Linking genome wide association studies (GWAS) and genomic selection (GS) to better utilize natural variation in rice

Susan McCouch
Plant Breeding & Genetics
School of Integrative Plant Science
Cornell University

Rice: a mosaic of diversity

Most is rice produced on farms < 1 ha in size.

92% consumed w/in country where produced.

Consumer quality preferences diverse & inelastic.

Building a road map of natural variation in rice

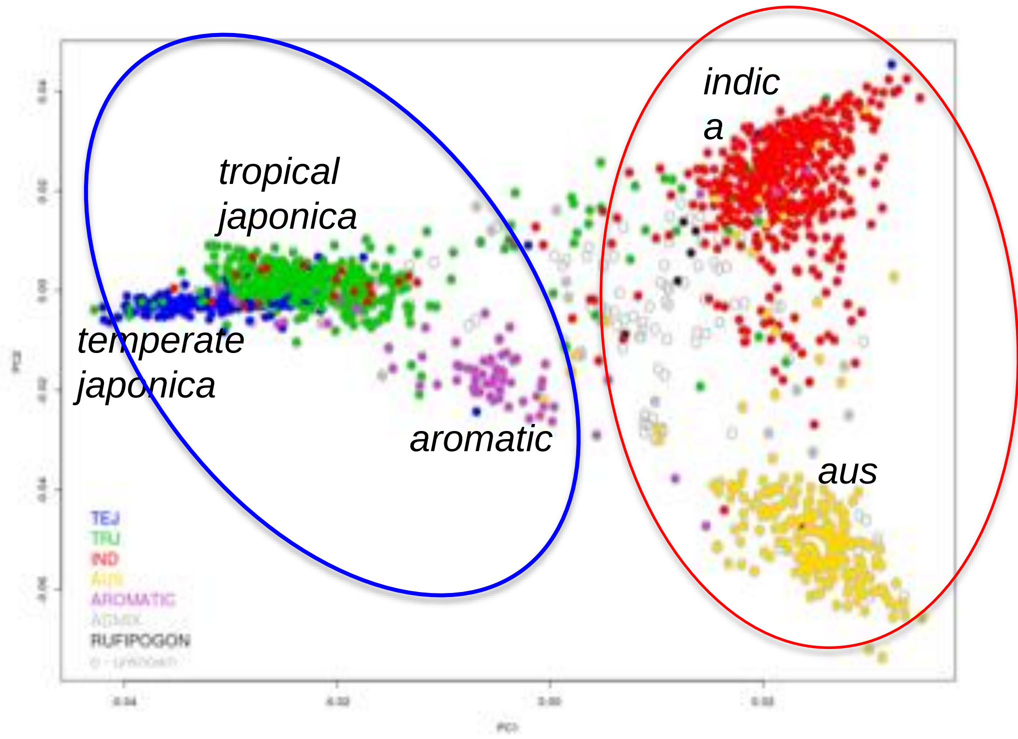
- How much genetic variation is there in *O. sativa,* how is it partitioned and where is it found?
- How can we use that diversity to identify genotype-phenotype associations for traits of interest?
- How can we efficiently select on favorable alleles at thousands of loci across the genome to increase the overall rate of genetic gain in rice improvement?

How is diversity organized in *O. sativa*?

Two major varietal groups

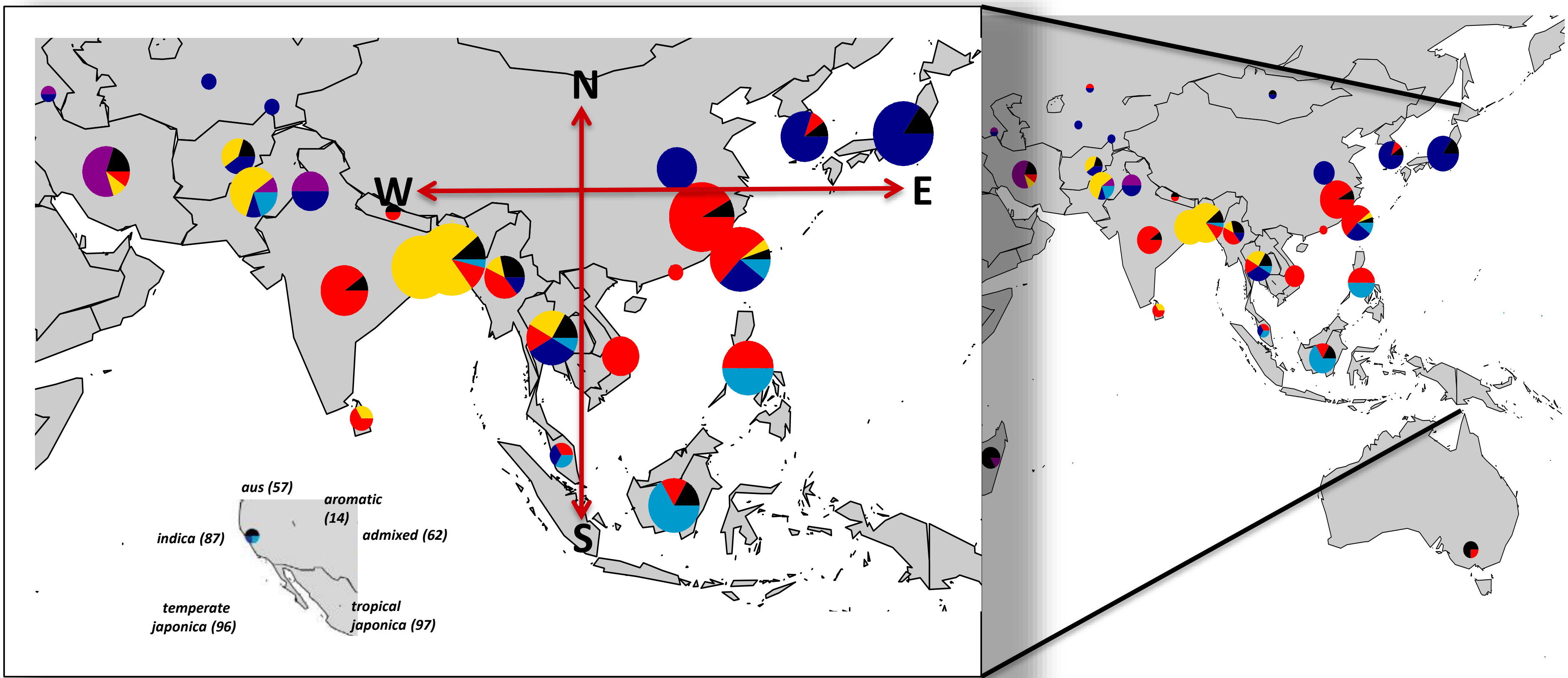
Five genetic sub-populations

- ***JAPONICA***
  - *tropical japonica*
  - *temperate japonica*
  - *aromatic (Group V)*
- ***INDICA***
  - *indica*
  - *aus*



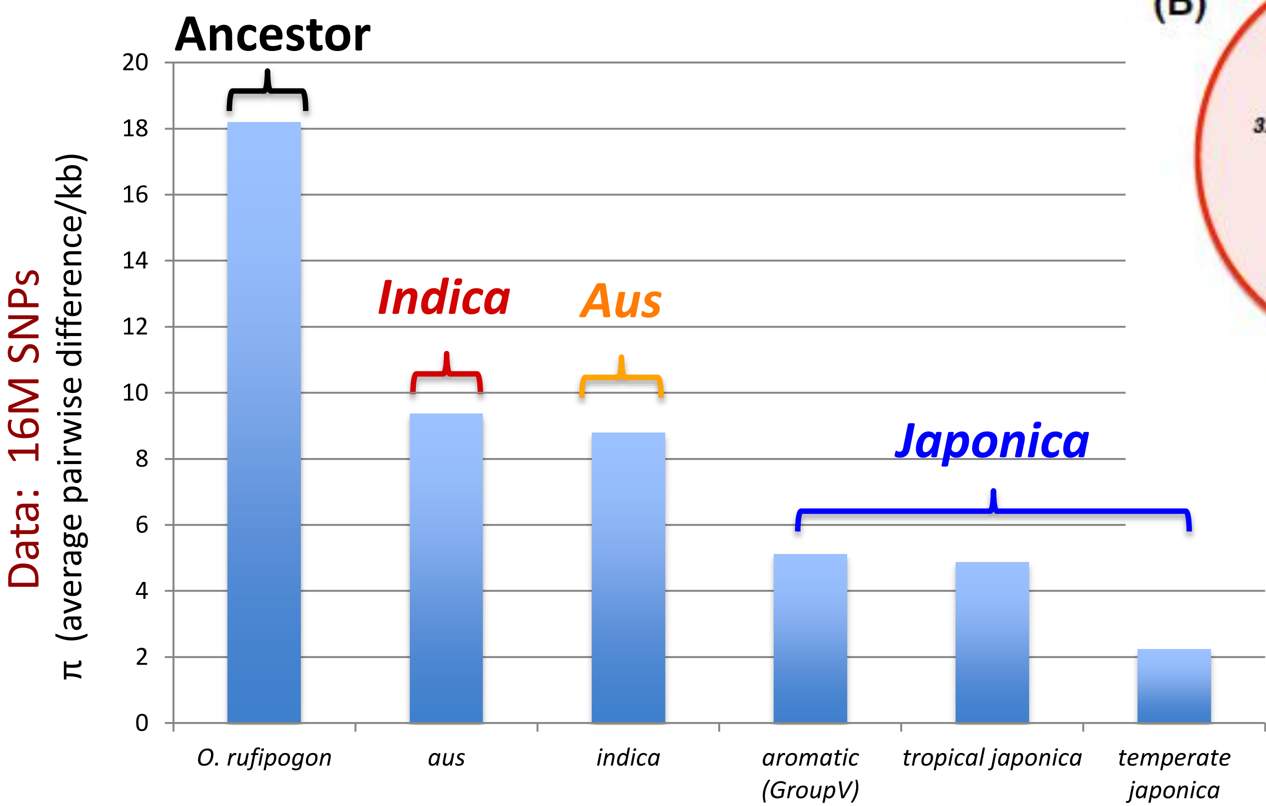
McCouch et al. (2016) Open access resources for genome-wide association mapping in rice. Nature Comm. 7:10532

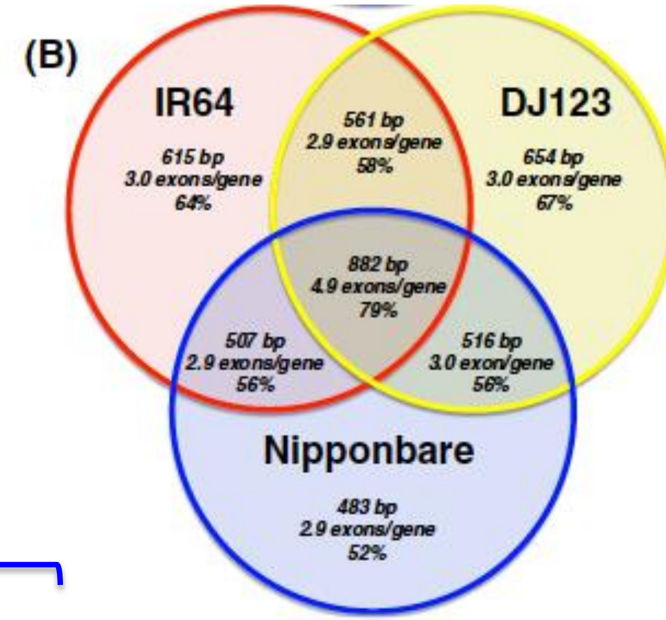
Geographic distribution of rice subpops



Zhao et al. (2011) Nature Communications 2:476

Variation within & between subpopulations





Gene Space

Median CDS length

Av No. Exons

% InterPro/homology

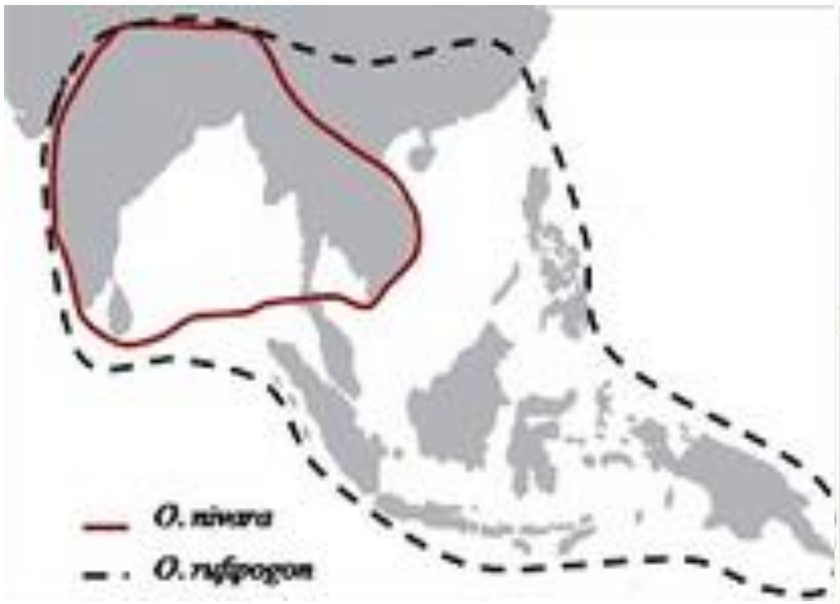
LD = 5-50 kb **LD = 50-120 kb** **LD = 100-500 kb**

Rice SNP Consortium; analysis by M. Wright

Schatz et al. (2014) Genome Biology

Oryza rufipogon species complex

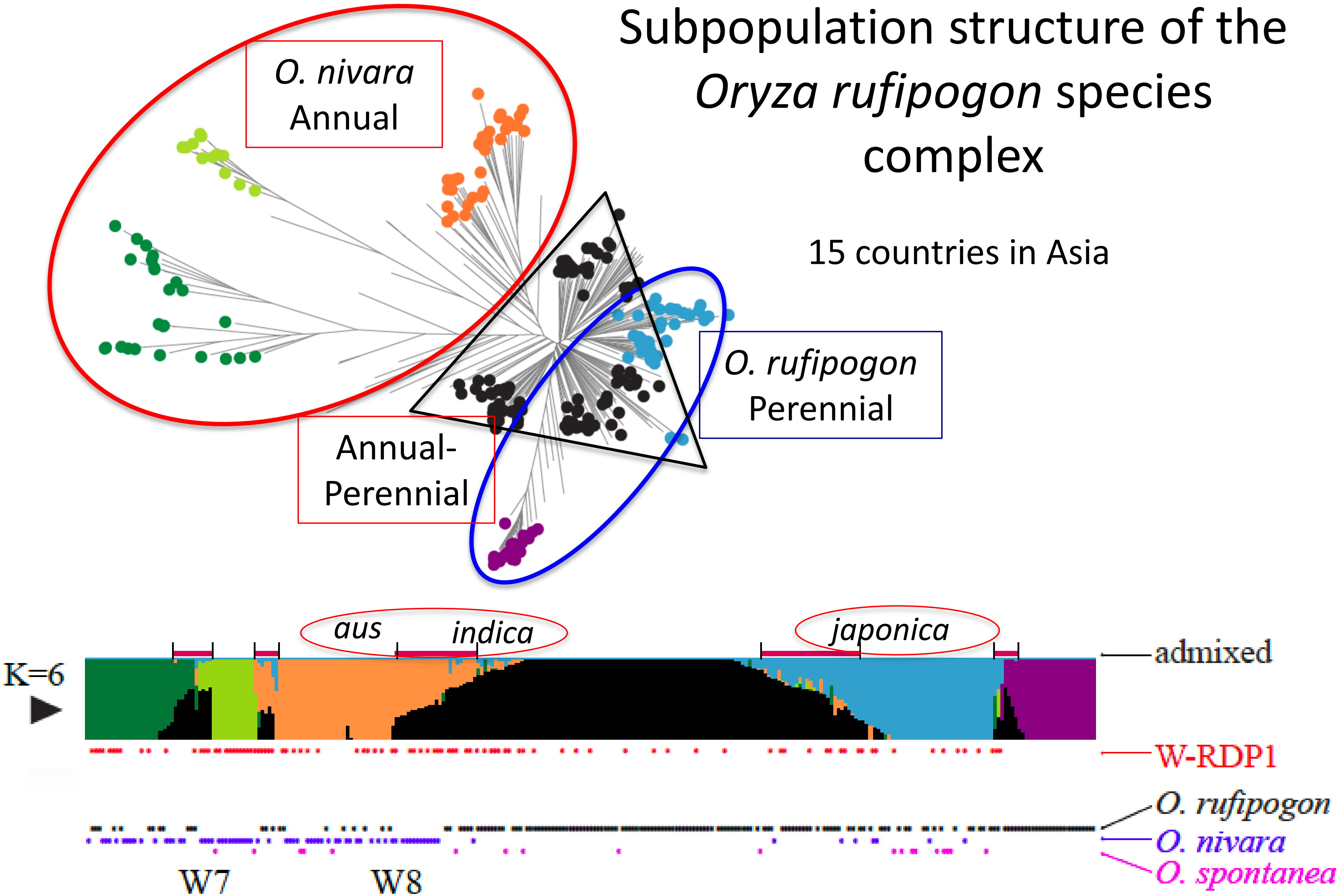
Inbreeding annuals and outcrossing perennials



Features	***O. rufipogon***	***O. nivara***
Habitat	**Deep water**	**Seasonally dry**
Life history	**Perennial**	**Annual**
Photoperiod	**Sensitive**	**Insensitive**
Breeding system	**Outcrossing**	**Selfing**
Seed dispersal	**Less efficient**	**More efficient**

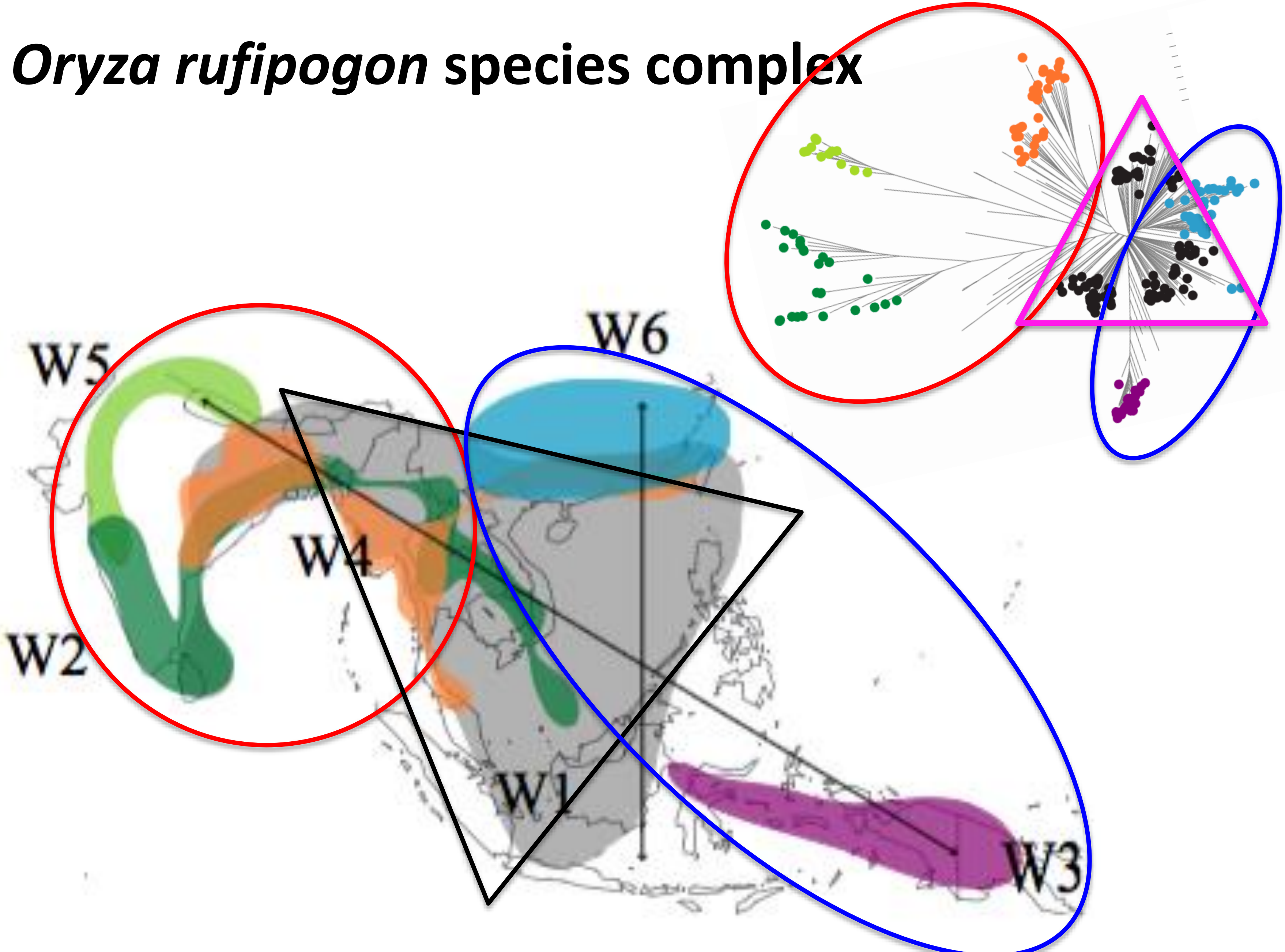
Subpopulation structure of the *Oryza rufipogon* species complex

15 countries in Asia

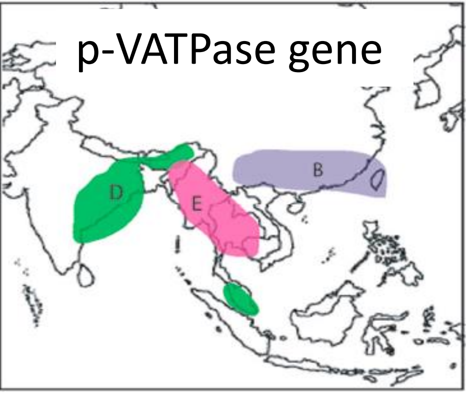


Kim et al. (2016) RICE 9:56

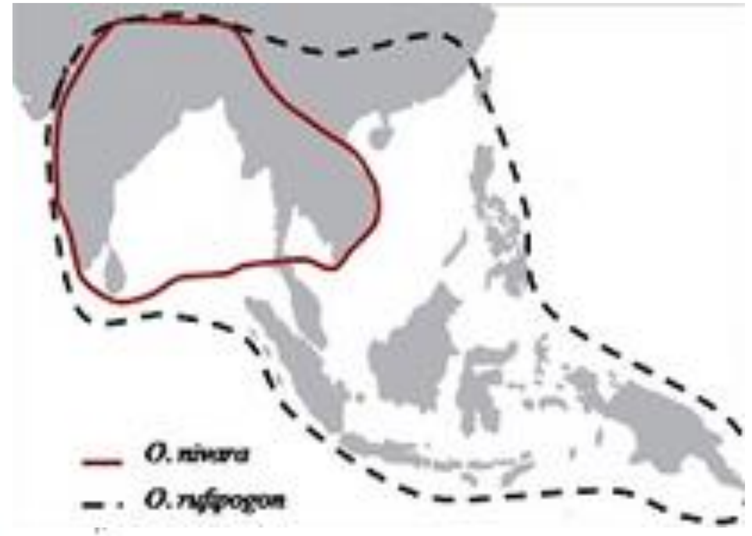
Oryza rufipogon species complex



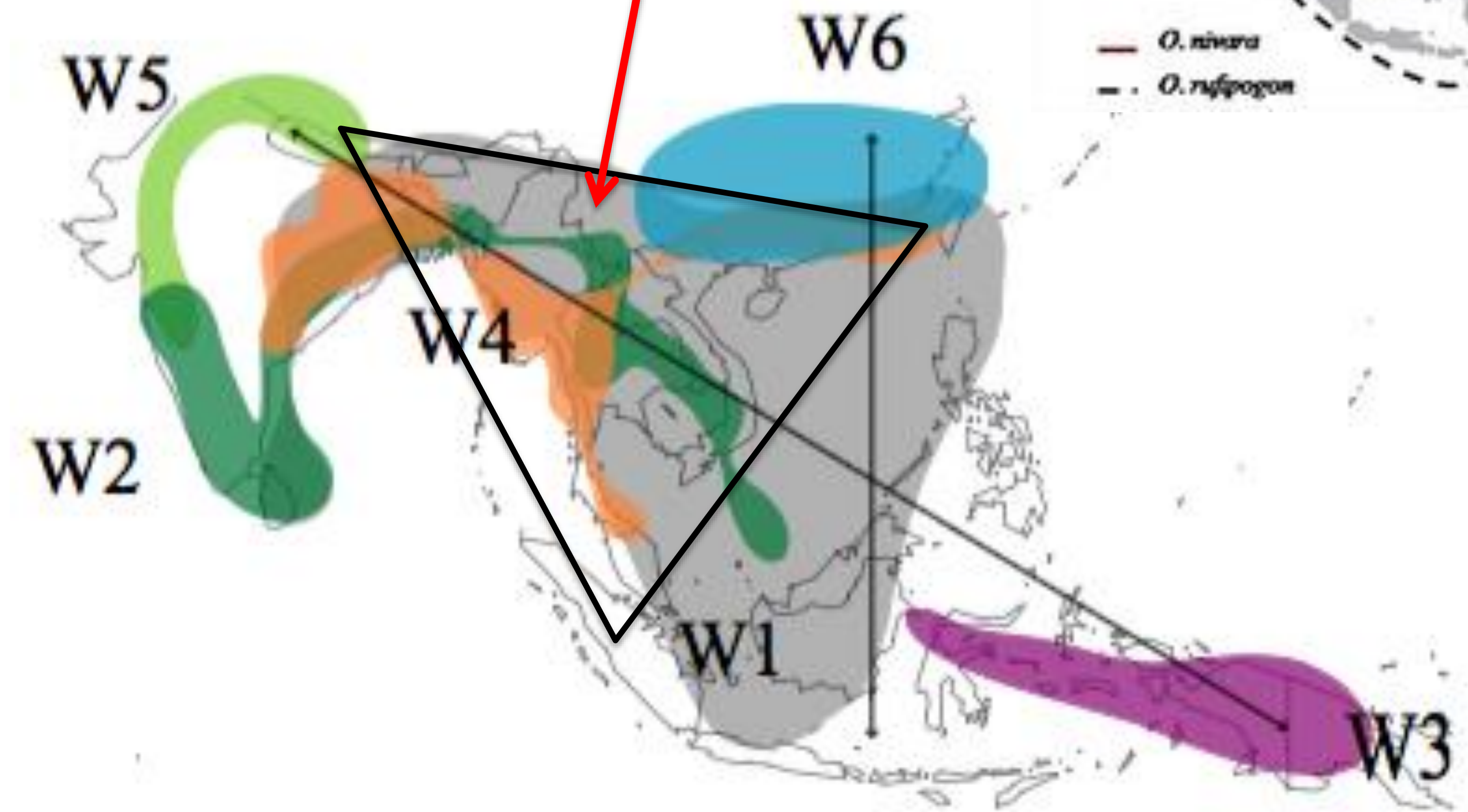
Kim/ Jung et al. (2016) Population Dynamics Among six Major Groups of the *Oryza rufipogon* Species Complex, Wild Relative of Cultivated Asian Rice. RICE 9:56



Londo et al. (2006)
PNAS 103(25): 9578



Greatest diversity



Kim/ Jung et al. (2016) Population Dynamics Among six Major Groups of the *Oryza rufipogon* Species Complex, Wild Relative of Cultivated Asian Rice. RICE 9:56

Arkansas, USA

Los Baños, Philippines

GWAS diversity panel

Rice Diversity Panel 1 "RDP1" (USDA)

~400 *O. sativa*

- 87 *indica*
- 57 *aus*
- 97 *tropical japonica*
- 96 *temp. japonica*
- 14 *aromatic*
- 49 admix
- **100 wilds (*O. rufipogon*)**

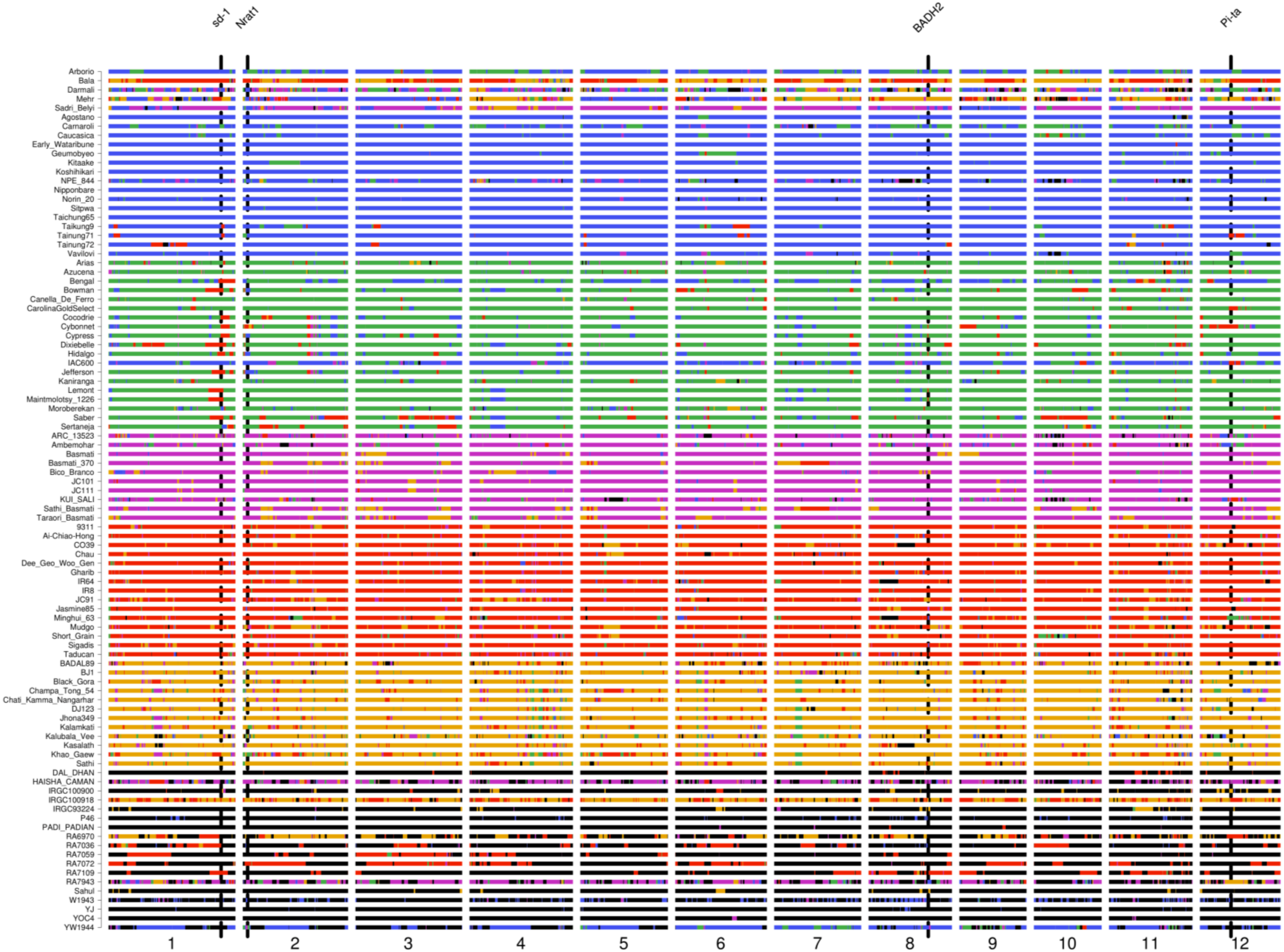
Rice Diversity Panel 2 "RDP2" (IRRI)

~1200 *O. sativa*

- 571 *indica*
- 203 *aus*
- 428 *tropical japonica*
- 152 *temp. japonica*
- 83 *aromatic*
- 7 admix

Total: ~1600 homozygous *O. sativa* accessions genotyped with 700,000 SNPs

SNP catalogue & admixture atlas



admixed

temperate japonica

tropical japonica

aromatic/ GroupV

indica

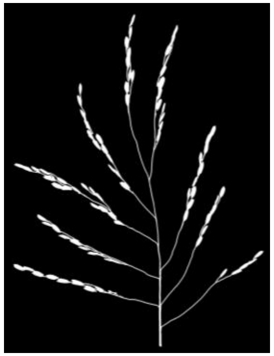
aus

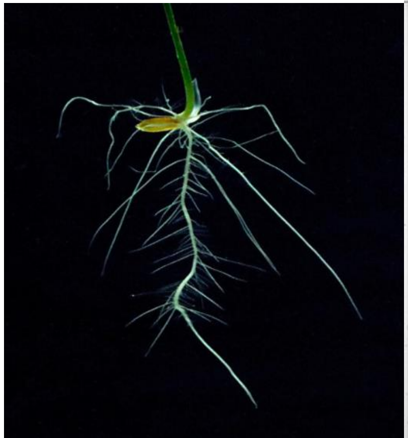
O. rufipogon

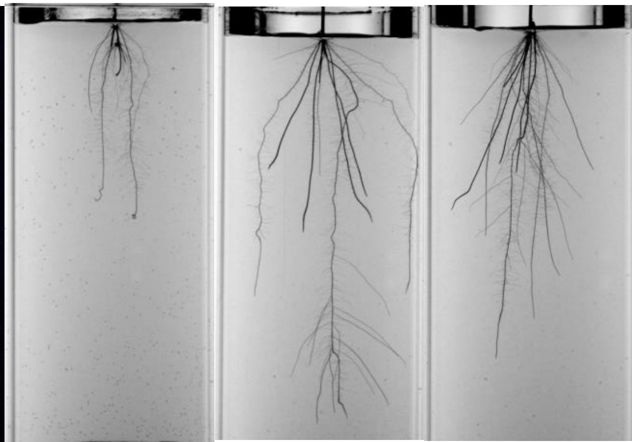
Phenotypic Evaluation

multiple locations, environments, collaborators

- Whole plant phenotypes in the field
- Seed & grain quality characters
- Disease and insect resistance
- Abiotic stress tolerance
- Root and panicle phenotypes
- Ionomics

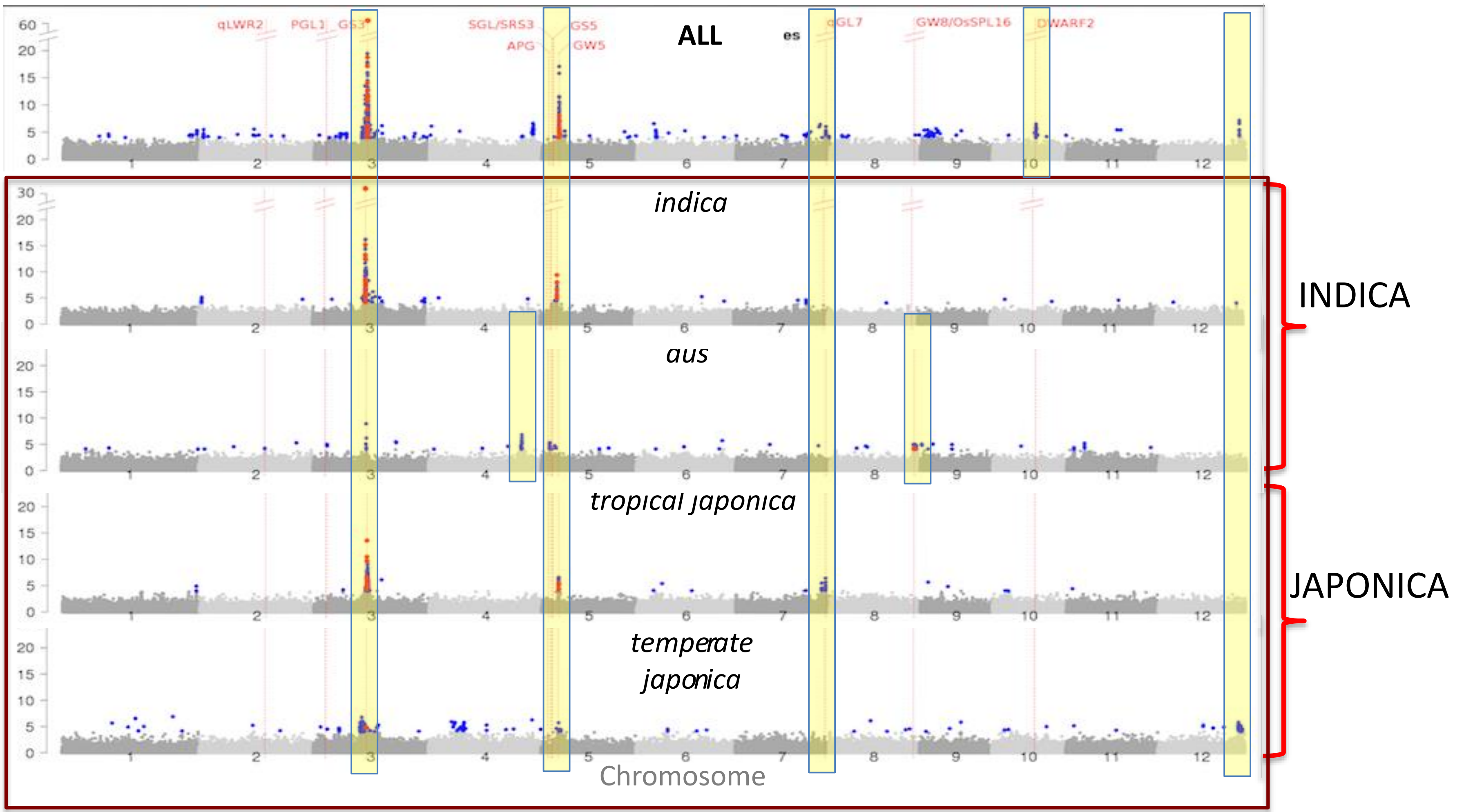






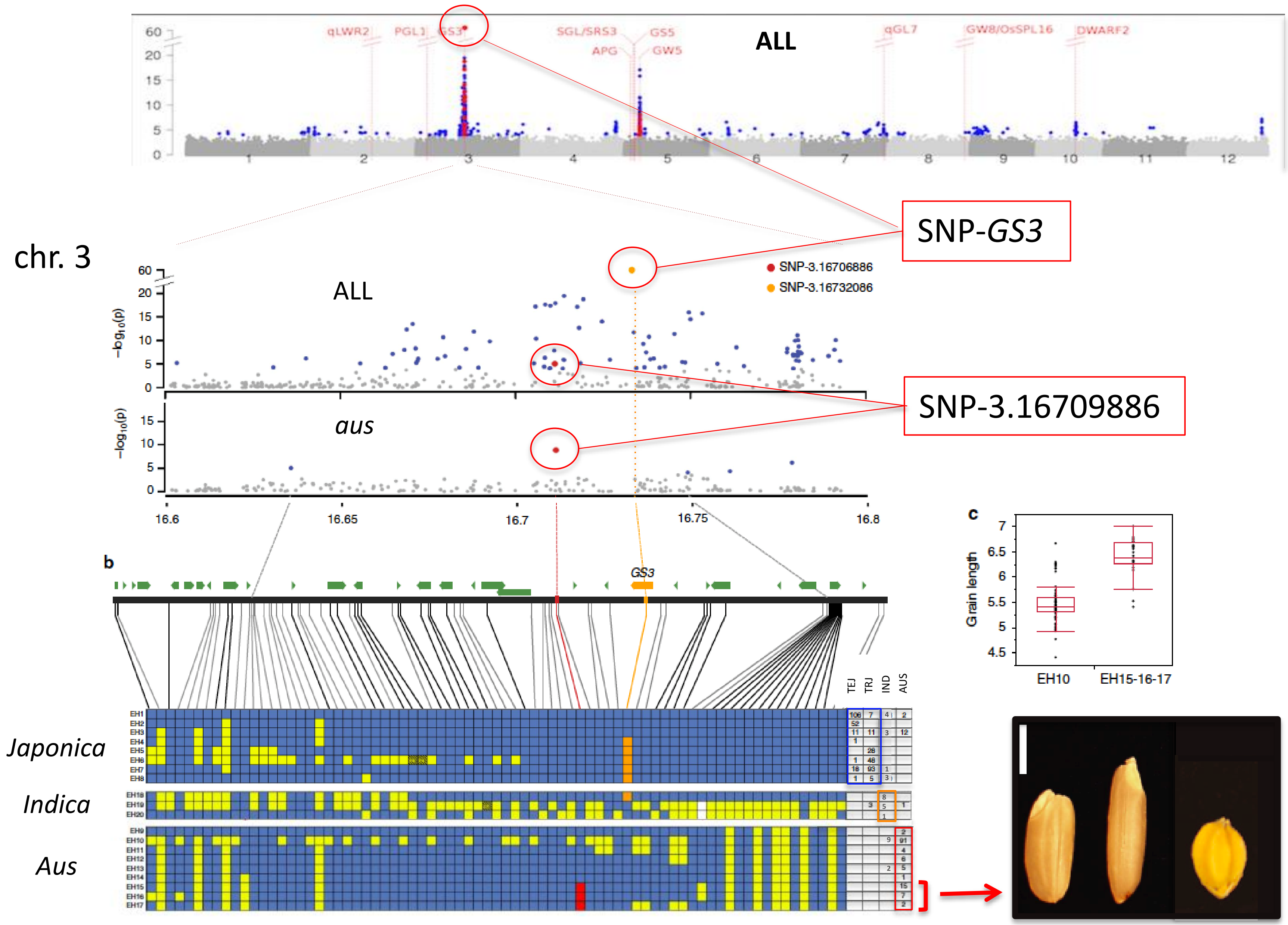
GWAS=> evidence of subpopulation-specific variation

Ex: grain length



McCouch et al., (2016) Open access resources for GWAS in rice. Nat.Comm.

Subpopulation-specific alleles – isolated or shared?



McCouch et al., (2016) Nat.Comm.,

Takano-Kai/Jiang et al., (2013) Breeding Science

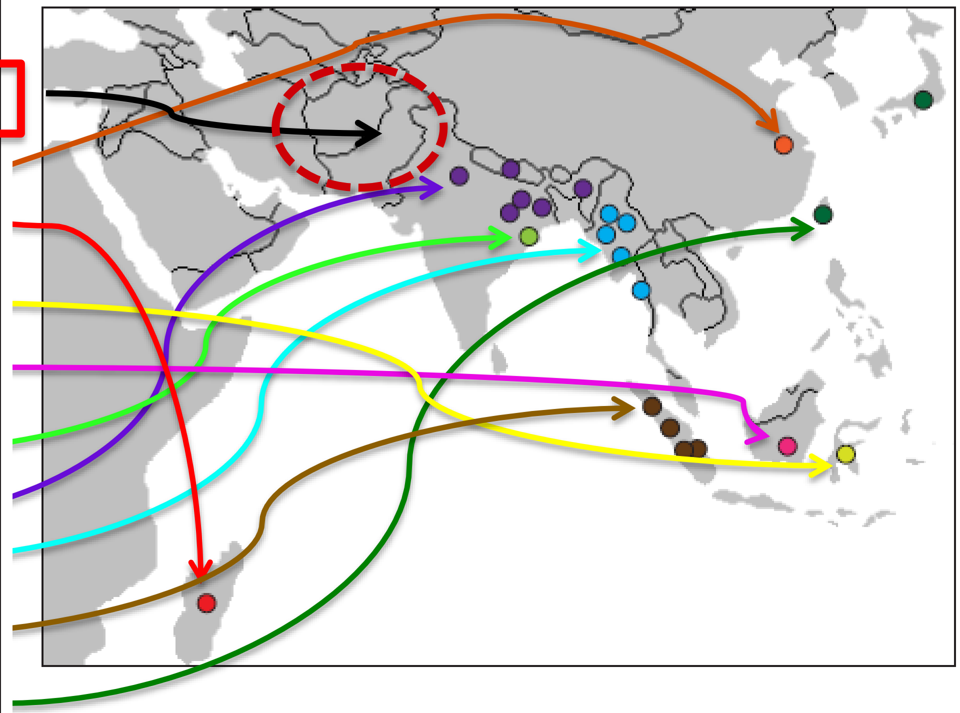
Isolated pockets of diversity persist in the hills & valleys

Many local varieties are maintained within a community, some are shared only through traditional networks, others are traded

Diverse origins of fragrance

Multiple *BADH2* alleles in locally adapted landraces

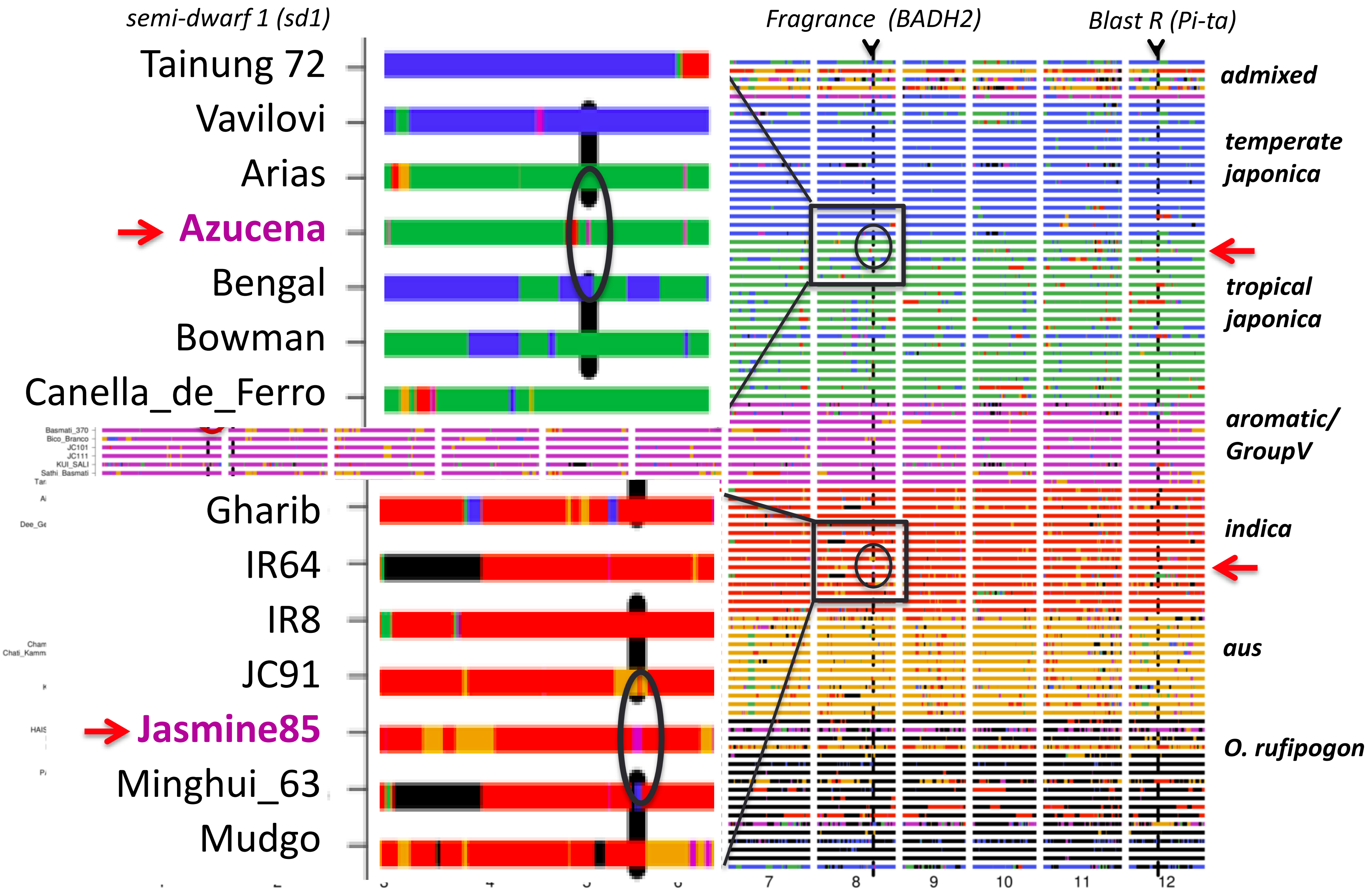
BADH2 allele	Mutation	N	Sub-populat'n	Aroma [2AP]
Wild Type		149	ALL	< 0.05
badh2.1	FNP	93	Group V (aromatic) *indica* tropical *japonica*	0.70 ± 0.41
badh2.2	7bp deletion	1	*tropical japonica*	0.66
badh2.3	2bp deletion	1	*tropical japonica*	0.74
badh2.4	1bp insertion	1	*tropical japonica*	0.59
badh2.5	1bp deletion	1	*indica*	0.70
badh2.6	G→T SNP	1	*aus*	0.41
badh2.7	1bp insertion	6	*aus*	0.43 ± 0.17
badh2.8	3bp insertion	5	Group V (aromatic)	0.36 ± 0.14
badh2.9	G→T SNP	6	*tropical japonica*	0.18± 0.08
badh2.10	C→T SNP	2	*tropical japonica*	0.35



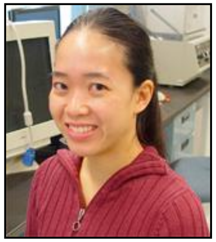
Predominant allele from *basmati* found in ~65% of aromatic accessions today

Kovach et al., (2009) The origin and evolution of fragrance in rice (*Oryza sativa* L.) PNAS 106(34):14444

Population structure & admixture in *O. sativa*



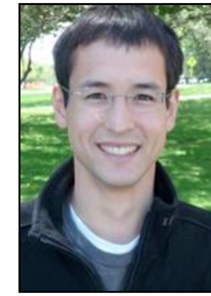
Analysis using NAKARA algorithm by Koni Wright



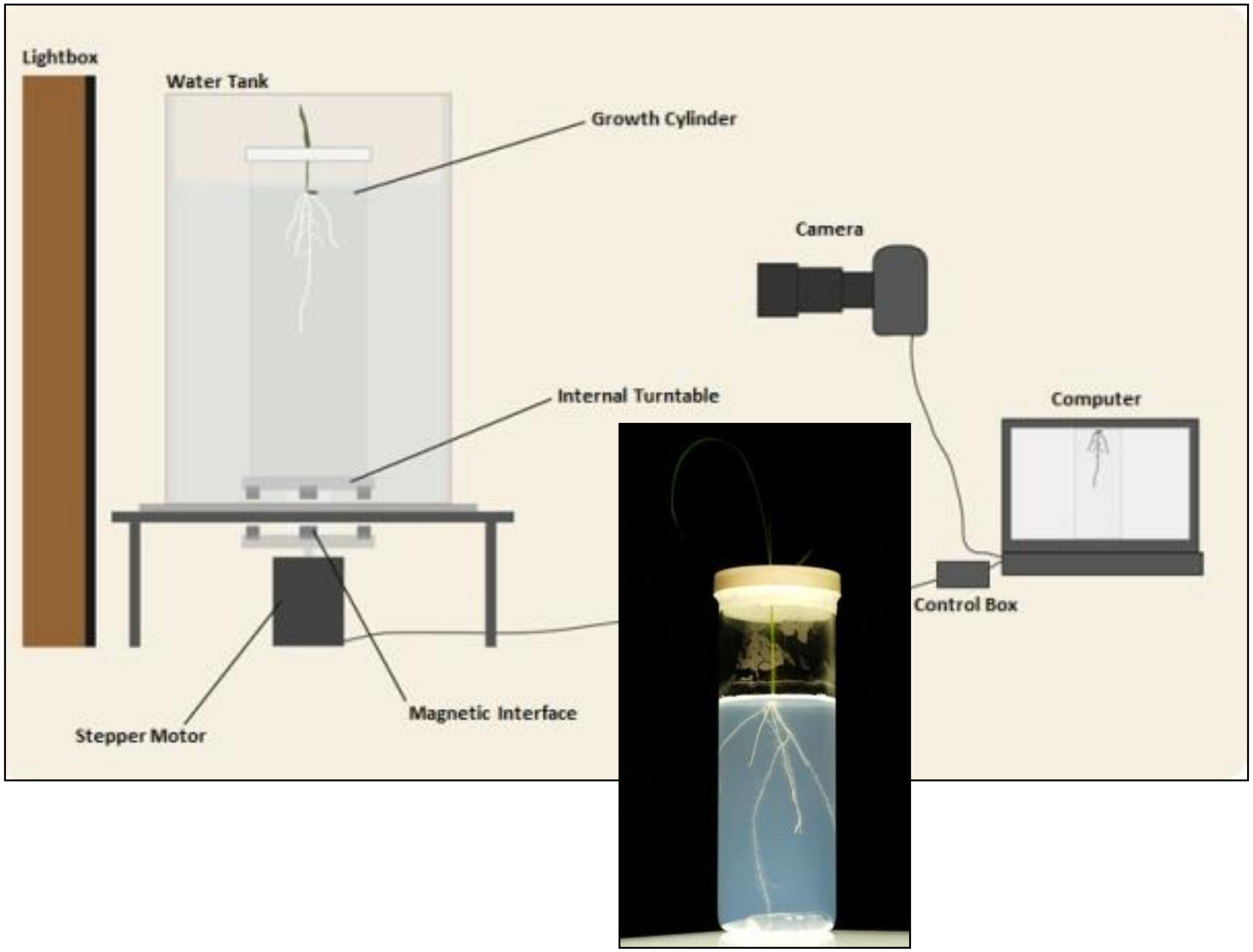
Janelle Jung

3D Root System Architecture

3D Phenotyping Platform



Randy Clark



- **Image & Analysis**

- Sequence of 40 images per plant

- Imaged at Day 3, 6, 9,

- RootReader3D Software

In collaboration with Kochian Lab, USDA/ARS

Clark *et al.*, 2011, Plant Phys.

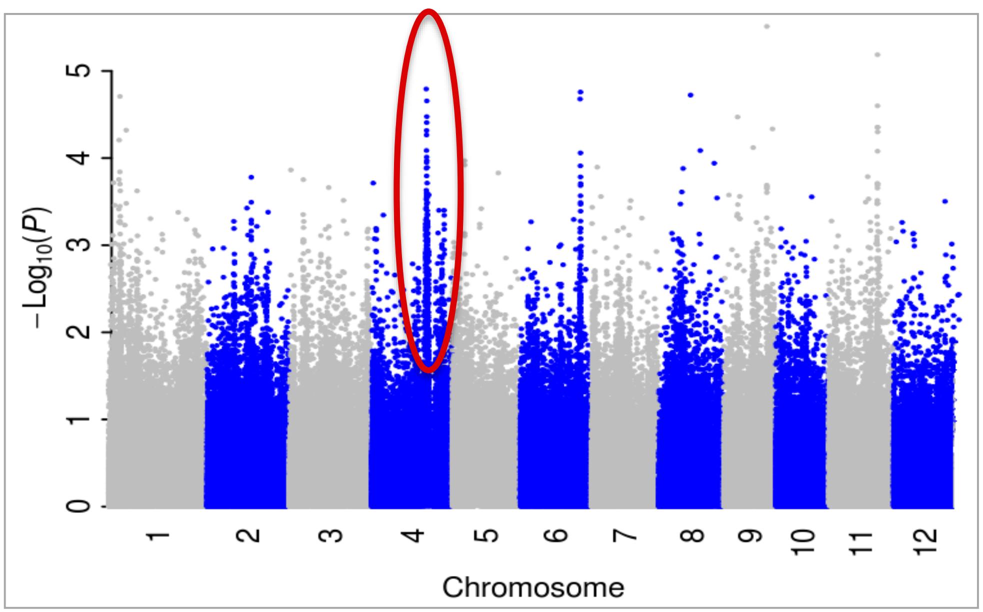
3D Root System Architecture (RSA)

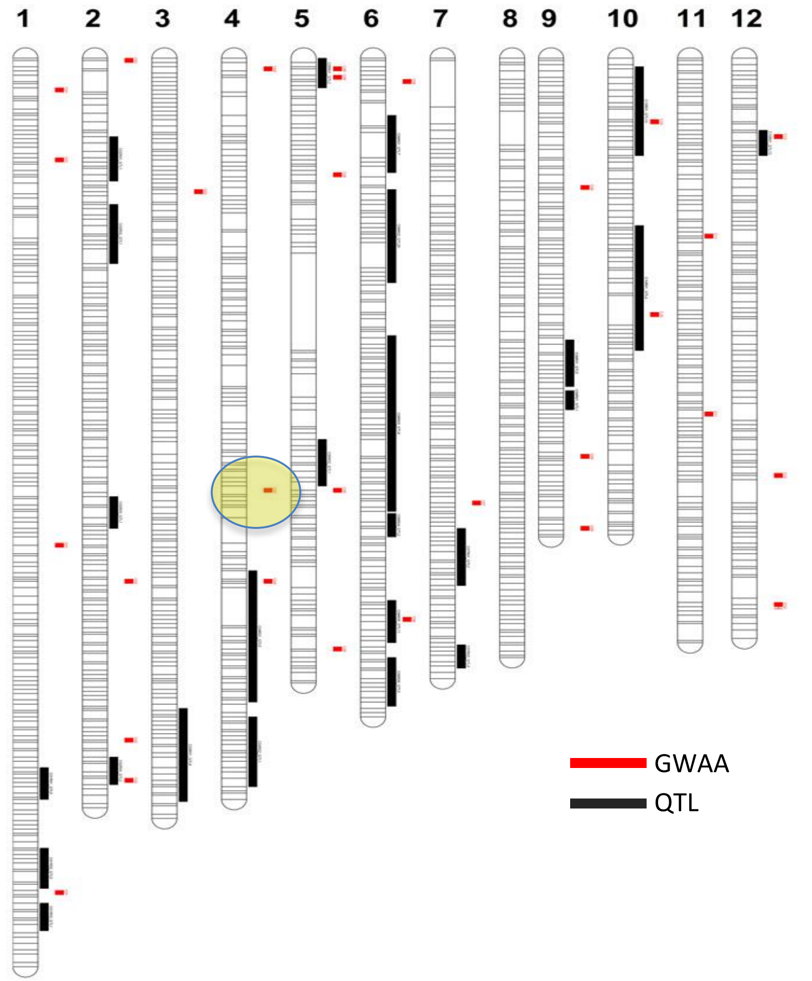
Randy Clark

Genome Wide Association Analysis

- 380 individual single trait analyses:
  - 13 traits x 3 days x 4 subpopulations
- Significant regions found for each analysis.
- Global, local and dynamic characteristics

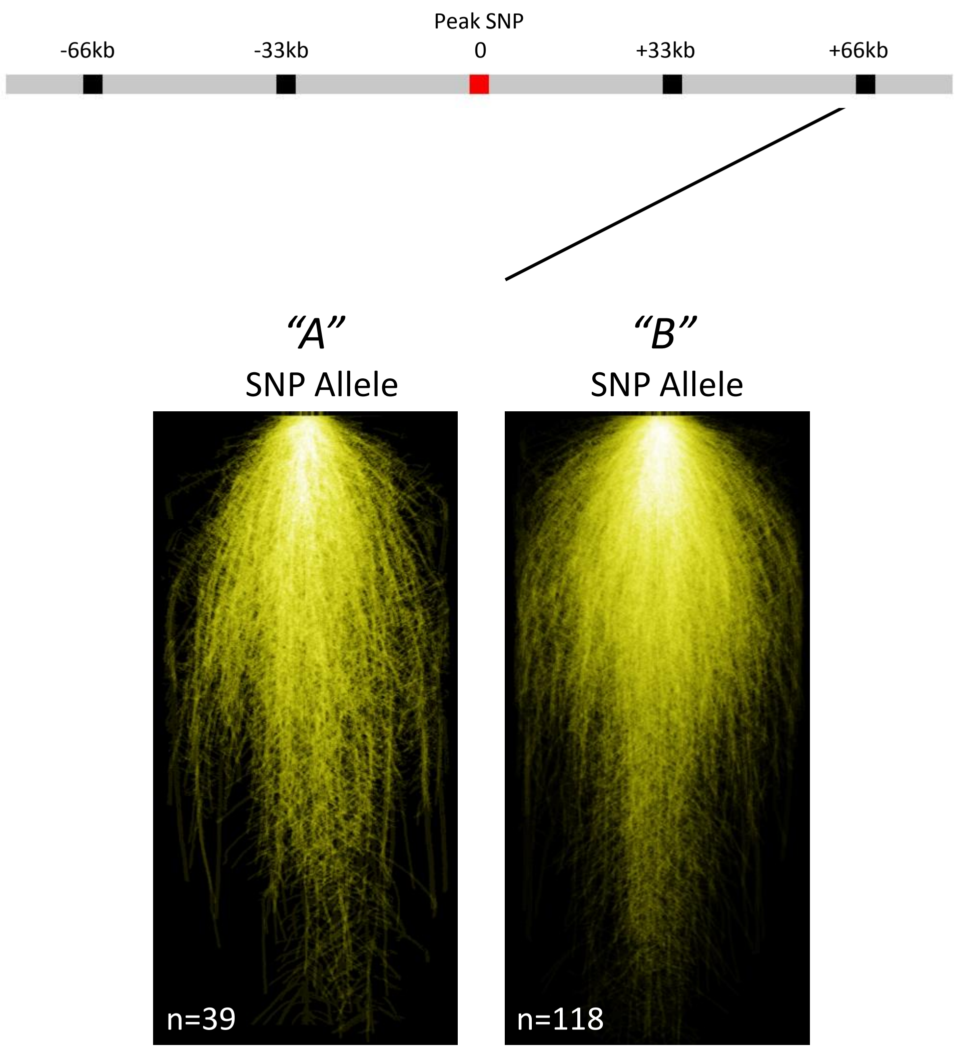
Region significantly correlated with rooting depth in *indica*

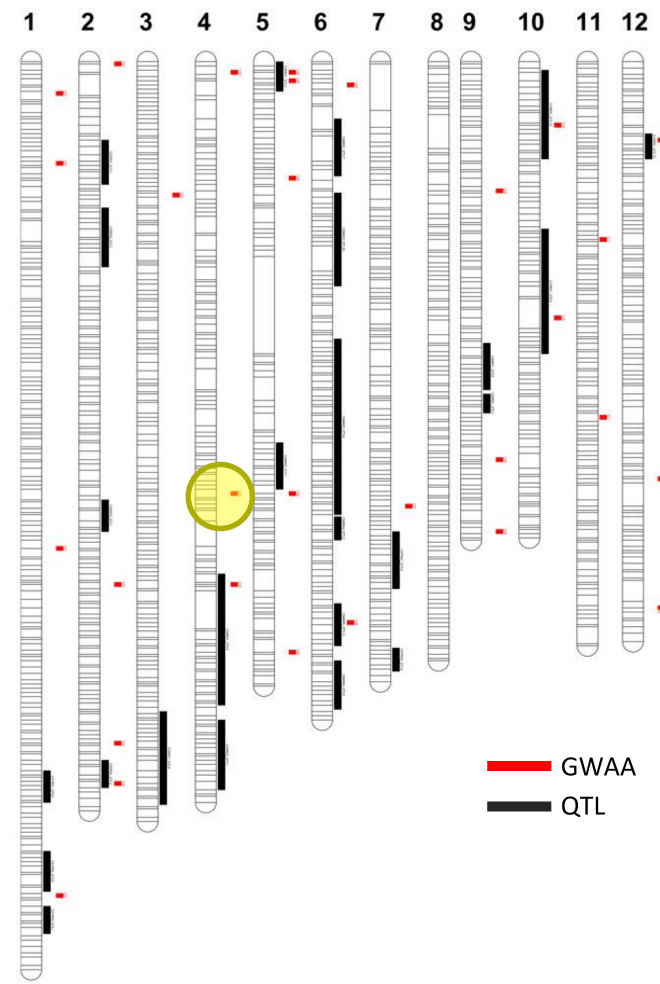




GWAS for Root System Architecture

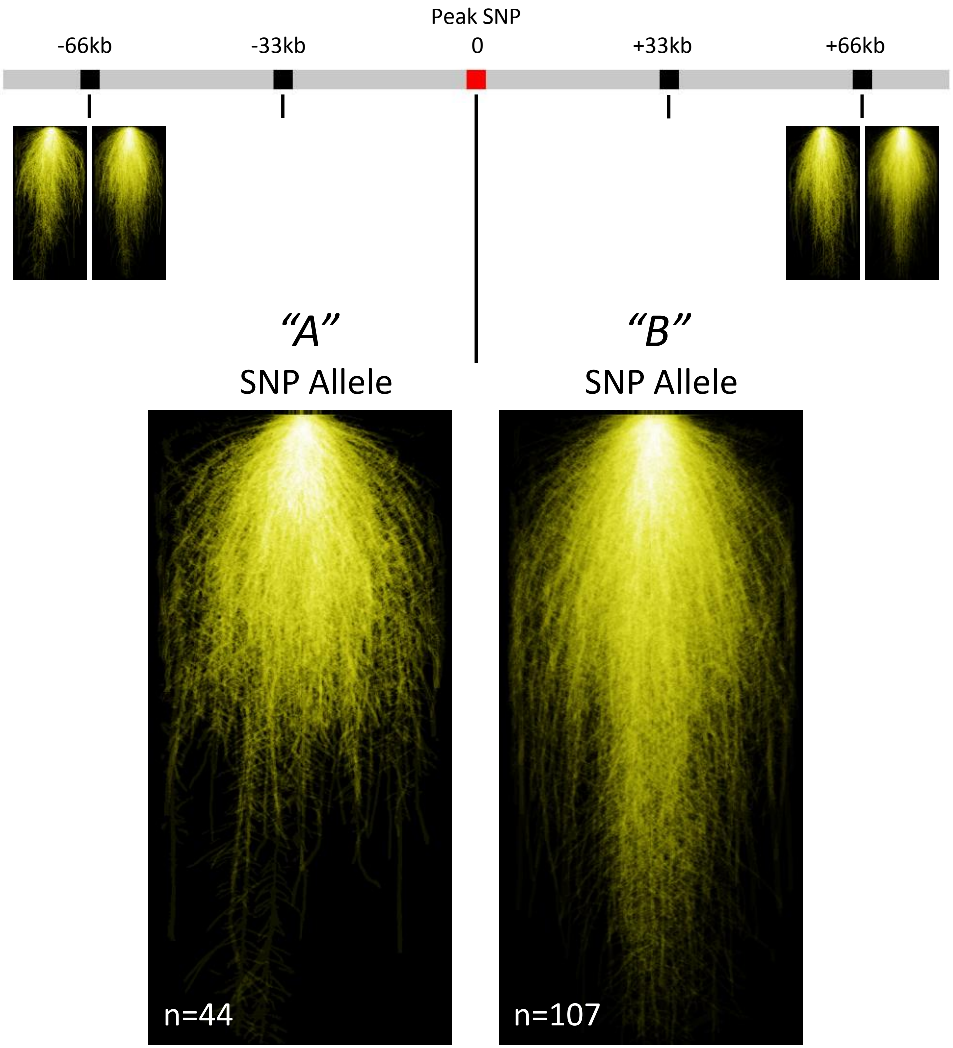
Indica subpopulation

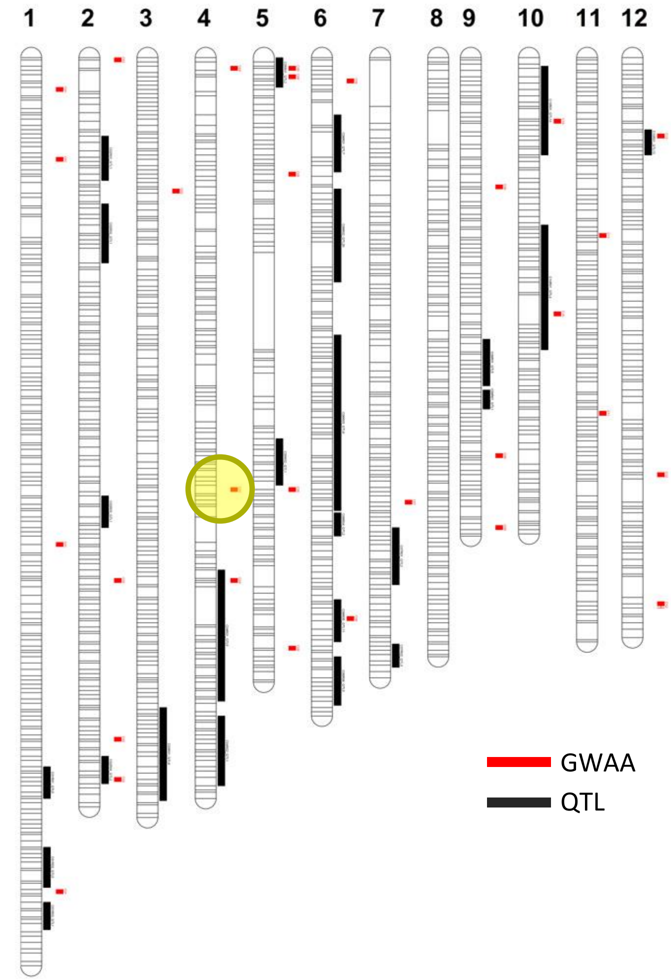




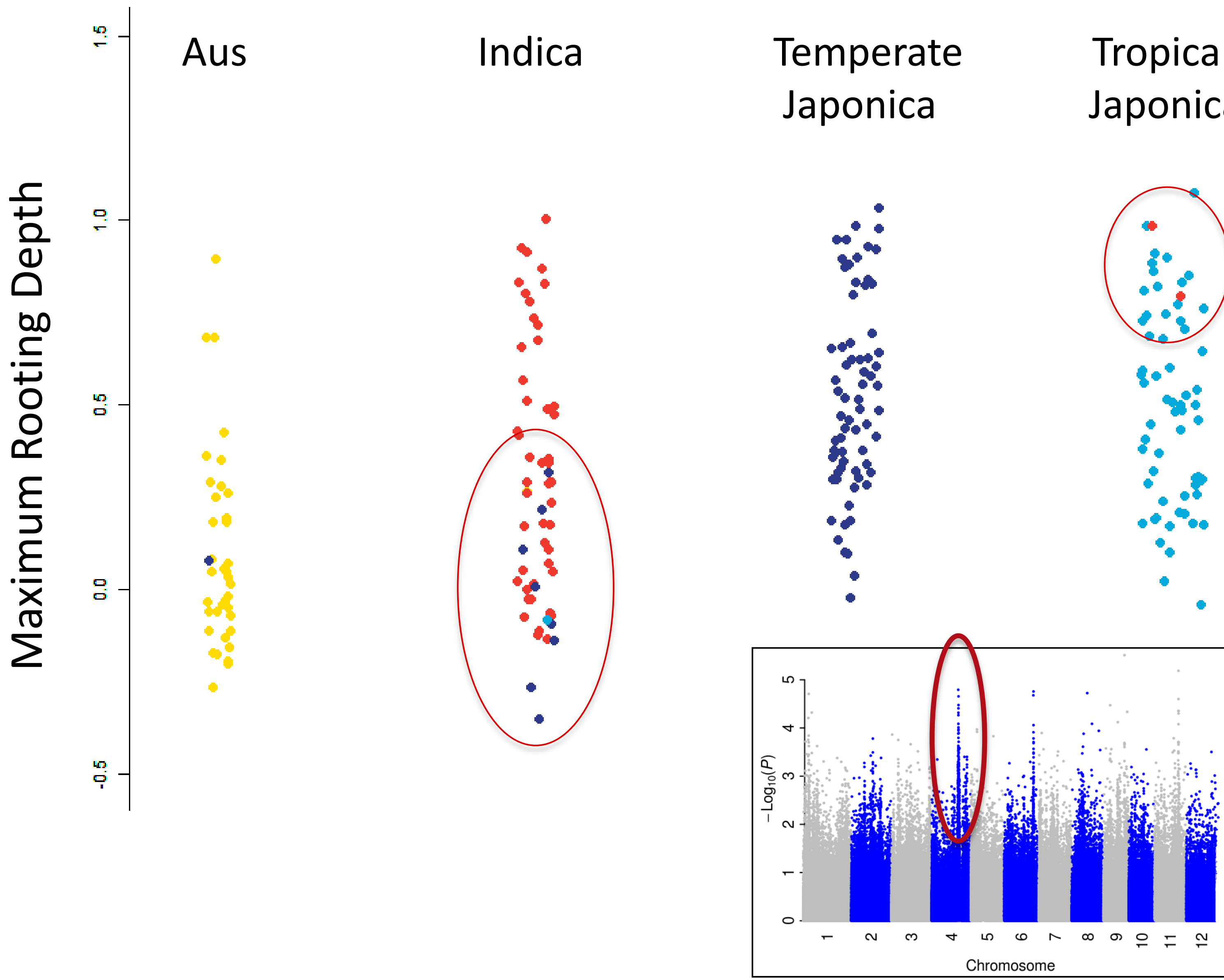
GWAS for Root System Architecture

Indica subpopulation



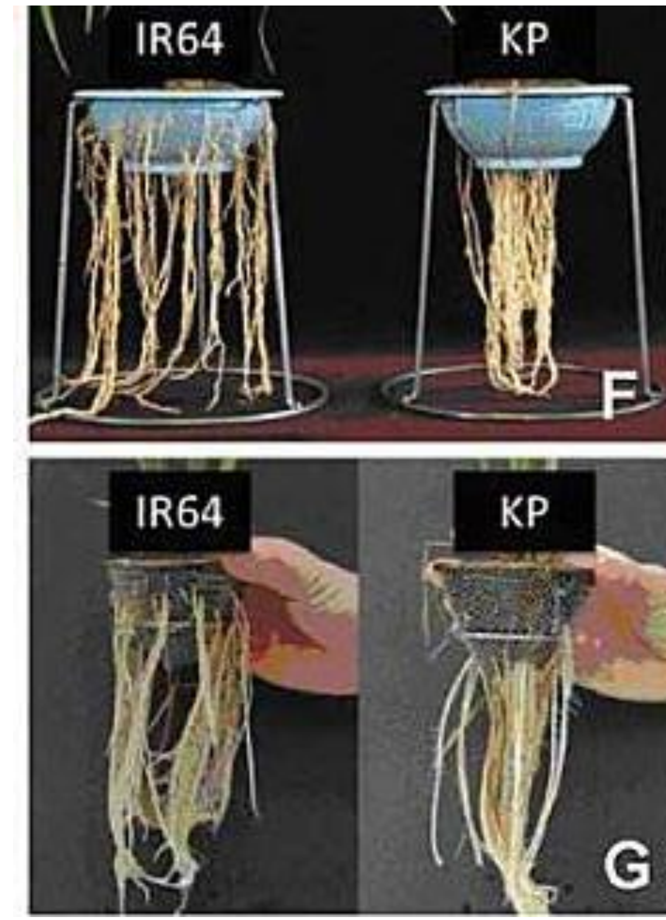


Sub-population introgression at target SNP



Co-localization of genes/QTLs - field & hydroponics

DRO1 & *Pistol1,* rice genes controlling root system architecture => enhance yield under drought & phosphorus uptake in low-fertility soils



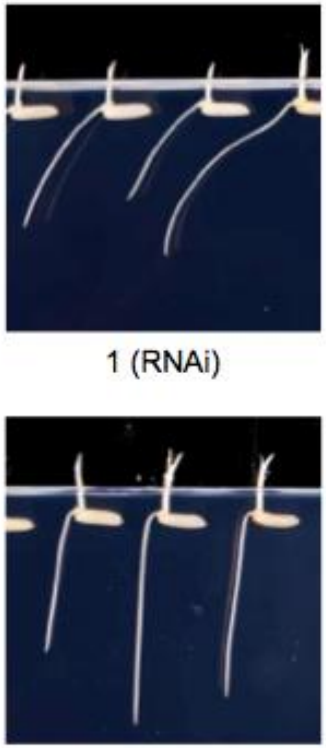
Uga et al. (2011) TAG;

Gamuyao et al. (2012) Nature

Uga et al. (2013) Nature Genetics 45(9):1097-102

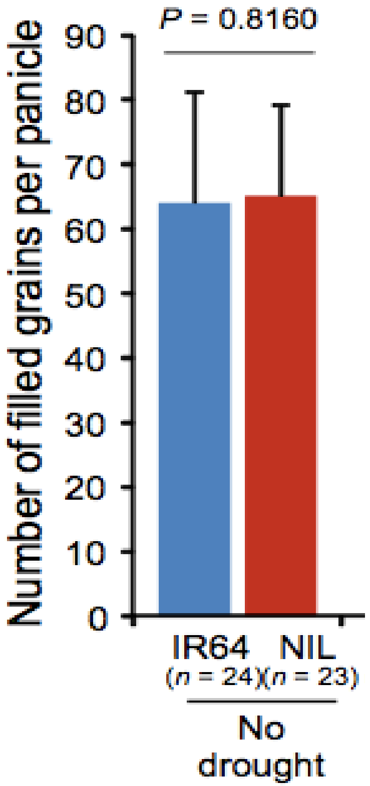
DRO1 = single bp deletion in exon 4 changes root angle & enhances grain yield under drought

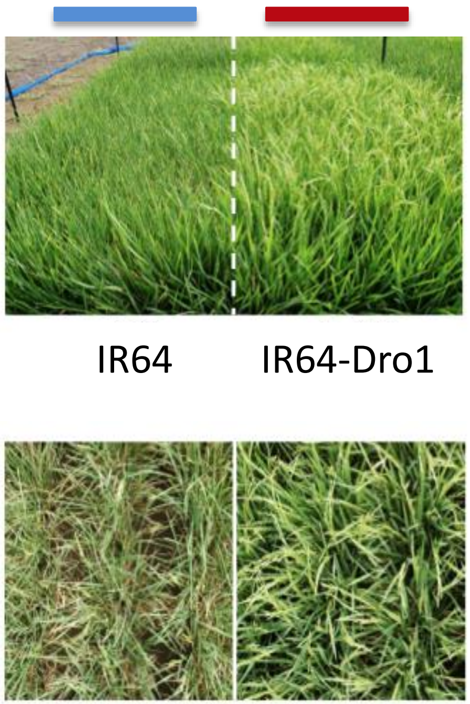
Marker-assisted selection (MAS) => NILs



1 (RNAi)

3 (KPg-Mc)





IR64 IR64-Dro1

IR64 IR64-Dro1

DRO1 expression positively correlated with root gravitropism, abundant in root meristems, responsive to auxin, localized to plasma membrane when introduced into protoplasts.

MAS vs Genomic Selection

Large-effect loci effectively targeted by MAS in rice; what about the many small-effect QTLs?

Yield

Grain quality

Disease and insect resistance

Plant architecture

Flowering time

Heat /cold tolerance

Submergence / drought tolerance

Micronutrient deficiency / toxicity

Developing GS models in rice

IRRI, Philippines

331 advanced *indica* inbred lines from the irrigated rice breeding program

Replicated Yield Trial (RYT):

4 years: 2009-2012

2 seasons: dry and wet

IRRI, Los Baños, Philippines

73K SNPs (GBS)
11 traits

INIA, Uruguay

311 *tropical japonica* elite Uruguayan varieties

- 12 half-sib families

Replicated Yield Trial (RYT)

- 3 years: 2010-2012
- Treinta y Tres, Uruguay

40K SNPs (GBS)
8 traits

Heritabilities & Prediction Accuracies

IRRI (tropics)

Trait	Season	h^2
YLD	DS 2012	0.3213
PH	DS 2012	0.3546
FL	DS 2012	0.4378

Trait	Training Population	Validation Population	Statistical Method	Accuracy
Yield	2009-2011 all	2012 DS	RR-BLUP	0.3044
Plant Ht	2009-2011 DS	2012 DS	Random Forest	0.3411
Flowering	2009-2011 DS	2012 DS	Multiple Linear Regression	0. 6270

Spindel et al. PLoS Genetics (2015)

Accuracy of models used for GS (IRRI)

Spindel et al. PLoS Genetics (2015)

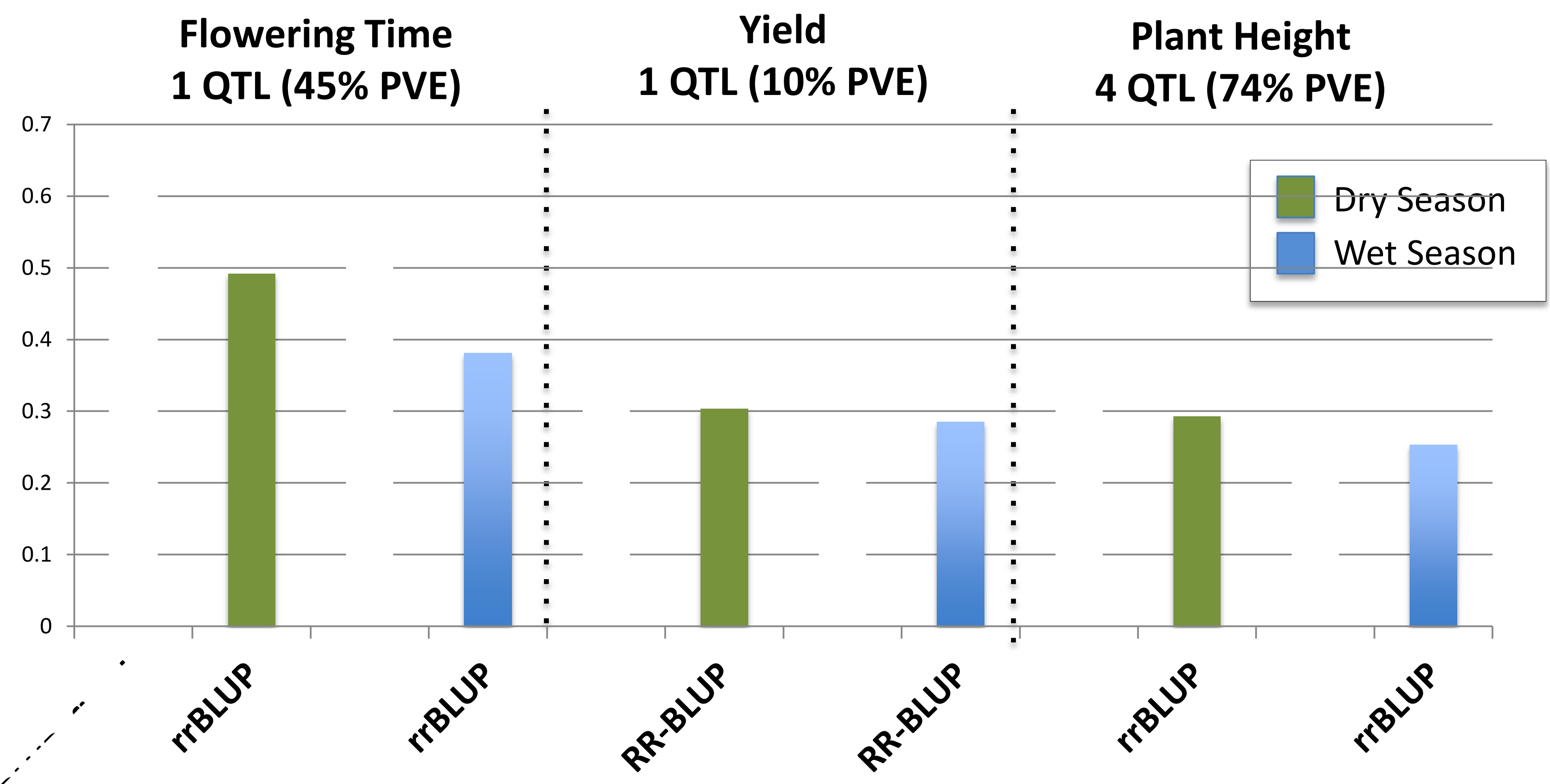
1. **RR-BLUP** – linear, parametric & frequentist
2. **Reproducing Kernel Hilbert Spaces (RKHS)** – linear, semi-parametric
3. **Random Forest** – machine learning
4. **Bayesian Lasso** – linear, parametric & Bayesian
5. **Multiple Linear Regression (MLR)** – a subset of significant markers chosen to fit a linear model
6. **Pedigree** –kinship BLUP used with pedigree A-matrix alone

Trait	VP	stat method	accuracy
YLD	2012 DS	A RR-BLUP	0.3044
YLD	2012 DS	A RKHS	0.2596
YLD	2012 DS	A RF	0.2458
YLD	2012 DS	B ped	0.2146
YLD	2012 DS	C BL	0.1358
YLD	2012 DS	D MLR	-0.0599
YLD	2012 WS	A RF	0.3136
YLD	2012 WS	A RR-BLUP	0.2852
YLD	2012 WS	A RKHS	0.2399
YLD	2012 WS	B ped	0.1904
YLD	2012 WS	C BL	0.0876
YLD	2012 WS	D MLR	0.0095
FL	2012 DS	D MLR	0.6270
FL	2012 DS	A RF	0.6093
FL	2012 DS	A RR-BLUP	0.4919
FL	2012 DS	A RKHS	0.4865
FL	2012 DS	C BL	0.4536
FL	2012 DS	B ped	0.3997
FL	2012 WS	D MLR	0.5400
FL	2012 WS	A RF	0.4187
FL	2012 WS	A RKHS	0.3872
FL	2012 WS	A RR-BLUP	0.3808
FL	2012 WS	C BL	0.3237
FL	2012 WS	B ped	0.2071

Spindel et al. Heredity (2016)

GS Prediction Accuracies

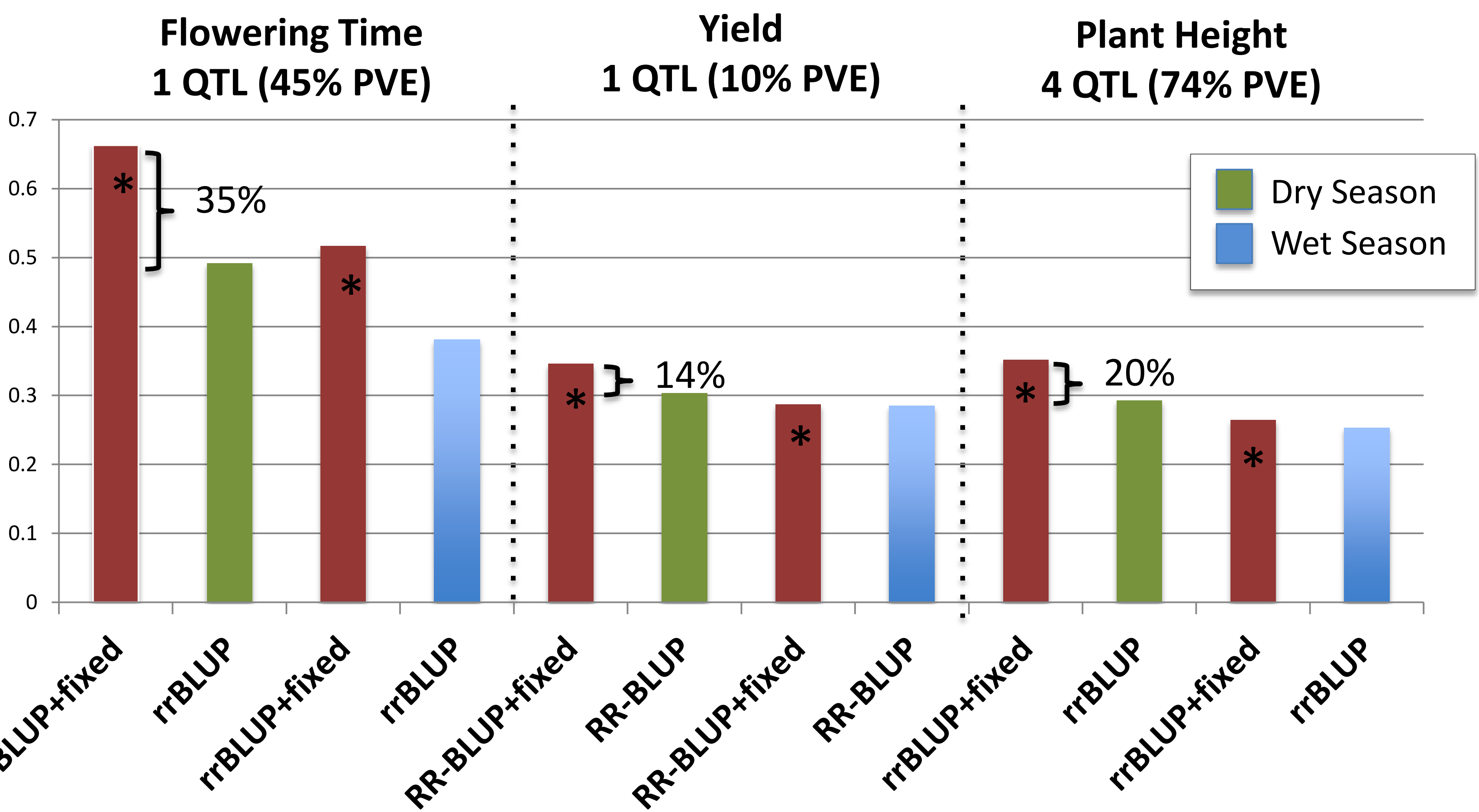
5-fold cross-validation (CV)



Prediction accuracies equivalent to trait heritabilities (h^2)

GS Prediction Accuracies

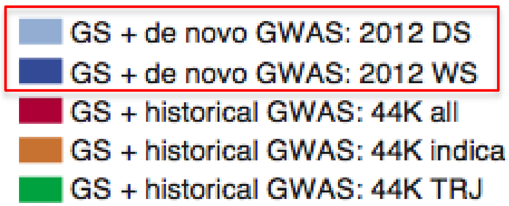
significantly improved using GWAS information

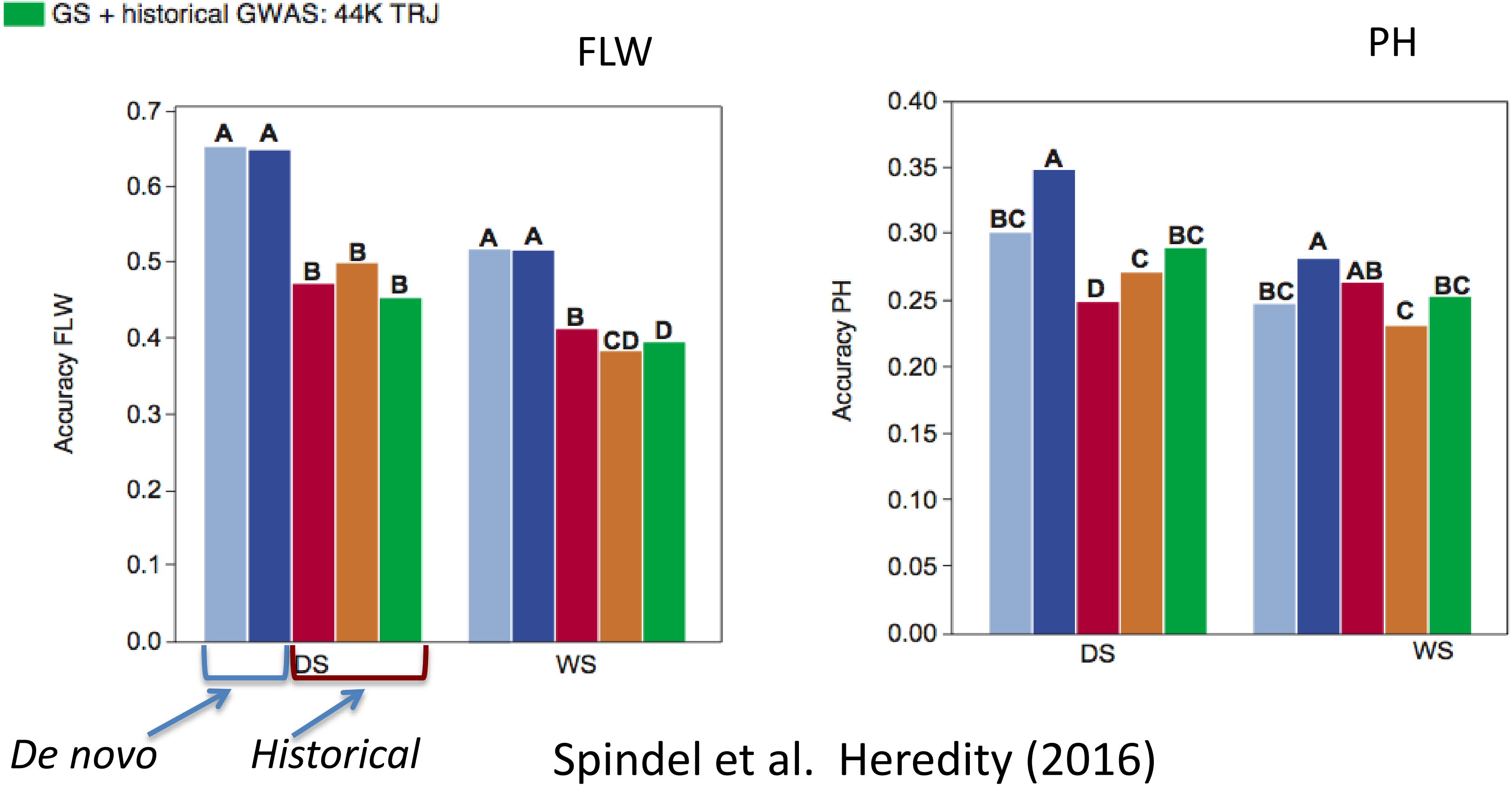


Spindel et al. Heredity (2016)

GS + *de novo* GWAS versus GS + historical GWAS

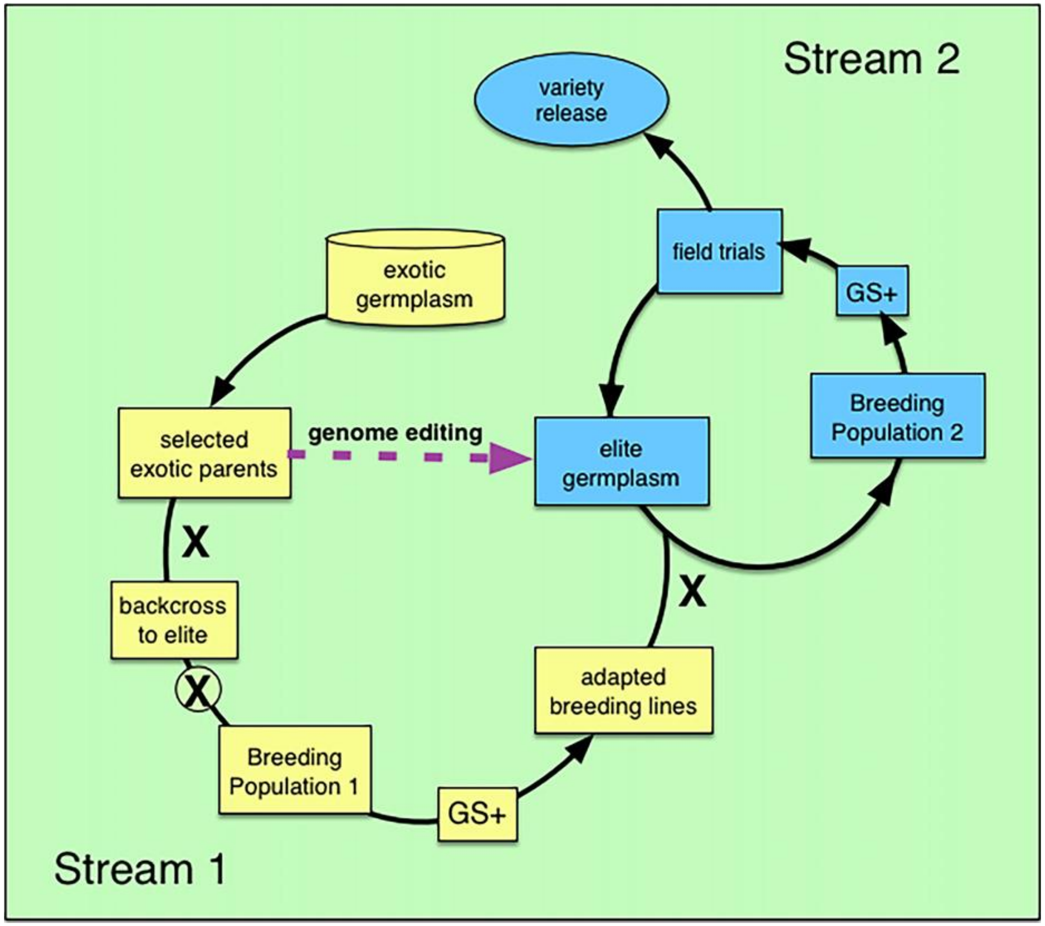
de novo GWAS on breeding population better than blindly using "candidate genes" from literature





Spindel et al. Heredity (2016)

GWAS + GS => facilitate/ accelerate introgression of 'exotic' variation into elite breeding pools



Spindel et al. Heredity (2016)

Linking GWAS & GS to better utilize natural variation in rice?

- GWAS is helping develop a Genotype => Phenotype road map to accelerate trait /gene discovery in rice
- Genetic architecture of most complex traits in rice involves alleles of large effect (GWAS) and genes of small effect (GS); different genes/ alleles important in different populations (wild & cultivated)
- Use of *de novo* GWAS + GS => weights alleles of large effect and enhances accuracy of genome-wide prediction.

National Science Foundation
WHERE DISCOVERIES BEGIN

USDA
United States Department Of Agriculture
Agricultural Research Service

Bill & Melinda Gates Foundation

Acknowledgements

PB&G- Cornell

Jennifer Spindel
Randy Clark
Lyza Marón
Koni Wright
Hyunjung Kim
Janelle Jung
Sandy Harrington
Kazi Akther
Moni Singh

Francisco Agosto
Pavel Korniliev
Jason Mezey

USDA

Leon Kochian
Jean Luc Janninck
Jon Shaff
Anna McClung
Georgia Eizenga

IRRI - Philippines

Bert Collard
Joshua Cobb
Ed Redoña
Hasina Begum
Parminder Virk
Eero Nissila
Ken McNally
Hei Leung

NIAS - Japan

Yusaku Uga
Masahiro Yano

INIA, Uruguay

Victoria Bonecarrere
Pedro Blanco
Akira Saito
Lucia Gutierrez
Juan Rosas