



OPEN

SUBJECT AREAS:
PLANT GENETICS
PLANT SCIENCES

Received
29 November 2013

Accepted
14 May 2014

Published
10 June 2014

Correspondence and
requests for materials
should be addressed to
B.K. (benjamin.kilian@
bayer.com)

* Current address:
Bayer CropScience
NV, Innovation Center,
BCS R&D, Trait
Research,
Technologiepark 38,
9052 Zwijnaarde
(Gent), Belgium.

† Current address:
BASF SE, 67117
Limburgerhof,
Germany.

Separating the wheat from the chaff – a strategy to utilize plant genetic resources from *ex situ* genebanks

Jens Keilwagen^{1,2}, Benjamin Kilian^{3*}, Hakan Özkan⁴, Steve Babben⁵, Dragan Perovic⁵, Klaus F. X. Mayer⁶, Alexander Walther⁷, C. Hart Poskar⁸, Frank Ordon⁵, Kellye Eversole⁹, Andreas Börner³, Martin Ganai¹⁰, Helmut Knüpfper³, Andreas Graner³ & Svetlana Friedel^{1†}

¹Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466 Seeland OT Gatersleben, Germany, ²Institute for Biosafety in Plant Biotechnology, Julius Kühn-Institut (JKI) - Federal Research Centre for Cultivated Plants, D-06484 Quedlinburg, Germany, ³Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466 Seeland OT Gatersleben, Germany, ⁴University of Çukurova, Faculty of Agriculture, Department of Field Crops, 01330 Adana, Turkey, ⁵Institute for Resistance Research and Stress Tolerance, Julius Kühn-Institut (JKI) - Federal Research Centre for Cultivated Plants, D-06484 Quedlinburg, Germany, ⁶MIPS/Institute for Bioinformatics and Systems Biology, Helmholtz Center Munich, D-85764 Neuherberg, Germany, ⁷University of Gothenburg, Department of Earth Sciences, SE-405 30 Göteborg, Sweden, ⁸Physiology and Cell Biology, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466 Seeland OT Gatersleben, Germany, ⁹International Wheat Genome Sequencing Consortium, Eversole Associates, MD 20816 Bethesda, USA, ¹⁰TraitGenetics GmbH, Am Schwabeplan 1b, D-06466 Seeland OT Gatersleben, Germany.

The need for higher yielding and better-adapted crop plants for feeding the world's rapidly growing population has raised the question of how to systematically utilize large genebank collections with their wide range of largely untouched genetic diversity. Phenotypic data that has been recorded for decades during various rounds of seed multiplication provides a rich source of information. Their usefulness has remained limited though, due to various biases induced by conservation management over time or changing environmental conditions. Here, we present a powerful procedure that permits an unbiased trait-based selection of plant samples based on such phenotypic data. Applying this technique to the wheat collection of one of the largest genebanks worldwide, we identified groups of plant samples displaying contrasting phenotypes for selected traits. As a proof of concept for our discovery pipeline, we resequenced the entire major but conserved flowering time locus *Ppd-D1* in just a few such selected wheat samples – and nearly doubled the number of hitherto known alleles.

Climate change and the rapidly growing human population trigger the demand for significantly improved crop plants^{1–7}. At present, new plant varieties are mainly developed through the reshuffling of alleles present in the elite gene pool (crossing the best – hoping for the best) resulting in a more or less constant repertoire of alleles or even an erosion of genetic diversity. As a result, genetic gains have gradually slowed for most of the major crop species including wheat^{8,9}. As a result, the relative increase in grain yield of wheat (*Triticum aestivum* L.), the most dominant crop species with a global acreage of 217 million ha (<http://faostat.fao.org/>), has fallen short of the global population growth rate in many parts of the world.

One of the most promising approaches to cope with this challenge is to valorize the rich genetic diversity conserved in large germplasm collections stored *ex situ* in genebanks for trait improvement by the introduction of novel alleles^{10–13}. However, the identification of plant samples (plant samples commonly referred to as ‘accessions’ in a genebank context) with specific combinations of traits from genebank collections is inherently difficult, because of varying regeneration intervals (due to improved storage conditions) and owing to the fact that collections were not regenerated in a systematic scheme (i.e. individual regeneration cycles did not comprise the same set of accessions). Furthermore, changes in conservation management over time (evolution of agricultural practices, equipment, storage and regeneration strategies, and genebank standards) impinge on the usefulness of phenotypic observations. Finally, environmental effects including inter-annual weather variability and long-term climate change affect plant phenology^{4,14–18}. This is reflected by a shift of up to three weeks towards earlier flowering within the last 60 years, correlated with a significant increase of average spring temperatures over



this period (Supplementary Fig. S1, Supplementary Fig. S2). As a result, technical and environmental effects lead to highly heterogeneous phenotypic data which makes valorization of genetic resources a daunting task.

Results and Discussion

The normalized rank product is less biased than other descriptive statistics. Against this backdrop, we aimed at exploiting long-term phenotypic data to enable systematic access to large genebank collections for upcoming genome resequencing and genome-wide association studies (Fig. 1). We propose a statistical approach, which allows for the consolidation of data sets that were collected over sixty years of seed multiplication. The approach aims at the comparison of accessions grown in different years using a standardized scale. This is achieved by calculating the Normalized Rank Product (NRP) to order accessions for a specific trait under consideration (Supplementary Table S1). Thus, accessions grown across multiple decades can be compared for particular traits using a common scale ranging from 0 to 1.

For validation, the NRP was compared against descriptive statistics like the mean, for ranking accessions based on historical data for flowering time. Analyzing this trait by using a 9-year sliding window for the annual median flowering time (Fig. 2A), it becomes obvious that the flowering time depends on the year of cultivation. More

specifically, early flowering was observed predominantly in recent decades (after 1980), while late flowering was observed in early decades. Based on this observation, we compare a varying number of accessions with best mean and NRP values by examining the percentage of accessions that have been cultivated for the first time after 1980 (Fig. 2B). For the mean, the percentage initially is very high and converges with increasing number of selected accessions to the percentage of accessions entered into the collection after 1980, whereas the percentage fluctuates around the global value for the NRP. This observation indicates that the NRP is less influenced by the shift of flowering time.

To further validate the power of the NRP, two winter wheat accessions showing contradicting mean and NRP were scrutinized as an example. A direct comparison was facilitated because of data from 2 years of common cultivation (1998, 2000). In both years, the accession with smaller NRP flowered earlier than the one with smaller mean flowering time (Fig. 2C) again confirming the accuracy of the NRP. Finally, mean and NRP were assessed on all pairs of accessions that have been cultivated together in at least two years as described in the specific example above. Based on this evaluation, the NRP was shown to be less biased than the mean (Supplementary Fig. S3). Thus, the NRP is a robust trait characteristics and can be used for instance for clustering and principal component analysis. As an example, every accession in the collection can be described by more than one normalized trait.

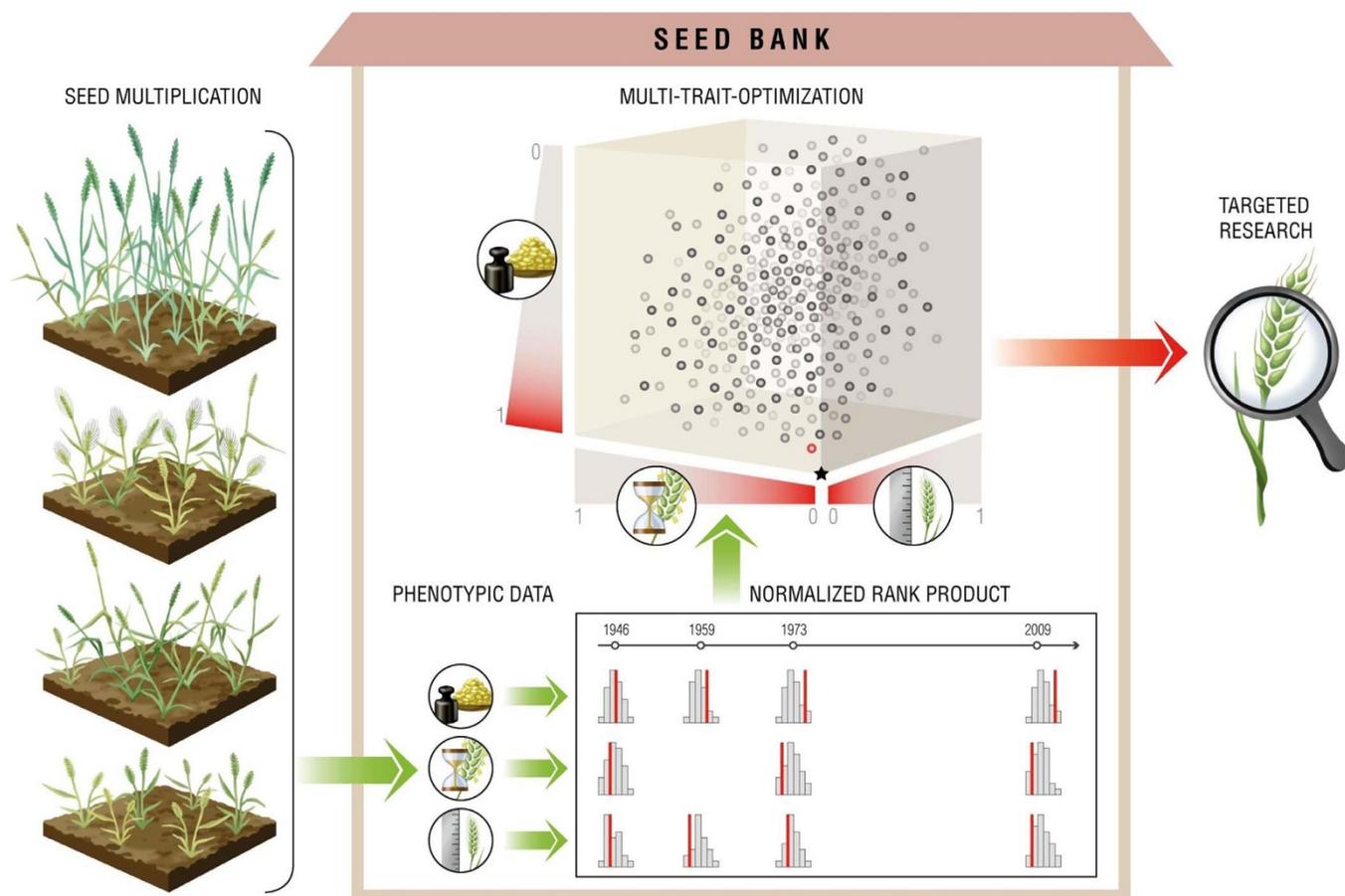


Figure 1 | A novel strategy for valorizing genetic diversity stored in genebanks. Every year genebank curators select accessions for seed multiplication. These accessions are grown in field trials and their phenotypic observations are recorded during the growing period. Annual ranking of these data is the first step in NRP analysis which allows comparing accessions that have been cultivated in different years and under different conditions. This example visualizes a wheat accession cultivated four times since 1946. The histograms indicate the distributions for all accessions cultivated in a given year. Missing histograms indicate missing phenotypic observations. Based on NRPs, MTO allows identifying accessions with specific combinations of traits that can be utilized for targeted plant research and breeding. The black asterisk and red dot in the cube represent the best virtual and one real accession, respectively, that simultaneously have early flowering time, small plant height and high thousand grain weight. In the histograms of absolute trait values, the actual accession is indicated by red lines. This figure has been generated using Adobe Photoshop CC and Adobe Illustrator CC.

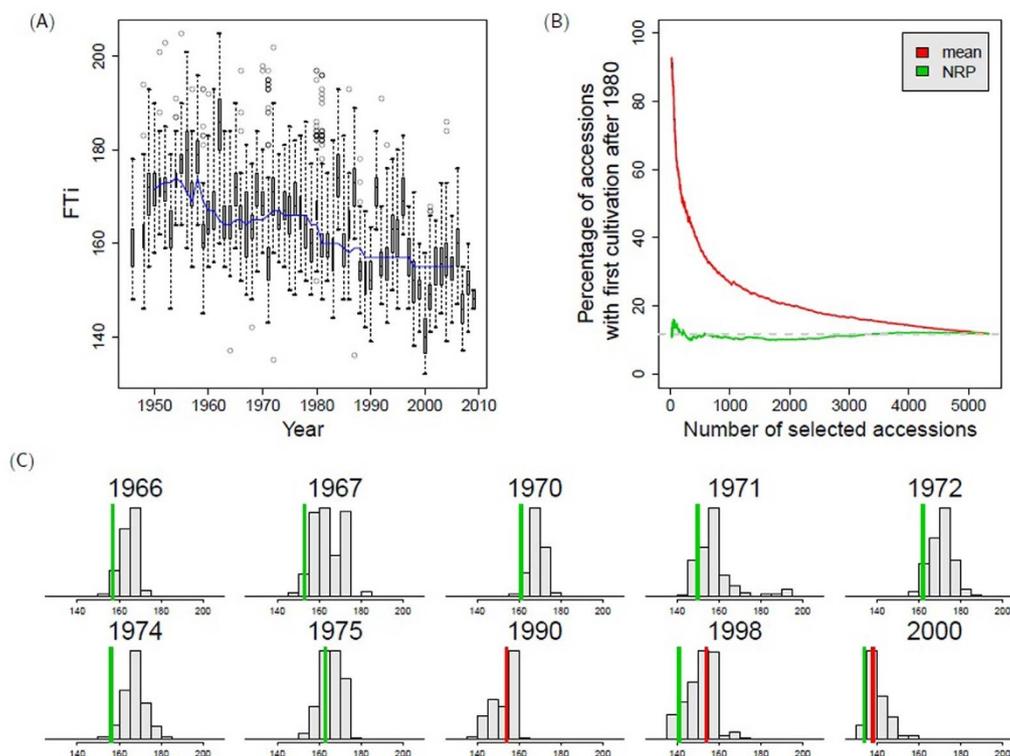


Figure 2 | Validation of the NRP for flowering time of winter wheat. (A) Box plots and trend of flowering time (FTi, in days of year, ordinate) for winter wheat between 1946 and 2010 (abscissa). The trend in blue is the 9-year sliding window for annual median flowering time indicating a clear shift towards earlier flowering. (B) Comparison of NRP with the naïve method of averaging the phenotypic observations (mean). The panel plots the percentage of accessions that have been cultivated for the first time after 1980 against the number of selected accessions. For the mean, the percentage is initially very high and converges to the global value. In contrast, the behavior of the NRP is almost constant. (C) Specific example of early flowering time for two winter wheat accessions. Green visualizes the accession TRI 7594 selected by NRP, while red visualizes the accession TRI 16575 selected by mean. Hence, the red accession has a smaller mean ($149 < 153$) but a higher NRP ($0.51 > 0.06$) than the green accession (Supplementary Table S1). However, the two common cultivations in 1998 and 2000 illustrate that the green accession typically flowers earlier than the red one. Plots were created with R (<http://r-project.org>).

Multi-trait optimization identifies promising accessions for further evaluation. In Fig. 1, the results are presented for three traits, namely, thousand grain weight, flowering time, and plant height. In general if one is interested in n different traits, one might span an n -dimensional hypercube. For the identification of promising accessions for further evaluation, we used a multi-trait optimization (MTO) approach based on the NRP to overcome an intrinsic structure caused by correlations between normalized traits and dependencies on passport data (Supplementary Text A, Supplementary Text B).

This approach selects accessions which are located near the corners of the cube corresponding to the most extreme trait combinations (Fig. 1, Fig. 3A, Supplementary Fig. S4A). Hence, opposing corners represent contrasting phenotypes (Supplementary Fig. S4). For example, the corner (1,0,0) indicated by an asterisk in Fig. 1 refers to the combination of high thousand grain weight, early flowering, and short plant height, whereas the corner (0,1,1) refers to the contrasting combination of low thousand grain weight, late flowering, and tall plant height. Based on that, MTO allows for identifying accessions breaking the existing correlations between traits, i.e., having combinations of traits that only rarely occur in a given gene pool (cf. empty corners in Fig. 3A, Supplementary Fig. S5, Supplementary Fig. S6).

To validate this MTO, we used long-term phenotypic data that were available for nearly 7,000 winter wheat accessions. For further validation, four groups of contrasting phenotypes were selected for extreme plant height and extreme flowering time (Supplementary Table S2) and grown in a field experiment. In total, 60 accessions (15

accessions per contrasting group) were grown together, which allowed for the direct comparison of their phenotypes. Comparison with the legacy data revealed highly similar patterns demonstrating the efficacy of MTO (Fig. 3B, Fig. 3C, Supplementary Table S2).

Proof of concept: resequencing at *Ppd-D1* locus nearly doubled the number of known alleles.

To investigate the potential of trait-based selection for discovering new alleles, all 60 contrasting accessions obtained from MTO (Fig. 3A), represented by 96 individual plants, were resequenced at the major photoperiod response locus *Ppd-D1* (Supplementary Table S4). For this, we utilized sequence information from the International Wheat Genome Sequencing Consortium (IWGSC)^{19–21}. Despite the small sample size and the high degree of conservation at this locus, three novel alleles were identified, nearly doubling the number of haplotypes known for bread wheat at this locus^{22–25} (Supplementary Table S5, Supplementary Fig. S8). Two of the polymorphisms were located in coding regions upstream of the CCT domain leading to premature stop codons and thus most likely cause a loss of function (Supplementary Text C, Supplementary Dataset).

In conclusion, we proposed a procedure to compare results from “non orthogonal” field trials which are a hallmark of the continuous reproduction schemes of genebank collections. The solution, illustrated here for wheat, may be applied to other crops and their wild relatives, making this strategy a standard approach to select for ‘wanted’ phenotypic combinations and to ‘widen’ the breeding bottleneck. This study demonstrates that the enormous amount of

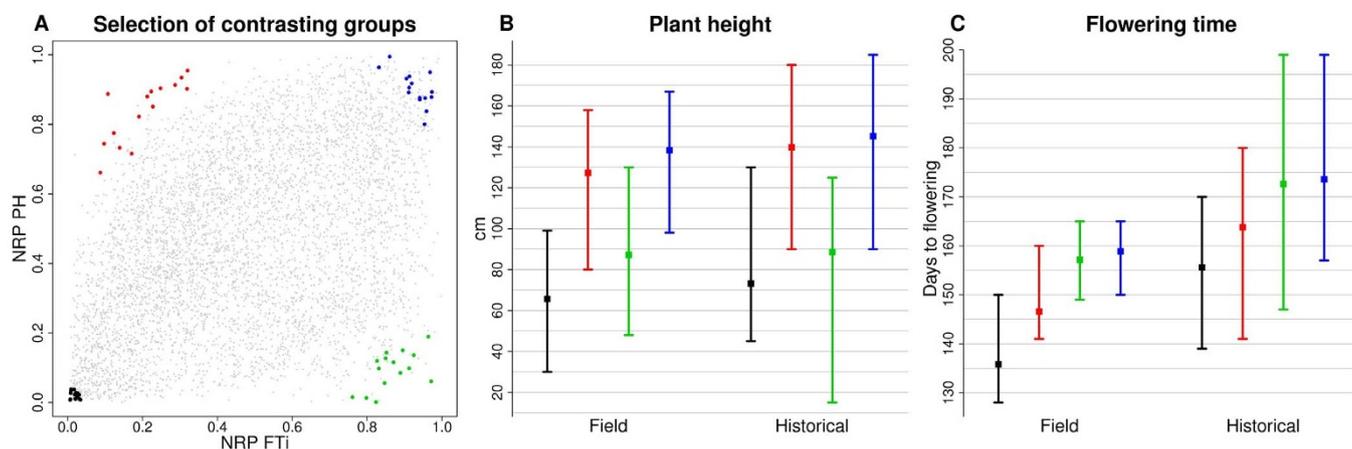


Figure 3 | Validation of MTO for a wheat collection comprising 6,959 accessions (Supplementary Table S6). (A) depicts the results of the MTO for the normalized traits flowering time (NRP FTI) and plant height (NRP PH) selecting four contrasting groups with 15 accessions each. (B) and (C) compare the phenotyping results (minimum, mean and maximum of pre-normalized values) of the field experiment 2010/2011 and the historical phenotypic data (1946–2009) for plant height and flowering time, respectively. In all three panels, the colors encode the selected contrasting groups, where black represents short and early flowering accessions, red represents tall and early flowering, green represents short and late flowering, and blue represents tall and late flowering (Supplementary Fig. S2).

phenotypic data, recorded in genebanks over decades, can indeed be used to select accessions with desired combinations of traits. Computational methods as the NRP can be applied immediately and at almost zero cost - harvesting the investment of time, energy and money made in genebanks over generations and proving their continued value.

Based on NRPs and MTO of long-term phenotypic data, it is feasible to accurately assess the genotypic diversity and identify divergent accessions from comprehensive collections of any size, thereby avoiding the limitations entailed by the use of “core collections”²⁶. Additional information including genotypic and pedigree information, if available, might be used to even better assess the genotypic diversity. Identification of a small set of contrasting phenotypes for flowering time immediately allowed for tapping into a substantial amount of novel allelic diversity at a major flowering time locus in wheat, which can be characterized and validated in future studies. Upcoming genome sequences for major cereal crops and new techniques including whole genome exome capture^{19–21,27} will allow for a much more extensive association of phenotypic and allelic diversity and thus help to bridge the genotype to phenotype gap.

Methods

Federal genebank of Germany. The Federal *ex situ* Genebank for Agricultural and Horticultural Plant Species of Germany maintained at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben hosts one of the largest barley and wheat germplasm collections in the world²⁸. The collection was established at its present location in 1945, and seed multiplication has been continuously performed since 1946.

The cultivation for seed multiplication has taken place every two to three years until 1976. After a cold storage facility was established in 1976, the average frequency of seed multiplication dropped drastically to every 10–30 years and accessions have been selected for seed multiplication if either the germination rate or the amount of available seeds dropped below a critical threshold²⁹.

Characterization and evaluation data. For IPK’s seed bank, Characterization and Evaluation (C & E) data comprises dates of various growth stages (e.g., sowing, emergence, heading, flowering, ripening, harvest), winter survival, lodging, infection scores for some of the most important fungal diseases, plant height, and thousand-grain weight. In this paper, we focused the analysis on three key traits, namely, days to flowering (FTi), plant height (PH), and thousand-grain weight (TGW) for wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.). Spring and winter types (growth habit) are distinguished for both cereals (Note: in Central and Northern Europe, winter types are sown in autumn - September until December -, while spring types are sown in spring - March until April).

Preprocessing of C & E data. Since 1946 the C & E data have been recorded in field books during the annual seed multiplication (Note: Since 2011, Personal Digital

Assistants (PDA) are used for data recording in the field making the repeated transfer of these data unnecessary.) Subsequently, these data have been transferred to main card files. However, the data had to be digitized prior to computational analysis. Thus, the data had been manually transferred twice, namely from field books to main card files and then to the computer, increasing the potential of introducing errors.

Hence we checked for spurious values in the recorded traits, and tested the temporal order of the recorded dates. The data was split into winter and spring types based on their classification of growth habit, commonly referred to as ‘annuality’ among genebanks, in IPK’s Genebank Information System (GBIS, <http://gbis.ipk-gatersleben.de>).

For the identification of potential errors, we subsequently performed an extreme outlier detection using three times the interquartile range. These outliers were checked against the field books and corrected wherever possible. To ensure that any accession provided at most one record per year, we checked for accessions with more than one record in any year. In case of an accession that has been recorded more than once in a year, the records were replaced by a single record with the median of the observations for the subsequent analysis.

For further analysis, the values for plant height (recorded at growth stage Z70 in cm³⁰) and thousand grain weight (in g) were used directly, while we computed time periods in days for the flowering time. Thus flowering time is recorded as day of year, i.e., the difference between January 1st and date of flowering (Z65³⁰). In Tab. S6, we present the amount of data records for barley and wheat classified according to their annuality. After outlier detection, each collection comprised several thousand accessions and up to five times more records.

Normalized rank product on C & E data. Due to the heterogeneity of the data, descriptive statistics cannot easily be utilized for the data analysis. Instead the recorded values of each trait in each year were ranked to overcome the strong effects of annual differences, for instance inter-annual climate variability, climate change or agricultural practice. Subsequently, we introduced the normalized rank product (NRP) that is based on the idea of the popular rank product analysis, originally proposed for selecting differentially expressed genes from microarray experiment data³¹.

Here, we extend the rank product to maintain the relation to the geometric mean of the normalized ranks. The NRP of an accession *a* for a specific trait denoted by NRP_a is defined as,

$$NRP_a = \sqrt[|Y_a|]{\prod_{y \in Y_a} \frac{r_{y,a}}{n_y}}$$

for a specific trait where Y_a is the set of years in which accession *a* has been cultivated and its trait has been recorded, $r_{y,a}$ is the rank of the trait value of accession *a* in year *y*, and n_y is the number of accessions with this trait recorded in year *y*. The NRP ranges between zero (exclusive) and one (inclusive) allowing for comparison between accessions, independent of their number of years of cultivation. The NRPs were computed for the three important traits: flowering time, plant height, and thousand grain weight.

In order to avoid overestimating the influence of accessions that have been infrequently cultivated and exhibiting interesting NRPs by chance, only accessions that have been cultivated and recorded in at least three different years were included in the present study. With this set of assumptions, we were able to compare large parts of the above mentioned collections comprising thousands of accessions.



As an illustrative example, a particular wheat accession has been cultivated in 1946, 1967 and 2009. This plant sample was the first to flower in 1946 and 2009 (rank 1), and the third in 1967 (rank 3). These ranks are divided by the number of all accessions of that species cultivated in each year. The geometric mean is taken over 1946, 1967 and 2009 yielding a NRP for flowering time of this accession ranging between 0 (early flowering) and 1 (late flowering).

Experimental validation under field conditions for winter wheat. Based on the NRP and multi-trait optimization (MTO), we were able to find plant samples with contrasting phenotypes, for example winter wheat with extreme flowering time and plant height (Fig. 3A). To validate the accuracy of this procedure, 60 winter wheat accessions (four groups with 15 accessions each) were selected and planted in a field experiment at Gatersleben (51° 49'N, 11° 16'E) during the growing season 2010/2011. We used plots of size 1.5 m × 2.5 m separated by equally sized plots of winter barley and followed local agricultural practice. In the field experiment, we performed three replicates arranged in three blocks. Each block consisted of one plot per selected accession in random order.

Each plot was planted at a density of 80 seeds per m². Subsequently, 10 individuals per plot – in total 1,800 individuals – were randomly selected and genetically purified as genebank accessions are not expected to be homogenous and must be purified for genomic studies. Thus all 1800 individuals were covered with bags to prevent cross pollination and phenotyped. The development of single seed descends (SSDs) is a procedure widely used in plant breeding³². All individuals were phenotyped for flowering time (Z60³⁰) as well as for plant height (Z70³⁰), and were finally harvested. Leaf samples for DNA isolation were taken from every plant at Z30³⁰.

Based on known geographical origin and pedigree information, a subset of 28 accessions (out of 60 accessions mentioned above) was then subselected (6–8 accessions per contrasting group) for an independent second field experiment in 2011/2012 – in order to validate the data from the experiment in 2010/2011. As in the year before, we used plots of size 1.5 m × 2.5 m separated by equally sized plots of winter barley and followed local agricultural practice. Three replicates were performed, arranged in three blocks. Each block consisted of one plot per selected accession in random order. Each plot was planted at a density of 80 seeds per m². Full plots were phenotyped for flowering time (Z60³⁰) as well as for plant height (Z70³⁰, Fig. S7).

Extraction of genomic DNA. Genomic DNA was isolated from silica-dried single leaves of each line with the Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions.

Genotyping major vernalization genes for wheat. Wheat can be divided into spring and winter habit varieties along with a group of intermediate lines known as facultative varieties. The vernalization-sensitivity (winter habit) alleles, *vrn-A1*, *vrn-B1* and *vrn-D1* are located on chromosomes 5A, 5B and 5D, respectively^{33,34}. Another vernalization gene, *Vrn-3*, located on chromosome 7B, was shown to encode a homolog of *Arabidopsis FT*, and was named *VRN3*³⁵.

Aiming at reducing any bias for flowering time analysis based on growth habit, in total, 60 purified individuals (1 to 3 plants per accession) were randomly selected among the 28 contrasting accessions subselected for the second field experiment (see above) and genotyped at four major vernalization loci (*Vrn-A1*, *Vrn-B1*, *Vrn-D1* and *Vrn-B3*) using allele-specific molecular assays. Following the experimental protocols^{22–24,35–40} (Tab. S3), all accessions were confirmed as true winter wheat harboring only vernalization-sensitive alleles providing strong support for the flowering time analysis.

Resequencing the major photoperiod response locus *Ppd-D1*. Sixty contrasting accessions (represented by 96 plants, 1 to 3 plants per accession) were resequenced at *Ppd-D1*. In order to develop wheat D genome specific primer combinations, sequence information was obtained from the International Wheat Genome Sequencing Consortium (IWGSC, <http://www.wheatgenome.org/>) and NCBI GenBank (DQ885766.1).

The Primer3 online software (Primer3 v. 0.4.0 (<http://frodo.wi.mit.edu/primer3/>)) was used to design primers. Oligonucleotides were purchased from Eurofins MWG Operon, Ebersberg, Germany. One c. 5920 bp genomic region completely covering *Ppd-D1* was amplified by locus-specific PCR primers from genomic DNAs (Tab. S4). 5'-UTR, exons, introns and 3'-UTR were analyzed, as start and stop codon of the reference (DQ885766.1) were located at alignment position 2150 and 5299, respectively. The exons were located at 2150 to 2321, 2439 to 2603, 2712 to 2845, 3041 to 3196, 3648 to 3825, 3932 to 4349, 4432 to 5089, 5200 to 5301. Specificity and chromosomal localization of PCR products were confirmed by Nulli-tetrasomic (NT) lines (N2A-T2B, N2B-T2D and N2D-T2B)⁴¹ (Fig. S8).

PCR amplification was performed in 20 µl reaction volume. Templates were purified and sequenced directly on both strands on an Applied Biosystems (Weiterstadt, Germany) ABI Prism 3730 xL sequencer using BigDye terminators as described in⁴². Their DNA sequences were determined using primers designed for amplification and internal primers (Tab. S4).

Single nucleotide polymorphism (SNP)-detection. DNA sequences were processed with AB DNA Sequencing Analysis Software 5.2 and later manually edited by Sequencher software v5.0 (Gene Codes Corp.). Sequence alignments were generated with MAFFT webserver (<http://www.ebi.ac.uk/Tools/msa/mafft/>) using default parameters except “perform ffts” which was set to “genafpair”. Subsequently, the multiple sequence alignment was manually modified. Filtering sequences with low

quality regions, 67 sequences were contained in the final alignment comprising c. 5920 bp. The heterozygous state at the deletion in the promoter region was indicated by inserting poly-N. In close analogy, the heterozygote state for the transposable element in the first intron was indicated by inserting poly-N.

Allelic haplotypes were defined by DNASP 5.10.01 as described in⁴² (Tab. S5). All identified singletons were confirmed afterwards by additional two independent amplifications and sequencing.

1. Takeda, S. & Matsuoka, M. Genetic approaches to crop improvement: responding to environmental and population changes. *Nat Rev Genet* **9**, 444–457 (2008).
2. Foley, J. A. *et al.* Solutions for a cultivated planet. *Nature* **478**, 337–42 (2011).
3. Esquinas-Alcázar, J. Protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nat Rev Genet* **6**, 946–953 (2005).
4. Fitter, A. H. & Fitter, R. S. R. Rapid changes in flowering time in British plants. *Science* **296**, 1689–1691 (2002).
5. Godfray, H. C. J. *et al.* Food Security: The challenge of feeding 9 billion people. *Science* **327**, 812–818 (2010).
6. Iglesias, A., Quiroga, S. & Diz, A. Looking into the future of agriculture in a changing climate. *Europ Rev Agricult Econ* **38**, 427–447 (2011).
7. Tester, M. & Langridge, P. Breeding Technologies to Increase Crop Production in a Changing World. *Science* **327**, 818–822 (2010).
8. Graybosch, R. A. & Peterson, C. J. Genetic Improvement in Winter Wheat Yields in the Great Plains of North America, 1959–2008. *Crop Sci* **50**, 1882–1890 (2010).
9. Brisson, N. *et al.* Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *Field Crops Res* **119**, 201–212 (2010).
10. Hoisington, D. *et al.* Plant genetic resources: What can they contribute toward increased crop productivity? *Proc Nat Acad Sci* **96**, 5937–5943 (1999).
11. Kilian, B. & Graner, A. NGS technologies for analyzing germplasm diversity in genebanks. *Brief Funct Genom* **11**, 38–50 (2012).
12. Snook, L. K., Dooloo, M. E., Jarvis, A., Scheldeman, X. & Kneller, M. [Crop germplasm diversity: The role of gene bank collections in facilitating adaptation to climate change]. *Crop Adaptation to Climate Change* [Yadav, S. S., Redden, R., Hatfield, J. L., Lotze-Campen, H. & Hall, A. (eds.)] [495–506] (Wiley-Blackwell, 2011).
13. Tanksley, S. D. & McCouch, S. R. Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* **277**, 1063–1066 (1997).
14. Amano, T., Smithers, R. J., Sparks, T. H. & Sutherland, W. J. A 250-year index of first flowering dates and its response to temperature changes. *Proc Royal Soc B-Biol Sci* **277**, 2451–2457 (2010).
15. Cleland, E. E., Chuine, I., Menzel, A., Mooney, H. A. & Schwartz, M. D. Shifting plant phenology in response to global change. *Trends Ecol Evol* **22**, 357–365 (2007).
16. Craufurd, P. Q. & Wheeler, T. R. Climate change and the flowering time of annual crops. *J Exp Bot* **60**, 2529–2539 (2009).
17. Estrella, N., Sparks, T. H. & Menzel, A. Trends and temperature response in the phenology of crops in Germany. *Global Change Biol* **13**, 1737–1747 (2007).
18. Hu, Q., Weiss, A., Feng, S. & Baenziger, P. S. Earlier winter wheat heading dates and warmer spring in the US Great Plains. *Agri Forest Meteorol* **135**, 284–290 (2005).
19. Jia, J. *et al.* *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**, 91–95 (2013).
20. Brenchley, R. *et al.* Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**, 705–710 (2012).
21. Ling, H.-Q. *et al.* Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* **496**, 87–90 (2013).
22. Beales, J., Turner, A., Griffiths, S., Snape, J. W. & Laurie, D. A. A pseudo-response regulator is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). *Theor Appl Genet* **115**, 721–733 (2007).
23. Wilhelm, E. P., Turner, A. S. & Laurie, D. A. Photoperiod insensitive *Ppd-A1a* mutations in tetraploid wheat (*Triticum durum* Desf.). *Theor Appl Genet* **118**, 285–294 (2009).
24. Nishida, H. *et al.* Structural variation in the 5' upstream region of photoperiod-insensitive alleles *Ppd-A1a* and *Ppd-B1a* identified in hexaploid wheat (*Triticum aestivum* L.), and their effect on heading time. *Mol Breed* **1–11** (2012).
25. Shaw, L. M., Turner, A. S., Herry, L., Griffiths, S. & Laurie, D. A. Mutant alleles of *Photoperiod-1* in wheat (*Triticum aestivum* L.) that confer a late flowering phenotype in long days. *PLoS One* **8**, e79459 (2013).
26. Brown, A. H. D. Core collections: a practical approach to genetic resources management. *Genome* **31**, 818–824 (1989).
27. Winfield, M. O. *et al.* Targeted re-sequencing of the allohexaploid wheat exome. *PL Biotech J* **10**, 733–742 (2012).
28. Knüpfper, H. [Triticeae genetic resources in *ex situ* genebank collections]. *Genetics and Genomics of the Triticeae* [Feuillet, C. & Muehlbauer, G. J. (eds.)] [31–79] (Springer, 2009).
29. Börner, A. Preservation of plant genetic resources in the biotechnology era. *Biotechnol J* **1**, 1393–1404 (2006).
30. Zadoks, J. C., Chang, T. T. & Konzak, C. F. Decimal Code for Growth Stages of Cereals. *Weed Res* **14**, 415–421 (1974).
31. Breiting, R., Armengaud, P., Amtmann, A. & Herzyk, P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *Febs Lett* **573**, 83–92 (2004).



32. Acquah, G. *Principles of Plant Genetics and Breeding*. (Wiley, 2012).
33. Worland, T. & Snape, J. [Genetic basis of worldwide wheat varietal improvement]. *The World Wheat Book: A history of wheat breeding*. [Angus, W. J., Bonjean, A. P. & Van Ginkel, M. (eds.)] [61–67] (Lavoisier Publishing, Paris, 2001).
34. Yan, L. [The Flowering Pathway in Wheat]. *Wheat Science and Trade* [Carver, B. F. (ed.)] [57–72] (Wiley-Blackwell, 2009).
35. Yan, L. *et al.* The wheat and barley vernalization gene *VRN3* is an orthologue of FT. *Proc Nat Acad Sci* **103**, 19581–19586 (2006).
36. Bentley, A. R. *et al.* Frequency of photoperiod-insensitive *Ppd-A1a* alleles in tetraploid, hexaploid and synthetic hexaploid wheat germplasm. *Pl Breed* **130**, 10–15 (2011).
37. Diaz, A., Zikhali, M., Turner, A. S., Isaac, P. & Laurie, D. A. Copy number variation affecting the *Photoperiod-B1* and *Vernalization-A1* genes is associated with altered flowering time in wheat (*Triticum aestivum*). *PLoS ONE* **7**, e33234 (2012).
38. Guo, Z. A., Song, Y. X., Zhou, R. H., Ren, Z. L. & Jia, J. Z. Discovery, evaluation and distribution of haplotypes of the wheat *Ppd-D1* gene. *New Phytol* **185**, 841–851 (2010).
39. Yan, L. *et al.* Allelic variation at the *VRN-1* promoter region in polyploid wheat. *Theor Appl Genet* **109**, 1677–1686 (2004).
40. Fu, D. *et al.* Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. *Mol Genet Genom* **273**, 54–65 (2005).
41. Sears, E. R. [Nullisomic-tetrasomic combinations in hexaploid wheat]. *Chromosome Manipulations and Plant Genetics* [Riley, R. & Lewis, K. R. (eds.)] [29–45] (Springer Science+Business Media, LLC, 1966).
42. Rozas, J. DNA sequence polymorphism analysis using DnaSP. *Meth Mol Biol* **537**, 337–350 (2009).

Acknowledgments

We thank M. Grau, M. Mildner, M. Oppermann, G. Schütze, and R. Selbig for technical assistance with the C & E data. We thank P. Abraham, S. Dreißig, B. Dubsky, H. Giraud, H. Harms, C. Kehler, U. Krajewsky, K. Neumann, M. Nix, P. Schreiber, and K. Wolf for excellent assistance in the field experiments. We thank E. Andeden and K. Niedung for DNA analysis and S. Kilian for kindly preparing Figure 1. We are grateful to T. Altmann, B. Bauer, G. Coupland, M. Friedel, W. Gruissem, H. Riegler, M. Röder, R. Schnee, M. Seifert, R.

Sharma, S. Singh, N. Stein, M. Strickert and N. von Wirén for discussions. The authors would like to acknowledge the support given by the Ministry of Culture of Saxony-Anhalt (grant XP3624HP/0606T) to JK and SF, and the German Science Foundation Priority Programme SPP1530 to BK.

Author contributions

J.K. and S.F. developed the basic ideas. A.B. and H.K. provided the legacy data. J.K. analyzed the data. B.K. conducted the field studies. H.Ö. performed the genotyping analyses. K.F.X.M. and K.E. provided sequence information. A.W. analyzed the climate data. S.B., D.P., J.K. and B.K. performed the resequencing analysis and analyzed the haplotypes. J.K., B.K. and S.F. wrote the manuscript. F.O., M.G., H.P. and A.G. provided supplementary information and edited the manuscript. All authors contributed to data analysis and approved the final manuscript.

Additional information

Accession codes. GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) accession numbers for *Ppd-D1* alleles are: KJ47477 (*Ppd-D1a*), KJ47478 (*Ppd-D1b*), KJ47483 (*Ppd-D1c*), KJ47481 (*Ppd-D1d*), KJ47482 (*Ppd-D1e*), KJ47479 (*Ppd-D1f*) and KJ47480 (*Ppd-D1g*). The multiple sequence alignment containing sequence information from 67 individuals belonging to 44 accessions is provided.

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Keilwagen, J. *et al.* Separating the wheat from the chaff – a strategy to utilize plant genetic resources from *ex situ* genebanks. *Sci. Rep.* **4**, 5231; DOI:10.1038/srep05231 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. The images in this article are included in the article's Creative Commons license, unless indicated otherwise in the image credit; if the image is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the image. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>