

# IRRI

INTERNATIONAL RICE RESEARCH INSTITUTE

## **3K genome and International Rice Informatics Consortium, Concept and Operation**

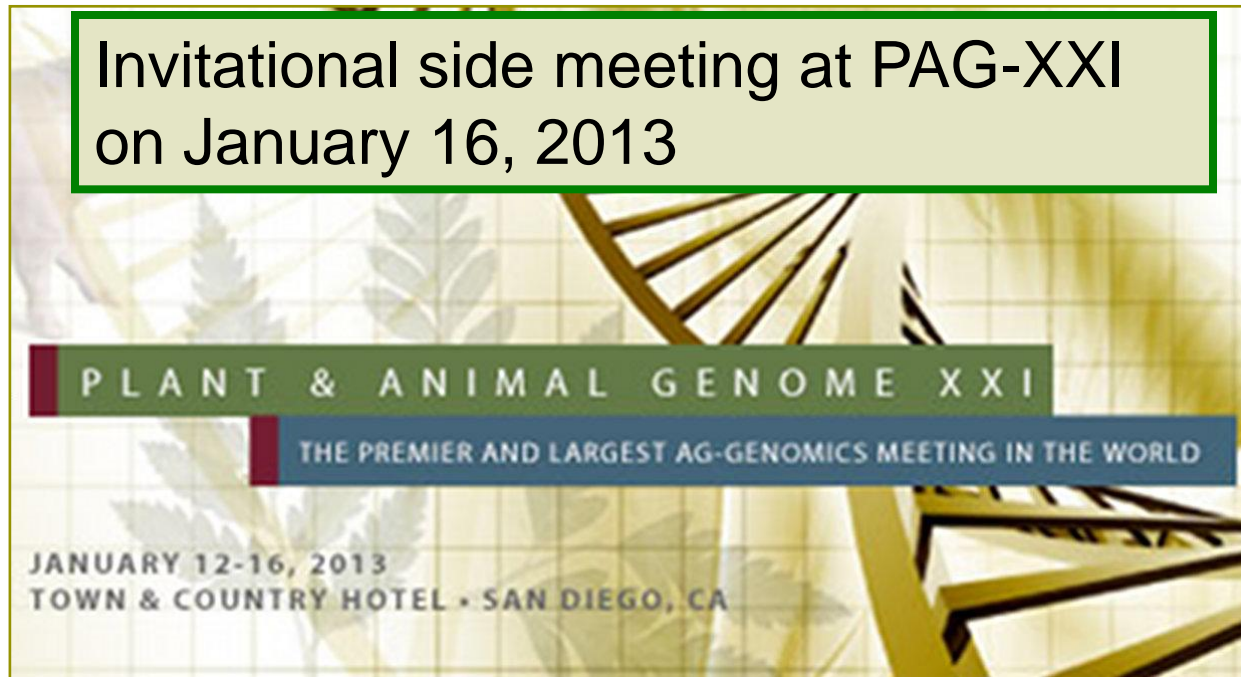
**Nickolai Alexandrov**

T.T.Chang Genetic Resources Center



# IRIC Kick-off Meeting

Invitational side meeting at PAG-XXI  
on January 16, 2013



56 people from 38 organizations

**Provide access to well organized information about rice,  
facilitate communication and collaboration for rice community  
having germplasm diversity as a focal entry point**

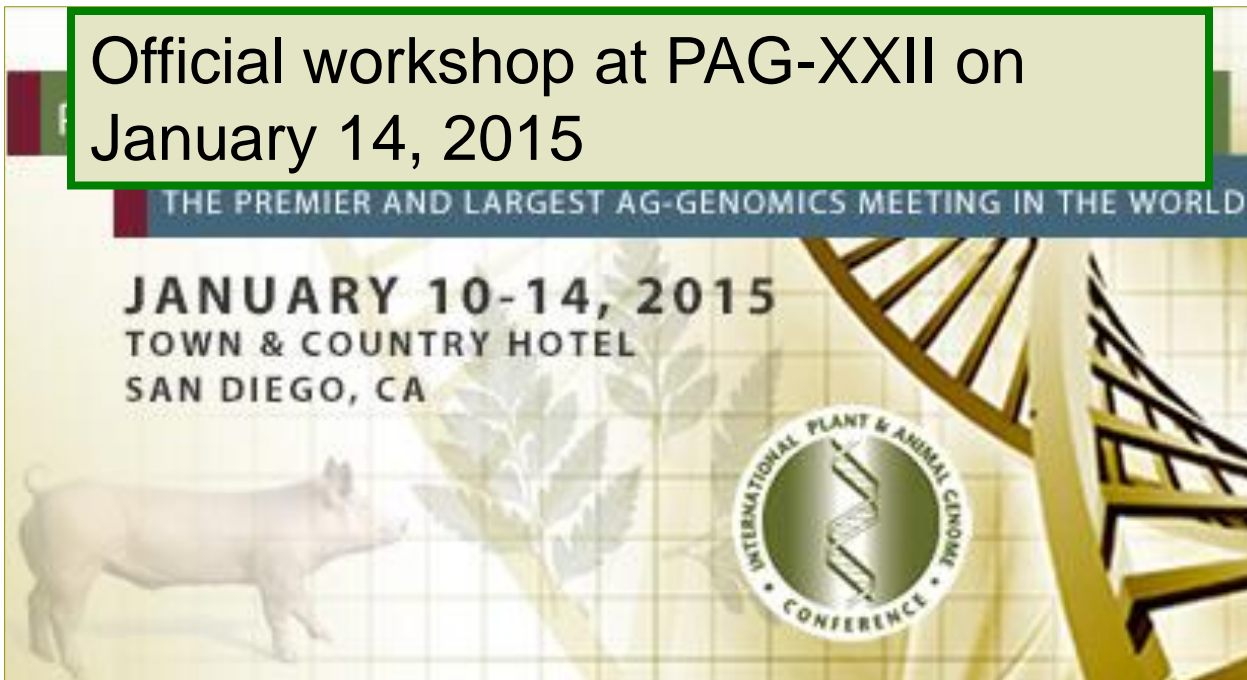


# IRIC PAG Workshop

Official workshop at PAG-XXII on  
January 14, 2015

THE PREMIER AND LARGEST AG-GENOMICS MEETING IN THE WORLD

JANUARY 10-14, 2015  
TOWN & COUNTRY HOTEL  
SAN DIEGO, CA



## IRIC progress

*Pierre Larmande, IRD, UMR DIADE, Institut de Biologie*

*Rod A. Wing, University of Arizona, AGI*

*Ruaraidh Hamilton, International Rice Research Institute*



# IRIC objectives

1. Organize available genotyping, phenotyping, expression and other information about rice into a linked, consistent and reliable source of knowledge
2. Provide user-friendly portal to browse, search and analyze the data
3. Support information sharing, public awareness and capacity building



# The 3000 Rice Genome Project

华大基因  
BGI

IRRI  
GRiSP



Rice  
Science  
for a Better  
World



# Our Project Team & Donors

## BGI

Shuai-Shuai Tai  
Xin Liu  
Jun Li  
Guo-Jie Zhang  
Bo Wang  
Xun Xu  
**Gengyun Zhang**

## IRRI

Ma. Elizabeth B. Naredo  
Sheila M. Q. Mercado  
Myla C. Rellosa  
Renato A. Reaño  
Grace Lee S. Capilit  
Flora C. de Guzman  
N. R. Sackville Hamilton  
Jauhar Ali  
Ramil P. Mauleon  
Nickolai N. Alexandrov  
Hei Leung  
**Kenneth L. McNally**

## CAAS

Wen-Sheng Wang  
Yong-Ming Gao  
Xiu-Qing Zhao  
Jian-Long Xu  
Fan Zhang  
Yong-Li Zhou  
Bin-Ying Fu  
**Zhikang Li**

华大基因  
**BGI**

**GRiSP & CAAS**



BILL & MELINDA  
GATES foundation 

**MOST: 科学技术部**

**IRRI**

[Top](#)[Abstract](#)[Data descript...](#)[Availability and requirements](#)[Discussion](#)[Availability ...](#)[The 3,000 ric...](#)[Abbreviations](#)[Competing interests](#)[Authors' contributions](#)[Acknowledgements](#)[References](#)

1,000,000  
antibodies  
ELISA kits  
proteins  
and more

**Data Note****Highly accessed****Open Access**

## The 3,000 rice genomes project

### The 3,000 rice genomes project<sup>†</sup>

Correspondence: The 3,000 rice genomes project

▼ Author Affiliations

<sup>†</sup> Equal contributors

Institute of Crop Sciences/National Key Facilities for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, 12 S. Zhong-Guan-Cun St, Beijing 100081, China

BGI, Bei Shan Industrial Zone, Yantian District, Shenzhen 518083, China

International Rice Research Institute, DAPO 7777, Metro Manila 1301, Philippines

*GigaScience* 2014, **3**:7 doi:10.1186/2047-217X-3-7

Correspondence: lizhikang@caas.cn; zhanggengyun@genomics.cn; k.mcnally@irri.org; <sup>†</sup> The list of project participants and their affiliations is given at the end of this paper.

The electronic version of this article is the complete one and can be found online at: <http://www.gigasiencejournal.com/content/3/1/7>

Received: 22 January 2014

Accepted: 12 May 2014

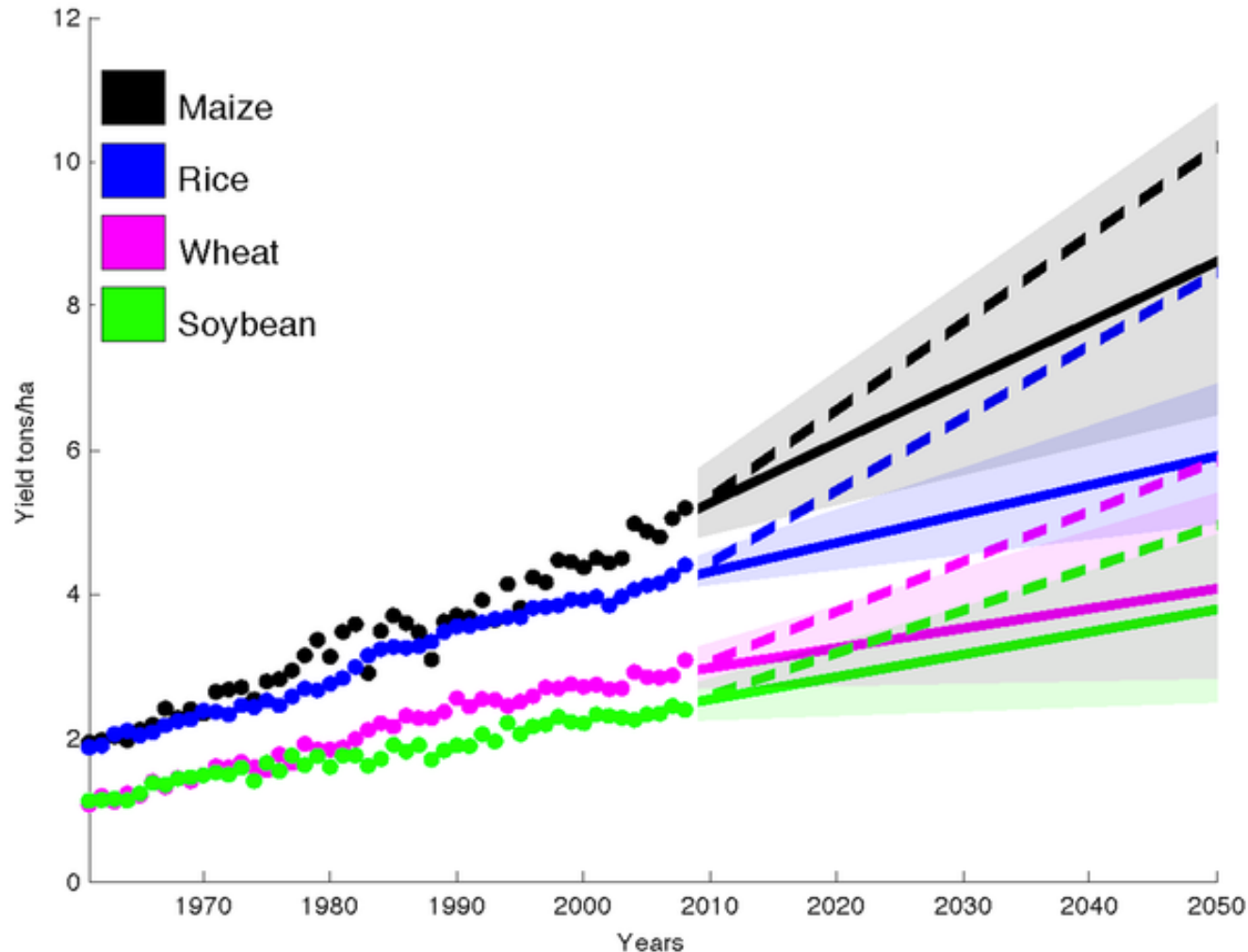
Published: 28 May 2014

© 2014 3K RGP; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>)

**GigaScience**  
Volume 3**Viewing options**[Abstract](#)**Full text**[PDF \(630KB\)](#)[ePUB \(113KB\)](#)[Additional files](#)**Associated material**[PubMed record](#)[Article metrics](#)[Readers' comments](#)[Pre-publication history](#)**Related literature**[Cited by](#)[Google blog search](#)[Other articles by authors](#)[► on Google Scholar](#)[► on PubMed](#)[Related articles/pages](#)[on Google](#)[on Google Scholar](#)[on PubMed](#)**Tools**[Download references](#)[Download XML](#)[Email to a friend](#)[Order reprints](#)

# “Terrifying Chart”



Ray DK, Mueller ND, West PC, Foley JA (2013) Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLoS ONE 8(6): e66428. doi:10.1371/journal.pone.0066428

<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0066428>



# Rice and Health

India, with 41 million obese people, ranks third after the US and China in having the highest number of overweight people in the world. Together, India and China represent 15% of the world's obese population.

**The Lancet May 29, 2014**

Replace white rice with brown to cut obesity and diabetes

Dr. V. Mohan

[World Diabetes Conference 2013](#)

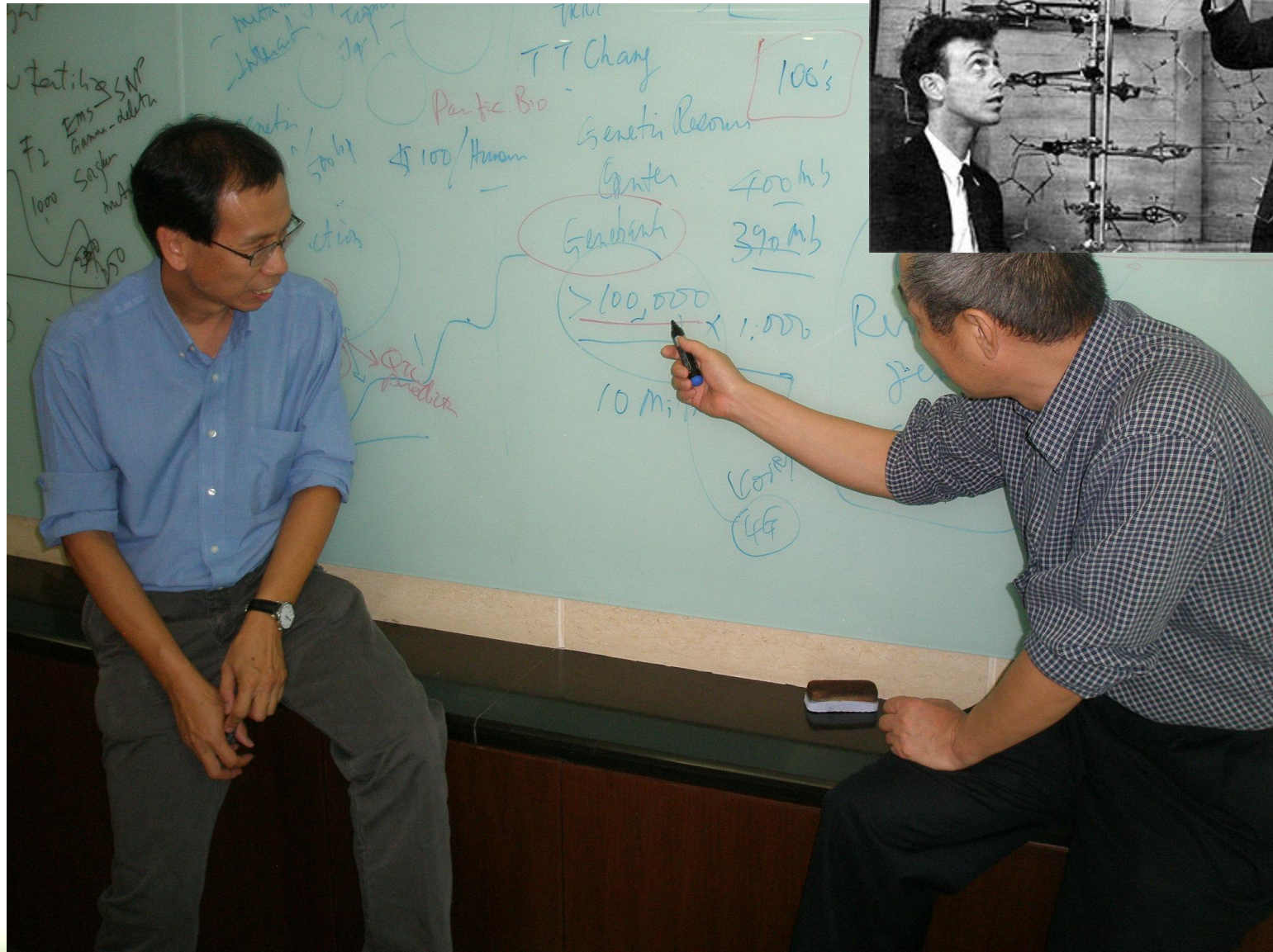
IRRI's mission: make rice healthier

# How can knowledge of 3,000 rice genomes help?

1. By discovering new locus-trait associations
2. By finding causative genome variations
3. By introducing new varieties to breeding programs



# Hei Leung (IRRI) and Wang Jian (BGI) Share



# IRIC deliverables for 2014

- ✓ Initial version of IRIC portal on-line
- ✓ Advisory Board is formed
- ✓ IRIC guidelines documents are finalized
- ✓ Established communication with IRIC members and supporters



# IRIC Progress for 2014

- IRRI hired an AXA Chair on Genome Biology and Evolutionary Genomics, two postdoctoral fellows and two senior software developers
- IRIC Guidelines for Partnership has been finalized. Information page is now available at <http://iric.irri.org/>
- IRRI purchased hardware, licensed Oracle database, has been granted access to iPlant computational infrastructures
- Data from the 3,000 rice genome project (CAAS, BGI, IRRI): QCed, aligned with Nipponbare 1.0 assembly, SNPs called, admixture analysis done, phylogenetic tree constructed. The data has been published at GigaScience on May 28, 2014
- Database schema has been populated with SNPs from the 3,000 project data, 700K chip data, other data
- The first version of IRIC portal (SNP-Seek database) is online <http://oryzasnp.org/iric-portal/> and is published in 2015 NAR database issue
- Collaborations are being established



# IRIC Advisory Committee

Mario Caccamo, Director, The Genome Analysis Center

Dave Edwards, Professor, The University of Western Australia

Harish Gandhi, Head – Rice Genetics and Traits Project, Syngenta

Hei Leung, Principal Scientist – Genetics, Program Leader – Genetic Diversity, IRRI

Takashi Matsumoto, Director – Agronomics Research Center, NIAS

Susan McCouch, Professor, Cornell University

Debashis Rana, Team Leader – Rice Genetics, Bayer

Rod Wing, Arizona Genomics Institute, University of Arizona

Gengyun Zhang, Vice President, Beijing Genomics Institute



# IRIC Information page

<http://iric.irri.org/>

IRRI

International Rice Informatics Consortium

Search this site

Home **IRIC** Members Projects News Resources Contact us



## Navigation

[About IRIC](#)  
[Members](#)  
[Projects](#)  
[News](#)  
[Resources](#)  
[Contact us](#)  
[Sitemap](#)

## Important dates

**14 January 2015** [IRIC Workshop at PAG XXIII](#) (Pacific Salon 2)

## Interact with us



IRRI



# IRIC LinkedIn group

[https://www.linkedin.com/groups?home=&gid=6722493&trk=groups\\_most\\_recent-h-log](https://www.linkedin.com/groups?home=&gid=6722493&trk=groups_most_recent-h-log)

**IRIC** (International Rice Informatics Consortium) 74 members Member

**IRIC** International Rice Informatics Consortium

Discussions Promotions Jobs Members Search Manage

Members of this Group

Jay Santos  
Specialist - Scientific Computing Support at International Rice Research Institute  
Unfollow Jay See all members

Your group contribution level

Start by commenting in a discussion. Group participants get 4x the number of profile views.

Getting Started

Ads You May Be Interested In

**Quantitative Research**  
Collect Critical Data Your Business Needs. \$75 Coupon for 1st Survey!

Start a discussion or share something with the group...

Popular Recent

**Community feedback on IRIC goals at Rice Functional Genomics Symposium**  
Nickolai Alexandrov  
Senior Scientist, Bioinformatics at International Rice Research Institute

Below is a summary of the feedback we received at the IRIC workshop at the ISFRG-12 meeting.  
Feel free to add more tasks or emphasize priorities.



# IRIC portal: target users

## **Breeders**

find germplasm with blast resistance

find germplasm with good Pi5 allele

## **Molecular biologists**

show genes associated with grain width

show all SNPs in Hd1 gene in given lines

## **Bioinformatics scientists**

give me all SNP calls in tropical japonica

API



# Initial sets of the IRIC data

## IRRI Genebank genetic stocks for trait discovery research

>25,000 purified genetic stocks, RILs, CSSLs, MAGIC lines

Annotated with standard unique names, origins, and identifiers

Purified seed of all lines can be requested via IRRI seed request forms

## Genotypic data

>10M SNPs from 3000 rice genomes (BGI/CAAS/IRRI)

700K SNP chip for ~1800 lines (Cornell/IRRI)

44K SNP chip for ~400 lines (Cornell/IRRI)

All SNPs will be mapped into annotated genome assemblies

## Phenotypic data

~60 morphological traits observed for genetic stock lines from the IRRI Genebank

~30 traits for ~400 lines genotyped with 44K SNPs

~60 traits for ~350 indica lines genotyped with 700K chip

~25 traits for ~300 tropical japonica and ~250 aus lines genotyped with 700K chip

grain quality traits for ~1600 lines genotyped with 700K chip

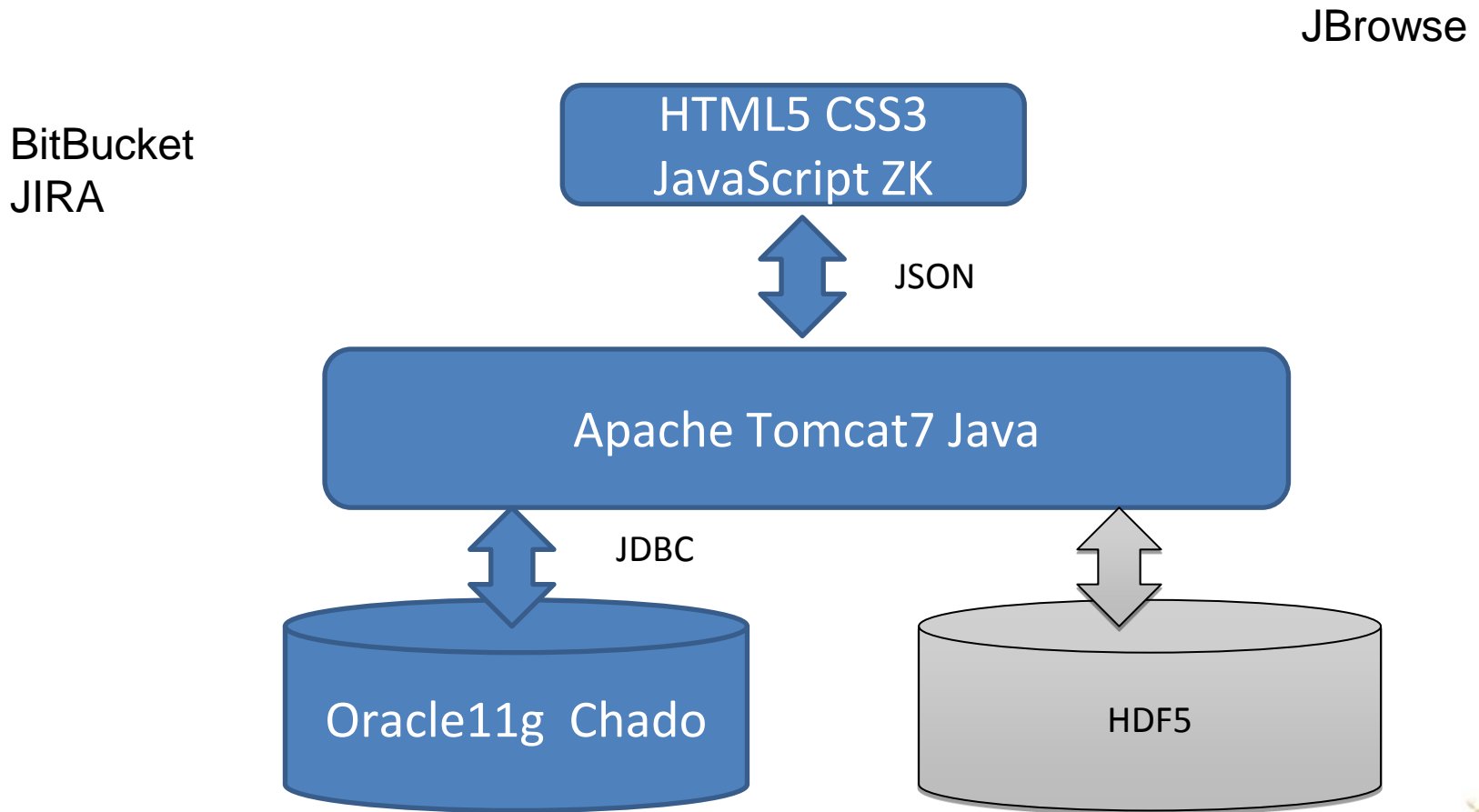
## Gene expression data

Basic developmental and selected stress experiments from NIAS, Japan

RNA-Seq data from C4 consortium



# IRIC portal: architecture



# Developers framework

Oracle: capable to store and quickly query 60B SNPs

JBrowse: interactive, easily customizable cool browser

ZK: enables smooth navigation in 1Mx1M table

HDF5: amazingly quick data retrieval

[https://bitbucket.org/irridev/iric\\_portal](https://bitbucket.org/irridev/iric_portal)

<https://riceinfo.atlassian.net/browse/IRIC-220?filter=-4&jql=project%20%3D%20IRIC%20>

Providing RESTful web APIs and web services for communications with other systems



# Existing rice portals

MSU Rice Genome Annotation Project  
<http://rice.plantbiology.msu.edu/>



Rice Genome Annotation Project  
Funded by the NSF



RAP-DB <http://rapdb.dna.affrc.go.jp/>



BGI-RIS Rice Information System  
<http://rice.genomics.org.cn/rice/index2.jsp>



**RIS<sup>e</sup>** -Rice Information Syst<sup>em</sup>

PlantGDB  
<http://www.plantgdb.org/OsGDB/>

**PlantGDB**



Gramene  
<http://www.gramene.org/>



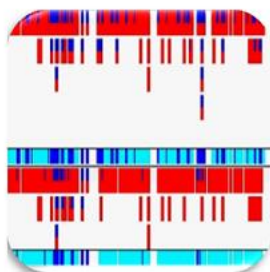
# Initial version of IRIC Portal



## Rice SNP-Seek Database

[Home](#) [Search](#) [Browse](#) [My Lists](#) [Help](#) [Contact Us](#)

This site provides Genotype, Phenotype, and Variety Information for rice (*Oryza sativa* L.). SNP genotyping data (called against [Nipponbare reference Os-Nipponbare-Reference-IRGSP-1.0](#)) came from [3,000 Rice Genomes Project](#). Phenotype and passport data for the 3,000 rice varieties came from the International Rice Information System (IRIS). We are a part of an ongoing effort by the International Rice Informatics Consortium (IRIC) to centralize information access to rice research data and provide computational tools to facilitate rice improvement via discovery of new gene-trait associations and accelerated breeding.



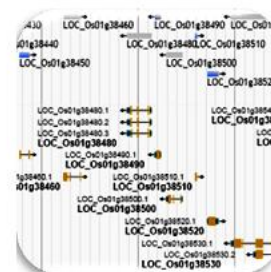
### Genotypes

Query for SNPs from the 3000 genome project



### Varieties

Query for Variety passport and phenotypes



### JBrowse

Rice Genome Browser



### Help

Help and documentation

# SNP-Seek database of SNPs derived from 3000 rice genomes

Nickolai Alexandrov<sup>1,\*†</sup>, Shuaishuai Tai<sup>2,†</sup>, Wensheng Wang<sup>3,†</sup>, Locedie Mansueto<sup>1</sup>, Kevin Palis<sup>1</sup>, Roven Rommel Fuentes<sup>1</sup>, Victor Jun Ulat<sup>1</sup>, Dmytro Chebotarov<sup>1</sup>, Gengyun Zhang<sup>2,\*</sup>, Zhikang Li<sup>3,\*</sup>, Ramil Mauleon<sup>1</sup>, Ruairaidh Sackville Hamilton<sup>1</sup> and Kenneth L. McNally<sup>1</sup>

<sup>1</sup>T.T.Chang Genetic Resources Center, IRRI, Los Baños, Laguna 4031, Philippines, <sup>2</sup>BGI, Shenzhen 518083, China and <sup>3</sup>CAAS, Beijing 100081, China

Received September 08, 2014; Revised October 10, 2014; Accepted October 10, 2014

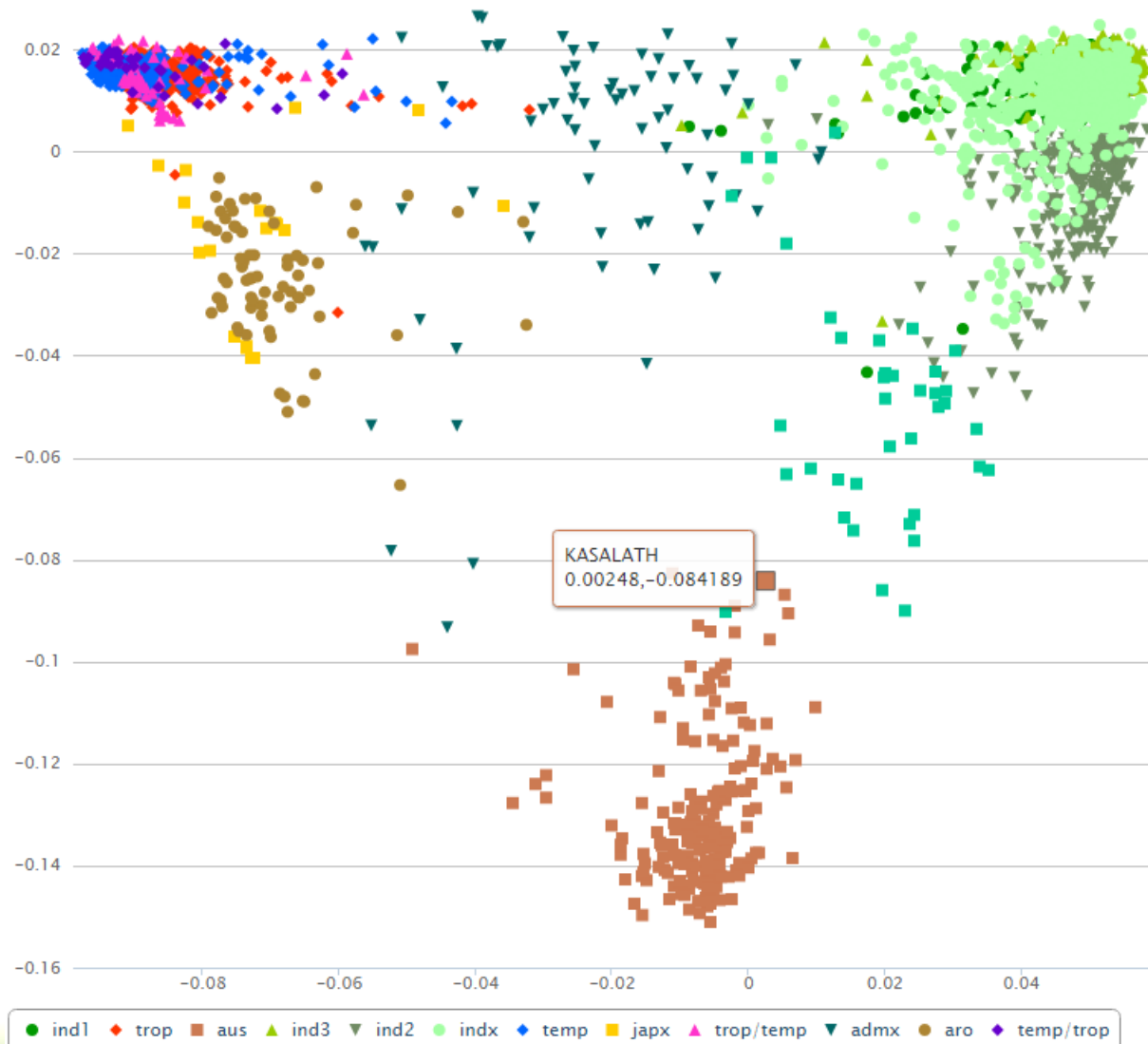
## ABSTRACT

We have identified about 20 million rice SNPs by aligning reads from the 3000 rice genomes project with the Nipponbare genome. The SNPs and allele information are organized into a SNP-Seek system (<http://www.oryzasnp.org/iric-portal/>), which consists of Oracle database having a total number of rows with SNP genotypes close to 60 billion (20 M SNPs × 3 K rice lines) and web interface for convenient querying. The database allows quick retrieving of SNP alleles for all varieties in a given genome

information, is an important step toward establishing gene-trait associations, building predictive models and applying these models in the breeding process. The 3000 rice genome project (6) produced millions of genomic reads for a diverse set of rice varieties. SNP-Seek database is designed to provide a user-friendly access to the single nucleotide polymorphisms, or SNPs, identified from this data. Short, 83 bp pair-ended Illumina reads were aligned using the BWA program (7) to the Nipponbare temperate japonica genome assembly (8), resulting in average of 14× coverage of rice genome among all the varieties. SNP calls were made using GATK pipeline (9) as described in (6).

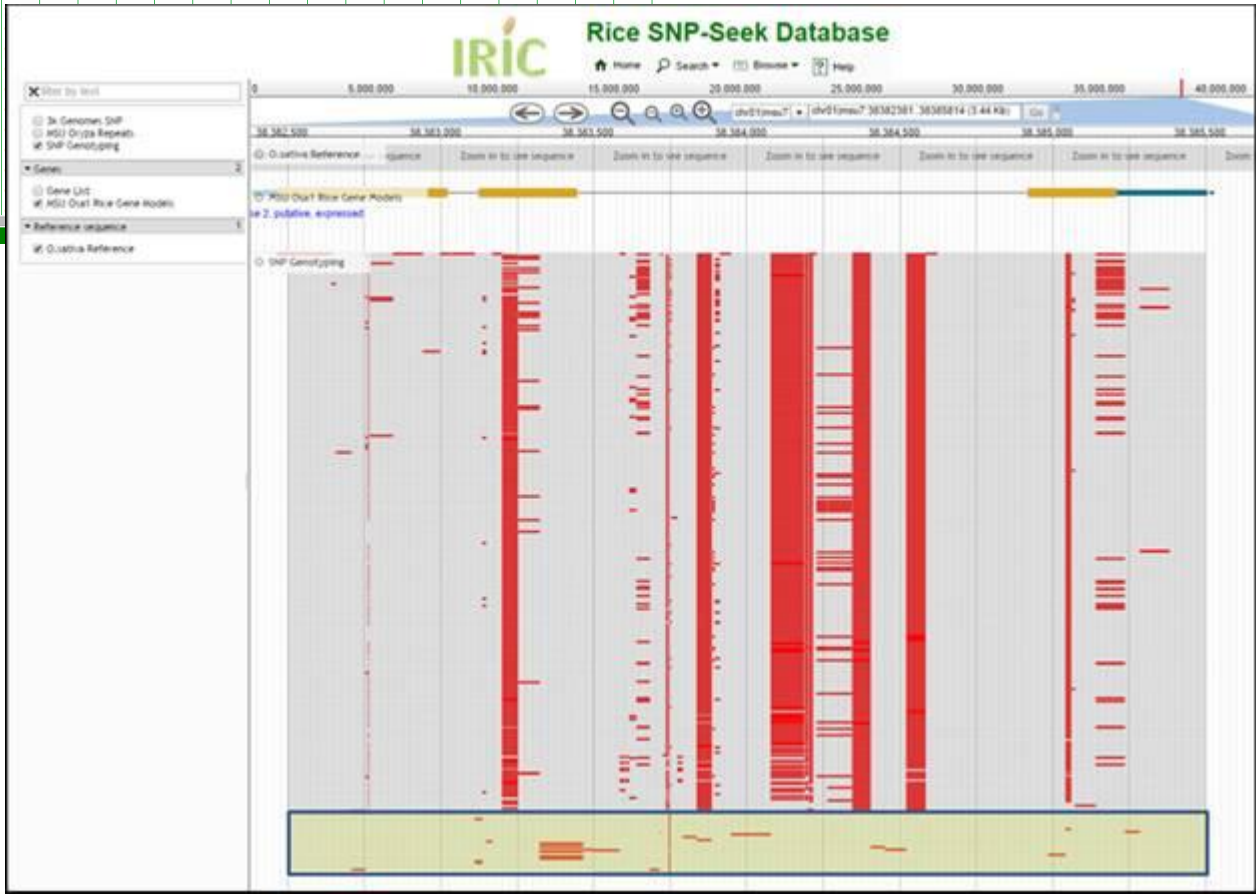


# MDS plot for 3000 genomes





VARPOS	MS	38382543	38382546	38382547	38382575	38382641	38382644	38382671	38382746	38382762	38382764	38382766	38382767	38382768	38382921	38382964	38383032	38383037	38383039	38383095	38383131	38383144	38383157	38383182	38383221	38383286	38383363	38383570	38383603	38383607
REFERENCE	C	C	C	G	C	C	G	G	C	A	C	A	A	C	A	G	G	T	T	T	A	A	A	C	C	C	C	C	A	
KHAO KHAM PAR	17	G	G	G	C	T	C	G	G	C	A	C	A	C	G	G	C	C	A	G	G	T	A	G	C	C	C	C	T	A
KONTOL 86	14	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	C	C	C	C	C	A
BR IRGA 409	13	C	C	C	G	C	C	G	G	A	C	A	G	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A
B 8136-3-TB-0-1-5	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	T	C	C	C	A
B 4414-FMR-6-3	11	C	C	C	G	C	C	G	G	C	C	A	C	A	G	A	G	G	T	G	T	A	A	T	T	C	C	C	C	A
PSBRC 88	11	C	C	C	G	C	C	G	G	C	G	A	C	A	G	A	G	G	T	G	T	A	A	T	T	C	C	C	C	A
PSBRC 50	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A
BAI HE	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	C	C	C	C	C	A
EX EBOKOZURU	11	C	C	C	G	C	C	G	G	C	G	A	G	C	A	G	A	G	G	T	G	T	A	A	C	C	C	C	C	A
TSMATAHOPOASA	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	C	C	C	C	C	A
MEKENZE SMALL	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	C	C	C	C	C	A
CHANDINA	11	C	C	C	G	C	C	G	G	C	A	G	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	C	C	A
MODDAI KARUPPAN	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A
LEAD	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A
AI ZI HUNG	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A
TAIPEI WOO CO	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A
SERAIUP 27	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A
UNGUL	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A
PA KABAKIE	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A
HOLDI GIRI	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A
ASHI BINI	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A
SADOK	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A
B 78-581	11	C	C	C	G	C	C	G	G	C	G	C	A	C	A	G	A	G	G	T	G	T	A	A	T	C	C	C	C	A

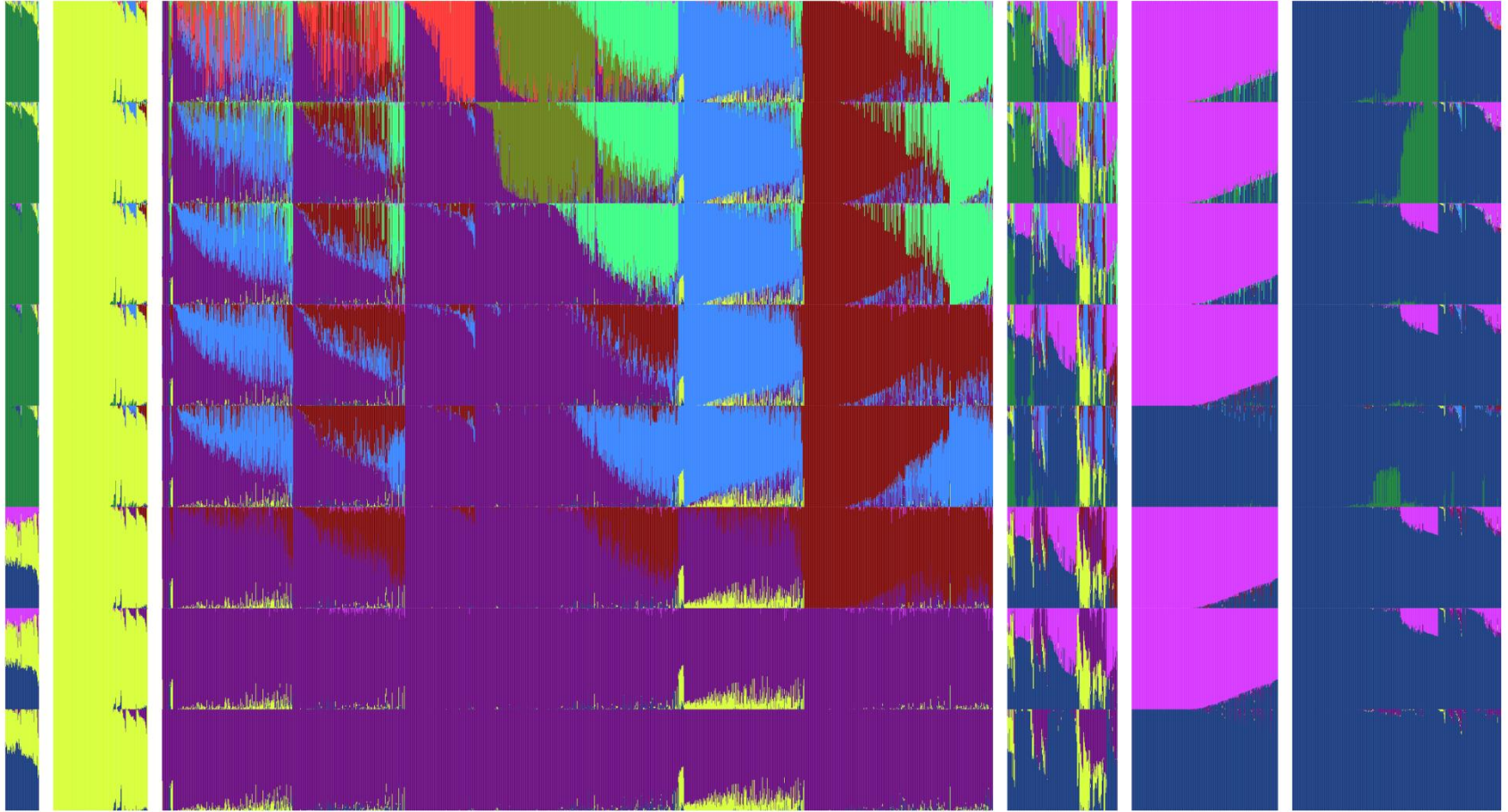


SNP-Seek online



# Admixture analysis of the 3K genomes

K=3 K=4 K=5 K=6 K=7 K=8 K=9 K=10



aro

aus

ind

Intermediate type

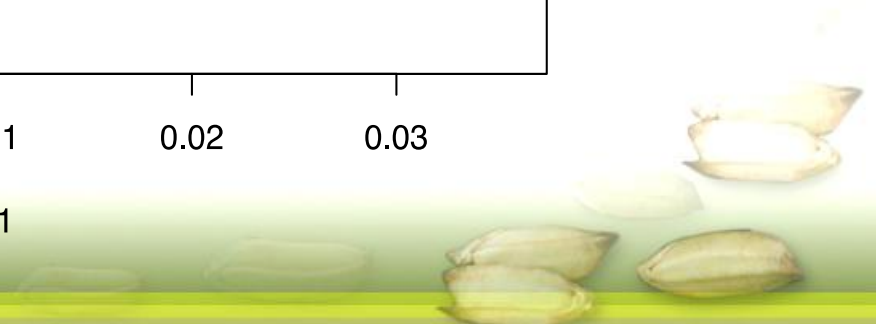
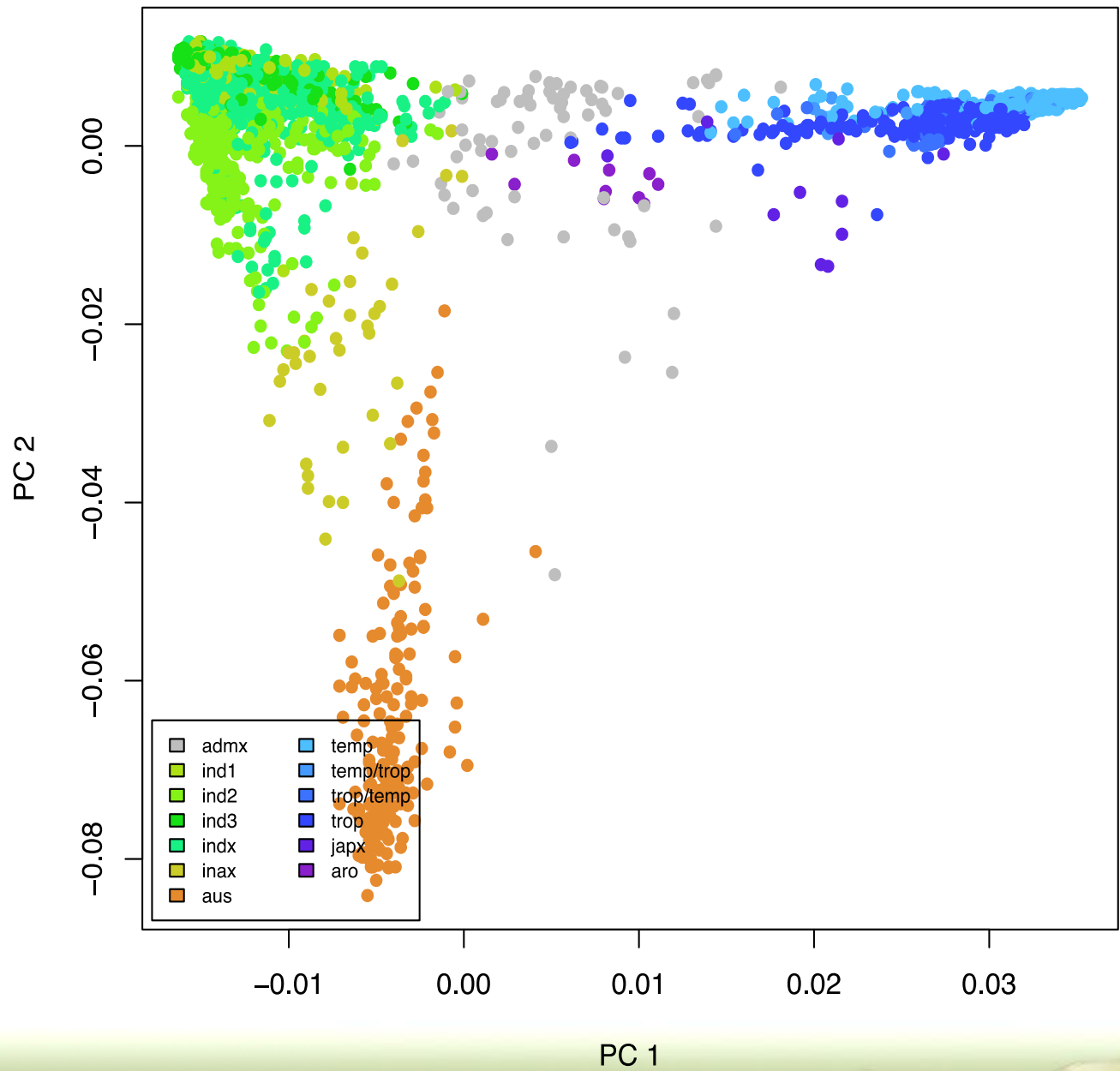
temp

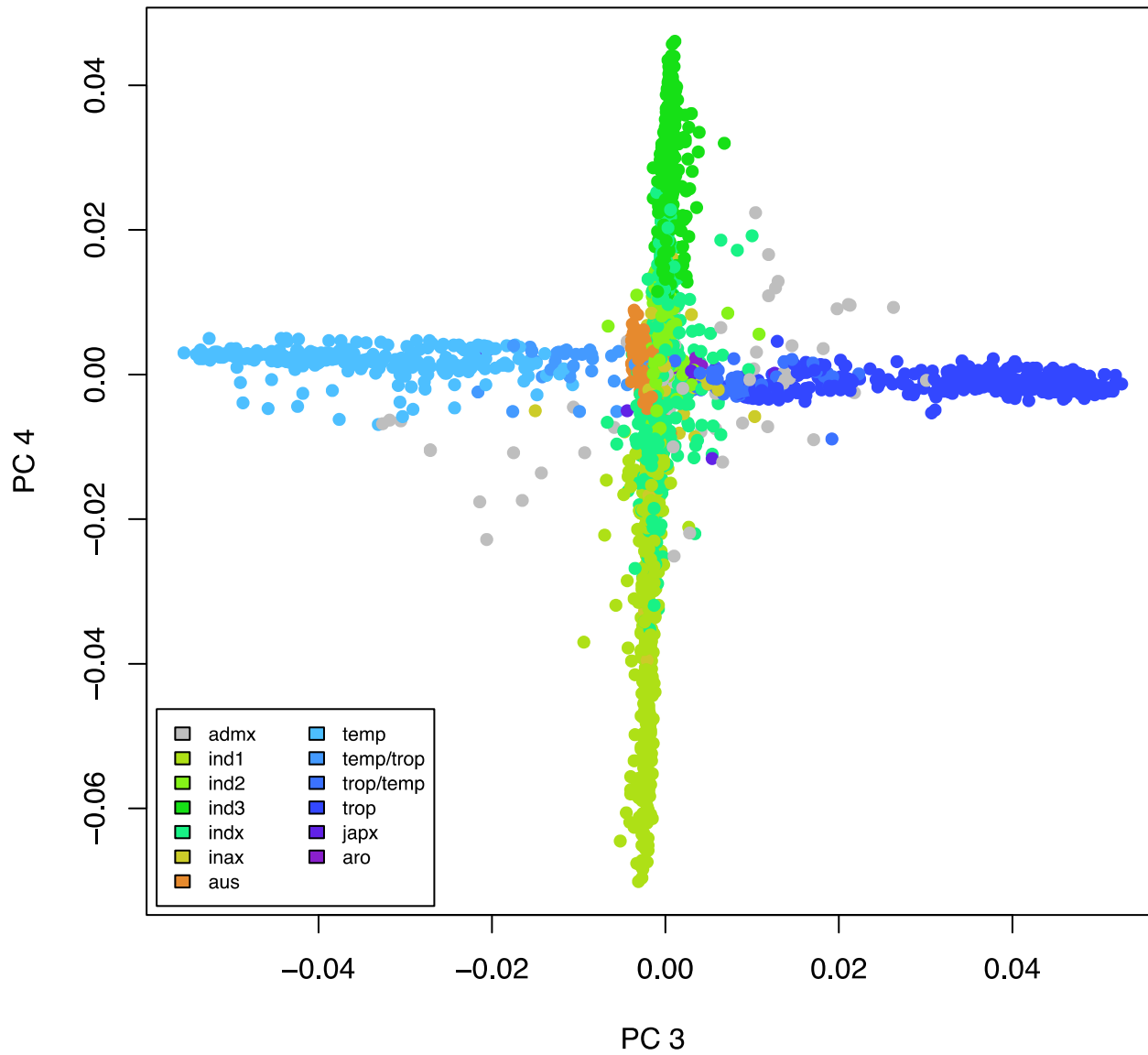
trop

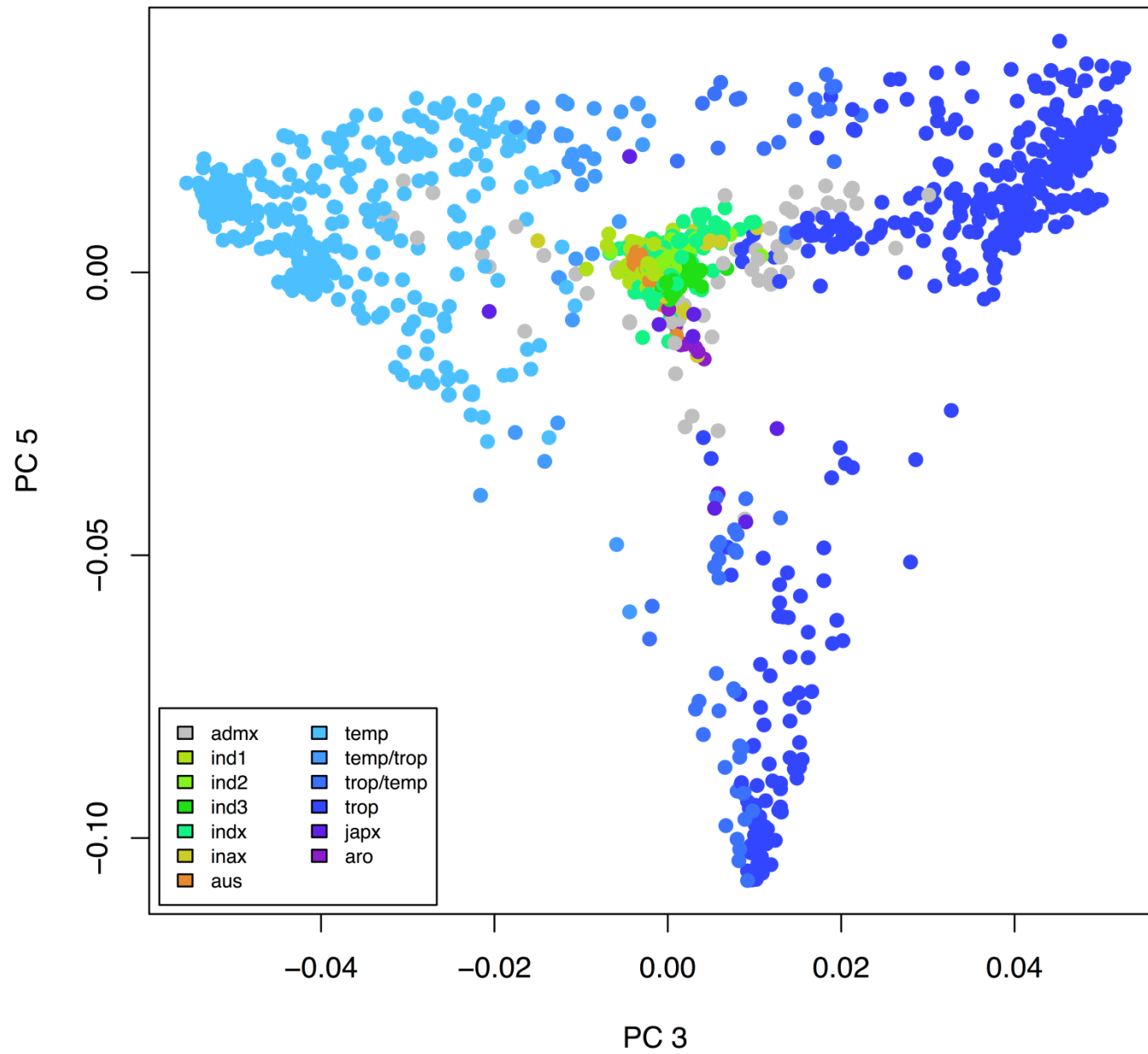
# Principal Component Analysis

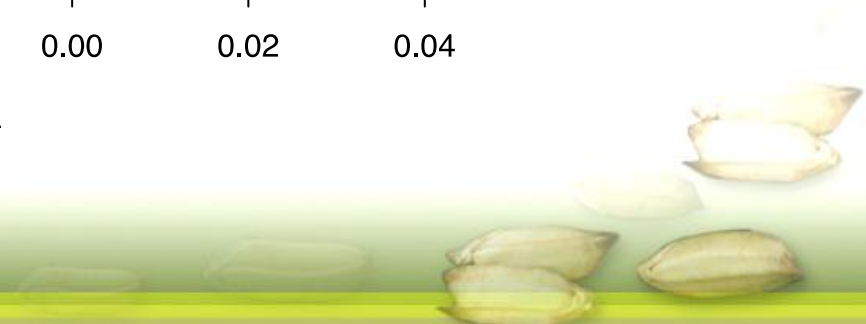
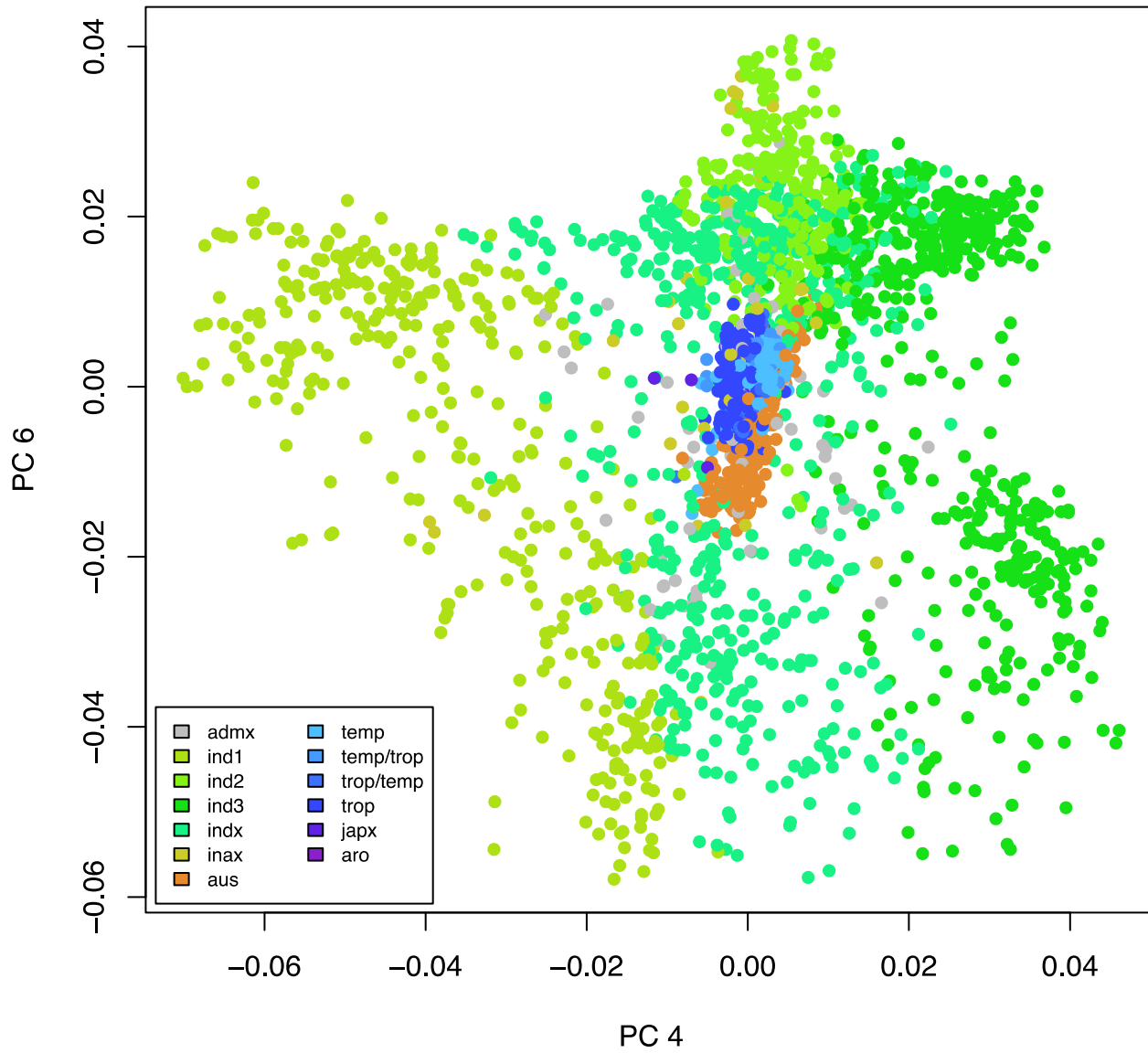
- PCA was run on 3k core SNP dataset using software EIGENSOFT
- 10 PCs were extracted
- The PCs suggest finer population structure ( both within indica and japonica)

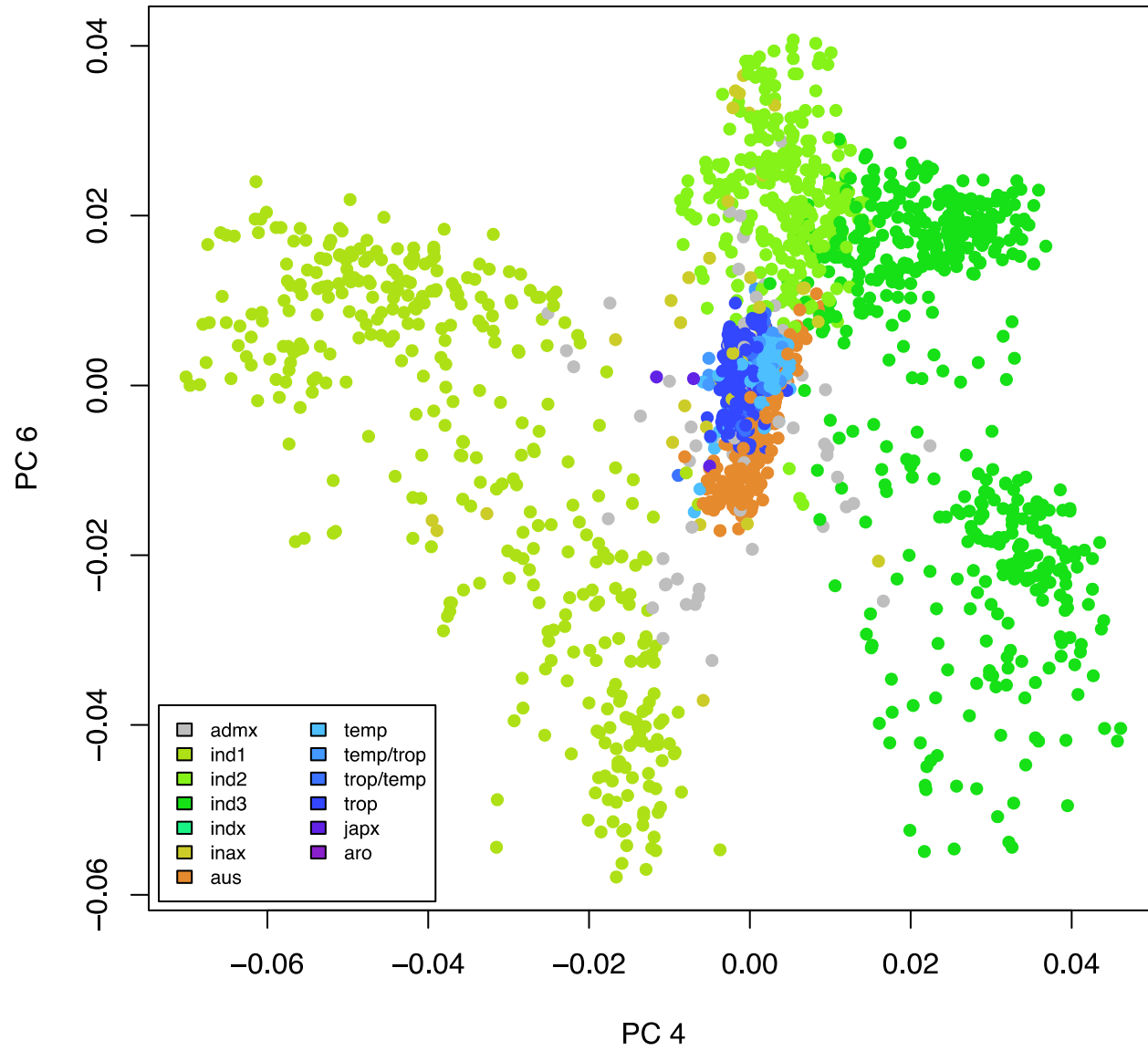




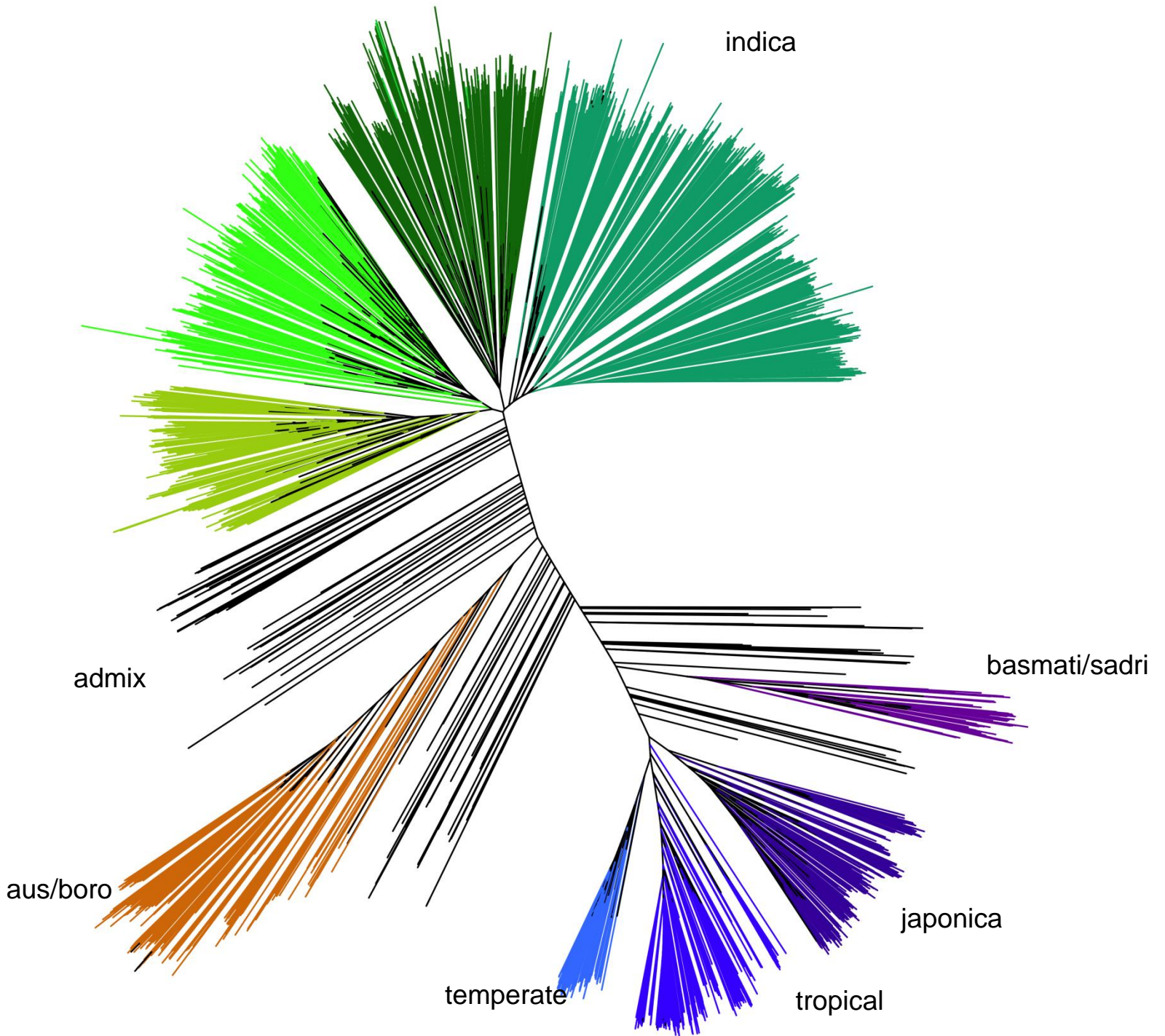




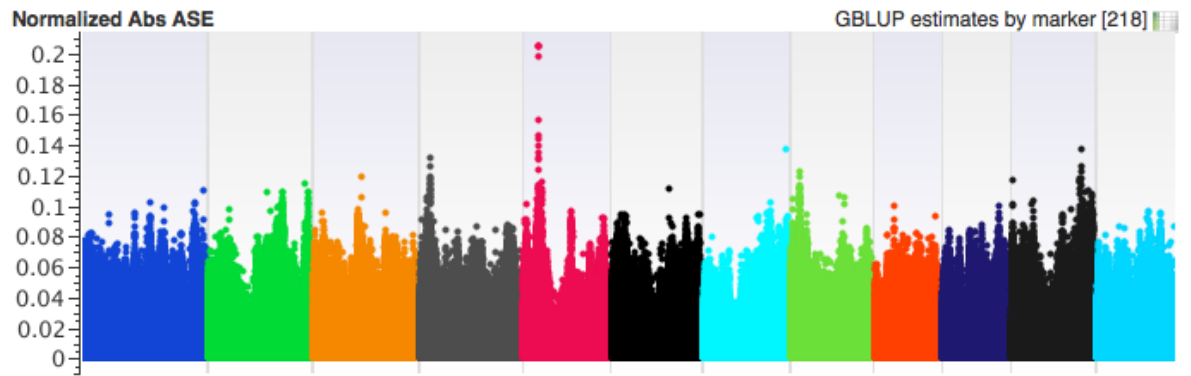
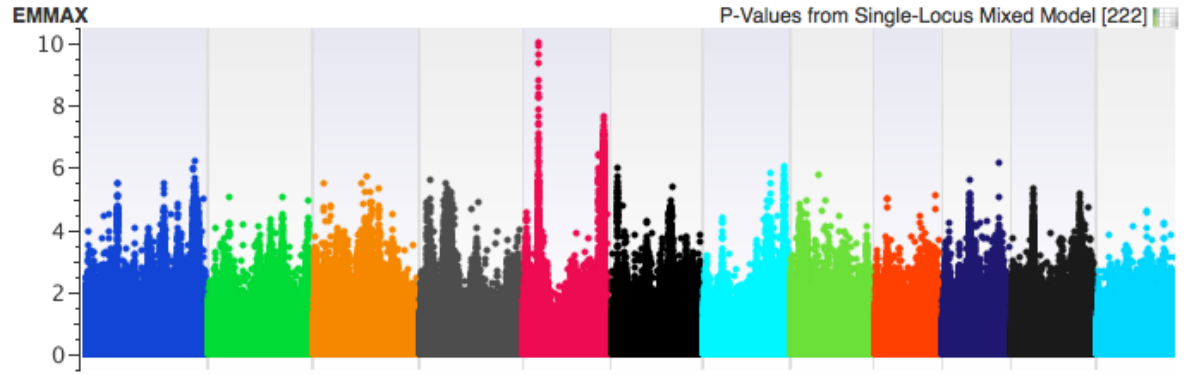
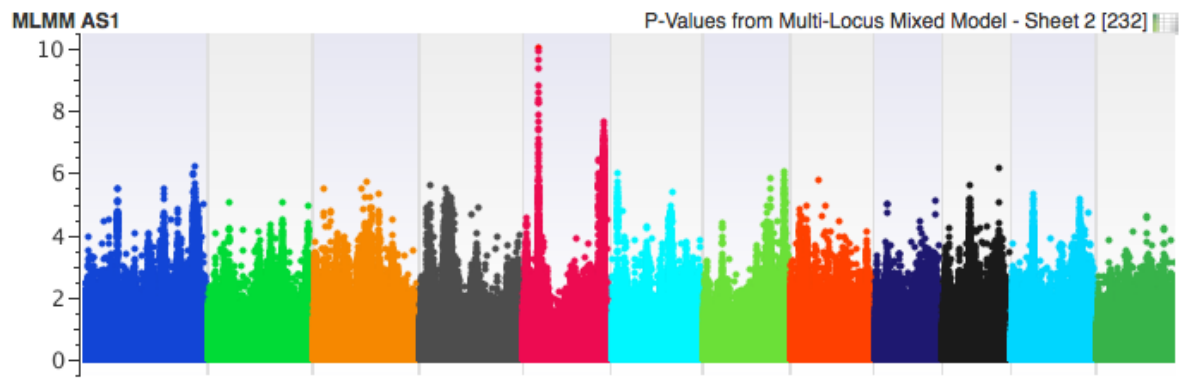
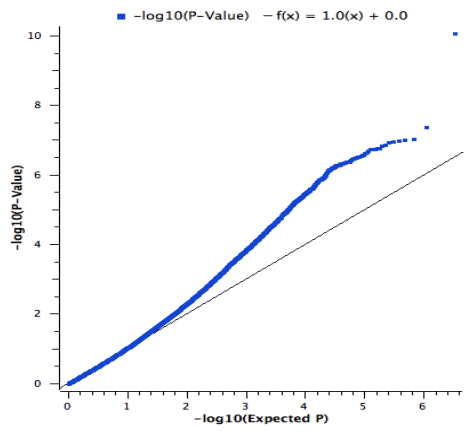




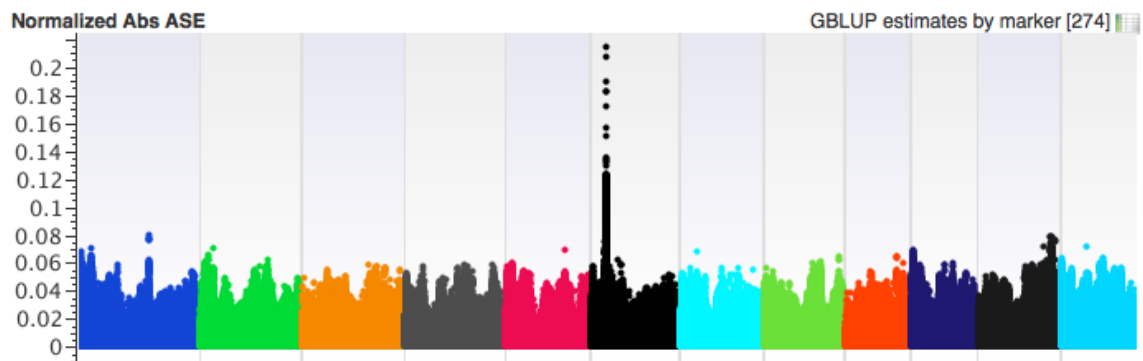
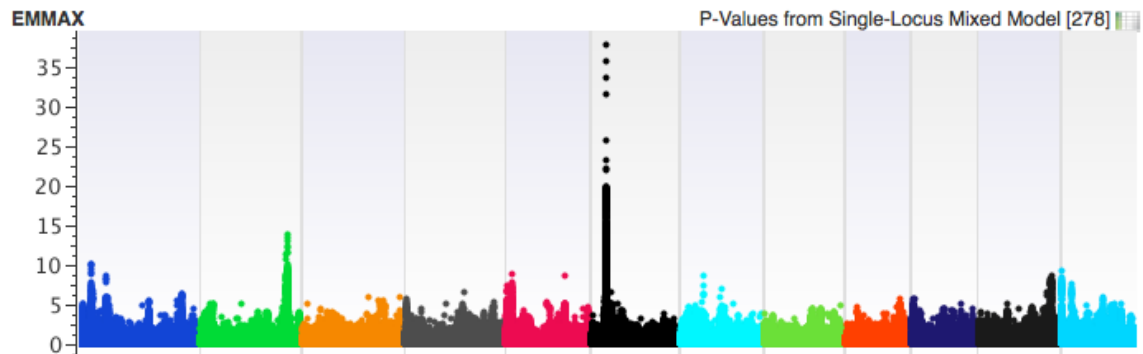
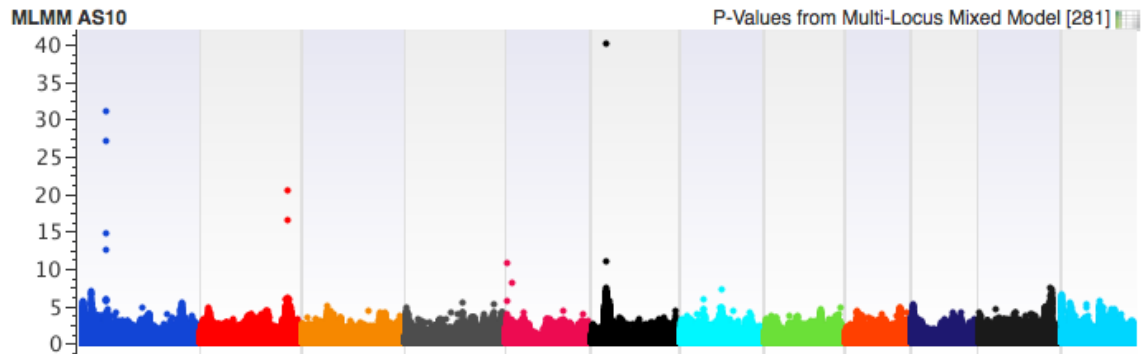
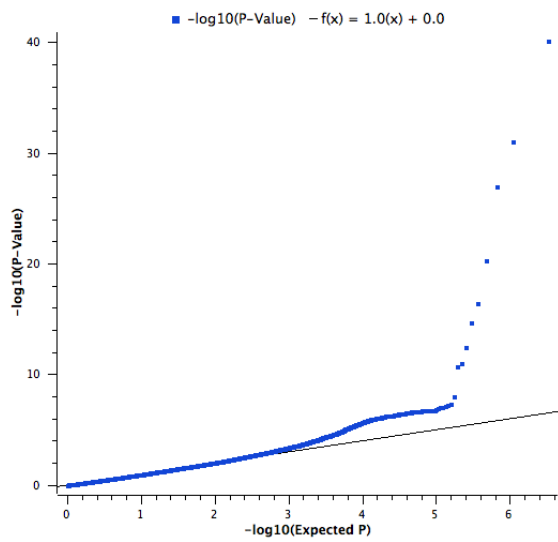




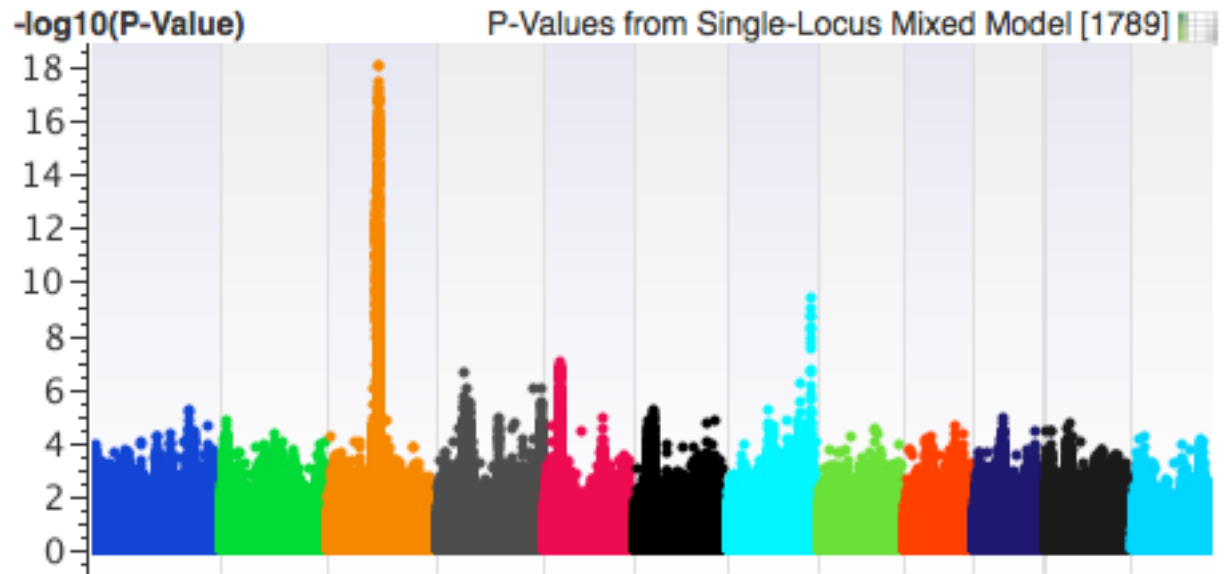
# Historical data for grain weight of 100 seeds from 1621 lines of 3K RGP (1.65M SNPS, CR>0.8, MAF<0.05)



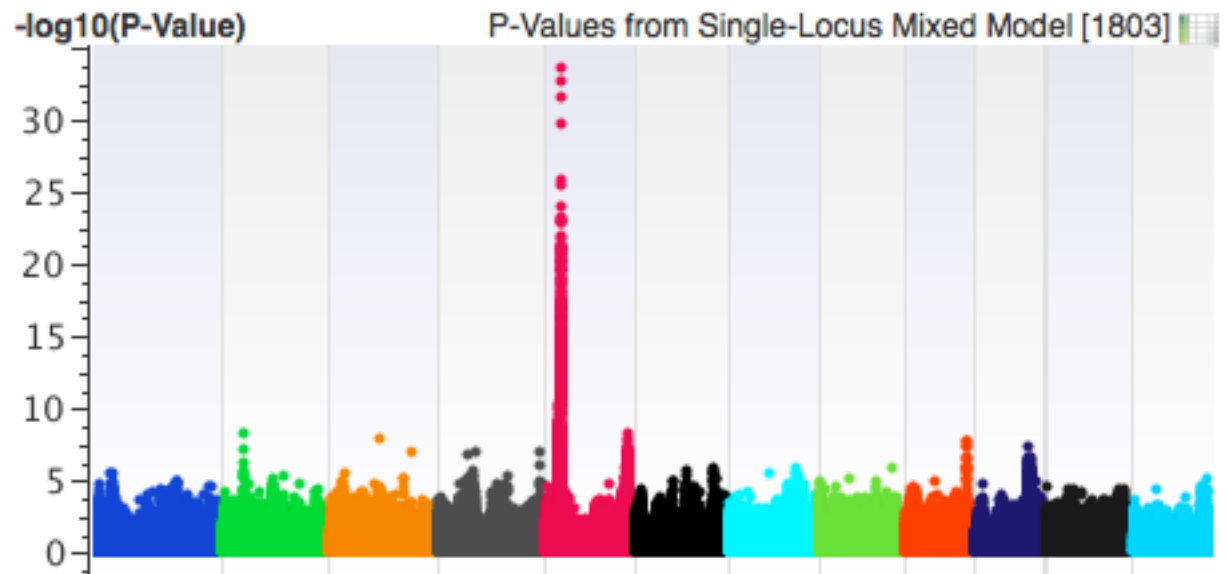
# Historical data for apiculus color from 1621 lines of 3K RGP (1.65M SNPS, CR>0.8, MAF<0.05)



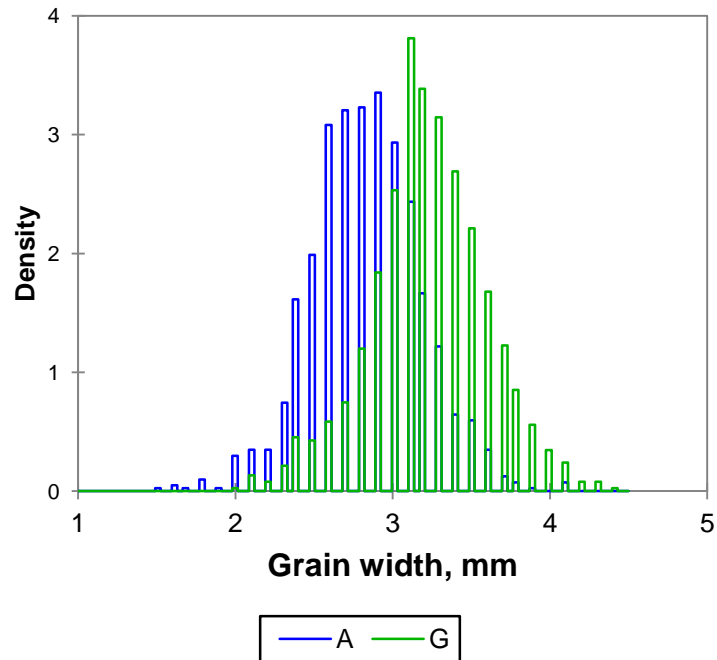
Grain length



Grain width



# Grain width distribution

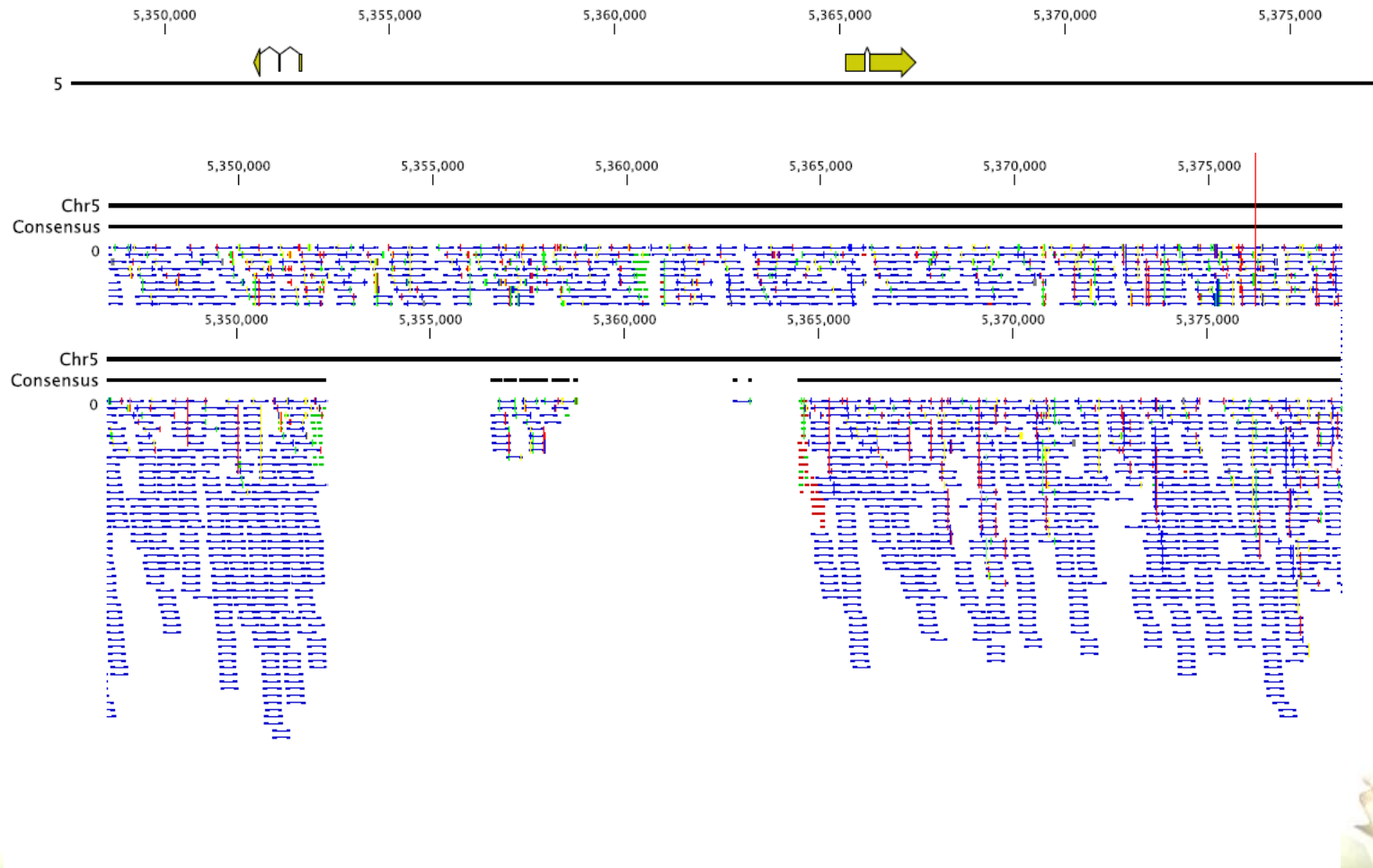


IRIS 313-8431 : G : 4.3 mm : trop : BOCAO

IRIS 313-12151 : A : 1.6 mm : ind3 : DAMNOEUB KRACHAK SESS



# Genome region at the GWAS peak



# GW5 rediscovered

*Cell Research* (2008) **18**:1199–1209. doi: 10.1038/cr.2008.307; published online 18 November 2008

## Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight

Jianfeng Weng<sup>1</sup>, Suhai Gu<sup>2</sup>, Xiangyuan Wan<sup>1</sup>, He Gao<sup>1</sup>, Tao Guo<sup>1</sup>, Ning Su<sup>2</sup>, Cailin Lei<sup>2</sup>, Xin Zhang<sup>2</sup>, Zhijun Cheng<sup>2</sup>, Xiuping Guo<sup>2</sup>, Jiulin Wang<sup>2</sup>, Ling Jiang<sup>1</sup>, Huqu Zhai<sup>2</sup> and Jianmin Wan<sup>2</sup>

<sup>1</sup>National Key Laboratory for Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China

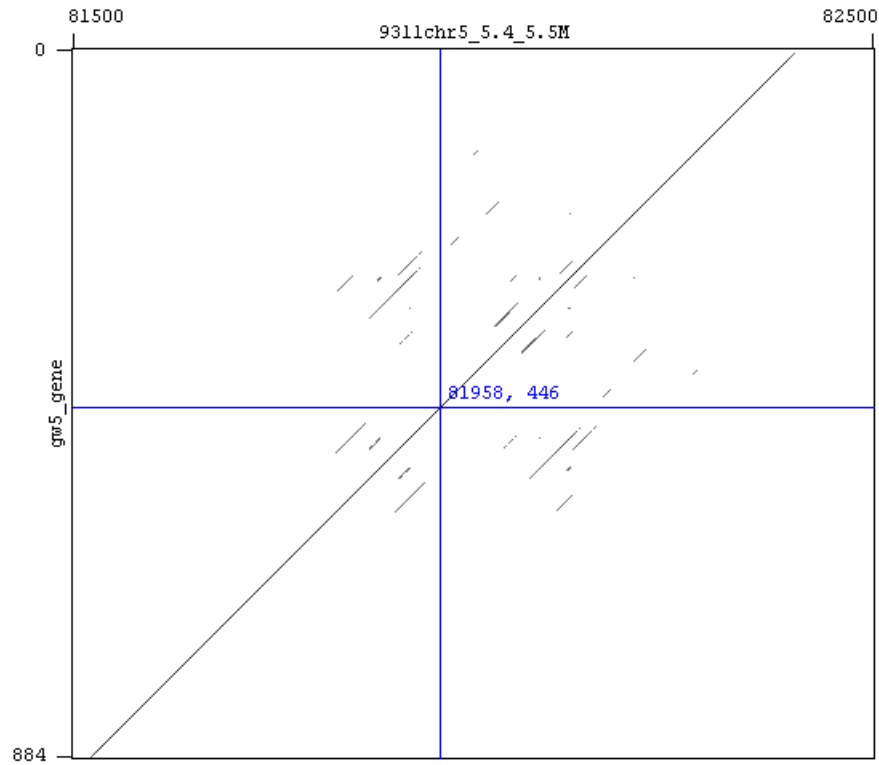
<sup>2</sup>Institute of Crop Science and the National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Correspondence: Jianmin Wan, Tel/Fax: +86-10-62186628 E-mail: [wanjm@caas.net.cn](mailto:wanjm@caas.net.cn)

Received 3 September 2008; Revised 6 October 2008; Accepted 6 October 2008; Published online 18 November 2008.



# Gw5 vs. 9311

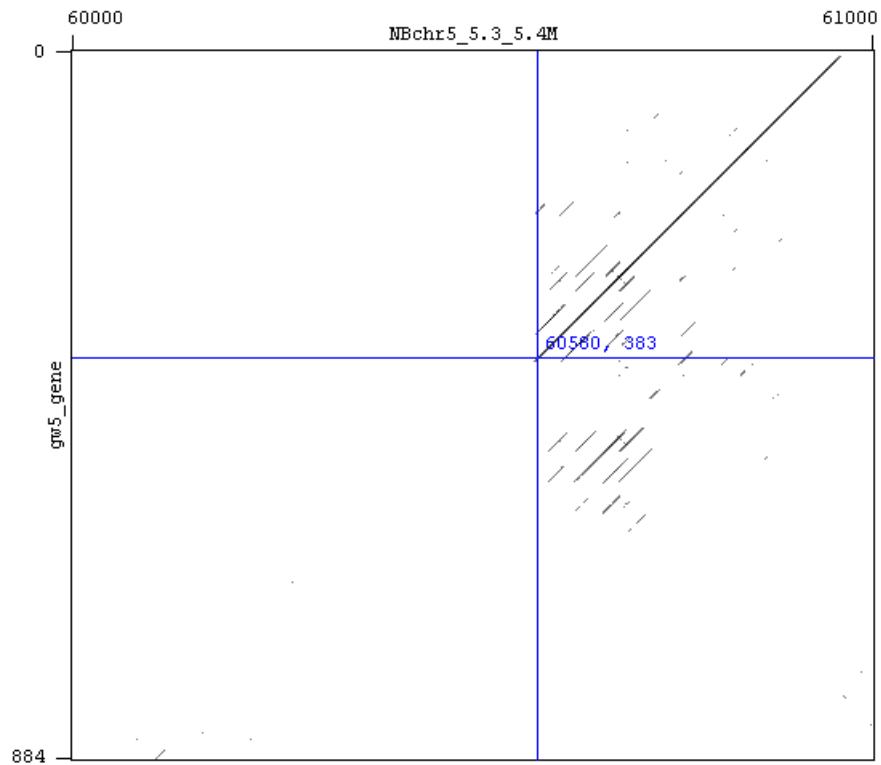


```
81864 | 81958 | 82051 |
| | |
TCTTCCTGCTCCCTCCCCCGATGCCGGTCCGTCCTCCTCCGCTGCTCCCTCCCCGGCGCCCGTCCCTCCTCCTGCCCCGCTCGCCAAGTTGCCGGGTGCACCGCCCGCTCGTGCACTCCTCCTGCCATCCGGCGCCGGTCCGCTGATCGCGCCGCTCCTCCTTCTGCTACTCCCTCCCCGGC
.....
TCTTCCTGCTCCCTCCCCCGATGCCGGTCCGTCCTCCTCCGCTGCTCCCTCCCCGGCGCCCGTCCCTCCTCCTGCCCCGCTCGCCAAGTTGCCGGGTGCACCGCCCGCTCGTGCACTCCTCCTGCCATCCGGCGCCGGTCCGCTGATCGCGCCGCTCCTCCTTCTGCTACTCCCTCCCCGGC
| | |
540 | 446 | 353 |
```





# Gw5 vs. NB

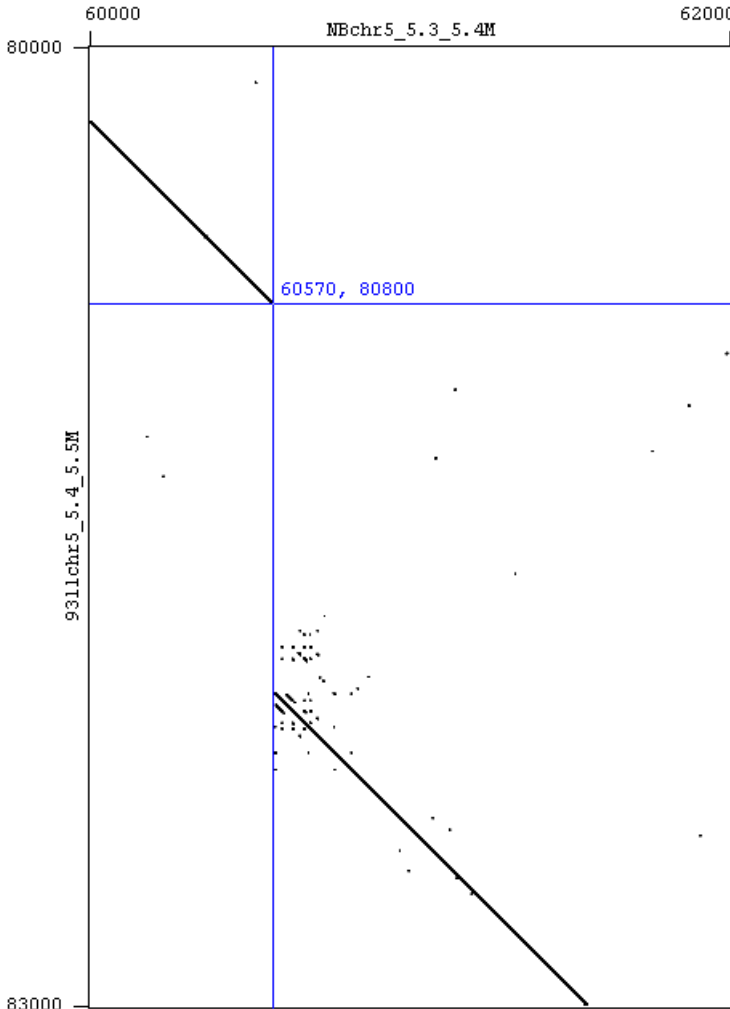


```

60486                                     60580                                     60673
|                                         |                                         |
TCCATCTATCCATCCGTCCACACACCTGCAAGTTTAAATTCATGCAGTTACTACATGAACAGTACTGCTGTCAAGTTTAAATTAATTTCCGCGCGCTCCTCTTCTGCTACTCCCTCCCCGGCACCAGTCCGTCCTCCTCCGCTTGCTCCCT
:   :  :  :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
GTCCCTCCTCCTGCCCCGCTCGCCAAGTTGCCGGCTGCACCGCCCGCTCGTGCAGTCTCCTGCCATCCGCGCCGCGGTCGCTGATCGCGCGCTCCTCTTCTGCTACTCCCTCCCCGGCACCAGTCCGTCCTCCTCCGCTTGCTCCCT
|                                         |                                         |
477                                         383                                         290
    
```



# NB vs. 9311



# Feedback from Rice Functional Genomics, Tucson AZ

② Have a central site for listing major genes/QTLs of interest to the community and adding information "Rice gene clearing house" for community annotation and prioritization for breeders  
 MIKE THOMPSON

② Set up Working Group for "Plant Breeding applications" to gather input from breeders on how to apply data for breeding (including linking with IBP and Breeding4RICE databases and tools)  
 MIKE THOMPSON

Serge Duitama ② I think that information of structural variation per sample is very important to design good markers. This information should be discovered and made available in the BRIC database.

② OARD  
 ① Get breeders involved early!  
 ② Make data easily accessible and understandable by and for breeders!  
 ③ Integrate other sequence data.

Tracking of germplasm beyond just "accession" seed source and generation are important when distributing seeds for phenotyping.

⑤

→ There should be an international support from Govts of Developed Countries to support Genome Sequencing Projects.

⑤

→ Immense diversity exists in form indig. But sequence are not a → we need

on the lead countries to facilitate Genome Seq

→ There should focus on of the best differential countries. (the germ)

# Feedback from Rice Functional Genomics, Tucson AZ

- Germplasm, mutants
  - Tracking
  - Access to seeds
  - Metadata (location, breeding history,
  - Mutants, knockouts and overexpression lines
  - Define subset(s) for phenotyping
- Phenotyping
  - Controlled vocabulary
  - Trait ontology, Crop ontology, Environment ontology, Method ontology
  - Global phenotyping



# Feedback from Rice Functional Genomics, Tucson AZ

- Sequencing
  - Information on the planned/ongoing projects
  - Version control
  - Correspondence between different assemblies (unified coordinates)
- Genes and loci
  - Versions of genome annotation
  - Gene names, Gene ontology
  - SNP
  - Location in different assemblies
  - Trait-loci associations
  - Gene expression
  - Gene networks



# Feedback from Rice Functional Genomics, Tucson AZ

- Tools
  - Web portal with restful queries
  - Mobile application
  - GWAS: population structure
  - Causative mutations, assay design
  - Predictive models, validation of trait-loci associations
  - Comparative genomics (e.g. for C4/C3 analysis)
  - Structural variant detection
  - Promoter analysis
  - Breeder assistance on selection of new lines, crosses
- Organizational
  - Monthly letters
  - Interest groups
  - IP management
  - Coordinating with other projects



# Priorities

- ✓ Use other reference genome sequences with **good** annotation
- ✓ Find important SNPs (merging with current GWAS/QTL results)
  - in CDSs
  - in promoters and other regulatory motifs
- ✓ Reconstruct large deletions/insertions/inversions in genome
- ✓ Find correlated SNPs
- ✓ Organize available GWAS/QTL data
- ✓ Create new genome assemblies
- ✓ Rice metagenome analysis
- ✓ Focus on known genes associated with traits
- ✓ Find conserved genome regions selected by breeders
- ✓ Population genetics analysis and predictions



# IRIC Goals for 2015

1. Develop a mechanism for unified germplasm tracking
2. Select a subset(s) of germplasm for detailed studies and organize corresponding metadata using existing ontologies
3. Create a curated database of functional loci annotation
4. Establish communication with similar plant consortia (Arabidopsis, wheat, soy, etc)
5. Approach funding agencies





# IRIC portal development team

Rolando Jay Santos  
Victor Jun Ulat  
Frances Nikki Borja  
Venice Margarete Juanillas  
Jeffrey Detras  
Roven Rommel Fuentes  
Dmytro Chebatorov  
Millicent Sanciangco  
Kevin Palis  
Locedie Mansueto

## **Technological advice**

Marco van den Berg

## **Visionary input**

Achim Dobermann  
Matthew Morell

## **Administrative support**

Vina Alvarez

## **Management team**

Hei Leung  
Ruairaidh Sackville Hamilton  
Kenneth McNally  
Ramil Mauleon  
Nickolai Alexandrov  
Rod Wing



