

Activating *ex situ* genebank collections of wheat and barley for  
breeding and research

**Dissertation**  
**zur Erlangung des**  
**Doktorgrades der Agrarwissenschaften (Dr. agr.)**

Der

Naturwissenschaftlichen Fakultät III  
Agrar- und Ernährungswissenschaften,  
Geowissenschaften und Informatik

der Martin-Luther-Universität Halle-Wittenberg



vorgelegt von

**Herrn Norman Philipp**

geb. am 07.08.1991 in Schleiz

Gutachter:

**Prof. Dr. Jochen C. Reif**  
**Prof. Dr. Tobias Würschum**

**Eingereicht am 20.02.2020**  
**Verteidigung am 26.10.2020**  
Halle (Saale)

*"Civilization as it is known today could not have evolved, nor can it survive, without an adequate food supply."*

Dr. Norman E. Borlaug

## Table of Contents

<b>1.</b>	<b>General introduction.....</b>	<b>5</b>
1.1.	Wheat and barley are key crops to feed a growing world population .....	5
1.2.	Need to exploit genetic resources to ensure a sustainable future food security.....	5
1.3.	Demands to make genebank collections accessible for breeding and research .....	6
1.4.	Wheat and barley are among the most prevalent accessions in genebanks .....	7
1.5.	Insufficient characterization of accessions or a limited access to this information hampers the exploitation of entire genebank collections.....	8
1.6.	Considerable interest for seed samples at the German Federal <i>ex situ</i> Genebank .....	9
1.7.	Historical data from seed regeneration can be used for the phenotypic characterization of genebank collections.....	10
1.8.	The FAIR principles as guideline for scientific data management .....	12
1.9.	Objectives.....	12
<b>2.</b>	<b>Peer-reviewed scientific articles.....</b>	<b>14</b>
2.1.	Grain number and grain yield distribution along the spike remain stable despite breeding for high yield in winter wheat .....	14
2.2.	Leveraging the use of historical data gathered during seed regeneration of an <i>ex situ</i> genebank collection of wheat .....	15
2.3.	Unlocking historic phenotypic data from an <i>ex situ</i> collection to enhance the informed utilization of genetic resources of barley ( <i>Hordeum sp.</i> ) .....	16
2.4.	Unbalanced historical phenotypic data from seed regeneration of a barley <i>ex situ</i> collection .....	17
2.5.	Historical phenotypic data from seven decades of seed regeneration in a wheat <i>ex situ</i> collection .....	18
<b>3.</b>	<b>General discussion .....</b>	<b>19</b>
3.1.	Exploiting wheat genetic resources for grain yield improvement .....	20
3.1.1.	Grain yield gap between wheat genetic resources and elite varieties .....	20
3.1.2.	Investigating grain yield components for useful diversity.....	21
3.2.	Leveraging the use of historical phenotypic data.....	22
3.2.1.	Finding the right key to unlock the value behind historical data .....	22
3.2.2.	From non-orthogonal historical observations towards high quality phenotypic characterization of genebank collections.....	22
3.2.3.	Rigorous quality assessment is essential for the analysis of historical data.....	23
3.3.	From conservation-driven genebanks towards bio-digital resource centers.....	25
<b>4.</b>	<b>Final outlook.....</b>	<b>28</b>
<b>5.</b>	<b>Summary .....</b>	<b>29</b>
<b>6.</b>	<b>Zusammenfassung .....</b>	<b>31</b>
<b>7.</b>	<b>General references.....</b>	<b>33</b>

---

<b>8.</b>	<b>List of general abbreviations .....</b>	<b>37</b>
<b>9.</b>	<b>Acknowledgements .....</b>	<b>38</b>
<b>10.</b>	<b><i>Curriculum vitae</i> .....</b>	<b>39</b>
<b>11.</b>	<b>List of publications .....</b>	<b>40</b>
<b>12.</b>	<b>Eidesstattliche Erklärung / Declaration under Oath.....</b>	<b>41</b>
<b>13.</b>	<b>Erklärung über bestehende Vorstrafen und anhängige Ermittlungsverfahren / Declaration concerning Criminal Record and Pending Investigations .....</b>	<b>42</b>

## 1. General introduction

### 1.1. Wheat and barley are key crops to feed a growing world population

The botanical family of *Poaceae* includes several economically relevant cereal plant species such as wheat (*Triticum aestivum* L.), barley (*Hordeum vulgare* L.), rye (*Secale cereal* L.), oat (*Avena sativa* L.), rice (*Oryza sativa* L.), maize (*Zea mays* L.), sorghum [*Sorghum bicolor* (L.) Moench], and pearl millet [*Pennisetum glaucum* (L.) R. Br.] (McKevith 2004). Cereals as an important source of carbohydrates, protein, lipids, fiber, and micronutrients play a central role in human diet and are staple food for large parts of the world's population (McKevith 2004; Wrigley et al. 2017). With a total production of 2.65 billion tons in the 2017-2018 growing season, cereals are the most produced commodity at the global food market with maize, wheat, rice, and barley accounting for 42.04%, 27.54%, 19.13%, and 5.27% of the entire production, respectively (FAO 2019). The main utilization of cereal production is food, feed, and other uses, corresponding to 42.60%, 35.60%, and 21.80%, respectively (FAO 2019). Alternative uses are for example the production of beverage, starch, and bioethanol (Wrigley et al. 2017). Wheat and barley are of high relevance in Europe, together accounting for 65.58% of the cereal production on the continent (FAO 2019).

### 1.2. Need to exploit genetic resources to ensure a sustainable future food security

The global human population is expected to rise to a level of 9 billion people by the middle of this century, which has to be compensated by an increase in agricultural production between 60% to 110% (Ray et al. 2013). Moreover, limited arable land, changing dietary and consumption behaviors in developing countries, the urgent need for a more resource efficient agriculture, the loss of biodiversity, as well as the increasingly occurring biotic and abiotic stress conditions in the course of climate change are putting additional pressure on future food production and food security (Godfray et al. 2010). In order to double the global food production by 2050, an annual yield increase of 2.4% in major crops would be necessary but, unfortunately, the actual rates vary between 1.6% for maize and 0.9% for wheat, which is

clearly insufficient to meet future demands (Ray et al. 2013). Therefore, Godfray et al. (2010) suggested a more holistic approach comprising in total five strategies to cope with the challenges of future food security by: (i) Expanding aquaculture as important source of animal protein; (ii) Changing the diets towards vegetarian nutrition to spare up to one third of the global cereal production, which is actually used as animal feed; (iii) Reducing the food waste considering that in both developing and developed countries roughly 30 to 40 % of food is lost due to various reasons; (iv) Closing the yield gap between the realized and potential productivity; and (v) Increasing the production limits by means of breeding, biotechnology and the exploitation of genetic variation harbored in old cultivars, landraces and wild relatives. Regarding the use of these genetic resources, the most important measure is to mine for valuable diversity in genebanks, which McCouch et al. (2013) elegantly summarized in the sentence: “The time is ripe for an effort to harness the full power of biodiversity to feed the world”.

### **1.3. Demands to make genebank collections accessible for breeding and research**

There are two main strategies for the preservation of plant biodiversity denoted as *in situ* and *ex situ*. *In situ* preservation is defined as the conservation of ecosystems and natural habits along with the maintenance and recovery of viable populations of species in their natural surroundings as well as in the surrounding where domesticated or cultivated species have developed their distinctive properties (FAO 2010). In contrast, *ex situ* preservation comprises the conservation of genetic resources outside their natural habits in specialized facilities, known as genebanks. With more than 1,750 facilities worldwide hosting 7.4 million accessions, genebanks represent the most significant and widespread mean of conserving plant genetic resources (FAO 2010).

The development of genebanks is an ongoing process, which can be traced back to the middle of the 19th century. Cohen et al. (1991) divided this process into four major time eras: the era (i) “of plant exploration and introduction”, (ii) “of conservation”, (iii) “of regeneration and new international linkages”, and (iv) “of more efficient utilization”. In the first era (from 1850 to 1950), accessions were gathered by global collection expeditions, studied

taxonomically and tested for utility. In the second era (from mid 1950s to the 1980s), the awareness of the need to conserve genetic resources in mid- and long-term storage facilities has been raised by the *Green Revolution*, which displaced local adapted varieties by high-yielding cultivars. In the third era (during the 1990s), the collection activities were reduced, long-term viability of the collections were ensured and international agreements as well as cooperation were enhanced. Finally, during the fourth era (2010 and beyond) the collections started to become more user friendly by intensive phenotypic and genotypic characterization, which can facilitate their exploitation in plant breeding by means of intensive research and pre-breeding activities supported by biotechnology and information networks.

#### **1.4. Wheat and barley are among the most prevalent accessions in genebanks**

Genebanks can be classified into national and international facilities. The largest four national collections are the National Center of Genetic Resources Preservation (NCPGR) in the United States of America, the Institute of Crop Germplasm Resources of the Chinese Academy of Agricultural Sciences (ICGR-CAAS) in China, the National Bureau of Plant Genetic Resources (NBPGR) in India, and the N.I. Vavilov Research Institute of Plant Industry (VIR) in Russia. These national collections host roughly 509,000, 392,000, 366,000, and 322,000 accessions of plant genetic resources, respectively. International genebanks are often dedicated to certain agro-climatic regions or they are specialized in certain crops. In particular, the Consultative Group on International Agricultural Research (CGIAR) put together several genebanks that represent major collections of their mandate crops in behalf of the world community. Among them, the International Maize and Wheat Improvement Center (CIMMYT) in Mexico holds 13% and 8% of the global wheat and maize accession, respectively, whereas the International Rice Research Institute (IRRI) in the Philippines holds 14% of global rice accessions. In contrast, the major barley collection comprising 9% of the global barley accessions is hosted by a national genebank: the Plant Gene Resources of Canada (PGRC) in Canada (FAO 2010). Cereals account for 45% of the accessions hosted in genebanks worldwide. Thereof wheat, rice, barley, and maize represent the four largest global collections of cereal species, amounting to roughly 856,000, 774,000, 467,000, and 378,000 accessions, respectively.

However, it has to be mentioned that only 25-30 % of the total global holdings of plant genetic resources are expected to be unique, with the remaining ones being duplicated either in the same or a different collection.

### **1.5. Insufficient characterization of accessions or a limited access to this information hampers the exploitation of entire genebank collections**

Modern information technologies play an important role in genebank management on an institutional, regional and global level. For example, the Genbank Information System (GBIS) of the German Federal ex situ Genebank of Agricultural and Horticultural Crop Species is a multi-functional domain supporting the wide range of workflows within the genebank (e.g., seed logistics, seed regeneration, phenotyping, documentation), providing accessions-related information to external users and enabling the online request of material worldwide (Oppermann et al. 2015). These institutional or national inventories are the basis for building up an international infrastructure preserving and representing the world's plant biodiversity. One example on a continental level is the European Search Catalogue for Plant Genetic Resources (EURISCO), which represents 1.8 million accessions of plant genetic resources preserved by 43 member countries (Weise et al. 2017). EURISCO together with two other multi-institutional inventories denoted as the Germplasm Resources Information Network (GRIN) and the System-wide Information Network of Genetic Resources (SINGER) form the major source for the Global Portal on Plant Genetic Resources (GENESYS). Together, they represent more than 4 million accessions hosted at 463 institutes worldwide (GENESYS 2019). Unfortunately, large genebanks such as the national collections of India and China are not yet represented in GENESYS. However, these information platforms are only sparsely populated with relevant data. Passport data describing the taxonomy and provenance of an accession are often incomplete, occasionally erroneous and of limited agronomic relevance overall (Mascher et al. 2019). Phenotypic and genomic characterization data are even absent in entire genebank collections, which severely hinders the selection of appropriate accessions by breeders and researchers (McCouch et al. 2013). Moreover, sometimes is the access to characterizations rather than the absence of them the limiting factor that makes this information in practice inexistent for users. The simplest example for that is the lack of



digitization of historical genebank records from field books and file cards (Keilwagen et al. 2014). In summary, this reveals that the availability of genetic resources is not the fundamental problem, but the identification of useful diversity (Mascher et al. 2019).

### **1.6. Considerable interest for seed samples at the German Federal *ex situ* Genebank**

The German Federal *ex situ* Genebank of agricultural and horticultural crops hosted by the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben is the most comprehensive collection in the European Union and ranks among the ten largest collections worldwide preserving 151,000 accessions of ~3,000 plant species from 756 genera (Oppermann et al. 2015). The predecessor of the IPK Genebank was founded during 1943 in Vienna, Austria, and was relocated after World War II to Gatersleben, Germany, in 1946. Starting from the early 1920s the accessions were accumulated by more than 150 collection expeditions worldwide. Further, the collection was enriched by extensive seed exchange with other genebanks, botanic gardens, research and breeding institutions. A significant growth of the collection took place in the early 1990s, when several collections held by former governmental breeding stations of the former German Democratic Republic were incorporated and during 2003, when 45.000 accession of the West German genebank (Federal Center for Breeding Research on Cultivated Plants, Braunschweig) were transferred to Gatersleben. The preservation of the accessions is warranted by cold storage facilities, permanent garden, *in vitro* culture, and cryo-preservation at ultra-low temperature. Wheat and barley comprise roughly one third of the entire collection (Börner 2006).

Compared to other genebanks, the IPK wheat and barley collection is the ninth and sixth largest collection worldwide, with around 27,000 and 22,000 accessions, respectively (FAO 2010). Furthermore, the IPK barley collection was particularly described as one of the best documented germplasm collection in terms of taxonomic classification (de Carvalho et al. 2013). The value of the entire collection is also confirmed by 600 user requests per year involving the distribution of approximately 30.000 seed samples (Philipp et al. 2018b). External users can request passport data from GBIS for each accessions including, among others, information on origin, collection date, donor, seed availability, growth habit, and

taxonomic classification (Oppermann et al. 2015). Nevertheless, the lack of comprehensive, publicly available phenotypic and genomic characterization data is again the major obstacle for the educated choice of accessions (Philipp et al. 2018b).

### **1.7. Historical data from seed regeneration can be used for the phenotypic characterization of genebank collections**

Next to the long-term storage the regeneration of the accessions is a fundamental task of genebank management. This is done routinely when (i) seed stocks dropped below a critical quantity threshold; (ii) the germination rate dropped below a critical quality threshold; (iii) large quantities of seeds are needed for research activities; or (iv) new accessions need to be multiplied and characterized before entering the collection. Due to these reasons, between 8 and 10 % of the entire collection have to be regenerated every year (Börner 2006). In order to guarantee the genetic integrity of the accessions across decades genebank curators follow strict quality guidelines during the regeneration cycles (Börner 2006; Lehmann and Mansfeld 1957). This quality assessment includes, among others, the observation of morphological and agronomic traits for the phenotypic comparison with previous regenerations cycles. The morphological classification system developed by Rudolf Mansfeld still plays a central role in determining the phenotypic integrity of wheat and barley accessions. This classification is focused on highly heritable morphological traits like grain color, awning, spike shape, and spike color. The entirety of these morphological characteristics is finally reflected in the taxonomic name of the accession (Mansfeld 1950, 1951). These taxonomic names in turn can be easily used to retrieve morphological traits of accessions, e.g., to extend large genetic studies (Milner et al. 2019). However, the relevance of morphological traits for crop improvement is limited.

Fortunately, also agronomic traits like flowering time, plant height and thousand grain weight are sometimes routinely recorded during regeneration cycles. The present work encompasses historical phenotypic records of these three traits gathered during 70 years of seed regeneration of the IPK wheat and barley collection. While for barley the entire collection was considered, for wheat the focus was hexaploid bread wheat (*Triticum aestivum* L.) as predominant wheat species in the collection comprising roughly 20.000 accessions.

Further, large parts of the incorporated collection of the former West German genebank have not been regenerated at the IPK so far, thus, no historical data was accessible for these accessions. In total, phenotypic historical data was available for 12,876 barley and 12,754 bread wheat accessions representing roughly 60 % and 64 % of their overall collections, respectively. Due to the growth habit, spring and winter sown barley and wheat has to be distinguished. While spring sown accessions (77 %) dominated the barley collection, the ratio was quite balanced in the wheat collection (51 %, spring sown). Since on average only 8 % of the wheat and barley collection was regenerated per year, the historical data sets are characterized by a high non-orthogonal data structure, a.k.a. “unbalancedness”, and the direct comparison of the phenotypic values among accessions across years is hardly possible. Further, accessions which entered the collection earlier went through more regeneration cycles compared to recently incorporated material. This had an impact on the phenotyping intensity per accession varying between 1 and 22 regeneration cycles. Additionally, changes in the regeneration intervals due to technological progress, historical events like the acquisition of material from large collection expeditions and through donations of other institutes as well as comprehensive research projects had an impact on the quantity of material assessed per year. In wheat, for instance, between 37 and 3,012 accessions were regenerated on a yearly basis. Last but not least, the changing climate conditions during the last 70 years also can have an impact on trait performance. Taken together, the analysis of historical phenotypic regeneration data is challenging. In contrast, modern high-throughput genotyping technologies enable a time and cost-effective genomic characterization of entire genebank collections (Kilian and Graner 2012; Milner et al. 2019; Sehgal et al. 2015; Wang et al. 2014).

## 1.8. The FAIR principles as guideline for scientific data management

The FAIR principles have been proposed to establish an infrastructure that promotes the sustainable use of scientific data. (Wilkinson et al. 2016). FAIR refers to the data and metadata, which have to be: (i) Findable; (ii) Accessible; (iii) Interoperable; and (iv) Reusable. With the continuous generation of phenotypic and genomic characterization data of genebank collections, the question of common data standards for sustainable data management arises. This is crucial to integrate information within and across genebanks in order to maximize the benefit from characterization investments. In particular the reuse and exchange of phenotypic data is often hampered by the large variability of phenotyping protocols, the multitude of phenotypic traits measured, and the environment-dependency of these traits (Krajewski et al. 2015). In order to overcome these problems a list of attributes for the description of phenotypic observations: Minimum Information About a Plant Phenotyping Experiment (MIAPPE) was proposed (Ćwiek-Kupczyńska et al. 2016). In addition, common guidelines for storing and exchanging these data and their corresponding metadata have been worked out. In particular the ISA-Tab, “Investigation-Study-Assay” format has become standard in the bioscience community to obtain interoperable phenotypic data (Ćwiek-Kupczyńska et al. 2016; Rocca-Serra et al. 2010; Sansone et al. 2012). Community standards for the stewardship of genomic data have not been agreed so far (Mascher et al. 2019). In order to ensure the findability and accessibility of data and metadata, these must finally be stored in appropriate public repositories (Arend et al. 2016; Wilkinson et al. 2016).

## 1.9. Objectives

The overall goal of this work was to develop strategies for the quality assessment and aggregation of historical phenotypic data in order to transform the IPK wheat and barley collection from living archives into a bio-digital resource center. In particular, the objectives were to:

- I. Examine approaches to exploit genebank collections for grain yield improvement (Philipp et al. 2018a).
- II. Elaborate strategies to analyze and validate non-orthogonal historical phenotypic data gathered during seven decades of seed regeneration in an *ex situ* genebank collection of wheat and barley (González et al. 2018a; Philipp et al. 2018b)
- III. Study challenges and opportunities of implementing the FAIR-principles for the release of quality assessed historical phenotypic data from the IPK wheat and barley collection as a blueprint for other genebanks to unlock their historical data treasures (Gonzalez et al. 2018b; Philipp et al. 2019).

## 2. Peer-reviewed scientific articles

### 2.1. Grain number and grain yield distribution along the spike remain stable despite breeding for high yield in winter wheat

Norman Philipp<sup>1\*</sup>, Heiko Weichert<sup>2</sup>, Utkarsh Bohra<sup>2</sup>, Winfriede Weschke<sup>2</sup>, Albert W. Schulthess<sup>1</sup>, Hans Weber<sup>2</sup>

<sup>1</sup> Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstr. 3, 06466, Gatersleben, Germany

<sup>2</sup> Department of Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstr. 3, 06466, Gatersleben, Germany

\*Corresponding author (philipp@ipk-gatersleben.de).

#### Abstract

Two winter wheat (*Triticum aestivum* L.) populations, i.e. 180 genetic resources and 210 elite varieties, were compared in a field trial to analyse how grain number and grain yield distribution along the spike changed during the breeding process and how this associates to yield-related traits. Elites showed in average 38% more yield compared to resources. This breeding improvement mainly derived from an increase in grains and yield per spike in addition to grains and yield per spikelet. These increments corresponded to 19, 23, 21 and 25%, respectively. Not much gain in thousand grain weight (4%) was observed in elites as compared to resources. The number of spikelets per spike was not, or even negatively, correlated with most traits, except of grains per spike, which suggests that this trait was not favoured during breeding. The grain number and grain yield distributions along the spike (GDAS and GYDAS) were measured and compared by using a novel mathematical tool. GDAS and GYDAS measure the deviation of a spike of interest from the architecture of a model spike with even grain and yield distribution along all spikelets, respectively. Both traits were positively correlated. Elites showed in average only a 1% improvement in GDAS and GYDAS values compared to resources. This comparison revealed that breeding increased grain number and yield uniformly along the spike without changing relative yield input of individual spikelets, thereby, maintaining the general spike architecture.

Published in:

*PLoS ONE* (2018) 13(10):e0205452. doi: 10.1371/journal.pone.0205452

The original article has been published in an international journal. Due to copyright restrictions, it is available online:

<https://doi.org/10.1371/journal.pone.0205452>

## 2.2. Leveraging the use of historical data gathered during seed regeneration of an *ex situ* genebank collection of wheat

Norman Philipp<sup>1</sup>, Stephan Weise<sup>2</sup>, Markus Oppermann<sup>2</sup>, Andreas Börner<sup>2</sup>, Andreas Graner<sup>2</sup>, Jens Keilwagen<sup>3</sup>, Benjamin Kilian<sup>4</sup>, Yusheng Zhao<sup>1</sup>, Jochen C. Reif<sup>1\*</sup>, Albert W. Schulthess<sup>1</sup>

<sup>1</sup> Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, 06466, Gatersleben, Germany

<sup>2</sup> Department of Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, 06466, Gatersleben, Germany

<sup>3</sup> Institute for Biosafety in Plant Biotechnology, Julius Kühn-Institut (JKI) – Federal Research Centre for Cultivated Plants, 06484 Quedlinburg, Germany

<sup>4</sup> Global Crop Diversity Trust, Platz der Vereinten Nationen 7, 53113 Bonn, Germany

\*Corresponding author (reif@ipk-gatersleben.de).

### Abstract

Genebanks are a rich source of genetic variation. Most of this variation is absent in breeding programs but may be useful for further crop plant improvement. However, the lack of phenotypic information forms a major obstacle for the educated choice of genebank accessions for research and breeding. A promising approach to fill this information gap is to exploit historical information gathered routinely during seed regeneration cycles. Still, this data is characterized by a high non-orthogonality hampering their analysis. By examining historical data records for flowering time, plant height, and thousand grain weight collected during 70 years of regeneration of 6,207 winter wheat (*Triticum aestivum* L.) accessions at the German Federal *ex situ* Genebank, we aimed to elaborate a strategy to analyze and validate non-orthogonal historical data in order to charge genebank information platforms with high quality ready-to-use phenotypic information. First, a three-step quality control assessment considering the plausibility of trait values and a standard as well as a weather parameter index based outlier detection was implemented, resulting in heritability estimates above 0.90 for all three traits. Then, the data was analyzed by estimating best linear unbiased estimations (BLUEs) applying a linear mixed-model approach. An *in-silico* resampling study mimicking different missing data patterns revealed that accessions should be regenerated in a random fashion and not blocked by origin or acquisition date in order to minimize estimation biases in historical data sets. Validation data was obtained from multi-environmental orthogonal field trials considering a random subsample of 3,083 accessions. Correlations above 0.84 between BLUEs estimated for historical data and validation trials outperformed previous approaches and confirmed the robustness of our strategy as well as the high quality of the historical data. The results indicate that the IPK winter wheat collection reveals an extraordinary high phenotypic diversity compared to other collections. The quality checked ready-to-use phenotypic information resulting from this study is the first brick to extend traditional, conservation driven genebanks into bio-digital resource centers.

Published in:

*Front Plant Sci* (2018) 9:609. doi: 10.3389/fpls.2018.00609

The original article has been published in an international journal. Due to copyright restrictions, it is available online:

<https://doi.org/10.3389/fpls.2018.00609>

### 2.3. Unlocking historic phenotypic data from an *ex situ* collection to enhance the informed utilization of genetic resources of barley (*Hordeum sp.*)

Maria Y. González<sup>1</sup>, Norman Philipp<sup>1</sup>, Albert W. Schulthess<sup>1</sup>, Stephan Weise<sup>2</sup>, Yusheng Zhao<sup>1</sup>, Andreas Börner<sup>2</sup>, Markus Oppermann<sup>2</sup>, Andreas Graner<sup>2</sup>, Jochen C. Reif<sup>1\*</sup>

<sup>1</sup> Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466, Gatersleben, Germany

<sup>2</sup> Department of Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466, Gatersleben, Germany

\*Corresponding author (reif@ipk-gatersleben.de).

#### Abstract

Precise estimates of trait performance of genetic resources are considered as an intellectually challenging, complex, costly and time-consuming step needed to exploit the phenotypic and genetic diversity maintained in genebanks for breeding and research. Using barley (*Hordeum sp.*) as a model, we examine strategies to tap into historic data available from regeneration trials. This is a first step towards extending the Federal *ex situ* Genebank into a bio-digital resource center facilitating an informed choice of barley accessions for research and breeding. Our study is based on historic data of seven decades collected for flowering time, plant height, and thousand kernel weight during the regeneration of 12,872 spring and winter barley accessions. Linear mixed models were implemented in conjunction with routines for assessment of data quality. A resampling study highlights the potential risk of biased estimates in second-order statistics when grouping accessions for regeneration according to the year of collection or geographic origin. Based on rigorous quality assessment we obtained high heritability estimates for the traits under consideration exceeding 0.8. Thus, the best linear unbiased estimations for the three traits are a valuable source to populate a bio-digital resource center for the IPK barley collection. The proposed strategy to leverage historic data from regeneration trials is not crop specific and can be used as a blueprint for other *ex situ* collections.

Published in:

*Theor Appl Genet* (2018) 131:2009. doi: 10.1007/s00122-018-3129-z

The original article has been published in an international journal. Due to copyright restrictions, it is available online:

<https://doi.org/10.1007/s00122-018-3129-z>



## 2.4. Unbalanced historical phenotypic data from seed regeneration of a barley *ex situ* collection

Maria Y. González<sup>1</sup>, Stephan Weise<sup>2</sup>, Yusheng Zhao<sup>1</sup>, Norman Philipp<sup>1</sup>, Daniel Arend<sup>1</sup>, Andreas Börner<sup>2</sup>, Markus Oppermann<sup>2</sup>, Andreas Graner<sup>2</sup>, Jochen C. Reif<sup>1</sup>, Albert W. Schulthess<sup>1</sup>

<sup>1</sup> Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466, Gatersleben, Germany

<sup>2</sup> Department of Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466, Gatersleben, Germany

\*Corresponding author (reif@ipk-gatersleben.de).

### Abstract

The scarce knowledge on phenotypic characterization restricts the usage of genetic diversity of plant genetic resources in research and breeding. We describe original and ready-to-use processed data for approximately 60% of ~22,000 barley accessions hosted at the Federal *ex situ* Genebank for Agricultural and Horticultural Plant Species. The dataset gathers records for three traits with agronomic relevance: flowering time, plant height and thousand grain weight. This information was collected for seven decades for winter and spring barley during the seed regeneration routine. The curated data represent a source for research on genetics and genomics of adaptive and yield related traits in cereals due to the importance of barley as model organism. This data could be used to predict the performance of non-phenotyped individuals in other collections through genomic prediction. Moreover, the dataset empowers the utilization of phenotypic diversity of genetic resources for crop improvement.

Published in:

*Sci Data* (2018) 5:180278. doi: 10.1038/sdata.2018.278

The original article has been published in an international journal. Due to copyright restrictions, it is available online:

<https://doi.org/10.1038/sdata.2018.278>

The data is published in the *e!DAL - Plant Genomics and Phenomics Research Data Repository (PGR)* and can be assessed here:

<https://dx.doi.org/10.5447/IPK/2018/10>

## 2.5. Historical phenotypic data from seven decades of seed regeneration in a wheat *ex situ* collection

Norman Philipp<sup>1</sup>, Stephan Weise<sup>2</sup>, Markus Oppermann<sup>2</sup>, Andreas Börner<sup>2</sup>, Jens Keilwagen<sup>3</sup>, Benjamin Kilian<sup>4</sup>, Daniel Arend<sup>1</sup>, Yusheng Zhao<sup>1</sup>, Andreas Graner<sup>2</sup>, Jochen C. Reif<sup>1\*</sup>, Albert W. Schulthess<sup>1</sup>

<sup>1</sup> Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, 06466 Seeland / OT Gatersleben, Germany

<sup>2</sup> Department of Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, 06466 Seeland / OT Gatersleben, Germany

<sup>3</sup> Institute for Biosafety in Plant Biotechnology, Julius Kühn-Institut (JKI) – Federal Research Centre for Cultivated Plants, 06484 Quedlinburg, Germany

<sup>4</sup> Global Crop Diversity Trust, Platz der Vereinten Nationen 7, 53113 Bonn, Germany

\*Corresponding author (reif@ipk-gatersleben.de).

### Abstract

Genebanks are valuable sources of genetic diversity, which can help to cope with future problems of global food security caused by a continuously growing population, stagnating yields and climate change. However, the scarcity of phenotypic and genotypic characterization of genebank accessions severely restricts their use in plant breeding. To warrant the seed integrity of individual accessions during periodical regeneration cycles in the field phenotypic characterizations are performed. This study provides non-orthogonal historical data of 12,754 spring and winter wheat accessions characterized for flowering time, plant height, and thousand grain weight during 70 years of seed regeneration at the German genebank. Supported by historical weather observations outliers were removed following a previously described quality assessment pipeline. In this way, ready-to-use processed phenotypic data across regeneration years were generated and further validated. We encourage international and national genebanks to increase their efforts to transform into bio-digital resource centers. A first important step could consist in unlocking their historical data treasures that allows an educated choice of accessions by scientists and breeders.

Published in:

*Sci Data* (2019) 6(1):137. doi: 10.1038/s41597-019-0146-y

The original article has been published in an international journal. Due to copyright restrictions, it is available online:

<https://doi.org/10.1038/s41597-019-0146-y>

The data is published in the *e!DAL - Plant Genomics and Phenomics Research Data Repository (PGR)* and can be assessed here:

<http://doi.org/10.5447/IPK/2019/11>

### 3. General discussion

Domestication and modern plant breeding imposed influential genetic bottlenecks on crop plants. Strong selection pressure on wild populations for traits like non-shattering seeds, the loss of germination inhibition, and growth habit by early farmers resulted in locally adapted landraces, which were finally replaced by modern high-yielding cultivars. During this domestication and breeding process, it is likely that beneficial allelic variations may have been lost (Purugganan 2019). The extensive collections of genetic resources hosted in genebanks worldwide are a great opportunity to exploit this untapped genetic potential for human needs (Tanksley and McCouch 1997).

Genetic resources are seen as a rich allelic source to improve biotic resistances, abiotic stress tolerance, quality, and grain yield in major crops (Hajjar and Hodgkin 2007). Therefore, the exploitation of plant genetic diversity has been promoted as a crucial approach to feed a growing world population under the impact of climate change (Godfray et al. 2010; McCouch et al. 2013). Following these needs, several exploitation strategies of wheat and barley genetic resources were described (e.g., Lopes et al. 2015; Mascher et al. 2019; McCouch et al. 2013; Mondal et al. 2016). In summary, they all conclude that the insufficient availability of phenotypic and genomic characterization and poor data infrastructure are the major obstacles to the exploitation of plant genetic resources by modern breeding and research technologies. In this vein, McCouch et al. (2013) proposed three consecutive steps to mine genebank diversity: (i) The genetic fingerprinting of the world's genebank accessions; (ii) The multi-environmental phenotyping of these accessions; and (iii) The creation of an internationally accessible informatics infrastructure to catalogue the diversity of global genebank collections. Among the three tasks, the evaluation of the phenotypic performance is considered to be "... the most intellectually challenging, complex, costly and time-consuming stage" (McCouch et al. 2013).

In the last decade much progress has been made in the genomic characterization of entire genebank collections (Cavanagh et al. 2013; Crossa et al. 2016; Mascher et al. 2019; Milner et al. 2019; Sehgal et al. 2015; Yu et al. 2016) and advances in the international data infrastructure are also substantial (Ćwiek-Kupczyńska et al. 2016; GENESYS 2019; Krajewski et al. 2015; Oppermann et al. 2015; Sansone et al. 2012; Weise et al. 2017; Wilkinson et al. 2016).

Nevertheless, the systematic exploitation of plant genetic resources is still very limited due to the lack of sufficient characterization for useful phenotypic traits. Therefore, the main goal of this work was to present strategies for the collection, analysis and publication of phenotypic data in order to improve the phenotypic characterization of wheat and barley genebank collections.

### **3.1. Exploiting wheat genetic resources for grain yield improvement**

#### **3.1.1. Grain yield gap between wheat genetic resources and elite varieties**

The economically most important trait for wheat is grain yield. Previous studies on mining for valuable diversity (Keilwagen et al. 2014), however, focused rather on less complex agronomic traits such as flowering time and plant height as a first step towards a better characterization of genebank accession. The genetic architecture of these agronomic traits is well investigated and can be explained by a few major genes. Furthermore, the contributions of these major genes to grain yield were more due to indirect effects, adapting wheat to specific eco-agronomic conditions, rather than improving grain yield performance *per se*. The dwarfing genes of the Green Revolution, for example, prevented lodging of elite varieties under excessive N-fertilization, allowing them to reach their full yield potential (Hedden 2003). The insensitivity to the day length for wheat (*Ppd-1*) is more a strategy of adaptation to hot and dry environments but has no impact on grain yield under optimal growing conditions (Dawson et al. 2015; Semenov et al. 2014; Worland 1996). Therefore, most of the gains in grain yield and grain yield potential were due to recombining minor genes in adapted background of elite material (Falk 2010). This causes a grain yield gap between genetic resources and elite varieties, which makes the identification of superior accessions based on their *per se* grain yield performance unlikely. The grain yield gap was assessed in this thesis, which revealed that European elite varieties performed on average 38 % better compared to a random sample of genebank accessions (Philipp et al. 2018a). Apart from the grain yield gap, the multi-environmental evaluation for grain yield of entire genebank collections would be severely limited by costs and resources. In order to exploit genetic resources for grain yield improvement in a resource-efficient manner, the possibility of splitting grain yield into its components and investigating yield components for useful diversity was studied.

### 3.1.2. Investigating grain yield components for useful diversity

The missing adaptation and the costs make the evaluation of genetic resources for their *per se* grain yield performance unfeasible. Alternatively, the investigation of grain yield components could be an interesting approach to gather grain yield related data of genetic resources in a cheaper and less biased manner. Thousand grain weight, grains per spike, and spikes per area are often considered as major grain yield components in wheat (Simmonds et al. 2014). Due to the high heritability thousand grain weight is often used in yield related studies. Especially, in low and intermediate-yielding environments thousand grain weight is associated to higher grain yield performance (Lopes et al. 2012). In addition, other yield-related traits might also be of interest, either to select accessions for breeding purposes or to study how these traits have been changed in the course of plant breeding in order to improve breeding strategies. In this context, a study was conducted to answer the question how and to what extent the grain yield distribution along the spike (GYDAS) was changed during the breeding process in European winter bread wheat. Therefore, a novel measure was developed to quantify GYDAS and compare a random sample of 180 genetic resources and 210 European winter wheat elite varieties for several yield-related traits (Philipp et al. 2018a). Despite the fact that the elite population outyielded genetic resources by 38%, there was only marginally improvement in GYDAS (1%) revealing that breeding increased grain number and grain yield uniformly along the spike of winter wheat. In accordance with other studies (Würschum et al. 2018), we found that breeding progress in European winter wheat is likely due to the increase of grains per spikelet, which can be interpreted as an enhanced spikelet fertility. This knowledge can be used to improve future breeding strategies like, for example, the establishment of hybrid breeding programs in wheat and barley based on cytoplasmatic male sterility (cms) systems that rely on the restoration of floret fertility by restorer genes (Ahokas 1979; Longin et al. 2012; Merfert et al. 1988; Mühleisen et al. 2013). In contrast to the improvement in spikelet fertility observed in the long term, instability of the restoration system can lead to reduced spikelet fertility (Merkert et al. 1988). Therefore, selection for spikelet fertility should be a particularly important task in wheat hybrid breeding.

## **3.2. Leveraging the use of historical phenotypic data**

### **3.2.1. Finding the right key to unlock the value behind historical data**

More than 120,000 plots of wheat and barley were grown and characterized at the IPK genebank during the past 70 years of seed regeneration. The re-evaluation of these accessions in field trials with similar phenotyping intensity and plot size would cost more than 5.75 million Euro. Therefore, the historical phenotypic data represent a valuable, untouched treasure that could be salvaged for the characterization of plant genetic resources. Nevertheless, the historical observations are characterized by a highly non-orthogonal data structure, which makes their analysis challenging. Keilwagen et al. (2014) suggested a Normalized Rank Product (NRP) approach to overcome the non-orthogonality in the historical data of the IPK wheat and barley collection. Unfortunately, the application of NRP leads to a reduction in the information content of phenotypic data (e.g., units get lost), which makes trait comparisons between data sets, e.g., from other genebanks or between the genebank and breeding programs, difficult. In the context of the present work, an analysis strategy based on linear mixed models was developed to overcome the previous problems.

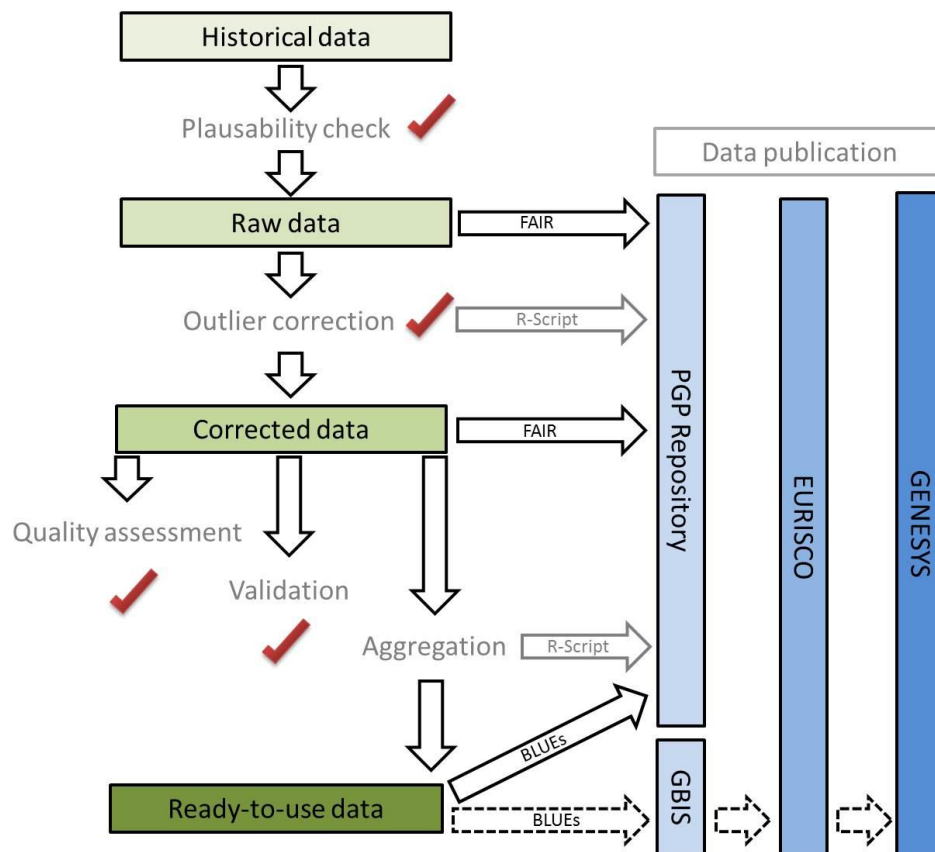
### **3.2.2. From non-orthogonal historical observations towards high quality phenotypic characterization of genebank collections**

A strategy for analyzing historical phenotypic data was elaborated based on flowering time, plant height, and thousand grain weight records of 6,207 accessions of the IPK winter bread wheat collection gathered during 70 years of seed regeneration (Philipp et al. 2018b) This data set was characterized by a multi-dimensional non-orthogonal structure. Considering this data as two-dimensional array with phenotypic observations of 6,207 accessions in 70 years, 92% of the data points correspond to missing values. Moreover, structured missing data due to historical seed regeneration patterns potentially biases the data analysis. Taken together, a rigorous quality assessment including plausibility checks, outlier corrections, bias estimation and the validation with independent data is indispensable when analyzing historical phenotypic data. A random validation sample of 3,083 winter wheat accessions, which was evaluated in multi-environmental orthogonal field trials, finally confirmed the high quality of the historical data and the robustness of the described approach. This was the motivation to

apply the elaborated strategy to further available historical data of the IPK wheat (Philipp et al. 2019) and barley (González et al. 2018a; Gonzalez et al. 2018b) collections, which also yielded high quality, ready-to-use phenotypic data of more than 25,600 wheat and barley accessions in total.

### **3.2.3. Rigorous quality assessment is essential for the analysis of historical data**

The workflow of analyzing historical data is shown in Fig. 1. A labor and time intensive task that should not be underestimated is the compilation of the raw data from historical observations, which were mainly recorded in the pre-digital era. The digitalization of historical data relies on a long-living and well-maintained genebank documentation system. The management of phenotypic data was manually at the IPK genebank until 2011, where first the traits were recorded in field books, then transferred to card files and finally digitized and stored in GBIS. Since 2011, field computers were used for recoding and storing the data directly from field. The manual transfer and digitalization of data carries the risk of errors ranging from individual typing mistakes to systematic misformatting of the data records. Moreover, records from off-season sowing, physiological absurd values, highly unadapted accessions, and wrongly classified species have to be identified and removed. Therefore, the compilation of the raw data from historical records was accompanied by a comprehensive, manual plausibility check, which demands deep knowledge of genebank management, taxonomy, agronomy, and cereal biology (Philipp et al. 2018b). After plausibility check, two steps of statistical outlier correction based on linear mixed models were performed using the raw data. In the first step, suspicious evaluation years were identified inspecting the coefficient of variation between the square root of the year specific error variances and the average effect of the year.



**Fig. 1** Workflow diagram for the quality assessment and publication of highly non-orthogonal historical phenotypic data gathered during seed regeneration of *ex situ* genebank collections. The raw data were compiled from historical data based on intensive plausibility checks before statistical outlier correction strategies were applied. After outlier correction the data quality was assessed by quantitative genetic parameters and validated with independent data. Finally, the outlier corrected and quality assessed data were aggregated by Best Linear Unbiased Estimates (BLUEs). The raw data, corrected data, ready-to-use data, and their corresponding R-software scripts for the reproduction of the results were deposited in the Plant Genomics and Phenomics Data Repository (PGP) hosted at the Leibniz Institute of Plant Genetics and Crop plant Research (IPK) using e!DAL as software infrastructure (Arend et al. 2016) and following the FAIR guiding principles for scientific data management [Findability, Accessibility, Interoperability and Reusability (Wilkinson et al. 2016)]. The ready-to-use data could potentially be integrated into the Genebank Information System (GBIS) of the IPK to promote the educated choice of accessions. GBIS as part of the European Search Catalogue for Plant Genetic Resources (EURISCO) and the Global Portal on Plant Genetic Resources (GENESYS) infrastructure could potentially provide global access to these phenotypic data.

Since excluding entire evaluation years is accompanied with the loss of information and statistical power, historical weather data was consulted to confirm and remove “outlier” years. The second step was a standard outlier correction based on studentized residuals and applying a Bonferroni-Holm test to correct for multiple testing (Philipp et al. 2018b). The quality of the outlier corrected data was assessed by estimating variance components and heritabilities. While outlier correction reduced the error variance up to 29% for flowering time in winter wheat, heritability estimates above 0.8 revealed an overall good data quality across



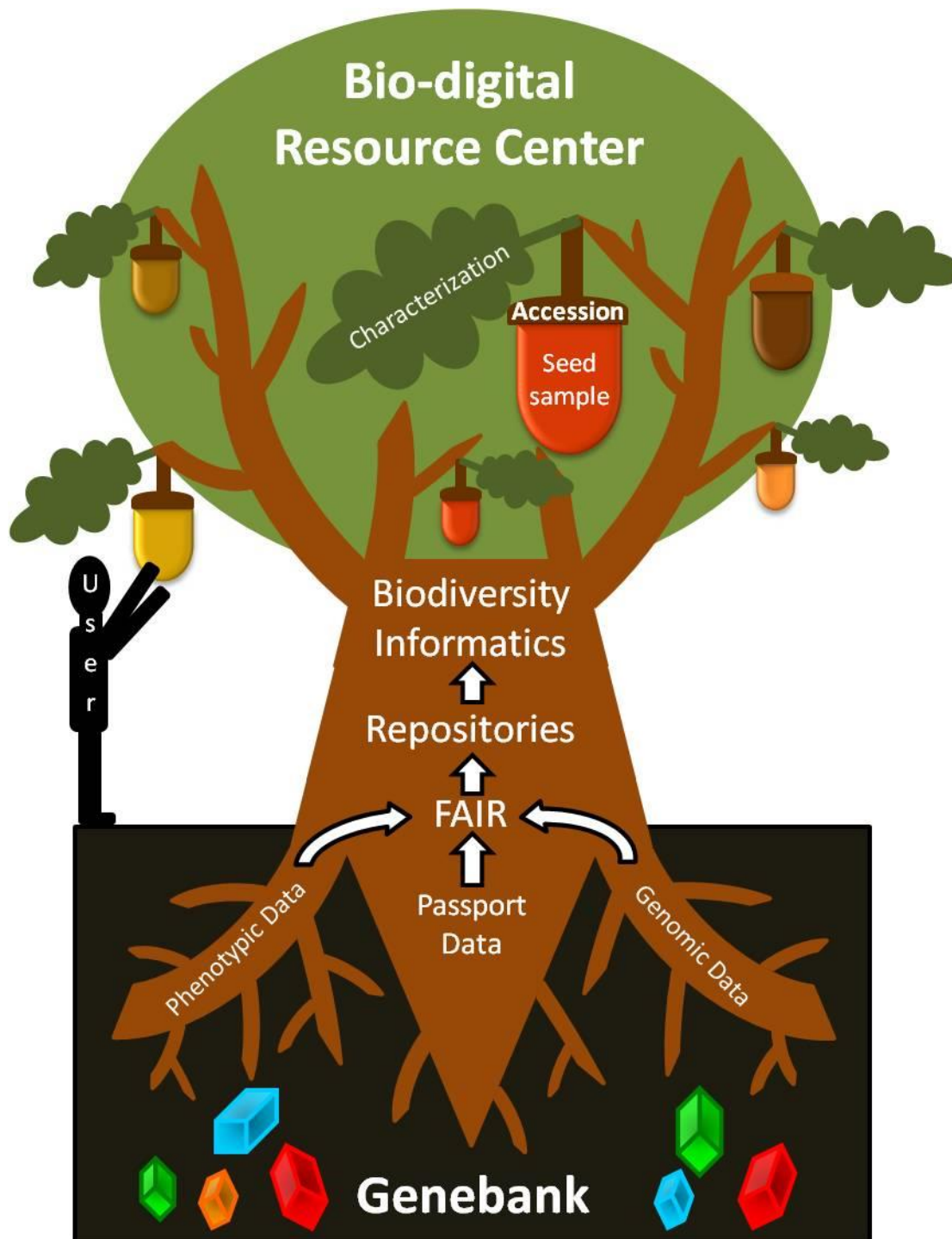
traits in wheat and barley. An *in silico* study based on an orthogonal subdata set revealed a potential bias in estimating quantitative genetic parameters due to a block-wise regeneration of accessions until 1976. However, these effects seem to be negligible when considering the entire data set. Nevertheless, in order to minimize the estimation bias from historical phenotypic data, it is recommendable that genebank curators regenerate accessions in a random fashion and not grouped by provenance or acquisition date. The quality assessment is complemented by the independent validation of historical data. While, for wheat a comprehensive validation set from orthogonal field trials was available, a two-fold cross validation was performed within the barley historical data set. Correlation coefficients equal to or greater than 0.83 reflect high data quality in both validation scenarios. Finally, the non-orthogonal, quality assessed data were aggregated by estimating the Best Linear Unbiased Estimates (BLUEs), resulting in intercomparable, ready-to-use phenotypic data.

### **3.3. From conservation-driven genebanks towards bio-digital resource centers**

The present work provides the phenotypic characterization for flowering time, plant height and thousand grain weight for more than 60% of the IPK bread wheat and barley collection. In combination with the ongoing progress in genomic fingerprinting of entire genebank collections (Milner et al. 2019) this phenotypic data can serve as a training set for genomic prediction of the missing phenotypes (Yu et al. 2016). As a result, genebanks could provide comprehensive information on agronomic traits, yield components, breeding values and genomic profiles of accessions in addition to seed samples and traditional passport data. Well-characterized genebank collections facilitate the informed choice of favorable accessions and thus promote the use of plant genetic resources for individual breeding and research projects. Considering only the historical data for flowering time and plant height, breeders and pre-breeders can, for example, estimate the phenotypic degree of adaptation of an accession to their target environments. Similarly, when compiling research populations, for example, to study ear and leaf diseases scientists can avoid bias in plant developmental stage caused by too much variation in flowering time and plant height. Moreover, the phenotypic and genomic characterization of entire genebank collections enables systematic approaches for

the identification of useful diversity, e.g., by means of association mapping and allele mining (Keilwagen et al. 2014; Milner et al. 2019). Finally, genomic prediction models can be applied to impute phenotypic characterization across genebanks (Mascher et al. 2019). Assuming two distinct genebank collections 1 and 2 fingerprinted with the same marker platform but phenotyped in two different eco-geographic environments A and B, respectively, than the performance of collection 1 in environment B and collection 2 in environment A could be predicted using the phenotypic data collected in the respective environment as training set. This approach would enable to screen global genebanks for accessions with favorable phenotypic characteristics, which are potentially adapted to the target environment. Nevertheless, integrating phenotypic, genomic and passport data from different spatial and temporal sources is a challenge, and common guidelines for data management are needed.

In order to promote the sustainable use of the historical phenotypic data in the future, the FAIR guiding principles (Wilkinson et al. 2016) were implemented in the frame of this thesis (Gonzalez et al. 2018b; Philipp et al. 2019). For this purpose, the e!DAL Plant Genomics and Phenomics Research Data Repository (PGP) hosted by the IPK was used as infrastructure for data and metadata publication (Arend et al. 2016). In summary, the data in its three stages: raw, corrected, and ready-to-use, were deposited in the PGP repository together with corresponding R-software scripts for outlier correction and estimation of BLUEs (Fig. 1). In this way, good data management leads to data and knowledge integration beyond the scope of this thesis and allows the maximum benefit to be derived from global research investments (Wilkinson et al. 2016). In order to activate genebank collections for human needs, the FAIR principles could be the common guideline for the worldwide integration of accession related data. Filling public repositories with these data would enable modern biodiversity informatics to systematically exploit the long-observed treasured biodiversity for crop improvement. These would be the steps to be taken to develop conservation-driven genebanks into bio-digital resource centers. (Fig. 2).



**Fig. 2** Extending conservation-driven genebanks towards bio-digital resource centers. Genebanks preserve a huge treasure of biodiversity, but identifying valuable diversity for human needs is like looking for a needle in a haystack. For the systematic exploitation of genebank collections traditional passport information have to be complemented by comprehensive phenotypic and genomic characterization of the hosted accessions. In order to warrant the maximum benefit of characterization investments a sustainable data integration following the FAIR guiding principles [Findability, Accessibility, Interoperability and Reusability (Wilkinson et al. 2016)] is needed. Filling public repositories with common-standard data enables modern biodiversity informatic tools to populate a bio-digital resource center and harnessing the full power of preserved biodiversity. Consequently, in addition to the seed samples, the genebank user also gets access to encyclopaedic information about the accessions of interest.

## 4. Final outlook

Seed regeneration at genebanks is an ongoing process. Each new regeneration cycle is accompanied by the production of new phenotypic data, which potentially could complement the existing data and further improve their quality. In addition, the improved characterization of genebank collections could be accompanied by an increased demand for seed samples, which in turn lead to an increased frequency in seed regeneration and a higher intensity of phenotyping. Therefore, the BLUEs which describe the phenotypic properties of an accession should be updated with new regeneration data on a regular basis. This would continuously improve the characterization of genebank collections. In the case of the IPK genbank, GBIS would be the ideal platform to maintain the BLUEs and make them available to genbank users. GBIS as part of the European (EURISCO) and global (GENESYS) genebank information infrastructure could potentially provide worldwide access to the advances in phenotypic characterization achieved in the present work (Fig. 1). Finally, it should be mentioned that the proposed strategy is neither crop-specific nor trait-specific and can be used by other genebanks as a blueprint for leveraging the use of their historical phenotypic data.

## 5. Summary

The growing human population in connection with changing dietary and consumption behaviors, limited arable land, the loss of biodiversity as well as the climate change and its consequences pose major challenges for future food production and food security. Wheat and barley are among the most important cereal crops that contribute to feeding the world. In order to adapt these crops to future demands the exploitation of genetic resources hosted in genebanks worldwide was suggested. Nevertheless, the lack of sufficient genomic and phenotypic characterization of entire genebank collections hampers the exploitation of this preserved biodiversity.

The economically most relevant traits, in particular grain yield, are quantitative inherited. Identifying beneficial genetic resources based on the evaluation of their *per se* grain yield performance is less promising, due to a huge yield gap compared to the elite material, biased yield measures in the course of missing adaptation, and the cost factor for the multi-environmental evaluation of entire genebank collections. In contrast, the evaluation of genetic resources in terms of their yield components would be a much cheaper and less biased approach to exploit genebank collections for useful diversity. A case study revealed that in particular increasing the number of grains per spikelet had an impact on improving grain yield during the winter wheat breeding in Europe.

In order to augment the amount of phenotypic information of genebank collections, a strategy was developed to analyze highly non-orthogonal historical phenotypic data gathered during 70 years of seed regeneration in an *ex situ* collection of wheat and barley. This strategy consists of a multi-level outlier correction, which is supplemented by comprehensive quality assessment and validation routines. The efforts resulted in high quality, ready-to-use phenotypic data for flowering time, plant height, and thousand grain weight for a total of more than 25,600 wheat and barley accessions. This corresponds to more than 60% of the wheat and barley collection of the German Federal *ex-situ* Genebank of Agricultural and Horticultural Crop Species and forms the basis for an informed choice of genebank accessions for breeding and research.

Finally, the FAIR (findable, accessible, interoperable, reusable) data guiding principles were implemented to ensure the sustainable use of the historical data in the future. In principle, raw data, outlier corrected data, aggregated ready-to-use data as well as the

corresponding R-software scripts for reproduction of the results were filed in a public data repository, whereby community standards for data formatting have been respected. Following the FAIR principles as common standard, good data management could lead to the integration of characterization data across genebanks worldwide. This would promote the extension of conservation-driven genebanks towards bio-digital resource centers and allows the maximum benefit to be derived from global research investments.

## 6. Zusammenfassung

Die wachsende menschliche Bevölkerung in Verbindung mit veränderten Ernährungs- und Konsumgewohnheiten, begrenzte Anbauflächen, der Verlust der biologischen Vielfalt sowie der Klimawandel und seine Folgen stellen große Herausforderungen für die zukünftige Nahrungsmittelproduktion und -sicherheit dar. Weizen und Gerste gehören zu den wichtigsten Getreidepflanzen, die zur Welternährung beitragen. Um diese Nutzpflanzen an die zukünftigen Anforderungen anzupassen, wurde die Nutzung von genetischen Ressourcen vorgeschlagen, die weltweit in Genbanken erhalten werden. Allerdings erschwert die unzureichende genomische und phänotypische Charakterisierung ganzer Genbanksammlungen die Nutzung dieser vorhandenen Biodiversität.

Die ökonomisch relevantesten Merkmale, insbesondere der Kornertrag, werden quantitativ vererbt. Die Identifizierung nützlicher genetischer Ressourcen auf der Grundlage der Bewertung ihrer *per se* Ertragsleistung ist wenig vielversprechend, da die Ertragslücke im Vergleich zum Elitematerial groß ist, die Ertragsmessungen im Zuge fehlender Anpassung verzerrt sind und die Kosten für die umweltübergreifende Bewertung ganzer Genbanksammlungen zu erheblich sind. Im Gegensatz dazu wäre die Bewertung der genetischen Ressourcen in Bezug auf ihre Ertragskomponenten ein viel billigerer und weniger verzerrter Ansatz, um Genbanksammlungen für nützliche Vielfalt zu durchsuchen. Eine Fallstudie zeigte, dass insbesondere die Erhöhung der Anzahl der Körner pro Ährchen einen Einfluss auf die Verbesserung des Kornertrags während der Winterweizenzüchtung in Europa hatte.

Um die Menge der phänotypischen Informationen der Genbanksammlungen zu erhöhen, wurde eine Strategie entwickelt, um hochgradig nicht-orthogonale historische phänotypische Daten zu analysieren, die während 70 Jahren der Saatgutregeneration in einer *ex situ* Sammlung von Weizen und Gerste erzeugt wurden. Diese Strategie besteht aus einer mehrstufigen Ausreiserkorrektur, die durch umfassende Qualitätsbewertungs- und Validierungsmethoden ergänzt wird. Die Bemühungen führten zu qualitativ hochwertigen, gebrauchsfertigen phänotypischen Daten für Blühzeitpunkt, Pflanzenhöhe und Tausendkorngewicht für insgesamt mehr als 25.600 Weizen- und Gerstenakzessionen. Dies entspricht mehr als 60% der Weizen- und Gerstensammlung der bundeszentralen Ex-situ-

Genbank für landwirtschaftliche und gärtnerische Kulturpflanzenarten und bildet die Grundlage für eine fundierte Auswahl von Akzessionen für Züchtung und Forschung.

Schließlich wurden die FAIR (aus dem Englischen: findable, accessible, interoperable, reusable) Datenleitlinien implementiert, um die nachhaltige Nutzung der historischen Daten in Zukunft zu gewährleisten. Im Prinzip wurden Rohdaten, ausreiserkorrigierte Daten, aggregierte Fertigdaten sowie die entsprechenden R-Software-Skripte zur Reproduktion der Ergebnisse in einem öffentlichen Datenspeicher abgelegt, wobei die Gemeinschaftsstandards für die Datenformatierung eingehalten wurden. Durch die Befolgung der FAIR-Prinzipien als gemeinsamen Standard könnte ein gutes Datenmanagement zur Integration von Charakterisierungsdaten über Genbanken weltweit führen. Dies würde die Erweiterung von erhaltungs-orientierten Genbanken in Richtung bio-digitaler Ressourcenzentren fördern und den maximalen Nutzen aus den globalen Forschungsinvestitionen ziehen.



## 7. General references

- Ahokas H (1979) Cytoplasmic Male Sterility in Barley. *Acta Agriculturae Scandinavica* 29:219-224
- Arend D, Junker A, Scholz U, Schüler D, Wylie J, Lange M (2016) PGP repository: a plant phenomics and genomics data publication infrastructure. *Database: the journal of biological databases and curation* 2016
- Börner A (2006) Preservation of plant genetic resources in the biotechnology era. *Biotechnology Journal* 1:1393-1404
- Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, Forrest K, Saintenac C, Brown-Guedira GL, Akhunova A, See D, Bai G, Pumphrey M, Tomar L, Wong D, Kong S, Reynolds M, da Silva ML, Bockelman H, Talbert L, Anderson JA, Dreisigacker S, Baenziger S, Carter A, Korzun V, Morrell PL, Dubcovsky J, Morell MK, Sorrells ME, Hayden MJ, Akhunov E (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences* 110:8057-8062
- Cohen JJ, Williams JT, Plucknett DL, Shands H (1991) Ex situ conservation of plant genetic resources: global development and environmental concerns. *Science* 253:866-872
- Crossa J, Jarquín D, Franco J, Pérez-Rodríguez P, Burgueño J, Saint-Pierre C, Vikram P, Sansaloni C, Petrolci C, Akdemir D, Sneller C, Reynolds M, Tattaris M, Payne T, Guzman C, Peña RJ, Wenzl P, Singh S (2016) Genomic prediction of gene bank wheat landraces. *G3: Genes|Genomes|Genetics* 6:1819-1834
- Ćwiek-Kupczyńska H, Altmann T, Arend D, Arnaud E, Chen D, Cornut G, Fiorani F, Frohberg 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
- Dawson IK, Russell J, Powell W, Steffenson B, Thomas WTB, Waugh R (2015) Barley: a translational model for adaptation to climate change. *New Phytologist* 206:913-931
- de Carvalho MAA, Bebeli PJ, Bettencourt E, Costa G, Dias S, Santos TMM, Slaski JJ (2013) Cereal landraces genetic resources in worldwide GeneBanks. A review. *Agronomy for Sustainable Development* 33:177-203
- Falk DE (2010) Generating and maintaining diversity at the elite level in crop breeding: "Exploiting genome-wide association in oilseed brassicas: a model for genetic improvement of major OECD crops for sustainable farming". *Genome* 53:982-991
- FAO (2010) The second report on the state of the world's plant genetic resources for food and agriculture. Available at: <http://www.fao.org/3/i1500e/i1500e.pdf>, 2019-07-20.
- FAO (2019) Food outlook: Biannual report on global food markets. Report of May. Available at: <http://www.fao.org/3/ca4526en/ca4526en.pdf>, 2019-07-20.
- GENESYS (2019) Global portal on plant genetic resources. Available at: <http://www.genesys-pgr.org>, 2019-07-20.
- Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C (2010) Food Security: the challenge of feeding 9 billion people. *Science* 327:812-818
- González MY, Philipp N, Schulthess AW, Weise S, Zhao Y, Börner A, Oppermann M, Graner A, Reif JC (2018a) Unlocking historical phenotypic data from an ex situ collection to enhance the informed utilization of genetic resources of barley (*Hordeum* sp.). *Theoretical and Applied Genetics* 131:2009-2019
- González MY, Weise S, Zhao Y, Philipp N, Arend D, Börner A, Oppermann M, Graner A, Reif JC, Schulthess AW (2018b) Unbalanced historical phenotypic data from seed regeneration of a barley ex situ collection. *Scientific Data* 5:180278

- Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156:1-13
- Hedden P (2003) The genes of the Green Revolution. *Trends in Genetics* 19:5-9
- Keilwagen J, Kilian B, Ozkan H, Babben S, Perovic D, Mayer KFX, Walther A, Poskar CH, Ordon F, Eversole K, Börner A, Ganai M, Knüpfper H, Graner A, Friedel S (2014) Separating the wheat from the chaff - a strategy to utilize plant genetic resources from ex situ genebanks. *Scientific Reports* 4:6
- Kilian B, Graner A (2012) NGS technologies for analyzing germplasm diversity in genebanks. *Briefings in functional genomics* 11:38-50
- Krajewski P, Chen D, Ćwiek H, van Dijk ADJ, Fiorani F, Kersey P, Klukas C, Lange M, Markiewicz A, Nap JP, 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. *Journal of Experimental Botany* 66:5417-5427
- Lehmann CO, Mansfeld R (1957) Zur Technik der Sortimentserhaltung. *Die Kulturpflanze* 5:108-138
- Longin C, Mühleisen J, Maurer H, Zhang H, Gowda M, Reif J (2012) Hybrid breeding in autogamous cereals. *Theoretical and Applied Genetics* 125:1087-1096
- Lopes MS, El-Basyoni I, Baenziger PS, Singh S, Royo C, Ozbek K, Aktas H, Ozer E, Ozdemir F, Manickavelu A, Ban T, Vikram P (2015) Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *Journal of Experimental Botany* 66:3477-3486
- Lopes MS, Reynolds MP, Jalal-Kamali MR, Moussa M, Feltaous Y, Tahir ISA, Barma N, Vargas M, Mannes Y, Baum M (2012) The yield correlations of selectable physiological traits in a population of advanced spring wheat lines grown in warm and drought environments. *Field Crops Research* 128:129-136
- Mansfeld R (1950) Das morphologische System der Saatgerste, *Hordeum vulgare* L.s.l. *Der Züchter* 20:8-24
- Mansfeld R (1951) Das morphologische System des Saatweizens, *Triticum aestivum* L.s.l. *Der Züchter* 21:41-60
- Mascher M, Schreiber M, Scholz U, Graner A, Reif JC, Stein N (2019) Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nature Genetics* 51:1076-1081
- McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, Burke JM, Charest D, Cloutier S, Cole G, Dempewolf H, Dingkuhn M, Feuillet C, Gepts P, Grattapaglia D, Guarino L, Jackson S, Knapp S, Langridge P, Lawton-Rauh A, Lijua Q, Lusty C, Michael T, Myles S, Naito K, Nelson RL, Pontarollo R, Richards CM, Rieseberg L, Ross-Ibarra J, Rounsley S, Hamilton RS, Schurr U, Stein N, Tomooka N, van der Knaap E, van Tassel D, Toll J, Valls J, Varshney RK, Ward J, Waugh R, Wenzl P, Zamir D (2013) Feeding the future. *Nature* 499:23
- McKevith B (2004) Nutritional aspects of cereals. *Nutrition Bulletin* 29:111-142
- Merfert W, Schilowa I, Faustmann M (1988) Stand der Hybridweizenzüchtung in der DDR. *Plant Breeding* 101:236-244
- Milner SG, Jost M, Taketa S, Mazón ER, Himmelbach A, Oppermann M, Weise S, Knüpfper H, Basterrechea M, König P, Schüler D, Sharma R, Pasam RK, Rutten T, Guo G, Xu D, Zhang J, Herren G, Müller T, Krattinger SG, Keller B, Jiang Y, González MY, Zhao Y, Habekuß A, Färber S, Ordon F, Lange M, Börner A, Graner A, Reif JC, Scholz U, Mascher M, Stein N (2019) Genebank genomics highlights the diversity of a global barley collection. *Nature Genetics* 51:319-326
- Mondal S, Rutkoski JE, Velu G, Singh PK, Crespo-Herrera LA, Guzmán C, Bhavani S, Lan C, He X, Singh RP (2016) Harnessing diversity in wheat to enhance grain yield, climate resilience, disease and insect pest resistance and nutrition through conventional and modern breeding approaches. *Frontiers in Plant Science* 7
- Mühleisen J, Maurer HP, Stiewe G, Bury P, Reif JC (2013) Hybrid breeding in barley. *Crop Sci* 53:819-824

- Oppermann M, Weise S, Dittmann C, Knüpffer H (2015) GBIS: the information system of the German Genebank. Database 2015
- Philipp N, Weichert H, Bohra U, Weschke W, Schulthess AW, Weber H (2018a) Grain number and grain yield distribution along the spike remain stable despite breeding for high yield in winter wheat. PLOS ONE 13:e0205452
- Philipp N, Weise S, Oppermann M, Börner A, Graner A, Keilwagen J, Kilian B, Arend D, Zhao Y, Reif J, Schulthess AW (2019) Historical phenotypic data from seven decades of seed regeneration in a wheat *ex situ* collection hosted at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK). e!DAL - Plant Genomics and Phenomics Research Data Repository (PGP), IPK Gatersleben, Seeland OT Gatersleben, Corrensstraße 3, 06466, Germany
- Philipp N, Weise S, Oppermann M, Börner A, Graner A, Keilwagen J, Kilian B, Zhao Y, Reif JC, Schulthess AW (2018b) Leveraging the use of historical data gathered during seed regeneration of an *ex situ* genebank collection of wheat. Frontiers in Plant Science 9
- Purugganan MD (2019) Evolutionary insights into the nature of plant domestication. Current Biology 29:R705-R714
- Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. PLoS One 8:e66428
- Rocca-Serra P, Brandizi M, Maguire E, Sklyar N, Taylor C, Begley K, Field D, Harris S, Hide W, Hofmann O, Neumann S, Sterk P, Tong W, Sansone S-A (2010) ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. Bioinformatics 26:2354-2356
- 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 CT, Forster MJ, Gaudet P, Gilbert J, Goble C, Griffin JL, Jacob D, Kleinjans J, Harland L, Haug K, Hermjakob H, Sui SJH, Laederach A, Liang S, Marshall S, McGrath A, Merrill E, Reilly D, Roux M, Shamu CE, Shang CA, Steinbeck C, Trefethen A, Williams-Jones B, Wolstencroft K, Xenarios I, Hide W (2012) Toward interoperable bioscience data. Nature Genetics 44:121
- Sehgal D, Vikram P, Sansaloni CP, Ortiz C, Pierre CS, Payne T, Ellis M, Amri A, Petroli CD, Wenzl P, Singh S (2015) Exploring and mobilizing the gene bank biodiversity for wheat improvement. PLOS ONE 10:e0132112
- Semenov MA, Stratonovitch P, Alghabari F, Gooding MJ (2014) Adapting wheat in Europe for climate change. Journal of Cereal Science 59:245-256
- Simmonds J, Scott P, Leverington-Waite M, Turner AS, Brinton J, Korzun V, Snape J, Uauy C (2014) Identification and independent validation of a stable yield and thousand grain weight QTL on chromosome 6A of hexaploid wheat (*Triticum aestivum* L.). BMC Plant Biology 14:191
- Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063-1066
- Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, Consortium IWGS, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo M-C, Dvorak J, Morell M, Dubcovsky J, Ganai M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnology Journal 12:787-796
- Weise S, Oppermann M, Maggioni L, van Hintum T, Knüpffer H (2017) EURISCO: The European search catalogue for plant genetic resources. Nucleic acids research 45:D1003-D1008
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer

- AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, 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. *Scientific Data* 3:160018
- Worland AJ (1996) The influence of flowering time genes on environmental adaptability in European wheats. *Euphytica* 89:49-57
- Wrigley C, Batey I, Miskelly D (2017) *Cereal grains: assessing and managing quality*, 2nd edn. Woodhead Publishing, Cambridge
- Würschum T, Leiser WL, Langer SM, Tucker MR, Longin CFH (2018) Phenotypic and genetic analysis of spike and kernel characteristics in wheat reveals long-term genetic trends of grain yield components. *Theoretical and Applied Genetics* 131:2071-2084
- Yu X, Li X, Guo T, Zhu C, Wu Y, Mitchell SE, Roozeboom KL, Wang D, Wang ML, Pederson GA, Tesso TT, Schnable PS, Bernardo R, Yu J (2016) Genomic prediction contributing to a promising global strategy to turbocharge gene banks. *Nature Plants* 2:16150

## 8. List of general abbreviations

Abbreviation	Explanation
BLUE	Best Linear Unbiased Estimate
CGIAR	Consultative Group on International Agricultural Research
CIMMYT	Maize and Wheat Improvement Center, Mexico
CMS	Cytoplasmatic Male Sterility
EURISCO	European Search Catalogue for Plant Genetic Resources
FAIR	Findable, Accessible, Interoperable, Reusable: data guiding principles
FAO	Food and Agriculture Organisation
GBIS	Genebank Information System
GENESYS	Global Portal on Plant Genetic Resources
GRIN	Germplasm Resources Information Network
GYDAS	Grain Yield Distribution along the Spike
ICGR-CAAS	Institute of Crop Germplasm Resources, the Chinese Academy of Agricultural Sciences; China
IPK	Leibniz Institute of Plant Genetics and Crop Plant Research
IRRI	International Rice Research Institute, Philippines
ISA-Tab	Investigation-Study-Assay: data and metadata format
MIAPPE	Minimum Information About a Plant Phenotyping Experiment
NBPGR	National Bureau of Plant Genetic Resources, India
NCPGR	National Center of Genetic Resources Preservation, United States of America
NRP	Normalized Rank Product
PGP	Genomics and Phenomics Research Data Repository
PGRC	Plant Gene Resources of Canada, Canada
SINGER	System-wide Information Network of Genetic Resources
VIR	N.I. Vavilov Research Institute of Plant Industry, Russia

## 9. Acknowledgements

First of all I would like to thank my supervisor Prof. Dr. Jochen C. Reif. Thank you, Jochen, for providing this interesting and challenging topic to me and for your continuous support and motivation. Thanks for the long discussions, the great ideas and the freedom I got for all my other projects besides this thesis.

Further, my sincere gratitude goes to all members of the Quantitative Genetics group, who have supported me in my research. Special thanks to Dr. Yusheng Zhao, Dr. Albert W. Schulthess and Maria Y. González for the intensive support during data analysis and the publication of the results. Gratitude is also owed to the technical team. In particular, I want to thank Christoph Martin, Andrea Kunze, Bettina Brückner, Jelena Perovic, Sonja Gentz and Martina Kühne for the pleasant cooperation.

Great thanks also to all co-authors who contributed to the publications in the frame of this work. Especially, I want to thank Andreas Graner, Andreas Börner, Stephan Weise and Markus Oppermann from the Department of Genebank, who made the work with the historical data possible, and Michael Grau for the deep insights into seed regeneration. Many thanks to Daniel Arend for the support with the data publications. Thanks also to Winfriede Weschke, Hans Weber and Heiko Weichert from the Department of Molecular Genetics for the scientific excursion into plant physiology.

Last but not least I want to thank my family and friends for their support during the course of my PhD thesis. I wouldn't have gotten this far without you.

## 10. Curriculum vitae

### Personal information

---

Name: Norman Philipp  
 Date of birth: 07.08.1991  
 Place of birth: Schleiz, Germany  
 Nationality: German

### Work experience

---

Since 04/2019 **Hybrid wheat breeder** at Syngenta Seeds GmbH, Hadmersleben, Germany

10/2015 - 03/2019 **Research assistant** at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany; Department of Breeding Research; Research group: Quantitative Genetics

### Education

---

Since 10/2015 **PhD Student** at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany; Department of Breeding Research; Research group: Quantitative Genetics

11/2015 Graduation as **M.Sc. Nutzpflanzenwissenschaften** at Martin-Luther-University Halle-Wittenberg (MLU), Germany; specialization in plant breeding; Master thesis: *“Genomic Prediction of Barley Hybrid Performance”*

10/2013 Graduation as **B.Sc. Agrarwissenschaften** at Martin-Luther-University Halle-Wittenberg, Germany; specialization in plant science; Bachelor thesis: *“Die genetische Kontrolle der Blattmorphologie in der Gersten-Nested Association Mapping (NAM) Population HEB-4”*

6/2010 **Abitur** at Staatl. Gymnasium Dr. Konrad Duden, Schleiz, Germany

### Honors

---

11/2015 **Award: “Best graduate”** of the year 2015 in crop science by MLU

10/2013 **Scholarship: “Plant breeding”** by Syngenta Seeds GmbH, MLU and IPK

## 11. List of publications

- Chu J, Zhao Y, Beier S, Schulthess AW, Stein N, **Philipp N**, Röder MS, Reif JC (2020) Suitability of SNP arrays versus genotyping-by-sequencing for genebank genomics in wheat. *Frontiers in Plant Science* DOI: 10.3389/fpls.2020.00042
- Longin F, Beck H, Gütler H, Heilig W, Kleinert M, Rapp M, **Philipp N**, Erban A, Brillhaus D, Mettler-Altmann T, Stich B (2020) Aroma and quality of breads baked from old and modern wheat varieties and their prediction from genomic and flour-based metabolite profiles. *Food Research International* 129:108748
- Philipp N**, Weise S, Oppermann M, Börner A, Keilwagen J, Kilian B, Arend D, Zhao Y, Graner A, Reif JC, Schulthess AW (2019) Historical phenotypic data from seven decades of seed regeneration in a wheat ex situ collection. *Scientific Data* 6:137
- Gonzalez MY, Weise S, Zhao Y, **Philipp N**, Arend D, Börner A, Oppermann M, Graner A, Reif JC, Schulthess AW (2018) Unbalanced historical phenotypic data from seed regeneration of a barley ex situ collection. *Scientific Data* 5:180278
- Philipp N**, Weichert H, Bohra U, Weschke W, Schulthess AW, Weber H (2018) Grain number and grain yield distribution along the spike remain stable despite breeding for high yield in winter wheat. *PLOS ONE* 13:e0205452
- González MY, **Philipp N**, Schulthess AW, Weise S, Zhao Y, Börner A, Oppermann M, Graner A, Reif JC (2018) Unlocking historical phenotypic data from an ex situ collection to enhance the informed utilization of genetic resources of barley (*Hordeum* sp.). *Theoretical and Applied Genetics* 131:2009-2019
- Philipp N**, Weise S, Oppermann M, Börner A, Graner A, Keilwagen J, Kilian B, Zhao Y, Reif JC, Schulthess AW (2018) Leveraging the Use of Historical Data Gathered During Seed Regeneration of an ex Situ Genebank Collection of Wheat. *Frontiers in Plant Science* 9
- Li Z, **Philipp N**, Spiller M, Stiewe G, Reif JC, Zhao Y (2017) Genome-Wide Prediction of the Performance of Three-Way Hybrids in Barley. *The Plant Genome* 10
- Philipp N**, Liu G, Zhao Y, He S, Spiller M, Stiewe G, Pillen K, Reif JC, Li Z (2016) Genomic Prediction of Barley Hybrid Performance. *The Plant Genome* 9



## 12. Eidesstattliche Erklärung / Declaration under Oath

Ich erkläre, an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. / I declare under penalty of perjury that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.

---

Datum / Date

---

Unterschrift / Signature

### **13. Erklärung über bestehende Vorstrafen und anhängige Ermittlungsverfahren / Declaration concerning Criminal Record and Pending Investigations**

Hiermit erkläre ich, dass ich weder vorbestraft bin noch dass gegen mich Ermittlungsverfahren anhängig sind. / I hereby declare that I have no criminal record and that no preliminary investigations are pending against me.

---

Datum / Date

---

Unterschrift / Signature