Report to the Wheat Genetic Improvement Network (WGIN)

Sub contract No. 9363 between ADAS UK Ltd and Rothamsted Research
The work reported was carried out between May 2010 and September 2011

Non-destructive screening of WGIN Paragon mutants for grain NUE traits

By:

Richard Weightman, Roger Sylvester-Bradley, Daniel Kindred
ADAS Boxworth, Battlegate Rd, Boxworth, Cambridge

Chris Piotrowski, Steven Tayfield
Aunir Ltd, The Byre, Pury Hill Business Park, Towcester

Simon Orford, Joe McNeice, Simon Griffiths
JIC, Research Park, Colney, Norwich

Stuart Swanston
James Hutton Institute, Invergowrie, Dundee

6th October 2011

Introduction

Background

The EMS mutated Paragon population (EMPP) is a major resource but its trait data are sparse. This is because screening of the full EMPP is a real challenge when the number of lines is large (~6500), seed samples are valuable and quantities for testing or field phenotyping are small (<50g). However, scanners are now available to characterise a wide range of traits on small quantities of grain, quickly and non-destructively.

Image analysers (e.g. the Marvin grain analyzer) provide rapid assessments of individual seed weight and size (length & width). Further, standard Near Infra-Red (NIR) calibrations predict hardness and contents of grain protein, moisture and starch. In addition, successful NIR calibrations have been developed in the GREEN grain project (LK0959; Sylvester-Bradley et al., 2010) for protein components as well as for alcohol processing yield (litres per tonne; L/t). An opportunity therefore existed to screen existing seed of the EMPP lines for grain quality traits, allowing lines with altered expression to be identified for future investigation. Any variants for protein traits could facilitate development of N efficient ideotypes such as those proposed by the GREEN grain project or the WGIN.

In the GREEN grain project it was predicted that wheat with genetically reduced nitrogen (N) storage in both the canopy and the grain would give reduced N fertiliser requirements. Gliadins are responsible for most grain N storage in wheat (Kindred et al., 2008), so NIR calibrations for gliadin were developed and lines with low gliadin were sought. Variation in gliadin contents amongst elite wheats was slight. A wider search amongst candidate UK elite germplasm (and mapping populations derived from them) also showed disappointing variation. However, based on previous mutant studies there is a good chance that within the Paragon mutant lines will exist with contrasting levels of gliadin. If mutants could be identified with much reduced gliadin content (and reduced protein content per se, but without major disruption of grain formation or starch per grain), this could enable further investigation of genetic control of gliadin synthesis and deposition, and tests of whether low grain N storage improves N use.

This project therefore aimed to provide as much grain trait information as possible for the EMPP lines using non-destructive scanning techniques, to identify potentially interesting lines and to conclude on the potential for future investigation. The NIR calibrations and approach developed here could subsequently be used with other WGIN germplasm, as well as in commercial breeding programs.

Approach in the current project

This work was led by ADAS in collaboration with the John Innes Centre (JIC), James Hutton Institute (JHI), and NIR specialists, Aunir.

The calibrations originally developed in the GREEN grain project were developed for the FOSS 'Infratec' machine – the standard machine for commercial use by the grain industry. The Infratec has a narrow spectral range and cannot be used with small samples, so the utility of these new NIR calibrations could be enhanced by their transfer to more advanced research platforms such as the FOSS 6500, 5000 or XDS. These have a wider spectral range so can predict traits such as starch content more accurately, as well as being capable of analysing small sample sizes (~10g).

Development of calibrations for the FOSS 6500 NIR instrument entailed scanning of existing GREEN grain samples which have known reference data provided by ADAS for protein composition and by Scotch Whisky Research Institute for alcohol yield.

An NIR platform was provided for the project by Aunir. Reference samples (552) were transferred from JHI to Aunir, along with associated reference data from ADAS. Aunir scanned the reference samples on the FOSS XDS and developed the calibrations for predicting quality traits. The calibrations was prepared were transferred to a FOSS 6500.

The 6500 machine with a small sample cell was then transported to JIC and samples scanned from the EMPP with the FOSS 6500, and with the Marvin seed analyzer (already available at JIC) to provide additional estimates of mean weight per grain, grain length and grain width. It was estimated initially that at least 3000 lines would be scanned. In the end 4,737 were scanned, giving 4,477 lines with useable data. The remaining c. 30% of the EMPP was not scanned due to lack of further resources.

Materials and Methods

The following approach was taken:

1. Grain samples transferred from JHI and reference data from ADAS to Aunir,
2. Reference samples scanned on FOSS 6500 NIR at Aunir,
3. Calibrations developed for grain protein fractions and alcohol yield (AY),
4. NIR calibrations transferred to FOSS 6500 NIR to JIC (Norwich) and train staff to use with small sample cell
5. At least 3,000 samples from EMPP with 6500 and Marvin analysers,
6. Data transferred to ADAS for analysis.

The following describes the procedures in more detail:

Production of EMPP grain samples

Paragon is a spring wheat, which can be late autumn sown. The field experiment was sown in spring (March 15th 2006) by the JIC, and the plot size was 1m x 1m, made up of 1m rows (6 rows per plot giving a row width of 20 cm) using a Hege 90 drill, with 50 cm gaps between plots. One line was sown per row, with 'a' and 'b' lines (taken from the same M2 plant) in adjacent rows. A total N application rate of 150 kg/ha was applied in the spring. The soil was light and the 2006 season was very dry, so some plots suffered from drought stress.

Yield per row was not measured, but for reference, a 1m x 1m bed of Paragon control would yield around 800 g so the 2006 1 row yield was about 130 g. The EMPP mutants would vary hugely from this, the great majority lower.

Other samples were included within the trial; C = Control Paragon and Rht marker lines. Of these, 14 Control samples were scanned by NIR.

Grain reference dataset and data

Grain samples (n=552) from the GREEN grain project were transferred from JHI and ADAS to Aunir. Approximately half the samples (ca. 500 g) came from JHI, but half of the original dataset had been contaminated with weevils, and so were replaced with replicate samples which had been stored at ADAS in a cold store, but smaller samples (ca. 100 g).

These reference samples represented a range of elite varieties and mapping populations developed during the GREEN grain project over four seasons 2005-2009. The identity of these lines has been kept confidential because of their ownership by the GREEN grain consortium.

Data comprising the following key traits were supplied in an excel database:

- Alcohol yield (litres/tonne)
- Residue viscosity (mPa.s)
- Gliadin content (% DM)
- LMW Glutenin content (%DM)
- HMW Glutenin content (%DM)
- Albumin & Globulin content (%DM)
- Gliadin fraction (proportion of protein)
- LMW fraction (proportion of protein)
- HMW fraction (proportion of protein)
- Albumin & Globumin fraction (proportion of protein)

In addition, other measurements which were made during the GREEN grain project such as grain size and width measured by SKCS and Marvin analyzer were also converted to NIR calibrations, and on a small subset, some traits like alpha amylase activity, and various predictions of AY from protein or TGW and L:W ratio. However since the current project included actual TGW/GD data recorded using the Marvin analyzer, these calibrations are not reported.

Development of calibrations

After all samples had been scanned, calibrations were developed to provide NIR predictions of 38 key traits. Performance of the calibrations was monitored by the normal parameters, specifically standard deviation (SD) and Standard error of Cross validation (SECV). Decisions on whether a particular calibration had utility as a predictive tool was based on the Ratio of standard error of prediction to sample standard deviation (RPD = SD/SECV). In this context, the following rules were applied: RPD >5, as good as ref chemistry; RPD >2 acceptable; RPD <2 useful as a guide only; RPD <1 of no use. The list of 9 traits with the calibration performances which were used finally are listed in Annex 1. In addition, the standard FOSS European grain network calibrations for protein, moisture and starch were loaded onto the FOSS 6500 machine.

Scanning of grain samples from EMPP with FOSS 6500 and Marvin analysers

Samples were removed from the long term store at JIC and placed in a laboratory at ambient temperature and relative humidity to allow the grain moisture content to equilibrate to a normal level, at which NIR analysis could proceed. Average MC of scanned samples was 9.15% (SD 0.92).

Grain dimensions (GD: mean length, mean width and mean cross sectional area, as well as maximum and minimum for each parameter in each sample) were determined using a Marvin digital seed analyser (GTA Sensorik GmbH). Grain L:W ratio was calculated from the primary data. As samples were weighed prior to scanning, thousand grain weight (TGW) was measured concurrently, using the digital seed analyser to count the grains of a known weight. TGW was corrected to 85% moisture, using the moisture determination provided by the NIR predicted moisture content.

In total 4,737 samples were scanned at JIC using the FOSS 6500 and/or the Marvin analyzer (for grain dimensions and weight). Data taken forward to the final database only included those for which there was a full set of both NIR and GD/TGW data. Of the total number of samples examined, there were 244 samples which had to be excluded for the following reasons:

a) 187 samples for which the NIR spectral files were corrupted and hence the data unavailable,
b) 30 samples for which checks on the spectral data suggested the measurements could not be trusted (this may have been due to the sample size being too small to fill the cell, for instance), and;
c) 27 samples where there were data only for NIR, or only GD/TGW, but not both.

This left 4,477 samples remaining in the full data set. In addition 14 Paragon control (C) samples were included and scanned from within the EMPP population.

The EMPP population comprised two sets of lines – an ‘a’ and ‘b’ population. The 'a' line and the 'b' lines were originally taken from the same M2 plant and so had an early genetic connection. The a and b samples were each grown side by side in the field, but scanned by NIR and Marvin in blocks – all ‘a’ samples followed by the ‘b’ samples.

Data analysis

Data were aligned from the two sets (NIR and GD/TGW) to make a full data set. The Control samples were removed to be analyzed separately. Data could then be sorted based on high and low values for any trait, whichever the WGIN deem to be of interest.

Based on the field trial design (not presented here), the plots were divided into sets of Eastings and Northings based on the plot identifier, and within a plot, each set of six lines was coded (1-3) from edge to edge in order 1,2,3,3,2,1 (i.e. ear rows allocated treatment 1 represented the outer rows, and ear rows allocated treatment 3 were the central rows). For key grain traits (protein and grain dimensions) an ANOVA was carried out to quantify the relative proportions of the total variance which could be allocated to Eastings and Northings (essentially spatial variability within the field trial area) and within-plot variation (to test whether position within-plot was significant). This was achieved by carrying out stepwise linear regression where a mixed model was used with Eastings, Northings and within-plot position as variables. There was no adjustment for missing values.

Results and Discussion

Calibration performance

The list of 9 traits (with calibration performances) which were used finally are listed in Annex 1. In addition, the standard FOSS European grain network calibrations for protein, moisture and starch were loaded onto the FOSS 6500 machine. From the Marvin analyzer, traits for TGW, average length, average width, and average grain cross sectional area were measured as well as the maximum and minimum value recorded in each line.

General description of datasets

The datasets contrasted in their overall distributions. Based on standard grain characters, the Green Grain dataset had larger grains (up to 60 g TGW), with lower grain protein (down to 5.7 g/100g DM). This reflects the fact that the reference dataset was designed to find predominantly low protein grain, suitable for alcohol production and also to represent a wide range of genetics via elite wheat varieties and Syngenta mapping populations. It also represented a wide range of sites and seasons and contained a number of samples which received no nitrogen, and these typically had large grains (Table 1). However in terms of average grain size, TGW (46.9 g) was broadly representative of that seen in commercial varieties (NB. The TGW range in RL is 41 to 51 g).

In contrast, the EMPP was derived from one variety, Paragon, and was grown with fertiliser nitrogen. The trial was also droughted which may have resulted in fewer heads, fewer grains and smaller grains (Table 1), although there are no data on yields per line. It should be noted that the Paragon parent included in the lines here had an average protein of 14.6 g/100g DM, and average TGW of 37.3 g @85%DM (Table 2) i.e. a higher average protein and lower average TGW than the GREEN grain reference dataset (and than in RL variety data).

Paragon is a Group 1 variety for both late autumn and spring sowing. Its protein content is reported as 13.9 g/100g spring sown and 12.5 g/100g autumn sown (Recommended List, 2011/12). Therefore the Paragon samples in the present study had slightly higher (0.5 g/100g) grain protein than a typical spring-sown Paragon sample according to the RL.

The distributions of results from the EMPP also appeared skewed: there was a longer ‘low TGW’ tail (Figure 1) and a longer ‘high protein’ tail (Figure 2). The Green Grain reference dataset was more normally distributed.

Table 1. Thousand grain weight and protein content of Green Grain ref dataset (n=552) and EMPP (n=4,477).

Population:	GREEN Grain Ref dataset	EMPP
	Thousand grain weight (g@85%DM)	
Mean	46.9	34.1
Median	46.6	34.3
Max	62.0	49.2
Min	25.3	21.3
SD	4.90	3.82
95%UCL	56.5	41.6
95%LCL	37.3	26.7
	Protein (g/100gDM)	
Mean	9.5	15.2
Median	9.4	14.9
Max	15.1	23.1
Min	5.7	11.3
SD	2.00	1.60
95%UCL	13.4	18.3
95%LCL	5.5	12.0

Table 2. Thousand grain weight and protein content of 10 Paragon control samples grown within the EMPP.

Sample	ID	Protein (g/100g DM)	N	TGW (g@85% DM)
141C	Control	13.7	2	39.6
142C	Control	14.5	4	37.4
143C	Control	15.6	4	34.8
141B	Control	13.2	4	*
	Avge protein:	*14.6*	*Avge TGW:*	*37.3*

Based on these initial results, caution is needed with respect to interpreting the predictions arising from the EMPP. This is because the calibration data set is built on predominantly low protein grain (5.7 to 15.1 g/100gDM), and for traits other than total grain protein and moisture (which derive from standard FOSS calibrations) the mutant population is predominantly outside the range of the calibration dataset (albeit with some overlap between 11.3 and 15.1 g protein/100gDM). The GREEN grain reference dataset also did not include Paragon within the varieties tested.

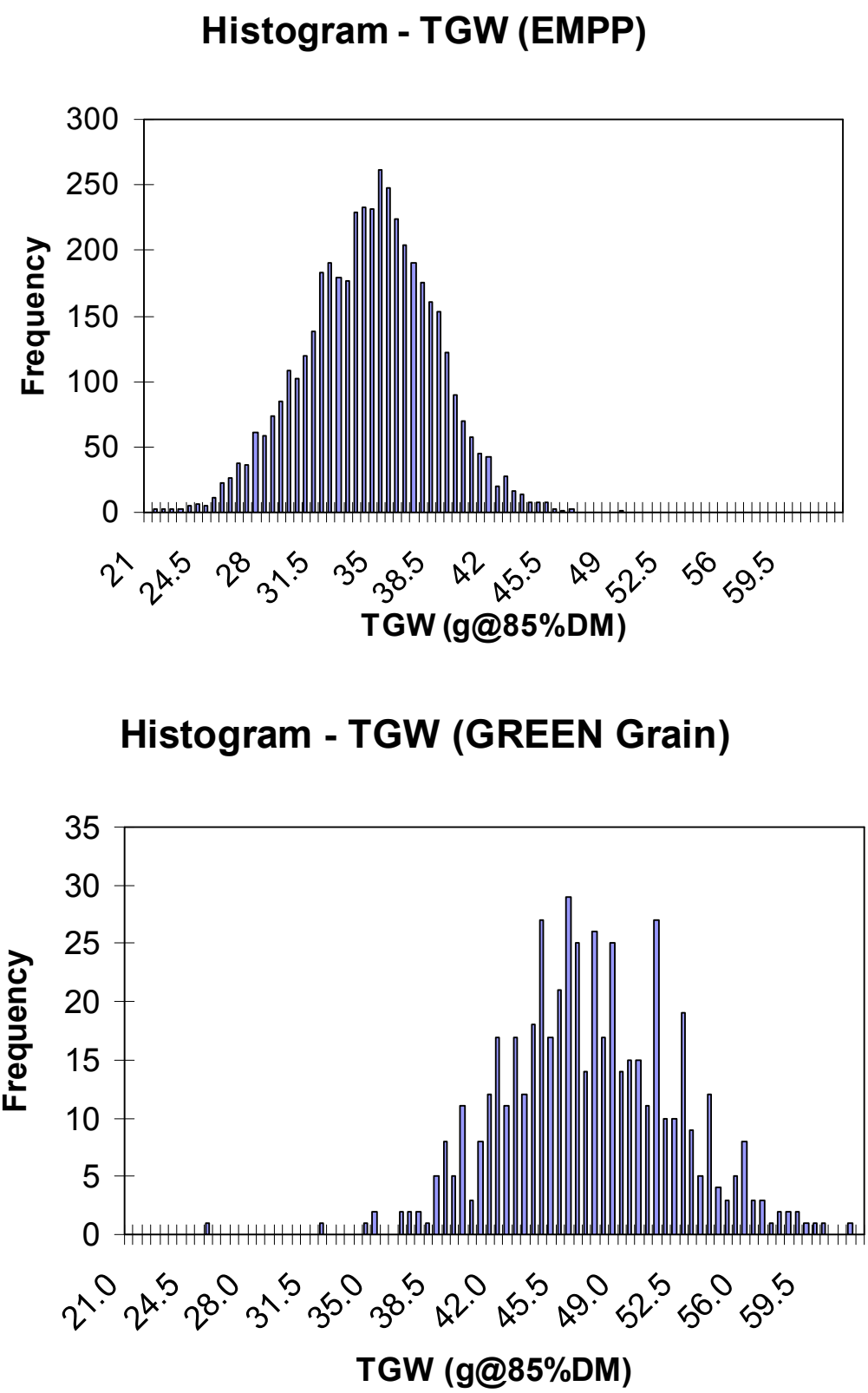
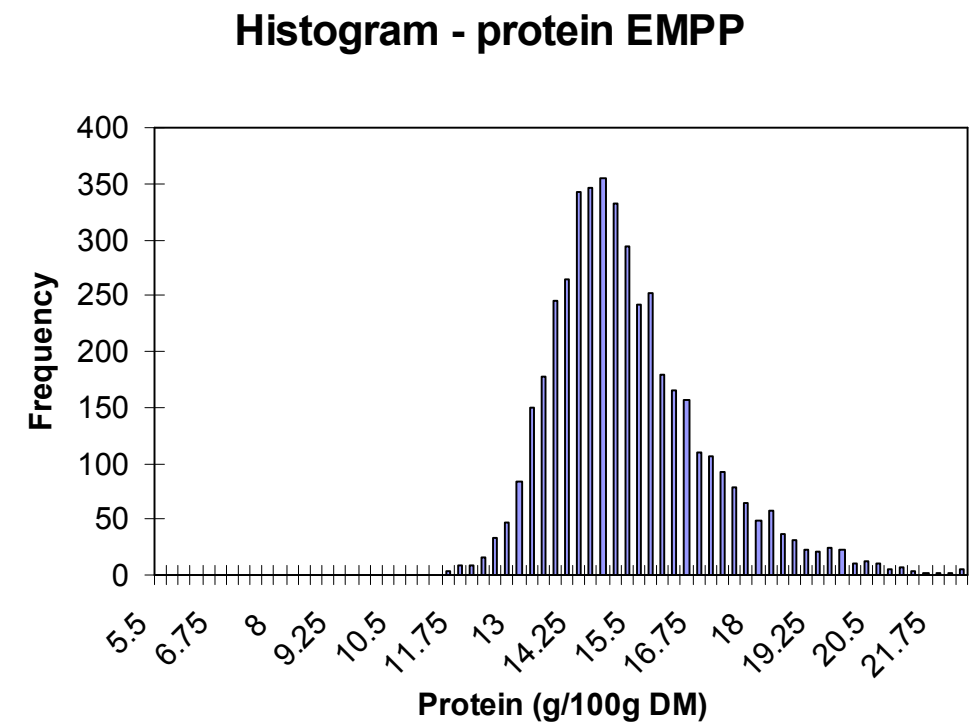


Figure 1. Thousand grain weight distributions for EMPP and GREEN Grain reference dataset (used to build NIR calibrations)



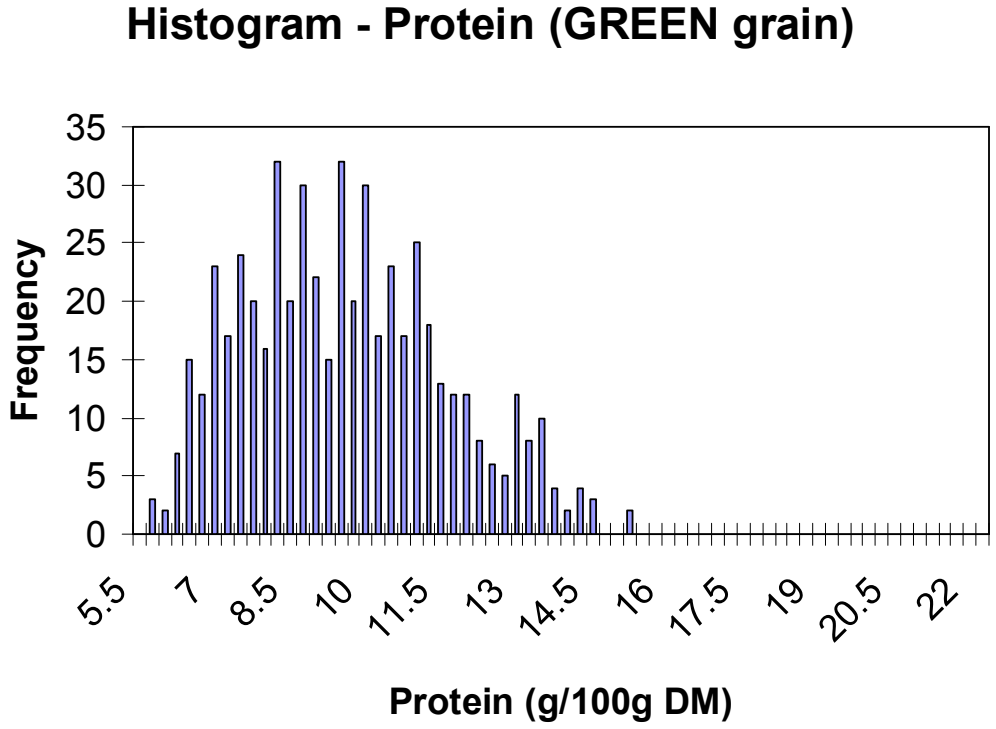


Figure 2. Grain protein distributions for EMPP and GREEN Grain reference dataset (used to build NIR calibrations)

Summary of key traits and relationships

Alcohol yield vs grain protein

Figure 3 shows NIR predicted AY plotted against grain protein. The original measurements of AY used to develop the calibration were made at the Scotch Whisky Research Institute using the 'wheat cook' method (Agu *et al*., 2006). The relationship between the two traits shows the typical negative linear relationship reported elsewhere (Smith *et al*., 2006; Kindred *et al*., 2008); very similar slopes (-6.3 and -6.8 L alcohol/10 kg protein respectively) are seen for EMPP and GREEN grain ref data. The relationship between AY and protein in the EMPP arises in part because the population encompasses a narrower range of genetics (i.e. all derived from Paragon), but it should be noted that the AY NIR calibration may well be measuring protein to a large extent, rather than genuinely predicting a chemical component directly related to alcohol, such as starch.

Alcohol yield vs starch

Starch was estimated using the standard FOSS starch calibration. The average starch content (68.1 g/100 gDM; SD 2.86) agrees well with that expected for 'typical' feed wheat grain but appeared wholly unresponsive to changes in grain protein content. The data plotted in Figure 4 demonstrate no relationship between AY and starch, or between starch and protein. Previous work on the Riband x Option dataset (Kindred *et al*., 2008) concluded that the NIR starch calibration typically failed for samples at very low protein concentrations, and it was decided that wet chemistry (specifically Ewers polarimetric method) was more reliable. Nevertheless, there does appear to be variation with some particularly high starch and high protein samples (Figure 4b) which may be of interest in the context of producing wheats which give high alcohol yield and high protein yields (Weightman et al., 2011) because they may produce a distillers grains (DDGS) co-product with low fibre content after processing. No further conclusions can be made about these samples until validation by wet chemistry for grain starch content is carried out.

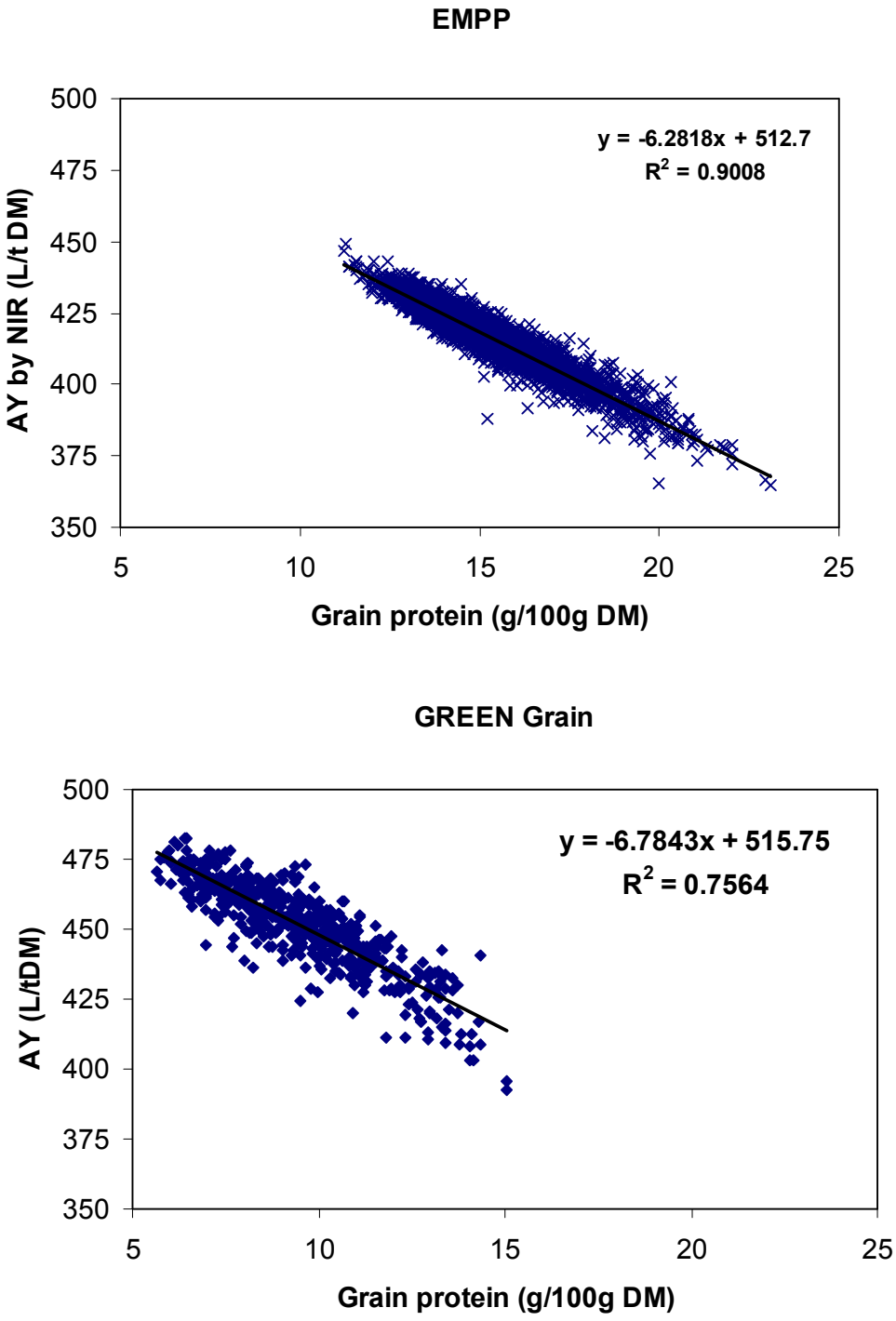
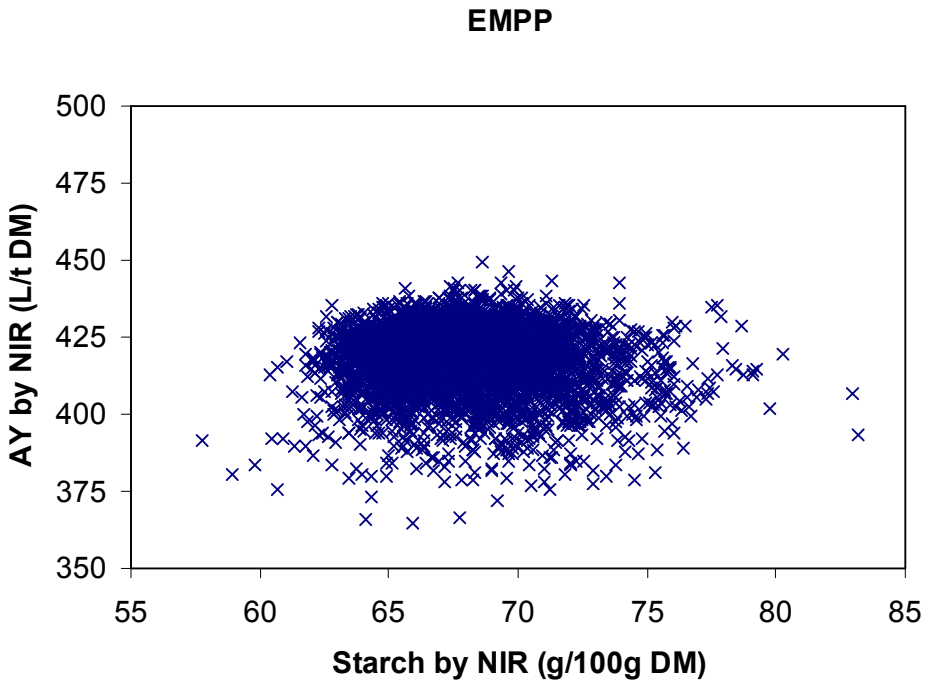


Figure 3. Alcohol yield (AY) vs grain protein for EMPP and GREEN Grain reference dataset. NB For GREEN grain dataset, AY data are based on wet chemistry measurements; all other data are based on NIR predictions.



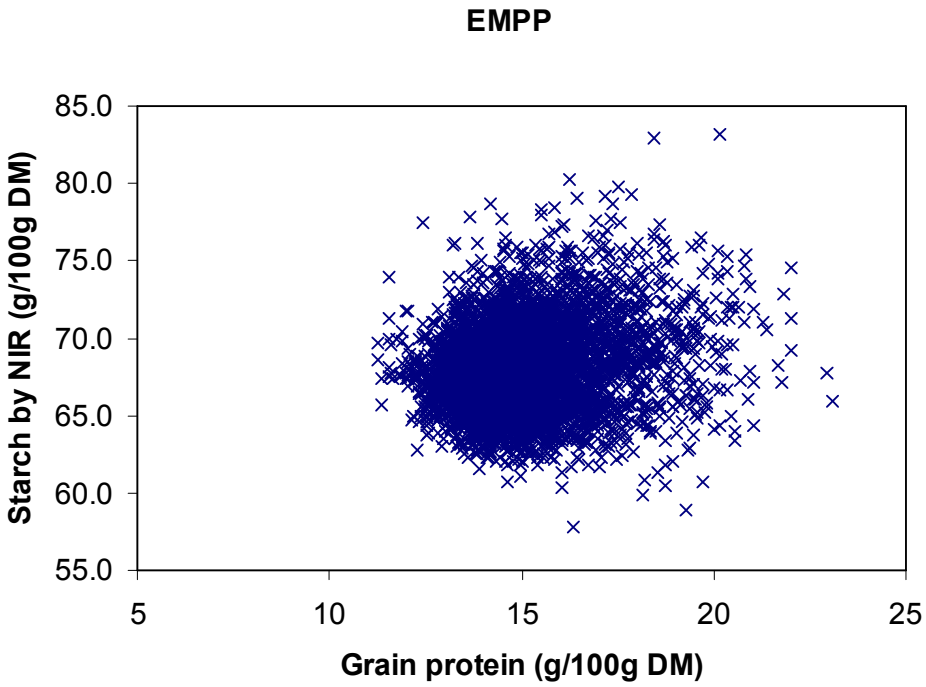


Figure 4. Alcohol yield (AY) vs starch (upper figure), and starch vs grain protein (lower figure) for EMPP dataset. NB All data are based on NIR predictions.

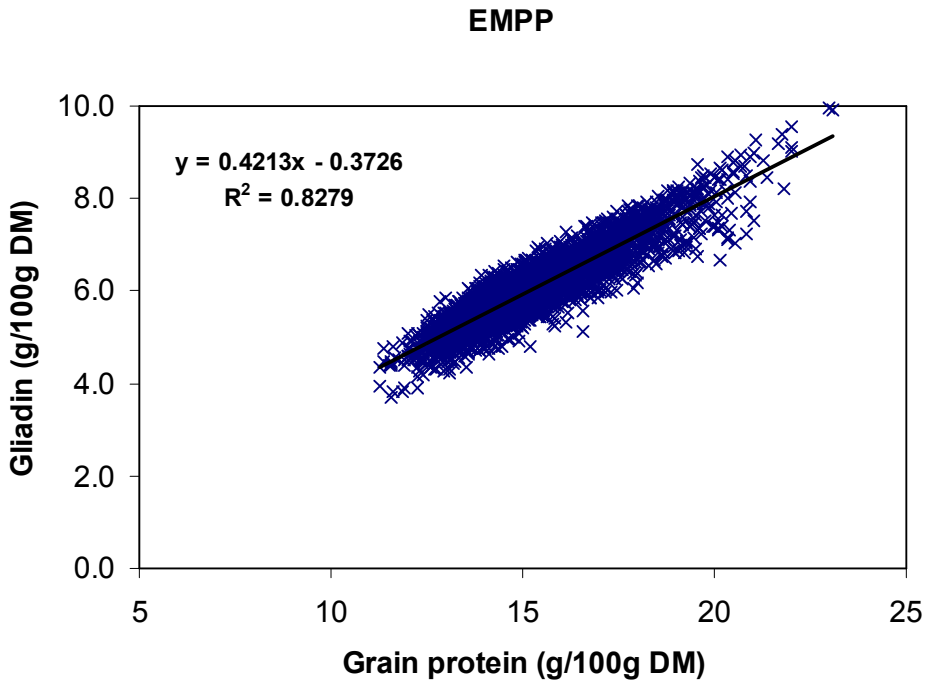
Gliadin content vs grain protein

Figure 5 shows the gliadin content in the EMPP population in relation to total grain protein. As might be expected, the two traits correlate closely and positively. The NIR calibration predicts the concentration of gliadin, based on the amount of total SDS-soluble protein extracted, and the proportion of gliadin within the extracted protein, determined by ADAS using the SE-HPLC method of Morel *et al*. (2000). This gliadin protein fraction was therefore estimated independently of the total grain protein figure (which is estimated by the standard FOSS NIR grain network calibration for both EMPP and GREEN Grain datasets). It can be seen from the slopes of the fitted equation in Figure 5, that the EMPP appears less responsive in terms of its gliadin than does the GREEN Grain reference dataset. It is not possible to say whether this is a feature of genetic and environmental changes in Paragon in particular. It may be that the grains in the EMPP have very high protein because gliadin deposition was curtailed prematurely, for instance if grain filling ceased when the crop was droughted.

The average gliadin content for the Paragon Control samples (described earlier at 14.6 g/100g average protein; Table 2) was relatively low at 38.6%, compared to a variety such as Riband, which would be expected to have a gliadin content of 47% at a similar level of total grain protein (using the equations in Kindred *et al*., 2008).

Until wet chemistry is carried out on the Paragon lines to validate gliadin content directly, and/or until the EMPP is grown in the field alongside some other reference material from elite varieties (like Riband) it is difficult to conclude whether these differences are significant.

Nevertheless, within the EMPP, there do appear to be a number of lines with low gliadin levels at a given grain protein content compared to the remainder of the population (Figure 5) e.g. sample 1195a with 32% gliadin at 11.6% total grain protein.



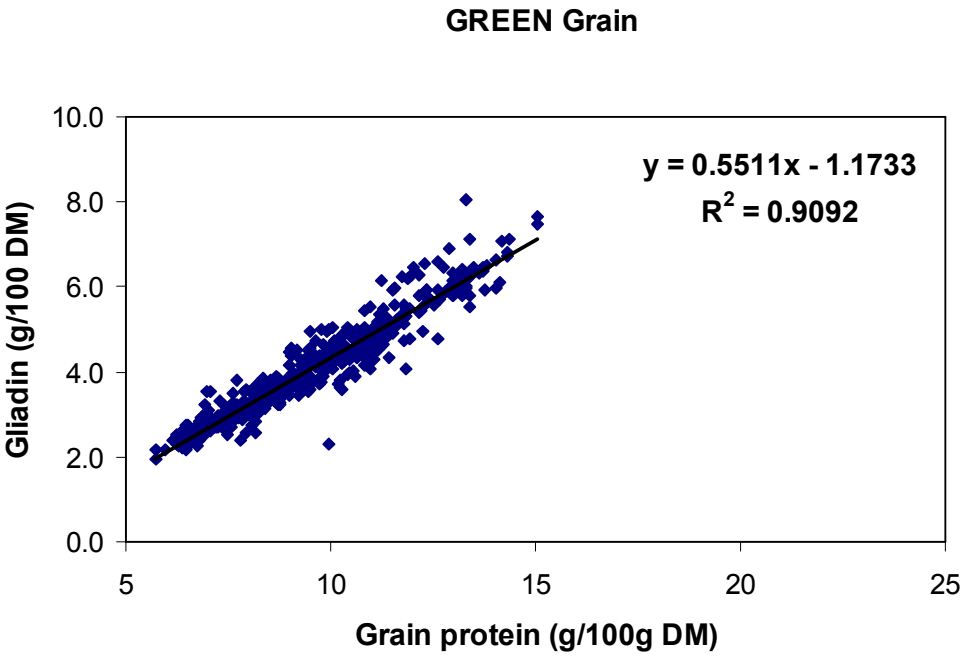


Figure 5. Gliadin protein vs total grain protein for EMPP and GREEN Grain reference dataset. NB For GREEN grain dataset, Gliadin data are based on wet chemistry (HPLC) measurements based on SDS-soluble protein; all other data are based on NIR predictions.

Grain protein

Table 3 summarises the lines with lowest and highest protein. In general, high protein lines were associated with small grains, suggesting that these might be shrivelled, or suffered from early senescence.

Table 3a. Five lowest protein samples in the EMPP dataset with associated grain dimensions.

Line	id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	Grain_ØArea (mm^2)	Grain_ØWidth (mm)	Grain_ØLength (mm)	TGW (g @85%DM)
2940	a	11.3	3.95	447	18.6	3.4	6.6	31.34
1157	b	11.3	4.34	449	18.6	3.5	6.5	34.35
578	b	11.4	4.77	441	18.3	3.5	6.5	34.14
1566	a	11.4	4.47	441	19.9	3.6	6.6	37.35
2900	a	11.5	4.38	442	19.6	3.5	6.7	35.01

Table 3b. Five highest protein samples in the EMPP dataset with associated grain dimensions.

Line	id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	Grain_ØArea (mm^2)	Grain_ØWidth (mm)	Grain_ØLength (mm)	TGW (g @85%DM)
2553	a	22.0	9.10	379	17.3	3.2	6.5	24.93
2045	a	22.0	9.01	376	17.4	3.3	6.4	24.87
2459	a	22.0	9.54	372	18.3	3.3	6.6	27.93
2096	a	23.0	9.94	367	16.8	3.2	6.2	22.97
2097	a	23.1	9.90	365	17.4	3.4	6.2	25.15

Alcohol yield and Residue viscosity

The extreme samples in terms of alcohol yield (AY) are shown in Table 4. The absolute AY values principally arose from samples with very high or low protein grain (e.g. the five lowest AY samples, not surprisingly had grain proteins >20 g/100g.

More discrimination was applied by using the regression equation from Figure 3 (AY = -0.628 x protein + 512.7) to give a predicted AY for each measured grain protein value, and this was then used to estimate the 'deviations' between observed and predicted values, to remove the dominant protein effect on AY. The deviations of AY are shown in Table 5. In this way, the results are less influenced by grain protein content. For example sample 997a (Table 5a) had a low protein content (for this dataset; 15.2 g/100g) but an AY which was 29 L/t less than predicted

by the common equation. Similarly, sample 1763a (Table 5b) had a very high grain protein content (20.3 g/100g), but an AY 15.9 L/t greater than predicted. Interestingly, despite the poor overall precision of the FOSS starch calibration as seen earlier (Fig 4), high starch content appeared to correlate with the positive deviations for alcohol yield as seen in Table 5b. For example line 2187a with the second highest AY for its grain protein content at +14.1 L/t had a high starch content at 79 g/100gDM.

Table 4a. Five lowest alcohol yield (AY) samples in EMPP dataset with associated grain dimensions and grain protein.

Line	id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	Grain_ØArea (mm^2)	Grain_ØWidth (mm)	Grain_ØLength (mm)	TGW (g @85%DM)
2097	a	23.1	9.90	365	17.4	3.4	6.2	25.15
450	a	20.0	8.44	366	21.6	3.8	7.0	34.13
2096	a	23.0	9.94	367	16.8	3.2	6.2	22.97
2459	a	22.0	9.54	372	18.3	3.3	6.6	27.93
3033	a	21.1	9.27	373	16.4	3.2	6.1	23.35

Table 4b. Five highest alcohol yield (AY) samples in EMPP dataset with associated grain dimensions and grain protein.

Line	id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	Grain_ØArea (mm^2)	Grain_ØWidth (mm)	Grain_ØLength (mm)	TGW (g @85%DM)
2882	a	11.6	4.38	443	19.3	3.5	6.8	32.91
2577	a	12.4	4.94	443	19.1	3.5	6.6	38.04
2923	a	11.6	4.45	443	19.1	3.6	6.5	35.26
2940	a	11.3	3.95	447	18.6	3.4	6.6	31.34
1157	b	11.3	4.34	449	18.6	3.5	6.5	34.35

Table 5a. Five lowest deviations for alcohol yield (AY) estimated as Observed AY-predicted AY using the regression equation in Figure 3, for samples in EMPP dataset, with associated protein, starch and TGW measurements.

Line	id	Grain protein (g/100g DM)	Starch (g/100gDM)	AY (L/t DM) measured	TGW (g @85%DM)	AY (L/tDM) predicted	AY deviation (obs-pred) (L/tDM)
997	a	15.2	68.3	388	32.38	417.2	-29.4
450	a	20.0	64.1	366	34.13	387.1	-21.5
474	a	16.4	57.8	392	32.61	409.9	-18.4
460	a	18.1	59.8	383	27.99	398.8	-15.4
575	a	18.5	68.4	381	38.57	396.6	-15.3

Table 5b. Five highest deviations for alcohol yield (AY) estimated as Observed AY-predicted AY using the regression equation in Figure 3, for samples in EMPP dataset, with associated protein, starch and TGW measurements.

Line	id	Grain protein (g/100g DM)	Starch (g/100gDM)	AY (L/t DM) Measured	TGW (g @85%DM)	AY (L/tDM) predicted	AY deviation (obs-pred) (L/t DM)
2246	a	18.7	75.3	408	32.77	395	+12.5
2573	a	17.5	70.1	416	30.82	403	+13.3
2935	a	14.5	77.7	435	33.33	422	+13.5
2187	a	17.9	79.3	414	39.30	400	+14.1
1763	a	20.3	72.0	401	22.52	385	+15.9

In addition to AY, the other variety factor of importance to the traditional distiller, is residue viscosity (RV) which causes problems during processing. Table 6 shows the highest and lowest RV samples withing the EMPP dataset. High RV is thought to be due to high arabinoxylan (AX) content. However, the distillers in general, do not like high protein grain as it is also known to give processing problems. Other work by ADAS (Davis-Knight et al., 2010) has shown that RV increases with N rate and protein content, indicating that RV could be driven by soluble protein concentration, rather than AX *per se*. The data here tend to confirm this, as the samples with low RV are lower protein (Table 6a) than those with high RV (Table 6b).

Table 6a. Five lowest Residue viscosity (RV) samples in EMPP dataset, with associated protein, starch and AY measurements.

Ref	id	Grain protein (g/100g DM)	Starch (g/100gDM)	Gliadin (g/100gDM)	RV (mPa.s)	AY (L/tDM)	AY deviation (obs-pred) (L/tDM)
425	a	16.9	60.5	6.32	1.229	408	1.376
570	a	14.3	67.2	5.86	1.240	420	-3.071
1042	a	12.3	64.0	4.16	1.247	430	-5.497
427	a	14.1	62.0	5.05	1.264	422	-2.438
83	a	12.5	59.1	4.66	1.271	432	-1.514

Table 6b. Five highest Residue viscosity (RV) samples in EMPP dataset, with associated protein, starch and AY measurements.

Ref	id	Grain protein (g/100g DM)	Starch (g/100gDM)	Gliadin (g/100gDM)	RV (mPa.s)	AY (L/tDM)	AY deviation (obs-pred) (L/tDM)
1990	a	20.1	76.4	6.65	1.635	393	7.067
3077	a	21.8	61.2	9.41	1.645	378	1.946
2530	a	18.8	65.4	7.65	1.647	398	3.835
3074	a	20.9	66.1	7.91	1.656	380	-1.094
2989	a	19.9	68.1	8.10	1.684	392	4.192

Grain dimensions

The following Tables (7-10) give the highest and lowest sets of five lines, for each of the grain dimensions TGW, cross-sectional area, length and width. For the latter three traits, the values represent the average for each sample, whereas in the database are also presented the maximum and minimum values scanned for each sample.

Table 7a. Five lowest thousand grain weight samples in the EMPP dataset with associated grain dimensions and protein.

Line	Id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	Grain_ØArea (mm^2)	Grain_ØWidth (mm)	Grain_ØLength (mm)	TGW (g @85%DM)
718	B	18.5	8.05	393	15.4	3.0	6.1	21.3
2457	A	18.8	8.16	387	16.1	3.1	6.2	21.3
2489	A	20.0	8.12	385	16.8	3.2	6.2	21.6
3353	A	19.4	8.34	392	15.6	3.0	6.0	21.6
1020	A	19.4	7.87	385	16.7	3.2	6.2	22.0

Table 7b. Five highest thousand grain weight samples in the EMPP dataset with associated grain dimensions and protein.

Line	id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	Grain_ØArea (mm^2)	Grain_ØWidth (mm)	Grain_ØLength (mm)	TGW (g @85%DM)
994	a	16.4	5.94	406	22.8	3.9	7.0	45.9
1916	a	16.6	6.34	408	23	4.0	7.1	46.1
2392	a	17.4	6.93	401	22.6	4.0	6.9	46.4
2634	a	17.3	6.40	411	21.5	4.1	6.6	46.4
1582	b	18.0	6.77	406	24.8	4.1	7.5	49.2

Table 8a. Five lowest grain cross sectional area samples in the EMPP dataset with associated grain dimensions and protein.

Line	id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	Grain_ØArea (mm^2)	Grain_ØWidth (mm)	Grain_ØLength (mm)	TGW (g @85%DM)
718	b	18.5	8.05	393	15.4	3.0	6.1	21.3
3	a	17.0	6.95	406	15.4	3.2	5.7	27.4
439	b	19.7	8.51	375	15.5	3.1	6.0	22.2
33	b	15.6	6.73	413	15.5	3.2	5.8	25.5
3353	a	19.4	8.34	392	15.6	3.0	6.0	21.6

Table 8b. Five highest grain cross sectional area samples in the EMPP dataset with associated grain dimensions and protein.

Line	id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	Grain_ØArea (mm^2)	Grain_ØWidth (mm)	Grain_ØLength (mm)	TGW (g @85%DM)
2062	a	20.1	7.46	393	23	3.9	7.1	43.4
1916	a	16.6	6.34	408	23	4.0	7.1	46.1
1961	a	16.1	6.35	411	23.5	3.9	7.4	40.6
776	a	17.6	6.29	408	23.6	4.0	7.1	43.5
1582	b	18.0	6.77	406	24.8	4.1	7.5	49.2

Table 9a. Five samples with lowest grain width in the EMPP dataset with associated grain dimensions and protein.

Line	id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	Grain_ØArea (mm^2)	Grain_ØWidth (mm)	Grain_ØLength (mm)	TGW (g @85%DM)
718	b	18.5	8.05	393	15.4	3.0	6.1	21.3
3353	a	19.4	8.34	392	15.6	3.0	6.0	21.6
28	a	19.0	8.19	390	15.9	3.0	6.2	23.7
439	b	19.7	8.51	375	15.5	3.1	6.0	22.2
2457	a	18.8	8.16	387	16.1	3.1	6.2	21.3

Table 9b. Five samples with highest grain width in the EMPP dataset with associated grain dimensions and protein.

Line	id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	Grain_ØArea (mm^2)	Grain_ØWidth (mm)	Grain_ØLength (mm)	TGW (g @85%DM)
624	a	19.1	7.00	388	22.7	4.0	6.9	45.1
1916	a	16.6	6.34	408	23	4.0	7.1	46.1
776	a	17.6	6.29	408	23.6	4.0	7.1	43.5
2634	a	17.3	6.40	411	21.5	4.1	6.6	46.4
1582	b	18.0	6.77	406	24.8	4.1	7.5	49.2

Table 10a. Five samples with lowest grain length in the EMPP dataset with associated grain dimensions and protein.

Line	Id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	Grain_ØArea (mm^2)	Grain_ØWidth (mm)	Grain_ØLength (mm)	TGW (g @85%DM)
732	a	21.4	8.45	377	16.6	3.8	5.5	25.6
3	a	17.0	6.95	406	15.4	3.2	5.7	27.4
2979	a	16.3	6.95	404	15.8	3.3	5.7	26.2
409	b	17.2	6.56	408	16.1	3.4	5.7	28.6
791	a	17.3	7.33	401	16.7	3.5	5.7	29.2

Table 10b. Five samples with highest grain length in the EMPP dataset with associated grain dimensions and protein.

Line	Id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	Grain_ØArea (mm^2)	Grain_ØWidth (mm)	Grain_ØLength (mm)	TGW (g @85%DM)
776	a	17.6	6.29	408	23.6	4.0	7.1	43.5
1381	a	19.5	7.47	388	21.7	3.6	7.2	35.1
1084	b	18.3	7.06	404	22.3	3.7	7.4	42.4
1961	a	16.1	6.35	411	23.5	3.9	7.4	40.6
1582	b	18.0	6.77	406	24.8	4.1	7.5	49.2

Gliadin analysis

The NIR predictions reported gliadin content as g/100g in grain dry matter, based on an original measurement of extracted protein quantified during the SE-HPLC analysis procedure. When this value was divided by the total grain protein (estimated independently using the standard FOSS grain network protein calibration) then the proportion of gliadin in the total protein could be

expressed as a percentage. Because gliadins increase in response to increasing grain protein, the data need to be treated in the same way as AY earlier, to look for deviations of the proportion of gliadins compared to the proportion which would be predicted from total protein alone.

More discrimination was therefore applied to the data by using the regression equation from Figure 5 (Gliadin = 0.4213 x protein - 0.3276) to give a predicted gliadin for each measured grain protein value. This was then divided by total grain protein and multiplied by 100 to generate a predicted value for the proportion of gliadin. This value was then used to estimate the 'deviations' between observed and predicted values, to remove the dominant protein effect. The deviations in the proportions of gliadin are shown in Table 11.

It can be seen that those lines in the EMPP with gliadins lower than predicted by the common equation, are associated with lower protein grain , with some exceptions e.g. line 1990a (Table 10a), which had a high grain protein content at 20 g/100g. Lines with 'high gliadin (i.e. observed proportion of gliadin was greater than that predicted by the common regression equation) tended to be associated with low to average protein (12.9-14.5 g/100g; Table 10b).

Table 11a. Five lowest deviations for the proportion of gliadins estimated as Observed-predicted gliadin using the regression equation in Figure 5, for samples in EMPP dataset, with associated protein, AY and TGW measurements.

Line	Id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	TGW (g @85%DM)	Prop. gliadin (Gliad/total protein x 100) measured	Prop. gliadin (Gliad/total protein x 100) predicted	Prop. gliadin deviation (obs-pred)
1077	a	16.6	5.13	412	38.19	31.0	40.2	-9.2
1094	a	15.2	4.78	411	40.05	31.4	40.0	-8.5
1159	a	12.3	3.90	438	35.24	31.8	39.5	-7.7
1990	a	20.1	6.65	393	26.79	33.0	40.5	-7.5
1865	a	13.5	4.36	433	36.95	32.3	39.7	-7.4

Table 11b. Five highest deviations for the proportion of gliadins estimated as Observed-predicted gliadin using the regression equation in Figure 5, for samples in EMPP dataset, with associated protein, AY and TGW measurements.

Line	Id	Grain protein (g/100g DM)	Gliadin (g/100gDM)	AY (L/t DM)	TGW (g @85%DM)	Prop. gliadin (Gliad/total protein x 100) measured	Prop. gliadin (Gliad/total protein x 100) predicted	Prop. gliadin deviation (obs-pred)
1513	b	14.3	6.43	419	32.94	45.0	39.8	+5.1
1239	b	12.9	5.77	429	32.89	44.9	39.6	+5.3
1275	b	14.5	6.55	417	31.71	45.2	39.9	+5.4
1493	b	13.0	5.85	433	31.80	45.0	39.6	+5.4
1259	b	13.9	6.35	416	38.13	45.6	39.8	+5.8

Other traits

For completeness, key grain quality parameters were matched with observations of a set of mutant phenotypes recorded in the field by JIC staff. These are presented in Table 12.

Table 12. Grain protein and dimensions for some key mutant lines identified by JIC staff in the field in 2006.

Line	Id	Mutation	Protein (g/100g DM)	ID check	TGW (g@85%DM)	Area (mm^2)	Width (mm)	Length (mm)	L:W ratio
12	a	leaf disease mimic	19.3	12a	26.2	17.9	3.2	6.4	2.00
16	a	tall line	16.9	16a	34.2	17.7	3.5	5.9	1.69
16	b	tall line	16.4	16b	40.0	20.3	3.8	6.6	1.74
170	b	early ear	14.2	170b	36.4	18.8	3.6	6.3	1.75
344	b	stem twist	14.6	344b	35.9	18.9	3.5	6.5	1.86
355	a	Square seed	18.2	355a	31.0	16.9	3.5	5.7	1.63
468	a	tall line	15.6	468a	34.0	19.2	3.5	6.5	1.86
468	b	tall line	13.4	468b	41.3	20.2	3.7	6.7	1.81
545	b	leaf twist	16.1	545b	30.3	17.5	3.3	6.4	1.94
561	a	early ear	13.9	561a	27.8	17.8	3.3	6.5	1.97
590	a	Lax ear	16.0	590a	34.0	19.6	3.6	6.6	1.83
670	a	aborted flag leaf	13.5	670a	35.6	19.8	3.6	6.7	1.86
725	b	Earliest ear	*	*	*	*	*	*	*
734	b	club ear	15.4	734b	38.5	20.8	3.7	6.9	1.86
864	b	club ear	*	*	*	*	*	*	*
1223	b	stripped leaf	14.3	1223b	30.8	18.1	3.4	6.4	1.88
1381	a	long seed	19.5	1381a	35.1	21.7	3.6	7.2	2.00
1421	a	round seed	*	*	*	*	*	*	*
1439	a	late ear	*	*	*	*	*	*	*
1926	a	white seed	*	*	*	*	*	*	*
1946	a	club ear	15.7	1946a	36.2	19.6	3.8	6.3	1.66
2172	a	black glumes	13.9	2172a	35.5	19.4	3.6	6.6	1.83
2206	a	Lax ear	17.2	2206a	35.4	18.6	3.7	6.0	1.62
2206	b	Lax ear	*	*	*	*	*	*	*
2218	a	sterile in mid ear	15.6	2218a	39.6	21.0	3.8	6.7	1.76
2484	a	club ear	*	*	*	*	*	*	*
2521	a	short line	*	*	*	*	*	*	*
2521	b	short line	*	*	*	*	*	*	*
2566	b	very club ear	*	*	*	*	*	*	*
2626	a	short line	*	*	*	*	*	*	*
2626	b	short line	*	*	*	*	*	*	*
2637	a	long seed	16.3	2637a	36.9	19.8	3.6	6.6	1.83
2644	a	tiny florets in mid ear	16.8	2644a	29.4	18.3	3.4	6.6	1.94
2734	a	large seed	16.0	2734a	33.7	19.1	3.5	6.5	1.86
2840	a	light green ear	16.3	2840a	36.0	20.7	3.6	6.8	1.89
2870	b	dark leaf	*	*	*	*	*	*	*
2939	a	late ear	*	*	*	*	*	*	*
3017	a	succulent leaf	*	*	*	*	*	*	*
3042	a	clean seed	*	*	*	*	*	*	*
3042	b	clean seed	*	*	*	*	*	*	*
3104	a	club ear	17.9	3104a	31.6	17.9	3.5	6.1	1.74
3143	a	white seed	*	*	*	*	*	*	*
3143	b	white seed	*	*	*	*	*	*	*
3279	a	small florets in mid ear	13.6	3279a	34.6	19.0	3.5	6.5	1.86

Note on spatial variability within the field trial area

There was some concern that, because of visible areas of poor fertility and/or droughtiness within the trial area, spatial variation might be larger than genetic variation. Moreover, with small plots (1 m^2) each containing six rows with only one line per row, it was possible that the plants in

the outer rows had access to more nutrients, water and light than inner rows. In this case any variation seen in quality traits may be incorrectly attributed to genetics, whereas in fact it could be due to environment. Spatial variability in grain protein content is shown in Figure 6.

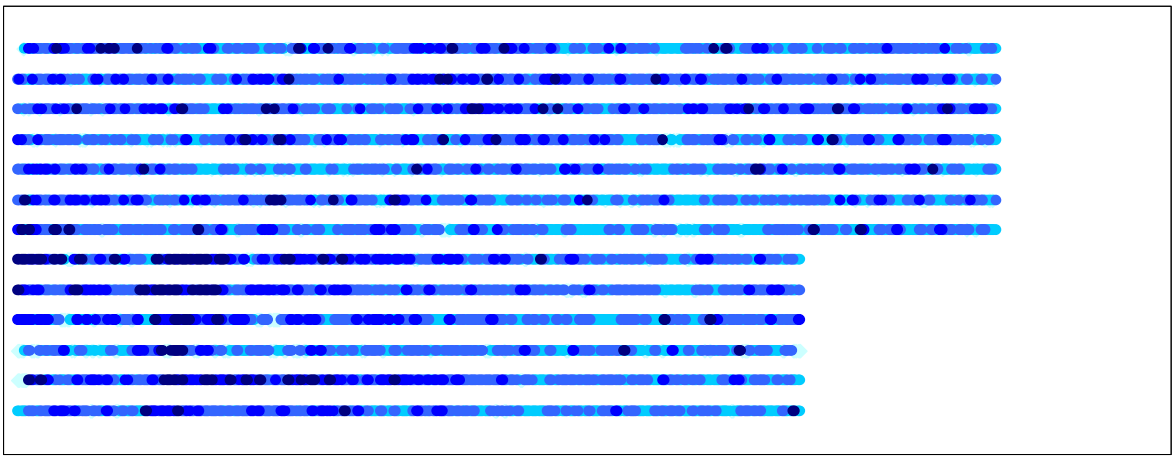


Figure 6. Schematic diagram representing spatial variation in protein content within the EMPP field trial 2006, where darkest blue plots represent high protein (>19 g/100gDM) and lightest blue, low protein (<13 g/100gDM).

Given the potential significance of spatial variability identified in Figure 6, a statistical analysis (ANOVA with stepwise linear regression) was carried out to partition the total variance for key grain traits, between Eastings, Northings and 'within plot' position. All three factors had highly significant effects on grain quality traits, although the sum of their effects never accounted for more than 15% of the total variance (Table 13). Thousand grain weight was most affected by position within the trial. It is desirable that these systematic effects be removed before more comprehensive analysis of this new dataset.

The results suggest that while spatial variation appears to be a relatively small part of the total variance, care should be taken to check the exact position of lines within the trial before these are selected for further analysis. Position within the plot appeared to be the least significant source of variation.

It is possible that inter-plot spatial variation could be analysed in a more sophisticated way, to identify and account for spatial effects that do not align as Eastings and Northings.

Table 13. Partitioning of total variance by ANOVA between areas within field trial (defined blocks of plots arranged as Eastings and Northings) and within plot position (from outer to central rows).

Trait		Partitioning of variance *Term* Eastings	Northings	Position within plot	Residual	Total	Variance accounted for (%)*
Protein (g/100g DM)	SS	737.6	64.0	63.8	19,462	20,328	4.22
	(% of total)	*3.63*	*0.31*	*0.31*			
	Sig	***	***	***			
TGW (g@85%DM)	SS	10,705	3,265	1,400	89,656	105,026	14.60
	(% of total)	*10.19*	*3.11*	*1.33*			
	Sig	***	***	***			
Area (mm^2)	SS	516	130	97	7,935	8,678	8.52
	(% of total)	*5.95*	*1.50*	*1.11*			
	Sig	***	***	***			
Length (mm)	SS	2.16	0.78	0.36	273	276	1.15
	(% of total)	*0.78*	*0.28*	*0.13*			
	Sig	***	***	**			
Width (mm)	SS	13.95	3.98	2.14	167	187	10.67
	(% of total)	*7.46*	*2.13*	*1.14*			
	Sig	***	***	***			

* Linear regression model with all three terms included.

Summary

This project directly supported a number of WGIN's objectives, specifically Resource development: Paragon gamma and EMS mutant lines (Objective 4) and Targeted traits: Improvements of nitrogen use efficiency and quality QTLs linked to NUE (Objective 8).

The NIR technology allowed phenotyping in a quick and non-destructive manner, without risking damage to the samples. Once spectral data had been collected, it was possible to make predictions for new variables, and to link these to grain dimensions measured by physical methods. This project has thus delivered phenotype data for a large subset (ca. 70%) of the Paragon mutant population, which is available for reference to the wheat breeding community.

Traits assessed in this project are important both for grain quality and for nitrogen efficiency. Reducing protein content, for example via reduced gliadin content, could be associated with

reduced N fertiliser requirement. Low protein and high alcohol yield are target traits for distilling and bioethanol markets. Protein composition is important for breadmaking quality, especially the content of gliadin and LMW glutenins.

The results presented here are inevitably a short summary of the available data. This report must be viewed in conjunction with the accompanying dataset. The whole EMPP and its controls (Paragon parents) appear to be relatively low gliadin for a given protein content, compared to the elite lines seen previously in the GREEN Grain project. However, it would be unwise to rely too strongly on this conclusion before confirmatory wet chemistry is carried out, because the GREEN grain reference dataset (on which the NIR calibrations were built) comprised primarily:

- Winter types,
- Grains of soft endosperm texture,
- Low grain protein samples (<15%) due to N deficiency,

and;

- Varieties other than Paragon.

Therefore it would be unwise to extrapolate too far in any conclusion based on an EMPP which is a quite different population from the one on which the NIR calibrations were built. In future it would be sensible to include standard reference varieties of elite lines (e.g, Riband, Glasgow, Solstice, etc.) about which normal performance is known. As Paragon is a spring wheat, this may be difficult to achieve in practice, unless the EMPP is drilled in late autumn.

Nevertheless, it is still valid to identify those lines within the EMPP which perform better or worse than their peers, based on the response to variation in a major trait like grain protein. There are several of these, as we have demonstrated for AY and the proportion of gliadins.

Real progress will only be made when users begin to interrogate the database, identify potentially interesting lines, and use reference (wet) chemistry to confirm actual grain quality.

Acknowledgements

The authors would like to thank the UK Department for Environment Food and Rural Affairs, and the Wheat Genetic Improvement Network and its Management Committee for financial support.

References

Agu, R. C., Bringhurst, T. A., Brosnan, J. M., 2006. Production of grain whisky and ethanol from wheat, maize and other cereals. *Journal of the Institute of Brewing* **112**(4), 314-323.

Davis-Knight, H., Weightman, R.M., Agu, R., Bringhurst, T. & Brosnan J. (2010). Effects of fertiliser nitrogen and variety on the concentrations of non-starch polysaccharides in winter wheat for alcohol production. *Aspects of Applied Biology* **101** Non-Food Uses of Crops, 155-162.

Kindred, D. R., Verhoeven, T. M. O., Weightman, R. M., Swanston, J. S., Agu, R. C., Brosnan, J.M. and Sylvester-Bradley, R. (2008). Effects of variety and fertiliser nitrogen on alcohol yield, grain yield, starch and protein content, and protein composition of winter wheat. *Journal of Cereal Science* **48**, 46-57.

Morel, M.-H., Dehlon, P., Autran, J. C., Leygue, J. P., 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**, 685-691.

Smith, T. C., Kindred, D. R.. Brosnan, J. M., Weightman, R. M., Shepherd, M. and Sylvester-Bradley, R. (2006). *Wheat as a feedstock for alcohol production*. HGCA Research Review, No. 61. Home Grown Cereals Authority, Caledonia House, 223 Pentonville Rd, London, N1 9NG, December 2006.

Sylvester-Bradley, R., Kindred, D.R., Weightman, R.M., Thomas, W., Swanston, J.S., Thompson, D., Feuerhelm, D., Creasy, T., Argillier, O., Melichar, J., Brosnan, J., Agu, R., Bringhurst, T., Foulkes, J.F., Pask, A., Cowe, I., Hemingway, D., Robinson, D., and Wilcox, S. (2010). *Genetic reduction of energy use and emissions of nitrogen through cereal production: GREEN grain*. HGCA Project report No. 468. Home Grown Cereals Authority, Stoneleigh Park, Kenilworth, Warwickshire, CV8 2TL, UK, June 2010.

Weightman, R.M., Kindred, D.R., Sylvester-Bradley, R., Clarke, S. and Berry, P.M. (2011). Productivity, sustainability and land use aspects of biofuel crops in UK arable rotations. *Aspects of Applied Biology* **112** Biomass and Energy Crops IV, 171-180.

Annex 1

Calibration performance of key traits

Constituent	N	Mean	SD	SEC	RSQ	SECV	1-VR	RPD
GLDDMHPLC	454	4.95	1.360	0.680	0.750	0.700	0.735	1.943
AlcYieldldryt	533	452	14.8	7.2	0.800	7.521	0.742	1.969
ResViscosy	545	1.48	0.137	0.065	0.774	0.069	0.744	1.977
LMWDMProt	463	1.96	0.554	0.263	0.774	0.274	0.755	2.021
GLDDMProt	448	4.06	1.117	0.471	0.822	0.488	0.809	2.287
Moisture	528	12.96	1.620	0.634	0.847	0.694	0.817	2.335
Protein	552	9.33	1.926	0.297	0.976	0.332	0.970	5.803
PAYProtein	552	458	13.5	2.1	0.976	2.323	0.970	5.803
Nitrogen	552	1.64	0.338	0.052	0.976	0.058	0.970	5.806

RPD = SD/SECV