

Rothamsted Repository Download

A - Papers appearing in refereed journals

Oddy, J., Chhetry, M., Awal, R., Addy, J., Wilkinson, M. D., Smith, D., King, R., Hall, C., Testa, R., Murray, E., Raffan, S., Curtis, T. Y., Wingen, L., Griffiths, S., Berry, S., Elmore, S. J., Cryer, N., Moreira de Almeida, I. and Halford, N. G. 2023. Genetic control of grain amino acid composition in a UK soft wheat mapping population. *The Plant Genome*.
<https://doi.org/10.1002/tpg2.20335>

The publisher's version can be accessed at:

- <https://doi.org/10.1002/tpg2.20335>
- <https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.20335>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/98956/genetic-control-of-grain-amino-acid-composition-in-a-uk-soft-wheat-mapping-population>.

© 3 May 2023, Please contact library@rothamsted.ac.uk for copyright queries.

Genetic control of grain amino acid composition in a UK soft wheat mapping population

Journal:	<i>The Plant Genome</i>
Manuscript ID	Draft
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Oddy, Joseph; Rothamsted Research, Sustainable Soils and Crops Chhetry, Monika; John Innes Centre Awal, Rajani; John Innes Centre Addy, John; Rothamsted Research Wilkinson, Mark; Rothamsted Research Smith, Dan; Rothamsted Research King, Rob; Rothamsted Research Hall, Chris; Rothamsted Research Testa, Rebecca; Rothamsted Research Murray, Eve; Rothamsted Research Raffan, Sarah; Rothamsted Research Curtis, Tanya; Curtis Analytics Ltd Wingen, Luzie; John Innes Ctr, Crop Genetics Griffiths, Simon; John Innes Centre Berry, Simon; Limagrain UK Elmore, J. Stephen; University of Reading Cryer, Nicholas; Mondelez UK Holdings and Services Ltd Moreira de Almeida, Isabel; Mondelez European Biscuit Research and Development Saclay Halford, Nigel G.; Rothamsted Research
Keywords:	Wheat, Seed physiology, Crop genetics

SCHOLARONE™
Manuscripts

Core Ideas

As part of the submission process, we ask authors to prepare highlights of their article. The highlights will consist of 3 to 5 bullet points that convey the core findings of the article and emphasize the novel aspects and impacts of the research on scientific progress and environmental problem solving.

The purpose of these highlights is to give a concise summary that will be helpful in assessing the suitability of the manuscript for publication in the journal and for selecting appropriate reviewers. If the article is accepted the highlights may also be used for promoting and publicizing the research.

Core Idea 1: High free asparagine and low lysine concentrations limit the nutritional value of wheat grain.

Core Idea 2: Investigation of a biparental mapping population formed from the UK soft wheats Claire and Robigus.

Core Idea 3: Breeding for lower free asparagine and higher lysine using Claire and Robigus diversity is possible but limited.

Core Idea 4: CUST_CORE_IDEA_4 :No data available.

Core Idea 5: CUST_CORE_IDEA_5 :No data available.

Genetic control of grain amino acid composition in a UK soft wheat mapping population

Joseph Oddy¹, Monika Chhetry², Rajani Awal², John Addy¹, Mark Wilkinson¹, Dan Smith¹, Robert King¹, Chris Hall¹, Rebecca Testa¹, Eve Murray¹, Sarah Raffan¹, Tanya Y. Curtis³, Luzie Wingen², Simon Griffiths², Simon Berry⁴, J. Stephen Elmore⁵, Nicholas Cryer⁶, Isabel Moreira de Almeida⁷, and Nigel G. Halford^{1*}

¹*Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*

²*John Innes Centre, Norwich Research Park, Colney Lane, Norwich, UK*

³*Curtis Analytics Limited, Discovery Park, Sandwich, UK*

⁴*Limagrain UK Ltd, Market Rasen LN7 6DT, UK*

⁵*Department of Food & Nutritional Sciences, University of Reading, RG6 6DZ, UK*

⁶*Mondelēz UK R&D Ltd, Bournville Lane, Bournville, Birmingham, B30 2LU, UK*

⁷*Mondelēz R&D International, 6 Rue René Razel, 91400 Saclay, France*

Received:

Corresponding author (e-mail): nigel.halford@rothamsted.ac.uk

Abbreviations: QTL (quantitative trait locus/loci); GS (genomic selection); HFN (Hagberg falling number); KHI (kernel hardness index)

Running title: Genetic control of grain amino acid composition in a UK soft wheat mapping population

Core Ideas:

- High free asparagine and low lysine concentrations limit the nutritional value of wheat grain.
- Investigation of a biparental mapping population formed from the UK soft wheats Claire and Robigus.
- Breeding for lower free asparagine and higher lysine using Claire and Robigus diversity is possible but limited.

For Review Only

Abstract

Wheat is a major source of nutrients for populations across the globe, but the amino acid composition of wheat grain does not provide optimal nutrition. The nutritional value of wheat grain is limited by low concentrations of lysine (the most limiting essential amino acid) and high concentrations of free asparagine (precursor to the processing contaminant acrylamide). There are currently few available solutions for asparagine reduction and lysine biofortification through breeding. In this study, we investigated the genetic architecture controlling grain free amino acid composition and its relationship to other traits in a Robigus × Claire doubled haploid population. Multivariate analysis of amino acids and other quality traits showed that the two groups are largely independent of one another, with the largest effect on amino acids being from the environment. Linkage analysis of the population allowed identification of QTL controlling free amino acids and quality traits, and this was compared against genomic prediction methods. Following identification of a QTL controlling free lysine content, wheat pangenome resources facilitated analysis of candidate genes in this region of the genome. These findings can be used to select appropriate strategies for lysine biofortification and free asparagine reduction in wheat breeding programmes.

Introduction

The nutritional quality of wheat has profound impacts on human health. As one of the largest sources of average daily calorie intake in the world (18.2% in 2019) (FAOSTAT, 2021), wheat is an essential source of macro and micronutrients. In 2019, 19.5% of average daily global protein intake was estimated to be provided by wheat-based foods (FAOSTAT, 2021). Similarly, between 2008 and 2017 in the UK, over 25% of average daily fibre intake was provided by wheat-based foods (Gressier & Frost, 2021). Wheat flour is often fortified to increase its nutrient content: in the UK, for example, wheat is fortified with calcium, iron, thiamine, niacin and, most recently, folate (DEFRA, 1998; DHSC, 2021). The quantities of different macro and micronutrients in wheat can have large impacts on population health because of the scale at which wheat products are consumed. For example, it is estimated that the addition of folate to UK flour will lead to a 20% decrease in neural tube defects in babies (DHSC, 2021). Consequently, it is essential to ensure that the nutritional profile of wheat is as beneficial as it can be for human health.

One way in which the nutritional profile of wheat can be improved is *via* optimisation of its amino acid composition, with the concentrations of lysine and asparagine most important. Free (soluble, non-protein) asparagine can be converted to the processing contaminant, acrylamide, during high-temperature cooking and processing, and this has led to ongoing efforts to reduce free asparagine concentration (Oddy *et al.*, 2022). Lysine, on the other hand, is not produced endogenously by humans or other monogastric animals, making it an essential amino acid in the diet, but it is present in only small quantities in wheat and other cereal grain and populations reliant on cereals for their nutrition may suffer from lysine deficiency (Galili & Amir, 2013). Indeed, fortifying wheat flour by adding lysine has been shown to improve indices of nutritional status in clinical trials in Pakistan, northern China, and Syria (Hussein *et al.*, 2004; Zhao *et al.*, 2004; Ghosh *et al.*, 2008). Flour fortification is

unlikely to be a sustainable solution in developing countries and it would be much cheaper and more efficient to increase the intrinsic lysine content of wheat grain. Therefore, the amino acid composition of wheat grain could be optimised both by decreasing grain free asparagine content and increasing lysine content.

In recent years, studies have investigated genetic strategies for the reduction of free asparagine content in wheat grain. Induced and natural variation in the asparagine synthetase 2 genes, for example, has been found to impact significantly on free asparagine content (Raffan *et al.*, 2021; Oddy *et al.*, 2021; Alarcon-Reverte *et al.*, 2022) and quantitative trait loci (QTL) for grain asparagine content have been identified from previous GWAS studies (Emebiri, 2014; Peng *et al.*, 2018; Rapp *et al.*, 2018). However, the small number of stable QTL available to breeders limits the progress that can be made to reduce grain asparagine content in breeding programmes and no genetic strategies for soft (biscuit) wheat specifically have been investigated. Similarly, there are limited strategies currently available for increasing lysine content in wheat grain. Lysine biofortification *via* QTL identification and marker-assisted breeding has been studied extensively in both rice (Wang *et al.*, 2008; Zhong *et al.*, 2011; Yoo, 2017; Jang *et al.*, 2020) and maize (Prasanna *et al.*, 2020), but only two studies have previously investigated lysine biofortification in wheat through association studies. Peng *et al.* (2018) successfully identified QTL controlling free lysine and Jiang *et al.* (2013) identified QTL for total lysine.

Consequently, the aim of this study was to investigate QTL, genomic prediction accuracy, and candidate genes controlling the free amino acid composition of wheat grain in a soft wheat mapping population developed from the varieties Claire and Robigus. Like many UK varieties, these parents both lack the B genome homeologue of the asparagine synthetase-2 gene, *TaASN-B2* (TraesLDM3B03G01566640 in variety Landmark), the presence/absence of which is a known source of grain asparagine content variation (Oddy *et al.*, 2021). This

67 mapping population, therefore, represents a useful resource for identifying additional
68 variation. Claire and Robigus are also represented by scaffold-level genome assemblies in the
69 wheat pangenome, facilitating candidate gene analysis. Furthermore, we investigated other
70 quality traits, such as grain size, hardness, and Hagberg falling number (HFN), to determine
71 whether QTL controlling nutritional traits overlapped with those controlling other quality
72 traits.

For Review Only

Materials and Methods

Production of Doubled Haploid lines

Doubled Haploid lines of Robigus x Claire were produced using a modified Knox *et al.* (2000) method. Wheat spikes were emasculated between growth stages GS55 and GS59. Once the stigma was receptive it was fertilised with freshly shed donor maize pollen. After one day, wheat florets were treated with Dicamba (20mgL⁻¹) (Sigma-Aldrich, D5417) and injected into the plant stem (100mgL⁻¹). Developing embryos were excised between 14 and 21 days. Under aseptic conditions, seeds were removed from the spikelets, surface sterilised with 70% (v/v) ethanol (EtOH) for 1 min, rinsed with sterile distilled water, and immersed in 20% (v/v) commercial bleach solution with a few drops of Tween® 20 for 20 mins. They were then rinsed with sterile distilled water three times.

Haploid embryos were excised and grown on 90mm Petri dishes in the dark on Gamborg's B5 media with minimal organics (Gamborg *et al.*, 1968), 2% (w/v) sucrose, pH 5.8, 9gL⁻¹ Difco bactoagar at 20°C. When showing signs of germination, embryos were transferred to a light incubator at 20°C. Any non-germinated 1 month old embryos were given cold shock treatment at 4°C for 7 days to promote germination. Germinated plantlets were vernalised for 4 weeks and were grown in the glasshouse until the 4-tiller stage. Plants were then given colchicine (Sigma-Aldrich, C9754) treatment for 5 to 6 hours in the light at room temperature, washed and transplanted to soil, acclimatised and grown in a glasshouse. The mapping population was genotyped by Limagrain using a proprietary SNP array. The genetic map comprising 872 loci was constructed using MSTMap Online (<http://mstmap.org/>).

The mapping population was grown in field trials at the John Innes Centre Morley Mill Hill field site (52°33'15.1"N 1°01'59.2"E), UK, in 2017 to 2018, and at the Church Farm field site (52°38'N 1°10'E) in 2018 to 2019, using an unreplicated, completely randomised

design. The H18 field trial was drilled on the 21st September 2017 and harvested on the 1st August 2018. The H19 field trial was drilled on the 14th September 2018 and harvested on the 12th August 2019. Growth habit, heading date, plant height, and yield traits were scored in the field.

Phenotyping

Grain diameter, kernel hardness index (KHI), and grain weight measurements were recorded for 300 kernels from each line in the population using a Perten Single Kernel Classification System (SKCS) 4100 (Calibre Control International Ltd., Warrington, UK). Grain length (mm), width (mm), and area (mm²) measurements were recorded in triplicate for each sample using a MARVIN Seed Analyser and software Marvin 4.0 (MARViTECH GmbH, Wittenburg, Germany). Grain samples were milled to wholemeal flour in a coffee grinder and flour moisture content was recorded using a Minispec nuclear magnetic resonance (NMR) analyser (Minispec Mq10, Bruker Inc., Germany). Hagberg falling number measurements were recorded using an FN 1000 as the average of two technical replicates (Perten, Sweden), adjusting for flour moisture content as required according to manufacturer's instructions. Amino acid analysis was performed on wholemeal flour samples by HPLC as described previously (Raffan *et al.*, 2021) by Curtis Analytics (Sandwich, UK). Briefly, free amino acids were extracted from 0.5g of wholemeal flour and underwent precolumn derivatisation (Curtis *et al.*, 2018). Samples were then run on an HPLC system identically to previously described (Raffan *et al.*, 2021). Three technical replicates were taken for each sample for amino acid measurement.

Phenotypic data analysis

Skewness and kurtosis were measured for all variables in each environment and normal plots visually inspected in Genstat (VSN International, 2021) to determine if variables required transformation. The data were appropriately transformed according to their distribution if necessary (see Tables S1 and S2 for details of transformations). Subsequent analyses were performed on transformed variables unless otherwise stated. Plotting was performed in R (R Core Team, 2021) with the packages ggplot2 (Wickham, 2016), tidyverse (Wickham et al., 2019), and cowplot (Wilke, 2020).

Broad-sense heritability for each trait was estimated as described in Covarrubias-Pazaran (2019) using the packages dplyr (Wickham et al., 2022) and lme4 (Bates et al., 2015). Kendall rank correlation coefficients were performed on non-transformed data and adjusted p values (Bonferroni correction) were calculated for plotting using R (R Core Team, 2021) and the package corrrplot (Wei & Simko, 2021). Principal component analysis was performed on untransformed, scaled variables using the package factoextra (Kassambara and Mundt, 2020). Correlation network analysis was performed and plotted by filtering for significant correlations where $p < 0.001$ using Kendall correlation with Bonferroni correction using the packages corrr (Kuhn, Jackson and Cimentada, 2020), igraph (Csardi and Nepusz, 2006), and ggraph (Pedersen, 2021).

Bayesian modelling was performed on untransformed variables in R using the package rstanarm (Goodrich *et al.*, 2020). Variables were scaled before modelling and individual linear models for each predictor variable were created to guide the selection of informative priors. Simulations of the posterior distribution were subsequently performed to check model fit and intervals were plotted using the package bayesplot (Gabry and Mahr, 2022). R^2 estimates were obtained by taking the median of leave-one-out cross validation adjusted estimates.

147 *Linkage analysis*

148 Multi-environment single trait linkage analysis was performed in Genstat for each trait to
149 detect QTL present in both environments, following selection of the most appropriate
150 variance-covariance model according to the Bayesian information criterion. Simple interval
151 mapping (SIM) was initially performed to identify putative QTL. These QTL were then used
152 as covariates in composite interval mapping (CIM). QTL identified from CIM were then used
153 to construct the final QTL models. Pseudo-markers were generated every 2 cM in the map.
154 The minimum cofactor proximity was set at 30 cM and the minimum separation for selected
155 QTL at 20 cM. Significance thresholds were determined by the Li and Ji method (Li & Ji,
156 2005) with a genome-wide significance level of 0.05.

157 Single-environment linkage analysis was performed in R using packages qtl (Broman
158 *et al.*, 2003) and qtl2 (Broman *et al.*, 2018). Single-environment linkage analysis was made
159 into an interactive app using the packages shiny (Chang *et al.*, 2021), plyr (Wickham, 2011),
160 and rsconnect (Atkins, McPherson & Allaire, 2021), accessible at [https://t9onwp-](https://t9onwp-wheatworker.shinyapps.io/QTL_Browser/)
161 [wheatworker.shinyapps.io/QTL_Browser/](https://t9onwp-wheatworker.shinyapps.io/QTL_Browser/) and in supplementary data file 1. As before, SIM
162 was performed first to identify covariates for use in CIM. Identified QTL from CIM were
163 then used to create single QTL models as well as additive QTL models. Upper and lower
164 95% confidence intervals for QTL location were calculated using the Bayesian credible
165 interval method in R/qtl and expanded to the closest markers. Pseudomarkers were generated
166 every 2 cM in the map and the minimum marker covariate proximity was set at 20 cM. A
167 logarithm of the odds (LOD) score of 3 was used as the significance threshold.

168

169 *KASP assays and statistical analysis*

170 Varieties were grown and DNA extracted as previously described (Oddy *et al.*, 2021). KASP
171 marker sequences for dwarf and wild-type alleles of *Rht-B1* were found on CerealsDB

(Wilkinson *et al.*, 2020). Assays were run in 96-well plates in an Applied Biosystems™ 7500 Real-Time PCR System. Primer mix was made using 46µL dH₂O, 30µL common primer (100µM), and 12µL of each tailed primer (100µM). Each reaction contained 0.14µL KASP primer mix, 2.86µL water, 5µl KASP low-ROX mix (PACE), and 2µL DNA sample. Cycling conditions were 95 °C for 15 min, followed by 10 cycles of 95 °C 20s, 61 °C 60s (reducing anneal 0.6 °C per cycle), followed by 30 cycles of 95 °C 20s, 55 °C 60s. Data were then read and analysed using KlusterCaller genotyping software (LGC Biosearch Technologies).

Statistical analysis was performed using a REML model, using asparagine data from two previous field trials (Curtis *et al.*, 2018). The analysis performed was the same as in Oddy *et al.* (2021), but with *Rht-B1* allele status included as an additional term nested within the variety factor. The fixed effects model was: Year * (TaASN-B2*TaRHT-B1/Variety) * Treatment. The random effect model was: Year/Block/MainPlot/SplitPlot.

Genomic prediction

Genomic prediction was performed for each trait *via* five-fold cross validation with 10,000 permutations using the R package rrBLUP (Endelman, 2011). The “mixed.solve” function within this package was used to estimate marker effects for each trait, with the identity matrix being left unspecified. Pearson correlation coefficients were calculated for the results from the training and testing datasets to estimate genomic prediction accuracy. For within year prediction estimates, training and testing datasets came from the same trial. For between year prediction estimates, training and testing datasets were from different trials. Further detail is available as R markdown in supplementary data file 2. Scripts were submitted to the high-performance computing cluster at Rothamsted Research via SLURM for execution.

Candidate gene analysis

197 The gene content of the lysine QTL was determined for all wheat pangenome varieties at
198 chromosome scale assembly by identifying the location of the markers in these varieties and
199 extracting genes from Ensembl Biomart (Howe et al., 2021). Genes residing within the region
200 in variety Chinese Spring v1.0 were submitted to KnetMiner
201 (https://knetminer.com/Triticum_aestivum/) (Hassani-Pak *et al.*, 2021) for ranking on
202 relevant keywords (“Lysine”, “Storage proteins”). Expression of the top hits was then
203 investigated in expVIP (Borrill, Ramirez-Gonzalez, & Uauy, 2016) to further narrow down
204 plausible candidate genes. Transcript per million (TPM) data for the Azhurnaya
205 developmental time-course experiment were extracted from expVIP for plotting in R using
206 the package pheatmap (Kolde, 2019). Corresponding Claire and Robigus genes were then
207 identified from these Chinese Spring candidate genes in Ensembl and pairwise aligned via
208 BLAST using Geneious Prime 2020.1.2 to identify variation.

Results

Phenotypic analysis

We measured free amino acid concentrations and other grain quality traits in the Robigus × Claire mapping population from field trials grown in 2017–2018 (H18) and 2018–2019 (H19) (Figure 1; Figure S1). Aspartic acid, asparagine, and glutamic acid were the most abundant of the free amino acids measured, with concentrations of free amino acids consistently higher in H19 than in H18 (Figure 1a). Principal component analysis revealed harvest year to be a key driver of variation in this dataset (Figure 1b) and, notably, the second harvest year (H19) also showed lower yield alongside the increased free amino acid content of the grain (Figure 1b). PCA and correlation network analysis revealed that most of the other quality traits measured here were uncorrelated with the amino acids (Figure 1b; Figure 1c; Figure S2; Figure S3), except for grain yield which showed negative correlations with a subset of amino acids (Figure 1b; Figure 1c; Figure 2a).

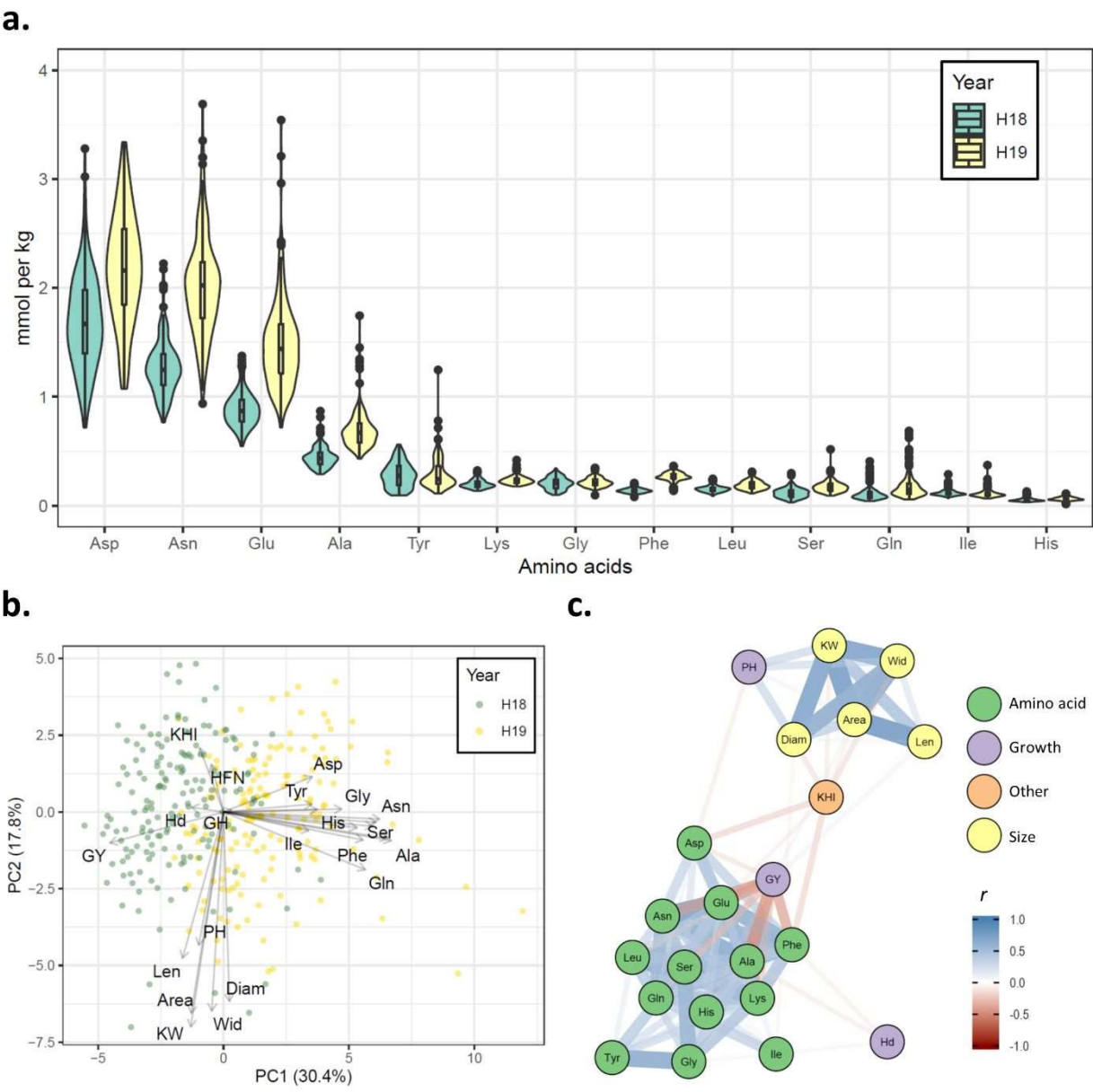


Figure 1. Characterisation of the Robigus × Claire mapping population. a. Measurements of amino acids in the 2017–2018 (H18) and 2018–2019 (H19) harvest years. **b.** Principal component analysis of all traits in both years along the first two principal components. **c.** Correlation network analysis of all traits across both years (GH omitted, Kendall correlation, only links with significance <0.001 shown).

To understand whether any of the traits we measured could predict free asparagine or lysine content in the grain, we constructed Bayesian linear models with the quality traits and harvest

year as explanatory variables (Figure 2b; Figure 2c). In both the free asparagine (Figure 2b) and lysine (Figure 2c) models, environment had the greatest effect whereas other variables had little explanatory power. Nevertheless, the variance explained in the models was still reasonable for asparagine at 56.5%, but only 22.2% for lysine.

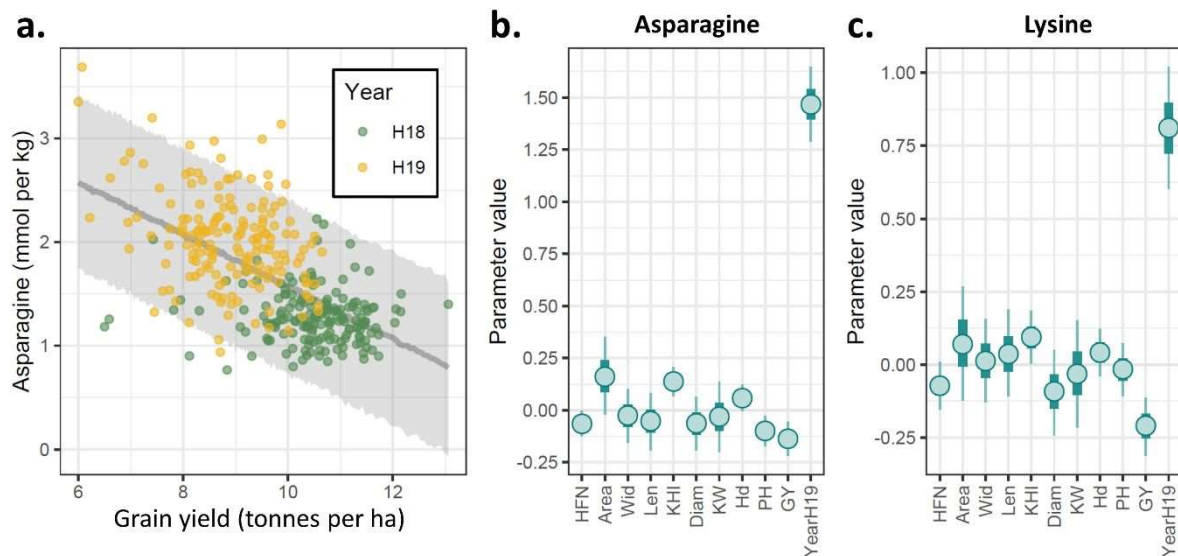


Figure 2. Relationships between free asparagine/lysine and other agronomic measurements. **a.** Linear modelling of free asparagine content against grain yield. The grey shaded ribbon shows 95% prediction intervals sampled from the posterior distribution. **b.** and **c.** Parameter values from multiple linear modelling of asparagine (**b.**) and lysine (**c.**) as explained by other quality traits measured in this population.

242 *QTL analysis*

243 Broad-sense heritability estimates varied substantially between the different amino acids,
244 with free asparagine and lysine showing heritability estimates of 0.60 and 0.45, respectively
245 (Table S1). Aspartic acid showed the highest heritability of the amino acids measured here,
246 with an estimate of 0.82. Heritability estimates for the size traits were generally very high, as
247 expected, and correlation of these values between years was also stronger than the correlation
248 of amino acids between years (Table S1).

249 We identified QTL for grain free asparagine content and lysine content on
250 chromosomes 4B and 1A, respectively (Figure 2a; Figure 2b; Table 1), which had significant
251 effects across both environments but were also affected by QTL by environment effects
252 (Figure 2c; Figure 2d; Table 1; Table S2). The asparagine QTL on 4B explained 2.6% of the
253 variance in H18, when free asparagine concentrations were lower overall, whereas it
254 explained 14.8% of the variance in H19, when free asparagine concentrations were elevated
255 (Table 1). In both years, the Robigus allele was associated with the higher free asparagine
256 concentrations. In contrast, the lysine QTL on 1A explained 12.1% of the variance in H18,
257 when free lysine was lower overall, and only 2.6% of the variance in H19, when free lysine
258 concentrations were elevated. The Claire allele was associated with higher free lysine
259 concentrations in both years in this case. Multi-environment linkage analysis of amino acid
260 and grain measurements revealed many QTL controlling the other amino acids and quality
261 traits as well (Table 1; Table S2; Table S3).

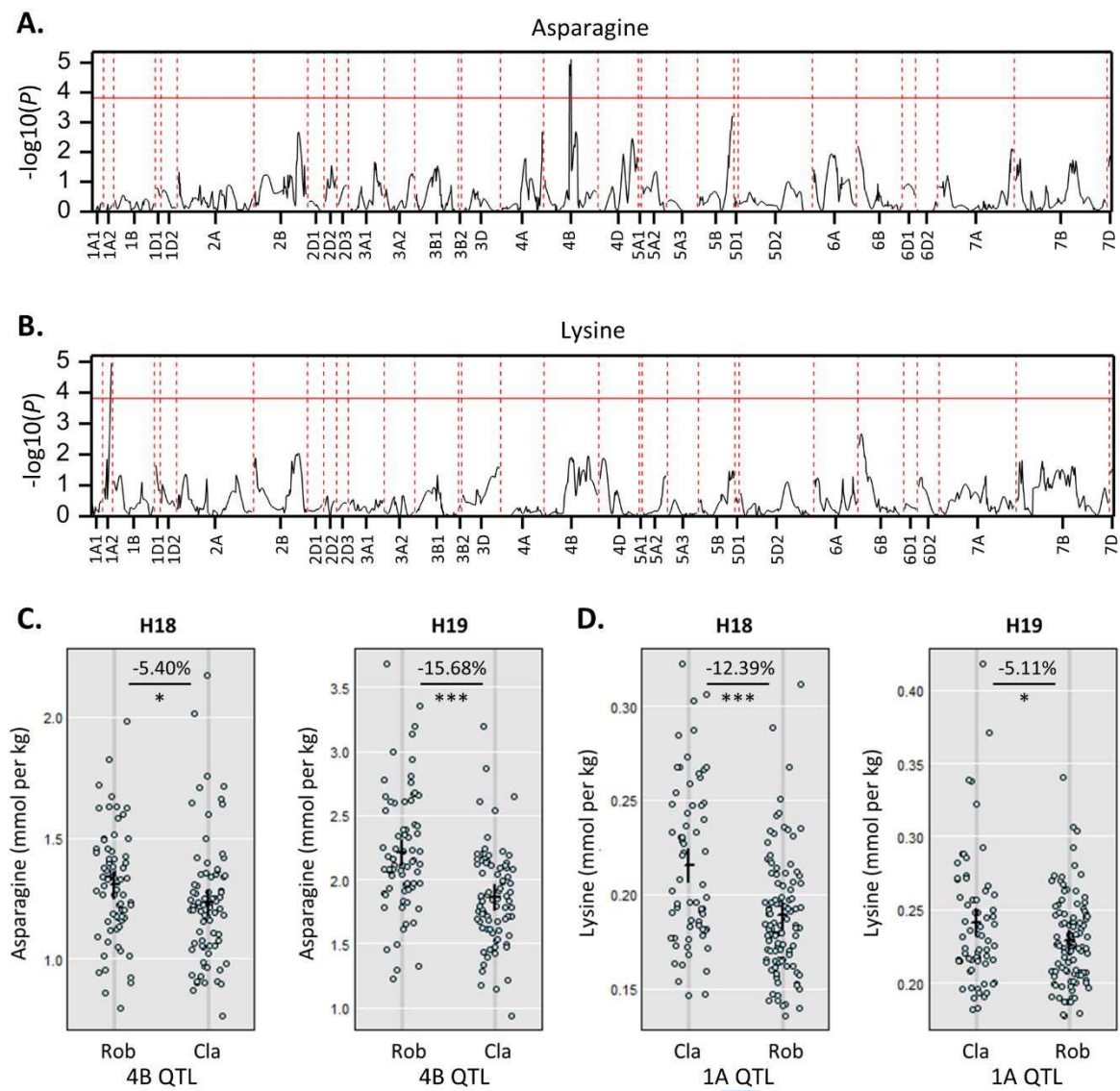


Figure 3. Identification of QTL controlling free asparagine and free lysine. **a.** Multi – environment genome scan plot for asparagine. **b.** Multi – environment genome scan plot for lysine. **c.** Impact of the asparagine QTL on free asparagine concentrations in both field trials. **d.** Impact of the lysine QTL on free lysine concentrations in both field trials. Error bars show plus and minus two times standard error of the mean. Significance values are taken from the corresponding years of the multi-environment linkage analysis.

The QTL controlling asparagine on chromosome 4B appeared to overlap with QTL for several other traits, including plant height, KHI, grain diameter, and grain weight (Figure

S4). As a result, we investigated whether variation in the *Rht-B1* dwarfing gene was associated, since Claire possesses the wild-type *Rht-B1a* allele, whereas Robigus possesses the dwarf *Rht-B1b* allele (Table S4). We screened cultivars that had been measured for grain asparagine content in two previous field trials for the different *Rht-B1* alleles (Table S5) and investigated whether this had any association with asparagine content through a REML analysis (Table S6). Our results indicate the *Rht-B1* status did not have any significant effect on asparagine content in these trials, suggesting that the cause of the asparagine QTL is something other than *Rht-B1* variation.

Table 1. Multi-environment QTL for measured amino acids. Chr. (Chromosome), cM (centimorgan), Mbp (megabase pair location in Chinese Spring v1.0).

Trait	Multi-environment single trait linkage analysis (H18 and H19)							
	Marker	Chr.	cM	Mbp	$-\log_{10}(p)$	H18 (%)	H19 (%)	High val.
Ala	WC.0223839	7B	211.2	719	5.03	7.1	5.7	Robigus
Asn	WC.0221262	4B	114.47	601	5.96	2.6	14.8	Robigus
Asp	WC.0218489	1B	54.4	530	5.4	8	5.9	Claire
	WC.0214359	3A2	2.3	738	7.95	7.3	15.3	Robigus
	WC.0221037	4A	148.8	703	8.08	12.6	9.3	Claire
	WC.0227146	4D	48.8	16	3.7	5.5	4.1	Claire
Gln	WC.0221302	4B	103.7	547	3.5	5.4	4.5	Robigus
	WC.0228471	6B	19.7	25	5.09	8.2	6.7	Claire
Glu	WC.0221329	4B	100.8	518	4.27	3.7	10.1	Robigus
Gly	WC.0226796	4B	155.2	327	4.26	3.2	5.3	Robigus
Iso	WC.0223785	7B	211.2	717	3.6	6.8	3.7	Robigus
Lys	WC.0218011	1A2	27.3	593	4.95	12.1	2.6	Claire
Phe	WC.0220622	3B1	78.1	116	3.83	6.2	5.6	Robigus

QTL for aspartic acid also appeared to overlap with QTL for other traits (Table 2). For aspartic acid on 4A and 4D, there are co-locating HFN QTL, suggesting that these two traits are under the control of the same locus. The location of the QTL on 4D matches the *Rht-D1* polymorphism between Claire and Robigus found at 18.78 Mbp in Chinese Spring.

Of all the amino acids measured in this study, we identified the most QTL controlling aspartic acid (Table 1).

Table 2. Multi-environment QTL impacting both amino acids and other traits on chromosomes 4A, 4B, and 4D. Chr. (Chromosome), cM (centimorgan), Mbp (megabase pair location in Chinese Spring v1.0).

Chr.	Multi-environment single trait linkage analysis (H18 and H19)							
	Trait	Marker	cM	Mbp	$-\log_{10}(p)$	H18 (%)	H19 (%)	High val.
4A	Asp	WC.0221037	148.8	703	8.08	12.6	9.3	Claire
	KHI	WC.0221037	148.8	703	8.26	14.8	14.7	Robigus
	HFN	WC.0188904	147.1	733	8.24	11.5	10.3	Robigus
	Area	WC.0220938	149.7	709	2.22	3	3.2	Claire
	Length	WC.0221119	149.7	702	7.12	1.8	6.5	Claire
4B	Asn	WC.0221262	114.47	601	5.96	2.6	14.8	Robigus
	KHI	WC.0226741	110.8	594	4.30	4.2	8.6	Robigus
4D	Asp	WC.0227146	48.8	16	3.7	5.5	4.1	Claire
	Width	WC.0227146	48.8	16	5.98	8	8.9	Robigus
	Diam	WC.0227146	48.8	16	7.84	7.6	8.1	Robigus
	HFN	WC.0227149	56.9	17	10.92	23.5	5.6	Robigus
	Height	WC.0213051	56.9	17	28.97	27.8	38.9	Robigus

296 *Genomic prediction*

297 Following our modelling of asparagine and lysine using agronomic measurements and
298 QTL models, we calculated the accuracy of genomic prediction (GP) for within and between
299 year prediction of traits (Figure 4a; Figure S6; Table S1). Prediction accuracy was more
300 consistent when performed across years rather than within years (Figure S6), so these were
301 used for further interpretation. Prediction accuracy for lysine was the lowest of all traits at a
302 mean accuracy of 0.10, whereas accuracy for asparagine was around 0.34. Of all amino acids,
303 aspartic acid had the greatest prediction accuracy results. Prediction accuracies for the other
304 functional traits were generally higher than the accuracies for amino acids, as expected from
305 the higher heritability of these traits. Comparing the amount of variation explained by
306 genomic prediction methods and additive QTL models, we can see that the GP models
307 explain more variance than the additive QTL models for all traits (Figure 4b).

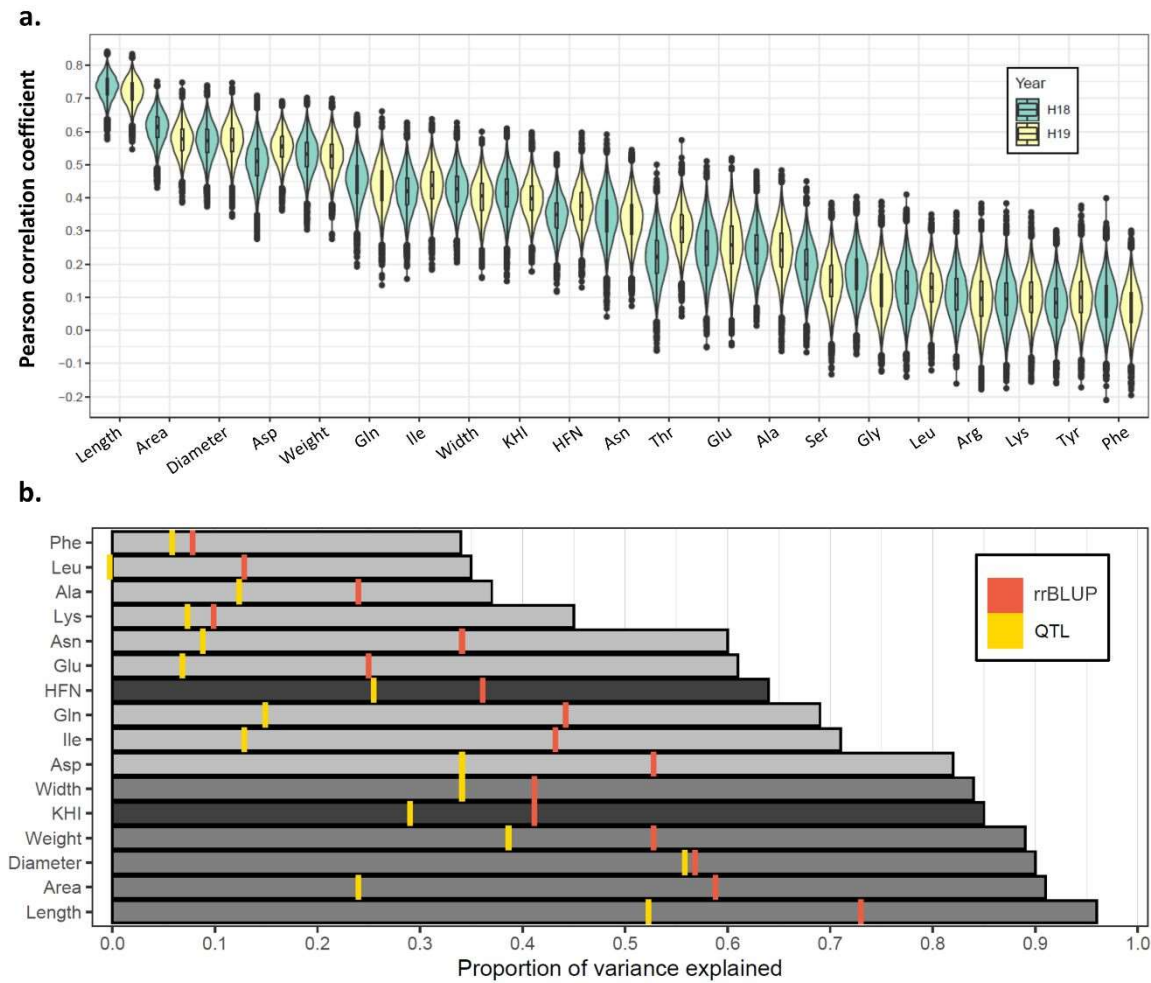


Figure 4. Variation explained by heritability, genomic prediction, and QTL. a. Genomic prediction accuracy between years. **b.** Additive QTL effects and genomic prediction (rrBLUP) accuracy (yellow and red marks, respectively) plotted alongside broad-sense heritability (shown as bars). Bars are shaded according to the trait group that they belong to (amino acid, size, or other).

315 *Lysine QTL candidate gene analysis*

316 The gene content and QTL size of the lysine QTL on 1A, the HFN/aspartic acid/KHI QTL on
317 4A, and the asparagine QTL on 4B differed substantially (Table S7). Due to the size of the
318 4A and 4B QTL, we were unable to plausibly narrow down candidate genes, whereas the
319 lysine QTL on 1A was much smaller so amenable to further analysis. We investigated the
320 gene content of the lysine QTL for all genomes assembled to chromosome scale in the wheat
321 pangenome and gene content varied to a small extent between the different varieties (Table
322 S8). Most notably, the QTL did not match any locations in variety Julius and matched to an
323 unanchored scaffold in Stanley.

324 KnetMiner analysis of the genes residing in Chinese Spring in the lysine QTL was
325 undertaken with relevant keywords to highlight possible candidate genes, and these genes
326 were subsequently investigated for their expression patterns from expVIP. Pairwise analysis
327 of the top KnetMiner hits in the lysine QTL showed that the top hit
328 (TRAESCS1A02G445700) differed between Claire and Robigus. TRAESCS1A02G445700,
329 or *TaHDT-A1*, has been identified as a member of the histone deacetylase family in wheat. A
330 deletion within the CDS of the gene in Robigus means that the most highly expressed
331 transcript cannot be expressed (Figure 5) and the two missing exons from this most highly
332 expressed transcript form a zinc finger/C2H2 DNA binding domain, which is important for
333 transcriptional regulation.

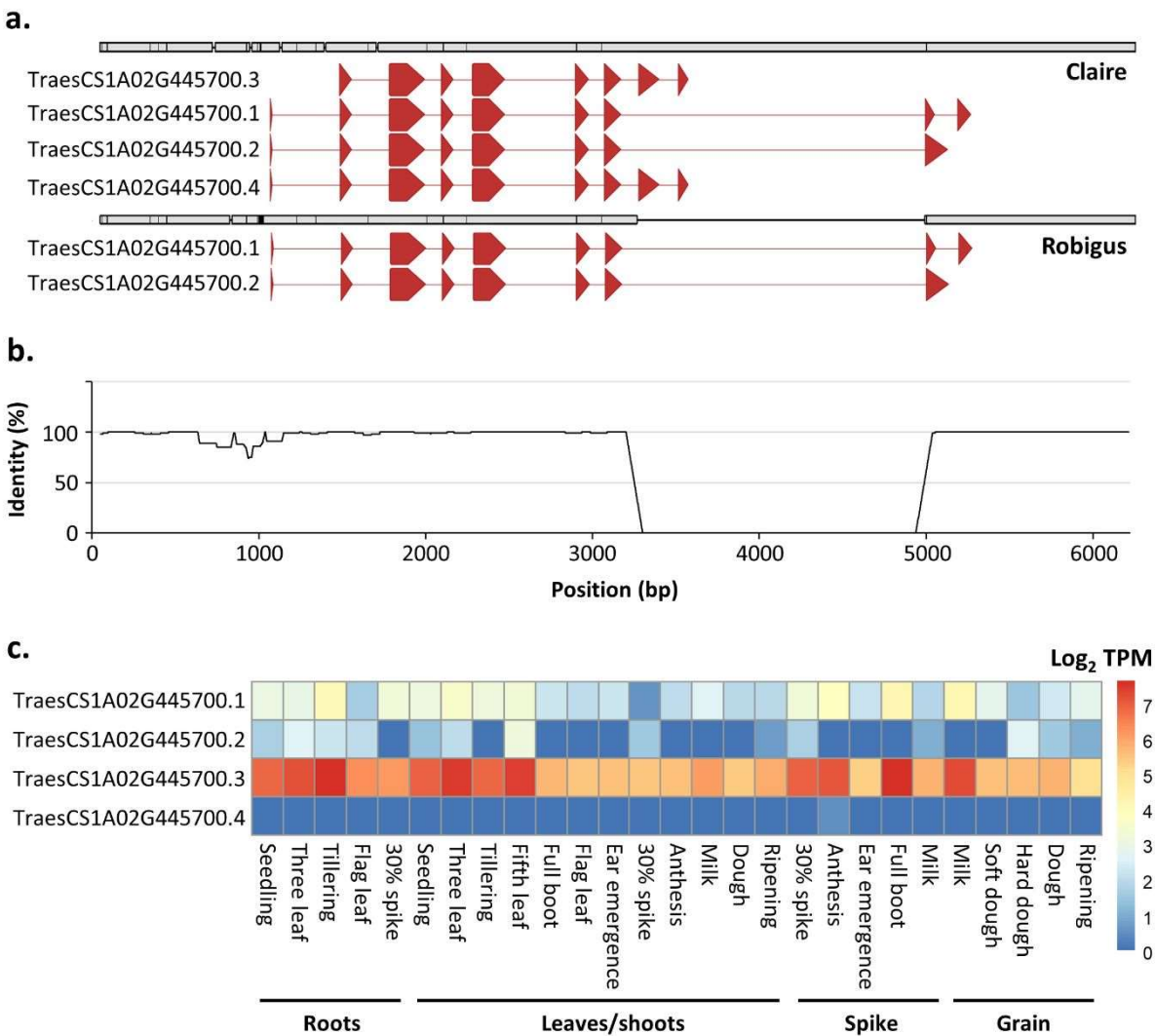


Figure 5. Analysis of the *TaHDT-A1* candidate gene for lysine QTL between parents Claire and Robigus. a. Pairwise alignment of the two genes. **b.** Percentage identity calculated as a sliding window average of 100 bp. **c.** Expression of the four transcripts throughout development in variety Azhurnaya.

Discussion

Limited variation in Claire and Robigus for asparagine and lysine improvement

Soft wheat breeding in the UK has relied heavily upon Claire and Robigus as parents since their development in 1999 and 2005, respectively. A recent study found that UK winter wheats developed between 2002 and 2017 could be clustered into four distinct populations, and two of these populations were characterised by their Claire or Robigus heritage (Shorinola et al, 2022). The varieties within these population groups characterised by Claire and Robigus heritage are also almost entirely soft wheat varieties, further emphasising the importance of these two varieties in UK soft wheat breeding. This large contribution of Claire and Robigus as parents to soft wheat breeding means that opportunities for nutritional improvement have often been limited to variation between these two parents.

Our analysis found that there is variation between Claire and Robigus and that this does impact asparagine and lysine content to a small extent. Asparagine had a moderate heritability (0.60) across both field trials in the study, whereas the heritability for lysine was lower (0.45). One QTL was found for asparagine and lysine each, both explaining less than 10% of the variance on average. The asparagine QTL identified here (peak at 601.4 Mbp in Chinese Spring) lies around 60 Mbp from another QTL (peak at 660.7 Mbp in Chinese Spring) identified by Peng et al. (2018), suggesting that these may coincide, whereas the lysine QTL does not overlap with previously identified QTL. Genomic selection had a predictive ability of 0.34 on average for asparagine, indicating that this method may be better suited for breeding because of the genetic architecture of this trait (many small-effect QTL). Rapp et al. (2018) also found that GS had a predictive ability of around 0.5 on average for asparagine, the higher estimate in this study likely due to within environment prediction and analysis of a more diverse mapping population. GS only achieved a predictive ability of 0.10

for lysine, indicating that only incremental advances in lysine content are possible using Claire and Robigus. Previous GWAS studies using more diverse panels have found more, larger-effect QTL controlling asparagine and lysine content (Peng et al., 2018; Rapp et al., 2018), indicating that there may be beneficial alleles in more diverse germplasm. Consequently, UK soft wheat germplasm will require diversity beyond Claire and Robigus to make changes to asparagine and lysine content beyond the incremental improvements found here.

Trade-offs between amino acid content and other traits

Another aspect we wanted to investigate in this population was whether there were any relationships between amino acids and other traits. Amino acids tended to correlate positively with one another and were mostly unrelated to the other measured traits, with the exception of grain yield and kernel hardness index. A negative correlation between grain yield and free asparagine has previously been documented (Xie et al., 2021), but in other experiments the association has been positive (Malunga et al., 2021; Xie et al., 2021). In our analysis, this association mostly arose because of the effect of environment on both yield and asparagine. Environmental stress can lead to decreases in yield whilst increasing free asparagine, whilst other variables (e.g., nitrogen fertiliser) can lead to increases in both yield and free asparagine (see Oddy et al. (2022) for review). Our modelling of asparagine through these variables mostly indicated environment as the driving force in our study, but there was still a slight negative association with yield and plant height as well as a slight positive association with kernel hardness. Kernel hardness, like grain free asparagine content, is known to increase with nitrogen application, which may underly this small association with asparagine.

A strong environmental effect on free asparagine concentration has been observed in response to many different stressors (see Oddy et al., 2020 for review) and it is under

stressful conditions that the highest asparagine levels are often observed. These increases in grain asparagine concentration vary massively, causing unexpected blips in acrylamide content in food products and posing the greatest threat to food safety and regulatory compliance, so elimination of this environmental response would be of great interest. A weak environmental effect was seen in this study: during the 2018–2019 season the average amino acid concentrations rose whilst the yields dropped. Interestingly, the asparagine QTL we identified here had greater effect in this season, enabling reductions of 15.68% in free asparagine concentrations in those lines possessing the Claire allele over those possessing the Robigus allele. This suggests that this QTL may be more effective under more stressful conditions, so selection of the Claire allele at this locus may prove beneficial for reducing the large free asparagine increases observed following environmental stress. This is in contrast to the effect of the *TaASN-B2* deletion, which has a greater effect when grain asparagine concentrations are lower (Oddy et al., 2021), when plants are not suffering from sulphur deficiency. Future work would therefore benefit from identification of similar QTL that are associated with lowering asparagine content from the high levels seen during stress. This would enable the stacking of alleles that are beneficial under both stress and non-stress conditions, to ensure that free asparagine concentrations are minimised in all environments.

We also wanted to understand whether any QTL controlling amino acid content had pleiotropic effects on other traits. The asparagine QTL we identified on chromosome 4B appeared to overlap with QTL for plant height in the first year, suggesting that there might be an impact of the *Rht-B1b* allele on asparagine. The *Rht* genes are dwarfing genes used during the green revolution that have many impacts on crop traits beyond height (Casebow et al., 2016) and Claire and Robigus both possess different *Rht* genes on 4B and 4D (www.cerealsdb.uk.net/cerealgenomics/CerealsDB/Excel/MAS_data_May_2013.xls).

However, this QTL overlap was not present in the second year of analysis and we found no

association between *Rht-B1* status and grain asparagine content in our analysis of previous field trials, suggesting that the QTL controlling height and asparagine may be distinct. The QTL controlling asparagine did overlap consistently with a QTL for KHI though, with the ‘increasing allele’ belonging to Robigus for both traits. Kernel hardness and free asparagine content are both known to correlate under certain conditions with nitrogen content (Oddy et al., 2022), so this QTL may be linked to nitrogen use efficiency/uptake. The KHI QTL on 4B also exhibited a similar genotype by environment effect pattern to the asparagine QTL, with a greater effect of the QTL observed in the second trial year. Selection for the Claire allele at this QTL would therefore be suitable in the context of soft wheat breeding, where both softer textures and lower asparagine content are desirable.

Interestingly, we found much more genetic control of free aspartic acid concentration in this population compared to the other amino acids. Heritability was high (> 0.8), genomic prediction accuracy was moderate (> 0.5 , same as grain weight), and there were four multi-environment QTL controlling the trait. Two of the QTL controlling aspartic acid also overlapped with QTL controlling HFN. One of these QTL was situated on 4D and overlapped with traits for plant height and grain size as well, indicating that this may be due to *Rht-D1* allele status, which is known to impact HFN as well as plant height (Fradgley et al., 2022). The second QTL controlling both aspartic acid and HFN was situated on 4A and also overlapped with traits for grain size and KHI. Previous work has identified a major QTL underlying pre-harvesting sprout (PHS) variation on 4A, but both Claire and Robigus share the same *MKK3-A* allele which underlies this QTL (Shorinola et al., 2017). Li et al. (2021) also identified a PHS QTL in a similar region on 4A but this does not overlap with the region identified here. One possible source of variation underlying the QTL controlling aspartic acid and HFN on 4A is the *Triticum dicoccoides* introgression in Robigus, which matches the region this QTL is found in (Przewieslik-Allen et al., 2021). The antagonistic relationship

between HFN and asparagine at this QTL could be a result of increased HFN reducing proteolysis, and thereby preventing accumulation of free amino acids.

Lysine candidate genes

Scaffold-level genome assemblies of Claire and Robigus (Walkowiak et al., 2020) enabled us to investigate the lysine QTL in greater depth, identifying the candidate gene *TaHDT-A1*, encoding a histone deacetylase. The wheat histone deacetylase family is very large, encompassing approximately 50 genes (Jin et al., 2020, Li et al., 2022). Histone deacetylases function mainly to inhibit gene expression because histone deacetylation causes chromatin condensation, with roles in many different developmental processes and environmental responses. In wheat, it is known that differences in grain lysine content can be caused by differential expression of lysine-poor storage proteins (prolamins). Gill-Humanes et al. (2014), for example, identified downregulation of gliadins (a class of prolamins) as a method of increasing lysine content in wheat, and Moehs et al. (2019) showed that mutation of wheat prolamin binding factor (*WPBF*), a DOF-class transcription factor, increased lysine concentration. Lower prolamin protein content is also associated with increased lysine content in barley (Rustgi et al., 2019). However, the prolamins confer the viscoelastic properties of wheat dough that are required for the manufacture of many products, including bread, so this must also be considered when trying to breed for higher lysine content.

In maize, grain lysine content is similarly affected by the abundance of lysine-poor proteins in the prolamin family called zeins. The expression of particular zein genes is determined by a bZIP transcription factor called *Opaque2* (Gavazzi et al., 2007), and the mutant line lacking a functional *Opaque2* gene is characterised by higher kernel lysine content (Mertz, Bates & Nelson, 1964). Interestingly, the lysine QTL identified in this study is situated upstream of an *Opaque2* orthologue on chromosome 1A: TraesCS1A02G329900,

otherwise known as SPA (storage protein activator), which is known to activate storage protein synthesis in wheat (Albani et al., 1997). The A genome homeologue of SPA does not differ in sequence between Claire and Robigus, but differential expression of SPA (through differences in HDT1 regulation) is a possible mechanism by which this QTL could affect lysine content.

Future work investigating *HDT1*, *SPA*, and other regulatory genes of storage proteins in wheat would help to elucidate their effects on grain lysine content and would be useful for expanding the germplasm available to increase lysine content, given the limited QTL and small effect of GS we found. Chromosome-level assemblies of Claire and Robigus would also enable further analysis of this mapping population in the future. Combining both increased diversity and pangenomes, sequencing of the Watkins collection and construction of genome assemblies will enable novel diversity to be identified that can be introgressed into elite soft wheat germplasm as well (Shewry et al., 2022).

Conclusions

The nutritional quality of UK soft wheat can be improved incrementally using diversity from Claire and Robigus, but greater diversity is required to make larger gains. The genetic architecture of different amino acids differs considerably, and they are often controlled by QTL that impact other quality traits as well. Future soft wheat breeding in the UK should therefore consider use of more genetic diversity and using pleiotropic QTL to the benefit of farmers and consumers.

485 **Acknowledgements**

486 JO is supported by a BBSRC Collaborative Training Partnership Studentship
487 (BB/T50838X/1) with partners: University of Reading and Mondelez UK R&D Ltd. JO is
488 also supported by a scholarship from the Society of Chemical Industry. NGH is supported at
489 Rothamsted Research by the BBSRC *via* the Designing Future Wheat Programme
490 (BB/P016855/1). SR is supported by a Biotechnology and Biological Sciences Research
491 Council (BBSRC) Super Follow-on Fund grant (BB/T017007/1), with partners: University of
492 Bristol, AHDB, KWS UK Ltd, Saaten Union UK Ltd, RAGT Seeds Ltd, Syngenta UK Ltd,
493 and Limagrain UK Ltd. BBSRC is part of UK Research and Innovation. Many thanks to
494 Steve Hanley, Steve Thomas, Andy Phillips, Friedrich Longin, Matthias Rapp, and Keywan
495 Hassani-Pak for their discussions and help in this study.

496

497 **Conflict of Interest**

498 The authors declare no conflict of interest.

499

500 **Supplemental Material**

501 Supplementary figures – Supplementary figures 1 to 6.

502 Supplementary tables – Supplementary tables 1 to 8.

503 Supplementary data file 1 – R shiny QTL analysis files and data.

504 Supplementary data file 2 – HPC rrBLUP R markdown files.

505

506 **Data availability**

507 Data generated in this study is available in supplementary data file 1.

References

- Alarcón-Reverte, R., Xie, Y., Stromberger, J., Cotter, J.D., Mason, R.E., & Pearce, S. (2022). Induced mutations in ASPARAGINE SYNTHETASE-A2 reduce free asparagine concentration in the wheat grain. *Crop Science*, <https://doi.org/10.1002/csc2.20760>
- Albani, D., Hammond-Kosack, M.C.U., Smith, C., Conlan, S., Colot, V., Holdsworth, M., & Bevan, M.W. (1997). The wheat transcriptional activator SPA: A seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. *Plant Cell*, . <https://doi.org/10.1105/tpc.9.2.171>
- Aron Atkins, Jonathan McPherson and JJ Allaire (2021).rsconnect: Deployment Interface for R Markdown Documents and Shiny Applications. R package version 0.8.25. <https://CRAN.R-project.org/package=rsconnect>
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- Borrill, P., Ramirez-Gonzalez, R., & Uauy, C. (2016). expVIP: a customizable RNA-seq data analysis and visualization platform. *Plant Physiology*, 170, 2172–2186.
- Broman, K.W., Gatti, D.M., Simecek, P., Furlotte, N.A., Prins, P., Sen, Ś., Yandell, B.S., & Churchill, G.A. (2019). R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics*, 211, 495–502.
- Broman, K.W., Wu, H., Sen, Ś., & Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19, 889–890.

- Casebow, R., Hadley, C., Uppal, R., Addisu, M., Loddo, S., Kowalski, A., Griffiths, S., & Gooding, M. (2016). Reduced height (Rht) alleles affect wheat grain quality. *PloS one*, *11*, e0156056
- Chang, W., Cheng, J., Allaire, J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B. (2021). shiny:Web Application Framework for R. R package version 1.7.1. <https://CRAN.R-project.org/package=shiny>
- Csardi, G., Nepusz, T. (2006) The igraph software package for complex network research, InterJournal, Complex Systems 1695. <https://igraph.org>
- Curtis, T.Y., Powers, S.J., Wang, R., & Halford, N.G. (2018). Effects of variety, year of cultivation and sulphur supply on the accumulation of free asparagine in the grain of commercial wheat varieties. *Food Chemistry*, <https://doi.org/10.1016/j.foodchem.2017.06.113>
- DEFRA (1998). Bread and Flour Regulations.
- Department of Health and Social Care UK Government. (2021). *Folic Acid Added to Flour to Prevent Spinal Conditions in Babies*. <https://www.gov.uk/government/news/folic-acid-added-to-flour-to-prevent-spinal-conditions-in-babies> (accessed February 24, 2022).
- Emebiri, L.C. (2014). Genetic variation and possible SNP markers for breeding wheat with low-grain asparagine, the major precursor for acrylamide formation in heat-processed products. *Journal of the Science of Food and Agriculture*, . <https://doi.org/10.1002/jsfa.6434>
- Endelman, J.B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, *4*

- 551 FAO. (2021). *FAOSTAT: Food Balances (2014-)*. <http://www.fao.org/faostat/en/> (accessed
552 February 1, 2022).
- 553 Fradgley, N.S., Gardner, K., Kerton, M., Swarbreck, S.M., & Bentley, A.R. (2022). Trade-
554 offs in the genetic control of functional and nutritional quality traits in UK winter wheat.
555 *Heredity*, 1–14.
- 556 Gabry J, Mahr T (2022). “bayesplot: Plotting for Bayesian Models.” R package version 1.9.0,
557 <URL: <https://mc-stan.org/bayesplot/>>.
- 558 Galili, G., & Amir, R. (2013). Fortifying plants with the essential amino acids lysine and
559 methionine to improve nutritional quality. *Plant Biotechnology Journal*, 11, 211–222
- 560 Gamborg OL, Miller R, Ojima K. (1968). Nutrient requirements of suspension cultures of
561 soybean root cells. *Experimental Cell Research*, 50, 151–158.
- 562 Gavazzi, F., Lazzari, B., Ciceri, P., Gianazza, E., & Viotti, A. (2007). Wild-type opaque2 and
563 defective opaque2 polypeptides form complexes in maize endosperm cells and bind the
564 opaque2-zein target site. *Plant Physiology*, 145, 933–945
- 565 Gil-Humanes, J., Pistón, F., Altamirano-Fortoul, R., Real, A., Comino, I., Sousa, C., Rosell,
566 C.M., & Barro, F. (2014). Reduced-gliadin wheat bread: an alternative to the gluten-free
567 diet for consumers suffering gluten-related pathologies. *PloS one*, 9, e90898
- 568 Giovanny E. Covarrubias-Pazaran. (2019). *Heritability: Meaning and Computation*.
569 [https://excellenceinbreeding.org/sites/default/files/manual/EiB-M2_Heritability_18-02-](https://excellenceinbreeding.org/sites/default/files/manual/EiB-M2_Heritability_18-02-20.pdf)
570 20.pdf (accessed March 1, 2022).

- 571 Goodrich B, Gabry J, Ali I & Brilleman S. (2020). rstanarm: Bayesian applied regression
572 modeling via Stan. R package version 2.21.1 <https://mc-stan.org/rstanarm>.
- 573 Gressier, M., & Frost, G. (2022). Minor changes in fibre intake in the UK population between
574 2008/2009 and 2016/2017. *European Journal of Clinical Nutrition*, 76, 322–327
- 575 Hassani-Pak, K., Singh, A., Brandizi, M., Hearnshaw, J., Parsons, J.D., Amberkar, S.,
576 Phillips, A.L., Doonan, J.H., & Rawlings, C. (2021). KnetMiner: a comprehensive
577 approach for supporting evidence-based gene discovery and complex trait analysis
578 across species. *Plant Biotechnology Journal*, 19, 1670–1678
- 579 Howe, K.L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean,
580 I.M., Azov, A.G., Bennett, R., & Bhai, J. (2021). Ensembl 2021. *Nucleic Acids*
581 *Research*, 49, D884–D891
- 582 Jang, S., Han, J.-H., Lee, Y.K., Shin, N.-H., Kang, Y.J., Kim, C.-K., & Chin, J.H. (2020).
583 Mapping and validation of QTLs for the amino acid and total protein content in brown
584 rice. *Frontiers in Genetics*, 11, 240
- 585 Jiang, X., Deng, Z., Ru, Z., Wu, P., & Tian, J. (2013). Quantitative trait loci controlling
586 amino acid contents in wheat ('Triticum aestivum' L.). *Australian Journal of Crop*
587 *Science*, 7, 820–829
- 588 Jin, P., Gao, S., He, L., Xu, M., Zhang, T., Zhang, F., Jiang, Y., Liu, T., Yang, J., & Yang, J.
589 (2020). Genome-wide identification and expression analysis of the histone deacetylase
590 gene family in wheat (*Triticum aestivum* L.). *Plants*, 10, 19

- 591 Kassambara, A., and Mundt, F. (2020). factoextra: Extract and Visualize the Results of
592 Multivariate Data Analyses. R package version 1.0.7. [https://CRAN.R-](https://CRAN.R-project.org/package=factoextra)
593 [project.org/package=factoextra](https://CRAN.R-project.org/package=factoextra)
- 594 Knox, R.E., Clarke, J.M., and DePauw, R.M. (2000). Dicamba and growth condition effects
595 on doubled haploid production in durum wheat crossed with maize. *Plant Breeding*, 119,
596 289–298
- 597 Kolde, R. (2019). pheatmap: Pretty Heatmaps. R package version 1.0.12. [https://CRAN.R-](https://CRAN.R-project.org/package=pheatmap)
598 [project.org/package=pheatmap](https://CRAN.R-project.org/package=pheatmap)
- 599 Kuhn, M., Jackson, S., and Cimentada, J. (2020). corrr: Correlations in R. R package version
600 0.4.3. <https://CRAN.R-project.org/package=corrr>
- 601 Li, H., Liu, H., Pei, X., Chen, H., Li, X., Wang, J., & Wang, C. (2021). Comparative
602 Genome-Wide Analysis and Expression Profiling of Histone Acetyltransferases and
603 Histone Deacetylases Involved in the Response to Drought in Wheat. *Journal of Plant*
604 *Growth Regulation*, 1–14
- 605 Li, L., Zhang, Y., Zhang, Y., Li, M., Tian, X., Song, J., Luo, X., Xie, L., Wang, D., & He, Z.
606 (2021). Genome-Wide Linkage Mapping for Preharvest Sprouting Resistance in Wheat
607 Using 15K Single-Nucleotide Polymorphism Arrays. *Frontiers in Plant Science*, 12
- 608 Malunga, L.N., Ames, N., Khorshidi, A.S., Thandapilly, S.J., Yan, W., Dyck, A., Waterer, J.,
609 Malcolmson, L., Cuthbert, R., Sopiwnyk, E., & Scanlon, M.G. (2021). Association of
610 asparagine concentration in wheat with cultivar, location, fertilizer, and their interaction.
611 *Food Chemistry*, 344. <https://doi.org/10.1016/j.foodchem.2020.128630>

- 612 Mertz, E.T., Bates, L.S., & Nelson, O.E. (1964). Mutant gene that changes protein
613 composition and increases lysine content of maize endosperm. *Science*, 145, 279–280
- 614 Moehs, C.P., Austill, W.J., Holm, A., Large, T.A.G., Loeffler, D., Mullenberg, J., Schnable,
615 P.S., Skinner, W., van Boxtel, J., & Wu, L. (2019). Development of decreased-gluten
616 wheat enabled by determination of the genetic basis of lys3a barley. *Plant Physiology*,
617 179, 1692–1703
- 618 Oddy, J., Alarcón-Reverte, R., Wilkinson, M., Ravet, K., Raffan, S., Minter, A., Mead, A.,
619 Elmore, J.S., de Almeida, I.M., Cryer, N.C., Halford, N.G., & Pearce, S. (2021).
620 Reduced free asparagine in wheat grain resulting from a natural deletion of TaASN-B2:
621 investigating and exploiting diversity in the asparagine synthetase gene family to
622 improve wheat quality. *BMC plant biology*, 21. [https://doi.org/10.1186/s12870-021-](https://doi.org/10.1186/s12870-021-03058-7)
623 03058-7
- 624 Oddy, J., Raffan, S., Wilkinson, M.D., Elmore, J.S., & Halford, N.G. (2020). Stress, nutrients
625 and genotype: understanding and managing asparagine accumulation in wheat grain.
626 *CABI Agriculture and Bioscience*, 1. <https://doi.org/10.1186/s43170-020-00010-x>
- 627 Oddy, J., Raffan, S., Wilkinson, M.D., Elmore, J.S., & Halford, N.G. (2022). Understanding
628 the Relationships between Free Asparagine in Grain and Other Traits to Breed Low-
629 Asparagine Wheat. *Plants*, 11, 669
- 630 Pedersen, TL. (2021). ggraph: An Implementation of Grammar of Graphics for Graphs and
631 Networks. R package version 2.0.5. <https://CRAN.R-project.org/package=ggraph>

- 632 Peng, Y., Liu, H., Chen, J., Shi, T., Zhang, C., Sun, D., He, Z., Hao, Y., & Chen, W. (2018).
633 Genome-wide association studies of free amino acid levels by six multi-locus models in
634 bread wheat. *Frontiers in Plant Science*, 9, 1196
- 635 Prasanna, B.M., Palacios-Rojas, N., Hossain, F., Muthusamy, V., Menkir, A., Dhliwayo, T.,
636 Ndhlela, T., San Vicente, F., Nair, S.K., & Vivek, B.S. (2020). Molecular breeding for
637 nutritionally enriched maize: status and prospects. *Frontiers in Genetics*, 1392
- 638 Przewieslik-Allen, A.M., Wilkinson, P.A., Burridge, A.J., Winfield, M.O., Dai, X.,
639 Beaumont, M., King, J., Yang, C., Griffiths, S., & Wingen, L.U. (2021). The role of
640 gene flow and chromosomal instability in shaping the bread wheat genome. *Nature*
641 *Plants*, 7, 172–183
- 642 R Core Team (2021). R: A language and environment for statistical computing. R Foundation
643 for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 644 Raffan, S., Sparks, C., Huttly, A., Hyde, L., Martignago, D., Mead, A., Hanley, S.J.,
645 Wilkinson, P.A., Barker, G., Edwards, K.J., Curtis, T.Y., Usher, S., Kosik, O., &
646 Halford, N.G. (2021). Wheat with greatly reduced accumulation of free asparagine in the
647 grain, produced by CRISPR/Cas9 editing of asparagine synthetase gene TaASN2. *Plant*
648 *Biotechnology Journal*, 19. <https://doi.org/10.1111/pbi.13573>
- 649 Rapp, M., Schwadorf, K., Leiser, W.L., Würschum, T., & Longin, C.F.H. (2018). Assessing
650 the variation and genetic architecture of asparagine content in wheat: What can plant
651 breeding contribute to a reduction in the acrylamide precursor? *Theoretical and Applied*
652 *Genetics*, . <https://doi.org/10.1007/s00122-018-3163-x>

- 653 Rustgi, S., Shewry, P., Brouns, F., Deleu, L.J., & Delcour, J.A. (2019). Wheat seed proteins:
654 factors influencing their content, composition, and technological properties, and
655 strategies to reduce adverse reactions. *Comprehensive Reviews in Food Science and*
656 *Food Safety*, 18, 1751–1769
- 657 Shewry, P.R., Lovegrove, A., Wingen, L.U., & Griffiths, S. (2022). Opinion exploiting
658 genomics to improve the benefits of wheat: Prospects and limitations. *Journal of Cereal*
659 *Science*, 105, 10344
- 660 Shorinola, O., Balcárková, B., Hyles, J., Tibbits, J.F.G., Hayden, M.J., Holuřova, K., Valárik,
661 M., Distelfeld, A., Torada, A., & Barrero, J.M. (2017). Haplotype analysis of the pre-
662 harvest sprouting resistance locus Phs-A1 reveals a causal role of TaMKK3-A in global
663 germplasm. *Frontiers in Plant Science*, 8, 1555
- 664 Shorinola, O., Simmonds, J., Wingen, L.U., & Uauy, C. (2022). Trend, population structure,
665 and trait mapping from 15 years of national varietal trials of UK winter wheat. *G3*, 12,
666 jkab415
- 667 Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M.T., Brinton, J., Ramirez-Gonzalez,
668 R.H., Kolodziej, M.C., Delorean, E., Thambugala, D., Klymiuk, V., Byrns, B.,
669 Gundlach, H., Bandi, V., Siri, J.N., Nilsen, K., Aquino, C., Himmelbach, A., Copetti, D.,
670 Ban, T., Venturini, L., Bevan, M., Clavijo, B., Koo, D.H., Ens, J., Wiebe, K., N'Diaye,
671 A., Fritz, A.K., Gutwin, C., Fiebig, A., Fosker, C., Fu, B.X., Accinelli, G.G., Gardner,
672 K.A., Fradgley, N., Gutierrez-Gonzalez, J., Halstead-Nussloch, G., Hatakeyama, M.,
673 Koh, C.S., Deek, J., Costamagna, A.C., Fobert, P., Heavens, D., Kanamori, H.,
674 Kawaura, K., Kobayashi, F., Krasileva, K., Kuo, T., McKenzie, N., Murata, K., Nabeka,
675 Y., Paape, T., Padmarasu, S., Percival-Alwyn, L., Kagale, S., Scholz, U., Sese, J.,

- 676 Juliana, P., Singh, R., Shimizu-Inatsugi, R., Swarbreck, D., Cockram, J., Budak, H.,
677 Tameshige, T., Tanaka, T., Tsuji, H., Wright, J., Wu, J., Steuernagel, B., Small, I.,
678 Cloutier, S., Keeble-Gagnère, G., Muehlbauer, G., Tibbets, J., Nasuda, S., Melonek, J.,
679 Hucl, P.J., Sharpe, A.G., Clark, M., Legg, E., Bharti, A., Langridge, P., Hall, A., Uauy,
680 C., Mascher, M., Krattinger, S.G., Handa, H., Shimizu, K.K., Distelfeld, A., Chalmers,
681 K., Keller, B., Mayer, K.F.X., Poland, J., Stein, N., McCartney, C.A., Spannagl, M.,
682 Wicker, T., & Pozniak, C.J. (2020). Multiple wheat genomes reveal global variation in
683 modern breeding. *Nature*, 588. <https://doi.org/10.1038/s41586-020-2961-x>
- 684 Wang, L., Zhong, M., Li, X., Yuan, D., Xu, Y., Liu, H., He, Y., Luo, L., & Zhang, Q. (2008).
685 The QTL controlling amino acid content in grains of rice (*Oryza sativa*) are co-localized
686 with the regions involved in the amino acid metabolism pathway. *Molecular Breeding*,
687 21, 127–137
- 688 Wei, T., and Simko, V. (2021). R package 'corrplot': Visualization of a Correlation Matrix
689 (Version 0.92). Available from <https://github.com/taiyun/corrplot>
- 690 Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of*
691 *statistical software*, 40, 1–29
- 692 Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New
693 York.
- 694 Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43),
695 1686, <https://doi.org/10.21105/joss.01686>
- 696 Wickham, H., François, R., Henry, L., and Müller, K. (2022). dplyr: A Grammar of Data
697 Manipulation. R package version 1.0.8. <https://CRAN.R-project.org/package=dplyr>

- 698 Wilke, CO. (2020). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R
699 package version 1.1.1. <https://CRAN.R-project.org/package=cowplot>
- 700 Wilkinson, P.A., Allen, A.M., Tyrrell, S., Wingen, L.U., Bian, X., Winfield, M.O., BurrIDGE,
701 A., Shaw, D.S., Zaucha, J., Griffiths, S., Davey, R.P., Edwards, K.J., & Barker, G.L.A.
702 (2020). CerealsDB - New tools for the analysis of the wheat genome: Update 2020.
703 *Database*, 2020. <https://doi.org/10.1093/database/baaa060>
- 704 Wilkinson, P.A., Winfield, M.O., Barker, G.L.A., Allen, A.M., BurrIDGE, A., Coghill, J.A., &
705 Edwards, K.J. (2012). CerealsDB 2.0: an integrated resource for plant breeders and
706 scientists. *BMC Bioinformatics*, 13, 1–6
- 707 Xie, Y., Malunga, L.N., Ames, N.P., Waterer, J., Khorshidi, A.S., & Scanlon, M.G. (2021).
708 Effects of growing environment, genotype, and commercial fertilization levels on free
709 asparagine concentration in Western Canadian wheat. *Cereal Chemistry*, 98.
710 <https://doi.org/10.1002/cche.10364>
- 711 Yoo, S.-C. (2017). Quantitative trait loci controlling the amino acid content in rice (*Oryza*
712 *sativa* L.). *Journal of Plant Biotechnology*, 44, 349–355
- 713 Zhong, M., WANG, L., YUAN, D., LUO, L., XU, C., & HE, Y. (2011). Identification of
714 QTL affecting protein and amino acid contents in rice. *Rice Science*, 18, 187–195

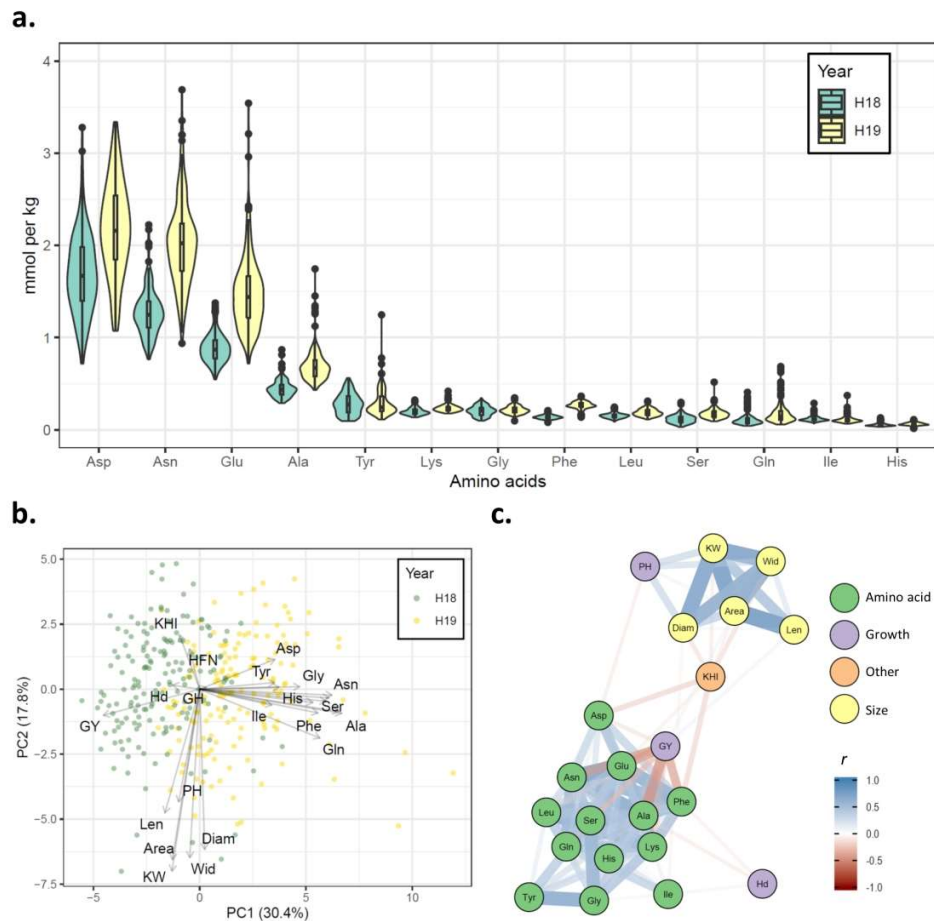


Figure 1. Characterisation of the Robigus \times Claire mapping population. a. Measurements of amino acids in the 2017–2018 (H18) and 2018–2019 (H19) harvest years. b. Principal component analysis of all traits in both years along the first two principal components. c. Correlation network analysis of all traits across both years (GH omitted, Kendall correlation, only links with significance <0.001 shown).

190x187mm (300 x 300 DPI)

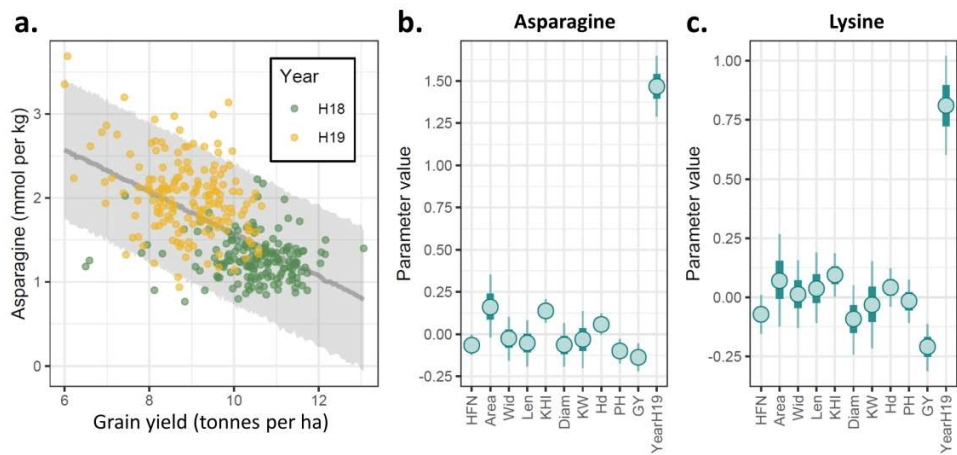


Figure 2. Relationships between free asparagine/lysine and other agronomic measurements. a. Linear modelling of free asparagine content against grain yield. The grey shaded ribbon shows 95% prediction intervals sampled from the posterior distribution. b. and c. Parameter values from multiple linear modelling of asparagine (b.) and lysine (c.) as explained by other quality traits measured in this population.

252x119mm (300 x 300 DPI)

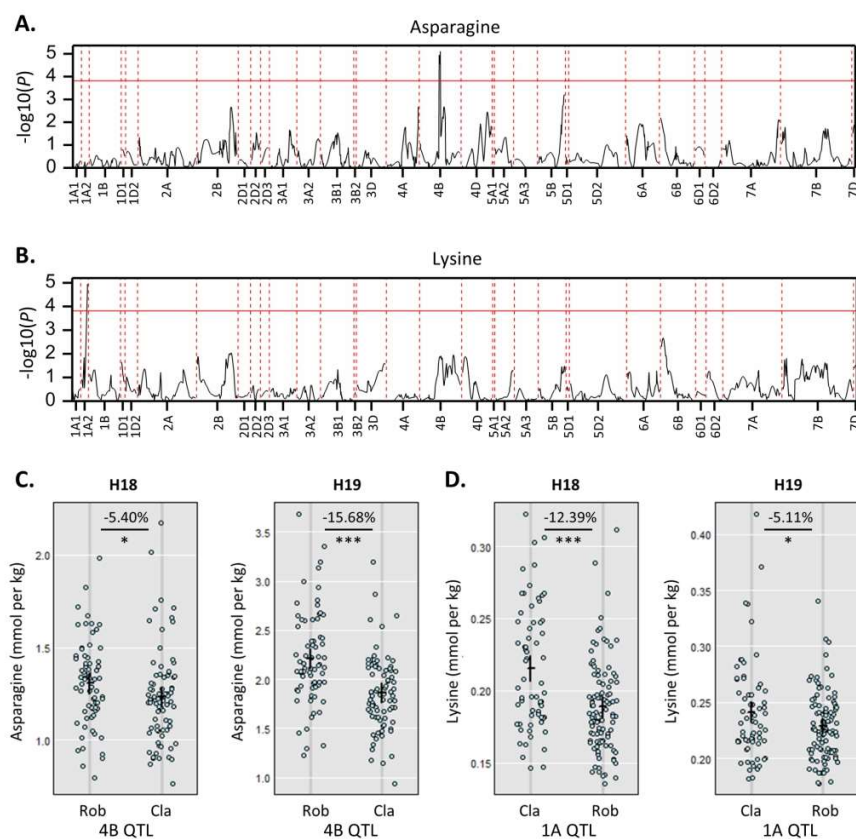


Figure 3. Identification of QTL controlling free asparagine and free lysine. a. Multi – environment genome scan plot for asparagine. b. Multi – environment genome scan plot for lysine. c. Impact of the asparagine QTL on free asparagine concentrations in both field trials. Error bars show plus and minus two times standard error of the mean. Significance values are taken from the corresponding years of the multi-environment linkage analysis.

190x175mm (300 x 300 DPI)

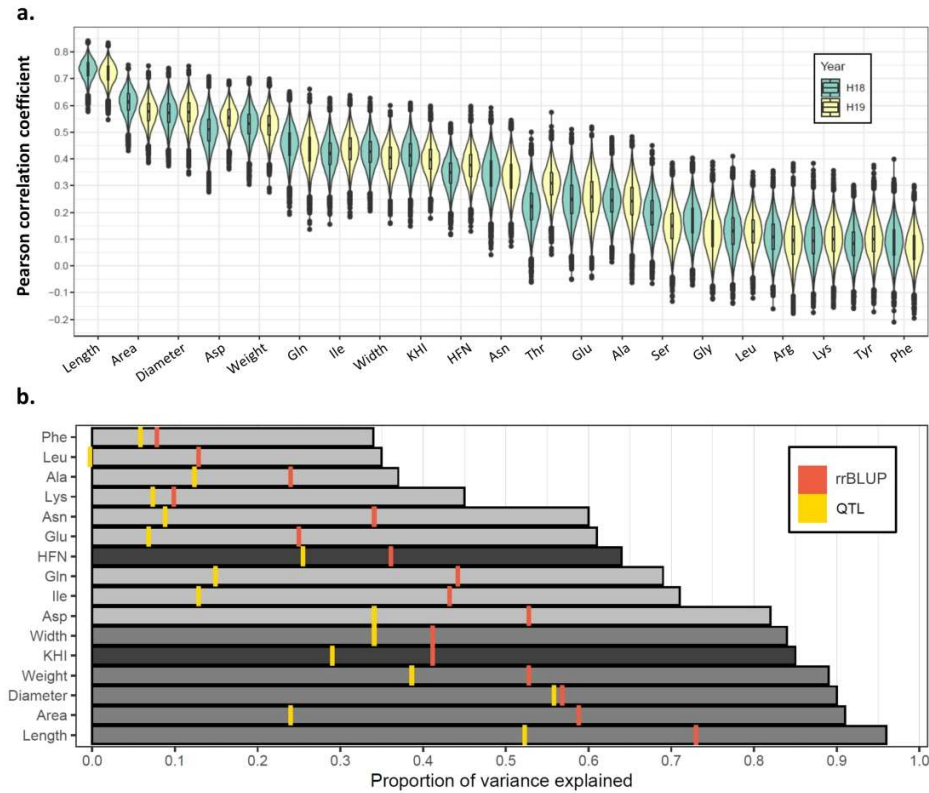


Figure 4. Variation explained by heritability, genomic prediction, and QTL. a. Genomic prediction accuracy between years. b. Additive QTL effects and genomic prediction (rrBLUP) accuracy (yellow and red marks, respectively) plotted alongside broad-sense heritability (shown as bars). Bars are shaded according to the trait group that they belong to (amino acid, size, or other).

252x209mm (300 x 300 DPI)

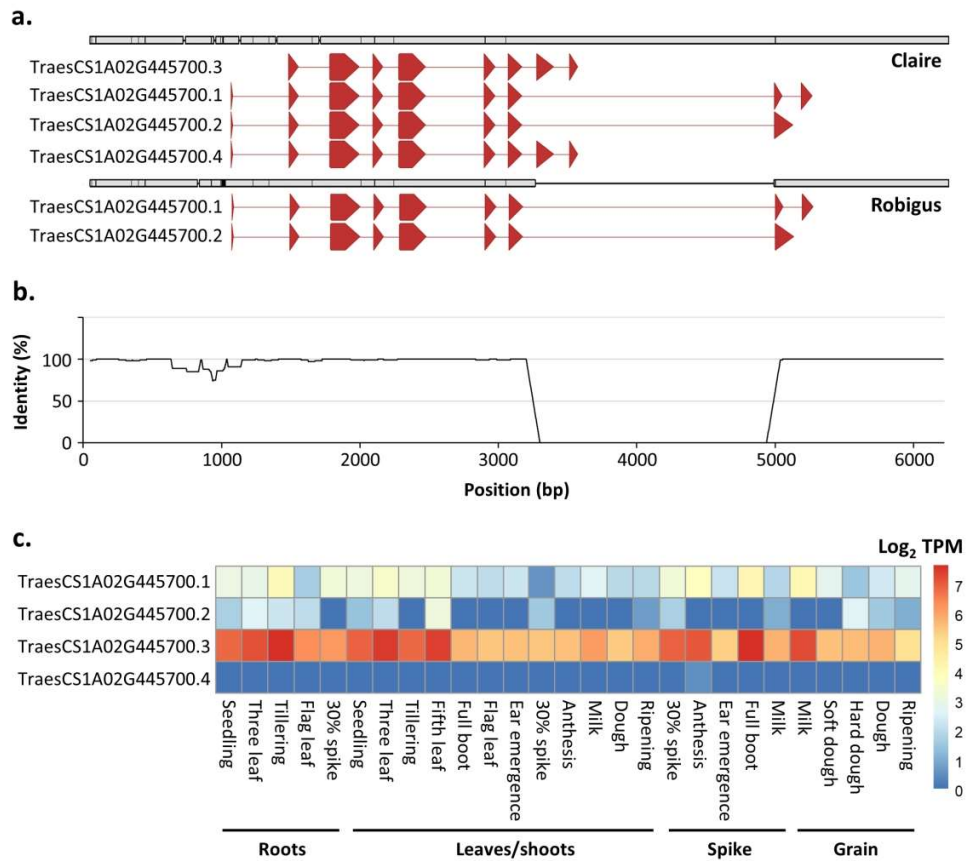


Figure 5. Analysis of the TaHDT-A1 candidate gene for lysine QTL between parents Claire and Robigus. a. Pairwise alignment of the two genes. b. Percentage identity calculated as a sliding window average of 100 bp. c. Expression of the four transcripts throughout development in variety Azhurnaya.

190x170mm (300 x 300 DPI)

Genetic control of grain amino acid composition in a UK soft wheat mapping population

Joseph Oddy¹, Monika Chhetry², Rajani Awal², John Addy¹, Mark Wilkinson¹, Dan Smith¹, Robert King¹, Chris Hall¹, Rebecca Testa¹, Eve Murray¹, Sarah Raffan¹, Tanya Y. Curtis³, Luzie Wingen², Simon Griffiths², Simon Berry⁴, J. Stephen Elmore⁵, Nicholas Cryer⁶, Isabel Moreira de Almeida⁷, and Nigel G. Halford^{1*}

¹*Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*

²*John Innes Centre, Norwich Research Park, Colney Lane, Norwich, UK*

³*Curtis Analytics Limited, Discovery Park, Sandwich, UK*

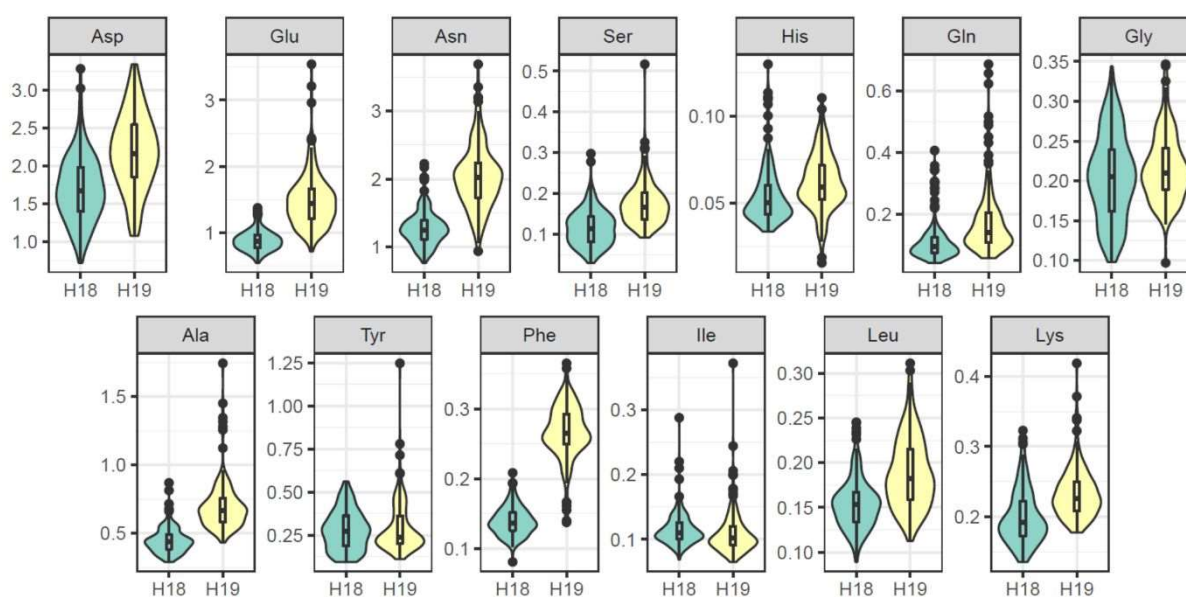
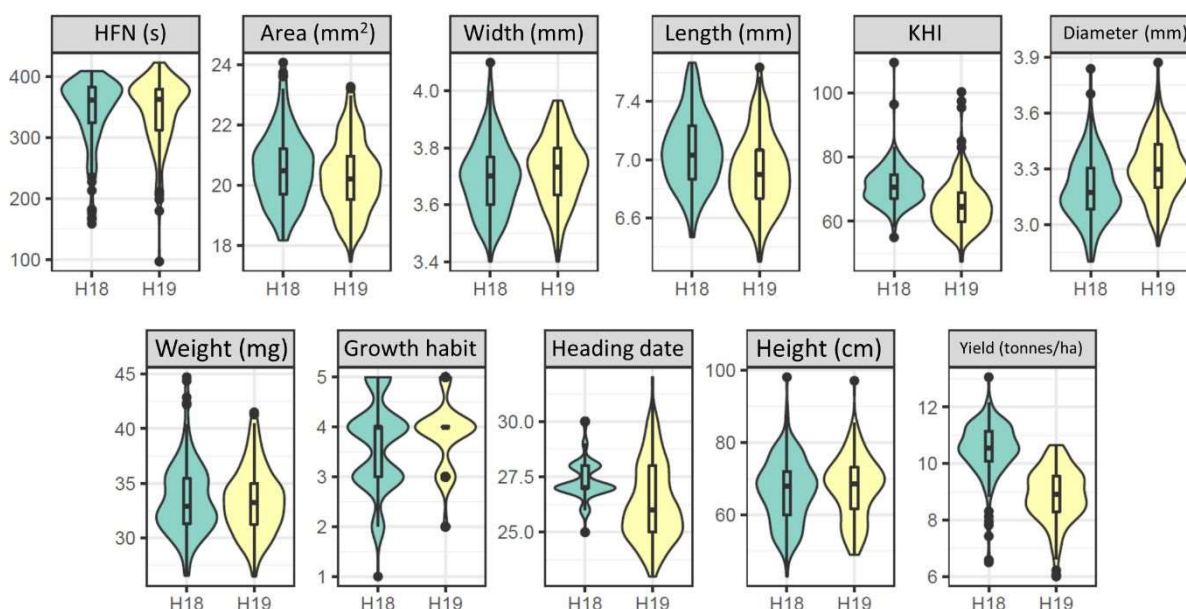
⁴*Limagrain UK Ltd, Market Rasen LN7 6DT, UK*

⁵*Department of Food & Nutritional Sciences, University of Reading, RG6 6DZ, UK*

⁶*Mondelēz UK R&D Ltd, Bournville Lane, Bournville, Birmingham, B30 2LU, UK*

⁷*Mondelēz R&D International, 6 Rue René Razel, 91400 Saclay, France*

Supplementary figures

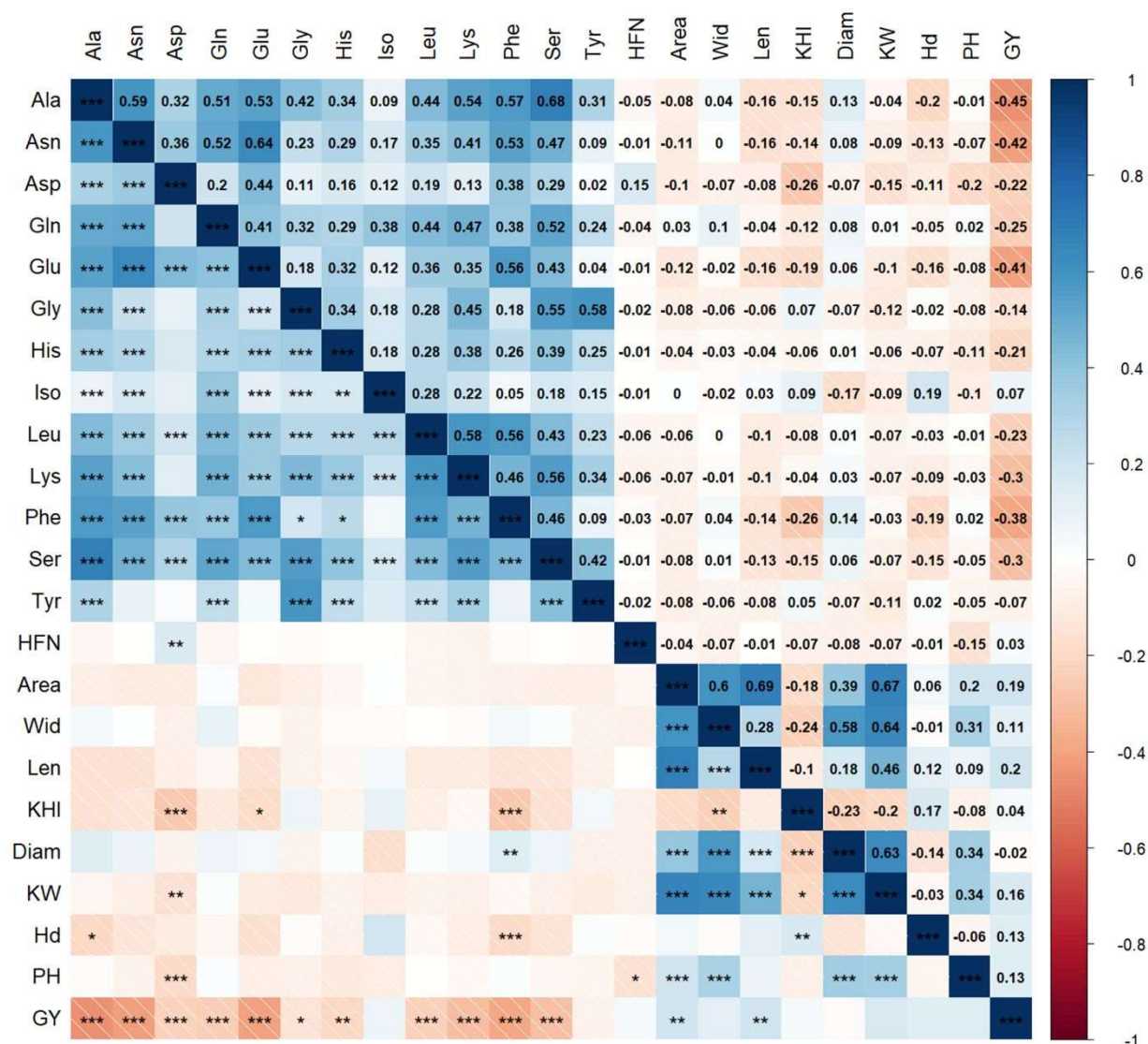
a.**b.**

Supplementary figure 1. Trait measurements from the Claire x Robigus mapping

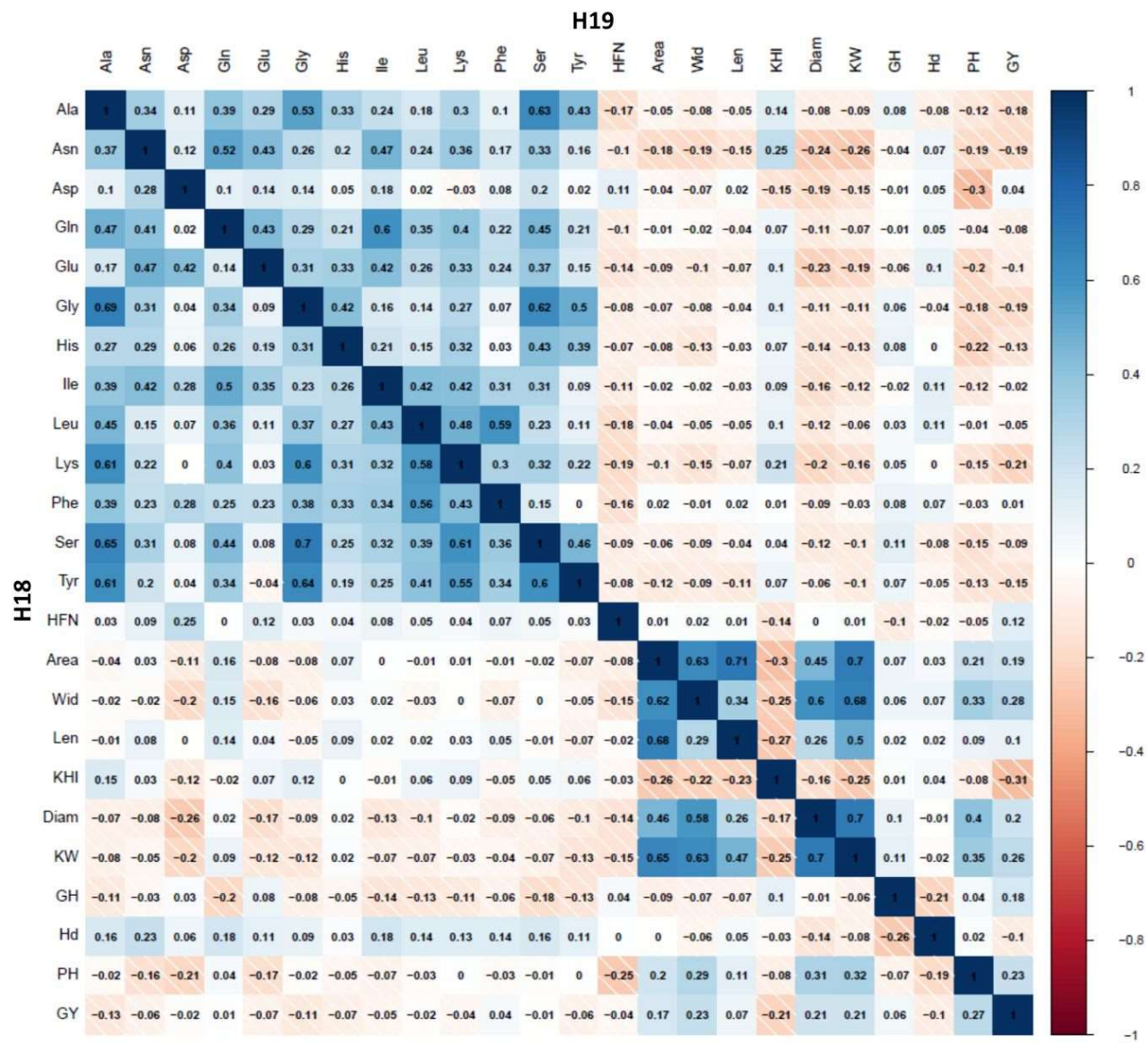
population grown across two years (H18 and H19). a. Concentration of amino acids

(mmol per kg) in wheat grain in both environments. **b.** Measurements of other quality and

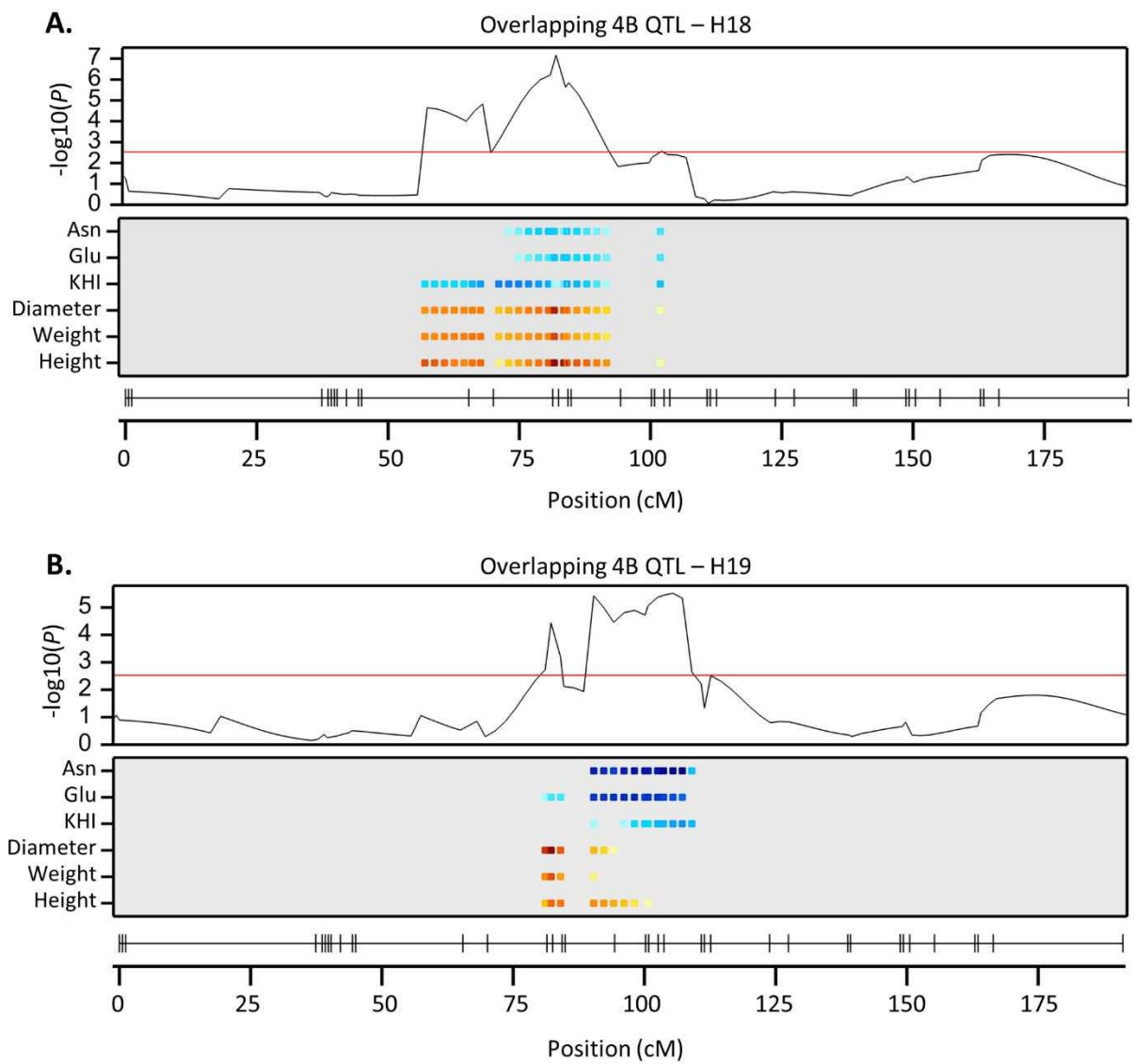
agronomic traits across both environments.



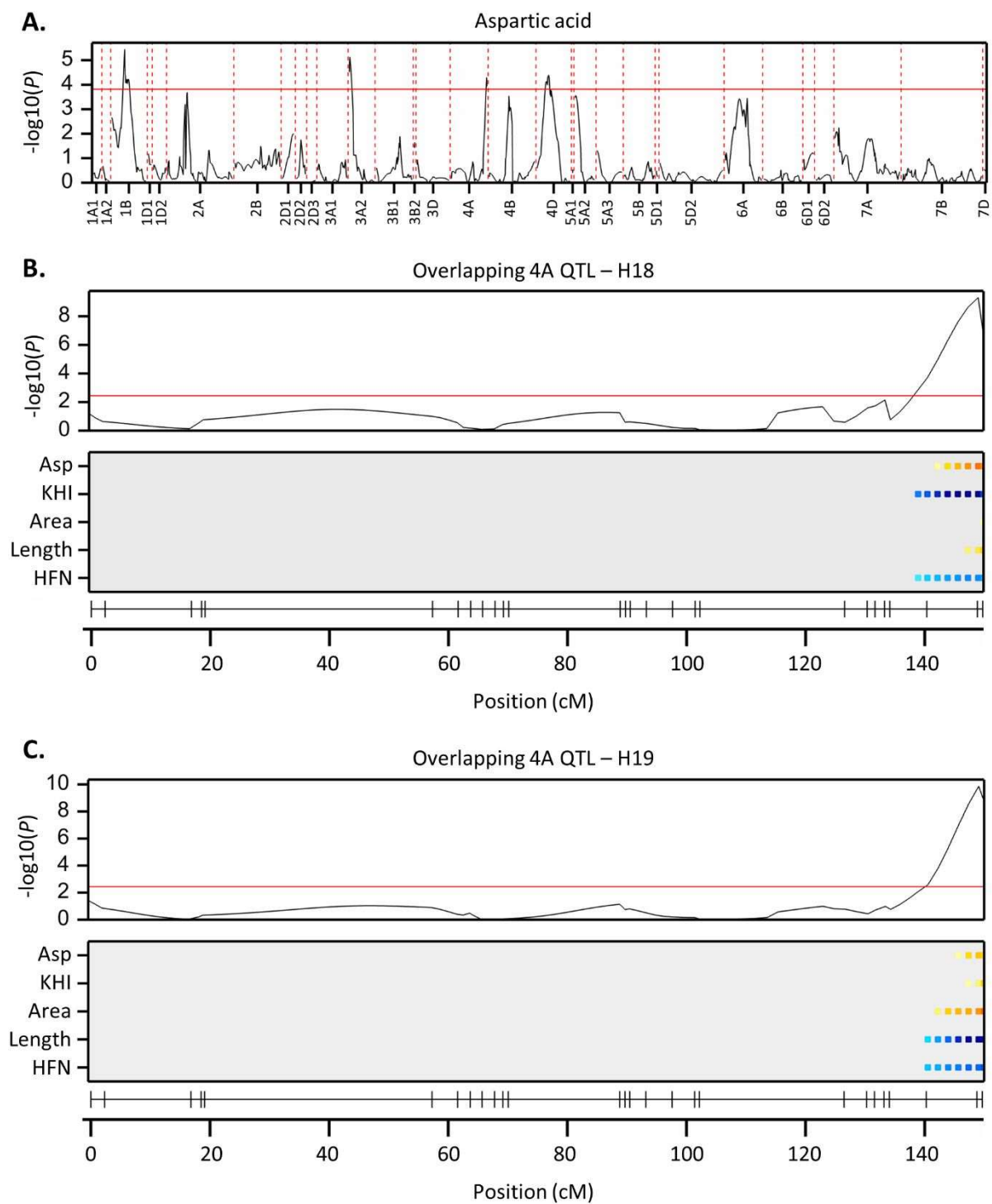
Supplementary figure 2. Correlation of traits taken from both H18 and H19 environments. Kendall correlation coefficients shown in upper right triangle and significance asterisks from adjusted p values (Bonferroni correction) shown in lower left triangle. HFN (Hagberg falling number), KHI (kernel hardness index), Diam (diameter), KW (kernel weight), Hd (heading date), PH (plant height), GY (grain yield).



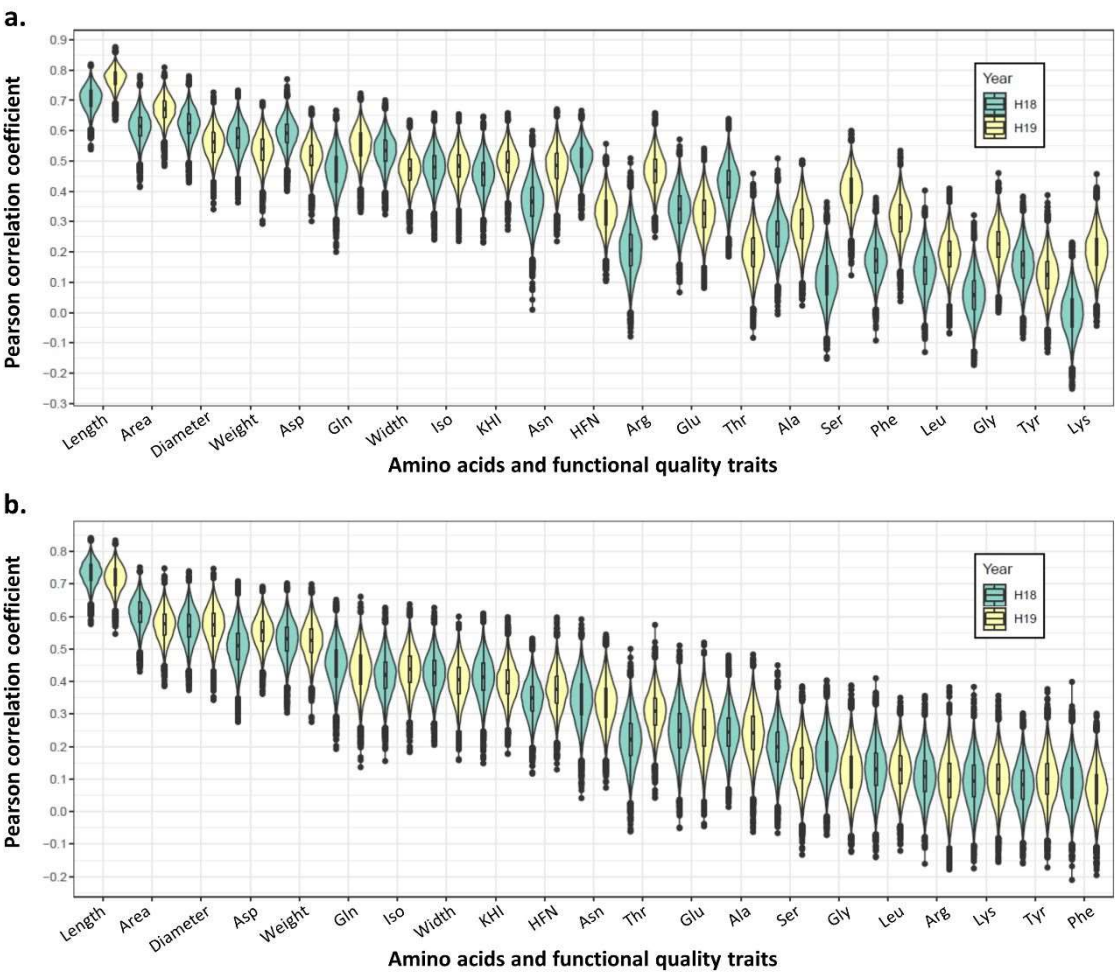
Supplementary figure 3. Kendall correlation coefficients between traits within each environment. H18 is the bottom left triangle, H19 is the upper right triangle.



Supplementary figure 4. Multi-trait analysis of the asparagine QTL on 4B in the Claire x Robigus mapping population. Blue indicates Robigus additive allele whilst red indicates Claire additive allele. The darkness of colour corresponds to the magnitude of the effect.



Supplementary figure 5. Multi-environment and multi-trait linkage analysis of aspartic acid QTL in the Claire x Robigus mapping population.



Supplementary figure 6. Accuracy of genomic selection for each trait measured in the mapping population using within (a.) and between (b.) year prediction.

Genetic control of grain amino acid composition in a UK soft wheat mapping population

Joseph Oddy¹, Monika Chhetry², Rajani Awal², John Addy¹, Mark Wilkinson¹, Dan Smith¹, Robert King¹, Chris Hall¹, Rebecca Testa¹, Eve Murray¹, Sarah Raffan¹, Tanya Y. Curtis³, Luzie Wingen², Simon Griffiths², Simon Berry⁴, J. Stephen Elmore⁵, Nicholas Cryer⁶, Isabel Moreira de Almeida⁷, and Nigel G. Halford^{1*}

¹*Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*

²*John Innes Centre, Norwich Research Park, Colney Lane, Norwich, UK*

³*Curtis Analytics Limited, Discovery Park, Sandwich, UK*

⁴*Limagrain UK Ltd, Market Rasen LN7 6DT, UK*

⁵*Department of Food & Nutritional Sciences, University of Reading, RG6 6DZ, UK*

⁶*Mondelēz UK R&D Ltd, Bournville Lane, Bournville, Birmingham, B30 2LU, UK*

⁷*Mondelēz R&D International, 6 Rue René Razel, 91400 Saclay, France*

Supplementary tables

Supplementary table 1. Broad sense heritability estimates, kendall correlation coefficients for traits across both environments, and within and between environments average genomic prediction accuracies (as Pearson correlation coefficients). Heritability estimates and genomic prediction performed on transformed data, kendall correlation coefficients performed on non-transformed data.

Traits	h^2	r	Within environment GP r			Between environments GP r		
			H18	H19	Mean	H18 train	H19 train	Mean
Amino acids								
Ala	0.37	0.00	0.26	0.29	0.27	0.24	0.24	0.24
Asn	0.60	0.34	0.36	0.48	0.42	0.35	0.33	0.34
Asp	0.82	0.51	0.59	0.52	0.55	0.51	0.55	0.53
Glu	0.61	0.33	0.34	0.33	0.33	0.25	0.26	0.25
Gln	0.69	0.30	0.47	0.55	0.51	0.45	0.43	0.44
Gly	0.00	-0.12	0.06	0.22	0.14	0.17	0.12	0.15
Iso	0.71	0.39	0.48	0.48	0.48	0.42	0.44	0.43
Leu	0.35	0.18	0.14	0.19	0.17	0.13	0.13	0.13
Lys	0.45	0.15	0.00	0.20	0.10	0.09	0.10	0.10
Phe	0.34	0.17	0.17	0.31	0.24	0.09	0.07	0.08
Ser	0.00	-0.14	0.11	0.40	0.25	0.20	0.15	0.17
Tyr	0.00	-0.23	0.16	0.12	0.14	0.08	0.10	0.09
Functional traits								
Area	0.91	0.63	0.61	0.67	0.64	0.61	0.58	0.59
Length	0.96	0.77	0.71	0.77	0.74	0.73	0.72	0.73
Width	0.84	0.55	0.53	0.47	0.50	0.43	0.40	0.41
Diameter	0.90	0.60	0.62	0.56	0.59	0.57	0.57	0.57
KHI	0.85	0.58	0.46	0.50	0.48	0.41	0.40	0.41
Weight	0.89	0.59	0.58	0.54	0.56	0.53	0.53	0.53
HFN	0.64	0.34	0.51	0.33	0.42	0.35	0.37	0.36

Supplementary table 2. Multi-environment QTL for measured amino acids. Chr. (Chromosome), Pos. (Position), QTL x E (QTL by environment interaction).

Trait	Single trait linkage analysis (H18 and H19)						H18			H19		
	Marker	Chr.	cM	Mbp	-log ₁₀ (p)	QTL x E	Ratio	Effect	S.E.	High value	p	% Expl.
Asparagine	WC.0221262	4B	114.47	601	5.96	yes	4.34	0.041	0.020	Robigus	0.040	2.6
Log _e alanine	WC.0218011	1A2	27.3	593	5.10	yes	0.34	0.065	0.016	Claire	<0.001	10.9
	WC.0223839	7B	211.2	719	5.03	no	1.00	0.052	0.012	Robigus	<0.001	7.1
Aspartic acid	WC.0218489	1B	54.4	530	5.40	no	1.00	0.123	0.027	Claire	<0.001	8
	WC.0214359	3A2	2.3	738	7.95	yes	1.68	0.118	0.029	Robigus	<0.001	7.3
	WC.0221037	4A	148.8	703	8.08	no	1.00	0.154	0.027	Claire	<0.001	12.6
	WC.0227146	4D	48.8	16	3.70	no	1.00	0.102	0.028	Claire	<0.001	5.5
Log _e glutamate	WC.0221329	4B	100.8	518	4.27	yes	2.41	0.034	0.013	Robigus	0.011	3.7
Log _e glutamine	WC.0218486	1B	117.93	660	2.16	yes	9.91	0.011	0.038	Robigus	0.760	0.1
	WC.0221302	4B	103.7	547	3.50	no	1.00	0.105	0.029	Robigus	<0.001	5.4
	WC.0228471	6B	19.7	25	5.09	no	1.00	0.129	0.029	Claire	<0.001	8.2
Glycine	WC.0218011	1A2	27.3	593	5.37	yes	0.53	0.017	0.004	Claire	<0.001	9.6
	WC.0226796	4B	155.2	327	4.26	no	1.00	0.010	0.002	Robigus	<0.001	3.2
-1/Isoleucine	WC.0221386	4B	94.3	172	7.31	yes	5.10	0.161	0.120	Robigus	0.180	1
	WC.0223785	7B	211.2	717	3.60	no	1.00	0.421	0.115	Robigus	<0.001	6.8
Log _e lysine	WC.0218011	1A2	27.3	593	4.95	yes	0.38	0.064	0.014	Claire	<0.001	12.1
Phenylalanine	WC.0220622	3B1	78.1	116	3.83	yes	1.80	0.005	0.002	Robigus	0.001	6.2
	WC.0226730	4B	123.8	632	3.39	yes	2.33	0.033	0.031	Claire	0.279	0.7
Log _e serine								0.077	0.020	Robigus	<0.001	7.9

Supplementary table 4. Sources of variation related to asparagine and falling number screened in this study in the Claire x Robigus mapping population. Chr. (chromosome).

Source of variation	Claire	Robigus	Chr.	Reference
ASN-B2 PAV	Absent	Absent	3B	Oddy et al., 2021
ASN-B1	Non-functional	Functional	5B	Oddy et al., 2021
ASN-A3.1	Non-functional	Non-functional	1A	Oddy et al., 2021
Rht-B1	Rht-B1a (WT)	Rht-B1b (Dwarf)	4B	Wilkinson et al., 2020
Rht-D1	Rht-D1b (Dwarf)	Rht-D1a (WT)	4D	Wilkinson et al., 2020
<i>T. dicoccoides</i> introgression	Absent	Present	4A	Przewieslik-Allen et al., 2021
TaMKK3A	A	A	4A	Shorinola et al., 2016
PM19-A1 promoter InDel	Deletion	Deletion	4A	Shorinola et al., 2016

For Review Only

Supplementary table 5. List of UK winter wheat varieties separated by *Rht-B1* allele status.

Type	Rht-B1 WT		Rht-B1 DWARF
G1	Avalon	Malacca	
	Cadenza	Shamrock	
	Crusoe	Skyfall	
	Gallant	Solstice	
	Hereward	Spark	
G2	Bonham	Evoke	Cashel
	Charger	Podium	
	Cordiale	Rialto	
	Cubanita	Shango	
	Einstein	Sterling	
G3	Claire	Invicta	Icon
	Cocoon	Scout	Monterey
	Croft	Tuxedo	Robigus
	Delphi	Warrior	Torch
	Diego	Weaver	Zulu
G4 - Hard	Badger	Icebreaker	Gator
	Buster	Kielder	Goldengun
	Dickens	Relay	Oakley
	Duxford	Savannah	Santiago
	Evolution		Solace
G4 - Soft	Alchemy	Leeds	Myriad
	Cougar	Revelation	Panacea
	Denman	Rowan	
	Horatio	Twister	
	Lancaster	Viscount	

Supplementary table 6. REML analysis of factors influencing asparagine content in field trials from 2011 – 2012 and 2012 – 2013.

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Year	125.04	1	125.04	<0.001
Rht_B1	0.17	1	0.17	0.676
ASN_B2	3.49	1	3.49	0.062
Treatment	125.4	1	125.4	<0.001
Year.Rht_B1	2.79	1	2.79	0.095
Year.ASN_B2	0.07	1	0.07	0.796
Rht_B1.ASN_B2	0.73	1	0.73	0.393
Year.Treatment	84.24	1	84.24	<0.001
Rht_B1.Treatment	3.09	1	3.09	0.079
ASN_B2.Treatment	6.57	1	6.57	0.010
Year.Rht_B1.ASN_B2	0.02	1	0.02	0.877
Rht_B1.ASN_B2.Variety	92.97	58	1.6	0.002
Year.Rht_B1.Treatment	3.34	1	3.34	0.068
Year.ASN_B2.Treatment	7.01	1	7.01	0.008
Rht_B1.ASN_B2.Treatment	0.73	1	0.73	0.394
Year.Rht_B1.ASN_B2.Variety	7.35	7	1.05	0.393
Year.Rht_B1.ASN_B2.Treatment	1.55	1	1.55	0.213
Rht_B1.ASN_B2.Variety.Treatment	64.83	58	1.12	0.251
Year.Rht_B1.ASN_B2.Variety.Treatment	6.54	7	0.93	0.478

Supplementary table 7. Physical locations of the HFN and Asn QTL in varieties Chinese Spring and Robigus.

Trait	Chr.	Location	Peak (Mbp)	Lower CI (Mbp)	Upper CI (Mbp)	QTL size (bp)	No. of genes
HFN	4A	CS	733	691	745	54,058,906	824
Asn	4B	CS	601	533	632	96,765,195	754
Lys	1A	CS	593	590	594	4,471,109	50

Supplementary table 8. Physical location of the lysine QTL in the wheat pangenome (chromosome level assemblies) and gene content.

Genome	Chr	Lower CI	Upper CI	QTL size	No. of genes
IWGSC	1A	590	594	4,471,109	50
Arinalrfor	1A	598	603	4,559,393	55
Jagger	1A	592	596	4,481,576	60
Julius	NA	NA	NA	NA	NA
Lancer	1A	591	595	4,548,456	59
Landmark	1A	593	595	2,101,319	38
Mace	1A	586	591	4,477,846	55
SY Mattis	1A	596	601	4,560,453	66
Norin61	1A	589	594	4,645,175	58
Stanley	scaffold_v3_2071	5	NA	NA	NA

For Review Only