Bioinformatics @ IPK Gatersleben
## Applied Bioinformatics Research @ IPK

### Departments and Research Groups

- **Dept. Genebank**
  - Programmes: Genebank Documentation (GDK) Dr. Helmut Krüpfel
  - Research Groups: Genome Diversity (GED) Prof. Dr. Andreas Graner, Genomics of Genetic Resources (GGR) Dr. Nils Stern

- **Dept. Breeding Research**
  - Programmes: Breeding Informatics Prof. Dr. Jochen C. Reif
  - Research Groups: Quantitative Genetics (QG) Prof. Dr. Jochen C. Reif, Bioinformatics and Information Technology (BIT) Dr. Uwe Scholz

- **Dept. Molecular Genetics**
  - Programmes: Chromosome Structure and Function (CSF) Prof. Dr. Andreas Houben
  - Research Groups: Heterosis (HET) Prof. Dr. Thomas Altmann, Seed Development (SE) Dr. Hans Weber (temp.), Gene Regulation (GR) Prof. Dr. Thomas Altmann (temp.)

- **Dept. Physiology and Cell Biology**
  - Programmes: Experimental Taxonomy (ETX) Dr. Frank Blattner, Genomic Analysis Prof. Dr. Patrick Schellhorn
  - Research Groups: Pathogen Stress Genomics (PSG) Dr. Patrick Schweizer, Gene and Genome Mapping (GGM) Dr. Marion Röder, Molecular Plant Nutrition (MPE) Prof. Dr. Nicolaus von Wirén, Applied Biochemistry (ABC) Dr. Hans-Peter Mock, Structural Cell Biology (SZB) Dr. Michael Meiner, Plant Reproductive Biology (PRB) Dr. Jochen Klinke, Yeast Genetics (YEG) Prof. Dr. Gotthard Kurose

- **Independent Research Groups**
  - Programmes: Management and Evaluation Prof. Dr. Andreas Bürner, Satellite Collections North (TEN) Dr. Klaus Dehmer
  - Research Groups: Cryo- and Stress Biology (CSB) Dr. Manuela Nagel, Characterisation and Documentation Prof. Dr. Andreas Bürner, Resources Genetics and Reproduction (RGR) Dr. Andreas Bürner, Experimental Taxonomy (ETX) Dr. Frank Blattner

### Systems Analysis and Modelling

- **Coordination:** N.N.

### Biodiversity Informatics

- **Coordination:** Uwe Scholz

---

28 research groups in total at IPK
(as of October, 2017)
Network Analysis and Modelling

Group Leader: Prof. Dr. Thomas Altmann (temporary)

model → test of model → revised model

- metabolic network
  - *Z. mays*
  - increased auxin

- venation regulatory network

- signaling network
  - in photorespiration mutants
  - metabolite mediated

- cryopreservation, root formation, iron toxicity, spike development, C4 metabolism, biomass, drought, ...

- more time

- PR
  - immunity response

- ROS
  - pathogen

Cluster 1 (375 genes)
Cluster 2 (696 genes)
Cluster 3 (979 genes)
Cluster 4 (155 genes)
Cluster 5 (448 genes)
Cluster 6 (471 genes)
Cluster 7 (543 genes)
Cluster 8 (1023 genes)
Cluster 9 (1046 genes)
Cluster 10 (258 genes)
Cluster 11 (208 genes)
Cluster 12 (598 genes)

Days after drought

Relative expression [z-score]
Image Analysis
Group Leader: Dr. Evgeny Gladilin

Research Areas
- Advanced image analysis algorithms for automated phenotyping
- Software system development (IAP)
- Cooperation with the biological research groups

IAP (Integrated Analysis Platform)
- Systems biology cloud storage, analysis and visualization system
- Web-based GUI, command line client
- Integration of data sets (LemnaTec, greenhouse climate data, ...)
- Statistic analysis (R-based diagrams, PDF reports)

Image Pipline

Biodiversity Informatics - Systems Analysis and Modelling

Slide # 4
Bioinformatics @ Dept. of Physiology & Cell Biology

Dr. Anja Hartmann

- Modelling, analysis, simulation and visualization of biological processes using systems biology standards
- Integration and exploration of multi-omics data within the functional context of biological networks

**Multi-omics data** → **Tools + Methods** → **Visualization + Exploration**

- MetaCrop
- VANTED
Bioinformatics and Information Technology
Group Leader: Dr. Uwe Scholz

Research Data Management:

Sequence Analysis:

Information Systems and Retrieval:

Biodiversity Informatics - Systems Analysis and Modelling
Genebank Documentation
Group Leader: Dr. Helmut Knüpffer

Information systems for Plant Genetic Resources

PGR-related data analysis

Participation in international networks for PGR and biodiversity informatics

Domestication Genomics
Group Leader: Dr. Martin Mascher

Biodiversity Informatics - Systems Analysis and Modelling
Quantitative Genetics
Group Leader: Prof. Dr. Jochen C. Reif

Biometric data analyses:

QTL and association mapping:

Genome-wide prediction:

\[ Y = \mu + X_{\text{gen}} \alpha + e \]

Estimation of marker effects

Prediction of genotypic values:

\[ \hat{Y} = X_{\text{unintegrated}} \hat{\alpha} \]
Computational Resources (as of October, 2017)

Hardware

• Compute power:
  11 SMP machines with total:
  o 912 physical cores
  o 9 Terabyte main memory

• Storage capacity:
  • Network attached storage capacity: 646 Terabyte
  • Storage and archive capacity: 1.1 Petabyte

• Network: 2 x 200 MBits internet connection, internal 10 GBits backbone

Software

• Oracle Database
• Oracle WebLogic Application Server
• Various web-based information systems
• Galaxy Infrastructure for IPK internal use
• Limsophy Laboratory information management system
• eIDAL Data publication infrastructure
• IAP The Integrated Analysis Platform for high throughput plant image analysis
Bioinformatics @ IPK - Training Activities

Recommendation of online trainings:
- Primer in Linux - http://linuxsurvival.com/
- Primer in R - http://tryr.codeschool.com/
- Primer in regular expressions - http://regexone.com/

Annual BioEXCEL courses:
- Excel-1: Basic excel functions in logic, string manipulation, calculation, and advanced logic for the analysis of -omics data
- Excel-2: Pivot tables for the analysis of -omics data

(Bi)-annual R courses:
- R-1: Basic R functions
- R-2: Statistics with R
- R-3: Quantitative genetics with R

Training courses in the framework of collaborative projects as example:
- deNBI User Training – PLANT 2030 ACADEMY Summer School: BBB – Basic Bioinformatics training for Biologists
Establishment of a data management task force

- **Goal**: Development of a data management plan
- **Composition**: experimentalists/bioinformaticians from all departments and scientific data management, LIMS and bioinformatics coordination

Support the pipeline from experiment to publication

- **Samples**
- **Measurement**
- **Quantification**
- **Analysis**
- **Publication**

Development of integrated infrastructure for standardized data

- **Assessment of Data Types**
  - Pheno-CE
  - Pheno-F
  - RNA-seq
  - Genome-seq
  - metabolomics
  - ...

- **Storage Solutions**
  - Experimental Descriptions
  - Raw Data
  - Processed Data
  - Interpreted Data

- **Access Solutions**
  - Search
  - Restricted Access
  - Material IDs
  - Data Publication

Limsophy
Bioinformatics Publications
2012 – 2017 (Jul)

# IPK Pubs with IF: 891
Sum IFs: 4,314.2

# Pubs with BI: 190 (21%)
Sum IFs with BI: 1,097.6 (25%)

BI = BA+BIT+DG+DI+DOK+NAM+PBI+SYS
Bioinformatics @ IPK - Networking Activities

Infrastructure: de.NBI

Crop Genome Sequencing Consortia:
- International Barley Sequencing Consortium
- International Wheat Genome Sequencing Consortium

Data Publication:
- DataCite

Plant Genetic Resources and Biodiversity:
- DivSeek
- iDiv
- Biodiversity Information Standards TDWG
- GBIF
- europeana

Plant Phenotyping Networks & Standards:
- DPPN
- EPPN 2020
- IPPN
- Emphasis
- miappe
- isatab

German Informatics Society:
- FaBI Fachgruppe Bioinformatik

Biodiversity Informatics - Systems Analysis and Modelling