October 2013

Project Report No. 521

Sustainability of UK-grown wheat for breadmaking

by

P. R. Shewry[1], Y. Wan[1], G. Chope[2], S. Penson[2], E. F.Mosleth[3] and M. J. Hawkesford[1]

[1] Rothamsted Research
Harpenden, Hertfordshire,
AL5 2JQ, UK

[2] Campden BRI
Chipping Campden,
Gloucestershire, GL55 6LD, UK

[3] Nofima AS
PBox 210
1431 Ås, Norway

This is the final report of a 36 month project (RD-2007-3409) which started in October 2009. The work was funded by BBSRC and a contract for £190,000 from HGCA.

CONTENTS

1. ABSTRACT 5

2. SUMMARY 6

2.1. Introduction/Background and aims 6

2.2. Materials and methods 8

2.2.1. Field trials and crop analysis 8

2.2.2. Transcriptomics 8

2.2.3. Protein analysis 9

2.2.4. Functionality 9

2.2.5. Baking tests 10

2.2.6. Statistical validation 10

1.2.7. Data integration 10

2.3. Results 10

2.3.1. Field trials 11

2.3.2. Transcriptomics 12

2.3.3. Protein analysis 17

2.3.4. Functionality 18

2.3.5. Baking tests 20

2.3.6. Data integration 21

2.4. Discussion/Conclusions and implications 24

3. TECHNICAL DETAIL 25

3.1. Introduction 25

3.1.1. Hypotheses 27

3.2. Materials and methods 28

3.2.1. Field trials 28

3.2.2. Transcriptomics 30

3.2.3. Protein analysis 31

3.2.4. Functionality 32

3.2.5. Baking tests 33

3.3. Results 33

3.3.1. Field trials 33
3.3.2. Transcriptomics 40
3.3.3. Protein analysis 48
3.3.4. Functionality 58
3.3.5. Bread baking quality tests 67
3.4. Discussion 85
3.5. References 87

1. ABSTRACT

The project aimed to establish how the interaction of genetic and environmental factors affect protein deposition during grain development and, in particular, to compare varieties which show grain protein deviation (GPD) with varieties which do not. It included comparisons of the protein content and composition of varieties under different growing seasons, locations and nitrogen (N) inputs; analysis of the gluten protein subunit and glutenin polymer profiles in relation to the dough mixing properties and breadmaking performance, and the identification of gene transcripts which are associated with the stability of wheat quality under reduced nitrogen inputs. Such tools will facilitate the development of new varieties that have high N-use efficiency and consistently good breadmaking quality.

The project focused on six varieties (Cordiale, Istabraq, Hereward, Malacca, Marksman and Xi19). Material was grown at multiple sites (Rothamsted in 2009, 2010 and 2011; RAGT, Limagrain, Syngenta and KWS in 2010 and 2011) and analysed for nitrogen, protein composition and processing quality, including milling and baking tests performed in four independent laboratories (Campden BRI, Warburtons, Premier Foods and Allied Technical Centre (ATC)). In addition, developing grain was harvested from the Rothamsted (2009, 2010 and 2011) and RAGT (2010 and 2011) sites at 21 days after anthesis (mid-grain filling) to measure gene expression using Affymetrix wheat microarrays.

Wheat yields and grain %N were responsive to N application at most sites, although little GPD was observed in 2011. An inverse relationship between yield and grain %N was consistent in all trials. Transcriptome data for N-responsive genes were compared for 2009 and 2010 (Rothamsted site). This showed that the same set of genes responded to N in the two years, but that the response was greater in 2010. Further transcriptome data were collected in 2011 for the plots receiving 200 kg/ha N at Rothamsted and RAGT. 159 significantly N-responsive genes were identified.

A detailed analysis of gluten protein gene expression was undertaken. For both 2009 and 2010, gliadin genes were responsive to N-application. A previously uncharacterised γ-gliadin gene was identified as strongly N-responsive. Amino acid sequences for this γ-gliadin from different wheat sources were compared and expression patterns across the six varieties analysed. The patterns of expression of monomeric and polymeric gluten proteins in all varieties at all N-levels were analysed by SDS-PAGE and SE-HPLC, respectively. Functionality testing was performed at Campden BRI and breadmaking at Campden BRI, Warburtons, Premier Foods and ATC.

All datasets were subject to multivariate data analysis, which allowed us to identify key parameters that predicted baking performance. In particular, specific genes were identified as related to GPD.

2. SUMMARY

2.1. Introduction/Background and aims

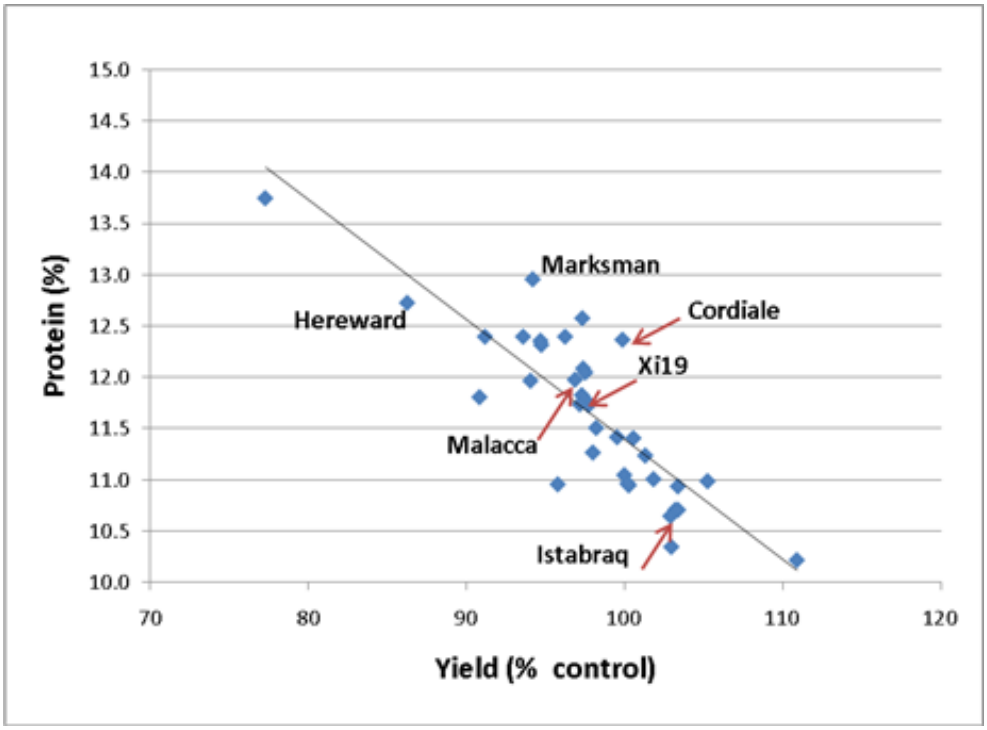


Figure 2.1 Grain protein deviation (deviation from a linear relationship between grain protein content and yield) in UK wheats, taken from the HGCA 2007 harvest survey covering east, west and north regions of the UK. The six selected varieties are indicated. Sixteen samples of each selected variety were analysed, except for Xi19 (12) and Hereward (6).

Wheat is the most important crop in the UK, giving average yields of about 8 tonnes per hectare and being used for food, distilling and livestock feed. However, high yields and the high protein contents required for breadmaking require high inputs of nitrogen fertiliser, which is not sustainable in terms of cost, energy requirement for fertiliser production and environmental footprint.

Furthermore, year to year variation in the weather conditions results in considerable variation in grain processing quality, which may necessitate the import of high volumes of wheat in some years with impacts on the cost of bread and other foods. It is, therefore, crucial that UK wheat production and quality are maintained to guarantee food security and maintain prosperity of the farming and food processing sectors. Data from field trials show significant variation in the response of currently grown wheat varieties to N fertiliser, and in particular in their ability to produce grain with high protein content at the same levels of N application, with Marksman and Cordial showing higher contents of grain protein (called Grain Protein Deviation, GPD) in 2007 (Figure 1.1). Furthermore, they also differ in the extent to which the composition and quality of the grain are affected by environmental fluctuations, with Hereward being recognised as showing unusually high stability.

We have, therefore, investigated the molecular basis for these differences:

1. By growing varieties known to differ in their response to N fertilisation and stability of quality in replicate field trials over several sites in the UK and three harvest years.
2. By performing transcriptome analysis of developing grain across germplasm, nitrogen input, site and year.
3. By collecting corresponding final grain samples for protein and functional analysis
4. By examining processes in canopy tissue contributing to yield/N remobilisation in selected lines.
5. By studying the expression of selected genes in greater depth.
6. By integrating all data sets to correlate functionality with underlying processes/genes.

Project aims:

The overall aim was to compare the expression of genes and the synthesis and accumulation of gluten proteins in the developing grain with the final composition and processing properties, and to relate this to wider aspects of nitrogen use efficiency in the whole plant.

1. To determine how the interactions of genetic and environmental factors affect the timing and spectrum of protein deposition during grain development, by conducting field trials of 6 selected UK wheat varieties, sampling developing grain during grain filling.
2. To determine how key functional proteins accumulate within wheat varieties under different growing seasons, locations and nitrogen inputs, using SDS-PAGE and HPLC analysis to separate the component proteins and polymers.
3. To determine how the assembly of protein polymers affect the dough mixing properties and breadmaking performance, by relating HPLC separated fractions to dough functionality and baking test results.
4. To identify and characterise transcripts which are associated with the stability of wheat quality under reduced nitrogen inputs by using transcriptome approaches (evaluation of level of expression of all genes) and correlating gene expression with the traits measured.

This should result in the identification of genes and proteins whose expression correlates with grain nitrogen content and composition and with processing quality (including stability of quality from year to year). Some of these genes and proteins may be directly involved in determining the traits of interest and hence the work will lead to better scientific understanding. Other genes and proteins may not be directly involved but could nevertheless be developed as markers which can be used by plant breeders to select for improved wheat varieties. The project, therefore, contributes to the target of more sustainable production of wheat in the UK.

2.2. Materials and methods

The project was based on multiple field trials over three years, extensive analysis of harvested grain, from N determination to protein and gene expression profiling, functionality testing and breadmaking by multiple bakers (Figure 2.2). All data have been cross-correlated to relate traits to one another and to specific patterns of gene expression.

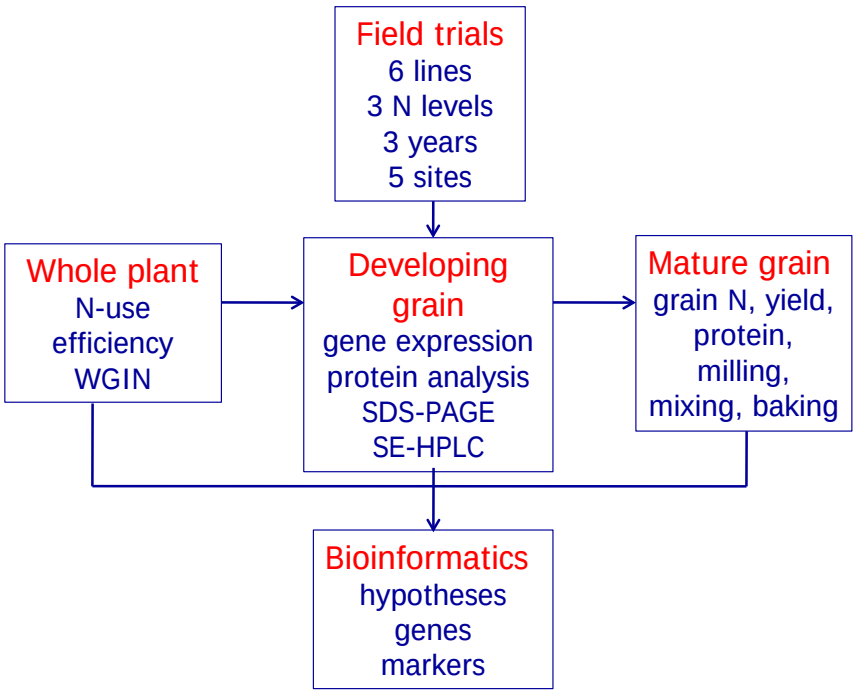


Figure 2.2 Practical components of the project.

2.2.1. Field trials and crop analysis

Six varieties were grown at Rothamsted in 2009, 2010 and 2011 and at sites run by the breeder partners in 2010 and 2011 (RAGT, Ickleton, Cambridge; Limagrain, Woolpit, Suffolk; Syngenta, Whittlesford, Cambridge; KWS-UK, Thriplow, Hertfordshire). Three N levels were used: 100kg/ha as a “low input” level, 200kg/ha to reflect modern practice for breadmaking wheats in the UK and 350 kg/ha as an extreme high input to achieve high grain protein. Nitrogen was applied in ‘splits’ as solid ammonium nitrate prills. All plots were randomised with three replicates.

2.2.2. Transcriptomics

Microarrays were used to profile gene expression. A specific time point of 21 dpa (days post anthesis) was chosen as a key developmental stage (mid-grain filling) in which grain storage proteins are being synthesised. Ears were tagged at anthesis and caryopses harvested 21 d later. Gene expression was determined by profiling RNA extracted from this material against a gene chip containing 55,000 elements. This is not a full genome chip but represents a large proportion of wheat genes and known alleles. Data from the profiling are semi-quantitative giving a good

indication of the relative levels of expression of all RNAs in the sample simultaneously. Data were collected for three years at Rothamsted and for 2010 and 2011 at the RAGT site, for the three N levels in 2009 and 2010, and for the 200 kgN/ha treatment in 2011.

2.2.3. **Protein analysis**

The protein composition of mature and developing grain samples was analysed using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and size-exclusion high pressure liquid chromatography (SE-HPLC). For SDS-PAGE analysis, protein was extracted from samples (35mg) of wholemeal flour and separated using gel electrophoresis. Proteins were visualised by staining with Coomassie Brilliant Blue and then scanned and quantified using Total Lab TL120 version 2006F (Nonlinear Dynamics, Newcastle-upon-Tyne, U.K.) with an optical density curve calculated from a Kodak T14 control scale (Tiffen LLC Rochester, NY). The bands were divided into three groups: high molecular weight subunits of glutenin, low molecular weight subunits of glutenin and gliadins. For SE-HPLC analysis (Morel *et al*., 2000), proteins were extracted from white flour (mature samples) or wholegrain flour (developing samples). The polymer size distribution was calculated from the chromatogram which has five identifiable peaks (Figure 2.3). The first peak to elute from the column is referred to as F1 and consists of high molecular weight (HMW) polymers enriched in HMW subunits. The F2 peak comprises low molecular weight (LMW) polymers and is enriched in LMW subunits. The F3 and F4 peaks are comprised principally of ω-gliadins and α-, β-, and γ-gliadins, respectively, while the F5 peak comprises low molecular weight proteins including albumins and globulins. The overall area under the trace is a measure of the total protein content of the flour and is termed AT.

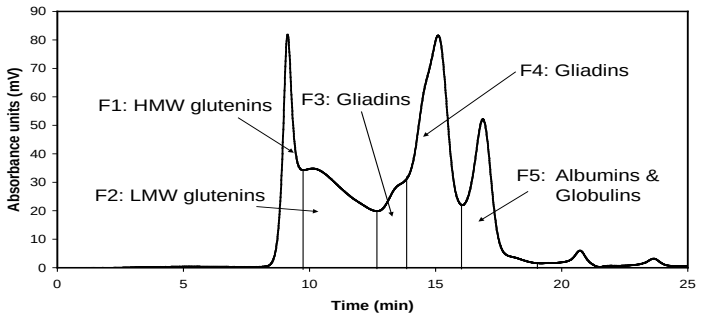


Figure 2.3 Typical SE-HPLC chromatogram of HMW and LMW glutenin polymers (F1 and F2, respectively), monomeric gliadins (F3 and F4) and smaller albumin and globulin proteins (F5).

2.2.4. **Functionality**

Milling

White flour was produced using a Bühler Laboratory Flour Mill MLU 202 at Campden BRI according to an internal Campden BRI method. Where replicates were pooled for analysis, wheat grain from each replicate was combined and blended together thoroughly prior to milling.

NIR analysis of grain and flour

Wholegrain and flour protein and moisture content were measured by NIR according to the Manual of methods for wheat and flour testing Guide Line No. 3 (Campden BRI).

Hagberg Falling number

The Hagberg Falling Number was determined according to the Manual of methods for wheat and flour testing Guide Line No. 3 (Campden BRI).

Dough rheology

Flour water absorption was measured using a Brabender Farinograph according to the Manual of methods for wheat and flour testing Guide Line No. 3 (Campden BRI). The extensibility and resistance of the dough was measured using a Brabender Extensograph according to the Manual of methods for wheat and flour testing Guide Line No. 3 (Campden BRI). Small scale dough rheology was determined using a Reomixer according to an internal method.

2.2.5. **Baking tests**

Bake tests were performed at Campden BRI using no time dough process (spiral mixer). A standard lean recipe was used to produce 800g, four piece loaves in unlidded tins proofed to time (Millar *et al*., 2008). Loaf volume was measured and crumb structure was analysed using a C-Cell imaging system (Calibre Control International Ltd). C-Cell uses high definition imaging and controlled illumination to ensure optimum image quality and consistent analysis. An image is analysed to provide data values which can provide valuable information about cells, circulation, cell elongation, faults and shape and size information. Bake tests were also performed by Warburtons, Allied Technical Centre (ATC) and Premier Foods according to their own procedures.

2.2.6. **Statistical validation**

Univariate statistical analysis was performed on all the data to verify their significance related to the experimental design consisting of the genotypic and the environmental factors and their interactions. The p-values were adjusted for multiple comparisons by False Discovery Rate using a statistical rotation test.

1.2.7. **Data integration**

The data were integrated by the multivariate data regression approach Partial Least Squares Regression as implemented in the software Unscrambler.

2.3. Results

Six varieties were selected for comparison based on the HGCA data shown in Figure 2.1. Marksman and Cordiale are Group 2 wheats which showed GPD (Figure 2.1), Hereward has been

the “gold standard” breadmaking wheat for over 20 years and shows high stability, Malacca and Xi19 are Group 1 breadmaking wheats and Istabraq is a Group 4 feed wheat. The varieties were grown and harvested at multiple sites (Rothamsted in 2009, 2010 and 2011; RAGT, Limagrain, Syngenta and KWS in 2010 and 2011) and analysed for nitrogen, protein composition and processing quality, including milling and baking tests performed in four independent laboratories (Campden BRI, Warburtons, Premier Foods and ATC). In addition, material from the Rothamsted (2009, 2010 and 2011) and RAGT (2010 and 2011) sites was sampled at 21 days post-anthesis (dpa) to determine gene expression using the Affymetrix wheat GeneChip microarray.

2.3.1. **Field trials**

Yields and grain N

In 2009, the yields at Rothamsted varied between 8.2 and 12.7 t/ha (at 85% dry matter), with grain %N varying from 1.4 to 2.4. All varieties responded positively to applied N in terms of yield and grain %N. Istabraq had the highest yields and lowest %N.

In 2010, all months had below average rainfall, with the exception of August which was very wet. The yield at Rothamsted varied from 7.3 to 10.2 t/ha., substantially lower than 2009, and grain %N varied from 1.4 to 2.8. Yield and grain %N both responded to applied N, with %N responding most strongly. Yield and grain %N also responded to N application at Limagrain, RAGT and Syngenta but not at KWS. Yields were lowest at RAGT and highest at KWS and Syngenta.

In 2011, March to May was characterised by exceptionally low rainfall followed by a relatively wet summer. Yields at Rothamsted ranged from 7.6 to 11.5 t/ha and grain %N from 1.6 to 3.2. Responses to applied N were generally poor for all sites except Rothamsted and Limagrain. The lowest yields were at RAGT, Syngenta and KWS, which were close geographically and had low rainfall.

Grain protein deviation

In 2009, Hereward showed a positive deviation at all N levels, Cordiale at 100 and 350, Marksman at 200 and Xi19 at 350 kg/ha (Figure 2.4). In 2010, the inverse relationship between yield and grain %N was consistent in all trials with Cordiale and Marksman showing positive GPD at the KWS site; Marksman and Hereward at Rothamsted; Hereward, Marksman and Cordiale at RAGT and Cordiale alone at Limagrain and Syngenta. Little GPD was observed in 2011 with the exceptions of positive GPD for Hereward at Syngenta and KWS, Xi19 and Hereward at KWS, the two lowest yielding locations

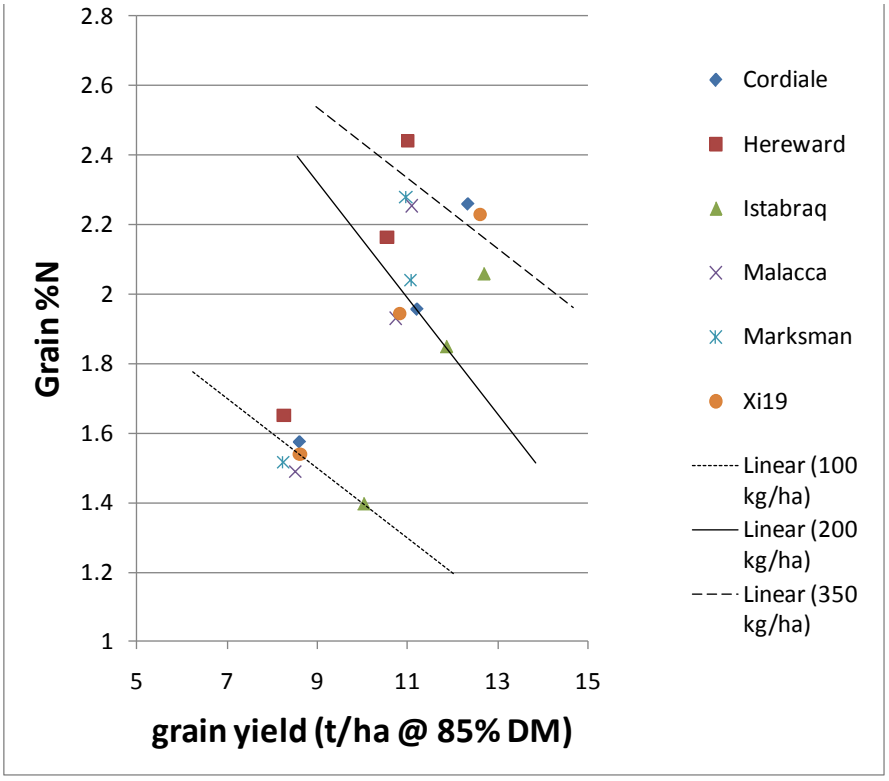


Figure 2.4 Relationship between grain yield and grain N for the six varieties grown at Rothamsted in 2009

2.3.2. Transcriptomics

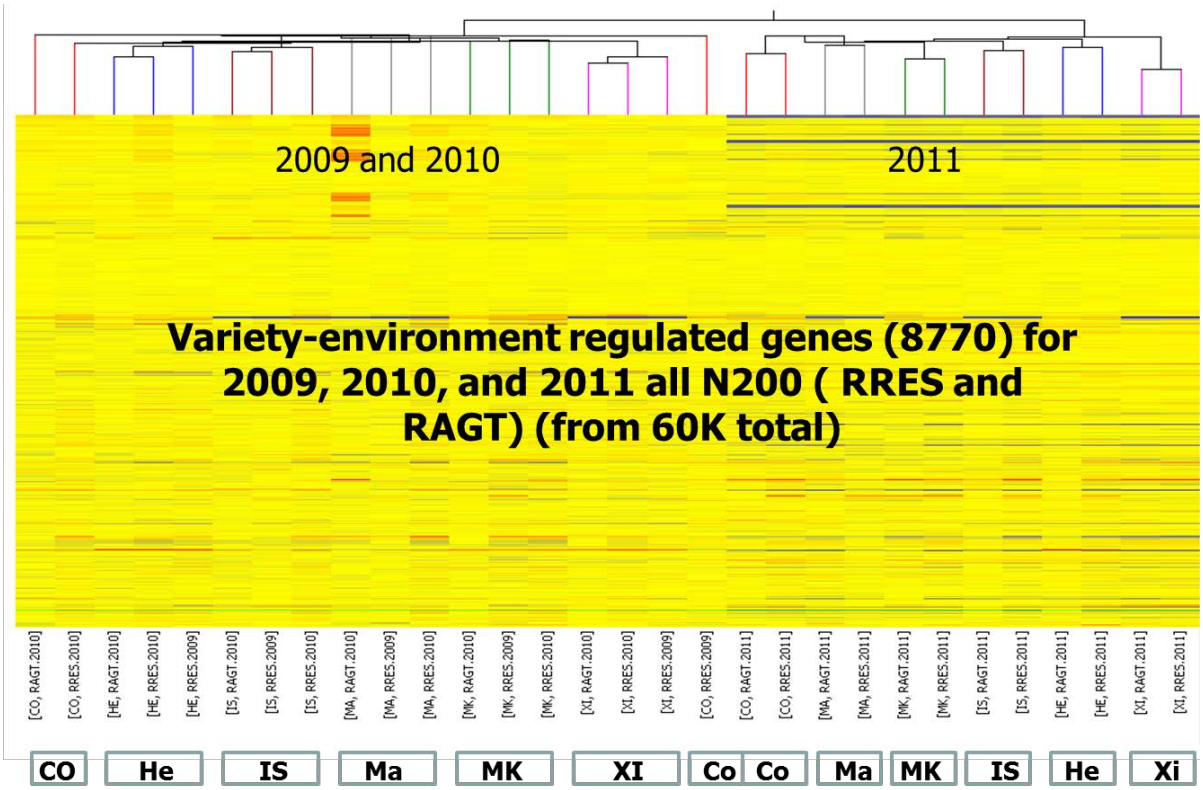


Figure 2.5 Transcriptomics overview: hierarchical cluster analyses of transcriptome profiles for 6 varieties in 3 years at the Rothamsted and RAGT sites at 200 kgN/ha only.

The gene expression profiles determined for whole caryopses at 21 days after flowering showed effects of year, site and variety as well as nitrogen (Figure 2.5). 8,770 of the 60,000 total features (i.e. different genes and alleles) represented on the microarray responded to the year, environment and genotype when comparing the 200 kg/ha treatment alone. The data for 2009 and 2010 were quite similar with the effect of variety overriding the differences due to year and site. The data for 2011 differed, showing strong effects of year.

By averaging the responses of varieties and comparing the profiles for the three nitrogen levels in 2009 and 2010, it was possible to identify multiple nitrogen-regulated genes: 107 genes were up-regulated significantly in both 2009 and 2010 in response to increasing N application, with a greater responsiveness being apparent in 2010; 52 genes were down-regulated in response to increasing N-application in both 2009 and 2010. Of the 107 up-regulated genes: 13 encoded storage proteins, 6 defence-related genes, 5 photosynthetic genes and 81 were un-annotated 'unknown' genes.

Relating GPD to the transcriptome

Firstly a value for GPD was determined for each variety and for each year (for all N-fertilisation rates), after normalisation for effects of N-fertiliser input and yield dilution effects (examples shown in Figure 2.6).

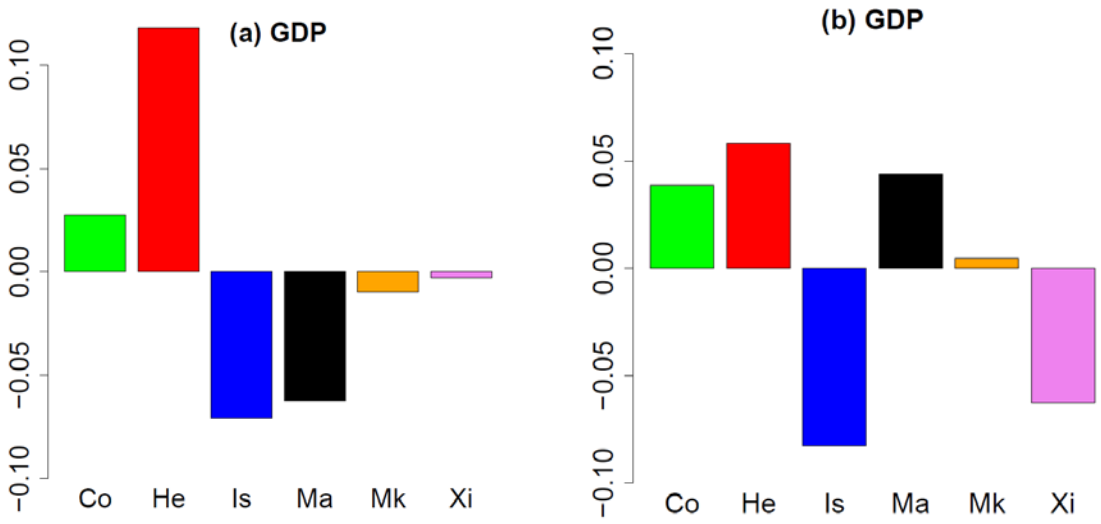


Figure 2.6 A measure of GDP for the 6 varieties along the x-axis; Cordiale (Co), Istabraq (Is), Hereward (He), Malacca (Ma), Marksman (Mk) and Xi19 (Xi)) grown at Rothamsted in 2009 (a) and 2010 (b) found by adjusting the protein content first by the direct effect of N-fertiliser level and secondly for the indirect effect of yield.

To identify genes related to the GPD, principal component analysis (PCA) was first performed on the gene expression data (the first 15 PCs are shown), and PCs related to GPD but not to yield

were identified (see Table 2.1). This table indicates which principal components in the analysis correlate most significantly with the various parameters (yield, grain N and GPD).

Table 2.1 Results of ANOVA (FDR adjusted p-values) showing the effect of the scores of PCA of the gene expression data (input of the model) on the phenotypic characteristics (output of the model), (a) for 2009, (b) for 2010.

(a)

	Yield	Protein	Yield corr N	Protein corr N	Protein corr N and Y
mean PC1	0.078	**0.031**	0.740	0.527	0.591
mean PC2	0.065	0.998	0.053	**0.005**	**0.042**
mean PC3	0.152	0.453	**0.008**	**0.006**	0.126
mean PC4	0.736	0.682	0.139	0.911	0.412
mean PC5	0.362	**0.036**	**0.000**	**0.014**	0.653
mean PC6	0.357	0.195	0.823	0.572	0.412
mean PC7	0.853	0.062	0.199	**0.000**	**0.000**
mean PC8	0.071	**0.034**	0.431	0.642	0.285
mean PC9	0.314	0.444	0.080	0.067	0.299
maen PC10	**0.000**	**0.000**	0.447	0.089	**0.010**
maen PC11	**0.024**	0.379	**0.000**	**0.000**	0.099
mean PC12	0.088	**0.018**	0.414	0.641	0.276
mean PC13	0.525	0.716	0.445	0.437	0.152
mean PC14	**0.000**	**0.000**	0.427	0.252	0.398
mean PC15	0.191	0.179	0.263	0.600	0.174

(b)

	Yield	Protein	Yield corr N	Protein corr N	Protein corr N & Y
mean PC1	**0.001**	**0.000**	0.293	0.392	0.756
mean PC2	0.603	**0.000**	0.170	**0.012**	**0.038**
mean PC3	0.942	**0.021**	0.285	**0.030**	0.062
mean PC4	0.966	0.085	0.449	0.081	0.111
mean PC5	0.143	0.255	0.369	0.837	0.708
mean PC6	**0.013**	**0.000**	0.860	0.244	0.120
mean PC7	0.313	**0.005**	0.641	0.491	0.608
mean PC8	0.283	**0.295**	0.740	0.084	0.061
mean PC9	**0.026**	**0.245**	0.136	0.087	0.301
mean PC10	0.836	0.948	0.704	0.561	0.329
mean PC11	0.204	0.014	0.812	0.483	0.306
mean PC12	0.840	0.384	0.762	0.549	0.604
mean PC13	0.639	0.798	0.380	0.752	0.317
mean PC14	0.966	0.916	0.963	0.855	0.849
mean PC15	0.135	0.338	0.221	0.913	0.473

For 2009, PC2 and PC7 showed a significant relation to GPD but not to the yield or the N-fertilisation level. And for 2010, PC2 and PC3 showed significant relation to GPD, but not to yield. The scores of these selected PCs are shown in Figure 2.7 as means for each of the cultivars.

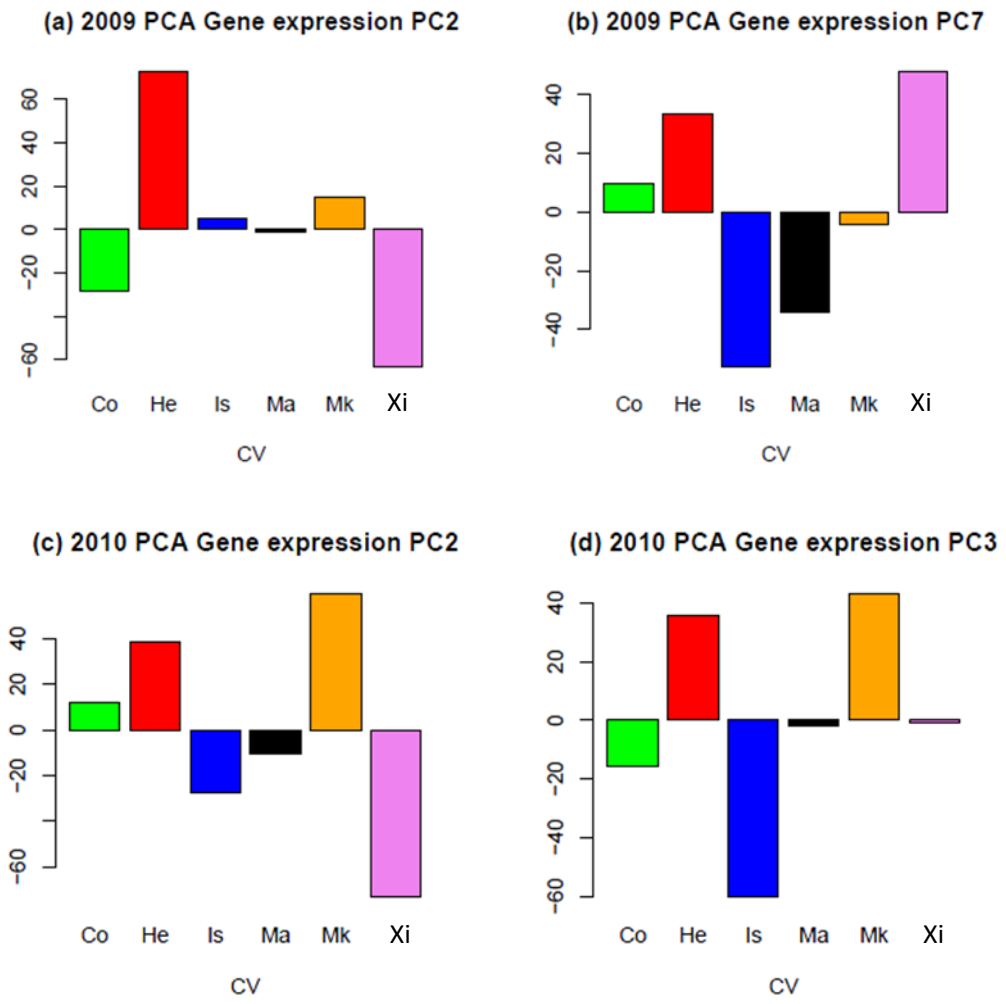


Figure 2.7 Cultivar means of scores of 4 selected principal components (PC2: a and c; PC3, b and d) PCA of the gene expression data obtained in 2009 (a and b) and 2010 (c and d) at Rothamsted. The PCs shown were those significant for GPD but not for grain yield.

Comparing Figures 2.6 and 2.7, there is a striking similarity in the patterns of GPD and the specific PC analysis of gene expression. For example, Hereward is positive and Istabraq is negative in both 2009 and 2010. The other cultivars vary in their position.

Genes do not usually act as individually, but as orchestrated multivariate patterns. The PC found to be significantly related to GPD reflects a multivariate pattern of genes related to GPD. All the genes will have a loading along each of these PCs. However, some genes are more important than others in spanning this PC with a consistent expression pattern across the biological replicates. To identify these genes an analysis of variance was performed using the means of the scores of the biological replicates as input in the model and the gene expression values as the response. By this approach the most important genes contributing to each of the selected PCs were identified. The different growth years were analyzed separately and consistent patterns across the years were regarded as the most relevant. To further zoom into a fewer selection of genes we also looked for genes significant in all the selected PCs.

Selection of significant genes across the growth year and PCs gave7669 genes that were significant for both PC2 and PC7 in 2009, and 9300 genes that were significant for both PC2 and PC3 in 2010. The total number of genes being identified in both years was 3207. These genes were therefore selected as good candidates for GDP, and the two dataset were combined. Partial Least Squares regression analysis was then performed to obtain a visual overview of the results, and to further focus on the most relevant genes for future studies.

Gene expression levels of selected genes with positive or negative relation to GPD are shown in Figure 2.8. The selection of the genes was performed on the data set from Rothamsted in 2009 and 2010. For the selected genes the gene expression profile in Figure 2.8 displays the gene profiles for all the three growth years (indicated by different symbols in the plot) and for both sites where gene expression pattern was available (Rothamsted and RAGT). Figure 2.8 shows that the genes selected are cultivar specific with a consistent pattern across growth environments. The genes in the upper rows of Figure 2.8, which were positively related to GDP, were genes with low expression in Istabraq (in blue) and high expression in Hereward (in red). Conversely, the genes in the lower row of Figure 2.8, which were negatively related to GDP, had low expression in Hereward (in red) and high expression in Istabraq (in blue).

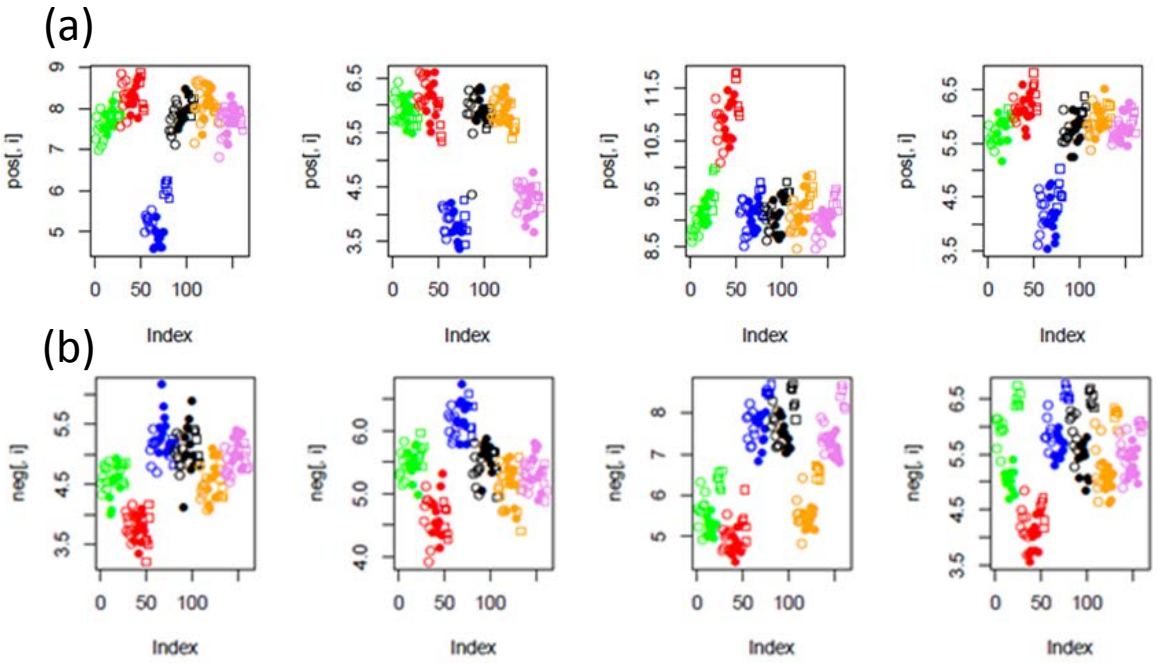


Figure 2.8 A gene expression profile of 4 of the genes with (a) the most positive and (b) 4 of the genes with the most negative relation to GDP in 2009 (open circles), in 2010 (closed circles) and 2011 (open squares). The samples are sorted along the x-axis according to the cultivars for clarity. Green is Cordiale, red is Hereward, blue is Istabraq, black is Malacca, yellow is Marksman and purple is Xi19.

2.3.3. **Protein analysis**

SDS-PAGE

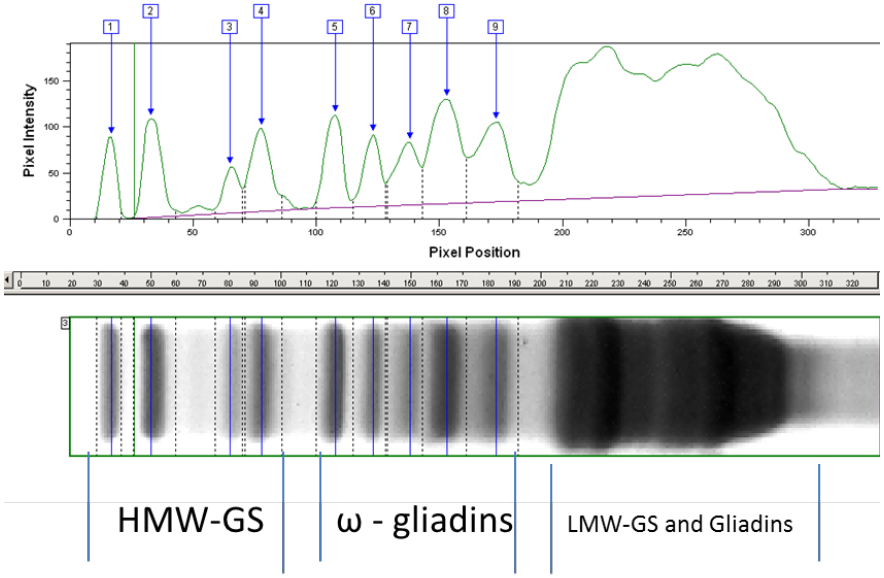


Figure 2.9 SDS-PAGE of storage proteins

Protein composition

SDS-PAGE analysis (Figure 2.9) was carried out on individual replicates of milled whole developing (21 dpa) and mature grain while SE-HPLC analysis was carried out on pooled replicates of ground whole developing grain and white flour from mature grain. Developing grain samples were from Rothamsted Research for 2009 and 2010 and mature grain was from Rothamsted Research alone in 2009 and from Rothamsted Research plus four additional sites in 2010 and 2011.

SDS-PAGE showed that in developing grain, the proportions of HMW subunits were not affected by N fertilisation, but the proportions of LMW subunits and gliadins increased and decreased, respectively, with increasing levels of N application. With increasing days after anthesis, the gliadins tended to accumulate at the expense of HMW subunits. Similarly, in mature grain the proportions of HMW subunits were generally not affected by N fertilisation, while LMW subunits decreased with increasing N fertilisation and gliadins increased. These patterns were consistent across the three years of the trial. Hence, it can be concluded that increasing fertilisation resulted in a higher content of grain protein but decreased intrinsic quality due to disproportional increases in gliadins (which contribute to dough extensibility rather than strength)..

SE-HPLC analysis (see Figure 2.3 for explanation of fractions F1-F5) showed that, in general, the proportion of polymeric proteins increased (fractions F1 and F2) during development, while that of the monomeric proteins decreased. The proportions of the the α- and γ-gliadins (F4) and non-

gluten protein (F5) fractions only were affected by the level of nitrogen fertilisation, where the proportion of the F4 fraction increased with increasing nitrogen addition and that of the F5 fraction decreased. For mature grain, only the the monomeric gliadins (F3, F4) and F5 fractions were affected by the level of nitrogen application, where the proportions of the F3 and F4 fractions increased with increasing nitrogen application, and the proportion of the F5 fraction decreased.

Both SDS-PAGE and SE-HPLC analysis showed differences between varieties that were consistent across different sites and years. Taken together, the results of the protein analysis show effects of nitrogen on protein composition that varied with variety.

2.3.4. **Functionality**

Sufficient samples for milling were not available for samples grown at 100N and 350N from the RAGT site in 2010. Both functionality and baking tests were undertaken on material from three sites in 2010 (Rothamsted Research, Limagrain and Syngenta), and functionality testing was additionally made on material from KWS.

NIR analysis of grain and flour

Determination of total grain N showed that the response to nitrogen fertilisation was dependent on the site. For samples grown at Rothamsted Research, the grain N increased with increasing levels of N fertilisation. The degree of response was lower at the Syngenta site, and lower still at the Limagrain site, while the samples grown at KWS showed no apparent response to N fertilisation in total grain N.

Hagberg Falling number

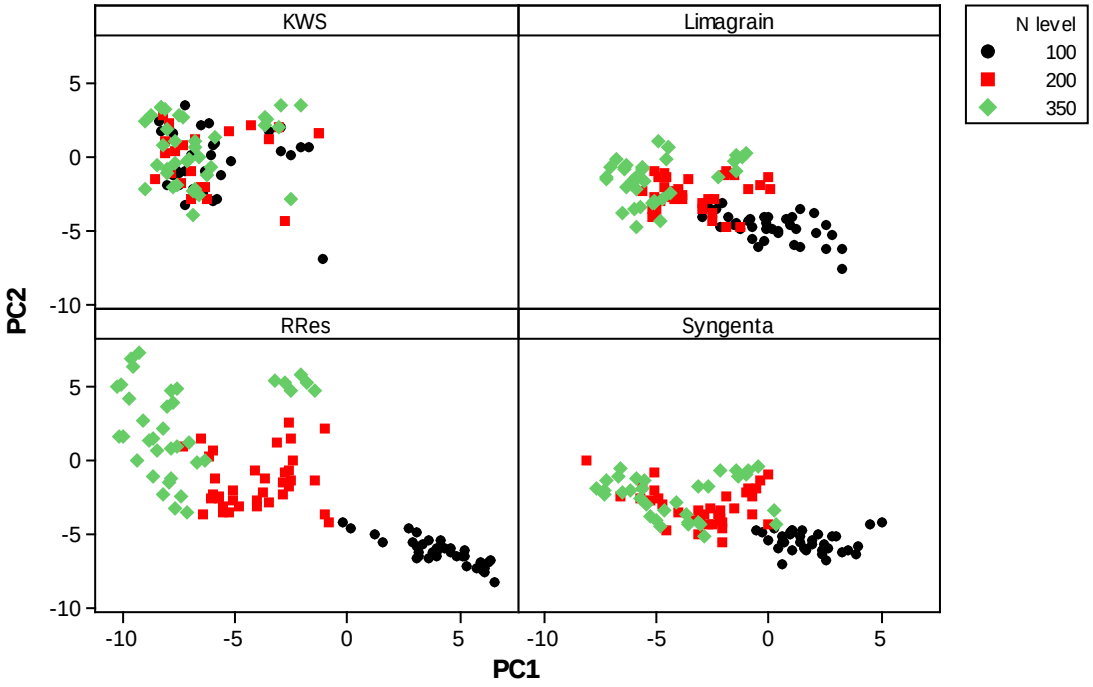
There was no effect of N fertilisation on Hagberg Falling Number (HFN). The HFN was generally lowest for Istabraq at each site, and overall, HFN values were low indicating the presence of pre-harvest sprouting for all varieties grown at Rothamsted Research in 2010.

Dough rheology

Dough rheology was determined on white flour samples made from pooled replicates. In general, Istabraq showed the lowest values for water absorption, development time, stability, resistance and extensibility. For samples grown at Rothamsted Research, there was an effect of nitrogen fertilisation, with increases in nitrogen fertilisation resulting in increases in water absorption, development time, stability, resistance and extensibility. There were also differences between varieties in their response to nitrogen fertilisation. A similar trend was observed for samples grown at the Limagrain and Syngenta sites, but not for the samples grown at KWS. Reomixer traces were collected in duplicate for each biological replicate and the traces reduced to two principal co-ordinates (PC1 and PC2) from which a quality map can be plotted. In this map negative PC1

values reflect higher dough strength, development time and consistency; negative PC2 values reflect greater dough stability. Plotting the data for all of the samples on the quality map shows a clear separation on the level of nitrogen fertilisation (Figure 2.8), which was most clearly seen for the samples grown at Rothamsted Research. By contrast, the samples grown with different levels of nitrogen fertilisation at KWS overlapped, showing that the applied fertiliser had less effect on grain quality.

A



B

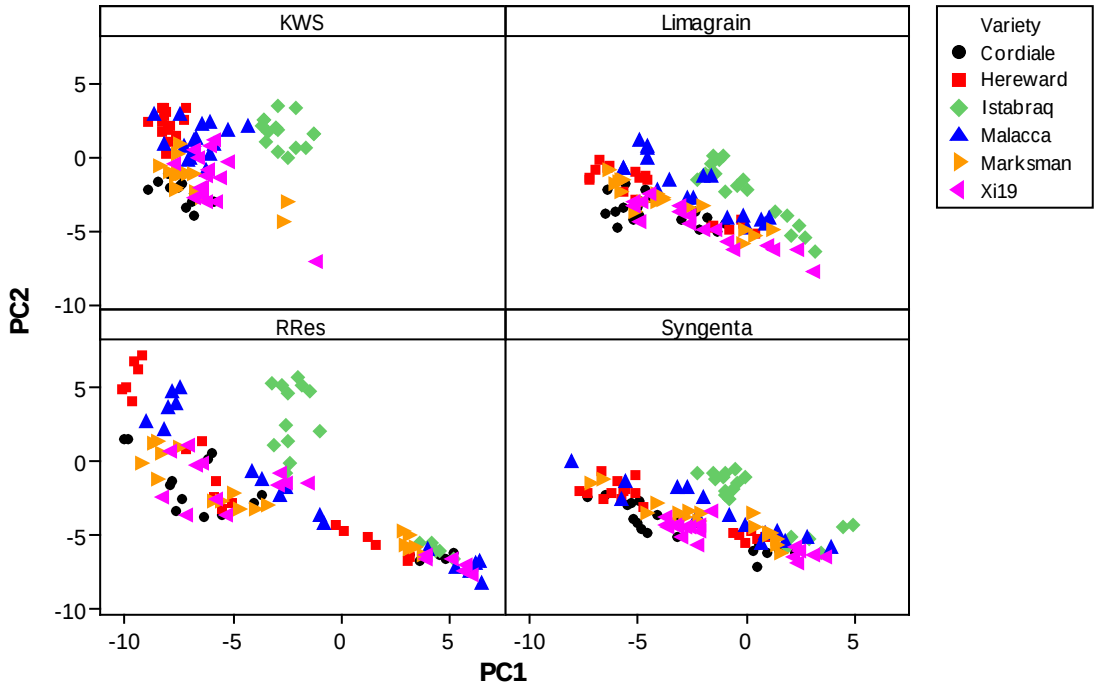


Figure 2.8 Reomixer quality map for white flour milled from wheat grown at four sites in 2010 at different levels of nitrogen fertilisation showing the effect of nitrogen fertilisation (A) and variety (B)

Taken together, the results for total grain N (protein) and flour rheology show no effect of N fertilisation at the KWS site, this suggests that the extra nitrogen applied at the KWS site was not taken up by the crop or that the residual N in the soil was high.. The results also show differences between the varieties in their response to nitrogen fertilisation.

2.3.5. **Baking tests**

The loaf volume of bread baked at Campden BRI generally increased with an increasing level of nitrogen (Figure 2.9). Bread made from Istabraq consistently had the lowest loaf volume and poorest crumb structure compared with bread made from the other varieties, which was expected as Istabraq is not a breadmaking variety. For samples grown at Rothamsted Research, there was an increase in the loaf volume with an increase in nitrogen fertilisation. This trend was seen for both 2009 and 2010 (in 2011, only samples from the 200N treatment were baked). There was also an increase in crumb score (subjective measurement). There was also evidence of a similar effect for bread baked from samples grown at Syngenta and Limagrain in 2010, but not for bread baked from samples grown at KWS.

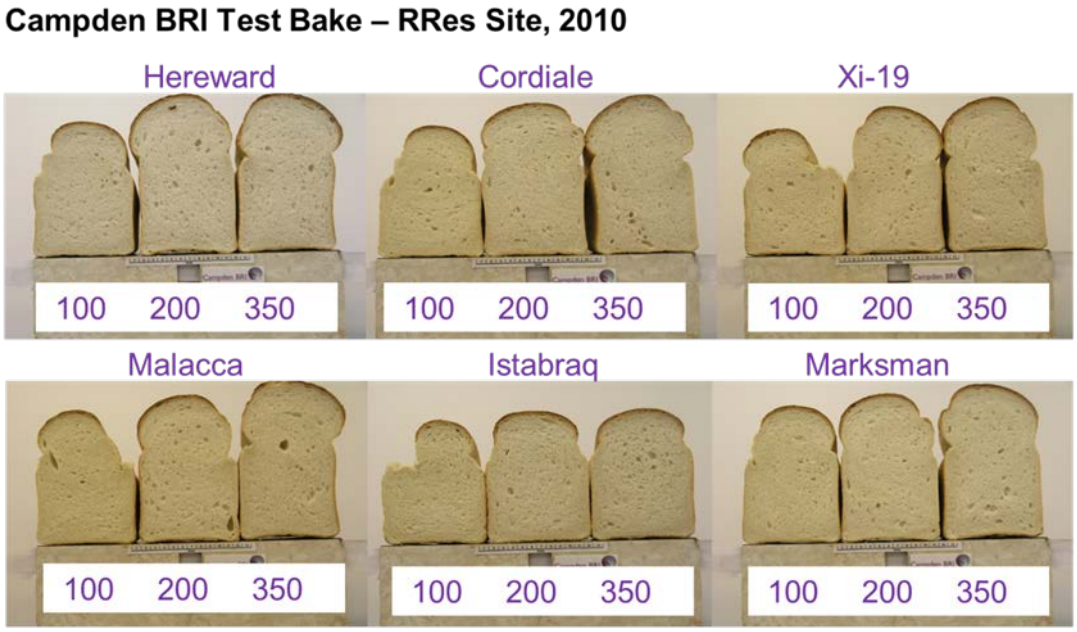


Figure 2.9 Bread baked from flour samples milled from wheat grown at Rothamsted Research in 2010 at three levels of nitrogen fertilisation (100, 200 or 350 kg/ha)

Industry partner data

Bread was baked by industry partners in 2009, 2010 and 2011. In 2010, the sample received from RAGT was not sufficient to allow for baking. The baking assessment was carried out by the industry partners according to their standard protocols and, therefore, the data recorded varied between partners.

The data from the industry partners also showed some differences between each other and from the Campden BRI analyses in relation to varieties and nitrogen treatments. The differences between the results obtained by Campden BRI and the industry partners are probably due to differences between the processes. The spiral bake performed by Campden BRI is a lean recipe and is baked in an unlidded tin, while at ATC and Warburtons the loaves were baked in lidded tins. The quality criteria demanded by industry will also include parameters other than simply a high loaf volume, such as handling properties and slicing.

2.3.6. **Data integration**

An objective of this study was to relate specific genes to function and wheat performance. This is targeted at grain protein deviation but also includes genes responsible for, or at least diagnostic of, baking parameters. It is also possible that these groups of genes may overlap.

The data were integrated by a multivariate data regression approach (Partial Least Squares (PLS) Regression). In the analysis performed here the focus is on genes related to baking quality which overlap the genes responsible for GPD.

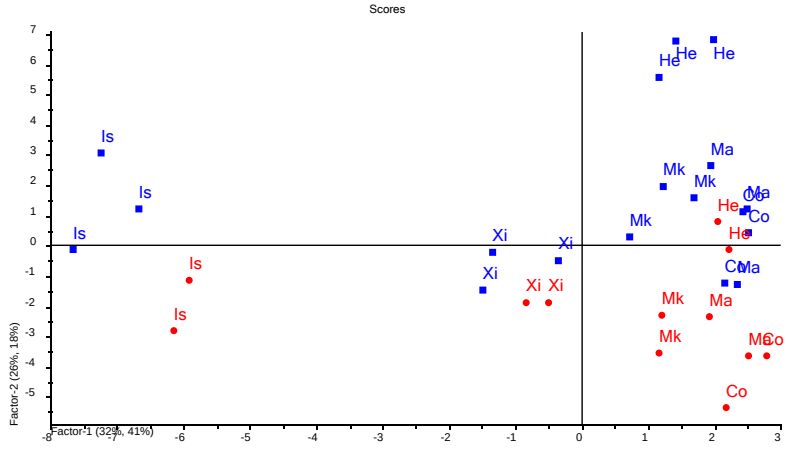
Analysis was first performed on the 2009 dataset which included all three biological replicates for the Campden BRI baking tests.

- Firstly, analysis was performed to study how the design parameters were related to the baking data (data not shown)
- Secondly, analysis was performed for the same growth year using the genes as input in the analysis and the baking data from Campden BRI as output. Only those genes found in the analysis above to be significant for the GPD were included (data not shown)
- Thirdly, genes selected as significant in this analysis performed on growth year 2009, were then subjected to similar analysis for the two other growth years (see Figure 2.10). Thus, we are then performing validation of the selected genes in new datasets.

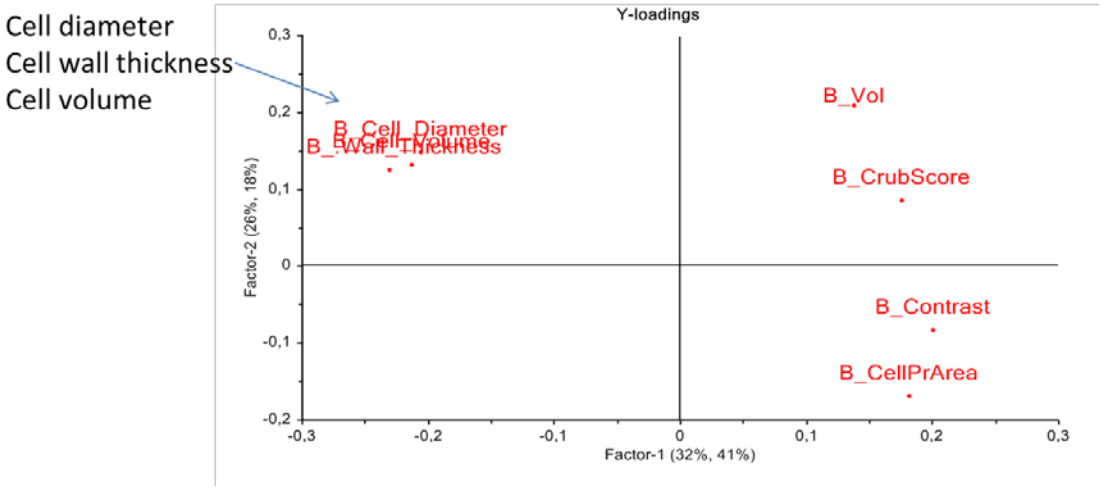
By PLS regression the data are projected down on a few new variables, called PLS factors, describing the main relation between the input data (x-variables) and the output data (y-variables). Visual inspection of plots of the first few PLS factors, viewing both a sample plot (score plot) and a plot of the x and y- variable, then reveal the main pattern of variation relevant for the prediction. The score plot and the loading plot correspond to each other, as the axes are the same PLS factors. Thus, samples located in one direction in the score plot have high value of the parameters in the loading plot located in the same direction, and low value of parameters located in the opposite direction.

An example is shown in Figure 2.10. Genes, chosen only from the GPD selected set (see section 3.2.3 above), and found to be significant in prediction baking performance in 2009, were selected and used in an analysis of the two remaining growth years. When analysing all genotypes, Istabraq was dominating the first PLS factor, located towards the lefthand side of the plot (Figure 2.10.a); this direction was characterised by high cell diameter, high wall thickness and high cell volume, and low value of loaf volume, crumb structure, contrast and cell per unit area (Figure 2.10.b). Genes responsible for this variation are shown in Figure 2.10c. The second PLS factor reflects primarily the growth year, where all the samples from 2010 (in blue) are located in the upper part of the plot, and the samples from 2011 in the lower half (in red). Along the second PLS factor loaf volume is located in the upper part of the loading plot of the y-variables (Figure 2.10.b), which reflects generally higher loaf volume of the 2010 samples compared with the 2012 samples. In the x-loading plot genes significant for the differences seen in the two first PLS factors are circled.

(a)



(b)



(c)

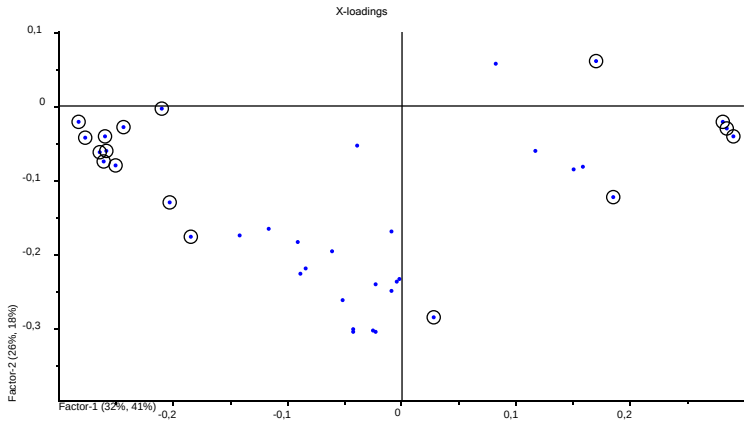


Figure 2.10 PLS regression on data from 2010 and 2011. Genes found in 2009 are used as input, and baking performed at Campden BRI as response, (a) score plot of the samples showing the cultivars (Co) Cordiale, (Is) Istabraq, (He) Hereward, (Ma) Malacca, (Mk) Marksman and (Xi) Xi19), (b) y-loading plot of the response parameters and (c) x-loading plot of the genes where genes circled are significant by statistical significance test.

2.4. Discussion/Conclusions and implications

Key messages to emerge from the project

- The inverse relationship between wheat grain yield and grain N is robust and consistent
- Grain protein deviation (GPD) is a much more variable trait: with Cordiale and Marksman showing consistent positive GPD, and sometimes Hereward (but with a lower yield)
- The three years studied had very different weather conditions, which were reflected in the wide variation between the results obtained. 2010 and 2011 were very dry with negative impacts on N-responses (either yield and/or grain N)
- Variety and nitrogen inputs have defined but complex impacts on dough quality
- N-responsive genes have been identified, including genes encoding gluten proteins
- Genes correlating with GPD were identified including some which also reflect baking performance

Implications for breeders/producers

- Methodologies used in this study and the extensive correlations with quality, compositional and baking traits will provide leads for high throughput selection protocols
- Further screening will identify varieties which make more efficient use of applied N, in terms of translocation to the grain and incorporation into quality-related grain proteins.
- The project has demonstrated the potential for determining the genes and mechanisms responsible for GPD and quality traits, which will facilitate the development of new wheat lines with improved N utilisation

Implications for millers/bakers

- Reducing the use of N on breadmaking wheats is important to reduce costs and environmental/energy footprints
- It is possible to identify varieties with stable and high processing quality when grown at low N inputs using the approaches described.