The Gramene Database

A closer look at the Genetic Diversity Module

Genevieve DeClerck

Gramene Database Curator

Cornell University, Ithaca, NY

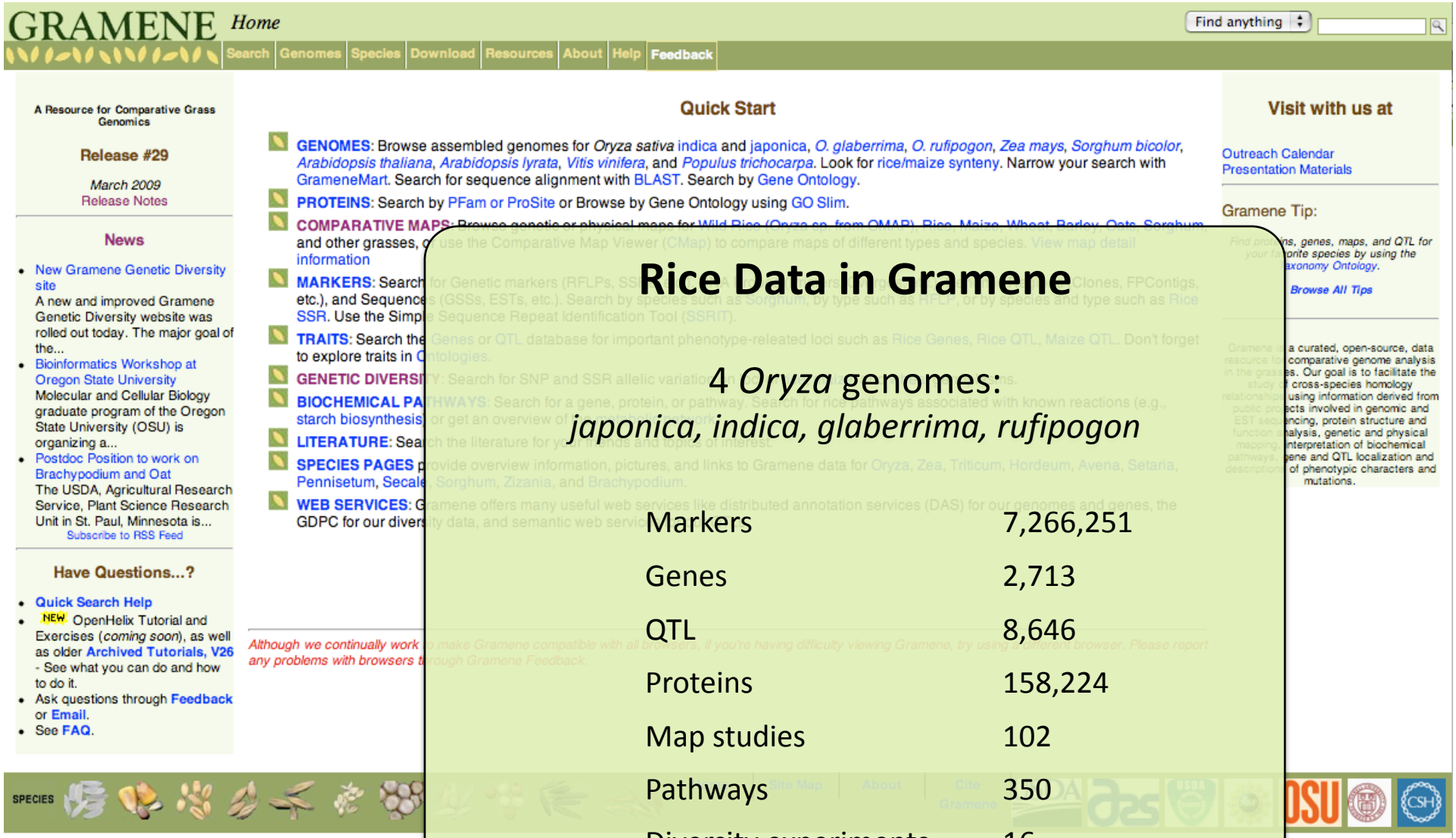
June 15,2009

USDA-ARS Dale Bumpers National Rice Research Center, Stuttgart, AR

www.gramene.org

www.gramene.org

www.gramene.org

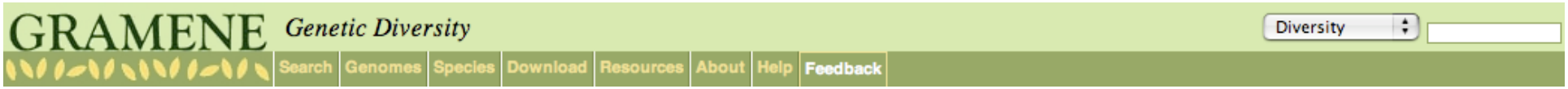


Rice Data in Gramene

4 *Oryza* genomes:
japonica, indica, glaberrima, rufipogon

Markers	7,266,251
Genes	2,713
QTL	8,646
Proteins	158,224
Map studies	102
Pathways	350
Diversity experiments	16

gramene.org/species/oryza/rice_stat.html



Gramene Genetic Diversity

- **Focus:** experiments which study the genetics of variation within populations of grasses and how this relates to observable traits, so that associations between genotype and phenotype can be discovered

- **Data types:** alleles, trait measurements, germplasm
- **Marker types:** QTL, SSR, RFLP, SNP
- **Species:** rice, maize, wheat, *Sorghum*, *Arabidopsis*

Gramene Genetic Diversity

GRAMENE *Genetic Diversity*

Diversity

Search | Genomes | Species | Download | Resources | About | Help | Feedback

Diversity Home | Tools | Tutorials | FAQ | Release Notes

Gramene: Genetic Diversity Database

Find: In: --Any-- Species: Rice

E.g., Germplasm "IRGC 3575", RA4969, Basmati 1, Marker/locus RM22.

Diversity Overview

The Gramene Diversity database contains genotype and phenotype experiments in rice, maize, wheat, sorghum, and soon, arabidopsis, which study the genetics of variation within populations and how this relates to quantitative and observable traits. Characterizing the genetic basis of variation within populations and linking this with observable traits provides an important framework for understanding population structure and evolutionary patterns, and for increasing the efficiency of plant breeding programs.

Database summary

	species	data types	experiments	germplasm	traits	genotype markers
rice	*Oryza sativa, O. glaberrima, O. rufipogon*	SNP, SSR, RFLP, phenotype data, QTL, passport descriptions	16	841	28	1617
maize	*Zea spp.*	SNP, SSR, RFLP, sequence data, phenotype data, passport descriptions	31	8037	150	13075
arabidopsis	*Arabidopsis thaliana*	*250K SNP chip, 362 lines assayed*	*2 pending*	*362 pending*	0	*1,126,177 pending*
wheat	*Triticum aestivum, Aegilops spp.*	SNP, sequence data, passport descriptions	1	48	0	4097
sorghum	*Sorghum* species	SNP, SSR, sequence data, Isozyme, RFLP, AFLP, INDEL polymorphism,CAPS (Cleaved amplified polymorphic site)	6	1993	0	544

Basic Search

- Search for rice experiments that contain a germplasm named Submit
- Search for rice experiments that contain a marker named Submit

Advanced Search

Perform an advanced search of the Genetic Diversity database with several parameters. Search scripts provided by the Panzea project.

- Germplasm Search
- Geographic Map Viewer
- Molecular Diversity Search
- Gene/Locus Search
- Phenotype Search
- Polymorphism Between Accessions

Tools

TASSEL - Java program for evaluating trait associations, evolutionary patterns, and linkage disequilibrium.

GDPC Browser - Java program for performing complex queries and data format manipulations.

Download data

- All Diversity datasets can be downloaded in bulk from the Gramene FTP archive. Files with 'diversity_' prefix contain the Diversity data.
- Also available are association data from the TASSEL program.

Gramene Genetic Diversity Module

GRAMENE *Genetic Diversity*

Diversity

Search | Genomes | Species | Download | Resources | About | Help | Feedback

Diversity Home | Tools | Tutorials | FAQ | Release Notes

Gramene: Genetic Diversity Database

Find: In: --Any-- Species: Rice

E.g., Germplasm "IRGC 3575", RA4969, Basmati 1, Marker/locus RM22.

Diversity Overview

The Gramene Diversity database contains genotype and phenotype experiments in rice, maize, wheat, sorghum, and soon, arabidopsis, which study the genetics of variation within populations and how this relates to quantitative and observable traits. Characterizing the genetic basis of variation within populations and linking this with observable traits provides an important framework for understanding population structure and evolutionary patterns, and for increasing the efficiency of plant breeding programs.

Database summary

	species	data types	experiments	germplasm	traits	genotype markers
rice	*Oryza sativa, O. glaberrima, O. rufipogon*	SNP, SSR, RFLP, phenotype data, QTL, passport descriptions	16	841	28	1617
maize	*Zea spp.*	SNP, SSR, RFLP, sequence data, phenotype data, passport descriptions	31	8037	150	13075
arabidopsis	*Arabidopsis thaliana*	*250K SNP chip, 362 lines assayed*	*2 pending*	*362 pending*	0	*1,126,177 pending*
wheat	*Triticum aestivum, Aegilops spp.*	SNP, sequence data, passport descriptions	1	48	0	4097
sorghum	*Sorghum* species	SNP, SSR, sequence data, Isozyme, RFLP, AFLP, INDEL polymorphism,CAPS (Cleaved amplified polymorphic site)	6	1993	0	544

Basic Search

- Search for rice experiments that contain a germplasm named Submit
- Search for rice experiments that contain a marker named Submit

Advanced Search

Perform an advanced search of the Genetic Diversity database with several parameters. Search scripts provided by the Panzea project.

- Germplasm Search
- Geographic Map Viewer
- Molecular Diversity Search
- Gene/Locus Search
- Phenotype Search
- Polymorphism Between Accessions

Tools

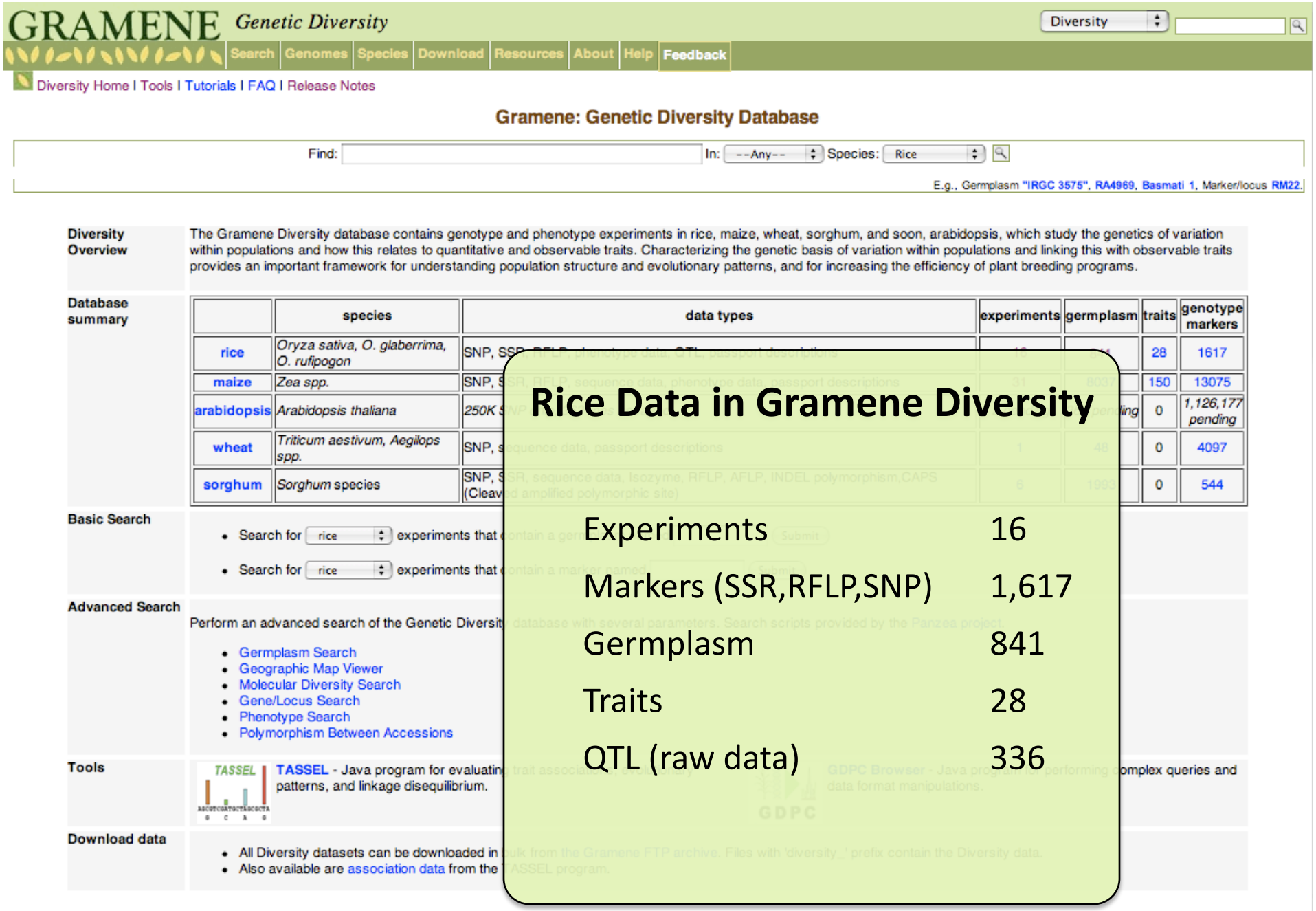
TASSEL - Java program for evaluating trait associations, evolutionary patterns, and linkage disequilibrium.

GDPC Browser - Java program for performing complex queries and data format manipulations.

Download data

- All Diversity datasets can be downloaded in bulk from the Gramene FTP archive. Files with 'diversity_' prefix contain the Diversity data.
- Also available are association data from the TASSEL program.

Gramene Genetic Diversity Module

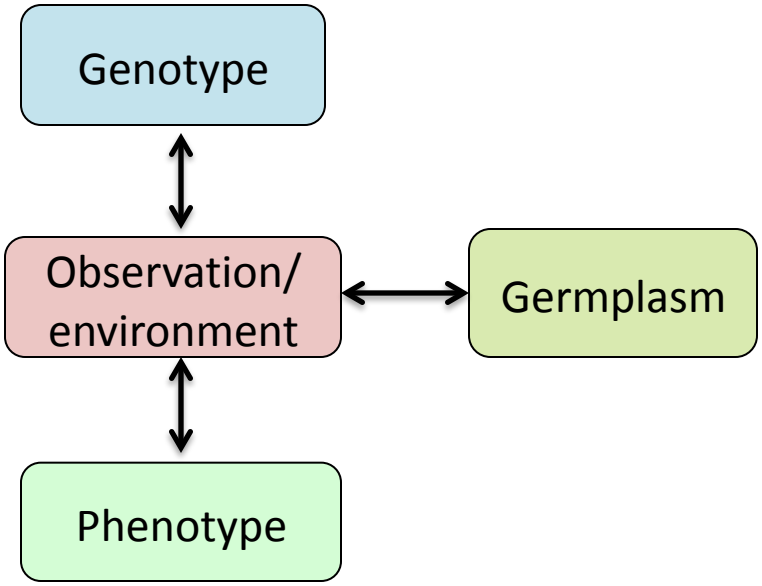


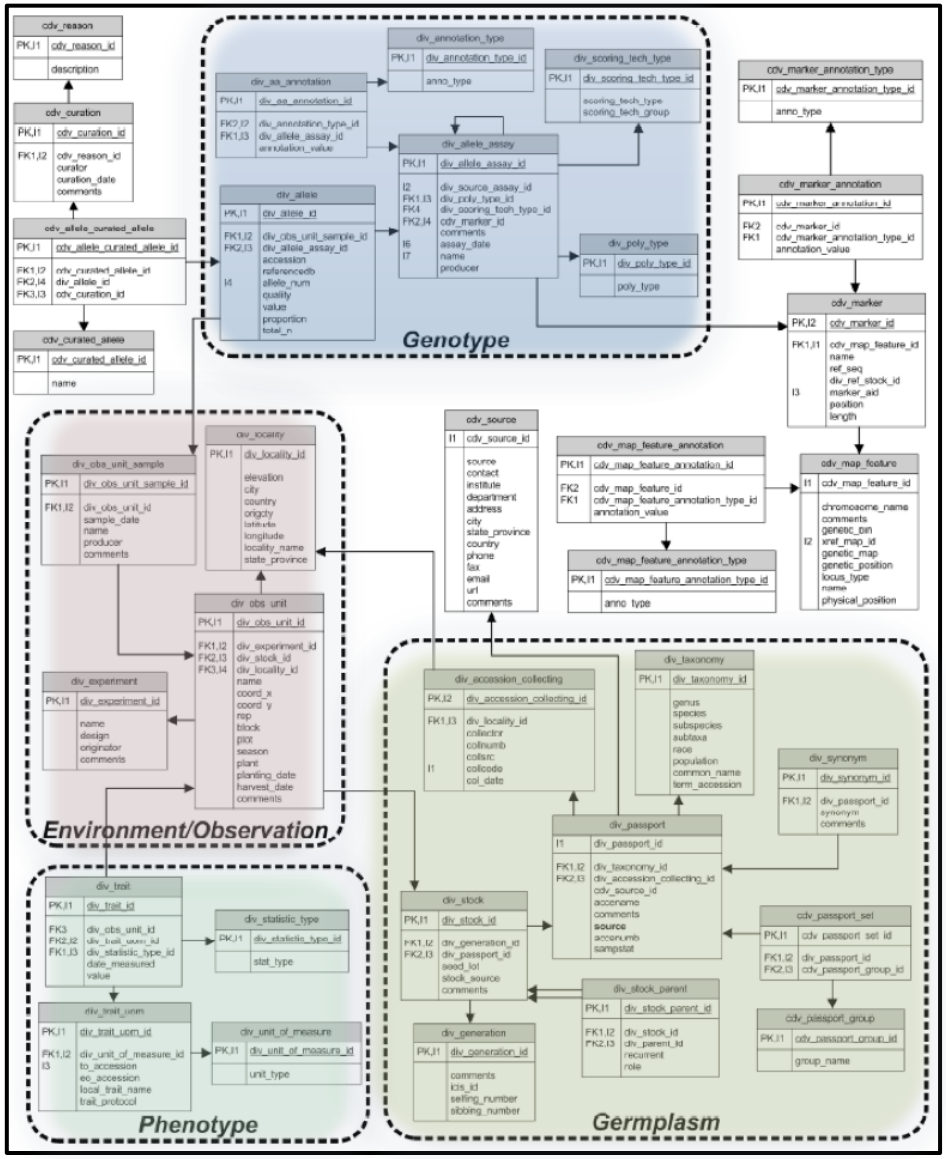
Rice Data in Gramene Diversity

Experiments	16
Markers (SSR,RFLP,SNP)	1,617
Germplasm	841
Traits	28
QTL (raw data)	336

Diversity Database Design

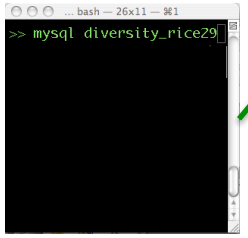
- **GDPDM schema** (**G**enomic **D**iversity and **P**henotype **D**ata **M**odel)





Diversity Database

Gramene website



Command line interface

Gramene Database MySQL RDBMS

Gramene modules

Gramene

Maps

Proteins

Markers

Pathways

Genes

Ontology

Diversity

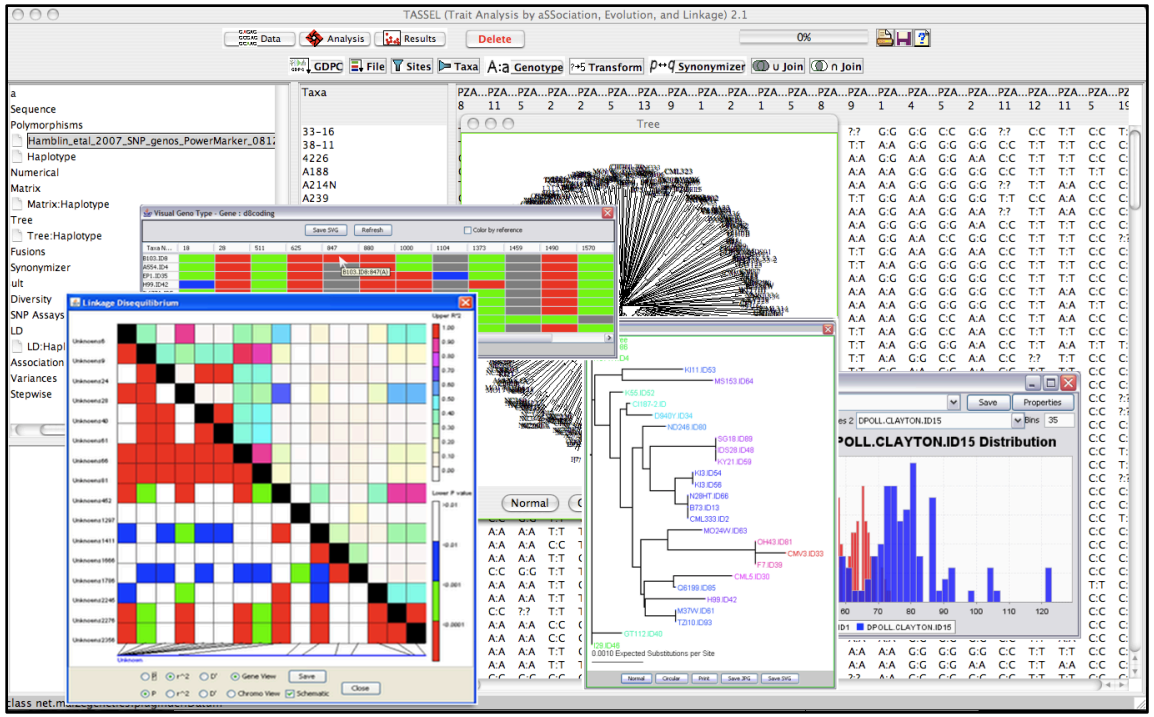
Literature

QTL

TASSEL

TASSEL

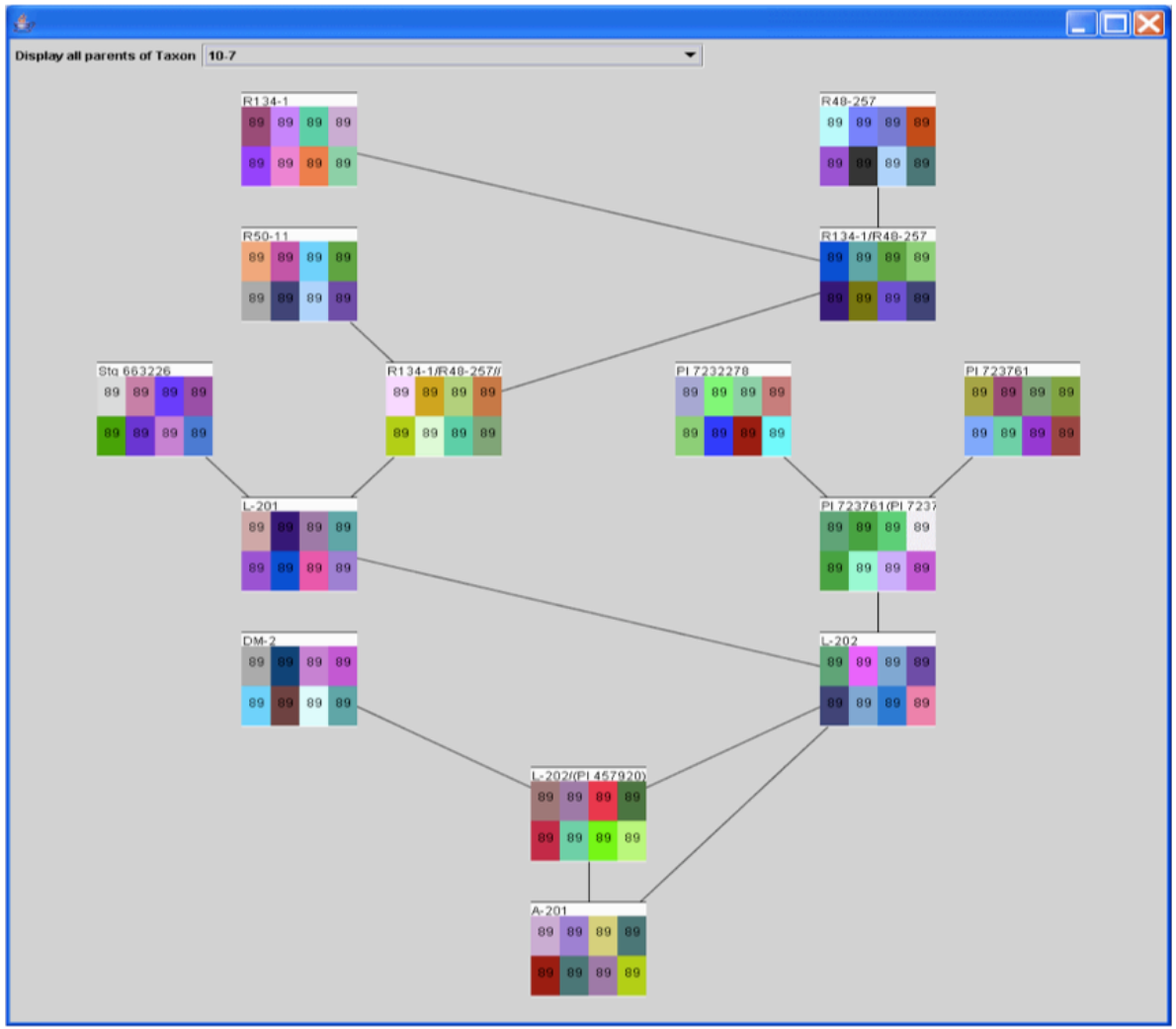
Trait **A**nalysis by a**SS**ociation, **E**volution and **L**inkage



- Evaluate trait associations
- Discover evolutionary patterns
- Calculate linkage disequilibrium
- Use new statistical approaches (GLM, MLM)*
- Stand-alone Java package
- Interfaces directly with Diversity database

* See paper "**Unified Mixed-Model Method for Association Mapping**", *Nat Genet*. 2006 Feb;38(2):203-8.

Pedigree viewer



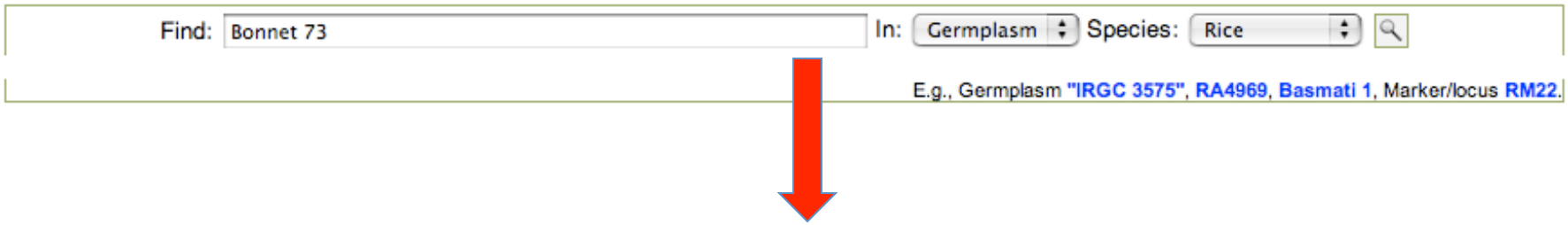
- Prototyped, proposed for TASSEL
- *network* diagram drawn on-the-fly based on data loaded in TASSEL
- Estimates of breeding values
- Highlight key polymorphisms

Demos…

- Diversity module
  - **Example #1**: *Search for allele profile of Bonnet 73*
  - **Example #2**: *View all rice germplasm records related to a particular trait*

Diversity example #1

*Search for allele profile of **Bonnet 73***



Oryza sativa Germplasm "Bonnet 73"

Accession Name	Bonnet 73
Institute(s) Holding Accession	IRRI, NSGC
Accession Number	Clor 9654
Synonym(s)	
Germplasm Type	Advanced/improved cultivar
Genus	Oryza
Species	sativa
Subspecies	japonica
Subtaxa	tropical japonica
Taxonomy Cross-Reference	GR_tax:013684

Germplasm record for "Bonnet 73"

Experiments (1)

Title	Design	Polymorphism Type	Allele Scoring Protocol	PI(s)	Comments	Data
Population Structure and Breeding Patterns of 145 U.S. Rice Cultivars Based on SSR Marker Analysis	115 U.S rice cultivars and 30 ancestral rice accessions introduced from Asia were genotyped by means of 169 genomic SSR markers.	N/A		McCouch-S-R, Tai-T-H	Population structure was inferred using genetic-distance method with 169 SSR markers. Reference: Lu et al., 2005.	Show Allele Data

Related Stocks

Diversity example #1

169 ***Bonnet 73*** *alleles found*

Items 1 to 25 of 169. Page 1 of 7 | Next

[Download]

Germplasm Accession Name	Germplasm Accession Number	Subsp. & subtaxa	Country of Origin	Stock Number	Locus name	Genotype	View All Genotypes on Marker
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	OSR13	98	View all "OSR13" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM1	88	View all "RM1" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM104	222	View all "RM104" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM105	129	View all "RM105" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM106	293	View all "RM106" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM108	81	View all "RM108" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM109	97	View all "RM109" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM11	129	View all "RM11" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM112	136, 142	View all "RM112" genotypes

Diversity example #1

*169 **Bonnet 73** alleles found*

Items 1 to 25 of 169. Page 1 of 7 | Next

[Download]

Germplasm Accession Name	Germplasm Accession Number	Subsp. & subtaxa	Country of Origin	Stock Number	Locus name	Genotype	View All Genotypes on Marker
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	OSR13	98	View all "OSR13" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM1	88	View all "RM1" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM104	222	View all "RM104" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM105	129	View all "RM105" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM106	293	View all "RM106" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM108	81	View all "RM108" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM109	97	View all "RM109" genotypes
Bonnet 73	Clor 9654	United States of America	japonica, tropical japonica	RA1792	RM11	129	View all "RM11" genotypes

```
germplasm_accession_name        country_of_origin       accession_number        stock_number    locus_name      genotype
Bonnet 73       United States of America        CIor 9654       RA1792  OSR13   98
Bonnet 73       United States of America        CIor 9654       RA1792  RM1     88
Bonnet 73       United States of America        CIor 9654       RA1792  RM104   222
Bonnet 73       United States of America        CIor 9654       RA1792  RM105   129
Bonnet 73       United States of America        CIor 9654       RA1792  RM106   293
Bonnet 73       United States of America        CIor 9654       RA1792  RM108   81
Bonnet 73       United States of America        CIor 9654       RA1792  RM109   97
Bonnet 73       United States of America        CIor 9654       RA1792  RM11    129
Bonnet 73       United States of America        CIor 9654       RA1792  RM112   136,
Bonnet 73       United States of America        CIor 9654       RA1792  RM116   279
Bonnet 73       United States of America        CIor 9654       RA1792  RM1167  175
Bonnet 73       United States of America        CIor 9654       RA1792  RM118   158
Bonnet 73       United States of America        CIor 9654       RA1792  RM1189  176
Bonnet 73       United States of America        CIor 9654       RA1792  RM119   164
Bonnet 73       United States of America        CIor 9654       RA1792  RM120   186
Bonnet 73       United States of America        CIor 9654       RA1792  RM122   229
Bonnet 73       United States of America        CIor 9654       RA1792  RM124   271
Bonnet 73       United States of America        CIor 9654       RA1792  RM125   126
Bonnet 73       United States of America        CIor 9654       RA1792  RM13    127
```

Text data dump of all 169 genotype measurements

Diversity example #1

All genotypes of marker 'RM1'

Items 1 to 25 of 145. Page 1 of 6 | Next

[Download]

Germplasm Accession Name	Germplasm Accession Number	Subsp. & subtaxa	Country of Origin	Stock Number	Locus name	Genotype	View All Genotypes on Germplasm
Badkalamakati		Bangladesh	indica, indica	RA4991	RM1	76	View all "Badkalamakati" genotypes
Lemont	PI 475833	United States of America	japonica	RA6270	RM1	86	View all "Lemont" genotypes
L-202	PI 483097	United States of America	japonica	RA6273	RM1	86	View all "L-202" genotypes
M-101	CI 9970	United States of America	japonica	RA5002	RM1	84, 88	View all "M-101" genotypes
Calrose 76		United States of America	japonica, temperate japonica	RA1810	RM1	84	View all "Calrose 76" genotypes
M-201	CI 9980	United States of America	japonica, temperate japonica	RA5008	RM1	88	View all "M-201" genotypes
Vegold	CI 9836	United States of America	japonica	RA5040	RM1	86	View all "Vegold" genotypes
Belle Patna	CI 9433	United States of America	japonica	RA1791	RM1	86	View all "Belle Patna" genotypes
Della	CI 9483	United States of America	indica	RA6389	RM1	86	View all "Della" genotypes

Diversity example #1

All genotypes of marker 'RM1'

Items 1 to 25 of 145. Page 1 of 6 | Next

[Download]

Germplasm Accession Name	Germplasm Accession Number	Subsp. & subtaxa	Country of Origin	Stock Number	Locus name	Genotype	View All Genotypes on Germplasm
Badkalamakati		Bangladesh	indica, indica	RA4991	RM1	76	View all "Badkalamakati" genotypes
Lemont	PI 475833	United States of America	japonica	RA6270	RM1	86	View all "Lemont" genotypes
L-202	PI 483097	United States of America	japonica	RA6273	RM1	86	View all "L-202" genotypes
M-101	CI 9970	United States of America	japonica	RA5002	RM1	84, 88	View all "M-101" genotypes
Calrose 76		United States of America	japonica, temperate japonica	RA1810	RM1	84	View all "Calrose 76" genotypes
M-201	CI 9980	United States of America	japonica, temperate japonica	RA5008	RM1	88	View all "M-201" genotypes
Vegold				RA5040	RM1	86	View all "Vegold" genotypes
Belle P				791	RM1	86	View all "Belle Patna" genotypes
Della				6389	RM1	86	View all "Della" genotypes

Oryza sativa Indica Group SSR "RM1"

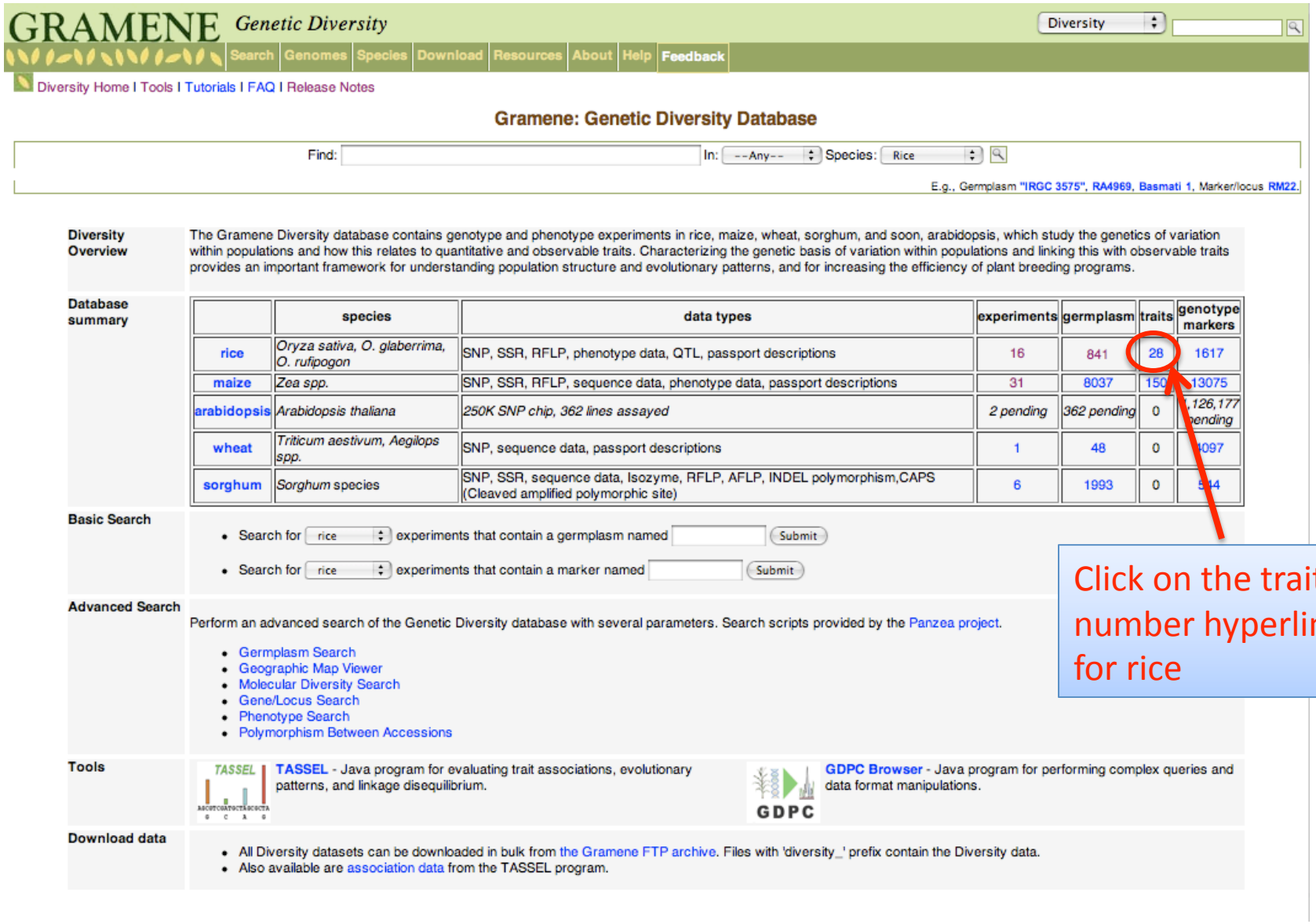
ID	24985466
Name	RM1
Synonyms (3)	GA12 RM001 RM1 (K)
Type	SSR
Species	Oryza sativa Indica Group (Indian rice) [GR_tax:013682]
Germplasm	IR36
Analysis	• fpc_loader • Rice_SSRs
Description	
Tags (0)	None

⊞ Source/Library
⊞ Details
⊞ Database Cross-References (1)
⊞ Sequences (5)
⊞ Map Positions (21)

Gramene Marker record for RM1

Diversity example #2

View all rice germplasm records related to a particular trait



Diversity example #2

View all rice germplasm records related to a particular trait

Traits 1 to 25 of 28. Page 1 of 2 | Next

[Download]

Local Trait Name	Trait Protocol	Trait Ontology Id	Unit Type	
Panicle number	Panicles per plant was calculated as the average number of panicles per plant at harvest.	TO:0000152	count	View
Germination rate	Percent germination was determined using 30 kernels that had been harvested 30 days after the panicle had flowered, dried at 45 °C for 4 days, soaked in water for 2 days, and then transferred to me...	TO:0000430	percent	View
Grains per panicle	Grains per panicle was measured as the average number of filled spikelets per panicle on the primary panicle of each plant.	TO:0000450	count	View
100-grain weight	Grain weight was measured, in grams, as the average weight of 100 kernels.	TO:0000269	g	View
Plant height	Plant height was measured in centimeters from the soil surface to the tip of the tallest panicle, excluding the awns.	TO:0000207	cm	View
Awn length	Awns were ranked in five categories, 1-5 scale, from no awns to awns over 1 cm in length.	TO:0000072	1-5	View

View trait details for "Grains per panicle"

Diversity example #2

View all rice germplasm records related to a particular trait

259 germplasm/stock records found related to “Grains per panicle”

Trait "Grains per panicle"

Trait Name	Grains per panicle
Trait Ontology ID	TO:0000450
Trait Protocol	Grains per panicle was measured as the average number of filled spikelets per panicle on the primary panicle of each plant.
Unit of measure	count

Experiments (1)

Related Germplasm (259)

Passport Accession Name	Seed Lot Name	Stock Source	Comments
Jefferson x Oryza rufipogon	BC2F2 Line 144	McCouch Lab	
Jefferson x Oryza rufipogon	BC2F2 Line 226	McCouch Lab	
Jefferson x Oryza rufipogon	BC2F2 Line 5	McCouch Lab	
Jefferson x Oryza rufipogon	BC2F2 Line 207	McCouch Lab	
Jefferson x Oryza rufipogon	BC2F2 Line 92	McCouch Lab	
Jefferson x Oryza rufipogon	BC2F2 Line 164	McCouch Lab	

Diversity example #2

View all rice germplasm records related to a particular trait

…and display the trait measurement data

On the same “Grains per panicle” record page, we can also view all the trait values measured for each stock

Trait "Grains per panicle"

Trait Name	Grains per panicle
Trait Ontology ID	TO:0000450
Trait Protocol	Grains per panicle was measured as the average number of filled spike … ch plant.
Unit of measure	count

Experiments (1)

Related Germplasm (259)

Trait Measurements (988)

Plant Observation Unit name	Experiment ID	trait_name	Value
BC2F2 Line 144	7	Grains per panicle	138.70
BC2F2 Line 226	7	Grains per panicle	94.50
BC2F2 Line 5	7	Grains per panicle	97.90
BC2F2 Line 207	7	Grains per panicle	155.00
BC2F2 Line 92	7	Grains per panicle	61.00
BC2F2 Line 164	7	Grains per panicle	156.00
BC2F2 Line 16	7	Grains per panicle	78.20
BC2F2 Line 127	7	Grains per panicle	101.60

Diversity example #2

View all rice germplasm records related to a particular trait

…and display the related experiments

Display all associated experiments with in Diversity with “Grains per panicle” trait measurements. There is only 1 related experiment in this case

Trait "Grains per panicle"

Trait Name	Grains per panicle
Trait Ontology ID	TO:0000450
Trait Protocol	Grains per panicle was measured as the average number of filled spikele plant.
Unit of measure	count

Experiments (1)

ID	Title	Design	Polymorphism Type	Allele Scoring Protocol	PI(s)	Comments	Data
7	Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson	Phenotypic evaluation was done in a greenhouse for 353 BC_2F_1 families and in three field trials for 258 BC_2F_2 families. Genotyping (104 SSRs and 49 RFLPs) was done on the 353 BC_2F_2 families	N/A		McCouch-S-R	Seventy-six QTLs were identified using interval mapping and composite interval mapping. A putative QTL was reported if detected in at least one environment at an experiment-wise significance threshold of $P < 0.01$. Reference: Thomson et al., 2003.	Show Trait Measurements

Related Germplasm (259)

Trait Measurements (988)

Your ideas

- What queries would you like to see?
- What tools do breeders need?
- Ideas about the website?
- What datasets would you like to see in Gramene Diversity?

Data formatting

- Regular (machine readable) format
  - Tab-delimited, csv
  - Uniform usage of delimiters

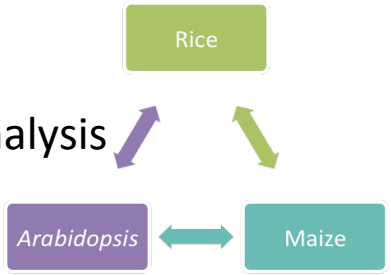
- Usage of controlled vocabulary/ontology
  - Gramene Trait Onotology (T.O.)
  - Universal identifiers (e.g. GRIN ID, dbSNP id)

“Bonnet 73” ≠ “bonnet73”

Gramene Diversity

upcoming

- Fine curation on all diversity data sets (ongoing!)
- Much more data!
  - **Rice**: NSF TV 44K SNP chip, 600K chip…
  - **Maize**: millions of SNPs from Maize HapMap project; large scale data from NAM population with over 5000 RILs. (panzea.org)
  - **Arabidopsis**: 250K SNP chip data; genotypes from 1001 genomes project
- Developing Germplasm module
- More analysis tools
  - Additional TASSEL modules (association stats, *pedigree networks*)
  - Ensembl Genome Browser (variation plots, heatmaps)
- Cross species association analysis
  - New Diversity member Charles Chen, statistical analysis



Acknowledgements…

- Susan McCouch, Ed Buckler, Doreen Ware, Pankaj Jaiswal
- The Gramene Team (especially Noel Yap, Ken Youens-Clark, Terry Casstevens, Qi Sun)
- Anna McClung, Georgia Eizenga, NSF-TV Project

- Gramene is a funded by a PGI grant from the National Science Foundation, an IFAFS grant from the USDA-CSREES, and was previously supported by a Cooperative Agreement from the USDA-ARS.