

**MIPS PlantsDB –  
infrastructure for the representation and  
analysis of plant genome data**

Klaus Mayer

# BioGreenformatics: From Models to Crops, from Pets to Beasts



*Arabidopsis thaliana* (Nature, 2000) *Medicago* (Nature, 2011) *Tomato* (Nature, 2012)

*Sorghum* (Nature, 2009) *Brachypodium* (Nature, 2010) *Maize* (Genome Research, 2006, Plant Phys. 2008, PNAS 2008)

*Arabidopsis lyrata* (Nature Genetics 2011) *Physcomitrella patens* (Science 2008) *Aegilops tauschii* (Nature, 2013)

*Oryza* (Genome Research 2001; Nature Genetics 2014) *Lolium* (Plant Phys 2013) *Festuca* (submitted) *Spirodella* (Nature

Comm., 2014) *Micromonas* (Science 2009) *Cardamine hirsuta* (submitted) *Rye* (PNAS 2012; Plant Cell 2014)

*Barley* (Plant Phys. 2009, Plant Cell 2011, Nature 2012) *Wheat* (Plant Cell 2011; Plant Journal 2012, Nature 2012, 4xScience 2014)

„Reference datasets are the infrastructure“ (P. Kersey)

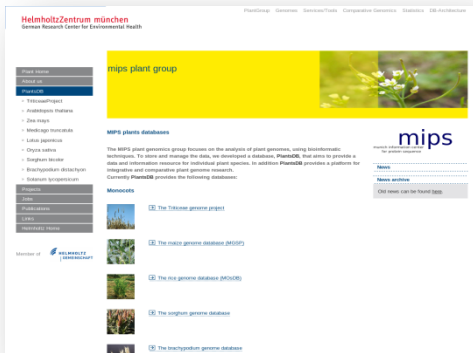
# What is MIPS PlantsDB?

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

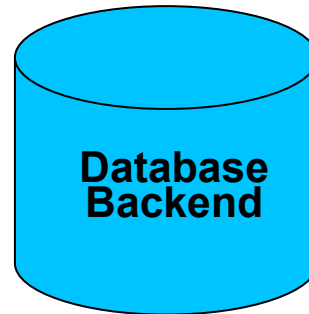
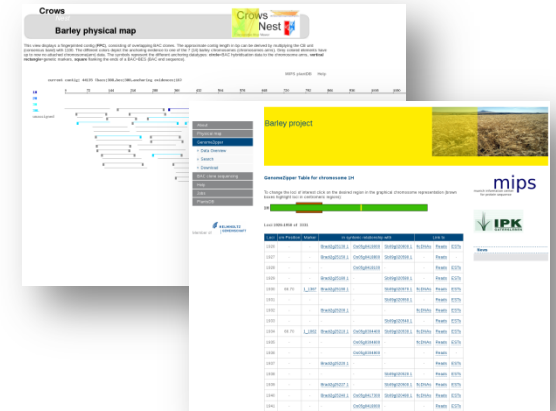
- modular design, extendable
- re-usable, flexible schema
- strong inter-connection with external databases
- standardized data retrieval
- tools available

# MIPS PlantsDB structure

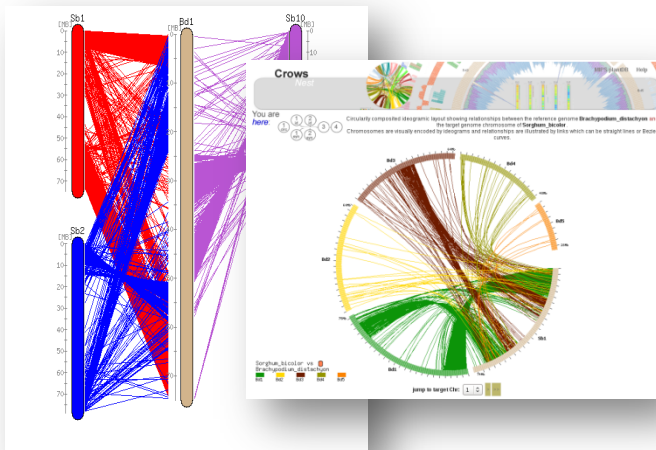
## Web interfaces & download



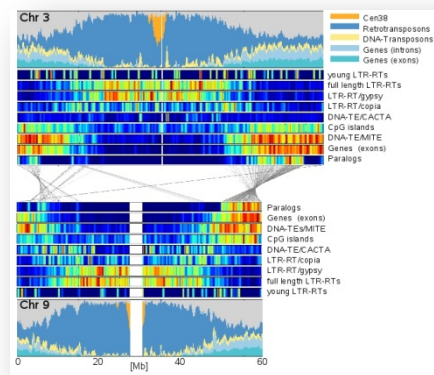
## Data integration & visualization



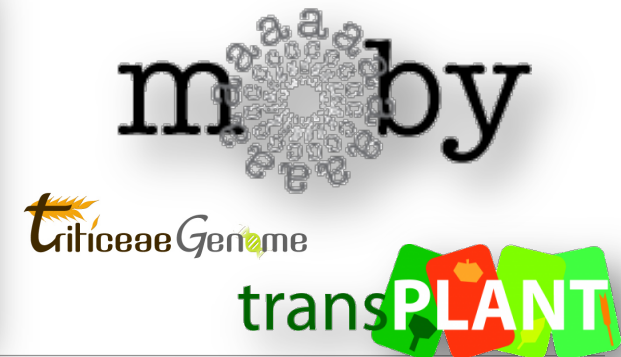
## Synteny & comparative genomics



## Genome Structure



## Web services & remote access

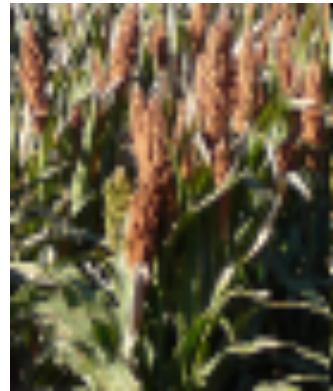




# What is stored in MIPS PlantsDB?

- 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

# MIPS PlantsDB – data content (Mar 2014)



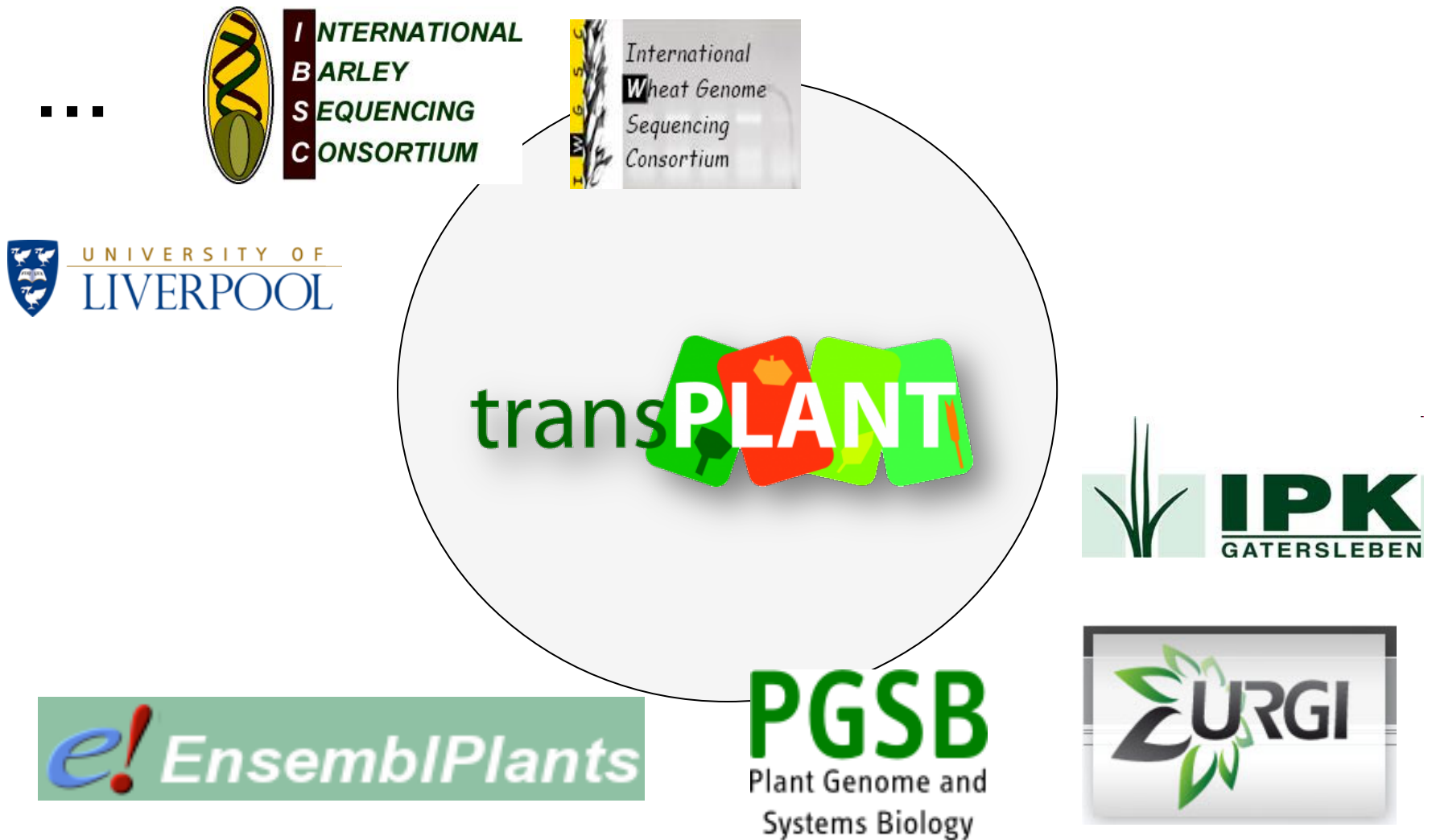
...and many more internal instances.

# MIPS PlantsDB – infrastructure for the representation and analysis of plant genome data

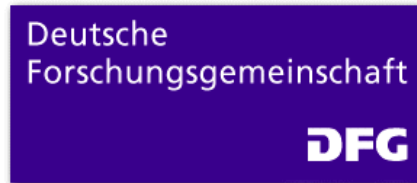
MIPS PlantsDB provides a framework for plant genome research via:

- Access to distributed resources via **BioMOBY web services** and **transPLANT** data exchange
- Portal for bulk download (**FTP**)
- Visualization of syntenic relationships in the **CrowsNest** tool
- Plant genome resources **registry**
- Data integration and visualization, especially in complex grass species such as wheat and barley: **Genome Zippers**, integrated physical and genetic maps, homology search, ...
- Search & Browse interfaces, ...

# MIPS PlantsDB – data exchange



# Funding



how do you plan to work to coordinate national activities?

what are you in the position to do for a European infrastructure?  
(either because you have funding in place, or because you are the recognised centre of expertise).

what is the green part of the german bioinformatics network planning to do?

what resources are you committed to maintaining?

what emergent demand do you anticipate?

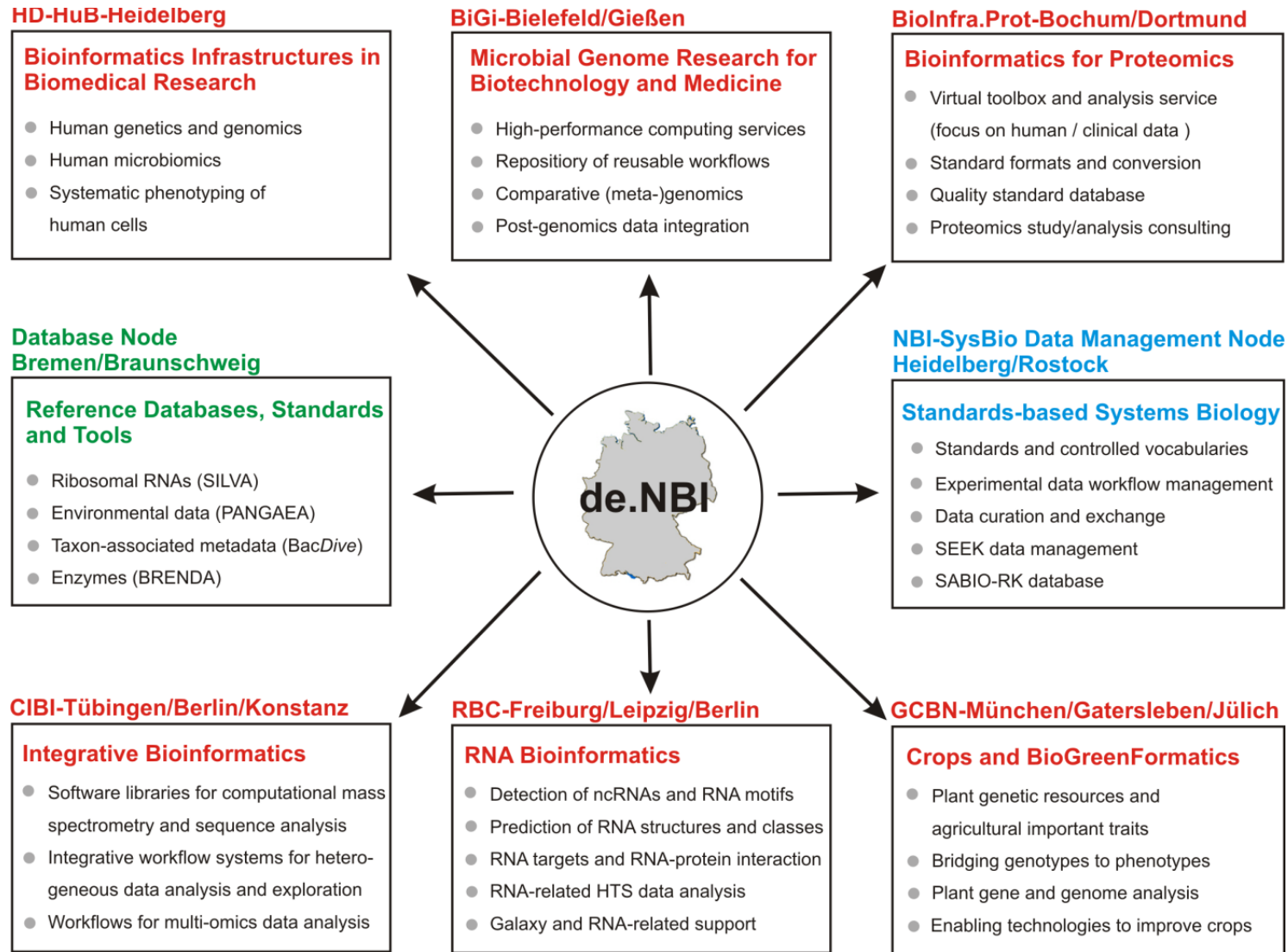
how do you plan to work to coordinate national activities?

what are you in the position to do for a European infrastructure (either because you have funding in place, or because you are the recognised centre of expertise).



## **De.NBI Mission statement**

*The ‘German Network for Bioinformatics Infrastructure’ provides comprehensive first-class bioinformatics services to users in basic and applied life sciences research. The de.NBI program coordinates bioinformatics training and education in Germany and the cooperation of the German bioinformatics community with international bioinformatics network structures.*



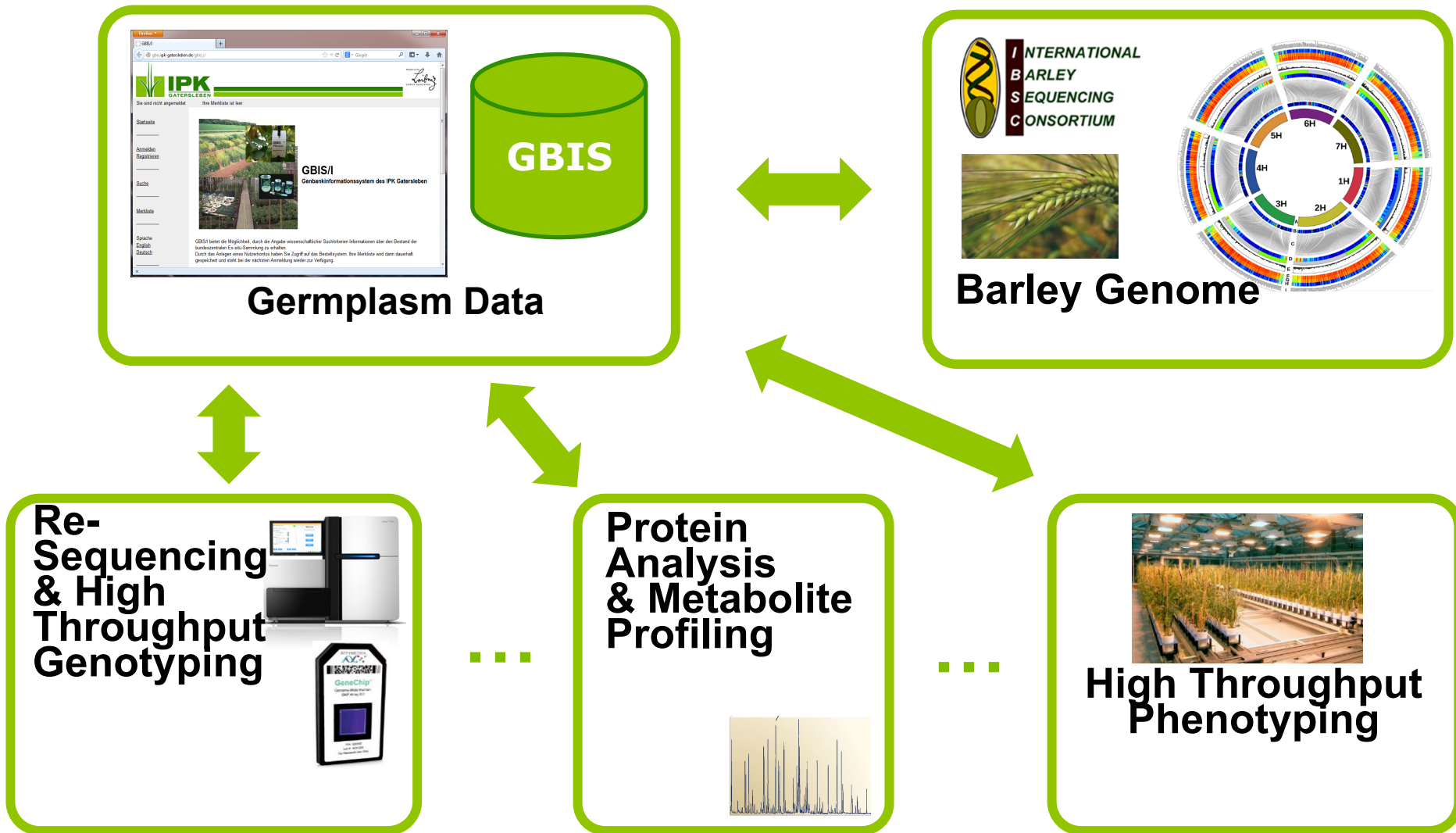
Overview of the research topics and the related services offered by the units of the ‘German Network for Bioinformatics Infrastructure (de.NBI)’. Each unit of the network, **including six service centers (red)** and two local data resource nodes (green, blue), provides bioinformatics services in a defined field of scientific expertise, thereby covering many areas of life sciences research.

## GCBN - Planned Work Packages

HMGU, IPK, FZJ/RWA

- WP: Transparent Access to Germplasms and Germplasm metadata (lead IPK)
- WP: Bridging Multiple Genotypes to Phenotypes (Lead HMGU)
- WP: Improved Workflows for Plant Gene Annotation (Lead HMGU)

# WP: Transparent Access to Germplasms and Germplasm metadata: Example Barley



what is the green part of the german bioinformatics network planning to do?

what resources are you committed to maintaining?

what emergent demand do you anticipate?

how do you plan to work to coordinate national activities?

what are you in the position to do for a European infrastructure (either because you have funding in place, or because you are the recognised centre of expertise).

# Some Hypothesis

- **Connectivity of infrastructure and research essential to develop need and research oriented platforms**
  - Embedding and tight connectivity of bioinfo platforms into nat./internat. genome-oriented/phenome-oriented/functional oriented etc. consortia; also as back-ends/front-end of the consortia
  - Need for a precise definition of „the community“. „The community“ often has a non-structured, amorphous swarm phenotype... => generic solutions often difficult or not adequate
- **Federation of genome, phenome etc. datasets to also buffer for funding bottlenecks and the death of dbs.**
  - However data asymmetry problems remain



# Some Hypothesis cont.

- **What is the limiting factor in a national and transnational platforms?**
  - Technology, ego`s, funding constraints...?
- **Is top-down preferable to bottom-up?**
  - Can a platform with nodes be decided top-down or is rather a process that generates new important resources at often unforeseeable places due to research projects that disseminate data in new and important areas. => Dynamic organisation of network to allow for „evolutionary novelty“
- **How to measure succesful service?**
  - What would be adequate means of „tipping“?

# 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](http://Biomoby.org)
- Java API (internal use)

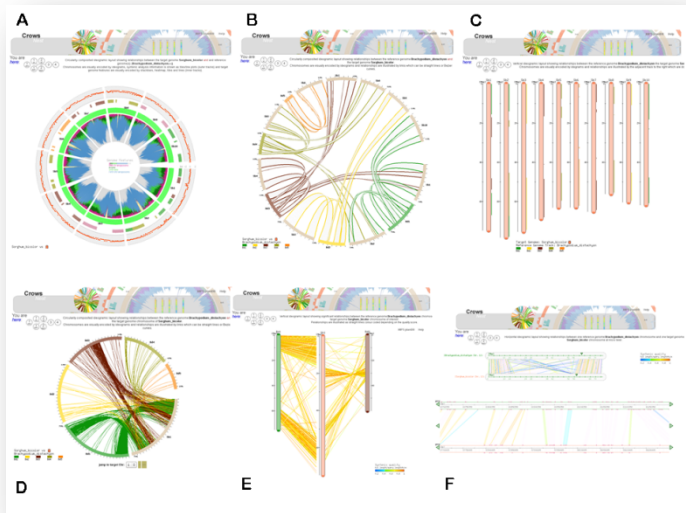
# MIPS PlantsDB – cereal genomes

**Barley project**

**GenomeZipper Table for chromosome 3H**

To change the list of interest (30k) on the desired region in the graphical chromosome representation (green boxes highlight list of cartimenes regions)

Contig	Start (kb)	End (kb)	Contig	Start (kb)	End (kb)	Region	Repeats	ESTs
1811			Brahtg25132.1	0	5813	30k	Repeats	ESTs
1812			Brahtg25132.1	5813	11626	30k	Repeats	ESTs
1813			Contig3418100			30k	Repeats	ESTs
1814			Brahtg25138.1			30k	Repeats	ESTs
1815	66.70	L_3387	Brahtg25138.1			30k	Repeats	ESTs
1816			Contig205378.1			30k	Repeats	ESTs
1817			Contig205358.1			30k	Repeats	ESTs
1818			Brahtg25208.1			30k	Repeats	ESTs
1819			Contig205443.1			30k	Repeats	ESTs
1820	66.70	L_1582	Brahtg25238.1			30k	Repeats	ESTs
1821			Contig204480			30k	Repeats	ESTs
1822			Contig204480			30k	Repeats	ESTs
1823			Contig204480			30k	Repeats	ESTs
1824			Brahtg25228.1			30k	Repeats	ESTs
1825			Contig205228.1			30k	Repeats	ESTs
1826			Brahtg25277.1			30k	Repeats	ESTs
1827			Contig205505.1			30k	Repeats	ESTs
1828			Brahtg25248.1			30k	Repeats	ESTs
1829			Contig418100			30k	Repeats	ESTs

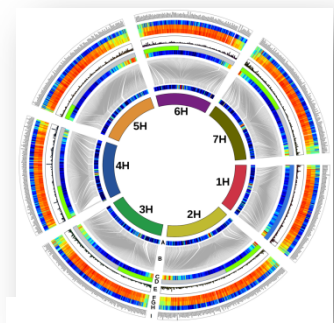


Genome Zippers for barley, wheat, rye and lolium

CrowsNest syteny browser for barley, *Ae. tauschii*, Brachypodium, sorghum and rice

Name	Größe	Zuletzt verändert
1112_P10500_2ACR08E_3M012_2121A	223081 kb	09.08.2012 00:00:00
112031_4ACR08E_2M012_2AC27A	3724 kb	09.08.2012 00:00:00
112032_4ACR08E_2M012_2AC27A	46 kb	09.08.2012 00:00:00
112033_4ACR08E_2M012_2AC27A	128841 kb	09.08.2012 00:00:00
112034_4ACR08E_2M012_2AC27A	293 kb	09.08.2012 00:00:00
112035_4ACR08E_2M012_2AC27A	128841 kb	09.08.2012 00:00:00
112036_4ACR08E_2M012_2AC27A	293 kb	09.08.2012 00:00:00
112037_4ACR08E_2M012_2AC27A	128841 kb	09.08.2012 00:00:00
112038_4ACR08E_2M012_2AC27A	293 kb	09.08.2012 00:00:00
112039_4ACR08E_2M012_2AC27A	128841 kb	09.08.2012 00:00:00
112040_4ACR08E_2M012_2AC27A	293 kb	09.08.2012 00:00:00
112041_4ACR08E_2M012_2AC27A	128841 kb	09.08.2012 00:00:00
112042_4ACR08E_2M012_2AC27A	293 kb	09.08.2012 00:00:00
112043_4ACR08E_2M012_2AC27A	128841 kb	09.08.2012 00:00:00
112044_4ACR08E_2M012_2AC27A	293 kb	09.08.2012 00:00:00
112045_4ACR08E_2M012_2AC27A	128841 kb	09.08.2012 00:00:00
112046_4ACR08E_2M012_2AC27A	293 kb	09.08.2012 00:00:00
112047_4ACR08E_2M012_2AC27A	128841 kb	09.08.2012 00:00:00
112048_4ACR08E_2M012_2AC27A	293 kb	09.08.2012 00:00:00
112049_4ACR08E_2M012_2AC27A	128841 kb	09.08.2012 00:00:00
112050_4ACR08E_2M012_2AC27A	293 kb	09.08.2012 00:00:00

Search, browse and download cereal genome data



Data integration layer  
Physical & genetic maps  
Variation data

...

# Data content in MIPS PlantsDB *triticeae* instances - **barley**

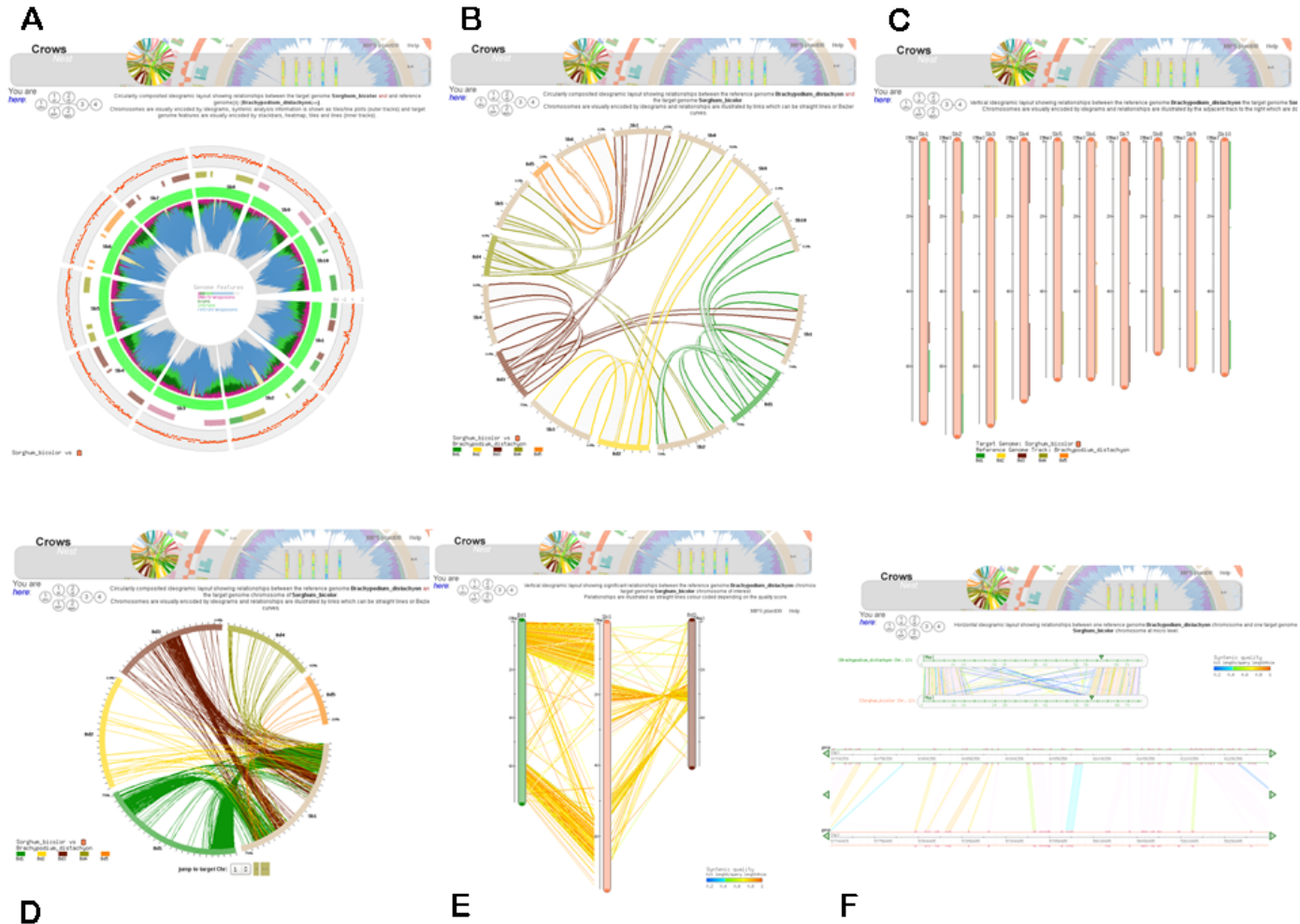
- Morex WGS sequence assembly (+ add. varieties)
- Gene annotation (+ functional annotation)
- Genome zipper results
- Marker maps, genetic and physical maps, gene anchoring
- Synteny to Brachy, Sorghum and Rice – CrowsNest tool
- Orthologous gene families
- BLAST server @ IPK
- Structured batch download 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 (under construction)
- IWGSC gene annotations (+functional annotation) (u.c.)
- IWGSC chromosome-sorted assembly sequences (u.c.)
- Batch download via FTP
- Synteny: *Ae. tauschii* - as a proxy for wheat - to barley, Brachy, Sorghum and Rice – CrowsNest tool



# Visualizing synteny – CrowsNest tool



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

# The MIPS Plant team...

**PGSB**  
Plant Genome and  
Systems Biology

