ARTICLES
https://doi.org/10.1038/s41477-019-0445-5
nature plants

# Breeding improves wheat productivity under contrasting agrochemical input levels

Kai P. Voss-Fels[1,2,11], Andreas Stahl[1,11], Benjamin Wittkop[1,11], Carolin Lichthardt[3], Sabrina Nagler[4], Till Rose[4], Tsu-Wei Chen[3], Holger Zetzsche[5], Sylvia Seddig[6], Mirza Majid Baig[7], Agim Ballvora[7], Matthias Frisch[8], Elizabeth Ross[2], Ben J. Hayes[2], Matthew J. Hayden[9], Frank Ordon[5], Jens Leon[7,10], Henning Kage[4], Wolfgang Friedt[1*], Hartmut Stützel[3*] and Rod J. Snowdon[1*]

**The world cropping area for wheat exceeds that of any other crop, and high grain yields in intensive wheat cropping systems are essential for global food security. Breeding has raised yields dramatically in high-input production systems; however, selection under optimal growth conditions is widely believed to diminish the adaptive capacity of cultivars to less optimal cropping environments. Here, we demonstrate, in a large-scale study spanning five decades of wheat breeding progress in western Europe, where grain yields are among the highest worldwide, that breeding for high performance in fact enhances cultivar performance not only under optimal production conditions but also in production systems with reduced agrochemical inputs. New cultivars incrementally accumulated genetic variants conferring favourable effects on key yield parameters, disease resistance, nutrient use efficiency, photosynthetic efficiency and grain quality. Combining beneficial, genome-wide haplotypes could help breeders to more efficiently exploit available genetic variation, optimizing future yield potential in more sustainable production systems.**

Winter wheat production in western Europe provides one of the highest-yielding staple food sources worldwide; however, realization of yield potential relies heavily on mineral fertilizer and chemical plant protection. Over the past 50 years, breeding of wheat and other major crops has focused primarily on selection for high grain yield in high-input cropping systems, which are frequently associated with ecological sustainability penalties[1–4]. For example, it is estimated that less than 50% of applied nitrogen is recovered in harvested crops[5,6], while up to 68% of global $N_2O$ emissions from croplands are attributed to nitrogen fertilization of wheat, maize and rice[7]. To feed the rapidly growing world population, cereal production must continue to increase[1], but environmental pollution and atmospheric emissions associated with excessive agrochemical inputs must simultaneously be reduced[2,4].

Despite confirmation that modern breeding maintains genetic diversity[8,9], a broadly held popular perception persists that intense selection during breeding of elite cultivars for high performance under optimal cropping conditions has depleted genetic variants required for adaptation to suboptimal environments or reduced-input cropping systems. For example, a recent study[10] suggested that climate resilience in European wheat is declining, based on observations that modern cultivars show more homogeneous responses to climatic uncertainty and variability. Although genetic data were not investigated, these observations were attributed to genetic erosion of allele diversity as a result of breeding[10].

Loss of adaptive diversity to secure crop performance in suboptimal conditions would have serious implications, particularly in the face of increasing political and environmental constraints on agrochemical inputs, along with climatic fluctuations impacting yield stability[11]. However, analyses focusing only on environmental responses, but ignoring genetic effects on yield performance[10], are unsuitable to draw conclusions about breeding potential[12]. To understand adaptive diversity associated with the effects of breeding progress, it is essential to evaluate genetic variance in empirical studies of cultivar performance and contributing traits in diverse, but clearly defined, growth environments.

## Results

**Assessing 50 years of wheat breeding progress under contrasting input scenarios.** In this study, we present one of the largest analyses to date of genetic and phenotypic changes associated with long-term breeding progress in a major global crop. We investigated a panel of elite winter wheat cultivars released during the past 50 years in western Europe, particularly Germany, where mean grain yield per production area ranked consistently among the highest in the world during this period. The panel included many of the most widely grown wheat cultivars during their period of release, including different grain quality classes (Supplementary Table 1). Grain yield, yield components, harvest index, plant height, plant biomass, flowering behaviour, grain quality characteristics, disease resistances and physiological parameters, including leaf area and

[1]Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Giessen, Germany. [2]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Queensland, Australia. [3]Institute of Horticultural Production Systems, Leibniz University Hannover, Hannover, Germany. [4]Department of Agronomy and Crop Science, Christian Albrechts University of Kiel, Kiel, Germany. [5]Julius Kuehn Institute (JKI), Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance, Quedlinburg, Germany. [6]Julius Kuehn Institute (JKI), Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance, Sanitz, Germany. [7]Institute of Crop Science and Resource Conservation, Chair of Plant Breeding, University of Bonn, Bonn, Germany. [8]Institute for Agronomy and Plant Breeding II, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Giessen, Germany. [9]School of Applied Systems Biology, AgriBio, La Trobe University, Melbourne, Victoria, Australia. [10]Field Lab Campus Klein-Altendorf, University of Bonn, Rheinbach, Germany. [11]These authors contributed equally: Kai P. Voss-Fels, Andreas Stahl, Benjamin Wittkop. *e-mail: wolfgang.friedt@agrar.uni-giessen.de; stuetzel@gem.uni-hannover.de; rod.snowdon@agrar.uni-giessen.de

photosynthetic potential, were assessed in large-scale field trials, in which cultivars representing five decades of breeding progress were planted side by side in experiments spanning multiple locations, years and treatments. To address questions about the effects of breeding on performance under reduced agrochemical applications, the entire cultivar panel was grown in a main trial series over two growing seasons, at six distinct locations, under three cropping intensities side by side at each location with strongly contrasting levels of nitrogen supply (soil mineral N + N fertilization) and plant protection chemicals (see Methods; Supplementary Tables 2 and 3). A high-intensity treatment, comprising nitrogen supply of 220 kgN ha$^{-1}$ (soil mineral N + N fertilization) along with best-practice fungicide, insecticide and growth regulator applications (HiN/HiF treatment), reflected standard conditions for intensive wheat production in western Europe. This high-intensity variant was directly compared with a fungicide-free treatment with the same level of nitrogen fertilization (HiN/NoF) and a fungicide-free treatment with only 110 kgN ha$^{-1}$ (LoN/NoF). Moreover, performance comparisons of the full cultivar panel under drought stress versus irrigated conditions (HiN/HiF) were also added at one location characterized by low rainfall and light soils.

For validation, we further tested a representative subset of 42 cultivars in a full treatment-factorial design, assessing yield responses to all combinations of the two nitrogen levels and the two fungicide variants in a total of eight independent environments, spanning five locations and four growing seasons (2014–2015 to 2017–2018). The cultivars in the validation panel were selected to be genetically and temporally representative for the larger test panel across all five decades of cultivar release. In addition to the three treatments from the main trials, the validation trials also included a fourth variant with best-practice fungicide and 110 kgN ha$^{-1}$ (LoN/HiF), which is unlikely to be applied in practical farming but addresses the potentially compounding effects of nitrogen and fungicide applications in the main experiment. Furthermore, at four of the locations in 2016–2017 and/or 2017–2018, the additional trials also included eight new cultivars, seven of which were registered between 2014 and 2016, to evaluate whether breeding trends observed in the main data set continued into subsequent years. Full details of all field trials, phenotypic measurements and data processing are provided in the Methods and Supplementary Tables 3 and 4, while the composition of the validation panel is indicated in Supplementary Table 1.

In total, the main trials and validation experiment comprised 18,844 full-sized yield plots, in which a total of 209,806 trait values were measured. These extensive, multi-location and multi-season field evaluations provided detailed insight into the long-term effect of intense phenotypic selection for maximum grain yield, under optimal cropping conditions, on the adaptive capacity of elite wheat cultivars to production scenarios with reduced agrochemical inputs or biotic and abiotic stress (Figs. 1 and 2). By comparing patterns of linkage disequilibrium (LD), calculated from 8,710 polymorphic, genome-wide single-nucleotide polymorphism (SNP) markers (see Methods; Supplementary Data File 1), we also assessed how breeding has impacted population genetic parameters over time in European winter wheat.

For each of the two different trial series, a trial-specific linear mixed model (see Methods, equations (1) and (5)) was used to estimate the adjusted mean phenotype values for each trait investigated, in each treatment, for each cultivar, across all years, locations and replicates in the respective trial (Supplementary Table 5). Supplementary Fig. 1 and Supplementary Table 6 describe the variance components and broad sense heritability for all assessed traits, estimated by a fully random model (Methods, equation (2)) with standard errors calculated according to Methods, equation (3)[13].

**Recent cultivars outperform older cultivars under both high-input and low-input scenarios.** Based on robust phenotype data from large-scale field experiments, the results of our study show that modern cultivars consistently perform best under both high and low agrochemical input. Also, genome-wide SNP data indicate that genetic diversity was not reduced in European wheat cultivars over the past five decades of breeding progress. An extended description of the results is provided as Supplementary Information. All traits investigated showed a low residual (error) variance, except for the total plant biomass and the number of spikes per m$^2$, which were measured on manually harvested samples, and radiation use efficiency, which is technically challenging to measure and was only assessed at one location. The very high proportion of explained variance for all other assessed traits, along with the correspondingly low proportion of variance explained by replicates and sub-blocks (with the exception of powdery mildew infection), indicate a high overall accuracy of the phenotyping and uniformity of the experimental conditions.

Cultivar × environment and cultivar × intensity interactions showed small effects on grain yield in both trials (1–4% of all explained variance). For most other traits, these interactions were also considerably lower than the variance explained solely by the cultivar. The high genetic variance in relation to the variance caused by interactions, in combination with the large size of the experiment, resulted in high broad-sense heritabilities ($H^2$; Supplementary Fig. 1 and Supplementary Table 6) ranging from $H^2 = 0.63$ for green-canopy duration to $H^2 = 0.94$ for plant height. Heritability for grain yield was $H^2 = 0.88$ in the main trial and $H^2 = 0.90$ in the validation trial.

We found strong genetic correlations ($r_g > 0.60$) for all pairwise comparisons between treatments for individual traits (Supplementary Fig. 2). Highly significant relationships (Pearson product moment correlation, $P < 0.001$) were seen within and between all treatments in both the main trial and the validation experiment (Supplementary Fig. 3). These results confirm a major contribution of the genotype to performance under all treatments (Supplementary Table 6). The number of kernels per m$^2$, the number of kernels per spike, total plant biomass, nitrogen use efficiency (NUE) and stripe rust resistance showed the highest positive relationships with grain yield under all tested intensity levels (Supplementary Fig. 2 and Supplementary Tables 7–9), whereas plant height showed a high negative correlation to grain yield under high nitrogen fertilization. For each of these traits, incremental genetic gain over time was demonstrated by strong linear improvement of cultivar performance in relation to the year of cultivar release, under all treatments, both in the main trial (Supplementary Fig. 4) and for grain yield in the validation trial (Supplementary Fig. 5). Genetic gains for grain yield were consistent across treatments regardless of the baking quality classification (Supplementary Fig. 6 and Supplementary Table 7), despite the overall negative correlation of protein content with grain yield ($R^2 = 0.58$) in the high-intensity treatment. As a consequence, total protein yield per ha and NUE increased over time under all treatments in all grain quality classes (Supplementary Fig. 6). The considerably lower thousand-kernel weight and NUE under the HiN/NoF treatment than under the HiN/HiF treatment reflects the strong difference in yield (in particular, seed filling) without fungicide applications due to pathogen infection (affecting the numerator in the NUE calculation, the harvested nitrogen). In the LoN/NoF treatment, this effect is negated by the overriding difference in the level of applied nitrogen (the denominator in the NUE calculation). Temporal changes in pathogen races as a direct result of cultivar resistance can compound retrospective comparisons of resistance levels. However, the accumulated resistances of modern cultivars are necessary to secure productivity in response to the disease pathotypes that are prevalent in present-day and future production systems.

Population structure was investigated by principal component analysis based on a modified Roger's genetic distance matrix

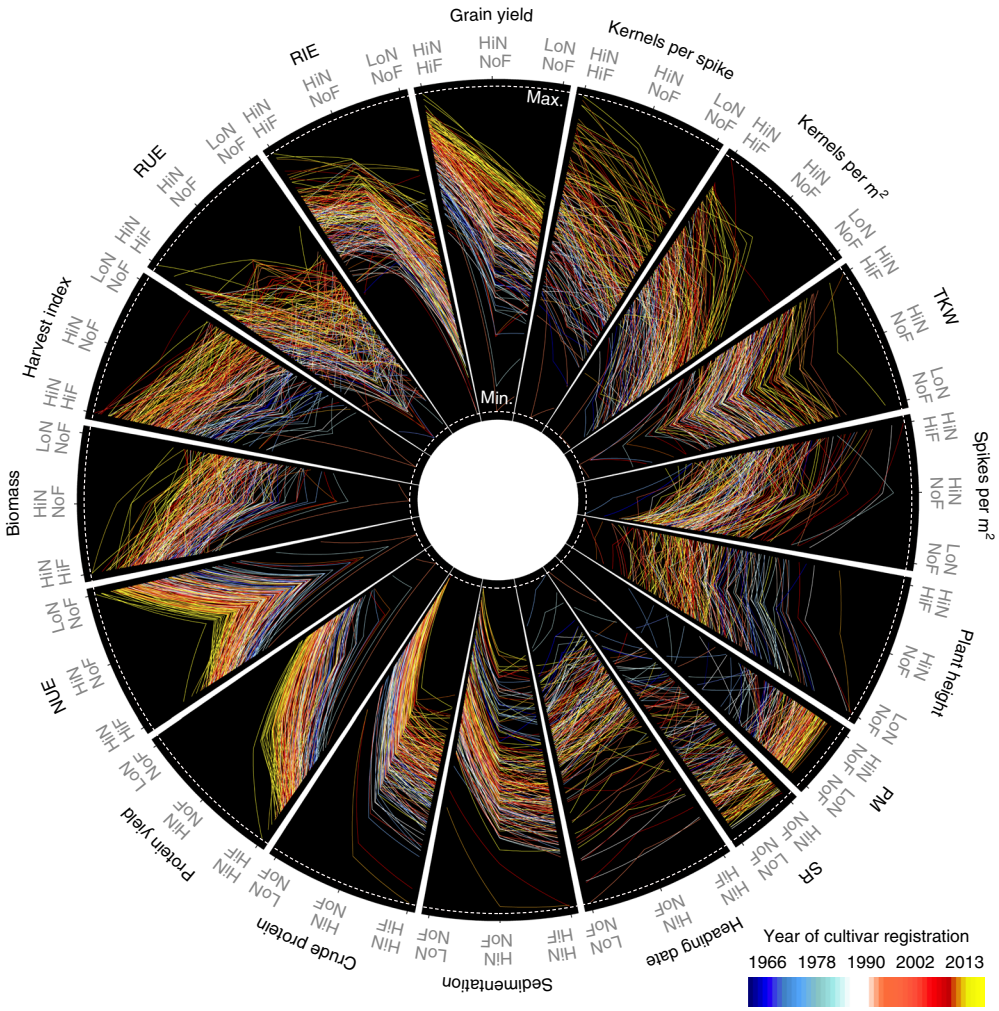


**Fig. 1 | Fifty years of breeding progress in European winter wheat.** Plots show trait values for grain yield, yield parameters, disease resistances, grain quality and physiological traits measured across cultivars released during the past five decades. Adjusted mean trait values for each of the cultivars are calculated for each of the three agrochemical treatment intensities, with varying nitrogen fertilization and/or fungicide application, at six locations over two years. Scales for each trait on the *y* axes represent the relative range from the minimum (innermost edge) to the maximum (outermost edge) mean value for the respective trait, calculated as [(cultivar value − minimum value)/(maximum value) − minimum value)]. PM, powdery mildew resistance; RIE, radiation interception efficiency; RUE, radiation use efficiency; SR, stripe rust resistance; TKW, thousand-kernel weight.

(Supplementary Fig. 7). Little evidence was found for genetic stratification in the cultivar panel, reflecting extensive germplasm exchange between breeding programmes under the 'breeder's privilege' facilitated by the International Convention for the Protection of New Varieties of Plants[14]. Temporal trends in gene diversity and LD within the A, B and D subgenomes were measured by a sliding window analysis, with 40-cultivar windows and single-cultivar increments according to the year of registration. This revealed fluctuating levels for both parameters over the five decades of the study period (Supplementary Figs. 8–10), but no evidence for long-term reduction in genetic diversity as a result of breeding. These results correspond to findings in earlier studies which also demonstrated retention of genetic diversity during long-term wheat breeding[8,15].

To account for pronounced LD patterns as a consequence of breeding[16], we analysed the size and distribution of genome-wide LD blocks consisting of SNPs in strong LD. Overall, 3,768 LD blocks were defined by 8,710 SNPs, with similar distribution patterns of LD blocks observed on all three subgenomes (Supplementary Figs. 11 and 12 and Supplementary Data File 2). For almost all traits, we observed strong linear relationships across all tested management intensities between the trait performance and the year of cultivar registration (Supplementary Figs. 4 and 5). The consistent improvement in cultivar performance across very different agricultural input scenarios suggests continual gains in ecovalence as a result of breeding, promoted by official cultivar testing procedures that require trait improvement and consistency across diverse environments as a prerequisite for cultivar registration. These findings strongly contradict the longstanding popular paradigm that intensive breeding for high performance leads to reduced adaptive diversity in modern cultivars and therewith reduces genetic potential for long-term genetic improvement[17–19].

**A novel haplotype-based approach to assess genome-wide local genomic estimated breeding values and trait variances.** To explore genetic factors that potentially underlie the significant performance improvement of modern elite wheat varieties, independently of crop management, we used a novel approach that focuses on trait effects conferred by chromosomal segments instead of single markers. Results from our data show strong linear temporal relationships between registration year and the number of detrimental chromosome segments. This suggests that breeding has incrementally eliminated negative genetic factors over time (Fig. 2 and Supplementary Fig. 4).

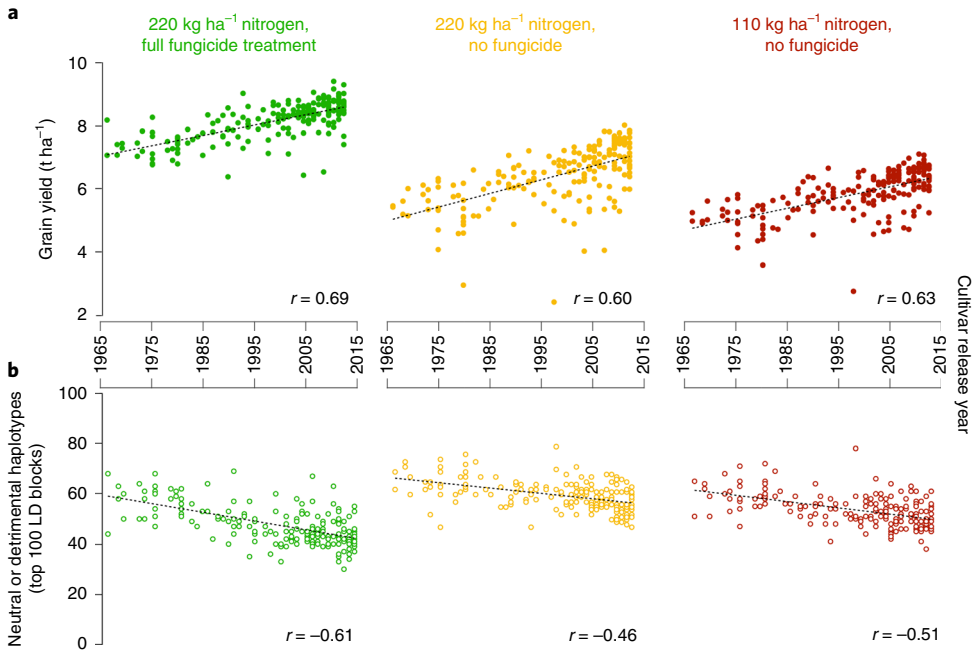


**Fig. 2 | Temporal trends in grain yield performance and detrimental haplotype counts across 50 years of wheat breeding progress. a**,**b**, Regression plots showing grain yield performance among 191 European winter wheat cultivars from the past 50 years in relation to their year of cultivar registration (**a**), and corresponding data showing the number of detrimental or neutral haplotypes per cultivar associated with grain yield (**b**). Performance was assessed for treatments under three contrasting agrochemical treatment intensities in six locations over two growing seasons. Only the 100 LD blocks with the highest trait variance were considered for the haplotype analysis. Linear regression lines and coefficients of determination (*r*) describe the strength of the relationship of the registration year to the trait values and to the detrimental or neutral haplotype counts of each cultivar, respectively.

To relate selection for favourable genetic factors to observed genome-wide LD patterns, we designed a novel haplotype-based procedure (see Methods; Supplementary Fig. 13), modified from classical genomic prediction methods[20], that estimates phenotypic effects conferred by individual LD blocks. The effects of observed haplotype variants on each respective trait were summarized for markers sharing the same LD block, enabling us to estimate local genomic estimated breeding values (GEBVs) for commonly inherited chromosome segments. These local GEBVs were used to calculate variances among haplotype effects for each genome-wide LD block, for all investigated traits (Methods). By grouping SNPs that are inherited in common, this novel approach overcomes one of the fundamental problems with single-marker effect estimates in structured breeding populations with pronounced LD. Thus, it improves effect estimations in chromosome blocks that might carry interacting alleles, for example, markers linked to the same quantitative trait locus. In all three subgenomes, we detected blocks with high contributions to variance for grain yield across the different treatments (Fig. 3, Supplementary Figs. 14 and 15 and Supplementary Data File 2), implying that significant variation exists in European elite winter wheat with relation to the genetic capacity for cultivar adaptation to strongly contrasting agricultural input scenarios. The most prominent LD block (2D_b003483), spanning 103 SNPs from 570,951,341 to 608,198,626 bp on chromosome 2D (Supplementary Fig. 1), exhibited the highest observed local GEBV variance for grain yield under optimum conditions, with estimated haplotype effects ranging from −40 kg ha$^{-1}$ to >30 kg ha$^{-1}$. This block was found to be under strong selection for number of grains per m$^2$ and total plant biomass, and haplotypes showing the strongest positive effects on grain yield were also fixed in the 20 cultivars with the highest mean values for both traits (Supplementary Fig. 16). Well-known wheat domestication and adaption genes lie outside the most prominent LD blocks (Supplementary Table 10 and Supplementary Fig. 11), consistent with recent selection during breeding in adapted elite germplasm. Sixteen smaller LD blocks comprising 20–40 SNPs were observed on chromosomes 1A, 2A, 4A, 5A, 2B, 4B, 5B, 6B, 1D and 5D (Supplementary Data File 2).

For the 100 LD blocks with the highest local GEBV variances, we investigated how many were shared for each trait across the three different treatments in the main trial (Supplementary Data File 2 and Supplementary Fig. 17). The results suggest that overall yield progress over time can be attributed to the accumulation of effects from parallel selection acting on multiple yield parameters, disease resistances, physiological traits and plant architecture. Correlations of individual traits to yield gains suggest that the greatest effect on breeding progress for grain yield came from an increase in the number of kernels per spike, which in turn was a major contributor to improvement of the harvest index. While these are traits that are relatively simple for breeders to assess and select for, the simultaneous strong improvements in radiation use and nitrogen use efficiency indicate that selection for grain yield over time has an overall positive influence on the efficiency of physiological resource allocation and nutrient remobilization. Common haplotypes with strong effects on different traits in different intensity levels (Fig. 3 and Supplementary Figs. 14, 15 and 17) provide valuable new information for targeted breeding of cultivars with consistently good performance across optimal and suboptimal environments.

**Breeding incrementally eliminates detrimental genetic variants.** Genetic effects of historical breeding in the context of genetic gain were investigated by comparing the coefficient of regression for each trait between the number of haplotypes with detrimental or neutral (≤0) effects within each cultivar to the year of cultivar registration (see Methods). For all traits exhibiting breeding gain over time, we found that breeding progress associates clearly with an incremental reduction in the number of haplotypes conferring detrimental

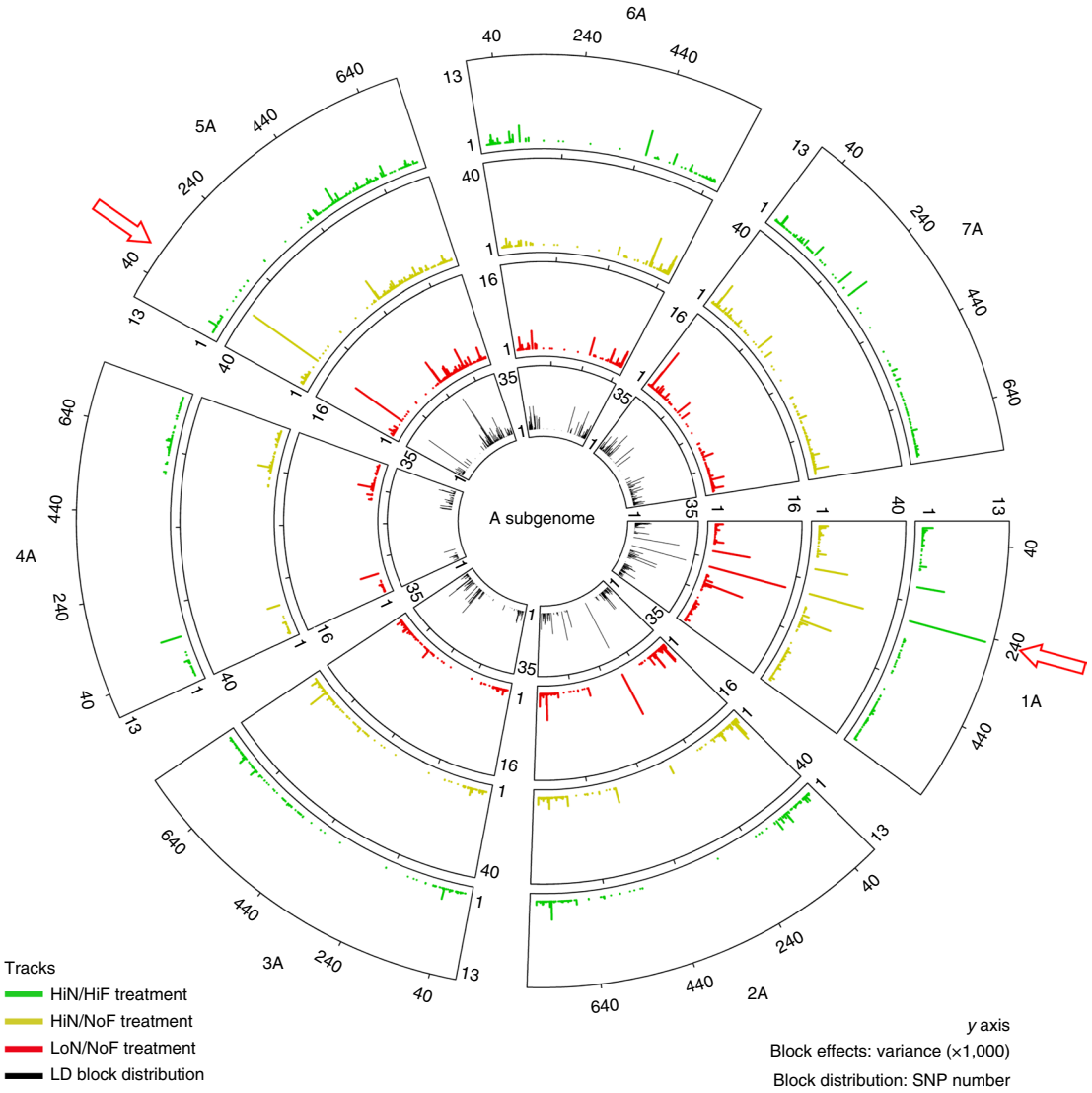


**Fig. 3 | Subgenomic patterns of LD block variance for grain yield in the A subgenome of winter wheat.** The coloured peaks show chromosomal positions in megabase (*x* axes) of LD blocks, along with the corresponding variance of each block (multiplied by 1,000) under three agrochemical treatments comprising HiN/HiF, HiN/NoF and LoN/NoF. Variances for haplotype effects were estimated for each block observed in the panel of 191 European winter wheat cultivars. The black peaks in the innermost ring show the number of SNPs per block. The arrow on chromosome 1A indicates an LD block with consistently high variance across all three intensity levels (block 1A_b000085), whereas the arrow on chromosome 5A indicates a resistance-associated LD block showing variance only in the absence of fungicide applications (block 5A_b000813). Corresponding patterns in the B and D subgenomes are shown in Supplementary Figs. 14 and 15.

or neutral phenotypic effects (Fig. 2, Supplementary Fig. 4 and Supplementary Table 11). As a consequence, the newest 20% of cultivars carried significantly fewer detrimental or neutral haplotypes than the oldest 80% of cultivars for almost all traits (Supplementary Fig. 18). Grain yield under full agrochemical application showed the strongest reduction of non-beneficial LD haplotypes, implying that selection for increased grain yield under optimal conditions was the major driver in European elite wheat breeding programmes over the past 50 years. However, corresponding patterns of allelic selection under all treatments (Supplementary Fig. 4) demonstrate a simultaneous positive selection effect on performance in reduced input production systems. These results suggest that modern elite cultivars are genetically more suitable than older cultivars to increase productivity in low-input wheat production systems and to minimize yield penalties from environmental or political constraints on agrochemical inputs. Interestingly, the regression coefficients between year of registration and number of detrimental alleles was distinctly lower when considering single SNP markers alone instead of LD blocks (Supplementary Table 11), implying that considering local GEBV variances of haplotypes is more powerful in detecting genome regions that have contributed to the genetic improvement of modern elite cultivars.

It may be argued that wheat cultivars specifically targeted for implementation in low-input organic production systems are subject to different breeding and cultivar testing systems, which consider specific requirements for organic production environments that are less relevant for conventional breeding and testing. Conversely, conventional cultivar registration procedures test candidates under a broad range of environments, including fungicide-free treatments, so that cultivars which fulfil registration requirements and perform well in independent regional tests can be expected to have a good relative performance under diverse conditions including stress

environments. Our observation that genetic gain for performance under high-input conditions simultaneously increases performance under suboptimal management conditions has important consequences for the design of appropriate breeding and cultivar testing strategies to enhance sustainability of future agricultural systems. Interestingly, under reduced nitrogen or fungicide-free conditions, cultivars registered specifically for organic agriculture rarely outperformed conventionally bred cultivars from the same year of release (Supplementary Fig. 19).

Climate change predictions forecasting increased drought incidence and severity in major wheat growing areas underline the need for cultivars better adapted to unpredictable water supply[21,22]. Yield performance under low-rainfall conditions at location Gross Gerau in 2015 (170 l per $m^2$ precipitation during the main growth and maturation period from March until July, on a light, sandy soil) showed a strong positive correlation between irrigated and non-irrigated conditions ($R^2 = 0.57$). A consistent mean irrigation effect on grain yield of +2.06 t $ha^{-1}$ across all genotypes confirmed the negative effect of drought in this environment. Surprisingly, the number of haplotype blocks with detrimental or neutral influence on grain yield showed a strong reduction through breeding under both the irrigated ($R^2 = 0.31$) and the non-irrigated ($R^2 = 0.22$) conditions (Supplementary Fig. 20a and Supplementary Table 11). This suggests that long-term breeding progress for maximum grain yield under optimal conditions has also improved the adaptive capacity of modern elite wheat cultivars to outperform older cultivars under low-rainfall conditions. Improved water use efficiency may relate to radiation interception efficiency, photosynthetic efficiency and an extended green-canopy duration (*staygreen* character), all features that we found to be improved in modern cultivars (Supplementary Results and Supplementary Figs. 4 and 20b).

**Exploiting genetic potential to improve sustainable productivity.** A haplotype-based simulation approach, using haplotype blocks weighted according to their estimated trait variances, facilitated forecasts of genetic potential within our cultivar panel in the context of future sustainable production. Estimated effects of stacking the most beneficial haplotypes for yield under optimum conditions suggested that replacing only 50 detrimental loci in the 20 highest-performing, most recent cultivars could improve grain yield potential by 2.6%, whereas accumulation of the most beneficial haplotypes at all 3,768 genome-wide LD blocks could increase yield potential by up to 23% compared with the best-current elite cultivars (Supplementary Fig. 21). Consideration of haplotype block effects can also facilitate design of complex crosses among complementary genotypes, to enrich breeding progenies with beneficial haplotypes. Induced targeted recombination has potential to facilitate this process in the future[23]. Combining accelerated breeding cycles with haplotype-based genomic selection methods could further accelerate the generation and recognition of superior recombinants to maximize exploitation of the available genetic potential.

**Countering common perceptions of breeding and crop diversity.** The results of this study clearly contradict, in one of the world's most important crops, the popular paradigm that intensive plant breeding results in cultivars that have high performance under optimal conditions but a poor capacity to perform under suboptimal growing conditions. In fact, we found the opposite to be true: the positive effects of genetic gain in modern cultivars for sustainability-related traits such as NUE or disease resistance were even more apparent under reduced input scenarios than under optimal agrochemical applications (Supplementary Figs. 4e,f,k and 22), and were also clearly seen under drought conditions (Supplementary Fig. 20). In light of these results, recommendations to use environmental response measures as a proxy for cultivar resilience[10], or to guide breeding decisions, appear misguided. Given that breeding, cultivar registration and post-registration testing procedures consider diverse environments, across many years or even decades, increased yield performance in conjunction with enhanced yield stability are the expected outcome of long-term cultivar breeding under increasingly erratic climatic conditions. Our study demonstrates that breeding promotes both cultivar performance and yield stability across diverse environments and management scenarios (Supplementary Fig. 23). Haplotype-based approaches for discovery and combination of beneficial variants associated with breeding progress provide a novel methodological basis to guide breeding and maintain future yield progress in the face of social, political and environmental constraints on agriculture.

## Methods

**Plant material and genome-wide SNP marker data.** Seeds from a collection of 191 wheat cultivars, registered in Europe between 1966 and 2013, were bulked under field conditions to generate sufficient seeds for multi-location, multi-year field trials. The 191 cultivars were selected for the study based on their economic and agronomic importance in wheat production in Germany during their respective period of release. Wheat productivity in Germany has been among the highest in the world during the past five decades (see global crop production statistics at http://www.fao.org/faostat/en/#data), and during their period of release, most cultivars in the test panel were market-leading varieties within the four main grain quality classes 'E', 'A', 'B' and 'C' (very high to very low baking quality) of the German wheat quality classification. Thus, the study panel represents long-term plant breeding progress at the very peak of global agronomic performance of the world's most widely grown food crop. Full details of cultivars, including year and country of registration and grain quality classification, are provided in Supplementary Table 1.

Leaf DNA samples from all cultivars in the panel were genotyped with a 15K SNP Illumina Infinium iSelect genotyping array[24] that was developed by TraitGenetics using selected, genome-wide, high-quality, polymorphic SNP probes from the 90K Illumina wheat genotyping array (Illumina)[25]. Raw genotype data were filtered to retain only markers with ≤10% missing values and a minor allele frequency of ≥5%.

Physical genetic SNP marker positions in the wheat genome were obtained using the alignment software GYDLE (Gydle Inc. Bioinformatics Service; http://www.gydle.com). The 50-mer nucleotide sequences corresponding to the SNP probes on the Illumina iSelect wheat 15K SNP bead-chip array were aligned to the genome sequence assembly for bread wheat cultivar Chinese Spring (IWGCS Reference Sequence v1.0; https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies). The GYDLE alignment parameters used allowed for reporting of all positions in the genome assembly where the 50-mer probe aligned with >80% sequence homology, to capture all potential hybridization sites for the probes. To determine which of the putative SNP probe hybridization sites in the wheat genome revealed polymorphism in the iSelect 15K bead-chip assay, previous knowledge about the chromosomal locations of genetically mapped SNP loci was used to filter the probe hybridization sites. This filtering step identified SNP probes that had only one putative hybridization site per chromosome and that had been previously genetically mapped on the same chromosome. Only the 8,710 high-quality, polymorphic SNP probes meeting these criteria were included in subsequent analyses. Physical genetic positions and calls for these 8,710 SNPs are provided in Supplementary Data File 1.

**Field trials and phenotype data analysis.** Field trials were conducted in full-sized yield plots (harvested area 4.5–12 $m^2$ depending on site-specific sowing and harvesting machinery), across a total of seven locations throughout Germany (Supplementary Table 2) characterized by diverse soil conditions (Supplementary Table 3), in four consecutive growing seasons from 2014–2015 to 2017–2018. Plots were sown with a sowing density of 330 viable seeds per $m^2$.

The main field trials, in which all 191 cultivars in the panel were tested, were performed at each of six locations over the growing seasons 2014–2015 and 2015–2016. In each of the 12 year × location environments in the main trials, the 191 cultivars were grown in at least two replicates, sown side by side under each of the three different cropping intensities, designated as HiN/HiF, HiN/NoF and LoN/NoF treatments. The HiN/HiF treatment received mineral fertilizer at a total nitrogen supply rate of 220 kgN $ha^{-1}$ (fertilization adjusted for soil mineral nitrogen, $N_{min}$) along with full intensity of fungicides, insecticides and growth regulators, representing standard agrochemical applications under intensive wheat production conditions in western Europe. The HiN/NoF treatment also received a total nitrogen supply of 220 kgN $ha^{-1}$; however, no fungicides were applied. The LoN/NoF treatment was supplied with only 110 kgN $ha^{-1}$ and no fungicides were applied. Full details of sowing dates are shown in Supplementary Table 3.

To reduce neighbour effects due to large differences among the cultivars in plant height and maturation period, the cultivars in the main trial were grouped within each treatment and replication into three sub-blocks considering previous knowledge about these parameters in the cultivar panel. Cultivars were completely

randomized within the experimental sub-blocks and the order of the treatments was completely randomized at each location in each year. To prevent confounding effects from weed contamination and basic nutrient deficiencies, standard weed control measures were applied across all treatments, and nutrients other than nitrogen were applied according to requirements determined individually at each location.

Levels of susceptibility to the fungal foliar diseases powdery mildew (*Blumeria graminis* (DC.) Speer f.sp. *tritici*) and stripe rust (*Puccinia striiformis* Westend f.sp. *tritici*), which showed moderate-to-strong infection levels over multiple locations in one or both years, were determined by visual scoring on all plots. The respective resistance score was expressed on a scale from 0% to 100% as the percentage of the non-infected leaf area. Disease scores for stripe rust were recorded in all environments except at location KIE, where infection levels were insufficient in both trial seasons. For the same reason, powdery mildew scores were recorded in both growing seasons only in the fungicide-free treatments at locations GGE and QLB, and only under the HiN/HiF treatment at GGE in the 2015–2016 growing season.

Before grain harvest from each plot, an aliquot of 0.5 m in the centre of the plot was removed to determine the number of spikes per $m^2$ and the above-ground plant biomass (corrected for the specific row width at each individual location). After threshing of these samples, the harvest index was calculated as the grain yield divided by the above-ground plant biomass (grain yield plus straw dry weight).

Grain yield was determined on the plot level by threshing from the mature standing canopy. Immediately after threshing, grain moisture was measured and grain yield was corrected to a standard moisture of 14%. A random grain subsample from each plot was used to determine the thousand-kernel weight. Grain quality parameters related to baking quality were measured according to standard procedures as follows: Hagberg–Perten falling number, ISO 3092:2009; near-infrared spectrometry estimates for Zeleny sedimentation index corresponding to ISO 5529:2007; crude protein, AACCI Method 39-11.01.

Protein yield was determined by multiplying the grain yield by the crude protein content of each plot sample. Nitrogen yield was calculated by dividing the protein yield by the wheat-specific protein factor of 5.7. NUE was expressed as the quantity of nitrogen in the harvested grains in relation to the quantity of fertilized nitrogen.

For the main field trials, a linear mixed model (equation (1)) was used to estimate adjusted means across locations and years for each cultivar in each treatment.

$$\begin{aligned} P_{ijklmo} &= \mu + g_i + t_j + (gt)_{ij} + Y_k + L_l + (YL)_{kl} \\ &\quad + (YLR)_{klm} + (YLRTB)_{jklmo} + e_{ijklmo} \end{aligned} \tag{1}$$

where $P_{ijklmo}$ is the observed phenotype of the $i$th variety, the $j$th treatment, the $k$th year, the $l$th location, the $m$th replication and the $o$th block, $\mu$ is the general mean, $g_i$ is the fixed effect of the $i$th variety, $t_j$ is the fixed effect of the $j$th treatment and $(gt)_{ij}$ is the fixed effect of the $i$th variety in the $j$th treatment. $Y_k$ represents the random effect of the $k$th year, $L_l$ is the random effect of the $l$th location, $(YL)_{kl}$ is the random interaction of the $l$th location, in the $k$th year, and $(YLR)_{klm}$ is the random effect of the $m$th replication, in the $l$th location, in the $k$th year. $(YLRTB)_{jklmo}$ represents the random effect of the $o$th block, in the $j$th treatment, in the $m$th replication, in the $l$th location, in the $k$th year, while $e_{ijklmo}$ is the error term. Fixed effects are denoted by lowercase letters, and random effects are denoted by uppercase letters. The average standard errors of the difference between the adjusted means were used to determine the least significant difference using the quantiles of the $z$-distribution. Data was analysed in the software R[26] with package asreml[27].

A fully random model (equation (2)) was used to estimate variance components.

$$\begin{aligned} P_{ijklmo} &= \mu + G_i + T_j + (GT)_{ij} + (GYL)_{ikl} + Y_k + L_l \\ &\quad + (YL)_{kl} + (YLT)_{jkl} + (YLR)_{klm} + (YLRTB)_{jklmo} + e_{ijklmo} \end{aligned} \tag{2}$$

where $G_i$ the random effect of the $i$th variety, $T_j$ is the random effect of the $j$th treatment and $(GT)_{ij}$ is the random effect of the $i$th variety in the $j$th treatment. $(GYL)_{ikl}$ is the random effect of the $i$th variety in the $l$th location and the $k$th year. $(YLT)_{jlk}$ is the interaction of the $j$th treatment with the $l$th location in the $k$th year.

For disease resistance traits, data from all replications per treatment were averaged into an arithmetic mean across all scoreable test locations and years.

A model containing only those interactions (equation (3)) that contribute to the heritability, with the cultivar as a random factor in addition to fixed main effects, was used to estimate the standard error of treatment means that was used subsequently to estimate the heritability.

$$P_{ijkl} = \mu + g_i + t_j + y_k + l_l + (GT)_{ij} + (GYL)_{ikl} + e_{ijkl} \tag{3}$$

where $P_{ijkl}$ is the observed phenotype of the $i$th variety in the $j$th treatment and the $k$th year at the $l$th location, $y_k$ is the fixed effect of the $k$th year and $l_l$ is the fixed effect of the $l$th location. $e_{ijkl}$ is the error term.

Genetic variance from equation (2) and the standard error from equation (3) were further used to estimate broad-sense heritability for each trait, as described in equation (4)[13].

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + SE^2} \tag{4}$$

where $\sigma_G^2$ is the genetic variance derived from the full random model (equation (2)) and $SE^2$ is the squared standard error of the difference between the means, derived from equation (3). Data analysis was performed with the R packages lsmeans[28] and lme4 (https://CRAN.R-project.org/package=lme4)[29].

Average annual gains in trait performance were quantified from the slopes of the linear regression curve for each trait, based on the adjusted mean trait values across all locations and years, separately for each of the three intensity levels. Figure 1 was created using the R package circlize[30].

**Validation trials.** To account for potential compounding effects of nitrogen application and fungicides, we conducted an additional experiment that implemented a full factorial combination of high or low nitrogen levels combined with the presence or absence of fungicides, respectively (Supplementary Table 3). A subset of 42 cultivars from the main panel was selected to be representative for the genetic diversity in the main panel throughout the entire five-decade cultivar release period (labelled 'V' in Supplementary Table 1 and highlighted in Supplementary Fig. 7). The validation panel was tested in eight contrasting environments, spanning five locations and four growing seasons (2014–2015 to 2017–2018), in two replications (with the exception of location KIE, for which three replications were performed). Four treatments were compared, comprising HiN/HiF, HiN/NoF and LoN/NoF (as described above for the main experiment) along with a LoN/HiF treatment comprising 110 kgN $ha^{-1}$ with fungicides. Furthermore, at four of the locations in 2016–2017 and/or 2017–2018, the additional trials also included eight new cultivars, seven of which were registered between 2014 and 2016, making a total of 50 entries in these environments (for details of locations and treatments, see Supplementary Tables 2 and 3).

Here, we report only grain yield performance for the validation trials. Because the composition of the validation trials enabled us to run the trials without grouping into sub-blocks, we used a slightly modified linear mixed model for the validation trial to estimate adjusted mean yields across locations and years for each cultivar in each treatment, using rows and columns rather than sub-blocks to correct for spatial effects, as shown in equation (5):

$$\begin{aligned} P_{ijklmwz} &= \mu + g_i + t_j + (gt)_{ij} + Y_k + L_l + (YL)_{kl} \\ &\quad + (YLR)_{klm} + (YLW)_{klw} + (YLZ)_{klz} + e_{ijklmwz} \end{aligned} \tag{5}$$

where $P_{ijklmwz}$ is the observed phenotype of the $i$th variety, the $j$th treatment, the $k$th year, the $l$th location, the $m$th replication, the $w$th column and the $z$th row. $(YLW)_{klw}$ represents the random effect of the $w$th column, in the $l$th location, in the $k$th year. $(YLZ)_{klz}$ represents the random effect of the $z$th row, in the $l$th location, in the $k$th year, while $e_{ijklmwz}$ is the error term. The average standard errors of the difference between the adjusted means were used to determine the least significant difference using the quantiles of the $z$-distribution. Data was analysed in the software R[26] with package asreml[27].

Accordingly, a lightly modified fully random model (equation (6)) was used to estimate variance components in the validation panel:

$$\begin{aligned} P_{ijklmwz} &= G_i + T_j + (GT)_{ij} + (GYL)_{ikl} + Y_k + L_l \\ &\quad + (YL)_{kl} + (YLT)_{jkl} + (YLR)_{klm} + (YLW)_{klw} + (YLZ)_{klz} \\ &\quad + e_{ijklmwz} \end{aligned} \tag{6}$$

where $(YLT)_{klj}$ is the interaction of the $j$th treatment with the $l$th location in the $k$th year.

Broad-sense heritability for grain yield of the validation trials was calculated as described for the main field trials, according to equations (3), (4) and (6).

**Evaluation of yield stability.** To evaluate yield stability of the 191 cultivars across all locations, years and management scenarios in the main trials, we calculated the coefficient of variation for grain yield across all treatment by location by year combinations in the main trial, along with $P_i$, the cultivar superiority index[31], measured according to equation (7):

$$Pi = \sum_{j=1}^{n} \frac{(X_{\max,j} - X_{ij})^2}{2n} \tag{7}$$

where $X_{ij}$ is the yield of cultivar $i$ in the year by location by treatment combination $j$, $X_{\max,j}$ is the highest yield of any cultivar across environments and treatments, and $n$ is the number of year by location by treatment variants. Thus, $P_i$ describes how closely a cultivar X comes to achieving the maximum yield potential ($X_{\max,j}$).

**LD and gene diversity.** Temporal trends in genome-wide LD and gene diversity in relation to the year of cultivar release were investigated in detail using a sliding window analysis. LD was calculated per chromosome as the average of $r^2$ values between adjacent markers separated by ≤20 Mb. Gene diversity was calculated from genome-wide SNP genotype scores according to Nei[32]. Because the panel contained different numbers of cultivars registered in each year, we applied a constant window size of 40 genotypes to avoid numerical bias from year effects. After randomizing the cultivar order in each year, LD and gene diversity were calculated for windows of 40 cultivars, beginning with the oldest 40 cultivars and incrementing in single-cultivar steps through to the most recently released cultivars tested in the main trials. Values were plotted against the average year in each window as a time reference.

**Genome-wide LD block construction.** Genome-wide SNP markers were assigned to LD blocks by grouping based on pairwise $r^2$ values with a minimum LD threshold of $r^2 = 0.7$. Calculations were performed in the statistical software R using an algorithm implemented in the R package SelectionTools (downloadable at http://population-genetics.uni-giessen.de/~software/). Pairwise $r^2$ values between SNP markers were first calculated across each chromosome, followed by selection of adjacent pairs on each chromosome with the highest LD among all pairs. If the $r^2$ of a pair exceeded the threshold of 0.7, these markers were defined as a new LD block. In the next step, flanking markers on each side of the LD block were considered and added to the same block if their pairwise $r^2$ values with the respective outer markers in the LD block also exceeded the threshold. A tolerance parameter per block of $t = 1$ was set to account for incorrectly positioned markers or biased LD estimates, meaning that, if a flanking marker did not fulfil the LD threshold, the block was still extended if at least $t$ adjacent flanking markers fulfilled the LD grouping threshold. If more than $t$ flanking markers had a lower LD than $r^2 = 0.7$, the block was completed. This procedure was repeated until all markers were assigned to blocks. SNPs that were not in LD with any other marker were assigned to individual LD blocks.

**Local genomic estimated breeding values, variance estimation and weighting of genome-wide LD blocks.** To identify chromosomal regions with a high effect on the respective traits in the panel, while simultaneously accounting for pronounced LD structures expected as a result of directional selection, we did not consider phenotypic effects of single markers, but rather assigned their effects to LD blocks (see above) consisting of groups of adjacent markers in LD, resulting in local GEBVs for haplotypes within chromosomal blocks. The principal workflow of the method is shown in Supplementary Fig. 13. We first predicted genome-wide marker effects for each trait in each cropping intensity level, using a ridge-regression best linear unbiased prediction (BLUP) model[20] that predicts all marker effects simultaneously. This procedure prevents the phenotypic effect of a SNP marker from being influenced by neighbouring SNPs in LD to it, which potentially overestimates the effect of a quantitative trait locus by repeatedly reassigning the effect to multiple SNPs in an LD block. By contrast, the calculation of local GEBVs overcomes this problem by assigning effects to LD blocks rather than to individual SNPs within each block. Because the BLUP model fits all SNPs simultaneously, the effect of a SNP in LD with a quantitative trait locus within a block is counted only once. We then summed up the predicted allelic effects of each observed haplotype variant for all genome-wide LD blocks. Finally, we estimated variances among local GEBVs for haplotypes within each LD block. Two strategies were applied to investigate relationships between trait-relevant haplotypes and the year of cultivar registration: first, we disregarded the most recent 20% of the cultivars and used the remaining 80% of the test panel to estimate LD block variances as described above. Based on this estimate, we selected the 100 LD blocks with the highest variance among local GEBVs of the haplotypes for each respective trait. For all of these selected blocks, we counted the total number of haplotypes with detrimental or neutral estimated effects (effect ≤ 0). We then counted the same haplotypes in the remaining 20% of cultivars, which were not used in the variance estimation of the LD blocks, and compared the two groups (oldest 80% versus newest 20%) using Student's $t$-test[33]. Second, we used the whole panel to estimate LD block variances, considered the 100 blocks with the highest variances and plotted the number of haplotypes conferring neutral or deleterious effects in each cultivar against its year of registration.

**Estimating available yield potential from haplotype stacking.** To estimate the potential for grain yield improvement (under optimum conditions) in the modern elite European winter wheat gene pool represented by the cultivar panel, we compared four scenarios that assume the ability to stack different total numbers of predicted 'best' haplotypes. The 20 cultivars with the highest yields under optimum conditions (HiN/HiF) were considered for simulated haplotype stacking. First, in silico genotypes were created from these selected cultivars by simulating the exchange of detrimental or neutral haplotypes by the highest positive-effect haplotype at each of the (1) 20, (2) 50 or (3) 100 blocks with the highest estimated variances for grain yield under the HiN/HiF treatment. In a fourth simulation scenario, (4) we derived new in silico genotypes from each of the 20 cultivars by assuming an exchange with the most favourable haplotype at each of the 3,768 genome-wide LD blocks. GEBVs for each of the 20 derived in silico cultivars were compared with the initially calculated GEBVs for the 20 most recent cultivars, and the relative change in predicted performance was calculated.

**Genetic correlations.** Genetic correlations between traits in the three different cropping intensities were estimated following equation (8):

$$r_g = \frac{\sigma_g(\text{trait1}, \text{trait2})}{\sqrt{\sigma_g^2(\text{trait1}) * \sigma_g^2(\text{trait2})}} \quad (8)$$

Where trait1 and trait2 represent the two phenotype sets for which the genetic correlation was estimated. The genetic covariance $\sigma_g$ and genetic variances $\sigma_g^2$ were estimated in asreml by fitting a multi-trait mixed model for the two traits under consideration, using an additive genomic relationship matrix calculated from 8,710 polymorphic SNP markers with the R package rrBLUP[34].

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Seed aliquots from all cultivars used in the study are available from the corresponding authors on request for research purposes. The complete set of adjusted mean trait data from all field locations, years, treatments and replications is available at the online data repository https://zenodo.org/record/1316947 with the digital object identifier number https://doi.org/10.5281/zenodo.1316947. All other data used in the analyses are provided in the Supplementary Information.

## Code availability

All codes used in the article are available from the corresponding authors. Please contact K.P.V.-F. via k.vossfels@uq.edu.au for code access and additional information.

Received: 17 August 2018; Accepted: 10 May 2019;
Published online: 17 June 2019

## References

1. Foley, J. A. et al. Solutions for a cultivated planet. *Nature* **478**, 337–342 (2011).
2. Garnett, T. et al. Agriculture. Sustainable intensification in agriculture: premises and policies. *Science* **341**, 33–34 (2013).
3. Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R. & Polasky, S. Agricultural sustainability and intensive production practices. *Nature* **418**, 671–677 (2002).
4. Tilman, D., Balzer, C., Hill, J. & Befort, B. L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl Acad. Sci. USA* **108**, 20260–20264 (2011).
5. Raun, W. R. & Johnson, G. V. Improving nitrogen use efficiency for cereal production. *Agron. J.* **91**, 357–363 (1999).
6. Sylvester-Bradley, R. & Kindred, D. R. Analysing nitrogen responses of cereals to prioritize routes to the improvement of nitrogen use efficiency. *J. Exp. Bot.* **60**, 1939–1951 (2009).
7. West, P. C. et al. Leverage points for improving global food security and the environment. *Science* **345**, 325–328 (2014).
8. van de Wouw, M., van Hintum, T., Kik, C., van Treuren, R. & Visser, B. Genetic diversity trends in twentieth century crop cultivars: a meta analysis. *Theor. Appl. Genet.* **120**, 1241–1252 (2010).
9. Fu, Y. B. Understanding crop genetic diversity under modern plant breeding. *Theor. Appl. Genet.* **128**, 2131–2142 (2015).
10. Kahiluoto, H. et al. Decline in climate resilience of European wheat. *Proc. Natl Acad. Sci. USA* **116**, 123–128 (2019).
11. Lobell, D. B., Schlenker, W. & Costa-Roberts, J. Climate trends and global crop production since 1980. *Science* **333**, 616–620 (2011).
12. Snowdon, R. J. et al. Reduced response diversity does not negatively impact wheat climate resilience. *Proc. Natl Acad. Sci. USA* **116**, 10623–10624 (2019).
13. Piepho, H. P. & Möhring, J. Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* **177**, 1881–1888 (2007).
14. *International Convention for the Protection of New Varieties of Plants* (International Union for the Protection of New Varieties of Plants, 1991).
15. Reif, J. C. et al. Wheat genetic diversity trends during domestication and breeding. *Theor. Appl. Genet.* **110**, 859–864 (2005).
16. Voss-Fels, K. et al. Subgenomic diversity patterns caused by directional selection in bread wheat gene pools. *Plant Genome* https://doi.org/10.3835/plantgenome2015.03.0013 (2015).
17. Tanksley, S. D. & McCouch, S. R. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**, 1063–1066 (1997).
18. Tripp, R. Biodiversity and modern crop varieties: sharpening the debate. *Agric. Hum. Values* **13**, 48–63 (1996).

19. Wilkes, G. & Williams, J. T. Current status of crop plant germplasm. *Crit. Rev. Plant Sci.* **1**, 133–181 (1983).
20. Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
21. Wheeler, T. & von Braun, J. Climate change impacts on global food security. *Science* **341**, 508–513 (2013).
22. Watson, J., Zheng, B., Chapman, S. & Chenu, K. Projected impact of future climate on water-stress patterns across the Australian wheatbelt. *J. Exp. Bot.* **68**, 5907–5921 (2017).
23. Bernardo, R. Prospective targeted recombination and genetic gains for quantitative traits in maize. *Plant Genome* https://doi.org/10.3835/plantgenome2016.11.0118 (2017).
24. Muqaddasi, Q. H., Brassac, J., Borner, A., Pillen, K. & Roder, M. S. Genetic architecture of anther extrusion in spring and winter wheat. *Front. Plant Sci.* **8**, 754 (2017).
25. Wang, S. et al. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol. J.* **12**, 787–796 (2014).
26. R Core Team *R: A Language and Environment For Statistical Computing* (R Foundation for Statistical Computing, 2008); http://www.R-project.org/
27. Butler, D. asreml: asreml() fits the linear mixed model. R package version 3.0. www.vsni.co.uk (2009).
28. Lenth, V. R. Least-squares means: the R package lsmeans. *J. Stat. Softw.* **69**, 1–33 (2016).
29. Bates, D., Maechler, M., Bolker, B. M. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
30. Gu, Z. G., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
31. Lin, C. S. & Binns, M. R. A superiority measure of cultivar performance for cultivar X location data. *Can. J. Plant Sci.* **68**, 193–198 (1988).
32. Nei, M. Analysis of gene diversity in subdivided populations. *Proc. Natl Acad. Sci. USA* **70**, 3321–3323 (1973).
33. Student The probable error of a mean. *Biometrika* **6**, 1–25 (1908).
34. Endelman, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* **4**, 250–255 (2011).

## Acknowledgements

Funding for this study was provided by the German Federal Ministry of Education and Research (BMBF) grant 031A354. Additional acknowledgements are provided in the Supplementary Information.

## Author contributions

W.F., H.K., F.O., J.L., H.S., R.J.S., A.S., K.P.V.-F. and B.W. conceived the study and the subsequent data analysis. B.W., A.S., K.P.V.-F., C.L., S.N., T.R., T.-W.C., H.Z., S.S., M.F., E.R., B.J.H., M.J.H., M.M.B., A.B., J.L. and H.K. generated and analysed the data. K.P.V.-F., A.S. and R.J.S. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41477-019-0445-5.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to W.F., H.S. or R.J.S.

**Journal peer review information:** *Nature Plants* thanks M. Reynold, Brande Wulff and other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): R. Snowdon, W. Friedt and H. Stützel

Last updated by author(s): May 9, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☐ | ☒ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Data collection was performed manually by technical staff at each of the field trials using standardized data scoring procedures and without the use of specific software. |
| Data analysis | The software 'GYDLE' (www.gydle.com) was used for SNP marker alignment to obtain physical marker positions in the wheat genome.<br><br>All remaining data analyses were performed in 'R' version 3.5.1. The following R packages were used for the different analyses:<br><br>- 'asreml' (fitting mixed and random linear models for variance component estimation and entry mean adjustment, genetic variance estimation for genetic correlations)<br>- 'lsmeans' (fitting mixed and random linear models for variance component estimation)<br>- 'rrBLUP' (calculation of the genomic relationship matrix from SNP data for fitting a multitrait model for calculation of genetic correlations)<br>- 'SelectionTools' (version 18.1: calculation of pairwise linkage disequilibrium (LD) and gene diversity among SNPs, definition of LD blocks, prediction of SNP effects using ridge regression BLUP)<br><br>References for the software 'R' and the packages 'asreml', 'lsmeans' and 'rrBLUP' are listed in the Supplementary Text reference list and the download link for the 'SelectionTools' package (http://population-genetics.uni-giessen.de/~software/) is also provided in the manuscript. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The complete set of adjusted mean trait data from all field locations, years, treatments and replications will be available in an online data repository via digital object identifier number doi:10.5281/zenodo.1316947 after acceptance of the manuscript. All other data used in the analyses is provided in the supplementary materials.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Study description | The study investigated a panel of 191 elite winter wheat cultivars released during the past 50 years in Western Europe, particularly Germany. Diverse phenotypic traits were analyzed in replicated, randomised field trials across six different locations and two years. In the main trial all 191 varieties were planted side-by-side under three different management intensity levels (high nitrogen, normal fungicide applications; high nitrogen without fungicides; low nitrogen without fungicides). A validation study including 1/4 of the cultivars was grown in a full treatment factorial (incuding the same three treatments plus a 4th treatment with low nitrogen and normal fungicide application).<br><br>To investigate genetic parameters, ~9,000 polymorphic, high-quality single nucleotide polymorphism (SNP) markers from the 90k SNP wheat genotyping array were used. Marker selection parameters are provided in the supplementary methods.<br><br>The main objectives of the study were to:<br>1) compare phenotypic measurements for different traits between the 191 cultivars from different years of variety registration and across a range of chemical input intensities, to investigate if modern varieties only perform well under high agrochemical inputs;<br>2) analyze genetic impacts of five decades of winter wheat breeding;<br>3) investigate the genetic potential in EU wheat germplasm.<br><br>To addressing objectives 2 and 3 we applied a novel framework to estimate haplotype-based trait variances from local genomic estimated breeding values (GEBV) for chromosome blocks consisting of multiple SNP markers in strong LD, thereby taking into account the directional selection history for specific fragments. |
| Research sample | The study population was a panel of 191 European winter wheat (Triticum aestivum) cultivars. The materials were selected based on their agronomic importance (in particular the area of cultivation and the performance during the specific period of release) to represent the most successful varieties being grown during the last five decades of wheat production in Western Europe. The focus was on the most important cultivars in Germany, where winter wheat yields were consistently among highest in the world during this time period. In this respect the sample population was selected to represent breeding progress at the very peak of global wheat production over the past 50 years. |
| Sampling strategy | To address the question of breeding progress over time we sampled representative, high-performing cultivars from each of the past 5 decades, using historical information about cultivar performance in national registration/performance trials, duration of cultivar listing and area of cultivation during the release period. Considering the need for best possible distribution of registration dates across the 5-decade study period, we limited the total number of cultivars to ~200 to avoid bias against earlier decades with (relatively) fewer high-performing cultivars. A small number of (mainly older) cultivars were eliminated during the seed production phase because they showed a strong degree of lodging (which leads to severe neighbor-plot effects in cereal field trials), resulting in a final panel of 191 cultivars. Nevertheless, the number of available cultivars was somewhat lower in the first two decades of the study period, hence fewer cultivars were sampled for this period. To address this imbalance we used a sliding-window approach for temporal analyses and performed regression analyses, considering each individual cultivar in association with its year of registration rather than comparing all cultivars between different registration decades. |
| Data collection | Data collection was performed manually by technical staff at each of the field trials using standardized data scoring procedures. Genotype data for all cultivars was generated using standard procedures with a commercial 90k SNP genotyping array for hexaploid wheat. |
| Timing and spatial scale | The main field trials were performed in 12 independent environments (2015/16 and 2016/17 at six locations) across Germany. The validation trials were performed in 8 independent environments across 5 locations over 4 growing seasons (2014/15 - 2017/18). |

| | |
|---|---|
| Data exclusions | No data were excluded in this study. |
| Reproducibility | Full details on the heritability of all traits investigated and on the contribution of variance from genotypes, locations, years, treatments and their interactions are provided in Supplementary Figure 1 and Supplementary Table 6. Due to a high number of replicated individual field trials (see Supp. Text) we obtained very high broad-sense heritabilities and very low error variances, indicating extremely high reproducibility of the trials. |
| Randomization | Full details on the trial design are provided as Supplementary Methods. In each replication and treatment all test cultivars were completely randomized within the experimental sub-blocks. Order of treatments was randomised for each per location and year. |
| Blinding | Due to the nature of the study, where large-scale data collection was necessary simultaneously at 6 different sites across Germany, all data collection at each individual trial location was performed completely independently by different investigators according to agreed data collection standards. No data was exchanged between the individual data collectors. Furthermore, all data was associated only with plot numbers from a randomized sowing plan and not with cultivar names, corresponding to sample blinding between replications and treatments at each study location. |

Did the study involve field work? ☒ Yes ☐ No

## Field work, collection and transport

| | |
|---|---|
| Field conditions | In all years and at all field trial locations, trials were sown between October 1st and November 4th depending on the location. Depending on year and location, the soil pH content before harvest ranged from 5.5 to 7.4. P2O5 and K2O concentrations ranged from 5.1-37 mg/100g soil and from 6.3-38 mg/100g soil, respectively. Depending on the trial, preceding crops were sugar beet, corn, oilseed rape, oat, spring barley, Faba bean or Avena. Full details about the soil, climate and field conditions of the test locations are provided in Supplementary Table 2. |
| Location | All field trials were performed in Germany. The main trials were performed at Gross Gerau, Klein Altendorf, Kiel, Rauischholzhausen, Quedlinburg and Hannover in the growing seasons 2015/16 and 2017/18, while the validation trials were grown at Kiel, Rauischholzhausen, Quedlinburg, Hannover and Weilburger Grenze between 2014/15 and 2017/18. Geo-references for all trial locations are provided in Supplementary Table 2. |
| Access and import/export | N/A – all samples were obtained from registered winter wheat cultivars which are freely available for breeding and research. |
| Disturbance | N/A – all trials were performed on agricultural field trial locations in the context of standard crop rotations, details about preceding crops are provided in Supplementary Table 2. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

### Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |