Norwegian University of Life Sciences
Faculty of Biosciences
Department of Plant Sciences

Philosophiae Doctor (PhD)
Thesis 2023:52

Grain yield in Norwegian spring wheat: past, present, and future. Historical genetic gains, their genetic basis, and prospects of future improvement using phenomics

Avlingsframgangen i norsk vårhvete: fortid, nåtid og framtid. Historisk framgang, genetisk grunnlag og nye muligheter med fenotypingsteknologier

Tomasz Mróz

Grain yield in Norwegian spring wheat: past, present, and future. Historical genetic gains, their genetic basis, and prospects of future improvement using phenomics

Avlingsframgangen i norsk vårhvete: fortid, nåtid og framtid. Historisk framgang, genetisk grunnlag og nye muligheter med fenotypingsteknologier

Philosophiae Doctor (Ph.D.) Thesis

Tomasz Mróz

Norwegian University of Life Sciences
Faculty of Biosciences
Department of Plant Sciences

Ås (2023)

Thesis number 2023:52
ISSN 1894-6402
ISBN 978-82-575-2081-6

Supervisors and Evaluation Committee

Supervisory team:

Prof. Morten Lillemo
morten.lillemo@nmbu.no
Faculty of Biosciences
Department of Plant Sciences
Norwegian University of Life Sciences (NMBU)
P.O. Box 5003 NMBU, NO 1432 Ås
Norway

Dr. Sahameh Shafiee
sahameh.shafiee@nmbu.no
Faculty of Biosciences
Department of Plant Sciences
Norwegian University of Life Sciences (NMBU)
P.O. Box 5003 NMBU, NO 1432 Ås
Norway

Dr. Jose Crossa
j.crossa@cgiar.org
Biometrics and Statistics Unit
CIMMYT
C.A.P. Plaza Galerías, Col. Verónica Anzures 11305
Ciudad de México
Mexico

Evaluation comitee:

Prof. Luc Janss
luc.janss@qgg.au.dk
Center for Quantitative Genetics and Genomics
Aarhus University
C.F. Møllers Allé 3,building 1535, 115, 8000 Aarhus C
Denmark

Prof. John Foulkes
john.foulkes@nottingham.ac.uk
Faculty of Science
University of Nottingham
Sutton Bonington, Leicestershire
LE12 5RD
United Kingdom

Table of Contents

Acknowledgements

This Ph.D. work was funded by the Norwegian University of Life Sciences and supported by projects vPheno (NFR grant no. 267806) and PhenoCrop (NFR grant no. 320090). The work was carried out from 2019 to 2023 at the Department of Plant Sciences (IPV) at the Norwegian University of Life Sciences.

First, I would like to thank my primary supervisor, Prof. Morten Lillemo. Your combination of hard work, persistence, intuition, experience, expertise, common sense and sharp wit were and remain genuinely inspiring. Thank you for exposing me to all aspects of academia: teaching, organization, project development, conferences and workshops, courses, the "Eureka!" moments, and the moments when reality does not match the hypothesis. Thank you also for allowing me to explore, make mistakes, learn from them, and make different mistakes.

I thank my co-supervisor, Dr. Sahameh Shafiee for sharing your views on my work and for discussions. I want to thank Dr. Jose Crossa and his team in CIMMYT, Mexico, for the kind, excellent, and swift feedback on my work and for an invitation to visit CIMMYT. I regret that I could not experience the excellent scientific environment in person, but I trust the chance will come again.

I want to thank the crew of SKP Friland and especially Svend Pung and Cecilie Yri for excellent help with experiments and being the silent heroes science needs. I want to thank students Oda Margarete Lund, Lavanathan Rathy, Peder Bukaasen, Guro Størdal, Simon Pedersen, and Shailaja Thapa for great help with data gathering and making days on the field even more enjoyable. A special thank you goes to Henrik Lassegård, who started as a student and stayed longer.

Special thanks go to the caffeine and coffee machines in the office, which suffered severe abuse. This work would not have been possible without you.

My family, especially my parents, Lidia, and Jacek Mróz, deserve my gratitude. I know that if the world gets too big, a place by a warm fire and a shoulder to cry on is always there waiting for me. For this, dziękuję Wam.

Last, I would like to express my sincere gratitude to min kjæreste, Mari Nordskog. Your patience and support were simply unparalleled. Tusen takk. I am one lucky man to have you.

Abstract

Bread wheat is among the most critical staples worldwide and is essential to Norway's food production. The aftermath of the green revolution nearly eradicated Norwegian wheat production in the 1950s and 1960s. The lack of fit between Norwegian wheat varieties at that time and the new agronomic practices rendered Norwegian wheat production uncompetitive, which spurred intense breeding efforts. The efforts resulted in the release of two landmark wheat varieties: Runar and Reno, released in 1972 and 1975, respectively, which brought back domestic wheat production in Norway. Since then, new varieties have been bred and released on the market, but we know little about the achieved breeding gains or their genetic background.

To fill this gap, we tested 21 historical and modern wheat varieties present on the Norwegian market since the release of Runar in a multi-year field trial under two fertilization levels (75 and 150 kg N ha^{-1}). We detected an annual genetic progress in grain yield (GY) of 18kg ha^{-1} (0.34%), which did not rely on agronomic input. The progress was primarily associated with increased number of grains per spike and area. New varieties in Norway can also take advantage of the changing climate (earlier spring) by extending their vegetative periods by three days on average.

Genome-wide association analysis on a diverse panel of over 300 lines could not fully explain the genetic progress in GY. However, it discovered several significant loci associated with GY and other traits. Validation of these loci using recent breeding lines confirmed their effects and showed potential for using them in marker-assisted selection. Two loci, representing the *Ta-Col5* and *Ta-GS5-3A* genes (associated with spike architecture and kernel size), showed a change in allele frequency over the last 100 years, indicating possible historical selection pressure.

We investigated the prospects of accelerating genetic gains using multispectral drone imaging data for standalone GY prediction and augmenting genomic selection models. The multispectral data allows for GY prediction with accuracy comparable to genomic data, and models using both genomic and phenomic variates appeared superior. We discussed the biological rationale behind this GY prediction and its practical aspects.

This work provides insight into historical changes in GY and other traits in Norwegian spring wheat, and describes a semi-novel way of using phenomic data for GY prediction and augmenting GP to accelerate genetic gains in GY.

Norsk sammendrag

Hvete er blant de viktigste matvekstene i verden, og spiller en nøkkelrolle i norsk matproduksjon. Mekaniseringen av jordbruket på 50- og 60-tallet nærmest utryddet norsk hveteproduksjon. Norske kornsorter var dårlig tilpasset de nye agronomiske dyrkingsmetodene, og gjorde norsk hveteproduksjon lite konkurransedyktig. Dette førte til intensivt foredlingsarbeid som resulterte i to milepælsorter: Runar fra 1972 og Reno fra 1975, og norsk hveteproduksjon tok seg sakte, men sikkert opp igjen. Siden den gang har nye og forbedrede sorter blitt utviklet og sluppet ut i markedet, men vi vet lite om det genetiske grunnlaget for den observerte avlingsframgangen.

For å fylle dette gapet har vi testet 21 historiske og moderne hvetesorter som er eller har vært tilgjengelige på det norske markedet siden Runar i et flerårig feltforsøk med to gjødslingsnivåer (7.5 og 15 kg N per dekar). Vi fant en årlig genetisk framgang i kornavling på 1.8 kg per dekar (0.34%), og denne var uavhengig av gjødslingsnivå. Økningen var i hovedsak assosiert med økt antall korn per aks og areal. Nye sorter kan også utnytte klimaendringene gjennom en i gjennomsnitt 3 dager forlenget vekstsesong.

Assosiasjonskartlegging i en samling med over 300 vårhvetelinjer kunne ikke helt forklare den genetiske avlingsframgangen, men avdekket flere viktige gener assosiert med kornavling og andre agronomiske egenskaper. Vi validerte disse assosiasjonene i nye foredlingslinjer, noe som bekreftet deres effekt og viste potensiale for å bruke dem i markør-assistert seleksjon. To kromosomområder som representerte genene *Ta-Col5* og *Ta-GS5-3A* (assosiert med aksstruktur og kornstørrelse) viste endringer i allelfrekvens i løpet av de siste hundre årene, noe som kan tyde på at de har vært under historisk seleksjonspress.

Vi så på muligheten til å øke den genetiske framgangen ved å bruke multispektrale dronebildedata til å predikere kornavling og forbedre modellene brukt i genomisk seleksjon. Multispektrale dronebilder gjør det mulig å forutse kornavling med nøyaktighet tilsvarende genomiske data, og modeller som benytter både genomiske og fenomiske data utpeker seg som mest effektive. Vi diskuterte biologiske sammenhenger som kan forklare hvordan denne metoden virker til å predikere kornavling og dens praktiske aspekter.

Dette arbeidet gir innsikt i historiske endringer i avling og andre egenskaper i norsk vårhvete, samt presenterer en delvis ny måte å bruke fenotypiske data på til å forbedre genomisk prediksjon av kornavling og øke den genetiske avlingsframgangen.

Abbreviations

AFLP – amplified fragment length polymorphism
AI – artificial intelligence
BLUP – best linear unbiased predictor
CIMMYT – International Center for Wheat and Maize Improvement
DArT – diversity array technology
DH – doubled haploid
DL – deep learning
GBLUP – genomic best linear unbiased predictor
GBS – genotyping by sequencing
GEBV – genomic-estimated breeding value
GEI – genotype by environment interaction
GP –genomic prediction
GPU – graphics processing unit
GrpA – grains per area
GrPS – grains per spike
GS – genomic selection
GWAS – genome-wide association study
GxE – see GEI
GY – grain yield
H^2 – broad sense heritability
HI – harvest index, ratio of grain yield to total plant biomass
HTP – high-throughput phenotyping
HYV – high-yielding variety
IRRI – International Rice Research Institute
IWGSC – International Wheat Genotyping/Sequencing Consortium
LD – linkage disequilibrium
LED – light-emitting diode
MAS – marker-assisted selection
ML – machine learning
MLM – mixed linear models
MQTL – meta-QTL analysis
NASA – National Space Agency
NDVI – normalized difference vegetation index
NMBU – Norwegian University of Life Sciences
QTL – quantitative trait locus (s.)/loci (pl.)
RAPD – random amplified polymorphic DNA
RFLP – restricted fragment length polymorphism
RGB – red, green, and blue
Rht – reduced plant height (genes)
SNP – single nucleotide polymorphism
SSR – simple sequence repeat
TKW – thousand kernel weight
TPE – target population of environments
UAV – unmanned aerial vehicle, aka. drone
VI – vegetation index

List of papers

I. **Historical grain yield genetic gains in Norwegian spring wheat under contrasting fertilization regimes**
Mróz, T., Dieseth, J.A., and Lillemo, M. (2022)
Crop Science, 62, 997–1010.
https://doi.org/10.1002/csc2.20714

II. **Grain yield and adaptation of spring wheat to Norwegian growing conditions is driven by allele frequency changes at key adaptive loci discovered by genome-wide association mapping**
Mróz, T., Dieseth, J.A., and Lillemo, M.
Manuscript.

III. **Multispectral-derived genotypic similarities from budget cameras allow grain yield prediction and genomic selection augmentation in single and multi-environment scenarios in spring wheat**
Mróz, T., Shafiee S., Crossa J., Montesinos-Lopez , OA., and Lillemo, M.
Manuscript.

1 Introduction

1.1 Current situation in small grain production in Norway

On the western side of the Scandinavian Peninsula, Norway stretches between latitudes 57°58' and 71°10'N with a total distance of 1,752 km from south to north. Out of its surface area of about 324,000 km^2, only 3% (about 1,000,000 ha) of Norwegian land is arable soil, with most of the land covered by mountains (44%), forests (38%), lakes (6%) and wetlands (6%). Unlike other lands situated at a similar latitude, the climate of Norway is relatively warm due to the presence of the Gulf Stream. The main agricultural area (in the country's southeastern part) is separated from the west coast by tall mountain ranges, creating conditions that resemble a continental climate with less rainfall and higher temperature differences between winter and summer. The area in the southeast constitutes over half of the country's arable land and is where most of the cereal production occurs. Other agriculture-intensive areas in Norway are found in Trøndelag in central Norway (with a mix of livestock and barley production) and Jæren on the southern part of the west coast, with intensive livestock production (Figure 1).

Glaciers covered Norway's surface entirely during the last ice age; therefore, most of today's soil is of recent origin, formed by glacial deposits created by the retraction of the glaciers some 10,000-20,000 years ago. The main agricultural area in the south rose around 200 m out of the sea when the glaciers melted, and thus, most of the produce is grown on clay soils of marine origin. In some areas, the clay is covered by moraines deposited during short periods of glacial expansion. Sandy soils are at the bottom of valleys, where rivers change their course (Lillemo & Dieseth, 2011).

Only 30% of Norway's arable land is suitable for grain farming, resulting in only 45.9% (on average) of the grain demand being satisfied by national production. Improving Norwegian self-sufficiency in grain production is an important political goal.

Barley, wheat, and oats are currently (in descending order) the most farmed crops in Norway, with a residual production of rye (Figure 2).

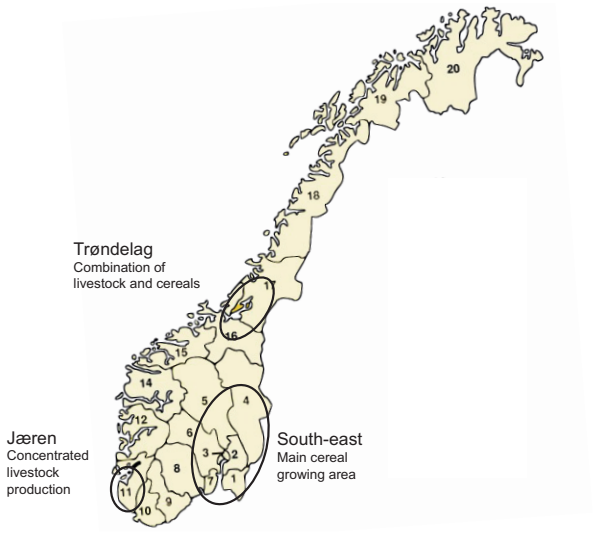


Figure 1 *Main agricultural areas in Norway: production type, arable land, and area share of wheat production. Taken from Lillemo and Dieseth (2011)*

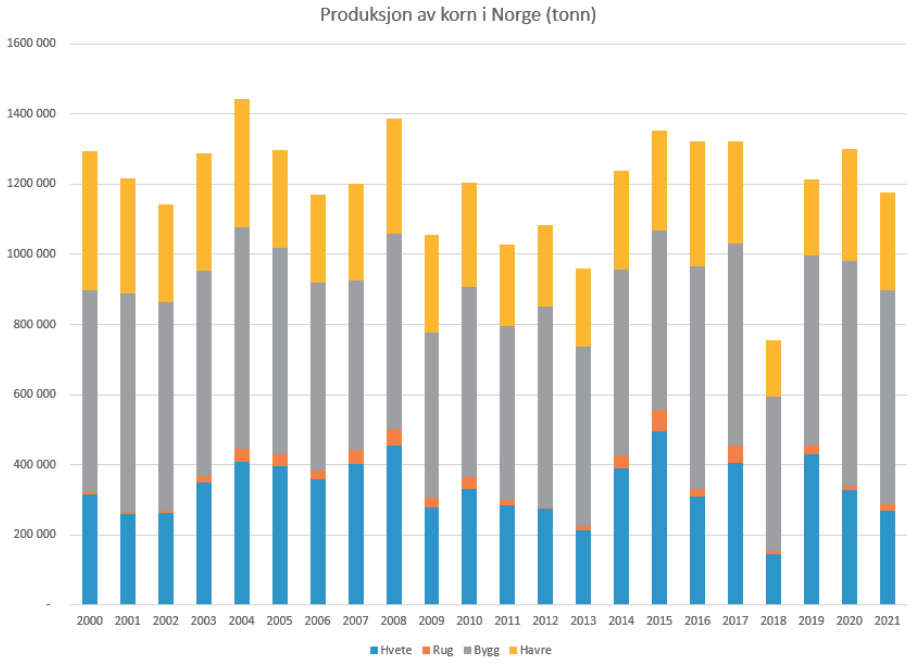


Figure 2 *Production of small grain crops in Norway in tons between 2000 and 2021. Hvete – wheat, rug – rye, bygg – barley, havre – oats. The color indicates the crop: blue – wheat, orange – rye, grey – barley, yellow – oats. Data source: Agriculture Authority of Norway. Figure source: https://brodogkorn.no/fakta/kornproduksjon-i-norge*

1.2 Grain yield in Norwegian small grains

According to the old records, the grain yield (GY) level in practical farming in Norway remained constant until the turn of the nineteenth century. The acreage of small grains and the agricultural practice were stable, with no significant development taking place for hundreds of years (Strand, 1964).

Experimental work in Norwegian agriculture was initiated in 1898, followed by the introduction of artificial fertilizers and the availability of improved varieties. Although these pioneering changes initially had minimal effect on the actual production, primarily due to slow work progress and the farming community's reluctance, the grain progress rate increased with time (Strand, 1964). The increase in GY was especially pronounced in the decade following stagnation during the Second World War (Figure 3).

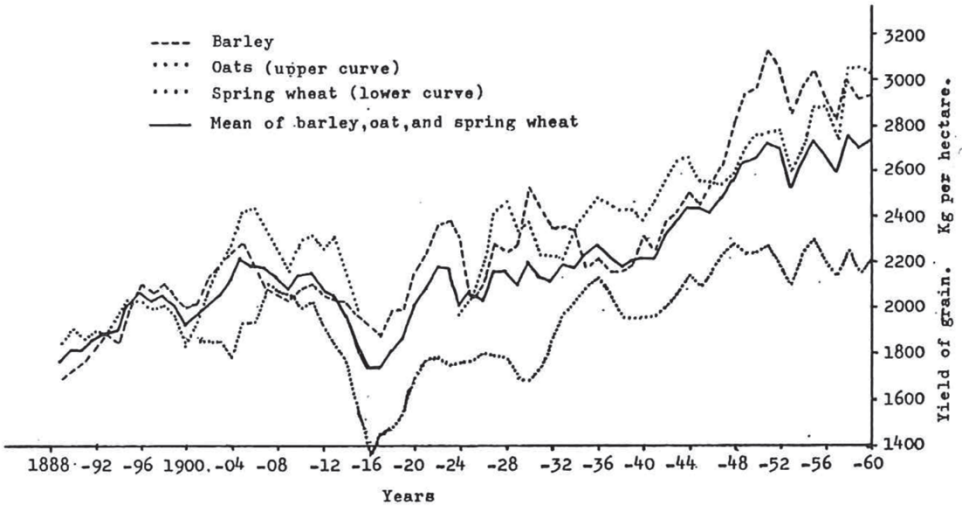


Figure 3 *Estimated productivity of seasons from 1888 until 1960 for barley (dashes), oats (dots, upper curve), spring wheat (dots, lower curve), and mean of the crops (solid line). Taken from Strand (1964)*

Between 1888 and 1960, the average GY for small grains rose from approximately 1.8 t ha^{-1} to almost 2.8 t ha^{-1} (Figure 3). The release of new varieties contributed to 34, 35, and 53% of the total progress in GY for barley, oats, and spring wheat, respectively (Strand, 1964).

The yields in Norwegian small grains continued to rise in the following period from 1960 to 1992. During this period, the average GY of barley, oats, and spring wheat increased by 1.8, 1.4, and 2.1 t ha^{-1}, respectively, corresponding to a linear increment

of 70, 56, and 74 kg ha^{-1} per year. The progress was attributed to new cultivars, agriculture practice, and interaction between those factors. The GY increase depended mainly on the introduction of new varieties (40-57%) and cultivation practice (31-60%), while the interaction term was less significant, especially in oats (Table 1) (Strand, 1994).

Table 1 *Gains in Norwegian small grains GY in the 1960-1992 period (GY increase) in barley, wheat, and oats, attributed to the release of new cultivars (G), improvements in management practice (M), and the interaction between G and M (G × M). Adapted from Strand (1994)*

Crop	GY Increase		G		M		G×M	
	Per year[a]	Total in period[a]	%	Kg in period[a]	%	Kg in period[a]	%	Kg in period[a]
Barley	7.0	180	40	72	50	90	10	18
Wheat	7.4	214	47	100	31	66	22	48
Oats	5.6	141	42	60	60	84	-2	-3

[a] *kg per daa*

Dissection of GY progress into the three primary statistical components underlined the importance of developing new cultivars and improvements in agricultural practice

It is important to mention that the 1960-1992 period coincided with the global Green Revolution, which changed the varieties and agronomic practices.

1.3 The Green Revolution

The term "Green Revolution" (also known as The Third Agricultural Revolution) was coined in 1968 by William S. Gaud of the US Agency for International Development. The term describes the whole of new technology introduction and agricultural policy implementation between the 1940s and 1960s., which resulted in drastic increases in yields in food crops, particularly in developing countries (Dalrymple, 1985; Eliazer Nelson, Ravichandran, and Antony, 2019).

To increase agricultural productivity, novel high-yielding wheat and rice varieties were introduced during the Green Revolution, developed by the International Maize and Wheat Improvement Center (CIMMYT), Mexico, and the International Rice Research Institute (IRRI), Philippines. These new varieties had doubled yield potential due to higher responsiveness to nitrogen fertilizers (Dalrymple, 1985; Davies, 2003).

The increase in genetic GY potential is associated with several traits in green-revolution varieties. For rice and wheat, the GY potential improvement was caused by increased biomass production and plant height reduction, thus improving the harvest index (HI). Pre-Green Revolution rice and wheat varieties were tall and leafy, with fragile stems and HI of around 0.3. They could achieve a total biomass of 10-12 t ha^{-1}, resulting in a yield potential of approximately 4 t ha^{-1}. However, when exposed to excess nitrogen fertilization, these varieties tillered intensely, grew excessively tall (which led to early lodging), and yielded less (Khush, 1999).

The new varieties, on the other hand, had HI of 0.5, closer to the theoretical limit of 0.6 (Slafer *et al*., 2023). Because of their shorter and stronger straw, they could withstand intensive nitrogen fertilization without the risk of lodging. The most critical factor behind HI improvement was the change in plant architecture, which doubled the GY potential in rice and wheat varieties (Khush, 1995). Incorporating disease, insect resistance, and abiotic stress resistance (for instance, moisture stress, salinity, alkalinity) genes increased yield stability in the new varieties.

The revolution in agriculture did not happen due to improved genetics alone. It was necessary to develop appropriate management practices to utilize the yield potential of the new cultivars, such as: (i) knowledge of plant nutrient requirements for each primary soil type, (ii) gathering data on agronomical traits needed to increase GY – optimum rates, dates and methods of sowing, (iii) irrigation practices, (iv) pest and weed control, and (v) mechanization in land preparation, sowing, and harvesting (Khush, 1999).

The best-described genes that played a significant role in the Green Revolution are *Rht-B1b* and *Rht-D1b* (Reduced height, previously known as *Rht1* and *Rht2*, respectively) (Flintham, Angus, and Gale, 1997; Y. Wang *et al.*, 2014). These two gibberellin-insensitive dwarfing genes reduce lodging by decreasing plant height and are associated with increasing grain number, GY, and HI (Chapman *et al.*, 2007). The *Rht* genes' positive effects on GY and grain number can be explained by reducing the internal competition for assimilates between the developing spike and the vegetative part of the plant, which allows the survival of a higher number of florets per ear to be filled after pollination (Fischer & Quail, 1990). The *Rht-B1b* and *Rht-D1b* genes originate from Japan. In the 20th century, a semi-dwarf wheat line, Daruma, was crossed with an American variety Glassy Fultz, producing "Futlz Daruma". "Fultz Daruma" was crossed with "Turkey Red" (a leading US cultivar imported from Russia) to produce "Norin-10" (Wilhelm *et al.*, 2013). "Norin-10" was, in turn, used to produce many high-yielding varieties, which gradually spread the *Rht-B1b* and *Rht-D1b* genes

worldwide. Nowadays, the Norin-10 dwarfing genes are present in more than 70% of commercial wheat varieties worldwide (Arnold, 1999).

"Daruma", which is thought to be the donor of both *Rht-B1b* and *Rht-D1b* genes, was present in Japan as early as 1894 (Kihara, 1983). However, the alleles might have first occurred much earlier, already in the 3rd-4th century AD, in a native Korean wheat population "Anzunbaengimil" ("crippled wheat") (Cho *et al.*, 1980). Both *Rht-B1b* and *Rht-D1b* genes are thought to be products of spontaneous mutations in wheat populations originating from Japan, China, and Korea (Kihara, 1983).

The Green Revolution echoed throughout the world and, with exceptions, tripled the average cereal crop yields with only a 30% of the cultivated land increase, leading to poverty reduction and lower food prices. The caloric availability would have declined by 11-13% without the Green Revolution (John & Babu, 2021). Thousands of hectares of land would otherwise have undergone agricultural transformation to meet food and fodder demand (Pingali, 2012).

The impacts of the Green Revolution were most pronounced in developing countries (Figure 4). For instance, India was able to transit from a "ship-to-mouth" system (dependency on imported grains) to self-sufficiency (Brainerd & Menon, 2014).

The Green Revolution, however, had negative environmental impacts, such as increased pesticide use (Choudhary *et al.*, 2018), water consumption (Davis *et al.*, 2018), air pollution and greenhouse gas emissions (de Miranda *et al.*, 2015), and the extinction of indigenous varieties of crops (Prasad, 2016), to name but a few. Those concerns, however, have been raised mainly during the last two decades, as most of the effects were not visible immediately, and the general awareness of environmental impacts had previously not been as high as today.

The Green Revolution laid the foundations for modern agriculture, allowing it to keep up with the ever-increasing demand for food worldwide due to increasing human population and at the same time preventing the agricultural area from expanding (Figure 5).

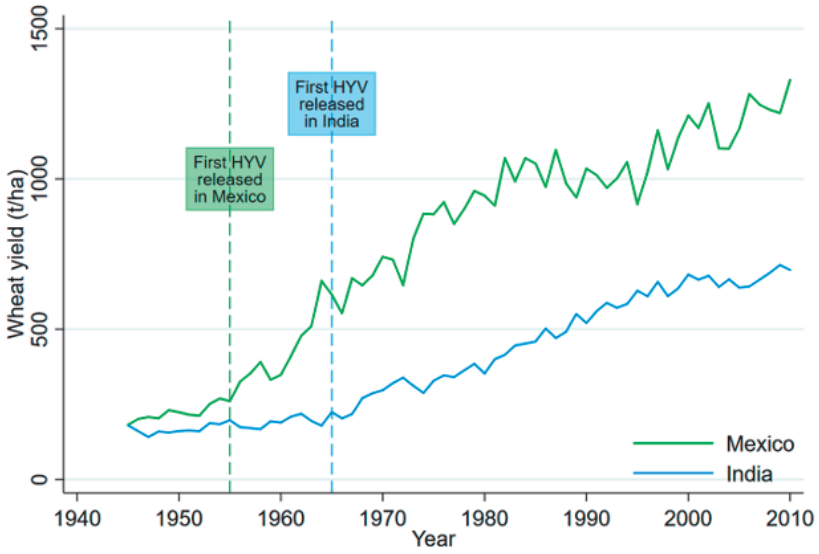


Figure 4 *Average grain yield in wheat from the 1940s to 2010 in Mexico and India, and the impact of introducing the Green Revolution's high-yielding varieties (HYVs). Taken from Hansen, Wingender and Gollin (2021)*

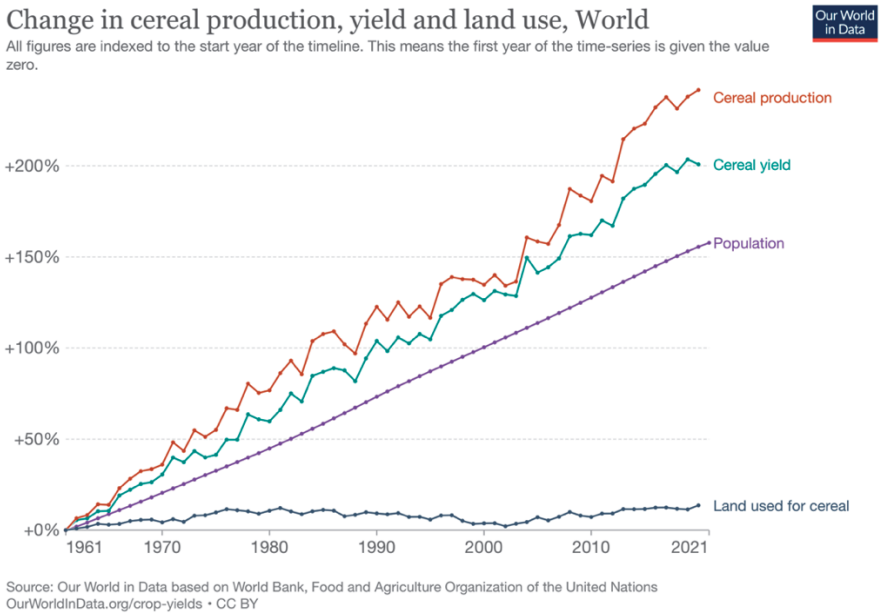


Figure 5 *Worldwide change in cereal production, yield, population, and land used for cereal production since the Green Revolution. Taken from Richie (2017)*

The overall global increase in cereal GY is a collective effort. Breeders and agronomists worldwide work to create genotypes achieving high GY in various environments. Even though the Green Revolution paved the general path towards achieving high GY, the resulting genetic gains in GY were created by many institutions worldwide, often independently. It is, therefore, worthwhile to investigate the GY increase in individual collections.

1.4 Genetic gains in grain yield

The ultimate goal of any breeding program is to produce superior genotypes in terms of adaptation, resistance, end-use qualities, and yield. Those improvements should be visible (measurable) in time. Estimating the rate of genetic gain allows for regular evaluation of breeding programs for their performance and improving their efficiency. There are several approaches to performing this estimation. For instance, Cargnin, de Souza, and Fronza (2008) used regression analysis on the GY of the top five varieties divided by field trial mean and multiplied by 100. Tadesse *et al*. (2010) developed and used the concept of "success rate" (when the newest variety performed best) in field trials. However, the most common method of estimating genetic gains is assigning each accession in a collection a value that places it on the timeline – most often the year when a line became a variety (became officially accepted for cultivation by local authorities) – and then performing regression analysis of this value against achieved GYs (Tadesse *et al.*, 2019).

Another question concerns the environment (management) used to test the varieties. Each variety is developed to be competitive under particular management where it can approach its yield potential. However, agricultural practices have changed over the years due to technical and regulatory reasons. Is it then appropriate to have a collection of varieties spanning a longer period in a single field trial under single management, as done in most studies? On the one hand, some varieties will have the advantage of being grown under their optimal management, while others not, possibly leading to conclusions not reflecting the cropping reality. On the other hand, a single field trial under single management allows one to perform a back-to-back comparison among the accessions, showing how efficient the varieties are in coping with the specific management (environment). There is no simple answer to this question. However, for practical reasons, most studies are carried out either under a single management (for example, Yao *et al.,* 2019) or a few managements (for example, Voss-Fels *et al.*, 2019).

Perhaps the most comprehensive study of genetic gains in GY was conducted by Voss-Fels *et al.* (2019), where 191 wheat cultivars registered as varieties in Western Europe between 1960 and 2013 were tested under three managements. This work extensively analyzed germplasm originating from one of the world's most essential and most intensive wheat production areas, deserving to be described here in detail.

The study showed clear progress in GY in the period due to the introduction of new varieties (approximately 800 kg ha^{-1} in total, 15 kg ha^{-1}, 0.2% per year), which was also visible in sub-optimal conditions (lack of fungicide treatment, low N fertilization) (Figure 6). The slopes of the regression lines for each environment are almost parallel, indicating very minor genotype x environment interactions (different ranking of varieties in a collection in different environments). New varieties are consistently better in terms of GY, even in suboptimal conditions (Voss-Fels *et al.*, 2019).

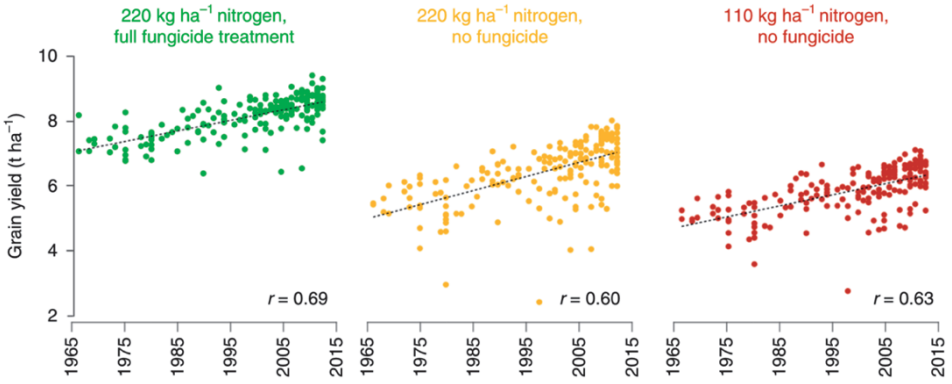


Figure 6 *Grain yield genetic gains in Western-European wheat between 1965 and 2013 under three managements (environments). Taken from Voss-Fels et al. (2019)*

There is strong evidence of GY's significant genetic progress in collections worldwide, spanning various time intervals during the 20th and 21st centuries (collected in Table 2). The studies use various genotype collections and methodologies but reach a similar conclusion: new varieties are better (at least in terms of GY) and can use available resources more efficiently. In wheat, the estimated value of the gains in GY varies between 5 and 115 kg ha^{-1} per year (0.19 - 2.80%, Table 2).

Table 2 *Genetic gains in wheat grain yield in various collections worldwide. GrPS – grains per spike, GrpA – grains per area, SppA – spikes per area, TKW – thousand kernel weight*

Country/Region	Period	Annual gain kg/ha/y(%)	Associated GY components	Reference
Argentina	1984-1994	23.6 (- %)	GrpA	Abbate *et al.* (1998)
Argentina	1934-2015	26.9 (1.35 %)	GrPS	Achilli, Roncallo, and Echenique (2022)
Asia/China	1950-2005	53.8 (1.25%)	GrPS, TKW	Tian *et al.* (2011)
Asia/China/Hebei	1964-2007	47.4 (0.72%)	GrPS, TKW	Yao *et al.* (2019)
Australia/NSW	1901-2014	26.0 (0.40%)	GrpA, TKW, GrPS	Flohr *et al.* (2018)
Brazil	1976-2005	48.0 (1.84%)	-	Cargnin, de Souza, and Fronza (2008)
Canada	1988-2018	38.0 (0.68%)	-	So *et al.* (2022)
Chile	1965-2019	70.2 (0.49%)	-	del Pozo *et al.* (2022)
CIMMYT	1994-2010	31.0 (1.00%)	-	Manès *et al.* (2012)
CIMMYT	1995-2009	27.8 (0.65%)	-	Sharma *et al.* (2012)
CIMMYT	1977-2008	34.7 (0.70%)	TKW	Lopes *et al.* (2012)
CIMMYT/ Afghanistan	2002-2016	115 (2.80%)	-	Sharma *et al.* (2022)
CIMMYT/ Worldwide	1979-1997	5.30 (0.19%)	-	Trethowan *et al.* (2002)
Europe/France	1980-2010	12.8 (2.60%)	-	Brisson *et al.* (2010)
Europe/Germany	1985-2007	28.0 (0.34%)	-	Ahrends *et al.* (2018)
Europe/Turkey	1964-2010	30.9 (0.62%)	SppA, GrPS	Akin *et al.* (2017)
Europe/UK	1972-1995	120 (1.20%)	GrpA	Shearman *et al.* (2005)
India	1900-2016	24.3 (0.8%)	-	Yadav *et al.* (2021)
Mexico	1950-1982	59.0 (1.10%)	GrpA, GrPS	Waddington *et al.* (1986)
Mexico	1962-1988	67.0 (0.88%)	GrpA, SppA, GrPS	Sayre, Rajaram, and Fischer (1997)
Mexico/CIMMYT	1966-2009	30.0 (0.59%)	TKW	Aisawi *et al.* (2015)
Northern Europe	1961-2005	35.3 (-%)	-	Peltonen-Sainio, Jauhiainen, and Laurila (2009)
South Africa	1998-2013	0.82%	TKW	Dube *et al.* (2019)
Western Europe	1960-2013	15.1 (0.21%)	GrPS, GrpA	Voss-Fels *et al.* (2019)

However, a question arises: how is the breeding progress achieved, and what are the traits associated with progress in GY?

To answer this question, attempting to dissect GY into more basic components is necessary. GY is a product of two primary components: the number of grains produced per unit area and grain weight. The number of grains per area is a product of the number of fertile tillers per area and the number of grains per spike. Likewise, grains per spike is a product of the number of florets per spike and floret fertility

(proportion of florets that successfully form a grain). Therefore, increasing yield components will increase GY (Tillett *et al.*, 2022). However, it is typical to observe that yield components compensate for each other.

The studies which attempted to point out the traits associated with GY (Table 2) most often report an increase in grains per spike (GrPS, 8 out of the analyzed 13 works), grains per area (GrpA, 6/13), and kernel weight (TKW, 6/13). It is also remarkable that a simultaneous increase in GrPS/GrpA and TKW was reported in 3 cases.

1.5 Grain yield plateau and perspectives for the future

Global agricultural production has to be increased by 60-110% by 2050 to provide for the changing world: to feed the growing population, to supply meat, dairy, and biofuel production, as well as to provide food security to over 870 million chronically undernourished people (Pingali, 2007; Charles *et al*., 2010; Liniger *et al*, 2007; Foley *et al.*, 2011; Tilman *et al.*, 2011). Doubling of the agricultural production translates to roughly a 2.4% increase per year, far higher than the current rate of yield increase in wheat (0.9%), indicating a strong need for yield increase without expanding the agricultural area (Ray *et al.*, 2013). At the same time, there are signs of stagnating yields in the major crops, especially in the most intensive and productive systems in the world (Grassini, Eskridge, and Cassman, 2013), which, together with rising challenges posed by climate change, calls for a drastic increase in yields (Braun, Atlin and Payne, 2010).

1.6 Genotype x environment interactions (GEI)

Genotype-environment interaction (GEI or GxE) is a universal phenomenon that relates to all living organisms: genotypes and environments interact to produce an array of phenotypes. When responses of two genotypes to different levels of environmental qualities are compared, GEI can be statistically described as the failure of the two response curves to be parallel (Baker, 1988). In practical terms, GEI can be defined as a difference in genotype ranking in different environments. GEI is of paramount importance in plant breeding - collections of genotypes are tested in an array of various environments, which implies the presence of GEI.

The importance of GEI is greatly summarized by Gauch and Zobel (1996):

> *"Were there no interaction, a single variety of wheat or corn or any other crop would yield the most the world over, and furthermore the variety trial need be conducted at only one location to provide universal results. And were there no noise, experimental results would be exact, identifying the best variety*

without error, and there would be no need for replication. So, one replicate at one location would identify that one best wheat variety that flourishes worldwide."

GEIs can be grouped in two broad categories: non-crossover and cross-over interactions (Figure 7). The crossover interaction is the most important in plant breeding (Kang, 2002).

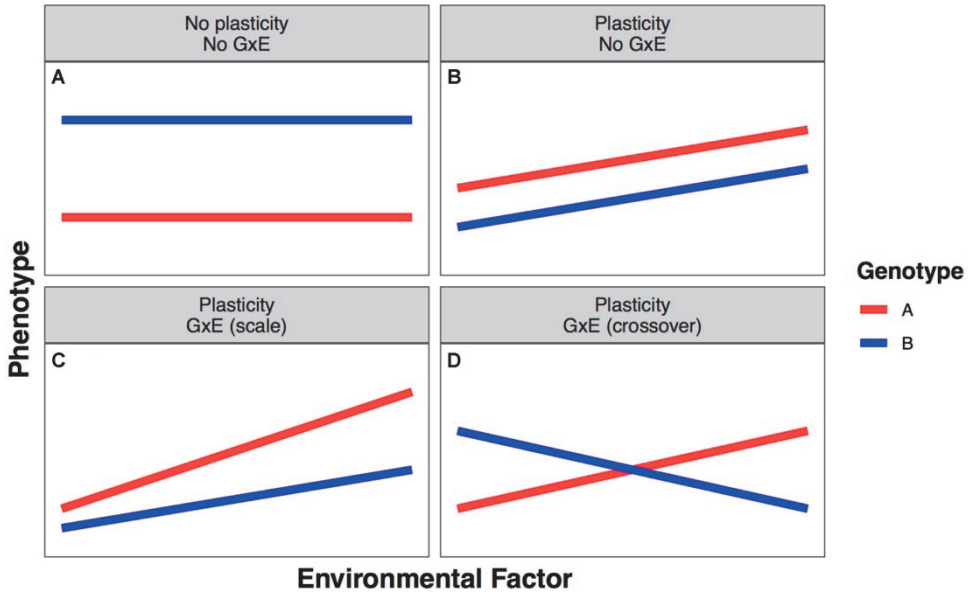


Figure 7 *Types of genotype x environment interactions (GEI, GxE): no GEI (AB), plasticity (C, scaling response without crossover), and crossover GEI (D). Taken from Kusmec, de Leon and Schnable (2018)*

A large presence of GEI indicates the need for testing the breeding material in numerous environments (locations over several seasons), therefore has a dramatic impact on all stages of a breeding program. Significant GEI also affects trait heritability negatively – the larger the GEI component, the smaller the heritability estimate. Therefore, GEI limits breeding progress in light of the Breeder's equation described in the next section (Kang, 2002).

Considering a large presence of GEI, a plant breeder faces a choice: to develop separate populations for each site type where environmental qualities cause different rankings, or to select genotypes that generally perform best across many sites. The first solution is expected to reach higher genetic gains but is also more costly. The second alternative yields lower gains, but at a lower cost.

1.7 Accelerating genetic gains through breeding

The sole purpose of breeding is to develop superior genotypes concerning target traits. Therefore, only traits influenced by genetics – heritable- are interesting for breeders.

Plant breeders often use heritability (repeatability) to quantify the precision of field trials or series thereof. In its simplest form, it is the ratio of trait variance explained by heritable genetic effects and the total trait variance (narrow-sense heritability). Analogically, broad-sense heritability is the proportion of trait variance associated with an effect for the whole genotype (including the sum of epistatic, dominance, and additive effects) (Nyquist, 1991). Heritability is a critical parameter in breeding as it directly influences the response to selection (Piepho & Möhring, 2007). Because the phenotype ultimately emerges from a complex interplay between many genes (and GEI), it was compelling to take a statistical approach to estimate the genetic gain (progress) achieved by breeding. It takes the form of the "Breeder's equation" (Equation 1, by Lush, 1943):

$$\Delta Z = H^2 S \qquad (Equation\ 1)$$

where Z is the mean of a trait in the population, ΔZ is the change in the mean over one whole generation, H^2 is the trait heritability, and S is the selection differential.

For practical purposes, the Breeder's equation can also be written as Equation 2:

$$\Delta G = \frac{i \cdot r \cdot \sigma_A}{L} \qquad (Equation\ 2)$$

where ΔG is the genetic gain, i is selection intensity, r is the selection accuracy, σ_A is the genetic variation, and L is the generation interval. It is essential to mention that Equation 2 is not strictly mathematically accurate, but helps to understand the main rules that govern the breeding progress.

A successful breeding program requires a broad additive genetic variation within the breeding population – the higher the genetic variance, the bigger the improvement potential. Selection intensity represents the difference between the trait mean of selected individuals and the population mean, expressed in standard deviations of the population mean. The selection intensity depends on the number of candidates available for selection, and selection response increases with intensity; however, at the cost of making fewer selections and reducing the genetic variance carried over to the next generation. The accuracy of selection describes how well the selection criteria represent the *true* breeding values of the selections and is directly connected

to trait heritability. Accuracy of selection is high for highly-heritable traits but has to be supplemented with extra information (such as genomic or pedigree) for traits with low heritability. Generation interval, in general, indicates the average age of the selections at the birth of their offspring (in plants: producing viable seeds). The shorter the breeding cycle interval, the faster the breeding progress (Falconer & Mackay, 1996).

The following sub-chapters will describe how to accelerate genetic gains by acting upon particular components of the Breeder's equation.

1.7.1 Need for speed in plant breeding

Shuttle breeding

A typical wheat variety takes over twelve generations of selfing to develop. The process includes three phases: crossing and inbreeding (approximately six generations of self-pollination to reach homozygosity and eliminate segregation), testing phase (screening for target traits in multi-environmental field trials), and bulking up the seed (Alahmad *et al*., 2022). Therefore, the progress of a breeding program is mainly limited by the number of generations that can be grown per year. The ability to conduct more than one generation of selfing in a single year would increase the genetic gains dramatically.

Shuttle breeding was introduced in 1946 by Norman Borlaug to shorten the time of developing varieties for Mexican farmers (Rajaram, Hettel, and International Maize and Wheat Improvement Center, 1995). This methodology attempts to turn over two plant generations per year by planting in contrasting environments (in terms of altitude, latitude, and precipitation). It has successfully reduced the time to complete a breeding cycle by 50% (Borlaug, 2007), doubling the genetic gains in light of the Breeder's equation (Figure 8).

This strategy is still being used today by CIMMYT, Mexico (International Maize and Wheat Improvement Center): during the winter season, populations are grown in the Sonora Desert under short days and are selected for disease resistance, grain quality, photoperiod insensitivity, grain yield, and agronomic type. During summer, the populations are planted at Toluca station at higher altitudes to subject the crop to lower temperatures during grain filling and are selected for disease resistance (Ortiz *et al.*, 2007). As an added benefit to cutting the breeding cycle length by half, this strategy also enables the selection of genotypes grown under different soil types, temperatures, photoperiods, and disease pressures. This strategy used by Borlaug led to the development of the high-yielding, disease-resistant, and photoperiod-

insensitive dwarf varieties, which laid the foundation of the Green Revolution (Rajaram, Hettel and International Maize and Wheat Improvement Center, 1995).

Figure 8 *Shuttle breeding strategy adopted by CIMMYT, Mexico, started by Norman Borlaug in the 1940's. Figure taken from Alahmad et al. (2022)*

Speed breeding

The initial cycles after crossing in a wheat breeding program eliminate heterozygosity by self-pollination. Only after breeding lines are stable (inbred) can they be reliably evaluated in multi-environment field trials. Selection under field conditions in those early generations is not the primary purpose, and being able to accelerate the time it takes to develop a stable line in artificial conditions would significantly increase the overall genetic gains of a breeding program (Watson *et al.*, 2018). Speed breeding aims to achieve this by growing multiple generations (cycles) under artificial conditions in a greenhouse, with prolonged photoperiod and elevated temperatures. Under those conditions, plants complete their life cycles faster. Breeders can grow up to 4-6 generations per year compared to usually 1 to 3 generations under field conditions or traditional greenhouses (Alahmad *et al.*, 2022), reducing the "L" in the Breeder's equation and effectively increasing the genetic gains (Figure 9).

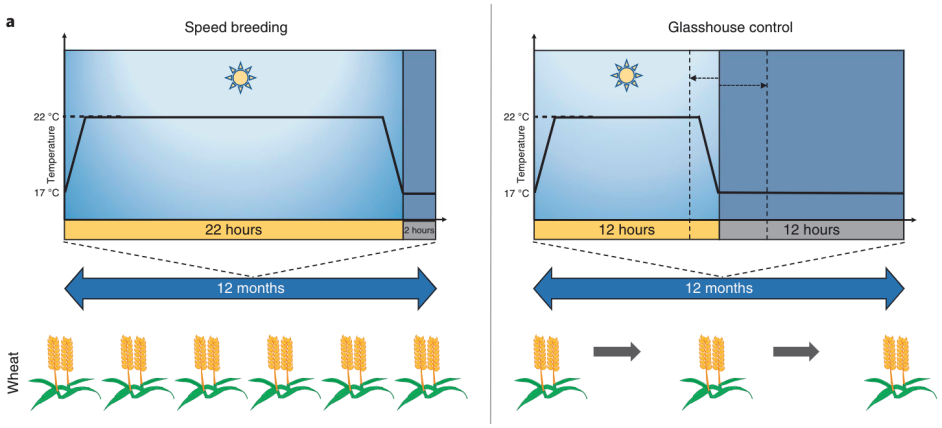


Figure 9 *Comparison between speed breeding approach and a traditional greenhouse number of wheat generations achieved per year. Taken from Watson et al. (2018)*

The idea of speed breeding was inspired by research conducted in the 1980s by NASA (National Aeronautics and Space Administration) and Utah State University, aiming to explore the possibilities of developing a wheat "fast crop", suitable for growing on orbital stations (Hickey *et al.*, 2019). This project yielded the "USU-Apogee" cultivar, which can flower within only 25 days after sowing when exposed to temperatures around 23 °C and continuous light (Bugbee & Koemer, 1997). The development of light-emitting diodes (LED) allows for cost-effective and crop-tailored lighting solutions for speed breeding. Nowadays, detailed protocols are widely available for various crops (Ghosh *et al.*, 2018).

Doubled haploids

The desired homozygosity for self-pollinating crops like wheat is typically achieved by several generations of selfing. Production of doubled haploids is an alternative way to achieve homozygosity immediately by "skipping" the selfing rounds.

A doubled haploid (DH) is created when chromosomes in a haploid cell (1n) are doubled. DHs can occur spontaneously or by colchicine-induced chromosomal doubling, directly producing completely homozygous lines from heterozygous plants in a single generation. Therefore, DHs save at least three to four generations of self-pollination to fix pure lines (Tadesse *et al.*, 2010; El-Hennawy *et al.*, 2011).

Some applied wheat breeding programs use DHs, especially in breeding of winter wheat due to the long generation interval including several weeks of vernalization. However, the problematic establishment of specialized DH laboratories and the cost of DH production limits their application (Rutkoski, Krause, and Sorrells, 2022).

1.7.2 Increasing accuracy – the genomic innovation

The genomic era started in the early 1980s with a breakthrough in the discovery of recombinant DNA technology. The following development of bioinformatics formed the so-called "Next Green Revolution" in plants (Araya *et al.*, 2017). Developing new technologies and tools led to cost reduction and popularizing of these techniques in both research and applied domains.

Hexaploid wheat genome sequencing

The International Wheat Genome Sequencing Consortium (IWGSC) was established in 2005 with the aim of sequencing wheat genomes. It delivered a complete assembly of allohexaploid wheat (cultivar "Chinese Spring") genome (IWGSC RefSeq v1.0, Appels *et al.*, 2018), recently followed by its refined version (RefSeq v2.1. Zhu *et al.*, 2021). The wheat genome sequencing was indeed a milestone, enabling the identification of genes underlying many traits and boosting wheat genetics research and development of selection tools for increasing breeding efficiency (Lukaszewski *et al.*, 2014). IWGSC was a worldwide effort with numerous participants. For instance, Norway, represented by NMBU, contributed by physical mapping and sequencing of chromosome 7B (Belova *et al.*, 2013, 2014).

Further advancements in genomics allowed for studies of pangenomes. For instance, extensive structural rearrangements, introgressions from wild relatives, and differences in gene content of wheat varieties were discovered due to complex breeding histories (Walkowiak *et al.*, 2020). Another study identified over 36 million intervarietal single nucleotide polymorphisms across 18 varieties, with over 140 000 genes predicted in the pangenome (Montenegro *et al.*, 2017).

Despite the decreasing cost, complete genome sequencing is still not widely used in plant breeding. A more technically and economically-viable alternative is molecular markers.

Molecular markers

Genetic markers are heritable biological features determined by allelic forms of genes or genetic loci used to track an individual. Genetic markers used in plant breeding consist of two main types: classical markers (not covered in this introduction) and DNA markers, including restricted fragment length polymorphisms (RFLPs), Amplified fragment length polymorphisms (AFLPs), Random amplified polymorphic DNAs (RAPD), Simple sequence repeats (SSRs), Single nucleotide polymorphisms (SNPs), and Diversity array technology (DArT) (Collard *et al.*, 2005). Recently, genetic

markers can also be discovered by sequencing methodologies (GBS, genotyping by sequencing) (Li *et al.*, 2015).

First discovered in the human genome, SNP markers proved to be an abundant and universally detectable form of genetic variation among individuals of the same species (Rafalski, 2002). Due to their biallelic nature, SNP markers are less polymorphic than SSR markers; however, their abundance and potential to be subjected to automation compensated for this shortcoming (Mammadov *et al.*, 2012). The genomic abundance and density of SNP markers allowed for the development of genetic maps for detecting and mapping quantitative trait loci (QTL). Even though the polymorphisms associated with a trait may not have a causative effect on the trait of interest, they are assumably tightly linked with a genetic structure with a causative effect (Ganal *et al.*, 2012).

The most common way to detect SNPs in plants is array genotyping. SNP arrays (chips) are commercially available and typically consist of thousands of SNPs selected to be highly polymorphic for a given species.

One of the first large-scale SNP genotyping arrays developed for wheat was the Illumina 90K SNP Chip, containing over 81 000 individual SNP markers, with nearly 47 000 placed on a consensus map (S. Wang *et al.*, 2014). Another project concerning SNP arrays was conducted by Allen *et al.* (2017) to provide a set of highly-informative markers for the wheat breeding community. A collection of 820 000 markers (Winfield *et al.*, 2018) was analyzed, and over 35 000 highly-polymorphic and informative markers were chosen to form what was later called the "Wheat Breeder's Array". Currently, the TraitGenetics 25K Illumina SNP chip, consisting of selected highly polymorphic markers from the Illumina 90K and "Wheat Breeder's" 35K array and gene-specific markers for important traits is currently used by many European wheat breeding programs, including Graminor. It was also used for genotyping the material used in this thesis research.

Results from SNP genotyping are an invaluable resource for wheat breeding, which enable marker-assisted selection, QTL discovery, and genomic selection, to name a few, all of which contribute to the rate of genetic gains.

Marker-assisted selection (MAS) – the concept

The concept of MAS was hinted already in the 1980s by Smith & Simpson (1986):

> *"It is unlikely that many of the polymorphisms identified by the new laboratory techniques will be the QTL themselves. However, many of them may be linked to the QTL and so will allow indirect selection".*

In practical terms, MAS relies upon linkage disequilibrium (LD) between a marker and a QTL – that a trait is associated with an allele of a gene, which is in high LD with the detectable marker. Therefore the allele is selected indirectly by selecting the marker as if the effect was caused by the marker (Ben-Ari & Lavi, 2011).

The usefulness of MAS stems from the fact that many traits of interest for breeders are challenging to assess. The possibility of selecting based on a linked marker is appealing. Additionally, selection based on DNA can be performed at an early plant age without the need to observe the actual phenotype (Ben-Ari & Lavi, 2011). However, MAS would not accelerate a classical breeding program as the genotypes still have to conclude their life cycle in the field. The true potential of MAS manifests in conjunction with new technologies such as speed breeding. In speed breeding, it is possible to advance the breeding material several generations in the time it would typically take to conduct one in field conditions. However, the ability to assess phenotypes under speed breeding conditions is limited, therefore selection of candidates based on their DNA rather than phenotypes is appealing. Predicting of phenotypes based on DNA coupled with short cycle time holds great promise to accelerate genetic gains (Rutkoski, Krause, and Sorrells, 2022). The prerequisite for deploying MAS is reliable detection of polymorphisms, performed nowadays primarily by association studies.

GWAS – Genome-Wide Association Studies

The development of genotyping technologies allowed for the gradual replacement of simple sequence repeats (SSR) and diversity array technology (DArT) markers with a significant number of single nucleotide polymorphisms in the form of an array. This provides an effective way to identify associations between various traits and loci through genome-wide association studies (GWAS) (Jin *et al.*, 2016; Li *et al.*, 2016; Quan *et al.*, 2021). GWAS' ability to use natural germplasms allowed for bypassing the time needed to develop biparental populations needed for more traditional linkage analyses (Shi *et al.*, 2017) and made GWAS adopted as the method of choice for searching for genotype-phenotype correlations (Myles *et al.*, 2009). The main advantage of GWAS over linkage analysis is the ability to exploit all the recombination

events that took place during the evolutionary history of the studied population rather than identifying only differences between the parents of a particular cross (Zhu *et al.*, 2008).

Various GWAS models successfully account for population structure and spurious associations while reducing computational effort, including MLM, CMLM (Zhang *et al.*, 2010), MLMM (Segura *et al.*, 2012), SUPER (Q. Wang *et al.*, 2014), GBLUP (Zhang *et al.*, 2007) and FarmCPU (Liu *et al.*, 2016). Except for the stable QTL with significant effects, loci discovered by GWAS for many traits are often population or environment specific. To tackle this challenge, an alternative method of QTL mapping, meta-QTL analysis (MQTL), was proposed. It attempts to collect independently discovered QTL from different populations and environments on a consensus map (Miao *et al.*, 2022), pointing out regions consistently associated with a trait. However, MQTL requires a wealth of GWAS results from various collections and environments.

Genomic selection and genomic prediction

For complex traits, the chances of identifying reliable molecular markers for single loci decrease with the number of genes involved, limiting the possibility of practical use of specific markers discovered in GWAS for making selections (Reynolds & Braun, 2022). Although MAS is a popular approach in molecular breeding (Lande & Thompson, 1990), its use has been limited due to the genetic complexity of many traits controlled by a large number of genes with minor effects (Riedelsheimer *et al.*, 2012). Genomic selection (GS) (Meuwissen, Hayes, and Goddard, 2001), bearing similarity to the BLUP method (Henderson, 1975), takes a vastly different approach than MAS. It aims to improve the breeding germplasm as a whole for all traits of interest over multiple breeding cycles. GS uses genomic-estimated breeding values (GEBVs), which indicate the value of a genotype as a parent based on genomic markers.

Calculating GEBVs requires a training set, that is genotypes that are both genotyped and phenotyped. Based on known genotypes and phenotypes, GEBVs are *predicted* for not phenotyped genotypes (Rutkoski, Krause, and Sorrells, 2022) (Figure 10).

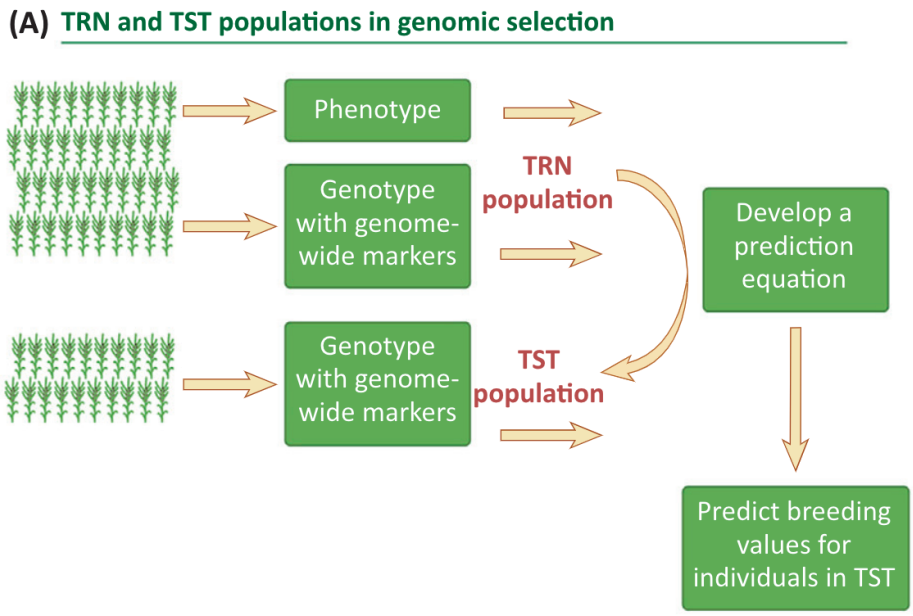


Figure 10 *Idea scheme of genomic selection/prediction. TRN – training, TST – testing populations. Taken from Crossa et al. (2017)*

GS gives a theoretical advantage in plant breeding by accelerating the genetic gains through cost reduction per cycle and shortening the cycle duration (Hickey *et al.*, 2014). A theoretical requirement of GS is that at least one marker is linked with each QTL (Hayes & Goddard, 2001), and all markers' effects are estimated simultaneously using known phenotypes of the training population (Heffner, Jannink, and Sorrells, 2011).

Many GS methods have been developed to improve computing efficiency, speed, and to handle the increasing sizes of available genomic data. Variation and the "curse of dimensionality" (when there are far more variables than records) resulting from thousands of markers can be controlled by either variable selection or shrinkage methods. GS methods differ in their assumptions about the distribution of marker effects and variances (Kaler *et al.*, 2022). Table 3 collects relevant examples of GS methods.

Methods such as ridge regression assume homogenous distribution of marker effects across the genome. In contrast, heterogeneity among markers is allowed in Bayesian methods, with part of the markers having effects drawn from different underlying distributions than the rest (Table 3).

Table 3 *An overview of some of the most popular GS methods and their features. Adapted from Kaler et al. (2022). Methods: Bayes A, Bayes B, Bayes C, Bayesian LASSO (BL) using Bayesian Generalized Linear Regression R package (BGLR_R), and Bayesian Ridge Regression (BRR) using BGLR_R, Bayesian LASSO (BLBL) using Bayesian Linear Regression R package (BLR_R), and Bayesian Ridge Regression (BLRR) using BLR_R, ridge regression–best linear unbiased prediction (rrBLUP), Reproducing Kernel Hilbert Spaces (RKHS) regression, Mixed model using Newton-Raphson algorithm (NR), and Genomic best linear unbiased prediction (GBLUP).*

Method	Main features	Ref.
Bayes A	Utilizes an inverse chi-square (x2) on marker variances yielding a scaled *t*-distribution for marker effects. Similar to BL and in contrast to BRR, it shrinks tiny marker effects towards zero and larger values survive.	Meuwissen, Hayes, and Goddard (2001)
Bayes B	Similar to Bayes A, uses an inverse x2 resulting in scaled t-distribution. Unlike Bayes A, utilizes both shrinkage and variable selection.	Meuwissen, Hayes, and Goddard (2001)
Bayes C	Applies both shrinkage and variable selection methods. Characterized by a Gaussian distribution. Bayes B and Bayes C consist of point of mass at zero in their slab priors.	de Los Campos *et al*. (2013)
BLBL_BLR	Bayesian lasso uses the Laplace (double exponential, DE) distribution, where the prior assigned to marker effects and all marker effects are assumed to be independently and identically distributed. This prior assigns the same variance or prior uncertainty to all marker effects. This prior possesses thicker tails than the normal prior. Bayesian lasso removes markers from the model, contrary to what happens in variable selection approaches. Bayesian lasso is expected to shrink effects more strongly toward zero than the Gaussian prior, as opposed to inducing sparsity in the strict sense of the Lasso.	Pérez & de Los Campos (2014)
BLRR_BLR	Induces homogeneous shrinkage of all marker effects towards zero and yields a Gaussian distribution of marker effects. Similar to RR-BLUP, there is a problem of QTL linkages to the marker.	de Los Campos *et al*. (2013)
BL_BGLR	Modified version of BLR, has same assumptions as BLBL_BLR.	Pérez & de Los Campos (2014)
BRR_BGLR	Modified version of BLR, has same assumptions as BLRR_BLR.	de Los Campos *et al*. (2013)
NR	Fit mixed models with the advantage of specifying the variance-covariance structure for the random effects and specify heterogeneous variances using more than one variance component and allowing specification of covariance structures. Uses Direct-Inversion Newton-Raphson algorithm.	Covarrubias-Pazaran (2016)

Table 3 *Cont.*

Method	Main features	Ref.
RKHS	Based on genetic distance and a kernel function with a smoothing parameter to regulate the distribution of QTL effects. Effective for detecting nonadditive gene effects.	Gianola & van Kaam (2008)
rrBLUP	Assumes markers have equal variances with small but non-zero effects. Applies homogeneous shrinkage of predictors towards zero but allows for markers to have uneven effects. Computed from a realized-relation matrix based on markers. Some QTL are in LD to marker loci, whereas others are not.	Meuwissen, Hayes, and Goddard (2001)
GBLUP	Assigns common variance to all loci and treats them as equal. Uses a genomic relationship matrix instead of the conventional pedigree-derived numerator relationship matrix.	VanRaden (2008)

Breeding programs use GS in two ways. The first focuses on predicting additive effects in early generations (F_2:F_3) to achieve a rapid selection cycle with a short interval. In this application, it is of interest to predict the additive values rather than the total genetic value (summarizing marker effects using linear additive models is sufficient). The second approach attempts to predict the total genetic values of individuals, considering both nonadditive and additive effects. Thereby, the genotypes' performance (commercial values) is predicted (Crossa *et al.*, 2017).

With the rising accessibility of genomic data and the development of analytic methods, acquiring plant phenotypes has become a bottleneck. Although high-throughput genotyping is expanding exponentially, the acquisition and processing of phenotypes constrain the ability to use this data for crop yield improvement (Ruiz-Guzman *et al.*, 2016; Li *et al.*, 2021).

Increasing capacity – high-throughput phenotyping

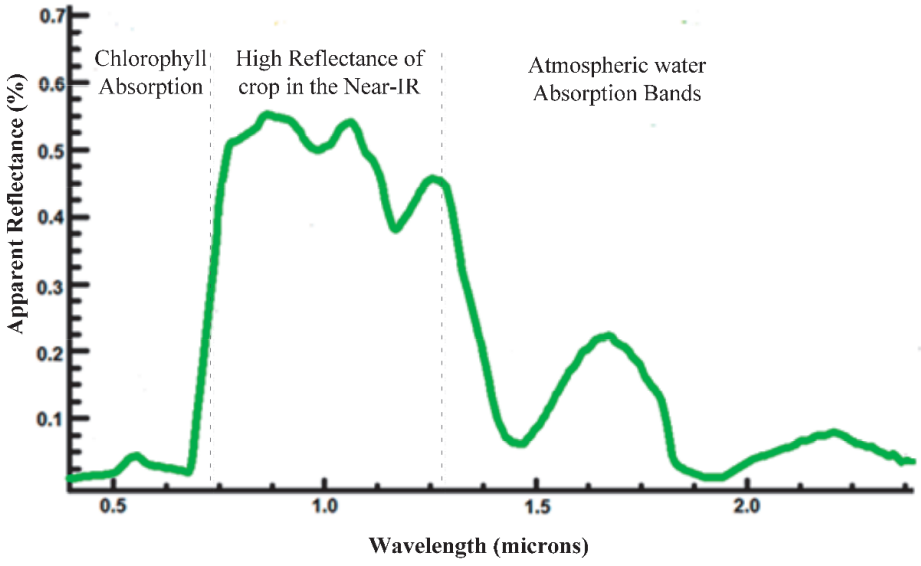
Phenotyping is the foundation of any breeding program and is traditionally done manually. Modern plant phenotyping often measures complex traits on different scales of organization from organs to whole canopies, increasing the workload, need for resources, and cost (Fiorani & Schurr, 2013). A more comprehensive and recent definition of plant phenotyping is the assessment of complex plant traits such as growth, development, tolerance, resistance, architecture, physiology, ecology, yield, and the essential measurement of individual quantitative parameters that form the basis for complex trait assessment (Li, Zhang and Huang, 2014; Costa *et al.*, 2019). The expanding plant phenotyping recently became an independent research field.

Current acquisition of plant characteristics relies primarily on expert visual scoring, which is expensive and generates bias. The latest developments of both hardware (for example, cameras) and software open new possibilities of both automating, improving, and expanding the capture of phenotypes for the needs of both research and plant breeding (Li, Zhang and Huang, 2014). In the recent years, many platforms using robotics and imaging technologies were deployed to automatically assess plant growth and performance, both under controlled and field conditions (Araus & Cairns, 2014). The new approach to phenotyping using technology and analytics forms the domain of high-throughput phenotyping (HTP).

Plant canopy reflectance

Electromagnetic waves/photons are carriers of information for any imaging technique. Their interaction with the plant canopy depends on its optical properties, which correspond to many aspects of physiology (Figure 11). Therefore imaging shows great potential for plant phenotyping (Li, Zhang, and Huang, 2014).

Figure 11 *Typical plant canopy reflectance spectrum with regions important from plant phenotyping's perspective. Taken from* Ollinger *(2011)*



Interactions between plant canopies and incident radiation are incredibly complex due to the diversity in size, shape, composition, and arrangement of cells and plant organs (Ollinger, 2011). However, the biophysical basis for nearly all of those interactions can be assigned to two categories: absorbance and scattering, with

scattering being divided further into transmittance and reflectance. Absorbance includes waves absorbed by pigments, liquid water, and other plant components (Blackburn, 1998; Kokaly *et al.*, 2009). By contrast, scattering occurs when radiation crosses a boundary between two media with different refractive indices. A well-known example is the illusion of a "bent stick" when dipped in water due to different refraction of water (1.33) and air (1.0) (Ollinger, 2011).

Machine learning and deep learning for plant phenotyping

Images of whole plants and organs can be used either for semantic analysis or for quantitative assessment of reflectance parameters (Burud *et al.*, 2017; David *et al.*, 2020). Deep learning (DL) is a sub-domain of machine learning (ML) and, more generally, artificial intelligence (AI). An excellent introduction to DL for plant phenotyping can be found in Arya *et al.* (2022).

With recent advances in GPU (graphical processing unit) performance and increasing availability of large-scale datasets (Russakovsky *et al.*, 2014; Alom *et al.*, 2019), deep learning (DL) has become the state-of-the-art method for numerous computer vision tasks, including object detection (Ren *et al.*, 2017), instance segmentation (He *et al.*, 2020), semantic segmentation (Ronneberger, Fischer and Brox, 2015), and image regression (Aich & Stavness, 2018; Xiong *et al.*, 2019). Several HTP-tailored DL models were also developed (David *et al.*, 2020). A synthetic overview of DL applications in HTP is shown in Figure 12.

A specific niche of HTP is multispectral phenotyping, which uses multispectral cameras equipped with "invisible" wavelengths like near-infrared on top of RGB. Images captured using multispectral cameras allow for deriving vegetation indices (VIs). VIs are constructed from reflectance values for two or more wavelengths to analyze the specific characteristic of vegetation. VIs as simple and effective proxies of surface vegetation conditions are widely used in vegetation monitoring via remote sensing (Araus & Cairns, 2014).

VIs can be calculated from satellite images or unmanned aerial vehicle (UAV) imagery. UAV remote sensing has many advantages: high image spatial resolution, instant information acquisition, convenient operation, high maneuverability, freedom from cloud interference, and low cost (Iizuka *et al.*, 2018).

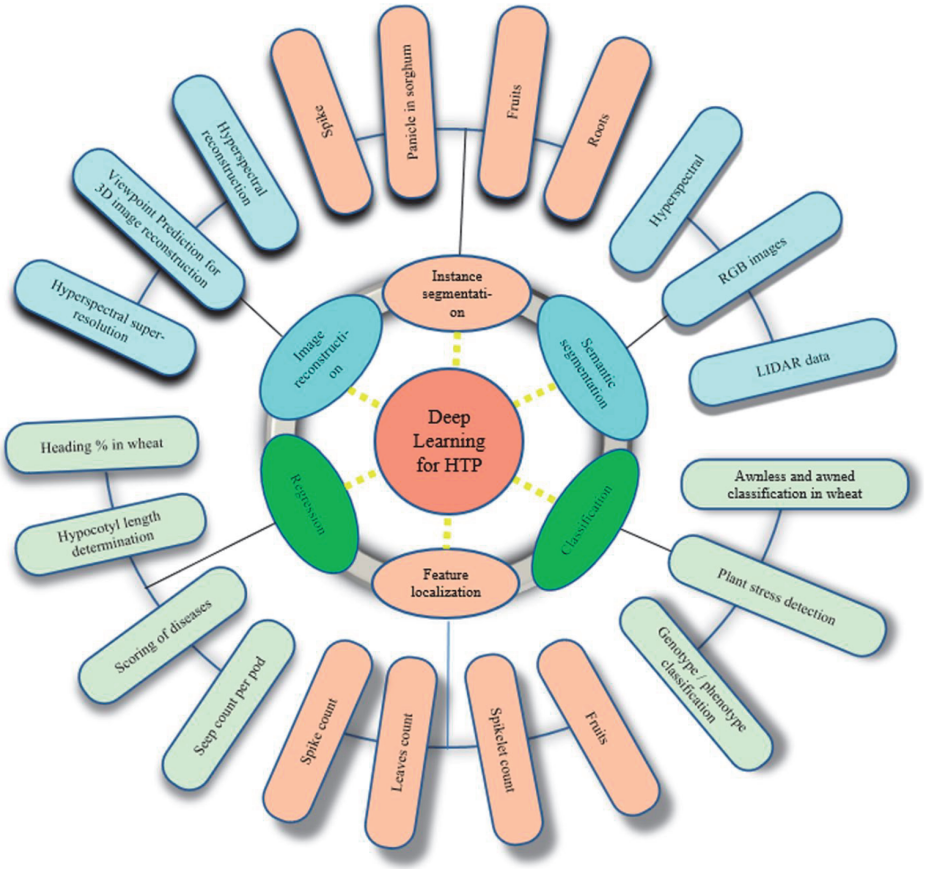


Figure 12 *An overview of DL applications in high-throughput plant phenotyping. Taken from Arya et al. (2022)*

In the recent years, field phenotyping has seen considerable advances due to improvements in camera technology, the popularization of cost-effective UAV platforms used as vectors for the cameras, and developments in AI methodologies. Multispectral UAV imagery (Figure 13) has been successfully used to *predict* or *estimate* numerous essential crop traits, such as grain yield (Duan *et al.*, 2017; Zhou *et al.*, 2017; Maimaitijiang *et al.*, 2020; Suab & Avtar, 2020; Shafiee *et al.*, 2021), above-ground biomass (Han *et al.*, 2019; Lu *et al.*, 2019; Li *et al.*, 2020), plant height (Hu *et al.*, 2018; Hassan *et al.*, 2019; Tirado, Hirsch and Springer, 2020), the maturity date (Zhou *et al.*, 2019; Trevisan *et al.*, 2020) and crop emergence (Li *et al.*, 2019), to name but a few.

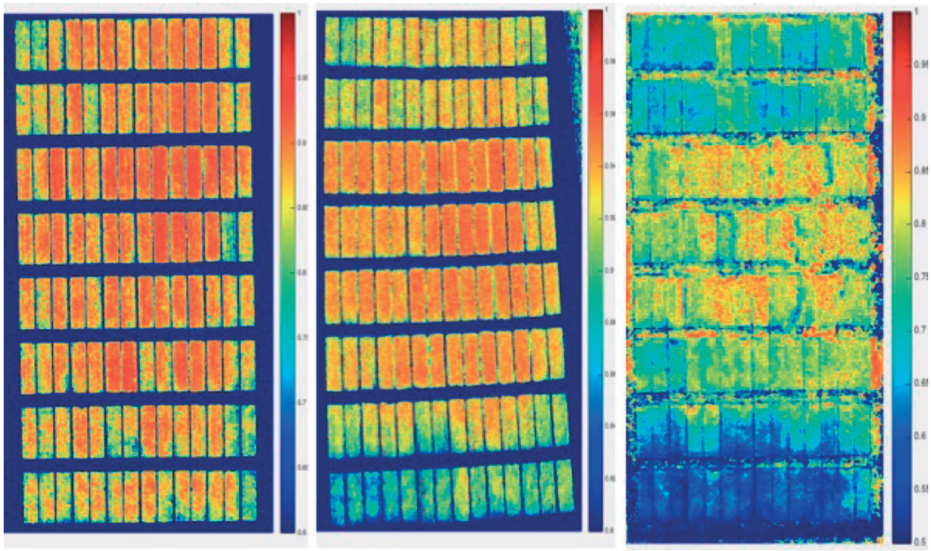


Figure 13 *An example of false-color visualization of NDVI (normalized vegetation index) map of a field trial at NMBU. Taken from Burud et al. (2017)*

Linking the "omics" – potential of GS supplemented with HTP data

While HTP data has shown great potential in standalone grain yield prediction using ML, the models are often environment or population-specific (Crossa *et al.*, 2017). Using HTP platforms with VIs as predictor traits in pedigree and GS models increases the prediction accuracy for grain yield (Rutkoski *et al.*, 2016). The authors found that within-environment secondary VIs increased prediction accuracies for grain yield by over 50% using pedigree relationships and 70% using genomic relationships (GS), indicating that traits measured by HTP paired with GS can improve prediction accuracy during the early stages of breeding. HTP techniques also effectively evaluate genetic resources for complex trait expression (Reynolds & Langridge, 2016). Predicting breeding values of genotypes concerning grain yield can also be improved by 70% on average by including VIs as covariates in a model with multivariate pedigree and genomic models (Sun *et al.*, 2017).

A recent study used statistical models to assess wavelength-by-environment interactions in HTP by incorporating genomic and pedigree GxE interactions (Montesinos-López *et al.*, 2017). Although the authors observed little gains in GS accuracy, critical hyperspectral wavelength by environment interactions were detected, demonstrating that GS coupled with HTP is a powerful tool for application to early-generation testing of many selection candidates.

2 The thesis

2.1 Background and motivation of the thesis

Norway initially was not able to fully benefit from the Green Revolution. Wheat cultivars and varieties in the 1950s and 1960s were tall and susceptible to lodging. It was a common practice at that time to windrow the yields, manually harvest, and even dry the yields indoors due to unfavorable weather conditions towards the end of the growing season. The old cultivars could not withstand increased N-fertilization without lodging and were not suitable for mechanized harvest. This lack of fit between varieties and mechanization nearly eradicated domestic wheat production in the 1960s due to favorable imported wheat prices from countries that could benefit from the technological advancements (Lillemo & Dieseth, 2011). This dependency on imported small grains spurred intense breeding efforts to produce varieties capable of benefiting from the new cultivation methods. These efforts yielded two "landmark" varieties: Runar and Reno (released as varieties in 1972 and 1975) and revitalized Norwegian wheat cropping. Both Runar and Reno could be machine-harvested and could take advantage of increased N fertilizer input without lodging. Their release marked the modern era of wheat breeding in Norway. In the following years, Norwegian wheat breeding became more international by, for instance, cooperating with CIMMYT and using their germplasm for crossing. Such efforts led to the development of important varieties such as Bastian, released in 1989. The scale and extent of nowadays' wheat breeding in Norway are more significant than three decades ago, continuously working to produce varieties of high value to satisfy domestic demand for grain.

Progress in spring wheat grain yield in Norway has been documented by Erling Strand in three of his works, concerning the three periods: 1889-1962, 1960-1974, and 1960-1992 (Strand, 1964, 1975, 1994). These studies gave valuable insight into the contributions of new genotypes and growing techniques to the grain yield progress in Norway and captured the echoes of the Green Revolution. However, little is known about later gains in grain yield until today.

Selection of breeding candidates is traditionally based on phenotypes (phenotypic selection) and "the Breeder's eye". Recently, with the advancements in DNA-based techniques, it became possible to pin down essential genes for traits of interest. These genes are often discovered in retrospective analyses. Breeding progress in grain yield in Norwegian spring wheat, bearing similarity to many collections worldwide,

remains largely undissected, with little knowledge of the underlying genes under historical selection pressure.

The ”Next Green Revolution” caused by the wide adoption of genomic and phenomic techniques in modern wheat breeding also happens in Norway. For instance, genomic selection has become nearly routine in many breeding programs due to lower genotyping costs and the extent of available phenotypic data. Efforts are also made to utilize more novel technologies in the breeding programs, such as high-throughput phenotyping; however, the use is mainly in the experimental phase. Both GS and HTP have the potential as standalone tools to accelerate genetic gains in grain yield but have their shortcomings. For instance, grain yield prediction using machine learning and UAV (Unmanned Aerial Vehicle) HTP multispectral data tends to be cost-efficient and decently accurate but does not work in cross-environment scenarios. GS protocols open new avenues for speed breeding and efficient selection using marker data for traits that are expensive or difficult to score. However, large datasets and more complicated models are needed to reach satisfactory accuracy in cross-environment scenarios. Integration of “omics” - for instance, adding HTP data layer in GS models - usually leads to improved accuracy and error reduction. So far, no such attempts were done for Norwegian spring wheat.

Therefore, based on the motivation described above, the main objectives of this thesis were to:

I. Determine, describe, and document grain yield and other agronomic and quality traits’ progress in Norwegian spring wheat since 1972 concerning changing management practices.
II. Perform genetic dissection of grain yield and associated traits in Norwegian spring wheat, considering the time dimension – breeding progress in the traits over time.
III. Assess the prospect of future improvement in grain yield in Norwegian spring wheat by exploring the use of UAV multispectral data for standalone grain yield prediction and its synergy with genomic prediction protocols.

2.2 Materials and Methods

2.2.1 Paper I

Breeding progress can be "measured" in several methods, as described in the Introduction. To study breeding progress in grain yield and associated traits in Norwegian spring wheat, we gathered a collection of representative varieties for the period (listed in Table 4) and tested them in multi-year field trials. We assembled a relatively small panel of initially 24 lines, including both historical varieties (such as the landmark varieties Runar and Reno), more recent varieties, and some of the most recent variety candidates at the time of the start of the trials in 2016. Varieties were placed on the timeline using their official year of release. Throughout the trials, three of the most recent variety candidates were rejected from the official field trials and their records were removed from the analysis. The historical varieties were chosen by their relevance to the Norwegian spring wheat market (ones with the most significant market share at a time). In contrast, the newest lines were chosen to sample the current diversity in variety candidates. The collection includes mostly varieties bred by either Norwegian or Swedish companies/institutes, with one exception (Arabella).

Table 4 *Varieties used to investigate the breeding progress in Norwegian spring wheat. Accessions present in the experiment, but not released as varieties are not included in the list. YOR – year of release*

Line	Cultivar	Country/breeder	YOR
1	Runar	Norway/IPK	1972
2	Reno	Norway/IPK	1975
3	Tjalve	Sweden/Weibull	1987
4	Bastian	Norway/IPK	1989
5	Polkka	Sweden/Lantmännen SW Seed	1992
6	Avle	Sweden/Lantmännen SW Seed	1996
7	Zebra	Sweden/Lantmännen SW Seed	2001
8	Bjarne	Norway/Graminor	2002
9	Demonstrant	Norway/Graminor	2008
10	Krabat	Norway/Graminor	2010
11	Mirakel	Norway/Graminor	2012
12	Rabagast	Norway/Graminor	2013
13	Seniorita	Norway/Graminor	2014
14	Arabella	Poland/Danko	2014
15	Willy	Norway/Graminor	2016
16	Caress	Sweden/Lantmännen SW Seed	2017
17	Zombi	Norway/Graminor	2018
18	Alarm	Norway/Graminor	2019
19	Betong	Norway/Graminor	2019
20	Eleven	Sweden/Lantmännen SW Seed	2019
21	Felgen	Sweden/Lantmännen SW Seed	2019

The collection "starts" with the landmark varieties – Runar and Reno – with the aim to anchor the progress to the current breeding era. Pre-Green Revolution varieties are dramatically different and would cause practical problems with lodging under the used fertilization rates.

The changing agronomic practice over the decades was also of interest, and therefore the collection was tested under two fertilization rates: 75 and 150 kg N ha^{-1}. The latter represents current fertilization practice, while the former is closer to the rates used in Norway in the early 1970s. The experiment aimed to observe if genetic progress in grain yield and associated traits depended on high agronomic input.

2.2.2 Paper II

Traits which exhibited progress over the period in Paper I were followed up in a genetic study. For this purpose, a panel of 301 hexaploid spring wheat varieties and cultivars was used, referred to as the NMBU spring wheat panel or MASBASIS. This panel also includes the varieties from the historical collection. The NMBU spring wheat panel consisted of 186 Norwegian, 40 Swedish, and 37 lines originating from CIMMYT, Mexico, and presented considerable genetic diversity in most traits. Multi-environment field data (7 years, 2 locations) was available for this panel, making a solid foundation for genetic studies. A peculiar aspect of the panel is strong population structure, which results from different backgrounds of lines (Nannuru *et al*., 2022). Accessions originating from the Nordic countries are considered well-adapted to the Norwegian growing environment, while lines from other climate zones (such as Mexico and China) show poor adaptation (and, therefore, more variable phenotypes). This population structure allowed for investigating the adaptation to the Norwegian growing conditions by comparing results obtained for the adapted and non-adapted lines.

By analyzing the documentation of lines in the NMBU spring wheat panel, it was possible to place a large part of the adapted lines on the timeline, similar to the historical panel. However, as only a part of the lines were released varieties, a different approach was taken: varieties were assigned to a discrete period based on their year of creation. Year of creation was estimated by analyzing the documentation of the lines.

The NMBU spring wheat panel was genotyped using Illumina 25K chip, yielding 19874 high quality polymorphic markers (minor allele frequency > 0.05 and less than 10% missing scores). This data was used in a Genome-Wide Association Study (GWAS) to discover regions associated with grain yield, days to maturity, days to

heading, and plant height. The population structure was addressed by introducing population structure correction in the models and by analyzing the adapted part separately from the entire panel. With the added time dimension of the part of the NMBU spring wheat panel, it became possible to confront findings from the historical collection and look at the discovered significant loci from a time perspective. The population structure additionally allowed for investigating the loci associated with adaptation to the Norwegian growing conditions.

A wide selection of GWAS methods became available during the last two decades (as reviewed in the Introduction). For the genetic study, a hybrid approach was taken. First, GWAS was conducted using the FarmCPU method (Liu *et al.*, 2016) with a strict significance criterion based on Bonferroni threshold and repeatability to see the most important and stable associations. Regions surrounding the associations were investigated further using the MLM method (Zhang *et al.*, 2010) to build more robust haplotypes. This approach was motivated by the fact that the FarmCPU method tends to "distillate" the associations, providing fewer hits but with lower p-values. On the other hand, MLM tends to provide more hits but with higher p-values. By linking the two methods, it was possible to increase the confidence of discovered regions.

Allele frequency of the discovered polymorphisms was analyzed over time. Collections of the most recent spring wheat breeding lines from Graminor AS breeding program were used for validation and estimating the usefulness of loci in future breeding efforts.

2.2.3 Paper III

High-throughput field phenotyping solutions range from budget devices (such as consumer-grade drones with cameras) to complicated, extensive, and expensive field installations (for instance, gantry systems with many sensors). The latter's cost limits the application of such solutions in commercial breeding programs. We explored the "low-cost phenotyping" technology by deploying commercially-available UAVs equipped with multispectral cameras for regular data capture during the growing seasons. Such data was already successfully used in our group for grain yield prediction using machine learning (Shafiee *et al.*, 2021). Genomic selection (GS) is already a routine in many breeding programs, and increasing its accuracy is one of the main goals. Some studies attempted to include HTP multispectral data as covariates (Crossa *et al.*, 2017), but few attempts were made to use "multispectral similarities" instead of fixed covariates. Therefore, we deployed UAV multispectral phenotyping of the NMBU spring wheat panel. Taking advantage of the already-available genotyping and field data, we built a series of models based on GBLUP

(VanRaden, 2008) using different combinations of genomic, multispectral, and environment relationship matrices. We compared them back to back for their GY prediction ability in both single and multi-environment scenarios.

2.3 Main results

In **Paper I**, historical genetic gains in Norwegian spring wheat grain yield (GY) and associated traits were investigated under two fertilization regimes. Using multi-season field trial data of the collection of 21 (initially 24) historical and current spring wheat varieties planted under 75 and 150 kg N ha^{-1} fertilization regimes, we discovered highly significant progress in GY between 1972 and 2019 of 17.8 kg ha^{-1} (0.34%) per year. The gains were clearly visible under both fertilization regimes (21.2 and 14.3 kg ha^{-1} per year under 150 and 75 kg N ha^{-1}, respectively), showing minimal genotype by environment interactions. New varieties consistently had higher GY than the older ones. New varieties under the 75 kg N ha^{-1} treatment yielded almost similar to old varieties under 150 kg N, showing that improved genetics leads to more efficient resource utilization (Figure 14).

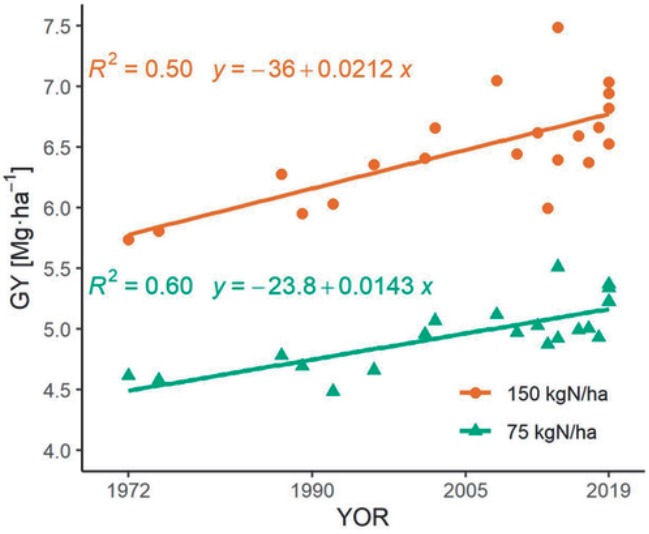


Figure 14 *Grain yield genetic gains over the 1972-2019 period under 75 (green, triangles) and 150 kg N ha^{-1} fertilization (orange, dots). GY – grain yield, YOR – year of release of a variety.*

The GY increase over the years coincides with an increase in the number of grains per spike and area, hinting that increments in these traits contributed to the progress. No signs of increments in kernel weight were found. Higher GY usually means lower grain protein content, and this was also the case in Norwegian spring wheat.

However, due to the substantial increase in GY, new varieties still had superior protein yield compared to old varieties.

New varieties have extended vegetative and grain-filling periods, reaching physiological maturity approximately three days later than old varieties.

In **Paper II**, grain yield, plant height, earliness, and heading time were dissected in GWAS. NMBU spring wheat panel of 301 lines genotyped with Illumina 25K SNP chip was used to discover genetic structures associated with genotype adaptation. Furthermore, placing lines on the timeline allowed to uncover loci associated with genetic progress in GY and the other agronomic traits.

The study identified twelve QTL associated with GY (three), days to maturity (two), days to heading (two), and plant height (five). Two loci significant for heading time (unknown locus) and grain yield (the *Vrn-A1* locus) were strongly associated with line adaptation, hinting that it mostly consists of a phenological response. The *Vrn-A1* locus on chromosome 5A explained a similar proportion of variance as line adaptation status.

Table 5 *QTL regions discovered in the study for days to heading (DH), days to maturity (DM), grain yield (GY), and plant height (PH), their genomic locations, QTL effects in the full dataset (E_F) and in the adapted part (E_A), trait variance explained in the full dataset ($\%PVE_F$) and in the adapted part ($\%PVE_A$), candidate well-known loci (Locus)*

Trait	QTL/Chr	Span Mbp	E_F	$\%PVE_F$	E_A	$\%PVE_A$	Locus
DH	*QHd.nmbu-1B*	1-2	2.04	14.1	Mono	Mono	
	QHd.nmbu-7B	606	0.91	11.4	0.91	11.4	
DM	*QMat.nmbu-6B*	132-136	5.93	25.8	4.85	24.8	
	QMat.nmbu-6D	6	0.82	8.80	0.81	10.0	
GY	*QYld.nmbu-3A*	267	38.1	22.0	16.4	14.3	*Ta-GS5-3A*
	QYld.nmbu-5A	683-708	192.6	67.2	Ns	Ns	*Vrn-A1*
	QYld.nmbu-7B	701-703	106.2	16.3	83.6	30.9	*Ta-Col5*
PH	*QHt.nmbu-2A*	524-543	7.91	9.21	7.97	12.6	
	QHt.nmbu-4A	570-603	9.77	9.36	5.45	9.08	
	QHt.nmbu-4B	13-59	8.20	27.7	7.22	23.9	*Rht-B1*
	QHt.nmbu-4D	19-26	8.29	32.5	9.32	42.9	*Rht-D1*
	QHt.nmbu-6B	202	2.00	5.21	1.25	2.14	

The study discovered two loci associated with GY on chromosomes 3A (267Mbp) and 7B (around 700Mbp), showing a decline over the years in the frequency of the unfavorable allele , possibly responsible for the breeding progress described in Paper I. The loci correspond to *Ta-GS5-3A* and *Ta-Col5* genes, respectively, associated with spike architecture (number of grains per spike) and kernel size. The locus on chromosome 7B substantially affects GY with an effect of nearly 1 t ha^{-1}. Its unfavorable allele has been eliminated over the years, showing no prospect of future

use for increasing GY. Although the locus on 3A has a minor effect of around 300 kg ha^{-1}, it still shows large diversity in the panel, showing room for future improvement.

The sudden drop in plant height after 1970 could not be explained by the *Rht* genes, which still segregate in today's germplasm. A locus on chromosome 2A explained the sudden drop. However, this association could be spurious since no QTL was previously identified in this locus. No genetic explanation could be found for the increments in the length of the vegetative period over the years, hinting that it must have occurred due to the accumulation of favorable alleles at numerous small-effect loci which could not be discovered using the methodology. Almost all the discovered QTL were validated using independent sets of breeding lines, providing wheat breeders in Norway and the Nordics a valuable base for developing marker-assisted selection protocols to improve GY and other traits.

Paper III elaborated on using UAV multispectral-derived genotypic relationships for GY prediction. Using multi-environment GY data from the NMBU spring wheat panel of 301 genotyped lines (used in Paper II) paired with UAV multispectral data gathered using budget cameras, a series of GS models based on the GBLUP model (VanRaden, 2008) were developed and tested in single and multi-environment scenarios. UAV multispectral phenotypes were used to derive multispectral relationships (M matrices) instead of using them as simple covariates in the model. The study demonstrated that in single-environment scenario GY prediction using the M matrix yields comparable accuracies to GS (using the G matrix) and that models using both G and M matrices are superior to models with individual matrices (Table 6).

M matrix, assumed to be environment-specific, shows high prediction ability in multi-environment scenarios. This ability increases with the number of data-gathering sessions. Practical aspects of GY prediction using the M matrix were also investigated, showing that the optimal time for data capture occurs during grain filling. GY prediction is also possible using a conventional RGB camera with a slight loss in accuracy. Paper III provided a simple yet effective framework for using budget UAV data and existing software to improve the accuracies of GS protocols and to lower their error. It also attempted to understand the mechanisms governing the predictive ability of the M matrix and overcomes the environment-specificity problem of machine learning GY prediction protocols.

Table 6 *Comparison between the predictive abilities of models with covariates: G (genomic relationship matrix) and M (multispectral relationship matrix). Comparison performed in a single-environment scenario using 200 iterations and 80/20 dataset partitioning to training and testing datasets, respectively. Metrics: rTRN – accuracy in training test, rTST – accuracy in testing set, rmseTRN – root mean squared error in training set, rmseTST – root mean squared error in testing set*

Metric	Model with covariates:		
	G	M	G+M
rTRN	0.98	0.79	0.99
rTST	0.68	0.71	0.79
rmseTRN	15.61	41.30	10.96
rmseTST	48.02	47.80	40.52

2.4 Additional work not included in the manuscripts

2.4.1 Use of soil parameter sensors in field trials.

Soil is a vital component of plant growing environment and directly influences GY and reflectance parameters, to name but a few. The ability to quantify dynamic soil properties such as moisture and temperature in real-time can be essential for linking canopy multispectral reflectance data with the environment and genotype that caused it. In 2020, we started using wireless soil sensors across field trials based on this rationale. Data gathered using those sensors allows for classifying environments (seasons) based on their properties. Dense soil and climate data is now routinely gathered in our field trials; however, more seasons are required before this data can be used.

2.4.2 Genetic study of spike parameters in spring wheat.

As a follow-up to the findings from Papers I and II, we gathered spike samples from the NMBU spring wheat panel and the validation panels of current breeding lines. For each line replicate, ten representative spikes were gathered and analyzed for spike morphological parameters (length, width, number of spikelets, number of fertile spikelets). Based on data obtained from those samples and previously available data described in the thesis, a genetic dissection of spike parameters is currently being conducted to verify the associations of the discovered loci with GY.

2.4.3 Historical progress in GY under controlled conditions.

The experiment aimed to determine if the historic yield increase in Norwegian spring wheat can be explained at physiological and metabolomic levels through combining novel HTP (high-throughput phenotyping) and semi-HTP biochemical methodologies

offered at PhenoLab, University of Copenhagen (Photo 1). The main goal was to study the differences in crop canopy development, photochemical performance and multispectral patterns supplemented with metabolomic / phytohormone signatures over the vegetation period in a set of spring wheat cultivars representing 5 decades of breeding progress in Norway (described in Paper I). Data obtained during the experiment is currently being analyzed.

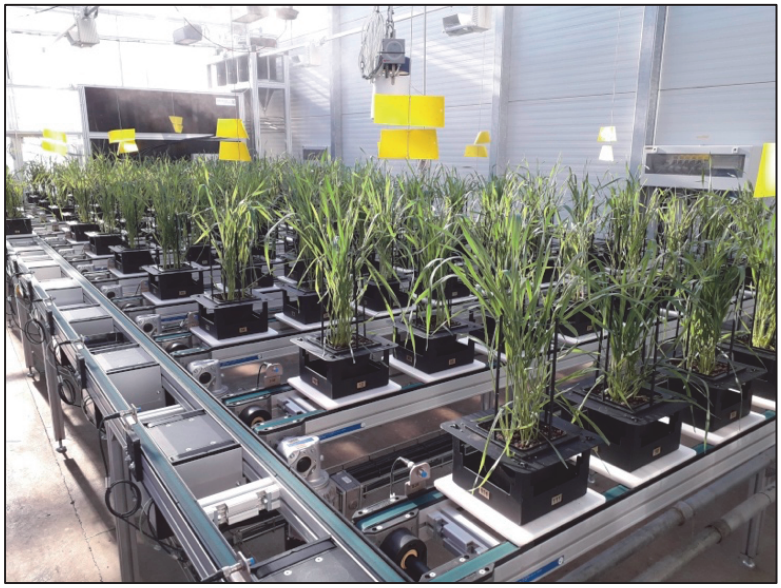


Photo 1 *Experiment with the historical and current variety collection under controlled conditions in PhenoLab, Copenhagen. Image taken at anthesis. Photo credit: Prof. Thomas Georg Roitsch*

2.5 Discussion

2.5.1 Did wheat breeding in Norway increase grain yield?

Significant genetic gains over time were repeatedly observed not only across numerous trials during representative seasons presented in Paper I, but also in 2018, when the whole region experienced severe drought. Similar numbers and conclusions about the genetic gains can also be drawn by extracting the data for the 21 historical and current varieties from the NMBU spring wheat panel, fertilized with 120 kg N ha^{-1}, as described in the supplementary of Paper I. In Paper II, a larger sample of varieties and breeding lines also showed GY progress over the periods. Significant gains in grain yield can be repeatedly observed across many environments and genotype sets, which is compelling.

Are not genetic gains in GY an inherent consequence of the legal requirements of variety acceptance by authorities and consumer demand? For a breeding line to become an official variety, the line has to be superior in at least one quality to the currently registered varieties. In Norway, those qualities include earliness, climate and disease resistance, winter survival (for winter crops), grain yield, and yield quality (Mattilsynet, 2022). Variety candidates are then tested across several environments (locations and seasons). Based on the results of these trials, a candidate either becomes a variety or is rejected if it fails to outperform already available material.

Varieties are commercial products of breeding companies, and their products have to be in demand for a breeding company to make profit. The demand is created by farmers, which are also profit-driven – buying seeds is an investment that has to bring return. Consumers, in turn, regulate this return through industry and their intermediates. Finally, the price fetched by farmers for their grain is a product of quantity and usefulness for the industry - determined by grain quality - on top of other investments like fuel, fertilizer, and fungicides. This perhaps oversimplified (legal incentives or inflation are not included) chain leads to an obvious conclusion that variety development and registration makes sense only if it is superior in yield quantity and quality. All requisites for a line to become a variety eventually boil down to the two factors under the challenging climate in Norway: high yield potential and the ability to resist yield and quality losses due to unfavorable weather conditions and plant diseases. Earliness is important in Norway due to erratic weather patterns during the late growing season and lowers the risk of quality and quantity loss due to lodging or pre-harvest sprouting.

Therefore, stating that Norwegian wheat breeding concentrates solely on increasing GY *per se* would not be complete. However, assuming that a variety candidate meets grain quality for a target class, trait superiority needed for registration also contributes to higher GY, directly or indirectly, which explains the progress in GY. Therefore, it is reasonable to state that genetic gains in GY are a logical consequence of economy and legislation.

Despite its genetic complexity as a highly quantitative trait, GY can be seen as an accessible target for selection. With many loci with often additive effects, breeding for GY becomes a true numbers game, the outcomes of which can be observed in linear GY increase in many post-Green Revolution collections worldwide. Estimated annual genetic gains in GY in Norwegian spring wheat (approximately 18 kg ha^{-1} per year) show similarity to many collections from many regions; however, they appear smaller than the median (as reviewed in the Introduction). It can be noticed that gains in collections originating from developing regions are usually higher than developed ones. Developing areas work primarily towards meeting their rising demand for grain by prioritizing GY. At the same time, developed regions focus more on end-user quality than on the magnitude of the yields. With its lion's share of bread wheat being used by industrialized bakeries today, Norway fits well in this picture.

2.5.2 Is this yield progress dependent on intensive management?

In the case studied in Paper I, the environments consisted of two fertilization regimes (75 and 150 kg N ha^{-1}, referred to as management – M). The experiment showed that N fertilization had a visible effect on GY, but rankings of accessions within each N treatment were similar, indicating only minor GxE (M) interactions Therefore, genetic gains in GY were not contingent on N input and were stable within these boundaries, although numerically slightly higher under 150 kg N ha^{-1}. The message from Paper I confirms findings from other recent works (as reviewed in the Introduction): breeding causes new varieties to be consistently better than older ones at utilizing available resources, and the degree of interactions among varieties and environments is minor. However, these conclusions are valid only if the studied environments remain within "reasonable" boundaries.

The concept of studying genetic progress in GY is extended in the work of Ahrends *et al.* (2018). The authors tested a collection of sixteen wheat varieties released in Germany between 1895 and 2007 under twenty-four long-term fertilization regimes, ranging from no fertilization to full intensive practice. The authors concluded that genetic gains in GY are observable only under particular management treatments, and overall, newer varieties tend to respond stronger to more intensive management.

These findings also align with Paper I, where the estimated genetic gains in GY were slightly higher under intensive management (21 *vs* 14 kg ha^{-1} per year). However, due to the small sample size used in this thesis, it is hard to go beyond speculation about responsiveness. Nevertheless, the slight difference between gains estimated under the managements does not invalidate the conclusions about the superiority of newer varieties, as their rankings remain similar.

Remarkably, the most recent varieties grown under 75 kg N ha^{-1} reached GY comparable to old varieties grown under 150 kg N ha^{-1}, clearly underlining that new varieties can utilize available resources more efficiently. This conclusion is further supported by the fact that despite the new varieties having significantly lower grain protein content, they still reach significantly higher protein yield.

2.5.3 What is the genetic cause of the GY increase?

Even though GY is controlled by many small-effect loci, there are known large-effect genes that either increase GY *per se* or allow increased GY indirectly. Stringent significance criteria used in Paper I were designed to detect rather large-effect stable loci and succeeded in identifying three QTL regions. Two loci (*QYld.nmbu-3A* and *QYld.nmbu-7B*) showed visible changes in allele frequency over time, with observable decay of negative effect allele frequencies. Both those regions coincide with chromosomal locations of two known genes associated with GY: *Ta-GS5-3A* and *Ta-Col5*, respectively, and have shown effects in the sets of independent lines.

QYld.nmbu-7B appears to have been under intense selection pressure: all studied accessions dated before 1970 carried an unfavorable allele, which was later gradually replaced by either of the favorable effect alleles so that no studied line dated after 2006 carries the allele associated with low yield. The estimated effect of this QTL (0.9 t ha^{-1}) explains most of the differences in GY between the oldest and more recent lines in the study. Interestingly, the gene behind *QYld.nmbu-7B, Ta-Col5,* affects spike parameters, including grains per spike (Dixon *et al.*, 2018). This aligns with the findings of Paper I - grains per spike and area were associated with the GY increase. However, it is unlikely that this QTL contributed to the GY gains described in Paper I due to its significant effect, no alleles with intermediate effect sizes, and near monomorphism in the collection from Paper I. Even though *QYld.nmbu-7B* does not explain the GY progress observed in Paper I, it sheds light on the broader historical changes in GY in Norwegian wheat germplasm and sources of beneficial alleles. The usefulness of this locus in modern breeding is limited, as the newest lines show little polymorphism concerning the favorable allele.

In Paper I, no trend over time was discovered for kernel weight, and GY was barely associated with kernel weight. However, in Paper II, the *QYld.nmbu-3A* locus, representing the *Ta-GS5-3A* gene associated with kernel size (Wang *et al.*, 2015), appears as the best candidate to explain the breeding progress in GY. With a moderate effect of approximately 0.3 t ha^{-1} this locus shows a shift in allele frequency with the favorable allele occurring more often in the more recent lines. This locus can also be used in future breeding, as it retains high polymorphism in current breeding material.

It was perhaps inappropriate to look for an explanation of a linear trait improvement using such stringent criteria for GWAS, suited to detect large-effect loci. However, using more relaxed criteria would, without a doubt, lead to the discovery of many QTL candidates, with many being spurious. Such an approach could explain the progress, but would it have any practical application? Voss-Fels *et al.* (2019) performed such an experiment, demonstrating that GY increase is linked with decay in the number of "detrimental haplotypes" (haplotypes with no or negative effect on GY) in a particular variety. However, the "detrimental haplotypes" remained a statistical concept, and it remains a question of the identical "detrimental haplotypes" would behave the same way when analyzed in another population.

Reliable detection of smaller-effect loci in GWAS requires large population sizes, and the smaller the assumed effect, the larger the population should be. Population size is usually constrained by practical causes, and the population used in this study does not allow for reliable detection of small-effect loci or at least to separate them from false positives.

Even though the genetic progress in GY described in Paper I remains unexplained - the discovered loci are not plausible candidates due to effect size or contradictory trait association - the results from Paper II are valuable. Almost all the discovered loci were validated using the current breeding lines and showed enough polymorphism in current breeding material to be developed further for marker-assisted selection. The two genes that showed effects on GY and frequency change over time are also candidates for a follow-up study on kernel and spike parameters in the population.

2.5.4 Did breeders breed for a changing climate?

According to the literature, growing season length in South-Eastern Norway was extended by approximately seven days since 1970 (Nordli *et al.*, 2008). In Paper I, the newer varieties tend to mature later (approximately three days) with a resulting extension of the grain filling period by two days on average. This increase in vegetative period length correlates well with the increase in GY and is causative: an

extended vegetative season means more time to build necessary source biomass and fill grains (in the case of this thesis, to fill more grains). However, was this a deliberate choice to pick up the gradual environmental changes and use them to the breeder's advantage?

Breeding aims to develop varieties that will thrive in their respective TPEs (target population of environments), and therefore breeding programs are conducted in those TPEs. Selections are traditionally made based on the performance of candidates in the TPE(s), or, with more modern approaches like MAS or GS, based on models built on data from the TPEs. Therefore, it is reasonable to assume that as long as changes in the TPE are stable and gradual, the selection process will ensure the best possible adaptation to the new conditions – including using new TPE features to the varieties' advantage. This selection amidst a gradually changing environment does not require a deliberate choice of the breeders to work.

Is then the cereal production in light of Climate Change safe? Is it then enough to "just" keep breeding new varieties, and the process will ensure adaptation?

The answer is no, at least partially. The merit behind this answer lies in the Climate Change itself, which is not only linear. The predictions for the future include gradual changes and more erratic weather patterns, and the latter threatens grain production the most (Reynolds & Braun, 2022). Small changes in climate can and will be picked up by breeding, just as it took place in Norway, but the challenges do not end there. Many traits can be reliably selected for only during a favorable season. For instance, a season promoting pre-harvest sprouting occurs approximately once every five years in Norway, hampering selection for the trait. Similarly, conditions allowing for selection for overwintering occur once every couple of years. With the expected year-to-year variability increase connected to the Climate Change, having reliable testing conditions for making selections can be challenged. With the projected increased demand caused by population growth, arable land reduction, and productivity loss in the existing systems, the need for better varieties is high.

Nowadays, in a conventional breeding program, it takes around 12-15 years to develop a new variety and bring it to the market. Selection, especially during later filial generations, ensures adaptation to the changing conditions, as outlined in the previous paragraphs. Is the tempo fast enough though? According to climate projections, the pace at which the climate will change will increase, meaning that the breeding process must also pick up the pace. How to achieve that? The most significant promise is held by DNA-based methods, such as GS or MAS, which recently

has become routine in many breeding programs. These promises are also even bigger with advancements in phenomics and phenotyping technologies, which aim to mitigate the bottleneck of phenotypic data collection.

2.5.5 Does grain yield prediction using multispectral data make sense?

After years of slumber (the "AI winter"), AI made a comeback recently taking advantage of new and abundant data sources and increased computational power. This comeback could also be seen in plant phenotyping, where the idea of GY prediction using multispectral data became viral in many crops. Although the concept of using remote sensing and multispectral reflectance for the estimation of ground covered by vegetation was not new (present already in the 1970s), this time it was paired with ML with the aim to predict grain yield in field trials before the harvest. The ML models are able to predict GY based on reflectance in usually five spectral regions: red, green, blue, red edge (rapid increase in sub-red reflectance), and near infra-red. Based on those bands, VIs are derived as linear combinations of the spectral regions, for instance NDVI – normalized difference vegetation index. VIs are then fed into a model, which associates their values with plant output. This methodology works, but why?

One fundamental problem with ML models is that they, unlike, for instance, linear least-squares models, can hardly be understood and interpreted - the relationships between input variables and model output are abstract. Even though such ML models for GY prediction may use many VIs are predictor variables, the discussion here can be narrowed down to the five spectral bands. VIs are their linear combinations, which do not increase the dimensionality of the dataset. Therefore, we know that five variables with the added time dimension can predict GY somehow.

GY, a genetically complex and highly-quantitative trait, can also be seen as a product of the plant's state or “health” throughout the growing season and genotype-specific yield potential. Factors which affect plant "health," such as water or nutrient deficiency or diseases, will ultimately deviate achieved GY from the potential maximum. Therefore, GY can be viewed as a function of genotypic yield potential and plant health over the season, or rather its calculus over time. This view allows us to assume that if we had a proxy of plant "health", it would be theoretically possible to predict GY. Multispectral canopy data can be used as such a proxy. For instance, NDVI is widely used today to assess plant health. Spectra such as red edge and near infra-red are associated with photosynthetic capacity and react to water status changes, which are related to "plant health”. By knowing the function of plant health over the season, mathematically estimating GY should be possible. Why is then ML involved?

The first problem is the quality of estimation of the plant's health function. For best estimation, the number of time points (data capture sessions) should be the highest possible, but is limited by economic and technical factors. The best automated facilities can deliver image data with a temporal resolution of hours, and cost-effective and mobile platforms such as UAVs have resolution measured in days. Moreover, lighting conditions further challenge data capture (as most sensors are passive), lowering the temporal resolution. Therefore, precisely enough estimating such a function is not possible using such equipment. Another problem is genotypic effects on multispectral reflectance, as GY prediction is often applied in field trials with multiple genotypes. Different genotypes can have different canopy reflectance patterns, potentially resulting in different plant health functions. This increases the complexity of the analysis, which becomes degree of freedom-deficient when conducted using "traditional" methods. However, the rationale behind GY prediction appears sound, and ML can tackle the data imperfections through a more abstract association of spectral reflectance and GY.

Despite their success in predicting GY, ML models have another problem: a model developed and successfully validated in environment X rarely works in environment Y, even if tested on identical genotypes. This environment-specificity can be explained by the fact that each environment is unique. Taking a snapshot of "plant health" at any given time-space can hardly be practically repeated under field conditions. Such a snapshot is practical for comparing genotypes grown in a single field trial under the same conditions, explaining the prediction ability. Attempting to use this snapshot in another environment theoretically has to fail due to presence of GEI. Providing a detailed description of the environmental conditions experienced by the plants while being phenotyped would probably increase the cross-environment prediction ability.

2.5.6 Why can M matrix predict grain yield?

GY prediction using M, in its essence, is equivalent to the GBLUP method of genomic prediction. In GBLUP, the most critical component is the so-called G matrix, which can be interpreted as a genomic relationship among studied genotypes. G matrix, in statistical terms, is a variance/covariance matrix computed based on SNP marker scores for each genotype. Its interpretation as a relationship matrix is possible due to the scaling and centering of marker values. Having the G matrix computed, we know approximately how related genotypes of interest are, similar to the information in the pedigree-based kinship matrix (A). This information is used to predict breeding values/phenotypes of genotypes that were not phenotyped but whose relationship with phenotyped records is known. The power of the G matrix comes from the fact

that markers are fully heritable and therefore are good proxies for estimating relationships.

Using multispectral phenotypes to derive genotypic relationships presents a different approach: multispectral phenotypes are far from fully heritable. They respond to the environment, unlike SNP markers. Multispectral relationships are calculated analogically to the G matrix, the difference being that SNP marker scores are replaced with multispectral phenotypes. What is the point of computing such relationships?

Considering a single drone data capture session with a multispectral camera, we obtain a snapshot of genotypes that experience similar conditions. If multispectral phenotypes are proxies of "plant health", then genotypes with a similar response to the conditions would then be related stronger than genotypes with different responses. Genotypes with similar responses will assumably handle the environment similarly and reach a similar proportion of their respective GY potentials as their products. This process repeated for multiple data capture sessions over the season yields the similarity of genotypic response to the particular environment that carries information about the final GY.

The critical difference between multispectral relationships and GY prediction using ML is that ML focuses on numerical phenotype values. In contrast, the M matrix derives similarities among genotypes based on those numerical phenotypes.

2.5.7 Why does M matrix GY prediction work?

The conclusions about spectral band importance challenge the theoretical rationale for the usefulness of the M matrix described in the previous paragraph. Based on the rationale, the bands which react to "plant health" should be the most critical (near-infrared and red edge). The results have shown that all bands but near infra-red have similar and high prediction ability and that removing the red edge and near infra-red bands causes only a slight loss in accuracy. Therefore, it is reasonable to conclude that the prediction ability of this method *does not* rely on the similarity of response to the environment. What can be then the reason for the M matrix GY prediction to work?

There are two possible explanations: either the bands carry information linked to GY, or the band information is valuable only because its heritable. Before discussing further, it is necessary to ponder what UAV multispectral imagery actually captured in the experiment.

Images were taken directly above experimental plots at 20 m with a multispectral camera, and the resolution allowed for identifying fine details of the plots (as

described in Paper III Supplementary). The ground was not masked in the images; therefore, a part of the pixels in each plot was soil, which has vastly different reflectance properties than plant canopy. Each plot was later cropped to size from the images and summarized by extracting median values for each spectral band. Therefore, the median values capture two semantic elements: plant canopy and soil. The values then represent both canopy coverage and canopy properties of a genotype at a given time.

It could be speculated that, for instance, the green reflectance points to chlorophyll parameters or content, similarly to the red edge band, but then why do red and blue bands also show almost identical accuracies? It is unlikely that different biological merits behind different bands would yield so similar accuracies. Maybe it is a single trait linked with GY that is captured by all bands?

The only trait captured by all bands in UAV multispectral imagery is crop canopy coverage. It is a likely culprit to be the underlying trait that allows for GY prediction. This hypothesis is supported by the fact that near infra-red showed lower accuracy due to its less heritable values (due to its reaction pattern to water content in both canopy and soil). The experiment showed that data capture time during the growing season affects accuracy and that the highest accuracy is during grain filling. If canopy coverage is the driver of the GY prediction accuracy, then canopy coverage during grain filling explains differences in GY to the largest degree. It is a reasonable guess, as higher canopy coverage allows for higher absorption of photosynthetically-active radiation and, therefore, more resources to form the final GY.

Another possibility is that plants that "look alike, yield alike", meaning that the underlying association is correlative, not causative, and is usable only because it is heritable. However, more experimentation and analysis are needed to go beyond speculation about the nature of the M matrix. Literature on the subject is scarce and primarily concentrates on practical aspects of the approach rather than its scientific merit.

2.5.8 Where does grain yield prediction using M matrix fit?

First, GY prediction using the M matrix is *not* an alternative to GP (genomic prediction), even though the core statistical protocol is standard.

GP's main "selling point" is the ability to predict breeding values of genotypes concerning a trait or selection index based on their DNA without the need to phenotype. This approach presents excellent opportunities, especially with methods like speed breeding. For instance, selection for GY can be made based on a single plant

in a greenhouse during early filial generations. At the same time, traditionally, GY estimation requires reasonably-sized trial plots available only in later filial generations. This advantage led to the great popularity of the method and its integration into many breeding programs.

At the same time, the protocol used for GY prediction using the M matrix requires observation of multispectral phenotypes *in situ*; therefore does not fit well for early generations in breeding programs. Especially in light of the uncertain biological merit outlined in the previous paragraph, it is doubtful if the M matrix would work on a single-plant or row level.

The M matrix has a prediction ability comparable to the G matrix in the setting with larger field trial plots. More importantly, it showed a synergistic effect with the G matrix - models with G and M matrices were superior to G or M alone. Therefore, including multispectral data in GP protocols provides an interesting additional layer of information to increase the accuracy of the already available GP protocols. The experiment also showed that even as little as a single data capture during the growing season has value and can be acquired using very budget-friendly hardware solutions. An added benefit of multispectral data captured using UAVs is its scaling. Any sample genotyping or analysis scales linearly with the number of samples, while UAV phenotyping of more material does not increase the workload linearly. The UAV phenotyping baseline workload may be significant, but phenotyping 100 or 200 plots are nearly similar in resources and effort needed. Many plant breeding companies and institutes have already shown interest in UAV technology for plant phenotyping, and this work presents a way to utilize their data for both standalone GY prediction as well as augmenting existing GP protocols at low cost. Additionally, the methodology and software used is already available, well-documented, and does not require much customization.

2.6 Conclusion

This thesis provided insight into Norwegian spring wheat's past, present, and potential future improvements in grain yield and other traits.

Genetic progress in grain yield from 1972 to 2019 was documented, describing significant genetic gains that were not dependent on intensive agronomic input over time. Genetic gains in grain yield were mainly associated with incrementing grains per spike and grains per area and were compared with other wheat historical collections worldwide. New varieties are consistently better adapted to the changing environment and outperform old varieties in resource use efficiency.

It was possible to pinpoint several robust loci associated with grain yield, earliness, and plant height. Several previously described and novel loci were detected to aid marker-assisted selection. Two well-described genes associated with spike architecture and kernel size were discovered in Norwegian wheat with changing frequency over the decades and with validated effects in current wheat breeding lines, providing insight into genomic regions under historical selection pressure.

A semi-new methodology of using drone multispectral phenotypes for grain yield prediction analogous to genomic prediction was developed, tested, and compared in single and multi-environment scenarios. The methodology is not an alternative to genomic prediction, but the standalone prediction accuracies of the two methods are comparable. Even though the rationale behind the ability to predict grain yield by multispectral data was challenged, the usefulness of the developed protocol is significant in the later stages of breeding programs. The methodology and existing genomic prediction protocols can allow for the much-needed acceleration of genetic gains. This work also laid the foundation for follow-up studies, including the genetic basis for historical yield increase connected with changes in spike architecture and the biological merit behind the ability of multispectral data to predict grain yield in crops.

References

Abbate, P.E. *et al*. (1998). Grain yield increase in recent Argentine wheat cultivars. *Crop Science*, 38(5), pp. 1203–1209.
Available at: https://doi.org/10.2135/cropsci1998.0011183X003800050015x

Achilli, A.L., Roncallo, P.F. and Echenique, V. (2022). Genetic Gains in Grain Yield and Agronomic Traits of Argentinian Durum Wheat from 1934 to 2015. *Agronomy*, 12(9).
Available at: https://doi.org/10.3390/agronomy12092151

Ahrends, H.E. *et al*. (2018). Genetic yield gains of winter wheat in Germany over more than 100 years (1895-2007) under contrasting fertilizer applications. *Environmental Research Letters*, 13(10). Available at: https://doi.org/10.1088/1748-9326/aade12

Aich, S. & Stavness, I. (2018). Global Sum Pooling: A Generalization Trick for Object Counting with Small Datasets of Large Images. Available at: http://arxiv.org/abs/1805.11123.

Aisawi, K.A.B., *et al*. (2015). The physiological basis of the genetic progress in yield potential of C.I.M.M.Y.T. spring wheat cultivars from 1966 to 2009. *Crop Science*, 55(4), pp. 1749–1764. Available at: https://doi.org/10.2135/cropsci2014.09.0601

Akin, B. *et al*. (2017). Genetic gains in grain yield in spring wheat in Turkey. *Turkish Journal of Agriculture and Forestry*, 41(2), pp. 103–112.
Available at: https://doi.org/10.3906/tar-1611-50

Alahmad, S. *et al*. (2022). Accelerating Breeding Cycles. In: *Wheat Improvement*. Springer International Publishing, pp. 557–571.
Available at: https://doi.org/10.1007/978-3-030-90673-3_30

Allen, A.M. *et al*. (2017). Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal*, 15(3), pp. 390–401. Available at: https://doi.org/10.1111/pbi.12635.

Alom, M.Z. *et al*. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics,* 8(3):292. Available at: https://doi.org/10.3390/electronics8030292

Appels, R. *et al*. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403).
Available at: https://doi.org/10.1126/science.aar7191

Araus, J.L. & Cairns, J.E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science*, 19(1), pp. 52–61.
Available at: https://doi.org/10.1016/j.tplants.2013.09.008

Araya, A. *et al*. (2017). The challenges and opportunities for wheat production under future climate in Northern Ethiopia. *Journal of Agricultural Science*, 155(3), pp. 379–393.
Available at: https://doi.org/10.1017/S0021859616000460

Arya, S. *et al*. (2022). Deep learning: as the new frontier in high-throughput plant phenotyping. *Euphytica* 2018, 47.
Available at: https://doi.org/10.1007/s10681-022-02992-3

Baker, R.J. (1988). Tests for crossover genotype-environmental interaction. *Can J Plant Sci*, 68, pp. 405-410.

Belova, T. *et al*. (2013). Integration of mate pair sequences to improve shotgun assemblies of flow-sorted chromosome arms of hexaploid wheat. *B.M.C. Genomics,* 14(1).
Available at: https://doi.org/10.1186/1471-2164-14-222

Belova, T. *et al*. (2014). Utilization of deletion bins to anchor and order sequences along the wheat 7B chromosome. *Theoretical and Applied Genetics*, 127(9), pp. 2029–2040. Available at: https://doi.org/10.1007/s00122-014-2358-z

Ben-Ari, G. & Lavi, U. (2011). Marker-assisted selection in plant breeding. In: *Plant Biotechnology and Agriculture: Prospects for the 21st Century*. Elsevier, pp. 163–184. Available at: https://doi.org/10.1016/B978-0-12-381466-1.00011-0

Blackburn, A.G. (1998). Quantifying Chlorophylls and Carotenoids at Leaf and Canopy Scales: An Evaluation of Some Hyperspectral Approaches. *Remote Sensing of Environment*, 66(3), pp. 273-285. Available at: https://doi.org/10.1016/S0034-4257(98)00059-5

Borlaug, N.E. (2007). Sixty-two years of fighting hunger: Personal recollections. *Euphytica*, pp. 287–297. Available at: https://doi.org/10.1007/s10681-007-9480-9

Brainerd, E. & Menon, N. (2014). Seasonal effects of water quality: The hidden costs of the Green Revolution to infant and child health in India. *Journal of Development Economics*, 107, pp. 49–64. Available at: https://doi.org/10.1016/j.jdeveco.2013.11.004

Braun, H., Atlin, G. and Payne, T. (2010). Multi-location testing as a tool to identify plant response to global climate change. Climate change and crop production. In: *Climate change and crop production*. CABI, pp. 115–138. Available at: https://doi.org/10.1079/9781845936334.0115 (Accessed: 10 February 2023)

Brisson, N. *et al*. (2010). Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *Field Crops Research*, 119(1), pp. 201–212. Available at: https://doi.org/10.1016/j.fcr.2010.07.012

Bugbee, B. & Koemer, G. (1997). Yield comparisons and unique characteristics of the dwarf wheat cultivar 'USU-Apogee'. Adv Space Res, 20(10), pp. 1891-1894. Available at: https://doi.org/10.1016/s0273-1177(97)00856-9

Burud, I. *et al*. (2017). Exploring Robots and UAVs as Phenotyping Tools in Plant Breeding. *IFAC-PapersOnLine*, 50(1), pp. 11479–11484. Available at: https://doi.org/10.1016/j.ifacol.2017.08.1591

Cargnin, A., de Souza, M.A. and Fronza, V. (2008). Progress in breeding of irrigated wheat for the Cerrado region of Brazil. *Crop Breeding and Applied Biotechnology*, 8(1), pp. 39–46. Available at: https://doi.org/10.12702/1984-7033.v08n01a06

Chapman, S.C. *et al*. (2007). Relationships between height and yield in near-isogenic spring wheats that contrast for major reduced height genes. *Euphytica* 157, pp. 391–397. Available at: https://doi.org/10.1007/s10681-006-9304-3

Charles, H. *et al*. (2010). Food Security: The Challenge of Feeding 9 Billion People. *Science*, Vol. 327, Issue 5967, pp. 812-818. Available at: https://doi.org/10.1126/science.1185383

Cho, C.H., B. H. Hong, M. W. Park, J. W. Shim, and B. K. Kim (1980). Origin, dissemination, and utilization of wheat semi-dwarf genes in Korea. *Ann. Wheat Newsletter*, 27(67).

Choudhary, S. *et al*. (2018). A review: Pesticide residue: Cause of many animal health problems. *Journal of Entomology and Zoology Studies*, 6(3), pp. 330-333.

Collard, B.C.Y., *et al*. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* 142, pp. 169–196. Available at: https://doi.org/10.1007/s10681-005-1681-5

Costa, C. *et al*. (2019). Plant phenotyping research trends, a science mapping approach. *Frontiers in Plant Science*, Vol. 9. Available at: https://doi.org/10.3389/fpls.2018.01933

Covarrubias-Pazaran, G. (2016). Genome-Assisted prediction of quantitative traits using the R package sommer. *PLOS ONE*, 11(6).
Available at: https://doi.org/10.1371/journal.pone.0156744

Crossa, J. *et al*. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science*, Vol. 22, Issue 11, pp. 961–975.
Available at: https://doi.org/10.1016/j.tplants.2017.08.011

Dalrymple, D.G. (1985). The Development and Adoption of High-Yielding Varieties of Wheat and Rice in Developing Countries. *American Journal of Agricultural Economics*, 67, pp. 1067-1073. Available at: https://doi.org/10.2307/1241374

David, E. *et al*. (2020). Global wheat head detection (GWHD) dataset: A large and diverse dataset of high-resolution RGB-labelled images to develop and benchmark wheat head detection methods. *Plant Phenomics*, Vol. 2020.
Available at: https://doi.org/10.34133/2020/3521852

Davies, W.P. (2003). An Historical Perspective from the Green Revolution to the Gene Revolution. *Nutr Rev* 61, pp. 124-134.
Available at: https://doi.org/10.131/nr.2003.jun.S124-S134

Davis, K.F. *et al*. (2018). Alternative cereals can improve water use and nutrient supply in India. *Science Advances*, Vol. 4, Issue 7.
Available at: https://doi.org/10.1126/sciadv.aao1108

de Los Campos, G. *et al*. (2013). Genome-enabled prediction using the BLR (Bayesian Linear Regression) R-package. *Methods in Molecular Biology*, 1019, pp. 299–320.
Available at: https://doi.org/10.1007/978-1-62703-447-0_12

de Miranda, M.S. *et al*. (2015). Environmental Impacts of Rice Cultivation. *American Journal of Plant Sciences*, 06(12), pp. 2009–2018.
Available at: https://doi.org/10.4236/ajps.2015.612201

del Pozo, A. *et al*. (2022). Genetic Yield Gains and Changes in Morphophysiological-Related Traits of Winter Wheat in Southern Chilean High-Yielding Environments. *Frontiers in Plant Science*, 12. Available at: https://doi.org/10.3389/fpls.2021.732988

Dixon, L.E. *et al*. (2018). TEOSINTE BRANCHED1 regulates inflorescence architecture and development in bread wheat (*Triticum aestivum* L.). *Plant Cell*, 30(3), pp. 563–581.
Available at: https://doi.org/10.1105/tpc.17.00961

Duan, T. *et al.* (2017). Dynamic monitoring of NDVI in wheat agronomy and breeding trials using an unmanned aerial vehicle. *Field Crops Research*, 210, pp. 71–80.
Available at: https://doi.org/10.1016/j.fcr.2017.05.025

Dube, E. *et al*. (2019). Genetic progress of spring wheat grain yield in various production regions of South Africa. *South African Journal of Plant and Soil*, 36(1), pp. 33–39.
Available at: https://doi.org/10.1080/02571862.2018.1469793

El-Hennawy, M.A. *et al*. (2011). Production of doubled haploid wheat lines (*Triticum aestivum* L.) using anther culture technique. *Annals of Agricultural Sciences*, 56(2), pp. 63–72.
Available at: https://doi.org/10.1016/j.aoas.2011.05.008

Eliazer Nelson, A.R.L., Ravichandran, K., and Antony, U. (2019). The impact of the Green Revolution on indigenous crops of India. *Journal of Ethnic Foods*. 6.
Available at: https://doi.org/10.1186/s42779-019-0011-9

Falconer, D. & Mackay, T. (1996). An introduction to Quantitative Genetics. 4th ed. London: Prentice Hall.

Fiorani, F. and Schurr, U. (2013). Future scenarios for plant phenotyping. *Annual Review of Plant Biology*, Vol. 64, pp. 267–291.
Available at: https://doi.org/10.1146/annurev-arplant-050312-120137

Fischer, R.A. and Quail, K.J. (1990) The effect of major dwarfing genes on yield potential in spring wheats. *Euphytica* 46, pp. 51-56.
Available at: https://doi.org/10.1007/BF00057618

Flintham, J.E., Angus, W.J., and Gale, M.D. (1997). Heterosis, overdominance for grain yield, and alpha-amylase activity in F1 hybrids between near-isogenic *Rht* dwarf and tall wheats. *Journal of Agricultural Science*, 129(4), pp. 371–378.
Available at: https://doi.org/10.1017/S0021859697004899

Flohr, B.M. *et al*. (2018). Genetic gains in N.S.W. wheat cultivars from 1901 to 2014 as revealed from synchronous flowering during the optimum period. *European Journal of Agronomy*, 98, pp. 1–13. Available at: https://doi.org/10.1016/j.eja.2018.03.009

Foley, J.A. *et al*. (2011). Solutions for a cultivated planet. *Nature,* 478(7369), pp. 337–342.
Available at: https://doi.org/10.1038/nature10452

Ganal, M.W. *et al*. (2012). Large SNP arrays for genotyping in crop plants. *Journal of Biosciences,* 37(5), pp. 821–828.
Available at: https://doi.org/10.1007/s12038-012-9225-3

Gauch, H.G. & Zobel, R.W. (1996). AMMI Analysis of Yield Trials. In: Kang, M.S. and Gauch, H.G., Eds., *Genotype by Environment Interaction*, CRC Press, Boca Raton, 85-122.
Available at: http://dx.doi.org/10.1201/9781420049374.ch4

Ghosh, S. *et al*. (2018). Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. *Nature Protocols,* 13(12), pp. 2944–2963.
Available at: https://doi.org/10.1038/s41596-018-0072-z

Gianola, D. & van Kaam, J. (2008). Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics,* 178(4), pp. 2289–2303.
Available at: https://doi.org/10.1534/genetics.107.084285

Grassini, P., Eskridge, K.M. and Cassman, K.G. (2013). Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nature Communications,* 4, 2918. Available at: https://doi.org/10.1038/ncomms3918

Han, L. *et al*. (2019). Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods,* 15(1), pp. 1–19.
Available at: https://doi.org/10.1186/s13007-019-0394-z

Hansen, C.W., Wingender, A. and Gollin, D. (2021). When agriculture drives development: Lessons from the Green Revolution. *V.O.X.E.U. Column*.
Available at: https://cepr.org/voxeu/columns/when-agriculture-drives-development-lessons-green-revolution (Accessed: 14 February 2023).

Hassan, M.A. *et al*. (2019). Accuracy assessment of plant height using an unmanned aerial vehicle for quantitative genomic analysis in bread wheat. *Plant Methods,* 15(1), pp. 1–12. Available at: https://doi.org/10.1186/s13007-019-0419-7

Hayes, B. & Goddard, M.E. (2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Genet. Sel. Evol,* 33(3), pp. 209-2029.
Available at: https://doi.org/10.1186/1297-9686-33-3-209

He, K. *et al*. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 42(2), pp. 386–397.
Available at: https://doi.org/10.1109/TPAMI.2018.2844175

Heffner, E.L., Jannink, J. and Sorrells, M.E. (2011). Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. *The Plant Genome,* 4(1). Available at: https://doi.org/10.3835/plantgenome2010.12.0029

Henderson, C.R. (1975). Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics,* 31(2), pp. 423–447. Available at: https://doi.org/10.2307/2529430

Hickey, J.M. *et al*. (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Science*, 54(4), pp. 1476–1488. Available at: https://doi.org/10.2135/cropsci2013.03.0195

Hickey, L.T. *et al*. (2019). Breeding crops to feed 10 billion. *Nature Biotechnology*, 37 pp. 744–754. Available at: https://doi.org/10.1038/s41587-019-0152-9

Hu, P. *et al*. (2018). Estimation of plant height using a high throughput phenotyping platform based on unmanned aerial vehicle and self-calibration: Example for sorghum breeding. *European Journal of Agronomy*, 95, pp. 24–32. Available at: https://doi.org/10.1016/j.eja.2018.02.004

Iizuka, K. *et al*. (2018). Advantages of unmanned aerial vehicle (UAV) photogrammetry for landscape analysis compared with satellite data: A case study of postmining sites in Indonesia. *Cogent Geoscience*, 4(1). Available at: https://doi.org/10.1080/23312041.2018.1498180

Jin, H. *et al*. (2016). Genome-wide QTL mapping for wheat processing quality parameters in a Gaocheng 8901/Zhoumai 16 recombinant inbred line population. *Frontiers in Plant Science*, 7. Available at: https://doi.org/10.3389/fpls.2016.01032

John, D.A. & Babu, G.R. (2021). Lessons From the Aftermaths of Green Revolution on Food System and Health. *Frontiers in Sustainable Food Systems*, 5. Available at: https://doi.org/10.3389/fsufs.2021.644559

Kaler, A.S. *et al*. (2022). Genomic prediction models for traits differing in heritability for soybean, rice, and maize. *BMC Plant Biology*, 22(1). Available at: https://doi.org/10.1186/s12870-022-03479-y

Kang, M. S. (2002). Genotype-environment interaction: progress and prospects. CABI Books. CABI International. Available at: https://doi.org/10.1079/9780851996011.0221.

Khush, G.S. (1995). Modern Varieties-Their Real Contribution to Food Supply and Equity. *GeoJournal*, 35, pp. 275–284. https://doi.org/10.1007/BF00989135

Khush, G.S. (1999). Green revolution: preparing for the 21st Century. *Genome*, 42(4), pp. 646-55.

Kihara, H. (1983). Origin and history of "Daruma" – a parental variety of Norin 10. In: S. Sakamoto (ed.), *Proceedings of the Sixth International Wheat Symposium*, 13-19.

Kokaly, R.F. *et al*. (2009). Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sensing of Environment*, 113(Suppl. 1). Available at: https://doi.org/10.1016/j.rse.2008.10.018

Kusmec, A., de Leon, N., and Schnable, P.S. (2018). Harnessing phenotypic plasticity to improve maize yields. *Frontiers in Plant Science*, 9. Available at: https://doi.org/10.3389/fpls.2018.01377

Lande, R. & Thompson, R. (1990). Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. Genetics, 124(3), pp. 743–756. Available at: https://doi.org/10.1093/genetics/124.3.743

Li, B. *et al*. (2019). The estimation of crop emergence in potatoes by UAV RGB imagery. *Plant Methods*, 15(1), pp. 1–13. Available at: https://doi.org/10.1186/s13007-019-0399-7

Li, B. *et al*. (2020). Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *Journal of Photogrammetry and Remote Sensing,* 162, pp. 161–172.
Available at: https://doi.org/10.1016/j.isprsjprs.2020.02.013

Li, C. *et al*. (2016). Single nucleotide polymorphisms linked to quantitative trait loci for grain quality traits in wheat. *Crop Journal*, 4(1), pp. 1–11.
Available at: https://doi.org/10.1016/j.cj.2015.10.002

Li, D. *et al*. (2021). High-Throughput Plant Phenotyping Platform (HT3P) as a Novel Tool for Estimating Agronomic Traits From the Lab to the Field. *Frontiers in Bioengineering and Biotechnology, 8*. Available at: https://doi.org/10.3389/fbioe.2020.623705

Li, H. *et al*. (2015). A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genomics*, 16(1).
Available at: https://doi.org/10.1186/s12864-015-1424-5

Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors*, 14, pp. 20078–20111. Available at: https://doi.org/10.3390/s141120078

Lillemo, M. & Dieseth, J.A. (2011). Wheat breeding in Norway. In: World Wheat Book, Vol. 2(1432), pp. 45–75.

Liniger, H.P., R. Mekdaschi Studer, C. Hauert and Gurtner, M. (2011). Sustainable Land Management in Practice – Guidelines and best Practices for Sub-Saharan Africa. TerrAfrica, World Overview of Conservation Approaches and Technologies (WOCAT) and Food and Agriculture Organization of the United Nations (FAO).

Liu, X. *et al*. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLOS Genetics*, 12(2).
Available at: https://doi.org/10.1371/journal.pgen.1005767

Lopes, M.S. *et al*. (2012). Genetic yield gains and changes in associated traits of CIMMYT. spring bread wheat in a "Historic" set representing 30 years of breeding. *Crop Science*, 52(3), pp. 1123–1131. Available at: https://doi.org/10.2135/cropsci2011.09.0467

Lu, N. *et al*. (2019). Improved estimation of aboveground biomass in wheat from RGB imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system. *Plant Methods*, 15(1), pp. 1–16. Available at: https://doi.org/10.1186/s13007-019-0402-3

Lukaszewski, A.J. *et al*. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum* L.) genome. *Science*, 345(6194).
Available at: https://doi.org/10.1126/science.1251788

Lush, J. (1943). Animal Breeding Plans. 2nd edition, pp. viii + 437. Iowa State Pres.

Maimaitijiang, M. *et al*. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sensing of Environment*, 237.
Available at: https://doi.org/10.1016/j.rse.2019.111599

Mammadov, J. *et al*. (2012). SNP markers and their impact on plant breeding. *International Journal of Plant Genomics*, 2012. Available at: https://doi.org/10.1155/2012/728398

Manès, Y. *et al*. (2012). Genetic yield gains of the CIMMYT International semi-arid wheat yield trials from 1994 to 2010. *Crop Science*, 52(4), pp. 1543–1552.
Available at: https://doi.org/10.2135/cropsci2011.10.0574

Mattilsynet (2022) Verdiprøving av plantesorter (in Norwegian). Available at: https://www.mattilsynet.no/planter_og_dyrking/plantesorter/godkjenning/verdiproving_av_plantesorter.3090 (Accessed: 28 February 2023)

Meuwissen, T., Hayes, B.J. and Goddard, M.E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 157(4), pp. 1819-1829. Available at: https://doi.org/10.1093/genetics/157.4.1819

Miao, Y. *et al*. (2022). Major Genomic Regions for Wheat Grain Weight as Revealed by QTL Linkage Mapping and Meta-Analysis. *Frontiers in Plant Science*, 13. Available at: https://doi.org/10.3389/fpls.2022.802310

Montenegro, J.D. *et al.* (2017). The pangenome of hexaploid bread wheat. *Plant Journal*, 90(5), pp. 1007–1013. Available at: https://doi.org/10.1111/tpj.13515

Montesinos-López, O.A. *et al*. (2017). Genomic Bayesian functional regression models with interactions for predicting wheat grain yield using hyper-spectral image data. *Plant Methods*, 13(1). Available at: https://doi.org/10.1186/s13007-017-0212-4

Myles, S. *et al*. (2009). Association mapping: Critical considerations shift from genotyping to experimental design. *Plant Cell* 21(8), pp. 2194–2202. Available at: https://doi.org/10.1105/tpc.109.068437

Nannuru, V.K.R., *et al*. (2022). Genetic architecture of fusarium head blight disease resistance and associated traits in Nordic spring wheat. *Theoretical and Applied Genetics*, 135(7), pp. 2247–2263. Available at: https://doi.org/10.1007/s00122-022-04109-9

Nordli *et al*. (2008). Regional trends for bud burst and flowering of woody plants in Norway as related to climate change. *International Journal of Biometeorology*, 52(7), pp. 625–639. Available at: https://doi.org/10.1007/s00484-008-0156-5

Nyquist, W.E. (1991). Estimation of Heritability and Prediction of Selection Response in Plant Populations. *Critical Reviews in Plant Sciences*, 10(3), pp. 235–322. Available at: https://doi.org/10.1080/07352689109382313

Ollinger, S. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist*, 189(2), pp. 375–394. Available at: https://doi.org/10.1111/j.1469-8137.2010.03536.x

Ortiz, R. *et al*. (2007). High yield potential, shuttle breeding, genetic diversity, and a new international wheat improvement strategy. *Euphytica*, 157, pp. 365–384. Available at: https://doi.org/10.1007/s10681-007-9375-9

Peltonen-Sainio, P., Jauhiainen, L. and Laurila, I.P. (2009). Cereal yield trends in northern European conditions: Changes in yield potential and its realization. *Field Crops Research*, 110(1), pp. 85–90. Available at: https://doi.org/10.1016/j.fcr.2008.07.007

Pérez, P. & de Los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198(2), pp. 483–495. Available at: https://doi.org/10.1534/genetics.114.164442

Piepho, H.P. & Möhring, J. (2007). Computing heritability and selection response from unbalanced plant breeding trials. *Genetics*, 177(3), pp. 1881–1888. Available at: https://doi.org/10.1534/genetics.107.074229

Pingali, P. (2007). Westernization of Asian diets and the transformation of food systems: Implications for research and policy. *Food Policy*, 32(3), pp. 281–298. Available at: https://doi.org/10.1016/j.foodpol.2006.08.001

Pingali, P.L. (2012). Green revolution: Impacts, limits, and the path ahead. *PNAS*, 109(31), pp. 12302–12308. Available at: https://doi.org/10.1073/pnas.0912953109

Prasad, S.C. (2016). Innovating at the margins: The system of rice intensification in India and transformative social innovation. *Ecology and Society*, 21(4). Available at: https://doi.org/10.5751/ES-08718-210407

Quan, X. *et al*. (2021). Genome-Wide Association Study Uncover the Genetic Architecture of Salt Tolerance-Related Traits in Common Wheat (*Triticum aestivum* L.). *Frontiers in Genetics*, 12. Available at: https://doi.org/10.3389/fgene.2021.663941

Rafalski, J.A. (2002). Novel genetic mapping tools in plants: SNPs and LD-based approaches, *Plant Science*, 162(3), pp. 329-333. Available at: https://doi.org/10.1016/S0168-9452(01)00587-8

Rajaram, S., Hettel, G.P., and International Maize and Wheat Improvement Center (1995). Conference Proceedings :*Wheat breeding at CIMMYT: commemorating 50 years of research in Mexico for global wheat improvement*. Available at: https://repository.cimmyt.org/handle/10883/1203

Ray, D.K. *et al*. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLOS One, 8(6). Available at: https://doi.org/10.1371/journal.pone.0066428

Ren, S. *et al*. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), pp. 1137–1149. Available at: https://doi.org/10.1109/TPAMI.2016.2577031

Reynolds, M. & Langridge, P. (2016). Physiological breeding. *Current Opinion in Plant Biology*, 31, pp. 162–171. Available at: https://doi.org/10.1016/j.pbi.2016.04.005

Reynolds, M. P. & Braun, H. J. (2022). Wheat Improvement. In: *Wheat Improvement. Food Security in a Changing Climate*. Springer International Publishing, pp. 3–15. Available at: https://doi.org/10.1007/978-3-030-90673-3_1

Richie, H. (2017). Yields vs. Land Use: How the Green Revolution enabled us to feed a growing population. Available at: https://ourworldindata.org/yields-vs-land-use-how-has-the-world-produced-enough-food-for-a-growing-population (Accessed: 14 February 2023)

Riedelsheimer, C. *et al*. (2012). Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *PNAS*, 109(23), pp. 8872–8877. Available at: https://doi.org/10.1073/pnas.1120813109

Ronneberger, O., Fischer, P. and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Lecture Notes in Computer Science*, Vol. 9351. Available at: https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28

Ruiz-Guzman, H.A. *et al*. (2016). Field phenomics: A web based image analysis platform using open source tools. In: *Proceedings - 2015 International Conference on Computational Science and Computational Intelligence, C.S.C.I. 2015*, pp. 851–852. Available at: https://doi.org/10.1109/CSCI.2015.157

Russakovsky, O. *et al*. (2014). ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis*, 115, pp. 211-252. Available at: https://doi.org/10.1007/s11263-015-0816-y

Rutkoski, J. *et al*. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3: Genes, Genomes, Genetics*, 6(9), pp. 2799–2808. Available at: https://doi.org/10.1534/g3.116.032888

Rutkoski, J.E., Krause, M.R. and Sorrells, M.E. (2022). Breeding Methods: Line Development. In: *Wheat Improvement. Food Security in a Changing Climate*. Springer International Publishing, pp. 69–82. Available at: https://doi.org/ 10.1007/978-3-030-90673-3_5

Sayre, K.D., Rajaram, S. and Fischer, R.A. (1997). Yield potential progress in short bread wheats in northwest Mexico. *Crop Science*, 37(1), pp. 36–42. Available at: https://doi.org/10.2135/cropsci1997.0011183X003700010006x

Segura, V. *et al*. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics*, 44(7), pp. 825–830. Available at: https://doi.org/10.1038/ng.2314

Shafiee, S. *et al*. (2021). Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. *Computers and Electronics in Agriculture*, 183. Available at: https://doi.org/10.1016/j.compag.2021.106036

Sharma, R. *et al*. (2022). Plant breeding increases spring wheat yield potential in Afghanistan. *Crop Science*, 62(1), pp. 167–177. Available at: https://doi.org/10.1002/csc2.20653

Sharma, R.C. *et al*. (2012). Genetic gains for grain yield in CIMMYT spring bread wheat across international environments. *Crop Science*, 52(4), pp. 1522–1533. Available at: https://doi.org/10.2135/cropsci2011.12.0634

Shearman, V.J. *et al*. (2005). Physiological processes associated with wheat yield progress in the U.K. *Crop Science*, 45(1), pp. 175–185. Available at: https://doi.org/10.2135/cropsci2005.0175

Shi, W. *et al.* (2017). A combined association mapping and linkage analysis of kernel number per spike in common wheat (*Triticum aestivum* L.). *Frontiers in Plant Science*, 8. Available at: https://doi.org/10.3389/fpls.2017.01412

Slafer, G.A. *et al.* (2023). A 'wiring diagram' for sink strength traits impacting wheat yield potential. *Journal of Experimental Botany*, 74 (1), pp. 40–71. Available at: https://doi.org/10.1093/jxb/erac410

Smith, C. & Simpson, S.P. (1986). The use of genetic polymorphisms in livestock improvement. *Journal of Animal Breeding and Genetics*, 103(1–5), pp. 205–217. Available at: https://doi.org/10.1111/j.1439-0388.1986.tb00083.x

So, D. *et al*. (2022). Genetics, not environment, contributed to winter wheat yield gains in Ontario, Canada. *Theoretical and Applied Genetics*, 135(6), pp. 1893–1908. Available at: https://doi.org/10.1007/s00122-022-04082-3

Strand, E. (1964). Contributions by new varieties to the increase in yield level of small grain species in south-east Norway. (In Norwegian). *Scientific Reports from the Agricultural College of Norway*, 168.

Strand, E. (1975). Breeding and testing of spring wheat cultivars 1960-74. (In Norwegian) *Scientific Reports from the Agricultural College of Norway*, 191.

Strand, E. (1994). Yield progress and the sources of yield progress in Norwegian small grain production 1960-92. (In Norwegian). *Norsk landbruksforsking*, 8, pp. 111–126.

Suab, S.A. & Avtar, R. (2020). Unmanned Aerial Vehicle System (UAVS) Applications in Forestry and Plantation Operations: Experiences in Sabah and Sarawak, Malaysian Borneo. In: Avtar, R., Watanabe, T. (Editors) *Unmanned Aerial Vehicle: Applications in Agriculture and Environment*. Springer, Cham. Available at: https://doi.org/10.1007/978-3-030-27157-2_8

Sun, J. *et al*. (2017). Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. *The Plant Genome*, 10(2). Available at: https://doi.org/10.3835/plantgenome2016.11.0111

Tadesse, W. *et al*. (2010). Adaptation and performance of CIMMYT spring wheat genotypes targeted to high rainfall areas of the world. *Crop Science*, 50(6), pp. 2240–2248. Available at: https://doi.org/10.2135/cropsci2010.02.0102

Tadesse, W. *et al*. (2019). Genetic Gains in Wheat Breeding and Its Role in Feeding the World. *Crop Breeding, Genetics and Genomics*, 1. Available at: https://doi.org/10.20900/cbgg20190005

Tian, Z. *et al*. (2011). Effects of genetic improvements on grain yield and agronomic traits of winter wheat in the Yangtze River Basin of China. *Field Crops Research*, 124(3), pp. 417–425. Available at: https://doi.org/10.1016/j.fcr.2011.07.012

Tillett, B.J. *et al*. (2022). Genes Impacting Grain Weight and Number in Wheat (*Triticum aestivum* L. *ssp. aestivum*). *Plants*, 11(13). Available at: https://doi.org/10.3390/plants11131772.

Tilman, D. *et al*. (2011). Global food demand and the sustainable intensification of agriculture. *PNAS*, 108(50), pp. 20260–20264. Available at: https://doi.org/10.1073/pnas.1116437108

Tirado, S.B., Hirsch, C.N. and Springer, N.M. (2020). UAV-based imaging platform for monitoring maize growth throughout development. *Plant Direct*, 4(6), pp. 1–11. Available at: https://doi.org/10.1002/pld3.230

Trethowan, R.M., van Ginkel, M. and Rajaram, S. (2002). Progress in breeding wheat for yield and adaptation in global drought affected environments. *Crop Science*, 42(5), pp. 1441–1446. Available at: https://doi.org/10.2135/cropsci2002.1441

Trevisan, R. *et al*. (2020). High-throughput phenotyping of soybean maturity using time series UAV imagery and convolutional neural networks. *Remote Sensing*, 12(21), pp. 1–19. Available at: https://doi.org/10.3390/rs12213617

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), pp. 4414–4423. Available at: https://doi.org/10.3168/jds.2007-0980

Voss-Fels, K.P. *et al*. (2019). Breeding improves wheat productivity under contrasting agrochemical input levels. *Nature Plants*, 5(7), pp. 706–714. Available at: https://doi.org/10.1038/s41477-019-0445-5

Waddington, S.R. *et al*. (1986). Improvement in the Yield Potential of Bread Wheat Adapted to Northwest Mexico. *Crop Science*, 26(4), pp. 698–703. Available at: https://doi.org/10.2135/cropsci1986.0011183x002600040012x

Walkowiak, S. *et al*. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature*, 588, pp. 277-283. Available at: https://doi.org/10.1038/s41586-020-2961-x

Wang, Q. *et al*. (2014). A SUPER powerful method for genome-wide association study. *PLOS One*, 9(9). Available at: https://doi.org/10.1371/journal.pone.0107684

Wang, S. *et al.* (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal*, 12(6), pp. 787–796. Available at: https://doi.org/10.1111/pbi.12183

Wang, S. *et al*. (2015). A single-nucleotide polymorphism of *TaGS5* gene revealed its association with kernel weight in Chinese bread wheat. *Frontiers in Plant Science*, 6. Available at: https://doi.org/10.3389/fpls.2015.01166

Wang, Y. *et al*. (2014). Genetic effect of dwarfing gene *Rht13* compared with *Rht-D1b* on plant height and some agronomic traits in common wheat (*Triticum aestivum* L.). *Field Crops Research*, 162, pp. 39–47. Available at: https://doi.org/10.1016/j.fcr.2014.03.014

Watson, A. *et al*. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants*, 4(1), pp. 23–29. Available at: https://doi.org/10.1038/s41477-017-0083-8

Wilhelm, E.P., Boulton, M.I., Barber, T.E.S., Greenland, A.J. and Powell, W. (2013). Genotype analysis of the wheat semidwarf *Rht-B1b* and *Rht-D1b* ancestral lineage. *Plant Breed*, 132, pp: 539-545. Available at: https://doi.org/10.1111/pbr.12099

Winfield, M.O. *et al.* (2018). High-density genotyping of the A.E. Watkins Collection of hexaploid landraces identifies a large molecular diversity compared to elite bread wheat. *Plant Biotechnology Journal*, 16(1), pp. 165–175. Available at: https://doi.org/10.1111/pbi.12757

Xiong, H. *et al*. (2019). From Open Set to Closed Set: Counting Objects by Spatial Divide-and-Conquer. Available at: http://arxiv.org/abs/1908.06473

Yadav, R. *et al*. (2021). Genetic Gain in Yield and Associated Changes in Agronomic Traits in Wheat Cultivars Developed Between 1900 and 2016 for Irrigated Ecosystems of Northwestern Plain Zone of India. *Frontiers in Plant Science*, 12. Available at: https://doi.org/10.3389/fpls.2021.719394

Yao, Y. *et al*. (2019). Genetic gains in grain yield and physiological traits of winter wheat in Hebei Province of China, from 1964 to 2007. *Field Crops Research*, 239, pp. 114–123. Available at: https://doi.org/10.1016/j.fcr.2019.03.011

Zhang, Z. *et al*. (2007). Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. *Journal of Animal Science*, 85(4), pp. 881–885. Available at: https://doi.org/10.2527/jas.2006-656

Zhang, Z. *et al*. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42(4), pp. 355–360. Available at: https://doi.org/10.1038/ng.546

Zhou, J. *et al*. (2019). Estimation of the maturity date of soybean breeding lines using UAV-based imagery. Remote Sensing, 11(18). Available at: https://doi.org/10.3390/rs11182075

Zhou, X. *et al*. (2017). Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, pp. 246–255. Available at: https://doi.org/10.1016/j.isprsjprs.2017.05.003

Zhu, C. *et al*. (2008). Status and Prospects of Association Mapping in Plants. *The Plant Genome*, 1(1). Available at: https://doi.org/10.3835/plantgenome2008.02.0089

Zhu, T. *et al*. (2021). Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *Plant Journal*, 107(1), 303–314. Available at: https://doi.org/10.1111/tpj.15289

Paper I

Received: 27 August 2021 | Accepted: 10 January 2022 | Published online: 2 March 2022

DOI: 10.1002/csc2.20714

Crop Science

ORIGINAL RESEARCH ARTICLE

Crop Breeding & Genetics

Historical grain yield genetic gains in Norwegian spring wheat under contrasting fertilization regimes

Tomasz Mróz[1] | **Jon Arne Dieseth**[2] | **Morten Lillemo**[1]

[1]Dep. of Plant Sciences, Norwegian Univ. of Life Sciences, NO-1432 Ås, Norway

[2]Graminor AS, Bjørke Gård, Hommelstadvegen 60, NO-2322 Ridabu, Norway

Correspondence
Morten Lillemo, Dep. of Plant Sciences, Norwegian Univ. of Life Sciences, NO-1432 Ås, Norway.
Email: morten.lillemo@nmbu.no

Assigned to Associate Editor Jochum Wiersma.

Funding information
Norges Forskningsråd, Grant/Award Number: 267806

Abstract

Spring wheat is currently dominating wheat production in Norway. The introduction of combine harvesting in the 1950s spurred breeding efforts to improve lodging and preharvest sprouting resistance, and to integrate good breadmaking quality in the locally adapted germplasm. Release of landmark cultivars Runar and Reno in the 1970s revitalized the country's production and marked the onset of modern wheat cultivation in Norway. Since that time, new cultivars have been developed but little is known about the genetic basis of the achieved yield gains. We collected 21 representative cultivars released since 1972 in Norway and tested them in a multiyear field trial including two fertilization rates: 75 and 150 kg ha^{-1} N. We assessed grain yield, plant height, heading, maturity, length of grain filling period, grain protein content, protein yield, aboveground biomass, harvest index, grain weight, test weight, grains per spike, grains per square meter, and spikes per square meter and their response to fertilization. We document an annual increase in grain yield of 17.8 kg ha^{-1} (0.34%), at both rates of N fertilization. None of the traits exhibited significant genotype × management interaction. Wheat breeding has led to the development of higher-yielding cultivars with higher protein yield that mature later, have a prolonged grain-filling period, and produce more grains per spike and grains per unit area.

1 | INTRODUCTION

Wheat cropping in Norway is challenged by several factors. Severe winters limit winter wheat production and a short vegetation period causes moderate yields (on average 4.5 Mg ha^{-1}, data from 2003 to 2019; Statistics Norway, 2020), as compared with averages of other European countries with more productive systems (France, 6.98 Mg ha^{-1}; Germany, 7.52 Mg ha^{-1}; Ireland, 9.12 Mg ha^{-1}; and the United Kingdom, 7.92 Mg ha^{-1}) (FAOSTAT, data from 2003 to 2019; https://www.fao.org/faostat/en/#home). Additionally, the wet and windy climate, especially during late season, caused delayed harvest in many years and further promotes production challenges such as preharvest sprouting, lodging, and diseases like powdery mildew, fusarium head blight, and septoria nodorum blotch (Lillemo & Dieseth, 2011).

To deal with those limitations, experimental work in agriculture was initiated in Norway in 1889, quickly being followed by introducing artificial fertilizers and new cultivars. During the 1889–1962 period, wheat yields increased by approximately 13 kg ha^{-1} per year, of which 52.6% were estimated to come from introduction of new cultivars and 47.4%

Abbreviations: BM, biomass; DM, days to maturity; GF, grain filling period; GPC, grain protein content; GrPm2, grains per m^2; GY, grain yield; HI, harvest index; PC, principal component; PH, plant height; PY, protein yield; TKW, thousand-kernel weight; TW, test weight.

from improved growing technique (Strand, 1964). From 1960 to 1974, spring wheat yields were increased further by approximately 130 kg ha^{-1} per year (Strand, 1975), and the 1960–1992 period yielded, in total, an annual increase of 74 kg ha^{-1} per year, with 47% attributed to new cultivars, 22% to management, and 31% to a combination of new cultivars and management (Strand, 1994). However, little is known about recent yield progress in Norwegian spring wheat cropping.

Cultivars grown in Norway before the 1950s were susceptible to lodging, and thus had to be either windrowed or manually harvested and dried indoors due to the weather conditions in the late growing season. This rendered them not eligible to fully benefit from mechanization (single-pass harvesting) and increased use of N fertilization. The lack of fit between available cultivars and the new agronomic practice coupled with low import prices resulted in little stimulation of domestic production, nearly eradicating spring wheat in the 1960s. This spurred breeding efforts that eventually resulted in the release of two landmark cultivars, Runar and Reno (introduced in 1972 and 1975, respectively), which showed enough resilience to the Norwegian growing conditions to revitalize the wheat cropping and mark the beginning of the era of modern wheat cultivation in Norway (Lillemo & Dieseth, 2011).

Research on yield genetic gains in many wheat collections revealed that it is associated with an increase in the number of kernels per spike and kernels per unit area whereas the kernel weight has remained constant or decreased (Flohr et al., 2018; Lo Valvo et al., 2018; Sayre et al., 1997; Voss-Fels et al., 2019). The increase in kernels per spike is mostly due to the introduction of the Norin 10 dwarfing alleles *Rht-B1b* and *Rht-D1b* (Foulkes et al., 2007), which have been spread around the world with the CIMMYT germplasm used in breeding programs (Mjærum, 1992). Shortening of the straw had little effect on spike architecture; the number of spikelets per spike is unaffected, but the above-mentioned dwarfing genes are known to increase spike fertility because less assimilates are needed for the growth and elongation of stems in semi-dwarf wheat cultivars (Fischer & Stockman, 1986; Miralles et al., 1998). Yield gains are contingent on increasing biomass produced while maintaining or improving harvest index (HI) in winter wheat (Beche et al., 2014). The effects of dwarfing genes on yield and HI are mostly explained by the reduced competition for assimilates between the straw and the spike during stem elongation, resulting in an increased sink size, yielding an increased seed number (Uddin & Marshall, 1989). Yield genetic gain drivers from various collections often overlap, but it is still necessary to investigate each set because each set has distinct characteristics and pedigrees. This knowledge is essential to maintain breeding progress by evaluating gains achieved and pointing out traits that can be emphasized in future breeding (Reynolds et al., 2009; Wu et al., 2014).

Core Ideas

- Grain yield in the 1972–2019 period increased by 17.8 kg ha^{-1} (0.34%) per year due to improved cultivars.
- Grain yield gains do not rely on intensive N fertilization.
- Cultivars do not exhibit significant genotype × management interactions for any of the measured traits.
- Breeding in Norway since 1972 caused the cultivars to have a 2-d longer grain-filling period and reach physiological maturity 3 d later.
- New cultivars in Norway produce more grains per spike and grains per unit area.

Historically, yield gains have been attributed to genetic progress and crop management in equal measure. However, it is not uncommon to observe a significant contribution of the interaction between genotype and management to the yield progress (Strand, 1964). The annual genetic yield gain in high-intensity wheat systems since the 1960s has been approximately 1% per year (Abbate et al., 1998; Sayre et al., 1997; Shearman et al., 2005). However, there is an ongoing discussion as to whether the genetic gains continue or whether they are approaching a plateau phase (Grassini et al., 2013).

It has been shown that the development of new cultivars usually leads to improvement in yield, regardless of agronomic practice. The performance is consistently better under both high and low inputs (Ahlemeyer & Friedt, 2011; Ahrends et al., 2018; Voss-Fels et al., 2019). This defies the view that genetic gains are observed only under intense management and proves that novel cultivars are better adapted to their target environments.

The objectives of this study were: (a) to estimate and document grain yield (GY) progress in Norwegian spring wheat over the course of the last five decades, (b) to determine if this progress relies on N fertilization input, (c) to determine the yield components linked to this increase, and (d) to determine and document wheat agronomical trait changes over this period.

To achieve these goals, we performed a multiyear study of historical and current spring wheat cultivars present on the Norwegian market between 1972 and 2019. We assessed genetic gains in GY, yield-related and physiological traits over the course of five decades, their response to agronomical input, and the underlying traits associated with the GY increase.

TABLE 1 Overview of the cultivars used in the study, countries of origin, breeders, and years of release

Line	Cultivar	Country/breeder	YOR[a]
1	Runar	Norway/IPK	1972
2	Reno	Norway/IPK	1975
3	Tjalve	Sweden/Weibull	1987
4	Bastian	Norway/IPK	1989
5	Polkka	Sweden/Lantmännen SW Seed	1992
6	Avle	Sweden/Lantmännen SW Seed	1996
7	Zebra	Sweden/Lantmännen SW Seed	2001
8	Bjarne	Norway/Graminor	2002
9	Demonstrant	Norway/Graminor	2008
10	Krabat	Norway/Graminor	2010
11	Mirakel	Norway/Graminor	2012
12	Rabagast	Norway/Graminor	2013
13	Seniorita	Norway/Graminor	2014
14	Arabella	Poland/Danko	2014
15	Willy	Norway/Graminor	2016
16	Caress	Sweden/Lantmännen SW Seed	2017
17	Zombi	Norway/Graminor	2018
18	Alarm	Norway/Graminor	2019
19	Betong	Norway/Graminor	2019
20	Eleven	Sweden/Lantmännen SW Seed	2019
21	Felgen	Sweden/Lantmännen SW Seed	2019

[a]YOR, year of release. The year when a cultivar was listed in Norway after passing official trials.

2 | MATERIALS AND METHODS

2.1 | Plant material

We assembled a collection of 21 spring wheat cultivars released in Norway, covering historically the most widely cultivated material since 1972 as well as the current and recently released cultivars (Table 1). It represents the breeding progress achieved from 1972 (onset of modern wheat cultivation in Norway) to the present day. Except for the cultivar "Arabella," which is of Polish origin, all the other cultivars were developed in either Norway or Sweden. The year when a cultivar was officially approved by the Plant Variety Board (year of release) was used to place it on the timeline. At the start of the field experiment in 2016, cultivars 16–21 (Table 1) were still undergoing official trials and were released in the following years. Additionally, our trials included three breeding lines that were either withdrawn or rejected from official trials during the period. Those lines remained in our field trials but data from those was excluded from the analysis after lsmeans calculation.

The collection was assembled to maximize its relevance for the actual market situation over the period; therefore, the collection suffers from imbalance as the cultivars are not equally distributed on the timeline. Cultivars Runar and Reno had almost 100% market share until the release of Tjalve in 1987, which creates a 12-yr gap between 1975 and 1987. We did not attempt to forcingly fill this gap as it would decrease the collections' relevance. Recent years were marked with the release of more cultivars to the market; therefore, the 2008–2019 period includes a relatively large number of accessions (Table 1).

2.2 | Field trials

We conducted the experiment in field seasons 2016–2020 at Vollebekk Research Station (Ås, southeastern Norway, $59^{0}39'$ N, $10^{0}45'$ E). This location represents the most important southern wheat cropping region in Norway. Field season 2018 was excluded from the analysis presented in the main text due to drought (Supplemental Figure S1; Table 2) but is presented separately in the Supplemental Material. To evaluate the effect of fertilization rates on yield performance and physiological traits, two rates (managements) were applied before sowing: 75 and 150 kg ha^{-1} N (referred to as lowN and highN, respectively) of compound NPK fertilizer (YaraMila 22–3–10). The highN treatment reflects typical fertilization rates for spring wheat in Norway currently, whereas the lowN treatment was included to assess the performance of the cultivars under less intensive management. Field trials included the full set of 24 cultivars and were arranged in randomized incomplete block split-plot design with two replicates per management and block size of six, with the position of the main treatment (fertilization level) and subtreatments (cultivar) being randomized every year. Trial plots of 5 m by 1.5 m arranged in eight rows with 30-cm spacing between neighboring plots were seeded with 185 g of kernels (61, 7 g m^{-2}); 1-m alleys were sprayed out with glyphosate after emergence, leaving plots of 4-m length for harvest. Trials were sown on 12 May 2016, 24 May 2017, 3 May 2019, and 20 Apr. 2020. Following the seeding, standard local agronomic practice was followed to keep the trial plots free of weeds and plant diseases by use of herbicides (Tripali [active ingredients: florasulam + metsulfuron-methyl + tribenuron-methyl] and/or Duplosan Meko [mekoprop]), and fungicides (Proline [prothioconazole], Aviator Xpro [bixafen + prothioconazole], Forbel [fenpropimorph], and/or Comet Pro [pyraclostrobin]) at recommended doses according to needs. Border rows were planted with buffer cultivar (Bastian) to eliminate border effects. Following ripening, samples were gathered for yield component estimation and the remaining trial material was combine harvested during the first 2 wk of September.

Weather conditions throughout the field trial years were similar in terms of average monthly temperature, solar radiation, and rainfall except for the 2018 season (Table 2).

TABLE 2 Weather data for the field trial seasons 2016–2020 between April and September

Month	2016			2017			2018			2019			2020		
	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}
	°C	mm	W m^{-2}	°C	mm	W m^{-2}	°C	mm	W m^{-2}	°C	mm	W m^{-2}	°C	mm	W m^{-2}
Apr.	5.2	100	11.7	4.4	35	13.4	5.1	32	14.0	7.9	15	16.2	6.4	30	15.0
May	11.6	50	18.5	9.6	68	14.6	15.0	0	22.2	9.7	101	17.0	9.5	47	21.4
June	15.6	90	21.1	14.3	94	19.1	16.7	82	23.9	14.8	64	17.7	17.6	115	21.1
July	16.1	55	19.1	15.9	20	19.8	20.2	44	23.2	17.2	52	20.2	14.3	128	18.0
Aug.	14.6	140	14.1	14.5	104	14.9	15.4	21	14.2	16.2	110	13.7	16.2	51	15.7
Sept.	14.1	41	9.9	11.5	119	6.7	12.1	128	9.4	11.0	191	8.3	12.0	81	9.0

Note. Ir_{avg}, average solar radiation; Rf_{sum}, sum of monthly rainfall; T_{avg}, average temperature.

TABLE 3 Overview of the gathered traits, abbreviations, units, and seasons when the data was collected

Trait	Abbreviation	Unit	Collected
Grain yield	GY	Mg ha^{-1}	2016, 2017, 2019, 2020
Plant height	PH	cm	2016, 2017, 2019, 2020
Days to heading	DH	day	2017, 2019, 2020
Days to maturity	DM	day	2016, 2017, 2019, 2020
Grain filling period	GF	day	2017, 2019, 2020
Thousand-kernel weight	TKW	g	2016, 2017, 2019, 2020
Test weight	TW	g	2016, 2017, 2019, 2020
Grain protein content	GPC	%	2016, 2017, 2019, 2020
Protein yield	PY	g m^{-2}	2016, 2017, 2019, 2020
Biomass	BM	g 50 $stems^{-1}$	2019, 2020
Harvest index	HI	ratio	2019, 2020
Grains per spike	GrPS	$spike^{-1}$	2017, 2019, 2020
Grains per area	GrPm2	grains m^{-2}	2016, 2017, 2019, 2020
Spikes per area	SpPm2	spikes m^{-2}	2017, 2019, 2020

Season 2018 was marked with higher average temperatures, no rainfall, and high solar radiation from the second half of April until early June, which, despite irrigation efforts, caused severe drought stress to the trial and shortened the growing season by nearly a month. Data from season 2018 was unrepresentative compared with the "normal" growing seasons, and therefore, was analyzed separately.

Daily weather data were downloaded from the Norwegian Bioeconomy Institute weather service, station in Ås (https://lmt.nibio.no/station/5/).

2.3 | Measurements

Cultivars were evaluated for GY, plant height (PH), heading and physiological maturity, yield components (test weight [TW], thousand-kernel weight [TKW], and grains per spike), grain protein content (GPC), and aboveground biomass (BM) at maturity. Based on these variables, additional parameters such as protein yield (PY), HI, grains per area, and spikes per area were derived. Not every trait was assessed in every season (Table 3).

Assessment of heading and maturity stages was performed visually, recording the date when 50% of the plants were in the respective stage.

Plant height was measured manually at crop maturity as an average height of a sample of fertile stems, from soil bed to the top of a spike (excluding awns, if they were present).

Aboveground biomass was estimated by weighing 50 randomly selected, moisture equalized (dried at 30 ^{0}C for 5 d) fertile mature tillers. Those samples were manually threshed to estimate grains per spike.

Protein content was determined by near infrared reflectance spectroscopy on full kernels using Perten Inframatic 9200 spectrometer (Perten Instruments AB).

Grain yield per plot was dried to 13.5% moisture content, weighed, and converted to Mg ha^{-1}. A subsample of kernels was used to estimate TKW and TW.

Protein yield was calculated as GY multiplied by protein content, number of grains per m^2 (GrPm2) as GY divided

by TKW, and number of spikes per m^2 as number of GrPm2 divided by grains per spike. Harvest index was calculated as a ratio between the grain weight per 50 spikes and the biomass of 50 fertile tillers. Collecting only fertile tillers for HI estimation causes high HI values as the proportion of infertile stems is not considered.

2.4 | Data analysis

All calculations and analyses were performed in R version 4.05. Least square estimates for traits across trial years and designs were calculated using packages "lme4" and "lmerTEST" according to the following mixed model:

$$P_{ijklnos} = \mu + g_i + m_j + g \times m_{ij} + Y_k + Y : R_{kl} + Y : R : B_{kln} + Y : W_{ko} + Y : C_{ks} + e_{ijklnos}$$

where $P_{ijklnos}$ is the phenotype (trait value) of the ith cultivar in the jth management (fertilization) grown in the kth field year in the lth replicate in the nth block in the oth row and sth column; μ is the general mean, g_i is the fixed effect of the ith cultivar, m_j is the fixed effect of the jth management, $g \times m_{ij}$ is the fixed effect of the ith cultivar grown under jth management (interaction), Y_k is the random effect of kth field year, $Y : R_{kl}$ is the random effect of the lth replicate within kth field year, $Y : R : B_{kln}$ is the random effect of the nth block within the lth replicate within kth field year, $Y : W_{ko}$ is the random effect of the oth field row within the kth field year, $Y : C_{ks}$ is the random effect of the sth field column within the kth field year, and $e_{ijklnos}$ represents the error term. Fixed effects are denoted as lowercase letters, random effects are denoted by uppercase letters, interaction is indicated by "×", and nesting is indicated by ":". Row and column random effects (denoted W and C in the model, respectively) were added to additionally correct for variability within the field on top of the block effects if a spatial trend was apparent. Model 1 was used to calculate lsmeans averaged over genotypes or managements and to perform ANOVA based on the estimated fixed effects. Degrees of freedom were calculated according to Satterthwaite's method.

The lsmeans were calculated based on the full experiment with 24 cultivars (including the three lines that were rejected) to take full advantage of the trial design. The three rejected lines were removed from the analysis after lsmeans calculation.

For the estimation of trait changes over the 1972–2019 period, a linear model was used with lsmeans of trait value as a response variable and year of release as an independent variable. Other models were investigated (quadratic, cubic, and polynomial), but those more complex curves did not explain significantly more variance and were potentially overfit as our sample of cultivars is small and imbalanced. Therefore, we decided to use standard linear regression for the purpose of documenting and estimating the changes in traits over the period.

For the traits that showed improvement over the period (correlation with year of release >0.3), genetic gains per year were reported as absolute values and as percent of the predicted trait value for year 1972 (earliest cultivar in the collection) to standardize the results. Traits that showed improvement under at least one treatment level were displayed in Figure 4.

Principal component analysis was performed on least squares trait estimates for either cultivars alone or cultivars in particular management. To account for the effect of scale, variables were scaled as 1/SD.

Results were visualized using R packages: "corrplot," "ggplot2," "ggpubr," and "ggmisc."

3 | RESULTS

3.1 | Grain yield

Weather conditions throughout the field trial years were similar in terms of temperature and rainfall, except for the 2018 season (Supplemental Figure S1) when heat and drought stress reduced yields by nearly 70%, which can be seen in the average GY of 6.01, 5.81, 5.70, and 5.35 Mg ha^{-1} in seasons 2016, 2017, 2019, and 2020, respectively, compared with 1.91 Mg ha^{-1} in 2018 (Supplemental Table S1). Achieved GYs in the representative seasons are significantly higher than the national long-term average of approximately 4.5 Mg ha^{-1}. The highest-yielding cultivar in the collection is Arabella (released in 2014, 6.5 Mg ha^{-1}) and the lowest-yielding is Runar (released in 1972, 5.2 Mg ha^{-1}). Significant ($p < .05$) annual genetic gains in GY over the 1972–2019 period are observed for all the trial years and their mean except for field season 2017 (Table 4; Supplemental Table S1). Estimated annual genetic gain in GY vary from 16 kg ha^{-1} (0.33%, Season 2020) to 23.1 kg ha^{-1} (0.47%, Season 2019; Supplemental Table S1), averaging 17.8 kg ha^{-1} (0.34%) per year (Table 4). Correlation between GY and year of release of a cultivar is subject to variation among the years (ranging from 0.64 to 0.73) whereas being the strongest for the average values (0.74) (Table 3; Supplemental Table S1). Genetic gains in GY can also be observed under severe drought stress; on average 6 kg ha^{-1} (0.39%) per year, 10.3 kg ha^{-1} (0.60%) under highN, and 2.8 kg ha^{-1} (0.17%) under lowN (Supplemental Table S1).

Grain yield is significantly ($p < .001$) affected by fertilization level and cultivar. Estimated GY averaged over trial seasons for the 75 kg ha^{-1} (lowN) and 150 kg ha^{-1} (highN) fertilization levels are 4.99 and 6.52 t ha^{-1}, respectively. No significant ($\alpha = .05$) interaction among genotypes and fertilization levels is observed (Figure 1; Table 5) for GY. Genetic gains over the period are observed for both fertilization

TABLE 4 Trait estimates for cultivars in the trial averaged over fertilization levels, their mean, correlation with year of release of a cultivar, and achieved genetic gains per year

Cultivar	YOR	GY	PH	DH	DM	GF	TKW	TW	GPC	PY	BM	HI ratio	GrPS	GrPm2	SpPm2
		Mg ha^{-1}	cm	day			g		%	g m^2	g			10^3 m^{-2}	m^{-2}
Runar	1972	5.18	90.4	55.1	105.8	50.7	39.2	79.0	11.46	59.34	93.4	0.448	25.96	13.19	508.25
Reno	1975	5.19	95.3	55.8	106.9	51.1	38.3	78.8	11.30	58.66	104.9	0.429	29.44	13.54	460.05
Tjalve	1987	5.53	79.5	58.1	107.2	49.1	37.3	77.2	11.58	64.03	85.0	0.467	28.59	14.82	518.27
Bastian	1989	5.32	74.3	54.6	105.4	50.8	34.2	78.0	12.07	64.23	82.4	0.440	25.27	15.54	614.88
Polkka	1992	5.26	89.1	57.5	105.7	48.2	35.8	78.2	11.75	61.75	85.2	0.430	28.23	14.69	520.17
Avle	1996	5.51	79.8	57.4	107.8	50.4	34.5	76.5	12.08	66.53	86.0	0.472	29.32	15.95	544.11
Zebra	2001	5.68	87.5	56.2	109.4	53.3	40.0	78.2	11.46	65.14	94.4	0.436	28.67	14.19	495.02
Bjarne	2002	5.86	72.4	56.4	106.5	50.1	37.0	77.1	11.28	66.15	80.2	0.489	26.84	15.84	590.03
Demonstrant	2008	6.09	80.1	57.9	109.5	51.6	39.8	79.6	10.51	63.97	96.3	0.471	29.21	15.30	523.72
Krabat	2010	5.71	77.0	58.4	108.1	49.7	37.2	77.9	11.20	63.92	89.4	0.463	30.85	15.35	497.55
Mirakel	2012	5.82	91.0	57.6	107.2	49.6	37.7	77.6	11.16	64.72	91.0	0.439	28.46	15.43	542.18
Rabagast	2013	5.44	71.7	58.0	107.5	49.5	34.3	78.0	11.32	61.53	72.3	0.440	25.69	15.84	616.30
Seniorita	2014	5.66	85.5	58.7	109.3	50.5	35.3	79.4	11.58	65.55	87.6	0.433	29.27	16.04	547.91
Arabella	2014	6.50	82.6	55.1	110.2	55.0	39.8	77.5	10.70	69.51	95.7	0.485	30.75	16.34	531.24
Willy	2016	5.79	80.1	56.7	109.6	52.9	36.0	76.5	10.55	61.09	93.9	0.466	30.78	16.09	522.54
Caress	2017	5.69	76.8	57.l	108.7	51.7	37.3	78.3	10.75	61.15	87.1	0.475	29.27	15.25	521.00
Zombi	2018	5.80	75.2	56.2	106.8	50.5	38.0	80.9	12.17	70.52	90.8	0.470	29.02	15.24	525.30
Alarm	2019	5.88	82.8	56.8	109.6	52.9	34.6	78.2	11.29	66.31	91.9	0.443	30.07	16.96	564.17
Betong	2019	6.08	82.0	56.3	109.2	52.9	39.4	77.9	11.19	68.05	93.2	0.446	27.55	15.43	559.94
Eleven	2019	6.08	83.1	55.6	109.7	54.1	43.8	79.3	10.55	64.19	103.0	0.471	28.03	13.89	495.66
Felgen	2019	6.20	82.1	57.1	109.7	52.6	37.2	79.8	10.75	66.71	99.5	0.469	32.85	16.66	507.08
mean ± SD		5.73 ± 0.34	81.8 ± 6.02	56.8 ± 1.15	108.1 ± 1.53	51.3 ± 1.76	37.5 ± 2.38	78.3 ± 1.17	11.27 ± 0.49	64.43 ± 3.09	90.6 ± 7.4	0.456 ± 0.018	28.77 ± 1.79	15.31 ± 0.98	533.59 ± 39.05
r[a]		0.74***	0.40†	0.27†	0.70***	0.43*	0.08†	0.15†	–0.44*	0.55*	0.033†	0.31†	0.42†	0.64**	0.19†
Genetic gain[b]		0.0178–0.34%	-0.17–0.19%	–	0.072 ± 0.07%	0.051–0.1%	–	–	–0.015–0.13%	0.11–0.19%	–	0.0004–0.09%	0.052–0.19%	0.04—0.30%	–

Notes. BM, biomass; DH, days until heading; DM, days until maturity; GF, length of grain filling period; GPC, grain protein content; GrPm2, grains per square meter; GrPS, grains per spike; GY, grain yield; HI, harvest index; PH, plant height; PY, protein yield; SpPm2, spikes per square meter; TKW, thousand-kernel weight; TW, test weight; YOR, year of release.

[a]Correlation with year of release of a cultivar.

[b]Per annuum, expressed as an absolute value and as % of the predicted value for year 1972.

*Significant at the .05 probability level. **Significant at the .01 probability level. ***Significant at the .001 probability level. †NS, nonsignificant.

TABLE 5 Traits collected in the study, units, abbreviations, least squares estimates of an average of trait values within management levels and significance of model parameters genotype (g), management (fertilization, m) and genotype × management (g × m)

Trait	Unit	Estimate ± SE[a]		Significance of model parameters[b]		
				g	m	g × m
		75 kg ha^{-1} N	150 kg ha^{-1} N			
GY	t ha^{-1}	4.99 ± 0.16	6.52 ± 0.16	***	***	ns
PH	cm	80.09 ± 4.21	83.61 ± 4.21	***	***	ns
DH	day	56.95 ± 3.04	56.90 ± 3.04	***	ns	ns
GF	day	49.83 ± 3.55	52.73 ± 3.55	***	***	ns
DM	day	106.78 ± 1.45	109.63 ± 1.45	***	***	ns
TKW	g	37.29 ± 0.57	37.99 ± 0.57	***	**	ns
TW	g	78.09 ± 0.98	78.53 ± 0.98	***	***	ns
GPC	%	10.71 ± 0.32	11.76 ± 0.32	***	***	ns
PY	kg m^{-2}	53.51 ± 2.25	76.46 ± 2.25	***	***	ns
BM	g	79.76 ± 13.07	101.72 ± 13.07	***	***	ns
HI	ratio	0.442 ± 0.02	0.473 ± 0.02	***	***	ns
GrPS	spike^{-1}	26.52 ± 2.65	31.01 ±2.65	***	***	ns
GrPm2	10^3 m^{-2}	13.45 ± 0.55	17.27 ± 0.55	***	***	ns
SpPm2	m^{-2}	532.1 ± 51.8	540.5 ± 51.8	***	ns	ns

Notes. BM, biomass; DH, days until heading; DM, days until maturity; GF, length of grain filling period; GPC, grain protein content; GrPm2, grains per square meter; GrPS, grains per spike; GY, grain yield; HI, harvest index; PH, plant height; PY, protein yield; SpPm2, spikes per square meter; TKW, thousand-kernel weight; TW, test weight.

[a]Least squares estimate of trait value averaged over all cultivars and years.

[b]Two-way ANOVA.

*Significant at the .05 probability level. **Significant at the .01 probability level. ***Significant at the .001 probability level.

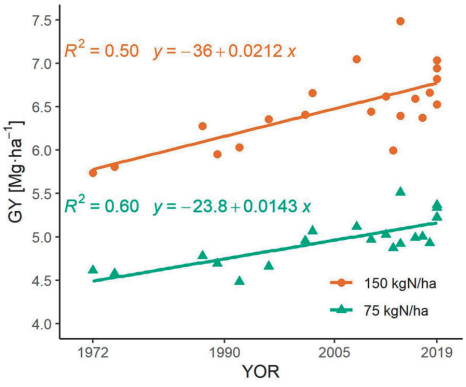


FIGURE 1 Grain yield genetic gains over the 1972–2019 period under 75 kg ha^{-1} N (green, triangles) and 150 kg ha^{-1} N (red, dots) fertilization regimes. GY, grain yield; YOR, year of release

levels, 21.2 kg ha^{-1} (0.37%) and 14.3 kg ha^{-1} (0.32%) per year for highN and lowN, respectively; however, the relationship between GY and year of release is stronger under lowN (R^2 = .6) than highN (R^2 = .5). Cultivars released recently (2014–2019) grown under lowN nearly approach GY values of the legacy cultivars (Reno and Runar) under highN (Figure 1). The genetic gains were positively validated in an independent multiyear field trial (Supplemental Figure S5).

3.2 | Agronomical and physiological traits

Significant genetic gains over the period were found for days to maturity (DM), length of grain filling period (GF), GPC, PY, and GrPm2 (Table 4).

None of the assessed traits were subject to genotype by management (g × m) interaction at 95% significance level. All traits, excluding DH and SpPm2, are strongly affected by fertilization, and significant differences among genotypes were found for every trait investigated (Table 5).

3.3 | Correlations among the traits

Under both managements and for their mean, GY is consistently and positively associated with GF and DM. A typical negative relationship between GPC and GY is observed; however, the relationship between GY and PY is positive and strong under both managements and their mean (Figure 2a,b).

Yield components associated with GY differ between managements: under lowN, GY is associated with TKW and

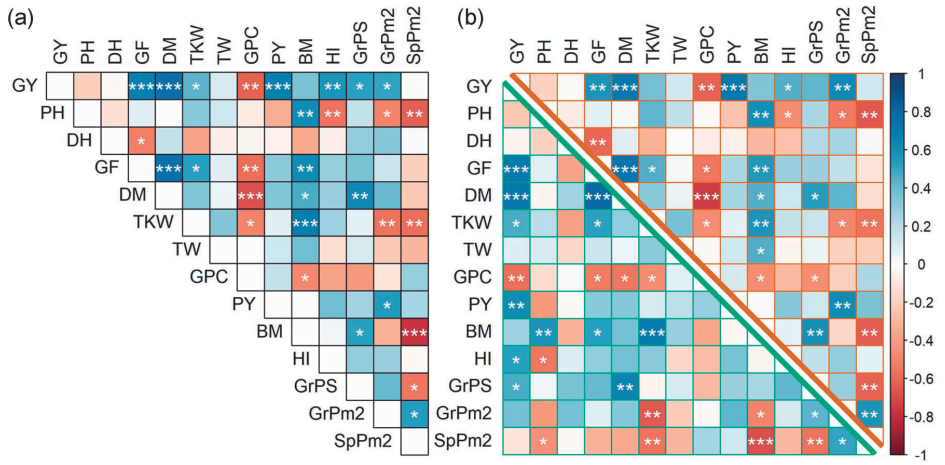


FIGURE 2 Pearson's correlation coefficient matrices for (a) genotypic means and (b) genotypic means for traits of cultivars grown under lowN (below diagonal, green border) and highN (above diagonal, red border). BM, biomass; DH, days until heading; DM, days until maturity; GF, length of grain filling period; GPC, grain protein content; GrPm2, grains per square meter; GrPS, grains per spike; GY, grain yield; HI, harvest index; PH, plant height; PY, protein yield; SpPm2, spikes per square meter; TKW, thousand-kernel weight; TW, test weight. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$

grains per spike (GrPS); and under highN, GY correlates with the number of GrPm2. For genotypic means, differences in GY can be best explained by grains per spike increase. The length of the GF is determined by DM rather than DH. GF is associated with GY and GPC due to the negative correlation between GY and GPC. Differences in biomass produced by the cultivars can be explained to a large degree by differences in PH, which, in connection with a lack of a significant relationship between GY and PH, suggests that the decrease in PH did not reduce GY achieved by the cultivars, and therefore, improvement of HI was driven by PH reduction. Number of spikes per square meter is associated under both managements with reduced PH, TKW, and BM, hinting that cultivars producing a large number of spikes per unit area tend to be shorter, produce smaller kernels, and accumulate less biomass during growing season, and therefore, possess higher HI. Longer GF is associated with higher TKW and BM production (Figure 2a,b).

3.4 | Principal component analysis

The first two principal components (PCs) explain a total of 59.4% of variance present in the dataset of genotypic means of the cultivars (Figure 3a). Biplot analysis reveals a similar correlation pattern to that in Figure 3b: a strong cluster of variables (GY, DM, GF, and GrPS) contributing to the PC with the most explanatory power. The year of release of a cultivar can be moderately explained by both PC1 ($r = .50$, $p = .021$) and PC2 ($r = .62$, $p = .002$).

Fertilization treatment clearly clusters the genotypic means under fertilization treatments (Figure 3b). The two first PCs explain a total of 65.6% of the variance present in the dataset with mostly PC1 (capturing 48.2% of the variance) determining the management clusters. Differences among the cultivars within the same cluster are determined mostly by PC2, explaining a total of 17.4% of the variance, with TKW, TW, and PH as its biggest contributors. The traits that characterize the management clusters are mainly GrPS, DM, GY, PY, and GrPm2, confirming the findings from Table 5.

3.5 | Genetic gains in agronomical and physiological traits under contrasting fertilization levels

Slopes for PH, GF, DM, PY, and GrPS are not significantly different between the treatments (Figure 4a,b,c,e,g), confirming the absence of detectable g × m interaction (Table 5). However, the slopes for GPC, HI, and GrPm2 (Figure 4d,f,h) vary between the treatments, which shows presence of minor interactions not detected by the ANOVA (Table 5).

Plant height shows a negative relationship with year of release under both treatments, but the relationship is significant only under lowN ($\alpha = .05$). This association is caused by the two old cultivars (Runar and Reno, released

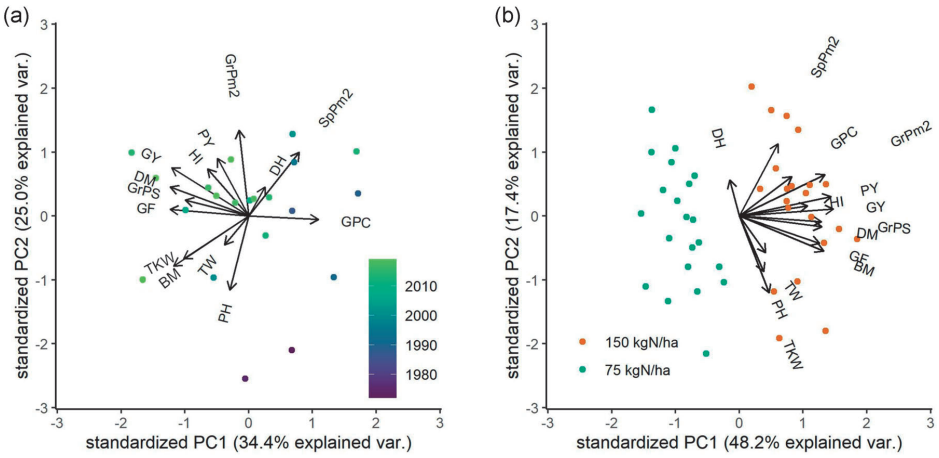


FIGURE 3 Principal component (PC) analysis on (a) trait genetic estimates and (b) genetic estimates under fertilization treatments. In (a), colors are mapped to the year of release of a cultivar, in (b), colors indicate fertilization treatments. BM, biomass; DH, days until heading; DM, days until maturity; GF, length of grain filling period; GPC, grain protein content; GrPm2, grains per square meter; GrPS, grains per spike; GY, grain yield; HI, harvest index; PH, plant height; PY, protein yield; SpPm2, spikes per square meter; TKW, thousand-kernel weight; TW, test weight

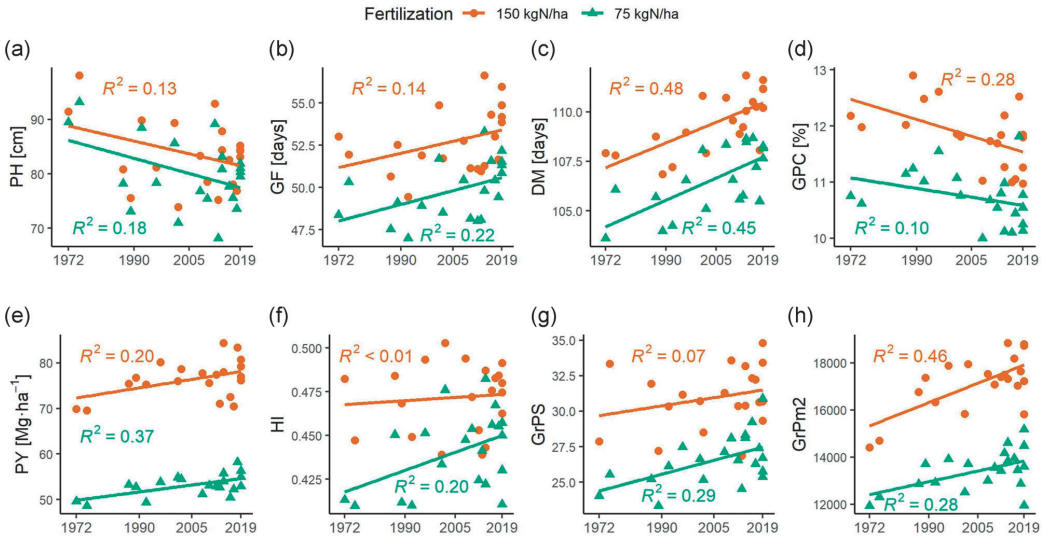


FIGURE 4 Relationships between (a) plant height, (b) grain filling period, (c) days to maturity, (d) grain protein content, (e) protein yield, (f) harvest index, (g) grains per spike, and (h) grains per m^2 cultivar estimates and year of release under 75 kg ha^{-1} N (green, triangles) and 150 kg ha^{-1} N (red, dots) fertilization. DM, days until maturity; GF, length of grain filling period; GPC, grain protein content; GrPm2, grains per square meter; GrPS, grains per spike; HI, harvest index; PH, plant height; PY, protein yield

in 1972 and 1975, respectively), showing that the newer cultivars are shorter than the old ones in Norway, but no consistent PH decrease can be documented from 1987 onward (Figure 4a).

The newer cultivars tend to have prolonged GF caused by longer DM (Figure 4b,c). Significantly ($\alpha = .05$) longer GF over the period can be observed under both fertilization treatments but is stronger under lowN. The DM was

consistently increased over the period under both treatments. Cultivars released since 2018 under lowN mature in a similar time as the old cultivars (released before 1985) under highN. The length of GF and DM were increased on average by 2 and 3 d, respectively, under both treatments.

Grain protein content shows a negative correlation with year of release under both treatments, which is stronger under highN (Figure 4d). The two old cultivars (released before 1985) decrease the correlation due to their low GPC and show that GPC consistently decreased from 1985 onward. Protein yield shows a positive relation with year of release under both treatments (Figure 4e). Despite a visible decrease in GPC (Figure 4d), GY increase (Figure 1) causes higher PY. The trend in PY is more visible under lowN.

Harvest index was increased over the period under lowN, but this relationship is not visible under highN (Figure 4f).

Grains per m^2 show an increase over time under both fertilization treatments, with a higher increase under highN (Figure 4h). Under lowN, the new cultivars tend to produce more grains per spike (Figure 4g).

4 | DISCUSSION

4.1 | Grain yield

For the present study, we took a deliberate choice to start our study with Runar, released in 1972, which marks the onset of the modern wheat cultivation era in Norway. Moreover, the older cultivars (cultivated before the 1970s) are substantially taller and more susceptible to lodging, which would have caused practical problems with the highN treatment without mechanically supporting the plants. The breeding period investigated is relatively short compared with many other studies of collections from regions with a long wheat cultivation history, including Ahrends et al. (2018), Akin et al. (2017), and Wu et al. (2014). The number of examined lines is, thus, relatively small, which limits statistical power to detect relationships.

The estimated annual genetic gains in GY for spring wheat in Norway since 1972 (on average 17.8 kg ha^{-1}, 0.34%) show similarity to gains determined in numerous collections worldwide (Ahrends et al., 2018; Crespo-Herrera et al., 2017; Dube et al., 2019; Evans et al., 1980; Oury et al., 2012; Rodrigues et al., 2007; Voss-Fels et al., 2019; Woyann et al., 2019), although they are slightly lower. The gains being relatively small show that GY was not the only priority for breeders. Wheat breeding in Norway has put a great emphasis on baking quality (due to the fact that most of the production is used for breadmaking), disease resistance (e.g., powdery mildew, *Fusarium* head blight, septoria nodorum blotch), and resistance to lodging and preharvest sprouting (Lillemo & Dieseth, 2011). Fungal diseases as well as lodging and preharvest sprouting are promoted by often rainy and windy weather during the season. The highest yielding cultivar, Arabella, is cultivated as a feed wheat due to its exceptional biomass and GY but does not meet the quality requirements for breadmaking in Norway.

The GY gains can be seen under both high and low fertilization inputs with similar cultivar ranking, showing that the source of improvement is of a genetic nature. The estimated gains are higher under high fertilizer input, which may indicate that the new cultivars are more responsive to increased N fertilization. This finding contradicts some of the results published (Ahrends et al., 2018), where the genetic progress relied strongly on the management applied (soil N availability). Our findings correspond to the results of Voss-Fels et al. (2019), where the progress was apparent across different management regimes. The new cultivars under low N input nearly approach the old ones under high input in terms of GY, which underlines genetic contribution to yield progress. Hypothesizing, it would be possible to reduce the fertilizer input by almost 50% and, by using new cultivars, still achieve yields close to those obtained with high input five decades ago. Despite the unrealistic nature of this scenario (constantly increasing national demand), it shows that breeding contributes to a more sustainable development of agriculture and leads to increased fertilizer use efficiency. This aspect is even more pronounced considering the higher PYs achieved by the newer cultivars. Based on our results, there are no apparent signs of the genetic gains approaching a plateau phase (Grassini et al., 2013).

4.2 | Agronomical and physiological traits

None of the investigated traits exhibited significant g × m interaction (at 95% confidence level), and similar findings can be found in the literature (Geren et al., 2019; Mandic et al., 2015). The small cultivar pool investigated and use of the split plot design to maximize cultivar comparisons within fertilization level leaves little statistical power to assess the minor g × m interactions that might be present. Increased fertilizer input has a positive influence on GY, PH, GPC, GF, DM, TKW, TW, GPC, PY, BM, grains per spike, and grains per area. Those results are in line with previous works (Amer, 2017; Asghar Ali et al., 2000; Mandic et al., 2015; Pradhan et al., 2018; Yu et al., 2018) and confirm the current view on fertilizer effects on crops.

4.3 | Grain yield and traits over the 1972–2019 period

The strong negative relation of GY and GPC, present under both fertilization rates, corresponds with the common view, for instance, of Monaghan et al. (2001) and Yu et al. (2018).

The increase in GY is achieved by accumulating more starch in the kernels at a cost of protein content. However, despite the decline in GPC, a positive relation of grain PY with year of release is observed. The negative relationship between GY and GPC is especially pronounced in high-yielding cultivars grown in European countries. A number of explanations for this phenomena have been proposed, but none is universally accepted (Simmonds, 1995). Bread making quality is a paramount breeding goal in Norway as the bulk of the domestic wheat production is being used by industrial bakeries. Newly released cultivars must be at least as good as cultivars already on the market in terms of yield, disease resistance, and agronomic properties, and match with the requirements for the different quality classes defined by the industry. It is a common practice nowadays in Norway to apply split fertilization with about two-thirds of N applied at sowing and the remainder at the heading stage. The latter amount is adjusted according to the yield potential of a cultivar to secure sufficient GPC, and therefore, achieve satisfactory baking quality. The trend observed in our data might be due to the fact that the highest-yielding cultivars have a higher grain set and the amount of N applied at sowing is not sufficient for them to reach their full potential in GPC. Grain protein content in wheat depends on the uptake of soil N before anthesis, its uptake during the GF, and finally, remobilization to grains of stored N in the plant. Larger N uptake before anthesis favors a higher grain number whereas late N uptake assures high GPC. It has been shown that up to 50% of total N in wheat plants at maturity may be taken up after anthesis (Austin et al., 1977; Ellen & Spiertz, 1980; Heitholt et al., 1990). Three of the cultivars present in our collection possess the wild-type allele of the *Gpc-B1* (*NAM-B1*) locus: Mirakel, Rabagast, and Polkka. The *Gpc-B1* wild-type allele is well documented to increase GPC, accelerate senescence, increase Fe and Zn content, and to reduce GY (Brevis & Dubcovsky, 2010; Uauy, Brevis, et al., 2006; Uauy, Distelfeld, et al., 2006). These three cultivars show no apparent difference in GPC, GY, and DM, showing that the general trend of decreasing GPC, delayed senescence, and increased GY was achieved by utilizing a larger number of small-effect quantitative trait loci rather than relying on a single, large effect locus.

Grain yield is strongly associated with DM. Longer GF results in an opportunity to gather more resources and allocate them in kernels, but this trait is difficult to balance under Norwegian growing conditions. Earliness is a desired trait as September (harvest time) is usually marked with frequent rainfall and wind, promoting lodging and preharvest sprouting. However, a shorter vegetative period comes with lower GYs and farmers in Norway must consider the balance between the possibility of growing later cultivars for higher yield with the risk of quality and yield loss due to difficult weather at the end of the season. Therefore, with the increase of latitude, earlier cultivars are desired in Norway due to the shortening of the available growing season. Changing climate with increased growing season temperature over the past five decades has indeed allowed for the introduction of later-maturing cultivars as seen in the trend over years for DM. The spring in Norway arrived on average 7 d earlier in 2005 than in 1971 (Nordli et al., 2008), which aligns with our findings of an average 3 d longer time period between sowing and physiological maturity.

The relationship between grains per spike and grains per area was found to be a driver of yield gains in European wheat collections (Voss-Fels et al., 2019); however, this was not present among CIMMYT (Aisawi et al., 2015) nor Chinese (Yao et al., 2019) accessions. In our study, GY is correlated with the number of grains per spike and kernel weight (under low fertilization input), and with the number of grains per area (under high input), which aligns well with other studies on European wheat (Voss-Fels et al., 2019). However, in contrast to what has been shown for CIMMYT wheat (Aisawi et al., 2015), kernel weight shows no signs of consistent improvement in the 1972–2019 period; it is grains per area and grains per spike that exhibit such an increase. By that, we conclude that the yield progress in our collection is driven by the increase in grain number both per spike and area, considering also reports from other wheat collections in Europe.

Grain protein content showed a declining trend over the year of release in this study, which is connected to the progress in GY (reverse relationship between protein content and GY). This tendency, however, is more pronounced under the high N treatment, suggesting that the highest-yielding cultivars need more N to reach their protein content potential. Protein content of the older cultivars appear to be strongly responding to soil N availability. The newer cultivars also respond to fertilization; however, the GPC difference between the treatments is smaller than for the legacy cultivars.

Harvest index was slightly improved over the years in our collection, which is visible mostly under low fertilization. No trend was observed for biomass production of the cultivars under any treatment, which hints that GY increase is driven by HI improvement rather than increases in total biomass.

5 | CONCLUSIONS

Spring wheat breeding progress represented by a collection of 21 cultivars released in Norway during the past five decades increased the average GY by 17.8 kg ha^{-1} (0.34%) per year under Norwegian growing conditions. Highly significant gains are present under both high and low fertilization treatments, providing grounds to conclude that breeding progress does not depend on intensive management. Increased fertilization has a significant positive effect on GY, PH, GPC, BM, grains per spike, TKW, TW, HI, and DM. Days to heading

and number of spikes per area are the only traits unaffected by fertilization. None of the traits were subject to a significant ($\alpha = .05$) g × m interaction. Grain yield in the collection is associated mostly with number of grains per spike and number of grains per area. Breeding led to a development of later-maturing cultivars with prolonged GF, producing more grains per spike and grains per area. Grain yield gains have been driven mostly by prolonged GF and increasing the number of kernels per spike and number of kernels per area.

ACKNOWLEDGMENTS

The main funding for this study was received from the Foundation for Research Levy on Agricultural Products (FFL) and the Agricultural Agreement Research Fund (JA) in Norway through NFR grant 267806 and Graminor. The PhD scholarship of the first author was funded by the Norwegian University of Life Sciences. We are grateful for the technical support from Cecilie Yri and Yalew Tarkegne as well as the staff at SKP Friland for the conduction of the field trials. We also acknowledge the data gathering and preliminary analysis conducted by Eivind Bleken and Bless Kufoalor from the 2016 and 2017 seasons as part of their master thesis projects.

AUTHOR CONTRIBUTIONS

Tomasz Mróz: Data curation; Formal analysis; Investigation; Methodology; Visualization; Writing – original draft. Jon Arne Dieseth: Conceptualization; Supervision; Writing – review & editing. Morten Lillemo: Conceptualization; Funding acquisition; Project administration; Supervision; Writing – review & editing.

CONFLICT OF INTEREST

Authors declare no conflicts of interest.

ORCID

Tomasz Mróz https://orcid.org/0000-0002-5469-4787
Morten Lillemo https://orcid.org/0000-0002-8594-8794

REFERENCES

Abbate, P. E., Andrade, F. H., Lázaro, L., Bariffi, J. H., Berardocco, H. G., Inza, V. H., & Marturano, F. (1998). Grain yield increase in recent Argentine wheat cultivars. *Crop Science*, *38*(5), 1203–1209. https://doi.org/10.2135/cropsci1998.0011183X003800050015x

Ahlemeyer, J., & Friedt, W. (2011). Progress in winter wheat yield in Germany: What's the share of the genetic gain? (In Austrian). *61st Conference of the Association of Plant Breeders and Seed Merchants of Austria* (Vol. *3*, pp. 19—23). Association of Plant Breeders and Seed Merchants of Austria

Ahrends, H. E., Eugster, W., Gaiser, T., Rueda-Ayala, V., Hüging, H., Ewert, F., & Siebert, S. (2018). Genetic yield gains of winter wheat in Germany over more than 100 years (1895-2007) under contrasting fertilizer applications. *Environmental Research Letters*, *13*(10), 104003. https://doi.org/10.1088/1748-9326/aade12

Aisawi, K. A. B., Reynolds, M. P., Singh, R. P., & Foulkes, M. J. (2015). The physiological basis of the genetic progress in yield potential of CIMMYT spring wheat cultivars from 1966 to 2009. *Crop Science*, *55*(4), 1749–1764. https://doi.org/10.2135/cropsci2014.09.0601

Akin, B., Sohail, Q., Ünsal, R., Dınçer, N., Demır, L., Geren, H., Sevım, İ., Orhan, Ş., Yaktubay, S., & Morgounov, A. (2017). Genetic gains in grain yield in spring wheat in Turkey. *Turkish Journal of Agriculture and Forestry*, *41*(2), 103–112. https://doi.org/10.3906/tar-1611-50

Amer, M. M. (2017). *Response of wheat yield to fertilization by nitrogen, potassium and biofertilizers in salt affected soils*. [Doctoral thesis, Department of Soil Science, Kafr EL-Sheikh]. https://doi.org/10.13140/RG.2.2.17572.88963

Ali A., Choudhry M.A., Malik M. A., Saifullah, R. A. (2000). Effect of various doses of nitrogen on the growth and yield of two wheat (*Triticum aestivum* L.) cultivars. *Pakistan Journal of Biological Sciences*, *3*(6), 1004–1005. https://doi.org/10.3923/pjbs.2000.1004.1005

Austin, R. B., Ford, M. A., Edrich, J. A., & Blackwell, R. D. (1977). The nitrogen economy of winter wheat. *The Journal of Agricultural Science*, *88*(1), 159–167. https://doi.org/10.1017/S002185960003389X

Beche, E., Benin, G., da Silva, C. L., Munaro, L. B., & Marchese, J. A. (2014). Genetic gain in yield and changes associated with physiological traits in Brazilian wheat during the 20th century. *European Journal of Agronomy*, *61*, 49–59. https://doi.org/10.1016/j.eja.2014.08.005

Brevis, J. C., & Dubcovsky, J. (2010). Effects of the chromosome region including the Gpc-B1 locus on wheat grain and protein yield. *Crop Science*, *50*(1), 93–104. https://doi.org/10.2135/cropsci2009.02.0057

Crespo-Herrera, L. A., Crossa, J., Huerta-Espino, J., Autrique, E., Mondal, S., Velu, G., Vargas, M., Braun, H. J., & Singh, R. P. (2017). Genetic yield gains in CIMMYT'S international elite spring wheat yield trials by modeling the genotype × environment interaction. *Crop Science*, *57*(2), 789–801. https://doi.org/10.2135/cropsci2016.06.0553

Dube, E., Kilian, W., Mwadzingeni, L., Sosibo, N. Z., Barnard, A., & Tsilo, T. J. (2019). Genetic progress of spring wheat grain yield in various production regions of South Africa. *South African Journal of Plant and Soil*, *36*(1), 33–39. https://doi.org/10.1080/02571862.2018.1469793

Ellen, J., & Spiertz, J. H. J. (1980). Effects of rate and timing of nitrogen dressings on grain yield formation of winter wheat (*T. aestivum* L.). *Fertilizer Research*, *1*(3), 177–190. https://doi.org/10.1007/BF01053130

Evans, L. T., Bingham, J., Blackwell, R. D., Ford, M. A., Morgan, C. L., & Taylor, M. (1980). Genetic improvements in winter wheat yields since 1900 and associated physiological changes. *The Journal of Agricultural Science*, *94*(3), 675–689. https://doi.org/10.1017/S0021859600028665

Fischer, R., & Stockman, Y. (1986). Increased kernel number in Norin 10-derived dwarf wheat: Evaluation of the cause. *Functional Plant Biology*, *13*, 767–784.

Flohr, B. M., Hunt, J. R., Kirkegaard, J. A., Evans, J. R., Swan, A., & Rheinheimer, B. (2018). Genetic gains in NSW wheat cultivars from 1901 to 2014 as revealed from synchronous flowering during the optimum period. In *European Journal of Agronomy*, *98*, pp. 1–13). https://doi.org/10.1016/j.eja.2018.03.009

Foulkes, M. J., Snape, J. W., Shearman, V. J., Reynolds, M. P., Gaju, O., & Sylvester-Bradley, R. (2007). Genetic progress in yield potential in wheat: Recent advances and future prospects. *Journal of Agricultural*

Science, *145*(1), 17–29. https://doi.org/10.1017/S0021859607006740

Geren, H., Dinçer, N., Orhan, Ş., Ünsal, R., Sevim, I., Morgounov, A., Guzman, C., Evlice, A., Yaktubay, S., Demir, L., Akin, B., Nehe, A., Sanal, T., & Ezici, A. (2019). Genotype x environment interaction and genetic gain for grain yield and grain quality traits in Turkish spring wheat released between 1964 and 2010. *PLOS ONE*, *14*(7). https://doi.org/10.1371/journal.pone.0219432

Grassini, P., Eskridge, K. M., & Cassman, K. G. (2013). Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3918

Heitholt, J. J., Croy, L. I., Maness, N. O., & Nguyen, H. T. (1990). Nitrogen partitioning in genotypes of winter wheat differing in grain N concentration. *Field Crops Research*, *23*(2), 133–144. https://doi.org/10.1016/0378-4290(90)90108-N

Lillemo, M., & Dieseth, J. A. (2011). Wheat breeding in Norway. In A. P. Bonjean, W. J. Angus, & M. Van Ginkel (Eds.), *World wheat book* (Vol. 2, pp. 45—75). http://www.lavoisier.eu/books/agriculture/the-world-wheat-book-a-history-of-wheat-breeding-volume-2/angus/description-9782743011024

Lo Valvo, P. J., Miralles, D. J., & Serrago, R. A. (2018). Genetic progress in Argentine bread wheat varieties released between 1918 and 2011: Changes in physiological and numerical yield components. *Field Crops Research*, *221*, 314–321. https://doi.org/10.1016/j.fcr.2017.08.014

Mandic, V., Krnjaja, V., Tomic, Z., Bijelic, Z., Simic, A., Muslic, D. R., & Gogic, M. (2015). Nitrogen fertilizer influence on wheat yield and use efficiency under different environmental conditions. *Chilean Journal of Agricultural Research*, *75*(1), 92–97. https://doi.org/10.4067/S0718-58392015000100013

Miralles, D. J., Katz, S. D., Colloca, A., & Slafer, G. A. (1998). Floret development in near isogenic wheat lines differing in plant height. *Field Crops Research*, *59*(1), 21–30. https://doi.org/10.1016/S0378-4290(98)00103-8

Mjærum, J. (1992). *Faginfo: Bastian brødhvete*.

Monaghan, J. M., Snape, J. W., Chojecki, A. J. S., & Kettlewell, P. S. (2001). The use of grain protein deviation for identifying wheat cultivars with high grain protein concentration and yield. *Euphytica*, *122*(2), 309–317. https://doi.org/10.1023/A:1012961703208

Nordli Ø, Wielgolaski, F. E., Bakken, A. K., Hjeltnes, S. H., Måge, F., Sivle, A., & Skre, O. (2008). Regional trends for bud burst and flowering of woody plants in Norway as related to climate change. *International Journal of Biometeorology*, *52*(7), 625–639. https://doi.org/10.1007/s00484-008-0156-5

Oury, F. X., Godin, C., Mailliard, A., Chassin, A., Gardet, O., Giraud, A., Heumez, E., Morlais, J. Y., Rolland, B., Rousset, M., Trottet, M., & Charmet, G. (2012). A study of genetic progress due to selection reveals a negative effect of climate change on bread wheat yield in France. *European Journal of Agronomy*, *40*, 28–38. https://doi.org/10.1016/j.eja.2012.02.007

Pradhan, S., Sehgal, V. K., Bandyopadhyay, K. K., Panigrahi, P., Parihar, C. M., & Jat, S. L. (2018). Radiation interception, extinction coefficient and use efficiency of wheat crop at various irrigation and nitrogen levels in a semi-arid location. *Indian Journal of Plant Physiology*, *23*(3), 416–425. https://doi.org/10.1007/s40502-018-0400-x

Reynolds, M., Foulkes, M. J., Slafer, G. A., Berry, P., Parry, M. A. J., Snape, J. W., & Angus, W. J. (2009). Raising yield potential in wheat. *Journal of Experimental Botany*, *60*(7), 1899–1918. https://doi.org/10.1093/jxb/erp016

Rodrigues, O., Lhamby, J. C. B., Didonet, A. D., & Marchese, J. A. (2007). Fifty years of wheat breeding in Southern Brazil: Yield improvement and associated changes. *Pesquisa Agropecuaria Brasileira*, *42*(6), 817–825. https://doi.org/10.1590/s0100-204X2007000600008

Sayre, K. D., Rajaram, S., & Fischer, R. A. (1997). Yield potential progress in short bread wheats in northwest Mexico. *Crop Science*, *37*(1), 36–42. https://doi.org/10.2135/cropsci1997.0011183X003700010006x

Shearman, V. J., Sylvester-Bradley, R., Scott, R. K., & Foulkes, M. J. (2005). Physiological processes associated with wheat yield progress in the UK. *Crop Science*, *45*(1), 175–185. https://doi.org/10.2135/cropsci2005.0175

Simmonds, N. W. (1995). The relation between yield and protein in cereal grain. *Journal of the Science of Food and Agriculture*, *67*(3), 309–315. https://doi.org/10.1002/jsfa.2740670306

Statistics Norway. (2020). *Cereals and oil seeds, area and yields*. Statistikkbanken. https://www.ssb.no/en/statbank/table/07480

Strand, E. (1964). Contributions by new varieties to the increase in yield level of small grain species in south-east Norway. *Scientific Reports from the Agricultural College of Norway*, *168*.

Strand, E. (1975). Breeding and testing of spring wheat cultivars 1960–74. *Scientific Reports from the Agricultural College of Norway*, *191*.

Strand, E. (1994). Yield progress and the sources of yield progress in Norwegian small grain production 1960–92. *Norsk Landbruksforsking*, *8*, 111–126.

Uauy, C., Brevis, J. C., & Dubcovsky, J. (2006). The high grain protein content gene Gpc-B1 accelerates senescence and has pleiotropic effects on protein content in wheat. *Journal of Experimental Botany*, *57*(11), 2785–2794. https://doi.org/10.1093/jxb/erl047

Uauy, C., Distelfeld, A., Fahima, T., Blechl, A., & Dubcovsky, J. (2006). A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science*, *314*(5803), 1298–1301.

Uddin, M. N., & Marshall, D. R. (1989). Effects of dwarfing genes on yield and yield components under irrigated and rainfed conditions in wheat (*Triticum aestivum* L.). *Euphytica*, *42*(1–2), 127–134. https://doi.org/10.1007/BF00042623

Voss-Fels, K. P., Stahl, A., Wittkop, B., Lichthardt, C., Nagler, S., Rose, T., Chen, T. W., Zetzsche, H., Seddig, S., Majid Baig, M., Ballvora, A., Frisch, M., Ross, E., Hayes, B. J., Hayden, M. J., Ordon, F., Leon, J., Kage, H., Friedt, W & Snowdon, R. J. (2019). Breeding improves wheat productivity under contrasting agrochemical input levels. *Nature Plants*, *5*(7), 706–714. https://doi.org/10.1038/s41477-019-0445-5

Woyann, L. G., Zdziarski, A. D., Zanella, R., Rosa, A. C., de Castro, R. L., Caierão, E., Toigo, M. D. C., Storck, L., Wu, J., & Benin, G. (2019). Genetic gain over 30 years of spring wheat breeding in Brazil. *Crop Science*, *59*(5), 2036–2045. https://doi.org/10.2135/cropsci2019.02.0136

Wu, W., Li, C., & Ma, B. (2014). Genetic progress in wheat yield and associated traits in China since 1945 and future prospects. *Euphytica*, *196*, 155–168. https://doi.org/10.1007/s10681-013-1033-9

Yao, Y., Lv, L., Zhang, L., Yao, H., Dong, Z., Zhang, J., Ji, J., Jia, X., & Wang, H. (2019). Genetic gains in grain yield and physiological traits of winter wheat in Hebei Province of China, from 1964 to 2007. *Field Crops Research*, *239*(March), 114–123. https://doi.org/10.1016/j.fcr.2019.03.011

Yu, Z., Islam, S., She, M., Diepeveen, D., Zhang, Y., Tang, G., Zhang, J., Juhasz, A., Yang, R., & Ma, W. (2018). Wheat grain protein accumulation and polymerization mechanisms driven by nitrogen fertilization. *Plant Journal*, *96*(6), 1160–1177. https://doi.org/10.1111/tpj.14096

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Mróz, T., Dieseth, J. A., & Lillemo, M. (2022). Historical grain yield genetic gains in Norwegian spring wheat under contrasting fertilization regimes. *Crop Science*, *62,* 997–1010. https://doi.org/10.1002/csc2.20714

Paper I Supplementary material

Historical Grain Yield Genetic Gains in Norwegian Spring Wheat Under Contrasting Fertilization Regimes

Tomasz Mroz[1], Jon Arne Dieseth[2], and Morten Lillemo[1]

[1]*Department of Plant Sciences, Norwegian University of Life Sciences, NO-1432 Ås, Norway*
[2]*Graminor, AS, Bjørke Gård, Hommelstadvegen 60, NO-2322 Ridabu, Norway*

Weather data

Weather data was downloaded from Norwegian Institute for Bioeconomy website (NIBIO, https://lmt.nibio.no/, weather station in Ås) and visualized using package "ggplot2". As the origin of the plot, the seeding date of the respective field trial year was taken.

Genetic relationship among the varieties and hierarchical clustering

Genomic relationship matrix according to Van Raden (2008) and hierarchical clustering were calculated and visualized using package "AGHmatrix", using genomic data already available in our laboratory.

Season 2018 (drought year) field trial

Trial was conducted in the same manner as in the remaining seasons as described in the main text in materials and methods.

Genetic gains in grain yield – validation

Genetic yield gains described were validated by comparison to the same varieties grown as a part of another field experiment of approximately 300 spring wheat lines conducted in the same years as the main trial (2016-2020, excluding 2018) in the same location. The experiment did not include the fertilization (management) factor, instead the varieties were grown under 120 kg N ha^{-1}, full fungicide, pesticide, and herbicide treatments following common practice. This fertilization, though slightly suboptimal, is a compromise to reduce risk of lodging. Due to large size of the experiment, it was arranged as an augmented alpha-lattice design. Least squares estimates for grain yield were calculated for each trial year with R package "lme4", as described in Materials and Methods, using an analogical mixed model:

$$P_{ik} = \mu + g_i + Y_k + Y{:}R_{kl} + Y{:}R{:}B_{klm} + Y{:}C_{kn} + Y{:}W_{ko} + e_{iklmno}$$

Where P_{ik} is a least square estimate of grain yield of the ith variety in kth field season (year), μ is the general mean, g_i is a fixed effect of the ith variety, Y_k is a random effect of kth field season (year), $Y{:}R_{kl}$ is a random effect of the lth rep in season k, $Y{:}R{:}B_{klm}$ is a random effect of the mth block within the lth rep in the kth season, $Y{:}C_{kn}$ is a random effect of the nth column within the kth season, $Y{:}W_{ko}$ is a random effect of the oth row within the kth season and e_{iklmno} is the corresponding error term. Results were visualized as described in Materials and methods.

Supplementary Results and Discussion

Weather data

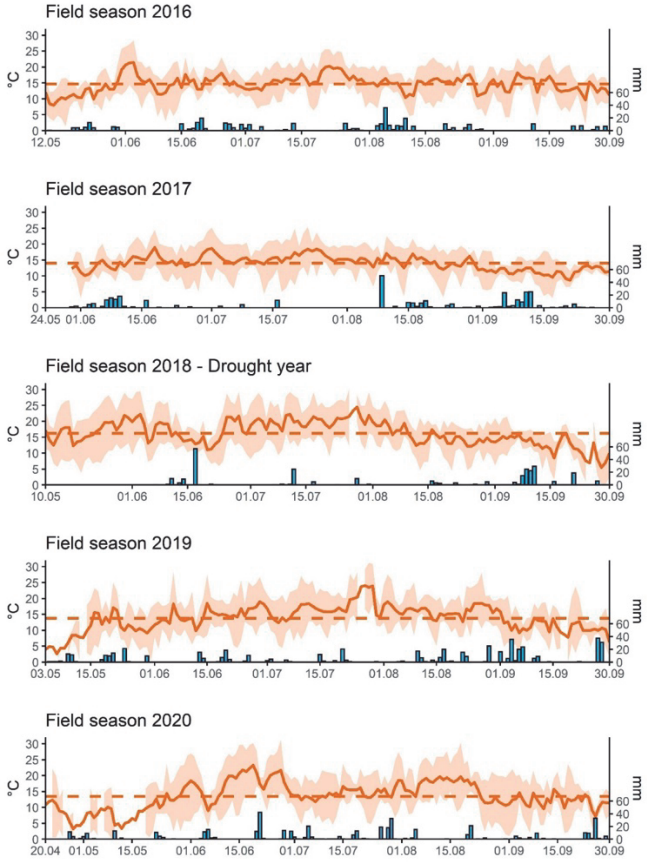


Figure S1 *Weather data for field seasons 2016-2019, from seeding the trial to the end of September. Solid line marks the average temperature for each day, dashed line marks the mean temperature across the season, bars mark the daily rainfall in mm per m². Primary y axis – temperature in ^{0}C, secondary y axis – rainfall in mm per m²*

Seasons 2016, 2017 and 2019 were approximately like one another in terms of temperature and rainfall, which has a reflection in the average yields in those years (Table 2). Season 2018 was characterized by slightly higher average temperature and very scarce (if any) rainfall during the intensive growth phase of plants (Figure S1), which resulted in severe drought stress and drastically reduced (approximately 60%) grain yield (Table S1). For this reason, we deemed 2018 trial heavily biased and therefore to not include it in the main analysis.

Table S1 *Summary of variety grain yield in tons per hectare) in the experiment years (field seasons) under fertilization treatments and their mean, correlations with year of release and annual estimated genetic gains in grain yield (in kg per hectare per year). Genetic gain (interpreted as the slope of the regression line of grain yield vs. year of release – YOR) expressed in kg per ha per year and as percentage of predicted average grain yield in year 1972.*

Variety	YOR	Year 2016			Year 2017			Year 2018 (drought)			Year 2019			Year 2020		
		75 kgN	150 kgN	mean	75 kgN	150 kgN	mean	75 kgN	150 kgN	mean	75 kgN	150 kgN	mean	75 kgN	150 kgN	mean
Runar	1972	4.72	5.89	5.31	5.06	6.06	5.56	1.70	1.32	1.51	4.31	5.50	4.90	4.37	5.35	4.86
Reno	1975	5.13	5.96	5.55	4.85	5.89	5.37	1.52	1.79	1.65	4.29	5.52	4.91	4.26	5.56	4.91
Tjalve	1987	4.79	6.43	5.61	4.74	6.44	5.59	1.89	2.11	2.00	5.22	5.93	5.57	4.55	6.32	5.43
Bastian	1989	5.40	6.46	5.93	4.81	5.96	5.39	1.46	1.66	1.56	4.68	6.07	5.38	3.88	5.32	4.60
Polkka	1992	4.36	6.13	5.25	4.94	6.94	5.94	1.79	1.89	1.84	4.07	5.64	4.86	4.16	5.66	4.91
Avle	1996	4.71	6.38	5.55	4.94	6.58	5.76	1.86	2.19	2.03	4.70	6.55	5.63	4.72	5.74	5.23
Zebra	2001	5.24	7.06	6.15	4.71	6.24	5.47	1.88	2.48	2.18	5.51	6.46	5.99	4.20	6.20	5.20
Bjarne	2002	5.22	6.77	6.00	5.04	7.08	6.06	1.68	1.79	1.73	5.60	6.53	6.07	4.47	6.08	5.27
Demonstrant	2008	5.54	7.59	6.57	5.38	7.60	6.49	2.19	2.37	2.28	5.13	6.41	5.77	4.57	6.47	5.52
Krabat	2010	5.28	6.61	5.95	5.17	6.83	6.00	1.66	2.14	1.90	5.02	5.70	5.36	4.63	6.68	5.65
Mirakel	2012	4.96	6.81	5.89	4.93	6.61	5.77	1.86	2.35	2.10	5.23	6.41	5.82	4.58	6.58	5.58
Rabagast	2013	5.07	6.31	5.69	4.78	5.19	4.99	1.74	2.31	2.03	4.74	6.45	5.60	4.53	6.25	5.38
Seniorita	2014	5.54	6.59	6.06	5.21	6.67	5.94	1.92	2.10	2.01	4.60	6.04	5.32	4.34	6.13	5.23
Arabella	2014	5.81	7.47	6.72	5.96	8.34	7.15	1.70	1.98	1.84	5.62	6.92	6.27	4.75	6.77	5.76
Willy	2016	5.34	6.82	6.08	5.47	6.76	6.11	1.84	2.20	2.02	4.46	5.77	5.11	4.89	6.66	5.77
Caress	2017	5.47	6.68	6.08	4.95	6.86	5.91	1.81	2.22	2.01	5.25	5.97	5.61	4.50	6.15	5.32
Zombi	2018	5.47	6.89	6.18	4.72	6.47	5.60	1.56	1.59	1.57	5.79	6.78	6.28	3.93	6.33	5.13
Alarm	2019	5.45	6.93	6.19	5.16	6.19	5.67	1.79	2.20	2.00	5.17	6.26	5.71	4.67	6.67	5.67
Betong	2019	5.88	7.47	6.67	4.61	6.17	5.39	1.58	1.97	1.78	5.82	7.62	6.72	4.93	5.92	5.43

Table S1 *Cont.*

Eleven	2019	5.83	7.26	6.55	4.70	6.72	5.71	1.67	1.79	1.73	6.16	7.15	6.65	4.50	6.42	5.46
Felgen	2019	5.22	7.19	6.21	5.21	7.23	6.22	2.09	2.57	2.33	5.56	6.66	6.11	5.16	6.64	5.90
Mean		5.26	6.76	6.01	5.02	6.61	5.81	1.77	2.05	1.91	5.09	6.30	5.70	4.50	6.19	5.35
Genetic gain		17.1 0.36%	23.8 0.40%	20.6 0.39%	4.4 0.09%	13.5 0.22%	9.0 0.16%	2.8 0.17%	10.2 0.60%	6.6 0.39%	23.5 0.55%	22.7 0.41%	23.0 0.47%	9.4 0.23%	22.8 0.42%	16.1 0.33%

Genetic relationship among the varieties with hierarchical clustering

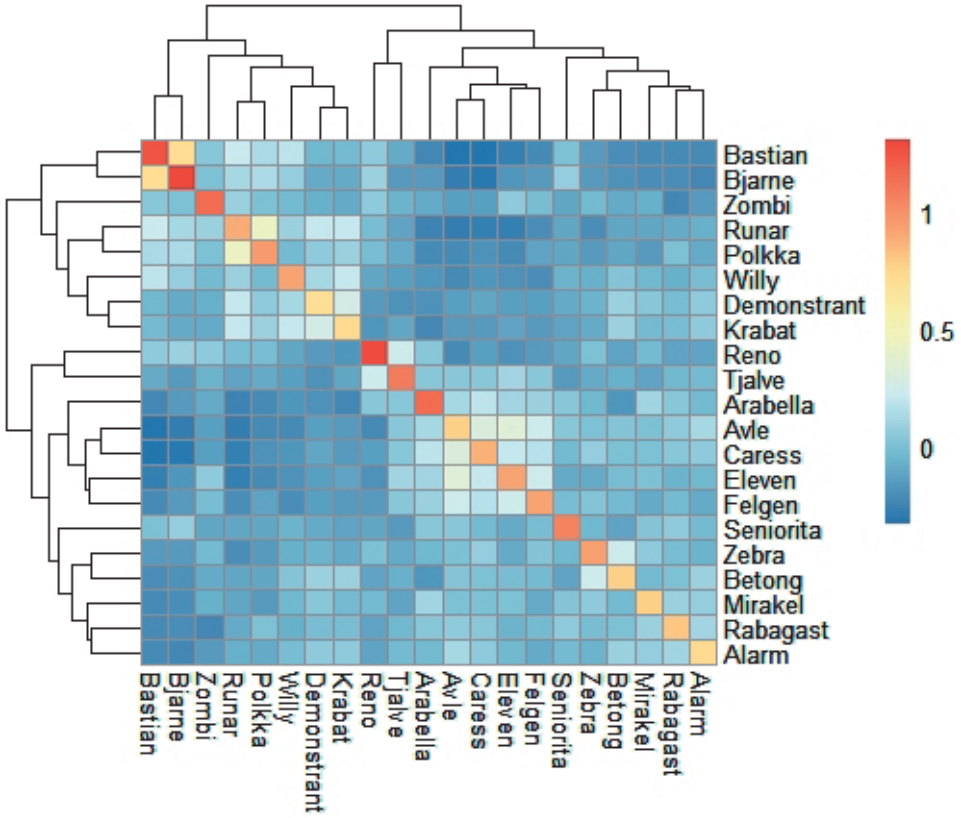


Figure S2*. Genomic relationship (G) with hierarchical clustering of the varieties present in the study, calculated according to Van Raden (2008)*

The most similar pairs of varieties are Bastian and Bjarne, Runar and Polkka, Demonstrant and Krabat, Avle and Caress. Two clusters containing highly genetically similar varieties are present: Runar, Polkka, Willy, Demonstrant and Krabat (cluster 1) and Avle, Caress, Eleven and Felgen (cluster 2). Cluster 1 contains mostly varieties developed by Graminor (Norway), while cluster 2 contains only varieties developed by Lantmännen (Sweden).

Trial data for field season 2018 (drought)

As mentioned before, grain yield in 2018 field season were reduced compared to the other years (by approximately 60%), which was caused by a combination of heat and drought stress. As a general effect seen across the wheat growing region in south-eastern Norway that year, high temperature and limited water availability during the

first month after seeding had a severe effect on tillering and grain set. Despite repeated irrigation, this combined stress of heat and drought could not be alleviated in our field trial. The data suggests that fertilization regime had a positive impact on yield but less pronounced than in the "normal" years (Table S1, Figure S3). Despite the drought and heat stress, statistically significant genetic gains in yield can still be observed, but only for the highN treatment (10.3 kg·ha^{-1} per year, $p < 0.05$).

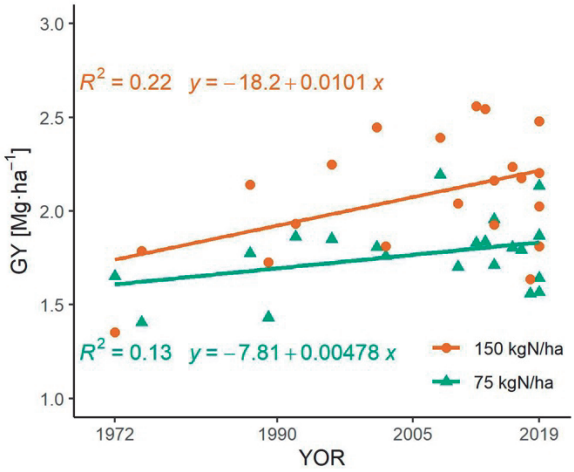


Figure S3 *Genetic grain yield gains over the 1972-2019 period under the two fertilization levels. Green – lowN (75kgN·ha^{-1}), red – highN (150kgN·ha^{-1}) fertilization*

Table S2 *Traits collected in the 2018 season, least square estimates of an average of trait values across management levels and significance of model parameters.*

Trait	Abbv.	Unit	Estimate ± SE†		Significance of model parameters††		
			75 kgN	150 kgN	genotype	management	g×m
Grain yield	GY	Mg ha^{-1}	1.77 ± 0.14	2.10 ± 0.14	*	***	ns
Plant height	PH	cm	56.41 ± 0.91	57.25 ± 0.91	**	ns	ns
Days to maturity	DM	day	94.29 ± 0.19	94.00 ± 0.19	***	ns	ns
Thousand kernel weight	TKW	g	32.75 ± 0.76	33.50 ± 0.76	***	**	ns
Test weight	TW	g	75.81 ± 0.57	76.30 ± 0.57	***	**	ns
Protein content	GPC	%	14.83 ± 0.20	16.94 ± 0.20	***	***	ns

*† Least squares estimate of trait value averaged over all varieties), signif. codes: *** p < 0.001, ** p < 0.01, * p < 0.05, ns p > 0.05*
†† Two – way ANOVA

The influence of genotype on all the measured traits in season 2018 (Table S2) is still significantly present, however, less apparent in cases of grain yield and thermal time to maturity ($p < 0.05$). Remarkably reduced plant height and increased protein content was observed in 2018 compared to the other years (Table S2, Table 3).

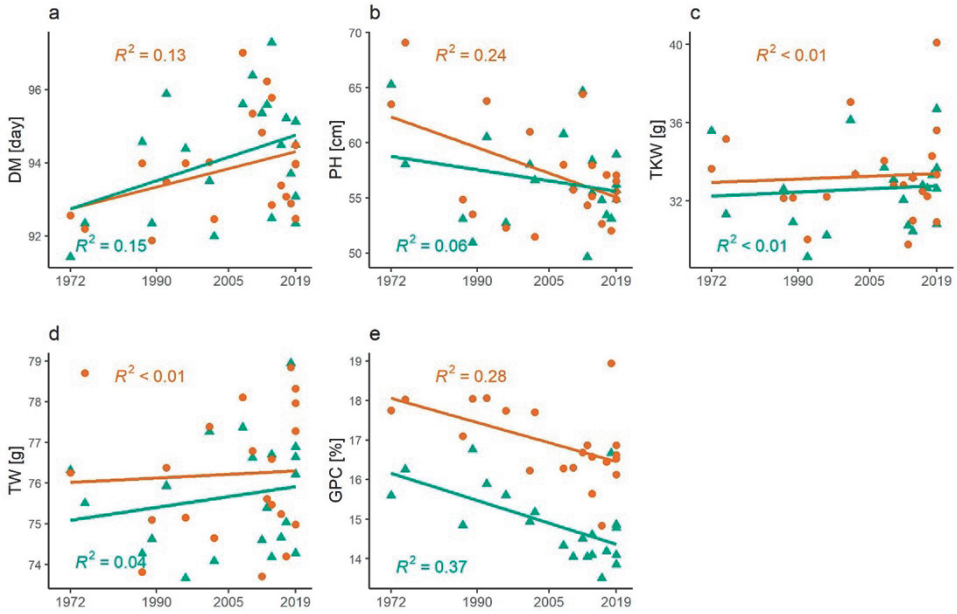


Figure S4 *Genetic gains over the 1972-2019 period in Norwegian spring wheat in season 2018 in (a) days to maturity, (b) plant height, (c) thousand kernel weight, (d) test weight and (e) grain protein content. Red – highN (150 kgN·ha^{-1}), green – lowN (75 kgN·ha^{-1})*

Under drought stress similar patterns over time can be observed as in "normal" conditions for most of traits measured in 2018.

To summarize, drought and heat reduced grain yield by approximately 60% and made the grain yield and plant height genetic gains dependent on fertilizer input. The stress conditions also reduced the effect of fertilization on yield and plant height but increased the average protein content of the kernels. Conclusions from this part should be considered approximate, as some of the results may be misleading due to the "uncontrolled" experiment conditions and uneven drought stress within the trial.

Validation of genetic gains in grain yield

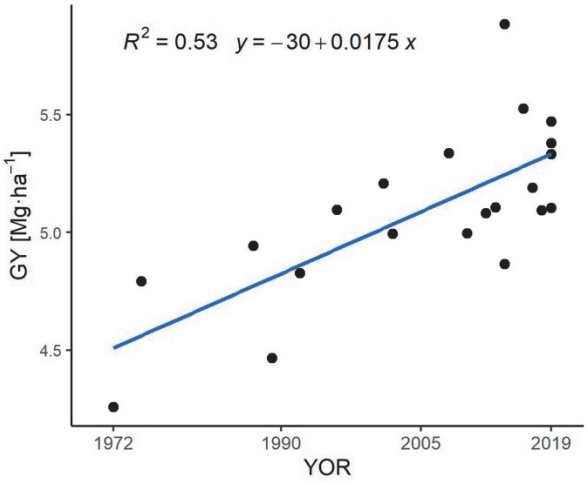


Figure S5 *Validation of grain yield genetic gains*

Estimated annual genetic gains from the validating experiment conducted over 4 years are approximately 17.52 kg·ha^{-1} (0.39%, Figure S5), which confirms the findings presented in the article.

Paper II

Grain yield and adaptation of spring wheat to Norwegian growing conditions is driven by allele frequency changes at key adaptive loci discovered by genome-wide association mapping

Manuscript

Tomasz Mróz[1], Jon Arne Dieseth[2], and Morten Lillemo[1]

[1]*Department of Plant Sciences, Norwegian University of Life Sciences, NO-1432 Ås, Norway*
[2]*Graminor, AS, Bjørke Gård, Hommelstadvegen 60, NO-2322 Ridabu, Norway*

Corresponding author:

Morten Lillemo
morten.lillemo@nmbu.no
Orcid-ID: 0000-0002-8594-8794

Abstract

Grain yields in Norwegian spring wheat increased by 18 kg ha^{-1} per year between 1972 and 2019 due to introduction of new varieties. These gains were associated with increments in the number of grains per spike and extended length of the vegetative period; however, little is known about the genetic background of this progress. To fill this gap, we conducted genome-wide association study on a panel consisting of both adapted (historical and current varieties and lines in the Nordics) and important exotic accessions used as parents in the Norwegian wheat breeding programs. The study concerned grain yield, plant height, and heading and maturity dates, and detected twelve associated loci, later validated using independent sets of recent breeding lines. Adaptation to the Norwegian cropping conditions is associated with the *Vrn-A1* locus, and a previously undescribed locus on chromosome 1B associated with heading date. Two loci associated with grain yield, corresponding to the *TaGS5-3A* and *Ta-Col5* loci, indicated historical selection pressure for high grain yield. A locus on chromosome 2A explained the tallness of the oldest accessions. We investigated the origins of the beneficial alleles associated with the wheat breeding progress in the Norwegian material, tracing them back to crosses with Swedish, German, or CIMMYT lines. This study contributes to the understanding of wheat adaptation to the Norwegian growing conditions, sheds light on the genetic basis of historical wheat improvement and aids future breeding efforts by discovering loci associated with important agronomic traits in wheat.

Key messages

Adaptation to the Norwegian environment is associated with polymorphisms in the *Vrn-A1* locus. Historical selection for grain yield in Nordic wheat is associated with *TaGS5-3A* and *Ta-Col5* loci.

Keywords

grain yield, days to maturity, days to heading, plant height, adaptation, breeding progress

Author contributions

TM: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing-original draft, and visualization.

JAD: Resources, data curation, writing – review & editing.

ML: Investigation, resources, data curation, writing – review & editing, supervision, project administration, and funding acquisition.

Conflict of interest

The authors declare no conflict of interest.

Availability of data and material

Data and material available upon reasonable request to the corresponding author.

Acknowledgements

We are grateful for the technical support from Cecilie Yri and Yalew Tarkegne as well as the staff at SKP Friland for the conduction of the field trials at Vollebekk and to Graminor for conducting the field trials at Staur.

Funding

The main funding for this study was received from the Foundation for Research Levy on Agricultural Products (FFL) and the Agricultural Agreement Research Fund (JA) through NFR grant 267806 and Graminor. The PhD scholarship of the first author was funded by the Norwegian University of Life Sciences.

Abbreviations

Chr - chromosome
CMLM – compressed mixed linear model
DArT – diversity array technology
DH – days to heading
DM – days to maturity
FarmCPU - Fixed and random model Circulating Probability Unification
FDR – false discovery rate
gBLUP – genomic best linear unbiased predictions
GP – genomic prediction
GWAS – genome-wise association study
GY – grain yield
HSD – honestly significant difference
LD – linkage disequilibrium
MAS – marker-assisted selection
MLM – mixed linear model
MLMM – multi-locus mixed model
MQTL – meta-QTL analysis
MTA – marker-trait association
PCA – principal component analysis
PH – plant height
QQ-plot – quantile-quantile plot
QTL – quantitative trait locus
QTN – quantitative trait nucleotide
SNP – single nucleotide polymorphism
SSR – simple sequence repeats
SUPER - Settlement of MLM Under Progressively Exclusive Relationship

Introduction

Hexaploid bread wheat (*Triticum aestivum* L.) is among the most essential and traded staple foods, providing around 20% of daily protein and calorie intake for approximately 4.5 billion people worldwide (Braun, Atlin and Payne, 2010). The annual genetic gain in wheat grain yield (GY) must be increased from the current levels to 1.0-1.6% to meet the food demands of the projected global population in the 2050. However, due to current and future challenges posed by climate change, such as reduced soil health, change in temperature, and erratic rainfall, maintaining and acceleration of the annual genetic gains is a challenge (FAO, 2017).

Today, wheat cropping in Norway is an integral part of the country's sustainability policy, despite being relatively small in size compared to the other Nordic countries (0.33 Mt in Norway versus 2.67 Mt in Sweden, 0.84 Mt in Finland and 4.47 Mt in Denmark) (FAOSTAT, data from 2010 to 2020; https://www.fao.org/faostat). From 1972 to 2019, wheat breeding in Norway increased grain yields by 17.8 kg ha^{-1} per year and prolonged grain filling and vegetation periods by 2 and 3 days, respectively, showing better adaptation of new cultivars to the changing climate (Mróz, Dieseth and Lillemo, 2022). To enable further genetic gains in Norwegian spring wheat by utilizing genomics-driven breeding approaches, such as marker-assisted selection (MAS) or genomic prediction (GP), a detailed association study of important agronomic and quality traits in the current breeding material is needed.

The success of a wheat crop is largely determined by adaptation, consisting of genes for phenology and development and their interactions with one another and the environment (Hyles *et al*., 2020). In order to reach maximum kernel size and number, wheat must develop biomass and flower at the time of optimal seasonal conditions (Trethowan, 2014). One of the key components of adaptation is the vernalization, defined as a requirement of low temperatures to flower. Vernalization requirement is mainly determined by the *Vrn-A1* loci on the long arm of chromosome 5 with copies in all the three wheat sub-genomes. Winter wheats usually carry the winter alleles at all three homoeologous loci, while it is not uncommon for spring wheat to carry the winter allele at *Vrn-A1* to extend the vegetative growth period (Yan *et al*., 2003, 2004, 2006). Another critical aspect of wheat adaptation is photoperiod sensitivity, largely determined by alleles of the *Ppd1* (*Photoperiod1*) gene, with homologous copies on chromosomes 2A, 2B, and 2D (Ramirez *et al*., 2018).

There have been many loci under historical selection pressure in wheat due to human-driven breeding efforts to improve yield, disease resistance, and other

desirable traits. The *Rht* genes are the main genes of the Green Revolution and contributed to dramatic gains in GY (Liu *et al*., 2017). The *Rht* gene alleles control plant height, and thus harvest index. Selection for favorable alleles in the *Rht* locus led to the development of semi-dwarf varieties with increased yield, suitable for intensive agriculture (Wang *et al*., 2014). Another selected locus is *TaGW2*, involved in regulating grain weight in wheat (Zhang *et al*., 2018). Recently, a locus on chromosome 3A – *Ta-GS5-3A* – was discovered to be under historical selection pressure in Chinese wheat, causing increments in grain size (Wang *et al*., 2016).

Significant progress in GY was discovered in Norwegian spring wheat over the last five decades due to introduction of new varieties, but little is known about its the genetic basis. The Norwegian growing environment is also distinct, and the genetic basis for genotype adaptation is largely unknown. To fill these gaps, the aims of this study were: i) to detect genomic regions associated with grain yield, plant height, days to heading and days to maturity; ii) to discover genomic regions associated with genotype adaptation under Norwegian growing conditions; and iii) to explore genetic explanations for the historical breeding progress in grain yield and plant height in Norwegian spring wheat.

Materials and methods

Plant material

The NMBU spring wheat panel, consisting of 301 hexaploid spring wheat varieties and breeding lines, was used for the primary association study. The same panel was recently used for genetic analyses of *Fusarium* head blight (Nannuru *et al*., 2022) and Septoria nodorum blotch (Lin *et al*., 2022) resistance. The collection encompasses 186 Norwegian, 40 Swedish, and 37 lines from CIMMYT, with several additional lines from Australia, Brazil, Canada, Czech Republic, Denmark, Finland, France, Germany, Netherlands, Poland, Russia, Slovakia, South Africa, Switzerland, UK, and the USA. Varieties from Norway or Sweden are adapted to the local growing conditions, while the remaining lines form the "exotic" (not adapted) part of the panel. The whole set encompasses historically significant and current varieties, covering the highlights of the last decades in wheat breeding in the Nordics and worldwide and representing a broad genetic and phenotypic diversity. This collection also contains historical varieties on the Norwegian market, described in Mróz, Dieseth, and Lillemo (2022). This collection is referred hereafter to as the main panel.

An independent set of 889 current breeding lines was used for QTL validation. This collection originates from the commercial spring wheat breeding program of Graminor (Ridabu, Norway) and is hereafter referred to as the validation panels. Not every genotype was tested in each season/location combination due to different genotype content and number in every field season (ranging from 90 to 397 lines, detailed overview in Table S3). The validation panels are considered adapted to the Norwegian growing conditions due to their origin.

Field trials

Field trials were carried out for the main panel during field seasons 2015-2021 in Vollebekk Research Station (Norway, Ås, $59^0$39'N, $10^0$45'E) and Staur farm (Norway, Stange, $60^0$43'N, $11^0$06'E). Those locations represent Norway's two main and economically important wheat-growing areas: the somewhat warmer and milder climate of south-eastern Norway and the slightly colder and temperate climate of inland Norway, respectively.

The trials were fertilized at sowing with 120 kg ha^{-1} of compound NPK fertilizer (YaraMila 22-3-10) and planted each season on the break of April and May in both locations (exact planting dates in Table S1). Following germination, trials were kept disease and weed free according to local management practices using herbicides (Tripali [active ingredients: florasulam + metsulfuron-methyl + tribenuron-methyl] and Duplosan Meko [mekoprop]) and fungicides (Proline [prothioconazole], Aviator Xpro [bixafen + prothioconazole], Forbel [fenpropimorph] and/or Comet Pro [pyraklostrobin]) in doses tailored to the needs. Irrigation was applied in case of drought that could affect the growth of the plants. Alleys within the trials were created by spraying Glyphosate shortly after seedling emergence. The trials were harvested each season towards the end of August after all varieties had reached full ripeness.

Season 2018 in both locations was marked by very little rainfall and high temperatures during the early growth stages of the plants (almost no rain from May to mid-June, Fig. S1, S2, Tables S4, S5), which, despite irrigation efforts, caused severe damage to the trials. This damage reduced grain yields by nearly 60% and caused many plant agronomical characteristics to be abnormal (data not shown). Therefore, we excluded the 2018 field season at both locations from the analysis.

Field trials of the validation panels were carried out following the same procedures as for the main panel in field seasons 2019-2022 at Staur and Vollebekk locations.

Field trial design

The trials were designed as an alpha-lattice with two replicates per genotype, a block size of 6, and positions of every accession randomized each year. Each column was planted with buffer variety at its start and end to eliminate border effects. Each field trial plot was 5 by 1.5 m in size at harvest, with gaps between the plots of 30 cm and a central alley of 1 m. Not every variety was tested for the main panel in each season/location. The number of genotypes tested varied from 100 to 295 per season/location, with 301 and 296 unique accessions in Vollebekk and Staur, respectively (Table S2).

Phenotypic data

The collection was phenotyped for days to heading (DH), days to maturity (DM), grain yield (GY), and plant height (PH). Not every trait was phenotyped in every environment (season/location combination) (Table 1).

Table 1 *Overview of when the phenotyped traits were captured in each location for the main panel. * Days since sowing*

Trait	Abb.	Unit	Vollebekk	Staur
Days to maturity	DM	dss*	2015, 16, 17, 19, 20, 21	2016, 17, 19, 20
Grain yield	GY	$g \cdot m^{-2}$	2015, 16, 17, 19, 20, 21	2016, 17, 19, 20
Days to heading	DH	dss*	2015, 16, 17, 19, 20, 21	2017, 19, 20
Plant height	PH	cm	2015, 16, 17, 19, 20, 21	2019, 20

DH and DM were assessed by recording when approximately 50% of the plants in an experimental plot had reached the respective stage. GY was measured by harvesting and threshing the trial plots, drying the yield until approximately 13.5% moisture, weighing it, and recalculating it to g m^{-2}. PH was assessed by measuring the distance between the ground and the top of spikes (excluding awns, if present) for a random tiller sample when plants reached their final height.

Data for plots that lodged early in the season was removed due to the heavy impact on their development. If lodging occurred later in the season (close to physiological maturity), data were double-checked for consistency and possible impact on the traits and judged if it should be included in the dataset.

Statistical analysis of the field trials

For each trait, three types of genotypic means (lsmeans) were calculated: location/season (field trial) means, location mean (all seasons from one location), and a global mean, where all the locations and seasons were combined.

As it was common to observe extra spatial variability within the trials (due to soil gradients) that could not have been captured by blocking, an additional covariate was introduced (columns) into the models to correct it.

The lsmeans were calculated using packages "lme4" and "lmerTEST" and custom scripts in R, version 4.2.1.

Field trial (environment) lsmeans were calculated using the mixed model (1):

$$P_{ilmn} = \mu + g_i + R_l + R{:}B_{lm} + C_n + e_{ilmn} \quad (1)$$

Cross-season lsmeans for each location were calculated using the mixed model (2):

$$P_{iklmn} = \mu + g_i + Y_k + Y{:}R_{kl} + Y{:}R{:}B_{klm} + Y{:}C_{kn} + e_{iklmn} \quad (2)$$

Global means (cross-season, cross-location) were calculated using the mixed model (3):

$$P_{ijklmn} = \mu + g_i + L_j + L{:}Y_{jk} + L{:}Y{:}R_{jkl} + L{:}Y{:}R{:}B_{jklm} + L{:}Y{:}C_{jkn} + e_{ijklmn} \quad (3)$$

Where P_{ijklmn} is the phenotype (trait) value for genotype g_i in location L_j in season Y_k planted within replicate R_l, block B_m, and column C_n. Small letters denote fixed effects, capitalized letters denote random effects and ":" denotes nesting of effects. μ is the general mean and e denotes the error, $IID(0, \sigma_e^2)$.

Field trial (season/location), location, and an overall means (across all field trials) are hereafter referred to as environment, location, and global means, respectively.

For the validation panel, only environment means were calculated due to varying genotype content in each environment.

Broad-sense heritability (H^2) was used to assess data quality (replicability), calculated for individual trials using equation (4):

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_e^2} \quad (4)$$

where σ_g^2 is the genotypic variance and σ_e^2 is the error variance.

Variance components for equation (4) were estimated using package "lme4" using a fully random model (5):

$$P_i = G_i + e_i \quad (5)$$

where P_i is the phenotype (trait) value of genotype G_i and e_i is the error term, $IID(0, \sigma_e^2)$.

Data visualization was performed in R using packages "ggplot2", "ggpubr" and "ggsci".

Genotyping data

Samples were prepared and genotyped with the TraitGenetics 25K SNP chip as described in Nannuru *et al*. (2022).

The physical positions of the markers were determined using the chip's documentation, and markers not mapped to any physical chromosome position were placed on a fictional chromosome Un.

Markers were filtered, leaving only the ones with less than 10% missing data and minor allele frequency (MAF) larger than 0.05. Heterozygous markers were treated as missing data. After the quality check, the dataset contained 19874 high-quality markers mapped to sub-genomes A (7999), B (7905), and D (2111) on chromosomes 1A (1156), 1B (1147), 1D (391), 2A (1232), 2B (1377), 2D (437) 3A (1074), 3B (1336), 3D (256), 4A (699), 4B (602), 4D (111), 5A (1340), 5B (1406), 5D (311), 6A (1126), 6B (1082), 6D (319), 7A (1372), 7B (955), 7D (285) and Un (1859).

Population structure, linkage disequilibrium, and GWAS

The main panel exhibits a strong population structure due to the presence of "adapted" and "exotic" groups of lines; therefore, additional correction for population structure was applied by including principal genomic components in the model. Due to their poor adaptation, the "exotic" lines often exhibit unusual phenotypes under Norwegian growing conditions. To mitigate the risk of confounding SNPs with line adaptation and to discover possible sources of adaptation, two series of GWA studies were carried out for each phenotype: on the whole collection and adapted lines only. A detailed description of the population structure present in the main panel can be found in Nannuru *et al*. (2022).

Genome-wide association studies (GWAS) and LD (linkage disequilibrium) analysis were performed using GAPIT v3.2 (Wang & Zhang, 2021) in R version 4.2.1. GWAS was performed on a series of phenotypes for each trait: all environment means,

location means, and global mean, computed as described in the section Statistical analysis of the field trials.

Detecting peak markers using the FarmCPU method

Several models' performance was considered, including CMLM (Zhang *et al*., 2010), MLMM (Segura *et al*., 2012), SUPER (Wang *et al*., 2014), GBLUP (Zhang *et al*., 2007), and FarmCPU (Liu *et al*., 2016).

The FarmCPU method was chosen to detect peak markers based on superior accordance with the null hypothesis and stronger signal compared to the other methods and its ability to "distillate" markers in a given significant locus by providing fewer MTAs (marker-trait associations), but with a stronger signal. FarmCPU is a multi-locus model based on the MLM method, which relies on iterative and alternative use of fixed and random effect models to minimize the proportion of false positives. Markers are tested one by one using random models. Then the resulting significant associations are used as covariates in a fixed model (random models allow for avoidance of the overfitting issue present with fixed effect models). FarmCPU is well-suited for highly quantitative trait analysis (Liu *et al*., 2016).

MTAs for each trait were considered based on Bonferroni – corrected p-values (effective threshold of $-\log_{10}(p) = 5.6$). MTAs that crossed the threshold at least for the global mean, and two other environments were reported. All the studied traits are highly quantitative and controlled by many small effect loci; therefore, the Bonferroni correction of p-values can be too conservative (Haikka *et al*., 2020). Therefore, a less stringent criterion of $\alpha = 0.001$ was also applied: a region was considered meaningful if two or more SNPs within 5 Mbp distance appeared significant for the global mean and at least two different environments/means.

Expanding QTL regions around peak markers

Consistent peak markers discovered using the FarmCPU method (section above) were used to anchor possible haplotypes. A window of 40 Mbp (chosen based on linkage disequilibrium, Fig. S3, section Genome-wide association study and linkage disequilibrium) around a peak marker's position was studied using the MLM method (also implemented in GAPIT v3.2). Markers were considered based on appearing significant ($-\log_{10}(p) > 3$) for the global mean and at least two environments/means. Peak markers alongside significant markers in the window were used to construct haplotypes. The search for MTAs with the MLM method was carried out in the adapted and complete datasets, similarly to the FarmCPU method.

Individual markers and QTL haplotypes were tested for their associations with the respective traits on the global mean without population structure correction in both the full main panel and its adapted part. Comparison between marker alleles and among haplotype alleles was performed using Tukey's Honestly Significant Difference (HSD) posthoc test, $\alpha = 0.05$. Rare haplotype alleles appearing in less than ten accessions were discarded from the analysis due to insufficient statistical power to detect their associations. The proportion of phenotype variance explained by each SNP and each putative QTL was estimated using linear models and reported separately as a percentage in the whole panel and its adapted part.

Allele frequency over time

Varieties in the adapted part of the main panel were assigned a breeding line year of creation by analyzing their documentation. For varieties for which it was only possible to establish the year they were released to the market, seven years were subtracted to obtain the year of creation (based on the average time it took from variety creation and release to the market in the collection). It was possible to establish the year of creation (YOC) for 180 lines in the adapted part of the main panel, assigned to the following seven periods: pre-1960 (2 lines), 1960-1969 (4), 1970-1985 (6), 1986-1995 (9), 1996-2005 (24), 2006-2010 (41), and 2011 onwards (94). Allele frequency for each discovered haplotype and MTA in GWAS was calculated for each period and analyzed for trends over the years.

Effect validation

Effects of the significant associations discovered in the main panel (full panel and its adapted part) were tested using an independent set of varieties and breeding lines originating from Graminor AS (Ridabu, Norway) spring wheat breeding program (section Plant Material).

Field trial data were analyzed using mixed models as described in the section "Statistical analysis of the field trials"; however, without calculating cross-environment means due to different genotype composition each year. Associations of markers and haplotypes were tested against each season's genotypic means without correcting for population structure using Tukey's Honestly Significant Difference (HSD) posthoc test, $\alpha = 0.05$. Rare haplotype alleles (appearing in less than ten accessions) were discarded from the analysis.

Results

Phenotype data

Differences between the two field trial locations are visible for every studied trait. Lines in the main panel grown in Staur, on average, tend to head nine days faster (DH), have vegetative period shorter by 2.5 days (DM), have higher grain yield by 90 g m^{-2} (GY) and reach lower plant height by 4.5 cm (PH) compared to Vollebekk. The Staur trials exhibit higher variability in DH, DM, and GY (Table 2).

The most substantial difference between the exotic and adapted lines of the main panel is seen for GY, with a severe reduction of GY in non-adapted lines accounting for 64% of the total phenotypic variance in yield. Adapted lines also tend to head earlier, mature earlier, and be shorter, but with the grouping explaining much less of the phenotypic variance (Table 2, Figure 1).

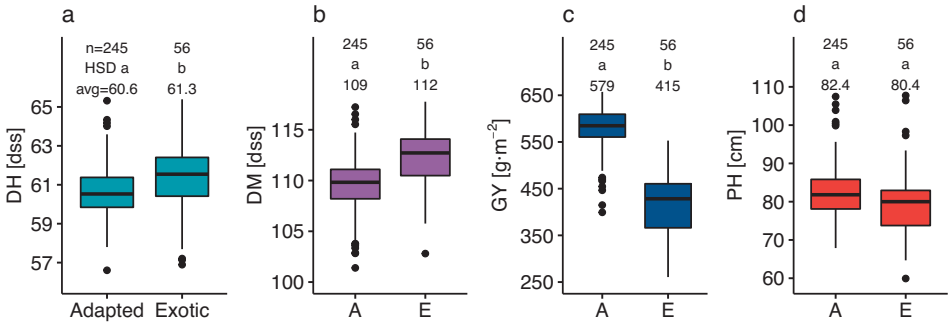


Figure 1 *Phenotypic differences among the genotypes due to line adaptation in (a) days to heading, (b) days to maturity, (c) grain yield, and (d) plant height. Groups with the same letter are not significantly different (HSD test, α = 0.05). A – adapted, E – exotic lines; n – number of records in the group, avg – average phenotype value in the group*

Traits analyzed in this study exhibited high heritability (H^2 > 0.63). DH showed the highest heritability (on average, 0.84), followed by DM, GY, and PH (0.63). PH achieved the highest variability in H^2 (0.39-0.9) compared to other traits (differences around 0.2) in different environments. Heritability estimates from Vollebekk and Staur experiment sites were comparable (0.73 and 0.69, respectively) except for PH (0.78 and 0.47 from Vollebekk and Staur, respectively) (Table 3).

Table 2 *Descriptive statistics and variance explained by cultivar adaptation for days to heading (DH), days to maturity (DM), grain yield (GY), and plant height (PH) in all the studied environments (Env) and means. Avg – mean value, SD – standard deviation, VarAD – fraction of variance explained by cultivar adaptation (in percent). For single field trials (environments): letter designates location (V – Vollebekk, S – Staur) and number denotes season (year). Vmean – cross-season trait means in location Vollebekk, Smean – cross-season trait means in location Staur, Gmean – cross-season, cross-location Global means.*

Env/ mean	DH, dss			DM, dss			GY, gm^{-2}			PH, cm		
	Avg	*SD*	Var_{AD}	*Avg*	*SD*	Var_{AD}	*Avg*	*SD*	Var_{AD}	*Avg*	*SD*	Var_{AD}
V2015	70.7	1.4	4.3	126.4	3.0	12.6	426	91	53.7	81.7	9.9	9.4
V2016	62.0	1.6	2.7	109.8	3.8	4.7	387	77	53.4	73.1	8.9	1.2
S2016	-	-	-	100.9	4.1	18.1	497	114	30.0	-	-	-
V2017	59.2	1.7	6.4	99.3	2.0	0.3	440	75	53.5	78.0	8.6	3.6
S2017	58.0	2.1	29.6	120.8	5.9	18.9	789	118	54.3	-	-	-
V2019	68.0	1.3	0.8	108.5	1.6	1.1	601	55	28.6	96.0	8.7	0.3
S2019	47.9	2.5	14.9	99.8	6.0	18.0	530	65	17.2	86.9	5.2	0.1
V2020	66.5	1.4	0.2	117.8	3.3	5.6	649	82	40.1	83.8	6.7	0.4
S2020	61.6	1.4	2.2	112.3	1.4	0.4	592	81	56.7	70.2	5.3	0.1
V2021	64.4	1.7	0.0	104.5	2.9	11.4	540	69	46.3	86.1	7.8	0.8
Vmean	65.1	1.4	0.5	111.1	2.3	10.6	509	72	63.7	83.0	7.7	1.7
Smean	55.9	1.7	16.7	108.6	3.7	13.8	600	97	56.3	78.5	5.0	0.1
Gmean	60.8	1.4	3.7	110.1	2.7	13.6	548	81	64.2	81.9	7.1	1.3

Table 3 *Broad-sense heritability estimates for each trait and each field trial (environment). DM – days to maturity, GY – grain yield, GPC – grain protein content, DH – days to heading, PH – plant height*

Trait	Vollebekk							Staur					Avg
	2015	*-16*	*-17*	*-19*	*-20*	*-21*	*Avg*	*2016*	*-17*	*-19*	*-20*	*Avg*	
DH	0.86	0.67	0.84	0.81	0.79	0.89	0.81	-	0.88	0.92	0.77	0.86	0.84
DM	0.77	0.62	0.51	0.69	0.66	0.75	0.67	0.73	0.58	0.61	0.42	0.59	0.63
GY	0.85	0.61	0.49	0.68	0.57	0.73	0.66	0.61	0.69	0.51	0.53	0.59	0.63
PH	0.85	0.83	0.66	0.90	0.59	0.86	0.78	-	-	0.55	0.39	0.47	0.63

Trial means for all traits were highly positively correlated ($r > 0.5$) across all-season/location combinations, except for seasons 2015 in Vollebekk (lower correlations for GY and DH) (Figures S4-7).

PCA of phenotypic data separates adapted from the exotic part of the panel. The exotic varieties form a spread-out cloud and have lower GY and higher PH than the adapted part, while differences in DH and DM are not strongly pronounced (Figure 2).

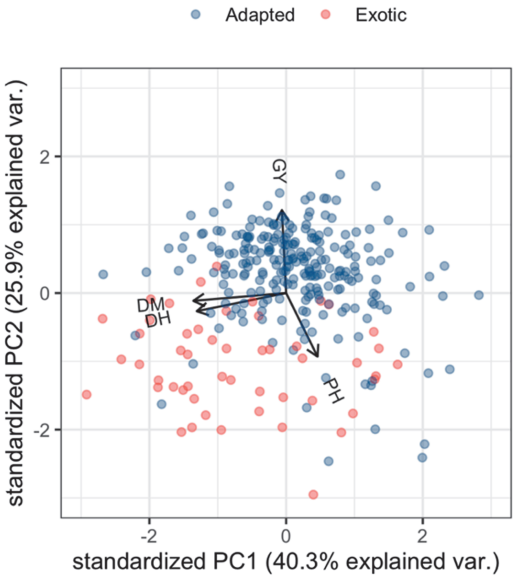


Figure 2 *Principal Component Analysis (PCA) of the main panel based on phenotypical data. Color indicates line adaptation: blue – adapted, red – exotic (not adapted). GY – grain yield, DM – days to maturity, DH – days to heading, PH – plant height*

The strong population structure could be seen in the lack of a significant relationship between DM and GY for the whole panel (Figure 3a) and its presence within both the adapted and exotic parts of the collection (Figure 3bc). PH showed a weak negative correlation with GY for adapted lines (Figure 3b), not observed in the entire panel or its exotic part.

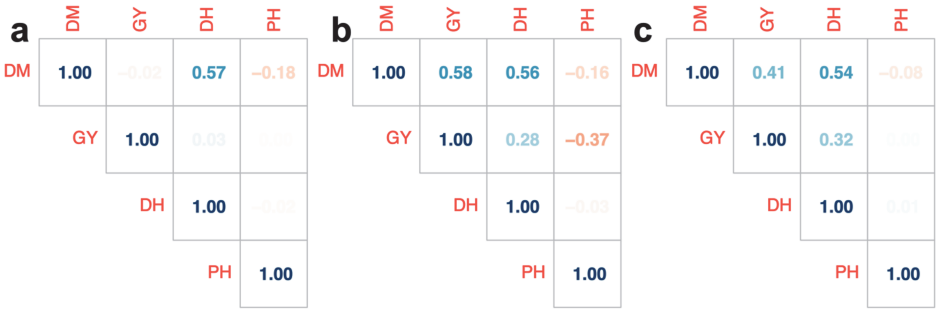


Figure 3 *Pearson's correlation coefficient matrices for genotypic means of (a) all accessions, (b) adapted accessions and (c) exotic accessions. DM – days to maturity, GY – grain yield, DH – days to heading, PH – plant height*

Accessions for which it was possible to establish the year of creation (YOC, 180 lines) were assigned to the following seven periods: pre-1960 (2 lines), 1960-1969 (4), 1970-1985 (6), 1986-1995 (9), 1996-2005 (24), 2006-2010 (41), and 2011 onwards (94). Significant trends over the years can be seen for all traits except days to heading. Lines belonging to the time periods from 1996-2005 onwards mature later than lines created in the 1960s, and old accessions (until 1970) are significantly taller than later lines. For GY, there has been consecutive increases with the lines created during 1971-1995 yielding significantly higher than older ones, but less than the most recent lines created after 1995 (Fig. 4).

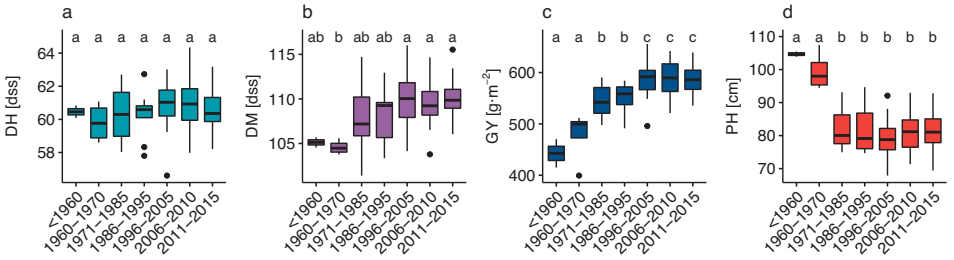


Figure 4 *Changes in the traits over the seven periods (x axis): days to heading (a, DH, blue), days to maturity (b, DM, violet), grain yield (c, GY, dark blue), and plant height (d, PH, red). Periods with the same letter are not significantly different (HSD test, α = 0.05)*

GWAS results

GWAS analyses were conducted for each trait for the whole panel and its adapted parts. QQ and Manhattan plots are available on Figures S8-15 and S16-23, respectively.

A total of thirteen consistent and highly significant MTAs were detected using the FarmCPU method, pointing to twelve QTL regions across all sub-genomes (three in A subgenome, five in B, and two in D). Two QTL were discovered for DH on chromosomes 1B and 7B, two for DM on chromosomes 6B and 6D, three for GY on chromosomes 3A, 5A, and 7B, and five for PH on chromosomes 2A, 4A, 4B, 4D, and 6B. All these regions met the stringent Bonferroni threshold for the most significant marker, except for the plant height QTL on 4D (*QHt.nmbu-4B*), where the peak markers were discovered using the criterion of $p<0.001$ across the global mean and two other environments and two MTAs within a 5 Mb window. The number of significant SNPs (detected using FarmCPU and MLM methods) in each QTL varied from 1 to 18, and no QTL was associated with more than one trait (Table 4, S6).

Table 4 *QTL regions discovered in the study for days to heading (DH), days to maturity (DM), grain yield (GY), and plant height (PH), their genomic locations (Chr, Span in Mbp), number of significant markers in the locus (n), peak marker(s), QTL effects in the full dataset (E_F) and in the adapted part (E_A) in respective units listed in Table 1, trait variance explained in the full dataset (%PVE_F) and in the adapted part (%PVE_A), and summary of validation results (Val) corresponding to Table S7.*

Trait	QTL/Chr	Span	n	Peak marker(s)	[1]E_F	[1]%PVE_F	[2]E_A	[2]%PVE_A	[3]Val
DH	*QHd.nmbu-1B*	1-2	2	BS00022180_51	2.04	14.1	Mono	Mono	Rare
	QHd.nmbu-7B	606	1	BobWhite_c3541_152	0.91	11.4	0.91	11.4	Rare
DM	*QMat.nmbu-6B*	132-136	10	Kukri_rep_c71420_511	5.93	25.8	4.85	24.8	+
	QMat.nmbu-6D	6	1	BS00022523_51	0.82	8.80	0.81	10.0	+
GY	*QYld.nmbu-3A*	267	1	BS00110129_51	38.1	22.0	16.4	14.3	+
	QYld.nmbu-5A	683-708	8	BobWhite_c8266_227	192.6	67.2	Ns	Ns	+
	QYld.nmbu-7B	701-703	6	BS00083578_51	106.2	16.3	83.6	30.9	+
PH	*QHt.nmbu-2A*	524-543	5	AX-95095516	7.91	9.21	7.97	12.6	Rare
	QHt.nmbu-4A	570-603	3	CAP11_c3631_75	9.77	9.36	5.45	9.08	+
	QHt.nmbu-4B	13-59	18	TG0010a; TG0010b	8.20	27.7	7.22	23.9	+
	QHt.nmbu-4D	19-26	3	BobWhite_s64797_152	8.29	32.5	9.32	42.9	+
	QHt.nmbu-6B	202	1	Ra_c10469_616	2.00	5.21	1.25	2.14	+

[1]Full difference between homozygotes without correcting for population structure using a linear model in the full dataset. [2]Full difference between homozygotes without correcting for population structure using a linear model in the adapted dataset.[3]Validation of effects without correcting for population structure in sets of independent lines in a total of six environments using a linear model. + - at least one SNP in the region showed significant effect (HSD test, α=0.05) in one of the validation sets. Rare – all markers had MAF < 0.05 (estimation not reliable), Mono – marker monomorphic (estimation not possible).

Two QTL regions were detected for DH on chromosomes 1B and 7B, explaining 14 and 11% of the variance and with two and one-day effects, respectively. Both consisted of a few SNPs (two and one, respectively). *QHd.nmbu-1B* was strongly associated with population structure, as it showed a significant effect only in the whole dataset due to a lack of haplotype diversity in the adapted part. *QHd.nmbu-7B* consistently showed the same effect and proportion of variance explained in both datasets (Table 4, Table S6), indicating a lack of association with line adaptation. *QHd.nmbu-1B* and *QHd.nmbu-7B* had low MAF in the validation sets, making the estimation of their effects unreliable (Table S7).

For DM, two QTL regions were detected on chromosomes 6B and 6D, consistently associated in the complete and adapted datasets. *QMat.nmbu-6B* consisted of 10 markers, while *QMat.nmbu-6D* included only one marker, which was discovered using the FarmCPU method. *QMat.nmbu-6B* showed a considerable effect of around five days, explaining approximately 25% of the variance in both datasets. Similarly, *QMat.nmbu-6D* was detected consistently across the datasets, explaining 9 to 10% of the variance, with a minor effect of 0.8 days. No adaptation-specific QTL was

discovered for DM (Table 4, Table S6). Validation of both *QMat.nmbu-6B* and *QMat.nmbu-6D* was successful, with multiple MTAs being confirmed for *QMat.nmbu-6B* and the one MTA comprising *QMat.nmbu-6D* (BS00022523_51) (Table S7).

Three QTL regions were detected for GY with effects (without correcting for population structure) ranging from 38 to 193 $g \cdot m^{-2}$ and explaining between 16 and 67% of the variance in GY in the complete set. *QYld.nmbu.5A* is purely associated with line adaptation, showing a very high effect and proportion of variance explained in the entire dataset (193 $g \cdot m^{-2}$and 67%, respectively) while exhibiting no significant effect in the adapted set. Also, all the MTAs of *QYld.nmbu.5A* were discovered in the entire dataset. However, the remaining two QTL (*QYld.nmbu.3A* and *QYld.nmbu.7B*) were consistently associated with GY in both datasets, with more negligible effects (Table 4, Table S6). Validation confirmed significant associations of at least one marker in each QTL region associated with GY (1 SNP for *QYld.nmbu.3A*, 3 for *QYld.nmbu.5A*, and 2 for *QYld.nmbu.7B*) (Table S7).

The highest number of QTL regions (5) was detected for PH, on chromosomes 2A, 4A, 4B, 4D, and 6B. All the QTL regions for PH consisted of more than 3 SNPs except for *QHt.nmbu-6B* (only one peak marker, discovered using the FarmCPU method). The QTL had effects ranging from 2 to 10 cm and explained 5 to 33% of the variance in PH in the entire dataset. None of the QTL appeared to be adaptation-specific, as all showed significant and comparable effects in both adapted and complete datasets. As the presence of exotic lines increases the variance in PH considerably (Figure 1), it is remarkable that *QHt.nmbu-4D*'s proportion of variance explained and effect are higher in the adapted than in the entire dataset. (Table 4). Validation confirmed the associations of *QHt.nmbu-4A*, *QHt.nmbu-4B*, *QHt.nmbu-4D*, and *QHt.nmbu-6B* with high confidence (multiple markers in the loci appeared significantly associated in multiple validation sets in multiple environments). *QHt.nmbu-2A* could not be validated due to residual minor allele frequencies of the SNPs comprising it in the validation sets (Table S7).

Adaptation to the Norwegian growing conditions

Two QTL regions associated with DH and GY (*QHd.nmbu-1B* and *QYld.nmbu-5A*) consistently appeared highly significant for their respective traits in the entire dataset while showing no polymorphism/effect in the adapted part of the dataset (Table 4, Figure 5). It was, therefore, reasonable to consider these QTL as pointers to genomic regions associated with genotype adaptation to the Norwegian growing conditions.

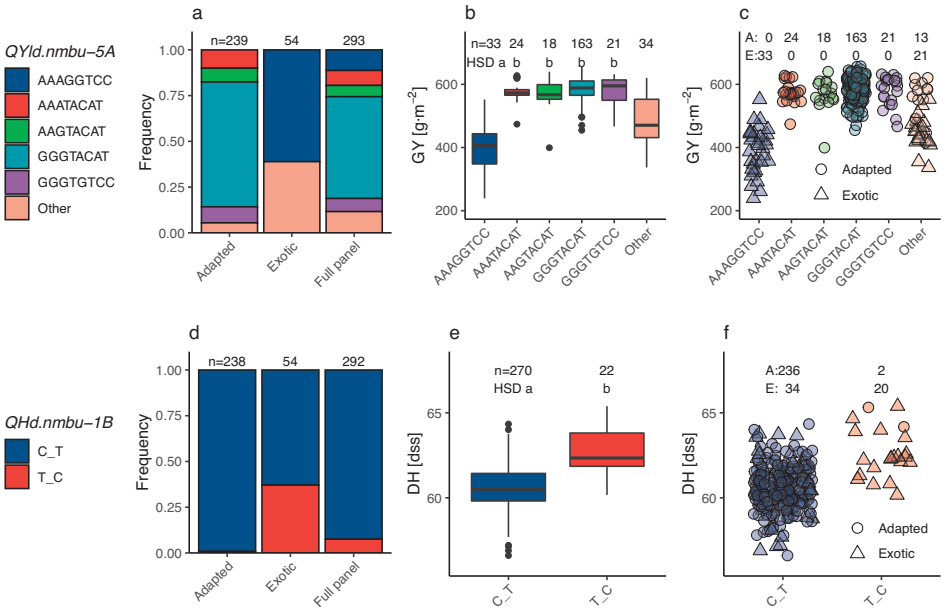


Figure 5 *Loci associated with adaptation to the Norwegian growing conditions: QYld.nmbu-5A (abc) and QHd.nmbu-1B (def). Allele frequencies in the full panel and adapted and exotic parts (ad), haplotype analysis of the loci (be), and presence of alleles in adapted and exotic parts of the panel (cf). Comparison among the alleles was performed using Tukey's HSD test. Alleles with the same letter are not significantly different (α = 0.05). Only alleles present in more than ten lines were considered for QHd.nmbu-1B. For QYld.nmbu-5A, alleles with low frequencies (in less than ten lines) were gathered into the "Other" bin. Association of QYld.nmbu-5A haplotypes with DH is shown on Fig. S24*

The *QYld.nmbu-5A* (683-708 Mbp) region (Table 4) showed an extremely high effect of 192 g m^{-2} (almost 2 t ha^{-1}) and captured 67% of GY variance in the entire dataset. The high proportion of variance explained by this QTL makes it likely that *QYld.nmbu-5A* is the leading cause for the observed differences between the adapted and exotic lines. Interestingly, even though *QYld.nmbu-5A* showed no significant effect in the main panel's adapted part, three SNPs belonging to it showed significant association with GY in the validation panels; however, with effects not nearly as high as in the entire main panel (Table S7). *QYld.nmbu-5A* is effectively represented by five haplotypes and several underrepresented variants. The allele associated with lower GY (AAAGGTCC) occurs exclusively in exotic lines, while high-GY alleles are present in adapted lines only. The rare alleles are present in 13 and 21 adapted and exotic

lines, respectively. Exotic lines carrying one of the underrepresented alleles also exhibit lower GY (Fig. 5abc). An analogous haplotype analysis of *QYld.nmbu-5A* revealed its significant association also with DH, which could not be detected by GWAS (Fig. S24).

QHd.nmbu-1B (1-2 Mbp), capturing 14% of the variance in DH in the entire dataset (Table 4), is a likely candidate to contribute to line adaptation concerning DH. However, the average proportion of DH variance explained by line adaptation is lower (4%, Table 2). The early haplotype (C_T) is present in almost all adapted lines, and in more than half of the exotic accessions. The late haplotype (T_C) occurs in only two adapted lines, and twenty exotic lines (Fig. 5def). *QHd.nmbu-1B* also exhibited almost no polymorphism in the validation sets.

Breeding progress in Norwegian spring wheat

Three of the detected QTL showed noticeable change in allele frequency over the periods: *QYld.nmbu-7B*, *QYld.nmbu-3A*, and *QHt.nmbu-2A* (Fig. 6).

QYld.nmbu-7B was consistently discovered in both adapted and complete datasets and captured nearly 31% of the variance in GY in the adapted lines (Table 2, Table S6) This QTL is represented in the adapted lines by four haplotypes: three with similar, positive effects, and one allele with strong negative effect (TCCT) explained mostly by SNP variation in the first marker (BS00083578_51) (Fig. 6c). The negative effect allele is present in all old accessions (before 1970). After 1971, its frequency decays gradually to zero in the 2006-2015 bracket, replaced by either positive effect allele. Two outliers carrying the negative effect allele are visible with high GY values (comparable to lines carrying one of the favorable effect alleles) (Fig. 6c).

QYld.nmbu-7B frequencies align well with the observed increase in GY over the periods. The diversity in the validation panels in the *QYld.nmbu-7B* locus was generally low. The new breeding lines were dominated by favorable effect alleles, limiting the prospects of exploiting this locus for further GY improvement. However, *QYld.nmbu-7B* still showed a significant association with GY in the validation panels (Table S7).

Another locus showing allele frequency change is QYld.nmbu-3A; however, not so radically as QYld.nmbu-7B. QYld.nmbu-3A comprises a single SNP, consistently detected using FarmCPU and MLM methods (Table S6) and explained 14 and 22% of GY variance in adapted and complete datasets, respectively (Table 4). This locus is polymorphic in all periods (except the oldest lines) and shows a decay in negative allele (G) frequency over time; however, it still retains a degree of polymorphism in

the most recent breeding lines (Fig. 6def). Although this region comprised only a single MTA, it was still significantly associated with GY in the validation sets. It is a potential GY improvement source in future breeding (Table S7).

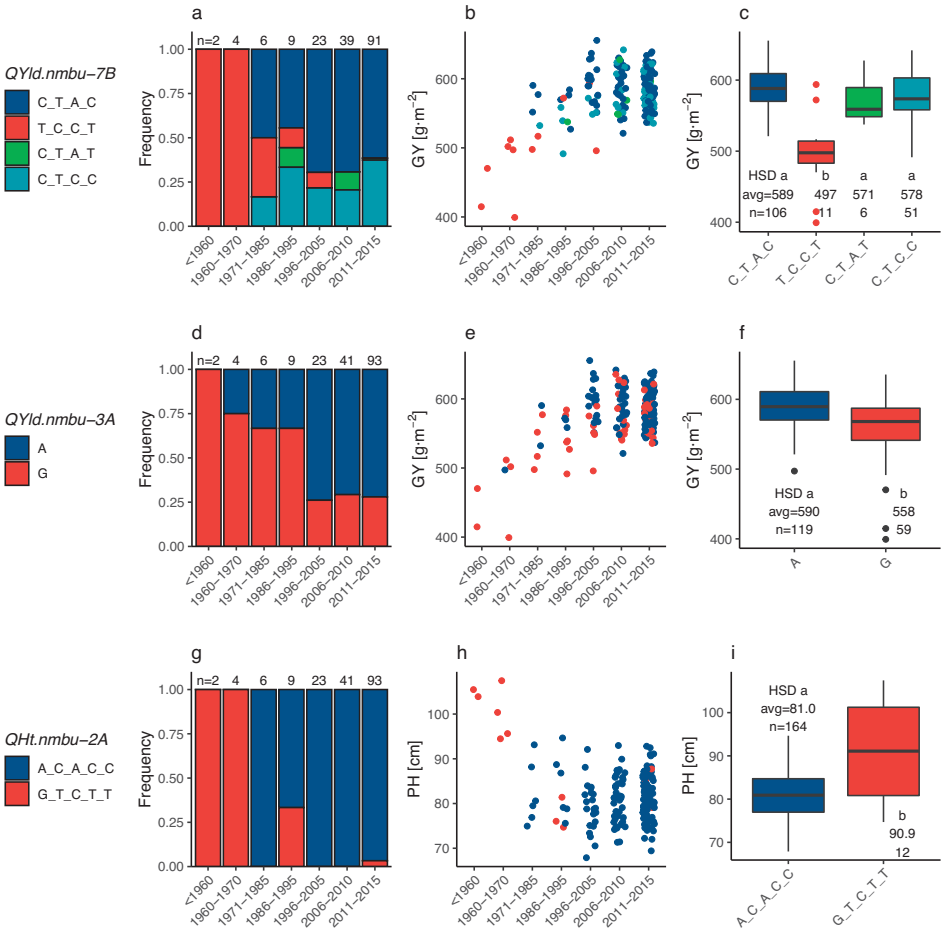


Figure 6 *Columns, from left to right: allele frequency over time, relationships between QTL alleles and their respective trait across the periods, and comparison among QTL alleles and the respective trait for QYld.nmbu-7B (abc), QYld.nmbu-3A (def), and QHt.nmbu-2A (ghi). Rare alleles occurring in less than 5 lines were removed from the analysis. Comparisons among QTL alleles were performed using HSD (Honestly Significant Difference) test. Alleles with the same letter are not significantly different (α = 0.05)*

Allele frequency change over the periods can also be observed for *QHt.nmbu-2A*, represented in the adapted lines effectively by only two alleles, despite five significant

MTAs comprising this locus. The lines dating before 1970 all carry the tall allele (GTCTT), while the later accessions mostly carry the shortening allele (ACACC). However, the tall allele could still be found in 3 lines created between 1986-1995 and two most recent breeding lines (Fig. 6ghi). The locus also showed almost no diversity in the validation panels (recent breeding lines) (Table S7).

By investigating the alle frequencies of SNPs in the validation panels (representing the most recent germplasm in Norway), it can be observed that SNP alleles associated with increased GY and higher DM are dominating in the population (Table S7).

Discussion

Crossing adapted germplasm with exotic parents is an essential source of variation and valuable alleles (Reynolds *et al.*, 2009) and has been actively used in Norwegian spring wheat breeding by introducing mainly CIMMYT lines into breeding programs, yielding many market-important varieties (Lillemo & Dieseth, 2011). Therefore, analyzing only lines adapted to the distinct Nordic environments would appear incomplete for this study. The exotic lines pose a statistical challenge, as those often exhibit extreme phenotypes paired with distinct genetic backgrounds. However, the statistical model used – FarmCPU – did effectively account for the population structure, as judged by the QQ plots. Although the population structure can be perceived as a shortcoming of the association panel it allows us to explore the genetic basis of line adaptation to the Nordic growing environment.

The number of discovered regions associated with the traits is relatively modest due to the stringent significance criterion applied. However, considering the presence of strong structure (line adaptation) and different backgrounds of lines in the studied population, it was necessary to reduce the risk of committing type 1 errors at the cost of a higher number of false negatives. The purpose of this study was to pinpoint the most important loci for each of the traits. Despite their highly quantitative nature, it was possible to discover several large-effect QTL, explaining the most important genetic variability in the traits in the panel. Apart from the two loci associated with adaptation, ten QTL regions were consistently discovered in the entire panel and its adapted part.

Line adaptation to the Norwegian growing conditions is most visible in achieved GY: more than half of the variance in GY can be attributed to line adaptation, with much smaller proportions for the other investigated traits. By comparing significant loci for the entire panel and its adapted part alone, two loci were highly significant in the

whole panel, with no effect in its adapted part, hinting that those regions may either play a role in adaptation to the Nordic growing environment, be a remnant of line origin, or be random (spurious). *QYld.nmbu-5A*'s polymorphism explains a significant part of the variance in GY (similar to the variance explained by line adaptation), is comprised of multiple SNPs, and coincides with the chromosomal location of the *Vernalization1* locus *Vrn-A1* . Vernalization response is a well-described, crucial component of genotype adaptation to particular growing conditions, as reviewed by Hyles *et al.* (2020). Lines of diverse origins will carry different alleles in this locus, favorable in their environments of origin. Considering the causative link between *Vrn-A1* and GY and line adaptation, it is reasonable to expect differences in flowering time. Indeed, the non-adapted lines, on average, head later than the adapted ones, but the differences are not as strongly pronounced as for GY. Despite that no signal of the *Vrn-A1* locus was detected for DH in GWAS (due to the stringent threshold), *QYld.nmbu-5A* still shows a significant effect on DH, as revealed by haplotype analysis. Additionally, in another round of late-planted trials of the same panel (sowing date in late June), the heading and flowering time differences are much more pronounced (data not shown), indicating that the vernalization genes are playing a crucial role in adaptation to the Nordic growing environment. The fact that *QYld.nmbu-5A* has effect on both GY and DH is convincing that GY benefits from phenological adaptation and not from some other linked gene in the vicinity of the *Vrn-A1* locus.

The second locus associated with line adaptation, *QHd.nmbu-1B*, captures over 14% of the variance in DH in the whole set. *QHd.nmbu-1B*'s peak marker (BS00022180_51) has been previously discovered for DH in the same panel (Sørensen, 2016), was found significant for drought stress adaptation (Kamruzzaman, 2022) and conferring sensitivity to the *Parastagonospora nodurum Tox1* effector (Cockram *et al.*, 2015), however, to the best of the authors' knowledge, does not align with any previously described locus for days to heading, except for the previous study on the same panel referred to above. The fact that line adaptation in this study is mainly associated with GY and, to a smaller extent, DH, highlights the importance of phenology for adaptation to the relatively short Norwegian growing season. Other important traits not considered directly in this study but documented elsewhere include the ability to withstand lodging and pre-harvest sprouting as well as disease resistance (Lillemo & Dieseth, 2011).

Unlike most studies on historical genetic gains, this study attempted to use a mix of registered varieties and advanced breeding lines based on the creation timeline rather than the year of release. This was unavoidable due to the scarcity of registered

varieties (especially in the earlier periods) and the resulting problems with reaching a large enough sample size. Despite this shortcoming, the GY increase over time shows similarity to many collections (Sayre, Rajaram and Fischer, 1997; Abbate *et al.*, 1998; Shearman *et al.*, 2005; Voss-Fels *et al.*, 2019) and our previous research on recent Norwegian varieties (Mróz, Dieseth and Lillemo, 2022). No trend over time was apparent for DH while a slight increase in DM and a sharp decline in PH (driven mainly by the oldest, tall accessions) could be observed. By investigating the genetic pool of the most recent wheat breeding germplasm in Norway (the validation panels), it becomes clear that favorable effect alleles of GY and DM were accumulated (all favorable effect alleles of GY and DM are dominating the population). This finding corresponds to Voss-Fels *et al*. (2019), where the authors showed that varieties gradually accumulated genetic structures associated with GY, resulting in their linear increase over the years. Interestingly, in the most recent breeding lines, alleles associated with reduced PH dominate (concerning the *Rht* loci (Pearce *et al.*, 2011) represented by *QHt.nmbu-4B* and *QHt.nmbu-4D* in this work); however, still retaining a high degree of diversity in the loci.

From the breeding progress standpoint, the most exciting finding is the *QYld.nmbu-7B* locus. The region is strongly associated with GY, and a sharp decline in the frequency of the unfavorable allele was observed over the studied periods. One could argue that its association could be spurious and linked more to genotype background rather than GY itself - all the oldest lines carry one allele and have lower GY, in contrast to the recent accessions with other alleles and higher GY. However, the SNPs constituting this locus still show a degree of polymorphism in the validation sets, enough to be significantly associated with GY. In a nearby region (chromosome 7B, 674 Mbp) to *QYld.nmbu-7B* (701 Mbp), the *TaCol-5* gene was discovered and experimentally confirmed to be associated with the number of spikelet nodes per spike in wheat (Zhang *et al.*, 2022). Transgenic plants overexpressing modified dominant *TaCol-5* allele showed increased GY due to higher tiller and spike number paired with more significant spike node number. In light of the previously mentioned GY progress in Norwegian spring wheat associated with an increase in the number of grains per spike (Mróz, Dieseth, and Lillemo, 2022), it is compelling to hypothesize that this locus contributed to the breeding progress in GY in Norway; however, a more detailed genetic study of yield components and spike parameters is needed before endorsing this hypothesis. The exact origins of the favorable effect alleles in Norwegian germplasm remain largely unknown. However, an examination of pedigrees revealed that the high GY allele (CTAC) of *QYld.nmbu-7B* occurred for the first time in line T7347 (YOC 1977), which was a product of a cross between Runar

(YOC 1965, first modern Norwegian landmark variety released in 1972) and the German variety Sirius. Since Runar carries the low GY allele (TCCT), the high GY allele must therefore come from Sirius, which indeed carries the CTAT allele. The high GY haplotype (CTAC) is also found in a group of Swedish varieties (Tjalve, Dragon, Avle, and Zebra). The other favorable allele (CTCC) is found in modern Norwegian varieties with CIMMYT parentage (Bastian, Bjarne, Bajass, and Berserk), indicating that this haplotype was introduced into Norwegian wheat breeding by crossing with CIMMYT parents. Indeed, the CTCC haplotype is present in varieties with CIMMYT pedigrees like Avocet, Kukri and Gamenya as well as the landmark Brazilian variety Frontana (all present in the exotic group of this study). This locus appears to be almost "fully utilized" in the most recent lines and shows no room for future improvement with the current collection of alleles present in the germplasm.

Another locus showing signs of selection in Norwegian wheat is *QYld.nmbu-3A*, which, in contrast to *QYld.nmbu-7B*, still shows a high degree of polymorphism in the most recent lines. Despite detecting only one significant SNP in the locus, its association was consistent in the validation sets. *QYld.nmbu-3A*'s only MTA appears in a similar region to the chromosomal location of the *TaGS5-3A* gene, a locus selected during breeding in Chinese wheat associated with increased kernel size, resulting in higher GY (Wang *et al.*, 2015; Ma *et al.*, 2016). No significant breeding-related progress in kernel weight was discovered in the 1972 – 2019 period in Norwegian spring wheat (Mróz, Dieseth, and Lillemo, 2022); however, considering the relatively small effect of *QYld.nmbu-3A*, it could easily have been missed (should it had occurred due to incorporation of this locus). *QYld.nmbu-3A* could also make a promising candidate for future improvement in GY due to the still-existing polymorphism in the most recent germplasm. However, to confirm *QYld.nmbu-3A*'s link with kernel weight, a follow-up study of kernel parameters in the collection is needed. The favorable allele (A) of *QYld.nmbu-3A* first appears with "Møystad", an old Norwegian variety released in 1966. "Møystad" is a product of a cross between a sister line (Mø043-40, not present in the panel) of Norrøna, old Norwegian variety, released 1952, with the Swedish variety Kärn II (not present in the panel). Since Norrøna carries the G allele (associated with negative effect on GY), the favorable allele (A) likely comes from Kärn II. The presence of the same allele in Swedish varieties like Dragon and Avle supports this hypothesis. The allele has been further transmitted from Møystad to important Norwegian varieties like Bastian and Bjarne, and later became dominating in the Norwegian spring wheat breeding material.

The tallness of the oldest lines investigated correlates well with the *QHt.nmbu-2A* locus, which became almost monomorphic over time with regard to the short allele. The oldest line with the short allele (ACACC) is the breeding line MS273-150 (YOC 1975), which is a progeny from the cross of Møystad with the landmark semi-dwarf variety Sonora 64. As Møystad carries the tall allele (GTCTT), it is likely that Sonora 64 was the source of the short allele. The short allele is also present in line T7347, which must have inherited it from Sirius. Moreover, the short allele is also found in Swedish varieties like Tjalve, Dragon, and Avle, which all have been used as crossing parents. Therefore, there are at least three plausible sources of the short allele in the Norwegian spring wheat breeding program. Significant MTAs with PH in a similar region were previously detected (Jamil *et al.*, 2019); however, to the authors' knowledge, no major known gene is situated in this locus.

Reduced height (*Rht*) genes originating from the Japanese line Norin-10 played a crucial role in the Green Revolution by introducing semi-dwarf posture to new varieties, which became spread worldwide during consecutive breeding efforts (Borojevic and Borojevic, 2005). *Rht-B1* and *Rht-D1* loci are well visible in Norwegian spring wheat, represented by the *QHt.nmbu-4B* and *QHt.nmbu-4D* regions, respectively. These loci still maintain high polymorphism in Norwegian spring wheat, with slight domination of the tall alleles, indicating that also other, mostly unknown, genetic mechanisms contribute to the desired plant height of present-day varieties.

Due to the risks associated with wet periods at the end of the growing season in Norway (lodging and quality loss due to pre-harvest sprouting), early-maturing varieties of spring cereals are generally desired. Due to climate change, it was estimated that from the 1970s until 2005, the vegetative season in Norway was extended by approximately seven days (Nordli *et al.*, 2008). Varieties released in this period did utilize that change by extending their vegetative periods by four days on average (Mróz, Dieseth, and Lillemo, 2022). The lack of evidence of consistent changes in alle frequencies of the discovered loci associated with DM indicates that this increase occurred due to the accumulation of several smaller-effect alleles rather than by incorporating fewer, big-effect alleles. The two discovered regions associated with DM (*QMat.nmbu-6B* and *QMat.nmbu-6B*) show polymorphism and significant effects in the validation panel, with the late alleles dominating the population.

Conclusions

A detailed GWAS analysis was conducted on multi-environment field trial data of grain yield, earliness, and plant height in a diverse panel of Nordic and exotic lines. The study detected twelve loci associated with the traits (two with heading time, two with maturity time, three with grain yield, and five with plant height), later validated using independent sets of recent Norwegian breeding lines. The results indicated that adaptation to the Nordic growing conditions seen mainly in changes in grain yield and is genetically associated with a phenological response due to polymorphisms in the *Vrn-A1* locus. The study also indicated that grain yield breeding progress in Norwegian spring wheat was associated with the incorporation of the *Ta-Col5* and *Ta-GS5-3A* loci, responsible for changes in spike architecture and kernel weight; however, a detailed follow-up study on spike and kernel traits is required. The radical drop in plant height since the 1970s was associated with a locus on chromosome 2A. Knowledge of these discovered QTL regions will be useful for breeding programs targeting high-latitude spring wheat growing regions with similar growing conditions to those in Norway.

References

Abbate, P.E. *et al*. (1998). Grain Yield Increase in Recent Argentine Wheat Cultivars. *Crop Science*, 38, pp. 1203-1209.
Available at: https://doi.org/10.2135/cropsci1998.0011183X003800050015x

Borojevic, Katarina & Borojevic, Ksenija (2005). The transfer and history of "reduced height genes" (*Rht*) in wheat from Japan to Europe. *Journal of Heredity*, 96(4), pp. 455–459.
Available at: https://doi.org/10.1093/jhered/esi060

Braun, H., Atlin, G. and Payne, T. (2010). Multi-location testing as a tool to identify plant response to global climate change. climate change and crop production. *Climate change and crop production*, CABI, pp. 115–138.
Available at: https://doi.org/10.1079/9781845936334.0115

Cockram, J. *et al.* (2015). Fine-mapping the wheat *Snn1* locus conferring sensitivity to the Parastagonospora nodorum necrotrophic effector SnTox1 using an eight founder multiparent advanced generation inter-cross population. *G3: Genes, Genomes, Genetics*, 5(11), pp. 2257–2266. Available at: https://doi.org/10.1534/g3.115.021584

FAO (2017). The future of food and agriculture and challenges. Rome.

Haikka, H. *et al.* (2020). Genome-wide association study and genomic prediction for *Fusarium graminearum* resistance traits in nordic oat (*Avena sativa* L.). *Agronomy*, 10(2), 174.
Available at: https://doi.org/10.3390/agronomy10020174

Hyles, J. *et al.* (2020). Phenology and related traits for wheat adaptation. *Heredity*, 125, pp. 417-430. Available at: https://doi.org/10.1038/s41437-020-0320-1

Jamil, M. *et al.* (2019). Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. *BMC Plant Biology*, 19(149).
Available at: https://doi.org/10.1186/s12870-019-1754-6

Kamruzzaman, M. (2022). Genetic analysis of drought stress adaptation in bread wheat diversity. Dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn.
Available at: https://nbn-resolving.org/urn:nbn:de:hbz:5-68894

Lillemo, M. & Dieseth, J.A. (2011). Wheat breeding in Norway. In: *World Wheat Book*, Vol. 2(1432), pp. 45–75.

Liu, X. *et al.* (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLoS Genetics*, 12(2).
Available at: https://doi.org/10.1371/journal.pgen.1005767.

Liu, Y. *et al.* (2017). Dwarfing genes *Rht4* and *Rht-B1b* affect plant height and key agronomic traits in common wheat under two water regimes. *Field Crops Research*, 204, pp. 242–248. Available at: https://doi.org/10.1016/j.fcr.2017.01.020

Ma, L. *et al.* (2016). *TaGS5-3A*, a grain size gene selected during wheat improvement for larger kernel and yield. *Plant Biotechnology Journal*, 14(5), pp. 1269–1280.
Available at: https://doi.org/10.1111/pbi.12492

Mróz, T., Dieseth, J.A. and Lillemo, M. (2022). Historical grain yield genetic gains in Norwegian spring wheat under contrasting fertilization regimes. *Crop Science*, 62(3), pp. 997–1010.
Available at: https://doi.org/10.1002/csc2.20714

Nannuru, V.K.R. *et al.* (2022). Genetic architecture of fusarium head blight disease resistance and associated traits in Nordic spring wheat. *Theoretical and Applied Genetics*, 135(7), pp. 2247–2263. Available at: https://doi.org/10.1007/s00122-022-04109-9

Nordli *et al.* (2008). Regional trends for bud burst and flowering of woody plants in Norway as related to climate change. *International Journal of Biometeorology*, 52(7), pp. 625–639. Available at: https://doi.org/10.1007/s00484-008-0156-5

Pearce, S. *et al.* (2011). Molecular characterization of *Rht-1* dwarfing genes in hexaploid wheat. *Plant Physiology*, 157(4), pp. 1820–1831. Available at: https://doi.org/10.1104/pp.111.183657

Ramirez, I.A. *et al.* (2018). Effects of photoperiod sensitivity genes *Ppd-B1* and *Ppd-D1* on spike fertility and related traits in bread wheat. *Plant Breeding*, 137(3), pp. 320–325. Available at: https://doi.org/10.1111/pbr.12585

Reynolds, M. *et al.* (2009). Raising yield potential in wheat. *Journal of Experimental Botany*, 60(7), pp. 1899–1918. Available at: https://doi.org/10.1093/jxb/erp016

Sørensen, E. (2016). Identification and validation of SNP markers for Fusarium head blight resistance in wheat. Master thesis. Norwegian University of Life Sciences, Ås.

Sayre, K.D., Rajaram, S. and Fischer, R.A. (1997). Yield potential progress in short bread wheats in northwest Mexico. *Crop Science*, 37(1), pp. 36–42. Available at: https://doi.org/10.2135/cropsci1997.0011183X003700010006x.

Segura, V. *et al.* (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics*, 44(7), pp. 825–830. Available at: https://doi.org/10.1038/ng.2314

Shearman, V.J. *et al.* (2005). Physiological processes associated with wheat yield progress in the UK. *Crop Science*, 45(1), pp. 175–185. Available at: https://doi.org/10.2135/cropsci2005.0175

Trethowan, R.M. (2014). Defining a genetic ideotype for crop improvement. *Methods in Molecular Biology*, 1145. Available at: https://doi.org/10.1007/978-1-4939-0446-4_1

Voss-Fels, K.P. *et al.* (2019). Breeding improves wheat productivity under contrasting agrochemical input levels. *Nature Plants*, 5(7), pp. 706–714. Available at: https://doi.org/10.1038/s41477-019-0445-5

Wang, J. & Zhang, Z. (2021). GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics, Proteomics & Bioinformatics*, 19(4), pp. 629-640. Available at: https://doi.org/10.1016/j.gpb.2021.08.005

Wang, Q. *et al.* (2014). A SUPER powerful method for genome wide association study. *PLOS ONE*, 9(9). Available at: https://doi.org/10.1371/journal.pone.0107684

Wang, S. *et al.* (2015). A single-nucleotide polymorphism of *TaGS5* gene revealed its association with kernel weight in Chinese bread wheat. *Frontiers in Plant Science*, 6. Available at: https://doi.org/10.3389/fpls.2015.01166

Wang, S. *et al.* (2016). Haplotypes of the *TaGS5-A1* gene are associated with thousand-kernel weight in Chinese bread wheat. *Frontiers in Plant Science*, 7. Available at: https://doi.org/10.3389/fpls.2016.00783

Wang, Y. *et al.* (2014). Genetic effect of dwarfing gene *Rht13* compared with *Rht-D1b* on plant height and some agronomic traits in common wheat (*Triticum aestivum* L.). *Field Crops Research*, 162, pp. 39–47. Available at: https://doi.org/10.1016/j.fcr.2014.03.014

Yan, L. *et al.* (2003). Positional cloning of the wheat vernalization gene *VRN1*. *PNAS,* 100(10), pp. 6263-6268. Available at: www.pnas.org/cgi/doi/10.1073/pnas.0937399100

Yan, L. *et al.* (2004). The Wheat *VRN2* Gene Is a Flowering Repressor Down-Regulated by Vernalization. *Science*, 303(5664), pp. 1640–1644. Available at: https://doi.org/10.1126/science.1094305

Yan, L. *et al.* (2006). The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. *PNAS*, 103(51), pp. 19581-19586. Available at: https://doi.org/10.1073/pnas.0607142103

Zhang, X. *et al.* (2022). *TaCol-B5* modifies spike architecture and enhances grain yield in wheat. *Science* 376, pp. 180-183. Available at: https://doi.org/10.1126/science.abm0717

Zhang, Y. *et al.* (2018). Analysis of the functions of *TaGW2* homeologs in wheat grain weight and protein content traits. *Plant Journal*, 94(5), pp. 857–866. Available at: https://doi.org/10.1111/tpj.13903

Zhang, Z. *et al.* (2007). Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. *Journal of Animal Science*, 85(4), pp. 881–885. Available at: https://doi.org/10.2527/jas.2006-656

Zhang, Z. *et al.* (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42(4), pp. 355–360. Available at: https://doi.org/10.1038/ng.546

Supplementary material

Grain yield and adaptation of spring wheat to Norwegian growing conditions is driven by allele frequency changes at key adaptive loci discovered by genome-wide association mapping

Manuscript

Tomasz Mróz[1], Jon Arne Dieseth[2], and Morten Lillemo[1]

[1]*Department of Plant Sciences, Norwegian University of Life Sciences, NO-1432 Ås, Norway*
[2]*Graminor, AS, Bjørke Gård, Hommelstadvegen 60, NO-2322 Ridabu, Norway*

Corresponding author:

Morten Lillemo
morten.lillemo@nmbu.no
Orcid-ID: 0000-0002-8594-8794

Table S1 *Planting dates for the main and validation panels' trials*

Season	Vollebekk		Staur	
	Main panel	**Validation panel**	**Main panel**	**Validation panel**
2015	2015-04-24	-	-	-
2016	2016-04-24	-	2016-05-10	-
2017	2017-05-04	-	2017-05-12	-
2018	2018-05-10	-	2018-05-09	-
2019	2019-05-19	2019-04-24	2019-06-04	-
2020	2020-05-15	2020-04-15	2020-04-21	2020-04-21
2021	2021-04-20	2021-04-19	2021-04-27	2021-04-27
2022	-	2022-04-26	-	-

Table S2 *Overview of the number of lines which were phenotyped in each season/environment combination in the main panel. Total – total number of unique genotypes present in the main panel*

Location	Field season						
	2015	**2016**	**2017**	**2019**	**2020**	**2021**	**Total**
Vollebekk	163	100	240	220	288	295	301
Staur	-	100	240	220	288	-	296

Table S3 *Overview of the number of lines which were phenotyped in each season/environment combination in the validation panel. Total – total number of unique genotypes present in the validation panel*

Location	Field season				Total
	2019	**2020**	**2021**	**2022**	
Vollebekk	309	397	267	265	889
Staur	90	354	-	-	

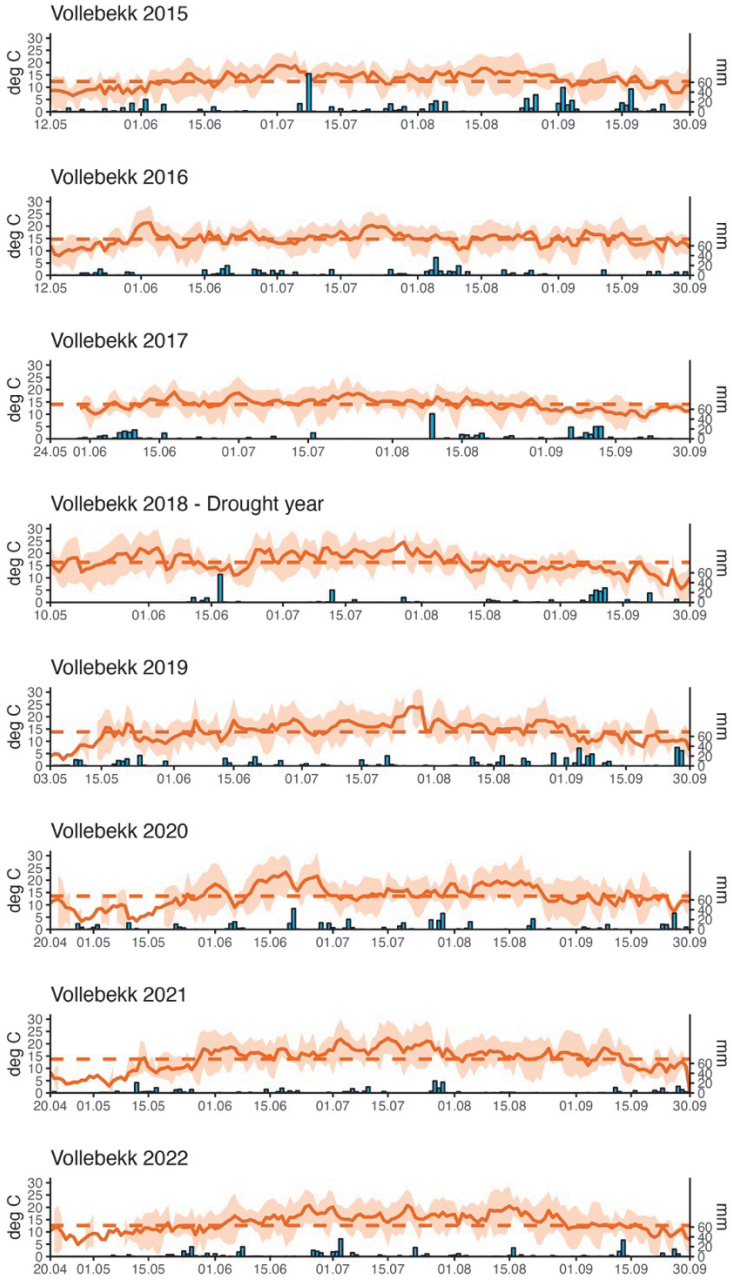


Figure S1 *Weather data for field seasons 2015-2022 in Vollebekk research station, from seeding the trial to the end of September. Solid line marks the average temperature for each day, dashed line marks the mean temperature across the season, bars mark the daily rainfall in mm per m². Primary y axis – temperature in ^{0}C, secondary y axis – rainfall in mm per m²*

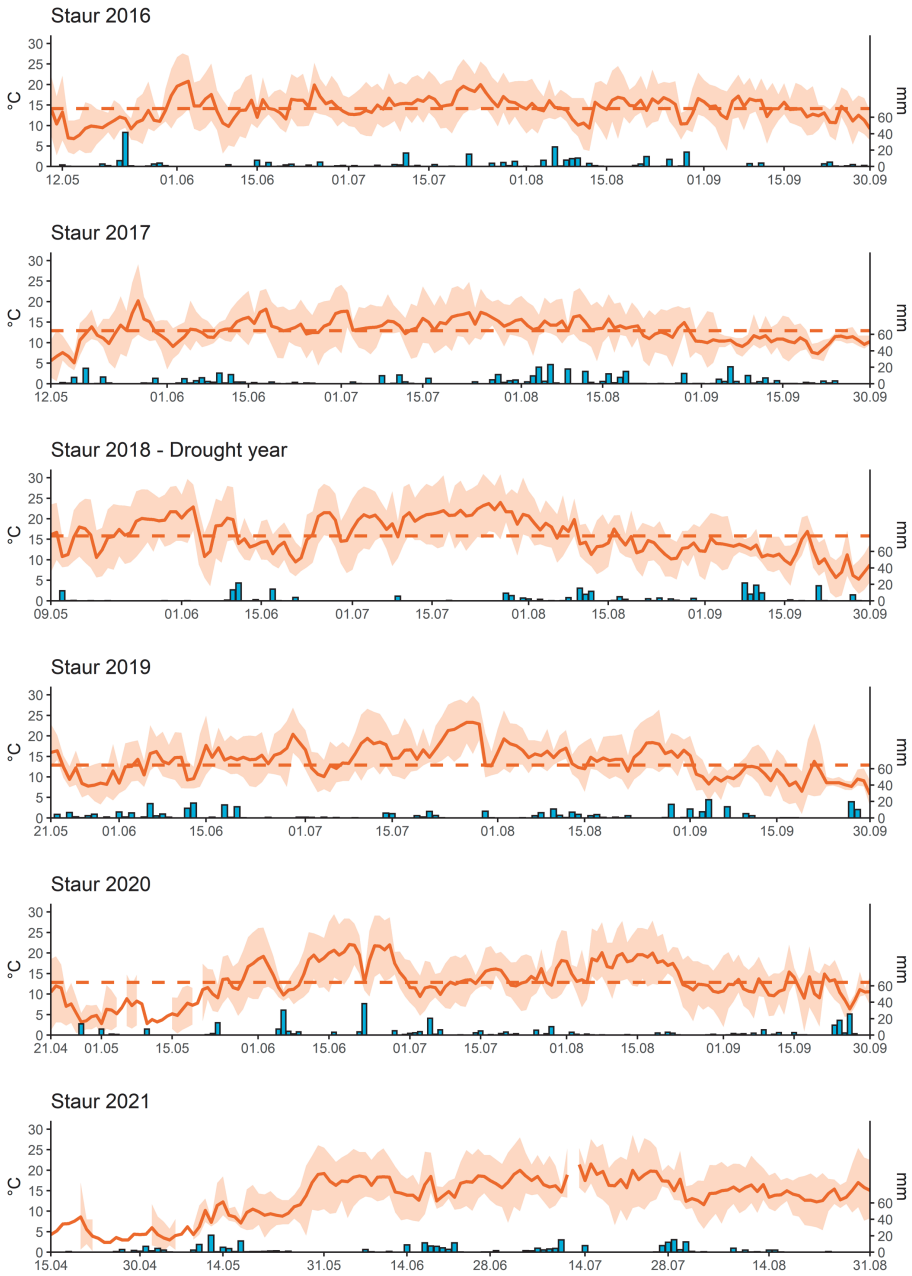


Figure S2 *Weather data for field seasons 2016-2021 in Staur, from seeding the trial to the end of September. Solid line marks the average temperature for each day, dashed line marks the mean temperature across the season, bars mark the daily rainfall in mm per m². Primary y axis – temperature in ^{0}C, secondary y axis – rainfall in mm per m²*

Table S4 *Descriptive statistics of weather conditions at Vollebekk research station over the analyzed growing seasons. T_{avg} – average temperature in ℃, Rf_{sum} – sum of rainfall in mm, Ir_{avg} – average solar irradiance in Wm^{-2}*

MON.	2015			2016			2017			2018 (DROUGHT)			2019			2020			2021			2022		
	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}
APR	5.5	8.8	15.4	5.2	100	11.7	4.4	35.0	13.4	5.1	32.4	14,0	7.9	14.6	16.2	6.4	30.2	15.1	4.7	18.2	17.1	5.4	10	15.2
MAY	8.3	116.6	16.3	11.6	50.2	18.5	*NA*	68.0	*NA*	15.0	0	22.2	9.7	100.6	17.0	9.5	47.2	21.4	9.7	72.2	15.4	10.7	63.2	17.8
JUN	13.1	65.6	21.4	15.6	89.8	21.1	14.3	94.0	19.1	16.7	82.4	23.9	14.8	63.6	17.8	17.6	115.4	21.2	16.1	34.6	19.9	16.0	72.8	21.5
JUL	15.0	158.8	19.4	16.1	54.8	19.1	15.9	20.2	19.8	20.2	44.6	23.2	17.2	52.0	20.2	14.3	127.9	18.0	18.9	95.4	19.8	16.4	91.8	19.7
AUG	15.3	140.2	16.0	14.6	140	14.1	14.5	104.0	14.9	15.4	21.4	14.2	16.2	110.2	13.7	16.2	50.6	15.7	15.3	7.8	15.2	16.2	34.8	16.0
SEP	11.5	220.2	8.8	14.1	41	9.9	11.5	119.2	6.7	12.1	128.4	9.4	11.0	190.8	8.3	12.0	81.0	9.0	12.5	75.2	8.7	11.1	76.8	8.79

Table S5 *Descriptive statistics of weather conditions at Staur research farm over the analyzed growing seasons. T_{avg} – average temperature in ℃, Rf_{sum} – sum of rainfall in mm, Ir_{avg} – average solar irradiance in Wm^{-2}*

MON.	2015			2016			2017			2018 (DROUGHT)			2019			2020			2021		
	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}	T_{avg}	Rf_{sum}	Ir_{avg}
APR	4.6	6.9	14.0	3.9	74.3	12.1	3.6	33.4	12.9	3.5	40.3	13.8	5.6	3.0	15.3	5.5	19.4	14.3	3.7	16.0	15.5
MAY	7.4	89.0	14.2	10.6	66.5	16.8	10.1	59.3	15.7	15.1	22.8	21.1	8.5	91.4	15.1	8.3	33.8	18.5	9.0	81.2	14.9
JUN	12.6	39.3	19.9	15.2	29.5	20.7	13.5	58.3	17.3	16.0	55.8	22.3	14.4	103.6	16.8	17.4	98.9	21.5	16.4	62.4	20.7
JUL	14.8	116.3	16.4	15.8	59.4	17.7	15.0	58.8	18.3	20.7	24.8	22.4	16.6	37.6	19.6	13.3	70.8	16.8	NA	93.4	NA
AUG	14.6	52.7	15.8	14.2	109.1	12.9	13.7	144.2	13.9	14.7	61.2	13.6	15.5	63.1	12.4	16.0	17.1	15.7	14.3	16.4	14.2
SEP	10.9	164.6	7.3	13.6	21.4	9.4	10.3	64.3	5.4	11.1	85.1	8.6	9.9	103.2	7.2	11.3	80.7	8.1	11.6	34.1	7.8

Table S6 *Significant markers belonging to the discovered QTL regions for days to heading (DH), days to maturity (DM), grain yield (GY), and plant height (PH). Pos – SNP position in Mbp, E – effect (without correcting for population structure) in the full dataset, %PVE – percentage of variance explained in the full dataset, without correcting for population structure, MAF – minor allele frequency*

Trait	QTL	Span	Marker	Pos	-logP	Dataset	Method	E	%PVE	MAF
DH	***QHd.nmbu-1B***	**1-2**	**BS00022180_51**	1	12.95	Full	FarmCPU, MLM	1.02	14.0	0.08
			BS00071161_51	2	3.65	Full/Adapted	MLM	0.95	13.3	0.08
	QHd.nmbu-7B	**606**	**BobWhite_c3541_152**	606	10.06	Adapted	FarmCPU, MLM	0.91	11.4	0.06
DM	***QMat.nmbu-6B***	**132-136**	wsnp_Ex_rep_c102159_87386822	131	4.22	Full/Adapted	MLM	0.63	3.51	0.21
			Kukri_rep_c71420_511	132	11.34	Adapted	FarmCPU, MLM	1.36	14.2	0.14
			Tdurum_contig14559_741	132	5.73	Full/Adapted	MLM	0.61	4.12	0.28
			Tdurum_contig14559_78	132	4.78	Full/Adapted	MLM	0.60	4.05	0.28
			wsnp_Ex_c12577_20022294	132	3.16	Full/Adapted	MLM	0.28	0.92	0.32
			GENE-3171_203	135	3.69	Adapted	MLM	0.97	9.74	0.21
			GENE-4074_537	135	4.22	Full/Adapted	MLM	0.67	3.57	0.17
			AX-158529874	136	3.11	Adapted	MLM	0.89	8.55	0.21
			AX-158552731	136	3.50	Adapted	MLM	1.06	10.4	0.18
			Tdurum_contig11700_1247	136	3.11	Adapted	MLM	0.89	8.55	0.21
	QMat.nmbu-6D	**6**	**BS00022523_51**	6	6.28	Full	FarmCPU, MLM	0.82	8.80	0.41
GY	***QYld.nmbu-3A***	**267**	**BS00110129_51**	267	7.16	Full	FarmCPU, MLM	38.1	22.0	0.42
	QYld.nmbu-5A	**683-708**	AX-158558760	685	3.50	Full	MLM	37.5	17.5	0.27
			AX-94387378	686	3.17	Full	MLM	42.2	24.0	0.31
			AX-95097524	694	3.82	Full	MLM	52.6	34.6	0.28
			BobWhite_c8266_227	699	36.78	Full	FarmCPU, MLM	79.4	61.5	0.20
			AX-158550750	708	3.55	Full	MLM	59.2	41.3	0.25
			wsnp_Ex_c2171_4072774	708	3.47	Full	MLM	59.3	41.3	0.25
			wsnp_Ex_c2171_4072995	708	3.47	Full	MLM	59.3	41.3	0.25
			wsnp_Ex_c2171_4073721	708	3.47	Full	MLM	59.3	41.3	0.25
	QYld.nmbu-7B	**701-703**	**BS00083578_51**	701	9.85	Full/Adapted	FarmCPU, MLM	48.3	12.8	0.10
			BobWhite_c12048_145	701	3.53	Full/Adapted	MLM	26.8	10.7	0.37
			BS00022522_51	701	5.24	Full/Adapted	MLM	46.1	11.8	0.10
			Excalibur_c50612_146	701	3.08	Adapted	MLM	31.5	21.6	0.11

Table S6 *Cont.*

PH	***QHt.nmbu-2A***	**524-543**	AX-110920715	524	3.47	Full	MLM	3.73	9.56	0.09
			AX-95095516	528	3.02	Adapted	FarmCPU, MLM	4.01	12.6	0.08
			AX-158561839	528	2.96	Full/Adapted	MLM	4.09	10.7	0.08
			AX-110399256	543	4.78	Full	FarmCPU, MLM	1.81	11.6	0.08
			AX-158540522	543	2.99	Adapted	MLM	4.00	12.6	0.08
	QHt.nmbu-4A	**570-603**	**CAP11_c3631_75**	570	7.52	Adapted	FarmCPU, MLM	1.98	6.97	0.22
			RAC875_c19303_228	578	5.73	Full/Adapted	MLM	2.02	7.70	0.33
			AX-95086844	603	3.14	Full	FarmCPU	0.89	0.38	0.33
	QHt.nmbu-4B	**13-59**	AX-158583339	13	3.43	Full	MLM	4.09	10.7	0.24
			BS00037094_51	13	3.25	Full	MLM	1.83	5.15	0.24
			Tdurum_contig93710_409	13	3.56	Full	MLM	1.79	4.70	0.23
			AX-110588223	23	3.96	Full/Adapted	MLM	1.84	7.08	0.44
			wsnp_Ex_c7362_12622736	23	3.40	Adapted	MLM	1.86	8.01	0.34
			TG0010a; TG0010b	31	5.50	Full/Adapted	FarmCPU	2.87	14.6	0.28
			AX-94685504	31	5.65	Full/Adapted	MLM	2.66	15.1	0.49
			AX-89380014	32	3.84	Full/Adapted	MLM	2.29	10.7	0.39
			AX-111081978	36	5.82	Full/Adapted	MLM	2.61	12.2	0.30
			AX-158564543	37	6.10	Full/Adapted	MLM	2.60	12.0	0.29
			AX-158564633	37	4.34	Full/Adapted	MLM	1.82	6.86	0.42
			AX-158537142	38	6.42	Full/Adapted	MLM	2.63	12.2	0.29
			AX-158564542	38	5.55	Full/Adapted	MLM	2.02	7.68	0.33
			Tdurum_contig33737_157	38	6.04	Full/Adapted	MLM	2.53	11.2	0.29
			IAAV971	41	7.00	Full/Adapted	MLM	2.76	12.5	0.26
			AX-158618805	52	6.12	Full/Adapted	MLM	2.23	8.02	0.25
			Excalibur_c56787_95	59	6.61	Full/Adapted	MLM	2.71	11.4	0.24
	QHt.nmbu-4D	**19-26**	TG0011b	19	25.89	Full (MLM) Adapted	FarmCPU, MLM	3.45	22.0	0.25
			TG0011a	19	10.53	Full/Adapted	MLM	2.53	8.97	0.22
			BobWhite_s64797_152	26	26.06	Full, Adapted (MLM)	FarmCPU, MLM	3.18	11.7	0.29
	QHt.nmbu-6B	**202**	**Ra_c10469_616**	202	6.68	Full	FarmCPU	2.00	5.21	0.19

Table S7 *Validation of associated markers using sets of independent lines in four seasons in two locations. V – Vollebekk, S – Staur, for instance 2019_S indicates field season 2019 in Staur. Validation results coded as follows: "-" – phenotype data not available, Mono – marker monomorphic (validation not possible), Rare - minor allele frequency < 0.05, validation not reliable, Rare* - marker with minor allele frequency < 0.05, but visible signs of possible association, Ns – marker polymorphic in the population (minor allele frequency > 0.05), but no significant effect (p <= 0.05), number and unit - marker polymorphic in the population (minor allele frequency > 0.05) and visible significant effect (p <= 0.05)*

QTL	Span	Marker	Pos	Alleles Pos/Neg	Dominant allele	Validation results					
						2019_S	2019_V	2020_S	2020_V	2021_V	2022_V
QHd.nmbu-1B	**1-2**	**BS00022180_51**	1	T/C	C	Rare	Rare	-	Ns	Rare*	Rare
		BS00071161_51	2	-	T	Rare	Rare	-	Rare	Rare	Rare
QHd.nmbu-7B	**606**	**BobWhite_c3541_152**	606	G/A	G	Rare	Rare	-	Rare*	Rare	Rare*
QMat.nmbu-6B	**132-136**	wsnp_Ex_rep_c102159_87386822	131	G/A	G	-	1.29dss	-	2.23dss	2.19dss	0.94dss
		Kukri_rep_c71420_511	132	A/G	A	-	1.19dss	-	2.33dss	2.32dss	0.89dss
		Tdurum_contig14559_741	132	G/A	G	-	0.72dss	-	1.27dss	1.92dss	0.95dss
		Tdurum_contig14559_78	132	G/A	A	-	0.74dss	-	1.27dss	1.89dss	0.95dss
		wsnp_Ex_c12577_20022294	132	-	C	-	Ns	-	Ns	Ns	Ns
		GENE-3171_203	135	C/T	C	-	Ns	-	1.83dss	2.38dss	1.72dss
		GENE-4074_537	135	G/A	G	-	Ns	-	1.52dss	Rare	Rare
		AX-158529874	136	G/A	G	-	Ns	-	1.11dss	1.85dss	0.78dss
		AX-158552731	136	G/A	G	-	Ns	-	1.70dss	2.38dss	1.72dss
		Tdurum_contig11700_1247	136	T/C	T	-	Ns	-	1.10dss	1.83dss	0.92dss
QMat.nmbu-6D	**6**	**BS00022523_51**	6	A/G	A	-	Ns	-	-	-	1.61dss
QYld.nmbu-3A	**267**	**BS00110129_51**	267	A/G	A	Ns	Ns	35.41g	-	-	27.71g

Table S7 *Cont.*

QYld.nmbu-5A	**683-708**	AX-158558760	685	A/G	G	Ns	Ns	Ns	-	-	18.50g
		AX-94387378	686	A/G	G	Ns	Ns	Ns	-	-	18.52g
		AX-95097524	694	G/A	G	47.63g	Ns	Ns	-	-	Rare
		BobWhite_c8266_227	699	-	-	Mono	Mono	Mono	-	-	Mono
		AX-158550750	708	-	A	Ns	Rare	Rare	-	-	Ns
		wsnp_Ex_c2171_4072774	708	-	C	Rare	Rare	Rare	-	-	Ns
		wsnp_Ex_c2171_4072995	708	-	A	Ns	Ns	Rare	-	-	Ns
		wsnp_Ex_c2171_4073721	708	-	T	Rare	Rare	Rare	-	-	Ns
QYld.nmbu-7B	**701-703**	**BS00083578_51**	701	T/C	T	Rare*	Rare	Rare*	-	-	Mono
		BobWhite_c12048_145	701	A/C	A	Ns	Ns	21.56g	-	-	23.62g
		BS00022522_51	701	C/T	C	Rare	Rare	Rare*	-	-	Rare
		Excalibur_c50612_146	701	C/T	C	Rare	Rare	33.81g	-	-	Rare*
QHt.nmbu-2A	**524-543**	AX-110920715	524	G/A	A	Mono	Rare	Rare	Rare	Rare	Rare*
		AX-95095516	528	A/C	A	Mono	Rare	Rare	Rare	Rare	Rare*
		AX-158561839	528	-	C	Mono	Rare	Rare	Rare	-	Rare
		AX-110399256	543	-	C	Mono	Rare	Rare	Rare	Rare	Rare
		AX-158540522	543	-	C	Mono	Rare	Rare	Rare	Rare	Rare
QHt.nmbu-4A	**570-603**	**CAP11_c3631_75**	570	T/C	T	7.23cm	6.41cm	3.87cm	4.81cm	6.53cm	7.07cm
		RAC875_c19303_228	578	A/G	A	7.24cm	6.98cm	3.89cm	5.53cm	6.77cm	8.02cm
		AX-95086844	603	C/T	T	Ns	1.89cm	Ns	Ns	Ns	Ns
QHt.nmbu-4B	**13-59**	AX-158583339	13	G/A	G	7.49cm	4.65cm	2.94cm	4.14cm	5.74cm	5.48cm
		BS00037094_51	13	A/G	A	7.04cm	4.42cm	2.95cm	3.75cm	5.57cm	5.55cm
		Tdurum_contig93710_409	13	A/G	A	7.49cm	4.38cm	3.01cm	3.84cm	5.66cm	5.93cm
		AX-110588223	23	G/T	G	Ns	1.98cm	2.30cm	2.15cm	1.96cm	3.89cm
		wsnp_Ex_c7362_12622736	23	T/G	T	Ns	2.07cm	2.21cm	1.97cm	1.82cm	3.26cm
		TG0010a; TG0010b	31	C/T	C	9.36cm	7.89cm	4.21cm	5.73cm	6.99cm	8.05cm
		AX-94685504	31	G/A	Unclear	Ns	5.70cm	4.23cm	4.68cm	5.60cm	6.80cm
		AX-89380014	32	G/A	Unclear	6.25cm	6.04cm	4.01cm	4.01cm	5.06cm	5.69cm
		AX-111081978	36	C/T	C	6.43cm	6.89cm	4.08cm	5.44cm	6.80cm	7.84cm
		AX-158564543	37	A/G	A	7.24cm	7.04cm	4.02cm	5.48cm	6.72cm	8.02cm
		AX-158564633	37	A/G	G	6.65cm	4.33cm	2.22cm	3.63cm	1.50cm	2.43cm

Table S7 *Cont.*

		AX-158537142	38	C/T	C	7.24cm	7.01cm	3.88cm	5.40cm	6.73cm	7.99cm
		AX-158564542	38	A/C	A	7.24cm	7.01cm	3.86cm	5.51cm	6.67cm	8.02cm
		Tdurum_contig33737_157	38	A/G	A	7.24cm	6.98cm	3.95cm	5.54cm	6.74cm	8.02cm
		IAAV971	41	C/T	C	7.24cm	6.97cm	4.07cm	5.38cm	6.60cm	7.96cm
		AX-158618805	52	C/T	C	7.24cm	6.56cm	3.99cm	4.88cm	6.50cm	6.69cm
		Excalibur_c56787_95	59	G/T	G	7.24cm	6.74cm	4.11cm	5.18cm	6.70cm	6.72cm
QHt.nmbu-4D	**19-26**	TG0011b	19	G/T	G	Mono	5.59cm	Ns	4.76cm	6.14cm	10.44cm
		TG0011a	19	G/T	G	Mono	4.76cm	3.29cm	5.66cm	7.46cm	10.40cm
		BobWhite_s64797_152	26	A/C	A	7.24cm	7.01cm	3.85cm	5.50cm	6.72cm	8.03cm
QHt.nmbu-6B	**202**	**Ra_c10469_616**	202	G/A	A	Rare	Rare	2.39cm	2.24cm	Ns	4.32cm

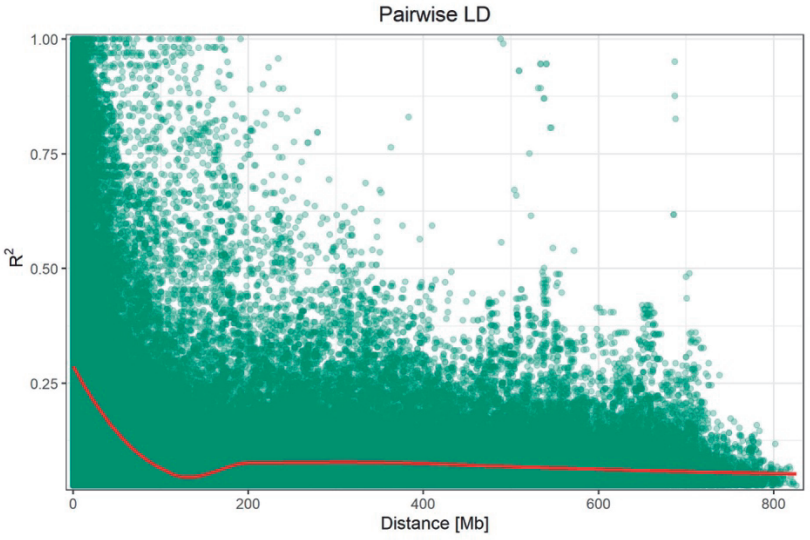


Figure S3 *Linkage disequilibrium (LD) between markers for the main panel*

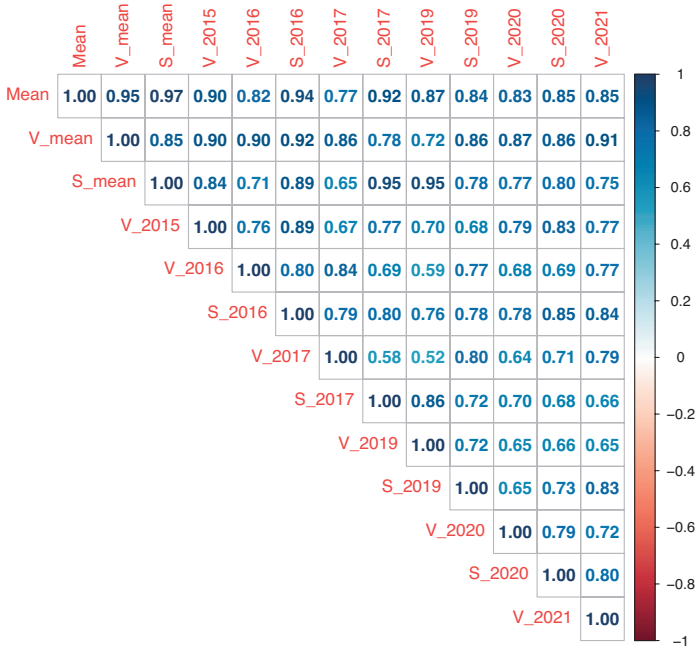


Figure S4 *Pearson's correlation matrix for* ***days to maturity*** *between trial means, environmental means, and the global mean. Variables named according to scheme: Location_season. V – Vollebekk, S – Staur*

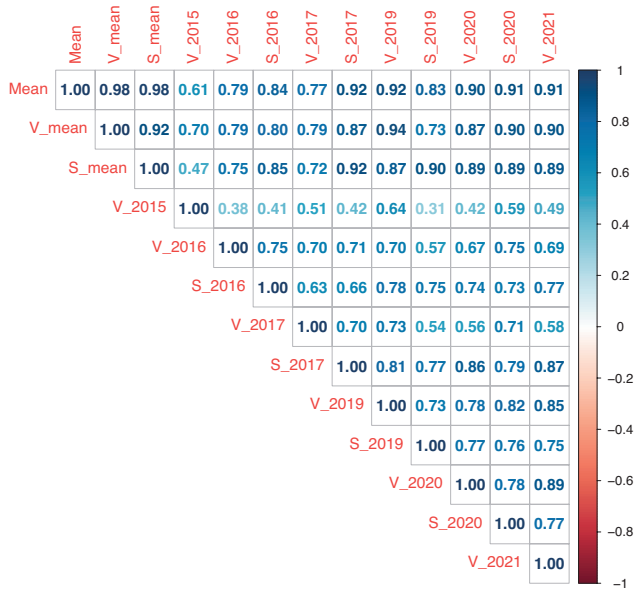


Figure S5 *Pearson's correlation matrix for* ***grain yield*** *between trial means, environmental means, and the global mean. Variables named according to scheme: Location_season. V – Vollebekk, S – Staur*

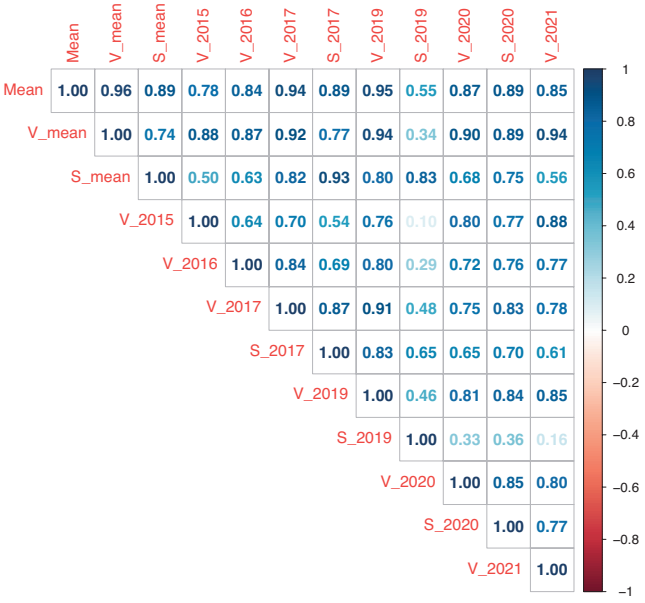


Figure S6 *Pearson's correlation matrix for* ***days to heading*** *between trial means, environmental means, and the global mean. Variables named according to scheme: Location_season. V – Vollebekk, S – Staur*

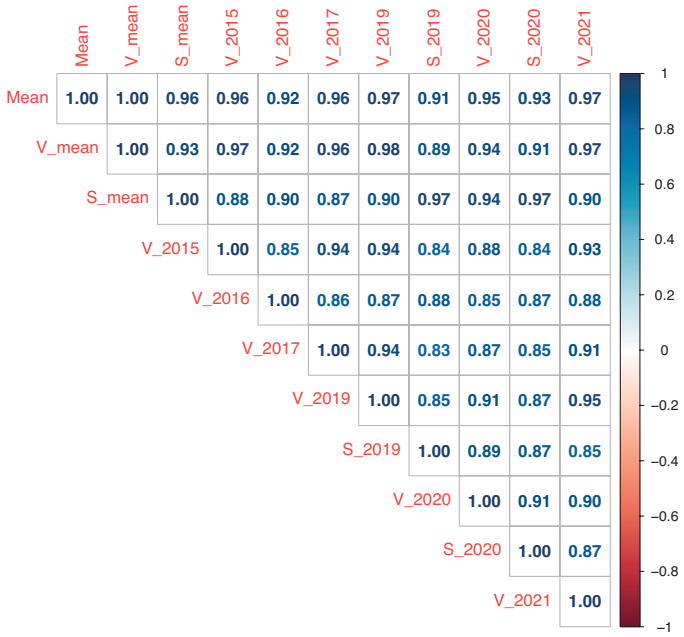


Figure S7 *Pearson's correlation matrix for* ***plant height*** *between trial means, environmental means, and the global mean. Variables named according to scheme: Location_season. V – Vollebekk, S – Staur*

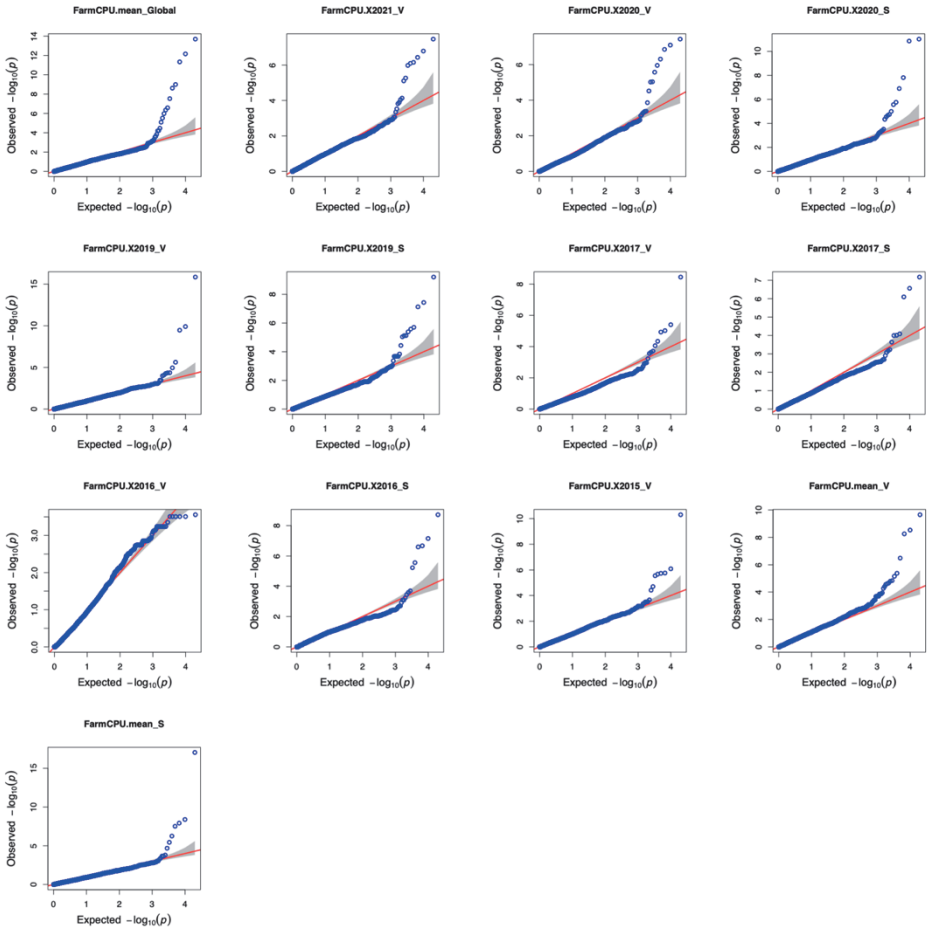


Figure S8 *Quantile-quantile (QQ) plots for GWAS analysis of* ***days to maturity*** *for the* ***adapted part*** *of the main panel. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

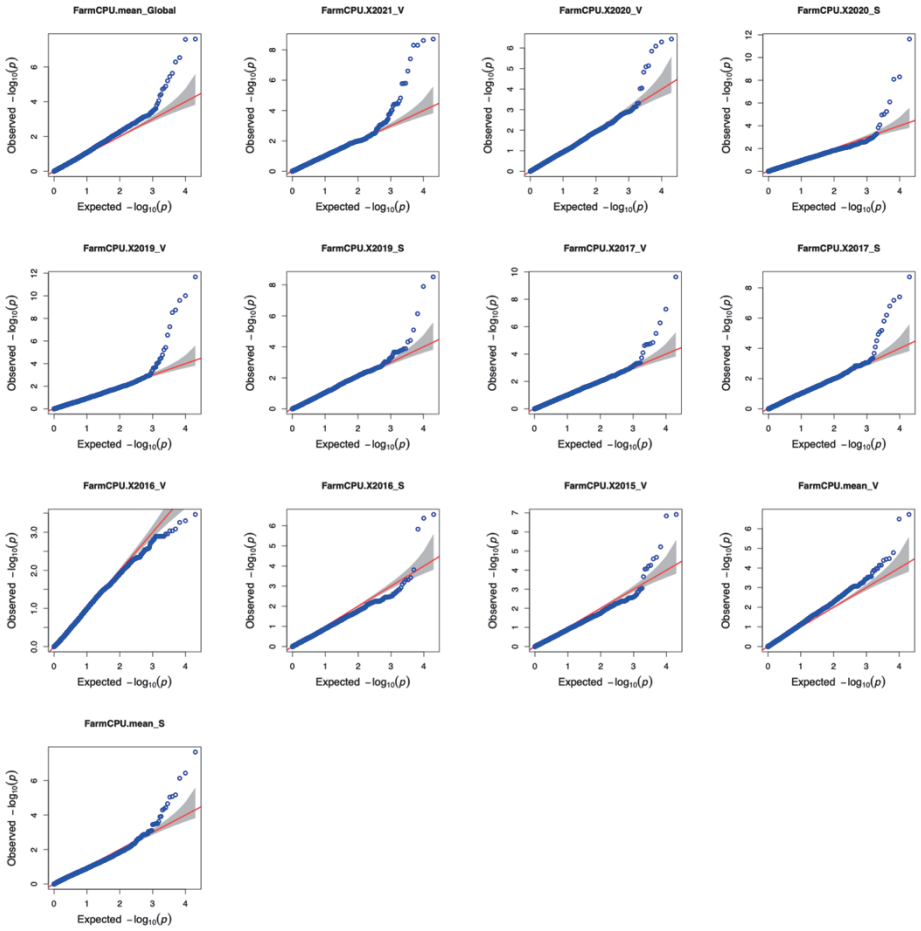


Figure S9 *Quantile-quantile (QQ) plots for GWAS analysis of* ***days to maturity*** *for the main panel* ***(all lines)****. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

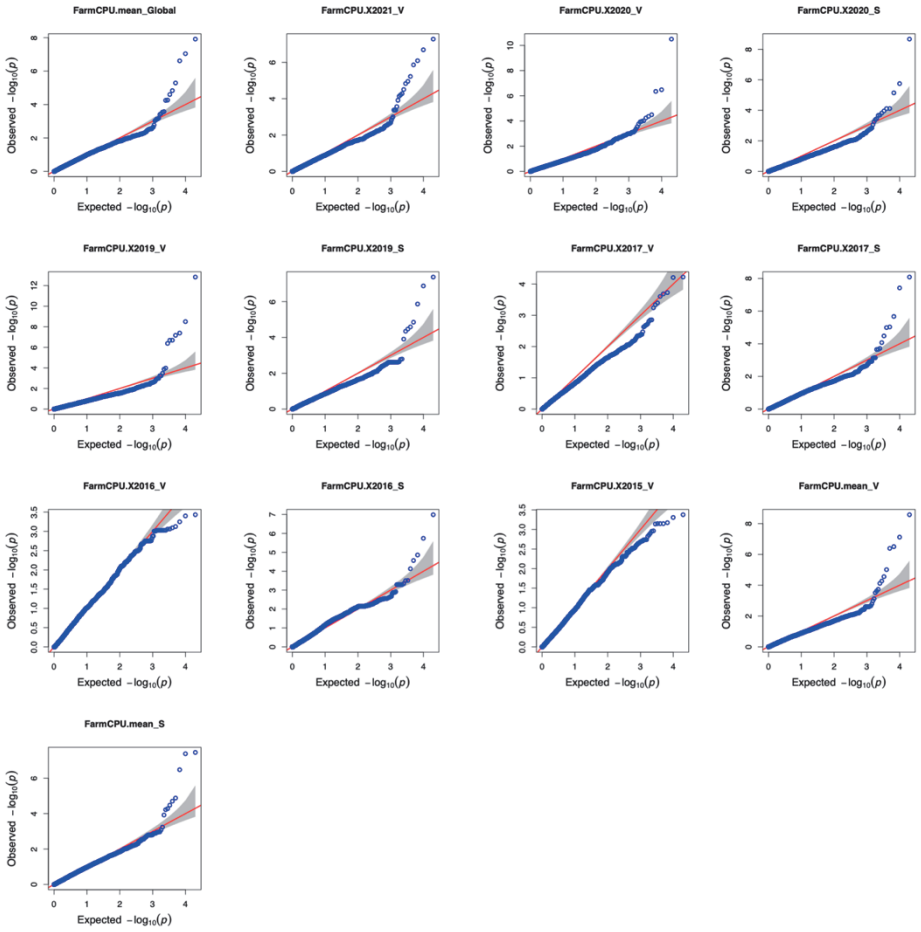


Figure S10 *Quantile-quantile (QQ) plots for GWAS analysis of* ***grain yield*** *for the* ***adapted part*** *of the main panel. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

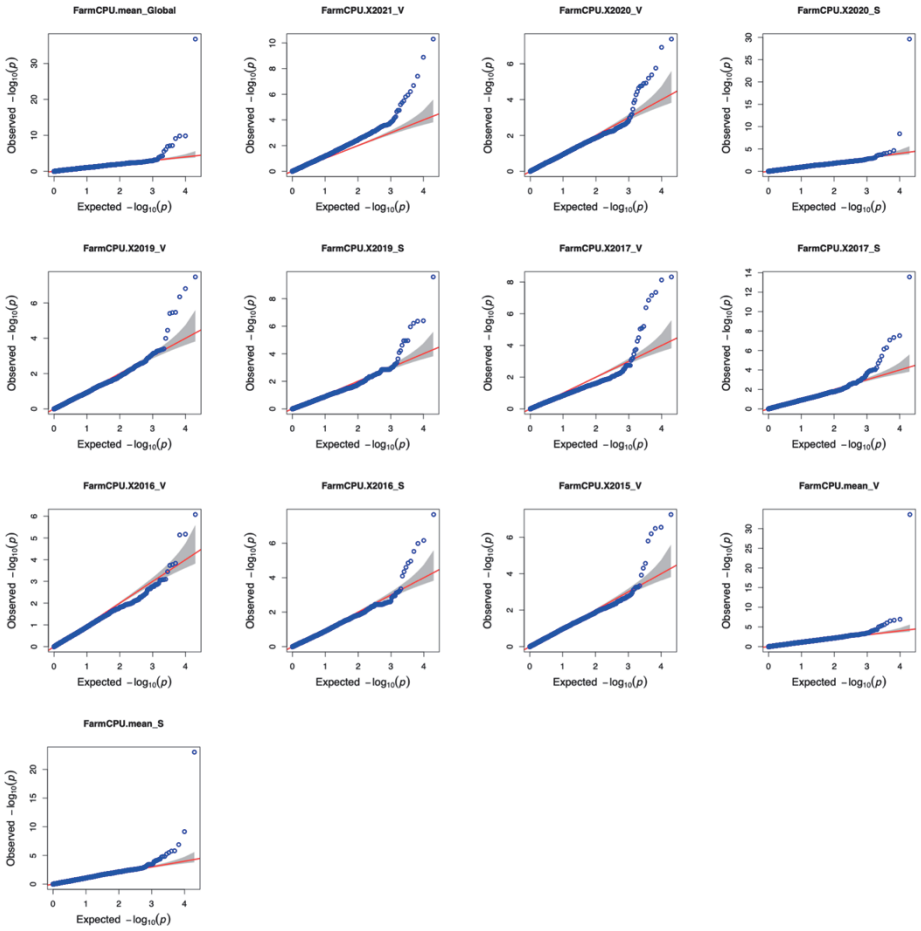


Figure S11 *Quantile-quantile (QQ) plots for GWAS analysis of* ***grain yield*** *for the main panel* ***(all lines)****. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

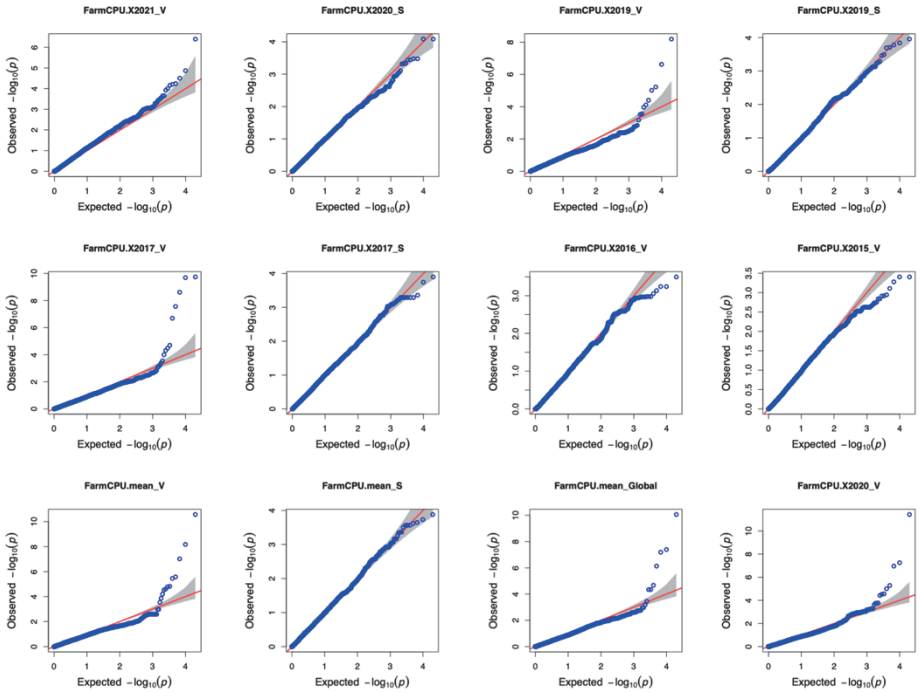


Figure S12 *Quantile-quantile (QQ) plots for GWAS analysis of* ***days to heading*** *for the* ***adapted part*** *of the main panel. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

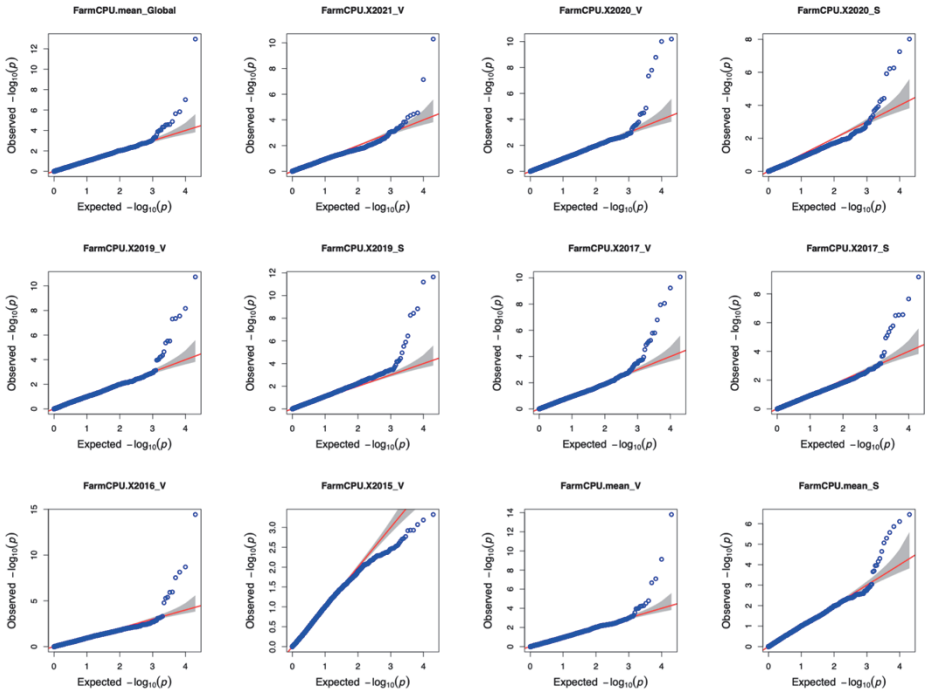


Figure S13 *Quantile-quantile (QQ) plots for GWAS analysis of* ***days to heading*** *for the main panel* ***(all lines)****. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

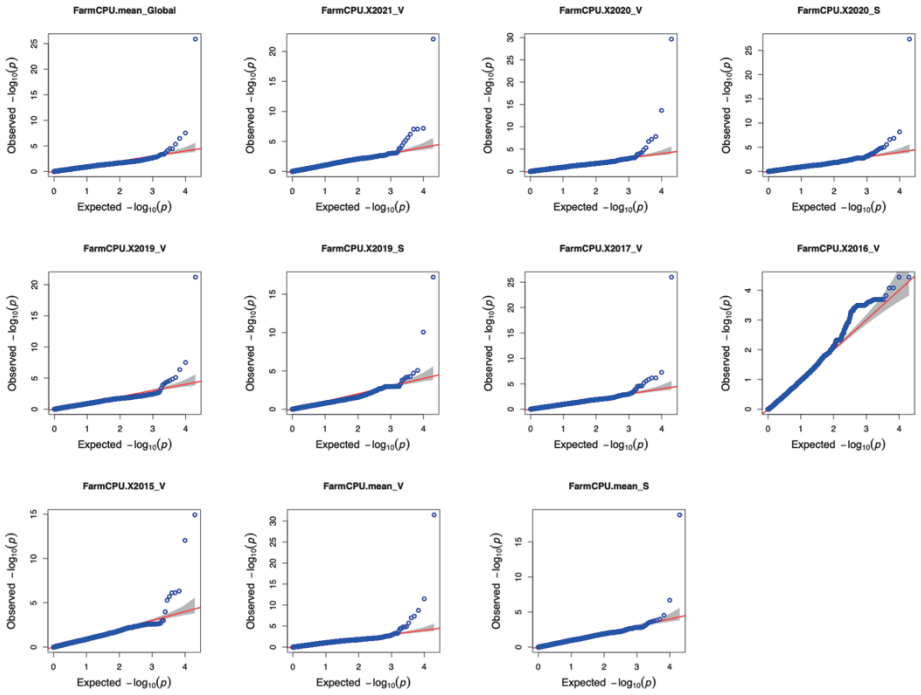


Figure S14 *Quantile-quantile (QQ) plots for GWAS analysis of* ***plant height*** *for the* ***adapted part*** *of the main panel. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

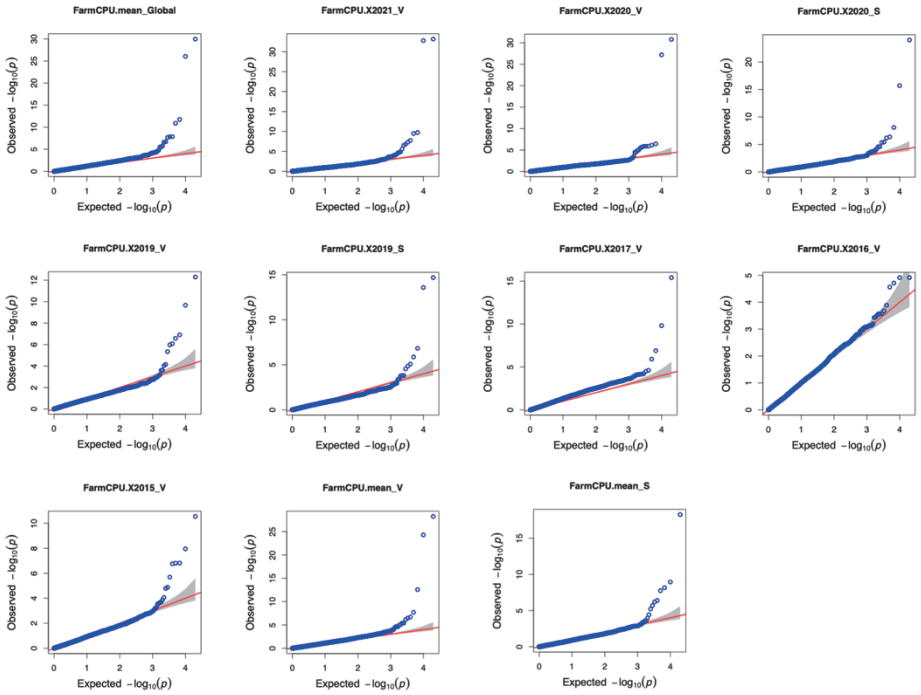


Figure S15 *Quantile-quantile (QQ) plots for GWAS analysis of* ***plant height*** *for the main panel* ***(all lines)****. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

Figure S16 *Manhattan plots for GWAS analysis of* ***days to maturity*** *for the* ***adapted part*** *of the main panel. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

Figure S17 *Manhattan plots for GWAS analysis of* ***days to maturity*** *for the main panel* ***(all lines)****. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

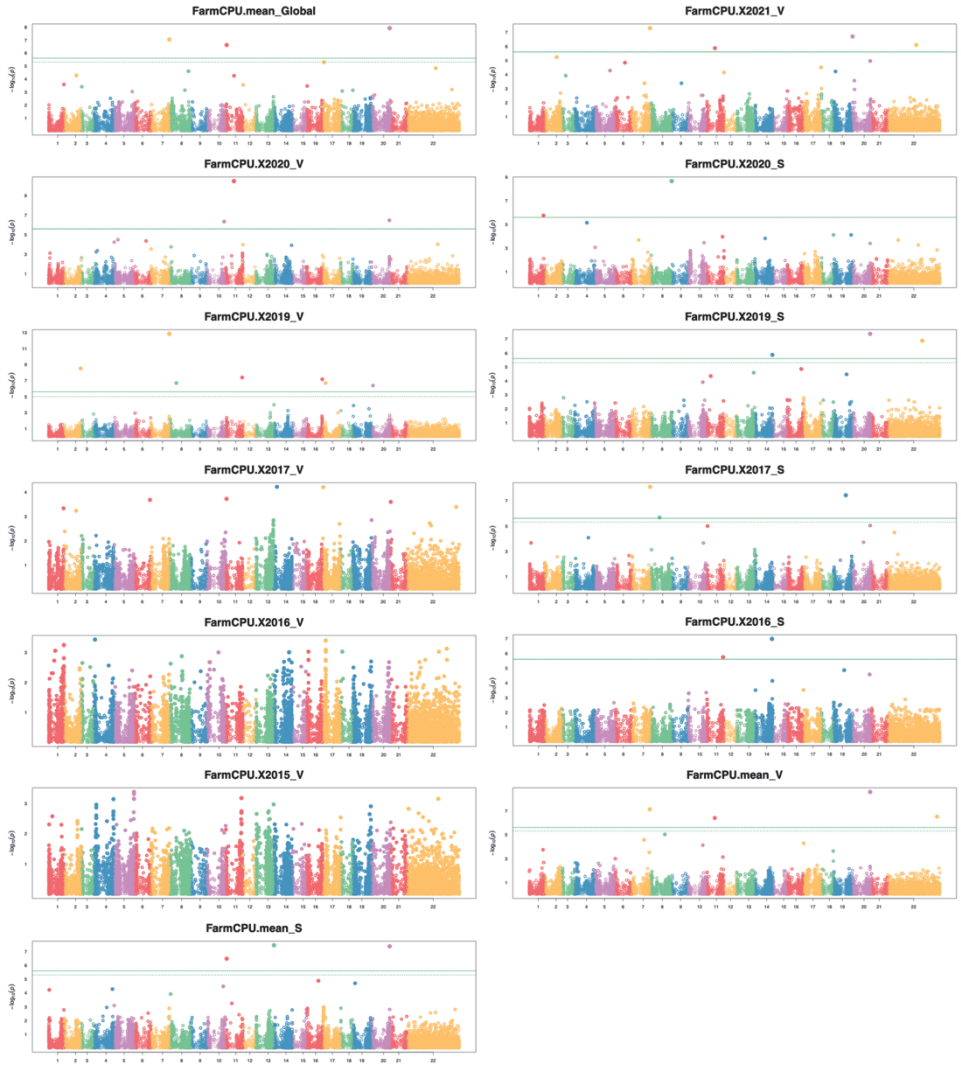


Figure S18 *Manhattan plots for GWAS analysis of* ***grain yield*** *for the* ***adapted part*** *of the main panel. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

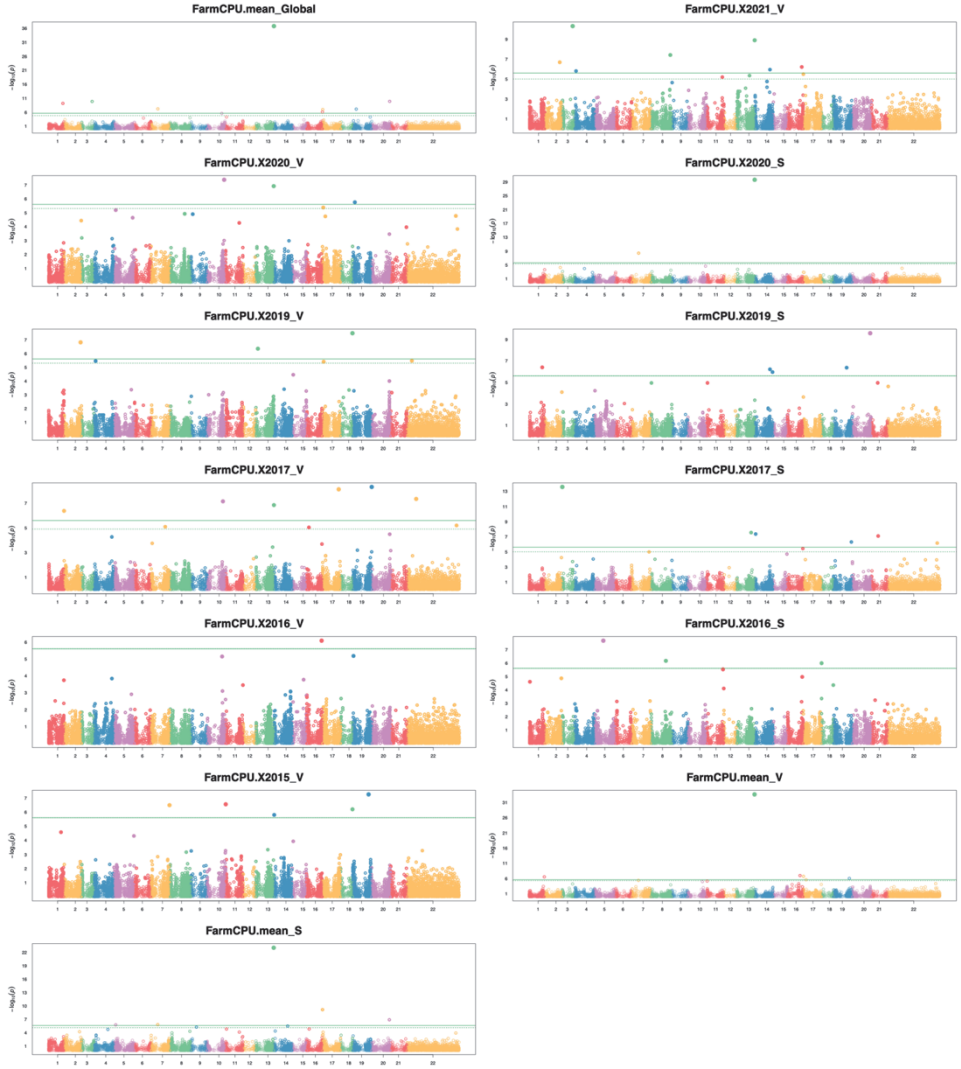


Figure S19 *Manhattan plots for GWAS analysis of* ***grain yield*** *for the main panel* ***(all lines)****. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

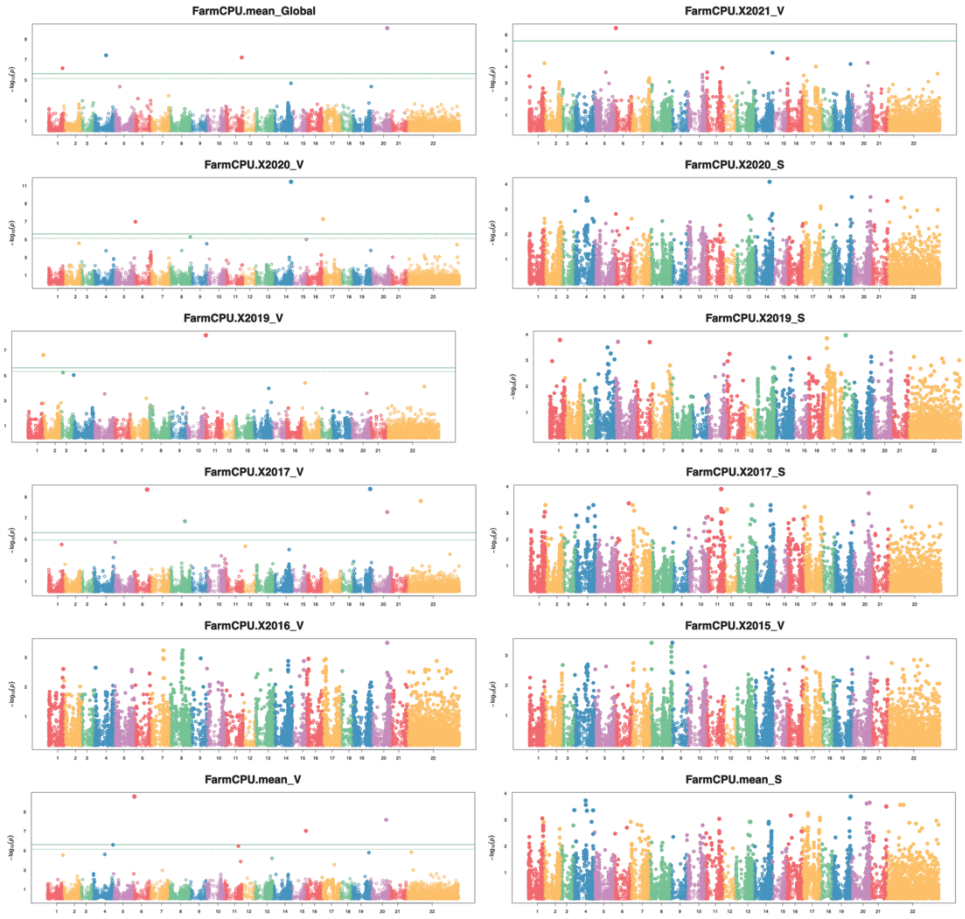


Figure S20 *Manhattan plots for GWAS analysis of* ***days to heading*** *for the* ***adapted part*** *of the main panel. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

Figure S21 *Manhattan plots for GWAS analysis of* ***days to heading*** *for the main panel* ***(all lines)****. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

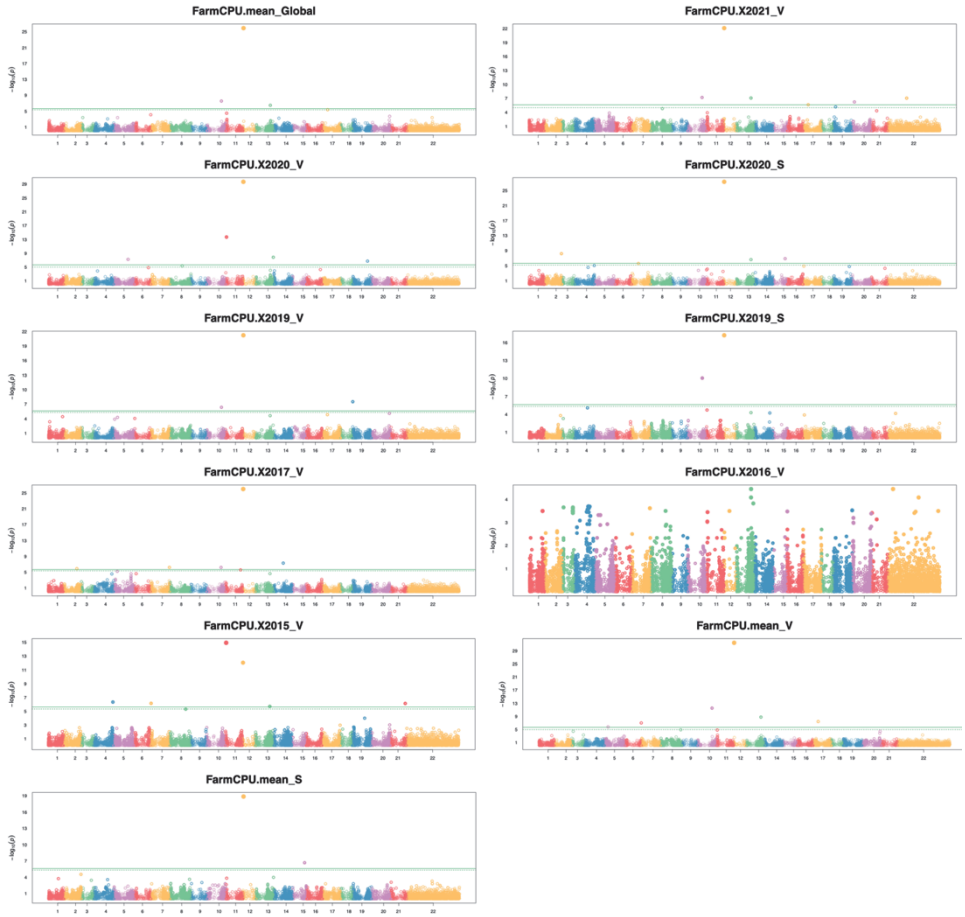


Figure S22 *Manhattan plots for GWAS analysis of* ***plant height*** *for the* ***adapted part*** *of the main panel. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

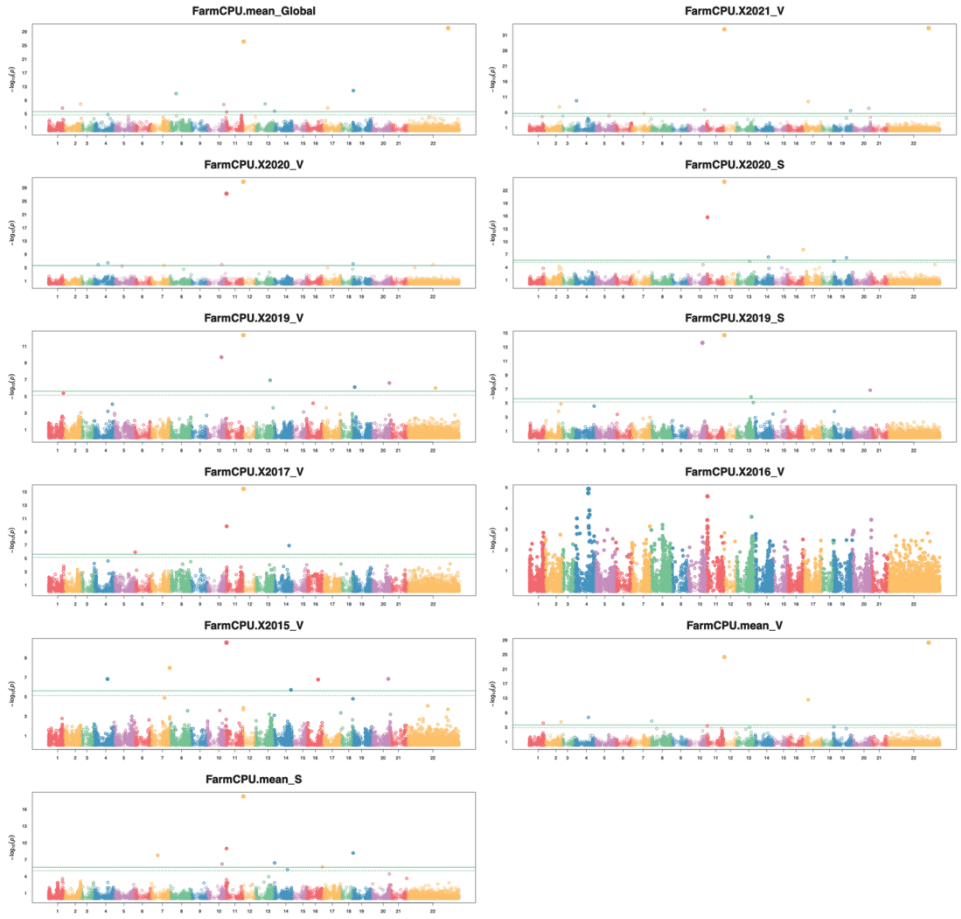


Figure S23 *Manhattan plots for GWAS analysis of* ***plant height*** *for the main panel* ***(all lines)****. Phenotypes named according to the scheme: FarmCPU.season_location. Location: V – Vollebekk, S – Staur*

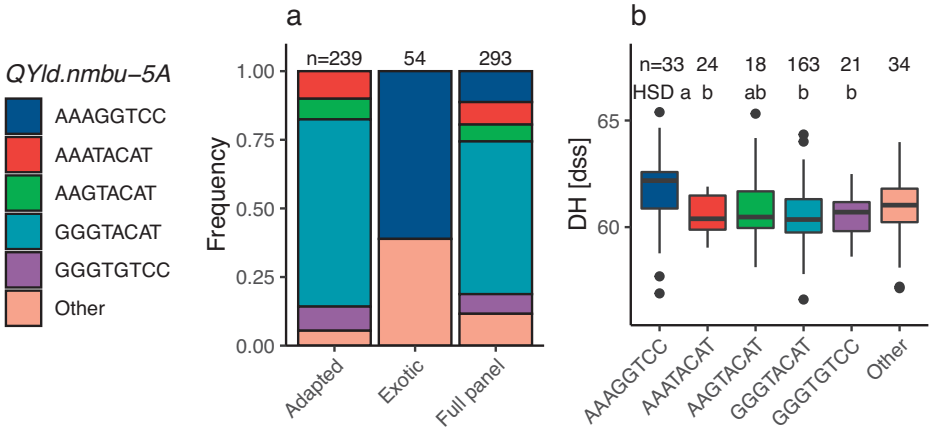


Figure S24 *Locus associated with adaptation to the Norwegian growing conditions: QYld.nmbu-5A. Allele frequencies in the full panel and adapted and exotic parts (a), haplotype analysis of the loci (b) on days to heading (DH)). Comparison among the alleles was performed using Tukey's HSD test. Alleles with the same letter are not significantly different ($\alpha = 0.05$). Alleles with low frequencies (in less than ten lines) were gathered into the "Other" bin*

Paper III

Multispectral-derived genotypic similarities from budget cameras allow grain yield prediction and genomic selection augmentation in single and multi-environment scenarios in spring wheat

Manuscript

Tomasz Mróz[1], Sahameh Shafiee[1], Jose Crossa[23], Osval A. Montesinos-Lopez[4], and Morten Lillemo[1]

[1]*Department of Plant Sciences, Norwegian University of Life Sciences, NO-1432 Ås, Norway*

[2]*International Maize and Wheat Improvement Center (CIMMYT), Km 45, Carretera Mexico Veracruz, CP 52640 Texcoco, Edo. de México, Mexico*

[3]*Colegio de Postgraduados, CP 56230 Montecillos, Edo. de México, Mexico*

[4]*Facultad de Telemática, Universidad de Colima, Colima, Mexico*

Corresponding author:

Morten Lillemo
morten.lillemo@nmbu.no
Orcid-ID: 0000-0002-8594-8794

Core ideas

1. Genotypic relationships derived using budget high-throughput multispectral data (the M matrix) allow for grain yield prediction in wheat
2. Using BLUP, the M matrix yields comparable accuracy in grain yield prediction to the G matrix
3. Adding the M matrix to GBLUP genomic prediction protocol increases its accuracy and reduces the error
4. Genotypic relationships derived using multispectral data (the M matrix) are not environment-specific, and the relationships become more robust with more data capture sessions
5. Data capture during the grain filling stage yields the best prediction accuracy using the M matrix
6. Grain yield prediction using the M matrix is also possible using a simple RGB camera, with a slight drop in accuracy compared to a multispectral camera

Author contributions

TM: Conceptualization, Methodology, Software, Formal analysis, Resources, Data curation, Writing – original draft, Visualization
SS: Resources, Data curation
JC: Writing – review & editing, Supervision
OAM: Writing – review & editing
ML: Writing – review & editing, Supervision, Project administration, Funding acquisition

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors acknowledge: Cecilie Yri and Svend Pung for excellent trial management and sample processing, students Henrik Lassegård, Lavanathan Rathy, Simon Pedersen, Guru Stordal and Peder Bukaasen for help with image acquisition and processing. The funding for the main author was provided by the Norwegian University of Life Sciences, Ås, Norway. The field trials were conducted in the framework of vPheno and PhenoCrop projects, funded by NFR.

Abbreviations

BLUE – best linear unbiased estimator
BLUP - best linear unbiased predictor
G (matrix) – genomic relationship matrix
GBLUP – genomic best linear unbiased predictor
GEBV – genomically-estimated breeding value
GS – genomic selection
GY – grain yield
HTP – high-throughput phenotyping
M (matrix) – multispectrally-derived relationship matrix
MAF – minor allele frequency
MS – multispectral (in relation to cameras)
NDVI – normalized differential vegetation index
NIR – near infrared
PS – phenomic selection
RGB – red, green, blue (camera bands)
UAV – unmanned aerial vehicle, aka drone
UV – ultraviolet
VI – vegetation index

Abstract

Using kinships and similarities among individuals forms the backbone of quantitative genetics. With abundant available genomic data, genomic selection has become routine in many plant breeding programs. Multispectral data captured by UAVs showed potential for grain yield prediction in many plant species using machine learning; however, the possibilities of utilizing this data to augment genomic prediction models still need to be explored. We collected HTP multispectral data in a genotyped multi-environment large-scale field trial using two cost-effective cameras to fill this gap. We tested back-to-back the prediction ability of GY prediction models, including genomic (G matrix), multispectral-derived (M matrix), and environmental (E matrix) relationships using BLUP methodology in single and multi-environment scenarios. We discovered that M allows for GY prediction comparable to the G matrix and that models using both G and M matrices show superior accuracies and errors compared with G or M alone, both in single and multi-environment scenarios. We showed that the M matrix is not entirely environment-specific, and the genotypic relationships become more robust with more data capture sessions over the season. We discovered that the optimal time for data capture occurs during grain filling and that camera bands with the highest heritability are important for GY prediction using the M matrix. We showcased that GY prediction can be performed using only an RGB camera, and even as little as a single data capture session can yield valuable data for GY prediction. This study contributes to a better understanding of multispectral data and its relationships. It provides a flexible framework for improving GS protocols at a low cost without significant investments or software customization.

Introduction

To develop new plant varieties and cultivars, breeders initially relied solely on recorded phenotypes of candidates paired with "the breeder's eye". With increasing pressure caused by climate change, increasing world population, and diminishing arable land, developing new and adapted germplasm is more urgent than ever (Hickey *et al*., 2019). Nowadays, plant breeders have an abundance of new and innovative tools at their disposal to aid their quest for better-adapted germplasm, focusing on selection accuracy, breeding cycle shortening, and maximizing the genetic pool to be screened – therefore effectively accelerating genetic gains in all aspects of the breeder's equation (Houchmandzadeh, 2014).

Genomic selection (GS), first proposed by Meuwissen, Hayes, and Goddard (2001), aims to estimate breeding values (GEBVs) of individuals that have been genotyped-but not phenotyped—based on prediction equations developed from a collection of phenotyped and genotyped individuals. New parents for crossing are then selected based on GEBVs, which shortens the breeding cycle since late filial generations do not need to be phenotyped for quantitative traits such as grain yield (Bassi *et al*., 2015). Due to the cost reduction of genotyping and well-elaborated methodologies, GS has become routine in many breeding programs (Bhat *et al*., 2016).

With abundant genomic data, plant phenotype registering became a bottleneck in plant research and breeding, stimulating the development of high-throughput phenotyping (HTP) methodologies. HTP involves automating the evaluation of plant phenotypes and was enabled by recent advancements and the popularization of sensor and computing technologies paired with data analytics (White *et al*., 2012); it allows to cover large numbers of genotypes in a fraction of the time needed for manual measurements (Araus & Cairns, 2014; Burud *et al*., 2017). HTP has shown considerable potential by enabling grain yield prediction using machine learning, as reviewed by van Klompenburg, Kassahun, and Catal (2020). HTP data has proven useful also in predicting above-ground biomass (Han *et al*., 2019; Lu *et al*., 2019; Li *et al*., 2020), plant height (Hu *et al*., 2018; Hassan *et al*., 2019; Tirado, Hirsch and Springer, 2020), earliness (Zhou *et al*., 2019; Trevisan et al., 2020), and crop emergence (Li *et al*., 2019) to name but a few.

A specific branch of HTP uses unmanned aerial vehicles (UAVs) equipped with multispectral or hyperspectral cameras, which record light spectrum above and beyond the visible spectrum. The usefulness of recording wavelengths outside the visible spectrum lies in their link with various aspects of crop physiology or

chemistry. For instance, near-infrared (NIR, 760-1400nm) is linked to crop water status; RedEdge (around 730nm) is arguably a proxy of chlorophyll content (Peñuelas & Filella, 1998); and Ultra-Violet A (UV-A, 200-380nm) can be used to monitor stress in plants (Brugger *et al*., 2019). This extra information can help construct vegetation indices (VIs), which are linear combinations of reflectance values such as NDVI (normalized difference vegetation index, Beisel *et al*., 2018) and, in turn, can be used for primary trait prediction (Montesinos-López *et al*., 2017; Shafiee *et al*., 2021).

HTP data gathered using multispectral and hyperspectral cameras has also been used to improve the accuracy of GS, as first demonstrated by (Rutkoski *et al*., 2016), where secondary VIs increased grain yield prediction accuracy by 70%. HTP can help measure genetically correlated secondary traits, which can be introduced into multivariate prediction models (Sun *et al*., 2017; Sakurai *et al*., 2022). Likewise, HTP data was also discovered to help evaluate genetic resources for the expression of complex traits (Reynolds & Langridge, 2016). In a recent study, NIR spectra of grain samples were used to construct spectral relationship matrices to enable phenomic selection (PS) and to aid GS, showing that the hyperspectral matrix-aided best linear unbiased prediction (H-BLUP) model performed at least as well as the standard genomic best linear unbiased prediction (GBLUP) model. A model combining both spectral and genomic information (GHBLUP) was superior to both G and HBLUP alone (Robert *et al*., 2022), showing similar results to (Krause *et al*., 2019). The PS based on the NIR spectra was also a promising, low-cost alternative to genotyping and a viable approach for predicting complex traits in perennial species such as grapevine (Brault *et al*., 2022). However, those approaches use expensive hardware like hyperspectral cameras or spectrometers, which limits their applications in practical breeding. To the authors' knowledge, no attempt has been made to utilize genetic relationships derived from low-cost multispectral imagery for grain yield prediction in wheat and augmenting GS protocols.

To fill this gap, we deployed HTP in a multi-environment spring wheat trial using two cost-effective multispectral cameras mounted on commercial UAVs. We tested various back-to-back grain yield prediction models using genomic (G) and multispectral (M) relationships combined with environment-specific phenotypical covariates. We investigated the applicability and flexibility of environment-specific M relationships in single and multi-environment scenarios and their synergy with the GS-GBLUP model. As such, the main objectives of this study were to:

1. Investigate the prediction ability of multispectral-derived genetic relationships for grain yield in single and multi-environment scenarios

2. Verify the possibility of augmenting GS with multispectral-derived genetic relationships
3. Study which multispectral band(s) are the most important for grain yield prediction
4. Examine the most informative data capture time for grain yield prediction under Norwegian growing conditions

Materials and methods

Plant material

The Norwegian University of Life Sciences (NMBU) spring wheat panel, consisting of 301 hexaploid spring wheat cultivars and breeding lines, was used for the study. The same panel was recently used for genetic analyses of Fusarium head blight (Nannuru *et al*., 2022) and Septoria nodorum blotch (Lin *et al*., 2022) resistance. The collection encompasses 186 Norwegian, 40 Swedish, and 37 lines from CIMMYT, with several additional lines from Australia, Brazil, Canada, Czech Republic, Denmark, Finland, France, Germany, Netherlands, Poland, Russia, Slovakia, South Africa, Switzerland, UK, and the USA. The whole set presents a broad genetic and phenotypic diversity.

Field trials

Trials were carried out during field seasons 2015-2022 between April and August in Vollebekk Research Station (Norway, Ås, $59^0$39'N, $10^0$45'E) and Staur farm (Norway, Stange, $60^0$43'N, $11^0$06'E), which represent the two principal economically important wheat growing areas in Norway due to the somewhat warmer and milder climate of south-eastern Norway and the slightly colder and temperate climate of inland Norway.

The trials were fertilized at sowing with 120kg·ha^{-1} of compound NPK fertilizer (YaraMila 22-3-10) and planted each season in both locations in late April or early May (exact planting dates in Table S3). Following germination, trials were kept disease and weed free according to local management practices using herbicides (Tripali [active ingredients: florasulam + metsulfuron-methyl + tribenuron-methyl] and Duplosan Meko [mekoprop]) and fungicides (Proline [prothioconazole], Aviator Xpro [bixafen + prothioconazole], Forbel [fenpropimorph] and Comet Pro [pyraklostrobin]) in doses tailored to the needs. Irrigation was applied in case of drought that could affect the growth of the plants. Alleys within the trials were created by spraying Glyphosate shortly after seedling emergence. The trials were

harvested each season towards the end of August after all varieties had reached full ripeness.

Field trial design

The trials were designed as an alpha-lattice with two replicates per genotype and a block size of 6 with positions of every accession randomized each year. Each column was planted with buffer variety at its start and end to eliminate border effects. Each field trial plot was 5 x 1.5mm in size at harvest, with gaps between the plots of 30 cm and a central alley of 1m. For the main panel, not every variety was tested in each season/location, and the number of genotypes tested varied from 100 to 295 per season/location.

Grain yield data

Grain yield was measured in two locations over seven field seasons (a total of 11 environments - year/location combinations): Vollebekk Research Station in 2015, 2016, 2017, 2019, 2020, 2021, and 2022; Staur Farm: 2016, 2017, 2019, and 2020.

Grain yield (GY) was measured by harvesting and threshing the trial plots, drying the yield until approximately 13.5% moisture, weighing it, and recalculating it to g per square meter.

Data for plots lodged early was removed due to the heavy impact on their development. If lodging occurred late in the season (close to maturity), data were double-checked for consistency and possible impact on the traits.

Statistical analysis of the field trial data

For GY, three types of genotypic means (lsmeans) were calculated: location/season (environment) mean; location mean (all seasons from one location); and a global mean (all the locations and seasons were combined).

As it was not uncommon to observe extra spatial variability within the trials (due to soil gradients) that was not fully captured by blocking, an additional covariate was introduced (columns) into the models to correct for it. The lsmeans were calculated using packages ‘‘lme4” and ‘‘lmerTEST” and custom scripts in R, version 4.2.1.

Environment (field trial) lsmeans were calculated using the mixed model (1):

$$P_{ilmn} = \mu + g_i + R_l + R{:}\,B_{lm} + C_n + e_{ilmn} \quad (1)$$

Where P_{ilmn} denotes the response variable measured in the ith genotype, lth replication, mth block and nth column, μ denotes a general mean or intercept, g_i denotes the fixed effect of genotype i, with $i = 1, \ldots, I$, R_l denotes the random effect of

replication effect distributed as normal with mean zero and variance σ_R^2, that is, $R_l \sim N(0, \sigma_R^2)$, $R{:}\,B_{lm}$ denotes the random effect of block m nested in replication l, also normally distributed as $R{:}\,B_{lm} \sim N(0, \sigma_{b(R)}^2)$, C_n denotes the random effect of column effect distributed as normal with mean zero and variance σ_C^2, that is, $C_n \sim N(0, \sigma_C^2)$, e_{ilmn} is the error random term normally distributed as $e_{ilmn} \sim N(0, \sigma^2)$.

Cross-season lsmeans for each location were calculated using the mixed model (2):

$$P_{iklmn} = \mu + g_i + Y_k + Y{:}\,R_{kl} + Y{:}\,R{:}\,B_{klm} + Y{:}\,C_{kn} + e_{iklmn} \quad (2)$$

Where P_{iklmn} denotes the response variable measured in the ith genotype, kth year, lth replication, mth block and nth column. Y_k denotes the random effect of year effect normally distributed as $Y_k \sim N(0, \sigma_Y^2)$, $Y{:}\,R_{kl}$ denotes the random effect of replication l nested in year k, also normally distributed as $Y{:}\,R_{kl} \sim N(0, \sigma_{R(Y)}^2)$, $Y{:}\,R{:}\,B_{klm}$ denotes the random effect of block m nested in replication l nested in year k, also normally distributed as $Y{:}\,R{:}\,B_{klm} \sim N(0, \sigma_{b(Y\times R)}^2)$, $Y{:}\,C_{kn}$ denotes the random effect of column n nested in year k, also normally distributed as $Y{:}\,C_{kn} \sim N(0, \sigma_{C(Y)}^2)$, and e_{iklmn} is the error random term normally distributed as $e_{iklmn} \sim N(0, \sigma^2)$.

Global mean (cross-season, cross-location) was calculated using the mixed model (3):

$$P_{ijklmn} = \mu + g_i + L_j + L{:}\,Y_{jk} + Y{:}\,L{:}\,R_{jkl} + Y{:}\,L{:}\,R{:}\,B_{jklm} + Y{:}\,L{:}\,C_{jkn} + e_{ijklmn} \quad (3)$$

Where P_{ijklmn} denotes the response variable measured in the ith genotype, jth location, kth year, lth replication, mth block and nth column. L_j denotes the random effect of location effect normally distributed as $L_j \sim N(0, \sigma_L^2)$, $L{:}\,Y_{jk}$ denotes the random effect of location j nested in year k normally distributed as $L{:}\,Y_{jk} \sim N(0, \sigma_{L(Y)}^2)$, $Y{:}\,L{:}\,R_{jkl}$ denotes the random effect of replication l nested in location j nested in year k normally distributed as $Y{:}\,L{:}\,R_{jkl} \sim N(0, \sigma_{R(Y\times L)}^2)$, $Y{:}\,L{:}\,R{:}\,B_{jklm}$ denotes the random effect of block m, nested in replication l nested in location j nested in year k normally distributed as $Y{:}\,L{:}\,R{:}\,B_{jklm} \sim N(0, \sigma_{b(Y\times L\times R)}^2)$, $Y{:}\,L{:}\,C_{jkn}$ denotes the random effect of column n nested in location j nested in year k normally distributed as $Y{:}\,L{:}\,C_{jkn} \sim N(0, \sigma_{C(Y\times L)}^2)$, e_{ijklmn} is the error random term normally distributed as $e_{ijklmn} \sim N(0, \sigma^2)$.

In the single-environment scenario (section Model performance assessment), environment (trial), location, and global means were used. In the multi-environment scenario (section Model performance assessment), only environment (trial) means were used. Broad-sense heritability (H^2) was calculated for individual trials using equation (4):

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_e^2} \quad (4)$$

where σ_g^2 is the genotypic variance and σ_e^2 is the error variance. Variance components for equation (4) were estimated using package "lme4" using the just described models but assuming the lines (genotypes) as normally distributed with mean zero and variance σ_G^2.

Genotyping data

Samples were prepared and genotyped as described in Nannuru *et al*. (2022).

Physical positions of the markers were determined using the chip's documentation and markers which weren't mapped to any physical chromosome position were placed on a fictional chromosome Un.

Markers were filtered, leaving only those with less than 10% missing data and minor allele frequency (MAF) larger than 0.05. Heterozygous markers were treated as missing data. After the quality check, the dataset contained 19874 high quality markers mapped to sub genomes A (7999), B (7905) and D (2111) on chromosomes 1A (1156), 1B (1147), 1D (391), 2A (1232), 2B (1377), 2D (437) 3A (1074), 3B (1336), 3D (256), 4A (699), 4B (602), 4D (111), 5A (1340), 5B (1406), 5D (311), 6A (1126), 6B (1082), 6D (319), 7A (1372), 7B (955), 7D (285), and Un (1859).

High-throughput phenotyping data

High-throughput phenotyping data were captured using two cameras (Micasense RedEdgeM https://micasense.com and DJI Phantom 4 Multispectral camera https://www.dji.com/p4-multispectral). In both locations, the RedEdgeM camera was used during field seasons 2019-2021, whereas Phantom 4 Multispectral was used during field season 2021 in Vollebekk Research Farm.

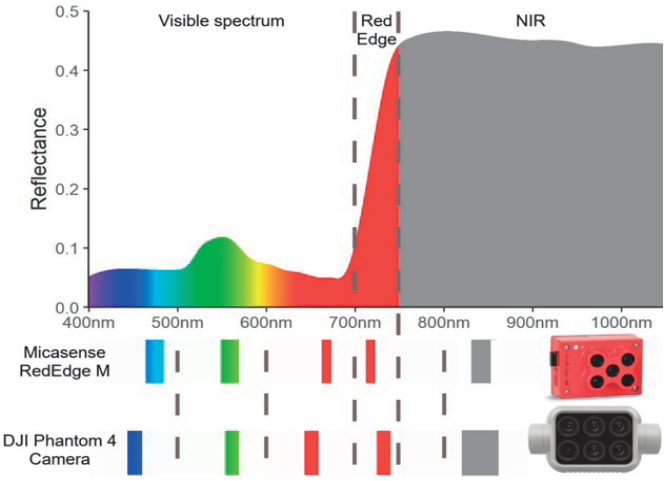
Detailed UAV specifications and the HTP data capture and processing description can be found in the Supplementary material.

High-throughput phenotyping data consisting of five color bands (red, green, blue, near-infrared, and red edge) was available for three field seasons in the two locations throughout the vegetation period, however, with varying temporal resolution: from 4 to 22 missions (Table 1).

Table 1 *HTP mission overview: number of data capture sessions for each season, camera, and location*

Season	Camera and Location			
	RedEdgeM		P4M	
	Vollebekk	Staur	Vollebekk	Staur
2019	7	4	-	-
2020	12	-	-	-
2021	8	-	22	-

The two cameras are fundamentally different regarding resolution and bandwidths/central bands (Figure 1), so they were analyzed separately. Only raw canopy reflectance values (red, green, blue, NIR, and RedEdge) were used for every part of the analysis, without calculating multispectral indices.



	P4 Multispectral	Micasense RedEdge M
Red	650 ± 16 nm	668 ± 5 nm
Green	560 ± 16 nm	560 ± 10 nm
Blue	450 ± 16 nm	475 ± 10 nm
Red Edge	730 ± 16 nm	717 ± 5 nm
NIR	840 ± 26 nm	840 ± 20 nm

Figure 1 *Top: typical plant canopy reflectance spectrum with graphical interpretation of light spectrum wavelengths; Bottom: visual interpretation and numeric values of central bands and bandwidths for the two tested cameras: Micasense RedEdgeM, and Phantom 4 (P4) Multispectral*

Analyzed models

A number of models described below were developed and tested using R package "lme4GS" (Caamal-Pat *et al.*, 2021) in R version 4.2.1.

1. G - Single-environment genomic prediction

To benchmark single-environment analysis, genomic prediction using G (genomic kinship) matrix (GBLUP, according to VanRaden, 2008) was performed in single-environment scenario (section Model performance assessment). G was calculated according to Equation 5:

$$G = \frac{WW'}{n} \tag{5}$$

Where *n* is the number of genotypes, *G* is the square genomic relationship matrix with *n* rows and *n* columns corresponding to the genotypes; W is a scaled (mean = 0, standard deviation = 1) matrix of SNP marker data with *n* rows and *m* columns (which equals number of quality checked markers, coded as 0s and 2s); and W' is its transpose.

For every environment (location/season combination), a random effect model was fitted using G as the definition of variance/covariance structure among the genotypes according to Model 1:

$$y = \mu 1 + g + e \qquad (Model\ 1)$$

Where y is the vector of lsmeans for a trait for n genotypes, μ is the intercept, 1 is a vector of ones, g is the vector of random genotypes effects distributed as $g \sim N(0, G\sigma_g^2)$ and e is the vector of residual effects distributed as $e \sim N(0, I\sigma^2)$.

Model 1 was trained and tested on environment (field trial), location, and global means.

2. G - Multi-environment genomic prediction

To benchmark multi-environment prediction using the G matrix, Model 1 was used in the multi-environment scenario (section Model performance assessment), using only environment means.

3. G+E - Multi-environment genomic prediction with environment covariance (E) matrix

To benchmark multi-environment predictions using the G matrix coupled with the environmental (phenotypical) variance/covariance matrix K_E, genomic prediction supplemented with K_E matrix was analyzed in a multi-environment scenario (section Model performance assessment). For this purpose, only environment means were used.

The K_E matrix was computed for grain yield according to equation 6:

$$K_E = \frac{PP'}{n} \quad (6)$$

Where n is the number of environments (environment/season combinations), K_E is the square environmental (phenotypical) variance/covariance matrix for grain yield of dimensions $n \times n$, P is a scaled rectangular matrix with n rows and m columns (representing scaled phenotype values for every genotype for every environment in rows) and $P'$is its transpose.

Using G and E, Model 2 was fitted:

$$y = \mu 1 + E + g + e \quad (Model\ 2)$$

All the terms of Model 2 are equal to Model 1 except for E – vector of random environment effects, $E \sim N(0, K_E \sigma_E^2)$.

4. M - Single-environment prediction using image-derived M matrix

For every season, based on least square mean values for every available raw band for each flight date and each genotype, a multispectral relationship matrix was computed according to Equation 7 and analogically to G and E matrices and similar to the work of Krause *et al.* (2019):

$$K_M = \frac{CC'}{n} \quad (7)$$

Where n is the number of genotypes, K_M is the multispectral variance/covariance matrix of dimensions $n \times n$ in a particular season, C is a scaled rectangular matrix with n rows and number of columns corresponding to genotypic BLUE reflectance values for each multispectral band at every flight within the season, and C' is its transpose.

As the reflectance values are assumably environment-specific, the K_M matrix was computed for each environment (season/location combination) separately, with no attempt to calculate a cross-environment K_M matrix.

Using the derived K_M matrix, an analogical analysis to single-environment genomic prediction was conducted by replacing G with K_M matrix in a single-environment scenario (section Model performance assessment) and fitting Model 3:

$$y = \mu 1 + g_* + e \quad (Model\ 3)$$

Where g_* is the vector of random genotypes effects distributed as $g \sim N\left(0, \boldsymbol{K_M \sigma^2_{g_*}}\right)$. Each K_M matrix (developed based on different environment data) was trained and tested on environment, location, and global means.

5. Multi-environment prediction using M matrix

To assess whether K_M matrix derived based on data from a single season possesses prediction ability in other environments (if it is environment-specific), an analogical analysis was carried out in a multi-environment scenario (section Model performance assessment) by replacing G with K_M in Model 1. Each environments' K_M matrix was tested for its multi-environment prediction ability using only environmental grain yield means. There was no attempt to calculate a cross-environment K_M because of the assumed environment-specificity of multispectral data.

6. Multi-environment prediction using M and E matrices

An analogical model to Model 2 was tested in a multi-environment scenario (section Model performance assessment) by replacing G with the K_M matrix. For this purpose, only environment (trial) means were used.

7. G+M - Single-season genomic prediction supplemented with M matrix

To assess the prospect of supplementing genomic prediction models with multispectral data, a Model 4 combining both G and K_M matrices was fitted for a single-environment scenario (section Model performance assessment):

$$y = \mu 1 + g + g_* + e \qquad (Model\ 4)$$

With terms identical as in Models 1 and 3. For the purpose, environment, location, and global GY means were used.

8. G+M - Multi-environment genomic prediction supplemented with M matrix

To evaluate the combined prediction ability of the G and K_M matrices, Model 4 was tested in the multi-environment scenario (section Model performance assessment) using only environment GY means.

9. G+M+E - Multi-environment genomic prediction supplemented with M and E matrices

To further evaluate the combined prediction ability of G and K_M matrices in multi-environmental scenario (section Model performance assessment), Model 5 was developed using G, M, and E matrices simultaneously:

$$y = \mu 1 + E + g + g_* + e \qquad (Model\ 5)$$

With terms identical as in the previous models. For this purpose, only environment (trial) means were used.

Model performance assessment

The models' performance was analyzed under two scenarios (described below): single (G, M, and G+M) and multi-environment (G, M, G+E, G+M, M+E, and G+M+E). The assessment was performed using the following metrics:

rTRN – prediction accuracy in the training set (in the dataset used to develop the model), defined as the Pearson correlation coefficient between predicted and observed values.

rTST - prediction accuracy in the testing set (the dataset not seen previously by the model), defined as the Pearson correlation coefficient between predicted and observed values.

rmseTRN – root mean squared error in the training set, defined as:

$$rmse_{TRN} = \sqrt{\frac{\sum_{n=1}^{N}(obs_{TRN} - pred_{TRN})^2}{N}}$$

Where obs_{TRN} are observed (ground truth) phenotypes, $pred_{TRN}$ are predicted phenotypes (output from the models), and N is the number of records (genotypes) in the training set.

rmseTST - root mean squared error in the test set (previously unseen data), defined as:

$$rmse_{TST} = \sqrt{\frac{\sum_{n=1}^{N}(obs_{TST} - pred_{TST})^2}{N}}$$

Where obs_{TST} are observed (ground truth) phenotypes, $pred_{TST}$ are predicted phenotypes (output from the models), and N is the number of records (genotypes) in the testing set.

The models were tested using cross-validation with 200 iterations in two scenarios:

Single-environment: the training set consisted of 80% of genotypes available in the respective environment/mean (20% as testing set). Genotypes were randomly assigned to training/test sets at every iteration.

Multi-environment: the testing set consisted of 20% of all the available genotypes in two environments not used for training the model. The training set comprised 80% of all the available genotypes in the remaining environments (9). Therefore, the testing set was double-blind: comprised of both environments and genotypes not

used for model training. Both genotypes and environments were randomly assigned to training/testing sets at every iteration.

The importance of camera bands for grain yield prediction

Model 3 was tested in a single-environment scenario with M matrices constructed based on all flight times with only a single camera band at a time (red, green, blue, RedEdge, and NIR) to verify the importance of particular camera bands for grain yield prediction using the M matrix.

The importance of timing of data capture

Model 3 was tested in a single-environment scenario with M matrices constructed on all camera bands but with only one date at a time to verify the effect of time of data capture on grain yield prediction accuracy.

Minimal setup for grain yield prediction

Based on the results mentioned in the previous paragraphs, a concept of minimal setup for grain yield prediction was formed: a single flight mission taken during July (grain filling stage). This concept was developed for multispectral cameras (with five bands) and a simple RGB camera (3 bands, red, green, and blue). The RGB camera was "simulated" using only three bands (out of the five available bands) for constructing M matrices.

Model 3 was tested in the single-environment scenario, constructing M matrices based on a random flight date in July in each environment with five (multispectral camera) or three (RGB camera) bands.

Results

Phenotypic data evaluation – grain yield

Mean genotypic GY values across all environments (year and location combinations) are similar (approximately 520 g m^{-2}), except for a field experiment in Staur in 2017 when the average GY value reached 789 g m^{-2}. The global mean is influenced mainly by trials conducted in Vollebekk and resembles the distribution of the Vollebekk environmental mean. The environmental mean in Staur is higher than the Vollebekk means by 70 g m^{-2}. In all environments and means, a long left tail can be observed in the distributions (Figure 2).

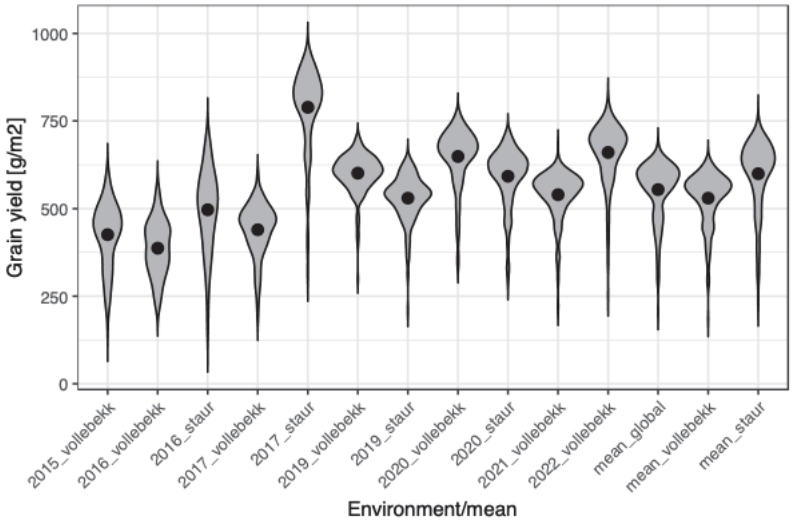


Figure 2 *Violin plot of distributions of grain yield genotypical means in each of the studied environments and means: global means (across all studied environments) and location specific means (across all environments within one location: Staur or Vollebekk). Black dots indicate mean values*

Table 2 *Broad sense heritability (H^2) of grain yield in each environment and number of genotypes (n genotypes) tested in each environment (field trial)*

Environment	n genotypes	H^2
2015 Vollebekk	157	0.86
2016 Staur	100	0.55
2016 Vollebekk	98	0.58
2017 Staur	240	0.46
2017 Vollebekk	240	0.71
2019 Staur	220	0.50
2019 Vollebekk	220	0.68
2020 Staur	288	0.51
2020 Vollebekk	288	0.57
2021 Vollebekk	293	0.72
2022 Vollebekk	296	0.82

Across the field trials (environments), broad-sense heritability for GY ranged between 0.46 (Staur 2017) to 0.86 (Vollebekk 2015) and the number of tested genotypes varied between 98 and 296 (Table 2).

Field trials (environments) and means were, on average, highly correlated (r = 0.77). The field trial from Vollebekk in 2015 is the most different from the remaining trials and means, with r ranging from 0.32 (with Staur 2019) to 0.64 (with Vollebekk 2019) and 0.67 with the Vollebekk environmental mean. The location means resemble more

recent trials (2019 onwards), which can also be observed for the global mean (Figure 3).

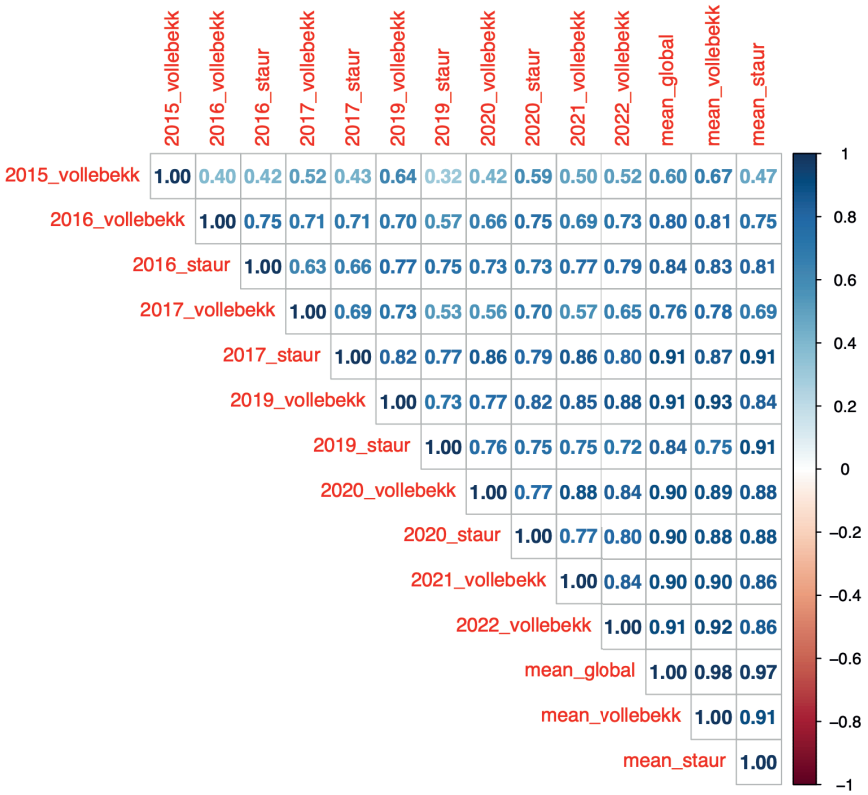


Figure 3 *Genotypic Pearson correlations for grain yield values among field trials (environments), their means, and the global mean*

High-throughput phenotyping data evaluation

Raw reflectance values for each band over field seasons in each environment are shown on Figure S3.

Broad-sense heritability of each band changed during the season with no apparent consistent trend; however, heritability values tended to be more stable later in the growing season (from July onwards). NIR and red were the least heritable bands, while RedEdge, green, and blue had higher heritability values. It was not uncommon to observe that during the same mission, different bands had very different heritability (Figure 4).

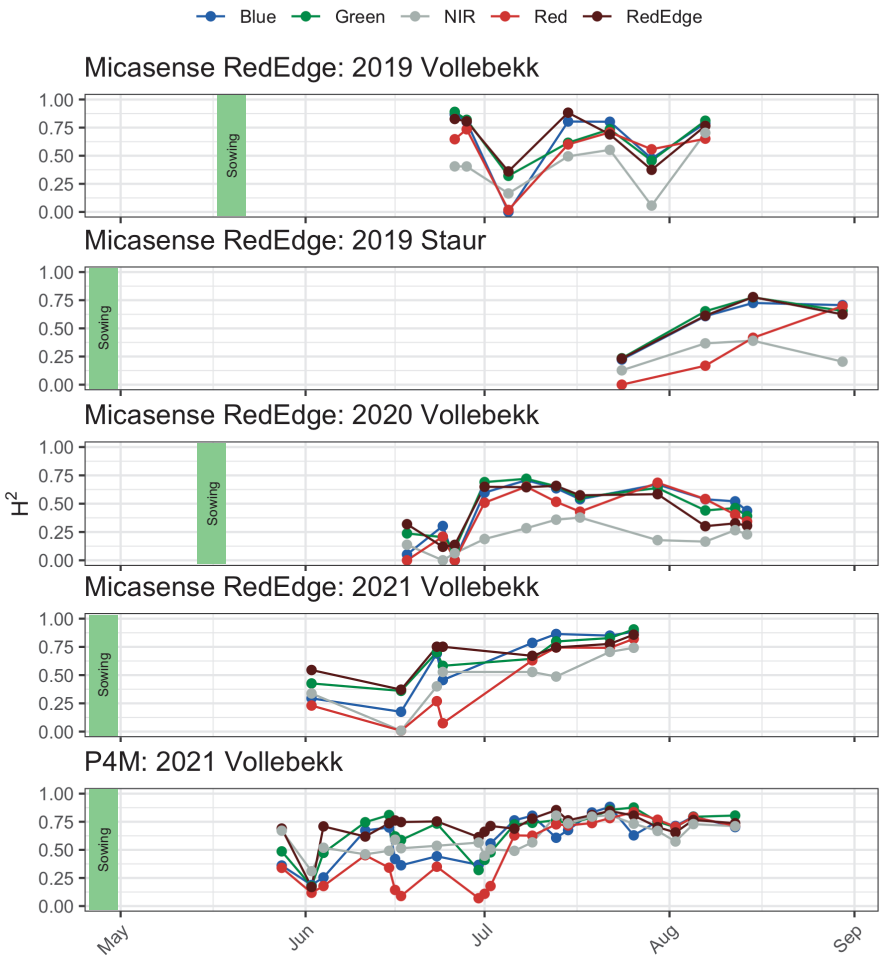


Figure 4 *Broad-sense heritability estimates for each band, in each environment, and for each flight and camera. The X axis – data capture date (flight date), Y axis – broad sense heritability. Line colors correspond to the bands they represent (RGB), grey – NIR, dark red – RedEdge*

Evaluation of single environment prediction using G matrix

Grain yield prediction using the G matrix in single-environment scenarios (model trained and validated on a single season) consistently showed high accuracy in the training set (on average 0.99). In contrast, accuracies in testing sets ranged from 0.59 to 0.81 in individual field trials, averaging to 0.75. In both location and global mean, where the genetic signal is stronger, testing accuracies (rTST) were higher than in the individual trials. Root mean squared error (rmse) in the testing set was

approximately four times higher than in the training set (53.0 and 13.9 for testing and training sets, respectively, Table 3).

Table 3 *Comparison of grain yield prediction accuracy and root mean squared error for G matrix in a single season scenario (models built and verified on a single environment) using cross-validation with 200 iterations. rTRN – accuracy in the training set, rTST – accuracy in the testing set, rmseTRN – root mean squared error in the training set, rmseTST – root mean squared error in the testing set*

Environment/mean.	G matrix			
	rTRN	rTST	rmseTRN	rmseTST
2015_V	0.99	0.69	15.7	64.9
2016_S	0.99	0.75	14.9	73.6
2016_V	0.98	0.72	14.7	52.0
2017_S	0.99	0.81	19.7	69.7
2017_V	0.98	0.70	16.2	52.4
2019_S	0.95	0.59	23.6	51.9
2019_V	0.98	0.63	11.1	40.7
2020_S	0.99	0.77	12.6	51.9
2020_V	0.98	0.75	17.9	53.5
2021_V	0.99	0.75	9.9	46.0
2022_V	0.99	0.79	14.4	51.5
mean_global	1.00	0.86	6.3	40.2
mean_S	0.99	0.83	14.5	54.4
mean_V	1.00	0.85	3.6	39.1
Avg.	**0.99**	**0.75**	**13.9**	**53.0**

Evaluation of single trait, single environment prediction using M matrix

M matrices showed the highest prediction ability on the environment they originated from; however, they often retained prediction ability when tested on other environments, especially those highly correlated with their environment of origin. M matrices developed in seasons 2019 and 2020 showed poor prediction ability in 2015 Vollebekk, 2016 Vollebekk, and 2017 Vollebekk due to low correlations with those environments. M matrices developed with data from 2021 Vollebekk (using both cameras) showed decent prediction abilities across all the tested environments, even in environments not strongly correlated with the M matrix's origin (Figure 3). Prediction accuracy was high (>0.5) for the global, and location means for all the M matrices (Figure 5).

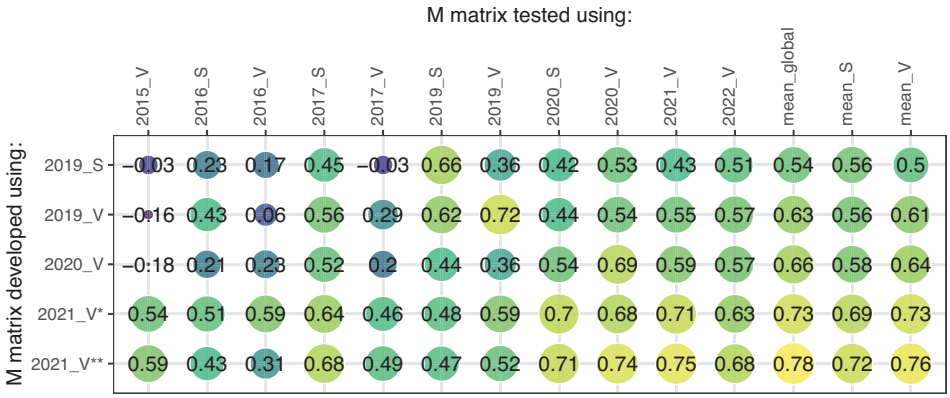


Figure 5 *Prediction accuracy achieved in single-environment scenarios by using M matrix constructed on multispectral data from each environment with multispectral data available. The y axis indicates from where the environment multispectral data originated, while the x axis indicates the environment on which the model was trained and tested. Numbers and colors indicate the prediction accuracy in the testing set defined as Pearson correlation between predicted and actual values. * - data gathered using Micasense RedEdgeM camera, ** data gathered using Phantom 4 Multispectral camera. S – Staur, V – Vollebekk*

Evaluation of G, M and G+M models in single-environment scenarios

Genomic prediction accuracies in testing sets ranged from 0.59 to 0.75, averaging 0.68 in the chosen environments, while training set accuracies reached nearly perfect (0.98). Predictions using the M matrix in the single-environment scenario showed, on average, lower training set accuracies than predictions using the G matrix (0.79 and 0.98, respectively); however, testing set accuracies were higher than those of the G matrix (0.71 and 0.68, respectively). By comparing the difference between training and testing sets accuracies, the M matrix model was less prone to overfitting than the model using the G matrix (difference of 0.08 and 0.30 for M and G matrices, respectively) (Table 4).

Table 4 *Comparison among prediction performance of M matrices originating from different seasons, prediction using only the G matrix (genomic prediction), and a combined model utilizing both G and M matrices in the same model in a single-environment scenario. Models were developed and tested on single environment means. S – Staur, V – Vollebekk, rTRN – accuracy in the training set, rTST – accuracy in the testing set, rmseTRN – root mean squared error in the training set, rmseTST – root mean*

*squared error in the testing set. *M matrix developed using Micasense RedEdge M camera data, ** M matrix developed using Phantom 4 Multispectral camera data*

Model	Metric	Season/M matrix origin (if M present)					Avg.
		2019_S	2019_V	2020_V	2021_V*	2021_V**	
G	rTRN	0.95	0.98	0.98	0.99		**0.98**
	rTST	0.59	0.63	0.75	0.75		**0.68**
	rmseTRN	23.61	11.06	17.87	9.88		**15.61**
	rmseTST	51.88	40.72	53.51	45.95		**48.02**
M	rTRN	0.74	0.81	0.75	0.81	0.84	**0.79**
	rTST	0.66	0.73	0.69	0.71	0.75	**0.71**
	rmseTRN	43.20	31.37	53.62	40.24	38.07	**41.30**
	rmseTST	48.84	37.03	58.13	49.11	45.90	**47.80**
G+M	rTRN	0.96	0.98	0.99	1.00	1.00	**0.99**
	rTST	0.74	0.79	0.83	0.80	0.81	**0.79**
	rmseTRN	19.51	11.02	11.23	6.87	6.19	**10.96**
	rmseTST	42.54	32.33	45.76	41.00	40.96	**40.52**

Supplementing genomic prediction (G matrix) with the M matrix in a single-environment scenario yielded similar accuracy (0.71 for M versus 0.68 for G). The G+M model exhibited traits of both individual matrix models and performed better than either G or M alone: very high training accuracy, high testing accuracy, low training error, and low testing set error (Table 4).

Evaluation of G, M, G+M, G+E and G+M+E models in multi-environment scenarios

GY prediction using the G matrix alone in multi-environmental scenarios achieved accuracies of 0.57 and 0.49 in training and testing sets, respectively. Prediction accuracy in testing sets using M matrices originating from different environments ranged from 0.27 to 0.44, averaging 0.36. Replacing G with the M matrix resulted in a significant reduction of accuracy (difference in testing accuracy of 0.13) and a slightly larger degree of overfitting of the model (difference of accuracy in training-testing sets of 0.12) (Table 5).

Table 5 *Comparison of prediction ability of different models including combinations of G, E and M matrices in multi-environmental scenarios (two environments as testing sets, drawn randomly at every iteration). M matrices were developed based on data originating from different environments, and each M matrix has been tested individually on randomly selected test environments over 200 iterations. S – Staur, V – Vollebekk, rTRN – accuracy in the training set, rTST – accuracy in the testing set, rmseTRN – root mean squared error in the training set, rmseTST – root mean squared error in the testing set. *M matrix developed using Micasense RedEdge M camera data, ** M matrix developed using Phantom 4 Multispectral camera data*

Model	Metric	If M present, M matrix developed on data from:					Avg.
		2019_S	2019_V	2020_V	2021_V*	2021_V**	
G	rTRN	**0.57**					-
	rTST	**0.49**					-
	rmseTRN	**111.1**					-
	rmseTST	**117.0**					-
M	rTRN	0.43	0.40	0.54	0.52	0.52	**0.48**
	rTST	0.27	0.32	0.36	0.42	0.44	**0.36**
	rmseTRN	110.6	111.6	112.2	113.4	113.5	**112.3**
	rmseTST	117.0	115.0	124.2	121.7	119.7	**119.5**
G+E	rTRN	**0.95**					-
	rTST	**0.83**					-
	rmseTRN	**40.59**					-
	rmseTST	**58.16**					-
G+M	rTRN	0.46	0.46	0.56	0.57	0.57	**0.52**
	rTST	0.44	0.46	0.62	0.57	0.63	**0.56**
	rmseTRN	107.0	107.0	110.0	108.0	109.0	**108.0**
	rmseTST	121.0	121.0	117.0	124.0	121.0	**120.8**
M+E	rTRN	0.90	0.92	0.90	0.91	0.93	**0.91**
	rTST	0.70	0.78	0.73	0.77	0.79	**0.75**
	rmseTRN	50.8	45.9	56.7	53.7	49.0	**51.3**
	rmseTST	62.3	57.4	68.6	63.4	60.3	**62.4**
G+E+M	rTRN	0.95	0.95	0.95	0.95	0.95	**0.95**
	rTST	0.76	0.79	0.84	0.84	0.85	**0.82**
	rmseTRN	36.9	37.2	40.1	40.2	40.5	**39.0**
	rmseTST	54.9	50.9	56.2	54.8	53.0	**53.9**

Supplementing genomic prediction with a phenotypically-derived E matrix drastically increased the prediction accuracy in training and testing sets (0.95 and 0.83, respectively) and reduced the errors. The G+E model achieved the highest accuracy among all the tested models (Table 5).

Aiding genomic prediction with M matrices also increased prediction accuracy, albeit smaller than adding the E matrix (testing sets accuracy difference of 0.27 between G+E and G+M models). Prediction based on M matrices coupled with the E matrix achieved an accuracy comparable with genomic prediction aided by the E matrix (testing set accuracies of 0.75 and 0.83 for M+E and G+E models, respectively). The

M+E model was similar to the G+E model in its degree of overfitting (difference between training and testing set accuracies of 0.12 and 0.16, respectively) (Table 5).

The most complex model, utilizing G, E, and M matrices, achieved accuracy almost identical to the G+E model (testing set accuracies of 0.82 and 0.83, respectively); however, adding multispectral information resulted in more minor errors both in training and testing sets, as compared to the G+E model. M matrix originating from the 2021 Vollebekk environment (with the highest temporal density) paired with G and E matrices showed the highest accuracy in the prediction of GY (testing set accuracy 0.85) (Table 5).

Which camera bands are the most informative for grain yield prediction using M matrix?

In single-environment scenario, grain yield prediction using a constructed M matrix based on only one band reduced accuracy by 35% compared to the entire M matrix (average testing set accuracy for the individual bands of 0.46 compared to 0.71 for the entire M matrix, Table 4 and 5). On average, bands exhibited the following ranking (descending prediction in the test set accuracy): RedEdge, red ex aequo green and blue, and NIR; however, these differed slightly among the environments. The bands with the highest prediction abilities were RedEdge and the three "basic" bands (red, green, and blue). Contrastingly, The least informative band was consistently NIR (except for Vollebekk 2019, where it ranked 4), with high variability in the testing set prediction accuracy reaching as low as -0.22 in the Staur 2019 environment. The remaining bands were consistent in their prediction abilities (Table 6).

Bearing similarity to the single-environment scenarios (Table 6), M matrices developed on single bands had poor and reduced prediction ability in multi-environment scenarios by 41% (average testing set prediction accuracy of 0.21) compared to the entire M matrices (Table 5, 7). The average ranking of bands also bared similarity to the single-environment scenarios: RedEdge ex aequo red and green, blue and NIR.

Table 6 *Comparison of prediction ability of constructed M matrices based on a single band captured during a single season in a single-environment scenario (models trained and tested on a single environment, 80/20 genotypes train/test set). S – Staur, V – Vollebekk, rTRN – accuracy in the training set, rTST – accuracy in the testing set, rmseTRN – root mean squared error in the training set, rmseTST – root mean squared error in the testing set. *M matrix developed using Micasense RedEdge M camera data, ** M matrix developed using Phantom 4 Multispectral camera data*

Band	Metric	In season:					Avg.
		2019_S	2019_V	2020_V	2021_V*	2021_V**	
Red	rTRN	0.59	0.58	0.64	0.59	0.73	**0.63**
	rTST	0.55	0.47	0.60	0.44	0.63	**0.54**
	rmseTRN	52.29	44.2	62.6	55.95	47.2	**52.4**
	rmseTST	52.7	46.5	66.0	70.2	54.8	**58.0**
Green	rTRN	0.59	0.62	0.59	0.59	0.71	**0.62**
	rTST	0.55	0.56	0.53	0.54	0.65	**0.53**
	rmseTRN	52.4	42.4	65.6	56.0	48.6	**53.0**
	rmseTST	53.1	44.8	68.1	59.0	52.2	**55.4**
Blue	rTRN	0.45	0.38	0.60	0.57	0.69	**0.54**
	rTST	0.40	0.25	0.53	0.47	0.65	**0.46**
	rmseTRN	57.4	50.2	65.2	56.5	50.1	**55.9**
	rmseTST	59.6	51.9	67.7	63.6	52.9	**59.1**
RedEdge	rTRN	0.63	0.63	0.58	0.60	0.69	**0.63**
	rTST	0.60	0.58	0.53	0.56	0.61	**0.58**
	rmseTRN	50.2	42.0	66.3	54.9	49.7	**52.6**
	rmseTST	51.4	44.0	68.6	59.4	55.8	**55.8**
NIR	rTRN	0.82	0.45	0.42	0.44	0.66	**0.56**
	rTST	-0.22	0.35	0.21	0.13	0.55	**0.21**
	rmseTRN	30.7	48.4	73.0	58.0	51.8	**52.4**
	rmseTST	64.6	50.6	81.2	68.7	58.8	**64.8**

Table 7 *Comparison of prediction ability of constructed M matrices based on a single camera band in multi-environment scenarios (two environments as testing set, drawn randomly at every iteration). The development of M matrices was based on data originating from different environments, and each M matrix has been tested individually on randomly selected test environments over 200 iterations. S – Staur, V – Vollebekk, rTRN – accuracy in the training set, rTST – accuracy in the testing set, rmseTRN – root mean squared error in the training set, rmseTST – root mean squared error in the testing set. *M matrix developed using Micasense RedEdge M camera data, ** M matrix developed using Phantom 4 Multispectral camera data*

Band	Metric	M matrix developed on data from:					Avg.
		2019_S	2019_V	2020_V	2021_V*	2021_V**	
Red	rTRN	0.46	0.44	0.56	0.56	0.55	0.51
	rTST	0.21	0.24	0.24	0.25	0.35	0.26
	rmseTRN	109.6	110.3	111.3	111.4	112.1	111.0
	rmseTST	119.7	118.1	128.5	134.3	124.7	125.0

Table 7 *Cont.*

Green	rTRN	0.46	0.45	0.56	0.57	0.55	0.52
	rTST	0.22	0.23	0.23	0.25	0.35	0.26
	rmseTRN	109.3	110.0	111.4	110.9	111.7	111.0
	rmseTST	119.2	118.6	128.8	130.4	124.7	124.0
Blue	rTRN	0.48	0.48	0.57	0.58	0.57	0.53
	rTST	0.10	0.12	0.18	0.12	0.26	0.16
	rmseTRN	109.1	109.3	111.1	110.6	111.1	110.0
	rmseTST	121.0	121.4	131.7	133.1	129.2	127.0
RedEdge	rTRN	0.44	0.45	0.56	0.56	0.56	0.52
	rTST	0.24	0.25	0.24	0.26	0.33	0.26
	rmseTRN	110.1	110.0	111.27	111.15	111.5	111.0
	rmseTST	118.1	117.7	128.8	129.3	125.9	124.0
NIR	rTRN	0.48	0.47	0.57	0.58	0.56	0.53
	rTST	0.00	0.18	0.08	0.10	0.31	0.13
	rmseTRN	109.2	109.5	110.9	110.5	111.5	110.0
	rmseTST	122.8	120.3	132.7	133.3	127.2	127.0

Effect of multispectral data capture on grain yield prediction accuracy

GY prediction in single-environment scenarios was possible, with accuracies ranging from 0.17 to 0.68. Based on all the environments, data capture sessions late in the growing season (when plants approach physiological maturity, data not shown) tended to be less informative. At the same time, the prediction accuracy dropped further as maturing progressed. It is difficult to conclude the informativeness of early season flights due to the scarcity of available records; however, based on Vollebekk 2020 and 2021 environments, early season flights are more informative than flights taken later, until approximately the end of June. Data capture sessions carried out in July showed the highest accuracy in all the seasons with stable accuracies (Figure 6). Based on the 2021 Vollebekk environment, no significant differences in prediction accuracy could be observed between the two used cameras (Figure 6).

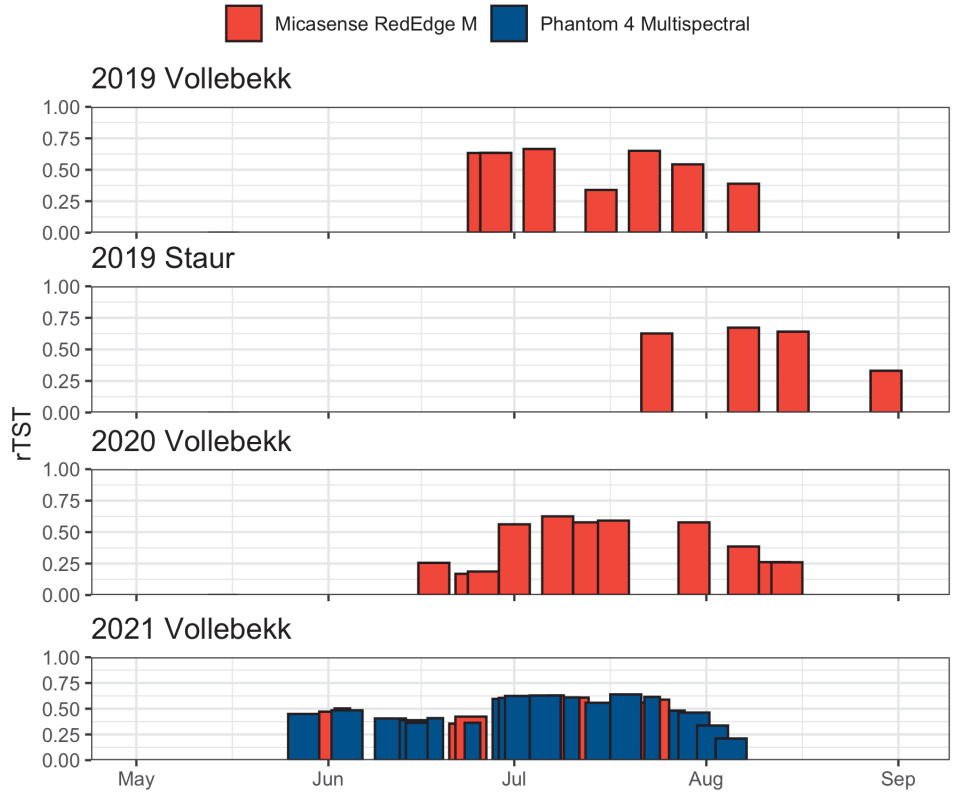


Figure 6 *Prediction accuracy using M matrix developed on all bands from a single date in single-environment scenario, expressed as Pearson correlation between predicted and actual values in testing sets. X axis – date of mission, y axis – prediction accuracy. Color indicates the camera used for data capture: red – Micasense RedEdgeM, blue – Phantom 4 Multispectral camera*

Minimal GY prediction setup in single-environment scenarios using M matrix

The prediction accuracy ranged from 0.51 to 0.58 and 0.55 to 0.62 for RGB and MS cameras, respectively - using MS instead of RGB cameras increased prediction accuracy only slightly (average difference of 0.04). Prediction performed using both cameras showed identical degrees of overfitting (prediction accuracy difference between training and testing sets of 0.03) (Table 8).

Table 8 *Comparison of grain yield prediction accuracy in single-environment scenarios using a minimal setup (single data capture during July) with M matrices developed based on RGB and multispectral data (cameras). S – Staur, V – Vollebekk, rTRN – accuracy in the training set, rTST – accuracy in the testing set, rmseTRN – root mean squared error in the training set, rmseTST – root mean squared error in the testing set. *M matrix developed using Micasense RedEdge M camera data, ** M matrix developed using Phantom 4 Multispectral camera data*

Setup	Metric	M matrix developed on data from:					Avg.
		2019_S	2019_V	2020_V	2021_V*	2021_V**	
RGB camera	rTRN	0.58	0.55	0.60	0.57	0.58	0.58
	rTST	0.56	0.51	0.58	0.54	0.55	0.55
	rmseTRN	53.0	44.3	65.0	56.4	56.1	55.0
	rmseTST	52.4	45.2	66.0	57.2	27.3	55.6
MS camera	rTRN	0.63	0.59	0.62	0.63	0.61	0.62
	rTST	0.62	0.55	0.59	0.60	0.58	0.59
	rmseTRN	50.4	42.4	63.7	53.4	54.3	52.9
	rmseTST	50.5	44.3	65.2	54.7	56.0	54.1

Discussion

Using relationships among objects or individuals has been present in plant breeding for over 100 years and has been the foundation for quantitative genetics. Those relationships can be derived based on various properties of the individuals, such as their pedigrees (A matrix) or dense genotyping data (Meuwissen, Hayes, and Goddard, 2001) and are widely used in breeding programs worldwide for both simple and complex traits. This study elaborates on utilizing multispectral phenotypes to construct genotypic relationships. The described methodology, bearing similarity to the G or A matrices or work of Krause *et al*. (2019), focuses not on individual numerical multispectral phenotype values and their possible abstract relationship with the complex trait of interest but rather on the similarities between genotypes.

Having standard GS in a single-environment scenario as a benchmark, M matrix-based prediction performed well. GS tended to reach almost perfect accuracy in the training set, with a significant drop in accuracy when tested on new lines. With an average GY heritability of 0.6 in this study, we can see that GS attempts to predict experimental error in individual trials. The accuracy difference between training and testing sets for the M matrix was four times smaller than for GS, indicating a much lower level of overfitting, probably due to being "closer" to the actual phenotypes. It is also seen in the fact that GS performs best when applied on multi-environment means, where the genetic signal is strongest and the experimental error is of lesser importance. The prediction accuracies using M matrices in a single-environment

scenario are comparable to those using H matrices (Krause *et al*., 2019). H matrix is analogical to the M matrix but developed on hyperspectral data with 62 bands, covering a spectrum between the 380-850 nm region. It indicates that introducing more narrower bands is less valuable than using fewer but broader bands available on commercial "low-cost" cameras such as those used in the study. The prediction accuracies achieved by the M matrices are also similar to other studies using linear and non-linear modeling approaches, including OLS (ordinary least squares), Bayesian methods, PLS (partial least squares), as well as functional regression (Aguate *et al*., 2017; Montesinos-López *et al*., 2017) or machine/deep learning methods (Shafiee *et al*., 2021).

The development of M matrices based on different environments should have lower prediction abilities when tested on the multi-environment means due to their inherent connection to the environment from which they originated. However, when tested on multi-environment means, temporally-dense data shows slightly higher accuracy than on trial means from which they originated. It could be partially because the means resemble the original environment but could also indicate that even the seemingly environment-specific similarity measure has the potential to generalize the genetic part of the phenotype. It is also highlighted by the M matrices originating from different environments, showing prediction power when tested on different environments (with exceptions). The temporally denser the data, the higher the M matrix's generalization ability. However, it is not easy to consult this hypothesis with available research.

G and M matrices complement each other – the GS model coupled with the M matrix (G+M) in a single-season scenario achieved higher accuracy than its components alone. The G+M model has the theoretical advantage of using both genetic information and the outcome of this information in a particular environment, capturing more of the crucial GxE interactions. However, the performance gain of adding M to GS was relatively small and came with valuable error reduction in the testing set. Considering the relatively low expense of acquiring multispectral information and its standalone prediction capacity, it can be a viable addition to the practical applications of GS protocols.

In multi-environment scenarios, an M matrix-based prediction was inferior to GS, a logical consequence of the inherent environment-specificity of the M matrices, as opposed to the "general" genetic nature of the G matrix. However, the prediction ability of M matrices in multi-environment scenarios tended to increase with the number of data capture sessions, which was not the case for single-environment

scenarios. It indicates that a temporally denser M matrix can describe the genetic component of grain yield, reaching accuracy almost as high as GS, even though this component is not as crucial for the prediction in a single-environment scenario.

GS supplemented with the M matrix shows overall slightly superior prediction accuracy compared to the GS or M matrix–based prediction alone; however, this appears to depend on the origin of the M matrix and, probably more importantly, the temporal density of data capture sessions (these two are confounded in this work). Despite higher accuracy, the G+M model shows higher error values, indicating that providing environment-specific information (M matrix) to GS in multi-environment prediction scenarios brings little value without providing further context.

Grain yield prediction in multi-environment scenarios using G or M matrices with environmental context (E matrix) shows high prediction accuracy, with GS's superiority in accuracy and error. It indicates that both layers of information prove informative when used in the environmental context. Although the model combining G, M, and E variates (G+M+E) is not superior to G+E in terms of accuracy, it minimizes the error, hinting that even only one field season of HTP data capture can aid GS protocols in providing more accurate genetic estimates of grain yield in multi-environment scenarios.

The camera bands' relative ranking of prediction accuracy indicates that heritability is essential. Both the least heritable and the least important band was near infra-red (NIR), despite its established link with plant physiology (multiple reflections of turgid cell structure (Josep Peñuelas & Iolanda Filella, 1998)). Hypothesizing, NIR reflectance could gain importance when water availability severely limits grain yield output (drought); however, it is impossible to verify this based on our available data. NIR band tends to be "unstable" and prone to differences in light conditions during data capture, bearing a significant challenge in field-based HTP. This problem is partially solved by introducing normalized vegetation indices (VIs, linear combinations of reflectance values in selected spectral regions) such as NDVI, which are more robust under variable lighting conditions.

The most important bands (RedEdge, Red, Green, and Blue) all link to chlorophyll and are more heritable than NIR. RedEdge points to chlorophyll content (Gitelson I, Merzlyak, and Lichtenthaler, 1996), and due to its photochemical properties, chlorophyll absorbs red and blue light while reflecting green. Therefore, it is reasonable to hypothesize that chlorophyll properties and content of a genotype govern the usefulness of the M matrix, following findings made by Krause *et al*.

(2019). It may also be that these associations are spurious – the most influential bands are highly heritable, and the M matrix models may therefore work on a "plants that look alike, yield alike" principle without an actual biological component to it.

The most informative data capture time occurs during the grain-filling period, which aligns well with the hypothesis that chlorophyll properties are captured by the M matrix and govern its predictive ability – during grain-filling, higher chlorophyll content means higher assimilation force and photosynthesis rate, resulting in higher grain yield. At the same time, inspecting drone imagery during grain filling indicates that the purely visual differences among trial plots are the smallest. Surprisingly, data capture sessions taken later in the growing season yield lower prediction accuracy. The moment when plants start maturing is easy to determine visually using HTP imagery due to the decay of chlorophyll and water content. At the same time, grain yield is generally highly correlated with earliness; hence, it should be reasonably possible to predict grain yield based on differences in genotype earliness. Our findings contradict this hypothesis, as a decay in prediction accuracy was observed as maturing progressed. These arguments also support the hypothesis of the M matrix using chlorophyll information proxies to predict grain yield rather than the "plants that look alike, yield alike" principle. Krause *et al*. (2019) did not observe a similar relationship: all flights taken during the vegetative season yielded comparable accuracy.

This study used two cameras for HTP data capture: Micasense RedEdge M and Phantom 4 Multispectral camera. They were analyzed back-to-back for their prediction accuracy using the M matrix in all models and scenarios. Our results show no evidence to conclude that there are significant differences in prediction accuracy between the cameras, despite the different technical specifications and numerical reflectance values obtained. This conclusion aligns with the authors' previous studies, comparing the same two cameras in parallel mission sessions for grain yield and biomass prediction using machine learning (Shafiee *et al*., in revision).

Based on our results, the prediction accuracy gains of using a multispectral camera over a simple RGB camera are incremental, despite multispectral cameras giving access to the informative RedEdge band. Considering the needed hardware, effort, and other resources for grain yield prediction, a simple RGB camera is more appealing from a purely economic standpoint. It was also exemplified that as little as a single flight mission with a simple RGB camera during the grain-filling period yields enough data to predict grain yield with accuracies over 0.5 in a single-environment scenario.

It shows the potential of the method and the potential of HTP in large-scale field trial applications.

The usefulness of GS and grain yield prediction using the M matrix can hardly be compared, as those two methodologies occupy different application niches in plant breeding: the purpose of GS is an early prediction of genotype's GEBV (genotypically-estimated breeding value) to enable efficient screening of early-generation progenies in breeding programs and being able to apply speed breeding. Therefore, the most significant advantage of GS is the ability to estimate GEBVs based on a sample of DNA of a single plant earlier. Grain yield prediction using the M matrix does not have this advantage. Genotypes must be put in field trials to collect their multispectral phenotypes, which can occur only in later-generation progenies in reasonably sized field trials. However, prediction using the M matrix scales very well, as adding more plots does not increase the workload linearly (which is the case in GS). Therefore, grain yield prediction using the M matrix fits well in the later stages of large-scale breeding programs, allowing the breeder to test a more significant number of variety candidates without expanding their technical base.

One disadvantage of the M matrix and ML protocols is their inherent connection with their environment of origin. Environment-specific trait estimates are of little use for breeders unless the environment closely resembles their target population of environments. Nevertheless, it was shown that a constructed M matrix based on dense data from a single environment could generalize (to "see" the heritable signal) and perform well when tested on a multi-environment mean. The fact that the M matrix works synergistically with GS makes it an affordable way to improve GS prediction accuracy and be used as a standalone tool. An added advantage of M matrix-based prediction or its inclusion into GS protocols is its purely statistical and comprehensible nature paired with using already available software without customization.

Conclusions

Developing genotypic relationships using high-throughput multispectral data (M matrix) gathered using consumer-grade equipment for grain yield prediction in wheat was elaborated. A back-to-back comparison of the prediction abilities of genomic selection models, including combinations of G, M, and E matrices, was conducted using multi-environment field trial data and mixed models (BLUP) in single and multi-environment scenarios. M matrix possesses a similar to the G matrix standalone prediction ability, and genomic selection models can be improved by including both G and M matrices. The importance of camera bands for grain prediction using the M matrix was discussed, showing that bands with the highest heritability are the most important. The importance of data capture was investigated, demonstrating that imagery taken during grain filling yields the best accuracy. The study also showed that grain yield prediction is possible using a simple RGB camera with a slight accuracy loss. The work contributes to expanding use cases for multispectral high-throughput phenotyping data and shows the potential of using this data for improving genomic selection protocols or standalone grain yield prediction in large-scale field trials.

References

Aguate, F.M. *et al.* (2017). Use of hyperspectral image data outperforms vegetation indices in prediction of maize yield. *Crop Science*, 57(5), pp. 2517–2524. Available at: https://doi.org/10.2135/cropsci2017.01.0007

Araus, J.L. & Cairns, J.E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science*, 19(1), pp. 52–61. Available at: https://doi.org/10.1016/j.tplants.2013.09.008

Bassi, F.M. *et al.* (2015). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp). *Plant Science*, 242, pp. 23–36. Available at: https://doi.org/10.1016/j.plantsci.2015.08.021

Beisel, N.S. *et al.* (2018). Utilization of single-image normalized difference vegetation index (SI-NDVI) for early plant stress detection. *Applications in Plant Sciences*, 6(10). Available at: https://doi.org/10.1002/aps3.1186

Bhat, J.A. *et al.* (2016). Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Frontiers in Genetics*, 7. Available at: https://doi.org/10.3389/fgene.2016.00221

Brault, C. *et al.* (2022). Interest of phenomic prediction as an alternative to genomic prediction in grapevine. *Plant Methods*, 18(1). Available at: https://doi.org/10.1186/s13007-022-00940-9

Brugger, A. *et al.* (2019). Extending hyperspectral imaging for plant phenotyping to the UV-Range. *Remote Sensing*, 11(12). Available at: https://doi.org/10.3390/rs11121401

Burud, I. *et al.* (2017). Exploring Robots and UAVs as Phenotyping Tools in Plant Breeding . *IFAC-PapersOnLine*, 50(1), pp. 11479–11484. Available at: https://doi.org/10.1016/j.ifacol.2017.08.1591

Caamal-Pat, D. *et al.* (2021). lme4GS: An R-Package for Genomic Selection. *Frontiers in Genetics*, 12. Available at: https://doi.org/10.3389/fgene.2021.680569

Gitelson I, A.A., Merzlyak, M.N. and Lichtenthaler, H.K. (1996). Detection of Red Edge Position and Chlorophyll Content by Reflectance Measurements Near 700 nm. *J Plant Physiol*, 148(3-4), pp. 501-508. Available at: https://doi.org/10.1016/S0176-1617(96)80285-9

Han, L. *et al.* (2019). Modelling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods*, 15(1), pp. 1–19. Available at: https://doi.org/10.1186/s13007-019-0394-z

Hassan, M.A. *et al.* (2019). Accuracy assessment of plant height using an unmanned aerial vehicle for quantitative genomic analysis in bread wheat. *Plant Methods*, 15(1), pp. 1–12. Available at: https://doi.org/10.1186/s13007-019-0419-7

Hickey, L.T. *et al.* (2019). Breeding crops to feed 10 billion. *Nature Biotechnology*, 37, pp. 744–754. Available at: https://doi.org/10.1038/s41587-019-0152-9

Houchmandzadeh, B. (2014). An alternative to the breeder's and lande's equation. *G3: Genes, Genomes, Genetics*, 4(1), pp. 97–108. Available at: https://doi.org/10.1534/g3.113.008433

Hu, P. *et al.* (2018). Estimation of plant height using a high throughput phenotyping platform based on unmanned aerial vehicle and self-calibration: Example for sorghum breeding . *European Journal of Agronomy*, 95, pp. 24–32. Available at: https://doi.org/10.1016/j.eja.2018.02.004

Peñuelas J. & Filella I. (1998). Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science*, 3(4), pp. 151-156. Available at: https://doi.org/10.1016/S1360-1385(98)01213-8

van Klompenburg, T., Kassahun, A. and Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177. Available at: https://doi.org/10.1016/j.compag.2020.105709

Krause, M.R. *et al.* (2019). Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3: Genes, Genomes, Genetics*, 9(4), pp. 1231–1247. Available at: https://doi.org/10.1534/g3.118.200856

Li, B. *et al.* (2019). The estimation of crop emergence in potatoes by UAV RGB imagery. *Plant Methods*, 15(1), pp. 1–13. Available at: https://doi.org/10.1186/s13007-019-0399-7

Li, B. *et al.* (2020). Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162, pp. 161–172.
Available at: https://doi.org/10.1016/j.isprsjprs.2020.02.013

Lu, N. *et al.* (2019). Improved estimation of aboveground biomass in wheat from RGB imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system. *Plant Methods*, 15(1), pp. 1–16. Available at: https://doi.org/10.1186/s13007-019-0402-3

Meuwissen, T.H., Hayes, B.J. and Goddard, M.E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 157(4), pp. 1819-1829.
Available at: https://doi.org/10.1093/genetics/157.4.1819

Montesinos-López, O.A. *et al.* (2017). Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods*, 13(1).
Available at: https://doi.org/10.1186/s13007-016-0154-2

Mróz, T., Dieseth, J.A. and Lillemo, M. (2022). Historical grain yield genetic gains in Norwegian spring wheat under contrasting fertilization regimes. *Crop Science*, 62(3), pp. 997–1010. Available at: https://doi.org/10.1002/csc2.20714

Nannuru, V.K.R. *et al.* (2022). Genetic architecture of *fusarium* head blight disease resistance and associated traits in Nordic spring wheat. *Theoretical and Applied Genetics*, 135(7), pp. 2247–2263. Available at: https://doi.org/10.1007/s00122-022-04109-9

Reynolds, M. & Langridge, P. (2016). Physiological breeding. *Current Opinion in Plant Biology*, 31, pp. 162–171. Available at: https://doi.org/10.1016/j.pbi.2016.04.005

Robert, P. *et al.* (2022). Phenomic selection in wheat breeding: identification and optimization of factors influencing prediction accuracy and comparison to genomic selection . *Theoretical and Applied Genetics*, 135(3), pp. 895–914.
Available at: https://doi.org/10.1007/s00122-021-04005-8

Rutkoski, J. *et al.* (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3: Genes, Genomes, Genetics*, 6(9), pp. 2799–2808.
Available at: https://doi.org/10.1534/g3.116.032888

Sakurai, K. *et al.* (2022). Time-series multispectral imaging in soybean for improving biomass and genomic prediction accuracy. *Plant Genome*, 15(4).
Available at: https://doi.org/10.1002/tpg2.20244

Shafiee, S. *et al.* (2021). Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV

imagery. *Computers and Electronics in Agriculture*, 183. Available at: https://doi.org/10.1016/j.compag.2021.106036

Shafiee, S. *et al*. (in review) "Multispectral camera, Phenological stage, and sun elevation angle impact on yield and biomass prediction in wheat – a comparative study of Micasense RedEdge-M and Phantom 4 sensors". Submitted to: *Computers and Electronics in Agriculture*

Sun, J. *et al.* (2017). Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. *The Plant Genome*, 10(2). Available at: https://doi.org/10.3835/plantgenome2016.11.0111

Tirado, S.B., Hirsch, C.N. and Springer, N.M. (2020). UAV-based imaging platform for monitoring maize growth throughout development. *Plant Direct*, 4(6), pp. 1–11. Available at: https://doi.org/10.1002/pld3.230

Trevisan, R. *et al.* (2020). High-throughput phenotyping of soybean maturity using time series uav imagery and convolutional neural networks. *Remote Sensing*, 12(21), pp. 1–19. Available at: https://doi.org/10.3390/rs12213617

White, J.W. *et al.* (2012). Field-based phenomics for plant genetics research. *Field Crops Research*, pp. 101–112. Available at: https://doi.org/10.1016/j.fcr.2012.04.003

Zhou, J. *et al.* (2019). Estimation of maturity date of soybean breeding lines using UAV-based imagery. *Remote Sensing,* 11(18). Available at: https://doi.org/10.3390/rs11182075

Supplementary material

Multispectral-derived genotypic similarities from budget cameras allow grain yield prediction and genomic selection augmentation in single and multi-environment scenarios in spring wheat

Manuscript

Tomasz Mróz[1], Sahameh Shafiee[1], Jose Crossa[23], Osval A. Montesinos-Lopez[4], and Morten Lillemo[1]

[1]*Department of Plant Sciences, Norwegian University of Life Sciences, NO-1432 Ås, Norway*

[2]*International Maize and Wheat Improvement Center (CIMMYT), Km 45, Carretera Mexico Veracruz, CP 52640 Texcoco, Edo. de México, Mexico*

[3]*Colegio de Postgraduados, CP 56230 Montecillos, Edo. de México, Mexico*

[4]*Facultad de Telemática, Universidad de Colima, Colima, Mexico*

Multispectral image acquisition and processing

Two multispectral cameras were used in this study. The RedEdge-M has five imaging sensors, including red, green, blue, red edge, and near InfraRed (NIR) bands (Figure 1). The camera has a 5.4 mm focal length, the sensor size is 4.8 mm x 3.6mm, and the image size is 1280 x 960 pixels. A sunshine sensor can record the illumination information of each image to calibrate the multispectral images. The camera was mounted on a fully programmable DJI Matrix 100 UAV with a maximum payload of 1.25 kg. The flight routes were planned in the Altizure application (Everest Innovation Technology).

The P4M camera has six imaging sensors, five narrow bands, and one combined RGB sensor. The multispectral bands include Red, Green, Blue, red edge, and (NIR). The camera focal length is 5.74 mm, the sensor size is 4.87 mm x 3.96 mm, and the image size is 1600×1300 pixels. This camera is integrated into its UAV platform. Flight routes were planned in the DJI GSPro application (https://www.dji.com/ground-station-pro). Flights were conducted around local noon time. The images were taken from a nadir view with 85% frontal and 80% side overlap.

The image pixel value depends on different factors, such as sensor setting, sensor properties, and scene condition, which must be corrected to get a radiometrically trusted measure of terrain reflectance. All related parameters are present in the image EXIF data and applied for radiometric corrections in Pix4D (refer to Shafiee et al., 2021, for more details). Reflectance targets were applied to do a radiometric calibration in field conditions. RedEdge-M has its calibration panel (with Albedo values of 0.58, 0.59, 0.60, 0.59, and 0.56, respectively, for the blue, green, red, red edge, and NIR bands). Our experiments showed that P4M generates more reliable spectral values when the radiometric calibration (data not shown) is applied, which was also noted by Di Gennaro et al. in their recent study (2022). Therefore, a calibration panel (SphereOptics, Diffuse Reflectance Target-53%R) was applied to correct P4M images (Albedo values of 0.54329, 0.54389, 0.54260, 0.54092, and 0.53788 respectively, for blue, green, red, red edge and NIR bands). Albedo values were determined for each central wavelength based on the datasheet the target provider enclosed. Images from both cameras were imported into Pix4D software (Pix4D SA, Lausanne, Switzerland) for processing. Different Processing steps of UAV images, including geometric correction, image mosaicking, and radiometric calibration, were conducted in Pix4D with a spatial resolution of 1.3 and 1.09 cm/pixel, respectively, for RedEdge-M and P4M. Orthomosaics were generated for each band separately. QGIS software (QGIS 3.4, Open-Source Geo-Spatial Foundation

Project. http://qgis.osgeo.org) was used to extract average spectral values for each experimental plot in the trial (refer to Shafiee *et al.*, 2021 for more details). Each plot was masked out in the middle part to avoid border effects using a polygon shape file in QGIS. A separate mask was generated for each camera. Since the pixel size is much smaller than the plot size and the canopy is well structured, the mixed pixel issue is insignificant in this study. The ZonalStatistic Tool was applied to calculate the median reflectance value per plot.

Weather conditions during the field trials

Please refer to supplementary information of Paper II.

Planting dates of the trials

Table S1 *Planting dates of the field trials at both locations*

Season	Vollebekk	Staur
2015	2015-04-24	-
2016	2016-04-24	2016-05-10
2017	2017-05-04	2017-05-12
2019	2019-05-19	2019-06-04
2020	2020-05-15	2020-04-21
2021	2021-04-20	2021-04-27
2022	2021-04-21	-

Band values during the season

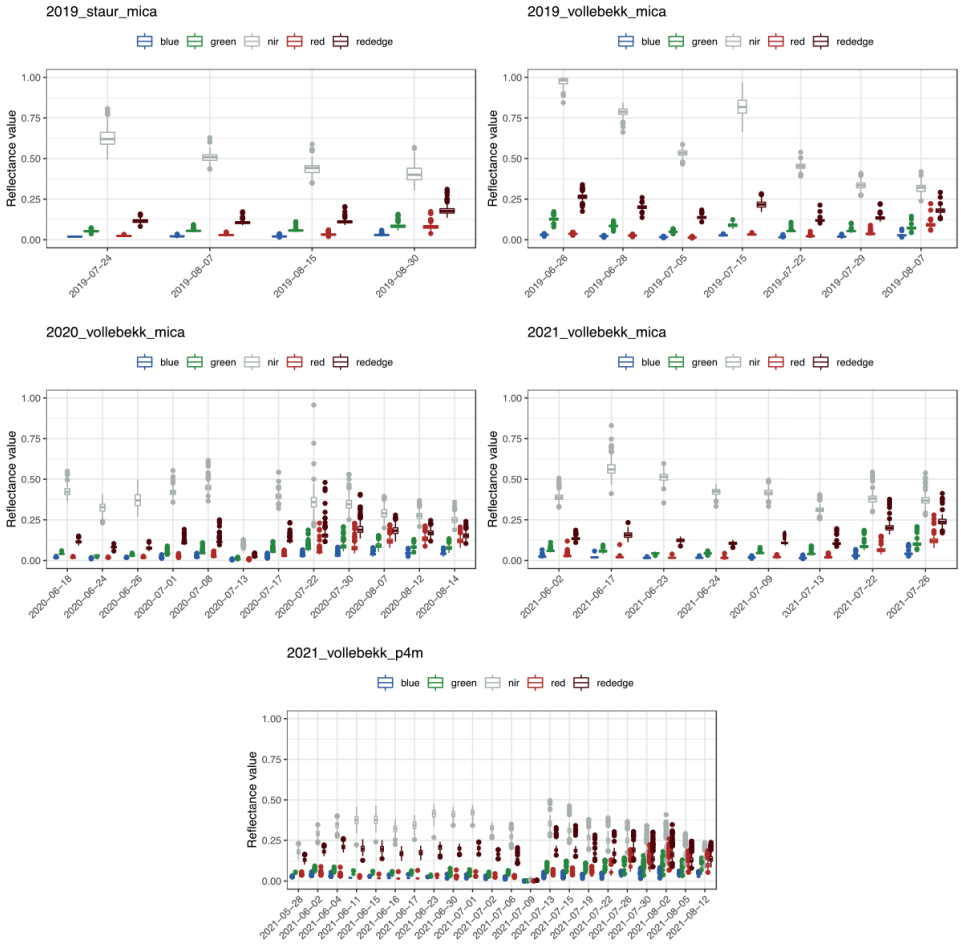


Figure S1 *Raw band reflectance during the season in all environments. For red, green, and blue, color on the chart represents the respective band; silver – NIR, dark red – RedEdge*

Correlations between the bands and grain yield

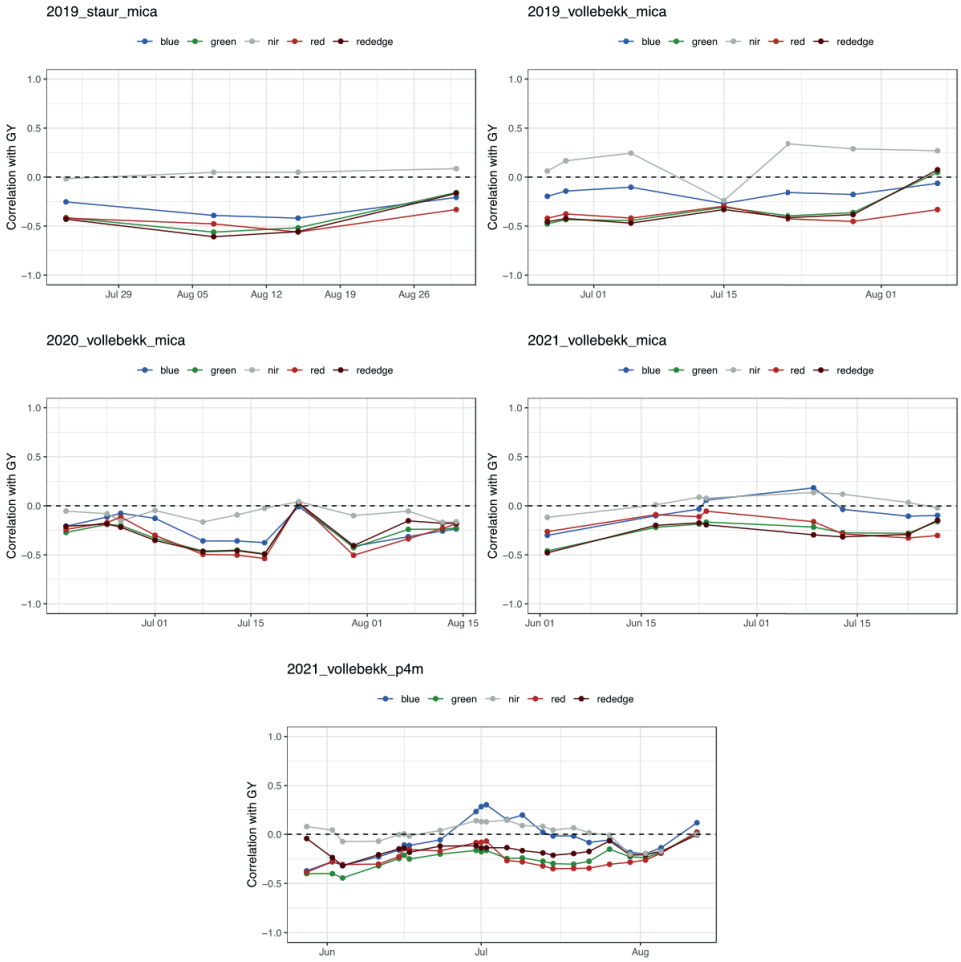


Figure S2 *Pearson correlation between the camera bands and grain yield for all the flight days. For red, green, and blue, the color on the chart represents the respective band; silver – NIR, dark red – RedEdge*

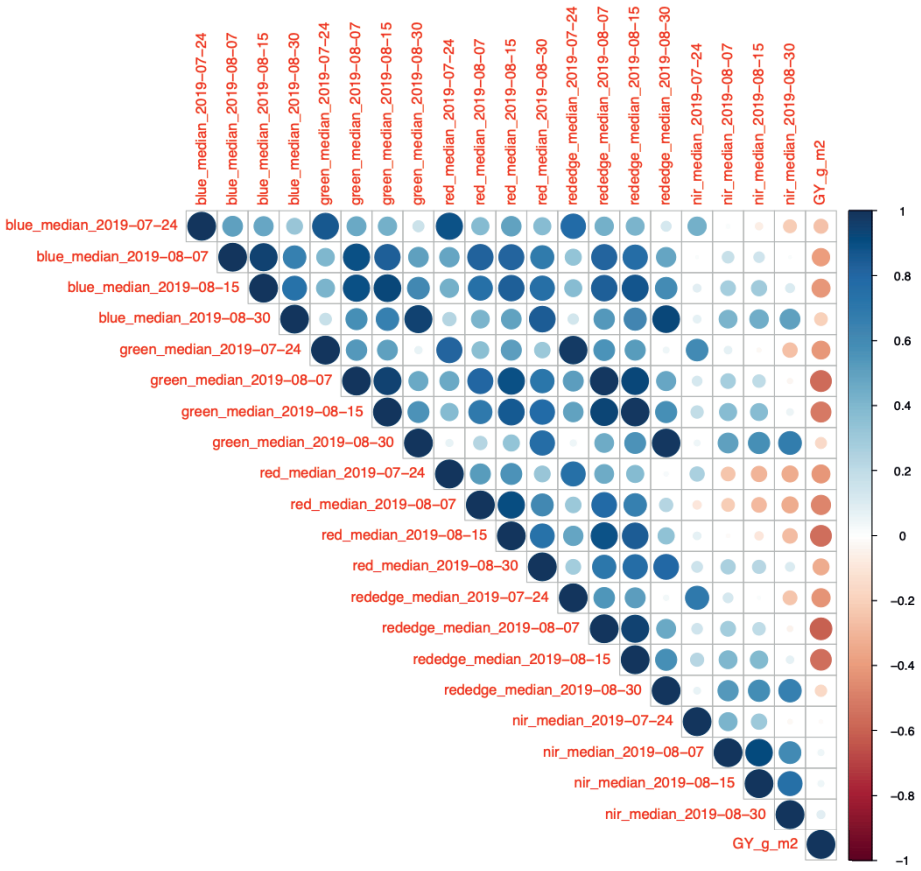


Figure S3 *Env: Staur 2019, Camera: Micasense. Pearson correlations among the bands at different days and grain yield.*

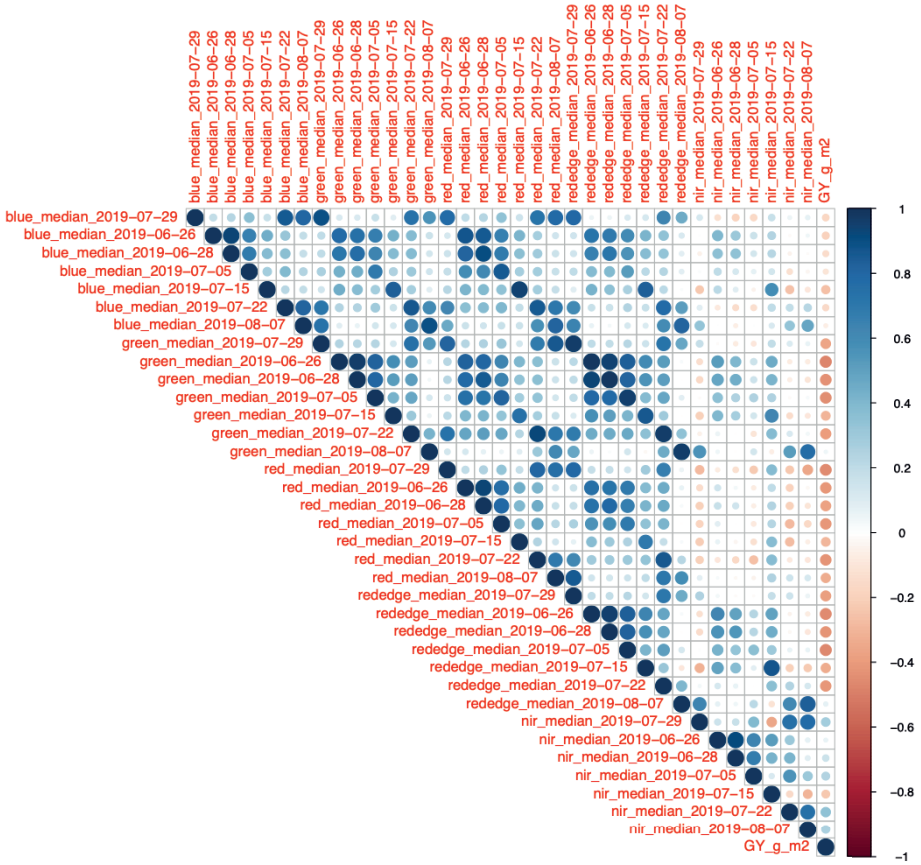


Figure S4 *Env: Vollebekk 2019, Camera: Micasense. Pearson correlations among the bands at different days and grain yield*

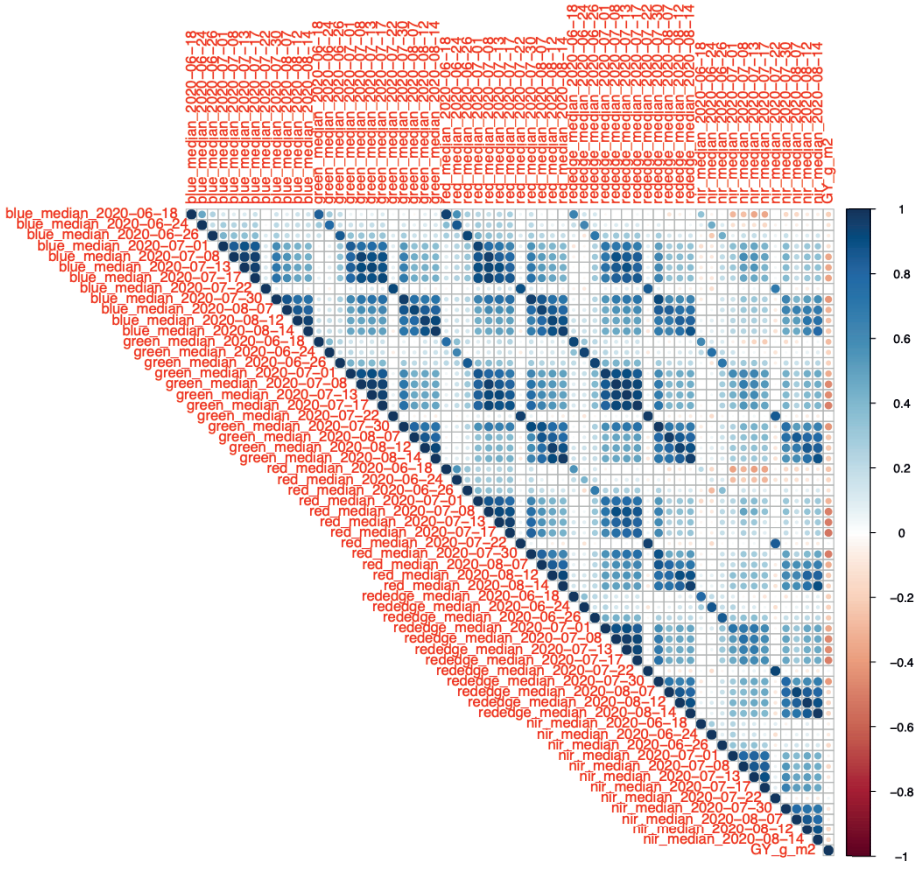


Figure S5 *Env: Vollebekk 2020, Camera: Micasense. Pearson correlations among the bands at different days and grain yield*

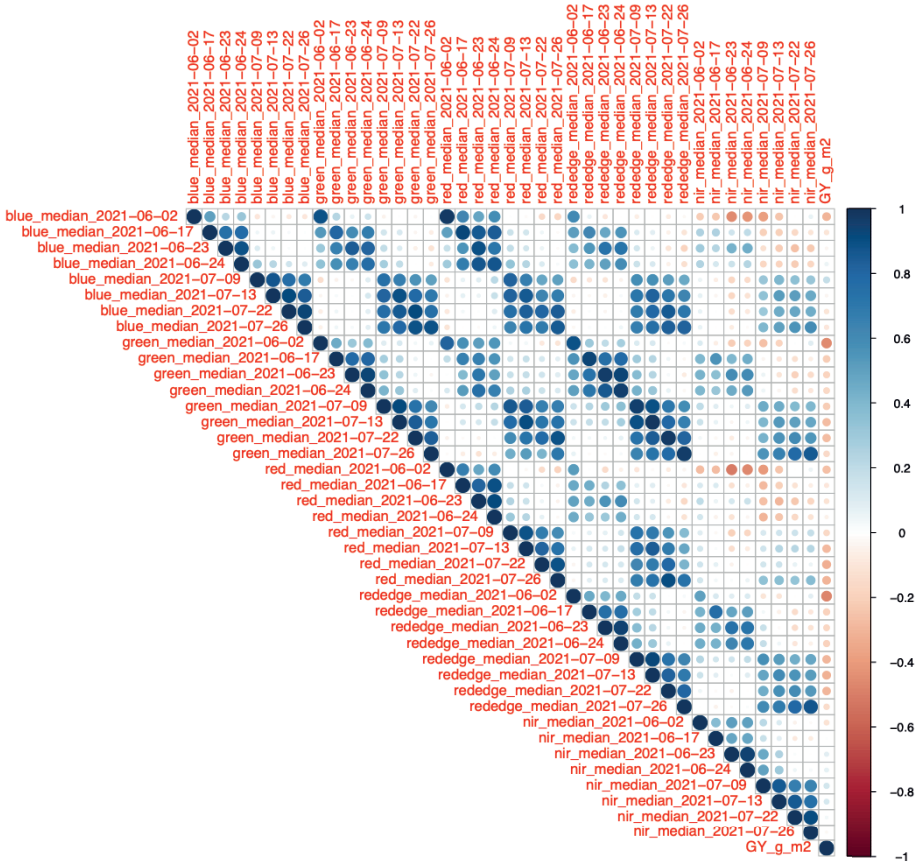


Figure S6 *Env: Vollebekk 2021, Camera: Micasense. Pearson correlations among the bands at different days and grain yield*

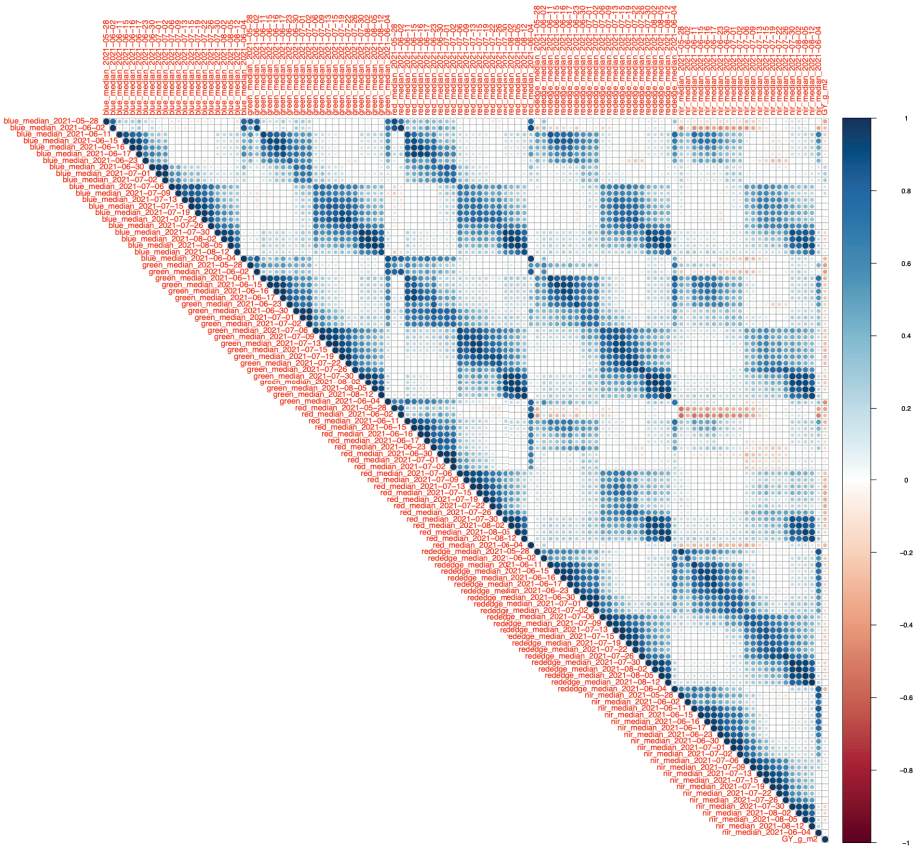


Figure S7 *Env: Vollebekk 2021, Camera: P4M. Pearson correlations among the bands at different days and grain yield*

Errata

Section and page	Modification	Change
References p.57	Added reference	Slafer, G.A. *et al.* (2023). A 'wiring diagram' for sink strength traits impacting wheat yield potential. *Journal of Experimental Botany*, 74 (1), pp. 40–71. Available at: https://doi.org/10.1093/jxb/erac410
Introduction p.5	Changed sentence	Old: The new varieties, on the other hand, had HI approaching the theoretical limit of 0.5. New: The new varieties, on the other hand, had HI of 0.5, closer to the theoretical limit of 0.6 (Slafer *et al*., 2023).
Introduction p.9	Changed sentence	Old: In wheat, the estimated value of the gains in GY varies between 5 and 115 kg ha-1 per year (0.19 - 2.80%), with values around 0.2% being most common in more reliable collections (Table 2). New: In wheat, the estimated value of the gains in GY varies between 5 and 115 kg ha-1 per year (0.19 - 2.80%, Table 2).
Introduction p.10	Changed table header	Old: Table 2 Genetic gains in wheat grain yield in various collections worldwide. GrPS – grains per spike, NUE – nitrogen use efficiency, GrpA – grains per area, BM – above- ground biomass, SR – stripe rust resistance, HI – harvest index, SppA – spikes per area, TKW – thousand kernel weight, DH – days to heading, Chl – chlorophyll content, PAR- increased capture of photosynthetically-active radiation New: Genetic gains in wheat grain yield in various collections worldwide. GrPS – grains per spike, GrpA – grains per area, SppA – spikes per area, TKW – thousand kernel weight
Paper II p.16, Table 4	Addition to table footnote	Mono – marker monomorphic (estimation not possible); (…) Full difference between homozygotes (…)
Paper II p.16, Table 4	Added to table caption	in respective units listed in Table 1

ISBN: 978-82-575-2081-6
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no

116852 / ANDVORDGRAFISK.NO