

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

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Oddy, J. ORCID: <https://orcid.org/0000-0001-6792-1152>, Chhetry, M., Awal, R., Addy, J., Wilkinson, M., Smith, D., King, R., Hall, C., Testa, R., Murray, E., Raffan, S., Curtis, T. Y., Wingen, L., Griffiths, S., Berry, S., Elmore, J. S. ORCID: <https://orcid.org/0000-0002-2685-1773>, Cryer, N. ORCID: <https://orcid.org/0000-0002-3684-143X>, Moreira de Almeida, I. and Halford, N. G. ORCID: <https://orcid.org/0000-0001-6488-2530> (2023) Genetic control of grain amino acid composition in a UK soft wheat mapping population. *The Plant Genome*. ISSN 1940-3372 doi: <https://doi.org/10.1002/tpg2.20335>  
Available at <https://centaur.reading.ac.uk/111946/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/tpg2.20335>

Publisher: Wiley

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in

the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)




## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

SPECIAL ISSUE: GRAIN QUALITY AND NUTRITIONAL GENOMICS  
FOR BREEDING NEXT GENERATION CROPS

# 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> | Nigel G. Halford<sup>1</sup> 

<sup>1</sup>Rothamsted Research, Harpenden, UK

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

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

<sup>4</sup>Limagrains UK Ltd, Market Rasen, UK

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

<sup>6</sup>Mondelēz UK R&D Ltd, Bournville, UK

<sup>7</sup>Mondelēz R&D International, Saclay, France

## Correspondence

Nigel G. Halford, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK.  
Email: [nigel.halford@rothamsted.ac.uk](mailto:nigel.halford@rothamsted.ac.uk)

Assigned to Associate Editor Manish Pandey.

## Funding information

Biotechnology and Biological Sciences Research Council, Grant/Award Numbers: BB/P016855/1, BB/T017007/1, BB/T50838X/1

## Abstract

Wheat (*Triticum aestivum* L.) 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 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 quantitative trait loci (QTL) controlling free amino acids and other 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 programs.

**Abbreviations:** GS, genomic selection; HFN, Hagberg falling number; KHI, kernel hardness index; PCA, principal components analysis; QTL, quantitative trait locus.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *The Plant Genome* published by Wiley Periodicals LLC on behalf of Crop Science Society of America.

## 1 | INTRODUCTION

The nutritional quality of wheat (*Triticum aestivum* L.) has profound impacts on human health. As one of the largest sources of average daily calorie intake in the world (18.2% in 2019; FAO, 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 (FAO, 2021). Similarly, between 2008 and 2017 in the UK, over 25% of average daily fiber intake was provided by wheat-based foods (Gressier & Frost, 2022). Wheat flour is often fortified to increase its nutrient content. In the UK, wheat is fortified with calcium, iron, thiamine, niacin, and, most recently, folate (DEFRA, 1998; Department of Health and Social Care UK Government, 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 (Department of Health and Social Care UK Government, 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 optimization of its amino acid composition, with the concentrations of lysine and asparagine as most important. Free (soluble, nonprotein) asparagine can be converted into 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 (Hussain 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 optimized 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 (Alarcon-Reverte et al., 2022; Oddy et al., 2021; Raffan et al., 2021) and quantitative trait loci (QTLs) for grain asparagine content have been identified from

### 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.

previous genome-wide association studies (GWAS) (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 programs 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 (*Oryza sativa* L.; Jang et al., 2020; Wang et al., 2008; Yoo, 2017; Zhong et al., 2011) and maize (*Zea mays* L.; 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 (GP) 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 homologue 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 mapping population, therefore, represents a useful resource for identifying additional variation. Claire and Robigus are also represented by scaffold-level genome assemblies in the wheat pangenome, facilitating candidate gene analysis. Furthermore, we investigated other traits, such as grain size, hardness, and Hagberg falling number (HFN), to determine whether QTL controlling nutritional traits overlapped with those controlling other traits of interest.

## 2 | MATERIALS AND METHODS

### 2.1 | Production of doubled haploid lines

Doubled haploid lines (171) of Robigus × 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 fertilized with freshly shed donor maize pollen. After 1 day, wheat florets were treated with Dicamba ( $20\text{mgL}^{-1}$ ) (Sigma-Aldrich, D5417) and injected into the plant stem ( $100\text{ mg L}^{-1}$ ). Developing embryos were excised between 14 and 21 days. Under aseptic conditions, seeds were removed from the spikelets, surface sterilized 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 min. They were then rinsed with sterile distilled water three times.

Haploid embryos were excised and grown on 90 mm Petri dishes in the dark on Gamborg's B5 media with minimal organics (Gamborg et al., 1968), 2% (w/v) sucrose, pH 5.8,  $9\text{ g L}^{-1}$  Difco bactoagar at  $20^\circ\text{C}$ . When showing signs of germination, embryos were transferred to a light incubator at  $20^\circ\text{C}$ . Any nongerminated 1-month-old embryos were given cold shock treatment at  $4^\circ\text{C}$  for 7 days to promote germination. Germinated plantlets were vernalized for 4 weeks and were grown in the glasshouse until the 4-tiller stage. Plants were then given colchicine (C9754; Sigma-Aldrich) treatment for 5–6 h in the light at room temperature, washed and transplanted to soil, acclimatized, and grown in a glasshouse. The mapping population was genotyped by Limagrain using a proprietary single-nucleotide polymorphism (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 ( $52^\circ33'15.1''\text{N}$   $1^\circ01'59.2''\text{E}$ ) in 2017–2018 (abbreviated as H18) and at the Church Farm field site ( $52^\circ38'\text{N}$   $1^\circ10'\text{E}$ ) in 2018–2019 (abbreviated as H19). All 171 Doubled Haploid lines of the mapping population were grown in each trial. Within each trial, one replicate of each line was drilled in  $6\text{ m}^2$  plots in a completely randomized design. The H18 field trial was drilled on September 21, 2017 and harvested on August 1, 2018. The H19 field trial was drilled on the September 14, 2018 and harvested on the August 12, 2019. Growth habit, heading date, plant height, and yield traits were scored in the field.

DoubleTop fertilizer ( $27\text{N } 30\text{SO}_3$ ) was applied at a rate of  $150\text{ kg/ha}$  on the March 20, 2018 for H18 and the February 23, 2019 for H19. In both trials, slug control pellets were applied at a rate of  $7\text{ kg/ha}$  after drilling to control slug pests (Gusto (metaldehyde) pellets on September 27, 2017 for H18 and SluXX (Ferric phosphate) pellets on September 24, 2018 for H19). Herbicide mixtures were applied in autumn (November 21, 2017 for H18 and September 24, 2018 for H19) and spring (May 23, 2018 for H18 and March 19, 2019 for H19) for both trials to control weeds.

## 2.2 | 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.). Grain length (mm), width (mm), and area ( $\text{mm}^2$ ) measurements were recorded in triplicate for each sample using a MARVIN Seed Analyser and software Marvin 4.0 (MARViTECH GmbH). Grain samples were milled to wholemeal flour in a coffee grinder and flour moisture content was recorded using a Minispec nuclear magnetic resonance (NMR) analyzer (Minispec Mq10, Bruker Inc.). Hagberg falling number measurements were recorded using an FN 1000 as the average of two technical replicates (Perten), 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. Briefly, free amino acids were extracted from 0.5 g of wholemeal flour and underwent precolumn derivatization (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.

## 2.3 | 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-Pazarán (2019) using the packages dplyr (Wickham et al., 2022) and lme4 (Bates et al., 2015). Kendall rank correlation coefficients were performed on nontransformed data and adjusted  $p$ -values (Bonferroni correction) were calculated for plotting using R (R Core Team, 2021) and the package corrplot (Wei & Simko, 2021). Principal component analysis was performed on untransformed, scaled variables using the package factoextra (Kassambara & 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 *corr* (Kuhn et al., 2020), *igraph* (Csardi & Nepusz, 2006), and *ggraph* (Pedersen, 2021).

Bayesian modeling was performed on untransformed variables in R using the package *rstanarm* (Goodrich et al., 2020). Variables were scaled before modeling 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 & Mahr, 2022).  $R^2$  estimates were obtained by taking the median of leave-one-out cross validation adjusted estimates.

## 2.4 | Linkage analysis

Single-environment linkage analysis was performed in R using packages *qtl* (Broman et al., 2003) and *qtl2* (Broman et al., 2019). Single-environment linkage analysis was made into an interactive app using the packages *shiny* (Chang et al., 2021), *plyr* (Wickham, 2011), and *rconnect* (Atkins et al., 2021) (accessible at [https://t9onwp-wheatworker.shinyapps.io/QTL\\_Browser/](https://t9onwp-wheatworker.shinyapps.io/QTL_Browser/) and in Supporting Information 1). As before, simple interval mapping (SIM) was performed first to identify covariates for use in composite interval mapping (CIM). Identified QTL from CIM were then used to create single QTL models as well as additive QTL models. Upper and lower 95% confidence intervals for QTL location were calculated using the Bayesian credible interval method in R/*qtl* and expanded to the closest markers. Pseudomarkers were generated every 2 cM in the map and the minimum marker covariate proximity was set at 20 cM. A logarithm of the odds (LOD) score of 3 was used as the significance threshold.

Multi-environment single trait linkage analysis was performed in Genstat for each trait to detect QTL present in both environments, following selection of the most appropriate variance–covariance model according to the Bayesian information criterion. SIM was initially performed to identify putative QTL. These QTLs were then used as covariates in CIM. QTL identified from CIM was then used to construct the final QTL models. Pseudomarkers were generated every 2 cM in the map. The minimum cofactor proximity was set at 30 cM, and the minimum separation for selected QTL was set at 20 cM. Significance thresholds were determined by the Li and Ji's (2005) method with a genome-wide significance level of 0.05.

## 2.5 | Genomic prediction

GP was performed for each trait via fivefold 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 cor-

relation coefficients were calculated for the results from the training and testing datasets to estimate GP 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 Supporting Information 2. Scripts were submitted to the high-performance computing cluster at Rothamsted Research via SLURM for execution.

## 2.6 | Candidate gene analysis

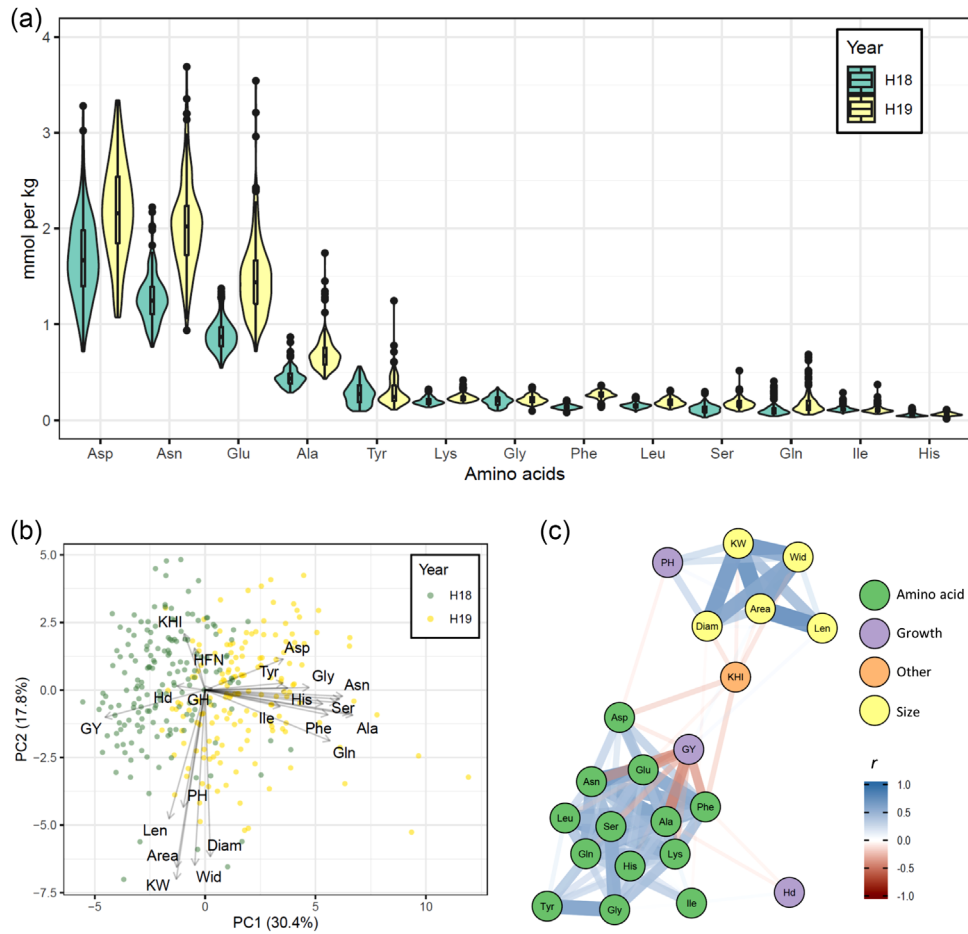
The gene content of the lysine QTL was determined for all wheat pangenome varieties at chromosome scale assembly by identifying the location of the markers in these varieties and extracting genes from Ensembl Biomart (Howe et al., 2021). Genes residing within the region in variety Chinese Spring v1.0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

## 3 | RESULTS

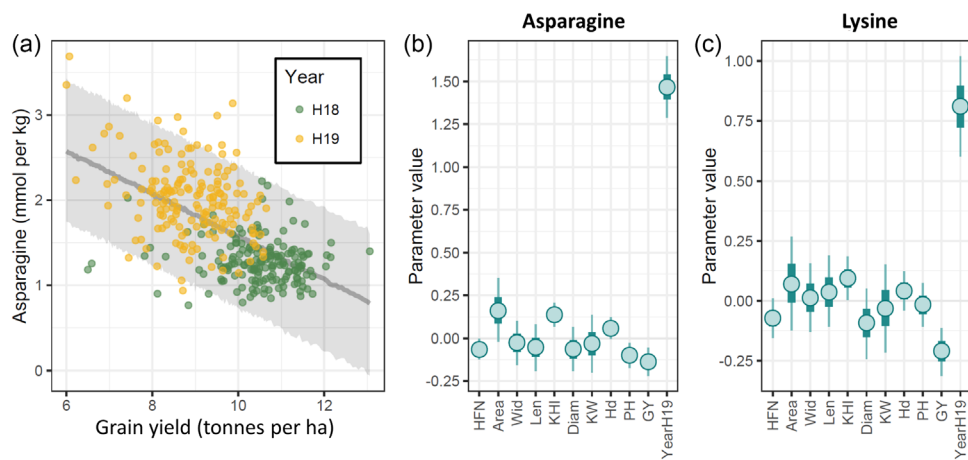
### 3.1 | Phenotypic analysis

We measured free amino acid concentrations and other traits of interest 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 (PCA) 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 traits measured here were uncorrelated with the amino acids (Figure 1b,c; Figures S2 and S3), except for grain yield which showed negative correlations with a subset of amino acids (Figure 1b,c; Figure 2a).

To understand whether any of the traits we measured could predict free asparagine or lysine content in the grain, we



**FIGURE 1** Characterization 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). GH (growth habit), GY (grain yield), Hd (heading date), HFN (Hagberg Falling Number), KHI (kernel hardness index), KW (kernel weight), PH (plant height).



**FIGURE 2** Relationships between free asparagine/lysine and other agronomic measurements. (a) Linear modeling of free asparagine content against grain yield. The gray shaded ribbon shows 95% prediction intervals sampled from the posterior distribution. (b and c) Parameter values from multiple linear modeling of asparagine (b) and lysine (c) as explained by other traits measured in this population. HFN, Hagberg falling number; KHI, kernel hardness index.

**TABLE 1** Multi-environment quantitative trait locus (QTL) for measured amino acids.

Multi-environment single trait linkage analysis (H18 and H19)								
Trait	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
Gln	WC.0227146	4D	48.8	16	3.7	5.5	4.1	Claire
	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

Abbreviations: Chr., chromosome; cM, centimorgan; H18/H19 (%), percentage of variation explained by QTL in respective year; High val., parental line carrying the allele responsible for the higher value; Mbp, megabase pair location in Chinese Spring v1.0.

constructed Bayesian linear models with the traits and harvest year as explanatory variables (Figure 2b,c). 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.

### 3.2 | QTL analysis

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

We identified QTL for grain free asparagine content and lysine content on chromosomes 4B and 1A, respectively (Figure 3a,b; Table 1), which had significant effects across both environments but were also affected by QTL by environment effects (Figure 3c; Figure 3d; Table 1; Table S2). The asparagine QTL on 4B explained 2.6% of the variance in H18, when free asparagine concentrations were lower overall, whereas it explained 14.8% of the variance in H19, when free asparagine concentrations were elevated (Table 1). In both years, the Robigus allele was associated with the higher free asparagine concentrations. In contrast, the lysine QTL on 1A explained 12.1% of the variance in H18, when free lysine was

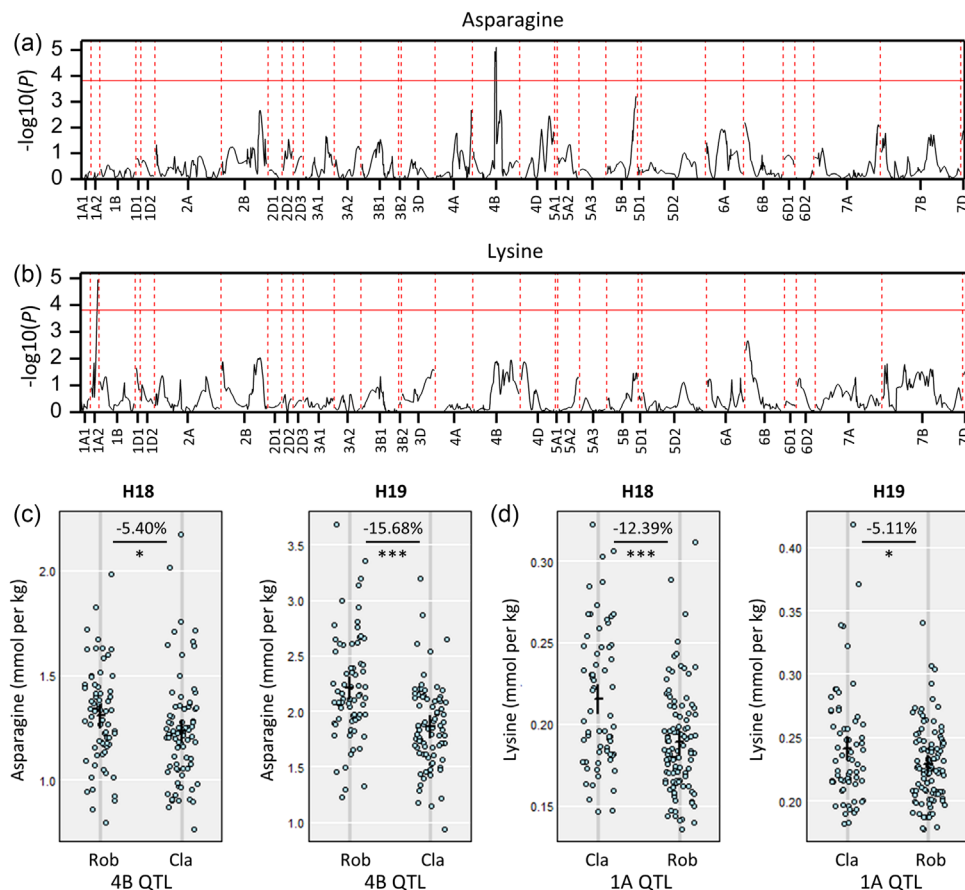
lower overall, and only 2.6% of the variance in H19, when free lysine concentrations were elevated. The Claire allele was associated with higher free lysine concentrations in both years in this case. Multi-environment linkage analysis of amino acid and grain measurements revealed many QTL controlling the other amino acids and traits as well (Table 1; Tables S2 and S3).

The QTL controlling asparagine on chromosome 4B appeared to overlap with QTL for glutamine, glutamic acid, and glycine (Table 1). Each of these QTL had a greater effect in H18 than in H19 and the Robigus allele was associated with the higher value in each case as well, suggesting that these QTL are caused by the same variant. These QTL were also located near to a QTL for KHI and more distantly to QTL for grain diameter, plant height, and grain weight (Table 2, Figure S4), which are likely caused by the *Rht-B1* polymorphism. 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 (Figure S5). 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). Other potential sources of variation underpinning the QTL in this study are presented in Table S4.

### 3.3 | Genomic prediction

Following our modeling of asparagine and lysine using agronomic measurements and QTL models, we calculated the



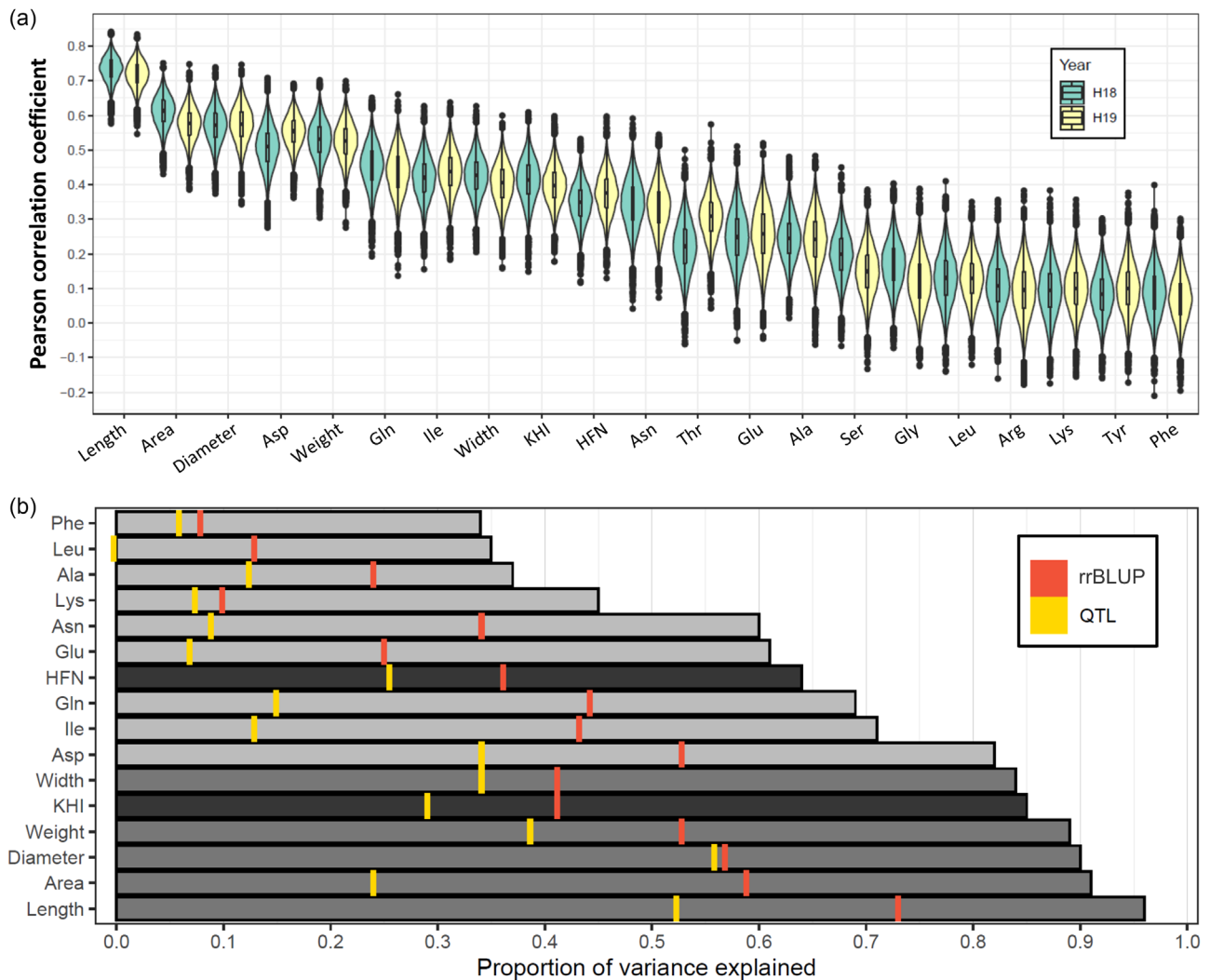


**FIGURE 3** Identification of quantitative trait locus (QTL) controlling free asparagine and free lysine. (a) Multi-environment genome scan plot for asparagine. (b) Multi-environment genome scan plot for lysine. (c) Impact of the asparagine QTL on free asparagine concentrations in both field trials. (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.

**TABLE 2** Multi-environment quantitative trait locus (QTL) impacting both amino acids and other traits on chromosomes 4A, 4B, and 4D.

Multi-environment single trait linkage analysis (H18 and H19)								
Chr.	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
	Diam	WC.0226868	82.5	32	18.53	17.7	18.7	Claire
	Height	WC.0226868	82.5	32	17.98	19.8	20.7	Claire
	Weight	WC.0226868	82.5	32	9.25	12.2	14.1	Claire
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

Abbreviations: Chr., chromosome; cM, centimorgan; H18/H19 (%), percentage of variation explained by QTL in respective year; High val., parental line carrying the allele responsible for the higher value; HFN, Hagberg falling number; KHI, kernel hardness index; Mbp, megabase pair location in Chinese Spring v1.0.

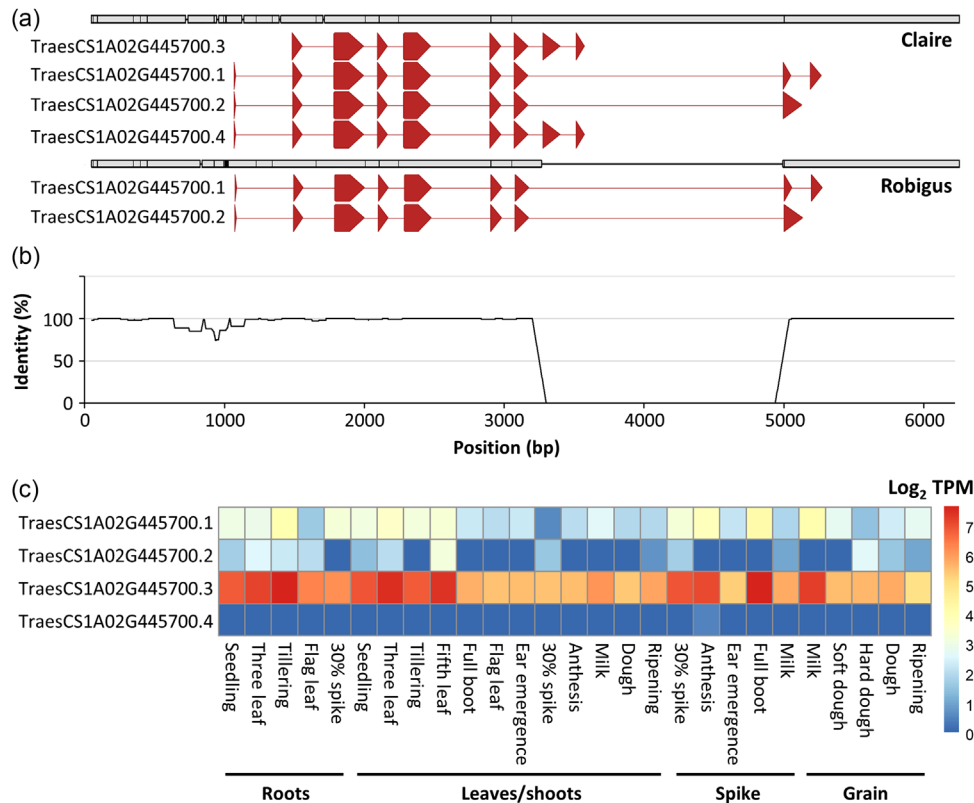


**FIGURE 4** Variation explained by heritability, genomic prediction, and quantitative trait locus (QTL). (a) Genomic prediction accuracy between years. (b) Additive QTL effects and genomic prediction (rrBLUP) accuracy (yellow and red marks, respectively) plotted alongside broad-sense heritability (shown as bars). Bars are shaded according to the trait group that they belong to (amino acid, size, or other). HFN, Hagberg falling number; KHI, kernel hardness index.

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

### 3.4 | Lysine QTL candidate gene analysis

The gene content and QTL size of the lysine QTL on 1A, the HFN/aspartic acid/KHI QTL on 4A, and the asparagine QTL on 4B differed substantially (Table S5). Due to the size of the 4A and 4B QTL, we were unable to plausibly narrow down candidate genes, whereas the lysine QTL on 1A was much smaller so amenable to further analysis. We investigated the gene content of the lysine QTL for all genomes assembled to chromosome scale in the wheat pangenome and gene content varied to a small extent between the different varieties (Table S6). Most notably, the QTL did not match any locations in variety Julius and matched to an unanchored scaffold in Stanley.



**FIGURE 5** Analysis of the *TaHDT-A1* candidate gene for lysine quantitative trait locus (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. TPM, transcript per million.

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

## 4 | DISCUSSION

### 4.1 | 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 characterized by their Claire or Robigus heritage (Shorinola et al., 2022). The varieties within these population groups characterized by Claire and Robigus heritage are also almost entirely soft wheat varieties, further emphasizing 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 genomic selection (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.

## 4.2 | 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 while increasing free asparagine, while other variables (e.g., nitrogen fertilizer) can lead to increases in both yield and free asparagine (see Oddy et al. (2022) for review). Our modeling 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. These environmentally induced increases pose 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 while 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 sulfur 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 minimized 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 (Wilkinson et al., 2020). However, this QTL overlap was not present in the second year of analysis when the asparagine QTL had a greater effect, suggesting that the QTL controlling height and asparagine may be distinct. However, a more detailed analysis is required to comprehensively assess the impact of *Rht-B1* alleles (and dwarfing genes in general) on grain asparagine content. 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), GP 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, Zhang 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.

### 4.3 | 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., 2021; 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 prolamins binding factor (WPBF), a DNA-binding with one finger (DOF)-class transcription factor, increased lysine concentration. Lower prolamins protein content is also associated with increased lysine content in barley (*Hordeum vulgare* L.) (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 prolamins 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 characterized by higher kernel lysine

content (Mertz et al., 1964). Interestingly, the lysine QTL identified in this study is situated upstream of an *Opaque2* orthologue on chromosome 1A: TraesCS1A02G329900, otherwise known as storage protein activator (SPA), which is known to activate storage protein synthesis in wheat (Albani et al., 1997). The A genome homologue of SPA does not differ in sequence between Claire and Robigus, but differential expression of SPA (through differences in *TaHDT-A1* regulation) is a possible mechanism by which this QTL could affect lysine content.

Future work investigating *TaHDT-A1*, 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).

## 5 | CONCLUSIONS

The nutritional quality of the 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 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.

### AUTHOR CONTRIBUTIONS

**Joseph Oddy:** Conceptualization; Investigation; Writing-original draft. **Monika Chhetry:** Investigation; Resources; Writing-review & editing. **Rajani Awal:** Investigation; Resources; Writing-review & editing. **John Addy:** Formal analysis; Investigation. **Mark Wilkinson:** Investigation; Supervision; Writing-review & editing. **Dan Smith:** Formal analysis; Investigation; Writing-review & editing. **Robert King:** Formal analysis; Investigation. **Chris Hall:** Investigation; Supervision. **Rebecca Testa:** Investigation. **Eve Murray:** Investigation. **Sarah Raffan:** Investigation; Supervision; Writing-review & editing. **Tanya Y. Curtis:** Formal analysis; Investigation; Writing-review & editing. **Luzie Wingen:** Formal analysis; Investigation; Supervision; Writing-review & editing. **Simon Griffiths:** Formal analysis; Investigation; Writing-review & editing. **Simon Berry:** Formal analysis; Investigation; Writing-review & editing. **J. Stephen Elmore:** Supervision; Writing-review & editing. **Nicholas Cryer:** Conceptualization; Project administration;

Resources; Supervision; Writing-review & editing. **Isabel Moreira de Almeida:** Conceptualization; Project administration; Resources; Supervision; Writing-review & editing. **Nigel G. Halford:** Conceptualization; Funding acquisition; Project administration; Resources; Supervision; Writing-review & editing.

## ACKNOWLEDGMENTS

Joseph Oddy is supported by a BBSRC Collaborative Training Partnership Studentship (BB/T50838X/1) with partners: University of Reading and Mondelez UK R&D Ltd. Joseph Oddy is also supported by a scholarship from the Society of Chemical Industry. Nigel G. Halford is supported at Rothamsted Research by the BBSRC via the Designing Future Wheat Programme (BB/P016855/1). Sarah Raffan is supported by a Biotechnology and Biological Sciences Research Council (BBSRC) Super Follow-on Fund grant (BB/T017007/1), with partners: University of Bristol, AHDB, KWS UK Ltd, Saaten Union UK Ltd, RAGT Seeds Ltd, Syngenta UK Ltd, and Limagrain UK Ltd. BBSRC is part of UK Research and Innovation. Many thanks to Steve Hanley, Steve Thomas, Andy Phillips, Friedrich Longin, Matthias Rapp, and Keywan Hassani-Pak for their discussion and help in this study.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Data generated in this study are available from Dryad (<https://doi.org/10.5061/dryad.b8gtht7hj>).

## ORCID

Joseph Oddy  <https://orcid.org/0000-0001-6792-1152>

Nicholas Cryer  <https://orcid.org/0000-0002-3684-143X>

Nigel G. Halford  <https://orcid.org/0000-0001-6488-2530>

## 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*, *62*, 1484–1496. <https://doi.org/10.1002/csc2.20760>
- Albani, D., Hammond-Kosack, M. C., 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. *The Plant Cell*, *9*, 171–184. <https://doi.org/10.1105/tpc.9.2.171>
- Atkins, A., McPherson, J., & Allaire, J. J. (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., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/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. <https://doi.org/10.1104/pp.15.01667>
- Broman, K. W., Gatti, D. M., Simecek, P., Furlotte, N. A., Prins, P., Sen, S., 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. <https://doi.org/10.1534/genetics.118.301595>
- Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*, 889–890. <https://doi.org/10.1093/bioinformatics/btg112>
- 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. <https://doi.org/10.1371/journal.pone.0156056>
- Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2021). *shiny: Web application framework for R*. R package version 1.7.1. <https://CRAN.R-project.org/package=shiny>
- Covarrubias-Pazarán, G. (2019). *Heritability: Meaning and computation*. [https://excellenceinbreeding.org/sites/default/files/manual/EiB-M2\\_Heritability\\_18-02-20.pdf](https://excellenceinbreeding.org/sites/default/files/manual/EiB-M2_Heritability_18-02-20.pdf)
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*, 1–9. <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*, *239*, 304–313. <https://doi.org/10.1016/j.foodchem.2017.06.113>
- DEFRA. (1998). *The bread and flour regulations 1998*. <https://www.legislation.gov.uk/uksi/1998/141/contents/made>
- 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>
- 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*, *94*, 1422–1429. <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*, 250–255. <https://doi.org/10.3835/plantgenome2011.08.0024>
- FAO. (2021). *FAOSTAT: Food balances (2014-)*. <http://www.fao.org/faostat/en/>
- Fradgley, N. S., Gardner, K., Kerton, M., Swarbreck, S. M., & Bentley, A. R. (2022). Trade-offs in the genetic control of functional and nutritional quality traits in UK winter wheat. *Heredity*, *128*, 420–433. <https://doi.org/10.1038/s41437-022-00503-7>
- Gabry, J., & Mahr, T. (2022). *bayesplot: Plotting for Bayesian models*. R package version 1.9.0. <https://mc-stan.org/bayesplot/>
- Galili, G., & Amir, R. (2013). Fortifying plants with the essential amino acids lysine and methionine to improve nutritional quality. *Plant Biotechnology Journal*, *11*, 211–222. <https://doi.org/10.1111/pbi.12025>
- Gamborg, O. L., Miller, R. A., & Ojima, K. (1968). Nutrient requirements of suspension cultures of soybean root cells. *Experimental Cell Research*, *50*, 151–158. [https://doi.org/10.1016/0014-4827\(68\)90403-5](https://doi.org/10.1016/0014-4827(68)90403-5)

- Gavazzi, F., Lazzari, B., Ciceri, P., Gianazza, E., & Viotti, A. (2007). Wild-type opaque2 and defective opaque2 polypeptides form complexes in maize endosperm cells and bind the opaque2-zein target site. *Plant Physiology*, *145*, 933–945. <https://doi.org/10.1104/pp.107.103606>
- Ghosh, S., Pellett, P. L., Aw-Hassan, A., Mouneime, Y., Smriga, M., & Scrimshaw, N. S. (2008). Impact of lysine-fortified wheat flour on morbidity and immunologic variables among members of rural families in northwest Syria. *Food and Nutrition Bulletin*, *29*(3), 163–171. <https://doi.org/10.1177/156482650802900302>
- Gil-Humanes, J., Pistón, F., Altamirano-Fortoul, R., Real, A., Comino, I., Sousa, C., Rosell, C. M., & Barro, F. (2014). Reduced-gliadin wheat bread: An alternative to the gluten-free diet for consumers suffering gluten-related pathologies. *PLoS One*, *9*, e90898. <https://doi.org/10.1371/journal.pone.0090898>
- Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2020). *rstanarm: Bayesian applied regression modeling via Stan*. R package version 2.21.1. <https://mc-stan.org/rstanarm>
- Gressier, M., & Frost, G. (2022). Minor changes in fibre intake in the UK population between 2008/2009 and 2016/2017. *European Journal of Clinical Nutrition*, *76*, 322–327. <https://doi.org/10.1038/s41430-021-00933-2>
- Hassani-Pak, K., Singh, A., Brandizi, M., Hearnshaw, J., Parsons, J. D., Amberkar, S., Phillips, A. L., Doonan, J. H., & Rawlings, C. (2021). KnetMiner: A comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across species. *Plant Biotechnology Journal*, *19*, 1670–1678. <https://doi.org/10.1111/pbi.13583>
- Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., Dodiya, K., El Houdaigui, B., Fatima, R., ... Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*, *49*, D884–D891. <https://doi.org/10.1093/nar/gkaa942>
- Hussain, T., Abbas, S., Khan, M. A., & Scrimshaw, N. S. (2004). Lysine fortification of wheat flour improves selected indices of the nutritional status of predominantly cereal-eating families in Pakistan. *Food and Nutrition Bulletin*, *25*(2), 114–122. <https://doi.org/10.1177/156482650402500202>
- Jang, S. u., Han, J.-H., Lee, Y. K., Shin, N.-H., Kang, Y. J., Kim, C.-K., & Chin, J. H. (2020). Mapping and validation of QTLs for the amino acid and total protein content in brown rice. *Frontiers in Genetics*, *11*, 240. <https://doi.org/10.3389/fgene.2020.00240>
- Jiang, X., Deng, Z., Ru, Z., Wu, P., & Tian, J. (2013). Quantitative trait loci controlling amino acid contents in wheat (*Triticum aestivum* L.). *Australian Journal of Crop Science*, *7*, 820–829.
- Jin, P., Gao, S., He, L., Xu, M., Zhang, T., Zhang, F., Jiang, Y., Liu, T., Yang, J., Yang, J., Dai, L., & Chen, J. (2021). Genome-wide identification and expression analysis of the histone deacetylase gene family in wheat (*Triticum aestivum* L.). *Plants*, *10*, 19. <https://doi.org/10.3390/plants10010019>
- Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and visualize the results of multivariate data analyses*. R package version 1.0.7. <https://CRAN.R-project.org/package=factoextra>
- Knox, R. E., Clarke, J. M., & Depauw, R. M. (2000). Dicamba and growth condition effects on doubled haploid production in durum wheat crossed with maize. *Plant Breeding*, *119*, 289–298. <https://doi.org/10.1046/j.1439-0523.2000.00498.x>
- Kolde, R. (2019). *pheatmap: Pretty heatmaps*. R package version 1.0.12. <https://CRAN.R-project.org/package=pheatmap>
- Kuhn, M., Jackson, S., & Cimentada, J. (2020). *corr: Correlations in R*. R package version 0.4.3. <https://CRAN.R-project.org/package=corr>
- Li, J., & Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix]. *Heredity*, *95*(3), 221–227. <https://doi.org/10.1038/sj.hdy.6800717>
- Li, H., Liu, H., Pei, X., Chen, H., Li, X., Wang, J., & Wang, C. (2022). Comparative genome-wide analysis and expression profiling of histone acetyltransferases and histone deacetylases involved in the response to drought in wheat. *Journal of Plant Growth Regulation*, *41*, 1065–1078. <https://doi.org/10.1007/s00344-021-10364-9>
- Li, L., Zhang, Y., Zhang, Y., Li, M., Xu, D., Tian, X., Song, J., Luo, X., Xie, L., Wang, D., He, Z., Xia, X., Zhang, Y., & Cao, S. (2021). Genome-Wide Linkage Mapping for Preharvest Sprouting Resistance in Wheat Using 15K Single-Nucleotide Polymorphism Arrays. *Frontiers in Plant Science*, *12*, 749206. <https://doi.org/10.3389/fpls.2021.749206>
- Malunga, L. N., Ames, N., Khorshidi, A. S., Thandapilly, S. J., Yan, W., Dyck, A., Waterer, J., Malcolmson, L., Cuthbert, R., Sopiwnyk, E., & Scanlon, M. G. (2021). Association of asparagine concentration in wheat with cultivar, location, fertilizer, and their interaction. *Food Chemistry*, *344*, 128630. <https://doi.org/10.1016/j.foodchem.2020.128630>
- Mertz, E. T., Bates, L. S., & Nelson, O. E. (1964). Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science*, *145*, 279–280. <https://doi.org/10.1126/science.145.3629.279>
- Moehs, C. P., Austill, W. J., Holm, A., Large, T. A. G., Loeffler, D., Mullenberg, J., Schnable, P. S., Skinner, W., Van Bostel, J., Wu, L., & McGuire, C. (2019). Development of decreased-gluten wheat enabled by determination of the genetic basis of lys3a barley. *Plant Physiology*, *179*, 1692–1703. <https://doi.org/10.1104/pp.18.00771>
- Oddy, J., Alarcón-Reverte, R., Wilkinson, M., Ravet, K., Raffan, S., Minter, A., Mead, A., Elmore, J. S., De Almeida, I. M., Cryer, N. C., Halford, N. G., & Pearce, S. (2021). Reduced free asparagine in wheat grain resulting from a natural deletion of *TaASN-B2*: Investigating and exploiting diversity in the asparagine synthetase gene family to improve wheat quality. *BMC Plant Biology*, *21*, Article 302. <https://doi.org/10.1186/s12870-021-03058-7>
- Oddy, J., Raffan, S., Wilkinson, M. D., Elmore, J. S., & Halford, N. G. (2020). Stress, nutrients and genotype: Understanding and managing asparagine accumulation in wheat grain. *CABI Agriculture and Bioscience*, *1*, Article 10. <https://doi.org/10.1186/s43170-020-00010-x>
- Oddy, J., Raffan, S., Wilkinson, M. D., Elmore, J. S., & Halford, N. G. (2022). Understanding the relationships between free asparagine in grain and other traits to breed low-asparagine wheat. *Plants*, *11*, 669. <https://doi.org/10.3390/plants11050669>
- Pedersen, T. L. (2021). *ggraph: An implementation of grammar of graphics for graphs and networks*. R package version 2.0.5. <https://CRAN.R-project.org/package=ggraph>
- Peng, Y., Liu, H., Chen, J., Shi, T., Zhang, C., Sun, D., He, Z., Hao, Y., & Chen, W. (2018). Genome-wide association studies of free amino acid levels by six multi-locus models in bread wheat. *Frontiers in Plant Science*, *9*, 1196. <https://doi.org/10.3389/fpls.2018.01196>
- Prasanna, B. M., Palacios-Rojas, N., Hossain, F., Muthusamy, V., Menkir, A., Dhliwayo, T., Ndhlela, T., San Vicente, F., Nair, S. K., Vivek, B. S., Zhang, X., Olsen, M., & Fan, X. (2020). Molecular breeding for nutritionally enriched maize: Status and prospects. *Frontiers in Genetics*, *10*, 1392. <https://doi.org/10.3389/fgene.2019.01392>

- Przewieslik-Allen, A. M., Wilkinson, P. A., Burr ridge, A. J., Winfield, M. O., Dai, X., Beaumont, M., King, J., Yang, C.-Y., Griffiths, S., Wingen, L. U., Horsnell, R., Bentley, A. R., Shewry, P., Barker, G. L. A., & Edwards, K. J. (2021). The role of gene flow and chromosomal instability in shaping the bread wheat genome. *Nature Plants*, 7, 172–183. <https://doi.org/10.1038/s41477-020-00845-2>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Raffan, S., Sparks, C., Huttly, A., Hyde, L., Martignago, D., Mead, A., Hanley, S. J., Wilkinson, P. A., Barker, G., Edwards, K. J., Curtis, T. Y., Usher, S., Kosik, O., & Halford, N. G. (2021). Wheat with greatly reduced accumulation of free asparagine in the grain, produced by CRISPR/Cas9 editing of asparagine synthetase gene *TaASN2*. *Plant Biotechnology Journal*, 19, 1602–1613. <https://doi.org/10.1111/pbi.13573>
- Rapp, M., Schwadorf, K., Leiser, W. L., Würschum, T., & Longin, C. F. H. (2018). Assessing the variation and genetic architecture of asparagine content in wheat: What can plant breeding contribute to a reduction in the acrylamide precursor? *Theoretical and Applied Genetics*, 131, 2427–2437. <https://doi.org/10.1007/s00122-018-3163-x>
- Rustgi, S., Shewry, P., Brouns, F., Deleu, L. J., & Delcour, J. A. (2019). Wheat seed proteins: Factors influencing their content, composition, and technological properties, and strategies to reduce adverse reactions. *Comprehensive Reviews in Food Science and Food Safety*, 18, 1751–1769. <https://doi.org/10.1111/1541-4337.12493>
- Shewry, P. R., Lovegrove, A., Wingen, L. U., & Griffiths, S. (2022). Opinion exploiting genomics to improve the benefits of wheat: Prospects and limitations. *Journal of Cereal Science*, 105, 10344. <https://doi.org/10.1016/j.jcs.2022.103444>
- Shorinola, O., Balcárková, B., Hyles, J., Tibbits, J. F. G., Hayden, M. J., Holuřová, K., Valárik, M., Distelfeld, A., Torada, A., Barrero, J. M., & Uauy, C. (2017). Haplotype analysis of the pre-harvest sprouting resistance locus *Phs-A1* reveals a causal role of *TaMKK3-A* in global germplasm. *Frontiers in Plant Science*, 8, 1555. <https://doi.org/10.3389/fpls.2017.01555>
- Shorinola, O., Simmonds, J., Wingen, L. U., & Uauy, C. (2022). Trend, population structure, and trait mapping from 15 years of national varietal trials of UK winter wheat. *G3 Genes, Genomes, Genetics*, 12, jkab415. <https://doi.org/10.1093/g3journal/jkab415>
- VSN International. (2021). *Genstat for Windows* (22nd ed.). VSN International.
- Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., Ramirez-Gonzalez, R. H., Kolodziej, M. C., Delorean, E., Thambugala, D., Klymiuk, V., Byrns, B., Gundlach, H., Bandi, V., Siri, J. N., Nilsen, K., Aquino, C., Himmelbach, A., Copetti, D., ... Pozniak, C. J. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature*, 588, 277–283. <https://doi.org/10.1038/s41586-020-2961-x>
- Wang, L., Zhong, M., Li, X., Yuan, D., Xu, Y., Liu, H., He, Y., Luo, L., & Zhang, Q. (2008). The QTL controlling amino acid content in grains of rice (*Oryza sativa*) are co-localized with the regions involved in the amino acid metabolism pathway. *Molecular Breeding*, 21, 127–137. <https://doi.org/10.1007/s11032-007-9141-7>
- Wei, T., & Simko, V. (2021). *R package 'corrplot': Visualization of a correlation matrix (version 0.92)*. <https://github.com/taiyun/corrplot>
- Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40, 1–29. <https://doi.org/10.18637/jss.v040.i01>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., François, R., Henry, L., & Müller, K. (2022). *dplyr: A grammar of data manipulation*. R package version 1.0.8. <https://CRAN.R-project.org/package=dplyr>
- Wilke, C. O. (2020). *cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*. R package version 1.1.1. <https://CRAN.R-project.org/package=cowplot>
- Wilkinson, P. A., Allen, A. M., Tyrrell, S., Wingen, L. U., Bian, X., Winfield, M. O., Burr ridge, A., Shaw, D. S., Zaucha, J., Griffiths, S., Davey, R. P., Edwards, K. J., & Barker, G. L. A. (2020). CerealsDB—New tools for the analysis of the wheat genome: Update 2020. *Database*, 2020, baaa060. <https://doi.org/10.1093/database/baaa060>
- Xie, Y. i., Malunga, L. N., Ames, N. P., Waterer, J., Khorshidi, A. S., & Scanlon, M. G. (2021). Effects of growing environment, genotype, and commercial fertilization levels on free asparagine concentration in Western Canadian wheat. *Cereal Chemistry*, 98, 89–99. <https://doi.org/10.1002/cche.10364>
- Yoo, S.-C. (2017). Quantitative trait loci controlling the amino acid content in rice (*Oryza sativa* L.). *Journal of Plant Biotechnology*, 44, 349–355. <https://doi.org/10.5010/JPB.2017.44.4.349>
- Zhao, W., Zhai, F., Zhang, D., An, Y., Liu, Y., He, Y., Ge, K., & Scrimshaw, N. S. (2004). Lysine-fortified wheat flour improves the nutritional and immunological status of wheat-eating families in northern China. *Food and Nutrition Bulletin*, 25(2), 123–129. <https://doi.org/10.1177/156482650402500203>
- Zhong, M., Wang, L.-Q., Yuan, D.-J., Luo, L.-J., Xu, C.-G., & He, Y.-Q. (2011). Identification of QTL affecting protein and amino acid contents in rice. *Rice Science*, 18, 187–195. [https://doi.org/10.1016/S1672-6308\(11\)60026-7](https://doi.org/10.1016/S1672-6308(11)60026-7)

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Oddy, J., Chhetry, M., Awal, R., Addy, J., Wilkinson, M., Smith, D., King, R., Hall, C., Testa, R., Murray, E., Raffan, S., Curtis, T. Y., Wingen, L., Griffiths, S., Berry, S., Elmore, J. S., Cryer, N., Moreira de Almeida, I., & Halford, N. G. (2023). Genetic control of grain amino acid composition in a UK soft wheat mapping population. *The Plant Genome*, e20335. <https://doi.org/10.1002/tpg2.20335>