The evolutionary patterns of barley pericentromeric chromosome regions, as shaped by linkage disequilibrium and domestication

Yun-Yu Chen1,2, Miriam Schreiber1,3, Micha M. Bayer1, Ian K. Dawson4,1, Peter E. Hedley1, Li Lei5, Alina Akhunova5,6, Chaochih Liu6, Kevin P. Smith5, Justin C. Fay7, Gary J. Muehlbauer5, Brian J. Steffenson8, Peter L. Morrell5, Robbie Waugh1,3 and Joanne R. Russell1,*

1The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK, 2Fios Genomics, BioQuarter, 13 Little France Rd, Edinburgh EH16 4UX, UK, 3Division of Plant Sciences, School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK, 4Scotland’s Rural College, Kings Buildings, West Mains Rd, Edinburgh EH9 3JG, UK, 5Department of Agronomy & Plant Genetics, University of Minnesota, 411 Borlaug Hall, 1991 Buford Circle, St Paul, MN 55108, USA, 6Department of Plant Pathology, Kansas State University, Throckmorton Hall, Manhattan, KS 66506, USA, 7Department of Biology, University of Rochester, 319 Hutchison, Rochester, NY 14627, USA, and 8Department of Plant Pathology, University of Minnesota, 495 Borlaug Hall, 1991 Buford Circle, St Paul, MN 55108, USA

Received 14 April 2022; revised 30 June 2022; accepted 13 July 2022; published online 14 July 2022. For correspondence (e-mail joanne.russell@hutton.ac.uk).

SUMMARY

The distribution of recombination events along large cereal chromosomes is uneven and is generally restricted to gene-rich telomeric ends. To understand how the lack of recombination affects diversity in the large pericentromeric regions, we analysed deep exome capture data from a final panel of 815 *Hordeum vulgare* (barley) cultivars, landraces and wild barleys, sampled from across their eco-geographical ranges. We defined and compared variant data across the pericentromeric and non-pericentromeric regions, observing a clear partitioning of diversity both within and between chromosomes and germplasm groups. Dramatically reduced diversity was found in the pericentromeres of both cultivars and landraces when compared with wild barley. We observed a mixture of completely and partially differentiated single-nucleotide polymorphisms (SNPs) between domesticated and wild gene pools, suggesting that domesticated gene pools were derived from multiple wild ancestors. Patterns of genome-wide linkage disequilibrium, haplotype block size and number, and variant frequency within blocks showed clear contrasts among individual chromosomes and between cultivars and wild barleys. Although most cultivar chromosomes shared a single major pericentromeric haplotype, chromosome 7H clearly differentiated the two-row and six-row types associated with different geographical origins. Within the pericentromeric regions we identified 22,387 non-synonymous SNPs, 92 of which were fixed for alternative alleles in cultivar versus wild accessions. Surprisingly, only 29 SNPs found exclusively in the cultivars were predicted to be ‘highly deleterious’. Overall, our data reveal an unconventional pericentromeric genetic landscape among distinct barley gene pools, with different evolutionary processes driving domestication and diversification.

Keywords: evolution, diversity, domestication, *Hordeum vulgare, pericentromeric regions, SNPs.*

INTRODUCTION

Continued improvements in crop productivity are critically founded upon the ability of breeders to identify new genotypes that outperform existing varieties when measured against an evolving set of agricultural challenges (Thomas, 2003). Recombination during meiosis is the process that has traditionally driven this, providing a mechanism by which existing parental alleles are shuffled in progeny into new and better combinations that are selected through phenotypic and genotypic screening. Meiotic recombination is typically unevenly distributed across chromosomes, being frequent in telomeric regions and suppressed in pericentromeric areas, which are characterized by high levels of linkage disequilibrium (LD) (Choulet et al., 2014; Gore et al., 2009; Higgins et al., 2014; Wu et al., 2003). In an extreme cereal crop example, all...
crossovers were observed to occur within the distal 13% of the physical length of chromosome 3B of *Triticum aestivum* (bread wheat) (Choulet et al., 2014). For plant breeding efforts, extended chromosomal regions with minimal recombination reduce the efficacy of selection (Hill & Robertson 1966), making it more difficult to remove deleterious mutations (Felsenstein, 1974), inhibiting the shuffling of alleles into favourable combinations (Baker et al., 2014) and reducing genetic diversity as a result of background selection (Charlesworth et al., 1993). Given the practical constraints that high levels of LD in pericentromeric areas can impose on crop improvement, much research effort has focused on molecularly dissecting the recombination machinery and using the resulting information to try to develop strategies to modify where and how frequently recombination occurs. In contrast, the evolutionary impacts of the lack of recombination have received only limited research attention, and interactions with other genetic processes, such as domestication, crop diversification and adaptation, remain largely unaddressed.

Here, to explore how a lack of regional recombination affects cereal crop genome evolution, we have performed an exhaustive genetic analysis of pericentric and non-pericentric regions in the primarily self-fertilizing crop plant *Hordeum vulgare* ssp. *vulgare* (barley), and its wild progenitor, *Hordeum vulgare* ssp. *spontaneum*. We chose barley as our model because extensive sequence analysis of formally bred homozygous genotypes (i.e. genotypes that are the end product of selection from directed bi- or multi-parental crosses, hereafter referred to as ‘cultivars’) sampled from across the globe has identified vast tracts of the genome with limited genetic diversity (Mascher et al., 2017; Beier et al., 2017; Bustos-Korts et al., 2019; Kono et al., 2019). In addition, parallel sequence analysis of extensive collections of wild barley sampled from its natural habitat in the Fertile Crescent and of landraces from across the eco-geographical expansion range of the crop has been undertaken (Feuillet et al., 2008; Morrell et al., 2014). The assembled knowledge of patterns of genotypic diversity, alongside evidence collected on the founding lineages of the barley crop that suggest a complex history with gene flow and introgression during the expansion of cultivation, provides an informed starting point for our analysis (Morrell & Clegg, 2007; Pankin et al., 2018; Poets et al., 2015; Russell et al., 2016; Saisho & Purugganan, 2007).

Estimates indicate that the low-recombining pericentric portion of barley chromosomes is among the largest of the cereal crops, covering around 48% of the physical genome (International Barley Genome Sequencing Consortium et al., 2012; Baker et al., 2014; Beier et al., 2017). During the evolution of the barley crop these pericentric regions will have, to a large extent, remained ‘locked’, with limited genetic exchange. We argue that these recombinationally inert expanses provide opportunities to explore the early domestication and diversification history of the crop. Of relevance to our analyses, previous studies of mutational load have not identified a greater proportion of deleterious variants in the pericentromeric regions of the barley chromosome, in contrast to other selfing crops such as *Oryza sativa* (rice) and *Glycine max* (soybean). The pericentromeric chromosomal regions of barley may therefore harbour unique features that are particularly worthy of exploration (Kono et al., 2016, 2019; Liu et al., 2017). As defined in the reference genome assembled previously by Mascher et al. (2017), each of the seven chromosomes has been spatially organized into distal (zone 1), interstitial (zone 2) and proximal (zone 3) compartments, based upon the frequencies of repetitive DNA (20 mers) and gene structure.

Here, by analysing genome-wide zonally partitioned variant data derived from exome sequences of a comprehensive panel of cultivar, landrace and wild barleys, we were able to trace the varied evolutionary histories of the pericentromeric regions for all seven barley chromosomes. We found that genetic bottlenecks and limited recombination underlie the unconventional pericentromeric genetic landscape observed in the barley gene pool, with different evolutionary processes in individual chromosomes and sub-chromosomal zones providing new evidence concerning founder events during domestication and diversification. By characterizing these genome-scale evolutionary patterns, our data provide an opportunity to comprehensively assess the extent to which the lack of recombination has been (and continues to be) a constraint on barley breeding, while lending further support to the potential value of exploiting barley genetic resources for future crop improvement.

RESULTS AND DISCUSSION

We assembled and analysed a collection of new and existing whole-exome capture sequence data from an initial panel of 879 accessions of cultivar, landrace and wild barleys sampled from across their eco-geographical ranges, identifying 93 849 112 variants (Figure S1; Table S1) (Bustos-Korts et al., 2019; Hübner et al., 2009; Russell et al., 2016; Steffenson et al., 2007). Following variant filtering and the removal of wrongly assigned accessions (Figures S2 and S3), a final data set was generated that comprised 3 082 873 high-quality single-nucleotide polymorphisms (SNPs), most of which had a minor allele frequency (MAF) of <0.05 \( n = 2 742 309 \), from a stringently curated and comprehensive set of 815 accessions (163 cultivars, 388 landraces and 264 wild barleys) (Table S2). For an initial check of the overall genetic relationship between these accessions, we conducted principal coordinate analysis (PCO) and inferred admixture using a randomly chosen genome-wide set of SNPs (Figure 1). A clear division
between wild and domesticated barleys was observed (Figure 1a), as had been expected from our prior work on smaller barley panels (e.g. Russell et al., 2016), with seven ‘subpopulations’ identified (Figure 1b) with designations corresponding to the groupings observed in the PCO. As expected from this earlier work, cultivar germplasm appeared to be derived from subsets of landraces, and a split was observed between two-rowed and six-rowed accessions.

We then explored different portions of the seven chromosomes of the barley genome. For this purpose, we partitioned each chromosome into three discrete zones using the physical positions reported by Mascher et al. (2017) (Table S4) that were reminiscent of the three compartments applied in an earlier analysis of bread wheat chromosome 3B (Choulet et al., 2014). Zone 1 covers the distal portions of each chromosome, characterized by high gene content and frequent recombination, zone 2 covers the interstitial regions with intermediate gene content and zone 3 approximates the pericentromeric regions, enriched in housekeeping genes with little or no recombination (Keller & Krattinger, 2017). We then generated a range of individual SNP- and chromosome-based diversity-related analyses for our barley germplasm groups (Figure 2). A clear genomic partitioning pattern between the zones (as defined in Figure 2) was observed, with the pericentromeric regions generally showing reduced genetic diversity (Figure 2a). In particular, the pericentromeric regions of domesticated accessions (cultivars and landraces) in our collection showed dramatically reduced diversity on chromosomes 1H, 2H and 4H, where the genetic diversity (π) values ‘flat-lined’ (more distal regions not only have higher diversity but the profiles revealed are ‘noisier’). Examining profiles of per-SNP differentiation (F_{ST}) between pairs of barley groups (Figure 2b–d), we observed distinctive patterns, sometimes including fixed differences, for pericentromeric regions. Intriguingly, F_{ST} values within zone 3 aligned into multiple horizontal ‘tracks’ that comprised long stretches of SNPs with shared F_{ST} values that sometimes extended in both directions into zone 2. The longest track, of approximately 200 Mbp, was located on chromosome 4H. Moreover, multiple ‘break points’ within tracks (creating multiple tracks with different F_{ST} values) were also observed. Zone-3 tracks with high F_{ST} values (0.8–1.0) were most noticeable in the cultivar–wild barley comparison (Figure 2b) for chromosomes 1H, 2H, 4H, 5H and 6H, indicating the close to complete, and sometimes complete, fixation of different allelic states between the two gene pools. Some of these large values may be associated with structural variants, as observed in previous studies in Zea mays (maize) and barley (Fang et al., 2012; Fang et al., 2014; Lei et al., 2019), but this was not explicitly tested here. Consistent with their similar π profiles, tracks of high F_{ST} appeared absent from the cultivar–wild barley comparison of zone-3 areas for chromosomes 3H and 7H. Extending this comparison, in the landrace–wild barley F_{ST} graph (Figure 2c) the horizontal track patterns within zone 3 were maintained, but generally with lower F_{ST}.
values and with no regions with complete differentiation ($F_{ST} = 1$). For the cultivar–landrace comparison (Figure 2d), features of the same pattern were retained, but less obviously and with even lower $F_{ST}$ values.

In the case of the cultivar–wild type comparison, the different $F_{ST}$ tracks are illustrated schematically for explanation purposes in Figure 3(a–d). The simple case of fixed alternate SNP states in cultivars and wild barleys is shown in Figure 3(a), which could represent an example where an early post-domestication allele is driven to fixation over the last 10,000 years of cultivation and expansion. Figure 3(b) represents a common run of shared states between the two barley categories (where the shared state in wild barley may indicate its progenitor status). In most of the pericentromeric regions, however, there are a mixture of completely and partly differentiated SNPs, presumably through the presence of multiple ancestral wild haplotypes, resulting in the ‘overlapping’ horizontal tracks of $F_{ST}$.

---

**Figure 2.** Extensive genetic differentiation in the pericentromeric regions among *Hordeum vulgare* (barley) groups, showing all single-nucleotide polymorphisms (SNPs) without minor allele frequency (MAF) filtering. The top track shows the chromosome diagrams, with the gradient of blue colours representing zone 1 (light blue), zone 2 (medium blue) and zone 3 (dark blue) regions, and the red bars representing the centromere, using the coordinates reported by Maechler et al. (2017) and physical distance. (a) Genetic diversity (c): red, wild barleys; orange, landraces; blue, cultivars. (b) Fixation index ($F_{ST}$) between cultivars and wild barleys. (c) $F_{ST}$ between landraces and wild barleys. (d) $F_{ST}$ between cultivars and landraces. In (b) and (c), sites with $F_{ST} \geq 0.8$ were coloured red (with no such sites in panel d).

**Figure 3.** Diagram of how different wild founder haplotypes give rise to horizontal $F_{ST}$ patterns. (a) In the simplest case, single-nucleotide polymorphisms (SNPs) in cultivars and wild barleys are fixed completely at two different states and a track of $F_{ST} = 1$ is formed. (b) Horizontal track with a lower $F_{ST}$ value is formed when some wild barleys share the fixed cultivated allele. (c) ‘Overlapping’ horizontal tracks of $F_{ST}$ formed when different wild barley alleles have varying degrees of differentiation from the cultivars. (d) ‘Break point’ variable horizontal tracks of $F_{ST}$ formed that represent rare recombination between two wild barley founder haplotypes. (e) Real exome sequence genotype data from a segment of barley chromosome 4H, zone 3, showing at least three wild barley founder haplotypes, separated by white space, in this region: the ancestors of the cultivars and one possible double crossover event between different wild founders (asterisk).
of Figure 3(c). Figure 3(d) shows the situation where a rare recombination event happens between wild barleys, causing a shift of allele frequencies at a chromosomal scale and forming the break points observed, as highlighted for the actual case of barley chromosome 4H in Figure 3(e).

We next analysed genome-wide linkage disequilibrium of cultivar, landrace and wild barley groups. Initial examination of genome-wide average $R^2$ estimates showed that LD decay in the cultivars was around 1.5$\times$ slower overall than in the wild barleys, and about 1.2$\times$ slower than in the landraces (Figure S4). Further examination of LD revealed contrasting haplotype block structures between the different germplasm categories (Table 1). The average block size in cultivars was 158,637 kbp, compared with only 26,284 kbp in wild barleys. Although blocks covered over 90% of chromosomes in cultivars, the value was only 50% for the wild barley group, although the wild barley blocks still contained many more SNP variants (almost double, with an average of 46,597 compared with 28,453). Levels of LD and block structure also varied between chromosomes, with 3H and 7H having markedly smaller block sizes in cultivars (80,843 and 89,407 kbp, respectively) than the average, for example. For all germplasm categories, chromosome 4H had comparatively few blocks and the greatest chromosome block coverage (94%).

We then extended our analysis to explore genes and gene haplotype features by chromosome and chromosome zone (Figure 4; Table S3). The greatest number of haplotypes per gene, accounting for different group sample size, was identified for wild barley (Figure 4a), with the median value of approximately 50 being about five times that of the cultivar group, which had the fewest number of haplotypes per gene. When we compared haplotype richness (randomly selecting 100 accessions for each of the three groups, then calculating the number of haplotypes for these, and repeating this analysis 100 times to generate averages) (Figure 4b), we found that zone 3 always had the lowest values and zone 1 had the highest values, consistent with earlier diversity profiles (Figure 2). Comparing wild and cultivar categories, zone 3 in wild barley had a much higher richness than zone 1 in the cultivar (about double). The frequencies of the major haplotype were higher for cultivars (approx. 60% median value for the major haplotype as a proportion of all haplotypes at each gene) than for landraces and wild barleys (50 and 25%, respectively) (Figure 4c). Corresponding with haplotype richness estimates by chromosome zone (Figure 4b), the dominance of a single haplotype was most prominent in zone 3 of each barley group (Figure 4d). In the cultivars the median frequency value for the major haplotype was over 80% in the zone-3 area. Data on block sizes (Figure 4e) were consistent with the patterns recorded in Table 1. The difference in block sizes between chromosome zones is much larger for cultivars than for wild barley, with

### Table 1: Linkage disequilibrium (LD) haplotype block structure for each group

| Group         | Chr. | Chr. length (bp) | No. blocks | Block coverage (kb) | Chr. block coverage (%) | Largest block (kb) | No. SNPs in blocks |
|---------------|------|------------------|------------|--------------------|-------------------------|--------------------|--------------------|
| Cultivars     | 1H   | 558,535,432      | 932        | 505,269            | 90                      | 161,870            | 22,405             |
| (n = 163)     | 2H   | 768,075,024      | 1418       | 707,970            | 92                      | 184,043            | 33,234             |
|               | 3H   | 699,711,114      | 1161       | 635,706            | 91                      | 80,843             | 31,218             |
|               | 4H   | 647,060,158      | 691        | 610,364            | 94                      | 258,652            | 20,001             |
|               | 5H   | 670,030,160      | 1351       | 615,478            | 92                      | 186,594            | 36,651             |
|               | 6H   | 583,380,513      | 1041       | 542,069            | 93                      | 149,055            | 26,062             |
|               | 7H   | 657,224,000      | 1221       | 597,940            | 91                      | 89,407             | 29,601             |
|               | Average |                | 1116       | 602,114            | 92                      | 156,637            | 28,453             |
| Landraces     | 1H   | 558,535,432      | 1843       | 485,605            | 87                      | 74,708             | 31,909             |
| (n = 388)     | 2H   | 768,075,024      | 2746       | 667,275            | 87                      | 125,158            | 49,418             |
|               | 3H   | 699,711,114      | 2613       | 611,705            | 87                      | 134,606            | 49,199             |
|               | 4H   | 647,060,158      | 1476       | 602,320            | 93                      | 185,970            | 34,126             |
|               | 5H   | 670,030,160      | 2457       | 591,257            | 88                      | 185,621            | 50,045             |
|               | 6H   | 583,380,513      | 2170       | 508,486            | 87                      | 130,284            | 40,095             |
|               | 7H   | 657,224,000      | 2501       | 572,238            | 87                      | 76,166             | 46,584             |
|               | Average |                | 2258       | 576,984            | 88                      | 130,359            | 43,054             |
| Wild barleys  | 1H   | 558,535,432      | 5769       | 275,005            | 49                      | 4476               | 41,438             |
| (n = 264)     | 2H   | 768,075,024      | 6835       | 373,400            | 49                      | 6847               | 52,933             |
|               | 3H   | 699,711,114      | 6686       | 364,791            | 52                      | 81,241             | 49,423             |
|               | 4H   | 647,060,158      | 5153       | 392,599            | 61                      | 10,417             | 45,920             |
|               | 5H   | 670,030,160      | 6684       | 326,365            | 49                      | 55,927             | 49,417             |
|               | 6H   | 583,380,513      | 4588       | 316,772            | 54                      | 17,857             | 35,907             |
|               | 7H   | 657,224,000      | 6932       | 306,958            | 47                      | 72,225             | 51,179             |
|               | Average |                | 6092       | 336,556            | 51                      | 26,284             | 46,597             |
landraces having intermediate differences (Figure 4f). To put these data into a practical context relevant for breeding, the block size observed in the most variable chromosomal region of the cultivars (zone 1) did not significantly differ statistically from that of the least diverse chromosomal region of wild barleys (zone 3) (Table S4).

These pericentromeric haplotype analyses provided indications of how evolutionary histories have varied among barley chromosomes. To evaluate further the factors involved, for each chromosome we studied the selection signals, structure and gene content of zone 3, compared with other zones. First, we used the $\mu$ statistic, which is a composite measure based on site variation, site frequency spectrum and LD profile (Alachiotis & Pavlidis, 2018), to identify potential signals of selective sweeps (Figure 5). For each barley group, we highlighted variants where $\mu$ scores were above our 95th percentile threshold, taken to suggest the presence of a selective sweep (Figure 5a). The calculated $\mu$ thresholds were $4.56 \times 10^{-5}$, $1.93 \times 10^{-5}$ and $1.26 \times 10^{-6}$ for cultivar, landrace and wild barleys, respectively. Analysis revealed the strongest evidence of selective sweeps in domesticated barleys on chromosome 4H (Figure 5a), although there was no significant difference in average $\mu$ scores between chromosomes for any barley group (Figure 5b). For each of the germplasm groups, zone-3 regions cumulatively showed the highest $\mu$ scores and zone-1 regions the lowest (Figure 5c), suggesting that, overall, pericentromeric regions are subjected to greater positive selection. An unusual feature, however, was the high $\mu$ scores found for a non-pericentromeric region of chromosome 6H in wild barleys (Figure 5a,b). Based on $\mu$ values in cultivars, even for zone 1 (lowest average score among zones), the evidence for selective sweeps is many orders of magnitude greater than for zone 3 in wild barleys (highest average score among zones).

We next assessed the structure of pericentromeric regions by exploring intraspecific relationships among samples for zone-3 SNPs in each barley chromosome and comparing the results with zone-1 and -2 SNPs combined. The zone-3-specific profiles showed the clustering of cultivars and landraces into one to three ‘monophyletic’ clades, separated by clusters of wild barley accessions, and contrasting pictures between chromosome zones and chromosomes (Figure 6a,b, examples of chromosomes 4H and 7H; Figure 4.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4}
\caption{Gene haplotype analysis for different barley chromosome zones. Haplotypes of 32,222 genes with variants covered by exome sequencing were characterized. (a) Gene haplotype count by chromosome. (b) Gene haplotype count by chromosome zone. (c) Major haplotype frequency by chromosome. (d) Major haplotype frequency by chromosome zone. (e) Block size (bp) by chromosome. (f) Block size (bp) by chromosome zone. Key: blue, cultivars; orange, landraces; red, wild barleys.}
\end{figure}

© 2022 The Authors. The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd., The Plant Journal, (2022), 111, 1580–1594
for the remainder of the chromosomes, see Figure S5). Polytomy, often observed only for cultivar and landrace zone-3 SNPs, indicated an inability to distinguish these accessions, whereas zone-3 SNPs on chromosome 7H split domesticated barley into two major clusters associated with different sets of wild barleys (Figure 6b) in a pattern not observed for 4H (other chromosomes except 3H showed a similar pattern to 4H, Figure S5).

To capture the variation characteristics of zone-3 ‘phylogenies’ visually, we assigned individuals to simplified ‘haplotype groups’ (haplogroups), which allowed the identification of subgroups of related haplotypes, where the genetic distance between accessions within groups was set at a maximum value of 0.045 according to the methods of Balaban et al. (2019). On this basis, we identified between nine and 21 haplogroups for the zone-3 region of each chromosome (Figures S6–S12; Table S5). By tracing the haplogroup identity of each accession, parallel plots revealed differences in the sample-wide diversity profiles of zone 3 between chromosomes for the different groups (Figure 7, each run of connected lines represents a summary of haplotype positions for a barley accession). These profiles show that the vast majority of cultivars share a single zone-3 haplogroup for each chromosome, except for 7H, with two major groups, one that represented primarily two-rowed types and the other that represented primarily six-rowed types (Figure 7b). This split for 7H was mirrored for two-rowed and six-rowed landraces (Figure 7c; evident also in Figure 6b). Of the 113 zone-3 haplogroups identified across all chromosomes and barley categories, 110 were present in wild barleys, with only 34 and 23 present in landraces and cultivars, respectively (Figures S6–S12). Several relatively common haplogroups in wild barley (e.g. 2H, 5H, 6H; Figures S7, S10 and S11, respectively) appeared to show a gradient of frequency occurrence across barley categories where landraces had intermediate frequencies higher than cultivars, possibly representing trails of founder events in the development of the modern crop. Summarized counts of haplogroups for cultivars and landraces showed the predominance of single haplogroups for most barley chromosome zone-3 regions, with this predominance being less pronounced for landraces than for cultivars (Figures S6–S12). Comparing these predominant domesticated zone-3 haplogroups with wild barley, only in two chromosomes (1H and 4H) were the same haplogroups the most common, whereas for other chromosomes the predominant domesticated haplogroup occurred in less than 10% of wild barleys. In the
case of chromosome 7H that showed row-type-related zone-3 haplogroups for domesticated barleys (Figure 7b, c), the two-row- and six-row-related haplogroups occurred in 20 and 13% of wild accessions (all wild types are two-row type), respectively (Figure S12). To explore this further, we plotted the geographical position of the common cultivar haplogroups that were present in wild barley, based on known collection coordinates (Figures S6–S12), observing considerable variation in distribution, depending on chromosome. For both chromosomes 1H and 4H, where all barley categories shared the same most common zone-3 haplogroup, these were observed across the geographic range of wild barley (Figures S6 and S9). Where the dominant domesticated haplogroup for a zone-3 region only occurred at low frequency in wild barley, however, geographic distributions – representing the putative ancestral origins of the crop – varied in wild barley by chromosome (Figures S7, S8, S10, S11 and S12). On chromosome 2H, for example, the most common domesticated haplogroup was present in only six wild barleys restricted to Israel and Jordan (Figure S7), whereas on chromosome 5H the most common domesticated haplogroup was again present in

Figure 6. Maximum-likelihood (ML) trees for barley constructed using single-nucleotide polymorphisms (SNPs) from zones 1 and 2, compared with ML tree constructed using zone-3 SNPs. (a) Chromosome 4H. (b) Chromosome 7H.

© 2022 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd., The Plant Journal, (2022), 111, 1580–1594
only six wild barleys but, in this case, these were distributed across the Fertile Crescent (Figure S10). The row-related zone-3 haplogroups observed in domesticated barley for chromosome 7H showed an interesting geographic distribution in wild barley, with the two-row-associated haplogroup restricted to the Fertile Crescent and the six-row-associated haplogroup distributed throughout the range (Figure S12).

Domestication bottlenecks and the effects of selection predict reductions in genetic diversity and the accumulation of deleterious alleles in a finite domesticated gene pool (Comeron et al., 2008; Lu et al., 2006; Makino et al., 2018). We were interested to explore whether potential deleterious alleles had, as a result of evident bottlenecks and a lack of recombination, become fixed in the barley crop gene pool. Based on SnpEff annotation (Cingolani et al., 2012), we located 22,387 non-synonymous SNPs within the zone-3 region across all tested barley accessions. Zone 3 of chromosome 4H had the highest count of non-synonymous SNPs, likely linked with being the physically largest such zone as well as the least diverse chromosome in domesticated barley (Table S6). The non-synonymous zone-3 SNPs were then filtered based on \( F_{ST} > 0.8 \) in both cultivar–wild barley and landrace–wild barley comparisons (see Figure 2b,c). After filtering, 92 SNPs remained and most were located on chromosomes 2H and 4H, with none in the zone-3 regions of chromosomes 3H and 7H, probably because chromosomes 3H and 7H have major splits in the pericentromeric haplogroups. The PROVEAN (Choi et al., 2012) scores of the 92 SNPs indicated that 29 cultivar alleles had values that were lower than the predefined threshold of −2.5, suggesting a deleterious effect (Table 2). Twenty-eight of the 29 were missense variants, with a single stop-loss variant on chromosome 6H. At least three genes that harboured ‘fixed’ deleterious alleles were of potential agricultural interest and are highlighted in Table 2. On chromosome 1H the affected gene was a galactosyltransferase, which could be related to the biosynthesis of arabinoxylan, a cell wall component and a main contributor of dietary fibre (Hassan et al., 2017); on chromosome 2H, the gene annotated as the E3 ubiquitin protein ligase NEURL1B is a candidate associated with grain weight in maize (Zhao & Su, 2019); and on chromosome 6H, an Xaa-Pro peptidase could relate to the mobilization of barley storage proteins during germination (Davy et al., 2000). The functional implication of these predicted deleterious alleles will require further verification.

Finally, we examined the function of genes within zone 3 to determine any over-representation of Gene Ontology (GO) terms (Table S7, with known agriculturally important genes highlighted). When analysis was performed on combined zone-3 gene sets compared with all genes (for all seven chromosomes), GO terms with housekeeping functions were enriched, such as nucleic acid binding, DNA integration and RNA-dependent DNA biosynthetic processes (Figure S13), as had previously been observed by Mascher et al. (2017). When our analysis was performed individually for chromosome zone-3 genes, varying GO terms were enriched (Figures S13 and S14). For example, pollen wall development was only found to be enriched for zone 3 of chromosome 1H, whereas root developmental genes (root morphogenesis and root hair tip) were over-represented for zone-3 regions of chromosomes 2H and 3H. For chromosomes 4H and 5H, zone-3 regions were enriched with plastid-related GO terms, including chloroplast organization, chloroplast fission and plastid translation. Zone 3 of chromosome 4H, which showed distinctive selective sweep signals in cultivars, also had translation-related terms over-represented, such as translational termination, translation release factor and mRNA splicing. It would be reasonable to speculate that human selection
| Chr. | Position | Effect | Wild seq. | Cultivar seq. | Gene affected | Transcript affected | PROVEAN score | Annotation | Morex v.3 gene ID |
|------|----------|--------|-----------|--------------|---------------|-------------------|---------------|------------|------------------|
| 1H   | 161 039 495 | Missense | Asn | Tyr | BART1_0-u02060 | 1 | -3.819 | Galactosyltransferase | HORVU.MOREX.r3.1HG0031180 |
| 1H   | 253 486 741 | Missense | Ser | Phe | BART1_0-u02519 | 1, 2, 3 | -5.483 to -5.800 | ABC transporter G family member 24 | HORVU.MOREX.r3.1HG0038900 |
| 1H   | 256 277 577 | Missense | Ala | Val | BART1_0-u02532 | 1, 3, 4 | -3 | n/a | HORVU.MOREX.r3.1HG0039050 |
| 2H   | 256 057 192 | Missense | Pro | Ser | BART1_0-u10642 | 11, 31 | -2.511 | Pre-mRNA-splicing factor ATP-dependent RNA helicase DEAH7 | HORVU.MOREX.r3.2HG0142550, HORVU.MOREX.r3.2HG0142540 (gene split in Morex v.3) |
| 2H   | 265 489 889 | Missense | Cys | Arg | BART1_0-u10590 | 2 | -3.955 | n/a | HORVU.MOREX.r3.2HG0142940 |
| 2H   | 271 533 763 | Missense | Pro | Ser | BART1_0-u10601 | 1, 2 | -6.607 to -6.973 | E3 ubiquitin protein ligase NEURL1B | HORVU.MOREX.r3.2HG0143100 |
| 2H   | 273 026 038 | Missense | Glu | Asp | BART1_0-u10619 | 4 | -2.911 | Peptide-(N4)-(N-acetyl-β-glucosaminyl)asparagine amidase | HORVU.MOREX.r3.2HG0143200 |
| 2H   | 288 348 617 | Missense | Asp | Val | BART1_0-u10701 | 1 | -7.99 | ATP-dependent DNA helicase | HORVU.MOREX.r3.2HG014360 |
| 2H   | 302 860 598 | Missense | Ser | Thr | BART1_0-u10798 | 2, 3 | -3 | n/a | no hit |
| 2H   | 325 288 183 | Missense | Ser | Arg | BART1_0-u10900 | 1, 2 | -5 | n/a | HORVU.MOREX.r3.2HG0146980 |
| 2H   | 327 156 323 | Missense | His | Arg | BART1_0-u10915 | 1 | -8 | n/a | no hit |
| 2H   | 342 024 777 | Missense | Cys | Tyr | BART1_0-u11010 | 1 | -10.236 | Tyrosine-sulfated glycopeptide receptor 1 AUGMIN subunit 3 | HORVU.MOREX.r3.2HG0149180 |
| 2H   | 352 826 802 | Missense | Lys | Met | BART1_0-u11071 | 1 | -5.78 | n/a | HORVU.MOREX.r3.2HG0150130 |
| 2H   | 365 683 330 | Missense | Gly | Asp | BART1_0-u11152 | 1 | -6.767 | n/a | HORVU.MOREX.r3.2HG0152460 |
| 2H   | 397 248 990 | Missense | Ser | Thr | BART1_0-u11344 | 3, 4 | -3 | P-loop containing nucleoside triphosphate hydrolase | HORVU.MOREX.r3.2HG0152460 |
| 2H   | 398 383 966 | Missense | Asn | Lys | BART1_0-u11335 | 1 | -6 | n/a | no hit |
| 4H   | 169 008 802 | Missense | Lys | Thr | BART1_0-u27962 | 1, 2 | -4.900 to -4.933 | GRAS family transcription factor containing protein, expressed | HORVU.MOREX.r3.4HG0357830 |
| 4H   | 195 116 684 | Missense | Pro | Leu | BART1_0-u28149 | 1 | -3.439 | Putative inactive leucine-rich repeat receptor-like protein kinase | HORVU.MOREX.r3.4HG0360590 |
| 4H   | 237 605 948 | Missense | Thr | Met | BART1_0-u28360 | 1, 2, 5, 6, 9, 10, 12, 15, 16, 17, 18, 19, 20, 21, 22 | -5.473 | n/a | HORVU.MOREX.r3.4HG0363910 |
| 4H   | 337 692 163 | Missense | Arg | Cys | BART1_0-u28832 | 1, 2, 5, 6, 9, 10, 12, 15, 16, 17, 18, 19, 20, 21, 22 | -6.000 to -6.233 | Rho GTPase activator | HORVU.MOREX.r3.4HG0372920 |
| 4H   | 340 149 652 | Missense | Leu | Val | BART1_0-u28824 | 1, 2, 6, 8, 9, 22 | -3 | n/a | no hit |
| 4H   | 366 230 980 | Missense | Ser | Leu | BART1_0-u29040 | 11, 12, 20 | -2.545 to -2.975 | β-Adaptin-like protein C | HORVU.MOREX.r3.4HG0374910 |
has been imposed on the variation that influences some of these biological processes. Still, more study is required to identify any beneficial alleles that are under selection. In the case of chloroplast-related genes, it may be that the nuclear chloroplast gene-related allelic composition has led to the selection or stochastic sampling of distinct chloroplast lineages during crop domestication and diversification (Molina-Cano et al., 2005).

CONCLUSION

Apart from revealing further details about the complex history of domesticated barley, our pericentromeric versus non-pericentromeric chromosomal comparisons have important practical applications. Modern, resilient barley production that ensures sustainable future harvests, in the light of challenges such as climate change (Dawson et al., 2015) and the need for greater resource-use efficiency (Cope et al., 2020), requires the recovery and exploitation of lost subsistence farming-derived (landrace) and naturally evolved (wild) traits through broad genomic access (Bailey-Serres et al., 2019). This is, however, restricted in the low-recombining pericentromeric regions of barley and other large genome cereals. Novel methods are being developed to alter the frequency and distribution of recombination and speed up the breeding process through the CRISPR/Cas9 manipulation of pro- and anticrossover (CO) genes, site-directed nucleases and/or epigenetic modifiers, among others (Taagen et al., 2020). However, their overall effectiveness in the context of crop improvement, including their potential for introducing deleterious unintended effects (e.g. increased mutation frequency or genome instability), remains to be assessed. Here, by using a large panel of cultivar, landrace and wild barleys, and chromosome zone-specific DNA sequence information, we have revealed in detail the extent to which the lack of recombination in pericentromeric regions has, and will likely continue to, constrain progress in barley breeding. Based on the measure of haplotype block size, we show that even the most recombination-accessible region of the cultivated barley genome (zone 1) has only around the same accessibility as the least recombination-accessible part of the wild barley genome (zone 3). Calculations of selective sweeps further indicate the consequences of linkage drag in cultivars, with the most accessible part of the barley cultivar genome having, overall, significantly higher selection scores than the least accessible genomic region of wild barley.

EXPERIMENTAL PROCEDURES

Sample selection, library preparation and exome sequencing

The germplasm chosen for this study is described in Table S1. Data on the majority of cultivars (163) and landraces (259) in the
starting panel were sourced from the European project Wheat and Barley Legacy for Breeding Improvement (WHEALBI), with the domestication status of accessions as described by Bustos-Korts et al. (2019). Other landraces (129) included in our initial panel were described by Russell et al. (2016) (known as “EXCAP” accessions). Data on wild barley accessions were obtained from several sources: for 98 accessions from EXCAP (Russell et al., 2016); for 75 accessions from barley B1K (Hübner et al., 2009); for 32 accessions from WHEALBI; for parents of a nested association mapping (NAM) population from Herzig et al. (2019); and for 61 accessions from the Wild Barley Diversity Collection (WBDC) (Steffenson et al., 2007). Library preparation and exome sequencing were described previously by Bustos-Korts et al. (2019) and Russell et al. (2016).

Reads mapping and variant calling, filtering and annotation

All sequence data were from paired-end Illumina sequencing (https://emea.illumina.com). Sequence lengths varied between 100 and 125 bp, depending on the source data set. Quality control of the raw data was carried out using FASTQC (Andrews, 2010). We followed the Genome Analysis Toolkit (GATK) Best Practices (Van der Auwera et al., 2013) for read mapping, BAM file pre-processing and variant calling. For the latter two steps, GATK 3.4.0 was used. The GATK Best Practices guidelines recommend the mapping of raw reads to enable barley arat duplication of paired-end read mappings. Consequently, no read trimming was carried out prior to mapping. In this scenario, read errors and adapter sequences are flagged up by the mapping tool through soft-clipping and are disregarded during downstream analysis.

BWA-MEM (Li, 2013) was used to separately map the raw reads from each barley line to the Morex 2017 reference genome (Mascher et al., 2017), with a comparatively strict mismatch rate of 4% applied to minimize the mis-mapping of reads to location and the consequent calling of false-positive variants (Ribeiro et al., 2015). In accordance with GATK Best Practices, the primary read mappings were then deduplicated using SAMDOOLS RMDUP (Li & Durbin, 2009) to remove both optical and PCR duplicates. In the next step, indel realignment was carried out with the GATK HaplotypeCaller tool and the resulting BAM file was used to produce an initial set of variants with the HAPLOTYPECALLER tool. These variants were then filtered (QUAL > 20) with VCF2 (https://github.com/vcftool/vcftool) and used as known sites for the base quality score recalibration. A second run of the HAPLOTYPECALLER was used to produce a final GVCF file for each barley line, and this was the basis for joint geno-type calling. Individual GVCF files were batched into cohorts of size 20 or fewer using the GATK COMBINEGVCFPS tool. Cohort files were then processed using the GATK GENOTYPINGVCFPS tool to produce the final variant calls. Mappings and variants were visually spot-checked using the TABLET assembly viewer tool (Milne et al., 2013).

To produce a robust set of variants for downstream analysis, we filtered the initial set of variants using custom JAVA code. The objective was to create a set of variants with a minimum of missing genotype calls and a minimum of false-positive variant calls, but with sufficient coverage of the genome. For a variant to be retained it had to pass the following filtering criteria:

- Read depth of ≥8 in at least 50% of the samples (removes variants with low read depth)
- ≤5% of samples with missing genotype calls (maximizes sample representation)
- At least one homozygous sample with the minor allele as its genotype (removes variants based on one or more heterozygous samples only)
- SNP QUAL score of >30 (removes low-confidence variants)
- ≤2% of samples being heterozygous (removes false-positive variants caused by mis-mapping)
- Number of alleles = 2
- Variant type is not insertion or deletion or multi-nucleotide polymorphism

The variants were then functionally annotated using SNPEFF (Cingolani et al., 2012), using the barley reference transcript data set BART 1.0 (Rapazote-Flores et al., 2019) as the basis for predictions.

Comparison of on/off-target variants and rare/non-rare variants

To allow a comparative analysis of variants that were on/off target with regards to the exome capture probes, the exome capture design file was obtained from the Nimblegen website (https://sftp.rch.cm/diagnostics/sequencing/nimblegen_annotations/ez_barley_exome/barley_exome.zip) and the capture probe sequences were mapped to the Morex 2017 reference genome using BLASTN (Altschul et al., 1990; Camacho et al., 2009), with an e-value cutoff of 1e-10 and a minimum percentage identity of >90. The BEDtools intersect method (Quinlan & Hall, 2010) was then used to compute the overlap between the filtered variants and the mapping positions of the exome capture probes, and variants overlapping the probes were classified as on target, whereas the remainder were classified as off target. Read depth and variant quality scores were then extracted from the VCF file using VCFTOOLS (Danecek et al., 2011).

‘Rare’ SNPs were defined as those with an MAF of <0.05. The averaged genotype quality score (GQ) was extracted for rare and non-rare SNPs from the VCF file using VCFTOOLS (Danecek et al., 2011). To compare GQ between major and minor alleles, the values for each called position were extracted across accesses using VCFTOOLS and grouped into major and minor alleles using a custom PYTHON script for distribution plot.

Genome-wide relatedness and ordination

A target of 10 000 SNPs (n = 9845) were randomly selected from the filtered variant data set using SELECTVARIANTS in GATK for the reconstruction of genome-wide relatedness and PCO. The PCO was performed using FAST 3.25 (Hammer et al., 2001) and the result visualized by CURLYWHIRL 1.19.03 (https://ics.hutton.ac.uk/curlwyhirl/).

Barley genetic landscape

Genetic diversity (τ) and pairwise FST values for SNPs were calculated using ‘-site-pi’ and ‘-weir-fst-pop’, respectively, in VCFTOOLS. The τ values were plotted using a moving average method with a window size of 10 000 bp, whereas the FST values were plotted on a per-base basis so that the fine-scale horizontal track patterns in pericentromeric regions could be observed. The zone-3 genotype heat map was visualized with FLAPJACK (Milne et al., 2010), with SNPs having MAFs of <0.05 being excluded to reduce noise, without altering the overall genetic variation pattern. The LD haplotype blocks were estimated using the `-blocks` function in PLINK 1.904 (Purcell et al., 2007), under default settings, following the block definition method mentioned in Gabriel et al. (2002), except that the limitation of block size was increased to allow large blocks that could potentially cover whole chromosomes (the `-blocks-max-kb` parameter was set to 800 000 kb). A similar approach had been used previously in wheat (Hao et al., 2017). The LD decay profiles (R2 vs distance) were calculated based on a thinned SNP data set
(thinned using the ‘thin’ function in vcf2tools), to keep only SNPs with at least a 10 000-bp interval distance. The thinned data were used for LD estimation via the plink ‘-r2’ function, with options applied to allow the calculation of $R^2$ for all pairwise SNPs within a given window size of 15 000 kb (--ld-window 100 000 -ld-window-kb 15 000), with $R^2$ values above 0.05 being reported. Distance information used for the final visualization was taken from the PLINK LD output file (BP_B).

Haplotype counts for chromosomes and chromosome zones were corrected estimates accounting for the different sample sizes of cultivar, landrace and wild barley categories. For each category, counts were based on randomly selected samples of 100 accessions. The randomization procedure was performed 100 times and average values were used. We applied this sample size correction specifically to haplotype richness estimation because of the potential high sensitivity of this parameter to sample size (when there are a large number of different haplotype states), which is not the case for individual SNP-based (i.e. biallelic) diversity estimates such as $\pi$.

Signatures of selective sweeps were detected using RAISR 2.4 (Alachiotis & Pavlidis, 2018), with the option to impute missing data (--M –1). The 95th percentile of $\pi$ was calculated for each population and used as the threshold to highlight outlier SNPs. All plotting was performed with R 3.6.0 and moving averages calculated using the ‘rollapply’ function of zoo 1.8-8 (Zeileis & Grothendieck, 2005). The chromosome containing unmapped contigs (chrU) was excluded from all analyses.

Zone-3 evolution comparison

We followed the zone-3 coordinates reported in the Morex 2017 reference genome paper (Mascher et al., 2017) and separated SNPs based on the coordinates for each chromosome. The ‘phylogenies’ and PCO analyses were performed as described in a previous article. For the intraspecific ‘phylogenetic’ relatedness analysis, the VCF file was first converted to PHYLIP format using VCF2PHYLIP.PY 2.0 (https://github.com/edgardomortiz/vcf2phylip). The GTR + G4 model was then selected under the Akaike information criterion (AIC) calculated via MODELTEST-NG 0.1.6 (Darriba et al., 2020), and the unrooted ML tree was estimated using RAxML-NG 0.6.0 (Kozlov et al., 2019). Trees were visualized using the interactive Tree Of Life (iTOL) web server (Letunic & Bork, 2019).

Identification of BarT1 homologues in Morex v.3

BART1 homologues in the Morex v.3 reference assembly (Mascher et al., 2021) were identified with BLASTP (Altschul et al., 1990) using BART1 proteins as queries and Morex v.3 proteins as subjects. Raw hits were sorted by percentage identity (descending) and query coverage per high-scoring segment pair (HSP) (descending) and then filtered by percentage identity (>=98%). This leaves the best hit topmost but still retains multiple transcripts for each query. We then removed duplicates by query gene and subject gene to leave the best hit for a given query-subject gene combination, while still allowing for split/fused genes. Some BART1 genes had no hits in Morex v.3 with the above approach, whereas others had multiple hits, presumably with genes having been fused or collapsed in BART1.

ACKNOWLEDGEMENTS

This article was a collaborative study and was funded from the following granting bodies: ERC project 669182 ‘SHUFFLE’ (to YYC, MS and RW); European Union’s Seventh Framework Programme (FP7/2007–2013) under grant agreement no. FP7-613556, WHEALBI (to JRR and IKD); Scottish Government Rural and Environment Science and Analytical Services (RESAS) (to JRR, RW, MMB and PEH); USDA-NIFA Triticeae Coordinated Agricultural Project 2011-68002–30029 (to GM, LL, AA, CL and BJS), and National Science Foundation (grant IOS-1339393) and the Minnesota Agricultural Experiment Station Variety Development Fund (to KPS, JCF and PLM). The authors would like to thank Professor Nils Stein and colleagues at Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben. The authors acknowledge the Research/Scientific Computing teams at The James Hutton Institute and the National Institute of Agricultural Botany (NIAB) for providing computational resources and technical support for the ‘UK’s Crop Diversity Bioinformatics HPC’ (BBSRC grant BB/S019669/1), the use of which has contributed to the results reported within this paper.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest associated with this work.

AUTHOR CONTRIBUTIONS

YYC carried out the statistical and genetic analysis and drafted the first version of the article. MS, MMB and PEH assembled the exome capture data and performed variant calling, filtering and annotation. IKD contributed to genetic interpretation and writing the article. LL contributed to the evolutionary interpretation and editing of final version for publication. AA, KPS and JCF generated exome capture data from a section of wild barley lines (Table S1, WBDC). GM and BJS collected and assembled the WBDC collection. PLM contributed to the genetic and evolutionary interpretation and drafting the article. RW conceived the project and assembled the collaborators. JR conceived part of the project and contributed to the interpretation and writing of the article.

DATA AVAILABILITY STATEMENT

All raw exome capture sequencing reads are publicly available as ENA/SRA accessions. Accession numbers are listed in Table S1. The variant data used for analysis is available as a BCF formatted file in the Zenodo data repository, DOI 10.5281/zenodo.6382440. Scripts used for data analysis and plotting of results are available at https://github.com/mb47/barleyvariomics.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Geographical distribution of the genotyped barley germplasm.

Figure S2. Comparison between rare (minor allele frequency, MAF < 0.05; $n = 2 742 309$) single-nucleotide polymorphisms (SNPs) and other ($n = 340 564$) SNPs.

Figure S3. Comparison between on-target ($n = 1 736 337$) and off-target ($n = 1 346 536$) single-nucleotide polymorphisms (SNPs).

Figure S4. Extent of linkage disequilibrium by groups ($\eta^2$).

Figure S5. Principal component analysis (PCA) plot of zone-3 regions, and comparison of maximum-likelihood (ML)
phylogenies derived from zone-1 + zone-2 regions with that derived from zone-3 regions.

Figure S6. Genetic diversity in chr1H pericentromeric regions.

Figure S7. Genetic diversity in chr2H pericentromeric regions.

Figure S8. Genetic diversity in chr3H pericentromeric regions.

Figure S9. Genetic diversity in chr4H pericentromeric regions.

Figure S10. Genetic diversity in chr5H pericentromeric regions.

Figure S11. Genetic diversity in chr6H pericentromeric regions.

Figure S12. Genetic diversity in chr7H pericentromeric regions.

Figure S13. Word clouds for the Gene Ontology (GO) enrichment results for zone-3 genes.

Figure S14. Gene Ontology (GO) terms in zone 3 of each chromosome.

Table S1. Information of 879 exome sequence Hordeum vulgare (barley) accessions.

Table S2. The filtered single-nucleotide polymorphism (SNP) set.

Table S3. Number of genes covered in exome capture sequenc- ing.

Table S4. The result from non-parametric analysis of variance (Kruskal–Wallis H-test) suggests at least one of the group shows significant difference in block size among all nine groups tested (P < 0.01).

Table S5. Haplotype grouping (haplgroup) for each accession.

Table S6. Summary of non-synonymous alleles in zone 3.

Table S7. Agriculturally important known Hordeum vulgare (barley) genes.

REFERENCES

Alachiotis, N. & Pavlidis, P. (2018) RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. Communications Biology, 1, 79. https://doi.org/10.1038/s42003-018-0085-8

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. Journal of Molecular Biology, 215, 403–410. https://doi.org/10.1016/0022-2836(90)90320-L

Andrews, S. (2010) FastQC. Available from: https://www.bioinformatics. babraham.ac.uk/projects/fastqc/

Bailey-Serres, J., Parker, J.E., Ainsworth, E.A., Oldroyd, G.E.D. & Schroeder, B.K. (2013) Genetic strategies for improving crop yields. Genetics, 193, 1193–1200. https://doi.org/10.1534/genetics.112.138578

Bailey-Serres, J., Parker, J.E., Ainsworth, E.A., Oldroyd, G.E.D. & Schroeder, B.K. (2013) Genetic strategies for improving crop yields. Genetics, 193, 1193–1200. https://doi.org/10.1534/genetics.112.138578

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. et al. (2009) BLAST: architecture and applications. BMC Bioinformatics, 10, 421. https://doi.org/10.1186/1471-2105-10-421

Charlesworth, B., Morgan, M.T. & Charlesworth, D. (1993) The effect of deleterious mutations on neutral molecular variation. Genetics, 134, 1289–1303. https://doi.org/10.1093/genetics/134.4.1289

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. & Chan, A.P. (2012) Predicting the functional effect of amino acid substitutions and indels. PLoS One, 7, e46888. https://doi.org/10.1371/journal.pone.0046888

Choulet, F., Alberti, A., Theil, S. et al. (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science, 345, 1249721. https://doi.org/10.1126/science.1249721

Cingolani, P., Platts, A., Wang Le, L. et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin), 6, 80–92. https://doi.org/10.4161/fly.19959

Comeron, J.M., Williford, A. & Kliman, R.M. (2008) The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity, 100, 19–31. https://doi.org/10.1038/hdy.2008.105

Cope, J.E., Russell, J., Norton, J.G., George, T.S. & Newton, A.C. (2020) Assessing the variation in manganese use efficiency traits in Scottish barley landrace Bere (Hordeum vulgare L.). Annals of Botany, 126, 289–300. https://doi.org/10.1093/abo/mca039

Danecek, P., Auton, A., Abecasis, G. et al. (2011) 1000 Genomes Project Analysis Group. The variant call format and VCFTools. Bioinformatics, 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Darriga, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B. & Flouri, T. (2020) ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Molecular Biology and Evolution, 37, 291–294. https://doi.org/10.1093/molbev/msz019

Davy, A., Thomsen, K.K., Juliano, M.A., Alves, L.C., Svendsen, I. & Simpson, D.J. (2000) Purification and characterization of Barley Dipeptidyl Peptidase IV. Plant Physiology, 122, 425–432. https://doi.org/10.1104/pp.122.2.425

Dawson, I.K., Russell, J., Powell, W., Steffenson, B., Thomas, W.T.B. & Waugh, R. (2015) Barley: a translational model for adaptation to climate change. The New Phytologist, 206, 913–931. https://doi.org/10.1111/nph.13266

Fang, Z., Gonzales, A.M., Clegg, M.T. et al. (2014) Two genomic regions contribute disproportionately to geographic differentiation in wild barley. Genetics, 193, 1193–1200. https://doi.org/10.1534/genetics.113.150951

Fang, Z., Pyhajarvi, T., Weber, A.L., Dawe, K. et al. (2012) Megabase-scale inversion polymorphism in the wild ancestor of maize. Genetics, 191, 883–894. https://doi.org/10.1534/genetics.111.138578

Felsenstein, J. (1974) The evolutionary advantage of recombination. Genetics, 78, 737–756. https://doi.org/10.1093/genetics/78.2.737Top

Feuillet, C., Langridge, P. & Waugh, R. (2008) Cereal breeding takes a walk on the wild side. Trends in Genetics, 24, 24–32. https://doi.org/10.1016/j.tig.2007.11.001

Gabriel, S.B., Schaffner, S.F., Nguyen, H. et al. (2002) The structure of haplotype blocks in the human genome. Science, 296, 2225–2229. https://doi.org/10.1126/science.1069424

Gore, M.A., Chia, J.M., Eslhize, R.J. et al. (2009) A first-generation haplotype map of maize. Science, 326, 1115–1117. https://doi.org/10.1126/science.1177837

Hammer, D., Harper, D.A.T. & Ryan, P.D. (2001) Past: Paleontological statistics software package for education and data analysis. Paleontologia Electronica, 4, 5 Available from: http://paleo-electronica.org/2001_1/past/issue1_01.htm

Hao, C., Wang, Y., Chao, S. et al. (2017) The iSelect 9 K SNP analysis revealed polyploidization induced revolutionary changes and intense human selection causing strong haplotype blocks in wheat. Scientific Reports, 7, 41247. https://doi.org/10.1038/srep41247

Hassan, A.S., Houston, K., Lahneht, J. et al. (2017) A Genome Wide Association Study of arabinoxylan content in 2-row spring barley grain. PLoS One, 12, e0182537. https://doi.org/10.1371/journal.pone.0182537

Herzig, P., Backhaus, A., Seiffert, U., von Wiren, N., Pillen, K. & Maurer, A. (2019) Genetic dissection of grain elements predicted by hyperspectral imaging associated with yield-related traits in a wild barley NAM population. Plant Science, 285, 151–164. https://doi.org/10.1016/j.plantsci.2019.05.008

Higgins, J.D., Osman, K., Jones, G.H. & Franklin, F.C.H. (2014) Factors underlying restricted crossover localization in barley meiosis. Annual Review of Genetics, 48, 29–47. https://doi.org/10.1146/annurev-genet-2013-092508

Hill, W.G. & Robertson, A. (1966) The effect of linkage on limits to artificial selection. Genetics, 98, 269–294. https://doi.org/10.1017/w0010673230010156

Hubner, S., Hoffken, M., Oren, E., Haseneyer, G., Stein, N., Graner, A. et al. (2009) Strong correlation of wild barley (Hordeum spontaneum)
