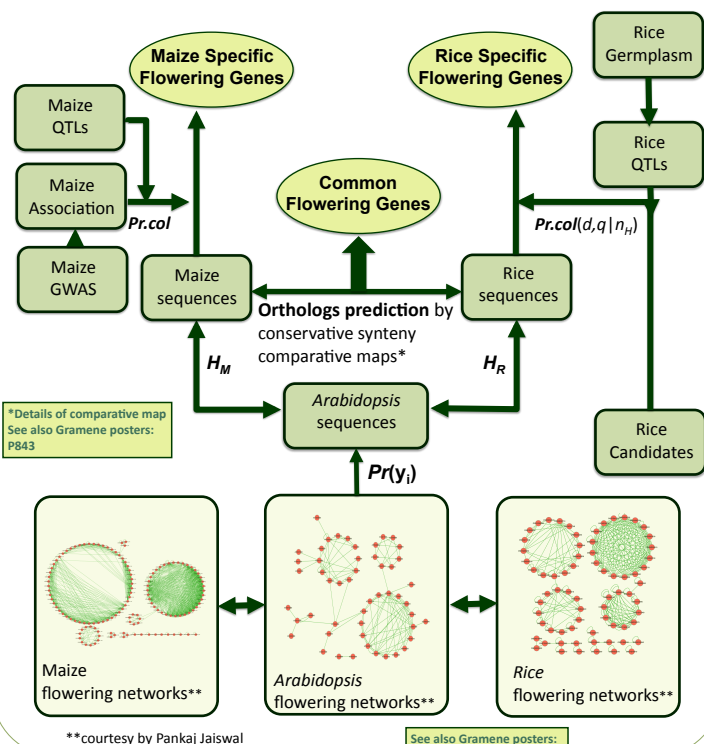Broad scale of phenotyping and association/QTL studies in maize

Low copy genomic sequences and an abundance of candidate genes available in rice

See also Gramene posters: P835

Model for Analyses

*Details of comparative map
See also Gramene posters: P843

**courtesy by Pankaj Jaiswal

See also Gramene posters: P822

Probabilistic Model of Gene Discovery

$L(x) \approx Pr(x, \pi, d, G, q \mid H,$ species, germplasms$)$
$Pr(y_1, y_2, ..y_i) = \prod p(y_i \mid pa(y_i))$
posterior $Pr\_col(d,q \mid n_R) = \Sigma\ Pr(\lambda_R) \times Pr.col(d,q)$
posterior $Pr\_col(q,G \mid n_M) = \Sigma\ Pr(\lambda_M) \times Pr.col(q,G)$

x : maize/rice genes
π : alignments
H: degree of homology given by *Arabidopsis* genes (given by $min(E_{y1}, E_{y2}, ..E_{yi})$)
$pa(y_i)$: local and global Markov property of gene y_i involved in flowering pathways
q: QTLs in maize/rice
G: genes mapped in GWAS study in maize
d: candidate genes discovered in Rice
λ_R: number of QTLs of a given map region in Rice
λ_M: number if QTLs of a given map region in Maize
n_R: gene richness with a given map region in Rice
n_M: gene richness with a given map region in Maize

Perspectives

- Gramene Diversity Module curates, stores and displays data used in the analyses and will provide statistics package as well as web-based query tools.

- *L(x)*, likelihood of is.gene, is a function of genome properties, including relationship bt/n phenotype and genome position encoded via QTL, similarity/homology among candidate genes associated with the same trait across different species, gene expression profiles, inferred orthology and evolutionary/genetic divergence within and among species lineages

- the probabilistic model allows "hot spots" to be discovered at a genomic scale, by exploitingintegrated genotype-phenotype information

SPECIES