

# 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://access.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](mailto: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:	<p>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</p>
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 Oddy<sup>1</sup>, Monika Chhetry<sup>2</sup>, Rajani Awal<sup>2</sup>, John Addy<sup>1</sup>, Mark Wilkinson<sup>1</sup>, Dan Smith<sup>1</sup>, Robert King<sup>1</sup>, Chris Hall<sup>1</sup>, Rebecca Testa<sup>1</sup>, Eve Murray<sup>1</sup>, Sarah Raffan<sup>1</sup>, Tanya Y. Curtis<sup>3</sup>, Luzie Wingen<sup>2</sup>, Simon Griffiths<sup>2</sup>, Simon Berry<sup>4</sup>, J. Stephen Elmore<sup>5</sup>, Nicholas Cryer<sup>6</sup>, Isabel Moreira de Almeida<sup>7</sup>, and Nigel G. Halford<sup>1\*</sup>

<sup>1</sup>*Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*

<sup>2</sup>*John Innes Centre, Norwich Research Park, Colney Lane, Norwich, UK*

<sup>3</sup>*Curtis Analytics Limited, Discovery Park, Sandwich, UK*

<sup>4</sup>*Limagrain UK Ltd, Market Rasen LN7 6DT, UK*

<sup>5</sup>*Department of Food & Nutritional Sciences, University of Reading, RG6 6DZ, UK*

<sup>6</sup>*Mondelēz UK R&D Ltd, Bournville Lane, Bournville, Birmingham, B30 2LU, UK*

<sup>7</sup>*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

1 **Abstract**

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

## 17 **Introduction**

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

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

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

46 In recent years, studies have investigated genetic strategies for the reduction of free  
47 asparagine content in wheat grain. Induced and natural variation in the asparagine synthetase  
48 2 genes, for example, has been found to impact significantly on free asparagine content  
49 (Raffan *et al.*, 2021; Oddy *et al.*, 2021; Alarcon-Reverte *et al.*, 2022) and quantitative trait  
50 loci (QTL) for grain asparagine content have been identified from previous GWAS studies  
51 (Emebiri, 2014; Peng *et al.*, 2018; Rapp *et al.*, 2018). However, the small number of stable  
52 QTL available to breeders limits the progress that can be made to reduce grain asparagine  
53 content in breeding programmes and no genetic strategies for soft (biscuit) wheat specifically  
54 have been investigated. Similarly, there are limited strategies currently available for  
55 increasing lysine content in wheat grain. Lysine biofortification *via* QTL identification and  
56 marker-assisted breeding has been studied extensively in both rice (Wang *et al.*, 2008; Zhong  
57 *et al.*, 2011; Yoo, 2017; Jang *et al.*, 2020) and maize (Prasanna *et al.*, 2020), but only two  
58 studies have previously investigated lysine biofortification in wheat through association  
59 studies. Peng *et al.* (2018) successfully identified QTL controlling free lysine and Jiang *et al.*  
60 (2013) identified QTL for total lysine.

61 Consequently, the aim of this study was to investigate QTL, genomic prediction  
62 accuracy, and candidate genes controlling the free amino acid composition of wheat grain in  
63 a soft wheat mapping population developed from the varieties Claire and Robigus. Like many  
64 UK varieties, these parents both lack the B genome homeologue of the asparagine synthetase-  
65 2 gene, *TaASN-B2* (TraesLDM3B03G01566640 in variety Landmark), the presence/absence  
66 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

73

## Materials and Methods

### 74 *Production of Doubled Haploid lines*

75 Doubled Haploid lines of Robigus x Claire were produced using a modified Knox *et al.*  
76 (2000) method. Wheat spikes were emasculated between growth stages GS55 and GS59.  
77 Once the stigma was receptive it was fertilised with freshly shed donor maize pollen. After  
78 one day, wheat florets were treated with Dicamba (20mgL<sup>-1</sup>) (Sigma-Aldrich, D5417) and  
79 injected into the plant stem (100mgL<sup>-1</sup>). Developing embryos were excised between 14 and  
80 21 days. Under aseptic conditions, seeds were removed from the spikelets, surface sterilised  
81 with 70% (v/v) ethanol (EtOH) for 1 min, rinsed with sterile distilled water, and immersed in  
82 20% (v/v) commercial bleach solution with a few drops of Tween® 20 for 20 mins. They  
83 were then rinsed with sterile distilled water three times.

84 Haploid embryos were excised and grown on 90mm Petri dishes in the dark on  
85 Gamborg's B5 media with minimal organics (Gamborg *et al.*, 1968), 2% (w/v) sucrose, pH  
86 5.8, 9gL<sup>-1</sup> Difco bactoagar at 20°C. When showing signs of germination, embryos were  
87 transferred to a light incubator at 20°C. Any non-germinated 1 month old embryos were  
88 given cold shock treatment at 4°C for 7 days to promote germination. Germinated plantlets  
89 were vernalised for 4 weeks and were grown in the glasshouse until the 4-tiller stage. Plants  
90 were then given colchicine (Sigma-Aldrich, C9754) treatment for 5 to 6 hours in the light at  
91 room temperature, washed and transplanted to soil, acclimatised and grown in a glasshouse.  
92 The mapping population was genotyped by Limagrain using a proprietary SNP array. The  
93 genetic map comprising 872 loci was constructed using MSTMap Online  
94 (<http://mstmap.org/>).

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

98 design. The H18 field trial was drilled on the 21<sup>st</sup> September 2017 and harvested on the 1<sup>st</sup>  
99 August 2018. The H19 field trial was drilled on the 14<sup>th</sup> September 2018 and harvested on the  
100 12<sup>th</sup> August 2019. Growth habit, heading date, plant height, and yield traits were scored in the  
101 field.

102

### 103 *Phenotyping*

104 Grain diameter, kernel hardness index (KHI), and grain weight measurements were recorded  
105 for 300 kernels from each line in the population using a Perten Single Kernel Classification  
106 System (SKCS) 4100 (Calibre Control International Ltd., Warrington, UK). Grain length  
107 (mm), width (mm), and area (mm<sup>2</sup>) measurements were recorded in triplicate for each sample  
108 using a MARVIN Seed Analyser and software Marvin 4.0 (MARViTECH GmbH,  
109 Wittenburg, Germany). Grain samples were milled to wholemeal flour in a coffee grinder and  
110 flour moisture content was recorded using a Minispec nuclear magnetic resonance (NMR)  
111 analyser (Minispec Mq10, Bruker Inc., Germany). Hagberg falling number measurements  
112 were recorded using an FN 1000 as the average of two technical replicates (Perten, Sweden),  
113 adjusting for flour moisture content as required according to manufacturer's instructions.  
114 Amino acid analysis was performed on wholemeal flour samples by HPLC as described  
115 previously (Raffan *et al.*, 2021) by Curtis Analytics (Sandwich, UK). Briefly, free amino  
116 acids were extracted from 0.5g of wholemeal flour and underwent precolumn derivatisation  
117 (Curtis *et al.*, 2018). Samples were then run on an HPLC system identically to previously  
118 described (Raffan *et al.*, 2021). Three technical replicates were taken for each sample for  
119 amino acid measurement.

120

### 121 *Phenotypic data analysis*

122 Skewness and kurtosis were measured for all variables in each environment and normal plots  
123 visually inspected in Genstat (VSN International, 2021) to determine if variables required  
124 transformation. The data were appropriately transformed according to their distribution if  
125 necessary (see Tables S1 and S2 for details of transformations). Subsequent analyses were  
126 performed on transformed variables unless otherwise stated. Plotting was performed in R (R  
127 Core Team, 2021) with the packages ggplot2 (Wickham, 2016), tidyverse (Wickham et al.,  
128 2019), and cowplot (Wilke, 2020).

129 Broad-sense heritability for each trait was estimated as described in Covarrubias-  
130 Pazaran (2019) using the packages dplyr (Wickham et al., 2022) and lme4 (Bates et al.,  
131 2015). Kendall rank correlation coefficients were performed on non-transformed data and  
132 adjusted  $p$  values (Bonferroni correction) were calculated for plotting using R (R Core Team,  
133 2021) and the package corrrplot (Wei & Simko, 2021). Principal component analysis was  
134 performed on untransformed, scaled variables using the package factoextra (Kassambara and  
135 Mundt, 2020). Correlation network analysis was performed and plotted by filtering for  
136 significant correlations where  $p < 0.001$  using Kendall correlation with Bonferroni correction  
137 using the packages corrr (Kuhn, Jackson and Cimentada, 2020), igraph (Csardi and Nepusz,  
138 2006), and ggraph (Pedersen, 2021).

139 Bayesian modelling was performed on untransformed variables in R using the  
140 package rstanarm (Goodrich *et al.*, 2020). Variables were scaled before modelling and  
141 individual linear models for each predictor variable were created to guide the selection of  
142 informative priors. Simulations of the posterior distribution were subsequently performed to  
143 check model fit and intervals were plotted using the package bayesplot (Gabry and Mahr,  
144 2022).  $R^2$  estimates were obtained by taking the median of leave-one-out cross validation  
145 adjusted estimates.

146

### 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-  
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

172 (Wilkinson *et al.*, 2020). Assays were run in 96-well plates in an Applied Biosystems™ 7500  
173 Real-Time PCR System. Primer mix was made using 46µL dH<sub>2</sub>O, 30µL common primer  
174 (100µM), and 12µL of each tailed primer (100µM). Each reaction contained 0.14µL KASP  
175 primer mix, 2.86µL water, 5µl KASP low-ROX mix (PACE), and 2µL DNA sample. Cycling  
176 conditions were 95 °C for 15 min, followed by 10 cycles of 95 °C 20s, 61 °C 60s (reducing  
177 anneal 0.6 °C per cycle), followed by 30 cycles of 95 °C 20s, 55 °C 60s. Data were then read  
178 and analysed using KlusterCaller genotyping software (LGC Biosearch Technologies).

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

184

#### 185 *Genomic prediction*

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

195

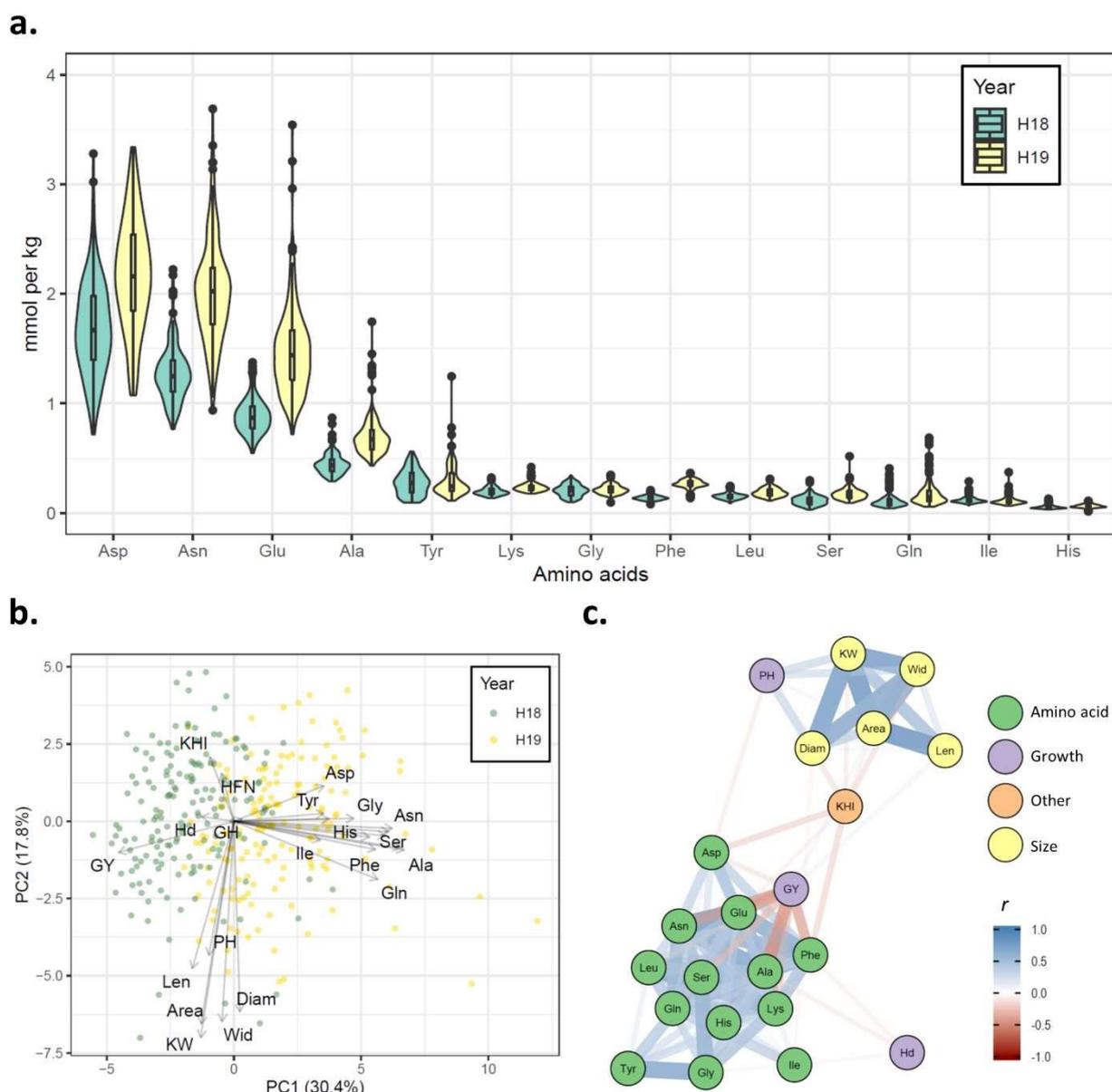
#### 196 *Candidate gene analysis*

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/](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.

209

**Results**210 *Phenotypic analysis*

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



222

223 **Figure 1. Characterisation of the Robigus × Claire mapping population. a.**

224 Measurements of amino acids in the 2017–2018 (H18) and 2018–2019 (H19) harvest years.

225 **b.** Principal component analysis of all traits in both years along the first two principal226 components. **c.** Correlation network analysis of all traits across both years (GH omitted,

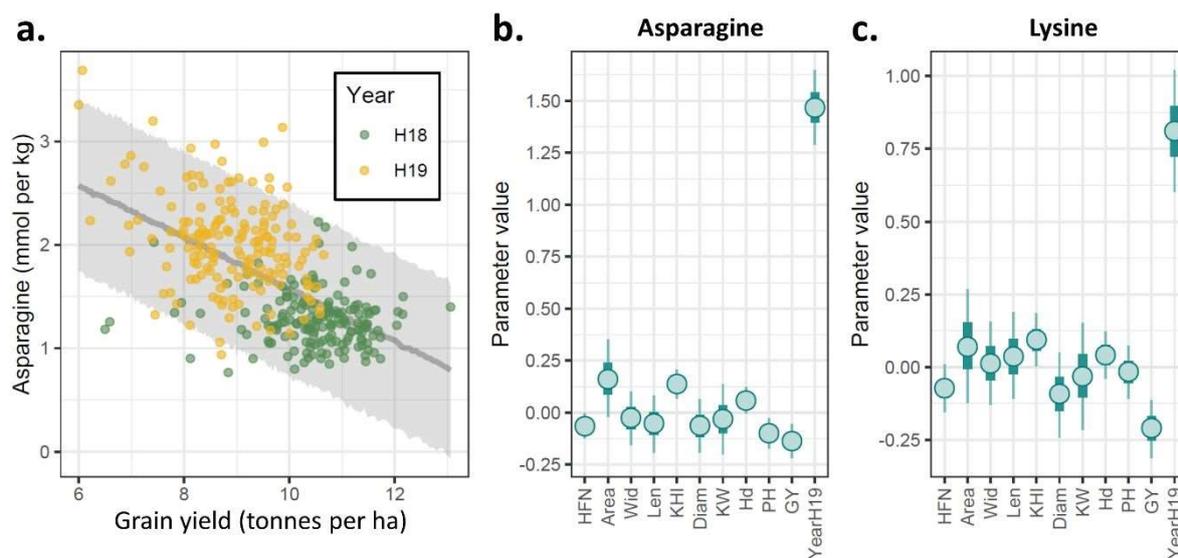
227 Kendall correlation, only links with significance &lt;0.001 shown).

228

229 To understand whether any of the traits we measured could predict free asparagine or lysine

230 content in the grain, we constructed Bayesian linear models with the quality traits and harvest

231 year as explanatory variables (Figure 2b; Figure 2c). In both the free asparagine (Figure 2b)  
 232 and lysine (Figure 2c) models, environment had the greatest effect whereas other variables  
 233 had little explanatory power. Nevertheless, the variance explained in the models was still  
 234 reasonable for asparagine at 56.5%, but only 22.2% for lysine.  
 235

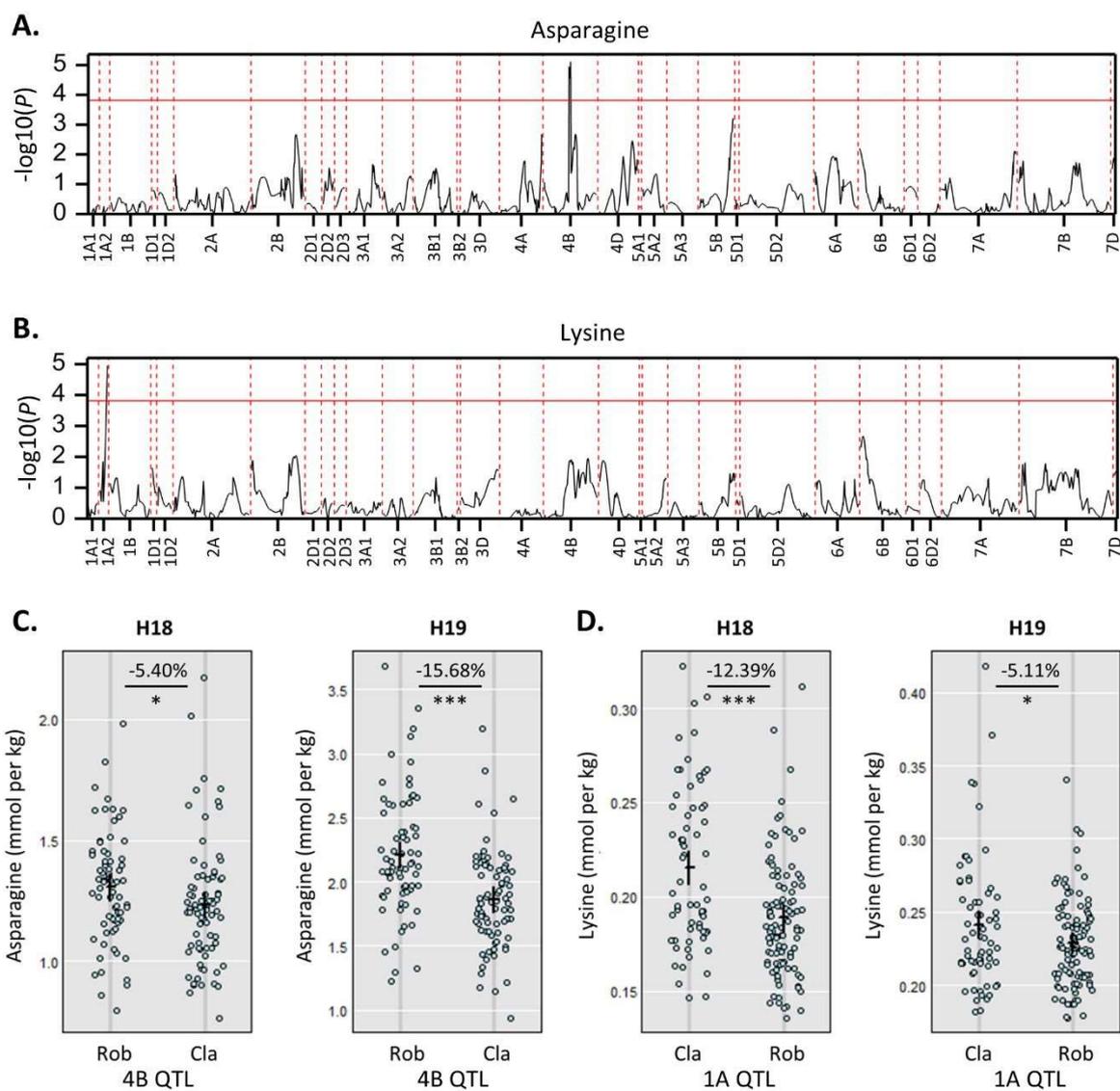


236  
 237 **Figure 2. Relationships between free asparagine/lysine and other agronomic**  
 238 **measurements. a.** Linear modelling of free asparagine content against grain yield. The grey  
 239 shaded ribbon shows 95% prediction intervals sampled from the posterior distribution. **b.** and  
 240 **c.** Parameter values from multiple linear modelling of asparagine (**b.**) and lysine (**c.**) as  
 241 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).



262

263

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

270

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

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

281

282 **Table 1.** Multi-environment QTL for measured amino acids. Chr. (Chromosome), cM  
 283 (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

284

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

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

291

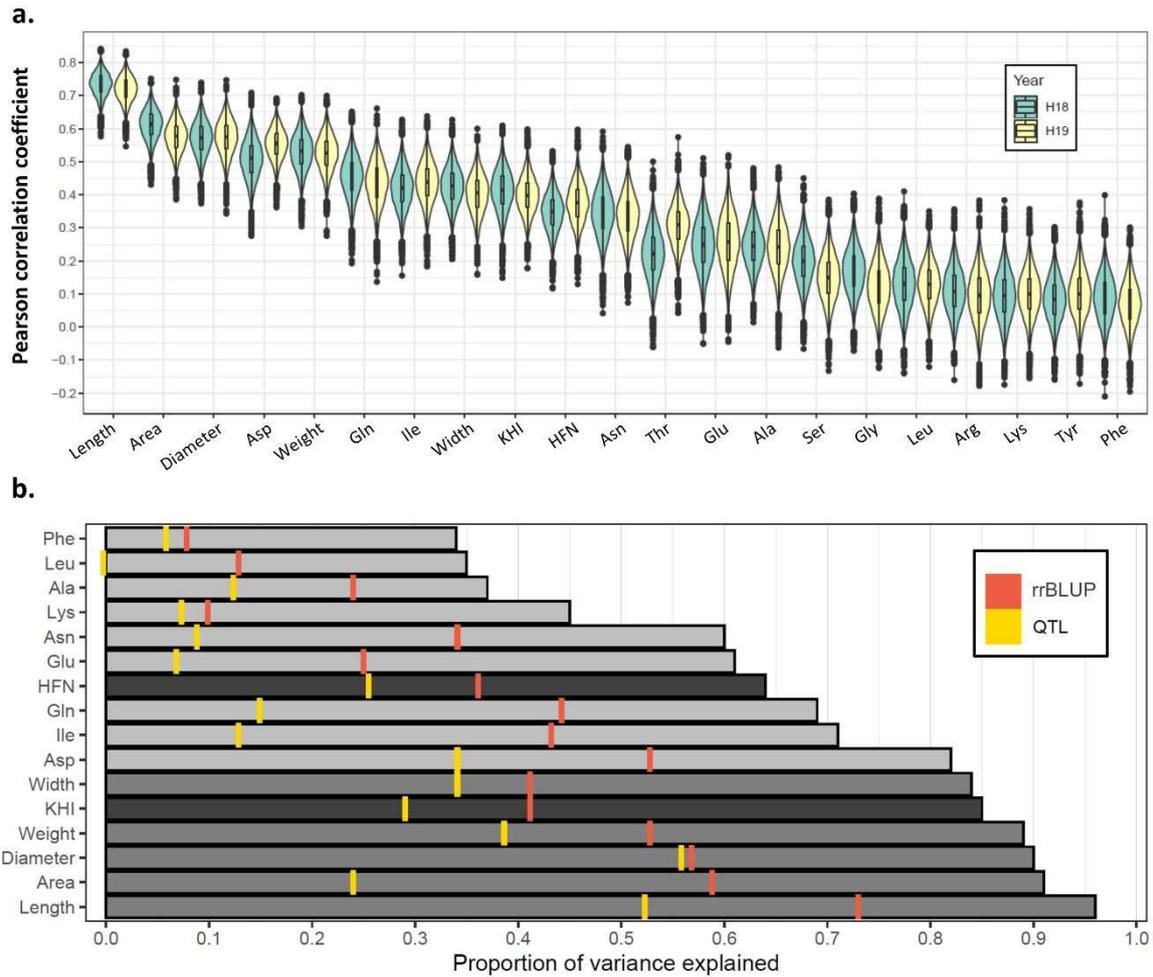
292 **Table 2.** Multi-environment QTL impacting both amino acids and other traits on  
 293 chromosomes 4A, 4B, and 4D. Chr. (Chromosome), cM (centimorgan), Mbp (megabase pair  
 294 location in Chinese Spring v1.0).

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

295

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).



308

309 **Figure 4. Variation explained by heritability, genomic prediction, and QTL. a.** Genomic310 prediction accuracy between years. **b.** Additive QTL effects and genomic prediction

311 (rrBLUP) accuracy (yellow and red marks, respectively) plotted alongside broad-sense

312 heritability (shown as bars). Bars are shaded according to the trait group that they belong to

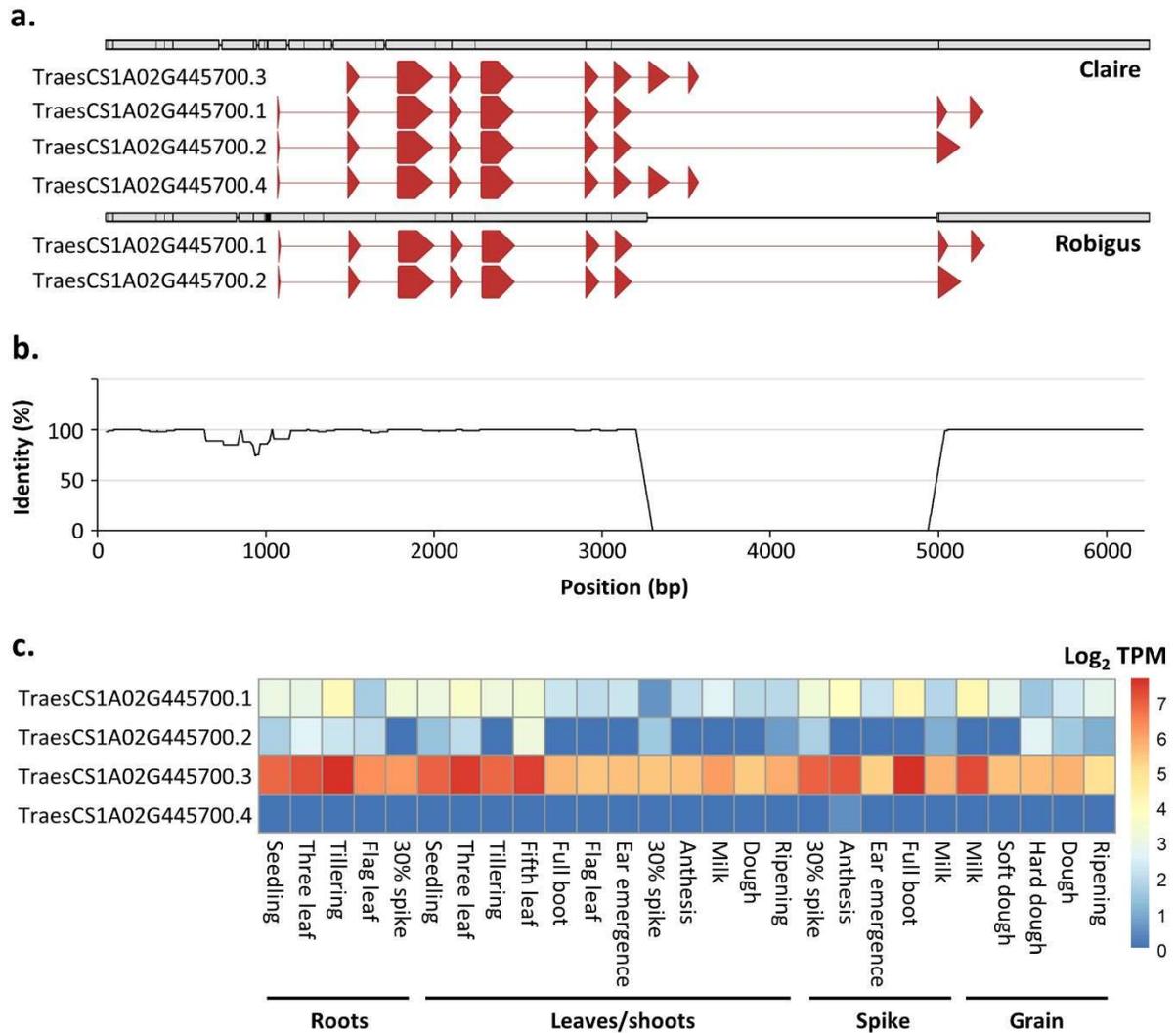
313 (amino acid, size, or other).

314

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.



339

## Discussion

340

### 341 **Limited variation in Claire and Robigus for asparagine and lysine improvement**

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

351 Our analysis found that there is variation between Claire and Robigus and that this  
352 does impact asparagine and lysine content to a small extent. Asparagine had a moderate  
353 heritability (0.60) across both field trials in the study, whereas the heritability for lysine was  
354 lower (0.45). One QTL was found for asparagine and lysine each, both explaining less than  
355 10% of the variance on average. The asparagine QTL identified here (peak at 601.4 Mbp in  
356 Chinese Spring) lies around 60 Mbp from another QTL (peak at 660.7 Mbp in Chinese  
357 Spring) identified by Peng et al. (2018), suggesting that these may coincide, whereas the  
358 lysine QTL does not overlap with previously identified QTL. Genomic selection had a  
359 predictive ability of 0.34 on average for asparagine, indicating that this method may be better  
360 suited for breeding because of the genetic architecture of this trait (many small-effect QTL).  
361 Rapp et al. (2018) also found that GS had a predictive ability of around 0.5 on average for  
362 asparagine, the higher estimate in this study likely due to within environment prediction and  
363 analysis of a more diverse mapping population. GS only achieved a predictive ability of 0.10

364 for lysine, indicating that only incremental advances in lysine content are possible using  
365 Claire and Robigus. Previous GWAS studies using more diverse panels have found more,  
366 larger-effect QTL controlling asparagine and lysine content (Peng et al., 2018; Rapp et al.,  
367 2018), indicating that there may be beneficial alleles in more diverse germplasm.  
368 Consequently, UK soft wheat germplasm will require diversity beyond Claire and Robigus to  
369 make changes to asparagine and lysine content beyond the incremental improvements found  
370 here.

371

### 372 **Trade-offs between amino acid content and other traits**

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

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

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

406 We also wanted to understand whether any QTL controlling amino acid content had  
407 pleiotropic effects on other traits. The asparagine QTL we identified on chromosome 4B  
408 appeared to overlap with QTL for plant height in the first year, suggesting that there might be  
409 an impact of the *Rht-B1b* allele on asparagine. The *Rht* genes are dwarfing genes used during  
410 the green revolution that have many impacts on crop traits beyond height (Casebow et al.,  
411 2016) and Claire and Robigus both possess different *Rht* genes on 4B and 4D  
412 ([www.cerealsdb.uk.net/cerealgenomics/CerealsDB/Excel/MAS\\_data\\_May\\_2013.xls](http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/Excel/MAS_data_May_2013.xls)).

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

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

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

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

441

#### 442 **Lysine candidate genes**

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

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

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

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

477

478

### Conclusions

479 The nutritional quality of UK soft wheat can be improved incrementally using  
480 diversity from Claire and Robigus, but greater diversity is required to make larger gains. The  
481 genetic architecture of different amino acids differs considerably, and they are often  
482 controlled by QTL that impact other quality traits as well. Future soft wheat breeding in the  
483 UK should therefore consider use of more genetic diversity and using pleiotropic QTL to the  
484 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.

508 **References**

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

- 529 Casebow, R., Hadley, C., Uppal, R., Addisu, M., Loddo, S., Kowalski, A., Griffiths, S., &  
530 Gooding, M. (2016). Reduced height (Rht) alleles affect wheat grain quality. *PloS one*,  
531 *11*, e0156056
- 532 Chang, W., Cheng, J., Allaire, J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson,  
533 J., Dipert, A., and Borges, B. (2021). shiny:Web Application Framework for R. R  
534 package version 1.7.1. <https://CRAN.R-project.org/package=shiny>
- 535 Csardi, G., Nepusz, T. (2006) The igraph software package for complex network research,  
536 *InterJournal, Complex Systems* 1695. <https://igraph.org>
- 537 Curtis, T.Y., Powers, S.J., Wang, R., & Halford, N.G. (2018). Effects of variety, year of  
538 cultivation and sulphur supply on the accumulation of free asparagine in the grain of  
539 commercial wheat varieties. *Food Chemistry*,  
540 <https://doi.org/10.1016/j.foodchem.2017.06.113>
- 541 DEFRA (1998). Bread and Flour Regulations.
- 542 Department of Health and Social Care UK Government. (2021). *Folic Acid Added to Flour to*  
543 *Prevent Spinal Conditions in Babies*. [https://www.gov.uk/government/news/folic-acid-](https://www.gov.uk/government/news/folic-acid-added-to-flour-to-prevent-spinal-conditions-in-babies)  
544 [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).
- 545 Emebiri, L.C. (2014). Genetic variation and possible SNP markers for breeding wheat with  
546 low-grain asparagine, the major precursor for acrylamide formation in heat-processed  
547 products. *Journal of the Science of Food and Agriculture*, .  
548 <https://doi.org/10.1002/jsfa.6434>
- 549 Endelman, J.B. (2011). Ridge regression and other kernels for genomic selection with R  
550 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 Giovanni 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](https://excellenceinbreeding.org/sites/default/files/manual/EiB-M2_Heritability_18-02-20.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](https://doi.org/10.1186/s12870-021-03058-7)
- 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

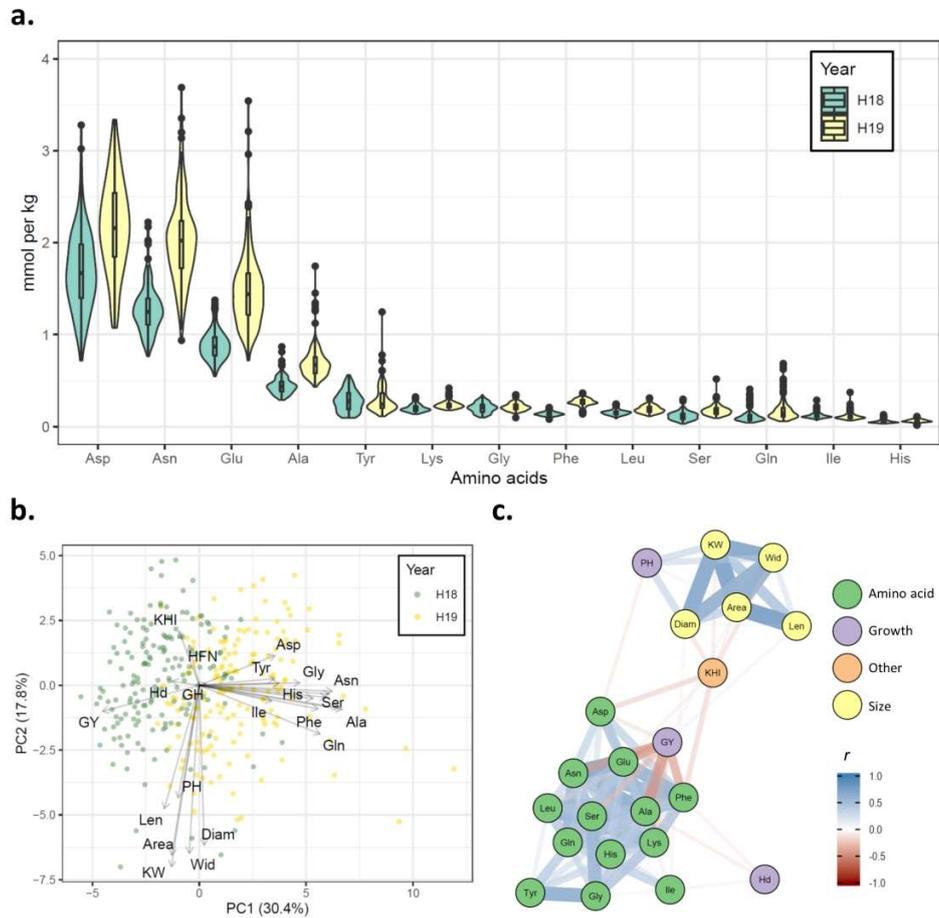


Figure 1. Characterisation of the Robigus x Claire mapping population. a. Measurements of amino acids in the 2017–2018 (H18) and 2018–2019 (H19) harvest years. b. Principal component analysis of all traits in both years along the first two principal components. c. Correlation network analysis of all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

190x187mm (300 x 300 DPI)

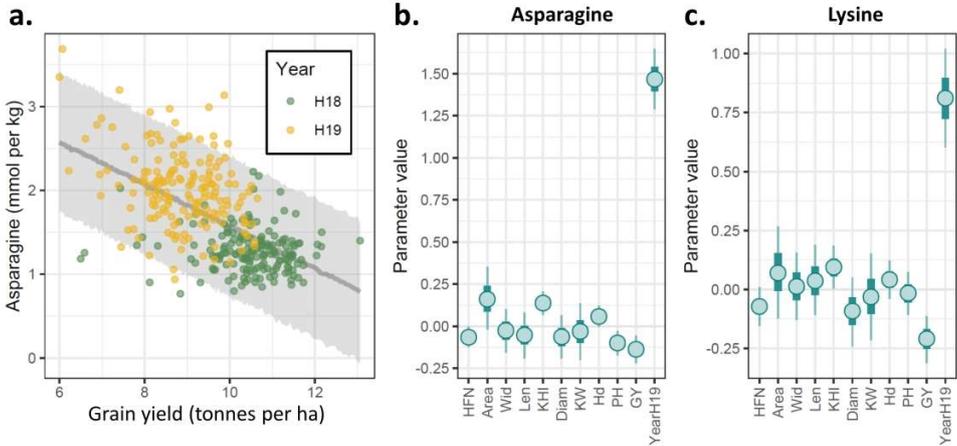


Figure 2. Relationships between free asparagine/lysine and other agronomic measurements. a. Linear modelling of free asparagine content against grain yield. The grey shaded ribbon shows 95% prediction intervals sampled from the posterior distribution, b. and c. Parameter values from multiple linear modelling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

252x119mm (300 x 300 DPI)

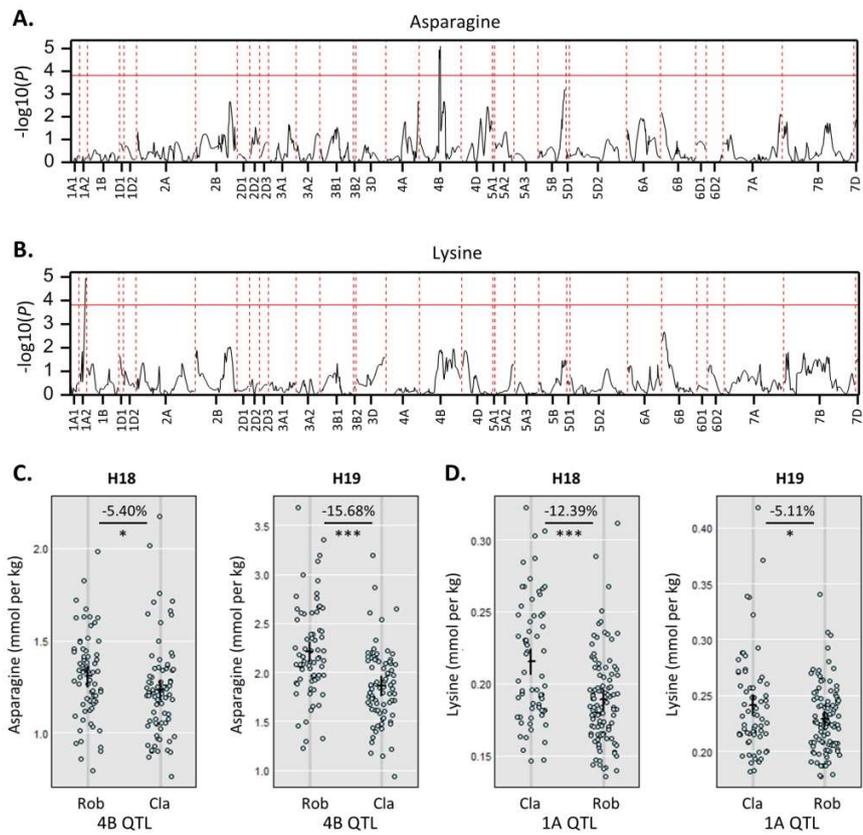


Figure 3. Identification of QTL controlling free asparagine and free lysine. a. Multi – environment genome scan plot for asparagine. b. Multi – environment genome scan plot for lysine. c. Impact of the asparagine QTL on free asparagine concentrations in both field trials. d. Impact of the lysine QTL 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.

190x175mm (300 x 300 DPI)

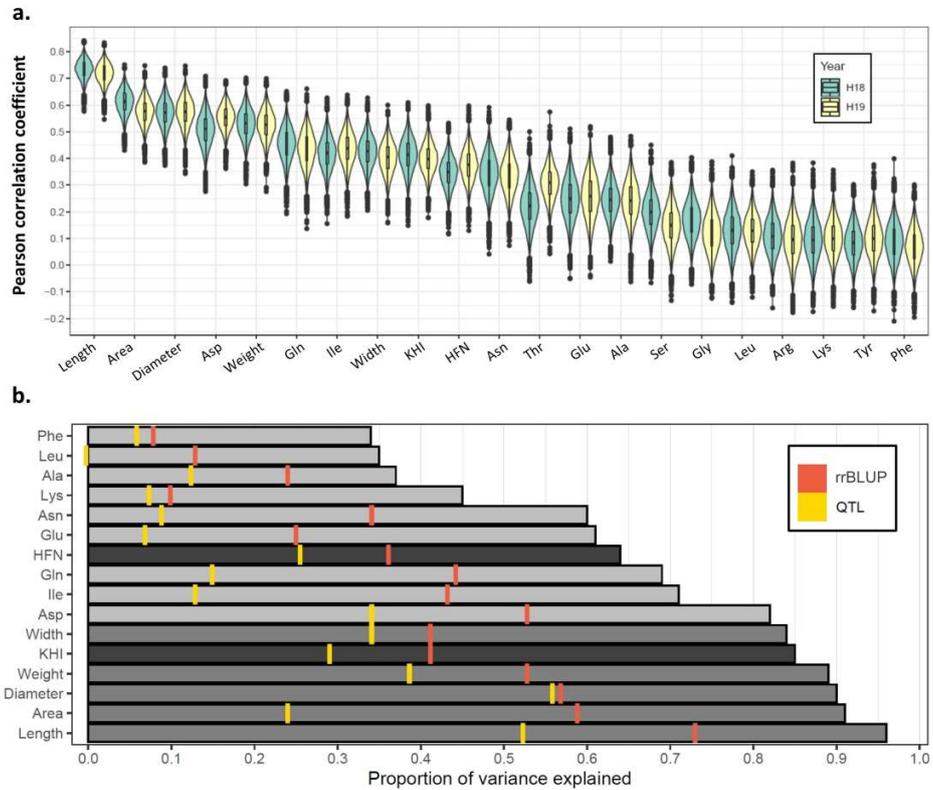


Figure 4. Variation explained by heritability, genomic prediction, and QTL. a. Genomic prediction accuracy between years. b. Additive QTL effects and genomic prediction (rrBLUP) accuracy (yellow and red marks, respectively) plotted alongside broad-sense heritability (shown as bars). Bars are shaded according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

252x209mm (300 x 300 DPI)

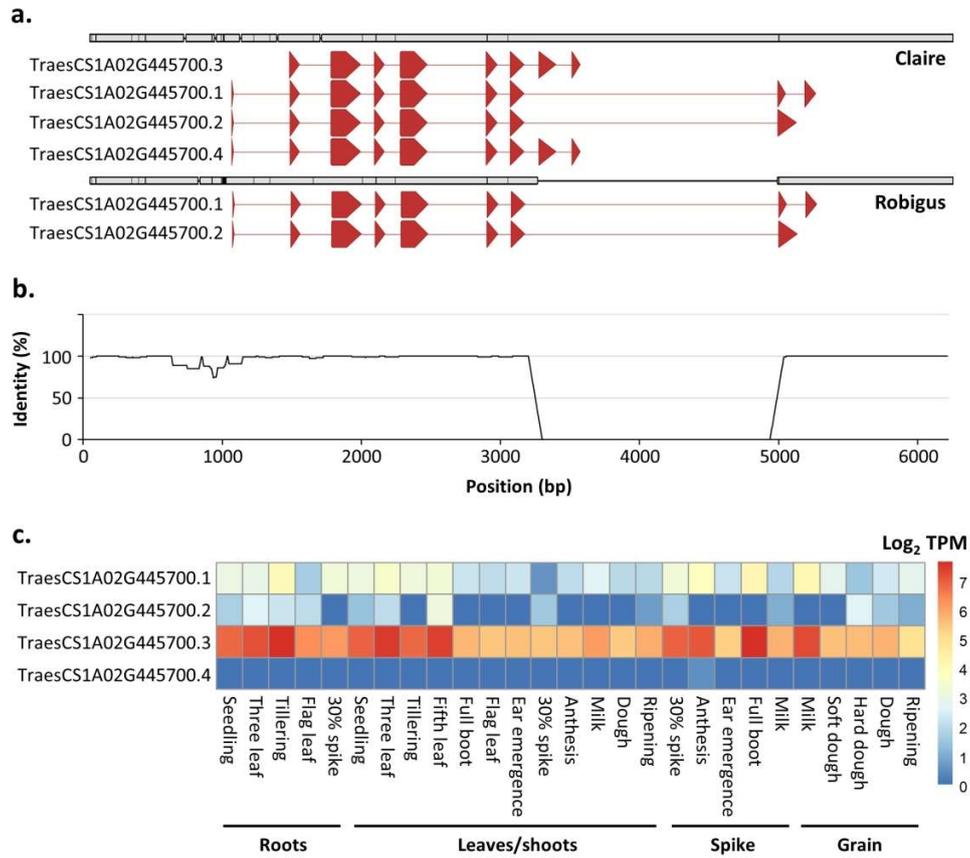


Figure 5. Analysis of the TaHDT-A1 candidate gene for lysine QTL between parents Claire and Robigus. a. Pairwise alignment of the two genes. b. Percentage identity calculated as a sliding window average of 100 bp. c. Expression of the four transcripts throughout development in variety Azhurnaya.

190x170mm (300 x 300 DPI)

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

**Joseph Oddy<sup>1</sup>, Monika Chhetry<sup>2</sup>, Rajani Awal<sup>2</sup>, John Addy<sup>1</sup>, Mark Wilkinson<sup>1</sup>, Dan Smith<sup>1</sup>, Robert King<sup>1</sup>, Chris Hall<sup>1</sup>, Rebecca Testa<sup>1</sup>, Eve Murray<sup>1</sup>, Sarah Raffan<sup>1</sup>, Tanya Y. Curtis<sup>3</sup>, Luzie Wingen<sup>2</sup>, Simon Griffiths<sup>2</sup>, Simon Berry<sup>4</sup>, J. Stephen Elmore<sup>5</sup>, Nicholas Cryer<sup>6</sup>, Isabel Moreira de Almeida<sup>7</sup>, and Nigel G. Halford<sup>1\*</sup>**

<sup>1</sup>*Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*

<sup>2</sup>*John Innes Centre, Norwich Research Park, Colney Lane, Norwich, UK*

<sup>3</sup>*Curtis Analytics Limited, Discovery Park, Sandwich, UK*

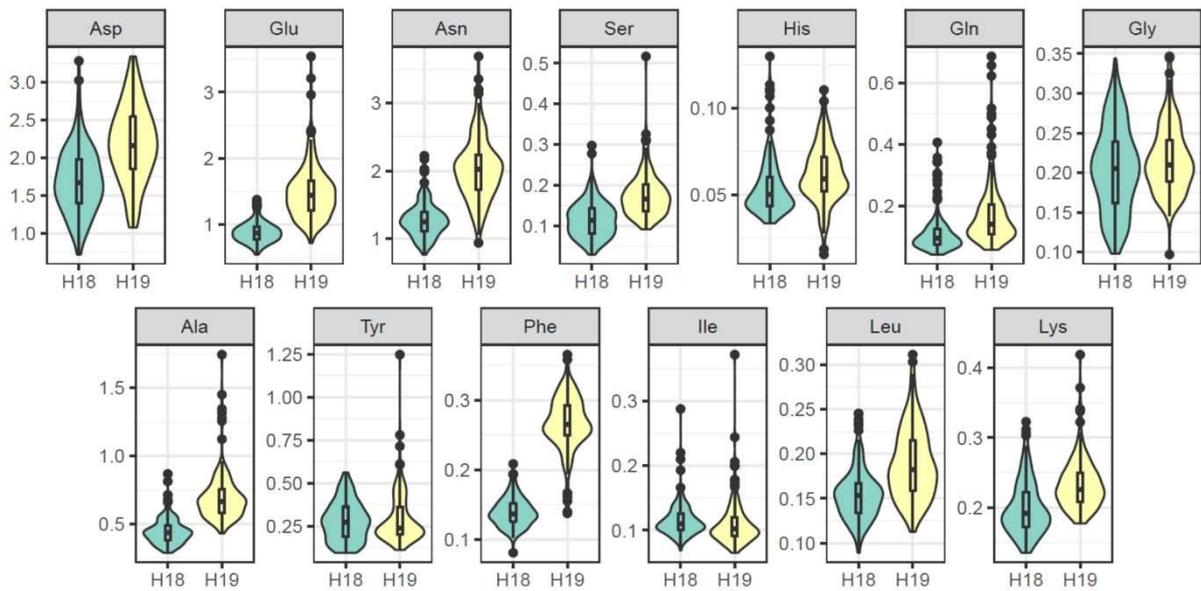
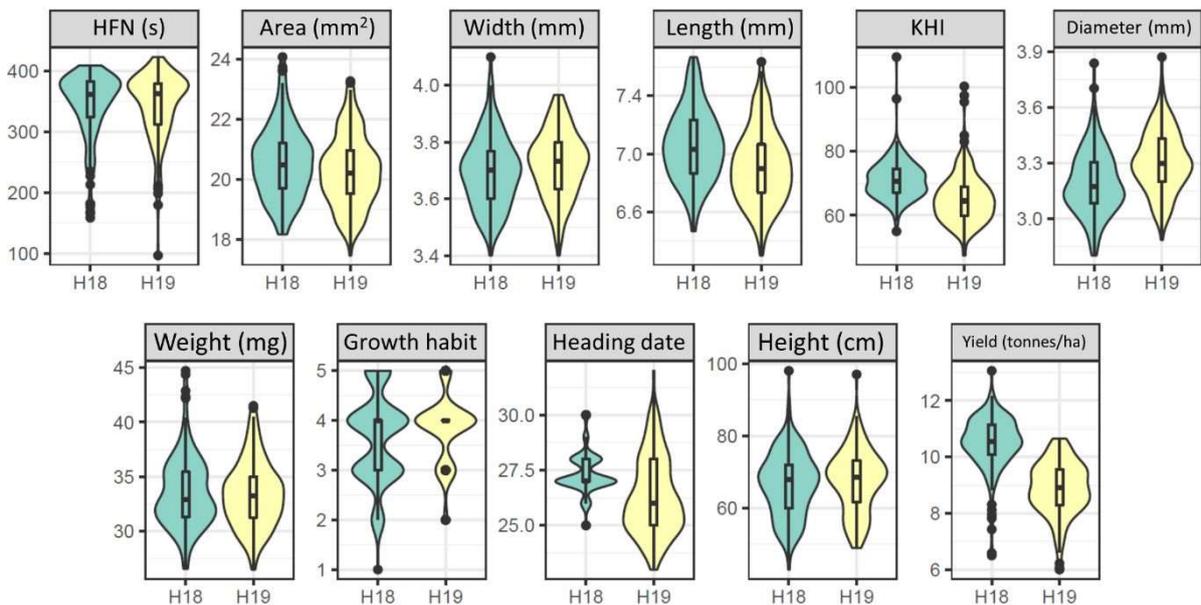
<sup>4</sup>*Limagrain UK Ltd, Market Rasen LN7 6DT, UK*

<sup>5</sup>*Department of Food & Nutritional Sciences, University of Reading, RG6 6DZ, UK*

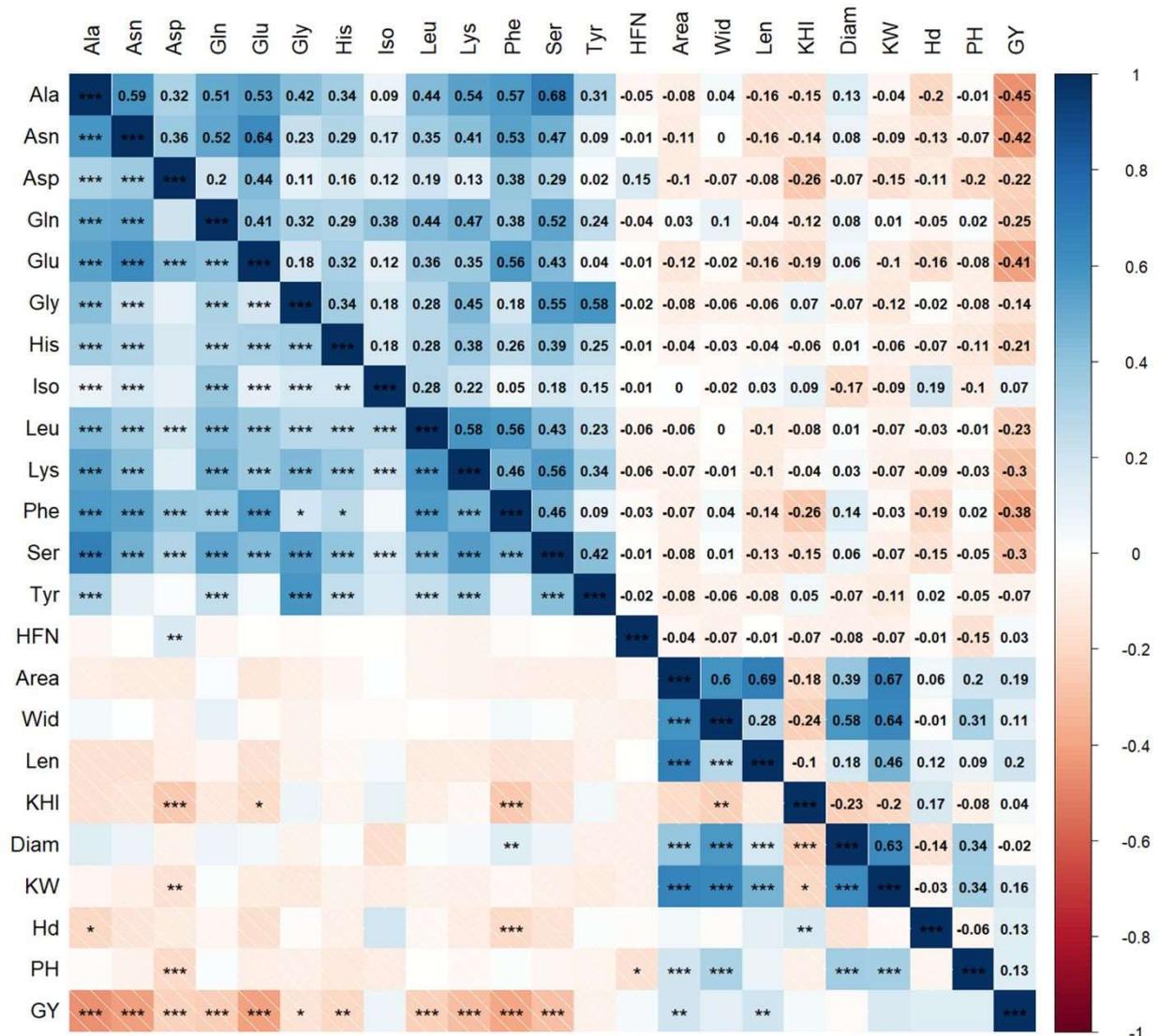
<sup>6</sup>*Mondelēz UK R&D Ltd, Bournville Lane, Bournville, Birmingham, B30 2LU, UK*

<sup>7</sup>*Mondelēz R&D International, 6 Rue René Razel, 91400 Saclay, France*

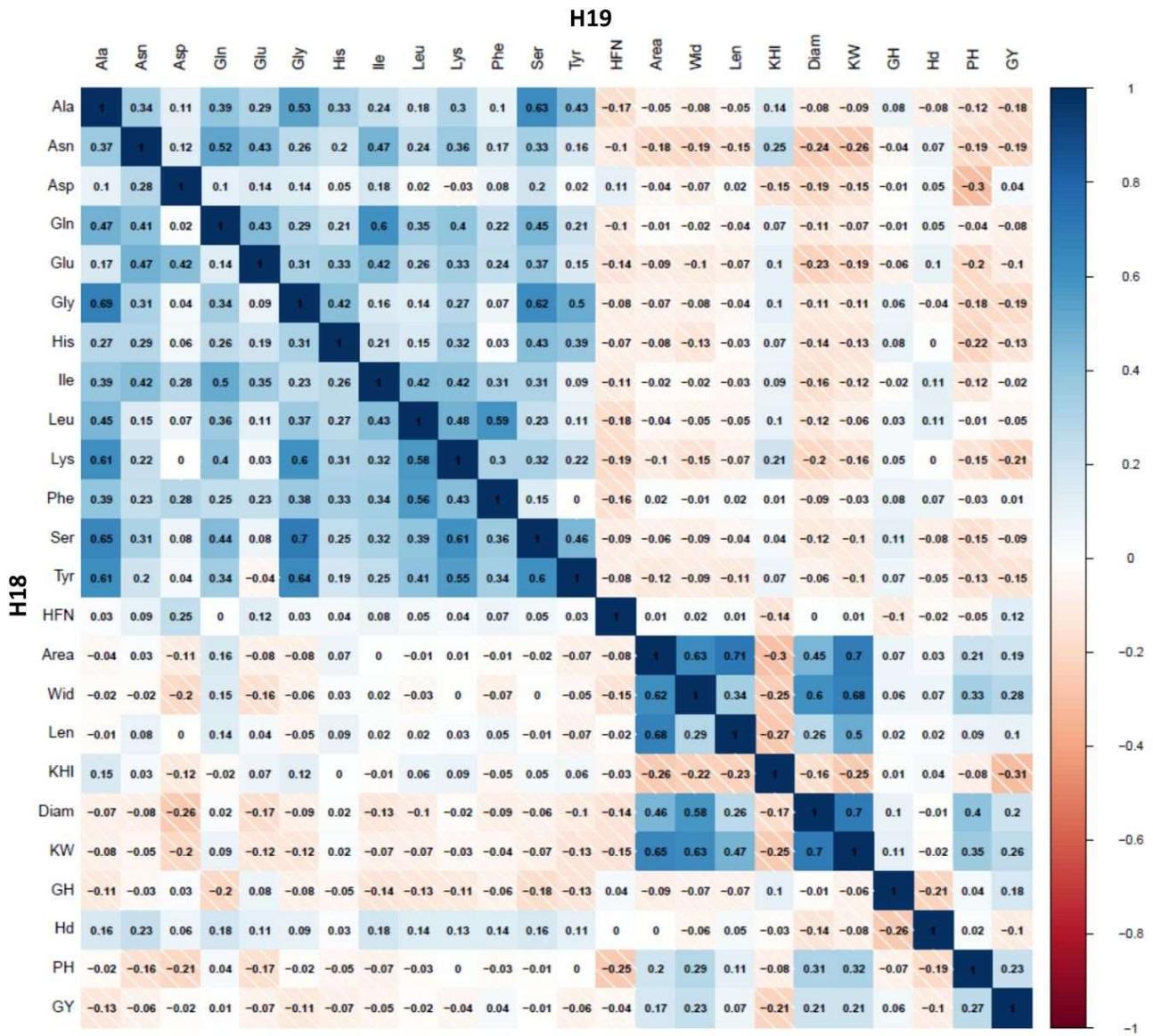
**Supplementary figures**

**a.****b.**

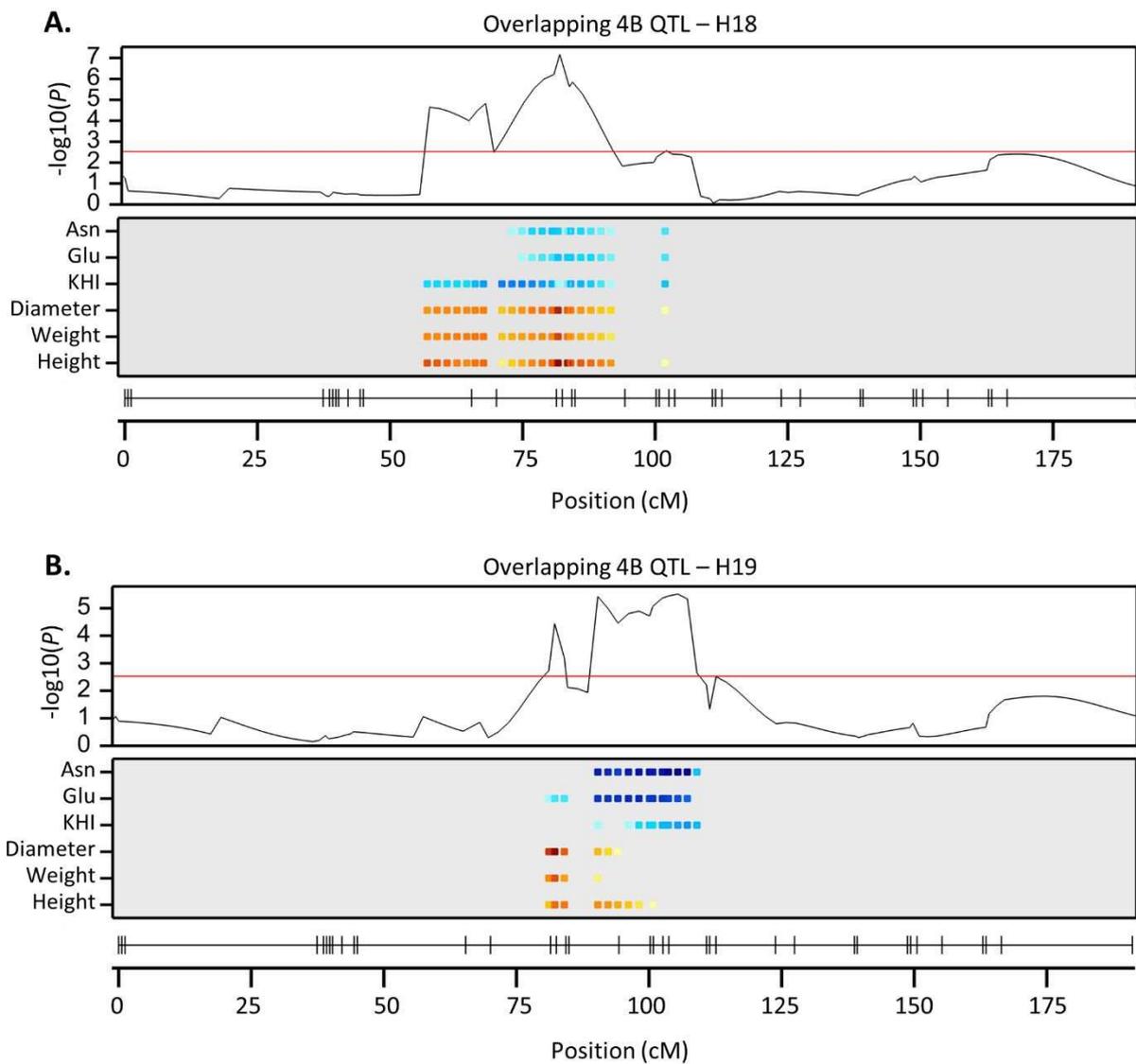
**Supplementary figure 1. Trait measurements from the Claire x Robigus mapping population grown across two years (H18 and H19). a.** Concentration of amino acids (mmol per kg) in wheat grain in both environments. **b.** Measurements of other quality and agronomic traits across both environments.



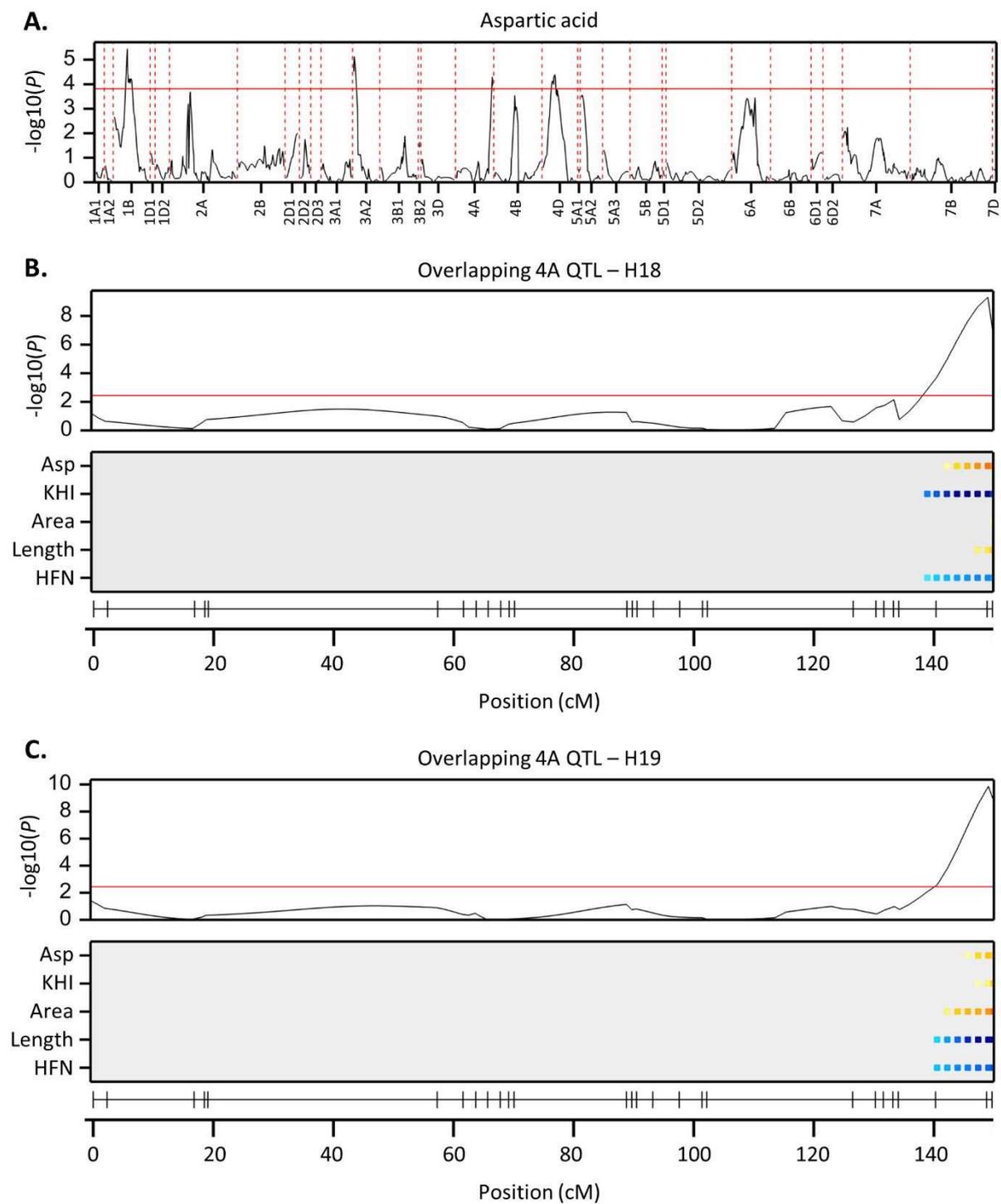
**Supplementary figure 2. Correlation of traits taken from both H18 and H19 environments.** Kendall correlation coefficients shown in upper right triangle and significance asterisks from adjusted p values (Bonferroni correction) shown in lower left triangle. HFN (Hagberg falling number), KHI (kernel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



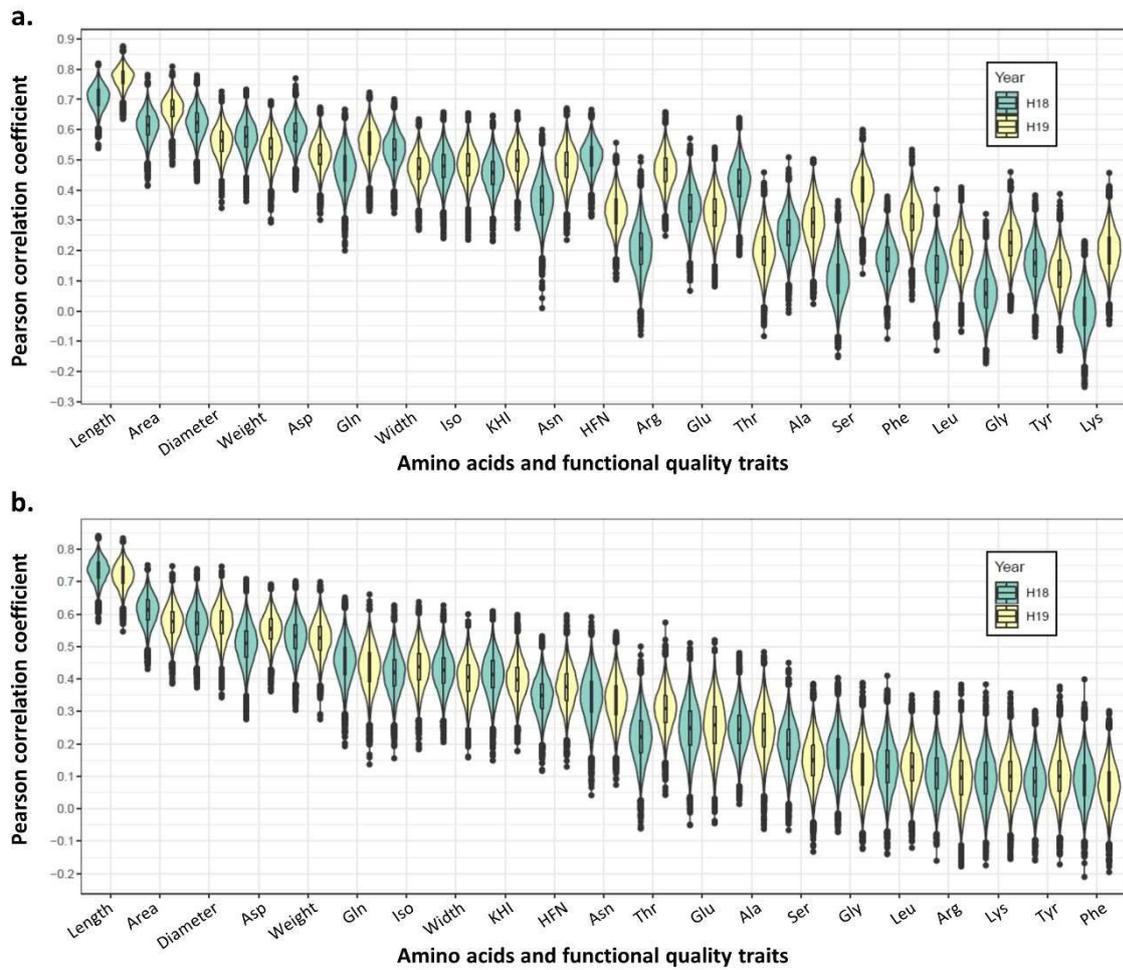
**Supplementary figure 3. Kendall correlation coefficients between traits within each environment. H18 is the bottom left triangle, H19 is the upper right triangle.**



**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 Oddy<sup>1</sup>, Monika Chhetry<sup>2</sup>, Rajani Awal<sup>2</sup>, John Addy<sup>1</sup>, Mark Wilkinson<sup>1</sup>, Dan Smith<sup>1</sup>, Robert King<sup>1</sup>, Chris Hall<sup>1</sup>, Rebecca Testa<sup>1</sup>, Eve Murray<sup>1</sup>, Sarah Raffan<sup>1</sup>, Tanya Y. Curtis<sup>3</sup>, Luzie Wingen<sup>2</sup>, Simon Griffiths<sup>2</sup>, Simon Berry<sup>4</sup>, J. Stephen Elmore<sup>5</sup>, Nicholas Cryer<sup>6</sup>, Isabel Moreira de Almeida<sup>7</sup>, and Nigel G. Halford<sup>1\*</sup>

<sup>1</sup>Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK

<sup>2</sup>John Innes Centre, Norwich Research Park, Colney Lane, Norwich, UK

<sup>3</sup>Curtis Analytics Limited, Discovery Park, Sandwich, UK

<sup>4</sup>Limagrain UK Ltd, Market Rasen LN7 6DT, UK

<sup>5</sup>Department of Food & Nutritional Sciences, University of Reading, RG6 6DZ, UK

<sup>6</sup>Mondelēz UK R&D Ltd, Bournville Lane, Bournville, Birmingham, B30 2LU, UK

<sup>7</sup>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
<b>Amino acids</b>								
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
<b>Functional traits</b>								
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.	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	0.178	0.034	Robigus	<0.001	14.8
Log <sub>e</sub> alanine	WC.0218011	1A2	27.3	593	5.10	yes	0.34	0.065	0.016	Claire	<0.001	10.9	0.022	0.015	Robigus	0.159	1
	WC.0223839	7B	211.2	719	5.03	no	1.00	0.052	0.012	Robigus	<0.001	7.1	0.052	0.012	Robigus	<0.001	5.7
Aspartic acid	WC.0218489	1B	54.4	530	5.40	no	1.00	0.123	0.027	Claire	<0.001	8	0.123	0.027	Claire	<0.001	5.9
	WC.0214359	3A2	2.3	738	7.95	yes	1.68	0.118	0.029	Robigus	<0.001	7.3	0.198	0.033	Robigus	<0.001	15.3
	WC.0221037	4A	148.8	703	8.08	no	1.00	0.154	0.027	Claire	<0.001	12.6	0.154	0.027	Claire	<0.001	9.3
	WC.0227146	4D	48.8	16	3.70	no	1.00	0.102	0.028	Claire	<0.001	5.5	0.102	0.028	Claire	<0.001	4.1
Log <sub>e</sub> glutamate	WC.0221329	4B	100.8	518	4.27	yes	2.41	0.034	0.013	Robigus	0.011	3.7	0.082	0.019	Robigus	<0.001	10.1
Log <sub>e</sub> glutamine	WC.0218486	1B	117.93	660	2.16	yes	9.91	0.011	0.038	Robigus	0.760	0.1	0.109	0.038	Robigus	0.004	4.9
	WC.0221302	4B	103.7	547	3.50	no	1.00	0.105	0.029	Robigus	<0.001	5.4	0.105	0.029	Robigus	<0.001	4.5
	WC.0228471	6B	19.7	25	5.09	no	1.00	0.129	0.029	Claire	<0.001	8.2	0.129	0.029	Claire	<0.001	6.7
Glycine	WC.0218011	1A2	27.3	593	5.37	yes	0.53	0.017	0.004	Claire	<0.001	9.6	0.009	0.003	Robigus	0.005	4.3
	WC.0226796	4B	155.2	327	4.26	no	1.00	0.010	0.002	Robigus	<0.001	3.2	0.010	0.002	Robigus	<0.001	5.3
-1/Isoleucine	WC.0221386	4B	94.3	172	7.31	yes	5.10	0.161	0.120	Robigus	0.180	1	0.821	0.151	Robigus	<0.001	14.1
	WC.0223785	7B	211.2	717	3.60	no	1.00	0.421	0.115	Robigus	<0.001	6.8	0.421	0.115	Robigus	<0.001	3.7
Log <sub>e</sub> lysine	WC.0218011	1A2	27.3	593	4.95	yes	0.38	0.064	0.014	Claire	<0.001	12.1	0.024	0.011	Claire	0.038	2.6
Phenylalanine	WC.0220622	3B1	78.1	116	3.83	yes	1.80	0.005	0.002	Robigus	0.001	6.2	0.009	0.003	Robigus	0.002	5.6
Log <sub>e</sub> serine	WC.0226730	4B	123.8	632	3.39	yes	2.33	0.033	0.031	Claire	0.279	0.7	0.077	0.020	Robigus	<0.001	7.9

Trait	Single trait linkage analysis (H18 and H19)							H18				H19					
	Marker	Chr.	cM	Mbp	$-\log_{10}(p)$	QTLx E	Ratio	Effect	High value	S.E.	P	% Expl.	Effect	High value	S.E.	P	% Expl.
Log <sub>e</sub> KHI	WC.0221037	4A	148.8	703	8.26	yes	1.41	0.032	Robigus	0.006	<0.001	14.8	0.045	Robigus	0.008	<0.001	14.7
	WC.0226741	4B	110.8	594	4.30	yes	2.06	0.017	Robigus	0.006	0.002	4.2	0.035	Robigus	0.008	<0.001	8.6
	WC.0222754	6A	83.22	108	6.09	yes	4.40	0.005	Claire	0.006	0.377	0.4	0.022	Robigus	0.008	0.007	3.4
	WC.0228678	7A	152.17	539	4.45	no	1.00	0.024	Claire	0.006	<0.001	8.1	0.024	Claire	0.006	<0.001	4.1
Area	WC.0217441	2A	70.8	176	7.48	yes	1.43	0.325	Robigus	0.080	<0.001	7.4	0.464	Robigus	0.080	<0.001	15.8
	WC.0220938	4A	149.7	709	2.22	no	1.00	0.208	Claire	0.076	0.006	3	0.208	Claire	0.076	0.006	3.2
	WC.0193228	7A	158.09	610	4.94	no	1.00	0.360	Robigus	0.082	<0.001	9.1	0.360	Robigus	0.082	<0.001	9.5
	WC.0225130	2A	67.2	166	8.15	no	1.00	0.077	Robigus	0.013	<0.001	9.6	0.077	Robigus	0.013	<0.001	9.1
Length	WC.0212864	3A1	92.1	63	4.86	no	1.00	0.058	Robigus	0.013	<0.001	5.5	0.058	Robigus	0.013	<0.001	5.1
	WC.0225952	3B1	128.8	739	7.18	no	1.00	0.072	Claire	0.013	<0.001	8.5	0.072	Claire	0.013	<0.001	8
	WC.0221119	4A	149.7	702	7.12	yes	1.91	0.034	Claire	0.014	0.019	1.8	0.065	Claire	0.014	<0.001	6.5
	WC.0221932	5A1	0.6	1	9.23	no	1.00	0.084	Robigus	0.014	<0.001	11.4	0.084	Robigus	0.014	<0.001	10.7
	WC.0223413	7A	171.6	641	7.23	no	1.00	0.072	Robigus	0.013	<0.001	8.4	0.072	Robigus	0.013	<0.001	7.9
	WC.0222406	7B	40.6	34	7.35	yes	1.54	0.048	Claire	0.014	0.001	3.7	0.074	Claire	0.014	<0.001	8.4
	WC.0217441	2A	72.25	176	4.88	no	1.00	0.031	Robigus	0.007	<0.001	6.1	0.031	Robigus	0.007	<0.001	6.9
	WC.0095497	2B	120.78	545	4.04	yes	3.44	0.009	Claire	0.008	0.257	0.5	0.031	Claire	0.008	<0.001	7.1
Width	WC.0227146	4D	48.8	16	5.98	no	1.00	0.035	Robigus	0.007	<0.001	8	0.035	Robigus	0.007	<0.001	8.9
	WC.0228194	6A	86.1	127	7.35	yes	0.59	0.044	Robigus	0.008	<0.001	12.9	0.026	Robigus	0.008	0.001	4.9
	WC.0228849	7A	174.8	641	3.04	no	1.00	0.024	Robigus	0.007	0.001	3.8	0.024	Robigus	0.007	0.001	4.2
	WC.0212528	7B	107.2	641	4.60	yes	0.11	0.028	Robigus	0.008	<0.001	5.2	0.003	Robigus	0.008	0.732	0.1
	WC.0217441	2A	72.25	176	3.62	no	1.00	0.030	Robigus	0.008	<0.001	3.1	0.030	Robigus	0.008	<0.001	3.2
	WC.0226154	3B1	81.6	237	9.72	no	1.00	0.051	Claire	0.008	<0.001	8.9	0.051	Claire	0.008	<0.001	9.4
	WC.0226868	4B	82.5	32	18.53	no	1.00	0.072	Claire	0.008	<0.001	17.7	0.072	Claire	0.008	<0.001	18.7
	WC.0227146	4D	48.8	16	7.84	no	1.00	0.047	Robigus	0.008	<0.001	7.6	0.047	Robigus	0.008	<0.001	8.1
Weight	WC.0221859	5A2	2.3	30	5.62	no	1.00	0.038	Claire	0.008	<0.001	5	0.038	Claire	0.008	<0.001	5.2
	WC.0212957	6A	99.37	499	8.73	no	1.00	0.050	Robigus	0.008	<0.001	8.4	0.050	Robigus	0.008	<0.001	8.8
	WC.0222406	7B	40.6	34	4.43	no	1.00	0.034	Claire	0.008	<0.001	3.9	0.034	Claire	0.008	<0.001	4.1
	WC.0217441	2A	72.25	176	6.45	no	1.00	0.936	Robigus	0.184	<0.001	8.4	0.936	Robigus	0.184	<0.001	9.7
-1/(500-HFN)	WC.0229571	2B	9.37	12	3.18	no	1.00	0.629	Claire	0.185	0.001	3.8	0.629	Claire	0.185	0.001	4.4
	WC.0226868	4B	82.5	32	9.25	no	1.00	1.127	Claire	0.182	<0.001	12.2	1.127	Claire	0.182	<0.001	14.1
	WC.0212957	6A	99.37	499	8.28	no	1.00	1.085	Robigus	0.186	<0.001	11.3	1.085	Robigus	0.186	<0.001	13.1
	WC.0188904	4A	147.1	733	8.24	no	1.00	0.001	Robigus	0.000	<0.001	11.5	0.001	Robigus	0.000	<0.001	10.3
Heading date	WC.0227149	4D	56.9	17	10.92	yes	0.00	0.001	Robigus	0.000	<0.001	23.5	0.000	Robigus	0.000	<0.001	5.6
	WC.0226616	4A	130.3	632	4.14	no	1.00	0.258	Claire	0.065	<0.001	8.6	0.258	Claire	0.065	<0.001	2.2
Plant height	WC.0221774	5A3	104.9	706	4.51	yes	3.88	0.146	Claire	0.065	0.024	2.8	0.566	Claire	0.124	<0.001	10.5
	WC.0226868	4B	82.5	32	17.98	no	1.00	3.921	Claire	0.444	<0.001	19.8	3.921	Claire	0.444	<0.001	20.7
Grain yield	WC.0213051	4D	56.9	17	28.97	yes	1.16	4.647	Robigus	0.471	<0.001	27.8	5.376	Robigus	0.471	<0.001	38.9
	WC.0227571	7B	16.5	328	5.62	yes	0.66	0.354	Robigus	0.070	<0.001	13.9	0.235	Robigus	0.070	0.001	6.2

SSA 558, Guilford, MA 01831, USA

**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.

<b>Trait</b>	<b>Chr.</b>	<b>Location</b>	<b>Peak (Mbp)</b>	<b>Lower CI (Mbp)</b>	<b>Upper CI (Mbp)</b>	<b>QTL size (bp)</b>	<b>No. of genes</b>
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