

Plant variation data - what's available, which new projects are expected

Martin Mascher

IPK Gatersleben

4th transPLANT Workshop, Hinxton, UK
July 1st, 2015

Member of the



IPK Gatersleben



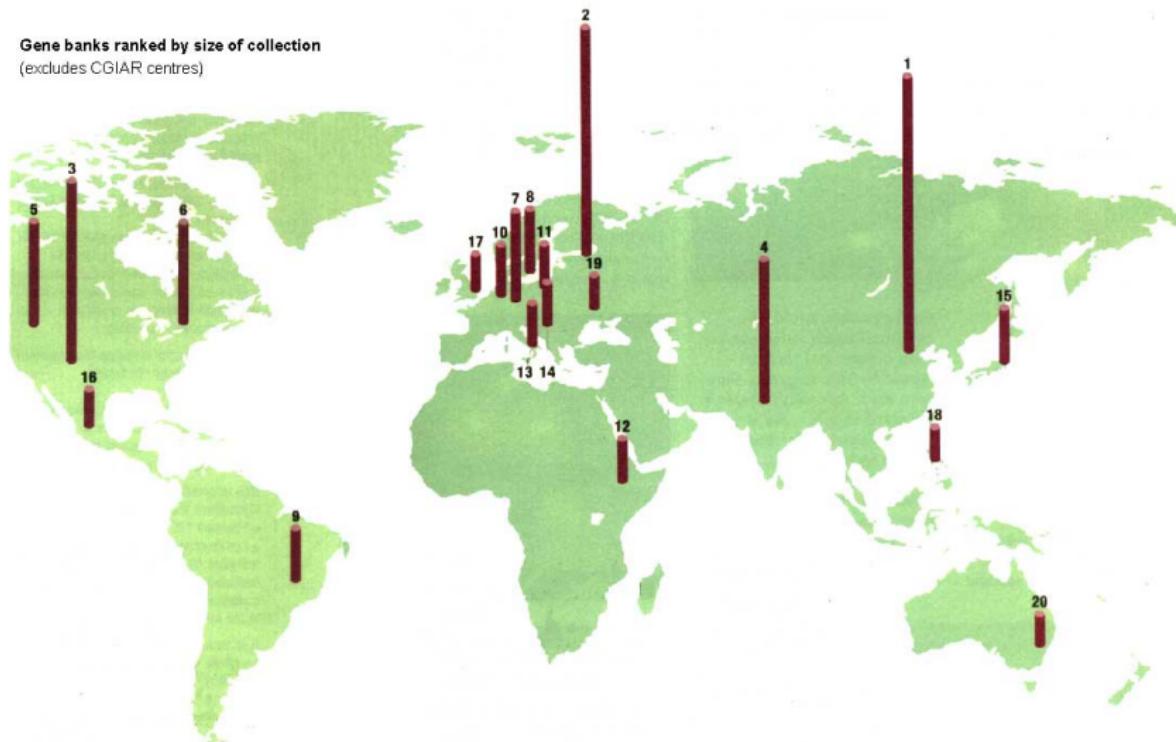
© IPK

Ex situ genebanks around the world

- ▶ Conserve, characterize and utilize plant genetic variation

Gene banks ranked by size of collection

(excludes CGIAR centres)



Barley diversity at IPK Gatersleben



Photos by Michael Grau, IPK genebank

Types of variation

- ▶ Phenotypic variation
 - "Visible" traits: plant height, flowering time, pathogen resistance, ...
 - Molecular phenotypes: metabolites, hormone, ionome, ...
 - Ecological data: geographic origin, environmental parameters, ...



<http://nativetraits.blogspot.de>

Types of variation

► Phenotypic variation

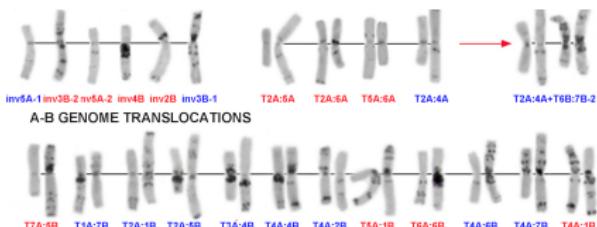
- "Visible" traits: plant height, flowering time, pathogen resistance, ...
- Molecular phenotypes: metabolites, hormone, ionome, ...
- Ecological data: geographic origin, environmental parameters, ...



<http://nativetraits.blogspot.de>

► Genetic variation

SNPs, Indels, structural variation, karyotypes, ploidy, genome size



Badaeva *et al.*, PLOS ONE, 2015

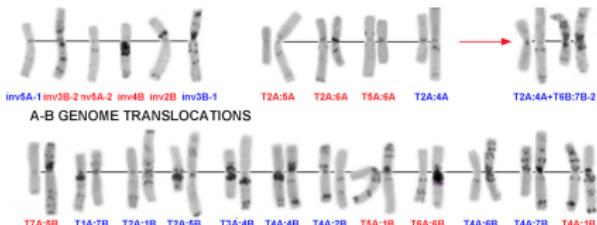
Types of variation

- ▶ Phenotypic variation
 - "Visible" traits: plant height, flowering time, pathogen resistance, ...
 - Molecular phenotypes: metabolites, hormone, ionome, ...
 - Ecological data: geographic origin, environmental parameters, ...



<http://nativetraits.blogspot.de>

- ▶ Genetic variation
 - SNPs, Indels, structural variation, karyotypes, ploidy, genome size



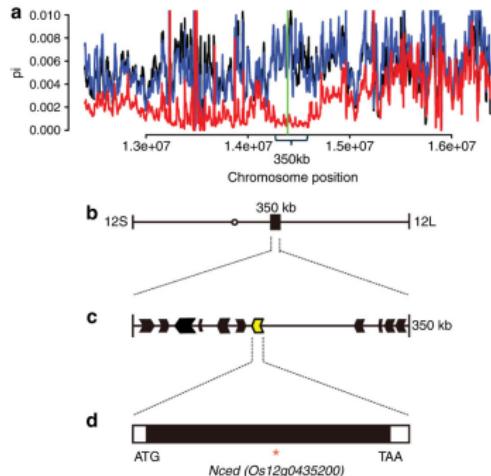
Badaeva *et al.*, PLOS ONE, 2015

- ▶ Inter-specific variation
 - Phylogeny, comparative genomics

Applications of variant data

► Population genetics

Nucleotide diversity, demographic inference, targets of selection



Lyu et al., Nat. Comm., 2013

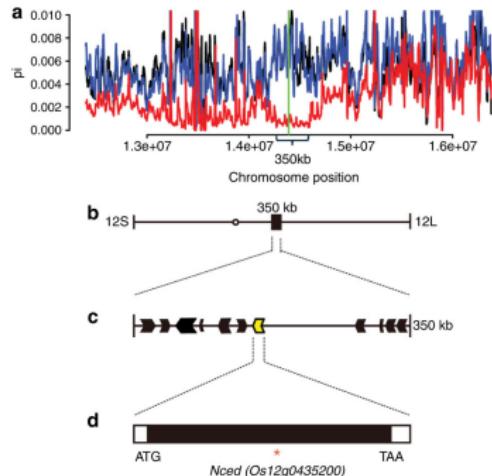
Applications of variant data

► Population genetics

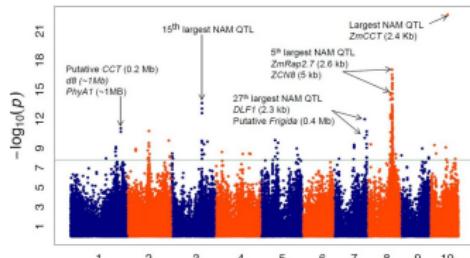
Nucleotide diversity, demographic inference, targets of selection

► Association genetics

Linking phenotype to genotype,
GWAS



Lyu et al., Nat. Comm., 2013



Romay et al., Genome Biology, 2013

Applications of variant data

► Population genetics

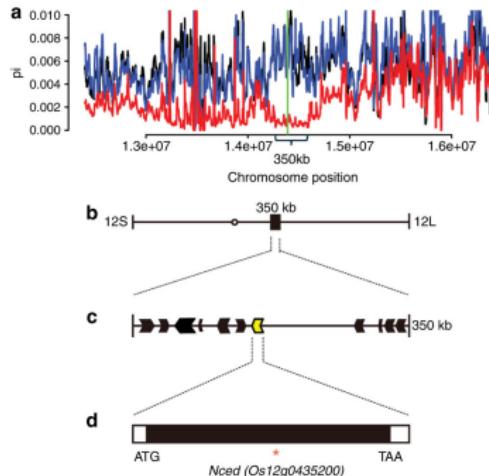
Nucleotide diversity, demographic inference, targets of selection

► Association genetics

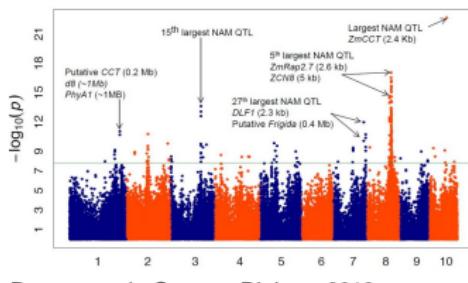
Linking phenotype to genotype,
GWAS

► Conservation management *in situ* and *ex situ*

Molecular passport data



Lyu *et al.*, Nat. Comm., 2013



Romay *et al.*, Genome Biology, 2013

Applications of variant data

► Population genetics

Nucleotide diversity, demographic inference, targets of selection

► Association genetics

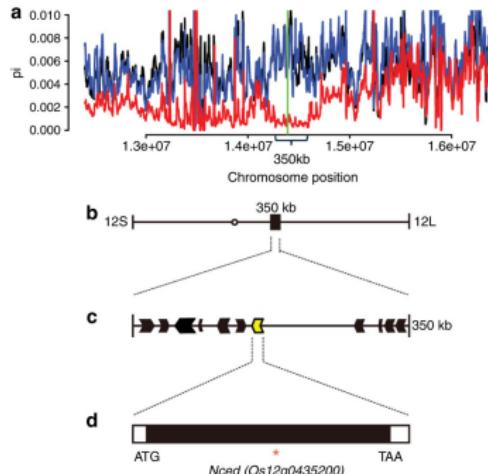
Linking phenotype to genotype,
GWAS

► Conservation management *in situ* and *ex situ*

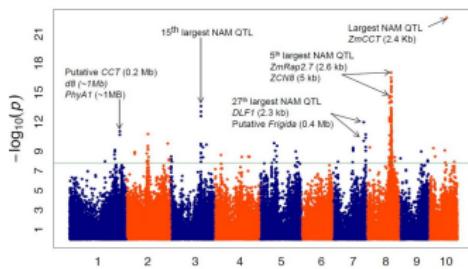
Molecular passport data

► (Pre)-breeding

Management of breeding programs,
characterization of exotic donors



Lyu *et al.*, Nat. Comm., 2013



Romay *et al.*, Genome Biology, 2013

Genotyping strategies

- ▶ SNP arrays

Genotyping of known variants,
ascertainment bias



<http://www.ibbl.lu>

Genotyping strategies

- ▶ SNP arrays
 - Genotyping of known variants,
ascertainment bias
- ▶ Whole-genome shotgun
 - Most comprehensive, most expensive



<http://www.ibbl.lu>



<http://www.illumina.com>

Genotyping strategies

- ▶ SNP arrays
 - Genotyping of known variants,
ascertainment bias
- ▶ Whole-genome shotgun
 - Most comprehensive, most expensive
- ▶ Genotyping-by-sequencing, RADseq
 - Cheap, tens of thousands of markers



<http://www.ibbl.lu>



<http://www.illumina.com>

Genotyping strategies

- ▶ SNP arrays
 - Genotyping of known variants,
ascertainment bias
- ▶ Whole-genome shotgun
 - Most comprehensive, most expensive
- ▶ Genotyping-by-sequencing, RADseq
 - Cheap, tens of thousands of markers
- ▶ RNAseq / Exome capture
 - Comprehensive discovery and
typing of variants in genes



<http://www.ibbl.lu>



<http://www.illumina.com>

Plant resequencing projects

Species	# acc	strategy	publication
Rice	3000	WGS	3000 RGP, 2014, GigaScience
Maize	2815	GBS	Romay <i>et al.</i> , 2013, Genome Biol
Rice	1529	WGS	Huang <i>et al.</i> , 2012, Nature
<i>A. thaliana</i>	1135	WGS	Cao <i>et al.</i> , 2011, Nat Genet [and others]
Sorghum	971	GBS	Morris <i>et al.</i> , 2013, PNAS
<i>P. trichocarpa</i>	544	WGS	Evans <i>et al.</i> , 2014, Nat Genet
Maize	503	RNAseq	Hirsch <i>et al.</i> , 2014, Plant Cell
Soybean	302	WGS	Zhou <i>et al.</i> , 2005, Nat Biotech
Barley	270	EC, GBS	unpublished
Maize	113	WGS	Chia <i>et al.</i> , 2012, Nat Genet
Wheat	62	EC, GBS	Jordan <i>et al.</i> , 2015, Genome Biol
<i>Mimulus</i> spp.	19	WGS	Brandvain <i>et al.</i> , 2014, PLOS Genet

Plant resequencing projects

Species	# acc	strategy	publication
Rice	3000	WGS	3000 RGP, 2014, GigaScience
Maize	2815	GBS	Romay <i>et al.</i> , 2013, Genome Biol
Rice	1529	WGS	Huang <i>et al.</i> , 2012, Nature
<i>A. thaliana</i>	1135	WGS	Cao <i>et al.</i> , 2011, Nat Genet [and others]
Sorghum	971	GBS	Morris <i>et al.</i> , 2013, PNAS
<i>P. trichocarpa</i>	544	WGS	Evans <i>et al.</i> , 2014, Nat Genet
Maize	503	RNAseq	Hirsch <i>et al.</i> , 2014, Plant Cell
Soybean	302	WGS	Zhou <i>et al.</i> , 2005, Nat Biotech
Barley	270	EC, GBS	unpublished
Maize	113	WGS	Chia <i>et al.</i> , 2012, Nat Genet
Wheat	62	EC, GBS	Jordan <i>et al.</i> , 2015, Genome Biol
<i>Mimulus</i> spp.	19	WGS	Brandvain <i>et al.</i> , 2014, PLOS Genet

Plant resequencing projects

Species	# acc	strategy	publication
Rice	3000	WGS	3000 RGP, 2014, GigaScience
Maize	2815	GBS	Romay <i>et al.</i> , 2013, Genome Biol
Rice	1529	WGS	Huang <i>et al.</i> , 2012, Nature
A. thaliana	1135	WGS	Cao <i>et al.</i> , 2011, Nat Genet [and others]
Sorghum	971	GBS	Morris <i>et al.</i> , 2013, PNAS
P. trichocarpa	544	WGS	Evans <i>et al.</i> , 2014, Nat Genet
Maize	503	RNAseq	Hirsch <i>et al.</i> , 2014, Plant Cell
Soybean	302	WGS	Zhou <i>et al.</i> , 2005, Nat Biotech
Barley	270	EC, GBS	unpublished
Maize	113	WGS	Chia <i>et al.</i> , 2012, Nat Genet
Wheat	62	EC, GBS	Jordan <i>et al.</i> , 2015, Genome Biol
Mimulus spp.	19	WGS	Brandvain <i>et al.</i> , 2014, PLOS Genet

Plant resequencing projects

Species	# acc	strategy	publication
Rice	3000	WGS	3000 RGP, 2014, GigaScience
Maize	2815	GBS	Romay <i>et al.</i> , 2013, Genome Biol
Rice	1529	WGS	Huang <i>et al.</i> , 2012, Nature
A. thaliana	1135	WGS	Cao <i>et al.</i> , 2011, Nat Genet [and others]
Sorghum	971	GBS	Morris <i>et al.</i> , 2013, PNAS
P. trichocarpa	544	WGS	
Maize	503	RNAseq	
Soybean	302	WGS	
Barley	270	EC, GBS	
Maize	113	WGS	
Wheat	62	EC, GBS	
Mimulus spp.	19	WGS	Brandvain <i>et al.</i> , 2014, PLOS Genet



Plant resequencing projects

Species	# acc	strategy	publication
Rice	3000	WGS	3000 RGP, 2014, GigaScience
Maize	2815	GBS	Romay <i>et al.</i> , 2013, Genome Biol
Rice	1529	WGS	Huang <i>et al.</i> , 2012, Nature
<i>A. thaliana</i>	1135	WGS	Cao <i>et al.</i> , 2011, Nat Genet [and others]
Sorghum	971	GBS	Morris <i>et al.</i> , 2013, PNAS
<i>P. trichocarpa</i>	544	WGS	Evans <i>et al.</i> , 2014, Nat Genet
Maize	503	RNAseq	Hirsch <i>et al.</i> , 2014, Plant Cell
Soybean	302	WGS	Zhou <i>et al.</i> , 2005, Nat Biotech
Barley	270	EC, GBS	unpublished
Maize	113	WGS	Chia <i>et al.</i> , 2012, Nat Genet
Wheat	62	EC, GBS	Jordan <i>et al.</i> , 2015, Genome Biol
<i>Mimulus</i> spp.	19	WGS	Brandvain <i>et al.</i> , 2014, PLOS Genet

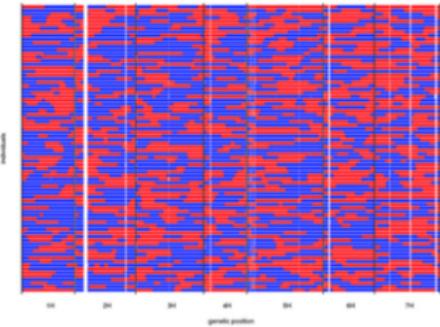
Plant resequencing projects

Species	# acc	strategy	publication
Rice	3000	WGS	3000 RGP, 2014, GigaScience
Maize	2815	GBS	Romay <i>et al.</i> , 2013, Genome Biol
Rice	1529	WGS	Huang <i>et al.</i> , 2012, Nature
<i>A. thaliana</i>	1135	WGS	Cao <i>et al.</i> , 2011, Nat Genet [and others]
Sorghum	971	GBS	Morris <i>et al.</i> , 2013, PNAS
<i>P. trichocarpa</i>	544	WGS	Evans <i>et al.</i> , 2014, Nat Genet
Maize	503	RNAseq	Hirsch <i>et al.</i> , 2014, Plant Cell
Soybean	302	WGS	Zhou <i>et al.</i> , 2005, Nat Biotech
Barley	270	EC, GBS	unpublished
Maize	113	WGS	Chia <i>et al.</i> , 2012, Nat Genet
Wheat	62	EC, GBS	Jordan <i>et al.</i> , 2015, Genome Biol
<i>Mimulus</i> spp.	19	WGS	Brandvain <i>et al.</i> , 2014, PLOS Genet

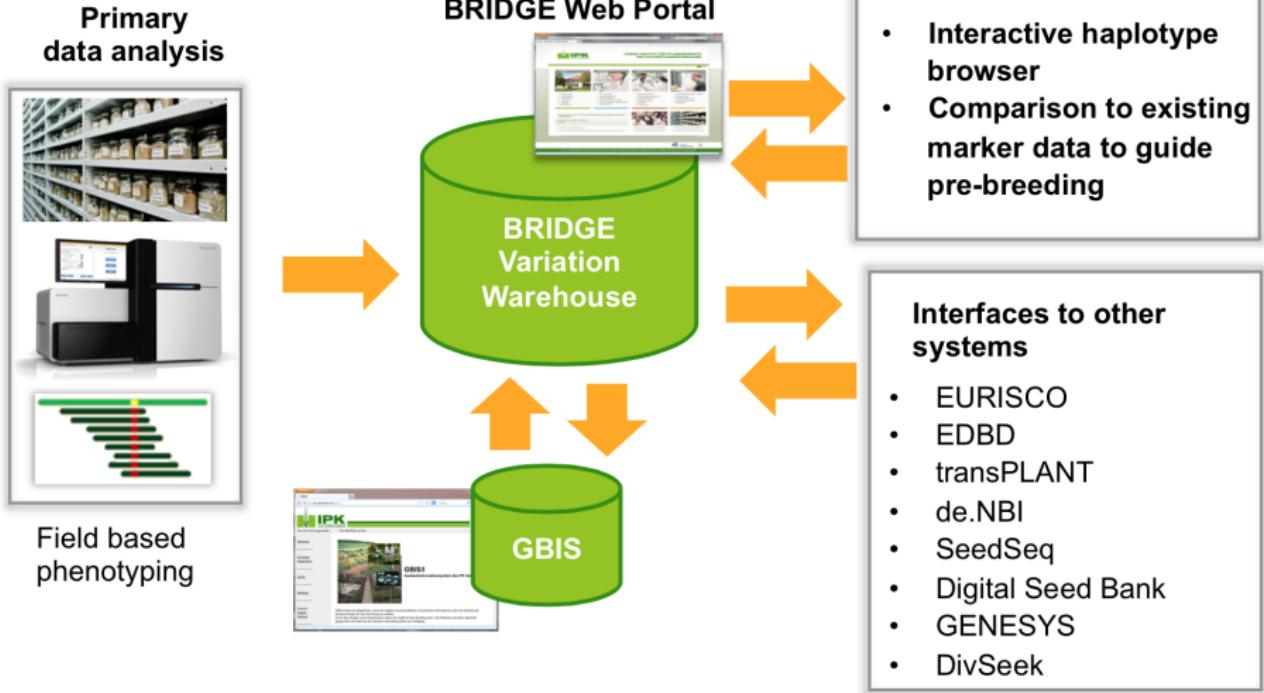
Plant resequencing projects

Species	# acc	strategy	publication
Rice	3000	WGS	3000 RGP, 2014, GigaScience
Maize	2815	GBS	Romay <i>et al.</i> , 2013, Genome Biol
Rice	1529	WGS	Huang <i>et al.</i> , 2012, Nature
<i>A. thaliana</i>	1135	WGS	Cao <i>et al.</i> , 2011, Nat Genet [and others]
Sorghum	971	GBS	Morris <i>et al.</i> , 2013, PNAS
<i>P. trichocarpa</i>	544	WGS	Evans <i>et al.</i> , 2014, Nat Genet
Maize	503	RNAseq	Hirsch <i>et al.</i> , 2014, Plant Cell
Soybean	302	WGS	Zhou <i>et al.</i> , 2005, Nat Biotech
Barley	270	EC, GBS	unpublished
Maize	113	WGS	Chia <i>et al.</i> , 2012, Nat Genet
Wheat	62	EC, GBS	Jordan <i>et al.</i> , 2015, Genome Biol
<i>Mimulus</i> spp.	19	WGS	Brandvain <i>et al.</i> , 2014, PLOS Genet

BRIDGE: genotyping the IPK barley collection



BRIDGE: biodiversity informatics



Field based phenotyping

New information infrastructures for genotypic datasets

D1018–D1022 *Nucleic Acids Research*, 2015, Vol. 43, Database issue
doi: 10.1093/nar/gku894

Published online 01 October 2014

RiceVarMap: a comprehensive database of rice genomic variations

Hu Zhao, Wen Yao, Yidan Ouyang, Wanneng Yang, Gongwei Wang, Xingming Lian,
Yongzhong Xing, Lingling Chen and Weibo Xie*

National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China

The screenshot shows the homepage of the RiceVarMap database. At the top, there is a banner featuring three images: a close-up of rice plants, a field of rice, and a building. Below the banner is a navigation bar with links: Home, SNP Information, INDEL Information, Cultivars, Tools, GBrowse, Documents & Downloads, and Contact. On the left, there is a sidebar titled "What's New" with a blue header. It lists three items: "27 Mar 2015" (Genome variation data (BCF format) can be downloaded at [Data downloads](#) page), "14 Sep 2014" (Cultivars can be searched by keywords in ["Cultivar Information"](#) page), and "14 Sep 2014" (Added an option "Download CSV File Only" in ["Search for Genotype With SNP ID"](#) and ["Search for Genotype With INDEL ID"](#) pages. This is the quickest way to obtain genotype data using SNP/INDEL). The main content area has a yellow header: "Welcome to Rice Variation Map, a comprehensive database of rice genomic variations." Below this, there is a section titled "Database contents:" which contains a detailed description of the database's features and data.

RiceVarMap

Home SNP Information INDEL Information Cultivars Tools GBrowse Documents & Downloads Contact

What's New

27 Mar 2015 Genome variation data (BCF format) can be downloaded at [Data downloads](#) page.

14 Sep 2014 Cultivars can be searched by keywords in ["Cultivar Information"](#) page.

14 Sep 2014 Added an option "Download CSV File Only" in ["Search for Genotype With SNP ID"](#) and ["Search for Genotype With INDEL ID"](#) pages. This is the quickest way to obtain genotype data using SNP/INDEL

Welcome to Rice Variation Map, a comprehensive database of rice genomic variations.

Database contents:

RiceVarMap provides comprehensive information of 6,551,358 single nucleotide polymorphisms (SNPs) and 1,214,627 insertions/deletions (INDELS) identified from sequencing data of 1,479 rice accessions. The SNP genotypes of all accessions were imputed and evaluated, resulting in an overall missing data rate of 0.42% and an estimated accuracy greater than 99%. The SNP/INDEL genotypes of all accessions are available for online queries and downloading. Users can search SNPs/INDELS by identifiers of the SNPs/INDELS, genomic regions, gene identifiers and keywords of gene annotation. Allele frequencies within various sub-populations and the effects of the variation that may alter the protein sequence of a gene are also listed for each SNP/INDEL. The database provides a tool to compare any two accessions and identify the polymorphisms between them. The database also provides geographical details and phenotype images for various rice accessions. In particular, the database provides tools to construct haplotype networks and design PCR-primers by taking into account surrounding known genomic variations.

What about the phenotypes?

OPEN  ACCESS Freely available online



Perspective

Where Have All the Crop Phenotypes Gone?

Dani Zamir*

Faculty of Agriculture, The Hebrew University of Jerusalem, Rehovot, Israel

What about the phenotypes?

OPEN  ACCESS Freely available online

 PLOS | BIOLOGY

Perspective

Where Have All the Crop Phenotypes Gone?

Dani Zamir*

Faculty of Agriculture, The Hebrew University of Jerusalem, Rehovot, Israel

nature
genetics

EDITORIAL



Growing access to phenotype data

DivSeek: Harnessing crop diversity to feed the future

► Who is DivSeek?

Currently 58 partners worldwide:
Gene banks, breeders,
plant scientists, database and
computational experts



DivSeek: Harnessing crop diversity to feed the future

- ▶ Who is DivSeek?

Currently 58 partners worldwide:
Gene banks, breeders,
plant scientists, database and
computational experts



- ▶ Aim:

Understanding and utilizing genetic variation in plant improvement

DivSeek: Harnessing crop diversity to feed the future

- ▶ Who is DivSeek?

Currently 58 partners worldwide:
Gene banks, breeders,
plant scientists, database and
computational experts



- ▶ Aim:

Understanding and utilizing genetic variation in plant improvement

- ▶ Mission statement:

"Characterize crop diversity and develop a unified, coordinated and cohesive information management platform to provide easy access to genotypic and phenotypic data associated with genebank germplasm."

<http://www.divseek.org>

Acknowledgments

- ▶ Andreas Graner
- ▶ Nils Stein
- ▶ Uwe Scholz
- ▶ Andreas Börner
- ▶ Helmut Knüppfer
- ▶ Michael Grau

- ▶ Robbie Waugh
- ▶ Joanne Russell
- ▶ Ian Dawson

- ▶ Gary Muehlbauer

