We have organised a bus transfer leaving from two central places in Quedlinburg (bus stop at Carl-Ritter-Platz and Gartenstraße near Hotel Acron, please check the map) to the institute in Gatersleben.

**Departure** times are: Carl-Ritter-Strasse: **8:30 a.m.**
Gartenstraße (close to hotel Acron): **8:35 a.m.**

After the meeting, the bus will drop all the participants at those two places.
Dear participants of the GRC2019-ABC,

it is a real pleasure to welcome you to the 15th Gatersleben Research Conference on Applied Bioinformatics in Crops at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben. It is the first time that the Gatersleben Research Conference is explicitly focusing on the importance of bioinformatics in plant science. In the age of big data, the analysis, storage and publication of the ever-increasing amounts of data in life sciences requires the power of bioinformatics and related information technologies and standards.

Likewise, crop plant research has to face the Big Data challenge. Current high-throughput technologies generate large quantities of high-dimensional data, which are scattered across hundreds of unstructured data sets, biological databases and thousands of scientific journals. It is the responsibility of Applied Bioinformatics to capture, model, integrate, analyze, visualize, and make those datasets accessible in a FAIR (Findable – Accessible – Interoperable – Reusable) way to provide new and more profound insights into complex biological systems like crops.

The GRC2019-ABC conference brings together Life Scientists, Bioinformaticians, Computer Scientists, Systems Biologists, Synthetic Biologists and others working or interested in the developing area of Applied Bioinformatics for crops. The meeting will provide an excellent environment and a range of opportunities to present and discuss methods, theoretical approaches, and their practical applications in the fields of ‘Breeding Informatics’, ‘Biodiversity and Information Systems’, ‘Distributed Computing, Tools and Infrastructures’, ‘Image-Based Data Analyses and Data Visualization’ and ‘Systems Biology and Modeling’.

We would like to thank Deutsche Forschungsgemeinschaft (DFG), the German Network for Bioinformatics Infrastructure (de.NBI), EXLIXIR Germany, KWS Saat SE, SGS TraitGenetics and Green Gate Gatersleben for their financial support to this meeting.

Please enjoy the scientific program of this meeting as well as the social event at the UNESCO World Heritage city Quedlinburg in Germany!

Uwe Scholz
on behalf of the Organizing Committee
Scientific Committee @GRC2019

Jan Baumbach @janbaumbach
Technical University of Munich

Micha Bayer
The James Hutton Institute

Evgeny Gladilin
IPK Gatersleben

Keywan Hassani-Pak @KeywanHP
Rothamsted Research

Theo van Hintum @TheOhAgain
WUR Wageningen

Astrid Junker @AstridJunker
IPK Gatersleben

Jens Keilwagen
Julius Kühn-Institut

Matthias Lange @lange_m_ger
IPK Gatersleben

Martin Mascher
IPK Gatersleben

Klaus Mayer
Helmholtz Zentrum München

Steffen Neumann @sneumannoffice
IPB Halle

Jesse Poland
Kansas State University

Jochen C. Reif
IPK Gatersleben

Falk Schreiber
Uni Konstanz

Uwe Scholz (chair) @UweScholz271
IPK Gatersleben

Heiko Schoof @heikoschoof
Uni Bonn

Burkhard Steuernagel @steuernb
John Innes Centre

Jedrzej J. Szymanski @JJSzymanski
IPK Gatersleben

Björn Usadel @usadellab
RWTH Aachen

Dirk Walther
MPI Golm

Stephan Weise
IPK Gatersleben

Bernd Weisshaar
Uni Bielefeld

Organizing Committee @GRC2019 @IPKGatersleben

Sibille Bettermann

Mary-Ann Blätke @MaryAnnBlaetke

Sebastian Beier @basti_beier

Regina Devrient @RDevrient

Nadine Küstermann

Uwe Scholz (chair) @UweScholz271

Uni Bielefeld
Monday, March 18th, 2019

09:00  Opening

SESSION 1 // BREEDING INFORMATICS
Chair: Jochen Reif

09:10  Keynote Lecture - Kelly Robins
Cornell University, Ithaca, NY, USA
Technology Driven Crop Improvement for Africa and South Asia

09:50  Keynote Lecture - Anne-Francoise Adam-Blondon
Université Paris-Saclay, Versailles, France
Developing Infrastructures for FAIR Data in Plant Biology

10:30  Coffee Break & Posters

11:00  Dario Copetti
Bielefeld University, Germany
Molecular Plant Breeding - ETH Zürich, Switzerland
A Phased Diploid Genome Assembly of Italian Ryegrass

11:20  Bianca Frommer
Bielefeld University, Germany
Targeting Haplotype Phasing: Assembling a Grapevine Rootstock Genome

11:40  Danara Ormanbekova
University of Bologna, Italy
Diversity Reduction and Selection Signature in Tetraploid Wheat
Germlasm from Wild Emmer to Modern Durum Wheat

12:00  Karin Iris Koehl
MPI of Molecular Plant Physiology, Potsdam, Germany
Tools for Drought Tolerance Selection in Potato

12:20  Lunch Break

13:20  Pick the speakers brain over coffee

13:40  Alice Fornasiero
Istituto di Genomica Applicata (IGA), University of Bologna, Italy
Single Primer Enrichment Technology as a Tool for Massive Genotyping: Bridging the Gap Between Arrays and Genotyping-by-sequencing

14:00  Jenny Hue/Tyng Lee
Justus Liebig University, Giessen, Germany
Genomic and Epigenomic Patterns in Novel Heterotic Pools of Winter Rapeseed

14:20  Sebastian Beier
IPK, Gatersleben, Germany
How to combine different molecular Marker Systems in Bread Wheat

14:40  Keynote Lecture - Andrea Bräutigam
Bielefeld University, Germany
Comparative transcriptional network analysis between model crop and model plant species

15:20  Keynote Lecture - Janine Felden
MARUM, University of Bremen, Germany
GFBio - A FAIR Infrastructure Network to Assist Scientists in Data Management

16:00  Coffee Break & Posters

16:30  Klaus Mayer
Helmholtz Center Munich, Germany
Applying Network-based Phylogenomics to Trace the Reticulate Ancestry of Modern Bread Wheats

16:50  Jeremy Harbison
Wageningen University, The Netherlands
Phenotyping and Genes for Photosynthesis

17:10  Thomas Schmutzer
Martin Luther University Halle-Wittenberg, Germany
Kmasker Plants – The In Silico Assistant To Avoid Drifting On The Slippery Path Of Repetitive Plant Sequences.

17:30  Ajit Singh
Rothamsted Research, Harpenden, UK
Update on KnellMiner FAIR developments

18:00  Poster session, Software demo with snacks & Finger Food

20:30  Transport to Quedlinburg

Tuesday, March 19th, 2019

08:30  Transport to Gatersleben

SESSION 3 // DISTRIBUTED COMPUTING, TOOLS AND INFRASTRUCTURES
Chair: Matthias Lange

09:00  Keynote Lecture - Barend Mons
Leiden University Medical Center, The Netherlands
Distributed learning by social machines in crop science, a peek in from health

09:40  Keynote Lecture - Björn Grüning
University of Freiburg, Germany
Integrative Bioinformatics with Galaxy Building Frameworks to Serve the 21st Century Data Science Problems

10:20  Coffee Break & Posters

11:00  David Marshall
James Hutton Institute & Scotslands Rural College (SRUC), UK
Software Development for Applied Crop Bioinformatics

11:20  Jens Keilwagen
Julius Kühn-Institut (JKI), Quedlinburg, Germany
Combining RNA-seq Data and Holmology-based Gene Prediction

11:40  Daniel Arend
IPK, Gatersleben, Germany
From FAIRbar Data Do Faster Discovery - A Comprehensive Infrastructure To Serve Plant Phenomic Research Data

12:00  Claudius Greih
Martin-Luther University Halle-Wittenberg, Germany
Detecting Food Fraud by Means of DNA Methylation

12:20  Lunch Break - Group Picture

13:20  Pick the speakers brain over coffee

SESSION 4 // IMAGE-BASED DATA ANALYSES AND DATA VISUALISATION
Chair: Astrid Junker

13:40  Keynote Lecture - Malia Gehan
Donald Danforth Plant Science Center, Creve Coeur, Missouri, USA
PlantCV 3.0: Open-Source High-throughput Image Analysis Across Platforms

14:20  Keynote Lecture - Softorios Tsafarlis
The University of Edinburgh and the Alan Turing Institute, UK
Machine learning in plant phenotyping: doing more with less

15:00  Evgeny Gladilin
IPK, Gatersleben, Germany
Automated Co-registration of Multimodal Plant Images in Context of High-throughput Data Analysis

15:20  Ricardo Humberto Ramirez Gonzalez
John Innes Centre, Norwich, UK
Expression-bias Visualisation in Polyploid Wheat.

15:40  Coffee Break & Posters

Wednesday, March 20th, 2019

08:30  Transport to Gatersleben

SESSION 5 // SYSTEMS BIOLOGY AND MODELING
Chair: Andrea Bräutigam

09:00  Keynote Lecture - Martin Mascher
IPK, Gatersleben, Germany
TRITEX: Chromosome-Scale Sequence Assembly of Triticeae Genomes with Open-Source Tools

09:40  Keynote Lecture - Zoran Nikoloski
Bioinformatics Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany
Integration of Metabolomics Data in Large-Scale Metabolic Models

10:20  Coffee Break & Posters

11:00  Amanda S. Câmara
IPK, Gatersleben, Germany
Polymer Simulations To Understand The Structure And Dynamics Of Mitotic Barley Chromosomes

11:20  Jedrzej Jakub Szymanski
IPK, Gatersleben, Germany
Transferring Quality Traits Between Domesticated Plants and their Wild Ancestors

11:40  Samuel Seaver
Argonne National Laboratory, Lemont, USA
Plant Informatics in DOE Systems Biology KnowledgeBase

12:00  Closing

12:20  Lunch Break

13:15  Pick the speakers brain over coffee
Talks

Technology Driven Crop Improvement for Africa and South Asia
Robbins, Kelly R..................................10

Developing infrastructures for FAIR data in plant biology
Adam-Blondon, Anne-Francoise ............11

A Phased Diploid Genome Assembly of Italian Ryegrass
Copetti, Dario .....................................12

Targeting Haplotype Phasing: Assembling a Grapevine Rootstock Genome
Frommer, Bianca.................................13

Diversity Reduction and Selection Signature in Tetraploid Wheat Germplasm from Wild Emmer to Modern Durum Wheat
Ormanbekova, Danara .........................14

Tools for Drought Tolerance Selection in Potato
Koehl, Karin Iris..................................15

Single Primer Enrichment Technology as a Tool for Massive Genotyping: Bridging the Gap Between Arrays and Genotyping-by-sequencing
Fornasiero, Alice .................................16

Genomic and Epigenomic Patterns in Novel Heterotic Pools of Winter Rapeseed
Lee, HueyTyng....................................17

How to combine different molecular Marker Systems in Bread Wheat
Beier, Sebastian...................................18

Comparative Transcriptional Network Analysis between Model Crop and Model Plant Species
Bräutigam, Andrea ...............................19

GFBio - A FAIR Infrastructure Network to Assist Scientists in Data Management
Janine Felden ....................................20

Applying Network-based Phylogenomics to Trace the Reticulate Ancestry of Modern Bread Wheats
Mayer, Klaus........................................21

Phenotyping and Genes for Photosynthesis
Harbinson, Jeremy ..............................22

Kmasker Plants – The In Silico Assistant To Avoid Drifting On The Slippery Path Of Repetitive Plant Sequences
Schmutzer, Thomas .............................23

Update on KnetMiner FAIR developments
Singh, Ajit ..........................................24

Distributed Learning by Social Machines in Crop Science, a Peek in from Health.
Mons, Barend .....................................25

Integrative Bioinformatics with Galaxy
Building Frameworks to Serve the 21st Century Data Science Problems
Björn Grüning ....................................26

Software Development for Applied Crop Bioinformatics
Marshall, David ...................................27

Combining RNA-seq Data and Homology-based Gene Prediction
Keilwagen, Jens ....................................28

From FAIRer Data Do Faster Discovery - A Comprehensive Infrastructure to Serve Plant Phenomic Research Data
Arend, Daniel ......................................29

Detecting Food Fraud by Means of DNA Methylation
Grehl, Claudius ...................................30

PlantCV 3.0: Open-Source High-throughput Image Analysis Across Platforms
Gehan, Malia .....................................31
Machine Learning in Plant Phenotyping: Doing More with Less
Tsaftaris, Sotirios ...........................................32

Automated Co-registration of Multi-modal Plant Images in Context of High-throughput Data Analysis
Gladilin, Evgeny ...........................................33

Expression-bias Visualisation in Polyploid Wheat.
Ricardo Humberto ...........................................34

Exploring Leaf Rust Resistance In Aegilops Tauschii Using Association Genetics Coupled With Resistance Gene Sequence Capture
Ghosh, Sreya ...........................................35

Transposons as Agents of Diversity
Gundlach, Heidrun ...........................................36

TRITEX: Chromosome-Scale Sequence Assembly of Triticeae Genomes with Open-Source Tools
Mascher, Martin ...........................................37

Integration of Metabolomics Data in Large-Scale Metabolic Models
Nikoloski, Zoran ...........................................38

Polymer Simulations To Understand The Structure And Dynamics Of Mitotic Barley Chromosomes
S. Câmara, Amanda ...........................................39

Transferring Quality Traits Between Domesticated Plants and their Wild Ancestors
Szymanski, Jedrzej Jakub ...................................40

Plant Informatics in DOE Systems Biology KnowledgeBase
Seaver, Samuel ...........................................41

Posters

Annotation of Resistance Genes in a Phased Grapevine Rootstock Genome Assembly
Frommer, Bianca*; Holtgraewe, Daniela; Hausmann, Ludger; Huettel, Bruno; Reinhardt, Richard; Toepfer, Reinhard; Weisshaar, Bernd ...........................................44

Whole Genome Sequencing and Annotation of Octoploid Strawberry (Fragaria × ananassa) Inbred Lines
Manivannan, Abinaya*; Lee, Sun Yi; Kim, Dosun; Lee, Eun-Su; Lee, Hye-Eun; Kang, Byoung-Cheorl; Kim, Jinhee .................45

Pan-transcriptome Analysis of Elite Durum Wheat Varieties
Ormanbekova, Danara*; Maccaferri, Marco; Twardziok, Sven O.; Scaglione, Davide; Vendramin, Vera; Scalabrin, Simone; Massi, Andrea; Mayer, Klaus F.X.; Morgante, Michele; Tuberosa, Roberto .................46

Big Data for Crop Breeding: Mining the Barley Variome with Exome Capture Data on a Terabyte Scale
Bayer, Micha*; Dawson, Ian; Hedley, Pete; Mascher, Martin; Morrell, Peter; Muehlbauer, Gary; Pillen, Klaus; Ramsay, Luke; Russell, Joanne; Steffenson, Brian; Stein, Nils; Thomas, Bill; Waugh, Robbie .47

Towards Identification Of Genome Regions Causal For Seed Protein Content In Brassica Napus
Schilbert, Hanna Marie*; Pucker, Boas; Holtgräwe, Daniela; Viehöver, Prisca; Beckmann, Katrin; Wolter, Frank; Micic, Zeljko; Weisshaar, Bernd .................48

Detecting large chromosomal Modifications using Short Read Data from Genotyping by Sequencing
Keilwagen, Jens*; Lehnert, Heike; Berner, Thomas; Beier, Sebastian; Scholz, Uwe; Himmelbach, Axel; Stein, Nils; Kilian, Benjamin; Hackauf, Bernd; Perovic, Dragan .................48
Predicting Maize Leaf Parameter from Leaf Reflectance Spectra  
Schlüter, Urte*; Denton, Alisandra K.; Weber, Andreas P. M. ..........................50

Comparative Analysis of Differential Gene Expression in Roots of Contrasting Wheat Genotypes Under PEG Induced Drought Stress  
Mathur, Priyanka*; Munjal, Renu; Swami, Pooja ................................................51

Identification and Comparative Analysis of Differential Gene Expression Revealed by RNA-Seq in Wheat Seedling under High Temperature Stress  
Munjal, Renu* ........................................52

Investigation of Machine Learning approaches for genomic-enabled prediction of phenotypic traits  
Jubin, Cathy * ......................................53

Stories from One Thousand and Two Secale Samples: Insights into the Evolutionary History of Domesticated Rye and Its Wild Relatives  
Schreiber, Mona*; Himmelbach, Axel; Börner, Andreas; Mascher, Martin........54

Tracking tree Frogs at high Canopy to Identify and Record their Tentative role in Pollination Biology  
Kumar Basu, Saikat; Zandi, Peiman*; Li, Qian; Możdżeń, Katarzyna .................55

Data Management For The Bean Adapt Project - Challenges And Solutions  
Rey Mazón, Elena*; Scholz, Uwe; Neuman, Kerstin; Graner, Andreas ..................56

DiSeMiNation: Digging into Sediments and Microbes for Nature Conservation  
Ulpinnis, Chris*; Fedder, Bevis; Grunewald, Karsten; Helfer, Véronique; Jeske, Olga; Neumann, Steffen; Overmann, Jörg; Reuter, Hauke; Syrbe, Ralf-Uwe; Zimmer, Martin .57

Pan-genome analysis of wheat D genome progenitor Aegilops tauschii  
Artmeier, Susanne*; Consortium, Open Wild Wheat; Mayer, Klaus ..........................58

Need for Developing a Suitable Insect Pollinator app for Education, Awareness and Building network for Facilitating Insect Pollinator Identification and Conservation  
Kumar Basu, Saikat; Zandi, Peiman* ......59

Genetic Variation Of SAD Genes, Involved In Fatty Acid Biosynthesis In Flaxseed  
Kezimana, Parfait*; Dmitriev, Alexey A.; Melnikova, Nataliya V.; Romanova, Elena V. 60

Reading Plant ITS2-Sequences with Nanopores and Polymerases: First Insights to Pros and Cons  
Leidenfrost, Robert*; Bänsch, Svenja; Prudnikow, Lisa; Westphal, Catrin; Wünschers, Röbbe .........................61

Discovery of Multi-megabase Polymorphic Inversions by Chromosome Conformation Capture Sequencing in Large Genome Plant Species  
Jayakodi, Murukarthick*; Padmarasu, Sudharsan; Himmelbach, Axel; Walde, Ines; Stein, Nils; Masher, Martin .................................62

Systems, Genome-Wide Association Determines Loci Underlying Lipid Metabolism In Arabidopsis Thaliana’s Seeds  
Ruß, Anne-Kathrin*; Luzarowska, Urszula; Nikoloski, Zoran; Brotman, Yariv ...............63

Revealing Organelle Genomes Diversity of Barley Iloplasmic Lines  
Makarevich, Anna*; Siniauskaya, Marina; Pankratov, Vasily; Liaudanski, Alek; Shymkevich, Andrei; Halayenka, Inessa; Danilenko, Nina; Davydenko, Oleg .........64

Influence Of Neighboring Small Sequence Variants On Functional Impact Prediction  
Baasner, Jan-Simon; Howard, Dakota; Pucker, Boas*........................................65

Bridge Web Portal - A Visual Analytics Software For Plant Genetic Resources  
König, Patrick*; Basterrechea, Martin; Lange, Matthias; Scholz, Uwe ....................66
A Novel Solution to Describe and Manage Maize Genomic Variation for High Resolution Genotyping

Reikav, Sharon* ..................................67

Structure-Preserving Image Enhancement for Contrast-Independent Segmentation of Plant Roots in a Noisy Soil Environment

Narisetti, Narendra*; Gladilin, Evgeny .....68

Prediction of drought-induced Phenotypes from Gene Expression Profiles in Arabidopsis thaliana

Erbe, Stephan*; Fitzek, Elisabeth; Altmann, Thomas; Bräutigam, Andrea; Junker, Astrid. ........................................................69

Omics-Based Prediction of Hybrid Performance, Dynamic QTL and Systems Genetic Analyses in Canola

Knoch, Dominic*; Abbadi, Amine; Grandke, Fabian; Himmelbach, Axel; Meyer, Rhonda Christiane; Riewe, David; Samans, Birgit; Werner, Christian Raimund; Snowdon, Rod; Altmann, Thomas............................................70

Unraveling Arabidopsis’ Response to Fluctuating Light Conditions by Genome-Wide Association of Time Resolved High-Throughput Phenotyping Traits

Heuermann, Marc Christian*; Meyer, Rhonda Christiane; Tschiersch, Henning; Altmann, Thomas............................................71

Integration Of GWAS And eQTL In A Wheat Population With 185 Genotypes Reveals Global Genomic And Expression Differences

Wagner, Christian; Omony, Jimmy*; Nosenko, Tetyana; Nussbaumer, Thomas; Seidel, Michael; Falter-Braun, Pascal; Steiner, Barbara; Mayer, Klaus F.X.; Bürstmayr, Hermann .........................72

Network Analysis in the Alga Chlamydomonas Reinhardtii

Wulf, Donat*; Lämmermann, Nina; Wobbe, Lutz; Kruse, Olaf; Bräutigam, Andrea ....73

Cell Wall Enzymes and ABI1 in Zygnema Circumcarinatum UTEX 1559 respond to osmotic Stress in a plant-like Fashion

Fitzek, Elisabeth*; Orton, Lauren; Entwistel, Sarah; Grayburn, Scott; Duvall, Melvin; Yin, Yanbin ........................................74

ATAC-seq derived Target Motif Prediction of Arabidopsis Thaliana Transcription Factors

Sielemann, Janik*; Bräutigam, Andrea ....75

Functional Annotation of Transcription Factors using Gene Regulatory Networks (GRN)

Halpape, Wiebke*; Härting, Nina; Wulf, Donat; Bräutigam, Andrea..........................76

A Constraint-based Modelling Approach to Study C4 Metabolism

Blätke, Mary-Ann*; Bräutigam, Andrea ....77

How to decipher the Drive of enigmatic B Chromosomes?

Boudichevskaja, Anastassia*; Bräutigam, Andrea; Houben, Andreas...............78

De novo transcriptome assembly and expression profiling in Allium sativum to gain insights into cryo-stress conditions

Ulagappan, Kamatchi *; Senula, Angelika; Nemati, Zahra; Blattner, Frank R.; Nagel, Manuela ................................79
Much of the world’s population growth will occur in regions where food insecurity is prevalent, with large increases in food demand projected in regions of Africa and South Asia. While improving food security in Africa and South Asia will require a multi-faceted approach, improved performance of crop varieties in these regions will play a critical role. Current rates of genetic gain in breeding programs serving Africa and South Asia fall well below rates achieved in other regions of the world. Given resource constraints, increased genetic gain in these regions cannot be achieved by simply expanding the size of breeding programs. New approaches to breeding are required. Several initiatives are currently underway to build capacity, develop breeding informatics capabilities and build computational infrastructure to enable routine use of new technologies that can improve the efficiency of breeding programs and increase genetic gains. A brief overview of several on-going projects will be provided with discussion focused on breeding strategies, cost effective genotyping, proximal sensing, data management and analytics. Collaborations with the Genomic Open-source Breeding Informatics Initiative (GOBii) for the implementation of genomic selection in CIMMYT Maize and ICRISAT Chickpea breeding programs will be presented, and the path forward for routine implementation of genomic selection discussed.
INRA is involved in several EU infrastructures (e.g. ELIXIR, EMPHASIS) or global initiatives (e.g. Wheat Initiative, Research Data Alliance) contributing to the development of: (i) community recommendations for data standardization (e.g. Dzale-Yeumo et al 2017, https://doi.org/10.12688/f1000research.12234.2), (ii) data standards for phenotyping data (www.miappe.org), (iii) crop specific ontologies in the frame of the CropOntology (http://www.cropontology.org/) and (iv) standard web services (www.brapi.org). These global resources are used to capture the data produced in large scientific projects with a standard and structured vocabulary and to store them into INRA’s central repository for plant genomic, phenomic and genetic data, GnpIS (https://urgi.versailles.inra.fr/gnpis/) under the FAIR principles (https://www.force11.org/group/fair-group/fairprinciples).

In parallel, we have developed tools to support federation of databases. The first one is a light data discovery tool based on text indexing technologies, which allows to find data across several distinct databases based on a common high level generic data model. The current federates 15 databases from 6 countries (https://urgi.versailles.inra.fr/wheatis/). Another tool allowing data search on phenotyping data and based on BrAPI web services is also about to be released in the frame of the H2020 ELIXIR-Excelerate project, n°676559. The long term resilience and success of such federations will highly depend on the capacity to animate and structure the community working on data for plant biology.
Grasses of the genera *Lolium* and *Festuca* are the main feed sources for a sustainable livestock production due to their high palatability and biomass production. Since decades, their importance for the agriculture of temperate regions led to the development of new varieties through traditional breeding programs. However, newer crop improvement methods such as genomic selection could benefit from a high-quality reference genome assembly. In the past, attempts at producing genomic resources have struggled due to the complexity of the genome and the outcrossing nature of the species. We sequenced an individual of the *L. multiflorum* (Italian ryegrass) cv. Rabiosa, producing a highly-contiguous and complete assembly. Due to the high heterozygosity of the genotype, the resulting assembly was as large as the diploid genome, thus presenting the sequence of both alleles in separate collinear scaffolds. The generation of large-scale scaffolding datasets (i.e. chromosome conformation capture data and optical maps) allowed to phase sequences, reaching chromosome-level contiguity. The comparison of the two allelic sequences for a region showed an extensive amount of intergenic sequence variation, confirming that ryegrass genomes are highly dynamic. The high-quality genome assembly of the cv. Rabiosa is the first phased diploid assembly of a plant genome and provides a high-quality reference for expediting ryegrass breeding and studying the genome biology of outcrossing species.
Due to the fact that third generation sequencing technologies provide increased read length and cost affordable high sequence coverage, the problem of generating genome assemblies comprising separated haplotype phases becomes solvable. Haplotype phase information can give insights into the relation between genotype and phenotype, disease susceptibilities, inherited diseases and many more. Here, we used long reads generated with single molecule real time (SMRT) DNA sequencing along with a binning and assembly approach to compute a potentially fully phased, high quality de novo genome assembly of the diploid and highly heterozygous interspecific grapevine rootstock cultivar ‘Börner’. Prior assembly, the long reads were divided into parental subsets using the software TrioCanu (Koren et al., 2018) and short Illumina read data of the parental lines of ‘Börner’, Vitis riparia GM183 and Vitis cinerea Arnold. With these subsets one genome assembly for each haplotype was constructed with the assembler Canu (Koren et al., 2017). An N50 of 4.45 and 4.89 Mb could be achieved for the assemblies. Furthermore, of 1,440 known plant core genes 1,383 complete genes could be identified in both assemblies. Very high phasing accuracy of the resulting haplotype assemblies was confirmed through mapping of various phased ‘Börner’ BAC sequences from the chromosome 1 and 14. The assemblies were further scaffolded with paired BAC end sequences also from ‘Börner’. Finally, sequences were assigned to chromosomes using the grapevine genome reference sequence PN40024 (Adam-Blondon et al., 2011; Jaillon et al., 2007; Velasco et al., 2007) and reciprocal best hits (RBHs) at the protein and nucleotide level. The resulting pseudochromosomes were used for gene annotation. Since ‘Börner’ is resistant to pathogens like phylloxera, downy and powdery mildew and black rot, the generated ‘Börner’ haplotype assemblies with their annotation will open the way to further investigate the underlying resistance genes of those diseases.
Diversity Reduction and Selection Signature in Tetraploid Wheat Germplasm from Wild Emmer to Modern Durum Wheat

Ormanbekova, Danara* (1); Maccaferri, Marco (1); Pasam, Raj K. (2); Mastrangelo, Anna Maria (3); Mazzucotelli, Elisabetta (4); Kilian, Benjamin (5); Tuberosa, Roberto (1); Cattivelli, Luigi (4); Durum Wheat Genome Sequencing Consortium, International (4)

1: Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy; 2: Department of Economic Development, Jobs, Transport and Resources, Agribio Centre, La Trobe Research And Development Park, Bundoora, Australia; 3: CREA - Research Centre for Cereal and Industrial Crops, Foggia, Italy; 4: CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d’Arda (PC), Italy; 5: Global Crop Diversity Trust, Bonn, Germany

The domestication of the wild emmer wheat nearly 10,000 years ago in the Fertile Crescent lead to the evolution of the free-threshing and easy to harvest tetraploid *Triticum durum* wheat. The recently assembled reference genome of durum wheat cultivar Svevo and a Global Tetraploid wheat Collection (GTC), consisted of 1,856 accessions, were used to produce the genotyping data based on the Illumina iSelect wheat 90K SNP array. The GTC was composed of up to ten different species and subspecies from various geographical regions and represented the four principal germplasm groups involved in the tetraploid domestication and selection history. In this study we aimed to identify the population structure of the GTC and detect the signatures of genetic divergence in tetraploid wheat germplasm from wild emmer wheat to domesticated modern wheat. We detected the genetic diversity and selection signature regions associated with wild emmer domestication and durum wheat evolution on the pericentromeric regions of chromosome groups 2, 4, 6 and on chromosomes 1A, 7B. Several diversity reduction and selection signature peaks overlapped with loci associated with domestication, disease resistance, yellow pigment content and seed dormancy. Population structure of the tetraploid diversity panel assessed by two model and two non-model based clustering methods subdivided wild emmer (WEW) in two main populations, domesticated emmer (DEW) and durum landraces (DWL) in six populations each, and durum cultivars (DWC) in five subpopulations. The modern durum wheat germplasm showed the highest relationship to the North African and Turkey to Transcaucasia DWL populations, while the Ethiopian and *T. turanicum* populations were the most differentiated with a minimal contribution to the modern durum germplasm. Our results suggest that modern durum wheat has been affected by genetic bottleneck/selection events leading to strong diversity depletions. The selection signature detection provided an insight into the dynamics of tetraploid wheat domestication.
Maintenance of yield stability requires efficient selection for drought tolerance, as climate models predict an increased likelihood of seasonal droughts. Potato is an important crop with high water-use efficiency but low drought tolerance. Breeding for drought tolerance by selection on yield in arid environments takes decades. This asks for marker-assisted selection (MAS) to accelerate breeding. However, drought tolerance is a polygenic trait with a strong gene × environment effect. This reduces MAS efficiency to identify the most promising lines in a breeding population. Efficiency may be increased by a stepwise combination of different metabolomics, transcriptomics and phenomics selection systems.

The hands-on development of such a system involved the integration of heterogeneous data, including field samples, sensor data and manual phenotyping information. We did this in a project aimed to find tools for drought tolerance selection in *Solanum tuberosum* ssp. *tuberosum*. Based on data from field and controlled environments on 34 cultivars, we developed a model that predicts drought tolerance from leaf metabolite and transcript levels independent of the agro-environment. Subsequently, we tested the model in a selection experiment.

From a population of 200 potato lines segregating for drought tolerance, a classic subpopulation was selected for superior tolerance based on tuber starch yield data from three trials. Metabolite and transcript levels in leaf samples from two trials were used to predict drought tolerance and select a second tolerant (MAS) subpopulation. Validation based on yield data from ten multi-environment drought trials indicated that both procedures identified lines with drought tolerance above the mid-parent mean. The MAS procedure was slightly less efficient but faster than the yield-based selection.

To further increase efficiency, we tested the predictive value of phenotypic markers gained from laser scanner imaging and thermometry and identified canopy temperature depression in defined screening windows as a promising drought tolerance predictor.
Several disciplines such as plant and animal breeding, conservation and ecology rely on the ability to massively sample the genetic diversity. This sampling exercise has a non-marginal importance in the feasibility and outcome of such studies. Costs, sample size, number of markers, as well as their type and distribution in the genome are driving factors of successful experiments. Genotyping-by-sequencing has become a widely adopted method for cost-effective genotyping with basically no initial investment for design as compared to array-based platforms which have shown to offer very robust assays, but with the drawback of being bound to the genetic diversity accounted during design. On the contrary, genotyping-by-sequencing with random sampling of genomic loci via restriction enzymes or random priming has shown to be fast and convenient but lacks the ability to target specific regions of the genome and maintain high reproducibility across laboratories. Here we present a first adoption of single-primer enrichment technology (SPET) which provides a highly efficient and scalable system to obtain targeted sequence-based large genotyping datasets, bridging the gaps between array-based system and traditional sequencing-based protocols. To fully explore SPET performance we conducted a benchmark study in ten Zea mays lines and an extensive application to study a natural black poplar population of 540 individuals with the aim of discovering polymorphisms associated to biomass-related traits. Our results showed the ability of this technology in providing dense genotype information on a customized panel of selected polymorphisms, while yielding hundreds of thousands of untargeted variable sites as an ideal resource for association analysis of natural populations harboring unexplored allelic diversities and structure such as in black poplar.
Exploitation of hybrid vigour in crops is simplified by distinct heterotic pools and breeding methods that facilitate effective prediction and use of heterosis. In crops which have traditionally been bred as open-polliated inbred line varieties, like oilseed rape/canola (*Brassica napus*), heterotic pools generally do not exist and systematic exploitation of heterosis is challenging. Using winter oilseed rape as a case study, we are investigating how genome-wide patterns of genomic and epigenomic variation may help distinguish and develop new heterotic pools. We sequenced two pools of 50 elite, winter type oilseed rape lines, and catalogued the genomic and epigenomic variants of each genotype. As expected from the breeding history, single nucleotide polymorphisms and methylation patterns were found to largely overlap between pools. However, variants unique to pools were detected at the genomic level, indicating strong potential for genomics-assisted separation of heterotic pools. By tracing these divergent variants throughout a breeding program in intercrossed pool offspring, we are generating a catalogue of genomic and epigenomic patterns which will serve as a basis for hybrid performance prediction once recombinants are successfully fixed in individual pools through genomics-assisted crossing designs. Our approach introduces a novel exploitation of heterotic patterns to enhance breeding process which bypasses the need for direct association of each variant to the trait performance.
Breeders and researchers alike apply molecular markers to genotype plant material. Most of these markers can be characterized as single nucleotide polymorphism (SNP) and can be used to analyse regions of interest. Different technologies like genotyping-by-sequencing (GBS) and exome capture sequencing can be utilised to generate novel SNP markers throughout the genome. Other genotyping platforms from array based systems are available. These arrays carry a large set of validated SNPs constructed by analysing diverse populations. Different technologies like competitive allele specific PCR (KASP) or simple sequence repeats (SSR) can be utilised to accurately genotype samples. Unfortunately, combining different molecular marker systems is not a trivial task. As point of reference a coordination system needs to be decided on, which usually is the most accurate version of genome reference sequence available. Through homology based searches of target regions many sequences can be placed. However, based on the genetic architecture of the organism of choice there might be a number of cases where it is not so clear.

In bread wheat this endeavour can be more complicated because of several different reference sequence that were published over the last few years, the allohexaploid nature of the genome, and the massive size of roughly 17 Gbp. Here, we demonstrate the strategy employed to integrate different marker systems in wheat for a winter wheat breeding panel that was genotyped using the 15k SNP array and exome capture data.
Transcriptional regulatory networks are at the core of plant responses to developmental, biotic, and abiotic cues. Yet, the majority of transcription factors remain without an associated function in model plants, crop models and crops. We have deployed supervised machine learning to predict target genes of all transcription factors of a species, annotate their functions based on these targets, test the functional annotation against database standards, and test the association of targets to transcription factors using RNA-seq. In general, data availability and data spread across conditions and tissues for a particular species have a larger influence on prediction success compared to genome complexity. In essence, predictive success hinges on sufficient privately and publicly available data. Some transcription factor families are more amenable to predictive success than others.

In wheat, the available data annotates half of all transcription factors functionally, and a between subgenome comparison reveals extensive conservation of function with limited exceptions. *Arabidopsis* transcription factor annotation is achieved for half of all transcription factors of which 72% are validated by database comparison. A case study matched 66% network edges emanating from a candidate to RNA-seq data of the respective knock-out. The comparison of networks from different species such as wheat, barley, maize, *Arabidopsis thaliana*, and tomato reveals extensive evolutionary conservation in the regulatory architecture.

Network predictions were used to inform binding motif studies in upstream regions and in ATAC-seq defined open chromatin upstream regions using motif prediction programs. Motifs extracted from purely computational approaches can be validated using DAP-seq and Jasper database motifs.

For details on methods and results of single species analyses, please see the authors at their posters.
The German Federation for Biological Data (GFBio) is implementing a national infrastructure for the preservation, integration, and publication of biological and environmental data generated in research projects. GFBio commits itself to the FAIR-data principles including technical, organizational, cultural, and policy aspects. It is based on the collective experience and expertise of leading researchers as well as on a network of German natural science collection data repositories, the data publisher for earth and environmental science - PANGAEA, as well as selected facilities from the molecular biology research community. GFBio addresses data management requirements of a wide array of stakeholders ranging from individual scientists to large research networks. GFBio must thus be capable of handling highly heterogeneous data. Services provided by GFBio will cover the full research data life cycle from field or real-time data acquisition to long term archiving and publication as well as analysis and re-use of these data. GFBio especially aims to encourage and support individual scientists to embed good data management practices in their daily scientific work. The goal is to promote open science by improving the availability, quality and thus re-usability of research data according to the FAIR data principles.

The GFBio project is currently in its third phase focusing on sustainability as well as on consolidation and hardening of services. In order to successfully manage the transition from a project to a national infrastructure network, the non-profit association GFBio e.V. was founded in 2016. It will provide a single entry point for all services and will assure cooperation of GFBio partners and availability of services beyond the project phase. GFBio is ready to impart its experiences in research data management with initiatives like the German national research data infrastructure (NFDI).
We used exome sequencing of almost 500 genotypes of the wheat species complex selected from all across the geographic range of the species to reveal genome-wide patterns of genetic variation. Implementing a network-based phylogenetic approach utilizing repeated random haplotype samples with maximum likelihood, subsequent graph reconstruction analysis and community clustering we traced the reticulate evolutionary history of modern hexaploid bread wheats from their di- and tetraploid progenitors. The resulting clustered consensus network comprises signals of both vertical, species relationships and horizontal, reticulation events. This allowed us to infer the ancestry of the investigated wheat genotypes within the *Triticum-Aegilops* species complex despite the presumable occurrence of multiple hybridization and introgression events and at the same time to discern three distinct aestivum communities that can be reconciled with distinct periods in human history.
High throughput phenotyping of photosynthesis, while challenging, is essential if we are to understand the natural diversity and the genetic basis of that diversity. Understanding these genetic underpinnings, while intrinsically interesting, will be important if we are to be able to breed for improved photosynthetic traits or manipulate these traits by means of transgenic or gene editing approaches. We have developed and been using for several years a high-throughput chlorophyll fluorescence based phenotyping system. With this we can measure, for example, the operating photosynthetic light-use efficiency of c. 1500 plants in 1 hour; in this case photosynthetic light-use efficiency is for photosystem II electron transport. With this system we can easily reveal the extent of natural variation for photosynthesis. In addition the data obtained is useful for further genetic analysis to allow us to localise and then identify the nuclear genes responsible for generating this variation. While the nucleotype generally dominates variation in photosynthesis the plasmotype has not insignificant effects. To allow us to get some idea of the role that plasmotypic variation has on photosynthesis we have developed a population of Arabidopsis in which the nuclear and cytoplasmic genomes have been swapped, allowing the effects of plasmotype - nucleotype interactions to be better assessed.
Many plant genomes display high levels of repetitive sequences. Underestimation or disregard of repeat complexity in these datasets can easily misguide downstream analysis. Detection of repeats by methods of \( k \)-mer counting were proven to be reliable. We re-implemented \textit{Kmasker} from scratch, leading to a decrease of processing time (10-fold), saving of storage space for constructed indices (\( \sim \)14-fold) and an improved usability. Newly integrated features of the command-line version are comparative studies between different cultivars or closely related species and methods estimating target specificity of guide RNAs for application in CRISPR/Cas9. In addition, the web-service of \textit{Kmasker plants} now maintains pre-computed indices for ten important cultivated plants (https://kmasker.ipk-gatersleben.de/).

To highlight its applicability, we present results from \textit{Aegilops speltoides}, studying the drift and origin of its extraordinary B chromosome. From standard low-coverage whole-genome shotgun sequencing (WGS) of two plants, one with and one without B chromosomes, \textit{Kmasker plants} directly revealed single-copy-sequences for fluorescence \textit{in situ} hybridization. From the WGS assembly more than 153.000 (\( \sim \)9.8\%) contigs with different \( k \)-mer patterns were identified and assigned to the B chromosome, including 481 gene derived sequences. Finally, this \textit{in silico} method was able to examine the composition of the \textit{Aegilops speltoides} B chromosome. With this successful use case, we present an easy-to-use procedure for comparative studies in plants using cheap and readily available WGS genomic datasets.
KnetMiner is a popular resource for gene discovery and biological hypothesis generation. Under the hood, KnetMiner is built on a large integrated knowledge graph and smart algorithms to index, search, rank and visualise the knowledge. We will present the latest developments of making KnetMiner tools and networks FAIR for the crop bioinformatics community. This includes our work on building a reusable BioJS component to visualise biological knowledge networks, mapping our knowledge concepts to BioSchemas and provide public SPARQL and Cypher endpoints to access KnetMiner networks programmatically. KnetMiner is available at: www.knetminer.org
With datasets frequently too large and/or too sensitive nowadays to be effectively ‘sent around’ to the analytics, more and more, the analytics move to the data. This approach opens up lots of opportunities and is very much related to data ownership as well. Examples of how this works in a highly sensitive field like human health will be given and some thoughts will be shred on how this all reflects on agriculture and the broader life sciences. The need for FAIR digital objects in the so called Internet of FAIR data and Services will be discussed and the latest developments in this field will be covered.
Advances in science increasingly rely on the analysis of large datasets. This produces new challenges to the scientific community as access to data and compute resources is becoming a crucial factor. Fortunately, more and more data is shared openly and is reusable for everyone. However, reusable data does not mean reproducible research and most of the published results are still hard to reproduce. Moreover, integration of different data sources to gain comprehensive insights into biological systems and bridging (scientific) communities opens up new challenges in communication.

Started in 2005, the Galaxy Project (https://galaxyproject.org) maintains a focus on enabling data-driven science by pursuing three goals: (a) accessible data analysis serving all scientists regardless of their informatics expertise and tool developers seeking a wider audience and broad integration of their tools; (b) reproducible analyses regardless of the particular platform and (c) transparent communication of analyses, which in turn enables reuse and extension of analyses across communities of practice.

Bioconda and BioContainers are empowering Galaxy to re-create and transfer execution environments, fulfilling the promises of reproducible research on the technical level. Both projects became practically the standard in scientific research in the last years, and have been adopted way beyond Galaxy.

This talk will introduce Galaxy, Bioconda, BioContainers and how they play together to create the foundation for UseGalaxy.eu, an integrative and scalable data analysis platform. We will further discuss how different user-groups with different skills are working together in one platform to make research more reproducible and efficient.

To glue all pieces together the Galaxy Training Network with its new training platform (https://training.galaxyproject.org) will be introduced.
The Bioinformatics Group at the James Hutton Institute develops software and information systems that are freely available and widely utilised by both the academic and breeding communities. We have been employing these tools in support of a number of international projects including the Crop Wild Relatives (CWR) project managed by the Crop Trust and Kew and supported by the Government of Norway where our informatics tools will be used to help provide access to pre-breeding data from 19 important crops and their wild relatives and the Seeds of Discovery (MasAgro Bioversidad) project funded by the Mexican Government and the BBSRC which is categorizing CIMMYT’s internationally important maize and wheat collections.

Collecting and managing data is difficult. Analysing data and presenting the results in a useful format can pose an even greater challenge. This, however, must be addressed if accessing information from genetic resource collections and pre-breeding projects is going to efficiently contribute to the development of improved crops better adapted to the demands of changing environments and the pressures for increased yields. The primary focus of the tool set that has been developed at the James Hutton Institute is improving the efficiency of information access.

Tools such as Flapjack allow visual exploration of genotype and trait data and analyses; CurlyWhirly plots 3D cartesian coordinate data such as that obtained from PCo and PCA analysis combined with selected querying by categorical information; Helium which can be used to visualize complex plant pedigrees and associated categorization or SNP data and Germinate which is a data warehouse offering standard query and search tools for data types typically associated with genetic resources collections and pre-breeding data.
Genome annotation is of key importance in many research questions. The identification of protein-coding genes is often based on transcriptome sequencing data, ab-initio or homology-based prediction.

Here, it was demonstrated that intron position conservation and RNA-seq data improves the homology-based gene prediction program GeMoMa. In addition, we demonstrate that using multiple reference organisms may help to further improve the performance of GeMoMa.

Finally, we present an extension of GeMoMa that replace (a) blastn by mmseqs and (b) allows to predict UTRs.

GeMoMa might be of great utility for annotating newly sequenced genomes but also for finding homologs of a specific gene or gene family. GeMoMa has been published under GNU GPL3 and is freely available at http://www.jstacs.de/index.php/GeMoMa.
In the context of a growing global demand for food and feed, the need for improved crop yield and the identification of more efficient as well as better-adapted crop plants is an important driving force for high-throughput phenotyping studies, which comprise comprehensive and data-intense experiments. One key challenge is the definition of appropriate data management procedures and infrastructures to preserve research data as valuable scientific asset. In line with that, funding agencies and scientific journals increasingly request to publish research data under the consideration of the FAIR data principles (findable-accessible-interoperable-reusable).

The FAIR-aware e!DAL infrastructure software was developed in order to lower the technical barriers and minimize the effort of research data publication. In contrast to the common publication process, it follows a “bringing the infrastructure to the data” approach and enables the publication of large volume in-house stored datasets by assigning Digital Object Identifiers (DOIs). Based on e!DAL, the Plant Genomics and Phenomics Research Data Repository (PGP, http://edal-pgp.ipk-gatersleben.de) [4] provides amongst others the first full MIAPPE-compliant phenotypic datasets. The embedded intuitive submission process supports researchers in sustainably describing and sharing their phenotypic data to exploit the full scientific potential of their data.

The all-in-one implementation of e!DAL provides a self-configuration procedure as well as support of the ELIXIR AAI, which allows an easy establishment of further repositories like the instance running at the research centre in Jülich (http://www.fz-juelich.de/ibg/ibg-2/edal). Furthermore e!DAL-PGP is a part of the diverse service portfolio of the de.NBI and ELIXIR Germany infrastructures.
Detecting Food Fraud by Means of DNA Methylation

Grehl, Claudius* (1,2); Kuhlmann, Markus (3); Fatma, Samar (2); Wagner, Marc (4); Altschmied, Lothar (3); Mäder, Paul (5); Himmelbach, Axel (3); Glaser, Bruno (1); Große, Ivo (2,6)

1: Martin-Luther University Halle-Wittenberg, Institute of Agronomy and Nutritional Sciences, Germany; 2: Martin-Luther University Halle-Wittenberg, Institute of Computer Science, Germany; 3: Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany; 4: Free University of Berlin, Germany; 5: Research Institute of Organic Agriculture, Switzerland; 6: German Centre for Integrative Biodiversity Research (iDiv), Germany

Organic farming stands for a sustainable use of natural resources, a higher biodiversity, and an environmentally friendly way to produce food and other products like clothes and cosmetics. Hence, there is a fundamental need for a reliable method to identify organic products to prevent food fraud. We aim to address this challenge by utilizing DNA methylation as an epigenetic biomarker for differentiation between organically grown and conventionally grown crop plants. Several studies have shown that the level of DNA methylation in specific genomic regions can be influenced by environmental conditions.

Motivated by the possibility that environmental conditions characteristically different between organic and conventional farming such as plant protection and fertilization might lead to differentially methylated regions in the genome, we apply whole-genome methylome studies based on whole-genome bisulfite sequencing (BS-seq) and Methylation Sensitive Restriction Digestion (MSRE) qPCR in soybean (*Glycine max.*) seed and potato (*Solanum tuberosum*) tuber samples derived from a long-term field trial with organic and conventional management. Using the same varieties grown at the same field in four different years with three replicates each offers the possibility to investigate if and how organic and conventional farming might differentially influenced DNA methylation patterns of these two crops.

We developed a modular pipeline consisting of multiple BS-seq read mappers such as bismark, bsmap, bsseeker2, gem3, gsnap, and segemehl as well as multiple tools for predicting differentially methylated positions (DMPs) and differentially methylated regions (DMRs) like metilene and defiant for the analysis and visualization of methylome differences in non-model species. Using this modular pipeline and simulated data, based on genomes of six different species, we performed a comprehensive benchmark study for choosing an optimal combination of modules and module parameters for a reliable prediction of differentially methylated regions in soybean and potato.
To tackle the challenge of producing more food and fuel with fewer inputs a variety of strategies to improve and sustain crop yields will need to be explored. These strategies may include: mining natural variation of wild crop relatives to breed crops that require less water; increasing crop temperature tolerance to expand the geographical range in which they grow; and altering the architecture of crops so they can maintain productivity while being grown more densely. These research objectives can be achieved with a variety of methodologies, but they will require both high-throughput DNA sequencing and phenotyping technologies. A major bottleneck in plant science is the ability to efficiently and non-destructively quantify plant traits (phenotypes) through time. PlantCV (http://plantcv.danforthcenter.org/) is an open-source and open development suite of image processing and analysis tools that analyzes images from visible, near-infrared, and fluorescent cameras. Here we present new PlantCV analysis tools available in version 3.0, which includes interactive documentation, color correction, and the development of thermal and hyperspectral imaging tools aimed at the identification of early abiotic stress response.
Advances in automation and imaging equipment have created a renaissance in the measurement of phenotyping traits in plants. I will describe how the use of machine learning can open the road towards relieving the bottleneck of extracting traits from imaging data. I will present solutions, stemming from our work and others on automated leaf counting, plant growth assessment, trait identification and root analysis. I will describe our efforts in training deep learning models with as less data as possible leveraging correlations across data sources and tasks and even using synthetic data. I conclude by describing our efforts on bridging the communities of computer vision, machine learning and plant sciences.
Co-registration of multi-modal images acquired with visible light (VIS), fluorescence (FLU), near-infrared (NIR) and 3D cameras is essential for automated processing and analysis of high-throughput data from multi-sensory plant phenotyping platforms. For establishment of image correspondences different algorithmic approaches based on different image features have been proposed. The particularity of plant image analysis consists in a large variation of forms and colors of plant species measured from different views in different optical scenes at different developmental stages. While adult plant shoots typically have a well recognizable and unique structure, young shoots may have a non-specific shape which can sometimes be hardly distinguished from the background structures. Due to naturally occurring shadows and reflections background regions may also contain similar colors as plant regions. Furthermore, dynamically measured plants may exhibit non-uniform movements that require application of non-rigid image registration. In this contribution, we compare and extend three common techniques for image registration that rely on finding correspondences between (i) feature-points, (ii) frequency domain features, and (iii) image intensity information. Our experimental results show that all three techniques are sensitive to structural image distortions and require additional pre-processing steps including adaptive filtering, structural enhancement and characteristic scale selection. To overcome the limitations of conventional approaches, an iterative algorithmic scheme is developed which allows to accurately perform affine as well as slightly non-rigid registration of high-throughput plant images in a fully automated manner.
The recent improvements in the assembly of the hexaploid wheat genome have the potential to accelerate the pace of genetic research in this important crop. At the same time, the amount and complexity of the data can feel overwhelming given its polyploid and large 16 Gb genome. We present visualisations to explore the relationship between homoeologous genes that intuitively represent the expression bias across the transcriptome. We analysed 850 public RNA-Seq expression samples across different varieties, tissues, ages and under different environmental conditions.

Ternary plots are used to get an overview of the whole transcriptome or for a single set of homoeologous genes (triad) across different tissues. The distance between the expression biases across tissues is used to categorise the expression of each triad as stable or dynamic. This visualisation technique helps to answer the longstanding question of dosage effects in polyploid organisms and characterises early steps towards sub-functionalisation. Including the genomic coordinates and predicted gene ontologies, we discovered that genes located in the most distal regions of the chromosomes and related to environmental response tend to be more dynamic. Likewise, housekeeping genes are located in proximal regions and show more stable expression across genomes. Using this novel metric, we discovered that dynamic genes are under more relaxed selection pressure than stable genes, linking variation in relative expression to sequence variation in coding regions.

To make the expression data more accessible to researchers and breeders in the wheat community, we developed expVIP, an expression browser. The interactive visualisation allows users to group the different experiments in high-level, intermediate, and fine-grained factors. Expression data can be represented for a group of genes, a single gene, or for all gene homoeologs. These features allow interactive hypothesis testing and have been used to narrow down candidate genes inside a target genetic interval.
Leaf rust of wheat, caused by the obligate biotrophic pathogen *Puccinia triticina*, is the most common among the three rusts of wheat worldwide. Wheat leaf rust epidemics can cause significant yield losses, and the most economical method to control them is breeding for genetic resistance to this pathogen. There are over 70 leaf rust resistance genes that have been mapped and designated in wheat, but the sequences of only a handful of genes have been identified so far. Many of these genes, including some of the cloned ones, have been transferred into cultivated wheat from the diploid wild wheat progenitor *Aegilops tauschii*. Access to the sequences of such resistance genes benefit breeders by allowing them to design gene-specific markers to quickly and cheaply track them in their breeding programs. It also helps researchers understand the genetic architecture of leaf rust resistance, and how it can be engineered to stay one step ahead of a rapidly evolving pathogen. Traditional map-based or mutational genomics-based gene cloning approaches can be slow and labour intensive, as they involve creating structured populations. A novel method for resistance gene-cloning, which combines association genetics with resistance-gene-enrichment sequencing (AgRenSeq), can be used instead to rapidly obtain candidate genes for resistance in a diversity panel of wild wheat. Here, I will describe how I am using AgRenSeq to identify candidate genes for leaf rust resistance using an *Aegilops tauschii* panel, and exploring their frequencies in the panel to dissect the genetic architecture of leaf rust resistance against several races from North America and Europe.
In the cereals like barley, rye or wheat - with genome sizes between 5-16 Gb - transposons and their deteriorated remnants add up to at least 85% of the genome. Differential transposon insertions and removals are thought to be one of the major causes for sequence variation between closely related lines. Their influence on phenotypic variation is well documented for single case studies. However, the sheer amounts of transposons and their high degree of repetitivity have for a long time massively impeded the creation of high quality assemblies and subsequent genome wide bioinformatic studies on exact transposon locations. Only recent breakthroughs in assembly technologies have led to highly contiguous chromosomal pseudo-molecules for the large Triticeae genomes (5-16 Gb) that correctly reconstruct most of the transposon space. This enabled for the first time to study the dynamics of transposon families within the Triticeae and their overall potential for phenotypic variation.
TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools

Mascher, Martin*
IPK Gatersleben, Seeland, Germany

Chromosome-scale genome sequence assemblies underpin basic and applied research in the cereal crops wheat and barley. Recent genome assembly efforts in these species have relied on the commercial closed-source assembly algorithm DeNovoMagic. We have developed TRITEX, an open-source computational workflow that combines paired-end, mate-pair, 10X Genomics linked-read and chromosome conformation capture sequencing data to construct sequence scaffolds with megabase-scale contiguity and arrange these into chromosomal pseudomolecules. We evaluated the performance of TRITEX on publicly available sequence data of emmer and bread wheat, and constructed an improved annotated reference genome sequence assembly of barley cultivar Morex as a community resource. TRITEX assemblies of 19 barley genotypes will form the backbone of the barley pan-genome project.
Metabolic functions are shaped by networks of biochemical reactions whose dynamics is in part shaped by the concentrations of the participating metabolites. Therefore, understanding the constraints on metabolite concentrations will help us elucidate the limits of other cellular phenotypes. In the first part of the talk, I will present our recently proposed constraint-based modelling approach to identify metabolites whose concentration ranges can be computed in metabolic networks endowed with mass-action kinetic. I will show that a network property which is at the core of this approach allows the computation of concentration ranges in agreement with simulated and measured concentrations. I will then present extensions to the approach that facilitate integration of metabolomics data, testing of tasks optimized by a cellular system, and understanding the concentration robustness of metabolites in different cellular compartments. In addition, I will briefly introduce few approaches that can be used to integrate time-series metabolomics data to predict growth and time-dependent metabolic fluxes. In the second part of the talk, I will show how metabolomics data from labelling experiments and biomass characterization have helped us to obtain a bottom-up metabolic reconstruction of *Jatropha curcas*. We use this model together with a series of approaches from the constraint-based modelling framework to pinpoint network reactions whose joint manipulation optimizes lipid production in Jatropha cell culture.
From interphase to metaphase the chromosomes go through intense condensation. This process, which is very conserved within eukaryotic organisms, leads to a two- to three-fold packing, enabling an accurate segregation of the chromosomes. Recent analysis from HiC data have brought light into this impressive process and are helping to unravel the more packed structure of the mitotic chromosome. Complementarily, polymer simulations can produce several models to be tested and their fit to experimental data be verified. Together, these two techniques were able to reconcile two apparently conflicting former views: condensation arising either from loop extrusion or from helical packing. Gibcus et al. have suggested that both models may be true. Mitotic chromosome may be formed by nested loops arranged side by side in a dynamical helical scaffold. This model is in good agreement with the contact probability calculated from Hi-C experiments with human cells. Our own Hi-C data from flow-sorted chromosomes suggest a similar structure for mitotic chromosomes of barley. Despite the differences between these two organisms regarding genome size and the presence of topologically associated domains during interphase, they both seem to share the same packing strategy, relying mainly on the different roles of condensins I and II. With polymer simulations, we can also infer structural aspects specific to barley, such as the loop lengths, the height of a helix turn, or the relative concentrations of condensins. These are all features that may help us understand how the large barley genome is organized during the cell cycle. Furthermore, de novo modelling may indicate which forces are acting on the polymer and are driving chromosome condensation. This may shed a light into the regulation of the process and how the involved proteins work and assemble during chromosome condensation.
Development of desired quality traits with parallel loss of wild type robustness during domestication of crop plants is reflected by remarkable change of phenotype. In our study we systematically evaluated transferability of complex phenotypic traits between the wild desert-adapted tomato *Solanum pennellii* and the domesticated cultivar *Solanum lycopersicum* using an integrated genomic, transcriptomic, metabolomic and phenomic analysis of a large introgression population. Our results reveal potential strategies of transferring quality traits such as pathogen resistance in crops, but also highlight a power of integrated multi-omic QTL analysis for pathway discovery and gene function annotation.
The U.S Department of Energy Systems Biology Knowledgebase (KBase; http://kbase.us) is an open-source, web-accessible platform designed for systems biology research of microbes, plants and their communities. It provides an extensible range of integrated biological data types and associated analytical tools (Apps) presently including gene expression and transcriptomics analysis, comparative genomics, genome annotation, metabolic simulation, and visualization. KBase has a rich set of computational methods and curated datasets for gene expression analysis based on RNA-seq, including a selection of preprocessed high-quality reference genomes and a variety of Apps supporting the Tuxedo tool suites, allowing users to generate an expression matrix of reads based on plant genomes. Apps for downstream analysis include clustering of expression profiles, reconstruction of primary metabolism, and simulation of metabolic pathways. KBase data and Apps are available from within an interactive notebook that supports the creation of dynamic workflow documents called Narratives, enabling experimental and computational biologists to work together to share and publish their data, approaches, workflows, and conclusions, leading to transparent and reproducible computational experiments.
Posters
Resistance genes are the key to overcome many diseases and thus to sustain animal and plant populations. Especially the perennial crop grapevine (*Vitis vinifera*) is highly susceptible for several pests and diseases. Therefore, determining genes that mediate resistances to those pathogens in resistant grapevines is of great interest. As we generated a potentially fully phased, high quality de novo genome assembly of the diploid grapevine rootstock cultivar ‘Börner’ that shows resistance to a variety of pathogens, we have a perfect basis to identify and to annotate known and also new resistance genes.

Our genome assembly is based on long PacBio reads. It is computed through dividing the two types of parental reads with TrioCanu (Koren et al., 2018) utilizing short Illumina read data of the parental lines of ‘Börner’, *Vitis riparia* GM183 and *Vitis cinerea* Arnold, and subsequently assembled with Canu (Koren et al., 2017) while using both read types separately as input. The result is one genome assembly per haplotype with an N50 of 4.45 and 4.89 Mb, respectively. Starting with both haplotype assemblies, resistance gene analogs were identified with hidden Markov model matrices and profile matrices. The genes were then classified into resistance gene analog (RGA) classes according to the found resistance domains. Moreover, a blast search against the plant resistance gene database PRGdb (Sanseverino et al., 2010) was performed. We also started to classify the resistance genes into those with a match on both haplotype assemblies, with an incomplete match and without a match on the other haplotype. With all data we hope to reveal new resistance genes and to identify candidate genes for e.g. phylloxera, downy and powdery mildew and black rot.

Partners:

Sequencing: Bruno Huettel, MP-GC MPIPZ, Cologne and Prisca Viehoever, Bielefeld University
Session 1 // Breeding Informatics

**Whole Genome Sequencing and Annotation of Octoploid Strawberry (Fragaria × ananassa) Inbred Lines**

Manivannan, Abinaya* (1); Lee, Sun Yi (1); Kim, Dosun (1); Lee, Eun-Su (1); Lee, Hye-Eun (1); Kang, Byoung-Cheorl (2); Kim, Jinhee (1)

1: Vegetable Research Division, National Institute of Horticultural and Herbal Science, Rural Development Administration, Republic of Korea.; 2: Department of Plant Science, Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, Republic of Korea.

Recent advances in next generation sequencing (NGS) approaches have facilitated the sequencing of various genomes with high quality and affordable prices. In addition, the reconstruction of chromosome level whole genome sequences of even non-model organisms has become affordable with third generation sequencing technologies such as “Single molecule real-time sequencing (SMRT) and nanopore sequencing” methods. In the present endeavor, the de novo genome sequencing and annotation of cultivated strawberry *Fragaria x ananassa* have been discussed. *Fragaria x ananassa* (2n = 8X =56) is a widely consumed crop species belonging to the Rosaceae family. The genus Fragaria consists of 24 wild species, including 13 diploids, 5 tetraploids, 1 hexaploid, 4 octoploids, and 1 decaploid. The origins of the wild strawberry species are distributed throughout Eurasia, North and South America, and Japan. Here, we report the genome sequencing and assembly of cultivated strawberry using PacBioRS II platform long read sequencing approach. The present draft genome consisted of high homozygosity in strawberry for the first time. The sequencing coverage of 80× was obtained with 40 single-molecule real-time cells in the present study. Further, sequences were assembled de novo and anchored to the genetic linkage map into pseudochromosomes. Gene prediction results suggested the identification of important genes involved in various biological processes. Thus, the draft genome sequenced for cultivated strawberry using cutting-edge genome sequencing approaches will facilitate the strawberry genomics research and enhance the strawberry breeding.

Acknowledgments: This research work was supported by the Cooperative Research Program for Agriculture Science and Technology Development [Project No. PJ0134412019], Rural Development Administration, Republic of Korea.
Variation of gene content and gene expression in terms of relative quantitative expression and tissue/organ specificity is a substantial factor affecting phenotypic diversity. In crops, particularly in cereals, the pan-transcriptome and pan-genome concepts are being increasingly investigated after the reference genomes were made available. Characterizing the gene expression presence-absence variation (ePAV) of tetraploid durum wheat (*Triticum turgidum* ssp. *durum*) enables to investigate the association between the genotypic and phenotypic variation at an unprecedented level of precision.

The current study presents the transcriptome analysis for 13 elite varieties from worldwide germplasm. Gene expression variation is described in relation to a high-quality reference genome sequence assembly of durum wheat cv. Svevo (c/o International Durum Wheat Genome Sequencing Consortium). cDNA libraries were produced from roots and leaves at the seedling stage and from developing grains. In order to study the gene expression pattern, these RNA-seq libraries were aligned to the durum wheat genome and the transcript abundance was calculated. Overall, 75.0, 70.5 and 74.5% of high-confidence Svevo genes were expressed in grain, leaf and roots, respectively. Principal Component Analysis (PCA) analysis showed a clear gene expression clustering lead by organs. Hierarchical clustering based on PC scores clearly differentiated up- and down-regulated gene clusters based on tissues and varieties. Variance expression analysis projected on the Svevo assembly allowed us to identify the chromosome regions that drove the major expression variation patterns. Interestingly, by clustering the gene expression profiles and the cultivar’s expression profiles several gene expression patterns related to the ancestry relationship among cultivars were evidenced, particularly for the grains. The functional annotation of these gene clusters is in progress. Towards assembly of a pan-transcriptome in durum, the cultivar-specific reads that could not be mapped on the Svevo genome (4-30% referred to the Svevo Illumina sequencing data) are being de novo assembled.
Modern crop breeding relies heavily on the use of large numbers of molecular markers. Single nucleotide polymorphisms (SNPs) have become the molecular marker of choice due to their abundance and because they are amenable to high-throughput genotyping with genotyping chips and KASP assays. We have produced an extremely large set of SNP markers for barley, making use of a diverse collection of barley exome capture datasets compiled from a variety of different projects and collaborating groups. The 1,336 barley lines used encompass representative cross-sections of the wild barley, landrace and cultivar gene pool. The raw data was mapped to the 2017 barley pseudomolecule reference sequence with BWA, and variant calling and genotyping was carried out using the Genome Analysis Toolkit (GATK). This produced a set of 107 million variants, with 101 million SNPs and 6 million small indels, equating to a change rate of 1 in 44 bases. The majority of variants were annotated as either being located in regions 5 kbp up-or downstream of genes (42.53%) or intergenic (26.46%), with the remainder being located in genic regions (31.01%). Three percent of variants were annotated as non-synonymous coding, while 0.6% occurred in splice sites. Variants were strongly concentrated towards chromosome ends as previously reported, reflecting the lack of recombination in the pericentromeric regions. A large proportion of reads mapped to off-target regions, and in conjunction with the large number of lines used this provided substantial read coverage that allowed the calling of robust variants and genotypes in regions not covered by exome capture probes, thus largely removing the constraint associated with a capture array. This dataset likely captures most of the barley variome and thus represents an invaluable resource for breeders and researchers alike.
Rapeseed is not only the most import home-grown oil crop in Germany, but also a very productive domestic protein crop. Its protein exhibits an excellent amino acid composition and nutritional value for human nutrition, but still awaits a broader utilization. The seed oil content (SOC) as well as the seed protein content (SPC) are two complex quantitative traits controlled by many genes with assumed additive and epistatic effects. SOC and SPC have a negative correlation and are frequently influenced by the environment. The complex regulatory mechanisms of both traits remain ambiguous. In order to promote food use of rapeseed protein, the identification of genomic loci causal for the SPC is one goal of the interdisciplinary project. To achieve this, we apply mapping-by-sequencing (MBS). The principle behind this approach is the identification of causal variations by crossing homozygous parents, pooling the DNA of phenotypically identical F2 plants, deep sequencing of the pooled DNA, and analysing the allele frequency differences between the pools by using a SNP marker table. For the molecular identification of genomic loci controlling the SPC trait we deep sequenced pools selected from a large segregating F2 population. Four pools of extreme genotypes were build based on phenotypic evaluation of the SPC (and SOC) content in the F2 individuals of two different environments. DNA from these pools was sequenced, mapped to a reference sequence and evaluated for polymorphisms. By analysing changes in allele frequencies between the high and low pools at each location, we were able to detect many genomic regions with a putative influence on SPC. The „consensus“ regions or intervals are probably the stable once. Repeating the MBS analysis with improved de novo assemblies of the parental lines is another strategy to decrease the number of detected intervals and to enhance the resolution in the remaining intervals.
Detecting large chromosomal Modifications using Short Read Data from Genotyping by Sequencing

Keilwagen, Jens* (1); Lehnert, Heike (1); Berner, Thomas (1); Beier, Sebastian (2); Scholz, Uwe (2); Himmelbach, Axel (2); Stein, Nils (2,3); Kilian, Benjamin (4); Hackauf, Bernd (1); Perovic, Dragan (1)

1: Julius-Kühn Institute, Germany; 2: Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany; 3: Georg-August-Universität Göttingen, Germany; 4: The Global Crop Diversity Trust, Germany

Genotyping by sequencing data has often been used to detect small polymorphisms including single-nucleotide polymorphisms and short insertions or deletions. Here, we aim at detecting larger modifications of chromosomal regions. These large chromosomal modifications could be duplications, deletions, substitutions including introgressions as well as modification of DNA methylation.

We demonstrate the applicability of the method using *H. bulbosum* introgression lines. Subsequently, we identify large chromosomal modifications in wheat and barley.
Harnessing natural variation in photosynthetic capacity is a promising route towards yield increases, but physiological phenotyping is still too laborious for large-scale genetic screens. Here, we evaluate the potential of leaf reflectance spectroscopy to predict photosynthesis related physiological parameters in the C4 crop *Zea mays*. The method is fast and non-destructive so that high numbers of plants can be evaluated. Preliminary experiments showed that a minimum of about 50 samples per species was required for reliable model development. Leaf reflectance spectra were collected in parallel with physiological parameters such as FW/DW ratio, specific leaf area, chlorophyll content, CN ratio, maximal assimilation, carboxylation capacity, delta13C, Rubisco, PEPC and MDH activity in plants grown in field as well as greenhouse trials. The experiments included the reference lines such as B73 and Mo17 and selected lines from an introgression library, some plants were also exposed to water or nitrogen stress. Suitable models for prediction of leaf parameters from the leaf reflectance spectra were tested including partial least square regression (PLS), PLS with recursive feature elimination (RFE), random forest and random forest with RFE. The most reliable models were obtained with PLS for FW/DW ratio, specific leaf area, chlorophyll content, maximal assimilation and Rubisco activity. Our results indicate that leaf reflectance phenotyping is an efficient method for improving crop photosynthetic capacity.
Wheat is a globally important cereal crop and a staple food source. For global food security, enhanced drought tolerance in wheat is critical for sustainable food production. Various key genes and transcription regulators governing morpho-physiological traits in drought tolerance have been revealed in recent years with the advancement in sequencing technologies. A lack of knowledge of the molecular mechanisms involved in drought stress has been a limiting factor for the effective management of wheat; therefore, it is imperative to assess the expression of genes involved in response to drought stress. In this study, differentially expressed genes (DEGs), under drought condition was investigated using Illumina RNA-Seq transcriptome profiling. A total of 65,722 and 69,074 DEGs were identified under drought in susceptible and tolerant genotype, respectively. Gene Ontology (GO) analysis revealed several sub-categories related to the stress, including response to stress, defense response and response to stimulus in the tolerant genotype WH 1025 as compared to the sensitive genotype WH 1105 under drought stress. Several Transcription factors (TFs) like NAC domain-containing protein, WRKY transcription factor and MYB transcription factors were identified which are responsible for expression of GRAM domain-containing protein/ABA-responsive, various metabolic pathways for the synthesis of proline and hydroxy proline and MDA receptors proteins. Drought tolerance dissection in the genotypes revealed that the genes and the pathways involved in roots of WH 1025 were the key factors to make a difference. The identified TFs/Genes up-regulated in roots of WH 1025 during drought, were potential candidates for enhancing tolerance and can be used to improve drought tolerance in elite wheat cultivars and other cereal crops.
Identification and Comparative Analysis of Differential Gene Expression Revealed by RNA-Seq in Wheat Seedling under High Temperature Stress

Munjal, Renu*
CCS Haryana Agricultural University, Hisar Haryana INDIA, India

Area under wheat is likely to increase in future and therefore, breeding efforts should be focused on developing varieties that have faster growth rate and tolerant to heat stress at reproductive stage of plant growth. An important factor when considering this new genetic resource is whether it may offer beneficial alleles for increased yield potential, therefore, it is imperative to assess the expression of genes involved in response to high temperature stress. In this study, differentially expressed genes (DEGs) under high temperature condition was investigated using Illumina RNA-Seq transcriptome profiling. A total of 357 and 455 DEGs were identified in the susceptible (WH 1105) and tolerant genotype (WH 730), respectively, 94 DEGs were detected as common DEGs related to high temperature response in both genotypes, and a total of 589 DEGs were detected as high temperature tolerance-associated genes. Gene Ontology (GO) analysis revealed several sub-categories related to the stress, including response to stress, defense response and response to stimulus in the tolerant genotype WH 730 as compared to the sensitive genotype WH 1105 under high temperature stress. By combining protein function clustering analysis and significantly enriched GO analysis suggest that transcription factors may play a dominating role in the high temperature stress tolerance. Furthermore, 74 differentially expressed transcription factors were identified, of those many genes involved in the metabolism and regulation of hormones. These results expand our understanding of the complex mechanisms involved in high temperature tolerance in wheat, and provide a set of candidate genes for further genetic analysis.
Efficient and accurate prediction of major phenotypic traits has become paramount to ensure the selection of the best individuals in the context of a breeding program. Various linear regression models have been developed to enable the estimation of breeding values based on a training set characterized both genotypically (e.g. with single nucleotide polymorphisms – SNPs) and phenotypically. Unfortunately, these statistical methods often fail to capture gene-gene and gene-environment interactions, oftentimes because of more complex non-linear relationships.

In plant and animal breeding, different machine learning methods have recently received attention for their ability to detect these types of interactions, and thereby to allow improving the prediction of complex polygenic traits. With the constantly increasing dimensionality of genomic and phenotypic data, deep learning approaches (based on Artificial Neural Networks – ANNs) appear as a tool of interest for marker-based predictions. Their key benefit is the absence of any assumption regarding the genetic architecture of the trait under study, unlike parametric statistical models.

Nonetheless, much examination is often required upstream, regarding the most appropriate model architecture and the fine-tuning of the hyperparameters (generally done using cross-validation) which strongly affect the performance of the ANN. We investigated different ANN models compared to the GBLUP method with plant traits datasets, and by using different input data.
As a minor crop with only local importance our understanding of the evolutionary history of domesticated rye and the relationship to its wild relatives are still limited. In contrast to wheat and barley, rye is not considered as a founder crop of Neolithic agriculture, even if the wild progenitors of all three cereal crops share the same area of distribution in Southwest Asia. Instead, rye is assumed to be a secondary domesticate that became widely used in Central and Eastern Europe only after its introduction as a weed in the course of the expansion of domesticated wheat and barley. The ability to thrive on poor soils and the high frost tolerance has enabled rye to become a suitable crop under harsh conditions as found in the northern areas of Europe.

Here, we use the combination of genotyping-by-sequencing and whole genome resequencing data on 1002 Secale samples mapped to the recently assembled high quality reference genome to study the population history of the small genus Secale. Overall, the weak genetic differentiation between wild and domesticated rye points to ongoing gene flow and a fairly recent speciation leading to incomplete lineage sorting and low fertility barriers.

The analysis of the population genomic history of domesticated rye within the complex population structure of the different wild Secale species points to the idiosyncratic characteristics of the domestication process of rye. Combining these newly available genomic resources with archaeological and linguistic evidence enables us to explore the differences in use and perception of rye in different geographic areas as a result of the underlying past processes.
Amphibians perform different ecological services in the natural world in addition to their role in food chains and food webs and in maintaining natural ecological stability and that includes dissemination of seeds of plants or providing nitrogen to epiphytic plants in their high canopy ecosystem through faeces and other organic matters. Very little or nothing is known around the globe regarding the role of amphibians as successful pollinators. We suggest that tree frogs completing their life cycles in high canopy ecosystems taking refuge in smaller epiphytic plants like bromeliads and orchids in tropical and subtropical ecosystems while breeding and foraging for food could act as potential natural animal pollinators. Primary insect pollinators visiting flowers for nectar and pollen fall prey to tree frogs foraging from one flower to another in search of suitable food (insect) sources. The moist skin of the tree frogs could easily facilitate pollens being dusted on their body and thereby achieve cross-pollination by visiting another flower of the same species on their high canopy home surviving inside epiphytes. However, there is no documentary proof available right now. Hence there is an opportunity to place camera traps at ideal positions on high canopy epiphytes to document tree frog mediated natural pollination in tropical ecosystems. Furthermore, GPS trackers could be attached on some males and females of selected tree frog species living in high canopy ecosystems to monitor their behaviour and movement through telemetry. Lastly, tree frogs could be observed in specially designed vivarium with ideal ecosystem, appropriate epiphytic species, pollinator suitable insect species with proper temperature, moisture and humidity regulations to record their natural behaviour under artificial environment through trap cameras, highly specialized videography, thermal image analysis and GPS tracking to identify and record any potential pollination activities.
The Bean Adapt project aims to dissect out the genetic bases and phenotypic consequences of adaptation to new environments by common bean and runner bean. The focus is to study their introduction into Europe from their respective centres of domestication and their subsequent expansion. The project will provide the scientific community, breeders and seed industry with knowledge of genes, quantitative trait loci, and genetic mechanisms that contribute to the phenotypic adaptation associated with environmental conditions, along with their mapping along the reference genome.

Our goal and task at the Bioinformatics and Information Technology Group (AG BIT) is to collect and organize all the data from the partners involved in the project in order to create a Data Warehouse that can be used to build a web access to the data and also data exportation for the bean research community. The data management of the project involve the curation of the data, template creation in order to homogenize the incoming data and also data storage.
Mangrove ecosystems provide a multitude of highly relevant ecological services such as storm protection, nursery, food provision and tourism. Still, there is a constant area loss of 2-8% annually. Therefore, we are developing a new approach for environmental protection planning by combining cutting-edge technologies and methods in biology, chemistry and socio-ecological sciences.

To reach our target, we analyze sediments and roots/rhizosphere of mangrove forests from six different sampling sites all over the world. Sediment contains numerous microbial microorganisms and deposits from dead and living organisms. Thus, investigating the abiotic and biotic sediment parameters facilitates the characterization of coastal ecosystems and the determination of their environmental conditions.

Another part of the project models the ecosystem, including the relations among biotic ecosystem compartments and external factors, such as use by the local human population and feedbacks as well as effects thereof.

The contribution of the Leibniz Institute of Plant Biochemistry is the development of methods to analyze and identify small chemical compounds of the sediment organic matter (SOM) by mass spectrometry. The chemical composition and structure of the SOM reflects sources (e.g. different mangrove species) and the processing of organic compounds, thus providing insight into interactions among ecosystem compartments.

Moreover, the development of new methods and software is necessary because the characteristics of such ecological experiments are very different from experiments designed with controlled conditions. For example, there are many more factors of influence, organisms and compounds to consider in environmental research. This diversity requires improvements to the already established methods, e.g. peak picking or retention time correction.

We present the concepts of algorithmic improvements, the status of implementation and evaluation results based on the project’s data sets.
Aegilops tauschii is the wild progenitor of the D genome of hexaploid wheat and is native to a region between Turkey and Pakistan. The publication of a high quality reference sequence gives the opportunity to investigate the genetic variances in this species. As a natural population Aegilops tauschii shows a wider range of phenotypic variation than domesticated species like bread and durum wheat. To capture this variation, a diverse panel of additional lines was sequenced. These lines were selected for their high variety of phenotypic traits such as flowering time and disease resistance. They originate from a wide geographical range and different altitudes. A. tauschii is divided into two sub-lineages (lineage 1 and 2 which roughly correspond to the subspecies A. tauschii tauschii and A. tauschii strangulata). Both sub-lineages are included in this panel. The sub-lineage 1 did not contribute to the D genome of modern wheat cultivars. The study of these accessions can give a better insight into the genetic basis underlying environmental adaptations in Triticeae. They can also be a pool for wheat improvement.

150 lines were sequenced by the Open Wild Wheat Consortium (http://www.openwildwheat.org/) of which 24 lines, spanning the phylogenetic diversity of these Aegilops accessions, had a coverage of 30x. In a reference guided approach the lines, having a higher coverage, are investigated for their gene presence-absence and copy number variations. They are also compared for their pseudogene pool as this can be a source of gene resurrection and regulation. It can be shown that accessions belonging to sub-lineage 1 show a distinct pattern of gene absence and variance compared to those of sub-lineage 2.
Anthropogenic factors are leaving big footprints in our natural ecosystems and environments around the planet negatively impacting global biodiversity. One of the worst impacted species around the planet with serious long-term consequences for agriculture, forestry and apiculture are the honey bees and native (indigenous) bees. Several anthropogenic factors such as excessive use of agricultural chemicals, pollution, parasitic diseases, Colony Collapse Disorder, rapid changes in land use patterns, habitat loss, lack of suitable bee foraging plants with adequate supply of pollen and nectar throughout the year to sustain bee colonies are directly or indirectly impacting bee diversity significantly. It is therefore important to conserve these important, farmer-friendly and eco-friendly insect pollinators from the real dangers of extinction. One of the biggest challenges towards successful insect pollinator/bee conservation is the lack of awareness and education among the public, lack of communication and networking between ordinary citizens and academics, researchers, students, journalists, lawyers, administrators, bureaucrats, politicians, agrologists, apiculturists, farmers, foresters, ecologist and conservators. It is therefore important to develop an app for these targeted communities to connect them together through a simple tool that everyone across the globe could easily download on to their phones and computers and access and/or use it as an information highway for exchanging pollinator insect/bee related information among them. Examples could be taking insect pollinator images across the world, identifying them and storing these into an online available pollinator insect image database/archive for public education, encouraging, promoting and exchanging bee-related research data, information and communication, questions and discussions among various interested groups to stimulate public debate, education and awareness. As more people are involved, bee education and awareness will increase; and help building consensus for bee conservation as a priority forcing governments to take suitable comprehensive actions and/or policies to help in successful insect pollinator conservation.
Flaxseed is one of the richest sources of essential omega-3 fatty acid, in particular, alpha-linolenic acid (ALA), which has been reported to dampen inflammatory reactions, thereby reducing the risk of heart attack or stroke, and other diseases. Given these health benefits of these fatty acids, several studies analyzed the genes involved in their biosynthesis. While the biosynthetic pathway of fatty acids in flaxseed involves different enzymes, genes, like stearoyl-ACP desaturase (SAD) have been identified to play an important role in the desaturation of fatty acids, as SAD genes are responsible for the conversion of stearoyl-ACP to oleoyl-ACP by introducing a double bond at the Δ9 position.

Taking advantage of these studies that identified genes involved in fatty acid biosynthesis, we wanted the extent of the genetic variability of these genes, their corresponding isoforms in a collection of flaxseed cultivars. The major aim was to study the polymorphism of sad genes by sequencing these genes from a collection of flaxseed cultivars. Next, we plan to correlate obtained polymorphisms to the fatty acid composition in these cultivars, which should help analyze the functionality of certain polymorphisms.
To meet the demand for future crop pollination services, it is important to understand the foraging ecology of wild and managed bees with respect to spatial and temporal changes in agricultural landscapes. In this respect, the identification of pollen resources can reveal part of their food plant preferences and requirements.

The internal transcribed spacer sequence (ITS2) is a commonly used genetic species barcode. We are sequencing ITS2-derived amplicons from plant pollen collected by bees in order to identify pollen source species. From this data we derive bees pollen foraging with respect to bee species under given environmental settings. Since ITS2-amplicons generated with common primer pairs typically exceed the length of polymerase-derived NGS-reads, we are evaluating full-length nanopore-based amplicon sequencing. Here we present the first insights from the comparison of microscopic, polymerase-based, and nanopore-based NGS pollen analyses. We evaluate these results with the ultimate goal in mind, to develop a system that allows "close-to-realtime" analyses in the field.
Chromosomal inversions occur in natural populations of many species and may underlie reproductive isolation and local adaptation. Traditional methods of inversion discovery are labor intensive and lack sensitivity. Here, we use three-dimensional contact probabilities between genomic loci as assayed by chromosome conformation capture sequencing (Hi-C) to detect multi megabase polymorphic inversions. These inversions were validated by fluorescence in situ hybridization and Bionano optical mapping. We applied our method for inversion discovery to a barley diversity panel comprising sixty accessions, mainly sampled from the German ex situ gene bank. In this panel, we identified dozens of inversions relative to the Morex reference genome, most of them occurring at the low frequency. Notably, one frequent inversion was detected on chromosome 7H. In conclusion, Hi-C is a powerful method for detecting large intrachromosomal inversions at a population-scale in large-genome crop species.
Lipidomics profiling of seeds of 320 *Arabidopsis thaliana* accessions was conducted using liquid chromatography mass spectrometry (LC-MS), with a focus on triacylglycerols (TAGs). Lipid levels were mapped using Efficient Mixed-Model Association (EMMA) and Mixed Linear Model (MLM), with 200k single nucleotide polymorphism (SNP) markers. The approaches were selected for the purpose of comparison, but also to increase robustness of the findings. The performance of each approach was assessed by comparing the identified genes with a reference list of lipid-related genes downloaded from http://aralip.plantbiology.msu.edu and by determining the significance of the overlap per permutation test.

Following this strategy, we expanded the list of lipid-related genes and provided new insights in the genetic architecture of lipid metabolism in *Arabidopsis*.
Revealing Organelle Genomes Diversity of Barley Alloplasmic Lines

Makarevich, Anna*; Siniauskaya, Marina; Pankratov, Vasily; Liaudanski, Aleg; Shymkevich, Andrei; Halayenka, Inessa; Danilenko, Nina; Davydenko, Oleg

Institute of Genetics and Cytology, National Academy of Sciences of Belarus (Minsk, Belarus)

A big collection of barley alloplasmic lines was previously created and maintained in our lab as a suitable model for studying nuclear-cytoplasmic genetic interactions. It was used to confirm significant effects of nuclear-cytoplasmic interactions on different phenotypic traits. Earlier studies revealed organelle genomes diversity within this collection using RFLP. NGS of complete chloroplast and mitochondrial genomes gave us wide picture of this variability.

In this study we used the organelle DNA of 3 varieties of *Hordeum vulgare* subsp. *vulgare* (Vezha, Roland, Vizit) and 9 alloplasmic lines with nuclei from these varieties and cytoplasm of 3 different lineages of *H. vulgare* subsp. spontaneum (W3, W4, W8), obtained by 7 rounds of substitution backcrossing. CpDNA was obtained from chloroplast fraction isolated by differential centrifugation and contained the admixture of mtDNA. A standard NGS data processing pipeline was changed and optimized to assemble full sequences of both genomes from such mixtures, considering the homology within and between them.

Comparison of 18 complete chloroplast genomes of *H. vulgare* (12 sequenced in this study and 6 downloaded from GenBank) revealed the presence of 102 polymorphic loci: 7 INDELs, 77 SNPs, 18 SSRs. Some of the SNPs are located in protein-coding regions (20) and lead the amino acid substitutions (6). The analysis of 14 complete mitochondrial genomes (12 sequenced, 2 downloaded) showed lower variability: 1 INDEL and 20 SNPs (2 in coding area, synonymous).

The phylogenetic trees for complete chloroplast and mitochondrial genomes were built using MrBayes. Topologies of the trees built based cpDNA and mtDNA data were consistent with each other, in accordance with parallel evolution of the two genomes. Both cp and mt trees separated on two large clades containing as wild as cultural forms, which consistent with hypothesis of more than one center of barley origin.
Once a suitable reference genome sequence and structural annotation are generated, genetic differences within a species are often assessed by re-sequencing. Variant calling processes can identify small sequence variants between two strains, accessions, genotypes, or individuals. These variants can be enriched with predictions about their functional implications based on available structural annotations. Although these predictions on a per variant basis are accurate in most cases, some challenging cases require the simultaneous incorporation of multiple adjacent variants into this prediction process. Examples are neighboring variants which interact by modifying each others’ functional impact. Neighborhood-Aware Variant Impact Predictor (NAVIP) considers all variants within a given protein coding sequence when predicting their effects. Possible applications of this tool are the annotation of variants identified through mapping-by-sequencing approaches or located in manually selected genes of interest. NAVIP is freely available on github: https://github.com/bpucker/NAVIP.
About seven million accessions of Plant Genetic Resources (PGR) are hosted globally in genebanks. Although some genebanks provide access to legacy data and to limited genotypic information, none of them offer comprehensive access to deep genotypic and phenotypic information of entire collections. Likewise, no genebank can so far support informed selection of accessions other than based on passport or legacy data. However, constantly decreasing cost for sequencing of plant genetic material allows extensive sequencing projects to be carried out. This allows to genotype entire genebank collections. In combination with phenotypic characteristics that have also been collected over decades, genebank databases need to be further developed towards information retrieval and data analysis systems. Cluster information, correlations between accessions and even causalities regarding genotype-to-phenotype relationships are often hidden in the highly multidimensional data potentially available for structured PGR collections. We contextualise passport, evaluation data and genotyping data in a broad scale to build an integrated platform for a facilitated utilisation of crop plant biodiversity. Several database and middleware strategies and technologies have be investigated to meet the requirements to store and access billions of data points and to provide an interactive, integrative and convenient web portal.
Next Generation sequencing technologies have opened the door to multiple genome analyses and an increased understanding of the variations present in populations. To date, most of the germplasm analyses have relied on the comparison of sequence reads to one reference genome assembly, limiting our understanding of genomic variation. NRGene has developed novel analytics and approaches to efficiently describe the relevant variation across germplasm using sequence-based haplotypes. These analytics along with new sequencing library methods from iGenomX and low-cost Illumina based sequencing have enabled a cost effective, high resolution, whole genome sequence method for genotyping maize populations. A genotyping example based on ultra-low sequence coverage using this method will be presented.
Plant roots are key drivers of plant development and growth. Quantitative characterization of complex root architecture and its development is important for the assessment of a complete plant phenotype. However, automated analysis of spatial-temporal root organization represents a challenging task which requires a combination of advanced methods and algorithms to perform robust and accurate image segmentation.

Conventional set-ups for root visualization are based on the cultivation of plants in artificial, optically transparent media such as gels or liquids that lack the natural physical properties of soils. Screening of roots in their natural opaque environment is, however, challenged by the low contrast of non-homogeneous soil background with scarce structures and a high level of noise. These problems make automated root phenotyping a non-trivial task.

Here, we present a novel approach to structure-preserving enhancement and segmentation of plant root images. Based on the application of adaptive curvilinear filters and phase congruency features, our method allows to detect roots in noisy images. The results of image enhancement and segmentation were evaluated by a direct comparison with ground-truth (i.e. manually segmented) image data. Our experimental results demonstrate improved accuracy of our adaptive algorithmic scheme in comparison to conventional segmentation using global image thresholding.
In the course of climate change it is important to know how crops behave under drought conditions in the field. Plant response to water limitation stress are very complex mechanism in order to prevent dehydration and retain all essential processes in plant.

We grew *Arabidopsis* plants of the Columbia0 genotype with very high replication in an automated plant phenotyping platform under well-watered and drought-recovery conditions. Covering the well-watered germination phase, the drought (and control) as well as recovery phases we non-invasively assessed growth dynamics and phenotypic features while sampling for RNA-Seq analyses with a high temporal resolution. With the aim of predicting plant phenotypic traits from gene expression profiles we implemented machine learning algorithms trained using gene expression data. To find top genes that best predict several architectural and color based traits we used the random forest algorithm in a two-way procedure. First, we identified phenotypic traits which provide the best classification into stress and non-stress categories. We found color-based traits to be the best classificator for abiotic stress. In the second step we used expression profiles to predict afore selected traits using a regression tree approach. These color traits are best predicted by genes enriched in stress- and ABA-response, anthocyanin/ flavonoid accumulation and water-transport. These candidate genes and related functional hubs in gene-trait-networks provide novel insights into the dynamics of drought-stress response, recovery processes and underlying mechanisms.
Omics-Based Prediction of Hybrid Performance, Dynamic QTL and Systems Genetic Analyses in Canola

Knoch, Dominic* (1); Abbadi, Amine (2); Grandke, Fabian (3); Himmelbach, Axel (1); Meyer, Rhonda Christiane (1); Riewe, David (1); Samans, Birgit (3); Werner, Christian Raimund (4); Snowdon, Rod (3); Altmann, Thomas (1)

1: Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany; 2: Norddeutsche Pflanzenzucht Innovation GmbH (NPZi), Holtsee, Germany; 3: Justus-Liebig-University Gießen, Research Centre for Biosystems, Land Use and Nutrition (iFZ), Giessen, Germany; 4: The Roslin Institute, University of Edinburgh, Easter Bush, Midlothian, Scotland, United Kindom

Based on previous work on biomass and heterosis prediction in Arabidopsis and maize, the project is built on the hypothesis that specific allelic combinations of regulatory genes, their downstream gene expression, as well as elicited metabolite profiles, are associated with improved vegetative growth and seed yield in hybrids. The project pursues two goals: on the one hand to effectively predict hybrid performance in spring oilseed rape by combining information of multiple omics-layers, and on the other hand to identify genetic loci causal for trait variation and to elucidate links between vegetative growth, transcript and metabolite levels. For this purpose, comprehensive datasets have been generated at an early vegetative stage (14 DAS) for a collection of 475 genetically diverse pollinator lines from a commercial canola breeding programme and two elite male-sterile testers. A F1 hybrid population with 950 individuals was generated and evaluated for yield and breeding-relevant traits at multiple locations in the field. Detailed phenotyping data were generated by growing the parental lines and selected hybrids in the automated IPK high-throughput phenotyping platform for large plants. Image-derived phenotype data were complemented by global metabolite (GC-MS) and transcriptome (RNA-Seq) profiles of pools of the same plants. These data were utilised for correlation analyses, and in combination with array-derived SNP and CNV data for genome-wide-association studies. Multiple co-localized marker-trait-associations for different omics-layers were detected and used for candidate gene identification. A time resolved analysis revealed dynamic contributions of loci for the accumulation of biomass and growth rates with certain loci being particularly active in either an early, intermediate and late phase. Finally, the individual and combined data sets of the parental lines were used to predict hybrid performance in the field using gBLUP and RHKS models.
In the last decades, plant research has focused on controlling the growth conditions in plant growth chambers to peel off environmental variation and to display differences that are solely due to the response to treatment and/or genotypic variation to get a better understanding of the function of specific genes. Such artificial cultivation conditions hardly unmask the effects of hitherto unknown genes potentially involved in the response and acclimation to rapidly changing environmental conditions. Thus, to discover new genes responsive to environmental fluctuations, there is a need for new ways to mimic natural growth conditions.

For this study, our non-invasive phenotyping system for small plants was upgraded with supplemental LED illumination, and the growth performance of 382 Arabidopsis accessions was evaluated simultaneously under fluctuating and standard light conditions, in two independent experiments. Time resolved high-throughput phenotyping allowed us to track changes in architectural and colour related traits and chlorophyll fluorescence incrementally over the course of the experiment. The best linear unbiased estimators (BLUEs), calculated over the two independent experiments under fluctuating and standard light, were associated with 214k SNPs distributed over all chromosomes in a genome-wide association study. Linkage disequilibrium was computed from a pairwise SNP matrix of 100k homozygous SNPs. Arabidopsis genome identifiers in full LD with a significant marker trait association (MTA) were annotated as candidates. Under fluctuating light the vegetative biomass of Arabidopsis was smaller than under standard light conditions. Significant MTAs were found specifically for fluctuating light and for standard light conditions, respectively.
Fusarium head blight (FHB) is a major pathogen in wheat, barley and of other small grain cereals - annually leading to dramatic yield losses. More than 250 QTL are associated with FHB resistance including the two major QTL \( Fhb1 \) and \( Qfhs.ifa-5A \) on chromosome 3BS and 5A, respectively. Here, we integrated information from a winter wheat diversity panel for a GWAS study which contains 185 genotypes and an eQTL panel from 96 genotypes to profile for novel and to verify existing resistance QTL. For eQTL and GWAS we incorporated genotypes with dominant and beneficial \( Fhb1 \) and \( Qfhs.ifa-5A \) alleles which enabled the detection of several genetic markers. The markers showed significance in explaining traits like plant height and anther retention but also diseases scores determined after several days of infection.

Eleven of thirty-two genes (34%) from the \( Fhb1 \) locus were among the 25% most variable genes in the eQTL dataset and genes that exhibited expression patterns associated with resistance to \( Fusarium \) and plant height were enriched on chromosomes 1B, 2B, 6A and 6B. Genes that correlated with resistance to FHB were significantly enriched for functional terms like primary and secondary cell wall biogenesis and cell growth. Lack of large overlaps between wheat genes associated with FHB-resistance and \( Fusarium \) responsive genes suggest that many candidate genes involved in FHB-resistance may be constitutively expressed in resistant lines. A population structure effect resulting in a strong linkage between FHB-resistance and other traits provides an alternative explanation to this overlap.
Supervised machine learning is a powerful tool for data analysis. GIENE3 uses this approach to construct a regulatory network from RNA-seq expression data. Using publicly available RNA-seq data of non-mutant strains which was mapped to the current genome annotation we constructed regulatory network of *Chlamydomonas reinhardtii*. *C. reinhardtii* is a photoautotroph green alga which is of interest for the production of biofuels and complex biological compounds. Since light is a renewable energy source it is ideal for the sustainable biotechnological production.

The transcriptional regulatory networks were scale free. The network was queried to identify targets of known regulatory factors and cross-validated using RNA-seq of mutants whose data was not used in network construction. Comparing network predictions with differential expression of mutants validated targets in the immediate vicinity of the regulatory factor. The analyses suggested additional regulatory factors downstream of the initial candidate.
The transition from wet-to-dry is accompanied with a sufficient stress management plan. In plants, the cell wall is the first protection barrier towards external stimuli, while the signaling pathway activates the response and adaptation from within. The plant hormone abscisic acid (ABA) is a key regulator in the development and response to environmental stresses in plants. When the early charophyte green algae (CGA) evolved, one of the biggest challenges was to overcome an extended exposure to drought, potentially leading to the development of ABA and CK response systems similar to land plants. Hence, we were eager to investigate how osmotic stress affects cell wall biosynthesis and stress signaling in *Z. circumcarinatum* UTEX 1559. De novo transcriptome sequencing of *Z. circumcarinatum* (Zygmenatophyceae) was conducted to help understand the genomic/genetic basis underlying the shift from the aquatic to terrestrial environment. Illumina HiSeq 2500 platform produced in total 55,575,710 RNA reads. Assembly was performed with the Trinity suite, yielding 43,573 predicted proteins. Reciprocal BLASTP exhibited a higher similarity of *Z. circumcarinatum* gene orthologs to land plants than to chlorophytes. Gene ontology analysis determined significant enrichment in cellular metabolic processes, response to stimuli, nucleotide binding and cytoplasm.

A total of 15 glycosyltransferase (GT) and 18 ABA/CK genes were selected and subjected to qRT-PCR analysis after 1 hr osmotic stress treatment. Cellulose synthase-like K (*ZcCslK*), a GT, showed the highest expression in response to osmotic stress. Whereas, ABI1 and XDH1 part of the ABA signaling mechanism, showed an increased expression after 1hr stress treatment. This study suggests that CGAs already have the framework for the responds to osmotic stress in a similar fashion as land plants.
Regulation of transcription is a key factor for dynamic adaptation to variable environmental conditions. The binding of transcription factors to cis-elements plays a leading role in transcriptional regulation in eukaryotes. However, most transcription factors cannot be assigned to their specific binding motif yet. RNA-seq data of *Arabidopsis thaliana* was analyzed to predict target genes of 2976 putative transcription factors using a supervised machine learning approach and used as the basis for motif prediction.

The recently published assay for transposase accessible chromatin (ATAC)-seq data provides new insights in open chromatin regions. Both, the 1.5kb upstream region of predicted target genes and open chromatin regions near target genes were used to generate sets of putatively targeted sequences for all transcription factors in *A. thaliana*. The motif prediction was calculated using motif-based sequence analysis tools (MEME-suite). Motifs from DAP-seq data were recalculated and predicted motifs from both upstream and ATAC-seq regions were verified against DAP-seq based motifs. Results from upstream sequences and ATAC-seq data were compared.
Transcription factors occupy key positions for the control of development in an organism. Yet, the majority of transcription factors still lack a functional annotation even in the model organism *Arabidopsis thaliana*.

The surge of RNA-seq experiments in recent years and the availability of genomes for crop and for model plants has led to unprecedented availability of data in the public domain. Data shared according to FAIR principles allowed us to determine non-mutant RNA-seq datasets for an organism. The datasets available on Sequence Read Archive (SRA) were mapped with Kallisto on reference transcriptomes from Phytozome and transcripts per million normalized datasets were merged by locus ID’s. Transcription factors were in each dataset were determined from plantco.de or based on genome annotation files provided with the genome sequence. Global regulatory networks for each selected organism were calculated with GENIE3, a regression-based supervised machine learning approach, to determine the connection between target genes and regulators. The networks were generated for crop plants i.e. *Triticum aestivum* and the model plant *A. thaliana*. Transcription factors were functionally annotated by gene ontology term enrichment of the target genes. Network visualization and comparison between species reveals the extent of evolutionary conservation between seed plants. A number of transcription factors for which a functional prediction is available are in the process of being validated.
Analysing the flux distribution in metabolic networks using constraint-based modelling is invaluable for unveiling the molecular principles underlying those networks. Particularly in plant science, an in-depth understanding of plant metabolism is of tremendous importance to engineer crops with high yield and yield stability to feed the world’s growing population. In this context, the C4 metabolism is of particular interest which employs a carbon-concentrating pump to supercharge the photosynthetic pathway leading to a highly productive metabolism. Genomic studies previously showed that the C4 trait evolved independently over 60 times in mostly hot and arid regions from its ancestral C3 state. Both types of metabolism rely on the same metabolic network and only differ in the abundance of enzymes and transporters due to different gene regulatory mechanisms. Therefore, we prepared a curated version of the Arabidopsis core model to construct a metabolic network of C4 metabolism. For high photorespiratory flux, the C4 cycle is an emergent property of our model without enforcing it by constraining the reactions of the C4 cycle itself. The C4 cycle using the decarboxylation enzyme NADP-ME represent the most optimal solution under resource limitation or photorespiration. PEP-CK and NAD-ME become optimal solutions only if NADP-ME is no longer available by constraint, albeit with a prediction of malate and pyruvate as the transfer acids. Analysing the flux variability identifies aspartate and alanine as less optimal transport metabolites compared to malate pointing to kinetic rather than stoichiometric constraints which drive this choice. All predicted flux distributions are conform to the current knowledge of the C4 metabolism. Even more, we were able to prove the occurrence of the intermediate C2 metabolism on the evolutionary C3-C4 trajectory. Further analysis of the conditions affecting the decarboxylation pathways of the C4 metabolism revealed light as a likely evolutionary driver.
When transmission rates of chromosomes are higher than 0.5, not obeying the Mendelian law of equal segregation, the resulting transmission advantage is collectively referred to as ‘drive’. Although B chromosomes (Bs) possibly show the most common form of drive known for genetic elements, little knowledge exists about the cellular and molecular mechanism behind their drive. In this study, we used the B chromosome of rye as a model and analyzed its variants based on the Next Generation Sequencing techniques. We carried out transcriptional profiling of plants possessing 0B, 2B and chromosomes with a different degree of B non-disjunction at the first pollen mitosis to identify the trans-acting coding and/or noncoding transcripts involved in the regulation of B chromosome drive. Our extensive genome-wide transcriptome study will allow increasing the resolution of B chromosome biology.
Garlic (*Allium sativum* L.) cryopreservation can be a safe and cost-effective tool for the long-term storage of garlic genetic resources. The major challenge for cryopreservation is exposing the explant to cryoprotectant-mediated dehydration and ultra-low temperature (−196 °C) thereby inducing a series of abiotic stress conditions which influence regrowth. The aim of this study is to understand the genetic background of abiotic stress sensitivity by comparing gene transcription differences between cryo-tolerant vs. cryo-sensitive garlic accessions during the cryogenic procedure. De novo transcriptome assembly after RNA-seq experiments resulted in a total of 12,755 unique transcripts (average length of 895 bp and N50 length of 1022 bp) that could be annotated. These genes belong to a wide range of biological pathways. Among them were 1,961 loci with microsatellite motives and 625 transcription factors could be identified. Differential expression analysis identified pronounced effects of the cryo procedures on transcription activity during dehydration and regrow 1 day after liquid nitrogen treatment. We identified significant changes in the abundance of genes encoding for transcription factors such as AP2-EREBP, bHLH, C2H2, Hap3/NF-YB, MYB-HB-like and WRKY, all involved in coping with biotic and abiotic stress. The results obtained in this study will contribute to support garlic cryopreservation and help to understand transcriptomic profiles of non-model plant species using next-generation sequencing technology.