



IRRI

# Rice Genetic Diversity, Gene Flow, and Population Structure

*Kenneth L. McNally*

*Senior Scientist, Molecular Genetics/Taxonomy*

T.T. Chang Genetic Resources Center

International Rice Research Institute

Los Baños, Laguna, Philippines

Rice  
Science  
for a Better  
World



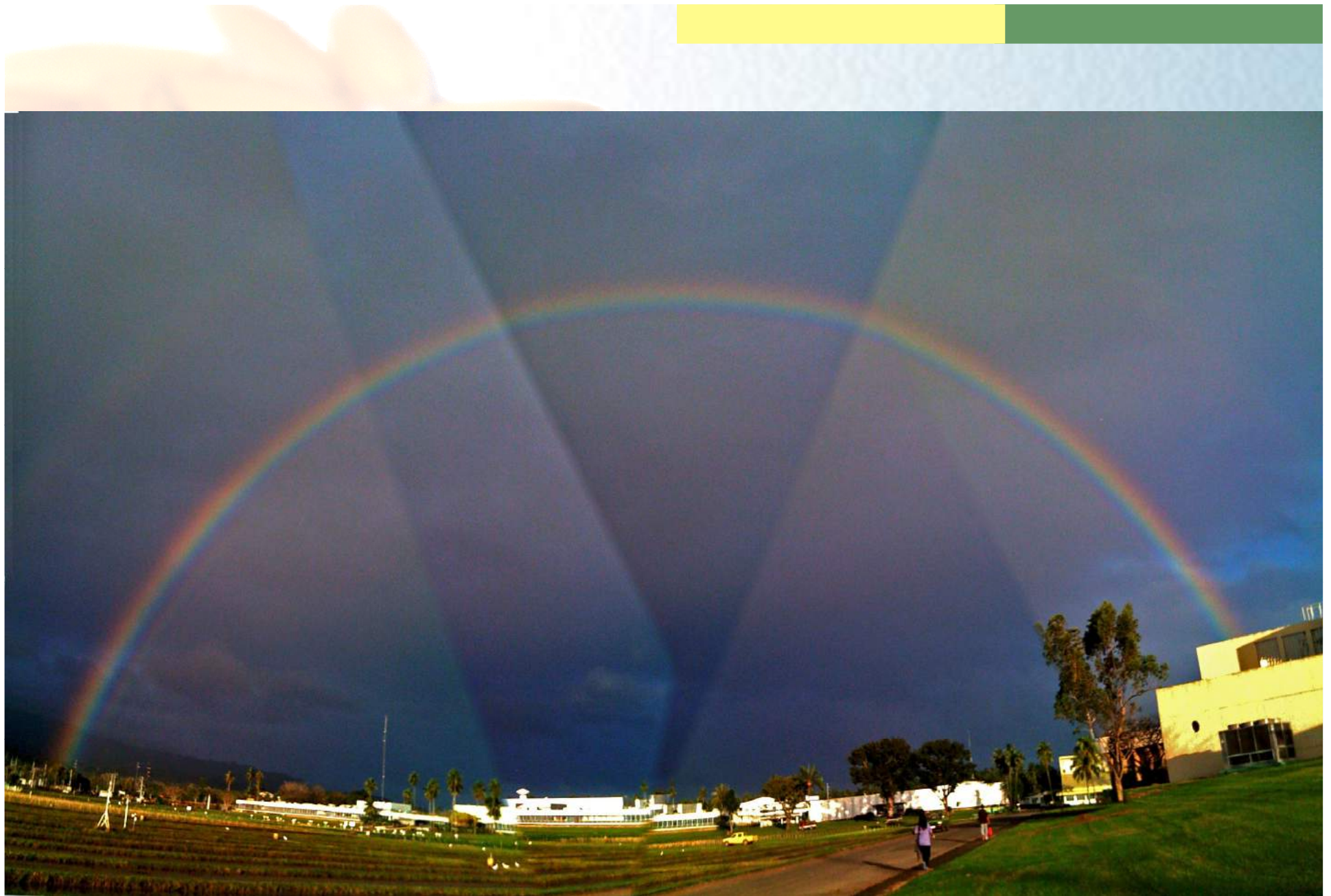
# IRGC - the International Rice Genebank Collection

World's largest collection of rice germplasm held in trust for the world community and source countries



- Over 108,000 registered accessions from 117 source countries
- Two cultivated species
  - Oryza sativa*
  - Oryza glaberrima*
- 22 wild species
- Relatively few accessions have donated alleles to current, high-yielding varieties
- <http://www.irri.org/GRC>





# How diverse is rice?

Approx. 22 wild species in *Oryza*

Two species cultivated

*O. sativa* and *O. glaberrima*

At least 3 separate domestication events

*O. glaberrima* in west Africa

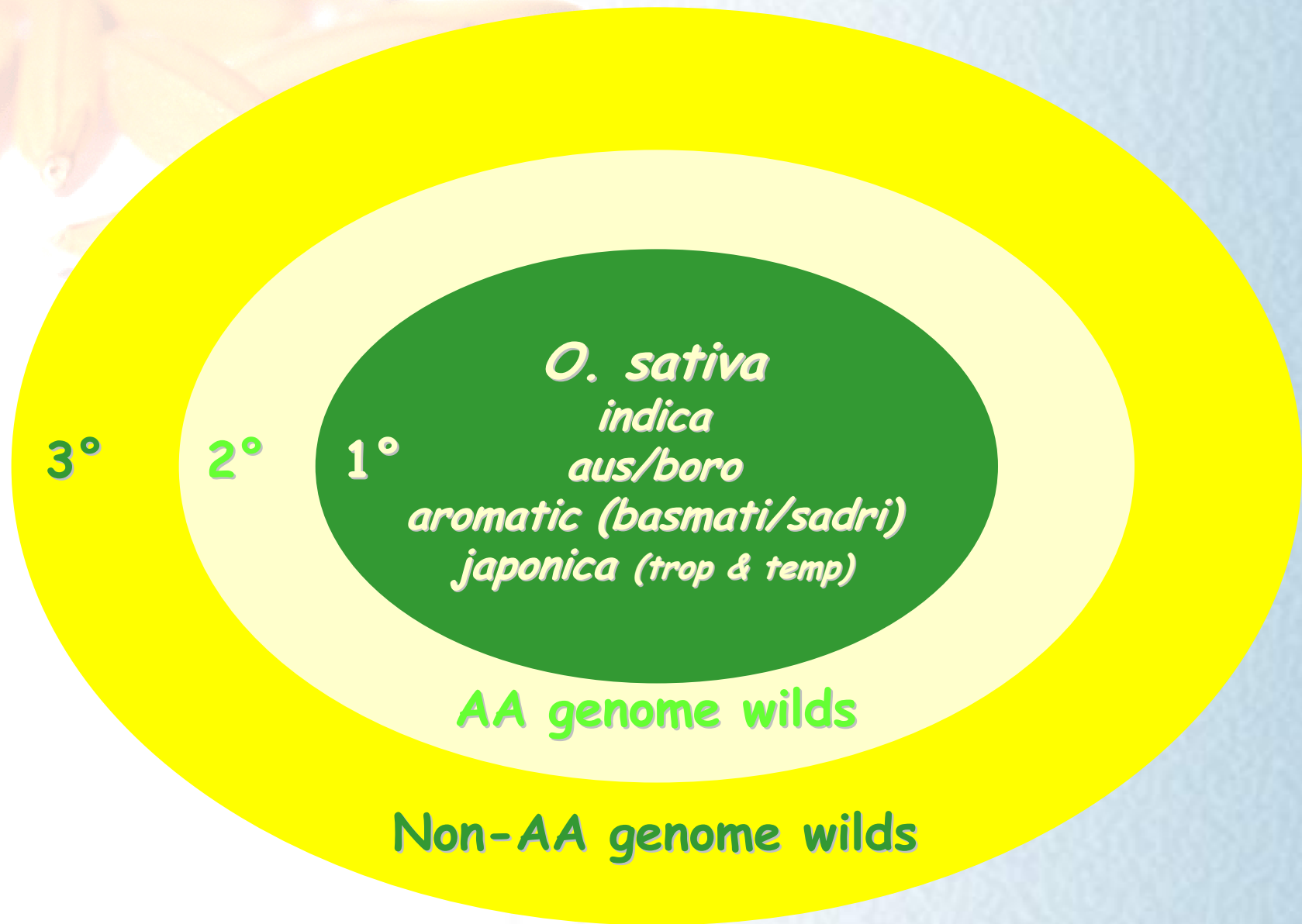
*O. sativa* twice? → two major "variety groups"

S Asia → Indica

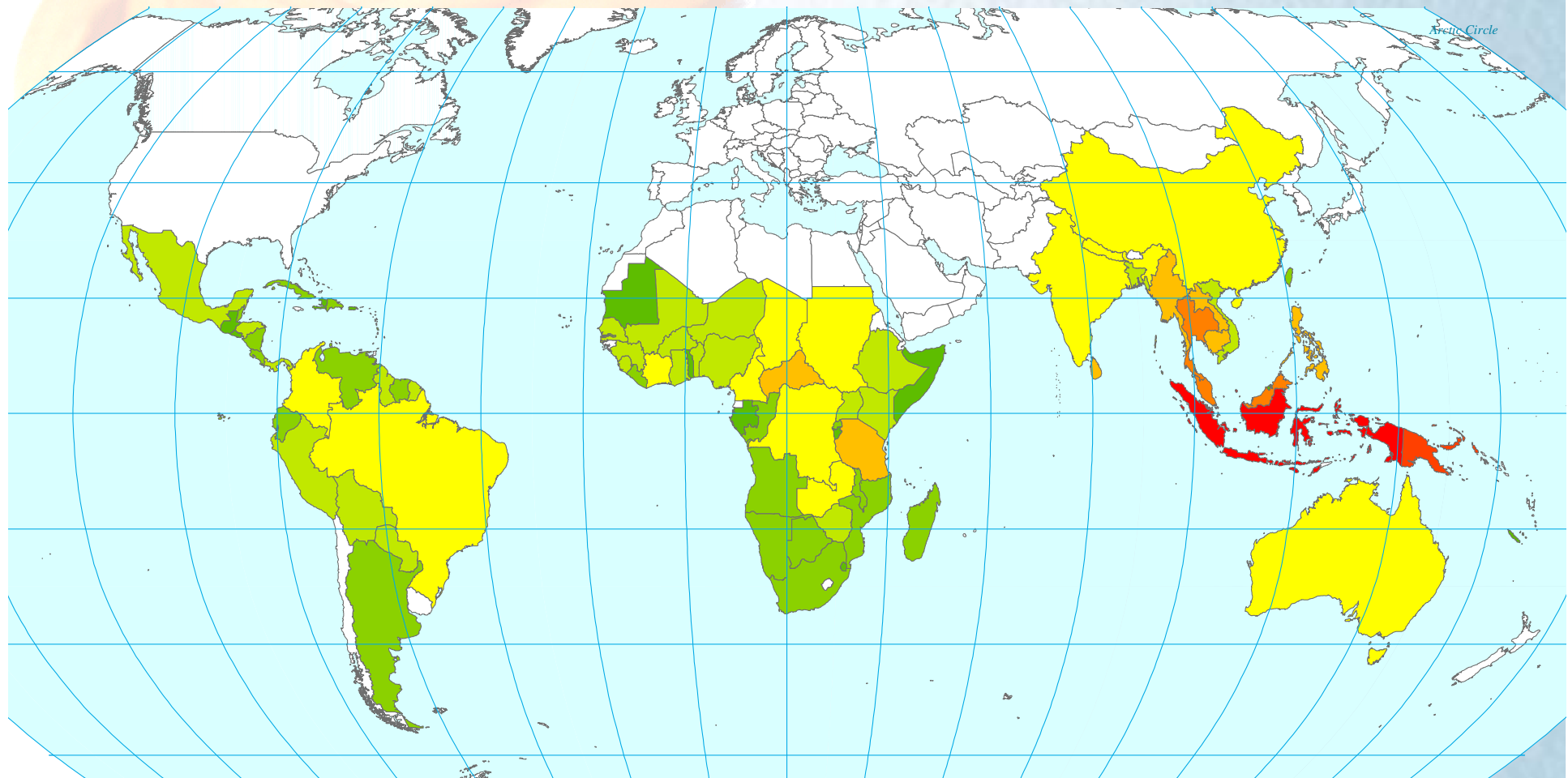
E Asia → Japonica


Possibly more?

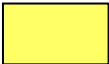
# *O. sativa* Gene Pools



# Wild rice distribution



 1 species

 4 species

 9 species



# Wild *Oryza* species with AA genome



**longistaminata**



**glumaepatula**



**barthii**



**rufipogon**

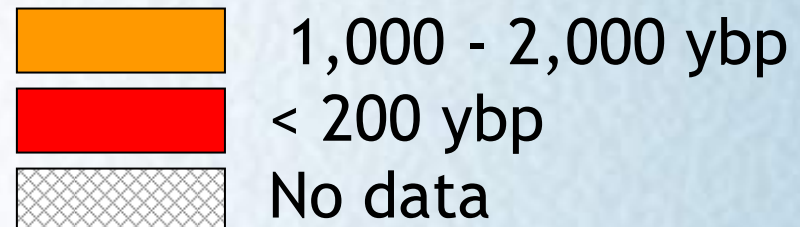
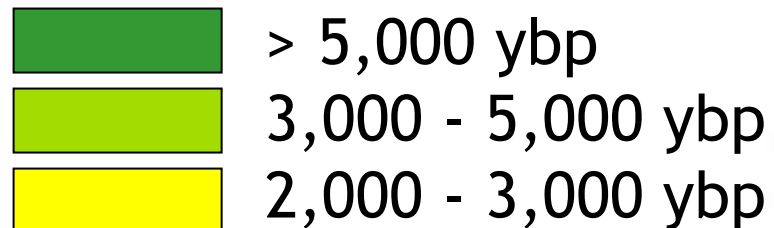
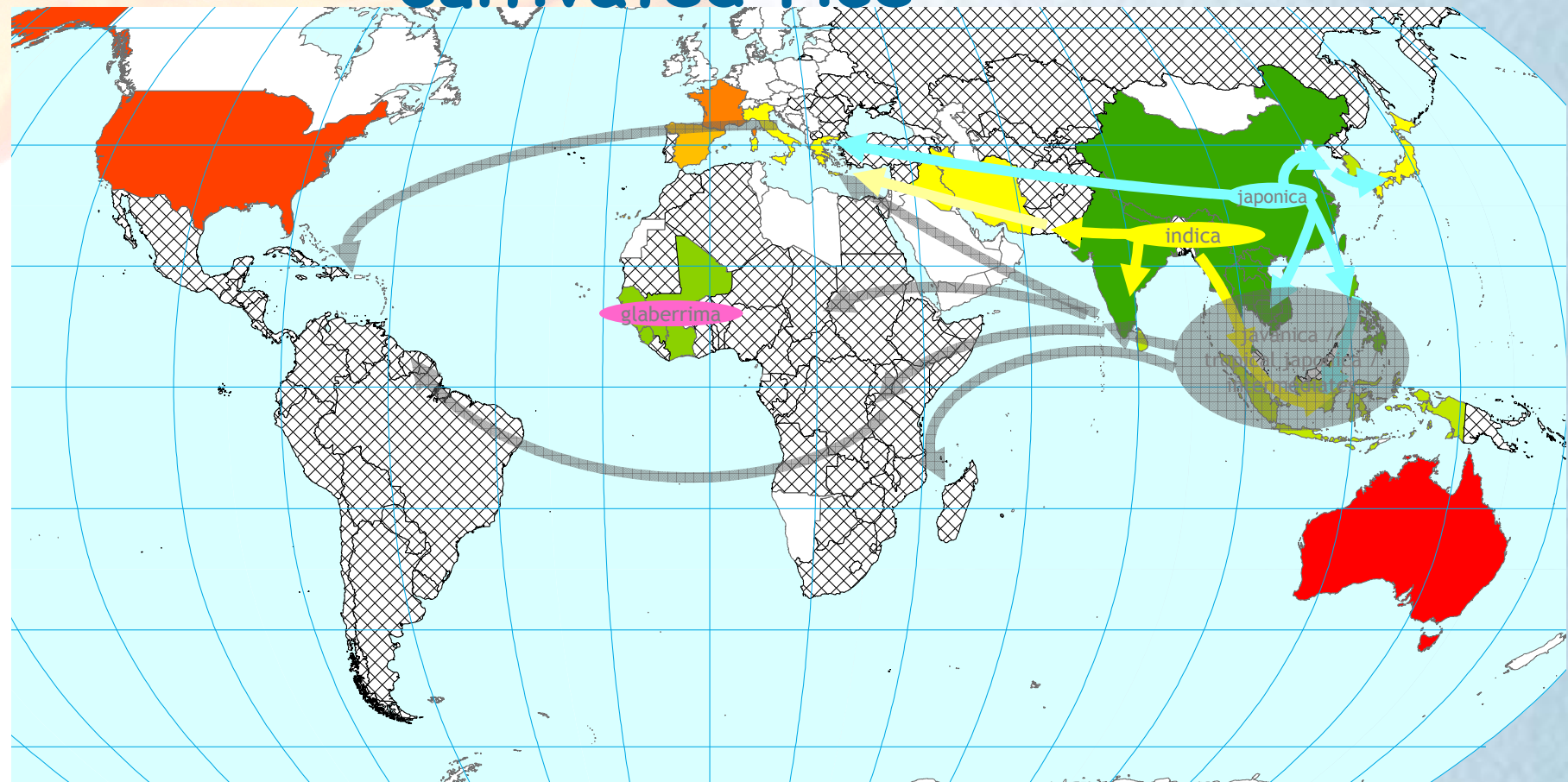


**nivara**



**meridionalis**

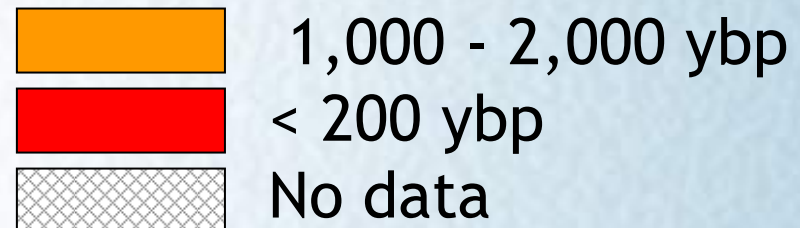
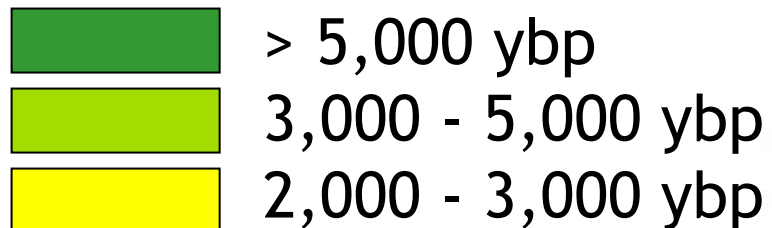
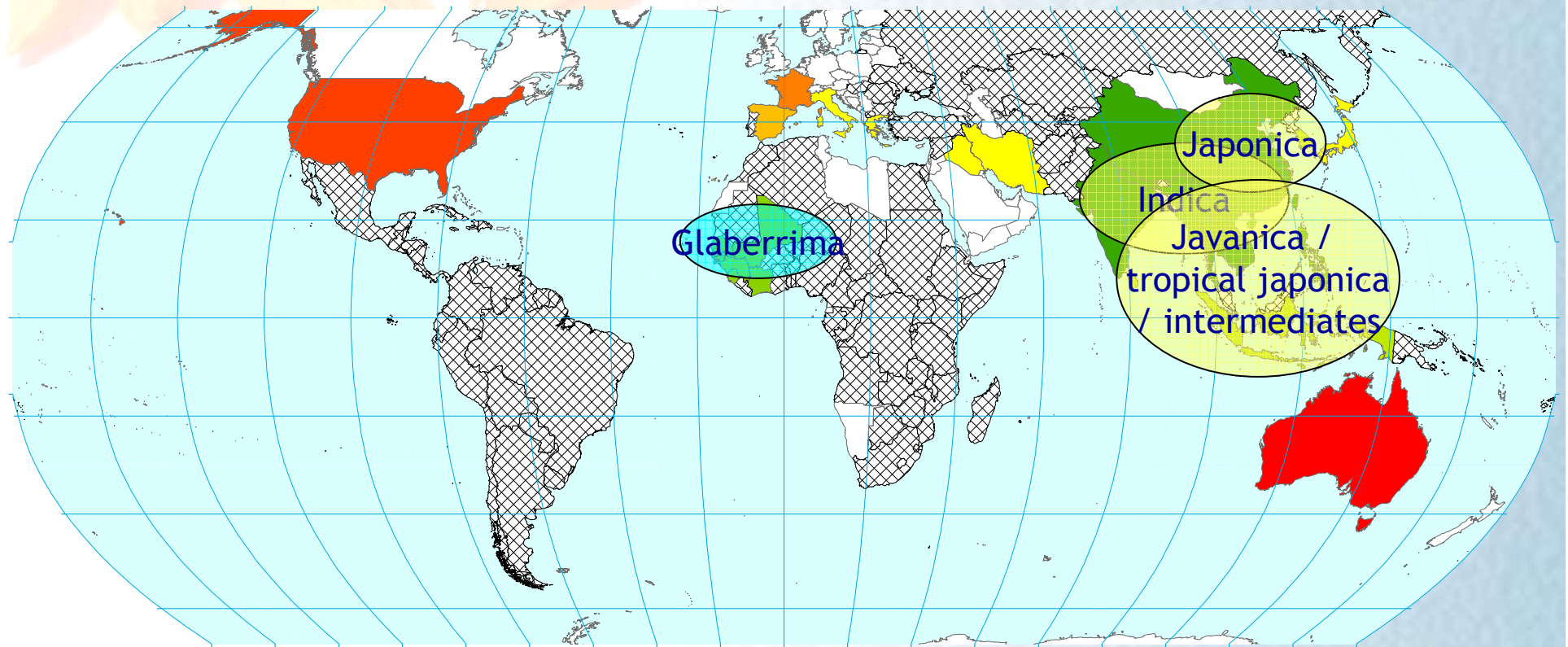
# Origin and spread of cultivated rice





# Cultivated rice:

## 1° Centres of diversity



# How many varieties of rice?

For the taxonomist: 0

"Variety" has a specific technical meaning in taxonomy.

No taxonomically valid descriptions of varieties of *O. sativa*

For the lawyer: 5,000-10,000?

New varieties only

IRIS database documents 5,000 new varieties

China releases about 200-300 a year

WTO requires a system of "Plant Variety Protection" (PVP)

Variety release requires DUS test

→ 1 named variety = 1 distinct genetic entity

For the genebank

How many traditional varieties/landraces?

Any accession from farmer's fields that doesn't have a modern breeding or selection history.



# How many traditional varieties?

MS Swaminathan Research Foundation estimates  
used to be 400,000 in India  
still 100,000 in use in India  
500,000 world wide?

How many varieties in IRRI genebank?

108,000 accessions of which 80% traditional and  
wild

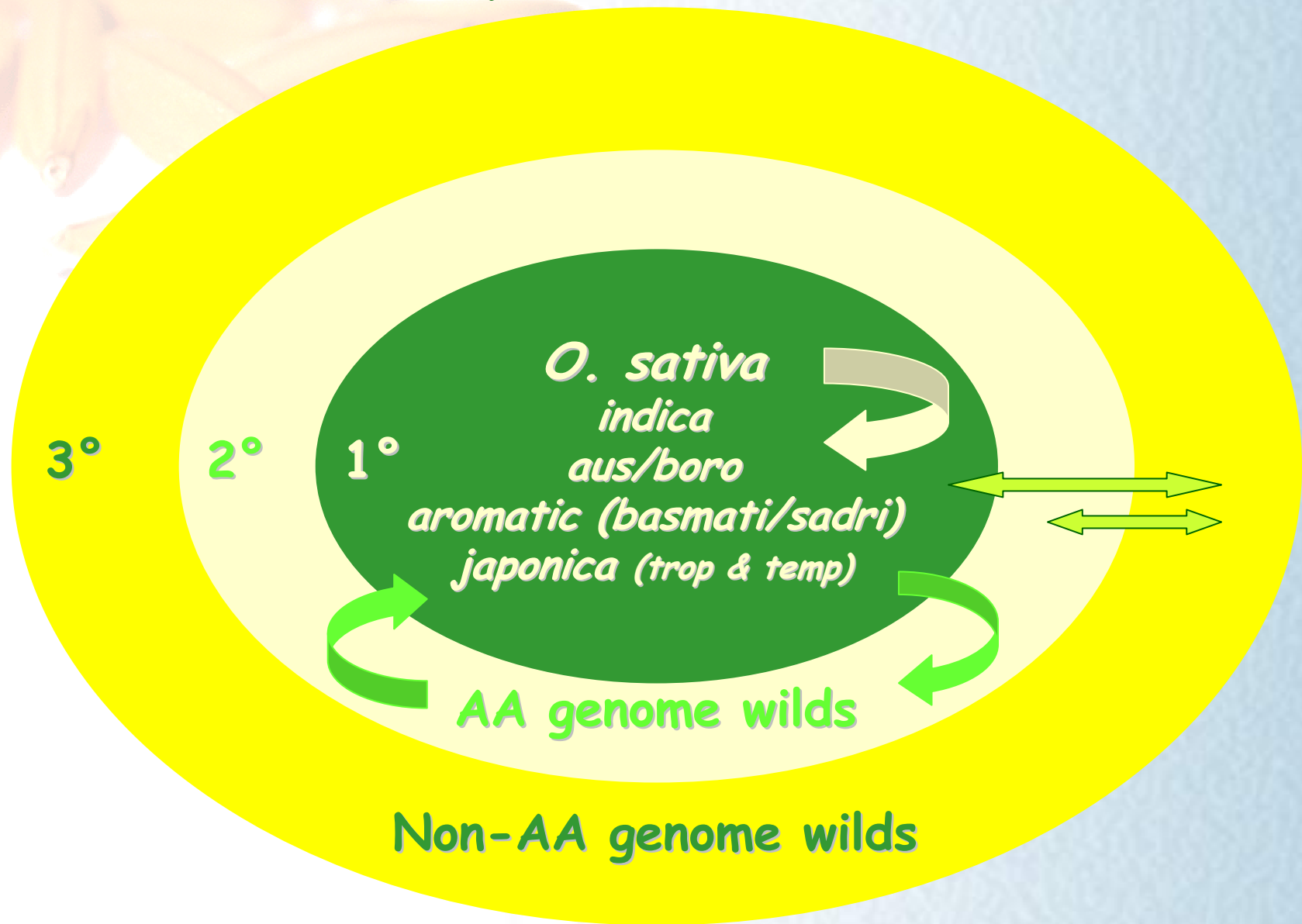
0 - 20,000 duplicates??

Depending on definition of duplicate





# Gene Flow in *Oryza*





# Molecular Markers in Rice



# Molecular Markers Applied to Rice

## Non PCR-based

### Isozymes -

classical marker, allozyme forms of enzymes such as SOD  
Time-consuming, difficult to score (mid 1980s)

RFLP- restriction fragment linked polymorphism  
Costly, time-consuming (early 1990s)

## PCR-based

RAPD - randomly amplified polymorphic DNAs  
Difficult to score, not-robust, dominant (mid 1990s)  
AFLP - amplified fragment length polymorphism (mid 1990s)

SSRs - simple sequence repeats (mid 1990s)  
Current marker of choice, >25,000 available  
Co-dominant  
Can be difficult to score (e.g. di-repeats)

→ **SNPs - single nucleotide polymorphism**  
**New marker of choice**  
**Non-gel based assays, High throughput** ←

Others:

SCAR

Intron Scanning/InDel

Inter-SSR



# SNPs as compared to other markers

## Qualitative (nucleotide difference) polymorphism

SNPs                      bi-allelic, co-dominant

less informative, but more reliable scoring

Allele may reflect functional differences

## Quantitative (size-based) polymorphism

SSRs                      multi-allelic, co-dominant

more informative, less reliable scoring

RFLPs                      usually bi-allelic, co-dominant

RAPDs                      bi-allelic, dominant

AFLPs                      bi-allelic, dominant

Alleles are usually not functional



# Understanding the Diversity of Rice

## *Population Structure*





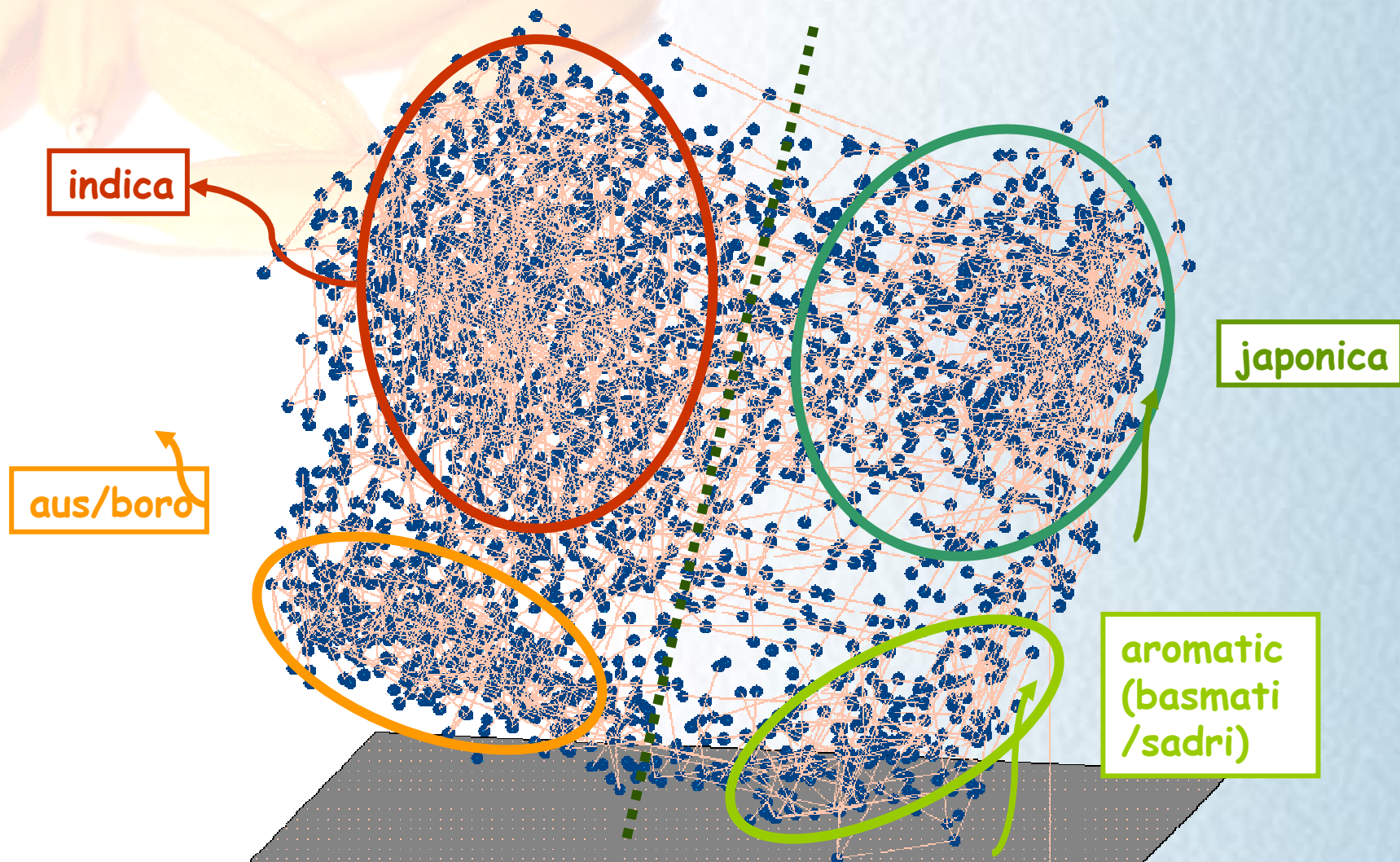
## Necessity for establishing population structure

- Group accessions according to shared common history
- Pre-requisite for effective association genetics
- Preclude spurious associations based on shared ideotypes
- Traits having a common or similar phenotype may not be a result of expression/regulation of the same genes
  - e.g. grain shape of indicas and javanicas



## Correspondence Analysis

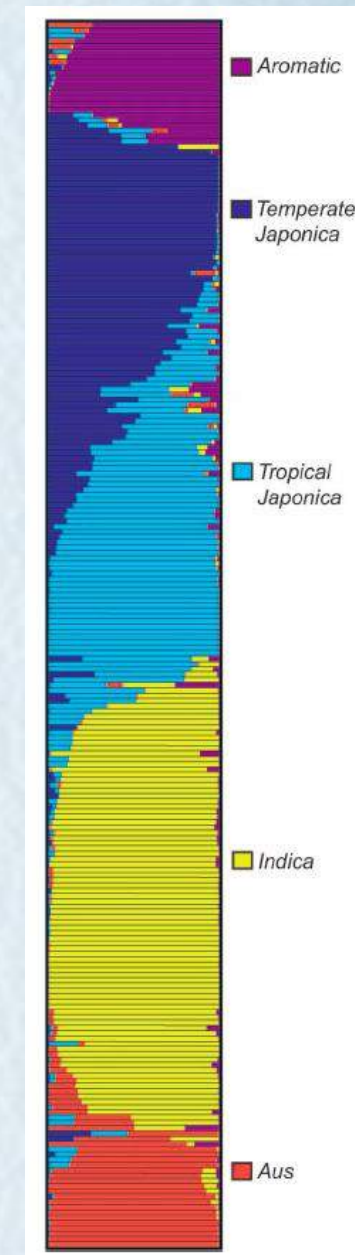
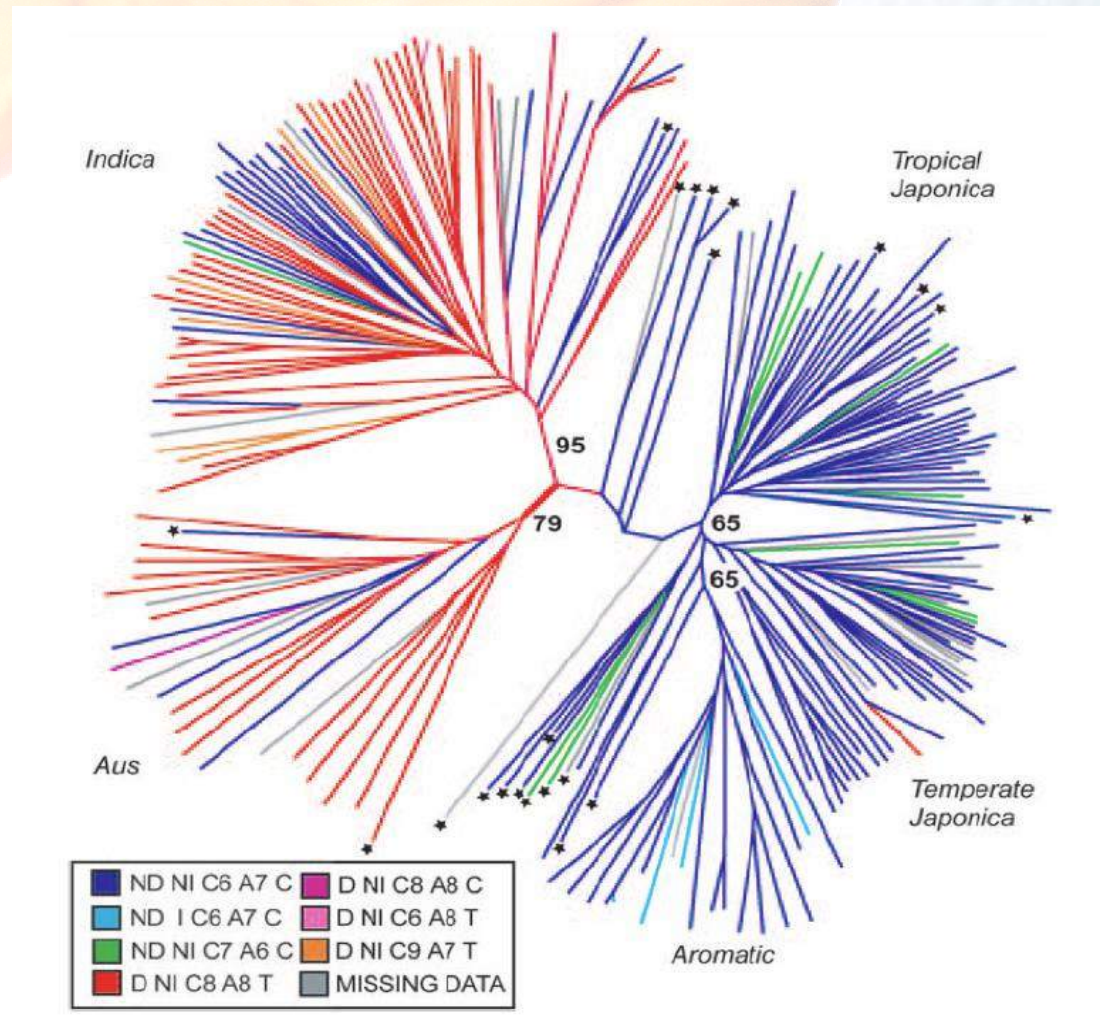
3,122 unique Isozyme patterns for 63 alleles at 15 loci



Summary of 24,157 accessions from  
IRGC (Glaszmann, Second, Juliano)  
PBGB (Brar et al., Virk, Courtois)

# Population Structure by SSRs

234 varieties, 169 nuclear SSRs



Garris et al 2005 Genetics 169:1631-1638

# GCP Rice Composite Set

IRRI



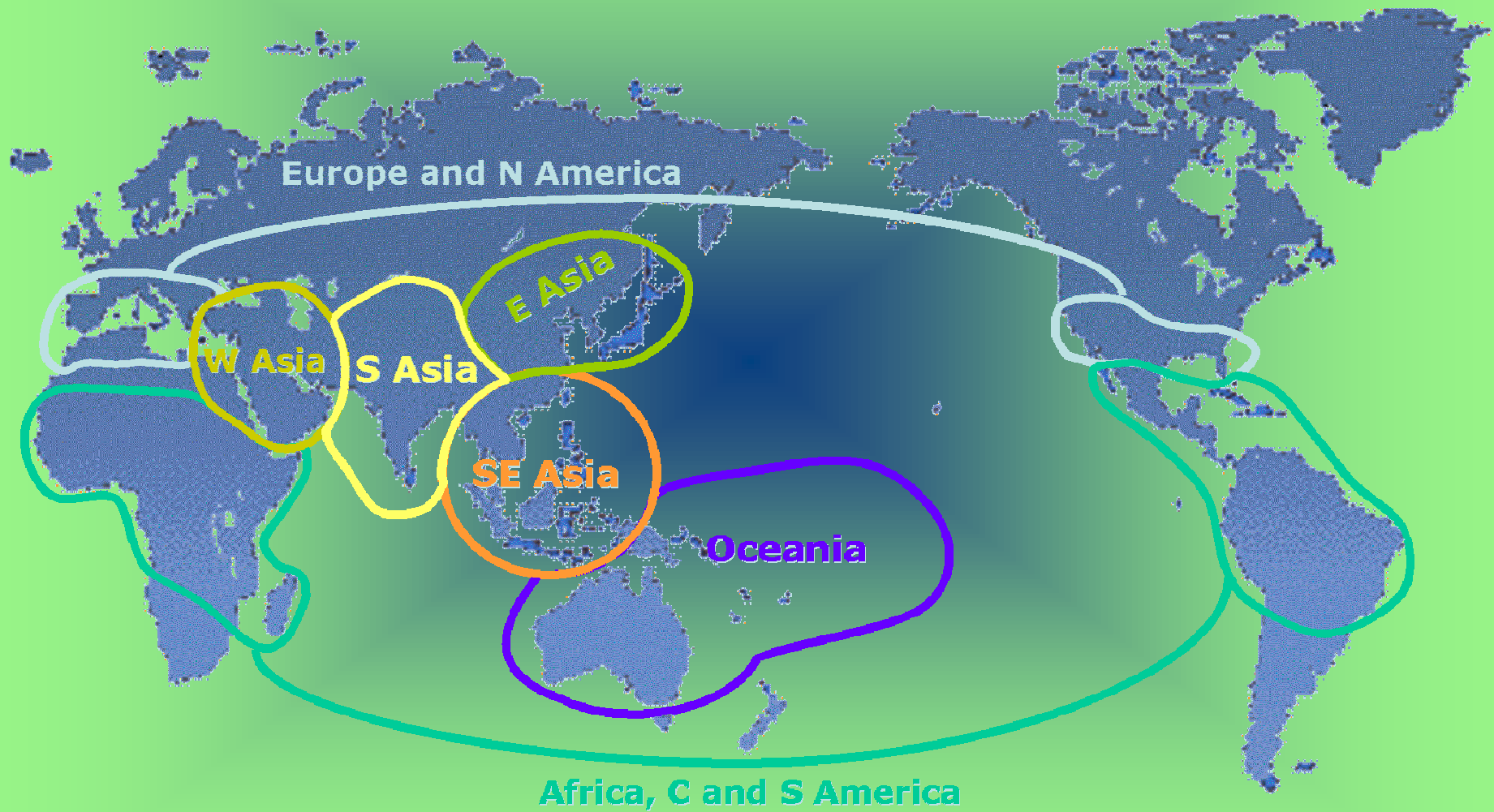
AGROPOLIS  
CIRAD

Embrapa

CORNELL



## Eco-regional Partitioning of Rice Genepools



# Species Composition

SPECIES	Frequency	%
<i>O. sativa</i> *	2338	84.80
<i>O. glaberrima</i> **	373	13.53
<i>O. rufipogon</i>	14	0.51
<i>O. nivara</i>	10	0.36
<i>O. glumaepatula</i>	9	0.33
<i>O. barthii</i>	7	0.25
<i>Oryza species</i> ***	1	0.04
<i>O. meridionalis</i>	5	0.18
<b>Total</b>	<b>2757</b>	

\* Lower than wanted due to lack of Chinese selections

\*\* Higher to compensate for lack of WARDA lines

\*\*\* Recently reclassified, previously *O. sativa*, aus?

# Marker System

SSR Panel from Cornell

Subset of those previously used to establish population structure by *Garris et al/2005*

50 SSRs having repeat types:

- 32 dinucleotide

- 11 trinucleotide

- 3 tetranucleotide

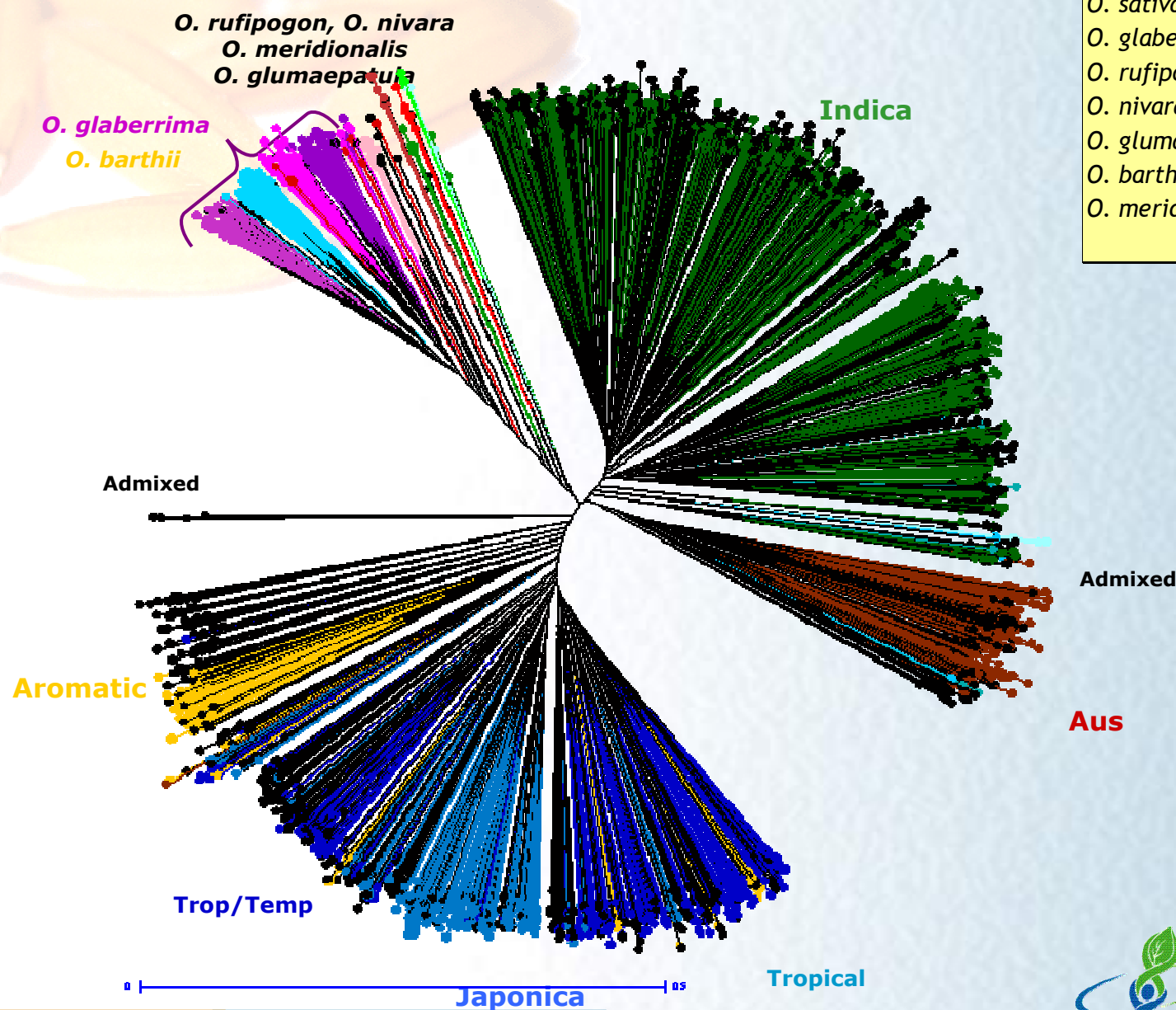
- 4 complex

Distributed across the genome

Optimized for use on automated genotypers

# Population Structure of the GCP Rice Composite Set

50 SSR Loci on 2757 *Oryza* accessions.



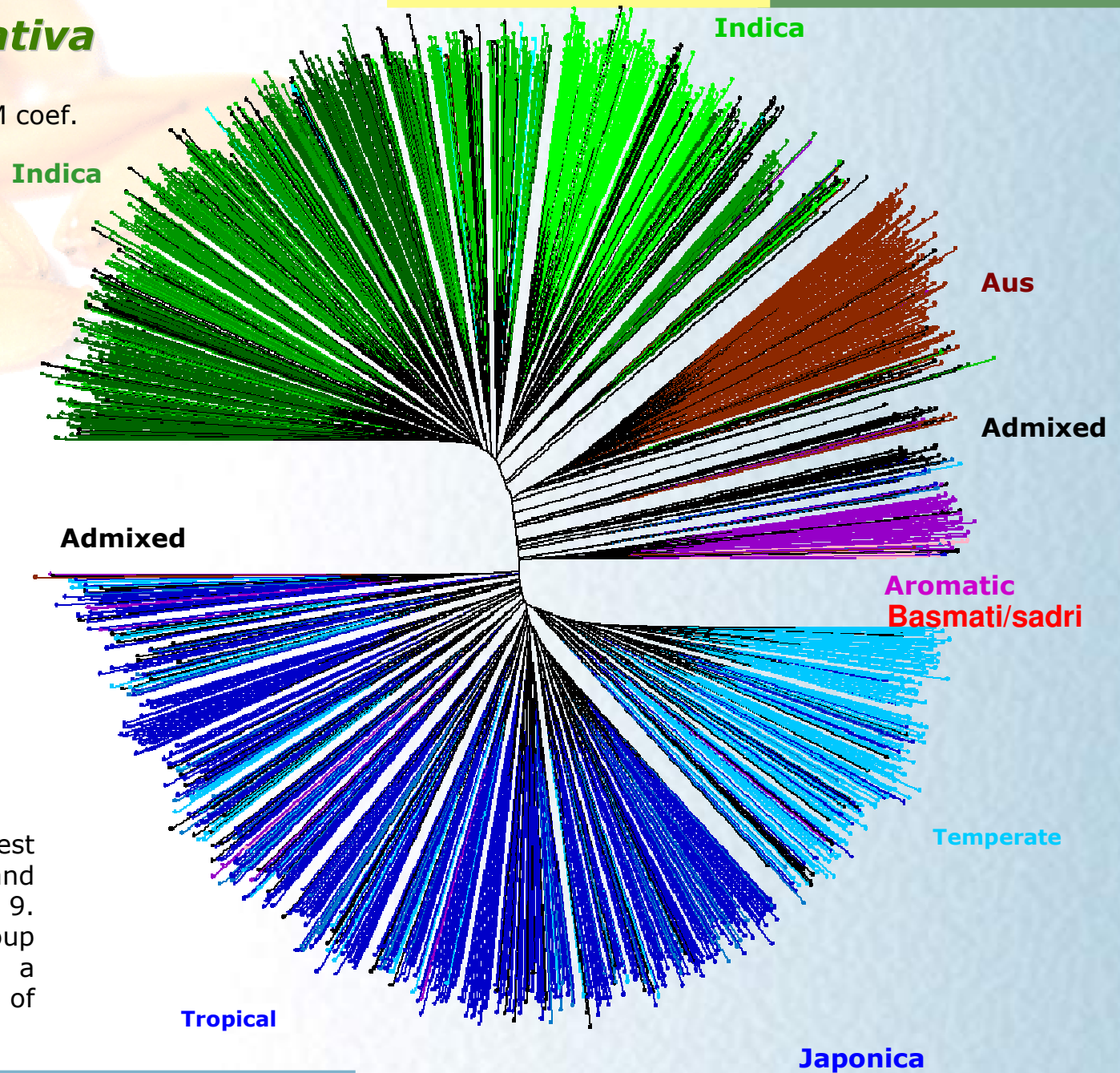
SPECIES	Total No.
<i>O. sativa</i>	2339
<i>O. glaberrima</i>	373
<i>O. rufipogon</i>	14
<i>O. nivara</i>	10
<i>O. glumaepatula</i>	9
<i>O. barthii</i>	5
<i>O. meridionalis</i>	5
	2757

IRRI (60%)  
 Cirad (14%)  
 Warda/  
 Cornell (12%)  
 CIAT (8%)  
 Embrapa (6%)



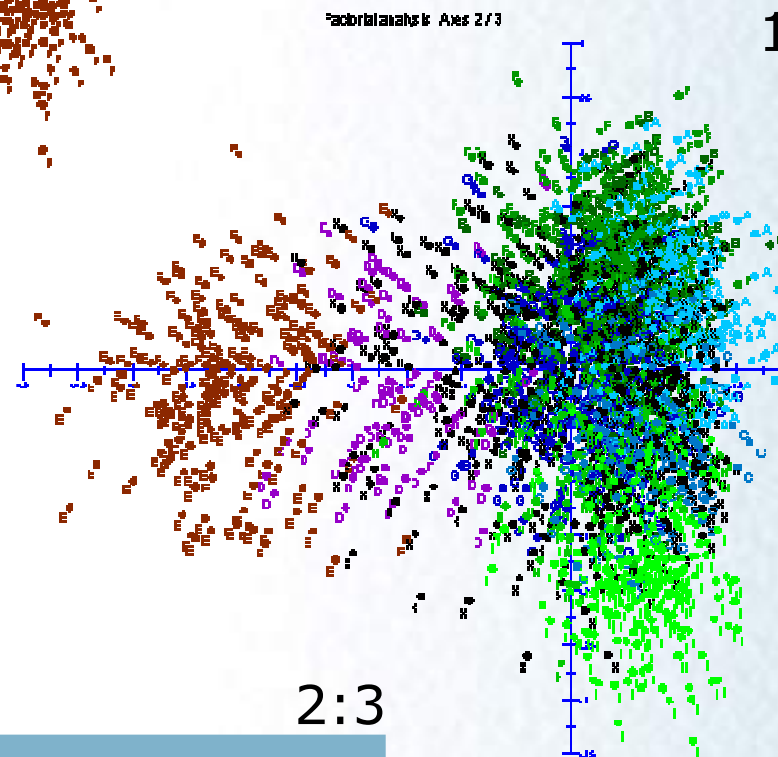
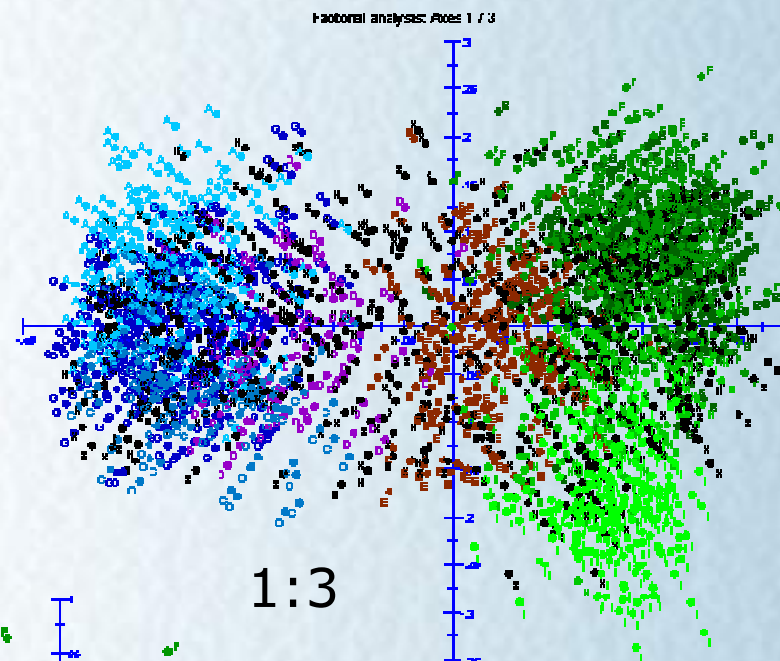
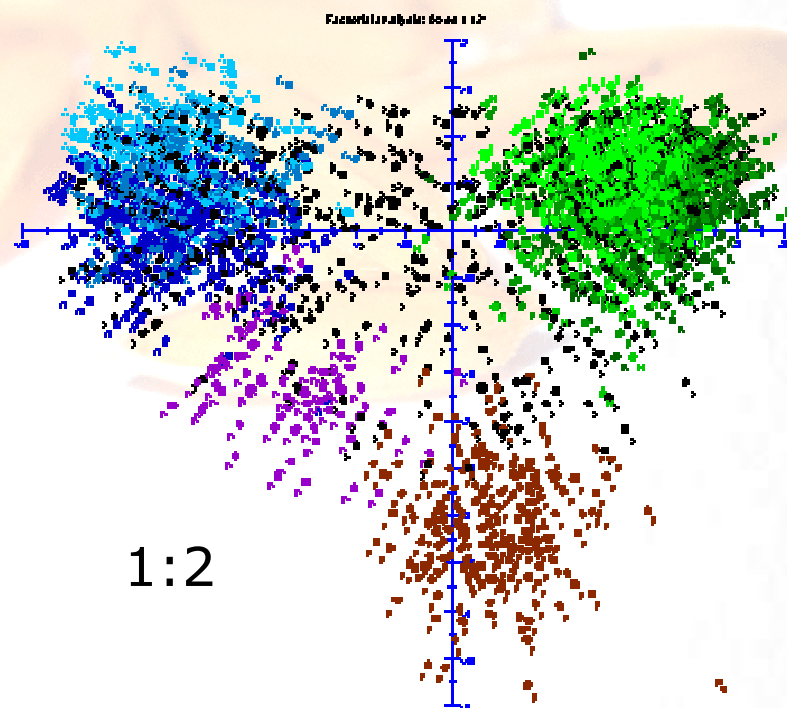
## Structure of *Sativa*

48 Loci on 2339 lines.  
(DARwin5, unwtcd NJ, SM coef.



InStruct indicated that the best DIC value was for  $K = 2$  and the second best was for  $K = 9$ . The coloring reflects the group assignment for  $K = 9$  with a minimum allele frequency of 0.65.

# Factorial Analysis on *Sativa*



48 Loci on 2339 lines.  
(with DARwin 5,  
Simple matching coefficient,  
distance matrix,  
no transformation

# Structure of *Glaberrima*

*sativa*

*glaberrima*

*glaberrima*

*glaberrima*  
*barthii*

*glaberrima*

*glaberrima*

*glaberrima*

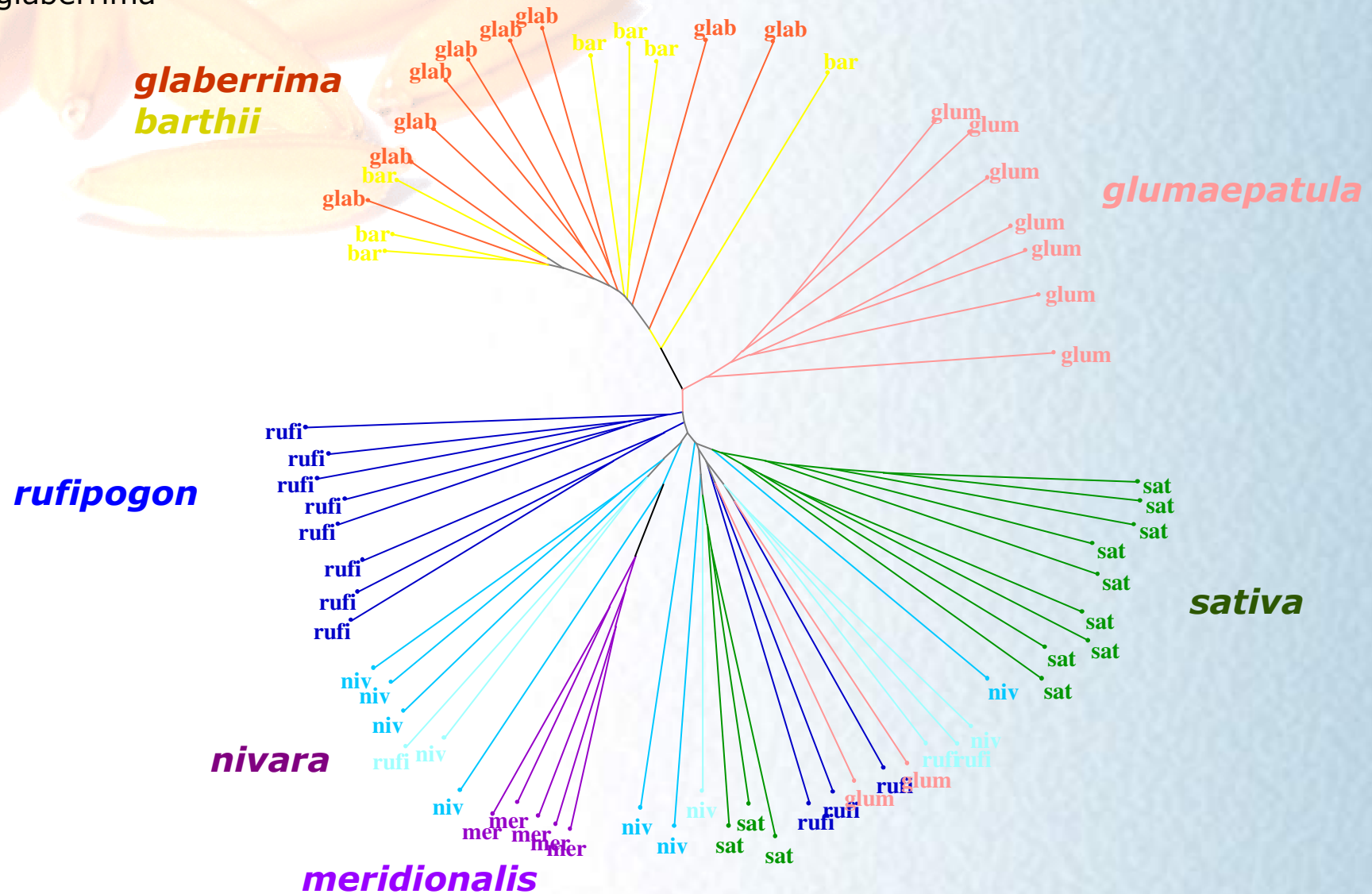
48 Loci on 12 *sativa*,  
373 *glaberrima*,  
and 5 *barthii*

Similar to Semon *et al*, 2005,  
Genetics **169**:1639–1647

0 | 0.2

# Structure of the Wild Relatives

48 Loci on 45 wild,  
12 sativa, and  
9 glaberrima





IR64-21

Swarna

Sadu-cho

Pokkali

SHZ2

MH63

ZS97 B

Aswina

Rayada

FR13A

N22

Dular

Dom-sufid

Moroberekan

Azucena

Cypress

LTH

M 202

Tainung 67

Nipponbare

PNAS

## Genomewide SNP variation reveals relationships among landraces and modern varieties of rice

Kenneth L. McNally<sup>a,1</sup>, Kevin L. Childs<sup>b</sup>, Regina Bohnert<sup>c</sup>, Rebecca M. Davidson<sup>d</sup>, Keyan Zhao<sup>e</sup>, Victor J. Ulat<sup>a</sup>, Georg Zeller<sup>c,f</sup>, Richard M. Clark<sup>f</sup>, Douglas R. Hoen<sup>g</sup>, Thomas E. Bureau<sup>g</sup>, Renee Stokowski<sup>h</sup>, Dennis G. Ballinger<sup>h</sup>, Kelly A. Frazer<sup>h</sup>, David R. Cox<sup>h</sup>, Badri Padhukasahasram<sup>e</sup>, Carlos D. Bustamante<sup>e</sup>, Detlef Weigel<sup>f</sup>, David J. Mackill<sup>a</sup>, Richard M. Bruskewich<sup>a</sup>, Gunnar Rättsch<sup>c</sup>, C. Robin Buell<sup>b</sup>, Hei Leung<sup>a</sup>, and Jan E. Leach<sup>d,1</sup>

[www.pnas.org/cgi/doi/10.1073/pnas.0900992106](http://www.pnas.org/cgi/doi/10.1073/pnas.0900992106)

PNAS | July 28, 2009 | vol. 106 | no. 30 | 12273–12278

IRRI



MICHIGAN STATE  
UNIVERSITY



CORNELL

McGill



IRRI

Klm '09



# OryzaSNP Platform

100.1 Mb of Nippon-bare reference genome

Unique plus some minimally repetitive (<10 hits), 57% coverage of annotated gene models

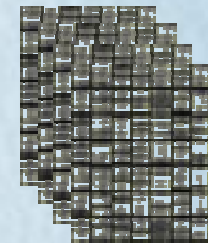
20 diverse varieties chosen for utility, function, diversity from all variety groups

4 temperate japonica, 3 tropical japonica, 1 aromatic,

2 deep-water, 3 aus/boro, 7 indica

HDOA re-sequencing technology (*Perlegen*)

1 chip + 5 wafers (5'x5') for 20 varieties



LR-PCR for target preparation

> 13,582 LR-PCR amplicons for each variety

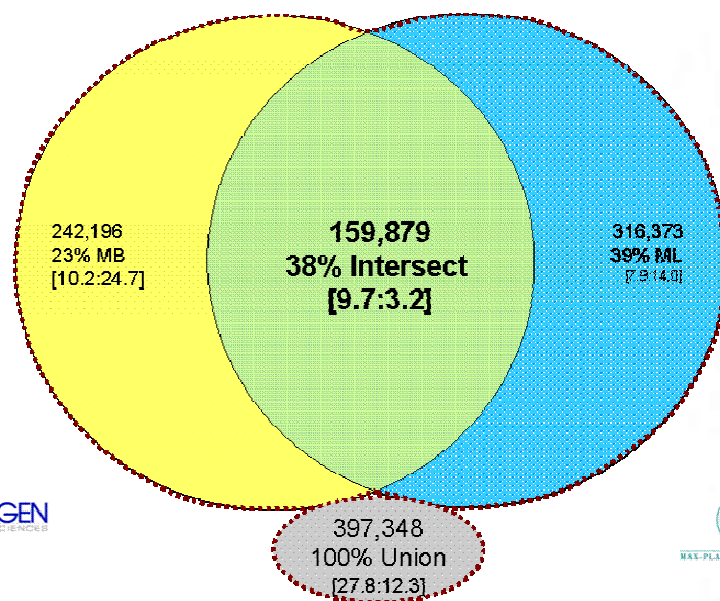
Hybridization, analysis,

Prediction by model-based (*Perlegen*) and Machine learning (MPI-Tubingen) algorithms

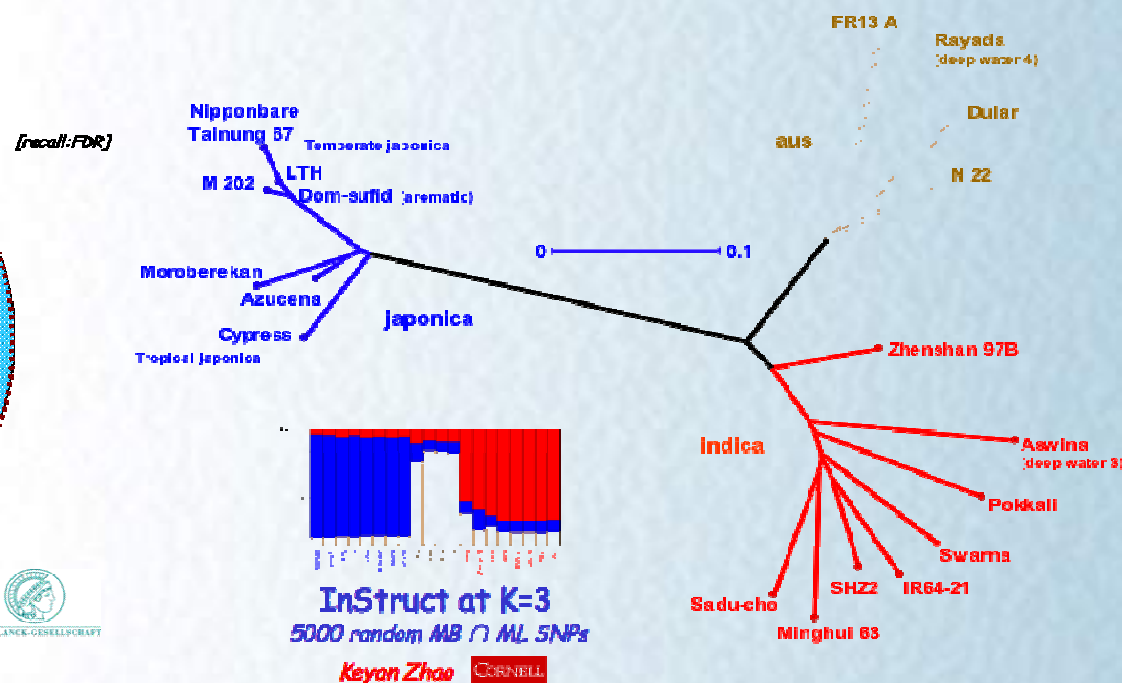
## Model-based (MB) and machine-learning (ML) SNP predictions

Dataset	All SNPs	Non-rep SNPs	Genotypes	Freq $\geq$ 0.15 (%)	Biallelic (%)	Transi/Transv
MB	259721	242,196	1242410	67.2	97.4	1.900
ML	326471	316,373	1349341	53.8	97.7	1.665
MB $\cup$ ML	422244	397,348	1824074	56.9	97.1	1.654
<b>MB <math>\cap</math> ML</b>	<b>162478</b>	<b>159,879</b>	<b>761606</b>	<b>64.9</b>	<b>99.7</b>	<b>2.072</b>

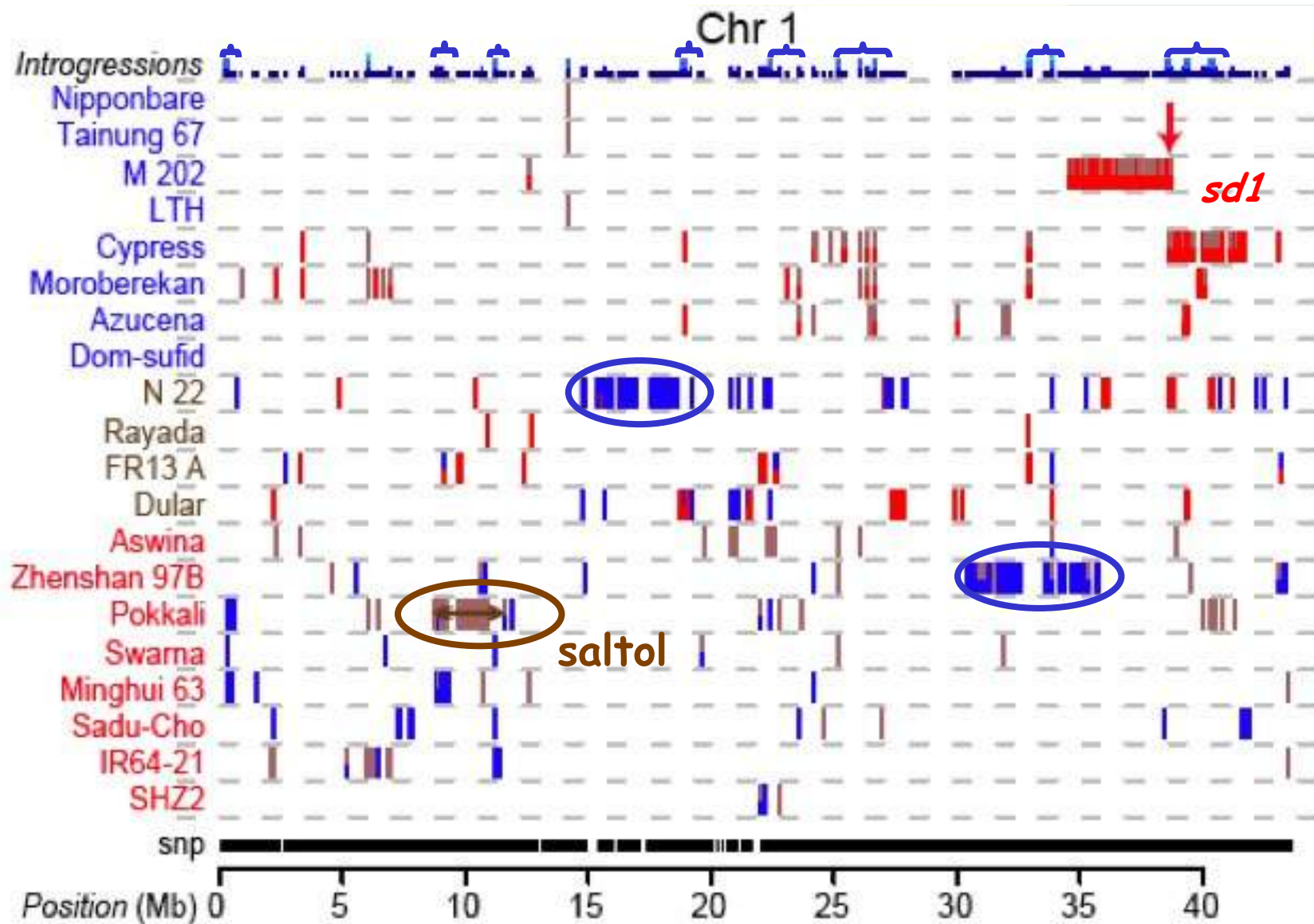
### Complementary methods



1.597 SNPs/kb, 20.7% singletons

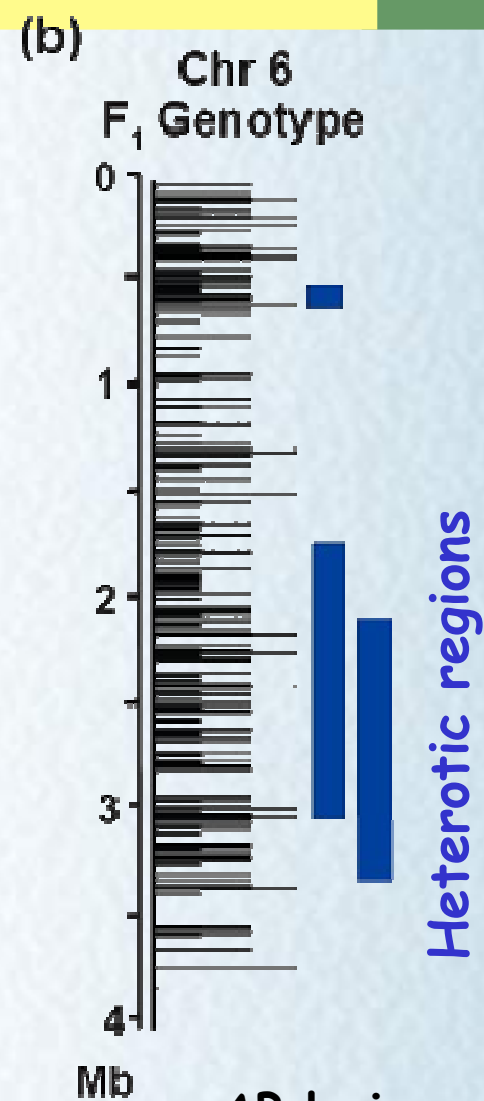
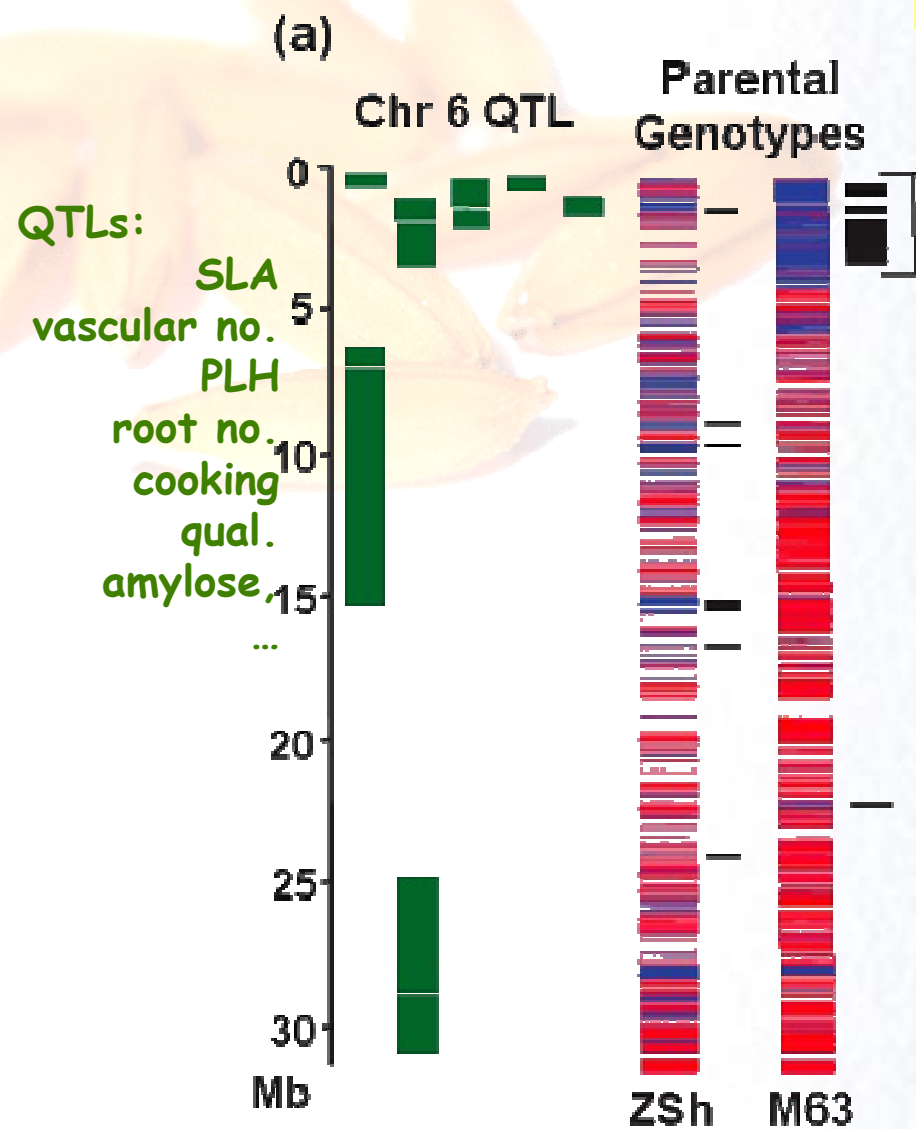


3 Groups: japonica, aus, indica



Regions of shared introgressions cover ~9% of genome,  
Associated with genes/QTLs related to domestication.

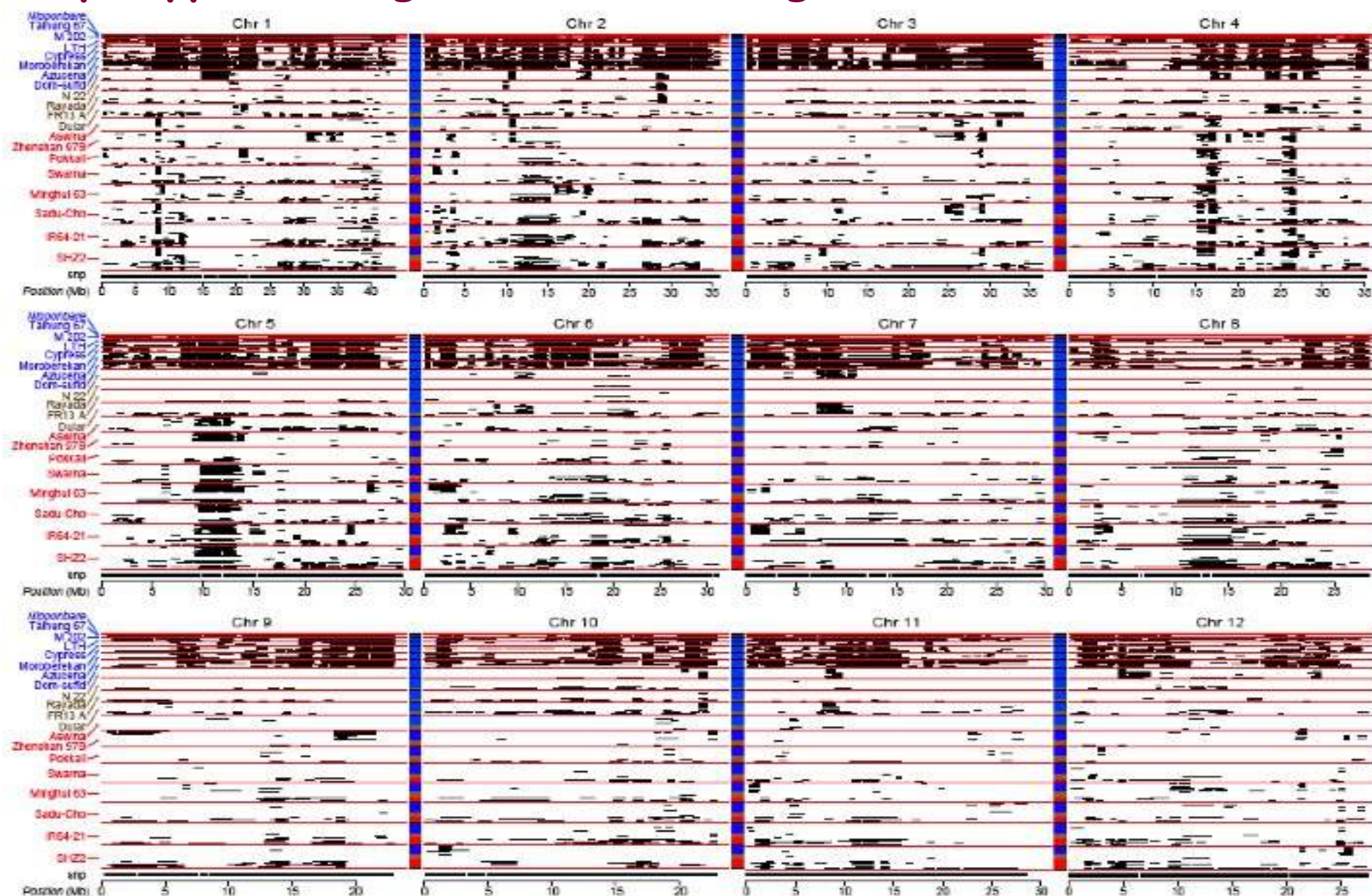




AB loci:

- Short - intergenic
- Med. - UTR/coding/synonymous
- Long - non-synonymous

# Haplotype sharing in 100 kb sliding window (MB ML)





# Links to OryzaSNP Database Search Forms - Mozilla Firefox

File Edit View History Bookmarks Tools Help

## Rice SNP Browser - TIGR Pseudomolecules: Chr01:4739360..4759359 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.oryzasnp.org/cgi-bin/gbrowse/osa\_snp\_tigr/#search

Links to OryzaSNP Database Sea...

Rice SNP Browser - TIGR Pse...

### OryzaSNP Genome Browser - TIGR Pseudomolecules

[OryzaSNP@MSU Home](#)

Showing 20 kbp from Chr01, positions 4,739,360 to 4,759,359

#### Instructions

**Searching:** Search using a sequence name, gene name, locus, or other landmark. The wildcard character \* is allowed.

**Navigation:** Click one of the rulers to center on a location, or click and drag to select a region. Use the Scroll/Zoom buttons to change magnification and position.

**Examples:** Chr01, Chr05:1475000..1490000, LOC\_Os10g28410, LOC\_Os02g02400.3.

[\[Bookmark this\]](#) [\[Upload your own data\]](#) [\[Hide banner\]](#) [\[Share these tracks\]](#) [\[Link to image\]](#) [\[High-res image\]](#) [\[Help\]](#) [\[Reset\]](#)

#### Search

Landmark or Region:

Chr01:4739360..4759359

Search

Reports & Analysis:

Download Decorated FASTA File

Configure...

Go

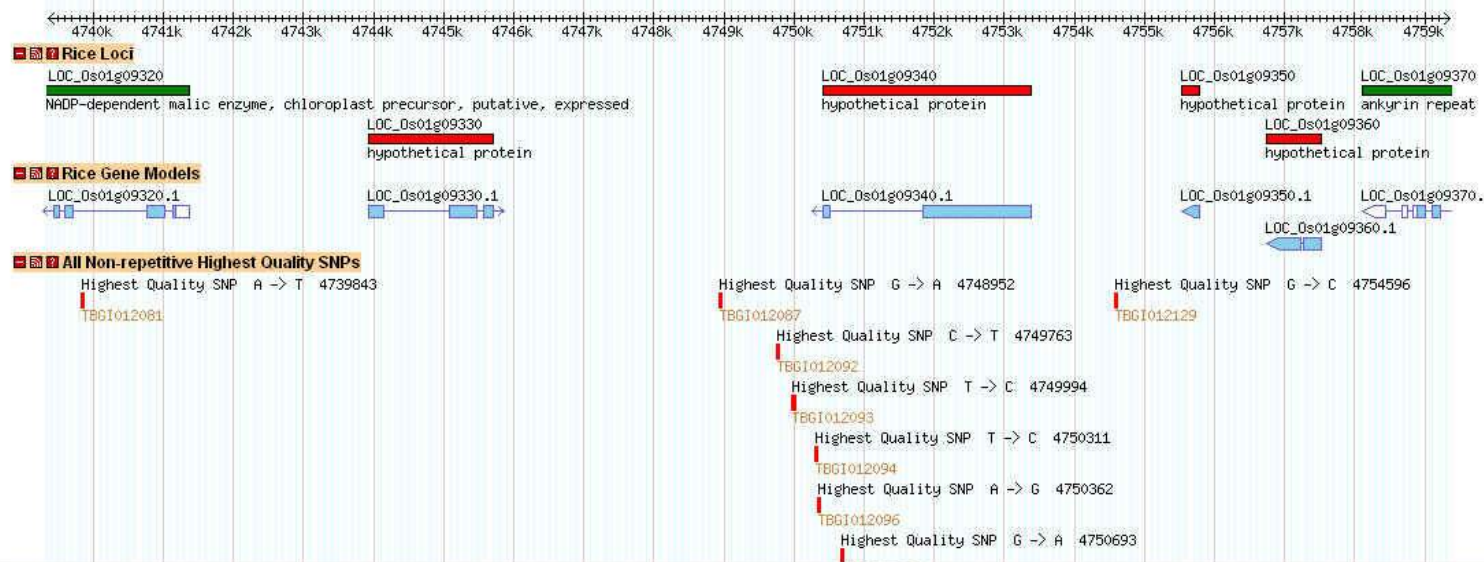
Data Source

Rice SNP Browser - TIGR Pseudomolecules

#### Overview

OryzaSNPdb r2

#### Details



# OryzaSNP

## Summary

- 160,000 HQ SNPs
  - 99.7% biallelic, 2 transitions/transversion; 0.8 million genotypes
- Three groups: japonica, aus, indica.
- LD ( $r^2$ ) of ~200 kb for indica, 500 kb for japonica,
  - High estimates possibly due to sampling/population structure.
- SNP categories vary across gene families.
- Transition/transversion ratios differ at major effect SNPs.
- Regions of shared haplotypes from natural or imposed selection.
  - Pattern of introgressions significantly different from random.
  - 9% of genome covered by clustered introgressions with highly significant associations with grain quality, panicle and yield QTLs
- 2nd version of OryzaSNP database open to public.
- > 40K of MB $\cap$ ML on 44K Affie array & 1536 on Illumina Goldengate array at Cornell have verified calls
- Application of subsets ongoing at IRRI and Cornell.



**Known  
edges  
elements**

**Ma. Elizabeth Naredo**

**Mario Rodriguez**

*Millicent Sanciangco*

*Reneeliza Melgar*

*Jeffrey Detras*

*Merlyn Rodonzo*

*Monique Barile*

*Jyoti Kaur*

*Olivia Listanco*

*Chay Bounphanousay*

*Don Pabale*

*Anselme Fournier*

**Renato Reano**

**Flora de Guzman**

**Reniela Prantilla**

**Sheila Mae Quilloy**

**Janice Rayco**

*Alicia Perez*

*Reflinur Basyirin*

*Genelou Atienza*

*Vanica Lacorte*

*Gilbert Mamiit*

*Hehe Wang*

*Michael Gamalinda*

*Eric Canicosa*

*Catherine Aquino*

*Zoe Lawson*

**Ma. Celeste Banaticla**

**Adelaida Alcantara**

**Belinda Caspillo**

*Minerva Macatangay*

*Rosalyn Angeles*

*Rhodesia Manzano*

*Maricris Zaidem*

*Leanilyn Castanar*

*Htay Htay Aung*

***\*Amita Juliano, RIP\****

*Jeffe Cadion*

*Xie Yong*

**Ma. Socorro Almazan**

**N.R. Sackville Hamilton**

**Teresita Santos**

**GCP SP1  
EcoTillers**

**CIRAD:**

L. Benoit

**CIAT:**

A. Garavito

**EMBRAPA:**

T. Borba

**WARDA:**

M. Cissoko

**Cornell:**

A. Famoso

**CAAS:**

L.Z. Han

R. Rivallan

M. Lorieux

R.V. Brondani

M.-N. Ndjiondjop

S.R. McCouch

Y.Y. Zhang

B. Courtois

C.P. Martinez

C. Brondani

C. Billot

**IRRI**



**The World Bank Group**



Bundesministerium für  
wirtschaftliche Zusammenarbeit  
und Entwicklung



OryzaSNP

Hei Leung  
R. Bruskiewich  
M. Macatangay  
M. Rodriguez  
E. Nelzo

Dave Mackill  
V.J. Ulat  
S.H. Zhang  
R. Reano  
R. Mauleon

K. McNally  
V. Pamplona  
Bin Liu  
J. Detras

IRRI



Jan Leach  
R. Davidson



C. Robin Buell  
Kevin Childs



Detlef Weigel  
Regina Bohnert

Richard Clark  
Gunnar Rättsch

Georg Zeller  
Gabriele Schweikert



Renee Stokowski  
Heng Tao  
Geoff Nielsen  
David Cox  
Dennis Ballinger

Julie Montgomery  
Cindy Chen  
Robert Gupta  
Eric Peacock  
Kelly Frazer

Laura Stuve  
Julee Kwon  
Matt Morenzoni  
Diana Starr  
Keiko Greenberg



T. Bureau  
D. Hoen



C. Bustamante  
Keyan Zhao  
B. Padhukasahasram

IRRI

