



Investigation of the stability of six breadmaking QTL and linkage  
mapping of GPD and senescence metrics in bread wheat.

Rohan Richard

Thesis submitted for the Degree of Doctor of Philosophy  
Department of Crop Science

April 2023

## **Declaration**

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Rohan Richard

## Acknowledgments

This PhD thesis was funded by the Agriculture and Horticulture Development Board (AHDB) and I warmly thank them for the financement of my thesis.

Fistly, I would like to express my gratitude to my supervisors, Professor Peter Shewry, Dr Alison Lovegrove, Professor Paola Tosi, Dr Simon Griffiths and Luzie Wingen for their advices and guidance during this PhD.

A special thanks for Peter Shewry for giving me the opportunity to come to the UK and do this PhD.

I would like to thank Liz Howes and Mervin Poole from Heygates Ltd. for hosting me and for our numerous meetings during the course of the project.

I express all my gratitude towards Richard Casebow, the field manager of the University of Reading for conducting the NDVI measurements.

A big thumbs up for the Rothamsted statistical team for the amazing (FREE!!) training courses and particularly Kirsty Hassall and Jess Hood, my statistician referents for their patience in answering my question and their reactivity.

I would like to say thank you to Yongfang Wan for showing me how to run SDS-PAGE gels and for her practical advices

I am also grateful toward Chris Hall for his kindness, the inductions, and the help with the use of the different machines in the Jenkinson building. It was great and reassuring to have you around!

Thanks to Heather Angus, Jack Dunn, Alison's daughter, and Michael Storkey for their temporary help with the grain processing. It was greatly appreciated!!

A huge cheer up for my Rothamsted friends (Rahul, Latifa, Parul, Mesfin, Gaetan, Yoann and all the others) for the company, and emotional support. It was great to discover new cultures, and I look forward to visit you in your respective countries in the future!

The byrianis and lamb nilgiri still make my mouth watering!

Before last, I have to say thank to Mom and Dad for passing me encouring me and supporting me during this PhD.

Last but not least, I would like to say thank you to my partner Katy Dow for her LOVE, her continuous support, her curiosity and our fascinating conversations about psychology, nature ...

## List of Abbreviations

<b>AHDB</b>	<b>Agriculture and Horticulture Development Board ANOVA Analysis of Variance</b>
<b>AMMI</b>	<b>Additive Main Effect and Multiplicative Interaction Model</b>
<b>ASV</b>	<b>AMMI Stability Values</b>
<b>BBIRA</b>	<b>British Baking Industries Research Association</b>
<b>BIBD</b>	<b>Balanced Incomplete Block Design</b>
<b>BU</b>	<b>Brabender Unit</b>
<b>CBP</b>	<b>Chorleywood Bread Process</b>
<b>cHATS</b>	<b>Constitutive High-Affinity Low Capacity Transport System</b>
<b>CI</b>	<b>Confidence Interval</b>
<b>CIM</b>	<b>Complex Interval Mapping</b>
<b>cM</b>	<b>centiMorgan</b>
<b>DDT</b>	<b>Dough Development Time</b>
<b>DEGs</b>	<b>Differentially Expressed Genes</b>
<b>DF</b>	<b>Dietary Fibre</b>
<b>DH</b>	<b>Doubled haploid</b>
<b>DS</b>	<b>Degree of Softening</b>
<b>G x E</b>	<b>Genotype by Environment</b>
<b>GEVES</b>	<b>Groupe d'Etude et de contrôle des Variétés Et des Semences</b>
<b>GOGAT</b>	<b>Glutamate synthase</b>
<b>GPC</b>	<b>Grain Protein Content</b>
<b>GPD</b>	<b>Grain Protein Deviation</b>
<b>GS</b>	<b>Glutamine synthetase</b>
<b>GWAS</b>	<b>Genome Wide Association Study</b>
<b>GY</b>	<b>Grain Yield</b>
<b>HC</b>	<b>Hierarchical Clustering</b>
<b>HD</b>	<b>Heading Date</b>
<b>HMW-GS</b>	<b>High Molecular Weight Glutenin Subunit</b>
<b>INRA</b>	<b>Institut National de la Recherche Agronomique</b>
<b>LATS</b>	<b>Low-Affinity Transport System</b>

<b>LMW-GS</b>	<b>Low Molecular Weight Glutenin Subunit</b>
<b>LOD</b>	<b>Logarithm Of the Odds</b>
<b>LSD</b>	<b>Least Significant Difference</b>
<b>M x H</b>	<b>Malacca x Hereward</b>
<b>MAGIC</b>	<b>Multiparent Advanced Generation Intercross</b>
<b>N</b>	<b>Nitrogen</b>
<b>NDVI</b>	<b>Normalized Difference Vegetation Index</b>
<b>NDVI_Max</b>	<b>NDVI before the onset of senescence</b>
<b>NDVI10</b>	<b>NDVI at onset or 90% of senescence</b>
<b>NDVI50</b>	<b>NDVI at onset or 50% of senescence</b>
<b>NDVI90</b>	<b>NDVI at onset or 10% of senescence</b>
<b>NH<sub>4</sub><sup>+</sup></b>	<b>Ammonium</b>
<b>NILs</b>	<b>Near-Isogenic Lines</b>
<b>Nir</b>	<b>Nitrite reductase</b>
<b>NIRS</b>	<b>Near Infrared Spectroscopy</b>
<b>NO<sub>2</sub><sup>-</sup></b>	<b>Nitrites</b>
<b>NO<sub>3</sub><sup>-</sup></b>	<b>Nitrates</b>
<b>NR</b>	<b>Nitrate reductase</b>
<b>NUE</b>	<b>Nitrogen-Use Efficiency</b>
<b>PCA</b>	<b>Principal Component Analysis</b>
<b>QTL</b>	<b>Quantitative Trait Loci</b>
<b>RCBD</b>	<b>Randomised Complete Block Design</b>
<b>R<sub>max</sub></b>	<b>Extensograph maximum resistance</b>
<b>RuBisCO</b>	<b>Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase</b>
<b>SDS-PAGE</b>	<b>Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis</b>
<b>SE-HPLC</b>	<b>Size-Exclusion High Performance Liquid Chromatography</b>
<b>SIM</b>	<b>Single Interval Mapping</b>
<b>SNP</b>	<b>Single Nucleotide Polymorphism</b>
<b>S-Poor</b>	<b>Sulphur-Poor prolamins</b>
<b>S-rich</b>	<b>Sulphur-Rich prolamins</b>
<b>SW</b>	<b>Specific Weight</b>
<b>TA</b>	<b>Total gluten protein</b>

<b>TCA</b>	<b>Trichloroacetic Acid</b>
<b>TKW</b>	<b>Thousand Kernel Weight</b>
<b>TT10</b>	<b>Thermal time at onset or 90% of senescence</b>
<b>TT50</b>	<b>Thermal time at onset or 50% of senescence</b>
<b>TT90</b>	<b>Thermal time at onset or 10% of senescence</b>
<b>WA</b>	<b>Water Absorption</b>

# Table of Contents

<b>Acknowledgments .....</b>	<b>iii</b>
<b>List of Abbreviations .....</b>	<b>iv</b>
<b>List of Figures.....</b>	<b>xiii</b>
<b>List of Tables.....</b>	<b>xvi</b>
<b>Abstract.....</b>	<b>xviii</b>
<b>Chapter 1. Introduction.....</b>	<b>1</b>
1.1. The genesis of bread wheat.....	1
1.2. Wheat in figures.....	2
1.3. Future challenges for wheat production .....	3
1.4. Breadmaking in the UK .....	5
1.4.1. Introduction.....	5
1.4.2. Classification of wheat according to end-use .....	5
1.4.3. The Chorleywood breadmaking process .....	6
1.5. Histology of the wheat caryopsis.....	7
1.6. Wheat caryopsis with health benefits or effect on technological properties .....	9
1.6.1. Starch and water absorption .....	9
1.6.2. Dietary fibre (DF) .....	10
1.6.3. Vitamins: essential nutrients .....	11
1.6.4. Lipids and gas cell stability.....	12
1.6.5. Carotenoids content and flour colour .....	12
1.7. The gluten proteins .....	13
1.7.1. Classification.....	13
1.7.2. Genetic architecture.....	14
1.8. Breadmaking quality .....	16
1.8.1. Physical dough testing.....	16
1.8.2. Gluten proteins quality .....	19
1.9. Effect of the environment on the synthesis of gluten proteins.....	22

1.9.1 Effect of temperature.....	22
1.9.2. Effect of nitrogen (N) fertilisation.....	24
<b>1.10. Improving the grain protein content .....</b>	<b>25</b>
1.10.1. N fertilisers cause environmental damages.....	25
1.10.2. The negative relationship between grain protein content and grain yield.....	26
<b>1.11. Grain protein deviation (GPD) .....</b>	<b>27</b>
Introduction .....	27
1.11.1. Stability of GPD across environment .....	28
1.11.2. Genetic architecture of GPD: QTL and candidate genes .....	29
1.11.3. GPD molecular pathway and its link to nitrogen metabolism .....	31
<b>1.12. Thesis Aims.....</b>	<b>34</b>
<b>Chapter 2. Materials and Methods .....</b>	<b>37</b>
<b>2.1. Creation of the Near Isogenic Lines .....</b>	<b>37</b>
2.1.1. Development of the Malacca x Hereward DH mapping population .....	37
2.1.2 Development of the NILs.....	37
<b>2.2. Selection of the NILs .....</b>	<b>39</b>
<b>2.3. Experimental field design in which the NILs were grown.....</b>	<b>39</b>
<b>2.4 Assessment of the kinship between the NILs using Hierarchical Clustering (HC) on the genotypic data. ....</b>	<b>42</b>
<b>2.5. Milling of the NILs .....</b>	<b>42</b>
2.5.1. Preparation of the white flour for rheology and baking tests .....	42
2.5.2. Preparation of whole grain mills for SDS-PAGE .....	43
<b>2.6. Preparation of 12 white flour bulks corresponding to the 12 NILs alleles .....</b>	<b>43</b>
2.6.1. Measurement of total N content on the white flour of the NILs (60 lines) .....	43
2.6.2. Creation of NILs allelic bulks by bulking together the five sister lines at each NILs QTL allele. .....	45
<b>2.7. Rheological assessments on the doughs prepared from the 12 NILs allelic bulks .....</b>	<b>47</b>
<b>2.8. Evaluation of the six NILs QTL effects. ....</b>	<b>47</b>

2.8.1. Bread height .....	47
2.8.2. Evaluation of flour and crumb whiteness using the L* a* b* colour space .....	48
2.8.3. Evaluation of the number of cells in the bread crumb .....	49
<b>2.9. Assessment of the NILs gluten protein .....</b>	<b>49</b>
2.9.1. Evaluation of size distribution of the gluten polymers using Size-Exclusion High Performance Liquid Chromatography (SE-HPLC) .....	49
2.9.2 Extraction of total protein and gluten proteins .....	51
2.9.3. Separation of the gluten protein on Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) .....	52
2.9.4. Relative quantification of the three gluten protein groups (HMW-GS, $\omega$ -gliadins, and LMW-GS and $\alpha$ -, $\beta$ - and $\gamma$ -gliadins) separable by SDS-PAGE by image analysis and comparison of their proportion between the allelic pairs .....	52
<b>2.10. Bioinformatic analysis .....</b>	<b>53</b>
2.10.1. Investigation of Lipoxigenase genes as a candidate for QTL 4D-2 whiteness .....	53
2.10.2. Blast of the NILs 1B QTL flanking markers on reference genome (RefSeq v1.0) of the wheat variety Chinese Spring. ....	54
<b>2.11. Statistical analysis .....</b>	<b>54</b>
2.11.1. Analysis of Extensograph data on the NILs QTL 4D-2 in the Rothamsted 2020-2021 trial.	54
2.11.2 Multiple regression on the flour colour (dry method) in the Rothamsted 2019-2020 trial. ....	54
2.11.3. Analysis of the QTL effects of the six NILs .....	55
<b>2.12. The DH field design .....</b>	<b>56</b>
<b>2.13. Phenotyping of the DH lines .....</b>	<b>57</b>
2.13.1 Measurement of grain nitrogen content by Near Infrared spectroscopy (NIRS) .....	57
2.13.2 Yield .....	58
<b>2.14. Calculation of the GPD genotypic means.....</b>	<b>58</b>
<b>2.15. QTL analysis.....</b>	<b>59</b>
2.15.1. Single environment QTL analysis .....	59
<b>2.16. Comparison of seven senescence metrics between the DH lines. ....</b>	<b>60</b>

2.16.1. Recording and conversion of heading dates (HD) .....	60
2.16.2. Measurement of canopy greenness from HD to crop maturity using a Normalized Difference Vegetation Index (NDVI) sensor.....	61
2.16.3 Modelling of senescence and calculation of seven parameters from the senescence curves .....	63
<b>Chapter 3. Study of the effect and stability of six NILs QTL .....</b>	<b>66</b>
3.1 Background .....	66
3.2. A short description of the cultivars.....	66
3.3. Project objectives.....	67
3.4. Results.....	68
3.4.1. Positioning of NIL 1B QTL on the reference genome (RefSeq v1.0.) of the wheat variety Chinese Spring.....	68
3.4.2. Assessment of the NILs kinship using hierarchical clustering on the genotypic data. ....	68
3.4.3. Analysis of the gluten subunits separated by SDS-PAGE in the six NILs sets .....	70
3.4.4. Assessments of the NILs QTL effects .....	73
3.4.5. Multiple regression analysis on flour whiteness in the 2019-2020 NILs trial. ....	77
3.4.6. Investigation of lipoxygenase gene in the QTL region of NILs 4D-2.....	78
3.4.7. Assessment of glutenin polymers size distribution using Size-Exclusion High Performance Liquid Chromatography (SE-HPLC). ....	81
3.4.8. Assessment of the dough mixing properties in the 2019-2020 NILs with the Farinograph. 82	
3.4.9. Assessment of the dough strength and extensibility on the NILs from the 2019-2020 and 2020-2021 trials using the Extensograph .....	83
3.4.10. Exploration of correlations between SE-HPLC ratios and dough properties determined by the Extensograph .....	85
3.5. Discussion and conclusion .....	86
3.5.1. Blast of the NIL 1B QTL on the reference genome (RefSeq v1.0) of the wheat variety Chinese Spring .....	86
3.5.2. Analysis of the NILs genetic kinship according to their genotypes on the Axiom 35K .....	86

3.5.3. Comparisons of the proportion of HMW-GS, $\omega$ -gliadins, LMW-GS and $\alpha$ -, $\beta$ -, $\gamma$ -gliadin, and the ratio HMW-GS/LMW-GS between the allelic pairs.....	86
3.5.4. Validation of the QTL breadmaking quality effects in the NILs.....	87
3.5.5. Is allele 4D-2a bleaching effect caused by a higher level or activity of lipoxygenase? .....	89
3.5.6. Correlations between SE-HPLC ratios and dough properties .....	89
3.5.7. Limits .....	91
<b>Chapter 4. Mapping of the trait grain protein deviation and its components in the M x H DH population.....</b>	<b>92</b>
4.1. Introduction .....	92
4.2. Project aim .....	92
4.3. Results.....	94
4.3.1. Description of the data .....	94
4.3.2 Calculation of GPD.....	96
4.3.3. Correlations of yield components and GPC with GPD in the three field trials. ....	100
4.3.4 Analysis of effects of environment using an Additive Main Effect and Multiplicative Interaction (AMMI) model.....	100
4.3.5. QTL analysis.....	104
4.4. Conclusion and Discussion .....	107
4.4.1. Simple Linear regressions GPC on GY showed low $R^2$ .....	107
4.4.2. AMMI analysis .....	108
4.4.3. Correlations of GPC, GY, and GPD among the environments .....	109
4.4.4. Correlations of GPD with GPC and yield components (GY and TKW, SW) .....	109
4.4.5. Single environment QTL analysis .....	109
4.4.6. Future prospects .....	111
<b>Chapter 5. Analysis and comparison of senescence metrics within the DH population .....</b>	<b>112</b>
5.1. Introduction .....	112
5.2. Results.....	112
5.2.1 Description of weather data.....	112

5.2.2. Description of HD data .....	113
5.2.3. Modelling of senescence .....	115
5.2.4. Comparison of senescence metrics between the DH lines.....	115
5.2.5. Exploration of correlations .....	127
5.2.6. Single environment QTL analysis with the seven senescence metrics .....	130
5.2.7. Principal component analysis (PCA) on the seven senescence metrics.....	133
<b>5.3. Discussion and Conclusions .....</b>	<b>137</b>
5.3.1 Variation in the length of the crop cycle between the Reading 2020-2021 and Reading 2021-2022 trials.....	137
5.3.2. Senescence metrics .....	138
5.3.3. Determination of correlations .....	138
5.3.4. QTL analysis .....	141
<b>5.4 Limitations to study .....</b>	<b>141</b>
5.4.1. Scoring of GS 55 in the Reading 2021-2022 .....	141
5.4.2. Insensitivity of NDVI to medium-high chlorophyll content.....	142
<b>Chapter 6. General discussion and future work .....</b>	<b>143</b>
<b>Chapter 7. References .....</b>	<b>147</b>

## List of Figures

Figure 1.1	Bread wheat and its tetraploid and diploid wild ancestors ear phenotypes	1
Figure 1.2	Wheat cultivation in the UK in percentage of total farmed area	3
Figure 1.3	Histology of the wheat grain	9
Figure 1.4	Separation of gluten protein on Lactate and SDS-PAGE	14
Figure 1.5	Chromosomal location of the gluten protein loci in bread wheat	15
Figure 1.6	Example of a farinogram of a dough expressed in BU per minutes	17
Figure 1.7	Example of an Extensograph trace of a dough expressed in BU per centimetres	18
Figure 1.8	Loaves heights in mm of seven Australian cultivars: Banks, Hartog, Osprey, Queens, Rosella, Sunbri, and Yanac	21
Figure 1.9	Effect of temperature and drought on accumulation of gluten protein transcripts in developing kernels supplied with post anthesis fertilizer	24
Figure 1.10	Example of a simple linear regression between GPC and GY	27
Figure 1.11	Assimilation of inorganic N into amino acids in wheat	32
Figure 2.1	Selection scheme used to produce the NILs	38
Figure 2.2	NILs split – plot design representing the organization of the NILs within the field trials	41
Figure 2.3	Bühler MLU 202 laboratory mill and its associated flow chart	44
Figure 2.4	Representation of a sample colour in the L*a*b* 3D colour space	48
Figure 2.5	Example of a gluten protein chromatogram obtained after separation by SE-HPLC using the Profilblé® method	50
Figure 2.6	Barley ear at heading date	61
Figure 2.7	Four-wheel platform equipped with a NDVI sensor to measure senescence	62
Figure 2.8	Example of a senescence curve fitting the NDVI against the thermal time	65
Figure 3.1	Pedigrees tree showing the ancestors of Hereward and Malacca	67
Figure 3.2	Dendrogram of the NILs Euclidian distances derived from their Axiom genotypes	69

Figure 3.3	SDS-PAGE gels showing the separation of the gluten proteins of the six NILs sets	71
Figure 3.4	Bar plots comparing the allelic pairs means for L*, cell number, and bake height of the six NILs set in Year 1	74
Figure 3.5	Structure of the lipoxygenase gene TRAESCS4D02G294100 ( <i>LOX3</i> )	78
Figure 3.6	Amount of TRAESCS4D02G294100 transcripts in the tissues of wheat	80
Figure 3.7.	Simple linear regressions of the rheological parameters extensibility and resistance to extension against SE-HPLC fractions and ratios F3+F4/F1, F3+F4/F1+F2, F1, and F1/F2, respectively.	90
Figure 4.1	Simple linear regressions between the percentage of nitrogen in the grain and the grain yield in tonnes/hectares	93
Figure 4.2	Boxplots of doubled haploids GY in Years 1, 2, and 3	94
Figure 4.3	Boxplots of doubled haploids GPC in Years 1, 2, and 3	95
Figure 4.4	Simple linear regressions GPC-GY in Years 1, 2, and 3	96
Figure 4.5	Barplots of the GPD means in Years 1, 2, and 3	98
Figure 4.6	Additive main effects and multiplicative interactions (AMMI) models for GPD, GPC, and GY	102
Figure 4.7	Co-locations of single-environment QTL on the M x H genetic map	106
Figure 4.8	Plot of standardised residuals from the regression GPC-GY against GY in q/ha	108
Figure 5.1	Line chart showing the average monthly temperatures, rainfalls, and solar radiations withstand by doubled haploids lines in Years 2 and 3	113
Figure 5.2	Barplots of the doubled haploids means HD in Years 2 and 3	114
Figure 5.3	Barplots of the doubled haploids means TT90 in Years 2 and 3	117
Figure 5.4	Barplots of the doubled haploids means TT50 in Years 2 and 3	119
Figure 5.5	Barplots of the doubled haploids means TT10 in Years 2 and 3	121
Figure 5.6	Barplots of the doubled haploids means NDVI90 in Years 2 and 3	123
Figure 5.7	Barplots of the doubled haploids means NDVI50 in Years 2 and 3	124
Figure 5.8	Barplots of the vmeans NDVI10 in Years 2 and 3	126
Figure 5.9	Co-locations of single-environment QTL on the M x H genetic map observed on linkage groups 1A and 3B	132

Figure 5.10	PCA plots of the seven senescence metrics in Year 2	134
Figure 5.11	PCA plots of the seven senescence metrics in Year 3	136

## List of Tables

Table1.1	Classification of common wheat varieties in the UK	6
Table 1.2	Histological composition of mature grain of two wheat cultivars	8
Table 1.3	Average percentage of total DF, arabinoxylans, $\beta$ -glucans cellulose, klason lignin, and fructan in the whole grains of 129 wheat varieties	11
Table 1.4	Effect of GPC on the loaf volume in cm <sup>3</sup>	19
Table 1.5	Quality score assigned to individual HMW-GS at the <i>Glu-1</i> loci	20
Table 1.6	Correlation coefficients for the relationship between heat stress and protein content, yield, dough rheology and bread quality measured on five cultivars grown in three environments	23
Table 1.7	Effect of N and N + P fertilisation on the grain yield reported in 11 studies	25
Table 1.8	Quantity of NO <sub>3</sub> <sup>-</sup> leached in different cropping systems	26
Table 1.9	Genotypes with significant GPD in bi-annual and multi-sites trials conducted between 1994 and 2002 by the “Groupe d'Etude et de contrôle des Variétés Et des Semences” (GEVES)	28
Table 1.10	Summary of GPD QTL found in two studies	30
Table 2.1	NIL crossing parents, QTL and markers informing the marker assisted selection and names and numbers of generated NILs streams and individual NIL number	40
Table 2.2.	Agronomy of the NILs including nitrogen fertilisation rate and pesticide application	41
Table 2.3.	Composition of the 12 NILs allelic bulks in the Rothamsted 2019-2020 field trial and of the 4D-2 NILs allelic bulks in the Rothamsted 2020-2021 field trial	46
Table 2.4.	Agronomy of the DH including nitrogen fertilisation rate and pesticide application	57
Table 2.5	Dates after heading where NDVI was recorded in Years 2 and 3	62
Table 3.1	Differences in the proportions of $\omega$ -gliadins, LMW-GS, HMW-GS and in the ratio HMW-GS/LMW-GS between the allelic pairs of the six NILs sets	72
Table 3.2	QTL chromosomal locations, associated quality traits and favourable alleles	73

Table 3.3	Assessment of QTL 4D-2 effect on flour and crumb whiteness	76
Table 3.4	R-Squares and adjusted R-squares of eight regression models	77
Table 3.5	SNP variants identified in the Lipoxygenase gene TRAESCS4D02G294100	79
Table 3.6	Results of the SE-HPLC performed on the six NILs sets flours in Year 1	82
Table 3.7	NILs Farinograph results for four parameters: dough development time, stability, water absorption and degree of softening	83
Table 3.8	Extensograph results for NILs 1B, 2D, and 4D-1 in Year 1	84
Table 3.9	Extensograph results for NIL 4D-2 in Year 2	85
Table 4.1	Summary statistics for the traits GPD measured in the three years	99
Table 4.2	Correlations between GPD and GPC, GY, SW and TKW in the three years	100
Table 4.3	Environmental means and variances of the doubled haploids population for the traits GPD, GPC and GY and correlation of the traits between the three years	101
Table 4.4	Genotypes ASV and mean across the three environments for the trait GPD	103
Table 4.5	Positions of the QTL mapped on the Malacca x Hereward genetic map	105
Table 4.6	Comparison of the QTL detected in the M x H doubled haploids population with the literature	111
Table 5.1	Meteorological data for the doubled haploids trials in Year 2 and 3	113
Table 5.2	Pearson correlations between HD, and the seven senescence metrics, GY, TKW, GPC and GPD in the Years 2 and 3	127
Table 5.3	Pearson correlations between GPC and the seven senescence metrics in Years 2 and 3	128
Table 5.4	Pearson correlations between yield (GY and TKW) and senescence metrics in Years 2 and 3	129
Table 5.5	Pearson correlations between GPD and the senescence metrics in Years 2 and 3	130
Table 5.6	Single environment QTL table of the seven senescence metrics	131

## Abstract

Achieving high wheat breadmaking quality and stability in environmental friendly farming systems will allow to optimize bread production and to reduce the negative impacts of inorganic nitrogen fertilisers on the environment.

On the quality aspect, six quantitative trait loci (QTL) effects associated with key quality traits such as the loaf volume and the crumb whiteness were investigated in six NILs of the cross Malacca x Hereward.

The six Near-Isogenic sets of lines (NILs) were grown under two field seasons (2019-2020 and 2020-2021) at Rothamsted in 2m<sup>2</sup> plots. The lines of each NILs sets (10) were arranged in vertical sub-blocks (6) with two repetitions (i.e. two sets of six sub-blocks forming two blocks). Standard farming operations were observed and nitrogen fertilisation was applied at a rate of 200 kg.ha<sup>-1</sup>.

Farinograph and Extensograph tests were carried out on doughs to assess whether the presence of the QTL affects the physical properties of the dough. The QTL traits were measured on bread (loaves volume) and slices of breads (crumb whiteness and number of cells) and the means of the allelic pairs were compared using ANOVA. QTL 4D-2 effect on the whiteness of the crumb was confirmed in the Rothamsted 2019-2020 trials and a candidate gene (TRAESCS4D02G294100) encoding a Lipoxygenase was identified in the QTL region. For the five remaining QTL, the comparisons of the allelic pairs were not significant at a threshold of 5% in the Rothamsted 2019-2020 trial.

Regarding the sustainability part, a complex interval mapping (CIM) QTL analysis was performed to dissect the genetic architecture of the traits grain protein deviation (GPD), grain yield (GY), and grain protein content (GPC), and seven senescence metrics using 109 doubled haploid (DH) lines of Malacca x Hereward as mapping population.

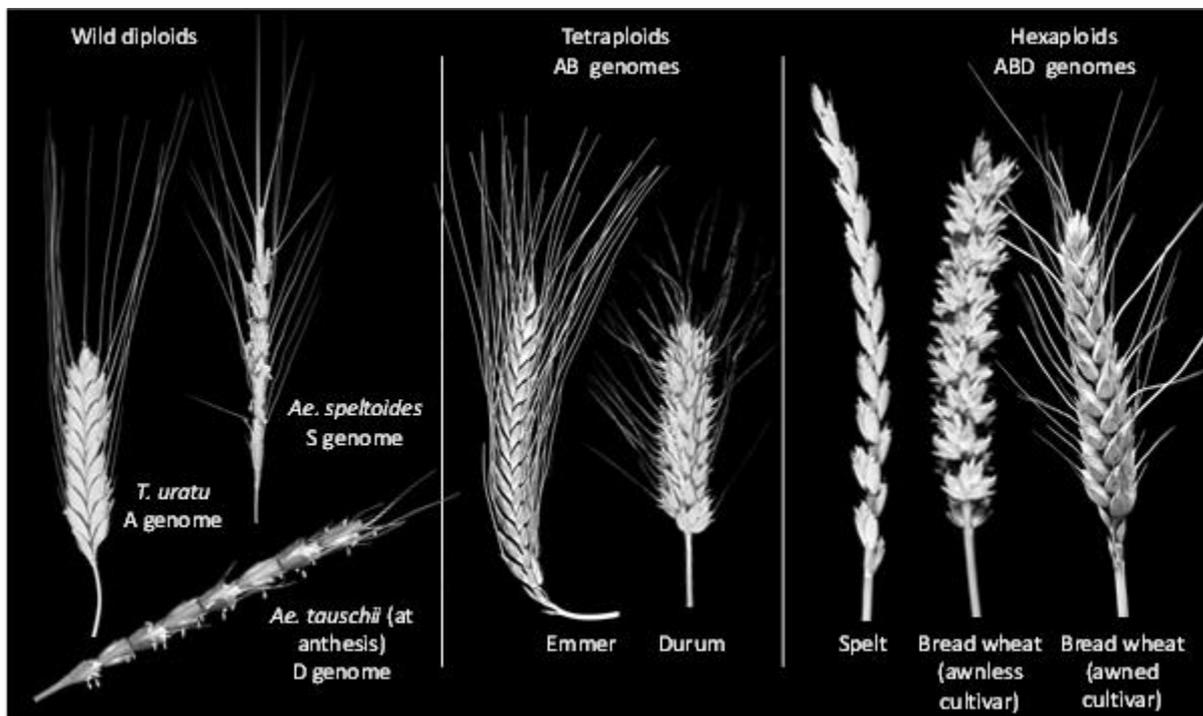
The 109 double haploid lines of the cross Malacca and Hereward were grown under three field seasons at Rothamsted in 2019-2020 and Reading 2020-2021 and 2021-2022 in 7.47m<sup>2</sup> plots. In Rothamsted, the lines were arranged in an balanced incomplete block design with three blocks of size 100 and in Reading they were laid out in a complete randomised block design with three blocks of size 109. The lines were sown at a density of 250 seeds.m<sup>-2</sup> and fertilised with 150 kg N.ha<sup>-1</sup>. A total of 52 QTL were detected for all traits in all environments.

Seven GPD QTL were discovered and one was detected over two field seasons without pleiotropic effect on GY.

## Chapter 1. Introduction

### 1.1. The genesis of bread wheat

Bread wheat (*Triticum aestivum*) has an allohexaploid ( $2n=6x=42$ ) genome which arose from two polyploidization events. The first one probably happened about half a million years ago by a natural cross-hybridization between two diploid species: *Triticum urartu* (AA genome) and an *Aegilops speltoides*-related species (BB genome) and resulted in the tetraploid species *Triticum turgidum* spp. *dicoccoides* (AABB). The second event which gave birth to *Triticum aestivum* (AABBDD), occurred also naturally 10000 years ago from a cross between cultivated emmer and *Aegilops tauschii* (DD) (Gooding and Shewry, 2022) (Figure 1.1).



**Figure 1.1** Bread wheat and its tetraploid and diploid wild ancestors ear phenotypes

Reference: Gooding and Shewry, 2022.

Afterwards, millennia of cultivation involving natural selection, visual selection practiced by farmer and genetic breeding have shaped the *Triticum aestivum* genome into its modern-day characteristics. At first, farmers applied empiric and unconscious selection by selecting traits which facilitated harvest, such as the loss of spike shattering at maturity and glume adherence (Shewry, 2009).

Today modern bread wheat varieties are the result of natural selection, which was driven by the environmental conditions, empiric selection carried out by farmers to facilitate harvest such as the selection of genotypes with a mutation preventing spike shattering or glume adherence or genotypes with an improved seed yield, and modern breeding.

## 1.2. Wheat in figures

Currently, two species of wheat predominate; bread wheat (*Triticum aestivum*), which accounts for 95% of the wheat production, and pasta wheat (*Triticum turgidum ssp. durum*), which represents about 5% of the production. In 2017, the global wheat production reached 772 million metric tons ranking making wheat the second crop after maize. Wheat production continues to increase to meet the demand of the population and is expected to reach 840 million metric tons in 2030 (OECD/FAO, 2021), with significant increase in production in Asia, especially India and Russia. In 2019, the world cultivated wheat area covered 218 million ha which corresponds approximately to the whole of western European countries (Giraldo *et al.*, 2019).

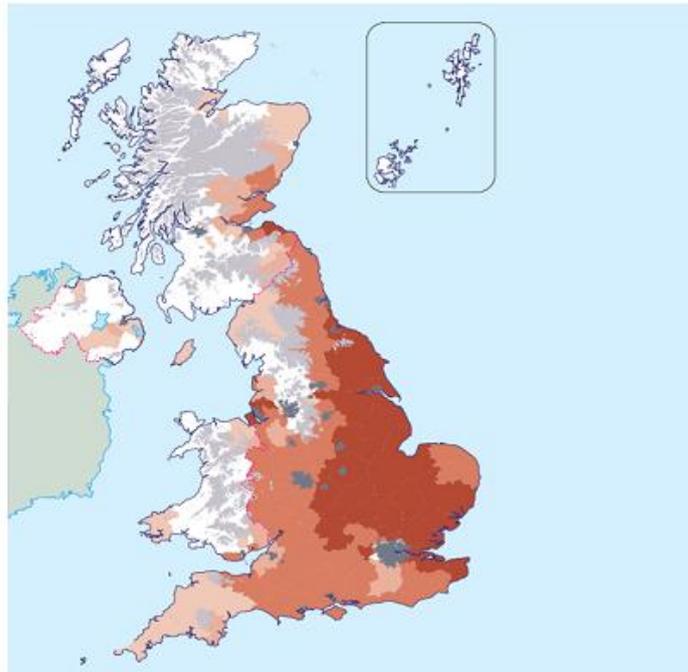
In the 2022-2023 period, China, the European Union, and India were the top wheat producers providing respectively 137, 134, and 103 million metric tons (Statista, 2023).

Regarding trade, USA, Canada, and France are major exporters while Russia, China and Japan are mostly importing to meet the demand of their growing population (FAO, 2021).

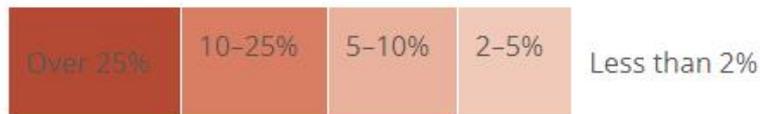
At the UK level, the importations of wheat grain have significantly decreased since 1961 where imports represented four million tons (OECD/FAO, 2021). In 2021, 80% of the wheat purchased by UK millers was home-grown and complemented by 20% of high-quality bread wheat from Canada or Europe (UK Flour Millers, 2021). The UK also exports low quality feed wheat with 261 thousands tons being shipped to the EU, the Republic of Ireland or the Netherlands in 2022 (AHDB, 2022).

Last year, the UK produced 15.5 million metric tons of wheat in 2022-2023 (Statista, 2023). The production is essentially localised in the east in an area delimited by East Anglia on the south and Yorkshire on the north (Figure 1.2, dark red). Wheat is also farmed to a lesser extent in the midlands, the west, and the north (Figure 1.2, light red).

## Wheat



### Area



**Figure 1.2.** Wheat cultivation in the UK in percentage of total farmed area.

Reference: AHDB, 2012

### 1.3. Future challenges for wheat production

Wheat is a staple food in many countries providing 20% of the daily calories and protein. It is also cultivable under a wide range of environments and conditions (Reynolds *et al.*, 2012).

Today, 8 billion humans are inhabiting the planet, around three-times as many as in 1950. Moreover, the population is forecasted to increase to 8.5 billion in 2030 and to 9.7 billion in 2050 (United Nations Department of Economic and Social Affairs, Population Division, 2022) and will therefore inflate the wheat demand by 12% in 2030 in comparison to the average 2018-2020 (OECD/FAO, 2021). To meet that demand, wheat yields will have to be increased as the surface area is limited. Tremendous progress has been achieved in the last century with

the introduction of high yielding semi-dwarf wheat cultivars resistant to lodging. This work was initiated in 1935 by the Japanese scientist Gonjiro Inazuka with the introduction of dwarf genes (*Rht1* and *Rht2*) from a Japanese semi-dwarf wheat landrace into elite American varieties and resulted in the improved variety Norin 10. Inazuka's work was later continued by Orville Vogel and Norman Borlaug; the latter who bred high yielding, short and stiff-strawed cultivars resistant to lodging (Ogihara *et al.*, 2016, p.13).

However, currently the wheat demand is increasing faster (1.7% per annum) than the productivity (1.1% per annum) (Reynolds *et al.*, 2012). In addition, climate change poses a real threat to the productivity (Reynolds *et al.*, 2012) impacting the wheat crop cycle. In the UK, for instance, the decade 2009-2019 was on average 0.3°C warmer than the 1981-2010 period and 0.9°C warmer than 1961-1990 average. By the end of the century, the UK may experience hotter and dryer summers as well as warmer and wetter winters according to the overland projection model UKCP18 forecast. Hot bursts (temperature above 30°C for two consecutive days) are also predicted (Met Office, 2021). In addition, diseases such as Fusarium head blight and rust can cause significant yield losses and quality issues. Fusarium can cause a global yield reduction up to 21.5% globally and threaten human health because of the production of mycotoxins (Buerstmayr *et al.*, 2019).

Finally, the increasing prices of agricultural inputs (i.e. fertiliser and pesticides), and the scarcity of water available for irrigation are also threatening the wheat production (Langridge *et al.*, 2022)

To meet the future demand, the Wheat Yield Consortium has announced a target of 50% increase in yield genetic gain for the period 2012-2032 which focus on three themes: improve the photosynthetic capacity and efficiency, optimize the nitrogen partitioning into the grain, and accumulate yield traits through breeding (Reynolds *et al.*, 2012).

## 1.4. Breadmaking in the UK

### 1.4.1. Introduction

In the following section we will refer to bread with high specific volume in distinction to low specific volume breads (e.g., flat bread), as it is the predominant form in which it is consumed in the UK.

The UK bread sector is dominated by large plant bakers who share 75% of the value and 85% of the production leaving the remaining part to in-store bakeries and craft bakers (Federation of Bakers Ltd, n.d.). The three major companies in the sector are Allied Bakeries (Kingsmill), Hovis, and Warburtons and account for three quarters of the bread sales (Federation of Bakers Ltd, n.d.).

Industrial UK breads are of various types such as white bread made with white flour, brown breads, a blend of white flour and bran), wholemeal (grinded grain), and seeded (Brinsden *et al.*, 2013). White bread is the principal type of bread consumed in the UK but consumption faced a steady decline from 686 grams per person per week in 2006 to 519 grams per person per week in 2016/2017. Brown bread sales are also decreasing but at a slower rate from 184 grams per person per week to 146 grams per person per week (DEFRA, 2018).

### 1.4.2. Classification of wheat according to end-use

Wheat variety grown and commercialised in the UK must be registered in the National List. This list is composed of four end-use groups: two breadmaking groups, one cake and biscuit group, and one feed group (Table 1.1.).

The two breadmaking groups (group 1 and 2) contain cultivars with a high protein content (>13%) and showing consistent milling and breadmaking performance over year in the case of group 1.

The cake and biscuit group encompass varieties with a protein content between 11% and 11.5% showing extensible gluten and the feed group (group 4) regroups varieties that did not meet criteria of group 1,2 or 3 and is intended for animal feed (Table 1.1.).

**Table1.1.** Classification of common wheat varieties in the UK.

Classification	Qualities and uses
UK Flour Group 1	Bread-making varieties with consistent milling and baking performance. They will achieve a premium if they achieve specified quality requirements of 13% protein, 250s Hagberg Falling Number and 76kg/hl specific weight.
UK Flour Group 2	Varieties with bread-making potential but not suited to all grists because of variability in performance or some undesirable traits.
UK Flour Group 3	Soft varieties used for biscuits, cakes etc. They are lower in protein (11.0 – 11.5%), have good extraction rates and extensible but not elastic gluten.
UK Flour Group 4	These are both hard and soft wheats used mainly for animal feed. Millers may use some varieties in general purpose grists.

Reference: Nabim, 2019

#### 1.4.3. The Chorleywood breadmaking process

The Cambridge dictionary defines bread as “a food made from flour, water, and usually yeast, mixed together and baked” (Cambridge University Press, 2023). The purpose of the breadmaking process is to achieve “a light, aerated and palatable food” (Cauvin and Young, 2006, p.1).

Nowadays, industrial UK bread is exclusively manufactured following the Chorleywood Bread Process (CBP).

CBP was developed in 1961 by Chamberlain and colleagues who worked for the British Baking Industries Research Association (BBIRA) based in Chorleywood, Hertfordshire, UK (Cauvin and Young, 2006, p.6).

The CBP method was swiftly adopted by many bakeries after its introduction for its important reduction of processing time and saving of space. In this procedure, the dough is mixed and developed in a single step lasting less than five minutes whereas the bulk fermentation method (which predominated before 1961) requires long fermentation times of 2-3 hours to develop the dough (Cauvin and Young, 2006, pp. 6,17). In CBP, the dough is developed under high-speed mixing for a set amount of time. The weight of the dough is used to estimate the energy input and mixing time. In addition, the mixer headspace pressure is adjusted to control air bubble size. The bread recipe is also enriched with improvers such as oxidising agents to increase the gas retention in the dough (Cauvin and Young, 2006, pp. 6,17).

After mixing the bulk dough is divided in equal portions by filling a defined volume in a divider chamber and rounded. This is followed by a step of proofing in which the dough is allowed to rest. The dough rheology is evolving during resting as the dough is relaxing. The aim of this step is to ensure a good gas bubble structure and to prepare the dough, giving it sufficient extensibility, for the final moulding in where the dough is incorporated into a mould fitting the end-product shape.

Finally, the dough is proofed a second time to increase its volume and baked. The baking will set the dough structure and convert it from a foam texture in which the air bubbles are separated to one another into an open crumb structure called a sponge with interconnected cells (Cauvin and Young, 2006, pp. 58-64).

### **1.5. Histology of the wheat caryopsis**

The wheat grain is a single seeded dry fruit, called a caryopsis because the pericarp tightly adheres to the seed coat (Khan and Shewry, 2009, p.52).

The histology of the wheat caryopsis is presented in Figure 1.3.

The caryopsis is composed of four parts: the pericarp, the seed coat, the endosperm, and the embryo. The pericarp is derived from the ovary cell walls and is the outermost tissue. It can be further divided into the outer (15-30µm thick) and inner pericarp. The seed coat is located just below the pericarp and is 5-8 µm thick. More internally, the endosperm is comprised of the aleurone layer, which is a single layer of cuboid-shaped cells of 65 µm in diameter, and

**Table 1.2.** Histological composition (% dry weight) of mature grains of two wheat cultivars (Caphorn and Crousty). Reproduced from Barron *et al.* (2007).

	Caphorn	Crousty
Embryo	3.0	3.2
embryonic axis	1.5	1.7
scutellum	1.5	1.5
Endosperm	89.2	90.1
starchy endosperm	82.7	83.7
aleurone layer	6.5	6.4
Outer layers (nucellar epidermis, testa, pericarp)	7.8	6.7

the starchy endosperm, which is a large storage tissue enriched in starch and protein and supplies the developing embryo in nutrients. Finally, the embryo is made of two tissues: the scutellum which is a storage tissue considered to be a single modified cotyledon, and the embryo axis (Khan and Shewry, 2009, pp.56-77).

Barron *et al.* (2007) analysed the tissue composition of two bread wheat cultivars: Caphorn and Crousty. The endosperm (aleurone layer plus starchy endosperm) was by far the heaviest tissue, accounting for 89-90 % of the total grain dry weight (dw), with the starchy endosperm accounting for about 83% dw and the aleurone for 6.5% dw (Table 1.2). The dry weight of the outer layers which include the pericarp and the seed coat ranged between 6.7% (Crousty) and 7.8% (Caphorn). The embryo was the smallest fraction and represented only 3- 3.2 % dw (Table 1.2).

## 1.6. Wheat caryopsis with health benefits or effect on technological properties

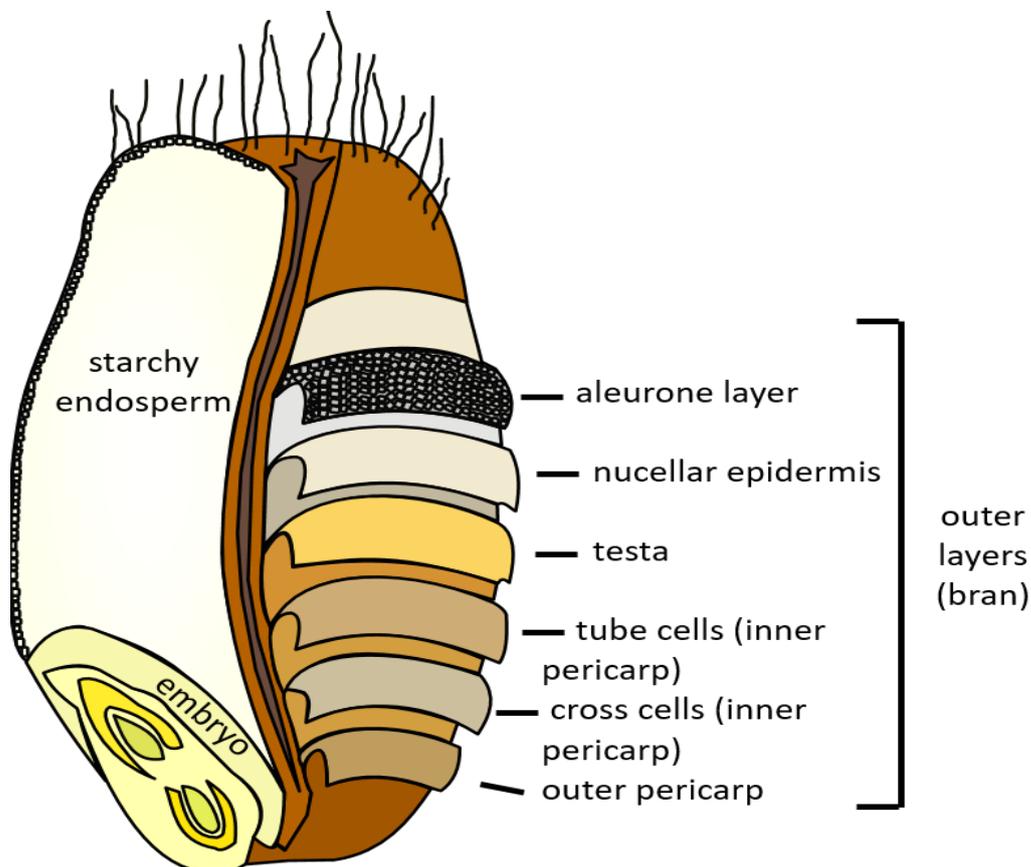
### 1.6.1. Starch and water absorption

The starchy endosperm contains between 70 and 80% dry matter of starch (Gooding and Shewry, 2022).

In wheat starch is present in the form of granules that can be of two types:

- A-granules have a lenticular shape and a diameter above 10  $\mu\text{m}$ ;
- B-granules have a spherical shape and a diameter below 10  $\mu\text{m}$ .

Wheat Starch is a combination of two polymers of glucose molecules: linear amylose linked in  $\alpha$ -1-4 (20-30% of the total polymers), and amylopectin highly branched with  $\alpha$ -1-6 linkages (70-80%).



**Figure 1.3.** Histology of the wheat grain

Reference: Surget and Barron, 2005.

Starch contributes to the water absorption of the flour along with protein and potentially dietary fibres. When undamaged, starch absorbs 0.5 times its weight in water, but damaged, it can absorb up to 3 to 4 times its weight in water (Gooding and Shewry, 2022). Damaged starch occurs during the milling process when the grain is broken between the roller of the mill. The granules size distribution also impacts water absorption as B granules have higher swelling power than A-granule (Park *et al.*, 2009).

### 1.6.2. Dietary fibre (DF)

In 2009, the CODEX Alimentarius Commission who was mandated by The World Health Organization (WHO) defined dietary fibres as: “carbohydrate polymers with ten or more monomeric units which are not hydrolysed by the endogenous enzymes in the small intestine of humans”.

In wheat, DF encompass mainly non-starch polysaccharides derived from the cell walls, but also Klason lignins and fructans (Andersson *et al.*, 2013; Shewry and Hey, 2015). The consumption of DF is associated with several health-benefits, such as reducing the glycaemic index, type 2 diabetes, and the risk of Colorectal cancer (Gooding and Shewry, 2022).

Andersson *et al.* (2013) quantified the total DF and their components in wholegrain of the HEALTHGRAIN diversity screen; a panel of 129 winter and spring wheat lines with landraces, cultivars, and breeding lines. Their analysis showed genetic variation for Total DF and its components. Total DF ranged from 11.5 to 15.5% of dry weight, with arabinoxylan being the major DF (5.53-7.42 % dw), with lower contents of cellulose,  $\beta$ -glucan, Klason lignin and fructan (0.5 to 3.05% of dw) (Table 1.3.). DF are essentially located in the pericarp and the aleurone layer being in very little amount in the starchy endosperm.

Arabinoxylan is unequally distributed within the caryopsis tissues, being enriched (40% of the tissue dry matter) while being scarce in the endosperm (2% of the tissue dry matter) (Gebruers *et al.*, 2008). Likewise,  $\beta$  – glucan is mainly concentrated in the aleurone layer (23% of the cell wall polysaccharide fresh weight) and present in minute proportion in the endosperm (0.3% dry matter) (Gebruers *et al.*, 2008) Cellulose is mainly present in the outer layer of the pericarp representing 25% of the polysaccharide fresh weight (Gebruers *et al.*, 2008).

**Table 1.3.** Average percentage in dry weight of total dietary fibre, arabinoxylans,  $\beta$ -glucans cellulose, Klason lignin, and fructan in the whole grains of 129 wheat varieties.

<b>Dietary Fibre</b>	<b>Whole grain</b>
Total Dietary fibre	11.5-15.5
Arabinoxylan	5.53-7.42
$\beta$ -glucan	0.5-1
Cellulose	1.67-3.05
Klason lignin	0.74-2.03
Fructan	0.84-1.85

Reference: Anderson *et al.*, 2013 (Adapted)

### 1.6.3. Vitamins: essential nutrients

B vitamins provide a range of benefits such as cancer prevention, antioxidant activities and regulation of the immune system.

Wheat provides an important source of B vitamins. Within the caryopsis, vitamins B are essentially located in the bran and the germ. They include thiamin (vitamin B1), riboflavin(B2), niacin (vitamin B3), pyridoxin (vitamin B6), folate (B8). Vitamins B1, B2, B3 and B6 were quantified by Shewry *et al.* (2011) on wholemeal flour samples of 24 wheat cultivars grown at four different sites. Vitamin B1 were present at a concentration of 5.53-13.55  $\mu\text{g/g}$  per dw(vitamin B1), vitamin B2 at 0.77-1.40  $\mu\text{g/g}$  dw, vitamin B6 at 1.27-2.97  $\mu\text{g/g}$  dw while the content of bioavailable vitamin B3 ranged from 0.16 to 1.74  $\mu\text{g/g}$  dw). The level of total folate in wholegrain varies considerably between studies and may be influenced by the cultivar and the protein content (Piironen *et al.*, 2008).

Vitamins are located in the internal tissues of the caryopsis: the aleurone layer, the starchy endosperm and the embryo. The aleurone layer is the richest tissue containing 80% of niacin, 60% of pyridoxin and 32% of thiamine. The starchy endosperm provides a significant part of riboflavin (32%) but is poor in niacin (12%), thiamine (<5%) and pyridoxin (6%) and the scutellum of the embryo essentially provides thiamine (60% of the total thiamine) (Piironen *et al.* 2008).

#### 1.6.4. Lipids and gas cell stability

Gas cells contribute to 70% to the final bread volume (Salt *et al.*, 2018).

Air cells are incorporated in the dough during the mixing process and expanded by the accumulation of carbon dioxide released during yeast fermentation. The stability of the gas cells present in the dough is mainly ensured by the gluten viscoelasticity determining cell retention and expansion, but lipids also contribute thanks to their surface-active properties (Salt *et al.*, 2018). The lipids of the wheat caryopsis can be classified in two categories according to their structure: neutral (acylglycerols and free fatty acids) and polar (glycolipids and phospholipids). Particularly, polar lipids were reported to influence gas bubble stability and loaf volume (Gonzalez-Tuillier *et al.*, 2015; Salt *et al.*, 2018). Salt *et al.*, (2018) found concentration of polar lipids in the dough liquor (liquid phase of the dough) of the cultivar Hereward of 21-67% of the total lipids.

#### 1.6.5. Carotenoids content and flour colour

The flour colour is a paramount quality trait (Parker *et al.*, 1998). For white bread manufacturing the whiteness of the crumb is essential as yellowness could be considered as a defect by the consumer. This was exemplified by the bad sales experienced for the wheat variety Malacca producing a slightly yellow flour (Personal communication, Dr Mervin Poole Heygates Ltd).

Two factors control the whiteness of the flour: the brightness and the yellowness. The flour particles size and the bran content generated by the milling process both influence the brightness (Leenhardt *et al.*, 2006). The content of yellow pigments called carotenoids present in the caryopsis endosperm govern the yellowness. Lutein and zeaxanthin are the two carotenoid pigments found in the caryopsis (Leenhardt *et al.*, 2006).

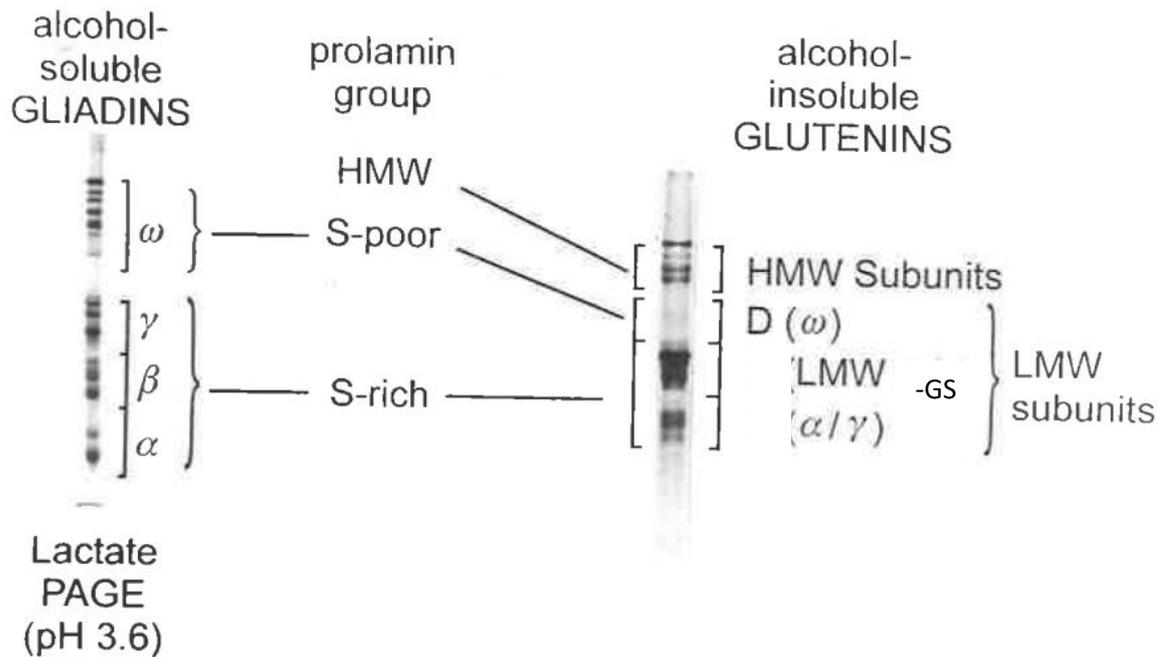
Carotenoid pigments initially present in the flour are degraded during the process by the action of lipoxygenase and heat. Leenhardt *et al.* (2006) monitored carotenoids degradation in a white organic flour occurring during the breadmaking process and observed an important reduction (66%) after kneading which they attribute to lipoxygenase activity. Lipoxygenases are known bleaching agent indirectly involved in carotenoid degradation. Lipoxygenases first react with polyunsaturated fatty acids to form hydroperoxides which in turn react with the carotenoid pigments causing bleaching.

## 1.7. The gluten proteins

### 1.7.1. Classification

Jacomo Becari was the first scientist to isolate and describe wheat gluten in 1745 (Shewry and Halford, 2002). Gluten proteins are seed storage proteins rich in proline and amide nitrogen and are thus also referred as prolamins. Shewry *et al.* (2003a) indicate that the gluten proteins cannot all be separated by a single technique owing to their amount (at least 50 in hexaploid wheat). However, they can be divided into broad groups according to their properties. Before protein sequencing technologies became available, the classification of gluten proteins was based on their solubility, isoelectric point, and molecular weight. Initially, the gluten proteins were classified in two groups by Osborne and Mendel in 1924 according to their solubility in alcohol-water solvent: the gliadins which are monomeric are soluble in ethanol at 60-70% v/v or propan-1-ol at 50-55% v/v whereas the glutenins, which are polymeric, are insoluble unless reduced into monomers (Shewry *et al.*, 2003a, pp. 221-223).

The gliadin and glutenin groups were further divided into four and two categories, respectively, according to their migration on lactate polyacrylamide gel electrophoresis pH 3.6) and sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gels (Figure 1.4.). Lactate-PAGE (Figure 1.4., left) separates the gliadins fraction into  $\omega$ -gliadins (44000-74000 Da on SDS-PAGE),  $\gamma$ - gliadins (38000-42000 Da), and  $\alpha$ -,  $\beta$ -gliadins (32000 Da) (Shewry *et al.*, 1986). SDS-PAGE (Figure 1.4., right) differentiates the glutenin into high molecular weight glutenin subunit (HMW-GS - 95000-136000 Da) and low molecular weight glutenin subunit (LMW-GS - 36000-44000 Da) (Shewry *et al.*, 1986). Thanks to the advances in sequencing, a simplified new classification of the gluten proteins was later proposed by Shewry *et al.* (1986). In this classification, the gluten proteins are classified in three groups: Sulphur-poor (S-poor) prolamins, Sulphur-rich (S-rich) prolamins, and HMW-prolamins with no further divisions. The  $\alpha$ -,  $\gamma$ - gliadins and LMW-GS have similar amino-acids composition (40-50% glutamine, 20-30% proline, and 8-9% phenylalanine) and high sequence homology, have comparable size (250-300 amino-acids) and are enriched in cysteine amino-acids residues (2-3mol.%). They were therefore classified in the S-rich group. In contrast,  $\omega$ -gliadins differ in their amino-acids composition (40-50% glutamine, 20-30% proline, and 8-9% phenylalanine) and lack cysteine residues; they were thus put in the S-poor category. HMW-GS protein sequence contains cysteine residues 0.4-1.5 mol.% but were put in a separate group HMW-prolamin because of their high molecular weight (Shewry and Tatham, 1997).



**Figure 1.4.** Separation of gliadins on Lactate-PAGE (pH 3.6) into four groups of bands (left) and separation of glutenin on SDS-PAGE into two groups of bands HMW-GS and LMM-GS (right).

Reference: Shewry *et al.* 2003a, p.223

### 1.7.2. Genetic architecture

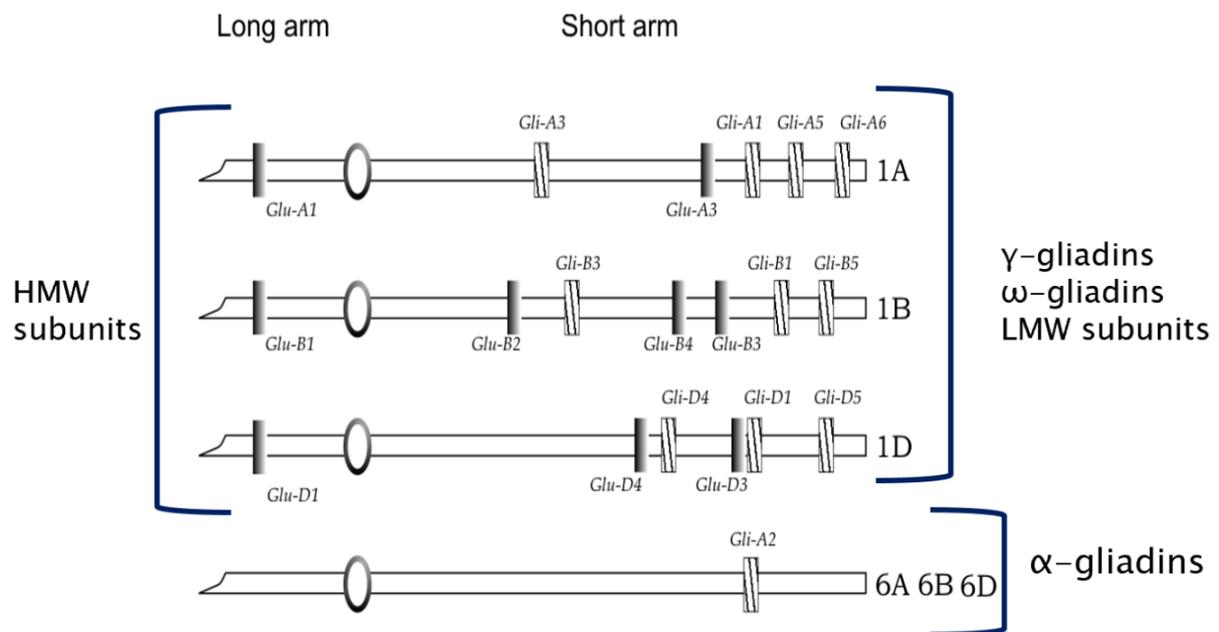
Gluten proteins are encoded at 12 major loci (*Glu-1*, *Glu-3*, *Gli-1*, and *Gli-2*) located on groups 1 and 6 chromosomes homeologues (Branlard *et al.*, 2020)

HMW-GS are encoded by a tandem of genes *Glu-1x* and *Glu-1y* at the complex loci *Glu-1* located on the long arm of chromosome 1 (Figure 6). *Glu-1* is formed of three simple loci named after their genome assignment: *Glu-A1* (Chromosome 1A), *Glu-B1* (Chromosome 1B), and *Glu-D1* (Chromosome 1D) (Payne *et al.*, 1980). The genes *Glu-1x* and *Glu-1y* are tightly linked and encode HMW-GS of high and low molecular weight, respectively. The six HMW-GS-encoding genes of the *Glu-1* loci (*Glu-A1x*, *Glu-A1y*, *Glu-B1x*, *Glu-B1y*, *Glu-D1x*, *Glu-D1y*) are not all expressed and *Glu-A1y* is always silent in hexaploid wheat. In fact, between three and five HMW-GS- encoding genes are expressed depending on the variety as NULL alleles were reported for either or both *Glu-D1* genes (Shewry *et al.*, 2003b, p123).

LMW-GS are encoded at the *Glu-3* loci formed of *Glu-A3*, *Glu-B3*, and *Glu-D3* and located on the short arm of chromosomes 1 homeologues (Figure 1.5.).

$\gamma$ - and  $\omega$ -gliadins are located at the *Gli-1* loci composed of *Gli-A1*, *Gli-B1*, and *Gli-D1* which was also mapped on the short arm of chromosomes 1 homeologues (Figure 1.5.). *Gli-1* and *Glu-3* are tightly linked (distance of 1.3 – 2 cM) although recombination was reported (Shewry *et al.*, 2003b, p127).

$\alpha$ - and  $\beta$ -gliadins are encoded at the *Gli-2* loci on the short arms of the group 6 chromosomes. In addition to these 12 major loci, minor gliadins (*Gli-3*, *Gli-5* and *Gli-6*) and LMW-GS (*Glu-3*, *Glu-4*) loci were also reported on the short arm of chromosome 1 (Sozinov and Poperelya, 1982) (Figure 1.5.).



**Figure 1.5.** Chromosomal location of the gluten protein loci in bread wheat. The position of the centromere is indicated by a circle.

Reference: Shewry *et al.*, 2003b, p.129

## 1.8. Breadmaking quality

### 1.8.1. Physical dough testing

Wheat dough exhibits visco-elastic properties that allow it to be processed in a variety of products such as cookies, puff pastries, pastas, and breads. Bread is made with highly elastic (strong) dough (Shewry *et al.*, 2002).

The physical properties of wheat dough and gluten are studied using rheology, which can be defined as “the study of how materials deform, flow or fail when a force is applied” (Amjid *et al.*, 2013). Typically, the material is subjected to a force (e.g., strain or distortion) for a given amount of time to deduce its properties (e.g., stiffness, elasticity, strength, viscosity). In the breadmaking industry, rheology may be applied either to predict the dough behaviour at various stages of the process (e.g., mixing, rounding, moulding, proving, and baking), to study the effect of an additive on the dough behaviour, or to categorise the flour and dough as strong, medium, or weak (Amjid *et al.*, 2013).

Wheat dough rheology is assessed by a Farinograph to determine its strength (i.e., resistance to mixing) and water absorption. The instrument is a blade mixer recording a change of strength over time as the dough develops or breakdown (Figure 1.6). As the dough is mixed, the flour constituents become hydrated which allow the gluten network to form. This is associated to a change in the dough viscosity which increase to a peak before progressively decreasing in the breakdown phase (destruction of the gluten network) (Oliver and Allen, 1992).

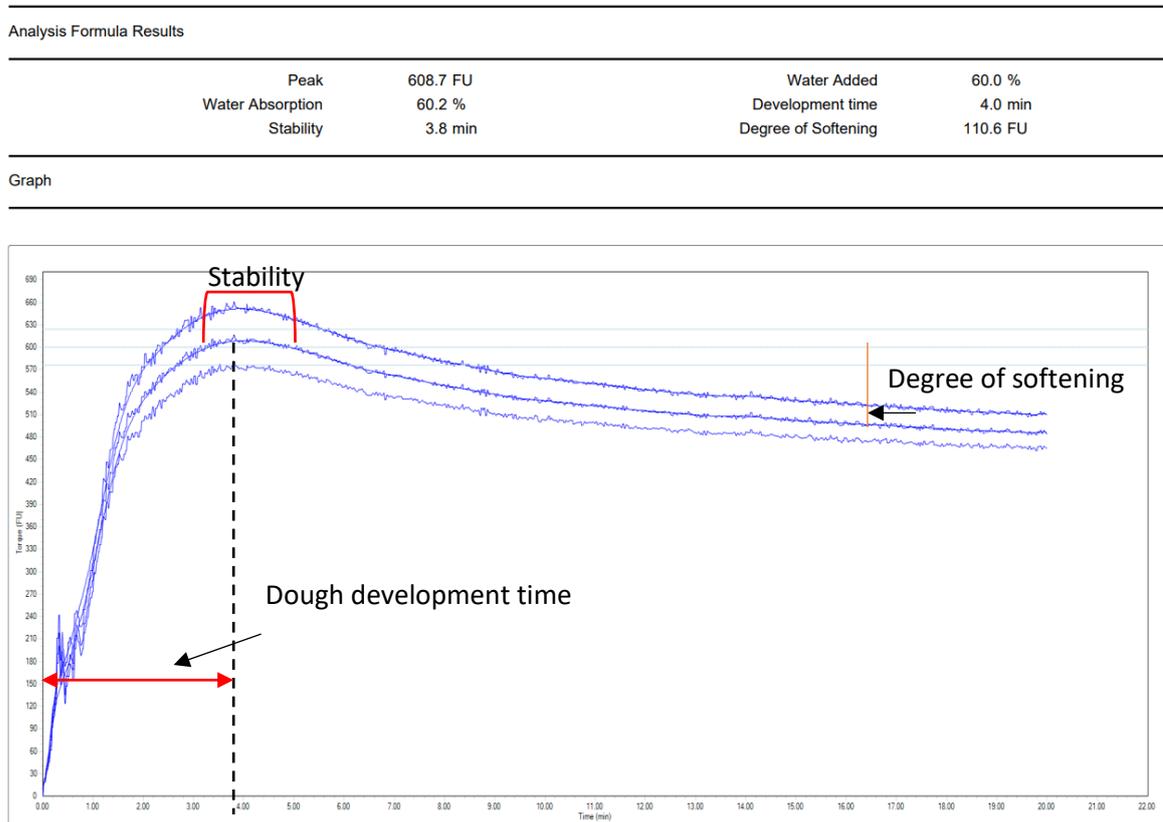
The Farinograph is employed for quality assessment of the dough and discriminates between dough types; a weak flour has a short dough development time (i.e., a sharp rising curve), a low stability and a high degree of softening (i.e., a quick and important decrease after the reach of the peak) in comparison to a strong flour (Wrigley *et al.*, 2004, p. 405). It is also used for prediction of the dough behaviour in the bakery (Oliver and Allen, 1992).

In addition to the Farinograph test, the dough is also analysed at the bakery using an Extensograph.

The Extensograph subjects the dough to a tensile stress or stretch to study its resistance to extension and extensibility (i.e., maximum length in centimetres the dough can be stretched before breaking). Like the Farinograph, the Extensograph can be used to differentiate among flours. For instance, weak doughs with high extensibility are suitable for biscuit making

whereas strong doughs are better suited for pasta making. For breadmaking, doughs should exhibit moderate strength and high extensibility (Bangur *et al.*, 1997).

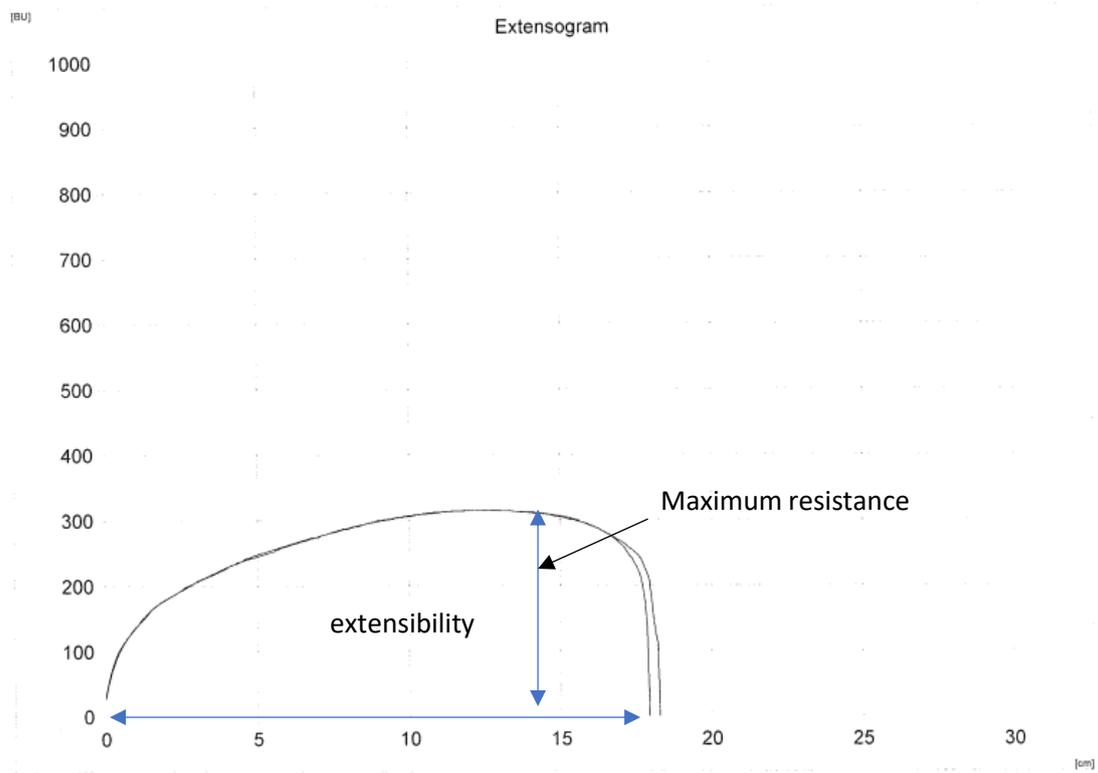
Figure 1.7. displays an Extensograph trace of a dough prepared from wheat flour at Heygates Ltd during this PhD.



**Figure 1.6.** Example of a farinogram of a dough expressed in BU (Brabender unit) per minutes. Three phases are shown: the dough development time which corresponds to the time elapsed in minutes between the first addition of water and the reach of the maximum consistency, the stability which is the amount of time in minutes the dough consistency stays above the 600 BU line and the degree of softening which measures the loss of dough consistency in BU 12 min after the peak. This farinogram was obtained at Heygates Ltd during this PhD.

Energy [cm <sup>2</sup> ]:	83
Resistance to Extension [BU]:	248
Extensibility [mm]:	181
Maximum [BU]:	315
Ratio Number:	1.4
Ratio Number (Max.):	1.7

Remarks:	Remark 1
	Remark 2



**Figure 1.7.** Example of an Extensograph trace of a dough expressed in BU per centimetres. The flour was analysed in duplicate. This Extensograph trace was obtained at Heygates Ltd during this PhD.

## 1.8.2. Gluten proteins quality

### Introduction

Wheat is the only cereal which can be baked into a bread of high specific volume (Shewry, 2019). This feature is conferred by the gluten proteins present in the flour which form a continuous network (the gluten) upon hydration ensuring cohesion and visco-elasticity (Shewry, 2019). Gliadins provide viscosity while glutenin give elasticity (Uthayakumaran *et al.*, 1999). Good dough physical properties are achieved by a combination of four parameters: the HMW-GS subunit composition, the gliadin-to-glutenin ratio, the molecular weight distribution of glutenin polymers and the protein content.

#### 1.8.2.1. Protein content

A positive linear relationship was reported between the protein content and the loaf volume for a range of protein between 8-18% (Finney, 1948; as cited in He and Hosney, 1992). This relation was later confirmed by He and Hosney, (1992) who prepared three flours with protein fraction of (7%, 8.5% and 10%) from the same commercial base flour (11.5% protein) by adding purified starch and measured their loaves volumes. They reported increase from 707 cm<sup>3</sup> (flour at 7% protein) to 932 cm<sup>3</sup> (11.5% protein) (Table 1.4.).

This difference of loaf volume occurred during the baking as the doughs with high protein content expanded faster than the doughs with a low protein content.

**Table 1.4.** Effect of protein content in % on the loaf volume in cm<sup>3</sup>

<b>Flour Protein (%)</b>	<b>Loaf Volume<sup>a</sup> (cm<sup>3</sup>)</b>
7.0	707 ± 12.6
8.5	818 ± 10.4
10.0	903 ± 5.8
11.5	932 ± 11.5

<sup>a</sup> ± Standard deviation.

### 1.8.2.2. HMW-GS subunit composition

HMW-GS composition is paramount and was shown to explain 55-77% of the breadmaking quality (Payne *et al.* 1987)

These authors developed a *Glu-1* quality score relating flour quality assessed by a SDS-sedimentation test to the HMW-GS subunit composition at the *Glu-1* loci (Table 1.5.).

The sedimentation test assesses the density and the swelling capacity of the gluten. The gluten of a strong flour absorbs more water at faster rate than the gluten of a weak flour and give larger SDS volume (Payne *et al.*,1981).

The HMW-GS subunits of 84 British-grown varieties were separated on SDS-PAGE and individual *Glu-1* loci rated on a 1 -10 depending on the flour sedimentation volume.

In Table 1.5, varieties with subunit 1 or 2\* at the *Glu-A1* loci had larger SDS-volume that those with the null allele and were therefore granted with a higher score.

**Table 1.5.** Quality score assigned to individual HMW-GS at the *Glu-1* loci. The HMW-GS of *Glu-A1*, *Glu-B1*, and *Glu-D1* are numbered according to their electrophoretic mobility on SDS-PAGE.

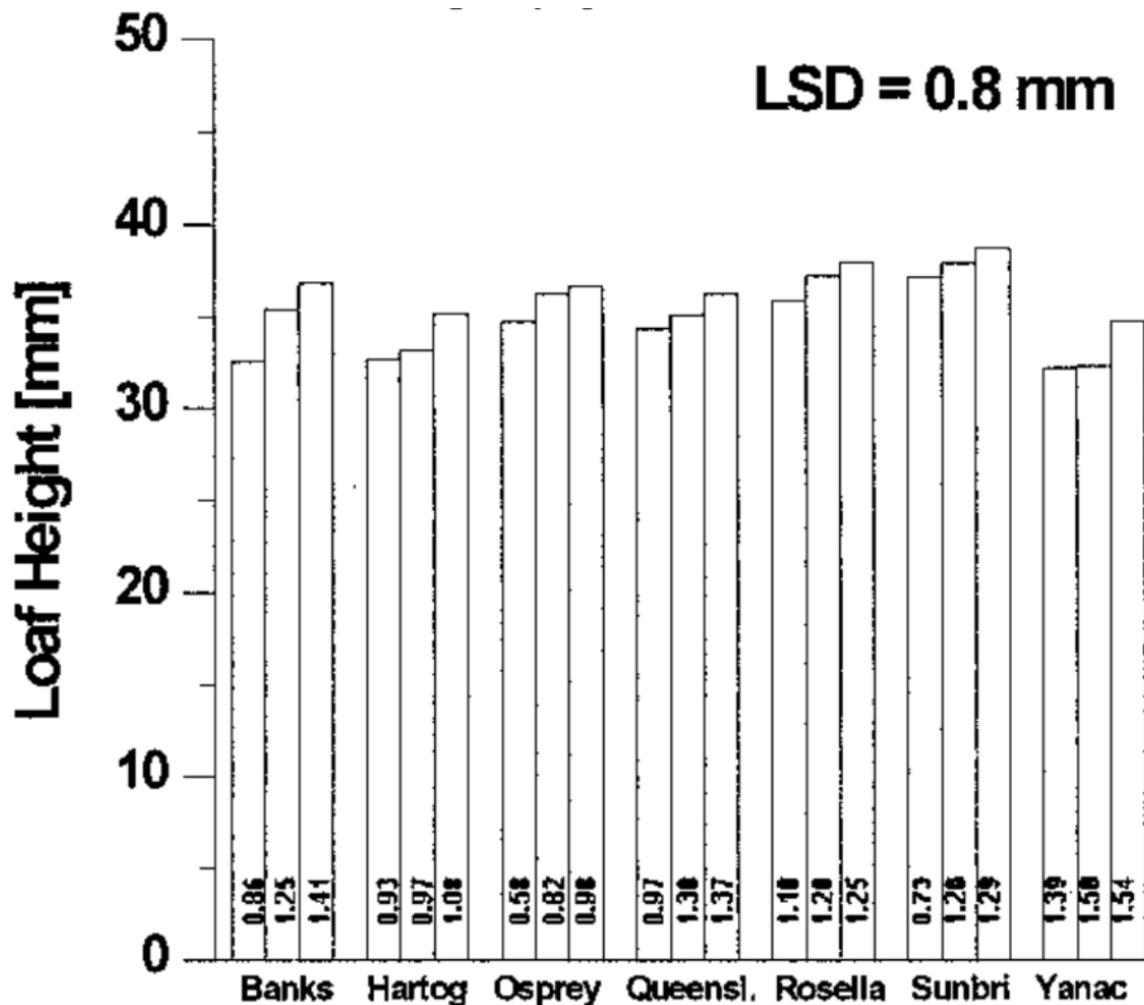
<b>Score</b>	<b>Chromosome</b>		
	<b>1A</b>	<b>1B</b>	<b>1D</b>
4	—	—	5+10
3	1	17+18	—
3	2*	7+8	—
2	—	7+9	2+12
2	—	—	3+12
1	null	7	4+12
1	—	6+8	—

Reference: Payne *et al.*, 1987

### 1.8.2.3. Ratio of glutenin-to-gliadin

Uthayakumaran *et al.*, (1999) observed a change in the dough rheology and in the loaf volume when the glutenin-to-gliadin ratio of the flour was altered for seven Australian cultivars: Banks, Hartog, Osprey, Queens, Rosella, Sumbri and Yanak. Increasing the glutenin content resulted in stronger and more stable doughs and increased the loaves volumes with variation

among the seven cultivars (Figure 1.8). This increase in the dough strength consequently to an increase of glutenin was also observed by MacRitchie, (1992).



**Figure 1.8.** Loaves height in mm of seven Australian cultivars: Banks, Hartog, Osprey, Queens, Rosella, Sunbri, and Yanac. Three glutenin-to-gladin ratios were tested for each cultivar and are provided in the bars. The LSD indicates the Least Significant Difference between the bars.

#### 1.8.2.4. Glutenin polymers size distribution

The glutenin polymers present in the flour are highly variable in size (Veraverbeke and Delcour, 2002).

Glutenin quality has been shown to be positively correlated to the relative proportion of large size polymers determined by gel permeation chromatography in many studies (Bottomley *et al.*, 1982; Godfrey *et al.*, 2010; Gupta *et al.*, 1993; Huebner *et al.*, 1976; Huebner *et al.*, 1985 as cited in Veraverbeke and Delcour, 2002) but one study found a negative correlation

(Dachkevitch *et al.*, 1989, as cited in Veraverbeke and Delcour, 2002). It is believed that the polymeric nature of glutenins allowing the formation of a large network is responsible for their elastic properties (Veraverbeke and Delcour, 2002).

Godfrey *et al.* (2010) used the Profilblé® method (Morel *et al.*, 2020) to separate and study the glutenin polymers size distribution of the cultivar Hereward grown under six nitrogen regimes from 0 to 288 kg N. ha<sup>-1</sup> on the dough rheology. The Profilblé® procedure separates the gluten protein in four fractions of decreasing molecular sizes: high and low molecular weight glutenin polymers (Fraction F1 and F2), high and low molecular weight gliadins (F3 and F4), and albumins and globulins (F5). They reported significant and positive correlations ( $r=0.74$  and  $0.68$ , respectively) between the large glutenin polymers fraction (F1) and the ratio large-to-small glutenin polymers (F1/F2) and the Extensograph strength. Similarly, the ratios gliadins-to-large glutenin polymers (F3+F4/F1) and gliadins-to-glutenin polymers (F3+F4/F1+F2) were significantly and highly positively correlated with the Extensograph extensibility ( $r=0.88$ ).

## 1.9. Effect of the environment on the synthesis of gluten proteins

### 1.9.1 Effect of temperature.

High temperatures were shown to impact both protein content and quality, and to affect the dough rheology.

Temperature above 30°C during the grain filling reduced both the absolute amount of protein and starch in comparison to temperatures ranging from 15-21°C. However, the decrease of starch content is more important than the decrease of protein content resulting in a higher protein content in % of dry weight (Blumenthal *et al.*, 1993).

Heat stress increased gliadin-to-glutenin ratio in the grain from 0.05 to 0.16 for six cultivars grown in a single environment where the heat stress occurred at the middle of the grain filling period in comparison to a non-stressed environment (Blumenthal *et al.*, 1993).

Heat stress (as cumulated hours above 35°C) was also shown to be negatively correlated with the dough strength (Extensograph  $R_{max}$  parameters) and the loaf volume in five Australian cultivars grown at three locations (Table 1.6.). The size of the correlation depends on the cultivar and the site and ranged from -0.22 to -0.68 for loaf volume (all sites).

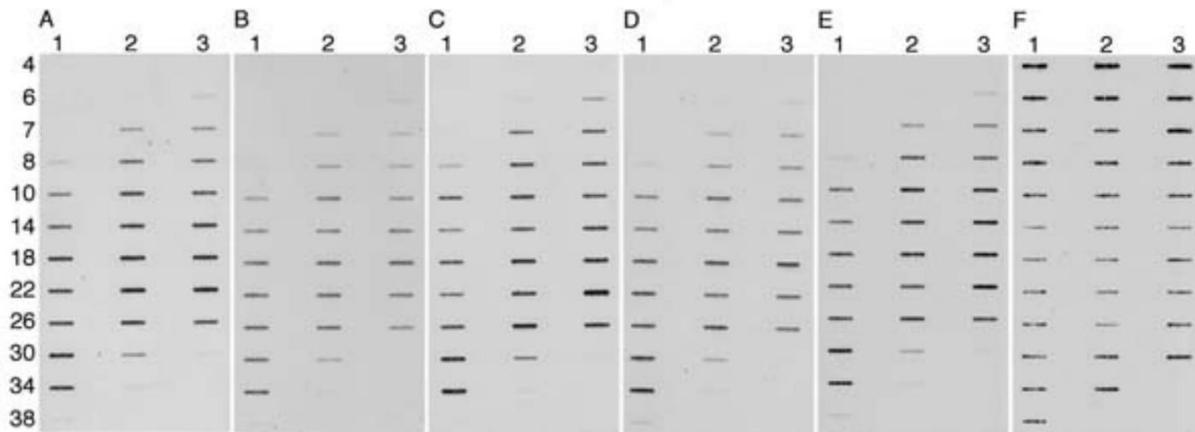
In addition to modifying protein content and quality, temperature also affect the temporal pattern of expression of gluten proteins genes. In a greenhouse experiment, transcript of  $\alpha$ -,

$\gamma$ -, and  $\omega$ -gliadins as well as those of HMW-GS and LMW-GS of the American cultivar Butte 86 were showed to accumulate a day earlier and within a shorter period (difference of five days in the accumulation) under post anthesis heat stress (37°C days/17°C night) than the control (24°C days/17°C night) (Altenbach *et al.*, 2002) (Figure 1.9.).

**Table 1.6.** Correlation coefficients for the relationship between heat stress (as hours above 35°C) and protein content, yield, dough rheology ( $R_{max}$  and extensibility) and bread quality (loaf volume) measured on five cultivars (Eagle, Oxley, Egret, Halberd and Cook) grown at three locations (Narrabri, Turretfield, and Wongan Hills).

	Narrabri	Turretfield	Wongan Hills	All sites
<b>Protein content</b>				
Eagle	0.72**	0.55	0.50	0.70**
Oxley	0.73**	0.53	0.31	0.63**
Egret	0.84***	0.68**	0.17	0.62**
Halberd	0.88***	0.62*	0.13	0.61**
Cook	0.80***	0.46	0.21	0.54**
<b>Grain yield</b>				
Eagle	-0.75**	-0.17	-0.17	-0.31*
Oxley	-0.83**	-0.26	-0.13	-0.36**
Egret	-0.77**	-0.16	-0.26	-0.57**
Halberd	-0.87**	-0.26	-0.17	-0.60**
Cook	-0.67*	-0.62	-0.17	-0.52**
<b><math>R_{max}</math></b>				
Eagle	-0.57	-0.82***	-0.72**	-0.23
Oxley	-0.60*	-0.26	-0.83***	0.43**
Egret	-0.77**	-0.54	-0.31	-0.42*
Halberd	0.32	-0.72**	-0.77**	-0.33
Cook	-0.38	-0.72**	-0.77**	-0.33
<b>Extensibility</b>				
Eagle	0.31	0.41	0.62*	0.52**
Oxley	0.17	0.55	0.57	0.44**
Egret	0.37	0.33	0.61*	0.59**
Halberd	0.23	0.58	0.35	0.59**
Cook	0.28	0.41	0.53	0.34
<b>Loaf volume</b>				
Eagle	-0.32	-0.43	-0.71**	-0.40**
Oxley	-0.25	-0.66*	-0.43	-0.56**
Egret	0.16	-0.53	-0.33	-0.22
Halberd	-0.44	-0.65	-0.41	-0.37**
Cook	-0.76**	-0.72**	-0.71**	-0.68**

Reference: Blumenthal *et al.*, 1993



**Figure 1.9.** Effect of temperature and drought on accumulation of gluten protein transcripts in developing kernels supplied with post anthesis fertilizer. The total RNA of gels A to F were hybridised with  $\alpha$ -,  $\gamma$ -, and  $\omega$ -gliadins, LMW-GS, HMW-GS, and 18S rRNA probes, respectively. Track 1 corresponds to the control treatment that received temperatures of 24°C/17°C after anthesis whereas tracks 2 and 3 correspond to the heat treatment (37°C /17°C) without or with associated drought, respectively. The numbers on the left refer to number of days after anthesis where the grains were sampled from 4 to 38 days.

Reference: Altenbach *et al.*, 2002

### 1.9.2. Effect of nitrogen (N) fertilisation

The nitrogen fertilization has been shown to impact both protein content and composition.

In their literature review based on 32 papers, Duncan *et al.* (2018), found that protein content generally increased with the amount of nitrogen fertiliser from 0.01 to 0.04% in dry weight per kg of N fertiliser applied in comparison to the unfertilised control. They also reported an increase in the mean grain yield following nitrogen application in 11 studies from 1000 to 3588 kg ha<sup>-1</sup> when no fertiliser was applied to 1100 to 4014 kg ha<sup>-1</sup> when fertiliser was used (Table 1.7).

Increasing nitrogen fertilisation levels result in an increase in the grain gliadin and glutenin content (Godfrey *et al.*, 2010; Johansson *et al.*, 2001; Wieser and Seilmeier, 1998) and in a higher gliadin-to-glutenin ratio (Godfrey *et al.*, 2010; Wieser and Seilmeier, 1998).

In Godfrey *et al.* (2010), incremental levels of N fertilisation of 48 kg N. ha<sup>-1</sup> between 0 and 288 kg N. ha<sup>-1</sup> resulted in an increase in the proportion of gliadins in the total protein from 38.4% to 52% in the flour of the breadmaking cultivar Hereward. Similarly, Wieser and Seilmeier (1998), reported a greater increase in the proportion of gliadin of 12 wheat cultivars

in comparison to the proportion of glutenin for the same cultivars when the nitrogen fertilisation was raised from 40 to 170 kg N. ha<sup>-1</sup>. The average proportion of gliadin increased by 73% whereas the proportion of glutenin increased by 53% between the low and high N treatments.

**Table 1.7.** Effect of N and N + P fertilisation on the grain yield in kg. ha<sup>-1</sup> reported in 11 studies.

Reference	Location	Residual soil P <sup>*</sup>	Fertiliser N range	Fertiliser P range	Grain Yield		
					No fertiliser	N	N + P
		mg kg <sup>-1</sup>	kg ha <sup>-1</sup>		kg ha <sup>-1</sup>		
Alston (1980)	Aus	9	0-60	0-50	2055	2470	2612
Brennan and Bolland (2009a)	Aus	2 to 10	0-138	0-40	1000	1100	2749
Colwell (1963)	Aus	NR	0-51	0-40	2063	2069	2513
Colwell (1963)	Aus	8 to 34	0-50	0-40	1689	1714	1601
Gates et al. (1981)	Aus	NR	0-160	0-400	1400	1450	3730
Grant and Bailey (1998)	Can	13 to 34	0-100	0-44	1316	2447	3262
Grant et al. (1985)	Can	15	0-120	0-22	1202	2927	3737
Leikam et al. (1983)	U.S.A	9 to 29	0-84	0-20	1922	3276	4173
May et al. (1991)	Can	4.5 to 8	41-140	0-17	2672	3066	6272
Usherwood and Segars (2001)	U.S.A	NR	0-60	0-26	2152	2556	4402
Widdowson et al. (1963)	U. K.	NR	0-75	0-62	3588	4014	4575

Reference: Duncan *et al.*, 2018

## 1.10. Improving the grain protein content

### 1.10.1. N fertilisers cause environmental damages

The UK breadmaking industry has set a threshold for breadmaking wheat of 13% protein in dry weight (UK flour millers, 2022). This target is difficult to meet without nitrogen fertilisation (Turner *et al.*, 2004). A typical UK soil contains in average 75 kg of available nitrogen per hectare (kg/N/ha), which is far below the 279 kg/N/ha required by the crop to match the average benchmark yield of 11t/ha. As a result, fertiliser may be applied up to a rate of 200 kg/N/ha to support the crop (AHDB, 2021).

Only a third of the N fertiliser applied is retrieved in the grain and the rest is either not absorbed by the root system and eliminated through leaching in the nitrate form (NO<sub>3</sub>), surface run-off, and volatilisation or absorbed but later eliminated in a gas form (NH<sub>3</sub>) or simply not stored in the grain (Raun and Johnson, 1999).

Nitrate leaching and run-off in the soil are important sources of pollution causing soil acidification, contaminating the ground water, nutrient imbalances, and eutrophication of surface water (Abdalla *et al.*, 2019). About 50% to 70% of NO<sub>3</sub><sup>-</sup> present in the soil is leached by high precipitation occurring in autumn in temperate countries representing 30-60 kg N. ha<sup>-1</sup>

<sup>1</sup> (Di and Cameron, 2002). Besides the environment, leaching is also influenced by the agronomy, the soil texture, and the cropping system (Table 1.8.).

**Table 1.8.** Quantity of NO<sub>3</sub><sup>-</sup> leached in kg N. ha<sup>-1</sup>. y<sup>-1</sup> in different cropping systems.

N applied (kg N ha <sup>-1</sup> y <sup>-1</sup> )	Soil texture	Cropping systems	Drainage (mm)	Leaching loss (kg N ha <sup>-1</sup> y <sup>-1</sup> )	Reference
Ammonium nitrate 200	Loamy sand	Cereal rotation: spring wheat	125–291	17–87	Shepherd and Lord (1996) (UK)
Ammonium nitrate 175	Loamy sand	Cereal rotation: winter sheat	85–226	4–45	Shepherd and Lord (1996) (UK)
Anhydrous ammonia 200	Loam	Continuous corn	90–336	11–107	Bjorneberg et al. (1996) (USA)
Anhydrous ammonia 170	Loam	Corn-soybean: corn phase	60–193	5–52	Bjorneberg et al. (1996) (USA)
None	Loam	Corn-soybean: soybean phase	50–320	5–51	Bjorneberg et al. (1996) (USA)
Urea + ammonium nitrate 200	Silt loam	No-till corn	110–470	8–77	Baker and Timmons (1994) (USA)
Urea + ammonium nitrate 125	Silt loam	No-till corn	140–390	8–36	Baker and Timmons (1994) (USA)
None	Silt loam	Mixed cropping: autumn ploughing, winter wheat	140–250	14–102	Francis et al. (1995) (NZ)

Reference: Di and Cameron, 2002

### 1.10.2. The negative relationship between grain protein content and grain yield

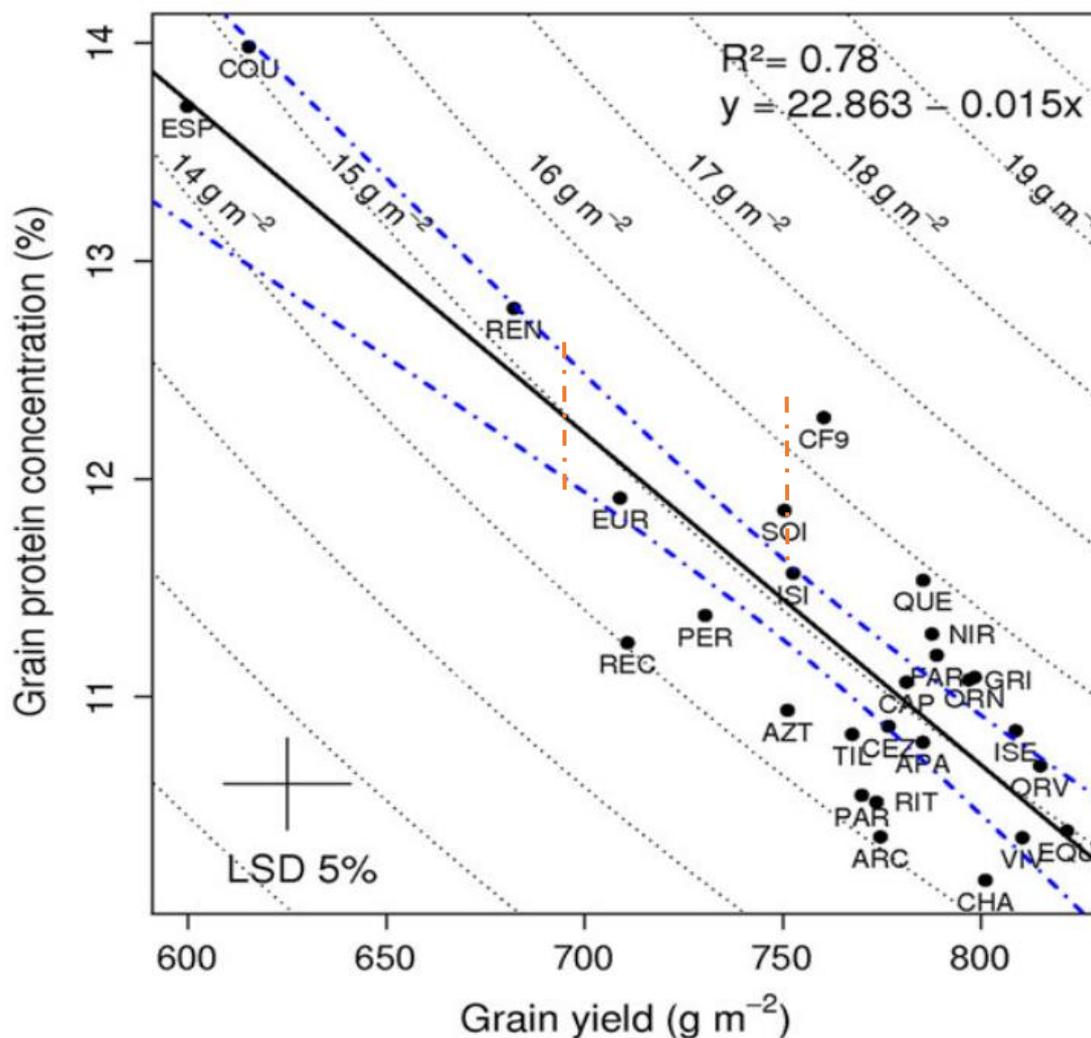
Alternatives to nitrogen fertilisers are therefore needed to alleviate their environmental impact. In this way, breeding has explored different solutions to increase the grain protein content (GPC) in wheat such as the introgression of high GPC genes from wild relatives, the use of a selection index or the defining of a grain yield (GY) for the increase of grain GPC but those attempts were not successful because of a negative relationship between GPC and GY (Bogard *et al.*, 2010). The dilution of protein by starch accumulation was suggested as a potential cause for this negative relationship (Slafer *et al.*, 1990; Bogard *et al.*, 2010).

The GPC-GY negative relationship was reported by many authors (Bogard *et al.*, 2010; Monaghan *et al.*, 2001; Oury *et al.*, 2003; Oury and Godin, 2007) and showed wide variation across environments. For instance, the investigation of the GPC-GY linear regression in 27 environments (combinations of sites, years, and nitrogen or density treatments) of 27 wheat varieties by Bogard *et al.* (2010) showed wide variation of coefficients of determination (R<sup>2</sup>) from 0.10 to 0.85. Regression slopes were all negatives and varied between -0.004 to -0.02 %protein.m<sup>2</sup>. g<sup>-1</sup>. According to that range, an increase of yield by 100 g.m<sup>-2</sup> will result in a decrease of protein content by 0.4 - 2%. This GPC-GY negative relationship is translated in a lower protein content in modern high-yielding wheat cultivars in comparison to older cultivars (Nigro *et al.*, 2019).

### 1.11. Grain protein deviation (GPD)

#### Introduction

To overcome this negative relationship, Monaghan *et al.* (2001) suggested the use of the positive residuals derived from the linear regression between grain yield and protein content, which they call Grain Protein Deviation (GPD), to select cultivars with both high yield and high protein content (Figure 1.10.).



**Figure 1.10.** Example of a simple linear regression between grain protein concentration (in %) and grain yield (in  $\text{g m}^{-2}$ ). Data are the means of 27 cultivars grown in 27 environments. The dotted blue line indicates the 95% confidence interval of the regression, and the black dotted lines are the iso-grain N yield. Few residuals are highlighted by yellow dashes to represent GPD.

Reference : Bogard *et al.*, 2010 (Adapted)

### 1.11.1. Stability of GPD across environment

Several authors (Oury and Godin, 2007; Bogard *et al.*, 2010; Mosleth *et al.*, 2020) showed that GPD was partially under genetic control, stable and therefore amenable to selection.

Oury and Godin (2007) calculated genotypic GPD mean values for candidate wheats varieties assessed by the “Groupe d'Etude et de contrôle des Variétés Et des Semences” (GEVES) for the registration on the French National List and reported the cultivars which showed consistent positive or negative GPD at a 5% (or with standardised residual greater than 1.64) over two consecutive field seasons (Table 1.9.).

Table 1.9. Genotypes with significant GPD at a threshold of 5% in bi-annual multi-sites trials conducted between 1994 and 2002 by the “Groupe d'Etude et de contrôle des Variétés Et des Semences” (GEVES) for the registration on the French National List. Candidate varieties were grown in target regions corresponding to Northern area (Northern France) or Southern area (Southern France) in multi-sites over two consecutive field seasons. 29-42 genotypes were tested at 4-13 sites in the northern area while 6-14 varieties were evaluated at 4-13 sites in the southern area.

Years of Evaluation	Genotypes with Positive GPD		Genotypes with Negative GPD	
	Northern area	Southern area	Northern area	Southern area
1994 + 1995	Record	Enesco	152269 Récital	
1995 + 1996	HynoKalia (F1)		Ami Récital	Tibet
1996 + 1997	Cockpit (F1)	Apache	158658 Ami Récital	Ami
1997 + 1998	Virtuose Auguste	Sidéral	Récital	
1998 + 1999	Runal	Sidéral	Châtelet Récital	Châtelet
1999 + 2000	HynoQuinta (F1)		287 307 351 Récital Trémie	
2000 + 2001	2535 2558		2520 Vercors	Récital
2001 + 2002	Grisby 4633 Astrakan		Récital	Récital
2002 + 2003	Hyval (F1) 6960 6965 Quebon Sankara		Récital	Récital Subtil
2003 + 2004		Exotic	9577 Honnor Wallaby	Raffy
2004 + 2005	1012215	Graindor	1012174 1012187 Trémie	

Reference: Oury and Godin, 2007

This threshold was set to limit the number of lines with inconsistent GPD values (e.g., lines with positive GPD in the first-year multi-sites trial and negative GPD in the second-year multi-sites trial) to 6.6%. In other words, at a threshold of 5% only 6.6% lines showed inconsistent GPD. In table 1.9., the authors note that cultivars Sideral and Recital show very robust positive or negative GPD values, respectively.

Based on the number of lines showing consistent positive or negative GPD (Figure x), Oury and Godin (2007), suggested that the GPD was partly under genetic control.

Other authors confirmed the genetic control, of GPD using either ANOVA, Pearson correlation or heritability values.

Bogard *et al.* (2010) used ANOVA to assess difference of GPD between genotypes and the influence of interaction genotype by environment (G x E). The ANOVA indicated that both factors were significant. In addition, they investigated the stability of the GPD values across environments using Pearson correlations. This resulted in 43 correlations being positive and significant out of the 45 correlations tested.

Mosleth *et al.*, (2020) studied the stability of GPD across nine environments located in the south-East of England for 30 wheat genotypes calculating a heritability score corresponding the proportion of the GPD variance explained by the genotype. They obtained a score of 0.44% of which 0.30 was attributable to the genotype factor and 14% to G x E interactions (nitrogen content and year).

In comparison, Geyer *et al.* (2022) obtained a much higher heritability score for GPD (0.88) which was based on eight field trials widely distributed across Germany and 394 recombinant inbred lines.

### 1.11.2. Genetic architecture of GPD: QTL and candidate genes

GPD is a complex trait in that it is controlled by many genes and environmental factors. The GPD physiological pathway has still not been fully unravelled and could involve processes such as nitrogen remobilisation and uptake as well as carbohydrates accumulation in the grain (Bogard *et al.*, 2010). However, multi-environments Pearson correlations showed that GPD may be associated to post-anthesis nitrogen uptake (Bogard *et al.*, 2010). Only two studies investigated the genetic architecture of GPD in durum and common wheats.

Nigro *et al.* (2019) performed a Genome Wide Association Study (GWAS) on a panel of 240 durum wheat varieties (*Triticum turgidum* L. spp. *durum*) to identify marker in association

with GPD. They identified four stable GPD QTL (i.e., detectable in at least three environments) located on short arm of chromosome 4A (1) and long arm of chromosome 5B (3) explaining 5.2-6.8% of the phenotypic variance and with additive effects of 0.44-0.72%protein.m<sup>2</sup>. g<sup>-1</sup>. (Table 1.10). The chromosome 4A QTL (*QGpd.mgb-4A*) peak marker was located inside the coding region of a Nitrogen-related gene (*GSr1-4A*) encoding a glutamine synthetase responsible for the incorporation of NH<sub>4</sub> in organic molecules. The chromosome 5B QTL(*QGpd.mgb-5B.2*) was located 1.9cM upstream of a candidate gene encoding a Glutamate dehydrogenase which is an enzyme catalysing the interconversion between glutamate and 2-oxoglutarate (Nigro *et al.*, 2019).

Geyers *et al.*, (2022) applied composite interval mapping (CIM) to find QTL in a Multiparent Advanced Generation Intercross (MAGIC) population derived from eight progenitors and formed of 394 bread wheat lines. They identified nine QTL, of which seven were environment specific (i.e., detected in only one environment), and two (*QGpd.lfl-2A.1*, and *QGpd.lfl-7A.3*) were stable in two environments (Table 1.10.). The QTL explained 1-8% of the phenotypic variance.

**Table 1.10.** Summary of QTL for the trait GPD found in two studies: Geyers *et al.*, (2022) and Nigro *et al.*, (2019) in bread wheat (*Triticum aestivum*) and pasta wheat (*Triticum turgidum* L. spp. *durum*). QTL: QTL identifier, -log<sub>10</sub>(p): significance threshold for the QTL detection, % variance: % of phenotypic variance explained by the QTL, confidence interval is the support interval around the peak marker and is given in centimorgans.

Reference	Specie	QTL	Chromosome	-log <sub>10</sub> (p)	% of Variance	Confidence Interval
Geyer <i>et al.</i> , 2022	<i>Triticum aestivum</i>	<i>QGpd.lfl-2A.1</i>	2A	9.1	8	122-131.9
Geyer <i>et al.</i> , 2022	<i>Triticum aestivum</i>	<i>QGpd.lfl-2A.2</i>	2A	6.6	6	227-234.7
Geyer <i>et al.</i> , 2022	<i>Triticum aestivum</i>	<i>QGpd.lfl-3A</i>	3A	5.6	8	109.9-125
Geyer <i>et al.</i> , 2022	<i>Triticum aestivum</i>	<i>QGpd.lfl-3B</i>	3B	4.6	1	254.4-266
Geyer <i>et al.</i> , 2022	<i>Triticum aestivum</i>	<i>QGpd.lfl-4D</i>	4D	6.9	4	0-17
Geyer <i>et al.</i> , 2022	<i>Triticum aestivum</i>	<i>QGpd.lfl-5B</i>	5B	4.7	3	52-63.7
Geyer <i>et al.</i> , 2022	<i>Triticum aestivum</i>	<i>QGpd.lfl-7A.1</i>	7A	10.4	4	78-82.5
Geyer <i>et al.</i> , 2022	<i>Triticum aestivum</i>	<i>QGpd.lfl-7A.2</i>	7A	6.2	2	215-227
Geyer <i>et al.</i> , 2022	<i>Triticum aestivum</i>	<i>QGpd.lfl-7A.3</i>	7A	4.8	4	322-326
Nigro <i>et al.</i> , 2019	<i>Triticum turgidum</i> L. spp. <i>durum</i>	<i>QGpd.mgb-4A</i>	4AS	2.9	5.2	37.8-39
Nigro <i>et al.</i> , 2019	<i>Triticum turgidum</i> L. spp. <i>durum</i>	<i>QGpd.mgb-5B.1</i>	5BL	3	2.4	53.1-54.4
Nigro <i>et al.</i> , 2019	<i>Triticum turgidum</i> L. spp. <i>durum</i>	<i>QGpd.mgb-5B.2</i>	5BL	3.4	5.8	140.7-146.5
Nigro <i>et al.</i> , 2019	<i>Triticum turgidum</i> L. spp. <i>durum</i>	<i>QGpd.mgb-5B.3</i>	5BL	3.6	6.8	172.9-

### 1.11.3. GPD molecular pathway and its link to nitrogen metabolism

#### 1.11.3.1. Introduction

GPD are the residuals from the GPC-GY regression and corresponds to variations of GPC left after accounting for the effect of GY. Therefore, GPD appears very highly positively correlated with GPC but weakly and not significantly correlated with GY (Mosleth *et al.*, 2020). This indicates that GPD is related to nitrogen metabolism rather than carbon metabolism.

A grain with a GPC of 13% dw contains 2.36% dw of nitrogen (Sosulski and Imafidon, 1990) which is supplied by direct root uptake, remobilisation from the vegetative parts after anthesis, and fixation of atmospheric ammonia (NH<sub>3</sub>) (Cormier *et al.*, 2016; Harper *et al.*, 1987). However, GPD was only found to be associated with post anthesis N uptake which was estimated as total N at harvest minus total N at anthesis and shown to be independent from Nitrogen remobilisation (Bogard *et al.*, 2010).

In their study, GPD was significantly correlated with post-anthesis N-uptake in 9 of the 12 environments assessed for 27 cultivars with the Pearson coefficient of correlation (r) ranging from 0.44 to 0.76. This suggests that GPD is partly linked to nitrogen uptake.

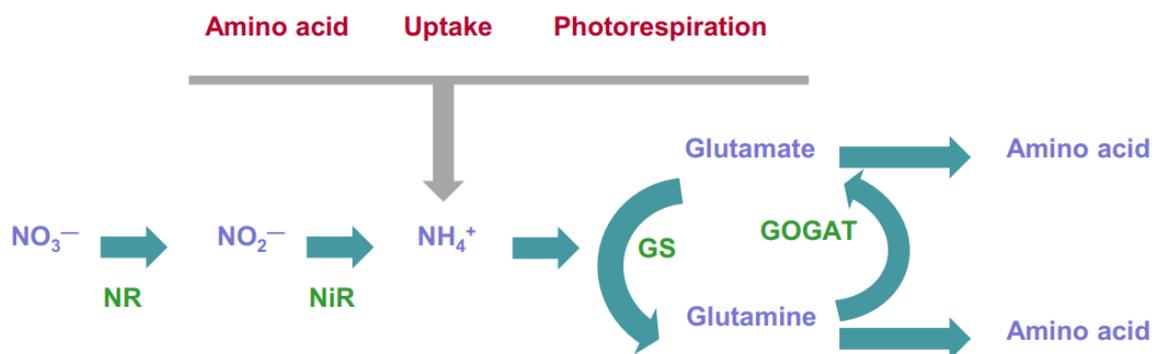
However, it should be also noted that no studies have assessed the correlation between GPD and the fixation of atmospheric NH<sub>3</sub>. Harper *et al.*, (1987) quantified the NH<sub>3</sub> atmospheric uptake at 0.1 kg N. ha<sup>-1</sup> during early vegetative stages (78-108 days after sowing) when the soil N supply is immobilised. In addition, the method used by Bogard *et al.*, (2010) to calculate post-anthesis N-uptake and N remobilisation is considered by some authors (Zhou *et al.*, 2018) to be biased and they indicate that the use of <sup>15</sup>N isotope labelling instead would give a better estimation.

#### 1.11.3.2. Root N uptake

Total N uptake by the root depends on both physiological and morphological characteristics (e.g., root age, root biomass, root morphology) (Glass, 2003).

The conversion of inorganic nitrogen in organic nitrogen (amino acids) follows a complex pathway involving enzymes (reductases and synthases), transporters (nitrate), and various co-substrates and co-factors (Figure 1.11.) and is regulated by environmental factors, mainly the concentration of nitrate in the soil and light conditions (Cormier *et al.*, 2016). At first, the nitrates (NO<sub>3</sub><sup>-</sup>) present in the soil are absorbed by the root and reduced into nitrites (NO<sub>2</sub><sup>-</sup>) in

the cytosol by the enzyme nitrate reductase (NR). Nitrites are then further reduced to ammonium ( $\text{NH}_4^+$ ) by the enzyme nitrite reductase (NiR) and subsequently introduced into amino-acids glutamine or glutamate by the glutamine synthetase (GS) or glutamate synthase (GOGAT), respectively (Cormier *et al.*, 2016) (Figure 1.11.).



**Figure 1.11.** Assimilation of inorganic nitrogen into amino acids in wheat. Enzymes involved in the assimilation of N are shown in green: NR (nitrate reductase), NiR (nitrite reductase), GS (glutamine synthetase), GOGAT (glutamate synthase).

Reference: Cormier *et al.*, 2016

The influx of  $\text{NO}_3^-$  through the root plasmic membrane is mediated by transporters. Three  $\text{NO}_3^-$  transporters differing by their activity and affinity for  $\text{NO}_3^-$  were discovered in barley. Their expression is influenced by the soil  $\text{NO}_3^-$  concentration. First, a constitutive high-affinity low capacity transport system (cHATS) is expressed independently of the  $\text{NO}_3^-$  or  $\text{NO}_2^-$  concentrations in the soil. Then, after prolonged exposure to  $\text{NO}_3^-$  or  $\text{NO}_2^-$  at a concentration below  $>250\mu\text{M}$ , an inducible high-affinity and high capacity transport system (iHATS) is expressed but for high  $\text{NO}_3^-$  concentration (above  $>250\mu\text{M}$ ), a low-affinity transport system (LATS) predominates (Glass, 2003).

### 1.13.3.3. N remobilisation into the grain

The other source of N provided to the grain comes from remobilisation of N accumulated in the biomass before anthesis (Bogard *et al.*, 2010).

The percentage of grain nitrogen provided by remobilisation diverges across studies: 50% in Harper *et al.* (1987), 66.7% in Zhou *et al.* 2018, and 84% in Bogard *et al.* (2010) which suggests genetic and environmental variations. Apart from being remobilised into the developing

grain, nitrogen can be lost in gaseous ammonia form (volatilization) after fertiliser application or during the senescence period. Following nitrogen application, up to 11.4% of the nitrogen could be lost at the soil-plant level as  $\text{NH}_3$ , while during senescence volatilization of nitrogen from the senescing organs may account for 9.8% of the nitrogen applied (Harper *et al.*, 1987). Monocarpic leaf senescence is an important process of N remobilisation. This is an irreversible process in which up to 80% of the leaf nitrogen is directed to the grain (Chapman *et al.*, 2021). Monocarpic terminal senescence is the final stage in wheat development which is initiated at anthesis at the whole-plant level and during which nitrogen is remobilised in the grain (Distelfeld *et al.*, 2014).

The onset of senescence is subject to strong environmental and genetic regulation and is affected by a modification of the sink (grain N demand)/source (soil N available) balance. Early senescence is triggered when the soil N available cannot meet the grain N demand, nitrogen is therefore remobilised from the leaves, the rachis, and the stem to complement (Bogard *et al.*, 2011; Cormier *et al.*, 2016). A late N application (before anthesis) re-equilibrates the sink/source balance and delays the senescence process as nitrogen can be provided from soil (Bogard *et al.*, 2011).

Protein degradation and especially chloroplast proteins such as the Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) are believed to be a major source of N remobilised in the grain although some evidence of N remobilised in inorganic form ( $\text{NO}_3^-$ ,  $\text{NO}_2^-$ , and urea), such as the presence of associated transporters upregulated during senescence has been provided in *Arabidopsis thaliana* (Havé *et al.*, 2017).

The negative relationship between GPC and GY is epitomised by the timing of senescence: a delayed onset favours GY through prolonged photosynthesis and carbon assimilation but delays N remobilisation in the grain thereby affecting GPC whereas an early senescence promotes GPC as nitrogen is remobilised in the grain but hinders photosynthesis efficiency and therefore GY (Bogard *et al.*, 2011; Havé *et al.*, 2017). For instance, positional cloning of the QTL *Gpc-B1* by Uauy *et al.* (2006) was associated with an increased in GPC but comes with a reduction of dry weight. This authors also found that the increase in GPC was caused by early senescence by the allele *NAM-B1* in the QTL region. The reduction of dry weight is coherent with the observations of Kindred *et al.* (2005) who observed that an increased supply of nitrogen to the grain (nitrogen applied at a rate of  $200 \text{ kg}\cdot\text{ha}^{-1}$ ) resulted in a

decrease of the grain cavity diameter (<0.1 mm) in comparison to a control (no nitrogen applied).

### 1.12. Thesis Aims

Despite tremendous progress in the understanding and improvement of breadmaking quality, notably on the importance of HMW-GS subunit composition and of the balance between the gliadins and the glutenins, there is still little understanding about the factors contributing to the variation of quality across years.

Environmental cues such as heat stress and the amount of nitrogen have been shown to impact dough physical properties through a modification of the glutenin-to-gliadin ratio but in this way some cultivars like Hereward are less sensitive to environmental change than others. Breadmaking cultivars showing variable quality over year are listed in the group 2 of the Agriculture and Horticulture Development Board (AHDB) reference list and are not suitable to use as a whole but in combination with a flour of higher quality and this generates extra work and testing at the bakery.

In addition, genetic variation in breadmaking quality is still not completely explained as the HMW-GS composition accounts for 55-67% of quality. It should therefore be theoretically possible to further improve the quality with the discovery of new components with minor effect. This was proved recently by the characterisation of a QTL associated with the loaf volume. This QTL was shown to be associated with a higher galactolipid content providing extra stabilisation of the gas cells in the dough and resulted in a higher loaf volume (Min *et al.*, 2020).

In that direction, the HGCA project: "Investigating wheat functionality through breeding and end use" spearheaded by Millar *et al.* (2008) ascertained the genetic architecture of key breadmaking traits (e.g. loaf volume and crumb colour) using different breadmaking processes (e.g. CBP and spiral white) in three biparental mapping populations of UK hard milling breadmaking wheat varieties: Malacca × Charger, Hereward × Malacca and Shango × Shamrock. In this project, six NILs derived from the Hereward x Malacca population were grown in two environments to test the following hypothesis:

- 1- Will the QTL effects be significant in the NILs backgrounds?
- 2- Will the QTL effects be significant or stable in the two environments studied?
- 3- Will the QTL effects be of the same magnitude as in the double haploid population?

The QTL to be tested and their associated traits are QTL 1B (finer crumb structure), QTL 2D (higher loaf volume), QTL 4D-1 (finer crumb structure and higher whiteness), QTL 4D-2 (higher crumb whiteness), QTL 6A (finer crumb structure) and QTL 7B (higher loaf volume).

In addition to assessing the QTL, each NIL was investigated for general quality parameters at each stage of the breadmaking process to assess the grain intrinsic quality and the dough physical properties and to verify that the presence of the QTL has no impact on other quality components.

A second aspect of this PhD deals with improving the sustainability of bread wheat farming using a linkage analysis technique for QTL detection to facilitate the breeding of cultivars with an improve flux of nitrogen to the grain and therefore reduce the use of nitrogen fertilisers and their associated negative impacts on the environment. The trait grain protein deviation (GPD) is an indicator of cultivar performance toward nitrogen partitioning into the grain. A double haploid mapping population of 109 lines derived from a cross between Malacca (negative GPD) and Hereward (positive GPD) was grown in three environments and phenotyped for GPC and GY to calculate and study the genetic architecture of the trait GPD. The following hypothesis will be tested:

- 1- Is there evidence of transgressive segregation for GPD in the double haploid population? In other words, are there any lines performing better or worse than the parents Malacca and Hereward?
- 2- Are there any pleiotropic QTL?
- 3- Are there any GPD QTL stable in at least two environments?

In the two Reading environments where the DH population was grown, senescence measurements were also taken at the whole plot level at eight time points from anthesis to maturity to model the senescence course and calculate metrics to correlate with GPD and use as phenotypes for a QTL analysis.

The senescence measurements were conducted in the aim of answering the following questions:

- 1- Is there any link between the senescence and GPD?
- 2- Are there any co-locations between the confidence intervals of GPD and the seven senescence metrics?

The two parts of the PhD described above are independent; they have their own plant materials and are designed to fulfil different objectives.

## Chapter 2. Materials and Methods

### 2.1. Creation of the Near Isogenic Lines

#### 2.1.1. Development of the Malacca x Hereward DH mapping population

DH lines are obtained by doubling the chromosome number of F<sub>2</sub> (i.e., second generation of plant after a cross) haploid embryos. In wheat, this can be done by fertilising emasculated florets with maize pollen. This results in a non-viable haploid embryo that needs to be rescued from the spike and cultivated in a growing media to develop. Diploidy is then achieved by treating the embryo with colchicine (Bhalla and Singh, 2017, pp. 235–249). With this method, 100% homogenous lines are produced after one generation instead of six to eight in conventional selfing, thereby considerably reducing the production time (Ren *et al.*, 2017).

In this project, 111 DH of the cross Malacca x Hereward and developed by the breeder RAGT (RAGT, UK) were used. These two parents were chosen to generate variability in breadmaking traits and in GPD after genetic recombination. They are both breadmaking cultivars; however, Hereward provides a high and stable breadmaking quality and has a positive GPD while Malacca has a negative GPD and a lower breadmaking quality (Millar *et al.*, 2008).

#### 2.1.2 Development of the NILs

Mia *et al.* (2019) define NILs as lines having “otherwise identical genetic backgrounds except at one or a few genetic loci” and note that these lines “have been used intensively for detailed mapping and characterization of individual loci.”

Indeed, NILs are developed to isolate individual QTL from mapping population such as DH to study their effect without interference from other QTL.

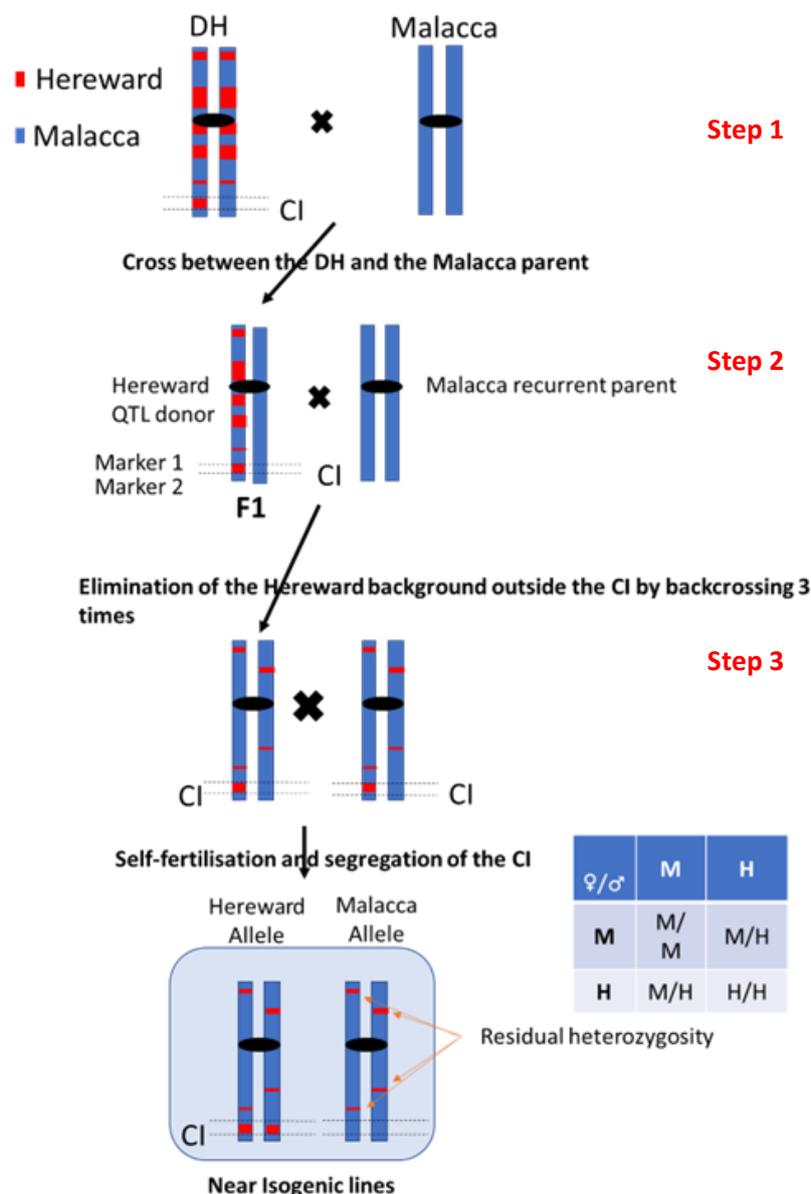
The Malacca x Hereward NILs were produced by Simon Griffiths group (John Innes Centre, Norwich, UK) using the QTL information of the HGCA project “Investigating wheat functionality through breeding and end-use” (Millar *et al.*, 2008).

They were produced by a crossing and selection scheme, starting with a DH line carrying the favourable allele at the QTL region which was then crossed with either the Malacca or Hereward as recurrent parent, whichever was contrasting to the favourable allele. Thereafter, heterozygous F<sub>1</sub> progenies were backcrossed three times to the recurrent parent (BC<sub>3</sub>F<sub>1</sub>) and were self-fertilised to produce BC<sub>3</sub>F<sub>2</sub> lines; that is, lines having been backcrossed three times and self-fertilised once (Figure 2.1).

Backcrossing is used to replace the segregating background of the DH lines by the parental genome of Malacca or Hereward retaining only the QTL region.

At each step of backcrossing, the donor genome content (i.e., the DH genome) of the next generation is reduced by half; after the first backcross, the donor genome content is 25%, but this reduces to 12.5% and 6.25% after the second and third backcrosses, respectively. Marker assisted selection (i.e., the genotyping of flanking marker at the QTL region) was employed at each generation to select lines heterozygous for QTL interval markers.

The final self-fertilisation step allowed the segregation of the heterozygous QTL into two homozygous alleles Malacca and Hereward allele forming a NIL within which the lines are 93.75% identical.



**Figure 2.1.** Selection scheme used to produce the NILs. The Hereward genome is highlighted in red and the Malacca in blue. The scheme is divided in three steps. NIL development starts by crossing a DH with a recurrent parent, such as Malacca. (Step 1), the F1 are 100% heterozygous. In step 2, three backcrosses are performed with the recurrent parent (Malacca) to eliminate the DH genetic background and Marker assisted selection used to genotype the QTL which was kept heterozygous and was segregated by selfing (step 3). The punnet square represents the segregation ratios of the QTL allele.

## 2.2. Selection of the NILs

Six NILs sets were formed by selecting randomly five lines per allele and QTL when there were enough lines available at the John Innes Centre. For QTL 2D and 7B; however, less than five lines were developed and the number of lines had therefore to be adjusted. The NILs sets are presented in Table 2.1. NILs 1B, 4D-1, and 6A contain a favourable allele for cell number from Hereward, NILs 2D, and 7B contain a favourable allele for loaf volume from Hereward, and NIL 4D-2 contains a favourable allele for crumb whiteness conferred by Malacca.

The genealogy of each individual NIL line within a set can be traced back to their parental DH line as shown by the red and blue colouring of table 2.1, column “Name of the NIL streams”. For example, the NIL 1B has five sister lines with the Malacca allele at the QTL region: MH100/Mal4-188-04-17, MH100/Mal4-188-04-18, MH100/Mal4-188-03-09, MH100/Mal4-188-03-14, MH100/Mal4-188-03-24). To produce these lines the DH (DH) line number MH100 was selected and crossed with Malacca to give a first generation (F1) (MH100/Mal4) where all the individuals are heterozygous and identical. Recombination occurs in the F1 gametes and after the first backcross (BC1), the lines are no longer genetically identical. One line, 188 (BC1), was therefore selected among the progeny to perform the second backcross. Finally, lines 04 and 03 (BC2) were both selected to carry out the final backcross. The individual lines MH100/Mal4-188-04-18, MH100/Mal4-188-04-17 are therefore more closely related to each other than individuals MH100/Mal4-188-03-09, MH100/Mal4-188-03-14, MH100/Mal4-188-03-24) as they have parent 04 in common.

## 2.3. Experimental field design in which the NILs were grown

The six NILs sets were grown at Rothamsted for two consecutive field seasons: 2019-2020 and 2020-2021 on 2m<sup>2</sup> plots arranged in a “split – plot” fashion (Figure 2.2.). The lines received 200kg/ha of nitrogen over the crop cycle. The crop husbandry is described in table 2.2.

Two randomisations were applied to ensure equal treatment of the lines: a randomisation of the lines within each NILs (set of 10 plots) and a randomisation of the NILs within the split-plot design (six columns of ten plots). The design was replicated once to assess biological variation and each replicate was grown in a block which was set in the direction of drilling to control for this external factor.

The split-plot model has the following equation:

$$Y_{ijk} = \mu + \alpha_i + \gamma_k + \eta_{ik} + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

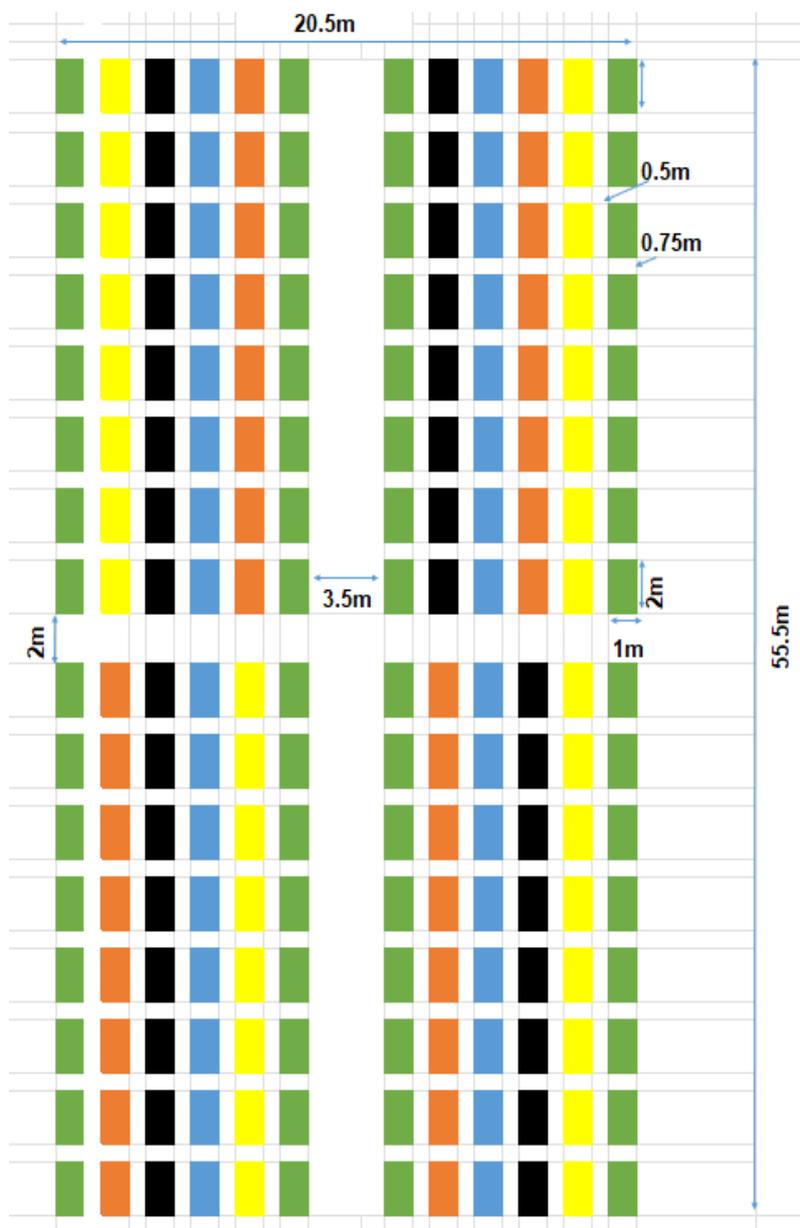
Where:

- $Y_{ijk}$  is the average response of a line with the QTL  $i$  ( $i=1:6$ ), the allele  $j$  ( $j=1:2$ ), in block  $k$  ( $k=1:2$ ).
- $\alpha_i$  is the fixed effect of QTL
- $\gamma_k$  is the fixed effect of block
- $\eta_{k(i)}$  is the whole plot error
- $\beta_j$  is the effect of alleles
- $(\alpha\beta)_{ij}$  is the interaction between QTL and allele
- $\varepsilon_{k(ij)}$  is the split - plot error.

**Table 2.1.** NIL crossing parents, QTL and markers informing the marker assisted selection and names and numbers of generated NILs streams and individual NIL numbers

Parent from DH population	Recurrent parent	QTL Chromosome	QTL trait	Markers used BC3F2	Name of NIL streams	NILs with Malacca allele (A)	NILs with Hereward allele (B)
MH100	Malacca	1B	Cell number	gwm264 and barc8	MH100/Mal4-188	04-17	04-05
						03-09	04-06
						03-14	03-04
						03-24	04-04
						04-18	03-20
MH9	Malacca	2D	Loaf volume	gwm102t,wmc18,gwm129t	MH9/Mal4-147 MH9/Mal4-163	01-06	02-22
						01-23	02-24
						01-15	02-18
							02-07
MH1	Malacca	4D-1	Cell number and crumb whitness	b98,gdm129t	MH1/Mal4-161	01-04	02-18
						01-01	02-17
						01-16	05-12
						01-15	02-16
						01-03	01-22
MH19	Hereward	4D-2	Crumb whitness	b98,gdm129t	MH19/Her4-207	06-22	06-05
						08-06	06-03
						06-18	06-06
						06-12	06-07
						08-02	06-01
MH70	Malacca	6A	Cell number	g334-b3	MH70/Mal4-168	01-15	09-06
						01-16	09-18
						03-05	09-09
						03-04	09-04
						01-18	09-08
MH39	Malacca	7B	Loaf volume	gwm537t,gwm577,barc182t	MH39/Her4-209 MH39/Her4-211	05-20	
						05-17	07-13
						06-08	07-22
						06-10	07-24
						06-13	

The red and blue colours connect each NIL with their corresponding NIL stream for QTL 2D and 7B



**Figure 2.2.** NILs split – plot design representing the organization of the NILs within the field trials. Each plot measures 2m<sup>2</sup> and the NILs are arranged vertically by colour with a randomization applied on the order of NILs (columns) within each block and on the order of the plots (sister lines) within each NILs. The split-plot is comprised of two blocks (rectangles) separated by 3.5m. The green plots are the guard rows.

**Table 2.2.** Agronomy of the NILs including fertilisation rate and pesticide applications

Location	Year	Plot Name	Sowing	Harvest	Fertilization	Pesticide
Rothamsted	2019-2020	OSIER	30/01/2020	11/08/2020	21/04/2020: 185 kg/ha (DoubleTop)	25/02/2020: 5 kg/ha (Ironmax pro)
					13/05/2020: 290 kg/ha (Nitram)	23/04/2020: 0.15 L/ha (X-clude)
					09/06/2020: 145 kg/ha (Nitram)	13/05/2020: 1L/ha (Cello)
Rothamsted	2020-2021	LONG HOOS	19/10/2020	13/08/2021	03/03/2021: 185 kg/ha (DoubleTop)	01/12/2020: 50mL/ha (Hallmark with Zeon Technology )
					20/04/2021: 290 kg/ha (Nitram)	02/12/2020: 0.5L/ha (Pontos)
					26/05/2021: 145 kg/ha (Nitram)	12/05/2021: 60 gr/ha (Presite SX)
						27/05/2021: 0.15 L/ha (Topik)
						10/06/2021: 1L/ha (Cello)

## 2.4 Assessment of the kinship between the NILs using Hierarchical Clustering (HC) on the genotypic data.

This method was used to cluster the NIL according to their genetic identity for a set of single nucleotide polymorphism (SNP) and calculate their genetic distance.

For this purpose, the lines were genotyped at University of Bristol using the Axiom Breeders 35K Axiom® array (Affymetrix product ID 550524), which is a DNA chip containing 35,143 SNP marker probes. The genotyping data were then analysed with the software Axiom Analysis Suite (version 5.1.1, ThermoFisher Scientific) and grouped into six categories: “call rate below threshold”, “mono high resolution”, “no minor homozygotes”, “poly-high resolution”, “off target variant (OTV)”, and “other”. The poly-high resolution category contained 3730 polymorphic (i.e., two alleles or more) SNP of high quality (i.e., with only few missing data) and was therefore selected to discriminate between the lines and conduct the HC. The other clusters were not considered further as they contained either monomorphic SNP or SNP with many missing values.

A Euclidian distance matrix was constructed with the poly-high resolution SNP group and the HC was built in R Statistical Software (v4.1.1; R Core Team 2021). The dendrogram was arbitrarily cut to separate the lines into four groups.

## 2.5. Milling of the NILs

The grain of the two field replicates of each line were bulked in equal proportion of one kilogram in plastic bags for milling.

### 2.5.1. Preparation of the white flour for rheology and baking tests

Before milling the grain were conditioned by adding water to soften the outer layers and facilitate the crushing of the grain. To this purpose, grain moisture content was determined using a DICKEY-john GAC® 2500-UGMA Grain Analysis Computer (Churchill Industries, Minneapolis, U.S.A) and the moisture was adjusted to 16%. The amount of water that was added depended on the grain initial moisture content and was read on the manufacturer conditioning chart (Bühler group, Uzwil, Switzerland). After water addition, the grains bags were thoroughly shaken to spread the water and set upon an agitating system overnight. A second moisture reading was taken to check the rise of moisture content to 16%.

After conditioning, the grains were milled with a Bühler MLU 202 laboratory mill (Bühler group, Uzwil, Switzerland) at Heygates Ltd. and the white flour fractions (breaks 1, 2, 3 and reductions 1, 2, 3) combined (Figure 2.3.). The fraction corresponding to the pericarp, the aleurone layer and the germ and called the bran which was separated from the starchy endosperm was further processed using a Bühler MLU 203 impact finisher (Bühler group, Uzwil, Switzerland) to detach the remaining white flour particles. The white flour retrieved from the bran was then added to the white flour fraction which was immediately stored at -20°C for preservation.

#### 2.5.2. Preparation of whole grain mills for SDS-PAGE

First, 10g of grain from each NIL was initially milled in a Retsch™ ZM 200 Model Ultra-Centrifugal Mills (Fisher Scientific, Leicestershire, UK) for 15 seconds at 18000rpm with a sieve size of 0.5mm. Then, the wholemeal flours were re-milled for 2min in a ball mill - 8000D Mixer/Mill (SPEX, Metuchen, NJ, U.S.A) to reduce the particle size and immediately stored in a freezer at -20°C for preservation.

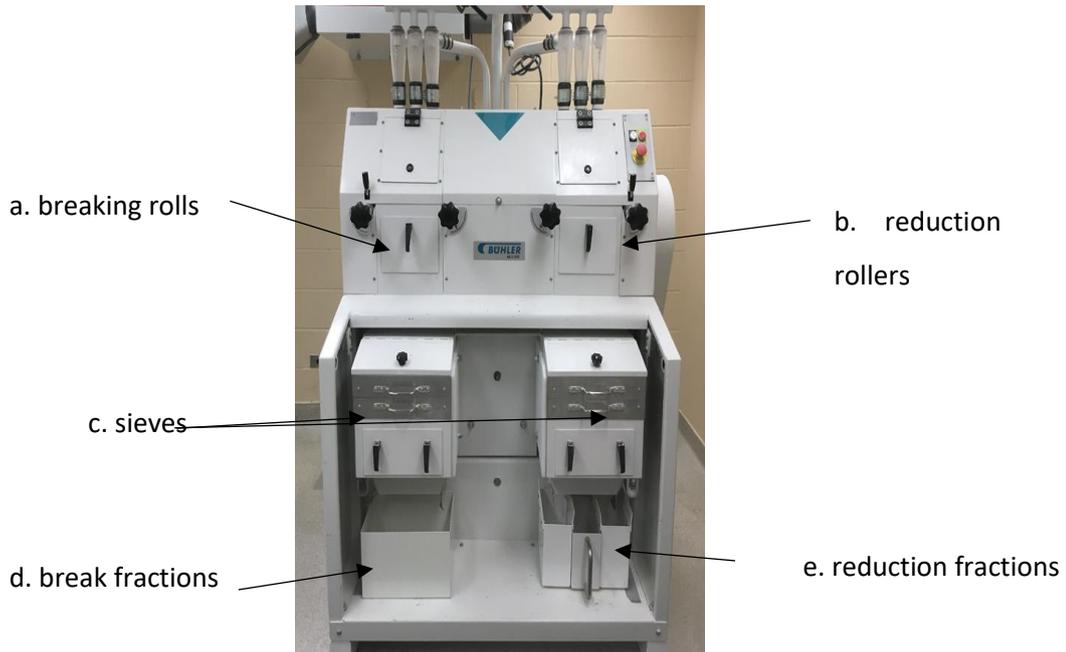
### 2.6. Preparation of 12 white flour bulks corresponding to the 12 NILs alleles

The bulking of the sister lines white flour fraction for each NILs QTL/allele fits two purposes: increasing the amount of flour available for rheological and baking tests and balancing the protein content between the sets of allelic pairs to remove its influence on breadmaking quality.

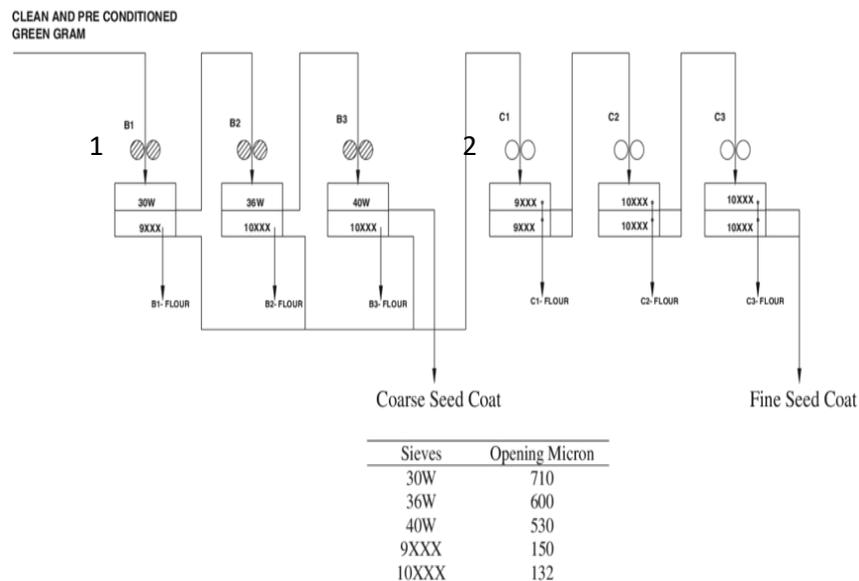
#### 2.6.1. Measurement of total N content on the white flour of the NILs (60 lines)

Total nitrogen content of the NILs white flour samples was determined by Heygates Ltd using a LECO CN628 combustion analyser (LECO corporation, MI, U.S.A) following the CCAT Method 019 (Campden BRI, Gloucestershire, UK) and the LECO FP-628 operations manual. Combustion of the sample at 900-1000°C with the presence of oxygen converts the nitrogen present in the samples in nitrogen oxides which are then reduced by a catalyst and detected by thermal conductivity. Nitrogen content was then converted to protein content by multiplying by a conversion factor of 5.7 (Maclean *et al.*, 2003).

a



b



**Figure 2.3.** Bühler MLU 202 laboratory mill (above) and its associated flow chart (below). The grain is first fed into a hopper at the back of the machine and then crushed successively by three pairs of rollers A) a). After each rolling step, the flour is sieved through two sieves of decreasing size. Coarse particles of bran do not pass the first sieve and undergo a second rolling step B)1), coarse flour particles enter the first sieve but are retained by the second and are reduced but the reduction rollers B)2). Fine flour particles exit the system and are retrieved in the break fraction A) d). The bran fractions that will be further processed (crushed and sieved) into a Bühler MLU 203 (not shown) to extract the remaining tightly bound flour particles.

2.6.2. Creation of NILs allelic bulks by bulking together the five sister lines at each NILs QTL allele.

An estimation of the bulk protein content was calculated using the individual lines protein content determined by LECO analysis and adjusting the proportion of their flour in the bulk using the following formula:

$$\text{Average Protein content (\%)} = \frac{\sum_i^n A_i X_i}{\sum_i^n A_i}$$

Where:

- n is the number of sister lines available for the allele at the QTL.
- $A_i$  is the amount in g of white flour for line i at the NIL QTL.
- $X_i$  the mean protein content in % for line i at the NIL QTL

Then, the proportion of each line were manually adjusted on Excel until the two NILs allelic bulk had the same protein content at three decimal points (Table 2.3.).

**Table 2.3.** Composition of the 12 NILs allelic bulks in the Rothamsted 2019-2020 field trial (a) and of the 4D-2 NILs allelic bulks in the Rothamsted 2020-2021 field trial (b).

**a**

NILs allelic bulks	Lines	Proportion	Protein content in %				
<b>1B-a</b>	188-3-9	14.3	11.581	<b>4D-2-b</b>	207-6-1	20.6	12.937
	188-3-14	14.3	12.343		207-6-3	14.7	12.946
	188-3-24	14.3	12.341		207-6-5	2.9	13.099
	188-4-17	14.3	12.255		207-6-6	20.6	12.906
	188-4-18	42.9	12.482		207-6-7	41.2	12.701
			<b>12.281</b>			<b>12.839</b>	
<b>1B-b</b>	188-3-4	15.9	12.238	168-1-15	20.5	12.151	
	188-3-20	36.5	12.064	168-1-16	20.5	12.638	
	188-4-4	15.9	12.626	168-1-18	20.5	12.661	
	188-4-5	15.9	12.373	168-3-4	20.5	12.331	
	188-4-6	15.9	12.385	168-3-5	17.9	11.160	
			<b>12.281</b>			<b>12.215</b>	
<b>2D-a</b>	147-1-6	2.8	11.872	168-9-4	20.0	12.306	
	147-1-15	19.4	12.226	168-9-6	20.0	12.292	
	147-1-23	19.4	11.950	168-9-8	20.0	12.500	
	209-5-17	38.9	12.717	168-9-9	20.0	11.997	
	209-5-20	19.4	12.313	168-9-18	20.0	11.980	
			<b>12.370</b>			<b>12.215</b>	
<b>2D-b</b>	163-2-7	30.6	11.991	209-6-10	26.3	13.020	
	163-2-18	23.1	12.719	209-6-13	26.0	12.506	
	163-2-22	23.1	12.304	211-7-13	47.7	13.239	
	163-2-24	23.1	12.592			<b>12.990</b>	
			<b>12.371</b>	209-6-8	35.8	13.280	
<b>4D-1-a</b>	161-1-1	16.7	12.336	211-7-22	42.8	12.646	
	161-1-3	16.7	11.999	211-7-24	21.4	13.198	
	161-1-4	16.7	12.014			<b>65.4</b>	
	161-1-15	16.7	11.770			<b>12.991</b>	
	161-1-16	33.3	12.080				
			<b>12.046</b>				
<b>4D-1-b</b>	161-1-22	20.7	11.847				
	161-2-16	17.2	11.914				
	161-2-17	17.2	11.986				
	161-2-18	25.9	12.204				
	161-5-12	19.0	12.229				
			<b>12.047</b>				

**b**

NILs allelic bulks	Lines	Proportion	Protein content in %
<b>4D-2-a</b>	207-6-22	14.7	11.6
<b>4D-2-a</b>	207-8-6	15.0	11.6
<b>4D-2-a</b>	207-6-18	23.7	11.2
<b>4D-2-a</b>	207-6-12	22.6	11.3
<b>4D-2-a</b>	207-8-02	23.9	11.1
			<b>11.318</b>
<b>4D-2-b</b>	207-6-05	23.4	11.3
<b>4D-2-b</b>	207-6-03	24.0	11.5
<b>4D-2-b</b>	207-6-06	15.1	11.2
<b>4D-2-b</b>	207-6-07	15.1	11.2
<b>4D-2-b</b>	207-6-01	22.4	11.3
			<b>11.318</b>

## 2.7. Rheological assessments on the doughs prepared from the 12 NILs allelic bulks

Rheology is a science studying the deformations and flows of liquid and solid materials. In the bread industry, it is used to predict the behaviour of the dough during the breadmaking process.

In this study, the dough rheology was assessed by a Farinograph: Perten doughLAB (Calibre Control International Ltd., Warrington, UK) and an Extensograph-E Brabender (Brabender, Duisburg, Germany).

The two instruments were operated following Heygates Ltd. internal protocol (Method Heygates 009 Determination of Dough Rheology), which is based on the AACC international Method 54-21. One flour sample of 300g was analysed for the 12 NILs allelic bulk with the Farinograph and Extensograph in the Rothamsted 2019-2020 field trial. The replication was increased to three flour samples of 300g which are technical replicates to re-assess the 4D-2 NILs allelic bulk in the Rothamsted 2020-2021 field trial.

## 2.8. Evaluation of the six NILs QTL effects.

The NILs and their associated quality traits were already described in Table 2.1.

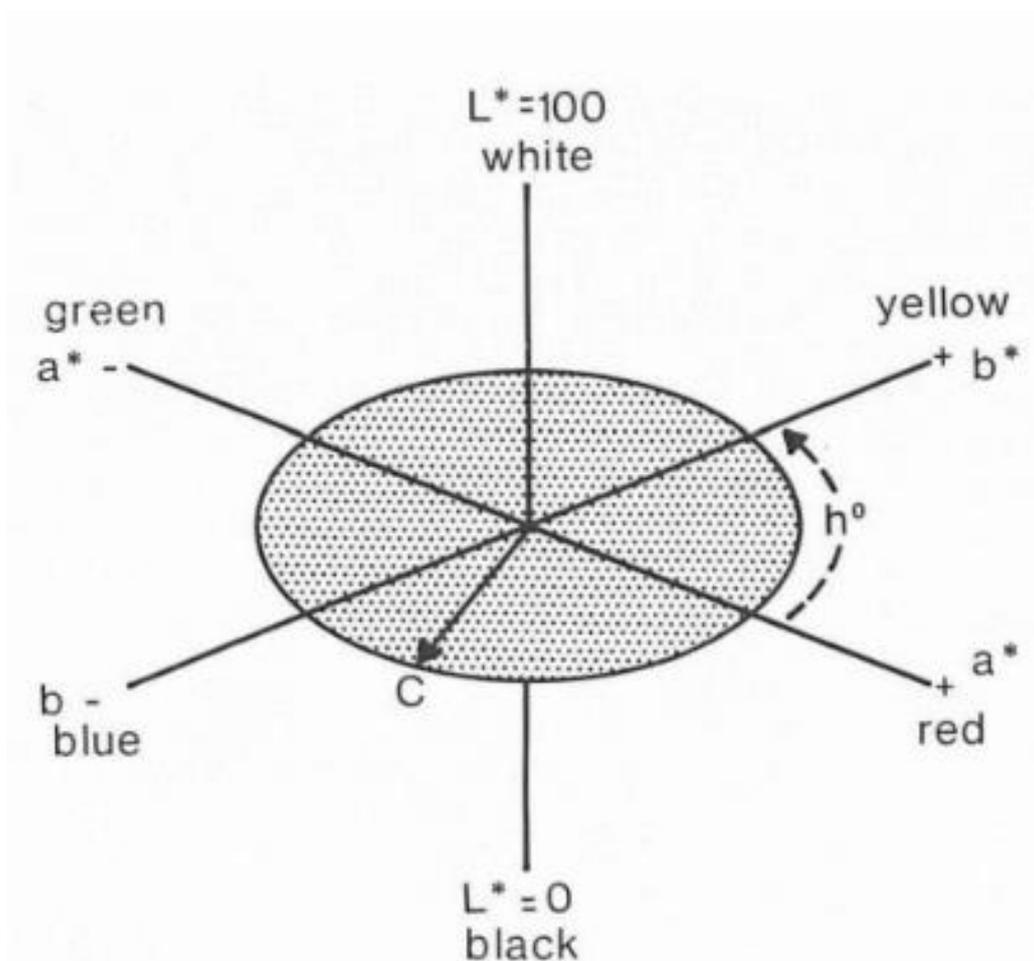
### 2.8.1. Bread height

The baking was performed by a qualified baker at Heygates following their internal protocol (Method Hey 00X 400g Test Baking). Four loaves of bread (technical replicates) were baked for each NILs allelic bulk giving a total of 48 loaves in the Rothamsted 2019-2020 field trial and eight loaves (only NILs 4D-2 was re-analysed) in the Rothamsted 2020-2021 field trial.

The bread recipe contained 400g of white flour, salt, and  $\alpha$ -amylase and the breadmaking process involved two bulk fermentations separated by a kneading phase. The bake height was measured immediately after baking before cooling starts and the bread starts to shrink and recorded in cm. Ten slices were then taken per loaf – 40 slices per NILs allelic bulk - and analysed for crumb whiteness, and crumb cell parameters.

### 2.8.2. Evaluation of flour and crumb whiteness using the $L^* a^* b^*$ colour space

Colour perception (e.g., brightness and contrast) depends on both the source of light under which the sample is studied (e.g., sample brightness might differ under sun light or office light) and the eyes of observer.  $L^* a^* b^*$  methods provide a reliable and reproducible measurement of colour by exposing the sample to a “standard illuminant” (a defined source of light) and by assessing its colour numerically on a three-dimensional space rather than by eye (Figure 2.4.).



**Figure 2.4.** Representation of a sample colour in the  $L^* a^* b^*$  3D colour space. The space has three axes: a  $L^*$  vertical axis for colour brightness (white upward and black downward), a  $a^*$  axis for green-red colour (red right, green left) and a  $b^*$  axis for the yellow-blue colour (yellow above, blue below). Each sample has therefore three coordinates:  $L^*$ ,  $a^*$ , and  $b^*$ )

Reference: Weatherall and Coombs, 1992

In our study, colour measurements were performed on white flour and bread slices following the Heygates protocol (Method Hey 017 LAB Tristimulus Colour) using a Satake Colour Grader NCG1A (SATAKE CORPORATION, Hiroshima-ken, Japan).

The colour of 40 slices of bread was analysed for each NILs allelic bulk in the Rothamsted 2019-2020 and Rothamsted 2020-2021 trials.

In addition, the colour of the flour of the NILs allelic bulk 4D-2 was analysed using a dry and wet method developed by Heygates Ltd. The wet method gives a more accurate estimation of flour colour by removing the effect of particle size, but it is the dry method that is favoured by millers as it takes less time to prepare the sample (Oliver *et al.*, 1993).

### 2.8.3. Evaluation of the number of cells in the bread crumb

The number of cells in the crumb was determined on slices of bread at Heygates Ltd using a C-CELL baking quality analyser (Calibre Control International Ltd., Warrington, UK) and following the manufacturer procedure. The number of cells were quantified on 40 slices of bread per NIL allelic bulk.

## 2.9. Assessment of the NILs gluten protein

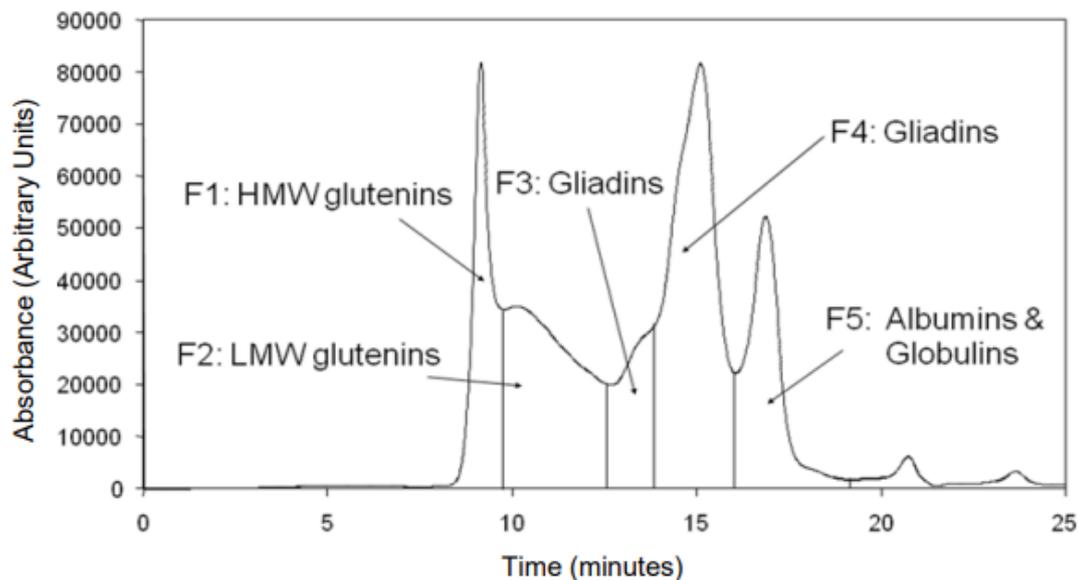
### 2.9.1. Evaluation of size distribution of the gluten polymers using Size-Exclusion High Performance Liquid Chromatography (SE-HPLC)

The SE-HPLC method is widely used to separate biological molecules on size. It was used in this study separate and quantify the native glutenin polymers that are too big to be separable by SDS-PAGE. SE-HPLC devices include a column with a stationary phase composed of porous beads of various diameters to separate the molecules of the sample according to their sizes and a detector which can be a spectrophotometer to measure the absorbance of the eluted molecules.

The gluten protein of the 12 NILs allelic bulks from Rothamsted 2019-2020 trial were analysed in duplicate by Campden BRI (Campden BRI, Gloucestershire, UK) following the SE-HPLC method Profilblé® (Morel *et al.*, 2020) co-developed by ARVALIS and l'INRA (Institut National de la Recherche Agronomique).

For each NILs allelic bulk, 160 mg of white flour was weighed and combined with 20 mL of 0.1 M phosphate buffer (pH 6.9) containing 1% (w/v) of SDS to dissolve the soluble gluten proteins. A mild sonication treatment was then applied to shear the insoluble large unextractable glutenin polymers into smaller and extractable polymers before a 10 min centrifugation at 5000 rpm. Aliquots of the supernatants were collected and separated for 25 minutes on an SE-HPLC column working at flow rate of 0.7 mL/min. The different protein components were detected at 214 nm by a UV detector.

The SE-HPLC chromatogram trace displays the absorbance at 214 nm in relation to the time elapsed in minutes after the introduction of the samples and consists of five peaks. Eluting first are the high molecular weight glutenin polymers or fraction F1, followed by the low molecular weight glutenin polymers (Fraction F2), the high and low molecular weight gliadins (F3 and F4) and finally non-gluten proteins of small molecular weight such as globulins and albumins (F5) (Figure 2.5.).



**Figure 2.5.** Example of a gluten protein chromatogram obtained after separation by SE-HPLC using the Profilblé® method and consisting of five peaks with different retention times and absorbances. From small to high retention time, the high and low molecular weight glutenins (F1 and F2), the high and low molecular weight gliadins (F3 and F4), and non-gluten proteins of small molecular weights (F5).

Reference: Morel *et al.*, 2020

### 2.9.2 Extraction of total protein and gluten proteins

The purity of seed materials of the six NILs (60 lines) was estimated beforehand to ensure their homogeneity.

This was done by assessing the HMW-GS pattern on three individual grains for each line.

Total protein extracts were prepared from individual grains with three grains extracted separately per line according to the following procedure: single grains were first crushed inside a piece of paper with a pair of pliers to give flour which was then poured into a 2mL Eppendorf tube containing 0.5mL of extraction buffer (50 mM Tris –HCl, pH 6.8, 2 % (w/v) SDS, 10 % (v/v) glycerol, 2 % (w/v) dithiothreitol) (DTT) and 0.1 % (w/v) bromophenol blue) and beads for sonication at a frequency of 30/s on a Qiagen MM300 Mixer RETSCH TISSUELYSER (Qiagen Str. 1, 40724 Hilden, Germany).

Gluten protein extracts were prepared to quantify the gluten protein groups (HMW-GS,  $\omega$ -gliadins, and  $\alpha$ -,  $\beta$ -, and  $\gamma$ -gliadins) of the 12 NILs allelic bulks.

The white flour of the 12 allelic bulks was extracted in triplicate following Wan *et al.* (2013) procedure to purify the gluten proteins.

Referring to Wan *et al.* (2013), 10 mg (+/- 0.5) of white flour was added to 200  $\mu$ L of extraction buffer (50 % (v/v) aqueous propan-1-ol, 2.5 % DTT) to solubilise the gluten protein. Then, the samples were shaken for 45 min at 50°C, centrifuged for 15 min at 14000 rpm and the supernatant was retrieved. The pellet was re-extracted using the same procedure to improve the extraction and the supernatants were combined and immediately freeze-dried for two days and stored at -20°C. Before SDS-PAGE, the samples were thawed and 200  $\mu$ L of loading buffer (50 mM Tris –HCl, pH 6.8, 2 % (w/v) SDS, 10 % (v/v) glycerol, 2 % (w/v) dithiothreitol) (DTT) and 0.1 % (w/v) bromophenol blue) was added.

This purity check was done to assess important contaminations or processing/handling errors on the NIL seed stock such as mislabelling for example but not small contaminations (e.g. mixing of grain between different NILs) as the number of grain extracted is too small. Pertaining to small contaminations (e.g. a NIL contaminated with a small number of grain from another NIL), we hope that they would be diluted and therefore would not impact the quality of the flour or the characteristic of the dough and the final baking product.

### 2.9.3. Separation of the gluten protein on Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE separates proteins according to their molecular mass after denaturation of their tertiary structure (linearisation) and masking of their intrinsic charge by a net negative charge with the combined action of the SDS and heating.

5 $\mu$ L of gluten protein extract (Chapter 2, 2.9.2) was introduced in a NuPAGE™ 4 to 12%, Bis-Tris, 1.0 mm, Midi Protein Gel, 26-well (Fisher Scientific, Leicestershire, UK) which had been immersed in a MOPS SDS Running Buffer (20X) (Fisher Scientific, Leicestershire, UK). A protein ladder (SeeBlue™ Plus2 Pre-stained Protein Standard, Thermofisher, UK) was included in the first well to control the migration which was set for 30 minutes at 30 mA followed by 3-4 h at 60 mA. The current set the negatively charged protein in motion toward the cathode (pole +) at different paces; small proteins migrate faster than larger ones as they are less retained by the gel matrix. After migration, the gels were stained overnight with a Coomassie blue solution (40% (v/v) methanol, 10%(w/v) trichloroacetic acid (TCA)) with gentle shaking to reveal the bands and de-stained for two days with a 10% (w/v) TCA solution to clarify the gel background. Finally, the gels were rinsed with distilled water and scanned.

### 2.9.4. Relative quantification of the three gluten protein groups (HMW-GS, $\omega$ -gliadins, and LMW-GS and $\alpha$ -, $\beta$ - and $\gamma$ -gliadins) separable by SDS-PAGE by image analysis and comparison of their proportion between the allelic pairs

After scanning the gels (Chapter 2, 2.9.3), the images were uploaded in the software ImageJ (Version 1.53k; Schneider *et al.*, 2012) to analyse the band intensities of three gluten protein groups (HMW-GS,  $\omega$ -gliadins, and LMW-GS and  $\alpha$ -,  $\beta$ - and  $\gamma$ -gliadins).

For each track (NIL), the band integrative density (i.e. area under the curve) of each gluten protein group and the integrative density of the whole track was calculated. The proportions of the three gluten protein groups were calculated by dividing their respective integrative densities by the integrative density of the whole track.

The mean proportions of the three gluten protein groups were then compared between the allelic pairs by one-way ANOVA with a factor QTL (12 levels) and six contrasts defined as follows:

- Contrast 1 compares allele 1B-a to allele 1B-b for the HMW-GS proportion
- Contrast 2 compares allele 2D-a to allele 2D-b for the for the HMW-GS proportion
- Contrast 3 compares allele 4D-1a to allele 4D-1b for the for the HMW-GS proportion
- Contrast 4 compares allele 4D-2a to allele 4D-2b for the for the HMW-GS proportion
- Contrast 5 compares allele 6A-a to allele 6A-b for the for the HMW-GS proportion
- Contrast 6 compares allele 7B-a to allele 7B-b for the for the HMW-GS proportion

The same reasoning was applied for the  $\omega$ -gliadins, and LMW-GS/ $\alpha$ -,  $\beta$ -,  $\gamma$ -gliadins.

## 2.10. Bioinformatic analysis

### 2.10.1. Investigation of Lipoxygenase genes as a candidate for QTL 4D-2 whiteness

Lipoxygenases are known bleaching agents (Leenhardt *et al.*, 2006). The Malacca allele at the 4D-2 QTL was found in the DH population to be associated with a whiter crumb than the Hereward allele. This analysis was performed to check the insertion of the peak marker of NILs 4D-2 QTL confidence interval in relation to a lipoxygenase encoding gene. For this purpose, the flanking sequence of the peak SNP marker AX-94454183 was retrieved on the database [CerealsDB](https://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php) (https://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php) (Wilkinson *et al.*, 2012) and was blasted against the reference sequence of Chinese Spring (IWGSC RefSeq v1.0) using the NCBI Blast web tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST\_SPEC=GeoBlast&PAGE\_TYPE=BlastSearch) (Sayers *et al.*, 2021).

The top hit giving the smallest E-value was selected and aligned against Chinese Spring (IWGSC RefSeq v1.0) on Ensembl plant for visualisation (Cunningham *et al.*, 2022). The expression pattern of the lipoxygenase gene (TRAESCS4D02G294100) collocating with marker AX-94454183 was retrieved from the Wheat Expression Browser exVIP (<http://www.wheat-expression.com/>) to verify its expression in the grain (Borrill *et al.*, 2016).

2.10.2. Blast of the NILs 1B QTL flanking markers on reference genome (RefSeq v1.0) of the wheat variety Chinese Spring.

The HMW-GS which are major quality determinants are encoded on the long arm of chromosome 1 (Branlard *et al.*, 2020). The position of the QTL 1B confidence interval was therefore checked in relation to the *Glu-B1* loci to verify that they did not co-locate, which otherwise would indicate that QTL 1B effect is linked to a mutation in the *Glu-B1* loci. The sequences of the markers gwm264 and barc8 flanking the NILs 1B QTL on the left and right sides respectively were retrieved from CerealsDB (Wilkinson *et al.*, 2012). The sequence of the *Glu-B1* loci, TraesCS1B02G329992, was retrieved from Ensembl plant (Cunningham *et al.*, 2022). Then, the markers and the gene TraesCS1B02G329992 sequences were both aligned simultaneously on the RefSeq v1.0 of Chinese Spring using Ensembl plant (Cunningham *et al.*, 2022) to compare their physical position.

## 2.11. Statistical analysis

All the analysis described thereafter were performed using the statistical software R (v4.2.0; R Core Team 2022). The graphics were made using the R package “ggplot2” (v3.3.6; Wickham and Winston, 2016)

2.11.1. Analysis of Extensograph data on the NILs QTL 4D-2 in the Rothamsted 2020-2021 trial. The mean maximum resistance ( $R_{max}$ ) and extensibility were compared between the two alleles by an unpaired t-test with a sample size of three pseudoreplicates.

2.11.2 Multiple regression on the flour colour (dry method) in the Rothamsted 2019-2020 trial. The impact of three known colour-altering factors: the bran and protein content, and the extraction rate was assessed using a multiple regression on the whole data set and t-tests to verify that the NILs 4D-2 QTL effect was not a consequence of the breadmaking process.

The multiple regression was performed on the whole dataset including the six NILs allelic pairs (60 lines) using the R package *olsrr* (v0.5.3; Hebbali 2022) to compute all possible regressions with three factors and to select the best regression model (i.e., with the highest  $R^2$  adjusted).

Three t-tests with a sample size of five biological replicates were carried out to compare the NILs 4D-2 allelic pairs means for bran and protein content and flour extraction rate.

### 2.11.3. Analysis of the QTL effects of the six NILs

#### 2.11.3.1. ANOVA on the six NILs sets allelic pairs to assess the QTL effects

Two ANOVA models were computed to assess the QTL effects. The loaf volume was analysed using one-way ANOVA with a treatment term “QTL” (12 levels). The number of cells and the whiteness of the crumb were analysed by two-way ANOVA with treatment term “QTL” (12 levels) and the term “QTL:Loaf” was added to the error term to take into account the nesting of the slices of bread into their corresponding loaf.

The two ANOVA models also test six contrasts comparing each allelic pair as follows:

- Contrast 1 compares allele 1B-a to allele 1B-b for the trait number of cells
- Contrast 2 compares allele 2D-a to allele 2D-b for the trait loaf volume
- Contrast 3 compares allele 4D-1a to allele 4D-1b for the trait number of cells
- Contrast 4 compares allele 4D-2a to allele 4D-2b for the trait crumb whiteness
- Contrast 5 compares allele 6A-a to allele 6A-b for the trait number of cells
- Contrast 6 compares allele 7B-a to allele 7B-b for the trait loaf volume

The ANOVA with contrasts method was chosen rather than a direct t-test to compare the allelic pairs to increase the sample size and to refine the estimation of the means.

#### 2.11.3.2. ANOVA on the six NILs sets allelic pairs to assess the evolution of colour during the breadmaking process

Besides colour measurements of the bread crumb, wet and dry flour colour were also measured (chapter 2, 2.8.2) to assess the evolution of the colour during the breadmaking process.

Firstly, the bread colour means of individual NILs allelic pair were estimated by a mixed linear model with a term “loaf” both included as fixed and random to account for the difference of loaves. This provides an unbiased estimation of the crumb colour means for the four replicates (set of 10 slices) of each NILs allele.

Secondly, a three-way ANOVA with three treatments: “QTL” (Malacca and Hereward), “Process of colour measurement” (dry, wet, and bread crumb) and “Year” (2019-2020 and 2020-2021) including the two and three-way interactions was computed to assess their effects on the whiteness.

Six pairwise contrasts were tested by the three-way ANOVA model to compare the 4D-2 allelic pairs within each process (bread, dry and wet) and year (2019-2020 and 2020-2021)). The contrasts were defined as follows:

- Contrast 1: compares allele 4D-2a to allele 4D-2b bread whiteness in 2019-2020
- Contrast 2: compares allele 4D-2a to allele 4D-2b bread whiteness in 2020-2021
- Contrast 3: compares allele 4D-2a to allele 4D-2b flour whiteness measured by the dry method in 2019-2020
- Contrast 4: compares allele 4D-2a to allele 4D-2b flour whiteness measured by the dry method in 2020-2021
- Contrast 5: compares allele 4D-2a to allele 4D-2b flour whiteness measured by the wet method in 2019-2020
- Contrast 6: compares allele 4D-2a to allele 4D-2b flour whiteness measured by the wet method in 2020-2021

## 2.12. The DH field design

The DH population, which comprises 111 lines including the parents Malacca and Hereward was grown in three field trials at Rothamsted in 2019-2020 (51°48'06"N, 000°23'42"W), and at Reading in 2020-2021(51°28'47"N, 000°53'59"W) and 2021-2022 (51°28'41"N, 000°54'06"W).

The lines were sown at a seed rate of 250 seeds.m<sup>-2</sup> and grown in 7.47 m<sup>2</sup> (4.15 m x 1.8 m) plots with application of standard agronomic treatments except nitrogen fertilisation which was 150 kg/ha which is slightly below the national average of 200kg/ha for breadmaking wheat (Table 2.4).

A low nitrogen fertilisation was chosen for sustainable reasons to identify cultivars performing well in low nitrogen environments.

The experiment was divided in three blocks of 100 plots each laid out in a square grid of 10 rows and 10 columns according to a Balanced Incomplete Block Design (BIBD) for Rothamsted and a Randomised Complete Block Design (RCBD) for Reading. The choice of the BIBD design was motivated by insufficient number of seeds for a few lines to sow three field replicates. The agronomy of the three field trials is described in table 2.4.

Table 2.4. Agronomy of the DH experiment including fertilisation rate and pesticide applications

Location	Year	Plot Name	Sowing	Harvest	Fertilization	Pesticide
Rothamsted	2019-2020	Meadow	29/10/2019	12/08/2020	16/03/2020: 185 kg/ha (DoubleTop)	04/11/2019: 5 kg/ha (Ironmax pro)
					14/05/2020: 290 kg/ha (Nitram)	04/11/2019: 1L/ha (Pontos + Firestarter + Velomax)
Reading	2020-2021	A1	15/10/2020	17/08/2021	29/09/2020: KCL (Potash) at 191kg/ha	07/07/2020: glyphosate at 4L/ha
					31/03/2021: 75kg N/ha + 40 kg SO <sub>3</sub> as ammonium nitrate and ammonium su	16/10/2020: pre emergence herbicide (Stomp Aqua) at 2.9L/ha
					29/04/2021: 75kg N/ha as ammonium nitrate	04/11/2020: Herbicide (Liberator) at 0.6L/ha
					30/04/2021: micronutrient by spraye; Manganese 15 at 3L per ha	30/03/2021: broadleaved herbicide (Ally Max) at 42g/ha
Reading	2021-2022	Broadmoor	44476	44764	28/09/2021: K2O (Potash) at 100 kg/ha	09/09/2021: glyphosate at 4 L/ha
					21/03/2022: 75kg N/ha + 40 kg S/ha as ammonium nitrate and ammonium si	15/10/2021: pre emergence herbicide (Stomp Aqua) at 2.9L/ha
					05/05/2022: 75kg N/ha as ammonium nitrate	02/11/2021: Herbicide (Liberator) at 0.6 L/ha
						24/03/2022: herbicide (Monitor) 25 g/ha
						28/03/2022: T0 fungicide (Axiator Xpro) at 1.25 L/ha
						08/04/2022: T1 fungicide (Ascra) at 1.2 L/ha
						14/05/2022: T2 fungicide (Revystar) at 1.2 L/ha
						02/06/2022: T3 fungicide (Tebucon) at 250 L/ha

## 2.13. Phenotyping of the DH lines

### 2.13.1 Measurement of grain nitrogen content by Near Infrared spectroscopy (NIRS)

Near Infrared spectroscopy (NIRS) is a rapid and non-destructive method that can be used to estimate the nitrogen content of wheat grains.

The NIRS data were previously calibrated against the reference method for N measurement (LECO). Here, NIRS was used to predict grain nitrogen content.

A small metallic plate was filled with cleaned grains and inserted into a FieldSpec®4 Standard-Res spectroradiometer (Malvern Panalytical, UK) which had been calibrated against the reference method for N measurement (AACC) Method 46-30. The absorption spectra were obtained with the software Indico Pro (Malvern Panalytical, UK) and analysed with the module IQ Predict of the GRAMS ActiveApp™ (Alphasoft, Dhaka, Bangladesh) to predict the nitrogen content. The protein content was calculated by applying a conversion factor of 5.7 to the N content.

### 2.13.2 Yield

The fresh grain weight (kg) from each plot was measured at harvest by a combine harvester and the fresh grain yield ( $\text{g}\cdot\text{m}^{-2}$ ) was calculated by dividing the fresh grain weight by the plot surface:

$$\text{Grain yield}(\text{g}\cdot\text{m}^{-2}) = \frac{\text{grain weight}(\text{kg}) \times 10^3}{\text{plot surface}(\text{m}^2)}$$

Fresh grain samples of 70-80 g were then taken out for each plot and dried overnight in an oven set at 105°C to determine their moisture content which was deduced from the following formula:

$$\text{Moisture content}(\%) = \frac{\text{fresh weight}(\text{g}) - \text{dry weight}(\text{g})}{\text{fresh weight}(\text{g})} \times 100$$

Finally, the grain yield at 15% moisture was calculated by the following formula:

$$\text{Yield}_{15\%} = \text{Yield}_{\text{ftw}} \times \frac{100 - \text{mc}_{\%}}{100 - 15}$$

A 15% moisture value was chosen as this is standard for the grain industry and allowed comparison with other studies such as Bogard *et al.* (2010).

## 2.14. Calculation of the GPD genotypic means

### Rothamsted 2019-2020 field trial

A simple linear regression was computed with the 300 lines (including the three field replicates and the parents Malacca and Hereward) using the statistical software R (v4.1.1; R Core Team 2021) and the residuals values (GPD) were recorded.

The GPD genotypic means were calculated using a mixed model with a fixed structure, Line, and a random structure, row\*column using the R package “lme4” (v1.1.30; Bates *et al.*, 2015) to account for the imbalance of the line treatment.

### Reading 2020-2021 and 2020-2022 field trials

Similar to the Rothamsted 2019-2020 trial, a simple regression was computed to obtain the individual residues values (GPD). The GPD genotypic means were this time estimated with a linear model with a treatment “variety” and a block structure and not with a mixed linear model as there was no need to correct for unequal replication of the treatments.

In 2020-2021, the homoscedasticity assumption of the residuals was violated and a  $\log_{10}(\text{GPD}) - 1.5$  transformation was therefore applied on the GPD values to improve the equality of variances among the genotypes.

## 2.15. QTL analysis

### 2.15.1. Single environment QTL analysis

The QTL analysis was performed in R using the package *qtl* (v.1.52; Broman *et al.*, 2003) and a script developed by Luzie Wingen (John Innes Centre, Norwich, UK). The script initially performs a Single Interval Mapping (SIM) detection that tests the presence of a QTL at the marker position and every 2cM between pairs of adjacent markers. Then, a CIM is run and the QTL effects and positions are refined by considering the SIM QTL as cofactors. Cofactors are introduced in the model to control to remove the influence of QTL outside the genetic interval which is tested. The QTL identified by CIM were then used to construct the final QTL model:

$$y_i = \mu + \sum_{q \in Q} \alpha_q x_{iq} + \varepsilon_i + e_i$$

Where:

- Q is the set of QTL,  $q = 1, \dots, Q$
- $y_i$  is the trait mean for genotype i
- $\mu$  is the overall mean

- $\alpha_q$  is the effect of QTL  $q$
- $x_{iq}$  is the genetic predictor of QTL  $q$  for genotype  $i$
- $\varepsilon_i$  is the genetic residual for genotype  $i$  (or residual if unit errors are omitted), assumed to follow a Normal distribution with mean 0 and variance  $\sigma^2$
- $e_i$  is the unit error for genotype  $i$ .

The significance of each individual QTL selected in the final model was assessed by backward multiple regression using the  $R^2$  criteria.

## 2.16. Comparison of seven senescence metrics between the DH lines.

### 2.16.1. Recording and conversion of heading dates (HD)

The HD of individual plots were scored by eye at the whole canopy level according to Zadoks decimal code at growth stage (GS) 55 (Zadoks *et al.*, 1974), that is when 50% of lines within the plot had half of their ear above the flag leaf ligule (Figure 2.6).

The monitoring of GS55 was done at least twice a week from the end of the booting stage (GS 49) until all the lines had reached GS55.

The HD were then converted into number of days elapsed between the 1<sup>st</sup> of January and the GS55 date to calculate the genotypic means.



**Figure 2.6.** Barley ear at GS 55 (heading date on the Zadoks scale) with half of the ear is located above the flag leaf ligule.

Reference: Alqudah and Thorsten, 2017

#### 2.16.2. Measurement of canopy greenness from HD to crop maturity using a Normalized Difference Vegetation Index (NDVI) sensor

During senescence the chlorophyll is progressively degraded causing gradual yellowing of the canopy that can be recorded using a NDVI sensor.

NDVI is a plant health indicator based on the crop differential light reflectance between the near infrared and red spectra which is calculated by the formula below:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Where:

- NDVI is a unitless number between -1 and 1. -1 to 0 values corresponds to water, clouds and soil observations whereas 0 to 1 values are observed for crop canopies.

- NIR is the reflectance in the near infrared spectrum
- RED is the reflectance in the red part of the spectrum

Green and healthy canopies strongly reflect NIR and absorb RED, resulting in a high NDVI value, while senescing canopies reflect less NIR and absorb less RED, resulting in a lower NDVI values.

In this study, the NDVI sensor was mounted on a four-wheel platform (Figure 2.7) developed at the University of Reading with the help of Richard Casebow and pushed above each plot to measure the greenness of the canopy at eight time points in the Reading 2020-2021 and 2021-2022 trials (Table 2.5). For each time point and each plot corresponds 19-21 NDVI measurements. The NDVI sensor was an Apogee S2-112 (Apogee Instruments, Logan, USA). The measurements of red and near-infrared (NIR) reflectances were done at 650 and 810 nm, respectively.

Table 2.5 Dates after Heading (GS 55) where NDVI was recorded for Reading 2020-2021 and 2021-2022

Environments	NDVI recording date							
Reading_2020_2021	02/06/2021	15/06/2021	30/06/2021	08/07/2021	13/07/2021	19/07/2021	23/07/2021	29/07/2021
Reading_2021_2022	01/06/2022	07/06/2022	14/06/2022	22/06/2022	29/06/2022	26/07/2022	13/07/2022	21/07/2022



Figure 2.7. Four-wheel platform equipped with a NDVI sensor (black arrow) to measure senescence.

### 2.16.3 Modelling of senescence and calculation of seven parameters from the senescence curves

The temperature data in Reading 2020-2021 and Reading 2021-2022 were retrieved from a field meteorological station, property of the University of Reading to calculate thermal time after HD. For each plot, the date of the eight NDVI measurements was converted in thermal time (in °C) by summing the mean daily temperatures from HD (included).

In each year, four logistic (inverse S-shape curve) models were applied to the data to represent the senescence of individual plot as NDVI over thermal time and were compared using the standard curves menu of Genstat to select the most parsimonious model.

The four models had the following equation:

$$NDVI = A + \frac{C}{1 + e^{-B(Degree\ Day - M)}}$$

And they are given in order of increasing complexity:

- Model 1: corresponds to a single sigmoid curve applied to all the lines with no grouping
- Model 2 or parallel curve model: is allowing one constant parameter (A) to change between groups; the other parameters are kept constant. There were 300 groups corresponding to the DH lines (including the block replicates).
- Model 3: is allowing parameters C and A to change between each group keeping B and M constant across the lines.
- Model 4 or separate lines model is the most complex and is allowing the four parameters (A, B, C, and M) to vary between the groups.

The most parsimonious of these four models was chosen using the accumulated analysis of variance function of GenStat based on a p-value inferior to 5%.

Seven parameters: thermal time at onset or 10% of senescence (TT90), thermal time at mid or 50% of senescence (TT50), thermal time at the end or 90% of senescence (TT10), their corresponding NDVI values (NDVI90, NDVI50, NDVI10), a NDVI value before senescence (NDVI\_Max), and the maximum rate of senescence (Figure 2.8) were then calculated from the model equation.

TT90, TT50, and TT10 are the thermal time values in degree days corresponding to a drop of 10%, 50% and 90% of maximum NDVI (predicted from the model), respectively and were calculated using the following formula:

$$\text{Degree day} = - \left( \ln \left( \frac{C}{(NDVI - A)} \right) \times \frac{1}{B} \right) + M$$

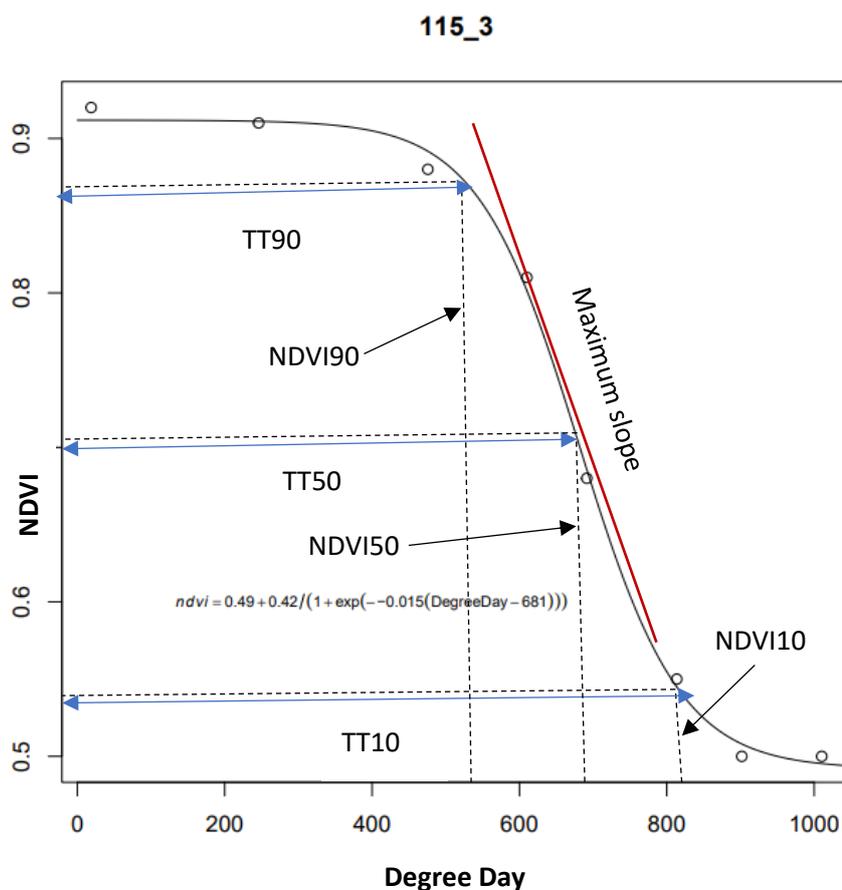
The maximum rate of senescence is given in decrease of NDVI per degree day and has a negative sign. It was calculated using the following formula:

$$\text{Maximum rate} = \frac{B \times C}{4}$$

NDVI90, NDVI50, NDVI10 were calculated as follows:

$$NDVI_i = \text{Minimum NDVI} + i \times (\text{Maximum NDVI} - \text{Minimum NDVI})$$

- Where i is the percentage of senescence



**Figure 2.8.** Example of a senescence curve fitting the NDVI (Y axis) against the thermal time (X axis) for line 115 in block 3. The maximum rate ( $B \times C/4$ ) is indicated by a plain red line and corresponds to the differential of the curve at the inflection point ( $A + C / 2$ ), the thermal times at onset (TT90), mid senescence (TT50), and senescence completion (TT10) are marked with horizontal blue arrows while their corresponding NDVI values (NDVI90, NDVI50, and NDVI10) which are vertical distances are marked with a black arrow.

Finally, the genotypic means (means of the three blocks replicates) of the seven parameters were estimated using a mixed linear model in R (package lme4, v1.1.30; Bates *et al.*, 2015).

## Chapter 3. Study of the effect and stability of six NILs QTL

### 3.1 Background

Within the framework of the HGCA project (Chapter 2, 2.1.), three DH populations (Malacca x Hereward, Shango x Shamrock, Malacca x Charger) segregating for major quality traits such as flour colour, flour brightness, extraction rate, dough performance, and final bake performance, were developed and grown over two years in 2005 and 2006. The DH lines were subsequently milled, and transformed into bread following three procedures: Chorleywood, spiral white, and puff pastry. A QTL detection was performed and identified 179 QTL in the Malacca and Hereward population combining the three bread processes. Eight of these QTL had a consistent effect on both years and were therefore selected and introgressed into NILs (Chapter 2, 2.1.2) in Simon Griffiths' lab at the John Innes Centre.

### 3.2. A short description of the cultivars

Hereward the Wake was an Anglo-Saxon nobleman who resisted the Normans during the invasion of 1066 organising a local rebellion against William the Conqueror. Perhaps as an homage, the name was latter given to a high standard breadmaking wheat variety. Hereward is a winter wheat variety developed in the UK in 1989, potentially (disputed) from a cross between Disponent and Norman (Figure 3.1, left) and was classified in the group 1 of the AHDB recommended list where it remained for an unusually long time before being outclassed in 2010 by higher yielding varieties (Shewry *et al.*, 2012).

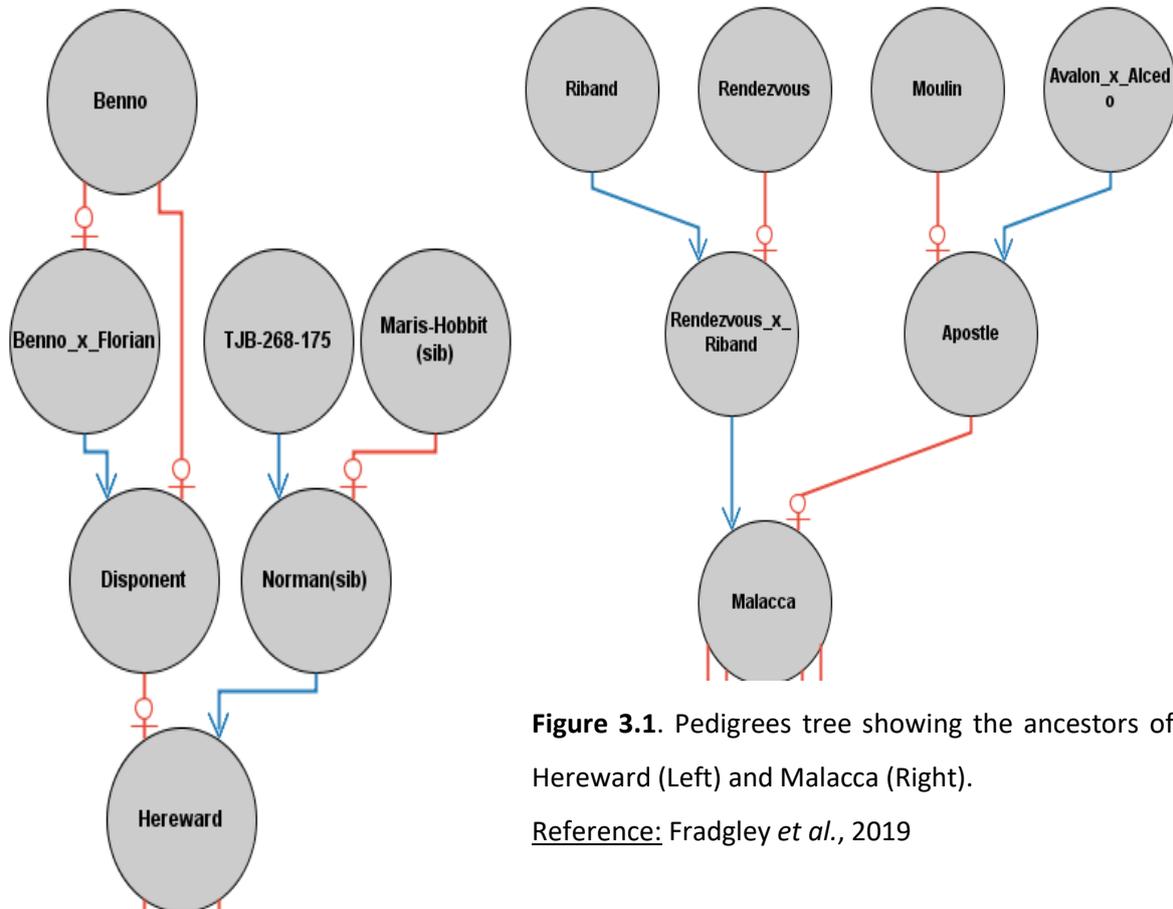
Its success was attributable to its high protein content (13.2% dry weight) and yield (9.48 t. ha<sup>-1</sup>) (HGCA, 2007). However, despite of its good performance, the intrinsic quality of Hereward is still poorly understood, and it lacks good high molecular weight glutenin subunits (HMW-GS) alleles at the major glutenin loci (*Glu-1*), notably *Glu-A1* (1), *Glu-B1* (7+8), and *Glu-D1* (5 + 10) (Chapter 1, 3.1.).

Instead, Hereward *Glu-1 loci* is composed of *Glu-A1* (Null), *Glu-B1*(7+9), and *Glu-D1*(2+12) which gives it a low *Glu-1* score of four out of ten (Payne *et al.*, 1987).

To this date, the Hereward quality paradox is still not full solved, but a first insight was brought by Min *et al.* (2020) with the discovery of a QTL for galactolipid, a component that stabilise the surface tension of air bubbles in the dough.

Malacca was bred in the UK by KWS in 1994 from a cross between Apostle and Rendez-vous x Riband (Figure 3.1., right) (Fradgley *et al.*, 2019). It was also placed in group 1 of the AHDB

recommended list and in a HGCA national trial conducted in 2007 it achieved a yield of 9.58 t. ha<sup>-1</sup> and a protein content of 12.4 %. Its *Glu-1* composition: *Glu-A1* (Null), *Glu-B1*(17+18), and *Glu-D1*(2+12) and its *Glu-1* quality score (5) are comparable to Hereward.



**Figure 3.1.** Pedigrees tree showing the ancestors of Hereward (Left) and Malacca (Right).

Reference: Fradgley *et al.*, 2019

### 3.3. Project objectives

This project is examining six sets of NIL in two environments (Rothamsted 2019-2020 and Reading 2020-2021) with the aim of assessing for each of them their quality-associated QTL effect. The NILs QTL are described in Table 2. (Chapter 2, 2.2).

Thereafter, the six NILs QTL will be referred by their chromosome number followed by the letter a or b indicating either the Malacca or the Hereward allele, respectively.

### 3.4. Results

#### 3.4.1. Positioning of NIL 1B QTL on the reference genome (RefSeq v1.0.) of the wheat variety Chinese Spring.

NIL 1B QTL may co-locate with the *Glu-B1* loci. If this was the case, the QTL effect may be caused by allelic variation at the *Glu-B1* loci.

To verify this, the position of the NIL 1B QTL was compared with the position of the *Glu-B1* locus on the reference sequence (RefSeq v1.0) of the wheat variety Chinese Spring (Chapter 2, 2.10.2)

The blast analysis located the NIL 1B QTL between 10104617 bp and 42329355 bp, that is approximately 513 Mb upstream the *Glu-1* loci which was situated between 555933489 bp and 555935716 bp.

#### 3.4.2. Assessment of the NILs kinship using hierarchical clustering on the genotypic data.

The NILs genotyping data were used to perform a hierarchical clustering and to group the NILs according to their kinship (Chapter 2, 2.4).

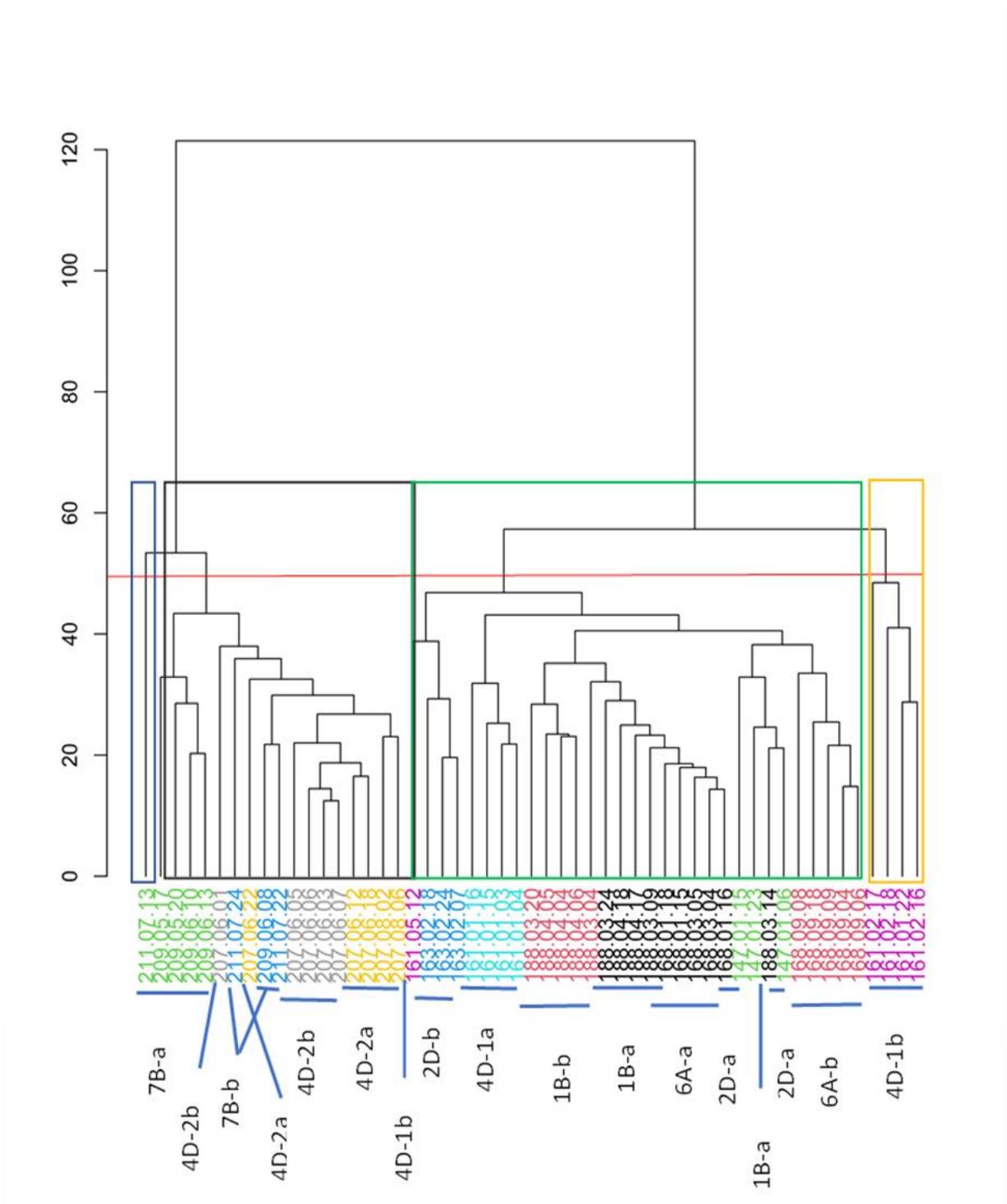
The resulting dendrogram is presented in Figure 3.2.

On the one hand, we would expect the NILs to be grouped by genetic backgrounds, Hereward or Malacca, which accounts for 94% of their genome.

On the other hand, we would anticipate the sister lines of each allele at the QTL to cluster together as they have the same ancestors (Grand-parents and/or parents) (Chapter 2, 2.1.2.)

The dendrogram (Figure 3.2.) displays the Euclidian distance on the vertical axis and the NILs names on the horizontal axis. The height at which any two objects (vertical lines) intersect corresponds to the Euclidian distance between the lines. Closely related NILs with high genetic resemblance will have a low Euclidian distance in comparison to distantly related ones. As expected, in most cases the sister lines of each allele at the QTL clustered together. Indeed, the NILs sister lines bearing the 7B-b, 4D-2a, and 4D-2b allele were grouped in the black cluster, and the NILs sister lines bearing the 1B-a, 1B-b, 2D-a, 2D-b, 4D-1a, 6A-a, and 6Ab

in the green group. However, for NILs 4D-1b and 7B-a, two sister lines (161-05-12 and 211-07-13) clustered apart in the green and blue groups respectively (Figure 3.2)



**Figure 3.2. Dendrogram of the NILs Euclidian distances derived from their Axiom genotypes.** The sister lines are named according to their QTL and parental allele which figures below their line number. The red line marks the threshold at which the dendrogram was cut to form four groups. The groups are highlighted by colour rectangles.

### 3.4.3. Analysis of the gluten subunits separated by SDS-PAGE in the six NILs sets

#### 3.4.3.1. Assessment of the HMW-GS band pattern

Visual assessments of the HMW-GS profiles of the six NILs sets were performed after separation of the gluten proteins by SDS-PAGE (Chapter 2, 2.9.3).

Three grains were therefore randomly selected from each NIL set and extracted to verify the absence of contamination of the seed material during harvest or handling.

NILs 1B, 2D, 4D-1, and 6A are in the Malacca background and should therefore display the Malacca HMW-GS allelic composition: *Glu-A1* (NULL), *Glu-B1* (17+18) and *Glu-D1* (2+12).

NILs 4D-2 and 7B were backcrossed with Hereward and are expected to show the Hereward HMW-GS composition: *Glu-A1* (NULL), *Glu-B1* (7+9) and *Glu-D1* (3+12). Malacca and Hereward parental controls were included in the test gels to compare with the NILs HMW-GS band profile but are not shown in Figure 3.3. Instead, NILs 6A-b and 7B-b were taken as references to show the Malacca and Hereward HMW-GS bands patterns, respectively (Figure 3.3., gels c and d).

The HMW-GS bands profiles of the three replicates of NILs 1B, 2D, 4D-1, and 6A were similar to the parent Malacca (Figure 3.3, gel c track 1).

Similarly, the three replicates of NILs 4D-2 and 7B HMW-GS bands profiles matched the parent Hereward (Figure 3.3, gel d track 2).

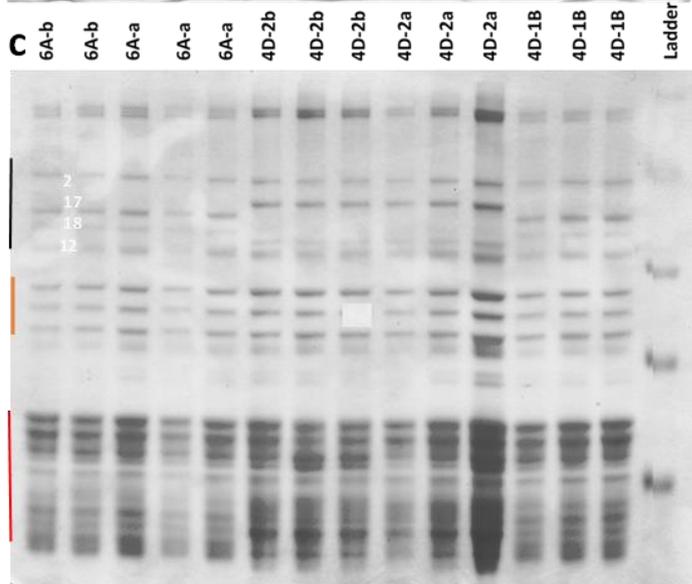
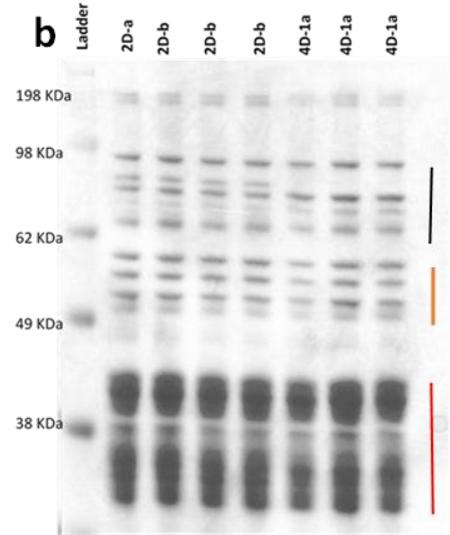
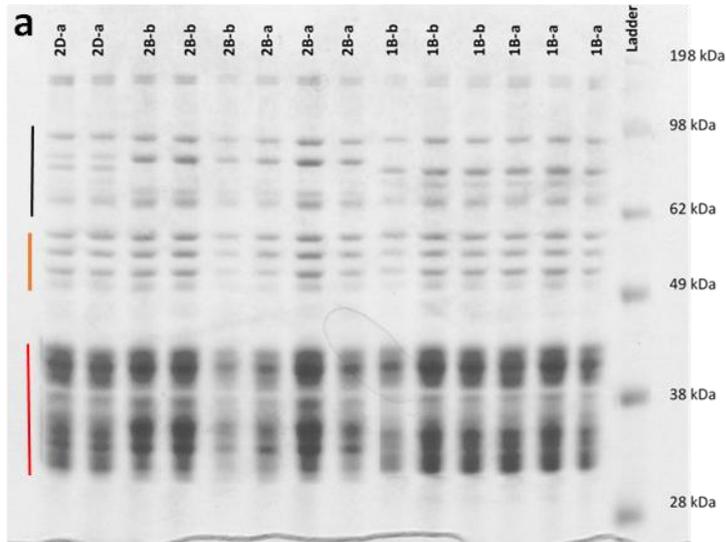
The three grains, sampled randomly from each NILs set, showed the expected HMW-GS parental profile and the entire seed set was therefore assumed to be pure.

#### 3.4.3.2 Comparisons of gluten proteins proportions of three groups separable by SDS-PAGE between the allelic pairs

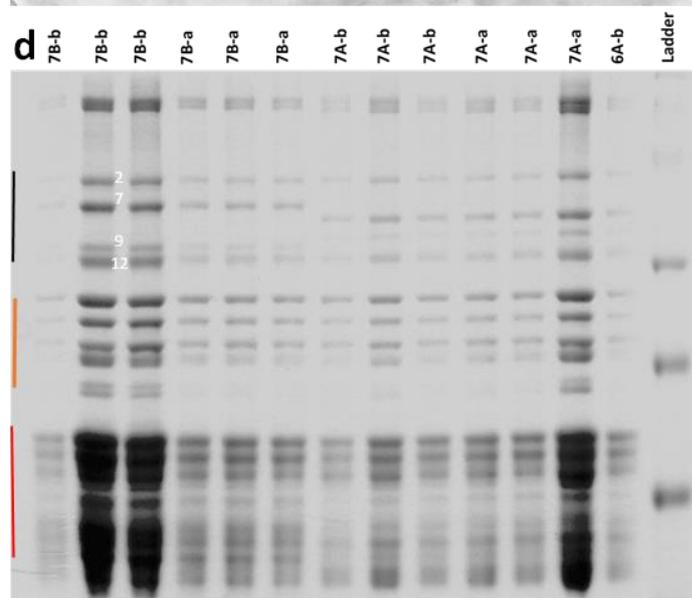
The relative proportion of HMW-GS and the ratio (HMW-GS/LMW-GS) were shown to be positively correlated with the dough rheological properties while the LMW-GS proportion was found to be negatively correlated with the dough rheological properties (Chapter 1., 4.2.3.)

Consequently, the mean proportions (expressed in percentage of total gluten protein – TA) of HMW-GS,  $\omega$ -gliadins, and LMW-GS/ $\alpha$ -,  $\beta$ -,  $\gamma$ -gliadins groups separated by SDS-PAGE (Chapter 2, 2.9.3) and the ratio HMW-GS/LMW-GS were compared between the allelic pairs grown at Rothamsted in 2019-2020 (Chapter 2, 2.1.14.).

The HMW-GS proportion ranged from 13% TA (4D-1a) to 0.22% TA (1B-a).



198 kDa  
98 kDa  
62 kDa  
49 kDa  
38 kDa



198 kDa  
98 kDa  
62 kDa  
49 kDa  
38 kDa

**Figure 3.3.** SDS-PAGE gels showing the separation of the gluten proteins for the six NILS sets.

The vertical lines indicate HMW-GS (black),  $\omega$ -gliadins (orange), and LMW-GS and  $\alpha$ -,  $\gamma$ -gliadins (red) band groups. A molecular weight ladder providing indication of band size was added in the last track of the gel on the right-hand side. The HMW-GS subunits of NILS 6A-b and 7B-b (gels c and d) were annotated and correspond to the parents Malacca and Hereward, respectively.

ANOVA (Table 3.1.) showed that the HMW-GS proportions differed significantly for QTL 4D-1 ( $F_{2,32} = 12$ ,  $p=0.001$ ), with a calculated difference in mean proportion of 5% TA and a 95% confidence interval of [0.6% TA - 9.3% TA].

The proportion of  $\omega$ -gliadins ranged from 11% TA (2D-b and 4D-1a) to 15% TA (6A-a and 4D-2a) and differed statistically for the QTL 1B ( $F_{2,32} = 10.90$ ,  $p=0.0002$ ), 2D ( $F_{2,32} = 12.91$ ,  $p=0.001$ ), and 4D-1 ( $F_{2,32} = 15.48$ ,  $p=0.0004$ ). The calculated differences for the means and the 95% confidence intervals were: 0.02% TA [-1.7% TA - 1.7% TA], 3% TA [1.04% TA - 4.9% TA], and 2% TA [-4.1% TA, 2.5% TA], respectively (Table 3.1).

The proportion of LMW-GS and  $\alpha$ -,  $\beta$ -,  $\gamma$ -gliadins were comprised between 43% TA (6A-b, and 7B-a) and 85% TA (1B-a) and varied significantly for QTL 2D ( $F_{2,32} = 4.35$ ,  $p=0.044$ ) and 4D-1 ( $F_{2,32} = 14.84$ ,  $p=0.0005$ ) with the calculated differences between the means and the 95% confidence interval being 16% TA [-7.5% TA – 39.5% TA] and 27% TA [-72% TA – 35% TA],

**Table 3.1.** Differences in the proportions of  $\omega$ -gliadins, LMW-GS, HMW-GS, and ratio HMW-GS/LMW-GS between the allelic pairs of the six NILs sets.

The significant comparisons are highlighted in red and the stars indicates thresholds of significance: \* ( $p<0.05$ ), \*\*( $p<0.01$ ), and \*\*\* ( $p<0.001$ ). The values are the means of three technical replicates which were sampled from the same bulk flour but extracted for protein and analysed by SDS-PAGE separately. The proportions are expressed in %TA and the ratio is unitless.

<b>NILs</b>	<b><math>\omega</math>-gliadin</b>	<b>LMW-GS</b>	<b>HMW-GS</b>	<b>HMW/LMW</b>
1B-a	0.02%***	6.19%	0.94%	0.08
1B-b				
2D-a	2.17%**	15.24%*	2.32%	0.11*
2D-b				
4D-1a	2.3%***	27.20%***	5.06%**	0.22***
4D-1b				
4D-2a	0.81%	1.54%	1.26%	0.04
4D-2b				
6A-a	5.92%	0.56%	1.25%	0.02
6A-b				
7B-a	0.89%	4.83%	0.99%	0.04
7B-b				

respectively (Table 3.1). Finally, the ratio HMW-GS/LMW-GS ranged from 0.18 (4D-1a) to 0.43 (6A-a) and differed significantly for QTL 2D ( $F_{2,32} = 5.04$ ,  $p=0.03$ ) and 4D-1 ( $F_{2,32} = 18.80$ ,  $p<0.001$ ).

#### 3.4.4. Assessments of the NILs QTL effects

The six NILs are presented in Table 3.2. with a description of their associated quality traits and their favourable alleles.

The effects of the six favourable alleles were assessed in Rothamsted in 2019-2020 using

**Table 3.2.** QTL chromosomal locations, associated quality traits and favourable alleles.

Chromosome	Trait	Favourable allele
1B	Number of cells	Hereward
2D	Bake height	Hereward
4D	Number of cells, crumb whiteness ( L*)	Hereward
4D	crumb whiteness ( L*)	Malacca
6A	Number of cells	Hereward
7B	Bake height	Malacca

ANOVA (Chapter 2, 2.4.3).

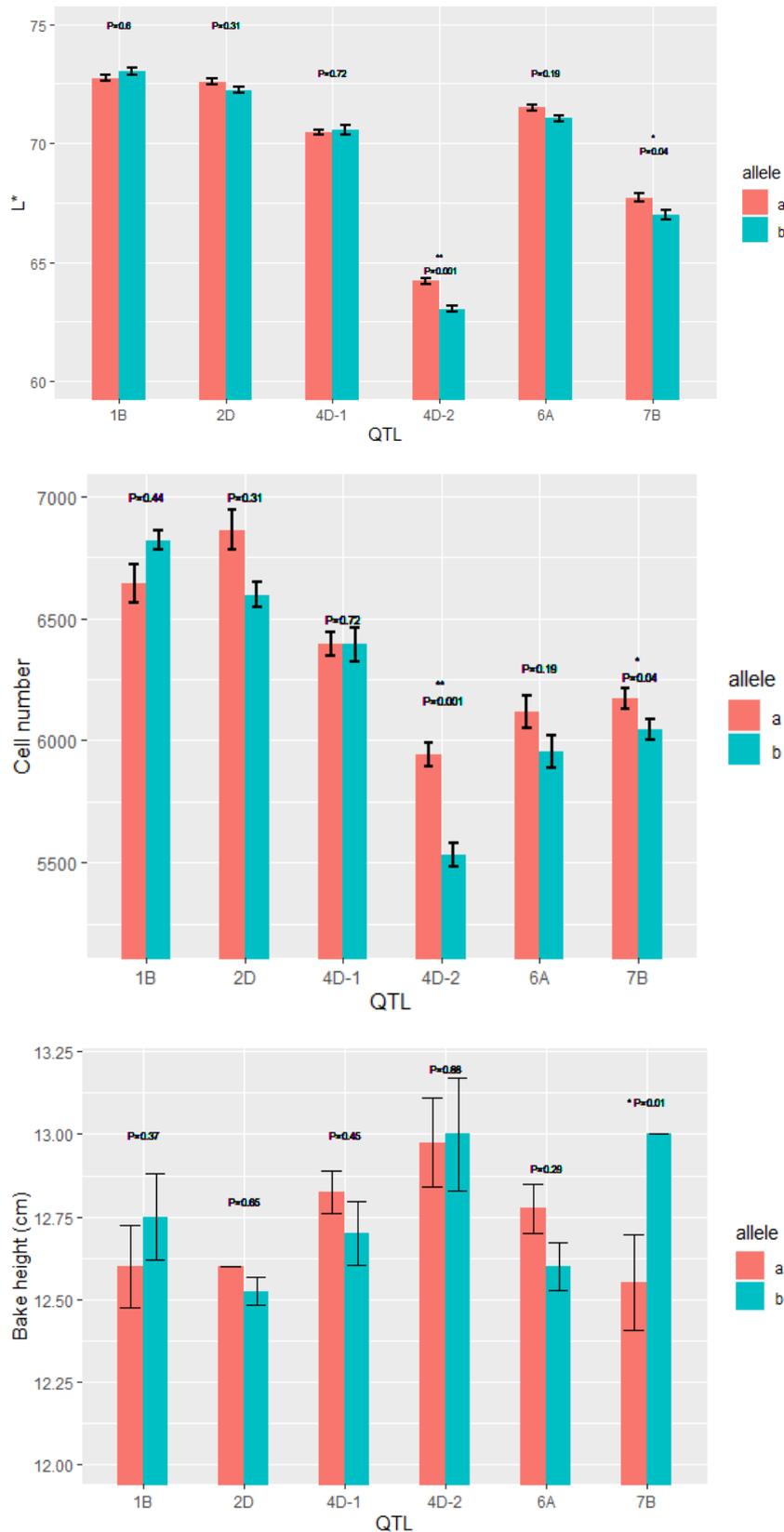
The results of the ANOVA are summarized in Figure 3.4 which compares the allelic pairs of the six NILs for three traits: crumb whiteness (L\*), loaf volume (bake height), and number of cells in the bread crumb (cell number).

The allelic pairs of QTL 4D-2 and 7B differed significantly ( $F_{1,36} = 11.76$ ,  $p < 0.001$ , and  $F_{1,36}=7.3$ ,  $p = 0.01$ ) for crumb whiteness (Figure 3.4., a). The favourable allele at QTL 4D-2 (4D-2-a) had an additive effect of +0.25 whereas the favourable 7B QTL allele (7B-a) had an effect of +0.11 on the brightness value.

The allelic pairs of QTL 4D-2 differed significantly ( $F_{1,36} = 7.66$ ,  $p < 0.001$ ) for the number of cells with the favourable allele (4D-2-a) showing an increase of +0.18 for the cell number value (Figure 3.4., b). The p-value for the comparison of QTL 2D allelic pairs were just above the significance threshold ( $\alpha=0.05$ ) ( $F_{1,36} = 3.2$ ,  $p = 0.08$ ).

The allelic pairs of QTL 7B significantly differed ( $F_{1,36}=7.32$ ,  $p=0.01$ ) for the bake height with the favourable allele (7B-a) having a negative effect of -0.17 cm (i.e. decreasing the height of 0.17cm) (Figure 3.4., d).The effect of QTL 4D-2 favourable allele (4D-2a) was re-assessed in Rothamsted 2020-2021 to confirm its effect. The five other NILs sets were not re-analysed as the comparisons were either not significant (1B, 2D, 4D-1, 6A) or significant (7B) but with a change of increasing allele between the DH (Malacca) and the NIL (Hereward). The effect of

QTL 4D-2 on the flour and crumb whiteness was determined for the 2019-2020 and 2020-2021 samples and the results were compared by ANOVA (Chapter 2, 2.11.3.2). They are presented in Table 3.3.



**Figure 3.4.** Bar plots comparing the allelic pairs means for L\* (a), cell number (b) and bake height (c) of the six NILs set grown in Rothamsted 2019-2020. The Malacca allele is coloured in red and the Hereward allele in blue. The vertical bars are standards errors of the means. P-values compare alleles a and b at each QTL and are displayed on the top of each bar. Above the p-values, the stars indicate significant differences at \*(p<0.05), \*\* (p<0.01), and \*\*\*(p<0.001). The samples sizes were n=4 for bake height and n=40 for L\* and cell number.

The calculated means from ANOVA (Table 3.3., a) showed that the colour became darker at the end of the breadmaking process for both alleles and years. This was illustrated by lower mean values of L\* for the bread crumb than for the white flour (dry and wet slurry methods), the highly significant main effect of process ( $p < 0.001$ ) and the interaction terms in the ANOVA (Table 3.3., b).

The ANOVA table (Table 3.3., b) showed highly significant main effects for the process and the year ( $p < 0.001$ ) but the QTL main effect was not significant ( $p = 0.90$ ). The two (QTL: process, QTL: year, and process: year) and three-ways interactions effects (QTL: process: year) were also highly significant.

This suggests that colour brightness is affected by a combination of factors interacting together, including the QTL allele, the process, and the year of measurement.

The t-tests (Table 3.3., c) showed no differences in crumb whiteness in the 2020-2021 trial as the comparison (4D-2A VS 4D-2B BREAD 2021) was not significant ( $t(44) = 0.39$ ,  $p = 0.70$ )

In addition, no significant differences in whiteness were detected the same year in the flour by either the dry or wet slurry methods ( $t(44) = 0.91$ ,  $p = 0.54$  and  $t(44) = 2.05$ ,  $p = 0.09$ , respectively) (Table 3.3. c). In 2019-2020, however, the comparison of the allelic pairs showed a significant difference of 0.43 (+/- 0.15) of flour whiteness with the dry method ( $t(44) = 2.88$ ,  $p = 0.01$ ) but not with the wet method ( $t(44) = -0.38$ ,  $p = 0.70$ ). The Hereward allele had the highest whiteness score recorded on the dry flour.

**Table 3.3. Assessment of QTL 4D-2 effect on flour and crumb whiteness.** A: table of predicted L\* ANOVA means for the two alleles within each year and processes; B: ANOVA table assessing the effect of the QTL, the year, the process, and their interaction on the crumb whiteness; C: pairwise comparisons of six contrasts using t-test. SE is the standard error of the mean, Df is the degree of freedom, CI LL and CI UL are respectively the lower and upper limits of confidence interval of the mean. The significance of the difference is indicated by \*\*\* ( $p < 0.001$ ), \*\* ( $p < 0.01$ ) or \* ( $p < 0.05$ ).

<b>a</b>	<b>QTL</b>	<b>Process</b>	<b>Year</b>	<b>Predicted means</b>	<b>SE</b>	<b>Df</b>	<b>CI LL</b>	<b>CI UL</b>
	4D_2A	Bread	2019	64.20175	0.119004	44	63.96191	64.44159
	4D_2A	Bread	2021	72.89	0.119004	44	72.65016	73.12984
	4D_2A	DRY	2019	89.21	0.10644	44	88.99548	89.42452
	4D_2A	DRY	2021	89.588	0.10644	44	89.37348	89.80252
	4D_2A	WET	2019	83.2	0.10644	44	82.98548	83.41452
	4D_2A	WET	2021	85.12	0.10644	44	84.90548	85.33452
	4D_2B	Bread	2019	63.05423	0.119004	44	62.81439	63.29406
	4D_2B	Bread	2021	72.956	0.119004	44	72.71616	73.19584
	4D_2B	DRY	2019	89.644	0.10644	44	89.42948	89.85852
	4D_2B	DRY	2021	89.726	0.10644	44	89.51148	89.94052
	4D_2B	WET	2019	83.142	0.10644	44	82.92748	83.35652
	4D_2B	WET	2021	85.43	0.10644	44	85.21548	85.6445

<b>b</b>	<b>Factor</b>	<b>Df</b>	<b>Sum Sq</b>	<b>Mean Sq</b>	<b>F value</b>	<b>Pr(&gt;F)</b>	<b>Significance</b>
	QTL	1	0	0	0.0134	0.908404	
	Process	2	4239.7	2119.86	37421.9	2.20E-16	***
	Year	1	170.5	170.45	3008.98	2.20E-16	***
	QTL:Process	2	1.7	0.83	14.6281	1.35E-05	***
	QTL:Year	1	0.5	0.49	8.5701	0.005392	**
	Process:Year	2	197.5	98.77	1743.54	2.20E-16	***
	QTL:Process:Year	2	1.3	0.63	11.174	0.000119	***
	Residuals	44	2.5	0.06			

<b>c</b>	<b>Contrasts</b>	<b>Estimate</b>	<b>SE</b>	<b>t value</b>	<b>df</b>	<b>Pr(&gt; t )</b>
	4D-2A VS 4D-2B Bread 2019	-1.1475	0.1683	-6.8185	44	0
	4D-2A VS 4D-2B Bread 2021	0.066	0.1683	0.3922	44	0.7019
	4D-2A VS 4D-2B DRY 2019	0.434	0.1505	2.8832	44	0.0182
	4D-2A VS 4D-2B DRY 2021	0.138	0.1505	0.9168	44	0.5464
	4D-2A VS 4D-2B WET 2019	-0.058	0.1505	-0.3853	44	0.7019
	4D-2A VS 4D-2B WET 2021	0.31	0.1505	2.0594	44	0.0908

### 3.4.5. Multiple regression analysis on flour whiteness in the 2019-2020 NILs trial.

In the literature, the bran content, the extraction rate and the protein content were reported to affect the whiteness of the flour and the bread crumb (Oliver *et al.*, 1993; Scanlon *et al.*, 1993).

Their contribution to the whiteness of the flour measured by the dry method was investigated in the six NILs sets by multiple regression (Chapter2, 2.4.4.).

None of the three predictors (bran content, extraction rate and protein content) were highly correlated with each other ( $r > 0.80$ ) (data not shown) and were therefore all included in the initial regressions models. Eight models were evaluated, and the optimal model selected had the highest adjusted  $R^2$  ( $R^2 = 0.64$ ) and contained all three variables (Table 3.4.). The model had the following equation:

$$Lstar = 93.74 - 0.50 \times branscan - 0.08 \times protein\ content - 0.02 \times extraction\ rate$$

Table 3.4. R-Squares and adjusted R-squares of eight regression models. All models have the flour whiteness measured by the dry method as dependant variable but different sets of predictors (Branscan, extraction rate, and protein content) as independent variables.

Predictors	R-Square	Adj.R-Square
Branscan	0.49	0.48
Extraction_rate	0.09	0.08
Protein_content	<b>0.08</b>	0.06
Branscan + Extraction_rate	0.59	0.58
Branscan + Protein_content	0.52	0.51
Protein_content + Extraction_rate	0.2	0.18
Branscan + Protein_content + Extraction_rate	0.65	0.64

The multiple regression was statistically significant ( $F_{3,72} = 45.63$ ,  $p < 2.10 \cdot 10^{-16}$ ) and explained 64% of the variation of flour colour. The branscan ( $\beta = -0.50$ ,  $p = 1.14 \cdot 10^{-14}$ ), the protein content ( $\beta = -0.08$ ,  $p = 5.10 \cdot 10^{-3}$ ), and extraction rate ( $\beta = -0.02$ ,  $p = 2.46 \times 10^{-6}$ ) significantly influenced the flour brightness (Lstar). However, t-tests performed on the QTL 4D-2 alleles showed no

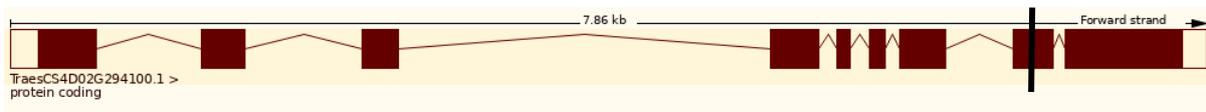
significant differences for these parameters: branscan ( $t(60) = -1.24$ ,  $p = 0.21$ ), protein content ( $t(60) = 0.252$ ,  $p = 0.80$ ) and extraction rate ( $t(60) = -0.53$ ,  $p = 0.59$ ).

### 3.4.6. Investigation of lipoxygenase gene in the QTL region of NILs 4D-2

Lipoxygenases are enzymes associated with flour bleaching (Chapter 1, 4.2.8) through oxidation of carotenoids (yellow pigments). Genes encoding for lipoxygenases were therefore searched within QTL 4D-2 confidence interval.

The blast of QTL 4D-2 peak marker AX-94454183 on the reference wheat genome (IWGSC RefSeq v1.0) gave a top hit with an extremely low E-value of  $4.5 \times 10^{-49}$ .

The marker flanking sequence completely aligned the target sequence, and marker AX-94454183 was located inside the TRAESCS4D02G294100 gene, which is predicted (based on sequence homology) to encode a lipoxygenase protein of 922 amino acids (The UniProt Consortium, 2019). TRAESCS4D02G294100 is formed of nine exons, and SNP AX-94454183 is located inside exon eight (Figure 3.5.) at 464537741 bp on the IWGSC RefSeq v1.0.



**Figure 3.5. Structure of the lipoxygenase gene TRAESCS4D02G294100 (LOX3).** This gene is 7.86 kb long and is formed of nine exons (red rectangles) and eight introns (red lines). The gene begins at 464531052 bp on the IWGSC RefSeq v1.0 and terminates at 464538915 bp. The black vertical line indicates the position of SNP AX-94454183 at 464537741 bp on the reference sequence.

Reference: Cunningham *et al.*, 2022

The consequence of the SNP marker AX-94454183 on the lipoxygenase protein sequence and function has not been studied experimentally but Cadenza tilling mutants with other SNP insertions at various locations in the TRAESCS4D02G294100 gene were developed and studied. Most of the variants (3) resulted in amino-acid substitutions (missense variants) while two of them did not change the protein sequence (synonymous mutations) (Table 3.5.). The SIFT code provides information on the consequence of an amino-acid substitution on the protein function. A SIFT code inferior to 0.05 indicates that the protein function is likely to be altered whereas a SIFT code superior to 0.05 suggests that the mutation is tolerated and the

protein function conserved. For the variants Cadenza0382 and Cadenza1443, substitutions of the amino acids histidine (basic amino acid) by tyrosine (Aromatic) and methionine (thiol group) by isoleucine, respectively, the function of the Lipoxygenase protein may be altered by a change of protein conformation caused by amino acids with different properties.

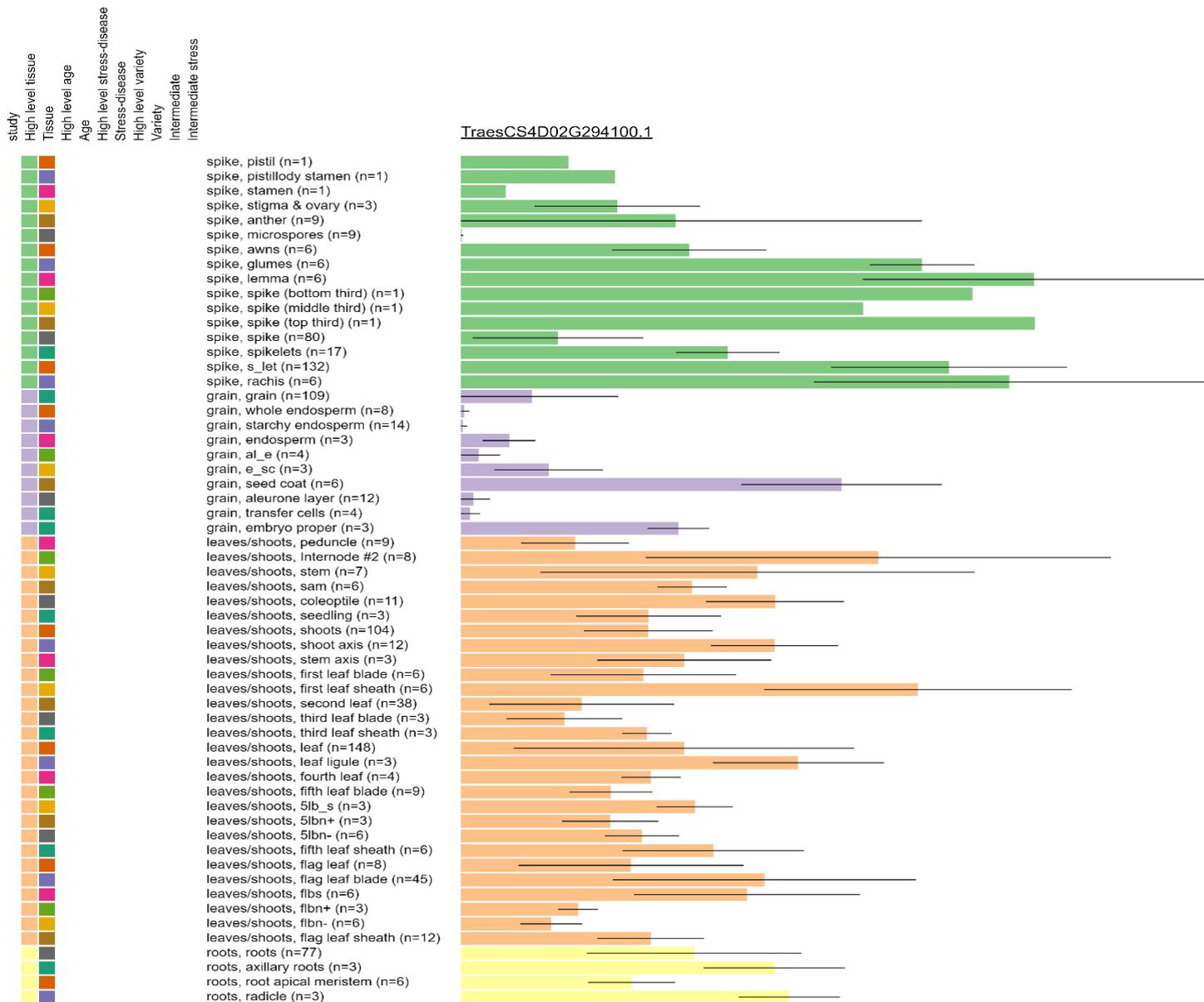
To alter the flour colour, TRAESCS4D02G294100 must be expressed in the grain and especially in the starchy endosperm, the main flour component. Therefore, the expression of TRAESCS4D02G294100 was assessed in wheat tissues at different developmental stages using the wheat expression atlas (Borill *et al.*, 2016).

Table 3.5. SNP variants identified in the Lipoxygenase gene TRAESCS4D02G294100 and their predicted consequences on the protein's function.

Reference: SNP data retrieved from Ensembl Plants (Cunningham *et al.*, 2022).

Variant ID	DNA location	SNP substitution	Source	Amino-acid change	Type of variant	SIFT	Consequence
Cadenza0382.chr4D.464537690	4D:464537690	C/T	EMS-induced mutation	H/Y	missense variant	0.04	deleterious
Cadenza0228.chr4D.464537695	4D:464537695	G/A	EMS-induced mutation		synonymous variant		
chr4D_scaffold24287_8836667	4D:464537706	C/T	Exome_Capture_Diversity	A/V	missense variant	0.34	tolerated
Cadenza1443.chr4D.464537710	4D:464537710	G/A	EMS-induced mutation	M/I	missense variant	0.03	deleterious
Cadenza0958.chr4D.464537758	4D:464537758	G/A	EMS-induced mutation		synonymous variant		

TRAESCS4D02G294100 was expressed in all plant tissues, including the spike, the leaves, the shoots, the stems, and the roots. It was highly expressed in the spikelet outer envelopes, the lemma (RNA transcript level of 8 log<sub>2</sub> (tpm)) and the glume (RNA transcript level of 6 log<sub>2</sub> (tpm)) (Figure 3.6.). In the grain, the gene was mainly expressed in the seed coat (6 log<sub>2</sub> (tpm)) and in the embryo (4 log<sub>2</sub> (tpm)) with little expression in the other grain tissues (<0.5 log<sub>2</sub> (tpm)), including the starchy endosperm (Figure 3.6).



**Figure 3.6. Amount of TRAESCS4D02G294100 transcripts in log<sub>2</sub>(tpm) in the tissues (spikes, grains, leaves, and roots) of the wheat variety Chinese Spring.** The different tissues are colour coded: in green the spike, in purple the grain, in orange the leaves, and in yellow the roots.

The horizontal bar represents the expression of the TRAESCS4D02G294100 in log<sub>2</sub>(tpm).

Reference: Borrill *et al.*, 2016

### 3.4.7. Assessment of glutenin polymers size distribution using Size-Exclusion High Performance Liquid Chromatography (SE-HPLC).

The SE-HPLC fraction F1% and the ratio F1/F2 were shown to correlate positively with the dough strength measured by the Extensograph parameter (resistance to extension). Similarly, the ratios  $(F3+F4)/(F1+F2)$  and  $(F3+F4)/F1$  were found to be positively correlated with the dough extensibility evaluated by the Extensograph (Godfrey *et al.*, 2010).

The glutenin polymers sizes distribution of the six NILs sets grown at Rothamsted in 2019-2020 were therefore analysed by SE-HPLC (Chapter 2, 2,9,1) and four fractions and ratios were compared between the allelic pairs: F1 (percentage of high molecular weight glutenin polymers), F1/F2 (ratio high molecular weight glutenin polymers to low molecular weight glutenin polymers),  $(F3+F4)/F1$  (ratio gliadin to high molecular weight glutenin polymers), and  $(F3+F4)/(F1+F2)$  (ratio gliadin to glutenin).

The protein content ranged from 28.70 (6A-b) to 31.30 (4D-2a) TA (Table 3.6) among the NILs. In term of protein composition, the  $\alpha$ - and  $\gamma$ -gliadins (F4) were the most abundant group accounting on average for 39.4% TA. The small glutenin polymers (F2), the large glutenin polymers (F1), and albumin and globulin (F5) represented 24.99 % TA, 14.25 % TA, and 12.55 % TA, respectively. The  $\omega$ -gliadins (F3) were the least represented protein group accounted for less than 10 % TA (Table 3.6.).

Regarding the SE-HPLC parameters that were found to correlate positively with the dough strength in Godfrey *et al.* (2010) study, the F1 ranged from 13.10% TA (1B-a) to 15.20% TA (2D-a) and the ratio F1/F2 varied from 0.50 (1B-a) to 0.61 (2D-a) (Table 3.6.).

As regard to those reported to be associated with the dough extensibility, the  $(F3+F4)/F1$  ratio were comprised between 3.12 (2D-a) and 3.80 (1B-a) and the  $(F3+F4)/(F1+F2)$  ratio varied between 1.16 (4D-1a) and 1.29 (1B-a) (Table 3.6.).

Comparison of the four SE-HPLC fractions and ratios between the allelic pairs showed the F1 differences ranging from 0.10 % TA (6A) to 1.10 % TA (4D-2), the F1/F2 differences varying between 0 (7B) and 0.06 (1B), the  $(F3+F4)/F1$  differences comprised between 0.04 (6A) to 0.28 (4D-2), and the  $(F3+F4)/(F1+F2)$  differences varying between 0.01 (6A) and 0.05 (2D) (Table 3.6).

**Table 3.6.** Results of the SE-HPLC performed on the six NILs sets flours. F1(large glutenin polymers), F2 (small glutenin polymers), F3 ( $\alpha$ -, and  $\gamma$ -gliadins), F4 ( $\omega$ -gliadins), and F5 (small albumin and globulin), F1/F2 (ratio large to small glutenin polymers), F3+F4/F1 (ratio gliadin to large glutenin polymers), F3+F4/F1+F2 (ratio gliadin to glutenin), and TA (total area under the SE-HPLC chromatogram which corresponds to the total protein content). Fractions F1-F5 are expressed in %TA.

QTL	F1	F2	F3	F4	F5	F1/F2	F3+F4/F1	F4+F3/F1+F2	TA
1B-a	13.10	25.10	9.20	39.90	12.60	0.50	3.80	1.29	29.40
1B-b	13.90	24.60	9.10	39.60	12.80	0.56	3.52	1.26	29.10
2D-a	15.20	25.10	8.60	38.80	12.30	0.61	3.12	1.18	29.30
2D-b	14.50	24.80	9.00	39.30	12.40	0.59	3.32	1.23	29.90
4D-1-a	15.00	25.30	8.60	38.20	12.90	0.59	3.13	1.16	29.50
4D-1-b	14.60	25.30	8.80	38.40	12.90	0.57	3.24	1.18	28.90
4D-2-a	15.20	24.90	8.30	39.70	12.00	0.61	3.16	1.20	31.30
4D-2-b	14.10	25.10	8.50	40.10	12.20	0.56	3.44	1.24	30.30
6A-a	13.80	25.00	8.90	39.30	12.90	0.55	3.49	1.24	29.30
6A-b	13.90	25.00	9.00	39.10	12.90	0.56	3.45	1.24	28.70
7B-a	13.70	24.70	9.00	40.20	12.40	0.56	3.58	1.28	31.10
7B-b	14.00	24.80	8.80	40.10	12.30	0.56	3.49	1.26	31.20
Range	13.10-15.20	24.60-25.30	8.30-9.20	38.20-40.20	12-12.90	0.50-0.61	3.12-3.80	1.16-1.29	28.70-31.30
Mean	14.25	24.98	8.82	39.39	12.55	0.57	3.40	1.23	29.83

### 3.4.8. Assessment of the dough mixing properties in the 2019-2020 NILs with the Farinograph

The Farinograph test is used to measure the water absorption (WA) and the strength of a flour (chapter 1, 4.1.). Four parameters were calculated from the Farinograph curve; three (dough development time, stability, and the degree of softening) measure dough strength and stability, and one measures the water absorption of the flour to reach a dough optimum consistency of 600 Brabender Unit (BU).

The WA is given in percentage and corresponds to the amount of water absorbed (in g) by 100 g of flour. The WA of the NILs flours ranged from 58.4 % (4D-2a) to 62.7% (7B-b) (Table 3.7).

The dough development time (DDT) is the time (in minutes) elapsed from water addition to the maximum strength. It measures the strength of the dough as strong doughs will take longer to develop than weak doughs.

The NILs DDT ranged from 3.2 minutes (4D-2b) to 4.1 minutes (2D-b) (Table 3.7.). This indicates that the gluten network formed quicker for NIL 4D-2b than for NIL 2D-b. When a dough is mixed beyond the DDT, its gluten network begins to break down thereby

decreasing its consistency. Dough stability measures the amount of time (in minutes) the consistency of the dough stays above the 600 BU lines which is the optimal consistency. The stability of the NILs ranged from 2.8 minutes (7B-a) to 3.8 minutes (1B-a) (Table 3.7.).

**Table 3.7.** NILs Farinograph results for four parameters: dough development time (in min), stability (in min), water absorption (in %) and degree of softening (in Brabender Unit – BU). The favourable allele at each QTL is highlighted in yellow. One flour sample (300g) was measured for each NILs allele.

QTL	WA (%)	DDT (min)	Stability (min)	DS (BU)
1B-a	60.2	4	3.8	110.6
1B-b	61.6	4	3.3	113.3
2D-a	60	4	3.8	110.8
2D-b	62.4	4.1	3.8	115
4D-1-a	61.1	3.8	3.3	135.3
4D-1-b	59.1	3.9	4	123.4
4D-2-a	58.4	3.7	3	149.7
4D-2-b	59.4	3.2	2.9	162.4
6A-a	62.1	3.8	3	139.7
6A-b	60.8	3.6	3	139.6
7B-a	60.3	3.4	2.8	171.7
7B-b	62.7	3.6	2.8	175.8

Finally, the degree of softening (DS) indicates the extent of breakdown occurring when the dough is overmixed. DS measures the drop of resistance in BU between the maximum resistance and 12 minutes after. The DS of the NILs ranged from 110.8 BU (2D-a) to 175.8 BU (7B-b) (Table 3.7.).

#### 3.4.9. Assessment of the dough strength and extensibility on the NILs from the 2019-2020 and 2020-2021 trials using the Extensograph

In 2019-2020, Extensograph measurements were carried out on three NILs (1B, 2D, and 4D-1) in first instance. The Extensograph results for two main parameters: resistance to extension and extensibility are provided in Table 3.8.

The doughs' resistance to extension ranged from 205 BU (4D-1a) to 248 BU (1B-a), and their extensibility varied between 18.1 cm (1B-a) and 20.9 cm (4D-1b). Differences between the

allelic pairs ranged from 4 BU (2D) to 27 BU (1B) in resistance to extension and from 0.1 mm (2D) to 1.2 mm (1B) in extensibility (Table 3.8).

**Table 3.8. Extensograph results for NILs 1B, 2D, and 4D-1 grown at Rothamsted in 2019-2020.** Two Extensograph parameters are reported to describe the physical properties of the dough: the resistance to extension in Brabender unit and the extensibility in cm. The NILs favourable alleles are highlighted in yellow. One flour sample (300g) was measured for each NILs allele.

QTL	Resistance to Extension (BU)	Extensibility (cm)
1B-a	248	18.1
1B-b	221	19.3
2D-a	210	19.6
2D-b	206	19.7
4D-1-a	205	20.5
4D-1-b	215	20.9
Range	205-248	18.1-20.9
Mean	217.5	19.68

In 2020-2021, Extensograph measurements were carried out in triplicate for the NIL 4D-2 allelic pairs (Table 3.9). NIL 4D-2 was selected as the QTL 4D-2 effect on the crumb whiteness was confirmed in Rothamsted 2019-2020.

The standard deviations were of 17.61 BU and 4.04 BU, and of 1.53 cm and 1.09 cm for the resistance to extension and extensibility of the allelic pairs 4D-2a and 4D-2b, respectively (Table 3.9). The sample means were of 148 BU and 25 cm for NIL 4D-2a and of 173 BU and 24.8 cm for NIL 4D-2b.

A t-test comparison of the means indicated no significant difference in extensibility between the 4D-2a and 4D-2b alleles ( $t(4) = 0.27$ ;  $p=0.79$ ) and no significant difference in resistance to extension between the alleles ( $t(4) = -2.49$ ;  $p = 0.08$ ).

**Table 3.9. Extensograph results for NIL 4D-2 grown at Rothamsted in 2020-2021.** The favourable allele is highlighted in yellow. Three samples of flour taken from the same bulk were analysed for each allele. They are labelled as “Rep” in the table below.

<b>QTL</b>	<b>Rep</b>	<b>Resistance to Extension (BU)</b>	<b>Extensibility (cm)</b>
4D-2a	1	128	22.8
	2	162	25.1
	3	153	25.7
	<b>Mean</b>	<b>148</b>	<b>25</b>
4D-2b	1	169	23.9
	2	174	26
	3	177	24.6
	<b>Mean</b>	<b>173</b>	<b>24.8</b>

#### 3.4.10. Exploration of correlations between SE-HPLC ratios and dough properties determined by the Extensograph

Pearson correlations of the SE-HPLC proportions F1 to F5 and ratios F1/F2, (F3+F4)/F1, and (F3+F4)/(F1+F2) with the Extensograph parameters resistance to extension and extensibility were evaluated for the three NILs (1B, 2D, and 4D1) measured in Rothamsted 2019-2020 (Table 3.10).

The large glutenin polymers (F1) were highly negatively and significantly correlated with the gliadin fractions F3 (high molecular weight gliadin) and F4 (low molecular weight gliadin) with  $r = -0.93$  and  $r = -0.84$ , respectively (Table 3.10). The two gliadin fractions F3 and F4 were highly significantly and positively correlated to each other ( $r = 0.90$ ).

One SE-HPLC fraction (F1) and three SE-HPLC ratios (F1/F2, (F3+F4)/F1, and (F3+F4)/(F1+F2)) were significantly correlated with the Extensograph parameter resistance to extension. The size of the correlations ranged from  $r = 0.80$  for the ratio (F3+F4)/(F1+F2) to  $r = -0.95$  for the ratio (F1/F2) (Table 3.10). One SE-HPLC proportion (F4), and two SE-HPLC ratios (F3+F4)/(F1+F2) and (F3+F4)/F1 were significantly correlated with the Extensograph parameter extensibility. All correlations were negatives and the strength of the correlations

varied between  $r = -0.83$  for the ratio  $(F3+F4)/F1$  and  $r = -0.91$  for the proportion F4 (Table 3.10).

None of the SE-HPLC fractions or the Extensograph parameters were significantly correlated with fraction F5 which contains non gluten proteins, namely globulins and albumins.

### 3.5. Discussion and conclusion

#### 3.5.1. Blast of the NIL 1B QTL on the reference genome (RefSeq v1.0) of the wheat variety Chinese Spring

The blast search located the NIL 1B QTL well upstream of the HMW-GS encoding loci *Glu-1B*. So consequently, the effect of the NIL 1B QTL observed in the DH population by Millar *et al.* (2008) could not be attributed to a genetic variation at the *Glu-1B* loci.

#### 3.5.2. Analysis of the NILs genetic kinship according to their genotypes on the Axiom 35K

As expected, most of the NILs clustered together in the same group. This was the case for NILs 7B-b, 4D-2a, 4D-2b, 1B-a, 1B-b, 2D-a, 2D-b, 4D-1a, 6A-a, and 6A-b.

Nevertheless, two lines: 161-05-12 and 211-07-13 clustered apart from their respective NILs groups (4D-1b and 7B-a) and were therefore more genetically distant from their four sister lines. This genetic difference could be explained by comparing the ancestry of the sister lines. The sister lines of QTL 4D-1b that cluster together have either parent 02 or 01 as ancestors; line (161-05-12), which clusters apart, has parent 05 and this is probably what caused it to separate apart. The same reasoning could be applied to lines 211.

#### 3.5.3. Comparisons of the proportion of HMW-GS, $\omega$ -gliadins, LMW-GS and $\alpha$ -, $\beta$ -, $\gamma$ -gliadin, and the ratio HMW-GS/LMW-GS between the allelic pairs

The ANOVA (Table 3.1) showed significant but small differences (5% TA) between the mean HMW-GS proportions of NIL 4D-1 allelic pairs. Significant but small differences (0.02% TA-2% TA) were also detected for the proportion of  $\omega$ -gliadins of NILs 1B, 2D, and 4D-1. Similarly, the LMW-GS proportions varied significantly for NILs 2D, and 4D-1 and the differences were higher than those observed for the HMW-GS and  $\omega$ -gliadins: 16% TA and 27% TA, respectively. The favourable allele of NIL 4D-1 QTL (4D-1b) was associated with an increase in the number of cells and in the whiteness of the crumb in the DH population. Here, in the 2019-2020 NILs

grown at Rothamsted, 4D-1b is associated with a higher proportion of HMW-GS and  $\omega$ -gliadins but with a lower proportion of LMW-GS than 4D-1a (Malacca).

The favourable allele of NIL 2D QTL (2D-b) increased the bread loaf volume in the DH population. Here, 2D-b had significantly lower  $\omega$ - gliadin and higher LMW-GS proportions than 2D-a.

Finally, the favourable allele of NIL 1B QTL (1B-b) improved the number of cells in the DH population. In this study, 1B-b had a significantly lower proportions of  $\omega$ - gliadin than 1B-a.

An attempt was made to quantify and compare the relative proportions of three gluten proteins groups by measuring their band intensity on SDS-PAGE gels using the software ImageJ. Caution must be exercised when interpreting these results, for the three protein groups: HMW-GS,  $\omega$ - gliadin, and LMW-GS reported on Table 3.1. did not add up exactly to 100%. This most likely would have occurred during the manual selection of the bands on the software ImageJ. The results are therefore to be treated with caution and should be used as an indication.

#### 3.5.4. Validation of the QTL breadmaking quality effects in the NILs

An ANOVA (Figure 3.4) was performed to compare the crumb whiteness, the number of cells and the bake height between the allelic pairs of six NILs grown at Rothamsted in 2019-2020. The ANOVA found no significant differences in the number of cells (1B, 4D-1, and 6A), in the bake height (2D, and 7B), and in the crumb whiteness (4D-1).

The effect of these QTL reported by Millar *et al.* (2008) were therefore not validated in the NILs grown at Rothamsted in 2019-2020. This absence of effect in the NIL population may be first attributed to the environment which may have impacted the ratio gliadin on glutenin or second to the NIL genetic background which is much more homogenous than in the DH. Indeed, in the DH population, the QTL was studied in a heterogenous background composed of Malacca and Hereward recombined DNA. The QTL effect might have been influenced by other QTL related to a similar trait. In the NIL, the QTL effect is studied individually in a homogenous parental background making comparison more reliable.

However, significant differences in the crumb whiteness and in the number of cells were detected between the allelic pairs of NILs 4D-2 and 7B (whiteness only). The loaves produced from the sister lines of NIL 4D-2 having the Malacca allele (4D-2a) had a higher mean crumb

whiteness than the loaves made from the sister lines of NIL 4D-2 bearing the Hereward allele (4D-2b). This confirms what was reported in the DH population (Millar *et al.*, 2008). In addition, the allele 4D-2a of NIL 4D-2 has the same effect on the whiteness (+0.25) than in the DH (+ 0.247).

Interestingly, the effects of QTL 4D-2 and 7B on, respectively, the number of cells and the brightness were not observed in the DH population (Millar *et al.*, 2008). In both cases, the sister lines having the Malacca allele (4D-2a and 7B-a) had the highest mean.

In 2020-2021, the comparison of crumb whiteness between NIL 4D-2 allelic pairs was not significant. Two hypotheses may explain the absence of confirmation of the NILs QTL effects:

1- The environmental conditions in which the NILs were grown

Bread is a complex product whose quality can be influenced by several factors such as the bread recipe, the breadmaking procedure and the flour intrinsic constituents.

Although the recipe and the breadmaking procedure can be controlled to some extent, the synthesis of grain components such as starch and gluten protein is partly under environmental control and is therefore subject to variation.

For instance, heat stress was shown to increase the gliadin-to-glutenin ratio through a higher rate of gliadin synthesis in comparison to glutenin, resulting in dough weakening and in a reduction of the loaf volume (Chapter 1, 5.1.). Differences in the weather conditions between 2006, when the DH population was phenotyped, and 2019-2021 when the NILs were grown, may explain the absence of effects for NILs 1B, 2D, 4D-1, and 6A in 2019-2020 and NIL 4D-2 in 2020-2021.

2- The sample size

In this study, the mean bake height was compared between the allelic pairs, with four loaves measured per allele. The amount of flour available for the rheological test and for baking was a limiting factor in increasing the sample size.

However, here, the small sample size may have prevented the detection of the QTL effect, especially for a small effect size. Ideally, a power analysis should be done to determine the appropriate sample size and to maximize the chance of detecting a difference between the allelic pairs.

In Rothamsted 2019-2020, three variables: the protein content, the extraction rate and the branscan were investigated for their impact on the NIL 4D-2 QTL flour whiteness. The primary

aim was to verify that the difference in the flour whiteness was not the consequence of a difference in the extraction rate, but instead was caused by an intrinsic flour component. T-tests performed on the 4D-2 allelic pairs showed no differences in protein content, extraction rate and bran, components known to impact the whiteness of the flour. This shows that the difference of whiteness between NIL 4D-2 alleles were not generated during the milling process (extraction rate) or caused by a difference of protein content in the flour.

In conclusion, the six NILs should be grown in additional environments to draw more firm conclusions regarding the stability of their QTL effects.

In addition, depending on the sample size suggested by the power analysis, the plot size may have to be increased to generate the amount of flour required.

### 3.5.5. Is allele 4D-2a bleaching effect caused by a higher level or activity of lipoxygenase?

My analysis showed that the peak marker of NIL 4D-2 QTL, AX-94454183, was located inside a coding sequence of the lipoxygenase gene TRAESCS4D02G294100. It is therefore possible that the alleles at this QTL differ in amount or activity of lipoxygenase protein.

The wheat gene expression atlas also shows that the expression of TRAESCS4D02G294100 is weak in the endosperm, but stronger in the seed coat and embryo; tissues not fully eliminated by the milling process and partly retrieved in the white flour, confirming the presence of the lipoxygenase protein in the white flour.

The next steps would be to determine the expression level of TRAESCS4D02G294100 in the endosperm of Malacca and Hereward grains, and to purify and quantify the product of TRAESCS4D02G294100 in their white flour. Finally, an enzymatic assay could be used to determine the lipoxygenase activity in the white flour of Malacca and Hereward.

### 3.5.6. Correlations between SE-HPLC ratios and dough properties

In this study, the ratios gliadin-to-large glutenin polymers  $((F3+F4)/F1)$  and gliadin-to-glutenin  $((F3+F4)/(F1+F2))$  were significantly negatively correlated with the Extensograph extensibility ( $r=-0.83$  and  $r=-0.87$ ,  $p<0.05$ , respectively) while the ratio large to small glutenin polymers  $(F1/F2)$  and the fraction F1 were significantly negatively correlated with the Extensograph resistance to extension parameter ( $r=-0.95$  and  $r=-0.91$ ,  $p<0.01$ , respectively). This disagrees with the study of Godfrey *et al.* (2010) who reported positive correlations for both

Extensograph parameters and was unexpected since the gliadins (fractions F3 and F4) are associated with dough extensibility and glutenin (fractions F1 and F2) with dough elasticity or

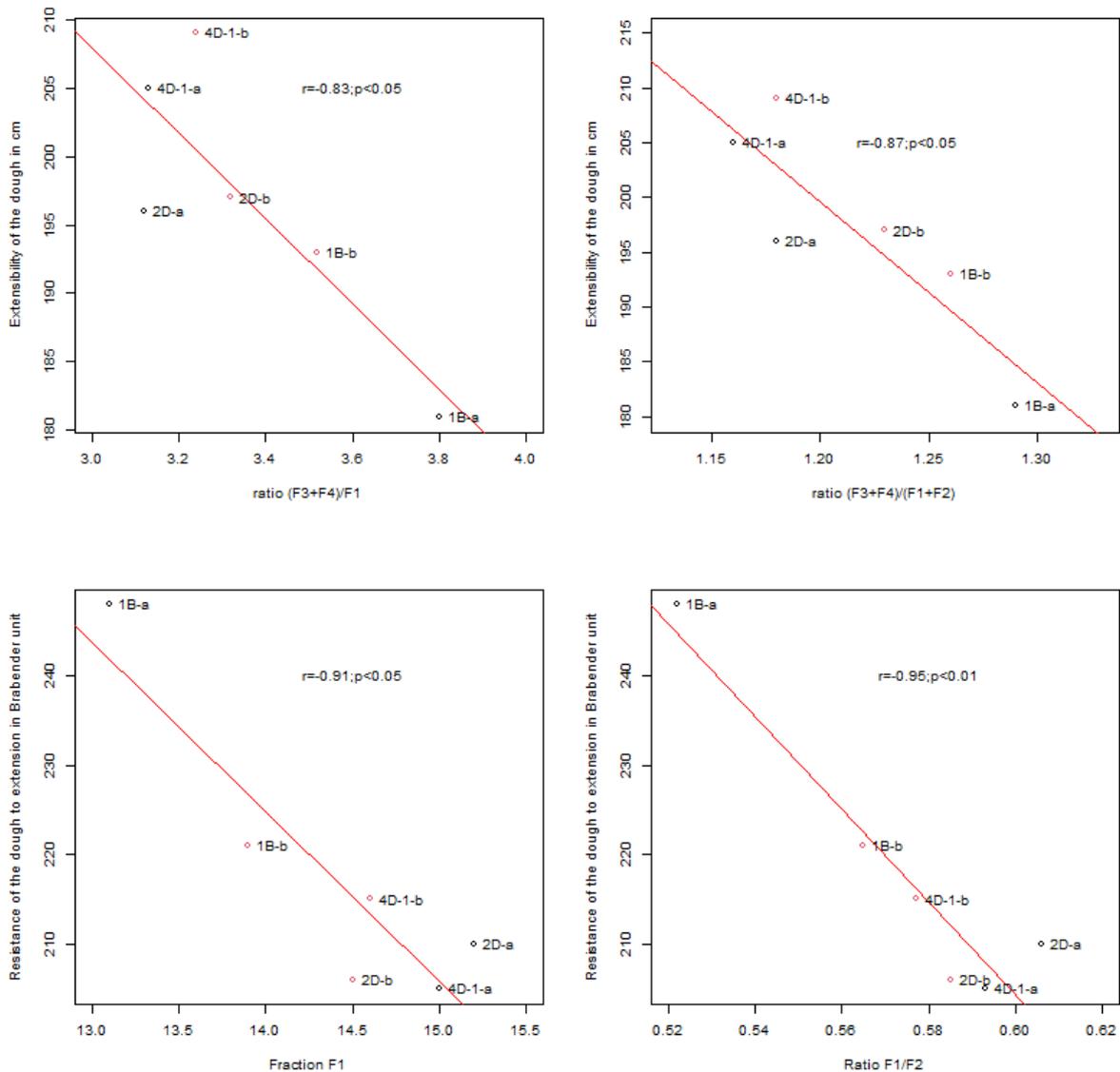


Figure 3.7. Simple linear regression curves showing on the top: the Extensograph extensibility parameter (in cm) against the SE-HPLC ratios F3+F4/F1 (top left) and F3+F4/F1+F2 (top right) and on the bottom: the Extensograph resistance to extension parameter against the fraction F1 (bottom left) and the ratio F1/F2 (bottom right). The fractions F1 and F2 correspond to large and small glutenin polymers and the fractions F3 and F4 correspond to the high and low molecular weight gliadin. The data points are colour-coded in red for the NILs with the Hereward allele at the QTL or black for the NILs with the Malacca allele at the QTL.

strength. Scatter plots were therefore made to look for potential clusters of points or outliers (Figure 3.7).

In Figure 3.7, NIL 1B-a clusters separately from the other NILs. This is particularly visible when considering the Extensograph parameter resistance to extension for which the dough of NIL 1B-a has high value associated with a low number of large glutenin polymers (Figure x, bottom left) or a low ratio of large glutenin polymers on small glutenin polymers (Figure x, bottom right).

Since all the NILs are in the Malacca genetic background, the Malacca allele at the QTL 1B may increase the dough resistance to extension in comparison to the Hereward allele (1B-b). However, the absence of replication of Extensograph and SE-HPLC measurements for each NIL prevent any firm conclusions and the extreme values of NIL 1B-b could also be attributed to an error of measurement either of SE-HPLC or Extensograph. Therefore, more replicates (e.g. three flour replicates for each NILs) would be needed to confirm the negative correlations observed.

### 3.5.7. Limits

The sister lines of each allelic pairs are not entirely genetically identical, sharing 94% of DNA (chapter 2, 2.1.2). Accordingly, the QTL effect might vary among the sister lines because of differences in genetic backgrounds. However, assessing this variation was prevented by the bulking of the sister lines, which was done to equalise the protein content between the allelic pairs and to produce enough flour for rheological and baking tests.

Likewise, differences of SE-HPLC ratios and fractions were observed between the allelic pairs (Table 3.6) but it was not possible to assess their significances because of the absence of replication, even though this was partly compensated by outsourcing the SE-HPLC analysis to Campden BRI ensuring high repeatability of measurements.

Extensograph analysis performed with the flour taken from QTL 4D-2 allelic pairs for Rothamsted 2020-2021 samples did not show differences of strength or extension which therefore indicates similar dough behaviours. Unfortunately, it was not possible to compare the rheology with the Rothamsted 2019-2020 flour samples because no Extensograph measurements were taken owing to time constraints at the bakery driven by the COVID-19 pandemic. The flour of NIL 4D-2 from the 2019-2020 field trial could be analysed by

Extensograph to confirm the absence of difference in the dough rheology between the allelic pairs in a different environment.

## Chapter 4. Mapping of the trait grain protein deviation and its components in the M x H DH population

### 4.1. Introduction

The nitrogen fertilisation requirements to achieve the target GPC of 13% for breadmaking wheat are high, which increases the cost of production (with nitrogen fertiliser being the major input cost) and can have an adverse environmental footprint if the fertiliser is leached into the environment which can occur when heavy rainfall follows nitrogen application or volatilised into the atmosphere in the form of NH<sub>3</sub> (Schreiber and Dowell, 1985; Raun and Johnson, 1999).

Fortunately, genetic variation in the efficiency of nitrogen uptake and remobilisation into the grain has been identified in bread wheat (Salim and Raza, 2020) and can be harnessed to improve nitrogen use efficiency (NUE) (i.e. the product of N uptake efficiency and N utilization efficiency) (Hawkesford, 2014). However, grain protein content is often negatively correlated with GY (Bogard *et al.*, 2010), meaning that increasing grain protein content may prove detrimental for yield and conversely. To offset this negative relationship, Monaghan *et al.* (2001) introduced the concept of GPD, which allows selection for both yield and protein content.

Comparisons of cultivars shows a clear negative relationship between GPC and GY, but some cultivars consistently have higher grain protein contents than would be predicted based on their yields (Figure 4.1.). GPD can be used to develop new varieties with higher grain protein contents without the need for additional fertiliser.

### 4.2. Project aim

Since Monaghan *et al.* (2001) first defined GPD, only few studies have been carried out on wheat (Bogard *et al.*, 2010; Mostleth *et al.*, 2015,2020; Nigro *et al.*, 2019), of which one (Nigro *et al.*, 2019) investigated the genetic architecture of GPD in durum wheat, *Triticum durum*. Therefore, I decided to investigate the genetic basis of GPD in bread wheat, *Triticum aestivum*. For this purpose, a mapping population of 111 DH (DH) lines from a cross between

Malacca (showing consistent negative GPD) and Hereward (showing consistent positive GPD) was used for QTL analysis of the trait. Monaghan *et al.* (2001) grew 8-13 cultivars for two field seasons, and one of them, Hereward, displayed consistent positive GPD.

Accordingly, these authors suggested the use of Hereward to study the genetic architecture of GPD. The expression of positive GPD by Hereward was confirmed by Shewry *et al.* (2013) in a study of six cultivars grown under three nitrogen conditions (low, medium, and high).

The Malacca x Hereward (M x H) DH population was grown in three environments, at Rothamsted 2019-2020 and at Reading in 2020-2021 and 2021-2022 to investigate the genetic control and stability of GPD.

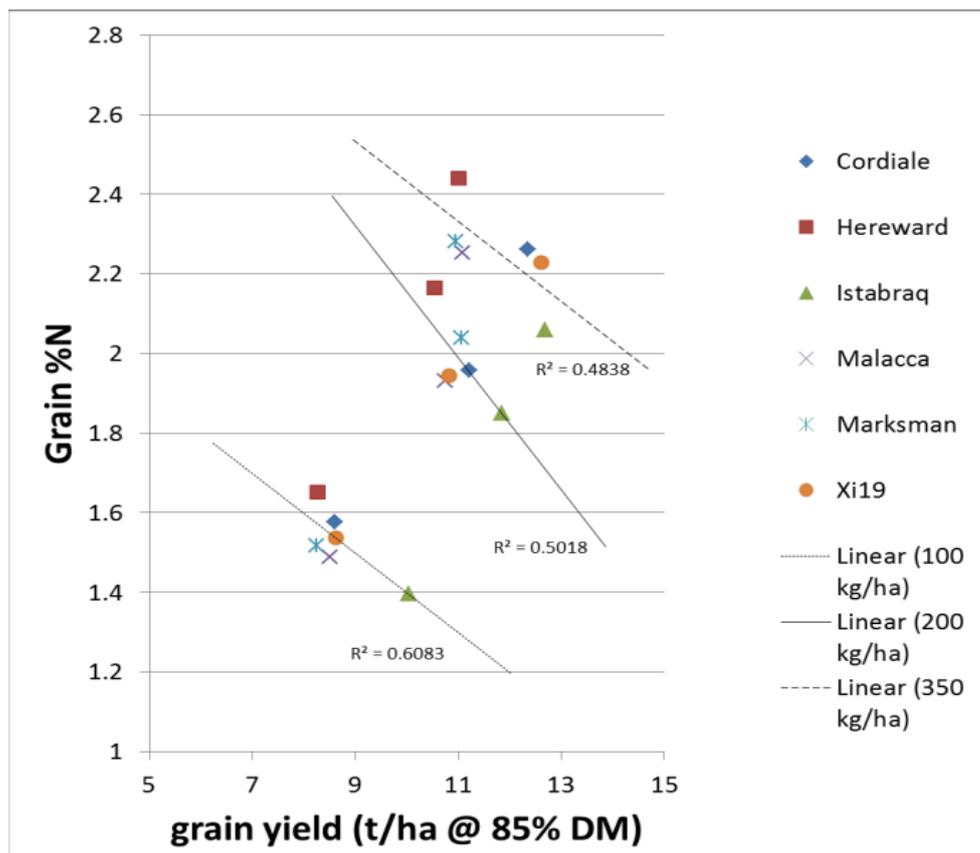


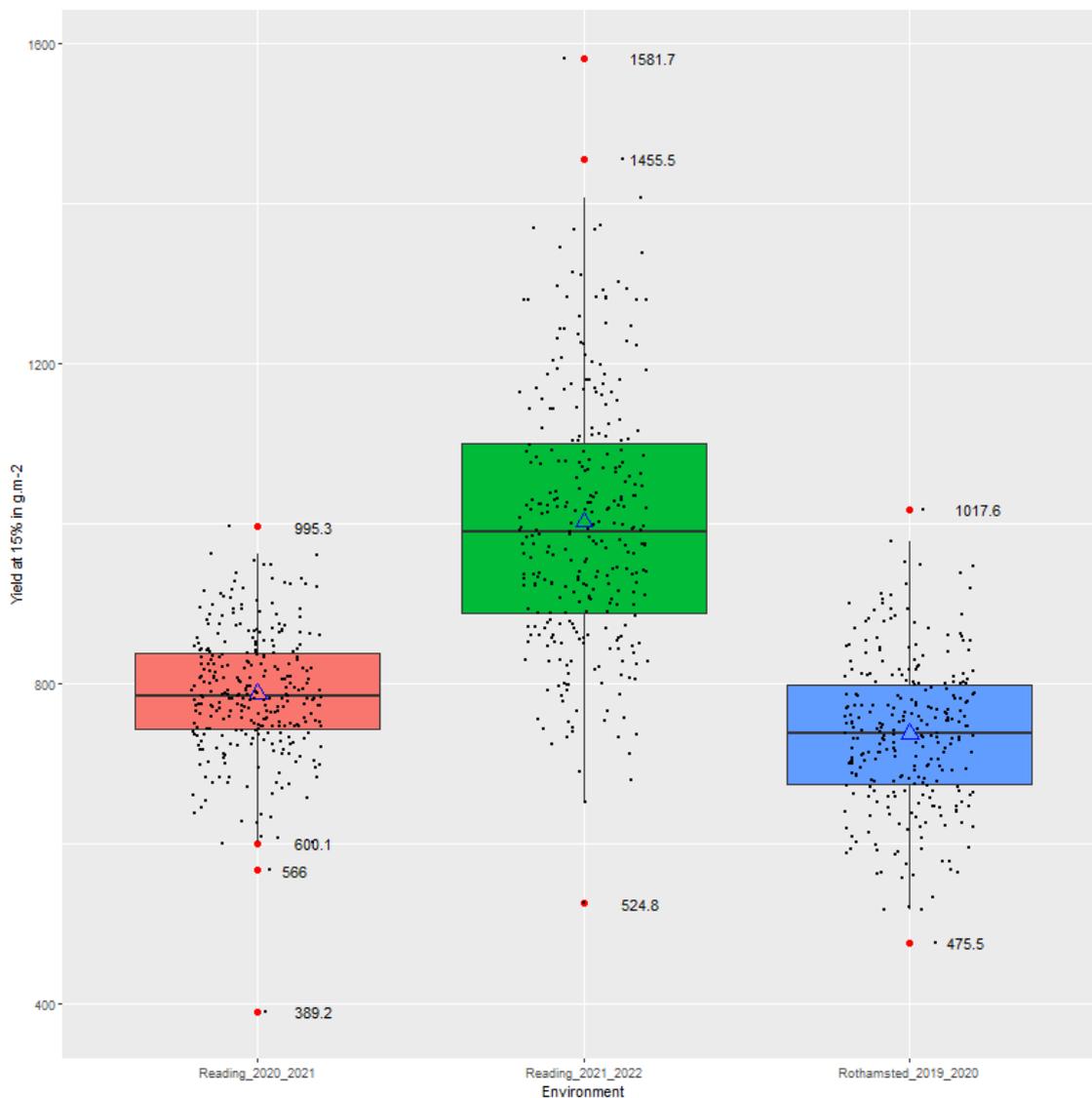
Figure 4.1. Simple linear regressions between the percentage of nitrogen in the grain and the grain yield in tonnes/hectares for six cultivars (Hereward, Cordiale, Istabraq, Marksman, Xi19, and Malacca) grown in 2009 at Rothamsted in three trials. The trials differed in nitrogen fertilisation (100, 200, and 350 kg N/ha).

Reference: Shewry *et al.*, 2013

### 4.3. Results

#### 4.3.1. Description of the data

##### 4.3.1.1. Variation of GY in the three field trials



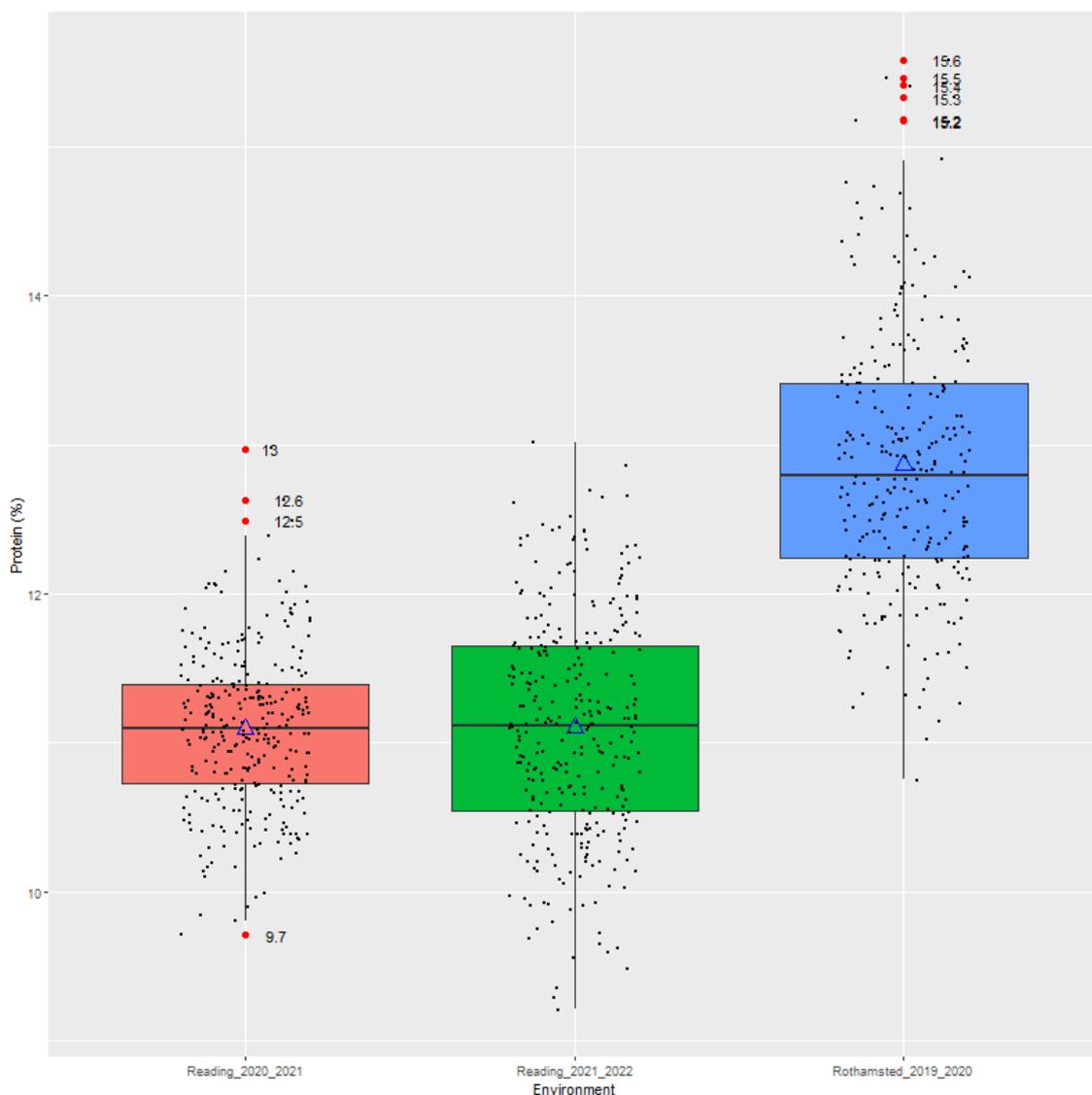
**Figure 4.2.** Boxplots of GY for the DH lines (n=300) in the three field trials: Reading 2020-2021 (red), Reading 2021-2022 (green) and Rothamsted 2019-2020 (blue). The mean of each boxplot is marked by a blue triangle and extreme values by a red dot.

The DH population grown at Rothamsted 2019-2020 and Reading 2020-2021 had comparable mean GY of 735 g.m<sup>-2</sup> and 785 g.m<sup>-2</sup> and similar variances of 3817 and 3638, respectively. In contrast, the DH lines grown in Reading 2021-2022 had a higher mean GY of 998 g.m<sup>-2</sup> and the GY varied more; variance of 815 (Figure 4.2). For each environment, some extreme values (i.e., values located above or below the maximum or minimum whiskers on the boxplot) were

identified. In Reading 2020-2021 M x H 85 in block 3 had a much lower yield ( $389.2 \text{ g. m}^{-2}$ ) than the other lines. Its GY was also considerably lower than those of the other two field replicates: M x H 85 in block 2 and line 85 in block 1. This plot was therefore excluded from the analysis, but all the other extreme values were kept.

#### 4.3.1.2. Variation of grain protein content in the three field trials

The DH lines grown in Rothamsted 2019-2020 had higher average GPC (12.86%) than those grown at Reading 2020-2021 and 2021-2022 (11.08% and 11.11%, respectively) (Figure 4.3.).

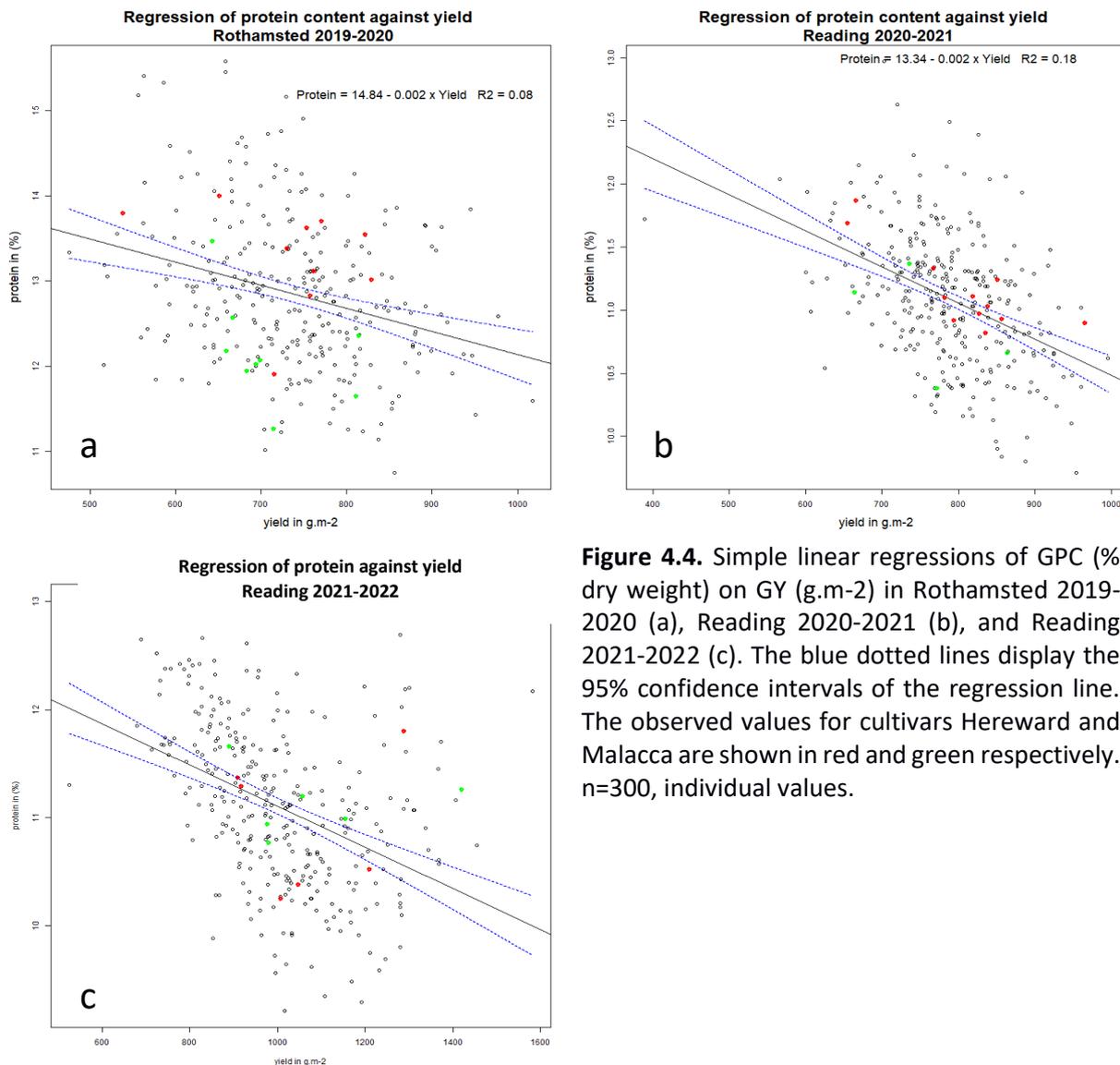


**Figure 4.3.** Boxplots of GPC (in %) for the DH population (n=300) in three field trials: Reading 2020-2021(red), Reading 2021-2022 (green) and Rothamsted 2019-2020 (blue). The mean of each boxplot is marked by a blue triangle and extreme values by a red dot.

The GPC variance was smaller at Reading 2020-2021 (0.19) than at Reading 2021-2022 (0.29) and Rothamsted 2019-2020 (0.60). Extreme values corresponding to high GPC (15-16%) were identified at Rothamsted 2019-2020 (Figure 4.3.) but not removed as they sat in the higher range of yield recorded in bread wheat (Shewry, 2009)

### 4.3.2 Calculation of GPD

#### 4.3.2.1 Simple linear regression between GPC and GY



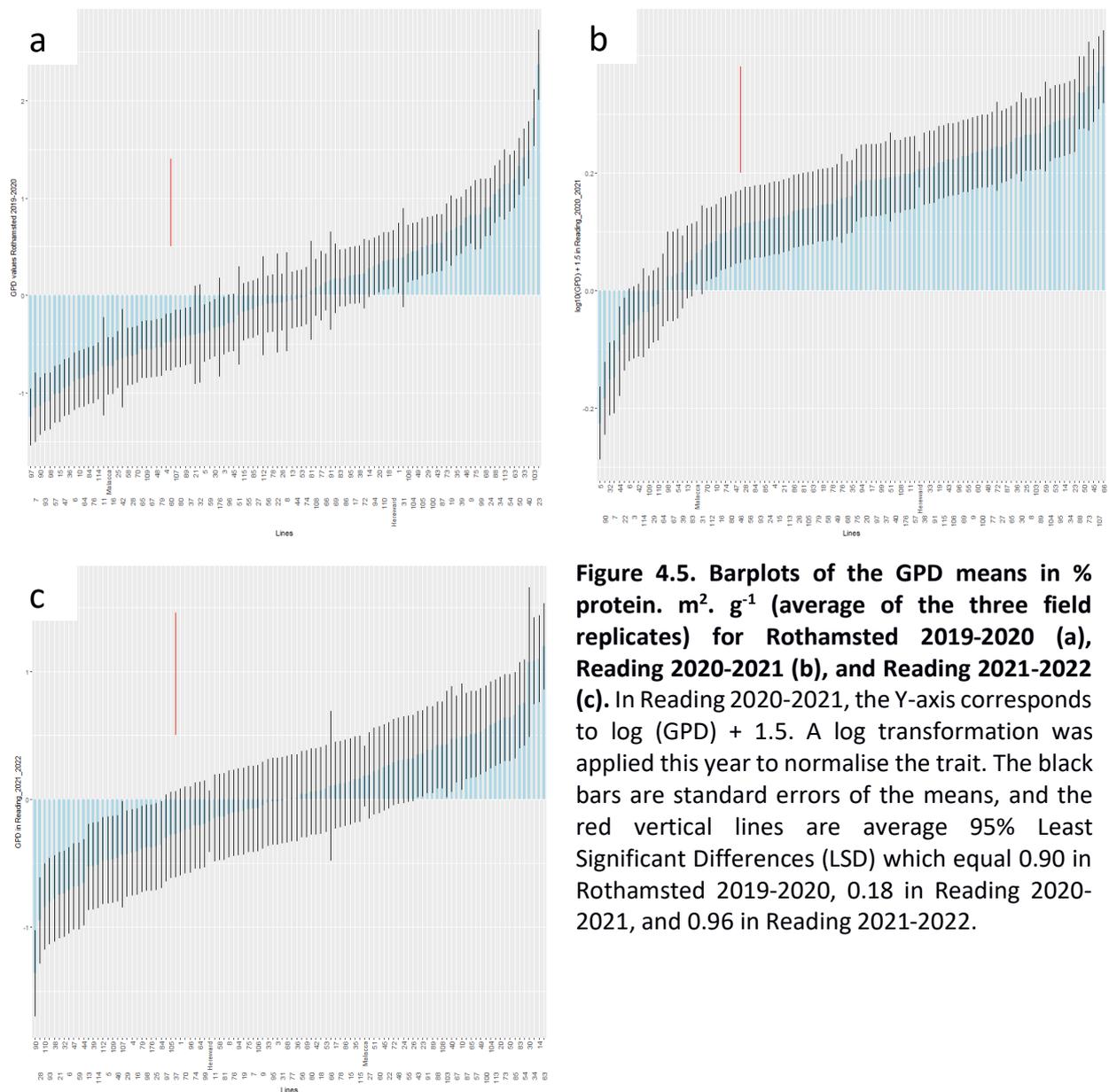
**Figure 4.4.** Simple linear regressions of GPC (% dry weight) on GY (g.m-2) in Rothamsted 2019-2020 (a), Reading 2020-2021 (b), and Reading 2021-2022 (c). The blue dotted lines display the 95% confidence intervals of the regression line. The observed values for cultivars Hereward and Malacca are shown in red and green respectively. n=300, individual values.

The regression diagnostic plot (standardised residuals vs leverage, not shown) was used to identify observations with a high impact on the regression coefficient using the Cook

distance. In the three regression models, no data points fell outside of the Cookdistance area and had therefore no significant impact on the slope of the regression. This means that even if outliers were still present after the first detection step (Chapter 4, 4.3.1), keeping them in the analysis won't influence the regression slope.

The simple linear regression between GPC and GY was significant ( $p < 0.05$ ) and slightly negative in the three field trials (Figure 4.4.). The slope coefficient ( $\beta$ ) was equal to  $-0.002$  in the three field trials; this corresponds to a 0.2% GPC decrease for every  $100 \text{ g.m}^{-2}$  increase in yield. The variation in GY explained more of the variation in GPC in Reading ( $R^2=0.18$ ) than in Rothamsted ( $R^2=0.08$ ). In Rothamsted 2019-2020 and Reading 2020-2021 (Figure 4.4., a and b), most of the observations (field plots) for Hereward were located above (positive GPD) the regression line and for Malacca below it (negative GPD). However, in Reading 2021-2022 there was no such clear separation between the individual observations (plots) of Malacca and Hereward (Figure 4.4., c). For all three environments, positive and high GPD values were more frequent at low yield than at high yield. Nevertheless, positive values at high yield were detected and the lines displaying them may prove useful for breeding. Figure 4.4 indicates possible transgressive segregation in the three environments as many individual data points are located either above Hereward (positive transgressive segregation) or below Malacca (negative transgressive segregation). The significance of the differences between the observations and the parents will need to be assessed separately for each lines by two-sample t-tests.

#### 4.3.2.2. Variation of GPD means within the DH population and between the three field trials



**Figure 4.5. Barplots of the GPD means in % protein. m<sup>2</sup>. g<sup>-1</sup> (average of the three field replicates) for Rothamsted 2019-2020 (a), Reading 2020-2021 (b), and Reading 2021-2022 (c). In Reading 2020-2021, the Y-axis corresponds to log (GPD) + 1.5. A log transformation was applied this year to normalise the trait. The black bars are standard errors of the means, and the red vertical lines are average 95% Least Significant Differences (LSD) which equal 0.90 in Rothamsted 2019-2020, 0.18 in Reading 2020-2021, and 0.96 in Reading 2021-2022.**

The variation in GPD values of the individual DH lines within the three environments are displayed in figure 4.5 and table 4.1.

In Rothamsted 2019-2020, the GPD ranged from -1.25%protein. m<sup>2</sup>. g<sup>-1</sup> (+/-0.29% protein. m<sup>2</sup>. g<sup>-1</sup>) for M x H 97 to 2.36 % protein. m<sup>2</sup>. g<sup>-1</sup> (+/-0.35%protein. m<sup>2</sup>. g<sup>-1</sup>) for M x H 23 (Table 4.1.). The median GPD was -0.07 % protein. m<sup>2</sup>. g<sup>-1</sup> which means that about the same number of lines display positive and negative GPD. Malacca exhibited negative GPD of -0.72 %protein.m<sup>2</sup>. g<sup>-1</sup> (+/- 0.29% protein. m<sup>2</sup>. g<sup>-1</sup>) whereas Hereward exhibited positive GPD of 0.37 %protein.m<sup>2</sup>. g<sup>-1</sup> (+/-0.29% protein. m<sup>2</sup>. g<sup>-1</sup>). However, their difference in GPD (0.51% protein.

$\text{m}^2 \cdot \text{g}^{-1}$ ) was smaller than the average LSD (0.90) and were therefore not statistically significant. The LSD calculation (Figure 4.5, a, red vertical line) suggests statistically significant differences of GPD within the DH population, especially between lines with negative and positive GPD values. This is further evidence by the high significance ( $p < 0.001$ ) of the genetic term of the mixed model design to calculate the GPD means (Chapter 2, 2.14).

In Reading 2020-2021, the GPD values (on the back transformed scale) ranged from -0.90% protein. $\text{m}^2 \cdot \text{g}^{-1}$  with a 95% confidence interval (CI) of [-1.05, -0.71] for M x H 5 and +0.89% protein.  $\text{m}^2 \cdot \text{g}^{-1}$  with a 95% CI [0.31, 1.67] for M x H 66 (Table 4.1). The median (-0.05) was also close to zero indicating a good balance between positive and negative GPD (Table 4.1). Malacca had negative GPD of -0.34% protein. $\text{m}^2 \cdot \text{g}^{-1}$  with a 95% CI [-0.59, -0.02] whereas Hereward had positive GPD of 0.10% protein.  $\text{m}^2 \cdot \text{g}^{-1}$  with a 95% CI [-0.10, 0.34]. However, the difference between the parents was not statistically significant, which can be deduced from their overlapping confidence intervals. The LSD calculation (Figure 4.5, b) and the genetic term of the mixed model again suggest some significant differences in GPD means between the lines.

In Reading 2021-2022, the GPD values ranged from -1.36 % protein.  $\text{m}^2 \cdot \text{g}^{-1}$  for M x H 90 (+/- 0.33% protein.  $\text{m}^2 \cdot \text{g}^{-1}$ ) and 1.19% protein.  $\text{m}^2 \cdot \text{g}^{-1}$  (+/- 0.33% protein.  $\text{m}^2 \cdot \text{g}^{-1}$ ) for M x H 63 (Table 4.1). Hereward had negative GPD of -0.17 % protein.  $\text{m}^2 \cdot \text{g}^{-1}$  (+/- 0.23% protein.  $\text{m}^2 \cdot \text{g}^{-1}$ ) and Malacca positive GPD of 0.18% protein. $\text{m}^2 \cdot \text{g}^{-1}$  (+/- 0.23% protein.  $\text{m}^2 \cdot \text{g}^{-1}$ ). However, the difference of GPD between the parents (-0.11% protein.  $\text{m}^2 \cdot \text{g}^{-1}$ ) was not statistically significant it was smaller than the average LSD (0.96) (Figure 4.5, c). The genetic effect of the mixed model was also highly significant implying at least one difference in the GPD means between the lines.

**Table 4.1.** Summary statistics for the traits GPD measured in the Rothamsted 2019-2020 and Reading 2020-2021, and 2021-2022 trials.

<b>Environments</b>	<b>Range</b>	<b>Mean</b>	<b>Median</b>	<b>Variance</b>
Rothamsted 2019-2020	-1.25; 2.36	0.004	-0.07	0.5
Reading 2020-2021	-1.05; 0.31	-0.03	-0.05	0.14
Reading 2021-2022	-1.36; 1.19	0.01	0.005	0.22

#### 4.3.3. Correlations of yield components and GPC with GPD in the three field trials.

GPD was highly positively and significantly correlated with GPC in the three trials. The correlation coefficient (r) ranged from 0.91 in Reading 2020-2021 to 0.97 in Rothamsted 2019-2020 (Table 4.2). By contrast, the GPD was weakly negatively correlated with the GY in the three trials with r ranging from -0.16 in Reading 2020-2021 to -0.26 in Rothamsted 2019-2020. However, the correlation was not significant in Reading 2020-2021 (p=0.11) but was nearly significant in Reading 2021-2022 (p=0.052). The specific weight (SW) was not significantly correlated with GPD in the three environments (p>0.05) but the thousand kernel weight (TKW) was weakly negatively and significantly correlated with GPD (p=0.01).

**Table 4.2.** Correlations between GPD and GPC, GY (in g.m<sup>-2</sup>), SW and TKW in the three field trials: Rothamsted 2019-2020, Reading 2020-2021 and 2021-2022. The r column corresponds to the correlation coefficient r and the p column gives the p-values of the correlations. Significant correlations (p<0.05) are highlighted in red.

Trait1	Trait2	r	p
GPD_2019_2020_Rothamsted	Protein_2019_2020_Rothamsted	0.974361	0
GPD_2020_2021_Reading	Protein_2020_2021_Reading	0.918336	0
GPD_2021_2022_Reading	Protein_2021_2022_Reading	0.949636	0
GPD_2019_2020_Rothamsted	Yield_2019_2020_Rothamsted	-0.26343	0.011176
GPD_2020_2021_Reading	Yield_2020_2021_Reading	-0.16509	0.115794
GPD_2021_2022_Reading	Yield_2021_2022_Reading	-0.20279	0.052534
GPD_2019_2020_Rothamsted	SW_2019_2020_Rothamsted	-0.07891	0.454637
GPD_2020_2021_Reading	SW_2020_2021_Reading	-0.16683	0.111947
GPD_2021_2022_Reading	SW_2021_2022_Reading	0.150011	0.153502
GPD_2021_2022_Reading	TKW_2021_2022_Reading	-0.24725	0.017497

#### 4.3.4 Analysis of effects of environment using an Additive Main Effect and Multiplicative Interaction (AMMI) model

The phenotypic variance varied between the three environments for all the traits (GPD, protein and yield) (Table 4.3, a). For example, it ranged between 0.14 (Reading 2020-2021) to 0.51 (Rothamsted 2019-2020) for the GPD. The phenotypic correlations between environments were positive and statistically significant but low for all the traits (Table 4.3, b). The weak phenotypic correlations across environments suggest the presence of G x E interactions. This was addressed in the data analysis by choosing a variance-covariance model that allows for heterogeneity of variances and covariances between the environments to model the G x E interactions.

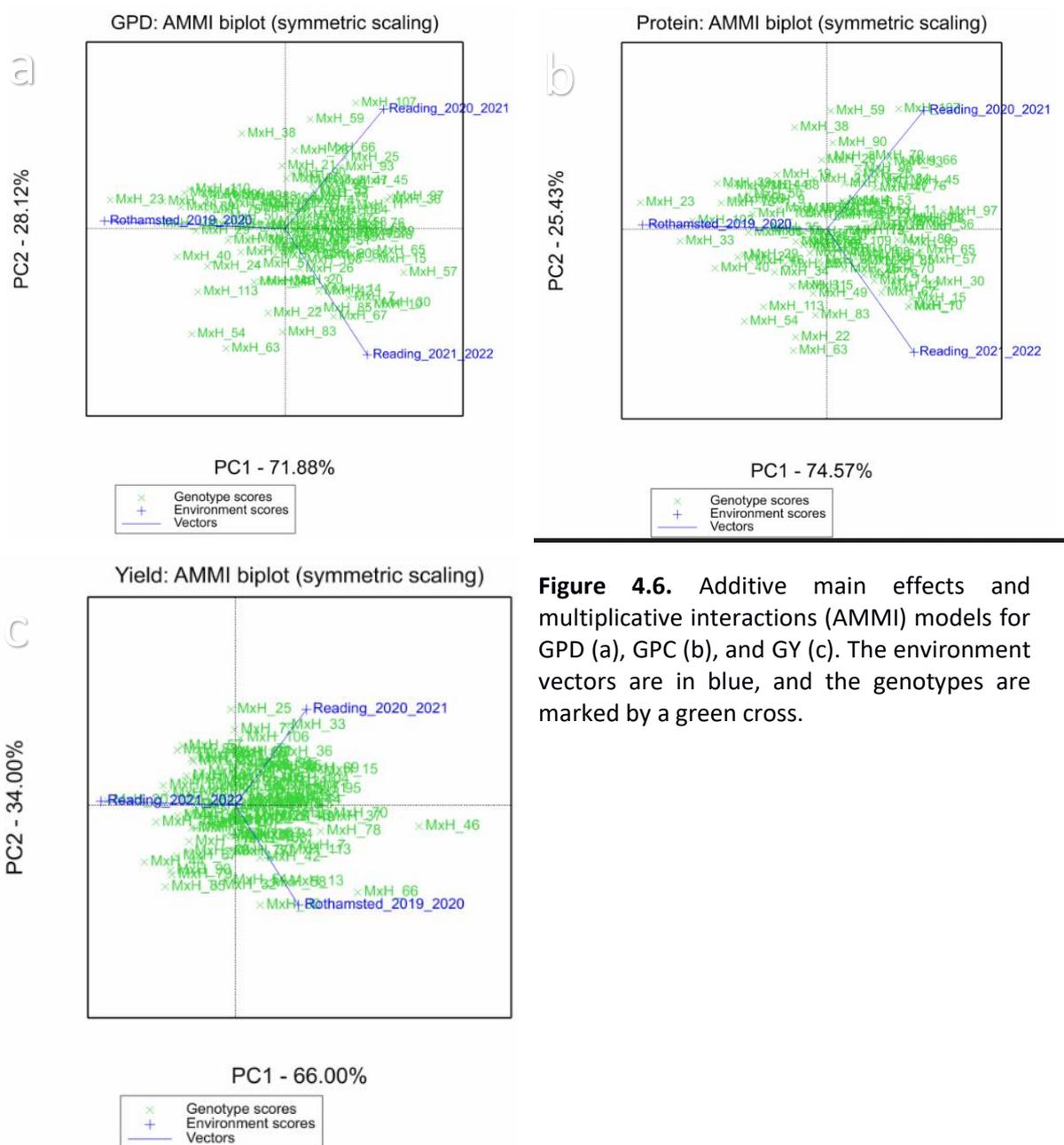
**Table 4.3.** Environmental means and variances of the DH population for the traits GPD, GPC and GY (a) and correlation of the trait between the three field trials (b).

<b>a</b>	<b>Environments</b>	<b>Number of genotypes</b>	<b>GPD mean</b>	<b>GPD variance</b>	<b>Protein mean</b>	<b>Protein variance</b>	<b>Yield Mean</b>	<b>Yield variance</b>
	Reading_2020_2021	102	-0.033	0.1462	11.09	0.2	785.6	3639
	Reading_2021_2022	102	0.0169	0.2336	11.13	0.3	998.8	8160
	Rothamsted_2019_2020	102	0.0124	0.5161	12.87	0.61	735.7	3817
	Overall mean		-0.001	0.2972	11.7	1.06	840	18229

<b>b</b>	<b>Trait 1</b>	<b>Trait2</b>	<b>r</b>	<b>p</b>
	GPD_Rothamsted_2019_2020	GPD_Reading_2020_2021	0.357954367	0.000459584
	GPD_Rothamsted_2019_2020	GPD_Reading_2021_2022	0.431000569	1.80E-05
	GPD_Reading_2020_2021	GPD_Reading_2021_2022	0.450867416	6.48E-06
	Protein_Rothamsted_2019_2020	Protein_Reading_2020_2021	0.456460637	4.80E-06
	Protein_Rothamsted_2019_2020	Protein_Reading_2021_2022	0.494126081	5.57E-07
	Protein_Reading_2020_2021	Protein_Reading_2021_2022	0.602620912	2.08E-10
	Yield_Rothamsted_2019_2020	Yield_Reading_2020_2021	0.399370369	8.02E-05
	Yield_Rothamsted_2019_2020	Yield_Reading_2021_2022	0.375988166	0.000221156
	Yield_Reading_2020_2021	Yield_Reading_2021_2022	0.355006278	0.000515882

Breeders are interested in genotypes that perform well for a trait (i.e., have a high genotypic value) in a wide range of environments. In this respect, Figure 4.6 shows an AMMI biplot of the G x E interactions for the three traits (GPD, GPC, and GY) where broadly adapted genotypes cluster at the origin of the biplot and environment-specific genotypes group near the blue cross of their environment vector. For all traits, many broadly adapted genotypes were identified at the origin of the biplot. This indicates that their genotypic value is stable across environment. Environment-specific genotypes were also identified, such as M x H 83 for GPD that is quite well adapted to the environment in Reading 2021-2022 (i.e., show good performance in that environment) but badly suited to the Rothamsted 2019-2020 and the Reading in 2020-2021 trials (Figure 4.6, a).



**Figure 4.6.** Additive main effects and multiplicative interactions (AMMI) models for GPD (a), GPC (b), and GY (c). The environment vectors are in blue, and the genotypes are marked by a green cross.

Table 4.4. gives the genotype AMMI stability values (ASV) derived from the AMMI biplot which were assessed simultaneously with the genotypic values for GPD to identify cultivars showing high GPD in the three environments.

Genotypes with low ASV values are more stable than genotypes with large ASV values. Table 4.4. shows that stability is not always paired with performance. Thus, M x H 109 is the most stable (rank 1) but also has a negative GPD genotypic value of  $-0.53 \text{ \% protein. m}^2. \text{ g}^{-1}$  and would therefore not be suitable for selection.

M x H 73 would be a good candidate for selection as it very stable in the three environments (stability rank 3/102) while having a high GPD average value across the environments (+0.66% protein. m<sup>2</sup>. g<sup>-1</sup>). In single environment, its GPD value was 0.63% protein. m<sup>2</sup>. g<sup>-1</sup> and 0.65% protein. m<sup>2</sup>. g<sup>-1</sup> in Reading 2021-2022 and Rothamsted 2019-2020, respectively, and 0.72 % protein. m<sup>2</sup>. g<sup>-1</sup> in Reading 2020-2021.

Table 4.4. Genotypes ASV and mean across the three environments for the trait GPD. n=102.

Genotype	ASV rank	GPD Genotypic value (n=3)	Genotype	ASV rank	GPD Genotypic value (n=3)	Genotype	ASV rank	GPD Genotypic value (n=3)
MxH_23	102	1.0743	MxH_83	50	0.1815	MxH_74	15	-0.151
MxH_34	47	0.8867	MxH_106	33	0.1701	MxH_38	63	-0.1537
MxH_50	68	0.8848	MxH_53	40	0.165	MxH_29	89	-0.1587
MxH_103	101	0.8673	MxH_69	5	0.1293	MxH_67	66	-0.1716
MxH_63	79	0.7434	MxH_77	9	0.1245	MxH_36	95	-0.2021
MxH_88	52	0.712	MxH_17	8	0.123	MxH_80	44	-0.2066
MxH_73	3	0.6692	MxH_55	48	0.1177	MxH_37	41	-0.2095
MxH_40	98	0.6654	MxH_18	43	0.11	MxH_10	88	-0.2103
MxH_14	58	0.6137	MxH_85	54	0.1074	MxH_59	59	-0.2107
MxH_113	90	0.5404	MxH_27	45	0.1019	MxH_25	72	-0.2207
MxH_33	100	0.509	MxH_89	80	0.095	MxH_11	74	-0.2662
MxH_54	94	0.496	MxH_65	86	0.0805	MxH_58	55	-0.2888
MxH_104	2	0.4818	MxH_94	29	0.0783	MxH_112	25	-0.294
MxH_87	12	0.4309	MxH_86	14	0.0587	MxH_3	19	-0.3139
MxH_100	13	0.4262	MxH_8	28	0.0542	MxH_76	71	-0.3229
MxH_66	56	0.3865	MxH_115	35	0.0459	MxH_110	91	-0.3335
MxH_24	82	0.3529	MxH_46	93	0.0442	MxH_13	32	-0.3347
MxH_30	92	0.3527	MxH_26	20	0.0408	MxH_15	87	-0.3376
MxH_43	24	0.3478	MxH_105	70	0.039	MxH_79	34	-0.3444
MxH_9	75	0.3378	MxH_51	39	0.0254	MxH_4	27	-0.3671
MxH_49	36	0.3139	MxH_31	53	-0.0011	MxH_70	42	-0.3888
MxH_20	18	0.3109	MxH_48	78	-0.0034	MxH_42	51	-0.4015
MxH_68	84	0.2827	MxH_60	73	-0.0145	MxH_21	26	-0.4285
MxH_35	64	0.2643	MxH_78	7	-0.0165	MxH_84	60	-0.4457
MxH_75	81	0.2629	MxH_107	85	-0.0297	MxH_16	49	-0.4635
MxH_19	65	0.2485	MxH_56	4	-0.0763	MxH_44	62	-0.4652
MxH_72	6	0.2478	MxH_81	11	-0.0797	MxH_97	97	-0.5083
MxH_45	77	0.2477	MxH_39	96	-0.0801	MxH_64	57	-0.5247
MxH_99	83	0.2368	MxH_96	37	-0.1168	MxH_109	1	-0.5321
MxH_91	10	0.2276	MxH_57	99	-0.1392	MxH_5	30	-0.5879
MxH_95	16	0.2221	MxH_22	46	-0.1394	MxH_28	38	-0.5904
MxH_108	17	0.1857	MxH_90	22	-1.1132	MxH_47	61	-0.6188
MxH_93	67	-0.6804	MxH_32	31	-0.6435	MxH_114	21	-0.6259
MxH_6	23	-0.7429	MxH_7	76	-0.6699	MxH_98	69	-0.6312

#### 4.3.5. QTL analysis

##### 4.3.5.1 Single-environment QTL analysis

QTL detection was performed for the traits GPD, GPC, GY, SW, and heading date for the three environments except for the TKW which was only recorded in Reading 2021-2022 and the heading date which was only scored in Reading. The results are displayed in Table 4.5.

The final QTL model identified seven GPD QTL in the three environments. Most of these QTL (2B, 3B, 5B, and 6B) were on the B genome but two QTL were on chromosomes 3A and 5D. In six of the seven QTL, Hereward provided the favourable allele, the allele increasing the phenotype, but for the 2B QTL mapped in the Reading 2020-2021 environment the favourable allele came from Malacca. The significance of the QTL is indicated by the logarithm of the Odds (LOD) values and a value of 3 (i.e., the likelihood of the presence of the QTL is 1000 times higher than its absence) is generally accepted as a high confidence threshold. Six of seven QTL (2B, 3A, 3B, 5B, 5D) had high confidence scores (LOD score above 3) while one (QTL 6B) was just below the LOD 3 threshold (LOD score of 2.9). The percentage of GPD variance explained by the QTL ranged from 5.9% (QTL 6B, Reading 2020-2021 trial) to 18.7% (QTL 5B, Rothamsted 2019-2020 trial) and the substitution effect, which is the additive effect of replacing one parental allele by the other, varied between +0.034% protein. m<sup>2</sup>. g<sup>-1</sup> (QTL 6B, Reading 2020-2021) and +0.554% protein. m<sup>2</sup>. g<sup>-1</sup> (QTL 5B, Rothamsted 2019-2020 trial) (Table 4.5.)

Eight QTL were detected for GPC in the three environments. Most of these QTL (3A, 3B, 5B, 5D) were on the same chromosomes as the GPD QTL, but two were detected on chromosomes 1A and 7A. Hereward provided seven favourable alleles whereas Malacca provided only one (QTL 1A, Reading 2021-2022). All but one QTL (3B, Rothamsted 2019-2020 trial: LOD = 1.5) were strong (LOD > 3) and the percentage of GPC variance ranged from 4.9% (3B Rothamsted 2019-2020 trial) to 17.7% (3A, Reading 2020-2021). The substitution effects varied between 0.22% protein (5D, Reading 2020-2021) and 0.57% protein (5B, Rothamsted 2019-2020 trial) (Table 4.5.)

13 QTL were detected for the three yield and yield related traits: GY, TKW, and SW in the three environments. Unlike GPD and GPC, the favourable alleles were balanced between Malacca (6) and Hereward (7). These QTL were located over a range of linkage groups (2A, 2B, 3A, 3B, 3D, 4A, 5A, 5B, 6B, 7D) and were all strong (LOD > 3). The percentage of variances

in the yield related traits ranged from 7.3% (2A, Reading 2020-2021, specific weight) to 20.8% (6B, Reading 2020-2021, specific weight) (Table 4.5.).

Three QTL were identified for HD on chromosome 1A, 5A, and 7B in both Reading environments. All were strong QTL (LOD>3) with moderate additive effects (0.4 days later flowering for the 5A and 7B Hereward alleles and 0.37 days later flowering for the Malacca 1A allele).

**Table 4.5.** Positions of the QTL mapped on the Malacca x Hereward genetic map. The positions and the confidence intervals (CI) on the M x H genetic map are given in centimorgans (cM). SW, specific weight in kg/hL; TKW, thousand kernel weight in g; GPD, grain protein deviation in %protein.m<sup>2</sup>. g<sup>-1</sup>; Heading, heading date (GS 55) in number of days elapsed since 1<sup>st</sup> January; Protein, grain protein content in %protein; Yield, grain yield in g.m<sup>-2</sup>.

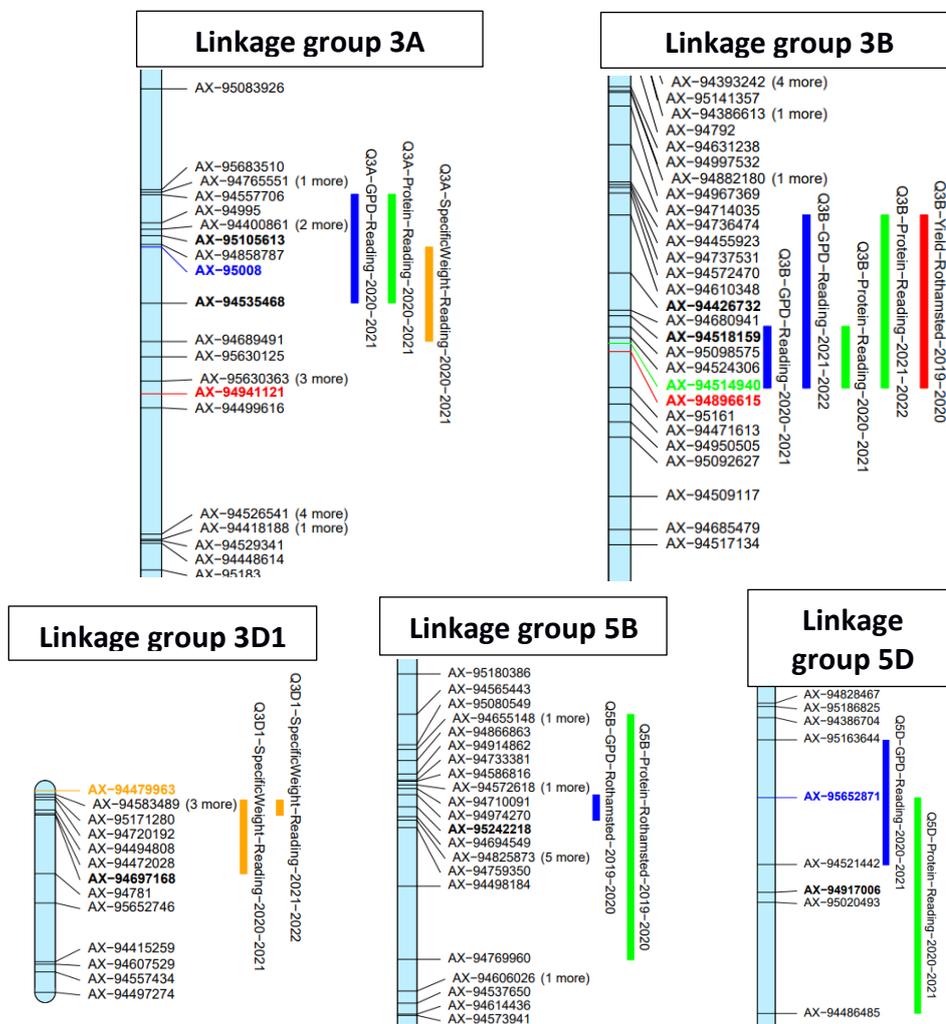
The column allele informs on the favourable allele provenance; A: Malacca or B: Hereward.

Chr	Position	LOD	%variance	additive effect	Location	Year	Allele	Trait	CI Start	CI end
5B	89	4.641	18.739	0.277	Rothamsted	2019-2020	B	GPD	87.52	89.6
2B	122	4.5	9.3	-0.131	Reading	2020-2021	A	GPD	119.31	136.47
3A	132	8.3	18.6	0.172	Reading	2020-2021	B	GPD	128.46	136.97
3B	157	3.2	6.5	0.121	Reading	2020-2021	B	GPD	155.27	159.74
5D	33	6.3	13.4	0.096	Reading	2020-2021	B	GPD	25.23	36.44
6B	122	2.9	5.9	0.017	Reading	2020-2021	B	GPD	118.81	123.83
3B	154.46	4.118	16.815	0.193	Reading	2021-2022	B	GPD	147.2	159.74
5A	47.29	4.6	15.8	0.446	Reading	2020-2021	B	Heading	32.05	50.74
7B	28	4.1	14	0.415	Reading	2020-2021	B	Heading	24.83	33.21
1A	55.57	3.437	14.245	-0.379	Reading	2021-2022	A	Heading	46.2	61.46
3B	68	1.5	4.9	0.24	Rothamsted	2019-2020	B	Protein	46.76	68.48
5B	88.49	3.1	10.5	0.286	Rothamsted	2019-2020	B	Protein	80.92	101.04
7A	66	3.5	11.8	0.229	Rothamsted	2019-2020	B	Protein	55.61	77.61
3A	132	5.8	17.7	0.171	Reading	2020-2021	B	Protein	128.46	136.97
3B	156.53	5.3	15.8	0.169	Reading	2020-2021	B	Protein	155.27	159.74
5D	38.92	3.7	10.8	0.108	Reading	2020-2021	B	Protein	30.41	49.78
1A	28	3.5	11.9	-0.201	Reading	2021-2022	A	Protein	22.73	29.55
3B	157.1	4.2	14.4	0.237	Reading	2021-2022	B	Protein	147.2	159.74
2A	3	4.2	7.3	-0.095	Reading	2020-2021	A	SW	0	5.98
3A	137	9.8	19.6	-0.094	Reading	2020-2021	A	SW	132.6	139.98
3D1	1.93	4.9	8.7	-0.793	Reading	2020-2021	A	SW	0.73	6.51
5A	13	4.3	7.4	0.359	Reading	2020-2021	B	SW	3.36	15.84
6B	140	10.3	20.8	0.802	Reading	2020-2021	B	SW	131.26	152.45
7D1	17	4.6	8.1	0.743	Reading	2020-2021	B	SW	5.01	34.77
2A	40	5.5	14.2	0.884	Reading	2021-2022	B	SW	35.77	42.28
3D1	1.93	5.8	15	-0.666	Reading	2021-2022	A	SW	0.73	1.93
4A	84.09	4.3	10.8	-0.124	Reading	2021-2022	A	SW	75.51	88.17
5B	31.23	4.9	12.5	0.808	Reading	2021-2022	B	SW	28.89	40.85
2A	114.24	5.109	20.42	1.692	Reading	2021-2022	B	TKW	107.94	117.32
3B	151.39	4.742	19.104	-26.942	Rothamsted	2019-2020	A	Yield	147.2	159.74
2B	174.68	3.546	14.662	36.42	Reading	2021-2022	B	Yield	167.41	182.91

Five co-locations of QTL intervals were observed for different years and traits. These are shown in Figure 4.7.

Firstly, overlapping QTL confidence intervals of the same trait and same effect direction for at least two years were detected for GPC and GPD (Figure 4.7, linkage group 3B), and SW (linkage group 3D1).

Secondly, pleiotropic QTL (i.e., between traits co-locations) were observed on linkage groups 3A, 3B, 5B, and 5D. Most of the co-locations were observed between GPD and GPC (linkage group 3A, 3B, 5B, and 5D) but additional co-locations with yield components, namely SW and GY were present on linkage groups 3A and 3B.



**Figure 4.7.** Co-locations of single-environment QTL on the Malacca x Hereward genetic map. The confidence intervals of the QTL are represented by coloured rectangles; blue (GPD), green (protein content), red (yield in g.m<sup>-2</sup>), orange (specific weight) and peak markers of the QTL intervals are highlighted in bold. Centimorgan distance is shown on the left of each map.

## 4.4. Conclusion and Discussion

### 4.4.1. Simple Linear regressions GPC on GY showed low $R^2$

A DH population comprising 111 genotypes was grown in three environments (Rothamsted 2019-2020 and Reading 2020-2021 and 2021-2022). The lines were phenotyped for GPC, three yield related traits (GY, SW, and TKW), and GPD calculated using simple linear regression between GPC and GY. The simple linear regression models showed significant negative ( $p < 0.05$ ) relationships in the three environments, but the  $R^2$  were low (0.08 – 0.18) (Figure 4.4).

The low  $R^2$  values are unlikely to be caused by incomplete removal of GY and/or GPC outliers as this was minimised by a two-step outliers' detection on the raw data (Chapter 4, 4.3.1) and on the regression model (Chapter 4, 4.3.2.1).

Moreover, the  $R^2$  values observed are consistent with the literature (Oury *et al.*, 2003; Oury and Godin, 2007) and fall in the higher range of values reported. The variability of GY and GPC in the DH population is demonstrated by a wide scatter in Figure 4.4 and is reflected in the poor fit between the regression lines and the data.

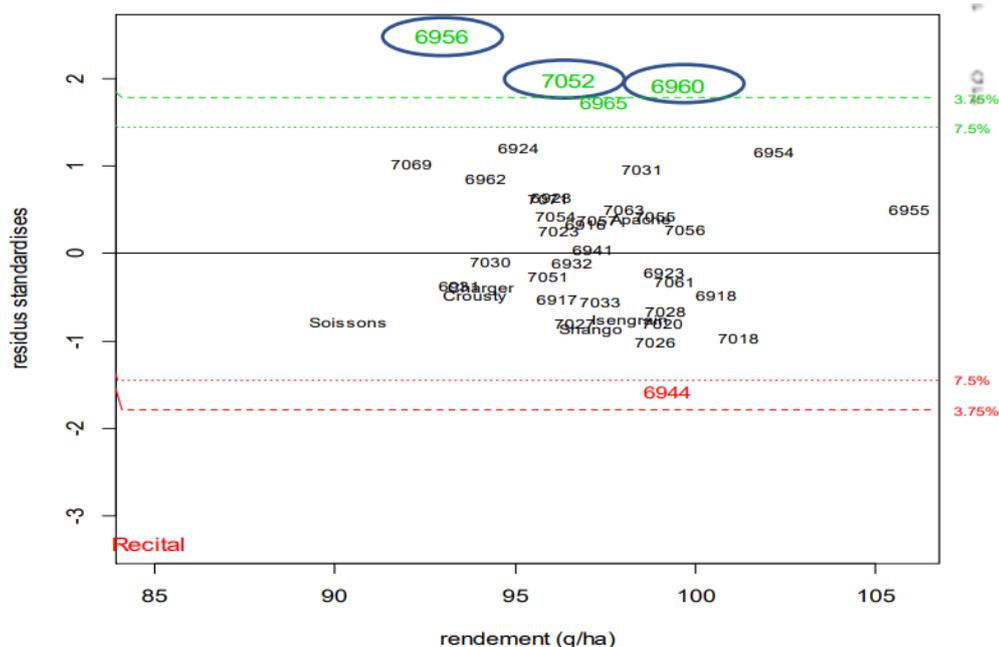
The difference of GPC (Figure 4.3.) and to a lesser extent of GY between the parents Malacca and Hereward generated a wide range of GPC-GY combinations (Figure 4.4) resulting in a low  $R^2$  for the simple linear regression GPC-GY. In contrast, breeding populations which are selected for both GPC and GY have a higher  $R^2$  for the simple linear regression GPC-GY (J. Le Gouis, personal communication, INRAe Clermont-Ferrand France, February 10, 2023)

Because of low  $R^2$ , large residuals (error) and consequently wide variation in GPD was observed in all environments: from -1.25% protein.  $m^2. g^{-1}$  to 2.36% protein.  $m^2. g^{-1}$  (Rothamsted 2019-2020), from -1.05% protein.  $m^2. g^{-1}$  to 1.67% protein.  $m^2. g^{-1}$  (Reading 2020-2021) and from -1.36% protein.  $m^2. g^{-1}$  to 1.19% protein.  $m^2. g^{-1}$  (Reading 2021-2022) which was promising for QTL mapping.

Simple GPC-GY regressions analysis on the individual environments highlighted several observations located above Hereward, parent of the DH with a positive high GPD. The significance of the differences of GPD between those observations and Hereward would need to be assessed by a series of two-sample t-tests before concluding to positive transgressive segregation.

#### 4.4.2. AMMI analysis

The AMMI analysis identified genotypes with high and stable GPD values in the DH population and in the three environments such as M x H 73 which varied from 0.63-0.72% protein. m<sup>2</sup>. g<sup>-1</sup> However, although GPD has been identified as a target trait by breeders in the UK, the threshold level required to be a viable economic target has not been established which renders the selection of high GPD genotypes difficult. By contrast, in France, the “RÈGLEMENT TECHNIQUE D'EXAMEN DES VARIÉTÉS DE CEREALES A PAILLE” (CTPS, 2020) specifies two GPD thresholds of 1.44 (7.5% quantile) and 1.78 (3.75% quantile) based on the GPD residuals graph (Figure 4.8.). High and very high GPD genotypes have positive standardised residuals and are located above the 7.5% and 3.75% thresholds (Figure 4.8, green dotted lines), respectively whereas low and very low GPD genotypes have negative standardised residuals and are situated below the 7.5% and 3.75% threshold (Figure 4.8, red dotted lines), respectively.



**Figure 4.8.** Plot of standardised residuals from the regression GPC-GY against GY in q/ha. The green and red dotted indicates positive and negative GPD thresholds corresponding to the quantiles 3.75% and 7.5% of the normal distribution

Reference: CTPS, 2020

#### 4.4.3. Correlations of GPC, GY, and GPD among the environments

The phenotypic correlations observed between the three trials were weak to medium for GPC ( $r=0.45-0.5$ ), GY ( $r=0.35-0.39$ ), and GPD ( $r=0.35-0.45$ ). These weak phenotypic correlations between the trials can be explained by well-known G x E effects of these traits (Basford and Cooper, 1998). GY outcome is known to be sensitive to high temperatures and water availability, especially during grain filling (Oury *et al.*, 2003). Post anthesis temperatures affects the rate and duration of grain filling with the optimum range being 15°C -20°C. Above 20°C the duration of grain filling is shortened and may or may not be compensated by a higher grain filling rate depending on the genotypes resulting in some cases in a lower kernel weight. Heat stress also shortens the grain filling period and interacts with temperature (Dupont and Altenbach, 2003).

GPC is mainly influenced by the environment and G x E interactions (Prasad *et al.*, 2003; Oury *et al.*, 2003). This was shown by a high variance ratio (7.34) between environment and genotype in a study of 27 cultivars grown at six locations carried out by Peterson *et al.* (1986). High temperatures increase the GPC (% protein) because of a decrease in starch accumulation in the grain but decrease the protein yield (in mg. grain<sup>1</sup>) (Zhao *et al.*, 2008).

#### 4.4.4. Correlations of GPD with GPC and yield components (GY and TKW, SW)

GPD was highly and positively correlated with GPC in the three environments with  $r > 0.91$  and  $p \ll 0.001$ . In contrast, GPD was weakly negatively correlated with GY in Rothamsted 2019-2020 ( $r = -0.26$ ;  $p = 0.01$ ), and with TKW in Reading 2021-2022 ( $r = -0.24$ ;  $p = 0.01$ ). The negative correlation with GY in Reading 2021-2022 was nearly significant ( $p = 0.052$ ). All the other correlations were not significant. In our study, GPD was mainly influenced by GPC with a small or no influence of GY (Reading 2020-2021).

#### 4.4.5. Single environment QTL analysis

Many individual QTL for GPC, SW, and TKW were detected on chromosomes 1A, 2A, 2B, 3B, 3D1, 4A, 5A, 5B,6B, 7A, 7B, and 7D1 (Table 4.4.). These QTL were only environment-specific and did not show pleiotropic effect.

In addition, the QTL analysis identified several partially overlapping confidence intervals for GPD and GPC (Figure 4.7, linkage groups 3A, 5B, and 5D) that could correspond to a single QTL with a pleiotropic effect on GPC and GPD or two closely linked QTL. On chromosome 3B,

in Reading 2020-2021 a QTL confidence interval (155.27 cM-159.74 cM) with a positive pleiotropic effect on GPD and GPC (0.12% protein.m<sup>2</sup>. g<sup>-1</sup>; 0.17 %protein) showed a 4cM overlaps with a QTL confidence interval (147.2 cM-159.74 cM) with positive pleiotropic effect on both GPD and GPC (0.19 %protein.m<sup>2</sup>. g<sup>-1</sup>, 0.23 % protein) detected in Reading 2021-2021. The multi-years and traits colocation on chromosome 3B may be explained by a single QTL with a stable (i.e., detected in two environments) pleiotropic effect on GPD and GPC. Importantly, the co-located QTL for GPD and GPC on chromosomes 3A and 3B described above also overlap with QTL with negative effect on SW (-0.1 kg. hL<sup>-1</sup>) in the Reading 2020-2021 trial and on GY (-27 g.m<sup>-2</sup>) in the Rothamsted 2019-2020 trial, respectively (Figure 4.7). This suggests that the exploitation of these QTL to increase GPC may be accompanied with a yield reduction in some environments (e.g., Rothamsted 2019-2020 where the GY QTL was detected). The 3B QTL region is nevertheless worth considering as the effects on GPD and GPC were stable in the Reading environments and did not co-locate with a GY QTL with a negative effect (Figure 4.7).

The QTL intervals chromosome locations identified in this study were compared with those of the literature (Table 4.6). For GPD, Nigro *et al.* (2019) identified three QTL at both ends of the long arm of chromosome 5B of durum wheat (*Triticum turgidum* L. ssp. *durum*) It is therefore possible that one of them co-locates with the 5B GPD QTL identified in this study. The chromosomal locations of the four GPC QTL (3A, 3B, 5B, 7A) may also coincide with those reported by Nigro *et al.* (2019) (Table 4.6).

The single TKW QTL detected on chromosome 2A may co-locate with one of the QTL reported by Wang *et al.* (2009) or Cui *et al.* (2014). Finally, the chromosomal locations of the HD QTL (1A, 5A, and 7B) could agree with those mapped by Zanke *et al.* (2014).

To confirm the QTL identified in other studies, one possibility would be to compare the position of the markers flanking the QTL confidence interval from both studies on the reference sequence of Chinese Spring (IWGSC RefSeq v1.0). If the QTL are the same, the confidence intervals should overlap.

Comparisons of the QTL from our study with others could be achieved by mapping the markers of the confidence intervals on the reference sequence of Chinese spring (RefSeq v1.0) to compare their physical position.

**Table 4.6** Comparison of the QTL detected in the M x H DH population with the literature

Author	Specie	Trait	Chromosome	Marker
Nigro <i>et al.</i> 2019	<i>Triticum durum</i>	GPD	5B	IWB34458
Nigro <i>et al.</i> 2019	<i>Triticum durum</i>	GPD	5B	IWB6634
Nigro <i>et al.</i> 2019	<i>Triticum durum</i>	GPD	5B	IWB2716
Nigro <i>et al.</i> 2019	<i>Triticum durum</i>	GPC	3A	IWB14495
Nigro <i>et al.</i> 2019	<i>Triticum durum</i>	GPC	3A	IWB35484
Nigro <i>et al.</i> 2019	<i>Triticum durum</i>	GPC	3B	IWB13886
Nigro <i>et al.</i> 2019	<i>Triticum durum</i>	GPC	5B	IWB6634
Nigro <i>et al.</i> 2019	<i>Triticum durum</i>	GPC	5B	IWB2716
Nigro <i>et al.</i> 2019	<i>Triticum durum</i>	GPC	7A	IWB65659
Prasad <i>et al.</i> 2002	<i>Triticum aestivum</i>	GPC	7A	Xgwm1171
Wang <i>et al.</i> 2008	<i>Triticum aestivum</i>	TKW	2A	-
Cui <i>et al.</i> 2014	<i>Triticum aestivum</i>	TKW	2A	-

#### 4.4.6. Future prospects

After discovering a stable QTL, the next step is to validate its effect alone (i.e., without a different QTL associated with the same trait) by comparing sets of NILs with the Malacca allele and the Hereward allele at the QTL (refer to chapter 2, 2.1.2, for a detail description on the construction of NILs). Practically, as an example, the 3B GPD-GPC QTL (4.3.5.2) will be introgressed in the Malacca genetic background using its flanking markers: AX-94610348 and AX-95161 to perform marker assisted selection. The assessment of the QTL effect should be done in at least three environments and the QTL effect should be significant in two environments before going any further into the validation process. The next step would be to fine map the QTL to narrow down its confidence interval. This can be done by genotyping SNP within the QTL region followed by a QTL detection to locate the QTL more precisely. In the resulting shorter QTL region, RNA-Seq could be performed to identify differentially expressed genes (DEGs) between the NILs parental lines. Ultimately, the function of the DEGs could be predicted to inform the selection of the candidate genes. In our case, any protein coding genes involved in nitrogen or carbon metabolism would be a good candidate for GPD.

## Chapter 5. Analysis and comparison of senescence metrics within the DH population

### 5.1. Introduction

Senescence is a process during which the leaf photosynthetic apparatus is degraded, and nitrogen is remobilised into the developing grain (Gregersen *et al.*, 2008). The timing of senescence had been shown to affect both GPC and GY. Functional stay-green phenotypes show delayed senescence and produce higher yields due to a prolonged photosynthesis period (Gregersen *et al.*, 2008). However, this is often associated with lower GPC as nitrogen is remobilised less efficiently (Gaju *et al.*, 2014). By contrast, fast senescing cultivars bearing the *GPC-B1* locus have enhanced GPC because of a higher rate of nitrogen remobilisation but show a yield reduction (Uauy *et al.*, 2006). Therefore, the present study assessed and compared the senescence patterns of lines within the DH population and explored its relationship with GPD.

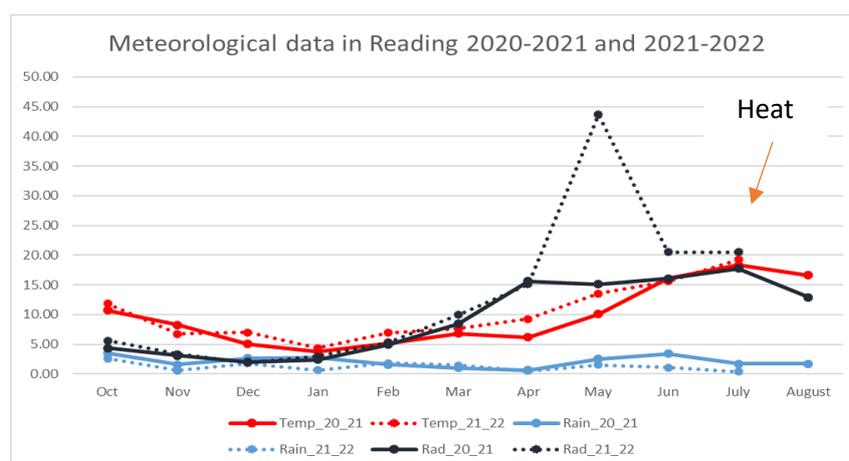
### 5.2. Results

#### 5.2.1 Description of weather data

The growth cycle of the DH population in Reading 2021-2022 was a month shorter (nine months) than the lines grown in Reading 2020-2021 (ten months) (Table 5.1). High temperatures were reported to accelerate the crop cycle (He *et al.*, 2015). The Reading 2021-2022 trial was on average warmer (+0.4°C) and dryer (-266mm) than the Reading 2020-2021 trial (Table 5.1.) and had much higher solar radiation (+2135 MJ/m<sup>2</sup>). The DH population grown in 2021-2022 experienced a period of unusual heat (individual temperatures above 35°C) at the end of the crop cycle in July during the UK heatwave. This heat wave was not noticeable when considering the average monthly temperatures and it did not appear to have an impact on the average GY and GPC in the DH population (Chapter 4, Table 4.2.) which can be explained by its occurrence after the grain filling period. The monthly average temperature, rain, and radiation curves superimposed well except in May where the solar radiations were much higher in the Reading 2021-2022 trial (Figure 5.1).

**Table 5.1.** Meteorological data for the Reading 2020-2021 and 2021-2022 trials. Rainfall corresponds to the cumulated daily precipitation in mm; temperature is the average temperature in degree Celsius and radiation is the cumulated daily incident radiation over the crop season expressed in MJ/m<sup>2</sup>.

Year	Site	Sowing	Harvest	Rainfall (mm)	Temperature (°C)	Radiation (MJ/m <sup>2</sup> )
2020-2021	Reading	15/10/2020	17/08/2021	625.8	9.8	1587.8
2021-2022	Reading	07/10/2021	22/07/2022	359.4	10.2	3722.9



**Figure 5.1.** Line chart showing the average monthly temperatures (°C), rainfalls (mm), and solar radiations (MJ/m<sup>2</sup>) in the Reading 2020-2021 (solid line) and Reading 2021-2022 (dotted line) trials.

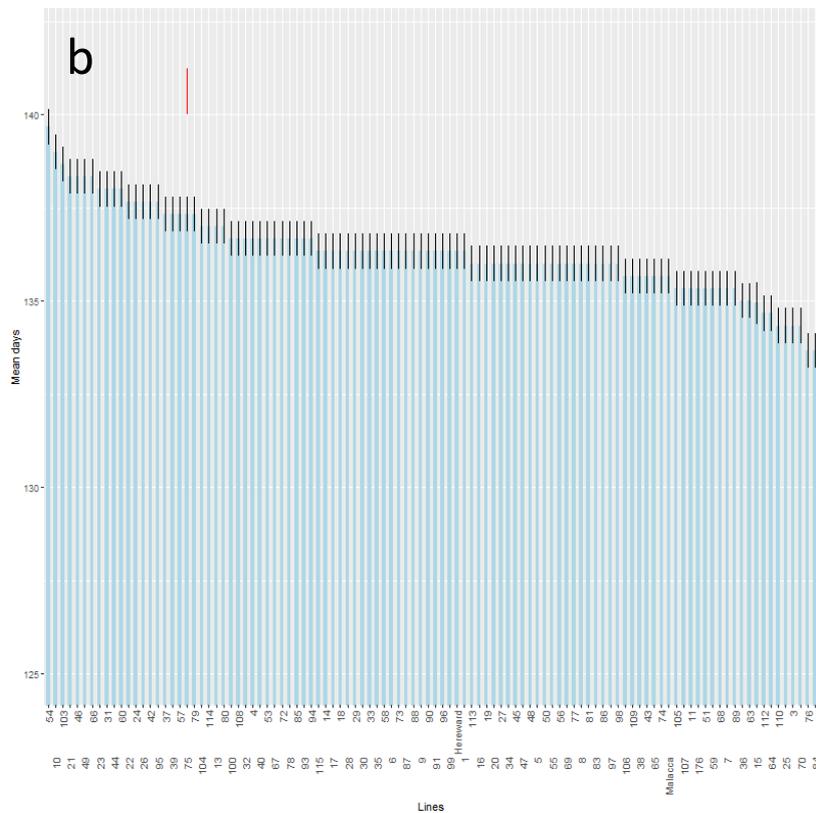
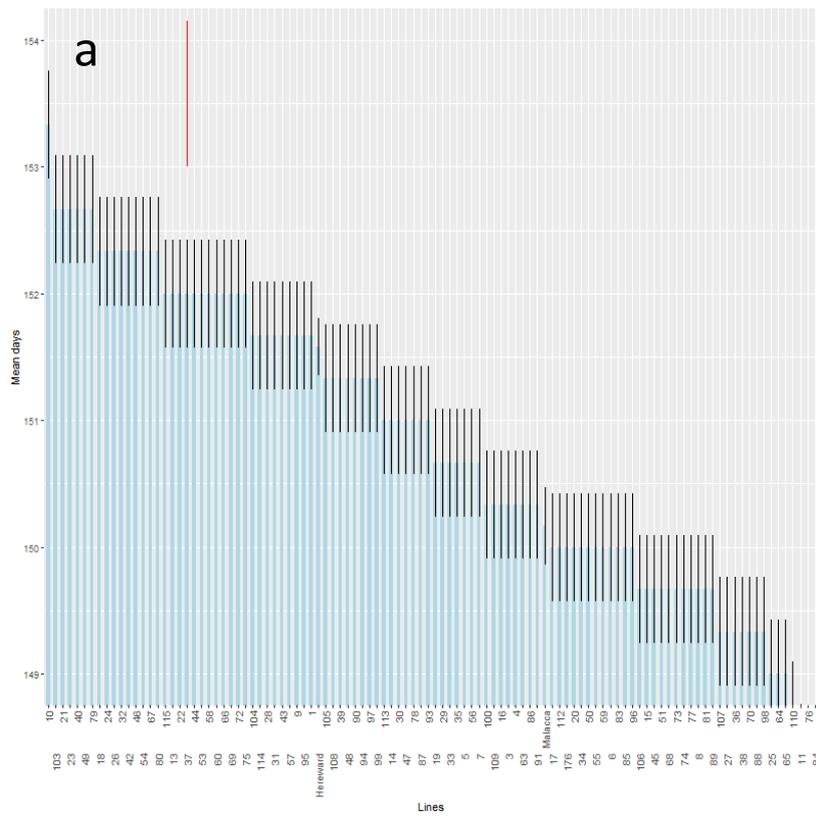
### 5.2.2. Description of HD data

The HD are expressed as days elapsed since January 1<sup>st</sup>, 2020, or 2021 and were scored during two field seasons (Chapter 2, 2.16.1), in 2020-2021 and 2021-2022 at Reading.

In the Reading 2020-2021 trial (Figure 5.2, a), the mean HD ranged from 147 (+/-0.42) days for lines 84 to 153 (+/- 0.42) days for line 10. Malacca headed (GS 55) in 150 days, that is one day before Hereward (151 days). However, the LSD calculation shows no significant difference between their mean heading dates (Figure 5.2., red vertical bars).

In the Reading 2021-2022 trial (Figure 5.2., b), the DH lines headed on average 14 days (136 days) before the those grown in the 2020-2021 trial (150 days), and the HD varied between 133 (+/- 0.48) days for line 76 and 139 (+/- 0.48) days for line 54. Malacca headed (GS 55) in 135 days, that is, one day before Hereward (136 days). Again, the LSD calculation shows no significant difference between the means of the two parents.

The HD were strongly positively correlated in the two environments ( $r=0.80$ ,  $p<2.2e-16$ ).



**Figure 5.2.** Barplots representing the mean HD as number of days elapsed since January 1<sup>st</sup> to reach GS 55 (heading date on the Zadoks scale) in the DH population grown at Reading in 2020-2021 (a) and 2021-2022 (b). The black vertical bars are standard errors of the means (n=3), and the red vertical lines represent the average Least Significant difference (LSD) which is of 1.74 days (a) and 1.23 days (b).

### 5.2.3. Modelling of senescence

Four logistic regression models were compared to represent the relationship between NDVI and degree days (Chapter 2, 2.16.3) in the Reading 2020-2021 and 2021-2022 trials and an ANOVA was performed to select the most parsimonious model. The four parameters controlling the shape of the logistic curve are A and C, which are linear parameters setting the level and controlling the scale of the curve, respectively, and  $\beta$  and M which are non-linear parameters controlling the curvature and position of the curve with respect to values of x.

In the Reading 2020-2021 trial, the full model with all four parameters (A, B, C, and M) (model 4, Chapter 2, 2.16.3) explained 99.2% of the variation of NDVI ( $R^2=0.992$ ) and had a highly significant p-value ( $p<0.001$ ). This model was therefore chosen to model the data.

In the Reading 2021-2022 trial, the p-value of model 3 (Chapter 2, 2.16.3) with parameters B and M constants, and A and C allowed to vary was non-significant ( $p=0.748$ ) indicating that the parallel line model ( $R^2=96.4\%$ ) (Chapter 2, 2.16.3, model 2) which allows only A to vary among the lines was sufficient to represent the data.

### 5.2.4. Comparison of senescence metrics between the DH lines

Seven senescence metrics corresponding to thermal time metrics (TT90, TT50, and T10) and NDVI indicators (Max\_NDVI, NDVI90, NDVI50, and NDVI10) were derived from the senescence curves ( $NDVI = f(\text{degree days after anthesis})$ ) to compare the senescence profile of the DH lines. The calculation of the senescence metrics is detailed in Chapter 2.16.3.

#### 5.2.4.1. Thermal time metrics

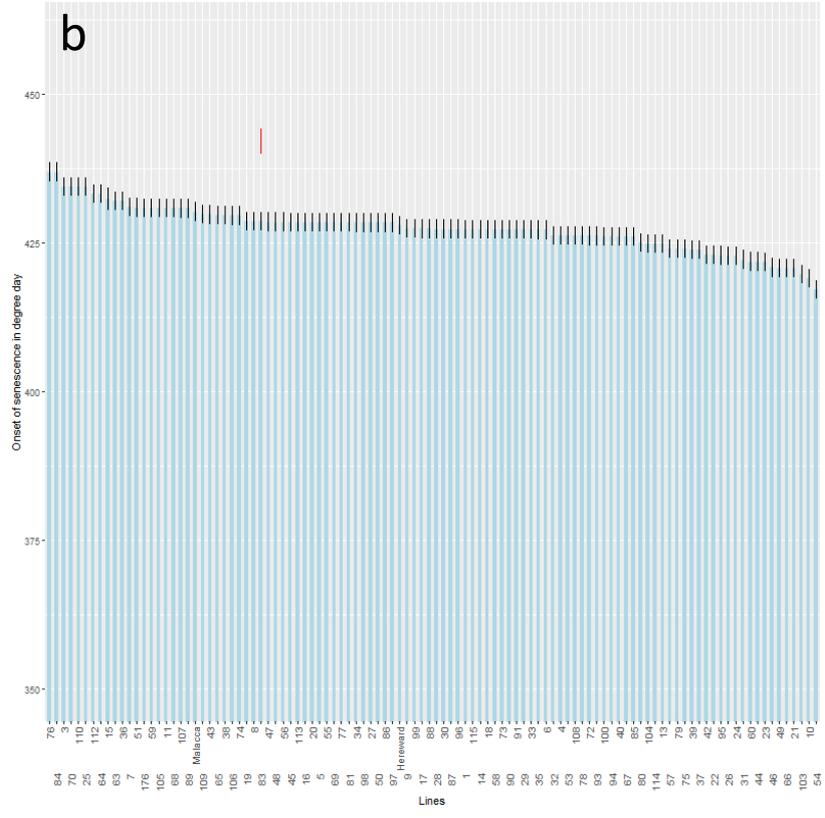
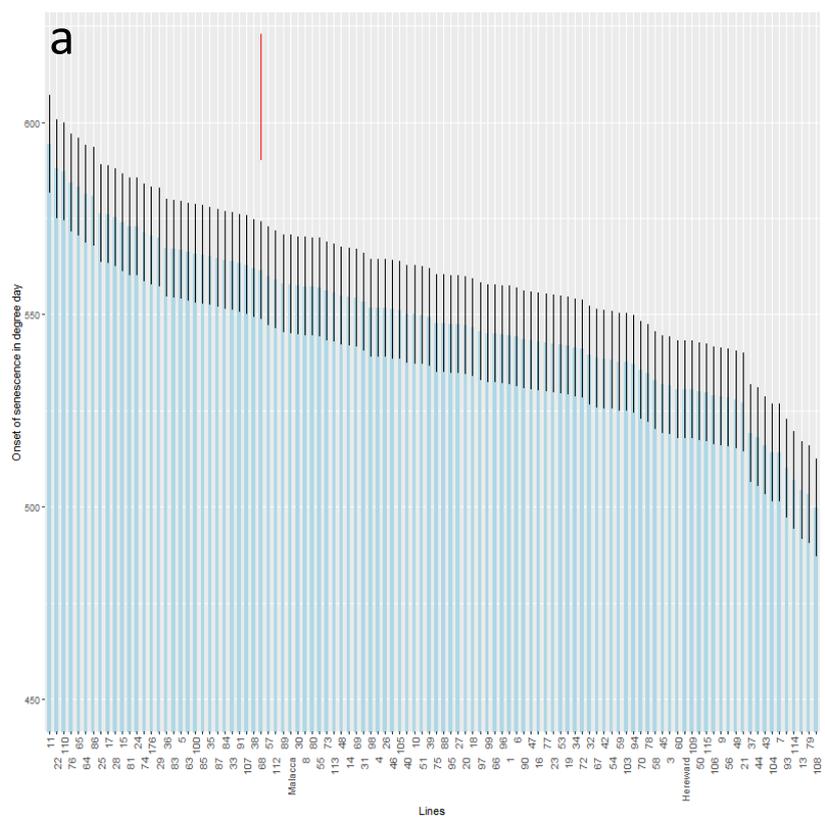
TT90 (onset of senescence) is the cumulated thermal time from heading expressed in degree days which corresponds to a 10% decrease of the maximum NDVI.

In the Reading 2020-2021 trial (Figure 5.3. a), the mean TT90 was 548 degree days and the TT90 values ranged from 500 (+/-12.73) degree days (line 108) to 594 (+/- 12.73) degree days (line 11). This interval corresponds to a difference of 5-6 calendar days. Hereward began senescing at 530 (+/-12.73) degree days which was 1-2 calendar days earlier than Malacca which started to senesce at 557 (+/- 12.73) degrees-day. However, the LSD calculation (Figure 5.3, a) and a comparison of their mean TT90 using a t-test indicated no significant differences ( $p=0.099$ ). Despite this, the LSD calculation shows significant differences of onset at the

extremes (i.e., between lines with low and high TT90 values), such as between line 108 and line 11 which suggest the presence of stay green cultivars.

In the Reading 2021-2022 trial (Figure 5.3, b), the mean TT90 was 427 degree days and the TT90 values ranged between 417 (+/-1.57) degree days (line 54) and 436 (+/- 1.57) degree days (line 84); a difference of approximately one calendar day. Hereward and Malacca began to senesce the same day at 427 (+/-1.57) degree days for Hereward and at 430 (+/- 1.57) degree days for Malacca. This was confirmed by the LSD calculation and a t-test on their means which showed no significant differences ( $p=0.27$ ). The LSD calculation shows differences between lines at extremes, for example, between line 54 (low TT90 value) and 76 (high TT90 value) which suggest the presence of stay green cultivars.

The TT90 Pearson correlation between the two years was weak ( $r=0.34$ ) but highly significant ( $p<<0.001$ ).



**Figure 5.3.** Barplots representing the mean TT90 expressed in degree days in the DH population grown at Reading 2020-2021 (a) and 2021-2022 (b). The black vertical bars are the standard errors of the means (n=3), and the red vertical lines represent the average Least Significant Difference (LSD) which is of 33.04 degree days (a) and 4.22 degree days (b).

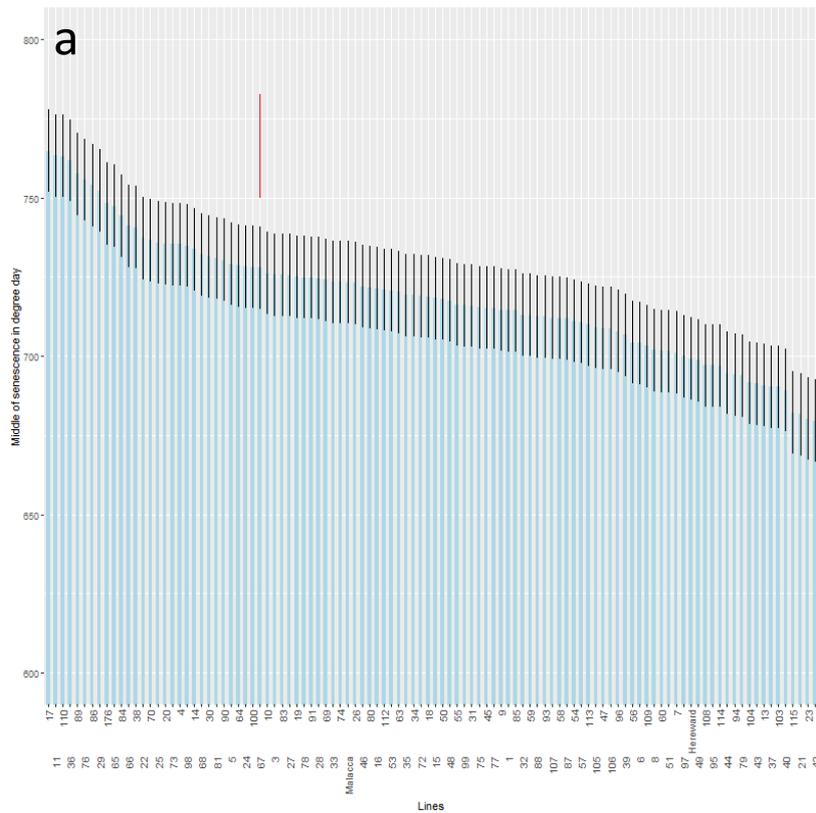
TT50 (mid senescence) is the cumulated thermal time from heading expressed in degree days which corresponds to a 50% decrease in the maximum NDVI.

In the Reading 2020-2021 trial (Figure 5.4, a), the average TT50 was 718 (+/- 13) degree days and varied between 679 (+/-13) degree days for line 52 and 764 (+/-13) degree days for line 17 corresponding to a difference of 5-6 calendar days.

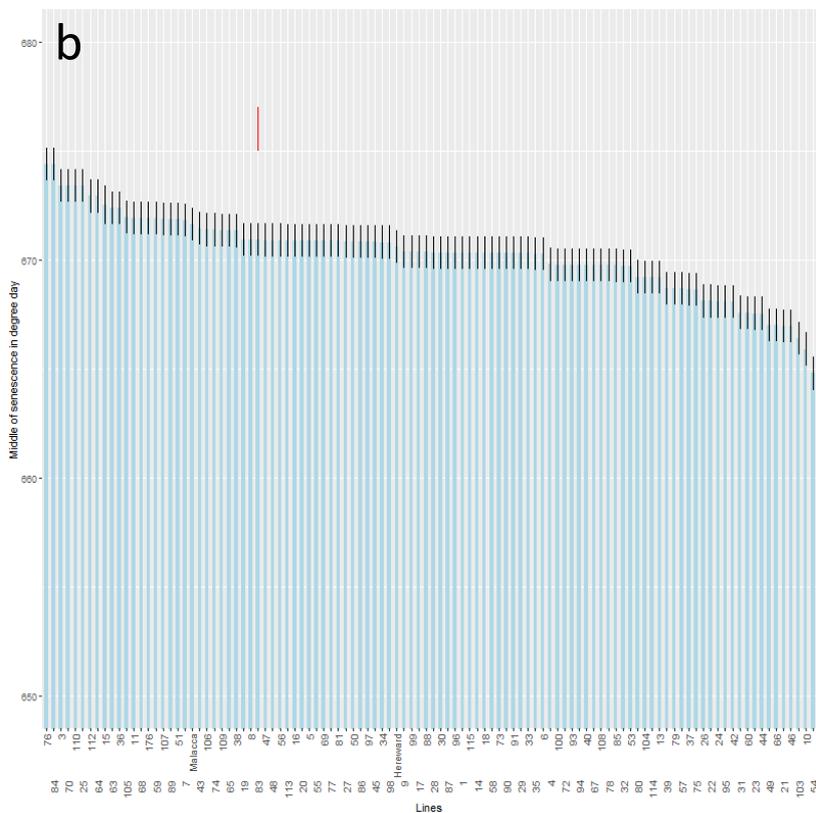
Hereward reached TT50 at 699 (+/-13) degree days while Malacca reached TT50 at 723 (+/- 13) degree days, which is one or two days after Hereward. The LSD calculation showed no significant difference between the mean TT50 of the two parents but suggests differences at extremes, for example, between line 42 (low TT50) and 17 (high TT50).

In the Reading 2021-2022 trial (Figure 5.4., b), the average TT50 was 670 (+/- 0.75) degree days and the variation in TT50 was much smaller than in Reading 2020-2021 as the lines varied between 664 (+/- 0.75) for line 54 and 674 (+/- 0.75) for line 84. The LSD calculation suggests differences in TT50 at extremes (e.g., between lines 54 and 76) but does not indicate a significant difference of TT50 between the parental lines.

The TT50 Pearson correlation between the two years was slightly higher ( $r=0.42$ ) than for the TT90 and highly significant ( $p<<0.001$ ).



**Figure 5.4.** Barplots representing the mean TT50 of the DH lines expressed in degree days in the Reading 2020-2021 (a) and 2021-2022 (b) trials. The black vertical bars are standard errors of the means (n=3) and the red vertical lines represent the average Least Significant difference (LSD) which is of 32.65 degree days (a) and 2 degree days (b).

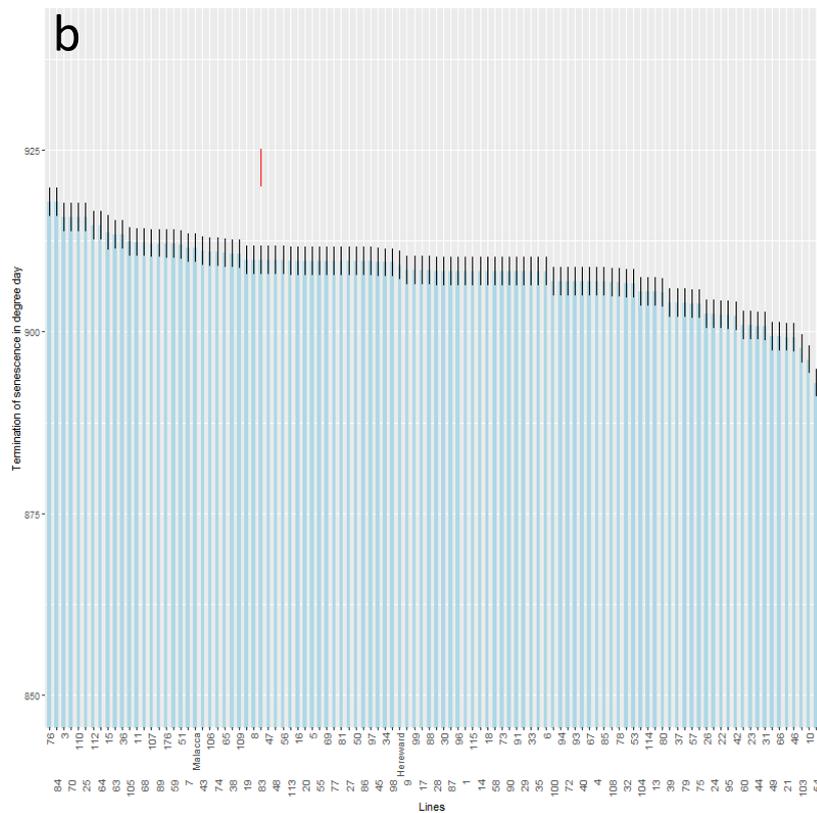
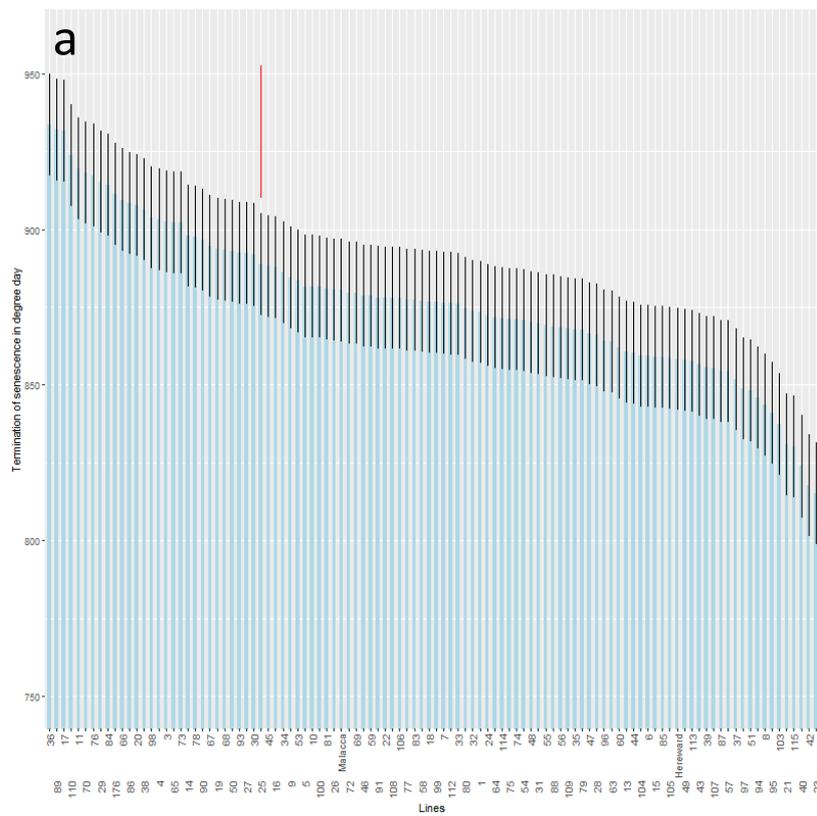


TT10 (termination of senescence or senescence duration) is the cumulated thermal time from heading date expressed in degree days which corresponds to a 90% decrease of the maximum NDVI.

In the Reading 2020-2021 trial (Figure 5.5, a), the average TT10 was 877 (+/- 16.37) degree days and there was variation of 115 degree days (6-7 calendar days) between lines with low (line 23, 815 (+/- 16.37) and high (line 36, 933 (+/- 16.37) TT10 values. Hereward reached TT10 at 858 (+/-16.37) degree days, that is one or two calendar days earlier than Malacca (880 (+/- 16.37) degree days). The LSD calculation does not show a significant difference in TT10 between the parental means but shows differences at the extremes (e.g., between lines 23 and 36).

In the Reading 2021-2022 trial (Figure 5.5, b), the average TT10 was 908 (+/- 1.94) degree days and the lines varied between 893 (+/- 1.94) degree days for line 54 and 917 (+/- 1.94) degree days for line 84, corresponding to a difference of 20 degree days or 1-2 calendar days. Hereward and Malacca showed similar durations of senescence (i.e., TT10 values) of 909 and 911 degree days, respectively. The LSD calculation does not show a significant difference in TT10 between the parental means but shows differences at the extremes.

The TT10 Pearson correlation between the two years was positive and highly significant ( $p < 0.001$ ) ( $r = 0.40$ ).



**Figure 5.5.** Barplots representing the TT10 means in degree days in the Reading 2020-2021 (a) and 2021-2022 trials (b). The black vertical bars are standard errors of the means (n=3), and the red vertical lines represent the average Least Significant Difference (LSD) which is of 42.36 degrees-day (a) and 5.17 degrees-day (b).

#### 5.2.4.2 NDVI metrics

NDVI90 is the NDVI value at 90% of the maximum NDVI (Chapter 2, 2.16.3).

In the Reading 2020-2021 trial (Figure 5.6, a), the average NDVI90 in the DH population was 0.86 and ranged between 0.84 (+/- 0.006) for line 48 to 0.88 (+/- 0.006) for line 4.

Malacca and Hereward had the same NDVI90 value (0.86, +/-0.006). The LSD calculation supports differences between the extremes, for example, between lines 48 and 4.

In the Reading 2021-2022 trial (Figure 5.6, b), the average NDVI90 in the DH population was 0.86 and the lines varied between 0.82 (+/- 0.018) for line 7 and 0.90 (+/- 0.018) for line 90.

Malacca and Hereward had the same value of NDVI90 (0.87, +/- 0.018) and the LSD calculation showed no significant differences in NDVI90 within the DH population.

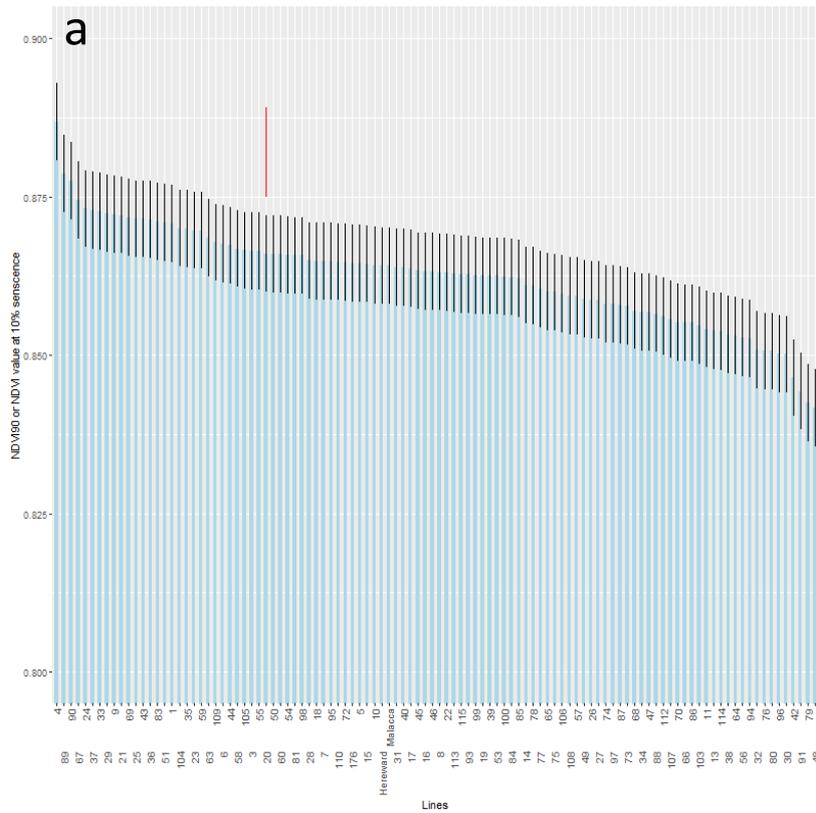
The Pearson correlation of NDVI90 between the two years was weak ( $r=0.23$ ) but significant ( $p=0.01$ ).

NDVI50 is the NDVI value at 50% of the maximum NDVI (mid senescence) (Chapter 2, 2.16.3).

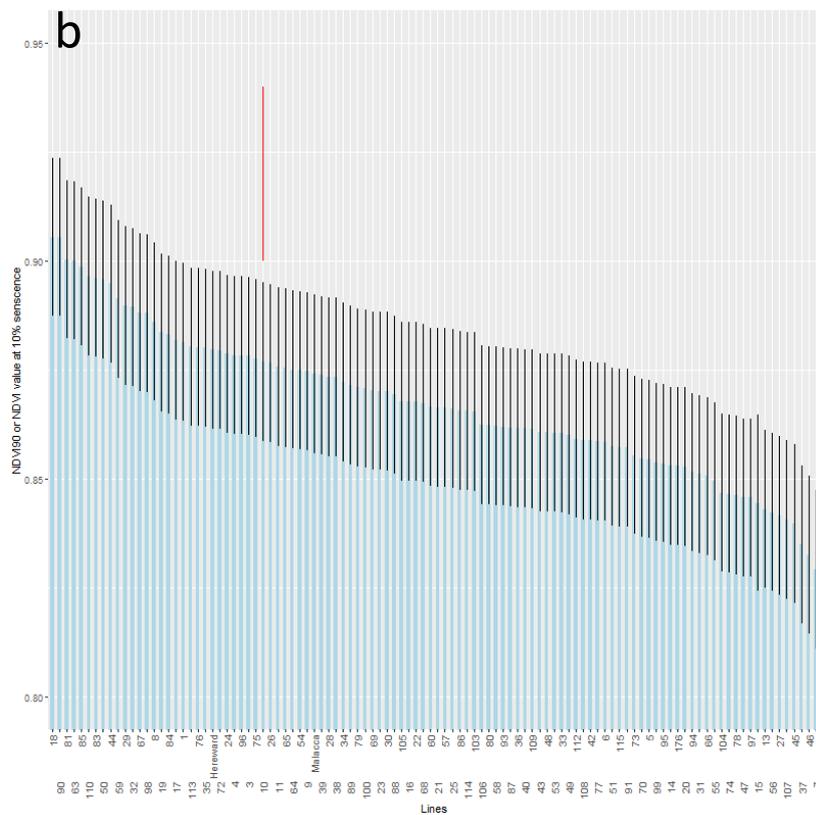
In the Reading 2020-2021 trial (Figure 5.7, a), Malacca and Hereward had both NDVI50 of 0.70 (+/- 0.007) while the lines had an average NDVI50 of 0.70 and varied between 0.68 (+/- 0.007) for line 79 and 0.73 (+/- 0.007) for line 4. The LSD value shows differences at extremes between lines with low and high NDVI50 values, for example between lines 4 and 79 (Figure 5.7, a)

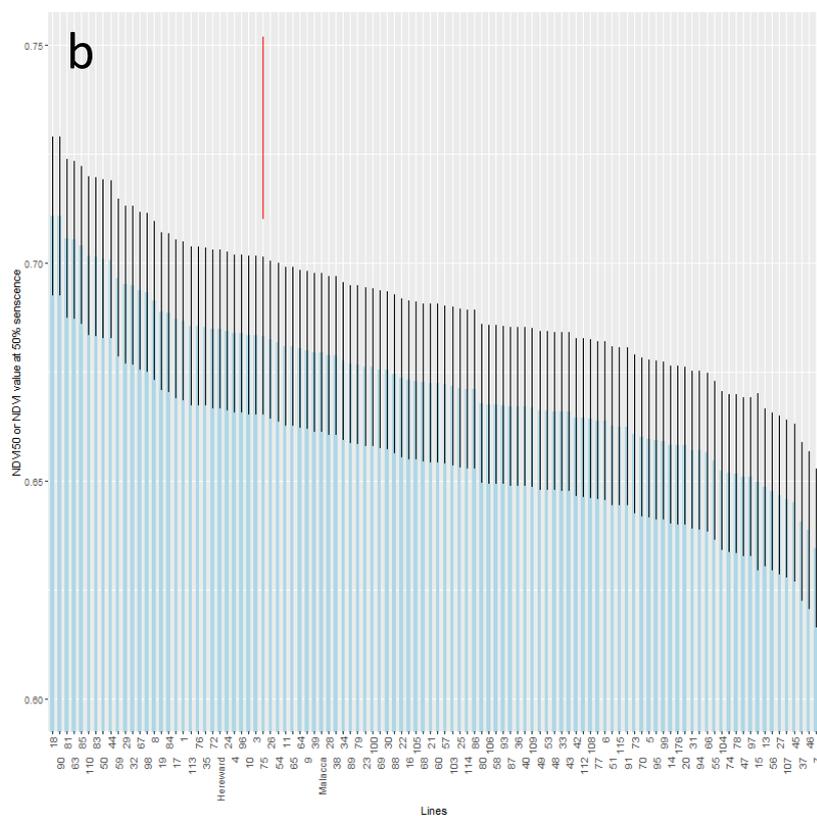
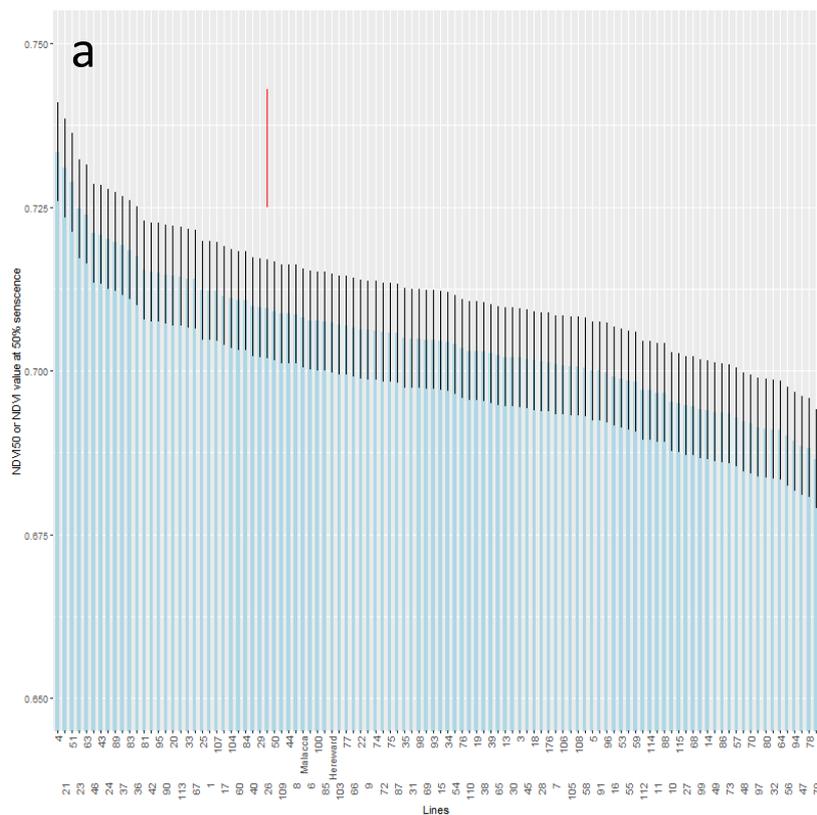
In the Reading 2021-2022 trial (Figure 5.7, b), the DH lines had an average NDVI50 of 0.67 and differed from 0.63 (+/- 0.018) for line 7 to 0.71 (+/-0.018) for line 90. Malacca and Hereward obtained similar NDVI50 of 0.67 and 0.68, respectively. The calculated LSD shows no significant differences in NDVI50 between the DH lines (Figure 5.7, b).

The Pearson correlation of NDVI50 between the two years was significant ( $p=0.02$ ) but weak ( $r=0.21$ ).



**Figure 5.6.** Barplots representing the doubled haploid lines mean NDVI90 values in Reading 2020-2021 (a) and 2021-2022 (b). The black vertical bars are standard errors of the means (n=3), and the red vertical lines represent the average Least Significant Difference (LSD) which is of 0.01 NDVI (a) and 0.04 NDVI (b).





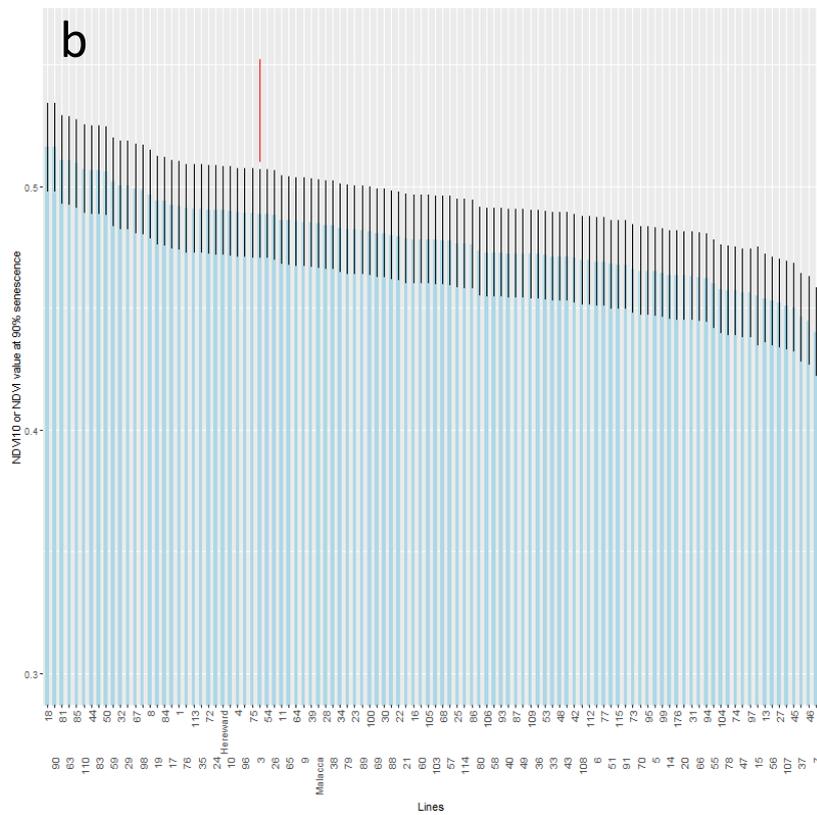
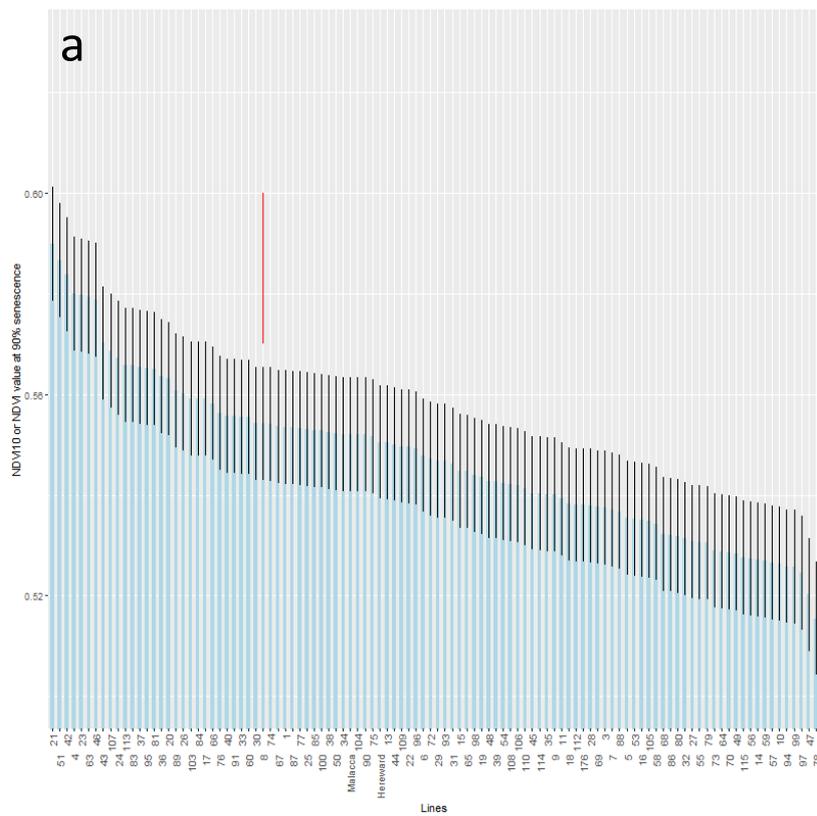
**Figure 5.7.** Barplots representing the NDVI50 means in the DH population grown at Reading 2020-2021 (a) and 2021-2022 (b). The black vertical bars are standard errors of the means (n=3), and the red vertical lines represent the average Least Significant Difference (LSD) which is of 0.018 NDVI (a) and 0.04 NDVI (b).

The NDVI10 is the NDVI value at 10% of the maximum NDVI (senescence completion) (Chapter 2, 2.16.3).

In the Reading 2020-2021 trial (Figure 5.8, a), the lines had an average NDVI10 of 0.54 (+/- 0.01) and varied between 0.51 (+/- 0.011) for line 78 and 0.58 (+/- 0.011) for line 21. The parents had the same NDVI10 value of 0.55 (+/- 0.01). The LSD value shows differences at extremes between lines with low and high NDVI10 values (e.g., lines 78 and 21) (Figure 5.8, a).

In the Reading 2021-2022 trial (Figure 5.8., b), the lines had an average NDVI10 value of 0.47 (+/- 0.01) and varied between 0.44 (+/- 0.01) for line 7 and 0.51 (+/- 0.01) for line 90. Malacca and Hereward had similar values of NDVI10 of 0.48 (+/-0.01) and 0.49 (+/- 0.01), respectively. The calculated LSD shows no significant differences between the lines (Figure 5.8., b).

The Pearson correlation of NDVI10 between the two environments was weak ( $r=0.15$ ) and non-significant ( $p=0.09$ ).



**Figure 5.8.** Barplots representing the NDVI10 means in the DH population grown at Reading 2020-2021 (a) and 2021-2022 (b). The black vertical bars are standard errors of the means (n=3), and the red vertical lines represent the average Least Significant Difference (LSD) which is of 0.02 NDVI (a) and 0.04 NDVI (b).

## 5.2.5. Exploration of correlations

5.2.5.1. Correlations between HD and, GPC, Yield (GY, TKW), GPD, and the seven senescence metrics (NDVI90, NDVI50, NDVI10; TT90, TT50, TT10; and Max\_NDVI) in the Reading 2020-2021 and 2021-2022 trials.

The Pearson correlations between HD, and GPD, GY, TKW, GPC, and the NDVI metrics were not significant at the 5% threshold. By contrast, HD were significantly and highly negatively correlated with all thermal time metrics (TT90, TT50, TT10) in both environments with strong correlations observed in the Reading 2021-2022 trial ( $r > 0.99$ ) (Table 5.2.).

Table 5.2. Pearson correlations between HD expressed in number of days elapsed since January the 1<sup>st</sup> (column “Trait 1”), and the seven senescence metrics, GY, TKW, GPC and GPD (column trait 2) in the Reading 2020-2021 and 2021-2022 trials. Significant correlations at a threshold of 5% are highlighted in red.

Trait1	Trait2	r	P
Heading_Date_Reading_2020_2021	TT10_2020_2021_Reading	-0.52692	7.73E-09
Heading_Date_Reading_2020_2021	TT50_2020_2021_Reading	-0.56707	2.85E-10
Heading_Date_Reading_2020_2021	TT90_2020_2021_Reading	-0.47957	2.27E-07
Heading_Date_Reading_2021_2022	TT10_2021_2022_Reading	-0.99626	0
Heading_Date_Reading_2021_2022	TT50_2021_2022_Reading	-0.99914	0
Heading_Date_Reading_2021_2022	TT90_2021_2022_Reading	-0.99835	0
Heading_Date_Reading_2021_2022	TKW_2021_2022_Reading	-0.2285	0.019052
Heading_Date_Reading_2020_2021	NDVI10_2020_2021_Reading	0.010472	0.915561
Heading_Date_Reading_2020_2021	NDVI50_2020_2021_Reading	0.024269	0.805884
Heading_Date_Reading_2020_2021	NDVI90_2020_2021_Reading	0.039309	0.690531
Heading_Date_Reading_2020_2021	Max_NDVI_2020_2021_Reading	0.039681	0.687759
Heading_Date_Reading_2020_2021	Yield_2020_2021_Reading	-0.00869	0.929927
Heading_Date_Reading_2020_2021	Protein_2020_2021_Reading	-0.10236	0.298795
Heading_Date_Reading_2020_2021	GPD_2020_2021_Reading	-0.12955	0.187789
Heading_Date_Reading_2020_2021	Max_Rate_2020_2021_Reading	-0.12878	0.1904
Heading_Date_Reading_2021_2022	NDVI10_2021_2022_Reading	-0.04968	0.614738
Heading_Date_Reading_2021_2022	NDVI50_2021_2022_Reading	-0.06502	0.50988
Heading_Date_Reading_2021_2022	NDVI90_2021_2022_Reading	-0.08015	0.416353
Heading_Date_Reading_2021_2022	Max_NDVI_2021_2022_Reading	-0.0841	0.393677
Heading_Date_Reading_2021_2022	Yield_2021_2022_Reading	-0.04745	0.630747
Heading_Date_Reading_2021_2022	Protein_2021_2022_Reading	0.154534	0.115484
Heading_Date_Reading_2021_2022	GPD_2021_2022_Reading	0.157829	0.107835

### 5.2.5.2. Correlations between GPC and the seven senescence metrics

Pearson correlation coefficients between GPC and the seven senescence metrics (TT90, TT50, TT10, NDVI90, NDVI50, NDVI10, and max\_NDVI) were weak ( $r < 0.50$ ) and ranged from -0.27 (Max\_NDVI\_2020\_2021\_Reading and Protein\_2020\_2021\_Reading) to 0.12 (Protein\_2020\_2021\_Reading and NDVI10\_Reading\_2020\_2021 (Table 5.3.). Of the 14 correlations of senescence metrics tested against GPC, six involving NDVI metrics Max\_NDVI, NDVI90, NDVI50, and NDVI10 (Table 5.3., red highlight) were significantly ( $p < 0.05$ ) and weakly negatively correlated with GPC in Reading 2020-2021 and 2021-2022 trials. This suggests that lines with high GPC values tend have low NDVI values and vice versa.

However, only the NDVI90 and Max\_NDVI were significantly correlated with GPC for both years while the other two NDVI metrics (NDVI50 and NDVI10) were not significantly correlated with GPC in the Reading 2020-2021 trial. The thermal time metrics were not significantly correlated with the GPC at the 5% threshold in either year.

Table 5.3. Pearson correlations between GPC (column trait 1) and the seven senescence metrics (NDVI90, NDVI50, NDVI10; TT90, TT50, TT10; and Max\_NDVI) in the Reading 2020-2021 and 2021-2022 trials Significant correlations at a threshold of 5% are highlighted in red. (column trait 2).

Trait1	Trait2	r	p
Protein_2020_2021_Sonning	TT10_Reading_2020_2021	-0.08	0.38
Protein_2020_2021_Sonning	TT50_Reading_2020_2021	-0.10	0.28
Protein_2020_2021_Sonning	TT90_Reading_2020_2021	-0.07	0.44
Protein_2020_2021_Sonning	NDVI50_Reading_2020_2021	0.00	0.96
Protein_2020_2021_Sonning	NDVI10_Reading_2020_2021	0.12	0.21
Protein_2020_2021_Sonning	NDVI90_Reading_2020_2021	-0.23	0.02
Protein_2020_2021_Reading	Max_NDVI_2020_2021_Reading	-0.27	0.01
Protein_2021_2022_Sonning	TT10_Reading_2021_2022	-0.15	0.11
Protein_2021_2022_Sonning	TT50_Reading_2021_2022	-0.15	0.11
Protein_2021_2022_Sonning	TT90_Reading_2021_2022	-0.15	0.11
Protein_2021_2022_Sonning	NDVI50_Reading_2021_2022	-0.21	0.02
Protein_2021_2022_Sonning	NDVI10_Reading_2021_2022	-0.21	0.03
Protein_2021_2022_Sonning	NDVI90_Reading_2021_2022	-0.22	0.02
Protein_2021_2022_Reading	Max_NDVI_2021_2022_Reading	-0.21	0.03

### 5.2.5.3. Correlations between GY and TKW, and the seven senescence metrics

Pearson correlations between GY (Yield) and the seven senescence traits ranged from 0.18 (Yield\_2020\_2021\_Reading and TT10\_Reading\_2020\_2021) to 0.63

(Yield\_2021\_2022\_Reading and NDVI90\_Reading\_2021\_2022) (Table 5.4). Seven correlations were significant at a threshold of 5%; five of them being related to NDVI traits and two of them to thermal time metrics (Table 5.4, red highlight). The four NDVI traits (Max\_NDVI, NDVI90, NDVI50, and NDVI10) were significantly ( $p < 0.001$ ) and highly positively correlated with yield with  $r$  ranging from 0.51 to 0.63. Max\_NDVI was the only NDVI metric to be significantly correlated in both environments. In contrast, the two thermal time metrics (TT10, and TT50) were less strongly associated with GY ( $r = 0.18$  and  $r = 0.21$ ).

All senescence metrics except Max\_NDVI were positively correlated with the TKW measured in the Reading 2021-2022 Reading trial with the correlation coefficients  $r$  ranging from 0.22 to 0.31 (Table 5.4, red highlight).

Table 5.4. Pearson correlations between yield (GY and TKW, column “trait 1”) and the senescence metrics in the Reading 2020-2021 and 2021-2022 trials (column “trait 2”). Significant correlations at a threshold of 5% are highlighted in red.

Trait1	Trait2	r	P
TKW_2021_2022_Reading	TT50_2021_2022_Reading	0.24	0.02
TKW_2021_2022_Reading	NDVI10_2021_2022_Reading	0.31	0.00
TKW_2021_2022_Reading	NDVI50_2021_2022_Reading	0.31	0.00
TKW_2021_2022_Reading	NDVI90_2021_2022_Reading	0.31	0.00
TKW_2021_2022_Reading	TT90_2021_2022_Reading	0.22	0.02
TKW_2021_2022_Reading	TT10_2021_2022_Reading	0.24	0.01
Yield_2021_2022_Reading	TT50_2021_2022_Reading	0.05	0.59
Yield_2021_2022_Reading	NDVI10_2021_2022_Reading	0.63	0.00
Yield_2021_2022_Reading	NDVI50_2021_2022_Reading	0.63	0.00
Yield_2021_2022_Reading	NDVI90_2021_2022_Reading	0.63	0.00
Yield_2021_2022_Reading	TT90_2021_2022_Reading	0.05	0.64
Yield_2021_2022_Reading	TT10_2021_2022_Reading	0.06	0.56
Yield_2021_2022_Reading	Max_NDVI_2021_2022_Reading	0.63	0.00
Yield_2020_2021_Reading	Max_NDVI_2020_2021_Reading	0.51	0.00
Yield_2020_2021_Reading	NDVI10_2020_2021_Reading	-0.13	0.17
Yield_2020_2021_Reading	NDVI50_2020_2021_Reading	0.04	0.43
Yield_2020_2021_Reading	NDVI90_2020_2021_Reading	0.45	7.72
Yield_2020_2021_Reading	TT10_2020_2021_Reading	0.18	0.05
Yield_2020_2021_Reading	TT50_2020_2021_Reading	0.21	0.03
Yield_2020_2021_Reading	TT90_2020_2021_Reading	0.14	0.13

#### 5.2.5.4. Correlations between GPD and the seven senescence metrics

The seven senescence metrics were negatively but weakly correlated ( $r=-0.15 - 0.08$ ) with GPD (Table 5.5.) in both years and the correlations were not statistically significant at the 5% threshold.

Table 5.5 Pearson correlations between GPD (column trait 1) and the senescence metrics in the Reading 2020-2021 and 2021-2022 trials (column trait 2).

<b>Trait1</b>	<b>Trait2</b>	<b>r</b>	<b>p</b>
GPD_2020_2021_Reading	TT10_Reading_2020_2021	-0.0102	0.916041
GPD_2020_2021_Reading	TT50_Reading_2020_2021	-0.0143	0.88178
GPD_2020_2021_Reading	TT90_Reading_2020_2021	-0.0083	0.931547
GPD_2020_2021_Reading	NDVI50_Reading_2020_2021	0.0345	0.720219
GPD_2020_2021_Reading	NDVI10_Reading_2020_2021	0.0793	0.410074
GPD_2020_2021_Reading	NDVI90_Reading_2020_2021	-0.0731	0.447763
GPD_2020_2021_Reading	Max_NDVI_2020_2021_Reading	-0.0962	0.328729
GPD_2021_2022_Reading	TT10_Reading_2021_2022	-0.1473	0.124581
GPD_2021_2022_Reading	TT50_Reading_2021_2022	-0.1501	0.117636
GPD_2021_2022_Reading	TT90_Reading_2021_2022	-0.1523	0.112126
GPD_2021_2022_Reading	NDVI50_Reading_2021_2022	-0.0116	0.904227
GPD_2021_2022_Reading	NDVI10_Reading_2021_2022	-0.0096	0.920608
GPD_2021_2022_Reading	NDVI90_Reading_2021_2022	-0.0136	0.887542
GPD_2021_2022_Reading	Max_NDVI_2021_2022_Reading	-0.0088	0.928618

#### 5.2.6. Single environment QTL analysis with the seven senescence metrics

##### 5.2.6.1 Description of the senescence QTL

A total of 19 QTL were detected for all senescence traits except for the maximum rate of senescence (Table 5.6.). More QTL were detected in the Reading 2020-2021 trial (12) in comparison to the Reading 2021-2022 trial (7). The QTL were distributed on every chromosome (1-7) but were absent of the D genome. Both Hereward and Malacca carried increasing alleles (noted A and B, respectively in table 5.6.) for thermal time or NDVI parameters.

Hereward carried six increasing alleles for thermal time located on chromosomes 1A, 3B, and 5B with small additive effects comprised between 0.59 and 9.8 degree days. For example, the Hereward allele at the QTL on chromosome 5B is conferring 9.8 extra degree days on the senescence duration than the Malacca allele, that is an increase of less than a day.

Malacca carried two increasing alleles for TT50 on chromosomes 3A and 7A with additive effects of 0.01 and 5.88 degree days.

11 QTL were detected for NDVI metrics, but their additive effects were very small (0.001-0.006) in comparison to the variations of NDVI observed in the DH which occur at the hundredth (e.g., 0.51-0.58 for NDVI10). Therefore, the practical significance of NDVI QTL is doubtful.

**Table 5.6.** Single environment QTL table showing QTL associated with the maximum rate of senescence (Max\_Rate), the thermal times to achieve different NDVI score (TT90, TT50, and TT10) and the NDVI value at different stage of senescence (NDVI90, NDVI50, and NDVI10). The letter A corresponds to the Malacca favourable allele whereas the letter B corresponds to the Hereward favourable allele

Chr	Position	LOD	%variance	additive effect	Location	Year	Allele	Trait	CI Start	CI end
1A	55.57	3	11.2	0.593	Reading	2021-2022	B	TT50	49.45	57.78
3A	69.89	2.4	8.8	-0.015	Reading	2021-2022	A	TT50	45.69	70.76
7A	47.4	3.243	13.496	-5.885	Reading	2020-2021	A	TT50	40.53	48.87
2B	16.27	5.907	23.21	-0.003	Reading	2021-2022	A	NDVI10	8.55	18.83
4B	55.48	3.6	11.1	0.004	Reading	2020-2021	B	NDVI10	50.62	55.48
5B	67.38	6.3	20.7	0.006	Reading	2020-2021	B	NDVI10	56.15	69.93
6B	114.34	2.1	6.3	-0.003	Reading	2020-2021	A	NDVI10	112.29	116.9
2B	16.27	5.916	23.242	-0.003	Reading	2021-2022	A	NDVI50	8.55	18.83
5A	116	5.8	18.1	0.004	Reading	2020-2021	B	NDVI50	113.67	123.82
5B	67.38	5.4	16.5	0.004	Reading	2020-2021	B	NDVI50	13.87	69.93
6B	36.81	1.8	5.1	-0.001	Reading	2020-2021	A	NDVI50	31.12	40.73
2A	112.47	2.4	8.8	0.003	Reading	2020-2021	B	NDVI90	31.83	117.32
2B	16.27	5.904	23.2	-0.003	Reading	2021-2022	A	NDVI90	8.55	18.83
5A	116	2.6	9.6	0.003	Reading	2020-2021	B	NDVI90	108.78	117.95
1A	55.57	3.333	13.845	1.234	Reading	2021-2022	B	TT90	48.58	61.46
3B	155.27	3.807	15.653	7.785	Reading	2020-2021	B	TT90	147.2	159.74
1A	55.57	3.805	15.646	1.531	Reading	2021-2022	B	TT10	48.58	61.46
3B	4	4.6	14.6	2.481	Reading	2020-2021	B	TT10	0	8.16
5B	52.67	7.4	25.4	9.827	Reading	2020-2021	B	TT10	49.01	62.43

#### 5.2.6.2. Co-locations of QTL for thermal metrics and NDVI metrics with GPC, GPC, GY, and HD.

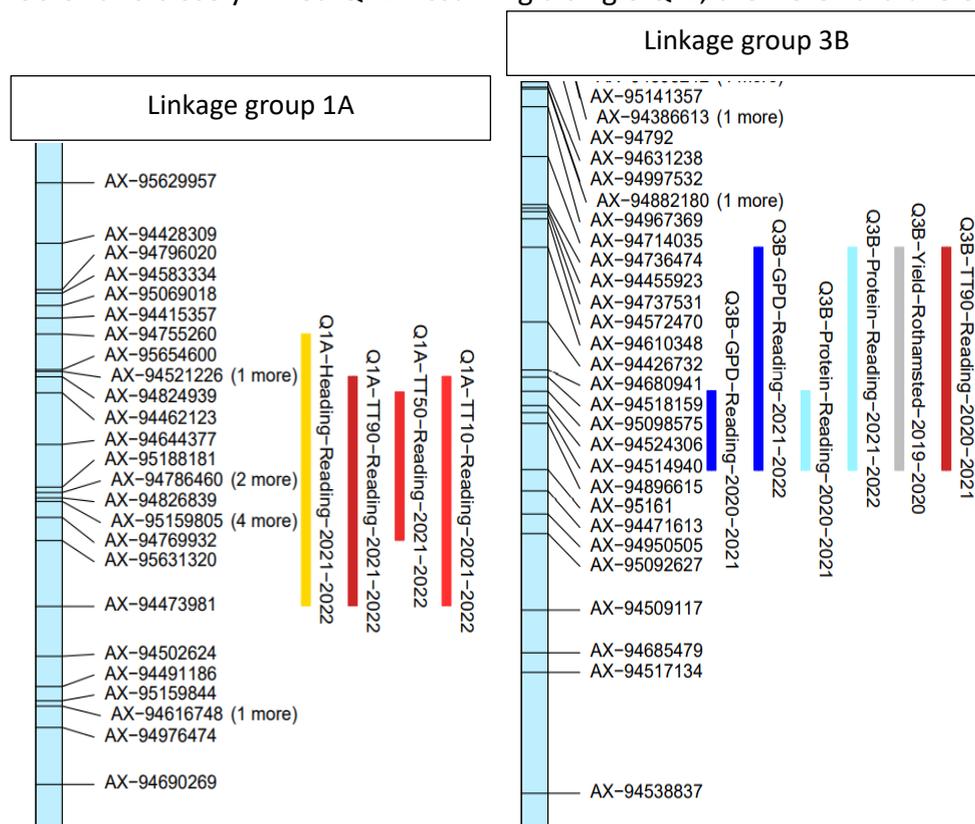
The thermal time and NDVI metrics QTL confidence intervals were positioned on the Malacca x Hereward genetic map together with QTL for GPD, GY, GPC, and HD to reveal co-locations. The GPC and GY QTL confidence intervals of linkage group 3B previously described in Chapter 4 (Figure 4.7) are coloured in light blue and grey, respectively (Figure 5.9., right). Two multiple co-locations were observed on linkage groups 1A and 3B (Figure 5.9).

On the linkage group 1A, the thermal time metrics (TT90, TT50, and TT10) QTL confidence intervals identified in the Reading 2021-2022 trial co-located with a QTL confidence interval for HD detected in the same trial (Figure 5.9., left). The size of the QTL confidence intervals

ranged between 8cM (TT50) and 15cM (HD). The overlap in the CI indicates that this region is likely to consist of one QTL located between 46.2 and 61.46 cM with pleiotropic effects on senescence timing and HD. This genomic region is associated with an earlier DH (-0.4 day), a delayed onset of senescence (+1.2 degrees-day) and an increased duration (+1.5 degree days) of senescence. These co-locations and the effects on thermal time metrics and HD QTL are consistent with the strong negative correlations between the traits reported in section 5.5.1. On the linkage group 3B (Figure 5.9, right), three QTL confidence intervals corresponding to GPD, GPC and TT90 traits were identified in the Reading 2020-2021 trial. The CI of the GPC and GPD QTL were smaller (4cM) in size than the one for TT90 (12cM).

The GPC-GPD CI were all included in the TT90 confidence interval and the 4cM overlap between the three CI was located between 155-159cM on the chromosome 3B (Figure 5.9, right).

This overlap may suggest the presence of either a single QTL with pleiotropic effects on GPD, GPC and TT90 or two closely linked QTL. Assuming a single QTL, the Hereward allele for this



**Figure 5.9.** Co-locations of single-environment QTL on the Malacca x Hereward genetic map observed on linkage groups 1A and 3B. The 95% CI of the QTL are represented by coloured rectangles: dark blue (GPD), light blue (GPC), grey (GY), yellow (HD), red gradient in increasing brightness (TT90, TT50, and TT10).

QTL would increase GPD, GPC, and TT90 by +0.12 % protein. m<sup>2</sup>. g<sup>-1</sup>, +0.17 % protein and +2.5 degree days, respectively in comparison to the Malacca allele.

#### 5.2.7. Principal component analysis (PCA) on the seven senescence metrics

A PCA was performed to determine whether a particular senescence pattern was associated with high GPD, GPC and GY.

The PCA was plotted using the seven senescence metrics (TT90, TT50, TT10, Max\_NDVI, NDVI90, NDVI50, NDVI10) set as active variables (i.e., contributing to the principal component axis) together with three additional qualitative variables (GPD, GPC, and GY) set as supplementary (i.e., no contribution to the principal components)

In the Reading 2020-2021 trial (Figure 5.10., a), the first principal component (PC1) explained 34.65% of the variance of the data and the second principal component (PC2) accounted for 30.07% of the variance of the data totalling to 65%.

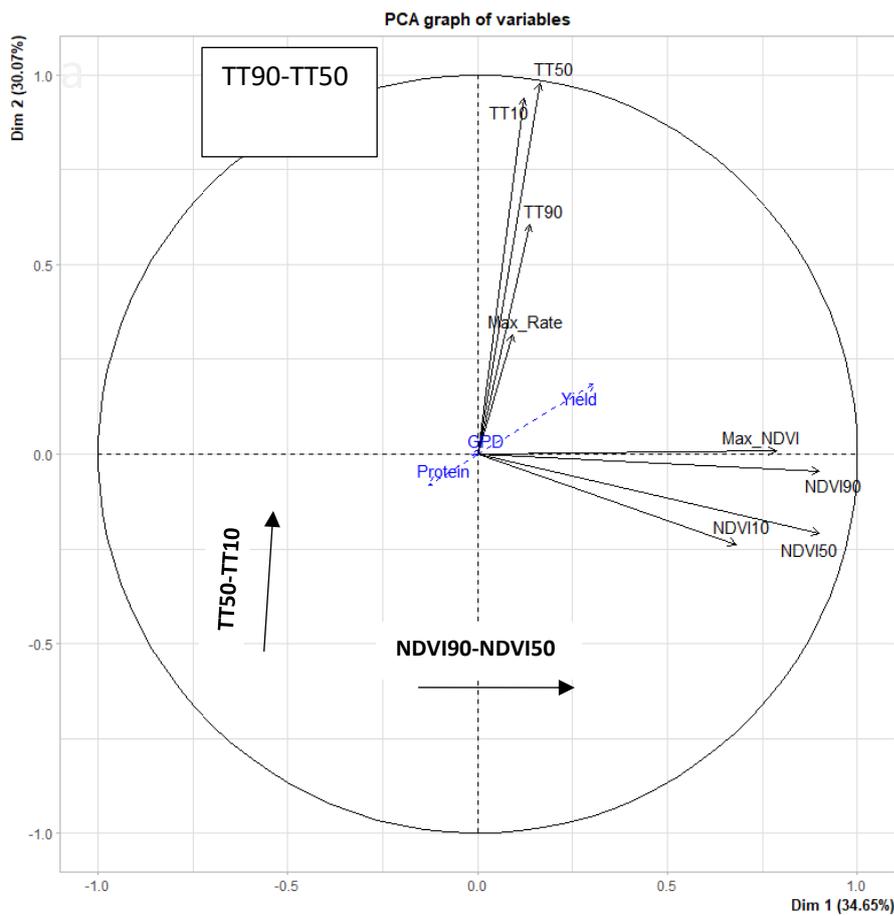
The principal components have the following equations:

$$PC1 = 0.90 \text{ NDVI50} + 0.90 \text{ NDVI90} + 0.78 \text{ Max\_NDVI} + 0.68 \text{ NDVI10}$$

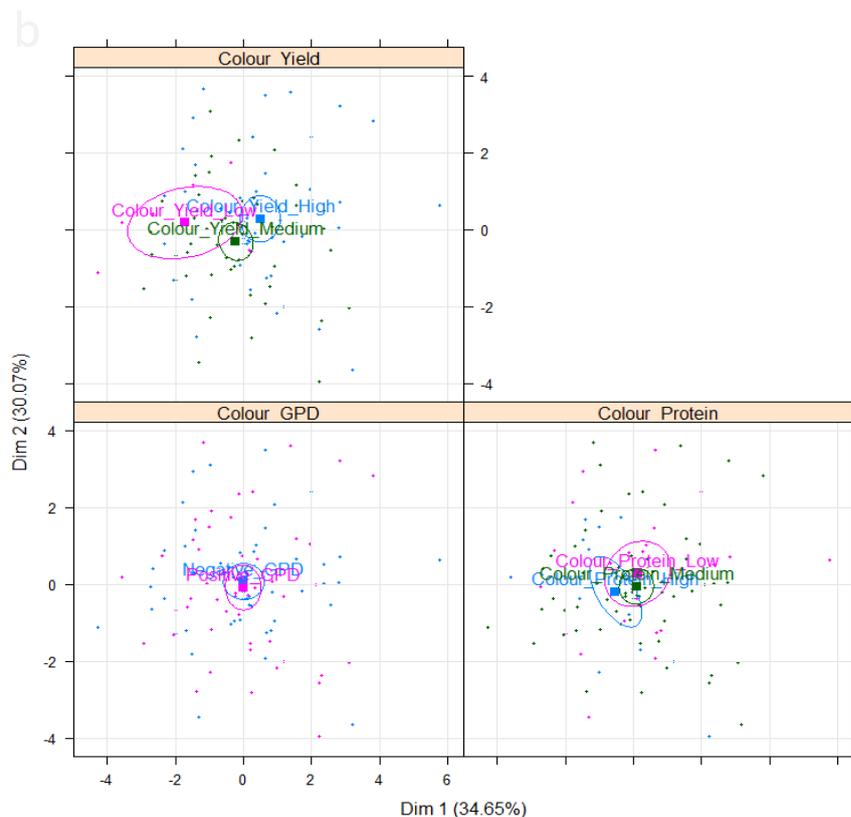
$$PC2 = 0.97 \text{ TT50} + 0.93 \text{ TT10} + 0.60 \text{ TT90} + 0.31 \text{ Max\_Rate} - 0.20 \text{ NDVI50} - 0.23 \text{ NDVI10}$$

For PC1, the PCA loadings of NDVI50 and NDVI90 (0.90 for both) and the magnitudes and directions of their vector (Figure 5.10., a) indicate that they are major contributors to this axis. TT50 and TT10 are the variable having the most important leverage on PC2 (loadings of 0.97 and 0.93, respectively). As a result, this axis differentiates the lines according to their thermal time value at mid senescence and at the end of senescence.

The PC1 and PC2 axis were therefore named “NDVI early-mid senescence (NDVI90-50)” and “thermal time mid-late senescence (TT50-10)”, respectively (Figure 5.10., a).



**Figure 5.10. Principal component analysis plots on the senescence metrics in Reading 2020-2021.** a: the variable plot showing the correlation between the variables and their contributions to the PCA axis (Dim 1 and Dim 2). The black vectors are the active variables, and the blue vectors are the additional variables.



b: the individual plots grouping the lines according to yield (Low: 601-685 g.m<sup>-2</sup>, medium: 700-799 g.m<sup>-2</sup>, high: 801-901 g.m<sup>-2</sup>), GPD (Positive, and negative), and Protein (low: 10.2-10.74%, medium: 10.76-11.52%, and high: 11.54-12.33%). The ellipses are the 95% confidence level for the categories.

The plot of individuals (Figure 5.10, b) with grouping according to yield shows that the three ellipses overlap slightly. The low yielding lines (601 - 685 g.m<sup>-2</sup>) centroid (Figure 5.10, b, pink square) is on the left of the y axis in an area corresponding to low NDVI90-50 values whereas the high yielding lines (801-901 g.m<sup>-2</sup>) centroid is located on the top right in an area corresponding to high NDVI90-50 and high TT50-TT10 values. The medium yielding lines (700-800 g.m<sup>-2</sup>) centroid is located toward the centre of the biplot (average NDVI90-50 and TT50-TT10).

The biplot indicates that high yielding lines take longer to senesce than the other (medium and low yielding lines) and have also higher NDVI values at the onset and mid senescence.

When grouped according to protein and GPD, there were no clear separations of the ellipses for the three protein groups (low:10.2-10.74%, medium: 10.76-11.52%, and high: 11.54-12.33%) and for the two GPD groups (negative or positive GPD) (Figure 5.10, b) indicating that there is no typical senescence pattern associated with a high GPC or GPD.

In the Reading 2021-2022 trial (Figure 5.11., a), the first principal component (PC1) explained 58% of the variance of the data and the second principal component (PC2) accounted for 42% of the variance of the data.

The principal components have the following equations:

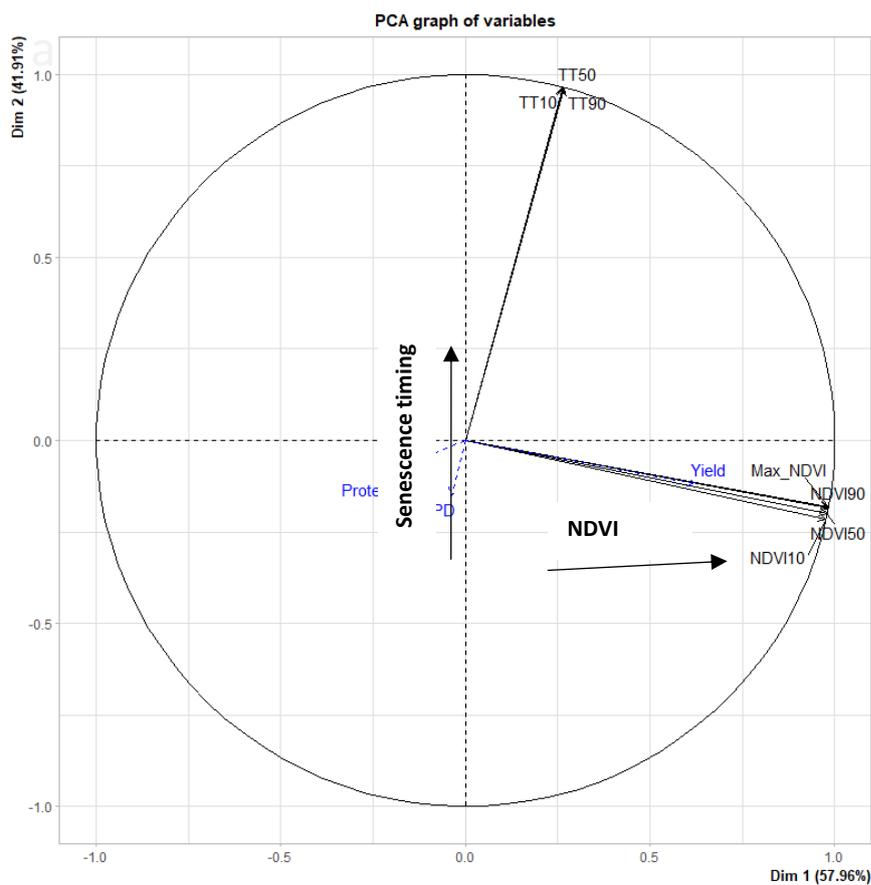
$$PC1 = 0.98 \text{ Max\_NDVI} + 0.98 \text{ NDVI90} + 0.97 \text{ NDVI50} + 0.97 \text{ NDVI10} + 0.26 \text{ TT90} + 0.26 \text{ TT50} + 0.26 \text{ TT10}$$

$$PC2 = 0.96 \text{ TT50} + 0.96 \text{ TT10} + 0.96 \text{ TT90} - 0.20 \text{ NDVI50} - 0.21 \text{ NDVI10}$$

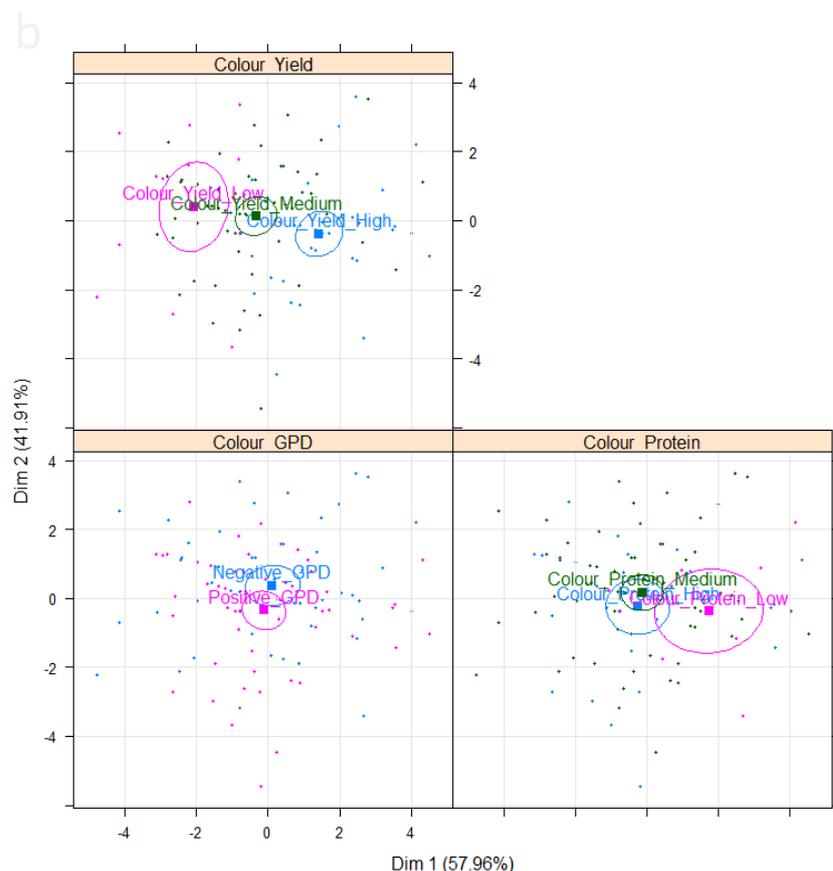
The four NDVI parameters contributed equally and significantly to PC1 (loadings of 0.97-0.98) and therefore this axis was used to separate the lines with high NDVI values (at all senescence times: TT90-TT50-TT10) from the lines with low NDVI values at all senescence times.

The three thermal time metrics contributed equally and significantly to PC2 (loadings of 0.96) and therefore this axis was used to separate the lines with high TT90, TT50, and TT10 from lines with low TT90, TT50, and TT10.

Hence, PC1 was named “NDVI value” and PC2 “senescence timing” (Figure 5.11, a).



**Figure 5.11. Principal component analysis plots with the senescence metrics in the Reading 2021-2022 trial.** a: the variable plot showing the correlations between the variables and their contributions to the PCA axis (Dim 1 and Dim 2). The black vectors are the active variables, and the blue vectors are the additional variables.



b: the individual plots grouping the lines according to yield (Low: 744-899  $\text{g}\cdot\text{m}^{-2}$ , medium: 902-1046  $\text{g}\cdot\text{m}^{-2}$ , high: 1054-1200  $\text{g}\cdot\text{m}^{-2}$ ), GPD (Positive, and negative), and Protein (Low: 9.40-10.40%, medium: 10.41-11.40%, and high: 11.45-12.44%). The ellipses are the 95% confidence level for the categories.

Figure 5.11.b shows three plots of individuals for GY: low (744-898 g.m<sup>-2</sup>), medium (902-1045 g.m<sup>-2</sup>), and high (1054-1200 g.m<sup>-2</sup>), GPD: positive and negative GPD; and GPC: low (9.41-10.41%), medium (10.41-11.40%), high (11.41-12.45%).

On the yield individual plot, the three categorial ellipses are clearly separated with high yielding lines located on the bottom right corner of the biplot in an area corresponding to high NDVI values and rapid senescence and low yielding lines situated on the top left corner; an area corresponding to low NDVI and slow and delayed senescence.

The two individual graphs for GPD and GPC show partial overlapping between the ellipses and no clear separation between the groups. It can be concluded that these two variables are not influenced by any specific senescence pattern.

### 5.3. Discussion and Conclusions

#### 5.3.1 Variation in the length of the crop cycle between the Reading 2020-2021 and Reading 2021-2022 trials

In the Reading 2021-2022 trial, the DH completed the growth cycle a month earlier than in the Reading 2020-2021 trial. The description of the weather data in part 5.2.1. shows that the Reading 2021-2022 trial received roughly three times as much solar radiation in May than the Reading 2020-2021 trial (Figure 5.1.).

Field and glasshouse experiments found that temperature influences the rate of plant development which occur faster at high temperatures than at low temperatures. For instance, at the onset of the wheat crop cycle, leaf initiation and development rate increase linearly over a range of temperatures from 0 to 15°C. Similarly, shoot elongation was shown to increase linearly by 1.5 mm.°C<sup>-1</sup>d<sup>-1</sup> for temperatures varying between 3 and 20.4°C (Porter and Gawith, 1999; Slafer and Rawson, 1995).

Therefore, the high solar radiation in May combined with the slightly higher temperatures during the crop season in the Reading 2021-2022 trial are the likely cause of this shorter crop development.

### 5.3.2. Senescence metrics

The parental lines Malacca and Hereward did not differ for any of the thermal time metrics (TT90, TT50, TT10) which suggests that they follow a similar pattern of senescence from onset to completion. The parents also had comparable NDVI90, NDVI50, and NDVI10 values which may indicate similar contents of nitrogen or chlorophyll at onset, mid and senescence termination (Pettorelli *et al.*, 2013, pp. 70-80). This may explain why the DH lines showed only small variation for the seven senescence metrics.

The Pearson correlations between the thermal time metrics (TT90, TT50, and TT10) in the two environments: Reading 2020-2021 and Reading 2021-2022 were significant and of medium strength ( $r=0.34-0.42$ ) supporting that the senescence dynamics are partly under genetic control. By contrast, the NDVI metrics (NDVI 90, NDVI 50, and NDVI10) were either significantly and weakly correlated ( $r=0.15-0.23$ ) or not significantly correlated in the two environments. This demonstrates that NDVI metrics are more strongly influenced by the environment and/or G x E interactions.

### 5.3.3. Determination of correlations

#### 5.3.3.1. Correlations of the DH heading dates with the GPC, GY, and GPD

The variable heading date was highly negatively correlated with the thermal time metrics in both environments but not significantly correlated with NDVI, GPC, GY, and GPD (Table 5.2.)

The negative correlations between the DH heading dates and the thermal time metrics indicate that lines with a late heading date senesce earlier (lower TT90) and/or over a shorter period (lower TT10) than lines with an early heading date. However, this shorter duration of senescence was not accompanied with a higher rate of senescence in 2020-2021 as the correlation between HD and Max\_Rate was not significant (Table 5.2.)

Surprisingly, the correlations between HD and GPC, and GPY were not significant in both environments which suggests that HD which was here associated with early senescence and/or a shorter senescence duration without increase of the senescence rate (variable Max\_Rate) in Reading 2020-2021 did not affect GY or GPC.

### 5.3.3.2. Correlations of senescence parameters with GPC

Some authors have shown a positive relationship between NDVI and chlorophyll content although the shape of the relationship (linear vs non-linear) remains unclear (Pettorelli *et al.*, 2013, pp. 70-80).

In my study, GPC was weakly and negatively correlated with NDVI metrics in both environments (Table 5.3). This implies that lines with high GPC also have low NDVI values, and conversely. One hypothesis to explain the negative correlations between GPC and NDVI metrics is that DH lines with low NDVI values (all thermal times considered) may have higher chlorophyll degradation and partitioning of chlorophyll-derived nitrogen into the grain.

GPC was not significantly correlated with the onset of senescence (TT90) in both environments within the DH population. This indicates that early senescence is not associated with a high GPC in the DH population.

This contrasts with Uauy *et al.*, (2006) who found that early senescence of 4-5 days was associated with an increase of GPC of 10-15% in wild emmer (*Triticum turgidum* ssp. *dicoccoides*) having the wild type allele (*TtNAM-B1*) in comparison to the durum wheat cultivar Langdon (*Triticum turgidum* ssp. *durum*).

In the Reading 2021-2022 trial, the variation of TT90 within the DH population was only one day and therefore, may not have been sufficient to generate a difference in GPC in comparison to the 4-5 days reported in Uauy *et al.* (2006). However, in the Reading 2020-2021 trial, the variation of senescence onset (TT90) between the DH lines was 5-6 days. This suggests that earlier chlorophyll degradation did not result in a greater nitrogen partitioning into the grain.

### 5.3.3.3. Correlations of senescence parameters with GY

In my study, GY was not significantly correlated with the onset of senescence (TT90) in the two Reading environments (Table 5.4.). This contrasts with Christopher *et al.* (2014) who found positive correlations ( $r > 0.40$ ) between GY and TT90 in three environments for a population of 184 DH lines. Borrill *et al.* (2015) reported that positive correlations between GY and a delayed onset of senescence are mainly observed in stress environments such as drought, heat, or low nitrogen inputs. However, under optimal growth conditions GY would be limited by its sink capacity with excess photoassimilates being directly stored in the stem.

Consequently, the absence of correlations between GY and TT90 in this study may be the consequence of sink saturation.

Another hypothesis would be non-functional stay greens or cosmetic stay-greens which retain green leaves, but do not photosynthesize because of an impairment in their chlorophyll catabolism pathway (Gregersen *et al.*, 2013). As a result, these stay green types may show delayed onset of senescence (TT90) without increase in GY.

In contrast to GY, which is determined by grain weight and number, TKW is determined only by grain weight.

In this study, TKW was weakly positively correlated with the onset of senescence (TT90) in the Reading 2021-2022 trial ( $r = 0.22$ ,  $p = 0.02$ ). The TKW was not measured in Reading 2020-2021 and therefore the correlation between TKW and TT90 in this environment could not be calculated. The positive correlation signifies that lines which exhibit a delay in the onset of senescence are associated with high TKW (grain weight). A possible explanation supported by authors such as Bogard *et al.*, (2011) and Cormier *et al.*, (2016) would be that a prolonged photosynthesis (i.e., delayed senescence) results in a greater carbon assimilation and therefore sugar accumulation in the grain. The positive correlation between TKW and TT90 in 2021-2022 suggests that late senescing lines continue to photosynthesize and that the sink capacity of the grain is not saturated.

The duration of senescence (TT10) was significantly positively correlated with GY in Reading 2020-2021 ( $r=0.18$ ,  $p = 0.05$ ) and with TKW in Reading 2021-2022 ( $r=0.24$ ,  $p=0.01$ ). This disagrees with Christopher *et al.* (2014) who did not find significant correlations between GY and TT10 in three environments.

#### 5.3.3.4. Correlations of senescence parameters with GPD

None of the seven senescence metrics studied were correlated with GPD in either environment. This was confirmed by the PCA biplots (figures 5.10, b and 5.11, b.) where no clear separations between lines with positive and negative GPD were observed.

As NDVI is correlated to the chlorophyll content, this shows that the contribution of positive GPD to GPC in the DH population does not come from a higher remobilisation of chlorophyll-derived nitrogen into the grain. Instead, positive GPD in this study could be caused by a higher post-anthesis nitrogen uptake from the soil or a greater remobilisation of non-chlorophyll N

in inorganic form (urea) or organic form (acid nucleic) (Bogard *et al.*, 2011; Cormier *et al.*, 2016).

#### 5.3.4. QTL analysis

Although Malacca and Hereward shared the same senescence pattern and that little variation in senescence traits was observed within the DH population, 19 QTL were mapped for both years.

The interpretation of the thermal time and NDVI QTL effects is complex as these traits are not direct breeding targets but have both been shown to correlate with GPC and GY.

In my study, the NDVI metrics were either only weakly negatively correlated or not correlated with GPC (Table 5.3.) but were strongly positively correlated with GY (Table 5.4.) suggesting that QTL with positive additive effects on NDVI metrics may be associated with an increase in GY. By contrast, the thermal time metrics were not correlated with GPC nor GY except for the metric TT50 which was positively correlated with GY in the Reading 2020-2021 trial. This indicates that selection for the thermal time metrics (TT10 and TT90) is unlikely to result in increases in either GPC or GY.

Despite the absence of correlations between thermal time and GPC or GPD, a multi-trait co-location GPC-GPD-TT90 was identified on chromosome 3B (Figure 5.9.). This shows that although two traits may not be correlated globally, co-locations and correlations can be revealed after they are decomposed into genomic regions. However, the TT90 genomic region on chromosome 3B (Figure 5.9.) was not detected in the Reading 2021-2022 trial. It is therefore not possible to establish a link between the onset of senescence (TT90) and GPC-GPD as only one co-location was detected. More environments would be needed, and additional colocations of GPD-GPC-TT90 would have to be mapped to establish a link.

### 5.4 Limitations to study

#### 5.4.1. Scoring of GS 55 in the Reading 2021-2022

In 2021-2022, many of the DH lines had already reached heading (GS 55) when the scoring started. This was partly because Reading was warmer in 2021-2022 (Chapter 5, 5.2) than the previous year. Warm temperatures were reported to boost crop development, especially the production of biomass. Therefore, the superior average temperature in Reading 2021-2022 may have accelerated the crop development and caused the DH to reach GS 55 earlier than

the previous field season which was taken as a reference to start the observations. Consequently, some HD were estimated at the start of recording. The quasi-linear negative correlations observed ( $r > 0.99$ ) between HD and thermal time metrics in 2021-2022 may therefore result from inaccurate estimation of some HD.

#### 5.4.2. Insensitivity of NDVI to medium-high chlorophyll content

NDVI is not able to discriminate between high and medium chlorophyll content (Pettorelli *et al.*, 2013, pp. 70-80), which corresponds in that study to NDVI measured before senescence (Max-NDVI) and NDVI measured at onset of senescence (NDVI90). Therefore, the size of correlations of Max\_NDVI and NDVI90 with GPC, and GY might be biased.

Alternatively, to improve quantification of chlorophyll content, the NDVI measurements at the onset of senescence could be replaced or corrected by direct measurement of chlorophyll content.

## Chapter 6. General discussion and future work

Despite being mainly influenced by the HMW-GS composition at the *Glu-1* loci, 23-45% of the variation of breadmaking quality is still not fully explained (Payne et al., 1987). Therefore, millers often have to adjust the quality of their flour which is time consuming and costly. Identifying new alleles associated with breadmaking quality and providing breeders with markers for marker assisted selection will allow them to screen for quality at early generations of breeding.

QTL mapping offers a powerful tool to dissect the genetic architecture of complex traits (e.g. breadmaking quality traits) and has successfully being used in wheat to identify QTL associated with grain yield, grain protein content and many other traits (Nigro *et al.*, 2019).

In Chapter 3, the individual effects of six QTL on rheology and baking parameters were investigated in six NILs over two field seasons. The quality tests which include milling, rheological measurements, and baking were all conducted at the bakery Heygates Ltd. which supported this project.

The effect the increasing allele 4D-2a on the crumb whiteness was confirmed in the Rothamsted 2019-2020 field trial but not in the Rothamsted 2020-2021 field trial and its effect was of similar magnitude than in the DH. This confirms hypothesis 1 “will the QTL effects be significant in the NILs backgrounds?” and 3: “will the QTL effects be of the same magnitude as in the double haploid population?” of the introduction for this QTL. However, hypothesis 2: “will the QTL effect be stable (significant) in a least two environment?” is rejected because the QTL 4D-2 effect on the crumb whiteness was not significant in Rothamsted 2020-2021.

A blast of the QTL 4D-2 peak marker AX-94454183 against RefSeq v1.0 identified the Lipoxygenase gene TRAESCS4D02G294100 as a candidate gene. This was confirmed by finding high level of expression for TRAESCS4D02G294100 in the wheat germ and seed coat and poor level (but present) in the endosperm using the expVIP database .

QTL 1B, 2D, 4D-1, 6A, and 7B effects on their respective quality parameters were not significant in the Rothamsted 2019-2020 and 2020-2021 field trials invalidating hypothesis 1, 2, and 3 for those QTL.

Confirming QTL effects on bread quality is challenging because of the experimental error that accumulates during the breadmaking process (i.e. milling, mixing, proving, baking), even though standard procedures were used in this study to minimize it. Should the QTL be re-

assessed, more replication would be needed to refine the estimation of the QTL effects. This could be achieved by increasing the plots size to increase the amount of flour retrieved after milling.

Although the whiteness of bread crumb prepared from NIL 4D-2a was not found to differ significantly from NIL 4D-2b in Rothamsted 2020-2021, it would be worth redoing the comparisons in other environments as the environment alone and its interaction with the genetic can influence the significance of the QTL effect. Ultimately, after confirmation of the QTL 4D-2 effect in at least two environments, fine mapping could be done to decrease the size QTL interval. Afterwards, targeted gene expression analysis can be carried out to find candidate genes whose levels of expression differ significantly between the allelic pairs.

To find an alternative way to the use of nitrogen fertilisers, this thesis dissected the trait grain protein deviation (GPD) within a DH haploid population grown under conventional farming practices (i.e. 200 kg.N.ha<sup>-1</sup>). The DH population was grown in three environments to find stable QTL for GPD, GPC, and GY.

Chapter 4 describes the 3-step method that was followed from the calculation of the GPD individual values to the QTL mapping including:

- 1- The computing of simple GPC-GY linear regressions to calculate GPD individual values
- 2- The mixed models used to retrieve the Best Linear Unbiased Estimators
- 3- The complex interval mapping (CIM) QTL mapping procedure

Many individual observations on the GPC-GY simple linear regressions had a GPD value higher than Hereward but since these are individual values and not means and that no statistical test was used to assess the significance of the differences, it was not possible to conclude to transgressive segregation and to answer to hypothesis 1: "Is there evidence of transgressive segregation for GPD in the double haploid population?". To this end, multiple testing may be carried out to assess the significance of the differences of the GPD means between each of the 109 DH lines and Hereward separately. This will allow to answer hypothesis 1.

For all the traits, the great majority of QTL were environment specific and located over a wide range of chromosomes. Three GPD-GPC QTL co-locations were found suggesting one single genomic region acting on both GPD and GPC. However, hypothesis 2: "are there any pleiotropic QTL?" is partly validated as the possibility of two closely related regions acting

separately on GPC and GPD cannot be ruled out. The three GPD-GPC co-locations will need to be confirmed in additional environments before drawing more firm conclusions.

On chromosome 3B, a genomic region in linkage with GPD detected in Reading 2020-2021 partially overlapped with a genomic region in linkage with GPD detected in Reading 2021-2022. This overlap may be attributed to either a single gene with a stable effect in the environments measured or to two closely linked genes; one with a significant effect on GPD in Reading 2020-2021 and another with a significant effect on GPD in Reading 2021-2022.

Therefore, at this stage of the analysis, hypothesis 3: “are there any stable GPD QTL?” remains inconclusive.

The course of monocarpic senescence was also monitored and analysed in the aim of exploring correlations between senescence metrics and GPD which would shed light on the physiology of the GPD.

The senescence progress was monitored within the same DH population and field trials (Reading 2020-2021 and Reading 2021-2022) that were used to dissect the genetic architecture of GPD.

Chapter 5 describes the methodology followed to convert the temperatures after heading in degree days, to model the senescence and to derive the senescence metrics from the senescence curves.

None of the seven senescence metrics were significantly correlated with GPD. Furthermore, no specific senescence pattern when the seven metrics were included in a PCA analysis could be related to a positive GPD value. Hypothesis 1: “Is there any link between the senescence and GPD” is therefore rejected in this study.

Graphical analysis of six metrics (TT90, TT50, TT10, NDVI90, NDVI50, and NDVI10) using the LSD showed small variation among the 109 DH lines. The parents Malacca and Hereward means did not differ significantly for the six metrics based on the LSD calculations.

Despite of the small variation observed, 24 genomic regions were detected for all the traits in the two environments measured. This includes one GPD-GPC-TT90 co-location mapped on chromosome 3B in Reading 2020-2021. This finding was at first surprising since TT90 was not significantly correlated with GPD in this environment but seemed possible since the QTL mapping technique is studying the effects of individual genetic variations (SNP) on the phenotype rather than considering the phenotype as a whole. Hypothesis 2: “Are there any

co-locations between the confidence intervals of GPD and the seven senescence metrics?" is validated.

## Chapter 7. References

Abdalla, M., Hastings, A., Cheng, K., Yue, Q., Chadwick, D., Espenberg, M., Truu, J., Rees, R. M. and Smith, P. (2019) 'A critical review of the impacts of cover crops on nitrogen leaching, net greenhouse gas balance and crop productivity', *Global Change Biology*, 25, pp. 2530-2543.

Alqudah, A. M. and Schnurbusch, T. (2017) 'Heading date is not flowering time in spring barley', *Frontiers in Plant Science*, 8, pp. 896.

Altenbach, S. B., Kothari, K. M. and Lieu, D. (2002) 'Environmental conditions during wheat grain development alter temporal regulation of major gluten protein genes', *Cereal Chemistry*, 79, pp. 279-285.

Amjid, M. R., Shehzad, A., Hussain, S., Shabbir, M. A., Khan, M. R. and Shoaib, M. (2013) 'A comprehensive review on wheat flour dough rheology', *Pakistan Journal of Food Sciences*, 23, pp. 105-123.

Andersson, A. A., Andersson, R., Piironen, V., Lampi, A.-M., Nyström, L., Boros, D., Fraš, A., Gebruers, K., Courtin, C. M. and Delcour, J. A. (2013) 'Contents of dietary fibre components and their relation to associated bioactive components in whole grain wheat samples from the HEALTHGRAIN diversity screen', *Food Chemistry*, 136, pp. 1243-1248.

Bangur, R., Batey, I., Mckenzie, E. and Macritchie, F. (1997) 'Dependence of extensograph parameters on wheat protein composition measured by SE-HPLC', *Journal of Cereal Science*, 25, pp. 237-241.

Barron, C., Surget, A. and Rouau, X. (2007) 'Relative amounts of tissues in mature wheat (*Triticum aestivum* L.) grain and their carbohydrate and phenolic acid composition', *Journal of Cereal Science*, 45, pp. 88-96.

Basford, K. and Cooper, M. (1998) 'Genotype × environment interactions and some considerations of their implications for wheat breeding in Australia This review is one of a series commissioned by the Advisory Committee of the Journal', *Australian Journal of Agricultural Research*, 49, pp. 153-174.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015). 'Fitting linear mixed-effects models using lme4', *Journal of Statistical Software*, 67(1), pp. 1–48.

Bhalla, P. L., Singh, M. B. (2017) *Wheat Biotechnology*. New York: Humana Press.

Blumenthal, C. S., Barlow, E. W. R. and Wrigley, C. W. (1993) 'Growth environment and wheat quality: The effects of heat stress on dough properties and gluten proteins', *Journal of Cereal Science*, 18(1), pp.3-21.

Bogard, M., Allard, V., Brancourt-Hulmel, M., Heumez, E., Machet, J.-M., Jeuffroy, M.-H., Gate, P., Martre, P. and Gouis, J. L. (2010) 'Deviation from the grain protein concentration–grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat', *Journal of Experimental Botany*, 61(15), pp.4303-4312.

Bogard, M., Jourdan, M., Allard, V., Martre, P., Perretant, M. R., Ravel, C., Heumez, E., Orford, S., Snape, J. and Griffiths, S. (2011) 'Anthesis date mainly explained correlations between post-anthesis leaf senescence, grain yield, and grain protein concentration in a winter wheat population segregating for flowering time QTLs', *Journal of Experimental Botany*, 62, pp. 3621-3636.

Borrill, P., Fahy, B., Smith, A. M. and Uauy, C. (2015) 'Wheat grain filling is limited by grain filling capacity rather than the duration of flag leaf photosynthesis: a case study using NAM RNAi plants', *PLoS one*, 10, p. e0134947.

Borrill, P., Ramirez-Gonzalez, R. and Uauy, C. (2016) 'expVIP: a customizable RNA-seq data analysis and visualization platform', *Plant Physiology*, 170, pp. 2172-2186.

Branlard, G., Faye, A., Rhazi, L., Tahir, A., Lesage, V. and Aussenac, T. (2020) 'Genetic and environmental factors associated to glutenin polymer characteristics of wheat', *Foods*, 9, pp. 683.

Brinsden, H.C., He, F.J., Jenner, K.H. and MacGregor, G.A. (2013) 'Surveys of the salt content in UK bread: progress made and further reductions possible', *BMJ Open*, 3(6), p.e002936.

Broman, K. W., Wu, H., Sen, S. and Churchill, G. A. (2003) 'R/qtl: QTL mapping in experimental crosses', *Bioinformatics*, 19, pp. 889-890.

Buerstmayr, M., Steiner, B. and Buerstmayr, H. (2020) 'Breeding for Fusarium head blight resistance in wheat—Progress and challenges', *Plant Breeding*, 139(3), pp. 429-454.

Cauvain, S. P. and Young, L. S. (2006) *The Chorleywood bread process*. Boca Raton: Woodhead Publishing.

Chapman, E. A., Orford, S., Lage, J. and Griffiths, S. (2021) 'Capturing and selecting senescence variation in wheat', *Frontiers in Plant Science*, 12, pp. 638738.

Christopher, J. T., Veyradier, M., Borrell, A. K., Harvey, G., Fletcher, S. and Chenu, K. (2014) 'Phenotyping novel stay-green traits to capture genetic variation in senescence dynamics', *Functional Plant Biology*, 41, pp. 1035-1048.

Cormier, F., Foulkes, J., Hirel, B., Gouache, D., Moënné-Loccoz, Y. and Le Gouis, J. (2016) 'Breeding for increased nitrogen-use efficiency: a review for wheat (*T. aestivum* L.)', *Plant Breeding*, 135, pp. 255-278.

Cui, F., Zhao, C., Ding, A., Li, J., Wang, L., Li, X., Bao, Y., Li, J. and Wang, H. (2014) 'Construction of an integrative linkage map and QTL mapping of grain yield-related traits using three related wheat RIL populations', *Theoretical and Applied Genetics*, 127, pp. 659-675.

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Austine-Orimoloye, O., Azov, A. G., Barnes, I. and Bennett, R. (2022) 'Ensembl 2022', *Nucleic Acids Research*, 50, pp. 988-995.

Di, H. and Cameron, K. (2002) 'Nitrate leaching in temperate agroecosystems: sources, factors and mitigating strategies', *Nutrient Cycling in Agroecosystems*, 64, pp. 237-256.

Distelfeld, A., Avni, R. and Fischer, A. M. (2014) 'Senescence, nutrient remobilization, and yield in wheat and barley', *Journal of Experimental Botany*, 65, pp. 3783-3798.

Duncan, E. G., O'sullivan, C. A., Roper, M. M., Biggs, J. S. and Peoples, M. B. (2018) 'Influence of co-application of nitrogen with phosphorus, potassium and sulphur on the apparent efficiency of nitrogen fertiliser use, grain yield and protein content of wheat', *Field Crops Research*, 226, pp. 56-65.

Dupont, F. M. and Altenbach, S. B. (2003) 'Molecular and biochemical impacts of environmental factors on wheat grain development and protein synthesis', *Journal of Cereal Science*, 38, pp. 133-146.

Finney, K., F (1948) 'Loaf volume and protein content of hard winter and spring wheat', *Cereal Chemistry*, 25, pp. 291-312.

Fradgley, N., Gardner, K. A., Cockram, J., Elderfield, J., Hickey, J. M., Howell, P., Jackson, R. and Mackay, I. J. (2019) 'A large-scale pedigree resource of wheat reveals evidence for adaptation and selection by breeders', *PLoS Biology*, 17, pp. 3000071.

Gaju, O., Allard, V., Martre, P., Le Gouis, J., Moreau, D., Bogard, M., Hubbart, S. and Foulkes, M. J. (2014) 'Nitrogen partitioning and remobilization in relation to leaf senescence, grain yield and grain nitrogen concentration in wheat cultivars', *Field Crops Research*, 155, pp. 213-223.

Gebruers, K., Dornez, E., Boros, D., Dynkowska, W., Bedő, Z., Rakszegi, M., Delcour, J. A. and Courtin, C. M. (2008) 'Variation in the content of dietary fiber and components thereof in wheats in the HEALTHGRAIN diversity screen', *Journal of Agricultural and Food Chemistry*, 56, pp. 9740-9749.

Geyer, M., Mohler, V. and Hartl, L. (2022) 'Genetics of the inverse relationship between grain yield and grain protein content in common wheat', *Plants*, 11, pp. 2146.

- Giraldo, P., Benavente, E., Manzano-Agugliaro, F. and Gimenez, E. (2019) 'Worldwide research trends on wheat and barley: bibliometric comparative analysis', *Agronomy*, 9(7), p.352.
- Glass, A. D. (2003) 'Nitrogen use efficiency of crop plants: physiological constraints upon nitrogen absorption', *Critical Reviews in Plant Sciences*, 22, pp. 453-470.
- Godfrey, D., Hawkesford, M. J., Powers, S. J., Millar, S. and Shewry, P. R. (2010) 'Effects of crop nutrition on wheat grain composition and end use quality', *Journal of Agricultural and Food Chemistry*, 58, pp. 3012-3021.
- Gonzalez-Thuillier, I., Salt, L., Chope, G., Penson, S., Skeggs, P., Tosi, P., Powers, S. J., Ward, J. L., Wilde, P. and Shewry, P. R. (2015) 'Distribution of lipids in the grain of wheat (cv. Hereward) determined by lipidomic analysis of milling and pearling fractions', *Journal of Agricultural and Food Chemistry*, 63, pp. 10705-10716.
- Gooding, M. J. and Shewry, P. R. (2022) *Wheat Environment, Food and Health*. Chichester: Wiley.
- Gregersen, P., Holm, P. and Krupinska, K. (2008) 'Leaf senescence and nutrient remobilisation in barley and wheat', *Plant Biology*, 10, pp. 37-49.
- Gregersen, P. L., Culetic, A., Boschian, L. and Krupinska, K. (2013) 'Plant senescence and crop productivity', *Plant Molecular Biology*, 82, pp. 603-622.
- Harper, L., Sharpe, R., Langdale, G. and Giddens, J. (1987) 'Nitrogen cycling in a wheat crop: soil, plant, and aerial nitrogen transport 1', *Agronomy Journal*, 79, pp. 965-973.
- Havé, M., Marmagne, A., Chardon, F. and Masclaux-Daubresse, C. (2017) 'Nitrogen remobilization during leaf senescence: lessons from Arabidopsis to crops', *Journal of Experimental Botany*, 68, pp. 2513-2529.
- Hawkesford, M.J., Kopriva, S. and De Kok, L.J., (2016) *Nutrient use efficiency in plants*. Springer.
- He, H. and Hosney, R. (1992) 'Effect of quantity of wheat flour protein on bread loaf volume', *Cereal Chemistry*, 69, pp. 17-19.
- Johansson, E., Prieto-Linde, M. L. and Jönsson, J. Ö. (2001) 'Effects of wheat cultivar and nitrogen application on storage protein composition and breadmaking quality', *Cereal Chemistry*, 78, pp. 19-25.
- Khan, K. and Shewry, P. (2009) *Wheat Chemistry and Technology. 4th edn*. St. Paul: Cereals and Grains Association.
- Langridge, P., Alaux, M., Almeida, N.F., Ammar, K., Baum, M., Bekkaoui, F., Bentley, A.R., Beres, B.L., Berger, B., Braun, H.J. and Brown-Guedira, G. (2022). 'Meeting the challenges facing wheat production: The strategic research agenda of the Global Wheat Initiative', *Agronomy*, 12(11), p. 2767.

Leenhardt, F., Lyan, B., Rock, E., Boussard, A., Potus, J., Chanliaud, E. and Remesy, C. (2006) 'Wheat lipoxygenase activity induces greater loss of carotenoids than vitamin E during breadmaking', *Journal of Agricultural and Food Chemistry*, 54, pp. 1710-1715.

Maclean, W., Harnly, J., Chen, J., Chevassus-Agnes, S., Gilani, G., Livesey, G. and Warwick, P., (2003), February. Food energy—Methods of analysis and conversion factors. In Food and agriculture organization of the united nations technical workshop report (Vol. 77, pp. 8-9). Rome, Italy: The Food and Agriculture Organization.

Macritchie, F. (1992) *Advances in Food and Nutrition Research*. Elsevier.

Mia, M. S., Liu, H., Wang, X. and Yan, G. (2019) 'Multiple near-isogenic lines targeting a QTL hotspot of drought tolerance showed contrasting performance under post-anthesis water stress', *Frontiers in Plant Science*, 10, pp. 271.

Millar, S., Snape, J., Ward, J., Shewry, P., Belton, P., Boniface, K. and Summers, R. (2008) *Investigating wheat functionality through breeding and end use (FQS 23) HGCA*, Project report no. 429.

Min, B., Salt, L., Wilde, P., Kosik, O., Hassall, K., Przewieslik-Allen, A., BurrIDGE, A. J., Poole, M., Snape, J. and Wingen, L. (2020) 'Genetic variation in wheat grain quality is associated with differences in the galactolipid content of flour and the gas bubble properties of dough liquor', *Food Chemistry*, 6, pp. 100093.

Monaghan, J. M., Snape, J. W., Chojecki, A. J. S. and Kettlewell, P. S. (2001) 'The use of grain protein deviation for identifying wheat cultivars with high grain protein concentration and yield', *Euphytica*, 122, pp. 309-317.

Morel, M. H., Dehlon, P., Autran, J., Leygue, J. and Bar-L'helgouac'h, C. (2000) 'Effects of temperature, sonication time, and power settings on size distribution and extractability of total wheat flour proteins as determined by size-exclusion high-performance liquid chromatography', *Cereal Chemistry*, 77, pp. 685-691.

Mosleth, E. F., Wan, Y., Lysenko, A., Chope, G. A., Penson, S. P., Shewry, P. R. and Hawkesford, M. J. (2015) 'A novel approach to identify genes that determine grain protein deviation in cereals', *Plant Biotechnology Journal*, 13, pp. 625-635.

Mosleth, E. F., Lillehammer, M., Pellny, T. K., Wood, A. J., Riche, A. B., Hussain, A., Griffiths, S., Hawkesford, M. J. and Shewry, P. R. (2020) 'Genetic variation and heritability of grain protein deviation in European wheat genotypes', *Field Crops Research*, 255, p.107896.

Nigro, D., Gadaleta, A., Mangini, G., Colasuonno, P., Marcotuli, I., Giancaspro, A., Giove, S. L., Simeone, R. and Blanco, A. (2019) 'Candidate genes and genome-wide association study of grain protein content and protein deviation in durum wheat', *Planta*, 249, pp. 1157-1175.

OECD/FAO (2021), *OECD-FAO Agricultural Outlook 2021-2030*, OECD Publishing, Paris.

Ogihara, Y., Takumi, S. and Handa, H. (2016) *Advances in wheat genetics: from genome to field: proceedings of the 12th international wheat genetics symposium*. Springer Nature.

Oliver, J. and Allen, H. (1992) 'The prediction of bread baking performance using the farinograph and extensograph', *Journal of Cereal Science*, 15, pp. 79-89.

Oliver, J., Blakeney, A. and Allen, H. (1993) 'The colour of flour streams as related to ash and pigment contents', *Journal of Cereal Science*, 17, pp. 169-182.

Oury, FX., Berard, P., Brancourt-Hulmel, M., Heumez, E., Pluchard, P., Rousset, M., Doussinault, G., Rolland, B., Trottet, M. and Giraud, A. (2003) 'Yield and grain protein concentration in bread wheat: a review and a study of multi-annual data from a French breeding program *Triticum aestivum L.*', *Journal of Genetics and Breeding (Italy)*, 57, pp. 59-68.

Oury, FX. and Godin, C. (2007) 'Yield and grain protein concentration in bread wheat: how to use the negative relationship between the two characters to identify favourable genotypes?', *Euphytica*, 157, pp. 45-57.

Park, S.-H., Wilson, J. D. and Seabourn, B. W. (2009) 'Starch granule size distribution of hard red winter and hard red spring wheat: Its effects on mixing and breadmaking quality', *Journal of Cereal Science*, 49, pp. 98-105.

Parker, G., Chalmers, K., Rathjen, A. and Langridge, P. (1998) 'Mapping loci associated with flour colour in wheat (*Triticum aestivum L.*)', *Theoretical and Applied Genetics*, 97, pp. 238-245.

Payne, P., Law, C. and Mudd, E. (1980) 'Control by homoeologous group 1 chromosomes of the high-molecular-weight subunits of glutenin, a major protein of wheat endosperm', *Theoretical and Applied Genetics*, 58, pp. 113-120.

Payne, P. I., Corfield, K. G., Holt, L. M. and Blackman, J. A. (1981) 'Correlations between the inheritance of certain high-molecular weight subunits of glutenin and bread-making quality in progenies of six crosses of bread wheat', *Journal of the Science of Food and Agriculture*, 32, pp. 51-60.

Payne, P. I., Nightingale, M. A., Krattiger, A. F. and Holt, L. M. (1987) 'The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties', *Journal of the Science of Food and Agriculture*, 40, pp. 51-65.

Peterson, C., Johnson, V. and Mattern, P. (1986) 'Influence of cultivar and Environment on mineral and protein concentrations of wheat flour, bran, and grain', *Cereal Chemistry*, 63, pp. 183-186.

Pettorelli, N. (2013) *The normalized difference vegetation index*. Oxford University Press.

Piironen, V., Edelmann, M., Kariluoto, S. and Bedo, Z. (2008) 'Folate in wheat genotypes in the HEALTHGRAIN diversity screen', *Journal of Agricultural and Food Chemistry*, 56, pp. 9726-9731.

Porter, J. R. and Gawith, M. (1999) 'Temperatures and the growth and development of wheat: a review', *European Journal of Agronomy*, 10, pp. 23-36.

Prasad, M., Kumar, N., Kulwal, P., Röder, M., Balyan, H., Dhaliwal, H. and Gupta, P. (2003) 'QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat', *Theoretical and Applied Genetics*, 106, pp. 659-667.

Raun, W. R. and Johnson, G. V. (1999) 'Improving nitrogen use efficiency for cereal production', *Agronomy Journal*, 91, pp. 357-363.

Ren, J., Wu, P., Trampe, B., Tian, X., Lübberstedt, T. and Chen, S. (2017) 'Novel technologies in doubled haploid line development', *Plant Biotechnology Journal*, 15, pp. 1361-1370.

Reynolds, M., Foulkes, J., Furbank, R., Griffiths, S., King, J., Murchie, E., Parry, M. and Slafer, G. (2012) 'Achieving yield gains in wheat', *Plant, Cell, and Environment*, 35, pp. 1799-1823.

Salim, N. and Raza, A. (2020) 'Nutrient use efficiency (NUE) for sustainable wheat production: a review', *Journal of Plant Nutrition*, 43, pp. 297-315.

Salt, L. J., González-Thuillier, I., Chope, G., Penson, S., Tosi, P., Haslam, R. P., Skeggs, P. K., Shewry, P. R. and Wilde, P. J. (2018) 'Intrinsic wheat lipid composition effects the interfacial and foaming properties of dough liquor', *Food Hydrocolloids*, 75, pp. 211-222.

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., Comeau, D. C., Funk, K., Kim, S. and Klimke, W. (2021) 'Database resources of the national center for biotechnology information', *Nucleic Acids Research*, 49, pp. 10.

Scanlon, M., Zhou, H.M. and Curtis, P. (1993) 'Tristimulus assessment of flour and bread crumb colour with flours of increasing extraction rate', *Journal of Cereal Science*, 17, pp. 33-45.

Schreiber, J. and Mc Dowell, L. (1985) 'Leaching of nitrogen, phosphorus, and organic carbon from wheat straw residues: I. rainfall intensity', *Journal of Environmental Quality*, 14(2), pp. 251-256.

Shewry, P. R., Tatham, A. S., Forde, J., Kreis, M. and Mifflin, B. J. (1986) 'The classification and nomenclature of wheat gluten proteins: a reassessment', *Journal of Cereal Science*, 4, pp. 97-106.

Shewry, P. and Tatham, A. (1997) 'Disulphide bonds in wheat gluten proteins', *Journal of Cereal Science*, 25, pp. 207-227.

Shewry, P. R. and Halford, N. G. (2002) 'Cereal seed storage proteins: structures, properties and role in grain utilization', *Journal of Experimental Botany*, 53(370), pp.947-958.

Shewry, P. R., Halford, N. G., Belton, P. S. and Tatham, A. S. (2002) 'The structure and properties of gluten: an elastic protein from wheat grain', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357, pp. 133-142.

Shewry, P. R., Halford, N. G., Tatham, A. S., Popineau, Y., Lafiandra, D. and Belton, P. S. (2003a) 'The high molecular weight subunits of wheat glutenin and their role in determining wheat processing properties', *Advances in Food and Nutrition*, 45, pp. 219-302.

Shewry, P. R., Halford, N. G. and Lafiandra, D. (2003b) 'Genetics of wheat gluten proteins', *Advances in Genetics*, 49, pp. 111-184.

Shewry, P. R. (2009) 'Wheat', *Journal of Experimental Botany*, 60, pp. 1537-53.

Shewry, P. R., Van Schaik, F., Ravel, C., Charmet, G., Rakszegi, M., Bedo, Z. and Ward, J. L. (2011) 'Genotype and environment effects on the contents of vitamins B1, B2, B3, and B6 in wheat grain', *Journal of Agricultural and Food Chemistry*, 59, pp. 10564-10571.

Shewry, P. R., Mitchell, R. A., Tosi, P., Wan, Y., Underwood, C., Lovegrove, A., Freeman, J., Toole, G. A., Mills, E. C. and Ward, J. L. (2012) 'An integrated study of grain development of wheat (cv. Hereward)', *Journal of Cereal Science*, 56, pp. 21-30.

Shewry, P., Wan, Y., Chope, G., Penson, S., Mosleth, E., Hawkesford, M., Campden, B., Campden, C. and Nofima, A. (2013), *Sustainability of UK-grown Wheat for Breadmaking*, Project Report No. 521.

Shewry, P.R. and Hey, S.J. (2015). 'The contribution of wheat to human diet and health', *Food And Energy Security*, 4, pp. 178–202.

Shewry, P. R. (2019) 'What is gluten? Why is it special?', *Frontiers in Nutrition*, 6, p.101.

Slafer, G. A., Andrade, F. H. and Feingold, S. E. (1990) 'Genetic improvement of bread wheat (*Triticum aestivum* L.) in Argentina: relationships between nitrogen and dry matter', *Euphytica*, 50, pp. 63-71.

Slafer, G. A. and Rawson, H. (1995) 'Photoperiod × temperature interactions in contrasting wheat genotypes: time to heading and final leaf number', *Field Crops Research*, 44, pp. 73-83.

Sosulski, F. W. and Imafidon, G. I. (1990) 'Amino acid composition and nitrogen-to-protein conversion factors for animal and plant foods', *Journal of Agricultural and Food Chemistry*, 38, pp. 1351-1356.

Surget, A. and Barron, C. (2005) 'Histologie du grain de blé', *Industries des Céréales*, 145, pp. 3-7.

Turner, A., Bradburne, R., Fish, L. and Snape, J. (2004) 'New quantitative trait loci influencing grain texture and protein content in bread wheat', *Journal of Cereal Science*, 40, pp. 51-60.

Uauy, C., Distelfeld, A., Fahima, T., Blechl, A. and Dubcovsky, J. (2006) 'A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat', *Science*, 314, pp. 1298-1301.

UniProt Consortium (2019) 'UniProt: a worldwide hub of protein knowledge', *Nucleic Acids Research*, 47, pp. 506-515.

United Nations Department of Economic and Social Affairs, Population Division (2022), *World Population Prospects 2022: Summary of Results*, United Nations, New York.

Uthayakumaran, S., Gras, P., Stoddard, F. and Bekes, F. (1999) 'Effect of varying protein content and glutenin-to-gliadin ratio on the functional properties of wheat dough', *Cereal Chemistry*, 76, pp. 389-394.

Veraverbeke, W. S. and Delcour, J. A. (2002) 'Wheat protein composition and properties of wheat glutenin in relation to breadmaking functionality', *Critical Reviews in Food Science and Nutrition*, 42, pp. 179-208.

Wan, Y., Gritsch, C. S., Hawkesford, M. J. and Shewry, P. R. (2013) 'Effects of nitrogen nutrition on the synthesis and deposition of the  $\omega$ -gliadins of wheat', *Annals of Botany*, 113(4), pp.607-615.

Wang, R., Hai, L., Zhang, X., You, G., Yan, C. and Xiao, S. (2009) 'QTL mapping for grain filling rate and yield-related traits in RILs of the Chinese winter wheat population Heshangmai  $\times$  Yu8679', *Theoretical and Applied Genetics*, 118, pp. 313-325.

Weatherall, I. L. and Coombs, B. D. (1992) 'Skin color measurements in terms of CIELAB color space values', *Journal of Investigative Dermatology*, 99, pp. 468-473.

Wickham, H., Chang, W. and Wickham, M. H. (2016) 'Create elegant data visualisations using the grammar of graphics', R package version, 3.0.

Wieser, H. and Seilmeier, W. (1998) 'The influence of nitrogen fertilisation on quantities and proportions of different protein types in wheat flour', *Journal of the Science of Food and Agriculture*, 76, pp. 49-55.

Wilkinson, P. A., Winfield, M. O., Barker, G. L., Allen, A. M., Burridge, A., Coghill, J. A. and Edwards, K. J. (2012) 'CerealsDB 2.0: an integrated resource for plant breeders and scientists', *BMC Bioinformatics*, 13, pp. 1-6.

Wrigley, C. W., Corke, H. and Walker, C. E. (2004) *Encyclopedia of Grain Science*. Academic Press.

Zadoks, J. C., Chang, T. T. and Konzak, C. F. (1974) 'A decimal code for the growth stages of cereals', *Weed Research*, 14, pp. 415-421.

Zanke, C., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G., Hinze, M. and Beier, S. (2014) 'Genetic architecture of main effect QTL for heading date in European winter wheat', *Frontiers in Plant Science*, 5, pp. 217.

Zhao, H., Dai, T., Jiang, D. and Cao, W. (2008) 'Effects of high temperature on key enzymes involved in starch and protein formation in grains of two wheat cultivars', *Journal of Agronomy and Crop Science*, 194, pp. 47-54.

Zhou, B., Serret, M. D., Pie, J. B., Shah, S. S. and Li, Z. (2018) 'Relative contribution of nitrogen absorption, remobilization, and partitioning to the ear during grain filling in Chinese winter wheat', *Frontiers in Plant Science*, 9, pp. 1351.

### Websites

AHDB (2012) *Where are cereals grown and processed in the UK?* Available at: <https://ahdb.org.uk/knowledge-library/where-are-cereals-grown-and-processed-in-the-uk> (Accessed: 3 April 2021).

AHDB (2021) *Wheat Growth Guide*. Available at: <https://ahdb.org.uk> (Accessed: 12 October 2021).

AHDB (2022) *Is UK wheat still export competitive? Grain market daily*. Available at: <https://ahdb.org.uk/news/is-uk-wheat-still-export-competitive-grain-market-daily> (Accessed: 15 June 2022).

Cambridge Dictionary (2023). Available at: <https://dictionary.cambridge.org/dictionary/english/bread> (Accessed: 20 Mars 2023).

DEFRA (2018) *UK - household purchases*. Available at: <https://www.gov.uk/government/statistical-data-sets/family-food-datasets> (Accessed: 6 January 2022).

Federation of Bakers (n.d.) *About the bread industry*. Available at: <https://www.fob.uk.com/about-the-bread-industry> (Accessed: 8 October 2022).

Comité Technique Permanent de la Sélection Des Plantes Cultivées (2020) *Règlement technique d'examen des variétés de céréales à paille*, Ministère de l'Agriculture, Beaucauzé. Available at: [https://cat.geves.fr/CAT\\_WEB/Data/RT\\_Cereales%20a%20paille%202020.pdf](https://cat.geves.fr/CAT_WEB/Data/RT_Cereales%20a%20paille%202020.pdf) (Accessed: 31 February 2021).

Hebbali, A. (2022) *Tools for Building OLS Regression Models*. Available at: <https://cran.r-project.org/> (Accessed: 18 May 2022).

HGCA (2007) *HGCA Recommended list winter wheat 2007*. Available at: <https://projectblue.blob.core.windows.net> (Accessed: 31 May 2022).

Met Office (2021) *UK Climate Projections: Headline Findings*. Available at: [https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18\\_headline\\_findings\\_v3.pdf](https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18_headline_findings_v3.pdf) (Accessed: 5 February 2023).

Nabim (2019) *Wheat Guide*. Available at: <https://www.fob.uk.com/wp-content/uploads/2019/03/Nabim-wheat-guide-2019.pdf> (Accessed: 27 June 2022).

R Core Team (2021) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Available at: <https://www.R-project.org/> (Accessed: 15 December 2021).

Statista (2023) *Leading 10 wheat producers worldwide in 2022/2023*. Available at: <https://www.statista.com/statistics/237912/global-top-wheat-producing-countries/#:~:text=Global%20leading%20wheat%20producers%202022%2F2023&text=In%20the%20marketing%20year%202022,over%20134%20million%20metric%20tons> (Accessed: 23 March 2023).

UK Flour Millers (2021) *Wheat Market Briefing*. Available at: [https://www.ukflourmillers.org/\\_files/ugd/329f2f\\_431423a4096d49c4aa6d66a197e39308.pdf](https://www.ukflourmillers.org/_files/ugd/329f2f_431423a4096d49c4aa6d66a197e39308.pdf) (Accessed: 15 June 2022).

UK Flour Millers (2022) *Wheat guide 2022*. Available at: [https://www.ukflourmillers.org/\\_files/ugd/329f2f\\_15f12039afa24482b77022911df7c2e2.pdf](https://www.ukflourmillers.org/_files/ugd/329f2f_15f12039afa24482b77022911df7c2e2.pdf) (Accessed: 3 March 2022).