

A tutorial introduction into the MIPS PlantsDB barley&wheat databases

Manuel Spannagl&Kai Bader

transPLANT user training Poznan – June 2013

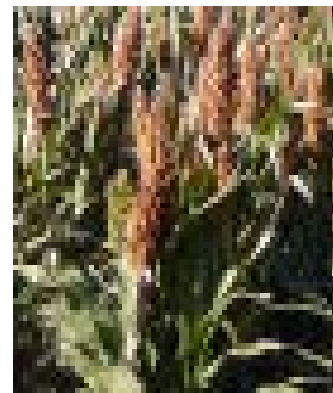


What is MIPS PlantsDB?

Generic database schema+system for the integration, management and (comparative) analysis of plant genome data

- modular design, extendable
- re-usable, flexible schema
- strong inter-connection with external databases
- standardized data retrieval
- tools available

MIPS PlantsDB – data content (Feb 2013)



...and many more internal instances.

What is stored in MIPS PlantsDB?

- whole genome sequences: pseudo-chromosome sequences if available, otherwise sequence contigs (also raw or assembled reads)
- gene and transcript annotation on these sequences
- non-coding transcripts/RNA, regulatory elements
- markers and genetic and physical maps
- repetitive elements and its annotation
- gene supporting evidences: fl-cDNAs, ESTs etc.
- transcriptomics data: RNA-seq
- comparative genome analysis results

What is NOT (yet) stored in MIPS PlantsDB?

- primary sequence data
- re-sequencing data and its results (under construction)
- phenotypic information
- metabolic/pathway information
- no community annotation interface

Access to MIPS PlantsDB data

- Web browser:
 - <http://mips.helmholtz-muenchen.de/plant/genomes.jsp>
- FTP:
 - <ftp://ftpmips.helmholtz-muenchen.de/plants/>
- BioMoby web services
 - Biomoby.org
- Java API (internal use)

Data content in MIPS PlantsDB triticeae instances - barley

- Morex WGS sequence assembly (+ add. varieties)
- Gene annotation (+ functional annotation)
- Genome zipper results
- Marker maps, genetic and physical maps, gene anchoring
- Synteny to Brachy, Sorghum and Rice – CrowsNest
- Orthologous gene families
- BLAST server @ IPK
- Batch download via FTP
- ...

Data content in MIPS PlantsDB triticeae instances - wheat

- UK 5x 454 Triticum aestivum WGS sequence assembly
- wheat gene sub-assemblies for ~ 20,000 grass orthologous gene representatives
- Sub-genome classification (A,B,D) for sub-assemblies
- BLAST server to search OGRs
- Genome zipper results for flow-sorted IWGSC wheat sequences (coming soon)
- Batch download via FTP
- ...

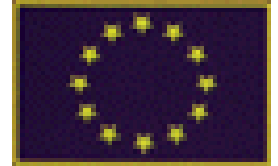
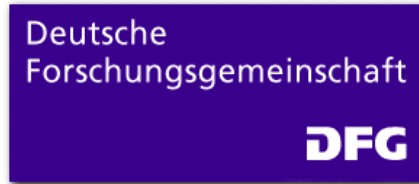
MIPS PlantsDB HELP

- General inquiries: Klaus Mayer
 - kmayer@helmholtz-muenchen.de
- Questions on MIPS PlantsDB content or functionality: Manuel Spannagl
- Inquiries for user tutorials/workshops etc.: Manuel Spannagl
 - manuel.spannagl@helmholtz-muenchen.de

The MIPS Plant team...



Acknowledgments&Funding



**All data providers and
collaboration partners...**



HelmholtzZentrum münchen
German Research Center for Environmental Health

MIPS PlantsDB tutorial

-

some exercises

Please go to:

<http://mips.helmholtz-muenchen.de/plant/genomes.jsp>

Tutorial objectives

- Search and navigate the MIPS PlantsDB instances
- Understand triticeae data concepts&contents and access
- Search and navigate the triticeae GenomeZippers
- Learn how to use the CrowsNest Synteny browser
- Use BLAST to identify candidate genes
- Download batch files to your local machine for further analysis

Plant Home

About us

PlantsDB

» transPLANT

» TriticeaeProject

» Lolium perenne

» Arabidopsis thaliana

» Zea mays

» Medicago truncatula

» Oryza sativa

» Sorghum bicolor

» Brachypodium distachyon

» Solanum lycopersicum

Projects

Jobs

Publications

Links

Helmholtz Home

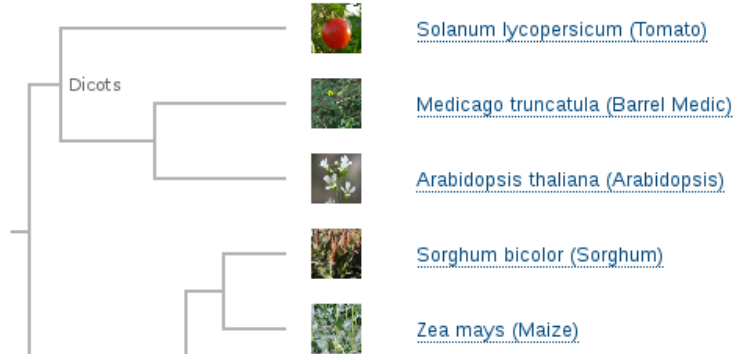
mips plant group

Access to species
databases

MIPS plants databases

The MIPS plant genomics group focuses on the analysis of plant genomes, using bioinformatic techniques. To store and manage the data, we developed a database, PlantsDB, that aims to provide a data and information resource for individual plant species. In addition PlantsDB provides a platform for integrative and comparative plant genome research.

PlantsDB databases:



mips
munich information center
for protein sequence

News

nature.com

An article about the repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres has been published online in [nature](#) on Dec. 20, 2012.

nature.com

Major Breakthrough in Deciphering

Exercise 1

Go to the MIPS PlantsDB home page at

<http://mips.helmholtz-muenchen.de/plant/genomes.jsp>

- Find all information available in PlantsDB about the **Barley** gene „MLOC_67600.1“
 - On what contig is it located?
- How many **Barley** genes do you find searching for „WUSCHEL-related homeobox“?

Exercise 2

Find details about a gene in the MIPS
PlantsDB barley DB

- Find all **barley** genes carrying the annotation „Alpha-glucosidase“ and open the gene report of the fourth hit (MLOC_68876.3)
- Identify the **putative orthologs** in other plant species and **putative paralogs** in barley – how many are there and for what organisms do you find some? (hint: use *gene family information*)
- **Download** the genes **protein and CDS sequence**

Exercise 3 - background

- *Brachypodium distachyon* is a model organism for related, but much more complex *triticeae* species such as wheat and barley.



Exercise 3

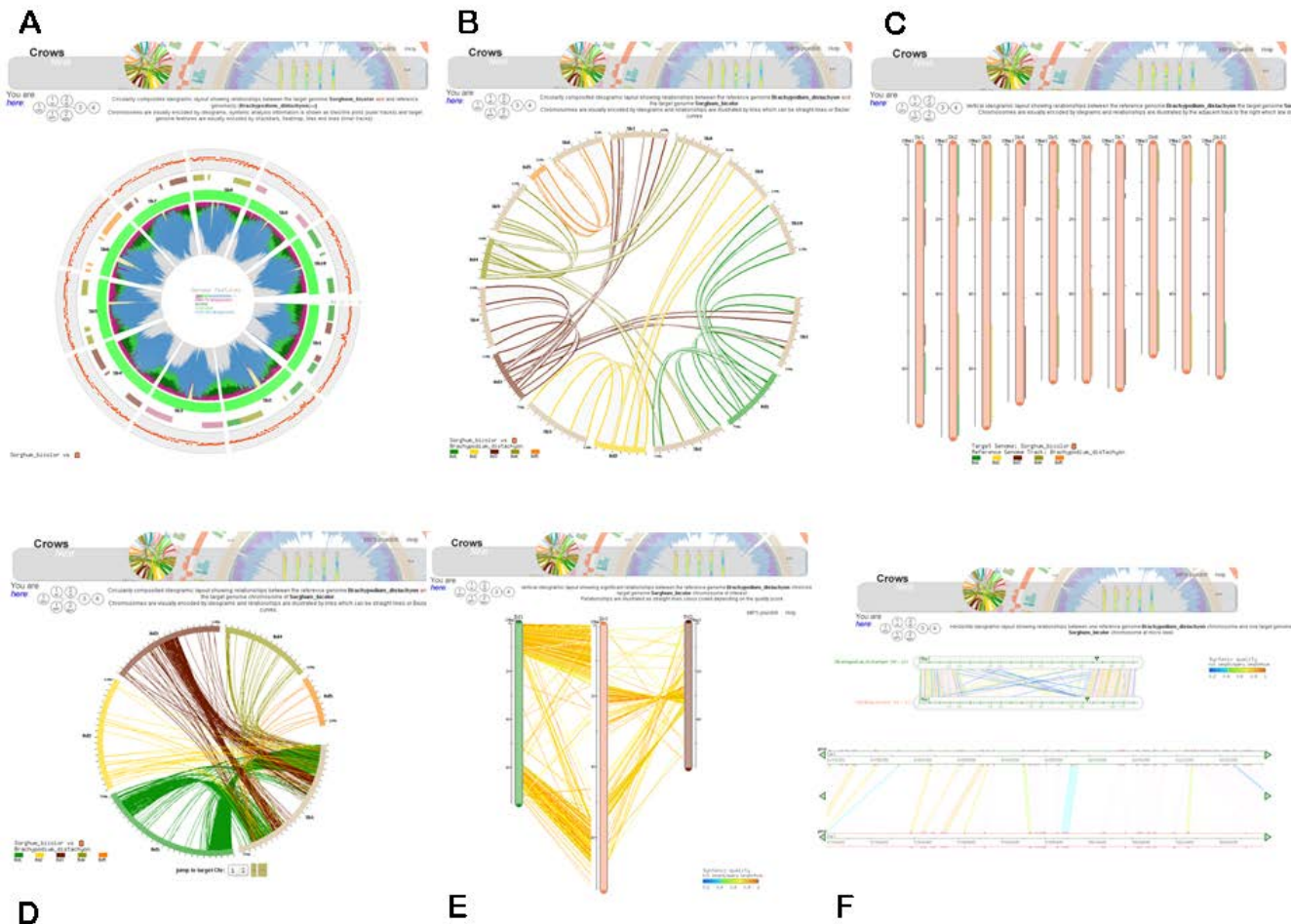
Query complex genomes using reference organisms – the barley „GenomeZipper“

- Can you identify the (approximate) location of the **barley** locus orthologous/syntenic to the **brachypodium gene „Bradi2g39900.1“** on the barley genome?
- What chromosome is it on, what is the appr. cM position? What marker is responsible for this anchoring?
- Are there any other genes in this syntenic position?

Exercise 4 - background

- Many related plant genomes share long stretches of conserved gene order: synteny
- This feature can be used to transfer knowledge from model to crop organisms
- Visualization of synteny is needed: CrowsNest tool
 - Displays pre-calculated syntenic regions
 - Browse from macro- to micro-synteny levels
 - Integration of add. features such as Ka/Ks, gene families, density and heatmap plots

Exercise 4 - background



<http://mips.helmholtz-muenchen.de/plant/crowsNest/index.jsp>

Exercise 4

Explore syntenic relationships between related plant genomes – the ***CrowsNest tool***

- Open the gene report of the Brachypodium gene from exercise 3: **Bradi2g39900.1**
- Start CrowsNest by clicking „*CrowsNest_SytenyToBarley*“ and explore the syntenic relationship of this gene region with barley
- What barley region is syntenic to the Bradi2g39900.1 gene region, what chromosome and what gene? Is there an inversion on that chr.?

Exercise 5

Search for homologous genes in **barley** for your own sequence – BLAST tool

- Download the DNA sequence from:

ftp://ftpmips.helmholtz-muenchen.de/plants/user_training/UnknownGrassSeq.fa

- Use BLAST to identify the best matching barley (HC) gene model for this sequence:

<http://webblast.ipk-gatersleben.de/barley/viroblast.php>

- Remember the found barley gene identifier and search for it in MIPS PlantsDB barley

Exercise 5 continued

Search for homologous genes in **barley** for your own sequence – BLAST tool

- What kind of function does the barley gene likely perform?
- What functional domains (PFAM, Interpro) are annotated for that gene?

Exercise 6

Query complex genomes using reference organisms – extract **wheat** genic sequence

- Download the DNA sequence from:

ftp://ftpmips.helmholtz-muenchen.de/plants/user_training/UnknownGrassSeq.fa

- Use BLAST to identify the best matching representative grass gene model for this sequence:

<http://mips.helmholtz-muenchen.de/plant/wheat/uk454survey/searchjsp/index.jsp>

- Extract&download the wheat sub-assembly sequences for the identified rep. Grass gene

Exercise 6 continued

Query complex genomes using reference organisms – extract **wheat** genic sequence

- Open the downloaded file
- What kind of information do you find encoded in the gene fragment identifiers?

Exercise 7

Download bulk genome data from MIPS PlantsDB

- Download all the protein sequences annotated for the barley genome
- Is there any categorization for these genes (such as confidence classes)?
- What kind of additional data is available as bulk download for barley?