Journal of Biotechnology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Bioinformatics in the plant genomic and phenomic domain: The German contribution to resources, services and perspectives

Thomas Schmutzer[a], Marie E. Bolger[b], Stephen Rudd[a], Jinbo Chen[a], Heidrun Gundlach[c], Daniel Arend[a], Markus Oppermann[a], Stephan Weise[a], Matthias Lange[a], Manuel Spannagl[c], Björn Usadel[b], Klaus F.X. Mayer[c,d], Uwe Scholz[a,*]

[a] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstraße 3, 06466 Seeland, Germany
[b] Forschungszentrum Jülich (FZJ), Institute of Bio- and Geosciences (IBG-2) Plant Sciences, Wilhelm-Johnen-Straße, 52425 Jülich, Germany
[c] Helmholtz Zentrum München (HMGU), Plant Genome and Systems Biology (PGSB), Ingolstädter Landstraße 1, 85764 Neuherberg, Germany
[d] School of Life Sciences Weihenstephan, Technical University of Munich, Alte Akademie 8, 85354 Freising, Germany

ARTICLE INFO

Keywords:
Plant genomics
Plant bioinformatics
Plant genetic resources
Plant data publication
Plant phenomics

ABSTRACT

Plant genetic resources are a substantial opportunity for plant breeding, preservation and maintenance of biological diversity. As part of the German Network for Bioinformatics Infrastructure (de.NBI) the German Crop BioGreenformatics Network (GCBN) focuses mainly on crop plants and provides both data and software infrastructure which are tailored to the needs of the plant research community. Our mission and key objectives include: (1) provision of transparent access to germplasm seeds, (2) the delivery of improved workflows for plant gene annotation, and (3) implementation of bioinformatics services that link genotypes and phenotypes. This review introduces the GCBN's spectrum of web-services and integrated data resources that address common research problems in the plant genomics community.

1. Introduction

The German Crop BioGreenformatics Network (GCBN) is a node of the German Network for Bioinformatics Infrastructure (de.NBI) (Pühler, 2016) and provides crop plant-related bioinformatics services. The three GCBN partners, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Helmholtz Zentrum München (HMGU) and Forschungszentrum Jülich (FZJ), have a strong focus on crop-plant research and provide tailored genomic and phenomic data to the plant research community, through its bioinformatics infrastructure. The main objectives of the GCBN-services include (1) providing transparent web access to germplasm-related data, (2) providing improved workflows for plant gene annotation, and (3) enabling the association of multiple genotypes with corresponding plant phenotypes and traits.

From a bioinformatics perspective, processing and analysing plant genomics data is notoriously problematic for many of the current tools. The widely used next generation sequencing (NGS) technologies are a prime example of such analytical challenges. Crop plants typically have large genomes, often with high ploidity, and typically contain long and complex repetitive regions. These characteristics confound NGS genome assembly tools and the sequencing of the wheat genome for example required the development of novel chromosomal separation techniques to reduce the NGS data to a manageable level which we discuss elsewhere in detail (Bolger et al., 2017).

In recent years, sequencing technologies have led to a considerable increase in the number of published reference genomes (Michael and Jackson, 2013). These genomic resources have accelerated the development of improved crop varieties (Bolger et al., 2014).

The agronomical importance of nutrition and energy traits for crops such as wheat, barley, and maize is a key motivator for the study of their genomes. Genomic research in such crops requires bespoke tools and contextual data repositories that integrate relevant and up-to-date datasets. The broad spectrum of plant genomic resources which are contained in the transPLANT portfolio (Spannagl et al., 2016a) for example, is notable.

Another challenge, from the bioinformatics perspective, stems from the huge phenotypic heterogeneity of the plant kingdom. As a consequence, heterogeneous species-specific databases and metadata standards have been developed to precisely record morphological, agronomic, developmental, pedigree, geographic, environmental and molecular properties of different crops. These aspects require the development of flexible, adaptive, and innovative bioinformatics services. Data standards are of critical importance and GCBN, in consultation with the user communities, addresses this issue to identify, develop and

---

[*] Corresponding author.
E-mail address: scholz@ipk-gatersleben.de (U. Scholz).

http://dx.doi.org/10.1016/j.jbiotec.2017.07.006
Received 14 February 2017; Received in revised form 30 June 2017; Accepted 4 July 2017

Please cite this article as: Schmutzer, T., Journal of Biotechnology (2017), http://dx.doi.org/10.1016/j.jbiotec.2017.07.006

Table 1
Overview of main services offered by GCBN.

With the broad spectrum capture by the seven main services and eight assigned sub services GCBN is providing a wide range of applications. With this we provide several established solutions for the plant science community to accomplish genomics and phenomics tasks.

No.	Service	Integrated Tool/Service	URL	Key Words	Short Description
1	GBIS		https://gbis.ipk-gatersleben.de	Plant Genetic Resources; Information System; Passport Data; Phenotypic Data; Ordering System	Information system of the German Federal *ex situ* Genebank of Agricultural and Horticultural Crop Species
2	EURISCO		http://eurisco.ecpgr.org	Plant Genetic Resources; Information System; Passport Data; Phenotypic Data	European search catalogue for plant genetic resources (on behalf of the European Cooperative Programme for Plant Genetic Resources)
3	Lailaps		http://lailaps.ipk-gatersleben.de	Search Engine, Relevance Ranking, Semantic Query Suggestion; Plant Science	Search engine for a specific, relevance based ranking of gene to trait associations.
4	PGP & e!DAL				
4.1		PGP	https://edal.ipk-gatersleben.de/repos/pgp/	Open Data, Plant Phenomics and Genomics, DataCite, DOI	Repository for a DOI featured publication of plant research data
4.2		e!DAL	https://edal.ipk-gatersleben.de/	Open Data, FAIR; API; Data Sharing; DOI; JAVA	System for generation of DOIs for data submissions
5	IPK Blast Server				
5.1		IPK Rye BLAST Server	http://webblast.ipk-gatersleben.de/ryeselect/	Rye; Blast; Sequence Homology	Web-based BLAST using recent rye genome sequence versions
5.2		IPK Barley BLAST Server	http://webblast.ipk-gatersleben.de/barley/	Barley; Blast; Sequence Homology	Web-based BLAST recent barley genome sequence versions
6	PlantsDB		http://pgsb.helmholtz-muenchen.de/plant/	Plant Genome Data	
6.1		CrowsNest viewer	http://pgsb.helmholtz-muenchen.de/plant/crowsNest/index.jsp	Synteny, Conserved Gene Order	Synteny viewer spanning from chromsome to gene level
6.2		chromoWIZ	http://pgsb.helmholtz-muenchen.de/plant/chromoWIZ/	Chromosomal Location	Visualizing the genomic positions for genes of interest
6.3		RNASeq ExpressionBrowser	http://pgsb.helmholtz-muenchen.de/plant/RNASeqExpressionBrowser/index.jsp	RNAseq, Gene Expression	A tool for searching and visualization of RNA-seq expression data sets for genes of interest to facilitate data interpretation
6.4		DroughtDB	http://pgsb.helmholtz-muenchen.de/droughtdb/drought_db.html	drought stress genes	A manually curated compilation of molecularly characterized genes that are involved in drought stress response and
6.5		REdat	http://pgsb.helmholtz-muenchen.de/plant/recat/index.jsp	Repeat Database, Transposon Classification	A large plant transposon collection focused on de novo detected LTR-retrotransposons and connected to a hierarchical repeat classification scheme (REcat)
7	PlabiPD	Mercator	http://www.plabipd.de/	Genome annotation. MapMan bin classificication.	An online tool to submit large protein or DNA FASTA files for automated annotation and classification of these records into MapMan bins.

promote the appropriate standards for each data type. These standards will include controlled vocabularies (e.g. Crop Ontology, Plant Trait Ontology), data formats (e.g. MCPD-Multi Crop Passport Descriptors), VCF (Variant Call Format), ISA-Tab (Sansone et al., 2012), minimal information standards e.g. MIAPPE (Ćwiek-Kupczyńska et al., 2016; Krajewski et al., 2015) and APIs e.g. BrAPI (http://docs.brapi.apiary.io/). All standards will be adopted from the scientific community in order to ensure maximum interoperability and acceptance of the data formats.

2. Overview of GCBN

The German Federal *ex situ* Genebank of Agricultural and Horticultural Crops at the IPK is one of the largest centres for plant genetic resources worldwide. This high-value resource is a repository of natural genetic diversity of crop plants and their wild relatives. Despite the size of genebank collections (on national and global level) and their agricultural value, such plant genetic resources are not yet amenable to systematic exploitation due to two bottlenecks:

(i) the lack of phenotypic information on the individual accessions, which is required to assess variation at the trait level and
(ii) the absence of sequence information for the vast majority of the samples, which is needed to index the allelic diversity.

Continuous technical improvements and changes in both automated DNA sequencing and automated phenotyping will drive an inestimable increase in the volumes of biological data that will be produced to augment and enrich genebanks (Cobb et al., 2013). The central challenges in the future reaping-of-benefit of genetic diversity rests in both the development of the computational resources to accommodate the expected data flow and in the generation of information from the two convergent streams of genomic and phenomic data. This is addressing the need for solutions tailored for crop-plant science whilst simultaneously scaling up with the massive increase of research data e.g. the number and complexity of phenotypic and genomic data sets (Muir et al., 2016).

GCBN aims to systematically collect germplasm data and integrate it with information from other service providers whilst making these derivative resources available for external analyses. The IPK develops information systems to integrate germplasm passport data, phenotypic data and genetic data. IPK also provides services to access germplasm collections (GBIS, EURISCO), provides information retrieval services over integrated genome annotations (LAILAPS), and provides data sharing and publication services e.g. Plant Genomics and Phenomics Research Data Repository (PGP) or e!DAL.

The key task of GCBN partner HMGU is to develop a generic gene annotation pipeline for high quality structural and functional gene annotations which in turn enable and improve the genotype-phenotype connections needed for efficient plant breeding. To explore plant genomes HMGU provides PlantsDB, a data and information resource for individual plant species that also crosslinks genes, their functions and syntenic relationships between different plant species.

A major goal of FZJ is to develop computational approaches to compare phenotypic observations between different labs, or different species. The first step towards this goal is the establishment of a common vocabulary in the form of standardized ontologies that will describe the nuances of phenotype and environmental data and to develop novel methods for protein functional annotation. Genotype data will be automatically linked to a phenotype based on e.g. genome-wide association and genomic selection methods. Subsequently, more elaborate QTL-based models will be introduced and delivered to the users. PlabiPD is a central portal provided by FZJ which provides integrated access to plant genomes, protein family data, and sequences of protein-coding genes.

The objective of GCBN is to provide strategies and solutions for accessing plant-related information (germplasm data, genetic and phenotypic data) for a wide range of applications within the plant research community. These pilot methods have the potential to be adapted by other facilities where GCBN assists with training and support.

3. Services

GCBN offers seven main services to support the plant research community with tools and information enriched resources e.g. gene sets that are functionally annotated and linked to our manually annotated resources or classifications in domain and gene families that are linked to multiple external databases. Several additional tools and services are offered which are grouped with one of the seven main services, but offer an independent feature or approach to handle certain bioinformatics problems. Table 1 contains the complete list of services including the main purposes for each tool.

3.1. GBIS

The IPK genebank is the largest collection of its kind in Western Europe and is amongst the ten largest collections worldwide, with over 151,000 plant accessions preserved and described. It has been continuously developed over the past 70 years. Providing access to genebank material is an overarching task of the institute.

The complex processes of managing such a huge collection are supported by the Genebank Information System (GBIS) (Oppermann et al., 2015). In addition to the internal management components of GBIS, an online search and ordering system has been established. GBIS is not only an important source of information for both researchers and breeders but the provision of access to genebank material is also of considerable interest to the general public. Thus, GBIS provides a user-friendly interface which allows users to search and filter accessions by various criteria and their combinations. GBIS provides both the passport data and phenotypic observations from genebank accessions. A shopping-cart function allows registered users to request material of interest online.

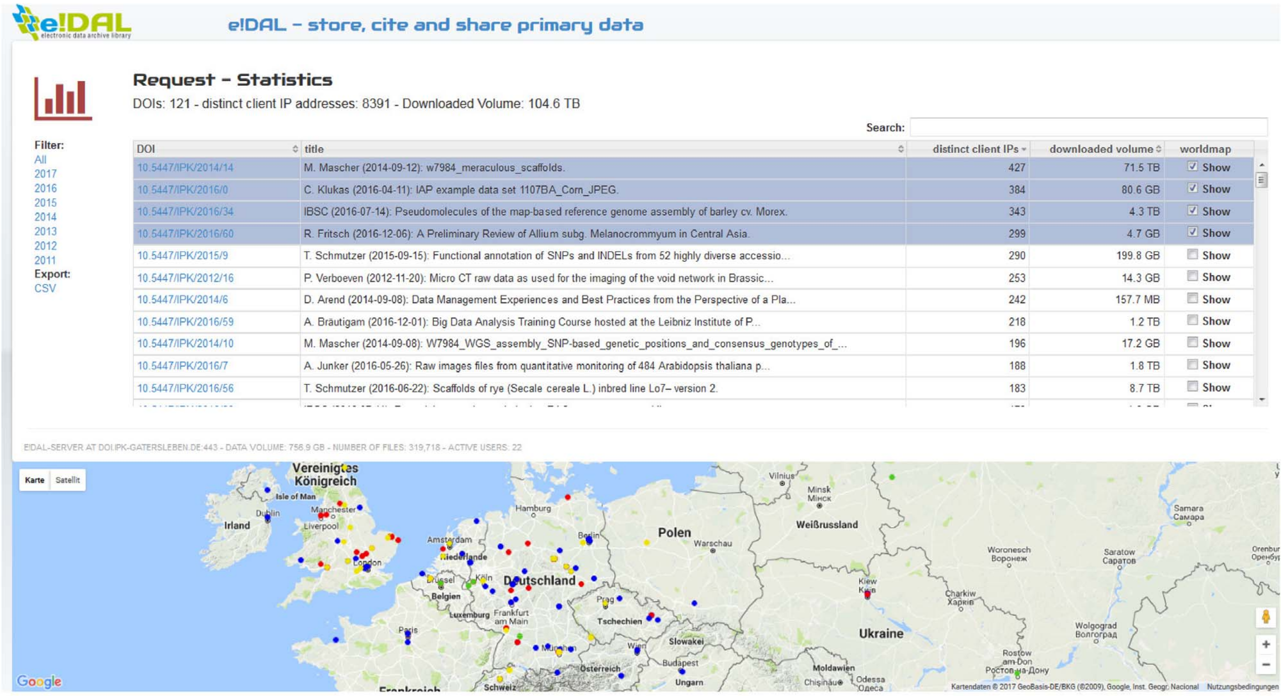
GBIS is also an important data source for global aggregators, such as the Global Biodiversity Information Facility (GBIF) or EURISCO (see below). Data managed by GBIS is used in various projects dealing with research on different cereals and other crops. Supporting the Standard Material Transfer Agreement allows GBIS to fulfil the requirements of the International Treaty on Plant Genetic Resources for Food and Agriculture.

3.2. EURISCO

EURISCO is a European search catalogue providing information on about 1.9 million accessions of crop plants and their wild relatives, preserved *ex situ* by almost 400 institutions (Weise et al., 2017) in Europe. EURISCO is maintained on behalf of the European Cooperative Programme for Plant Genetic Resources (ECPGR) and is based on a network of *ex situ* National Inventories from 43 member countries, which regularly update their data. The genebank accessions documented in EURISCO comprises 6334 genera and 42,938 species, respectively. EURISCO represents an important effort for the preservation of world's agrobiological diversity by providing information about the large genetic diversity maintained by the collaborating institutions.

The EURISCO system contains both passport data and phenotypic information about plant genetic resources. The major crops represented include wheat, barley and maize, which rank amongst the top five major cereal grains produced worldwide (Awika, 2011). Furthermore, EURISCO provides information about a large research collection of the Nottingham Arabidopsis Stock Centre, UK, comprising almost 670,000 *Arabidopsis thaliana* accessions.

The EURISCO infrastructure is based on Oracle™ technology; the



e!DAL – store, cite and share primary data

Request – Statistics

DOIs: 121 - distinct client IP addresses: 8391 - Downloaded Volume: 104.6 TB

Filter:
All
2017
2016
2015
2014
2013
2012
2011
Export:
CSV

Search:

DOI	title	distinct client IPs	downloaded volume	worldmap
10.5447/IPK/2014/14	M. Mascher (2014-09-12): w7984_meraculous_scaffolds.	427	71.5 TB	☑ Show
10.5447/IPK/2016/0	C. Klukas (2016-04-11): IAP example data set 1107BA_Corn_JPEG.	384	80.6 GB	☑ Show
10.5447/IPK/2016/34	IBSC (2016-07-14): Pseudomolecules of the map-based reference genome assembly of barley cv. Morex.	343	4.3 TB	☑ Show
10.5447/IPK/2016/60	R. Fritsch (2016-12-06): A Preliminary Review of Allium subg. Melanocrommyum in Central Asia.	299	4.7 GB	☑ Show
10.5447/IPK/2015/9	T. Schmutzer (2015-09-15): Functional annotation of SNPs and INDELs from 52 highly diverse accessio...	290	199.8 GB	☐ Show
10.5447/IPK/2012/16	P. Verboeven (2012-11-20): Micro CT raw data as used for the imaging of the void network in Brassic...	253	14.3 GB	☐ Show
10.5447/IPK/2014/6	D. Arend (2014-09-08): Data Management Experiences and Best Practices from the Perspective of a Pla...	242	157.7 MB	☐ Show
10.5447/IPK/2016/59	A. Bräutigam (2016-12-01): Big Data Analysis Training Course hosted at the Leibniz Institute of P...	218	1.2 TB	☐ Show
10.5447/IPK/2014/10	M. Mascher (2014-09-08): W7984_WGS_assembly_SNP-based_genetic_positions_and_consensus_genotypes_of_...	196	17.2 GB	☐ Show
10.5447/IPK/2016/7	A. Junker (2016-05-26): Raw images files from quantitative monitoring of 484 Arabidopsis thaliana p...	188	1.8 TB	☐ Show
10.5447/IPK/2016/56	T. Schmutzer (2016-06-22): Scaffolds of rye (Secale cereale L.) inbred line Lo7– version 2.	183	8.7 TB	☐ Show

E!DAL-SERVER AT DOI.IPK-GATERSLEBEN.DE:443 - DATA VOLUME: 756.9 GB - NUMBER OF FILES: 319,718 - ACTIVE USERS: 22

Kartendaten © 2017 GeoBasis-DE/BKG (©2009), Google, Inst. Geogr. Nacional Nutzungsbedingungen

Fig. 1. The e!DAL system for support of research data management.
Statistics of data accesses to DOI datasets published with the e!DAL system.

central entry point is through the web interface, which is available at http://eurisco.ecpgr.org. The web interface provides different possibilities of retrieving information, such as four different standard searches (by taxonomy, accession, status or site), as well as an advanced search functionality that allows users to freely combine all searchable fields within a single query. A variety of statistical reports are also provided. User-specific export functionalities, including a full dump in MS-Access format, are available.

3.3. LAILAPS

LAILAPS is an integrated information retrieval (IR) platform for plant genomics and phenomics data (Esch et al., 2015). It supports fuzzy querying for candidate genes which are associated to traits of interest. The basis is a loosely integrated system of indexed and interlinked databases. The web frontend features an integrated query assistance module and an evidence based relevance ranking system. This functionality is provided by an artificial neural network which encodes user feedback and domain expert knowledge to predict the relevance of search results. Currently, LAILAPS contains around 65 million indexed documents from thirteen major life science databases and provides approximately 80 million links to plant genomic resources.

IPK hosts two LAILAPS services. One LAILAPS service provides access to gene and protein databases, ontologies and compute associations to major European plant genomic resources, e.g. EnsemblPlants (for details on linked databases see Esch et al., 2015). A second, instance provides dedicated access to IPK plant genome databases, molecular databases and germplasm collections (e.g. GBIS). In 2016, the two instances were accessed by 2719 unique users what underlines its great importance for the plant research community. The user data is consistently used to improve performance and search quality by constantly updating the LAILAPS ranking model.

LAILAPS has been used as service for knowledge discovery in various studies. One example of LAILAPS usage is where it is used to provide information retrieval for OPTIMAS-DW, a comprehensive maize transcriptomics, metabolomics, ionomics, proteomics and phenomics data resource (Colmsee et al., 2012). Its underlying capabilities have assisted in the unravelling of relationships among data sets. Querying LAILAPS for 'low nitrogen in maize', for example, revealed nine proteins that were referenced by 13 OPTIMAS unigenes. While the regulation of a majority of these linked OPTIMAS unigenes has been experimentally validated, the functional relationship to a potential down-regulation of OptiV1S22914 by the genes BX7/ZRP4 was discovered by a LAILAPS.

3.4. PGP & e!DAL

e!DAL (https://edal.ipk-gatersleben.de/) is an open source software solution to publish and share research data in compliance with the open-data policies of research publications and funding agencies (Arend et al., 2014). The system is designed as a broker between in-house file storage and data registries. e!DAL is provided as a small all-in-one JAVA package that can be operated as micro service on any network accessible server. Its prerequisites are a sufficiently sized file system, access to an e-mail server, and open http(s) ports. The registration as a DataCite data centre is recommended to enable e!DAL's support for automatic registration of published data through citable digital object identifiers (DOI). Implemented features are:

- http(s) server to support web and DOI access to data sets,
- database to store and associate DataCite and Dublin core (ISO 15836) compliant meta data,
- data submission and review workflow,
- APIs for remote and local access,
- OAI-PMH meta data harvesting protocol,
- version tracking of stored data,
- text search engine

e!DAL is already the basis for the Plant Genomics and Phenomics Research Data Repository (Arend et al., 2016a) where 121 DOIs and more than 310.000 files are published (June 2017). The e!DAL driven PGP repository is accepted as data repository by the journals GigaScience and Scientific Data (Nature Publishing Group), is registered in re3data.org, biosharing.org, OpenAIRE and DataCite. This

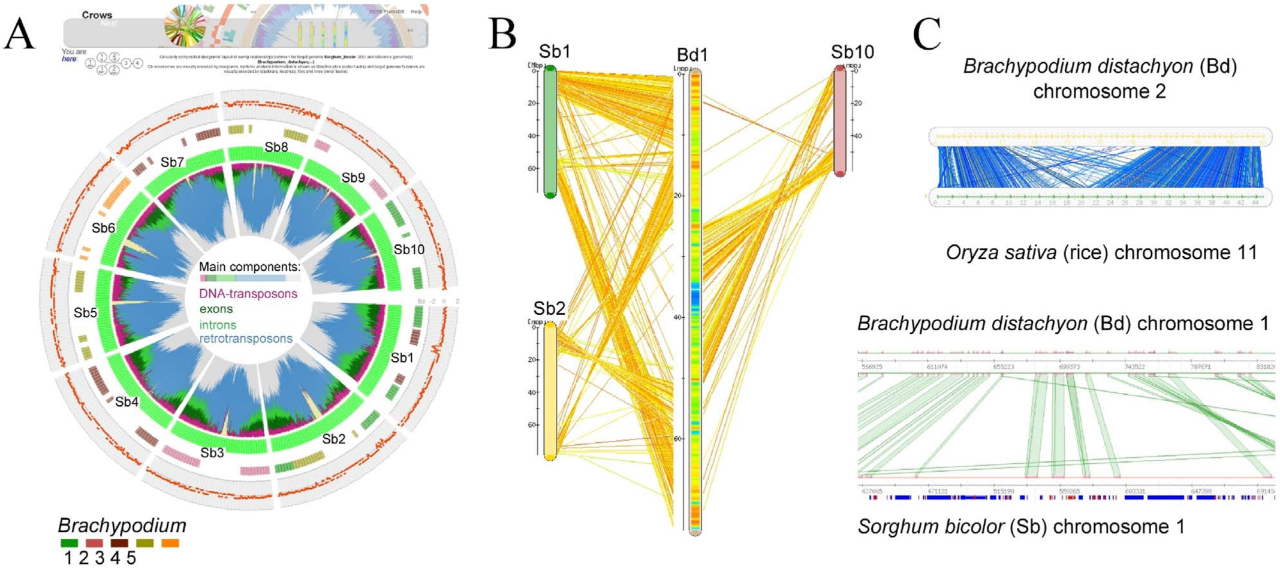


Fig. 2. CrowsNest visualization tool for plant genomic data.

Screenshots of the CrowsNest tool as part of PGSB PlantsDB to visualize synteny (conserved gene order) between plant genomes. The syntenic relationships between the species can be browsed in a hierarchical way from macro-synteny (A) over chromosomes (B) down to micro-synteny views (C). A: Circular display of synteny between *Sorghum bicolor* (Sb) as reference and *Brachypodium distachyon* (Bd) as target genome. The inner circle shows the distribution of the main genomic components, the syntenic regions to the five *Brachypodium* chromosomes are color coded, the outer circle displays the amount of syntenic conservation. B: *Brachypodium* chromosome 1 as reference with a heatmap for gene density and connections to homologs on three *Sorghum* chromosomes. C: upper panel: synteny between *Brachypodium* chromosome 2 and rice chromosome 11; lower panel: free adjustable zoomed in view between *Brachypodium* and *Sorghum* with the transposon annotation on Sb1 (blue retrotransposons, red DNA transposons). The larger genome size of *Sorghum* (734 Mb vs 355 Mb) is caused by an expansion of the transposon space between genes.

demonstrates the capabilities of e!DAL to describe individually configured research data. A brief overview of data access statistics is given in Fig. 1.

In the publication of Wilkinson et al. the authors emphasize the necessity that scientific data requires an infrastructure to support its reusability (Wilkinson et al., 2016). Furthermore, they propose guidelines to achieve this, formulated as the FAIR principles (Findability, Accessibility, Interoperability, and Reusability). These principles focus on enhancing the ability for machines to automatically find and incorporate data to increase reuse by scientists. The PGP repository supports this effort by providing a persistent data publication infrastructure. One example of a FAIR Principles aware data publication maintained in the PGP repository is the comprehensive datasets for a high-throughput plant phenotypic experiment (Arend et al., 2016b). The accompanying research article and its metadata follow the minimal information standard for plant phenotyping experiments MIAPPE, and fulfil the MIAPPE standard for a minimal information of plant phenotyping experiments (Ćwiek-Kupczyńska et al., 2016; Krajewski et al., 2015) and are formatted in ISA-Tab format (Sansone et al., 2012). This dataset is used as a first reference dataset for MIAPPE compliant metadata (http://www.miappe.org/data/) and benchmarks image analysis algorithms for biomass estimations.

3.5. IPK blast server

The IPK Blast Server is a simple and easy-to-use application to perform sequence comparisons against plant genomic resources. It is based on the stand-alone BLAST web interface ViroBLAST (Deng et al., 2007). The individual crops presented through the IPK Blast Server are currently barley (http://webblast.ipk-gatersleben.de/barley/) and rye (http://webblast.ipk-gatersleben.de/ryeselect/). This BLAST service provides users a versatile tool to perform sequence homology searches and to access the genomic resources of barley and rye in a feasible manner.

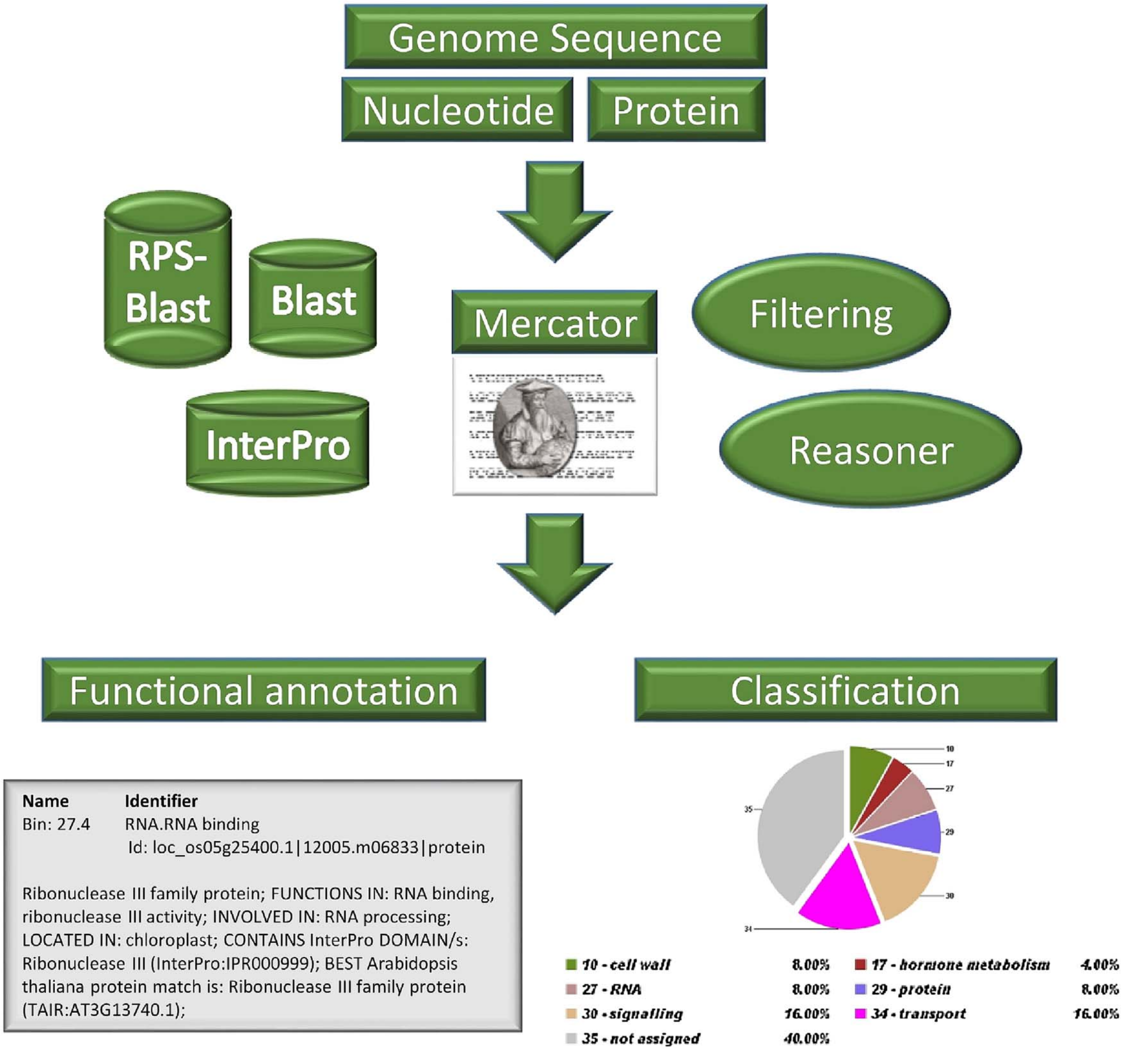
For barley the complete sequence data set comprises the whole genome shotgun assemblies of the cultivars Morex, Barke and Bowman as well as the high and low confidence gene sets (Mayer et al., 2012). In addition, POPSEQ anchoring data (Mascher et al., 2013) and barley exome capture targets are also integrated (Russell et al., 2016). The comprehensive resource is augmented with further sequence resources that include bacterial artificial chromosome (BAC) assemblies, BAC end sequences (BES) and full length cDNA sequences. Most of the integrated barley resources are also linked to the barley genome explorer BARLEX (Colmsee et al., 2015) allowing the user to investigate a matching sequence for further details.

The rye genomic dataset uses the whole genome shotgun assembly of inbred line Lo7 (Bauer et al., 2017). This resource is complemented with gene models' CDS and predicted protein sequences from the reference sequence assembly. The rye dataset also includes a transcriptome resource that was constructed from five diverse inbred lines (Haseneyer et al., 2011).

The IPK Blast Server is a central tool for crop scientists, especially for the *Triticeae* research community, and allows for an easy sequence searches against our plant specific resource, and to easily explore and navigate the resulting BLAST search output. The services is widely used e.g. for identification and study of the basic leucine zipper family of transcription factors in barley (Okay, 2016) or the study of a fungal pathogens in rye (Oeser et al., 2017).

3.6. PlantsDB

The PGSB PlantsDB system has been designed as an information resource for plant genome data providing flexible and intuitive access to complex data structures (Spannagl et al., 2016c) (http://pgsb.helmholtz-muenchen.de/plant/genomes.jsp). It was started in 2000 with *Arabidopsis,* the first plant genome project, and has since been maintained and further developed by the PGSB group (Plant Genome and Systems Biology) of Klaus Mayer at the Helmholtz Center Munich. The aim of PlantsDB is to structure and communicate plant genomic data and to assist comparative analyses of both model and crop plant genomes. The data resources provide a backbone against which experimental data, e.g., from high-throughput functional genomics, can be organized and analyzed. Biologically relevant subsets are overlaid and interconnected with PlantsDB. DroughtDB, for example, is a



Name	**Identifier**
Bin: 27.4	RNA.RNA binding Id: loc_os05g25400.1|12005.m06833|protein

Ribonuclease III family protein; FUNCTIONS IN: RNA binding, ribonuclease III activity; INVOLVED IN: RNA processing; LOCATED IN: chloroplast; CONTAINS InterPro DOMAIN/s: Ribonuclease III (InterPro:IPR000999); BEST Arabidopsis thaliana protein match is: Ribonuclease III family protein (TAIR:AT3G13740.1);

Fig. 3. The online annotation tool 'Mercator' (available on plabipd.de).
Overview of the Mercator pipeline which involves the submission of protein or DNA sequence for annotation and classification into MapMan bins. The pipeline includes sequence searches against reference databases followed by filtering and rule-based reasoning to generate the resulting output.

manually curated database of experimentally validated drought stress genes. The corresponding orthologous genes from whole-genome sequencing projects (Alter et al., 2015) are linked and aggregated. The PGSB database frontend is complemented by specialized tools such as the CrowsNest viewer to explore conserved gene order (Fig. 2) (Spannagl et al., 2017), an RNASeq ExpressionBrowser to facilitate data interpretation for genes of interest (Nussbaumer et al., 2014a) and chromoWIZ to depict the chromosomal locations for user defined gene sets (Nussbaumer et al., 2014b). PGSB also curates a database for plant transposable elements (REdat). REdat is connected to a hierarchical repeat classification scheme (REcat) and continuously updated with *de novo* detected LTR-retrotransposons from new genome projects (Spannagl et al., 2017). In plants, between 20% to 85% of the genome sequence is composed of transposons and their deteriorated remnants (Tenaillon et al., 2010). The ongoing insertions of transposons near or within genes can lead to phenotypic variation (Wei and Cao, 2016). Detailed studies of these transposition events requires a high-quality genome annotation for transposon sequences; PGSB provides curated and taxon-specific plant repeat libraries for this purpose.

PlantsDB currently hosts 18 plant species. Amongst these are plant model organisms such as *Arabidopsis thaliana, Brachypodium distachyon* (a grass reference) and *Medicago truncatula* (a nitrogen fixer), as well as major crop plants such as rice, maize, sorghum and tomato. Recent PlantsDB additions include members of the cereal clade *Triticeae*: barley (5Gbp, diploid), rye (8Gbp, diploid) and the hexaploid bread wheat (17 Gbp). Despite their agronomic importance the *Triticeae* genomes remained largely uncharacterized until recently (Bauer et al., 2017; Mascher et al., 2017; Mayer et al., 2014) due to their large size, their high transposon derived repeat content (> 80%) and their complex genetics. With the evolution of DNA sequencing technologies and the development of genome assembly strategies it has become increasingly feasible to obtain meaningful genomic sequence data from these large genomes. The improvements over the last 10 years towards those high quality references characterized by large contig N50 have been step wise from starting with individual sequence reads from the Roche 454 platform, followed by contig and scaffold assemblies up to the current chromosomal pseudomolecules (Phillippy, 2017). For the *Triticeae*, PGSB participated in this process as an early bioinformatics partner of the international barley (Mayer et al., 2012) and wheat (Mayer et al., 2014) genome projects. To make the most of the available data resources, PGSB developed new bioinformatics approaches such as the GenomeZipper concept, reference-guided exome assemblies and "chromosome genomics" based on flow cytometry sorted chromosomes (Spannagl et al., 2016b). The GenomeZipper provides a virtual synteny based gene ordering for sequence assemblies that lack chromosomal level detail. This has been a valuable resource for the identification of phenotype candidate loci in barley, wheat and several other grass genomes. To accommodate the new datatypes and results from these complex crop-plant genomes, a PlantsDB instance has been developed specifically for the *Triticeae* (Spannagl et al., 2017).

A second database independent service of PGSB is the detection and characterization of the gene and the transposon complements in newly assembled plant genomes. Over the last 10 years PGSB has generated annotation data for about 40 plant species/varieties such as *Physcomitrella*, *Micromonas*, duckweed, cotton, tomato and many grass species. The ongoing demand for such annotation services shows that structural and functional genome annotation is still not a trivial exercise for plant genomes. The de.NBI project allows us to improve our in-house pipelines and integrate them into an easy usable workflow system. It is planned to provide our compute-intensive gene detection pipeline through the de.NBI cloud as a "do it yourself" solution for draft plant genome annotations.

3.7. PlabiPD

PlabiPD is a portal which focuses on the integration and visualization of genomic data derived from crop species. PlabiPD currently hosts data from seven crop species including the model species, *Arabidopsis thaliana*. Protein coding genes from these species have been clustered using the MCL algorithm and visualizations provided to allow users to browse the resulting sequence group data. The provide web interface allows users to submit keywords or sequences (nucleotide or protein) to search the data for clusters. Consensus subcellular localisation data from a number of published tools is presented along with conserved structural domains and 'function context data' based on the MapMan classification system (Jaiswal and Usadel, 2016). PlabiPD also records published plant genomes and provides a visual timeline of plant genome publications and a visual taxonomic summary of the sequenced plant genome cladistics (split into monocotyledons and dicotyledons). This allows users to quickly view which clades contains plants which have already been sequenced.

PlabiPD also hosts the Mercator online tool for gene functional annotation (Lohse et al., 2014). This web-based tool enables the fast automatic annotation and classification of FASTA sequences into MapMan classification bins (Fig. 3). The underlying Mercator engine performs a search against the manually annotated protein clusters curated in PlabiPD. As such it eventually draws upon all sequenced plant genomes and computes the most likely MapMan classification bin. Each MapMan bin has been manually curated with published reference data. Users can submit either nucleotide or protein sequences and are provided with a number of parameters to optimize the results. The output consists of a mapping file and graphical overview charts. The mapping file generated is also a key element for the visualization of plant expression data on the MapMan standalone application. Recently this has been used to identify differences and commonalities in plant stress responses (Zhang and Sonnewald, 2017) as well as to start unravelling complex traits in cassava (Zou et al., 2017).

4. Discussion

Advances in biotechnology have large implications on plant research. Targeted genome editing or the shift from traditional breeding to new strategies such as genomic selection are some examples where the knowledge of genome context from plant sequencing projects results in direct applications (Morrell et al., 2011). Computational expertise is required to support the plant community with solutions on how to cope with the increasing amount of digital biological data.

GCBN strives to evolve as a central resource for plant specific questions by connecting the complementary expertise from its three partners. Establishing a bioinformatics infrastructure that provides concise training and services for the plant genetics community is our long-term vision. The next step is the implementation of APIs (application programming interfaces) which will enable a programmatic access to GCBN tools and plant data. The broad collection of services and tools captures a wide range of applications. All services are registered in the ELIXIR Tools and Data Services Registry (https://bio.tools/) and have a build-in user-feedback interface to collect the subjective assessments of the user, which we use to monitor and improve our services.

Scientific discoveries need to be reproducible and this is especially difficult for results derived from complex data analysis. To enable users to repeat such analysis, comprehensive training is required (Khan et al., 2014). Therefore, the demand is high for bioinformatics trainings on data analysis to achieve better scientific outcomes that are more accurate and reproducible. This trainings should be offered to scientists from graduate students over early-carrier postdocs to senior scientists (Baker, 2016). In GCBN, our goal is to provide plant-specific guidance which will enable users to achieve competent data analysis and robust scientific results. Our tailored training for plant researchers caters to all levels of expertise with fundamental to advance analysis techniques. They are designed as a mixture of lectures and hands-on trainings to mediate knowledge in an application-oriented manner. Upcoming training events are listed on the de.NBI website (https://www.denbi.de/training). The mission of de.NBI to establish a national bioinformatics infrastructure for life science is further embraced on international level. Founded in 2014, ELIXIR is an intergovernmental organization that brings together life science resources from across Europe to build a sustainable European infrastructure for biological research (Crosswell and Thornton, 2012). In 2016, de.NBI became the German ELIXIR node (https://www.elixir-europe.org/) and GCBN is tasked to mediate the plant bioinformatics expertize of this network.

5. Conclusions

Due to the inherent complexity and size of many plant genomes, the required analytic and bioinformatics solutions can be best handled through interdisciplinary collaborations between different research institutions from multiple research domains. With the mission to form a network between different life science fields, de.NBI is building hubs that can multiply their expertise throughout the network. These interdisciplinary collaborations provide great advantage that is beneficial for the different partners to solve grant problems as well as for the user community to learn in a cross-functional environment. The established synergies within GCBN, and with other partners, show that GCBN is well prepared to handle the current and upcoming challenges with crop-plant research data.

It is our mission to continuously look into user requirements and to accommodate these into existing services. In addition, it is our ambition to also provide new services for technological innovations relevant to applied plant science. With the GCBN service center, we provide tailored solutions and training for established, advanced, and evolving technologies.

Funding

This work was supported by different funding bodies. Uwe Scholz, Klaus F.X. Mayer and Björn Usadel acknowledge support from the German Ministry of Education and Research (BMBF) for grant 031A536 "de.NBI". Uwe Scholz and Klaus F.X. Mayer acknowledges support from BMBF for grant 0314000 "BARLEX", 0315954 "TRITEX" and the European Commission 7th Framework Infrastructure Programme 283496. Björn Usadel acknowledges support from BMBF for grant 0315961 "PPD".

References

Ćwiek-Kupczyńska, H., Altmann, T., Arend, D., Arnaud, E., Chen, D., Cornut, G., Fiorani, F., Frohmberg, W., Junker, A., Klukas, C., Lange, M., Mazurek, C., Nafissi, A., Neveu, P., van Oeveren, J., Pommier, C., Poorter, H., Rocca-Serra, P., Sansone, S.-A., Scholz, U., van Schriek, M., Seren, Ü., Usadel, B., Weise, S., Kersey, P., Krajewski, P., 2016.

Measures for interoperability of phenotypic data: minimum information requirements and formatting. Plant Methods 12, 44. http://dx.doi.org/10.1186/s13007-016-0144-4.

Alter, S., Bader, K.C., Spannagl, M., Wang, Y., Bauer, E., Schön, C.C., Mayer, K.F.X., 2015. DroughtDB: an expert-curated compilation of plant drought stress genes and their homologs in nine species. Database 2015, 1–7. http://dx.doi.org/10.1093/database/bav046.

Arend, D., Lange, M., Chen, J., Colmsee, C., Flemming, S., Hecht, D., Scholz, U., 2014. E!DAL—a framework to store, share and publish research data. BMC Bioinf. 15, 214. http://dx.doi.org/10.1186/1471-2105-15-214.

Arend, D., Junker, A., Scholz, U., Schüler, D., Wylie, J., Lange, M., 2016a. PGP repository: a plant phenomics and genomics data publication infrastructure. Database 2016, baw033. http://dx.doi.org/10.1093/database/baw033.

Arend, D., Lange, M., Pape, J.-M., Weigelt-Fischer, K., Arana-Ceballos, F., Mücke, I., Klukas, C., Altmann, T., Scholz, U., Junker, A., 2016b. Quantitative monitoring of Arabidopsis thaliana growth and development using high-throughput plant phenotyping. Sci. Data 3, 160055. http://dx.doi.org/10.1038/sdata.2016.55.

Awika, J.M., 2011. Major cereal grains production and use around the world. Advances in Cereal Science: Implications to Food Processing and Health Promotion, ACS Symposium Series. American Chemical Societypp. 1–13. http://dx.doi.org/10.1021/bk-2011-1089.ch001.

Baker, M., 2016. Seek out stronger science. Nature 537, 703–704. http://dx.doi.org/10.1038/nj7622-703a.

Bauer, E., Schmutzer, T., Barilar, I., Mascher, M., Gundlach, H., Martis, M.M., Twardziok, S.O., Hackauf, B., Gordillo, A., Wilde, P., Schmidt, M., Korzun, V., Mayer, K.F.X., Schmid, K., Schön, C.-C., Scholz, U., 2017. Towards a whole-genome sequence for rye (*Secale cereale* L.). Plant J. 89, 853–869. http://dx.doi.org/10.1111/tpj.13436.

Bolger, M.E., Weisshaar, B., Scholz, U., Stein, N., Usadel, B., Mayer, K.F.X., 2014. Plant genome sequencing – applications for crop improvement. Curr. Opin. Biotechnol. 26, 31–37. http://dx.doi.org/10.1016/j.copbio.2013.08.019.

Bolger, M., Schwacke, R., Gundlach, H., Schmutzer, T., Chen, J., Arend, D., Oppermann, M., Weise, S., Lange, M., Fiorani, F., Spannagl, M., Scholz, U., Mayer, K., Usadel, B., 2017. From plant genomes to phenotypes. J. Biotechnol. 0–1. http://dx.doi.org/10.1016/j.jbiotec.2017.06.003.

Cobb, J.N., DeClerck, G., Greenberg, A., Clark, R., McCouch, S., 2013. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor. Appl. Genet. 126, 867–887. http://dx.doi.org/10.1007/s00122-013-2066-0.

Colmsee, C., Mascher, M., Czauderna, T., Hartmann, A., Schlüter, U., Zellerhoff, N., Schmitz, J., Bräutigam, A., Pick, T.R., Alter, P., Gahrtz, M., Witt, S., Fernie, A.R., Börnke, F., Fahnenstich, H., Bucher, M., Dresselhaus, T., Weber, A.P., Schreiber, F., Scholz, U., Sonnewald, U., 2012. OPTIMAS-DW: a comprehensive transcriptomics, metabolomics, ionomics, proteomics and phenomics data resource for maize. BMC Plant Biol. 12, 245 (10.1186/1471-2229-12-245).

Colmsee, C., Beier, S., Himmelbach, A., Schmutzer, T., Stein, N., Scholz, U., Mascher, M., 2015. BARLEX – the Barley draft genome explorer. Mol. Plant. 1–3. http://dx.doi.org/10.1016/j.molp.2015.03.009.

Crosswell, L.C., Thornton, J.M., 2012. ELIXIR: a distributed infrastructure for European biological data. Trends Biotechnol. 30, 241–242. http://dx.doi.org/10.1016/j.tibtech.2012.02.002.

Deng, W., Nickle, D.C., Learn, G.H., Maust, B., Mullins, J.I., 2007. ViroBLAST: a standalone BLAST web server for flexible queries of multiple databases and user's datasets. Bioinformatics 23, 2334–2336. http://dx.doi.org/10.1093/bioinformatics/btm331.

Esch, M., Chen, J., Colmsee, C., Klapperstück, M., Grafahrend-Belau, E., Scholz, U., Lange, M., 2015. LAILAPS: the plant science search engine. Plant Cell Physiol. 56, e8. http://dx.doi.org/10.1093/pcp/pcu185.

Haseneyer, G., Schmutzer, T., Seidel, M., Zhou, R., Mascher, M., Schön, C.-C., Taudien, S., Scholz, U., Stein, N., Mayer, K.F., Bauer, E., 2011. From RNA-seq to large-scale genotyping – genomics resources for rye (Secale cereale L.). BMC Plant Biol. 11, 131. http://dx.doi.org/10.1186/1471-2229-11-131.

Jaiswal, P., Usadel, B., 2016. Plant pathway databases. In: Edwards, D. (Ed.), Plant Bioinformatics: Methods and Protocols. Springer, New York, NY, pp. 71–87. http://dx.doi.org/10.1007/978-1-4939-3167-5_4.

Khan, I.A., Fraser, A., Bray, M.A., Smith, P.J., White, N.S., Carpenter, A.E., Errington, R.J., 2014. ProtocolNavigator: emulation-based software for the design, documentation and reproduction biological experiments. Bioinformatics 30, 3440–3442. http://dx.doi.org/10.1093/bioinformatics/btu554.

Krajewski, P., Chen, D., Ćwiek, H., Van Dijk, A.D.J., Fiorani, F., Kersey, P., Klukas, C., Lange, M., Markiewicz, A., Nap, J.P., Van Oeveren, J., Pommier, C., Scholz, U., Van Schriek, M., Usadel, B., Weise, S., 2015. Towards recommendations for metadata and data handling in plant phenotyping. J. Exp. Bot. 66, 5417–5427. http://dx.doi.org/10.1093/jxb/erv271.

Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., Tohge, T., Fernie, A.R., Stitt, M., Usadel, B., 2014. Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. Plant Cell Environ. 37, 1250–1258. http://dx.doi.org/10.1111/pce.12231.

Mascher, M., Muehlbauer, G.J., Rokhsar, D.S., Chapman, J., Schmutz, J., Barry, K., Muñoz-Amatriaín, M., Close, T.J., Wise, R.P., Schulman, A.H., Himmelbach, A., Mayer, K.F.X., Scholz, U., Poland, J.A., Stein, N., Waugh, R., 2013. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). Plant J. 76, 718–727. http://dx.doi.org/10.1111/tpj.12319.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P.E., Russell, J., Bayer, M., Ramsay, L., Liu, H., Haberer, G., Zhang, Q., Zhang, Q., Barrero, R.A., Li, L., Taudien, S., Groth, M., Felder, M., Hastie, A., Sta, H., Vr, J., Chan, S., Ounit, R., Wanamaker, S., Bolser, D., Colmsee, C., Schmutzer, T., Aliyeva-Schnorr, L., Grasso, S., Tanskanen, J., Chailyan, A., Sampath, D., Heavens, D., Clissold, L., Cao, S., Chapman, B., Dai, F., Han, Y., Li, H., Li, X., Lin, C., Mccooke, J.K., Tan, C., Wang, P., Wang, S., Yin, S., Zhou, G., Poland, J.A., Bellgard, M.I., Borisjuk, L., Houben, A., Dole, J., Ayling, S., Lonardi, S., Kersey, P., Langridge, P., Muehlbauer, G.J., Clark, M.D., Caccamo, M., Schulman, A.H., Mayer, K.F.X., Platzer, M., Close, T.J., Scholz, U., Hansson, M., Zhang, G., Braumann, I., Spannagl, M., Li, C., Waugh, R., Stein, N., 2017. A chromosome conformation capture ordered sequence of the barley genome. Nature 544, 1–43. http://dx.doi.org/10.1038/nature22043.

Mayer, K.F.X., Waugh, R., Langridge, P., Close, T.J., Wise, R.P., Graner, A., Matsumoto, T., Sato, K., Schulman, A., Muehlbauer, G.J., Stein, N., Ariyadasa, R., Schulte, D., Poursarebani, N., Zhou, R., Steuernagel, B., Mascher, M., Scholz, U., Shi, B., Langridge, P., Madishetty, K., Svensson, J.T., Bhat, P., Moscou, M., Resnik, J., Close, T.J., Muehlbauer, G.J., Hedley, P., Liu, H., Morris, J., Waugh, R., Frenkel, Z., Korol, A., Bergès, H., Graner, A., Stein, N., Steuernagel, B., Scholz, U., Taudien, S., Felder, M., Groth, M., Platzer, M., Stein, N., Steuernagel, B., Scholz, U., Himmelbach, A., Taudien, S., Felder, M., Platzer, M., Lonardi, S., Duma, D., Alpert, M., Cordero, F., Beccuti, M., Ciardo, G., Ma, Y., Wanamaker, S., Close, T.J., Stein, N., Cattonaro, F., Vendramin, V., Scalabrin, S., Radovic, S., Wing, R., Schulte, D., Steuernagel, B., Morgante, M., Stein, N., Waugh, R., Nussbaumer, T., Gundlach, H., Martis, M., Ariyadasa, R., Poursarebani, N., Steuernagel, B., Scholz, U., Wise, R.P., Poland, J., Stein, N., Mayer, K.F.X., Spannagl, M., Pfeifer, M., Gundlach, H., Mayer, K.F.X., Gundlach, H., Moisy, C., Tanskanen, J., Scalabrin, S., Zuccolo, A., Vendramin, V., Morgante, M., Mayer, K.F.X., Schulman, A., Pfeifer, M., Spannagl, M., Hedley, P., Morris, J., Russell, J., Druka, A., Marshall, D., et al., 2012. A physical, genetic and functional sequence assembly of the barley genome. Nature 491, 711–716. http://dx.doi.org/10.1038/nature11543.

Mayer, K.F.X., Rogers, J., Doleel, J., Pozniak, C., Eversole, K., Feuillet, C., Gill, B., Friebe, B., Lukaszewski, A.J., Sourdille, P., Endo, T.R., Kubalakova, M., Ihalikova, J., Dubska, Z., Vrana, J., Perkova, R., Imkova, H., Febrer, M., Clissold, L., McLay, K., Singh, K., Chhuneja, P., Singh, N.K., Khurana, J., Akhunov, E., Choulet, F., Alberti, A., Barbe, V., Wincker, P., Kanamori, H., Kobayashi, F., Itoh, T., Matsumoto, T., Sakai, H., Tanaka, T., Wu, J., Ogihara, Y., Handa, H., Maclachlan, P.R., Sharpe, A., Klassen, D., Edwards, D., Batley, J., Olsen, O.-A., Sandve, S.R., Lien, S., Steuernagel, B., Wulff, B., Caccamo, M., Ayling, S., Ramirez-Gonzalez, R.H., Clavijo, B.J., Wright, J., Pfeifer, M., Spannagl, M., Martis, M.M., Mascher, M., Chapman, J., Poland, J.A., Scholz, U., Barry, K., Waugh, R., Rokhsar, D.S., Muehlbauer, G.J., Stein, N., Gundlach, H., Zytnicki, M., Jamilloux, V., Quesneville, H., Wicker, T., Faccioli, P., Colaiacovo, M., Stanca, A.M., Budak, H., Cattivelli, L., Glover, N., Pingault, L., Paux, E., Sharma, S., Appels, R., Bellgard, M., Chapman, B., Nussbaumer, T., Bader, K.C., Rimbert, H., Wang, S., Knox, R., Kilian, A., Alaux, M., Alfama, F., Couderc, L., Guilhot, N., Viseux, C., Loaec, M., Keller, B., Praud, S., 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science (80-.) 345, 1251788. http://dx.doi.org/10.1126/science.1251788.

Michael, T.P., Jackson, S., 2013. The first 50 plant genomes. Plant Genome 6. http://dx.doi.org/10.3835/plantgenome2013.03.0001in.

Morrell, P.L., Buckler, E.S., Ross-Ibarra, J., 2011. Crop genomics: advances and applications. Nat. Rev. Genet. 13, 85–96. http://dx.doi.org/10.1038/nrg3097.

Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D.J., Salichos, L., Zhang, J., Weinstock, G.M., Isaacs, F., Rozowsky, J., Gerstein, M., 2016. The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biol. 17, 53 (10.1186/s13059-016-0917-0).

Nussbaumer, T., Kugler, K.G., Bader, K.C., Sharma, S., Seidel, M., Mayer, K.F.X., 2014a. RNASeqExpressionBrowser-a web interface to browse and visualize high-throughput expression data. Bioinformatics 30, 2519–2520. http://dx.doi.org/10.1093/bioinformatics/btu334.

Nussbaumer, T., Kugler, K.G., Schweiger, W., Bader, K.C., Gundlach, H., Spannagl, M., Poursarebani, N., Pfeifer, M., Mayer, K.F.X., 2014b. chromoWIZ: a web tool to query and visualize chromosome-anchored genes from cereal and model genomes. BMC Plant Biol. 14, 348 (10.1186/s12870-014-0348-6).

Oeser, B., Kind, S., Schurack, S., Schmutzer, T., Tudzynski, P., Hinsch, J., 2017. Cross-talk of the biotrophic pathogen Claviceps purpurea and its host Secale cereale. BMC Genomics 18, 273. http://dx.doi.org/10.1186/s12864-017-3619-4.

Okay, S., 2016. Identification of gene families using genomics and/or transcriptomics data. In: Hakeem, K.R., Tombulo\uglu, H., Tombulo\uglu, G. (Eds.), Plant Omics: Trends and Applications. Springer International Publishing, Cham, pp. 217–254. http://dx.doi.org/10.1007/978-3-319-31703-8_9.

Oppermann, M., Weise, S., Dittmann, C., Knüpffer, H., 2015. GBIS: the information system of the German genebank. Database 2015, bav021. http://dx.doi.org/10.1093/database/bav021.

Pühler, A., 2016. German network for bioinformatics infrastructure – de.NBI.systembiologie.de 10, 8–13.

Phillippy, A.M., 2017. New advances in sequence assembly. Genome Res. 27, xi–xiii. http://dx.doi.org/10.1101/gr.223057.117.

Russell, J., Mascher, M., Dawson, I.K., Kyriakidis, S., Calixto, C., Freund, F., Bayer, M., Milne, I., Marshall-Griffiths, T., Heinen, S., Hofstad, A., Sharma, R., Himmelbach, A., Knauft, M., van Zonneveld, M., Brown, J.W.S., Schmid, K., Kilian, B., Muehlbauer, G.J., Stein, N., Waugh, R., 2016. Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. Nat. Genet. 1–10. http://dx.doi.org/10.1038/ng.3612.

Sansone, S.-A., Rocca-Serra, P., Field, D., Maguire, E., Taylor, C., Hofmann, O., Fang, H., Neumann, S., Tong, W., Amaral-Zettler, L., Begley, K., Booth, T., Bougueleret, L., Burns, G., Chapman, B., Clark, T., Coleman, L.-A., Copeland, J., Das, S., de Daruvar, A., de Matos, P., Dix, I., Edmunds, S., Evelo, C.T., Forster, M.J., Gaudet, P., Gilbert, J., Goble, C., Griffin, J.L., Jacob, D., Kleinjans, J., Harland, L., Haug, K., Hermjakob, H., Sui, S.J.H., Laederach, A., Liang, S., Marshall, S., McGrath, A., Merrill, E., Reilly, D., Roux, M., Shamu, C.E., Shang, C.A., Steinbeck, C., Trefethen, A., Williams-Jones, B., Wolstencroft, K., Xenarios, I., Hide, W., 2012. Toward interoperable bioscience data. Nat. Genet. 44, 121–126. http://dx.doi.org/10.1038/ng.1054.

Spannagl, M., Alaux, M., Lange, M., Bolser, D.M., Bader, K.C., Letellier, T., Kimmel, E., Flores, R., Pommier, C., Kerhornou, A., Walts, B., Nussbaumer, T., Grabmuller, C., Chen, J., Colmsee, C., Beier, S., Mascher, M., Schmutzer, T., Arend, D., Thanki, A., Ramirez-Gonzalez, R., Ayling, M., Ayling, S., Caccamo, M., Mayer, K.F.X., Scholz, U., Steinbach, D., Quesneville, H., Kersey, P.J., 2016a. transPLANT resources for triticeae genomic data. Plant Genome 9, 1–13. http://dx.doi.org/10.3835/plantgenome2015.06.0038.

Spannagl, M., Bader, K., Pfeifer, M., Nussbaumer, T., Mayer, K.F.X., 2016b. PGSB/MIPS plant genome information resources and concepts for the analysis of complex grass genomes. In: Edwards, D. (Ed.), Plant Bioinformatics: Methods and Protocols. Springer, New York, NY, pp. 165–186. http://dx.doi.org/10.1007/978-1-4939-3167-5_8.

Spannagl, M., Nussbaumer, T., Bader, K.C., Martis, M.M., Seidel, M., Kugler, K.G., Gundlach, H., Mayer, K.F.X., 2016c. PGSB plantsDB: updates to the database framework for comparative plant genome research. Nucleic Acids Res. 44, D1141–D1147. http://dx.doi.org/10.1093/nar/gkv1130.

Spannagl, M., Nussbaumer, T., Bader, K., Gundlach, H., Mayer, K.F.X., 2017. PGSB/MIPS plantsDB database framework for the integration and analysis of plant genome data. In: van Dijk, A.D.J. (Ed.), Plant Genomics Databases: Methods and Protocols. Springer, New York, NY, pp. 33–44. http://dx.doi.org/10.1007/978-1-4939-6658-5_2.

Tenaillon, M.I., Hollister, J.D., Gaut, B.S., 2010. A triptych of the evolution of plant transposable elements. Trends Plant Sci. 15, 471–478. http://dx.doi.org/10.1016/j.tplants.2010.05.003.

Wei, L., Cao, X., 2016. The effect of transposable elements on phenotypic variation: insights from plants to humans. Sci. China Life Sci. 59, 24–37. http://dx.doi.org/10.1007/s11427-015-4993-2.

Weise, S., Oppermann, M., Maggioni, L., van Hintum, T., Knüpffer, H., 2017. EURISCO: the European search catalogue for plant genetic resources. Nucleic Acids Res. 45, D1003–D1008. http://dx.doi.org/10.1093/nar/gkw755.

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen't, P.A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3http://dx.doi.org/10.1038/sdata.2016.18. (160018).

Zhang, H., Sonnewald, U., 2017. Differences and commonalities of plant responses to single and combined stresses. Plant J. 839–855. http://dx.doi.org/10.1111/tpj.13557.

Zou, M., Lu, C., Zhang, S., Chen, Q., Sun, X., Ma, P., Hu, M., Peng, M., Ma, Z., Chen, X., Zhou, X., Wang, H., Feng, S., Fang, K., Xie, H., Li, Z., Liu, K., Qin, Q., Pei, J., Wang, S., Pan, K., Hu, W., Feng, B., Fan, D., Zhou, B., Wu, C., Su, M., Xia, Z., Li, K., Wang, W., 2017. Epigenetic map and genetic map basis of complex traits in cassava population. Sci. Rep. 7http://dx.doi.org/10.1038/srep41232. (41232).