Supporting Information

Cockram et al. 10.1073/pnas.1010179107

SI Text

Genetic Stratification. Of the 620 cultivars included, clustering cultivars according to seasonal growth habit and ear row number within the principle component analysis (PCA) plot did not take into account the 78 cultivars for which phenotype data for one or both traits were missing. Of the cultivars with phenotype data for both traits, six are found within clusters that do not agree with their recorded phenotypic states. The cultivars Damas (UK) and Astrix (non-UK) belong to an infrequent and ill-defined seasonal growth habit class termed alternative, variously defined as cold-hardy spring types or those possessing a weak vernalization response. Both alternative cultivars are six-row and cluster within the winter six-row cloud. The cultivars Mosaic and Ayana are recorded as six-row winter types but cluster within the two-row winter cloud. Finally, Askanova (two-row winter) and accession HOR2937 (two-row spring, non-UK) cluster within six-row winter and six-row spring clouds, respectively.

Correction of Genetic Stratification. We evaluated various statistical methodologies to identify and account for genetic substructure within our varietal collection: structured association (1–3), a principal component analysis-based approach (4), and mixed linear modeling (5), using software developed in ref. 6, which we find to be ~10 times faster than a similar approach implemented by the software EMMA (5). Genomic control (7, 8) was also used, both on its own and after analysis by all other methods, to account for residual inflation of the test statistic.

***Corrected Between-Marker* Linkage Disequilibrium.** The *P* values calculated using the mixed model assume normally distributed errors, when a binomial distribution is more appropriate. However, the standard 2 × 2 contingency χ^2 test relies on a normal approximation to the binomial and returns *P* values very similar to those obtained from tests that model the binomial distribution directly (9). Although this approximation fails when expected numbers are rare, our experimental dataset is restricted to analysis with SNP minor allele frequency (MAF) > 0.1. Nevertheless, we further explored the possibility that breaching failure of the assumption of normality might result in a systematic excess of low *P* values (i.e., false discoveries). Under the null hypothesis, we expect a uniform distribution of *P* values. However, if there is a systematic problem giving an excess of low *P* values, this would manifest itself even if we analyze data in which there genuinely is no association (i.e., the null hypothesis holds). To show this, we used a randomized marker genotype of stated allele frequency (0.01, 0.05, 0.1, or 0.2) as a response variable in the mixed model. Marker genotypes from the experimental panel were used as explanatory variables. The probability of the null hypothesis was estimated for each marker for each of 1,000 randomized marker genotypes. To test the effect of markedly contrasting allele frequencies between response and explanatory variables, the data were partitioned to show *P* value distributions from marker panels with MAF ≥ 0.1, 0.2, or 0.45. Histograms of the resulting *P* value distributions are remarkably uniform (Fig. S1). Although a slight deficit of low *P* values is observed, this is particularly notable when the allele frequency of the response variable is low. However, there is never an excess of low *P* values, indicating that our results can only be conservative: by assuming normality, it seems that, although we have killed some power, we have not increased the false-discovery rate. These results have implications for the statistical validity of previously published loci for binomially scored traits detected using the mixed model. Although the issue is briefly acknowledged by Atwell et al. (10) in their supplementary material, we are not aware of a previous demonstration of robustness when applying the mixed model to binary traits. We have not studied loss of power here, but in backcross and F_2 mapping populations, there is very little such loss (11).

Mendelian Loci Previously Identified in Biparental Mapping Populations. The *ANTHOCYANINLESS 2* (*ANT2*) locus has been reported in numerous mapping populations; the greatest resolution was obtained by Lahaye et al. (12), who mapped *ANT2* as a single Mendelian locus with flanking markers 2.1 cM proximal (MWG503, orthologous to rice gene LOC_Os04g46200.1) and 0.2 cM distal (MWG892). Although no DNA sequence is available for MWG892, a sequenced barley bacterial artificial chromosome (BAC) clone anchored by *Rar1* (13), located another 0.7 cM distal to MWG892, contains *TIP1*. This latter gene is orthologous to LOC_Os04g46940.1, and is within the significant interval identified for all anthocyanin traits by genome-wide association (GWA) analysis (defined by barley transcripts orthologous to rice genes LOC_Os04g52380.1 and LOC_Os04g47690.1). As well as the *ANT2* locus, a subset of additional associations identified by GWA analysis are located within chromosomal regions previously reported as carrying relevant Mendelian loci. The *VRS1* locus controlling ear row number maps to chromosome 2HL (14) within the GWA intervals for sterile spikelet development and sterile spikelet attitude. Its genetic map location at 85.92 cM is estimated according to the position of marker 11_10287 derived from HarvEST (http://harvest.ucr.edu/) U32 Unigene 2371. Expressed sequence tags (ESTs) included within this transcript assembly correspond to restriction fragment length polymorphism marker cMWG699, which has previously been mapped 0.1 cM distal to *VRS1* (14). In addition, the peak marker for ear row number (11_20606, chromosome 4H, 26.19 cM) maps to a region broadly collinear with the *Int-c* locus, which modifies the degree of fertility of lateral spikelets (15, 16). Partitioning between two- and six-rowed barley for *Int-c* alleles that either prevent or promote anther development in lateral spikelets explains the strong association observed in our GWA scan. Mendelianly inherited loci for the following traits have also been previously genetically mapped: aleurone color (*Blx1*, chromosome 4H) (17), hairiness of leaf sheath (*Hsh*, 4H) (17), and grain rachilla hair type (*Srh*, 5H) (17). An SNP within the MADS [a highly conserved protein motif, named after the initials of the first four identified members of the family: MCMC1, AGAMOUS, DEFICIENS and SRF (serum response factor)] box gene thought to encode the flowering time locus *VRN-H1* is represented by marker 11_20188 (137.16 cM; HarvEST U32 Unigene 1501). Finally, the significant association for grain lateral nerve spiculation at 78.03 cM on the long arm of chromosome 2H is estimated to be 8 cM proximal to *VRS1*. A related phenotype controlled by the mutant locus *GTH1* determines the presence or absence of teeth on the lateral lemma nerves. Because *GTH1* is reported to map on the long arm of chromosome 2H, around 15 cM proximal to VRS1 (18, 19), it is predicted to approximate to the genetic region identified by GWA analysis for grain lateral nerve spiculation.

Interpretation of GWA *Peaks*. Because of the multiple origins of parental alleles in association mapping panels, regions of significant association detected during GWA studies invariably include peaks and troughs of association, within which a smaller subset of genetic markers are most highly associated with the trait. Where linkage disequilibrium (LD) is high (as was found in our population), these regions can extend over extended physical and genetic intervals. Arguably, this can make it difficult to determine whether such regions are because of one or more than one genetic determinant. We assume that the loci detected in this study are single genetic loci supported by the following observations. (*i*) Many of the associations are located in genomic regions previously shown in biparental mapping populations to contain major loci/quantitative trait loci controlling the relevant trait. (*ii*) The historic phenotypic data used in this study are collected during the submission of new varieties to ensure that they are phenotypicaly distinct from all previously released varieties. Accordingly, the phenotypes used are assumed to often be highly heritable and are frequently scored as two-state characters. (*iii*) Analysis of the proportion of the phenotypic variation accounted for by our GWA peaks suggests that, in the majority of cases, traits with significant associations are predominantly (proportion of V_p accounted $\geq$ 0.45 for 8 of 15 traits) controlled by a small number of loci of large effect. (*iv*) The predominant identification of single major loci in GWA scans agrees with our a priori power calculations, which indicate that, in a best case scenario, we only have power to detect associations for traits controlled by a small number of genetic determinants of major effect. (*v*) The wide peaks of association often observed could be caused by breeders' selection for specific traits, resulting in LD of surrounding markers with both allelic states at the causative locus and genomic background (as previously suggested for similar ranges of association observed in Arabidopsis) (10). It should be remembered that, just as in biparental crosses, it is possible that any genomic region detected by a GWA scan is controlled by more than one closely linked genetic locus. However, this can only be clarified by further investigation, such as increased mapping resolution or genetic complementation.

Comparison of GWAs with Previously Map-Based Cloned Genes. Four traits, for which GWA scans identified significant associations, map close to cloned genes of relevance. Of these, two traits controlling aspects of ear morphology (sterile spikelet development and sterile spikelet attitude) both locate a single highly significant locus on the long arm of chromosome 2H, which encompasses the homeodomain leucine-zipper homeobox gene *VRS1* known to control ear row number (20). The barley inflorescence (ear) is composed of a series of triplets, each with one central and two lateral spikelets arranged along a central rachis. The WT triplet consists of a central fertile spikelet flanked by two sterile spikelets (two-row barley). Three independent mutations within *VRS1* result in fertile lateral spikelets, conferring the six-row phenotype (20). Therefore, the strong associations observed for the two sterile spikelet traits are likely to be because of or dependent on allelic state at *VRS1*. In addition, a significant marker associated with ear attitude originates from *HvBM5A*, a MADS box gene thought to encode the flowering time locus *VRN-H1* (21); however, it is not necessarily obvious how the two phenotypes are related, and it could well be coincidental or represent a pleiotropic effect. As a major genetic component controlling seasonal growth habit, allelic state at *VRN-H1* (and a second major flowering time locus, *VRN-H2*, on chromosome 4H) has a strong effect on genome-wide genetic substructure (22). As such, there could be a higher risk that the association of the *VRN-H1* region marker with ear attitude represents a false-positive association. We note that, although GWA analysis of seasonal growth habit identifies a sporadic range of associations around *VRN-H1* (with single significant markers scattered at intervals >4 cM across a region of ~30 cM), the assayed SNP (11_20188) within the *VRN-H1* candidate gene did not show significant association with the trait that it controls. It has been suggested that such mountain ranges of association, within which causative genes/polymorphisms fail to return significant associations, could be common for loci/traits under strong selection, as was recently reported for the Arabidopsis flowering time locus *FRIGIDA* (10). Interestingly, seasonal growth habit returns a significant association at 96.92 cM on 1H (11_10396, orthologous to rice gene LOC_Os05g44760.1). Colinearity with rice predicts this marker to colocate with the barley short-day photoperiod locus *PPD-H2* (the rice ortholog of the *PPD-H2* candidate gene *HvFT3* is LOC_Os05g44180.1, located ~350 kb from LOC_Os05g44760.1). European winter varieties are essentially partitioned for predicted allelic state at *PPD-H2* (23), because the recessive *ppd-H2* allele found in almost all winter varieties delays flowering under short days, helping to prevent floral transition in autumn sown barley until favorable seasonal conditions in spring. Thus, it seems that, although GWA analysis of seasonal growth habit fails to precisely locate *VRN-H1* (or *VRN-H2*), we identify the *PPD-H2* locus as representing an important component of the winter/spring ideotype in UK germplasm. Although marker 11_10396 was also found to be significantly associated with hairiness of leaf sheath and growth habit, its proximity to *PPD-H2* could increase the chance that these represent false-positive associations because of insufficient correction for genetic substructure for traits closely correlated with the genetic determinants that control seasonal growth habit and ear row number. Similarly, we note that the major association ($-\log_{10} P = 117.60$) for hairiness of leaf sheath on chromosome 4H places this locus just 6 cM from the major seasonal growth habit locus *VRN-H2* (the estimated genetic map position of *VRN-H2* at 123.29 cM is based on the closely linked *Bmy1* SNP 11_11019).

Additional Gene-Based Markers. Putative barley orthologs of genes within selected rice and brachypodium (Release 1.0, http://www.brachypodium.org/) physical intervals were identified by BLASTn searches of public EST databases. To maximize the chances of identifying true orthologs in barley, genes with low copy number in rice and brachypodium were prioritized for marker development in barley. Alignment of selected barley EST sequences with genomic and cDNA gene models was performed manually using GeneDoc (http://www.nrbsc.org/gfx/genedoc/). To increase the probability of identifying DNA polymorphisms, barley primers were predominantly designed to flank predicted intron positions. The majority of genotyping was performed by direct Sanger sequencing of PCR products using BigDye kit v3.1 (Applied Biosystems) following the manufacturer's instructions. Fluorescently terminated extension products were separated using an ABI3730xl Bioanalyser (Applied Biosystems), and the resulting sequence was processed and analyzed using the VectorNTI Advance package v10.1.1 (Invitrogen). The 16-bp deletion within *HvbHLH1* was amplified using forward primer HvOs04g47080F-SSR in combination with the fluorescently labeled reverse primer HvOs04g47080R-SSR (Table S5). The resulting products were resolved using the ABI3730xl Bioanalyser, and alleles were scored using Genemapper (Applied Biosystems). All other exonic insertion/deletions (InDels) were resolved by electrophoresis of PCR amplification products across 1.5% agarose gels stained with ethidium bromide and visualized under UV light. The genetic map positions of the majority of genes were determined using JoinMap 3.0 (24) using the mapping populations listed in Table S5. Genetic map positions of the remaining genes were estimated by pair-wise association ($\hat{K}$) with previously mapped markers ranked by their χ^2 values (df = 1). Primers and associated information are listed in Table S5.

Phylogenetic Analysis. Unrooted phylogenies were estimated by maximum likelihood with 1,000 bootstrap replicates using the basic helix-loop-helix (bHLH) domain and downstream sequences. The relevant regions of the predicted proteins encoded by the following previously characterized *bHLH* genes were included for analysis: *AmDEL* (M84913),*AtEGL3* (At1g63650.1),*AtGL3* (At5g41315.1), *AtTT8* (At4g09820.1), *OrRc* (DQ204737), *PhAN1* (AF260919), *PhJAF13* (AF020545), *ZmB* (NM_001112236), *ZmLN1* (U57899), *ZmLc* (M26227), and *ZmR* (×15806). The maize *R* gene (*ZmR*) was used as a query for BLASTn searches for homologs within the genomes of rice, maize, and brachypodium. Protein alignments from the bHLH domain to the C terminus were constructed using ClustalW (25) and adjusted manually. *OsPlw-OSB2* (AB021080) represents an alternative ORF to that described by gene model LOC_Os04g47059.1, and we include only the former here. Although identified by BLASTn analysis, the following genes seem to be truncated upstream of the bHLH domain and therefore, were excluded from phylogenetic analysis: LOC_Os04g47059.1, LOC_Os01g39430.1, LOC_Os01g39560.1, and Sb02g00638.1. Although orthologous flanking genes were present, no homologous *bHLH* genes were identified in the sequenced *B. distachyon* genome (accession BD21), indicating species/accession-specific deletion.

BAC Library Screening. For the ongoing construction of the barley physical map, five BAC libraries were used: HVVMRXALLhA (26), HVVMRXALLeA, HVVMRXALLmA, HVVMRXALLhB, and HVVMRXALLrA. DNA extracted from BAC clones was pooled into 55 sequential Super Pools (SPs) consisting of seven 384-well plates per SP. Each SP was subsequently separated into independent sets of 5-plate, 8-row, and 10-column pools, forming 23 matrix pools. Screening the BAC libraries for genes of interest was performed by quantitative real-time PCR (qPCR) with primers designed using Universal Probe Library software (https://www.roche-applied-science.com/). Assays comprised a total reaction of 10 μL, with 2 μL BAC pool DNA (10 ng), 5 μL SYBR Green I qPCR Master Mix (Sigma-Aldrich), and gene-specific primers (500 nM each). Reactions were performed using the 7500 Fast Real-Time PCR System (Applied Biosystems) with 1 cycle at 95 °C for 10 min followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. After amplification, reaction product melting curve analysis was performed (95 °C for 15 s, 58 °C for 1 min, and 95 °C for 15 s) to confirm presence of a single reaction product. Deconvolution of qPCR screens enabled identification of individual positive BAC clones. BAC library screening primers are as follows: HvbHLH1-F 5′-ACTGTACTGTAGATCCGCGTCA-3′, HvbHLH1-R 5′-AGCCACCATCATTGTCAATCT-3′; HvOs04g47020-F 5′-GGGAGTTGTTGTGGAAGATGA-3′, HvOs04g47020-R 5′-TTCGCCTCTTCAAGTGATAATACA-3′.

RT-PCR. Total RNA was extracted using Tri Reagent (Sigma-Aldrich) from awn, auricle, and lemma tissue collected from the varieties Saffron and Retriever (30 plants each) at flowering time. Genomic DNA was removed by digestion with DNaseI (Invitrogen), and RNA was subsequently cleaned using the RNeasy Plant Mini Kit (Qiagen). First-strand cDNA synthesis from 2 μg RNA per tissue sample was performed using the M-MLV Reverse Transcriptase kit (Invitrogen) primed with oligo$(dT)_{15}$ (Sigma-Aldrich). Semiquantitative RT-PCR was carried out using a Veriti 96-well Thermo Cycler (Applied Biosystems) in the following reaction volumes: 1 μL cDNA, 1 μL DNTPs (5 mM each), 0.1 μL FastStart Taq DNA polymerase (5 U/μL; Roche), 1 μL Buffer (10×), 1 μL $MgCl_2$ (20 mM), 0.5 μL each primer (10 mM), 4.9 μL H_2O. Thermo cycling parameters: 38 cycles of 96 °C for 5 min (*HvbHLH1*), 25 cycles (*ACTIN*) of 96 °C for 40 s, 62 °C for 40 s, and 72 °C for 60 s, and final extension of 72 °C for 7 min. RT-PCR primers: HvbHLH1-F2 and HvbHLH1-R6 (Table S5). Primers for the control gene *ACTIN* are previously described (27). Amplicons were separated on 1.5% ethidium bromide stained agarose gels and visualized under UV. Amplicon intensity was measured using ImageJ v1.43 (http://rsbweb.nih.gov/ij/), with *HvbHLH1* expression normalized against *ACTIN*. Data reported are the mean of five technical replicates.

Estimations of Statistical Power, Heritability, and Phenotypic Variation. Power was estimated using our experimental markers and individuals by simulating phenotypes controlled by n_l loci randomly selected from the marker panel. Phenotypes were allocated (*i*) a genetic component, for which allele 1 was considered positive (contributing $9/n_l$ units), whereas allele 0 made no contribution and (*ii*) a structural component dependent on the seasonal growth habit and ear row number status of each variety (spring/two-row = +5; spring/six-row = +7; winter/two-row = +0; winter/six-row = +2). Values for gene and structure effects were arbitrarily chosen. The simulated genetic components were used to estimate the genetic variation (V_G) for the simulated trait. The value of environmental variation (V_E) necessary to achieve a heritability $[h^2 = \frac{V_G}{(V_E+V_G)}]$of 0.5 and 0.9 were obtained by drawing values from a distribution ~N(0, V_E). After removal of simulated causative markers from the analysis, simulated traits were subsequently used in GWA scans (after correction using genomic control or the mixed model). Simulations for n_l = 1, 2, and 10 were replicated 100, 100, and 375 times, respectively. Q values were used to determine significance against a false discovery rate threshold of 0.1. Discoveries were considered true if they were both significant and fell within a specified window of genetic distance from the known position of the simulated causative locus (±1, 2, 4, and 10 cM). Mean proportions of loci identified within each discovery window are reported.

Trait heritabilities (h^2) were estimated as $V_G/(V_G + V_E)$, where V_G and V_E are taken from a fit of the null (no marker association) mixed model. The variation-explained SNPs were estimated as the difference in residual phenotypic variation between the null and alternative (with marker association) models. Residual phenotypic variation was taken as the sum of the genetic and environmental variance estimates from the mixed model. The proportion of variation explained (V_P) was, therefore, estimated as $(V_G + V_E - V'_G - V'_E)/(V_G + V_E)$, where V_G and V_E are estimates from the null model and $V'_G - V'_E$ are estimates from the alternative model.

1. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
2. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. *Am J Hum Genet* 67:170–181.
3. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
4. Price AL, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909.
5. Yu J, et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208.
6. Astle W, Balding DJ (2009) Population structure and cryptic relatedness in genetic association studies. *Stat Sci* 24:451–471.
7. Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55: 997–1004.
8. Reich DE, Goldstein DB (2001) Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol* 20:4–16.
9. Snedecor GW, Cochran WG (1976) In *Statistical Methods*, eds Snedecor GW, Cochran WG (Iowa State University Press, Ames, IA), pp 119–223.
10. Atwell S, et al. (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:627–631.
11. Visscher PM, Haley CS, Knott SA (1996) Mapping QTLs for binary traits in backcross and F_2 populations. *Genet Res* 68:55–63.
12. Lahaye T, et al. (1998) High-resolution genetic and physical mapping of the *Rar1* locus in barley. *Theor Appl Genet* 97:526–534.
13. Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res* 10: 908–915.
14. Komatsuda T, Tanno K (2004) Comparative high resolution map of the six-rowed spike locus 1 (*vrs1*) in several populations of barley, *Hordeum vulgare* L. *Hereditas* 141:68–73.
15. Waugh R, Jannink J-L, Muehlbauer GJ, Ramsay L (2009) The emergence of whole genome association scans in barley. *Curr Opin Plant Biol* 12:218–222.
16. Komatsuda T, Mano Y (2002) Molecular mapping of the intermedium spike-c (*int-c*) and non-brittle rachis 1 (*btr1*) loci in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 105:85–90.
17. Lundqvist U, Franckowiak J, Konishi T (1996) New and revised descriptions of barley genes. *Barley Genet Newsl* 26:22–43.

18. Finch RA, Simpson E (1978) New colours and complementary colour genes in barley. *Z Pflanzenzuchtg* 81:40–53.
19. Wexelsen H (1934) Quantitative inheritance and linkage in barley. *Hereditas* 18: 307–308.
20. Komatsuda T, et al. (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc Natl Acad Sci USA* 104:1424–1429.
21. Distelfeld A, Li C, Dubcovsky J (2009) Regulation of flowering in temperate cereals. *Curr Opin Plant Biol* 12:178–184.
22. Cockram J, et al. (2008) Association mapping of partitioning loci in barley. *BMC Genet* 9:16.
23. Faure S, Higgins J, Turner A, Laurie DA (2007) The *FLOWERING LOCUS T*-like gene family in barley (*Hordeum vulgare*). *Genetics* 176:599–609.
24. Van Ooijen JW, Voorrips RE (2004) *JoinMap 3.0: Software for the Calculation of Genetic Maps in Experimental Populations* (CPRO-DLO, Wageningen, The Netherlands).
25. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
26. Yu Y, et al. (2000) A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. *Theor Appl Genet* 101:1093–1099.
27. von Zitzewitz J, et al. (2005) Molecular and structural characterization of barley vernalization genes. *Plant Mol Biol* 59:449–467.

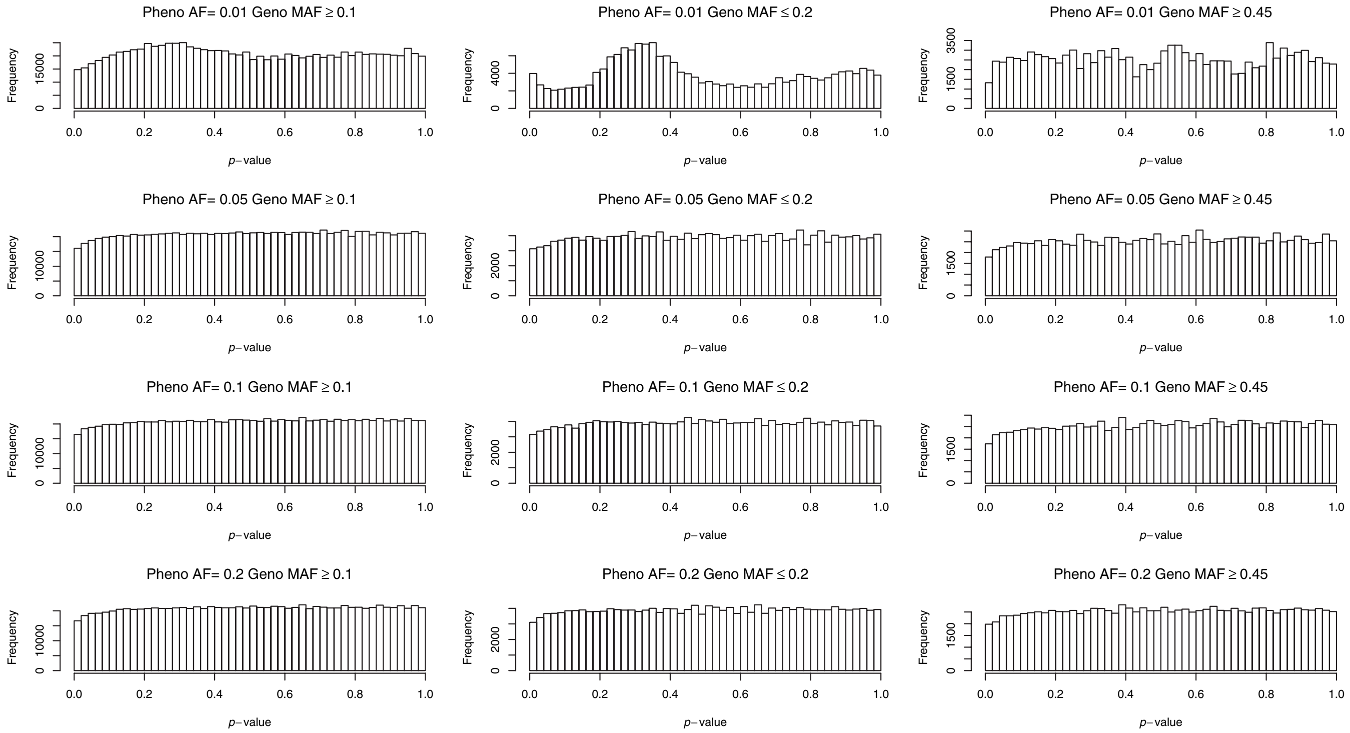

Fig. S1. Investigating the assumption of a normally distributed phenotype when using the mixed model. Illustrated are the observed *P* value distributions under the null hypothesis when association between binary phenotypes [allele frequency (AF) = 0.01, 0.05, 0.1, and 0.2] and marker genotypes [minor allele frequency (MAF) ≥ 0.1, 0.2, and 0.45] were estimated using the mixed model (based on 1,000 permutations). Note that, although the mixed model assumes a normally distributed phenotype, there is no excess of significant association when the phenotype is binomially distributed.

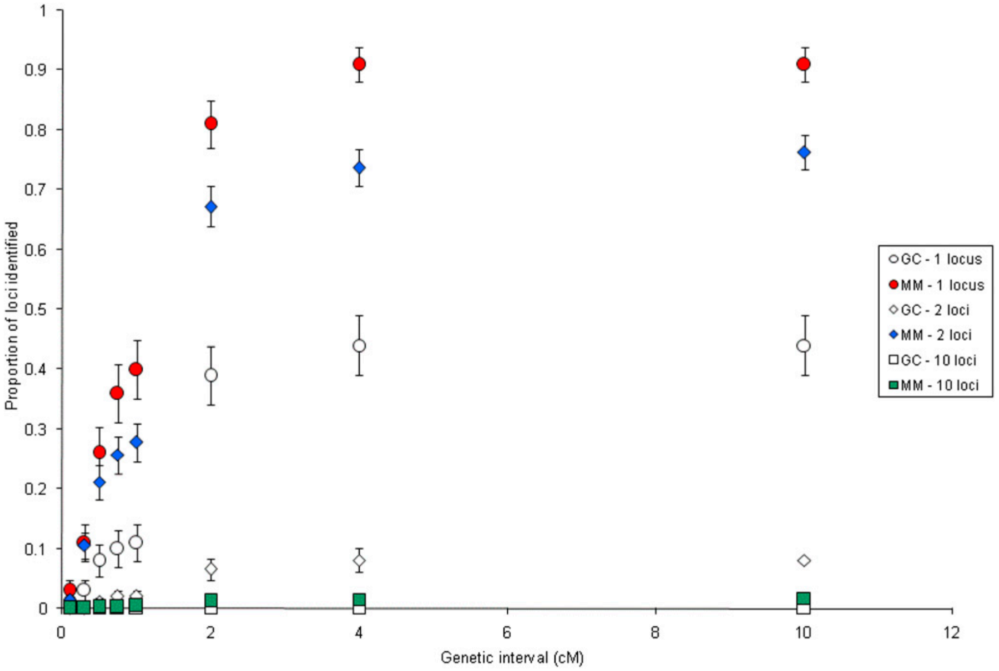

Fig. S2. Predicted experimental power to detect trait controlled by 1, 2, and 10 loci with heritability (h^2) = 0.5 over genetic distance (cM). Power is measured as the proportion of simulations in which at least one causative locus was detected (q value ≤ 0.1). Error bars denote ±1 SE.

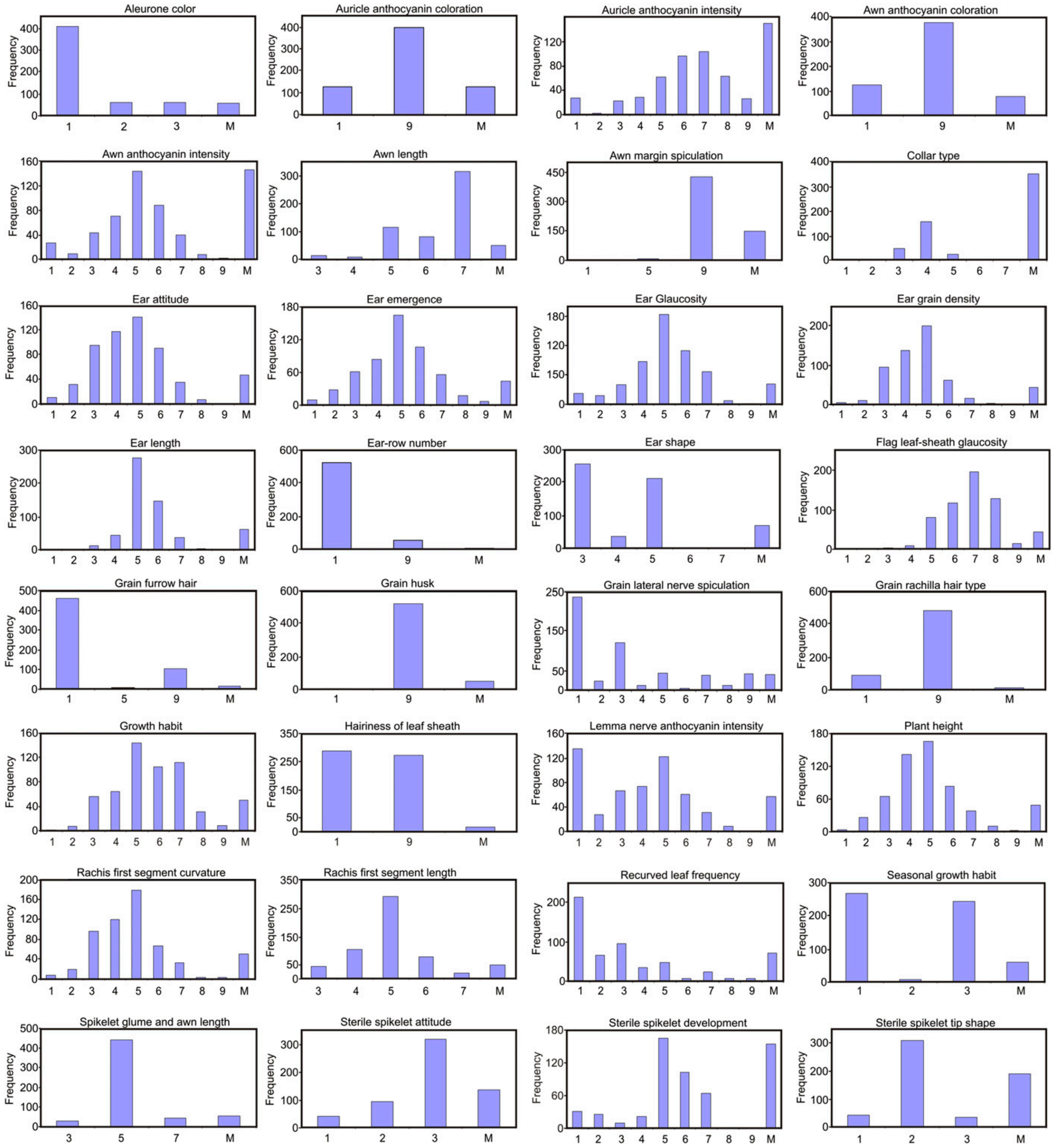

Fig. S3. Histograms of character fill for the 32 traits used for GWA analysis. The frequency of varieties that lack phenotypic data for any individual trait is sown in the missing (M) column.

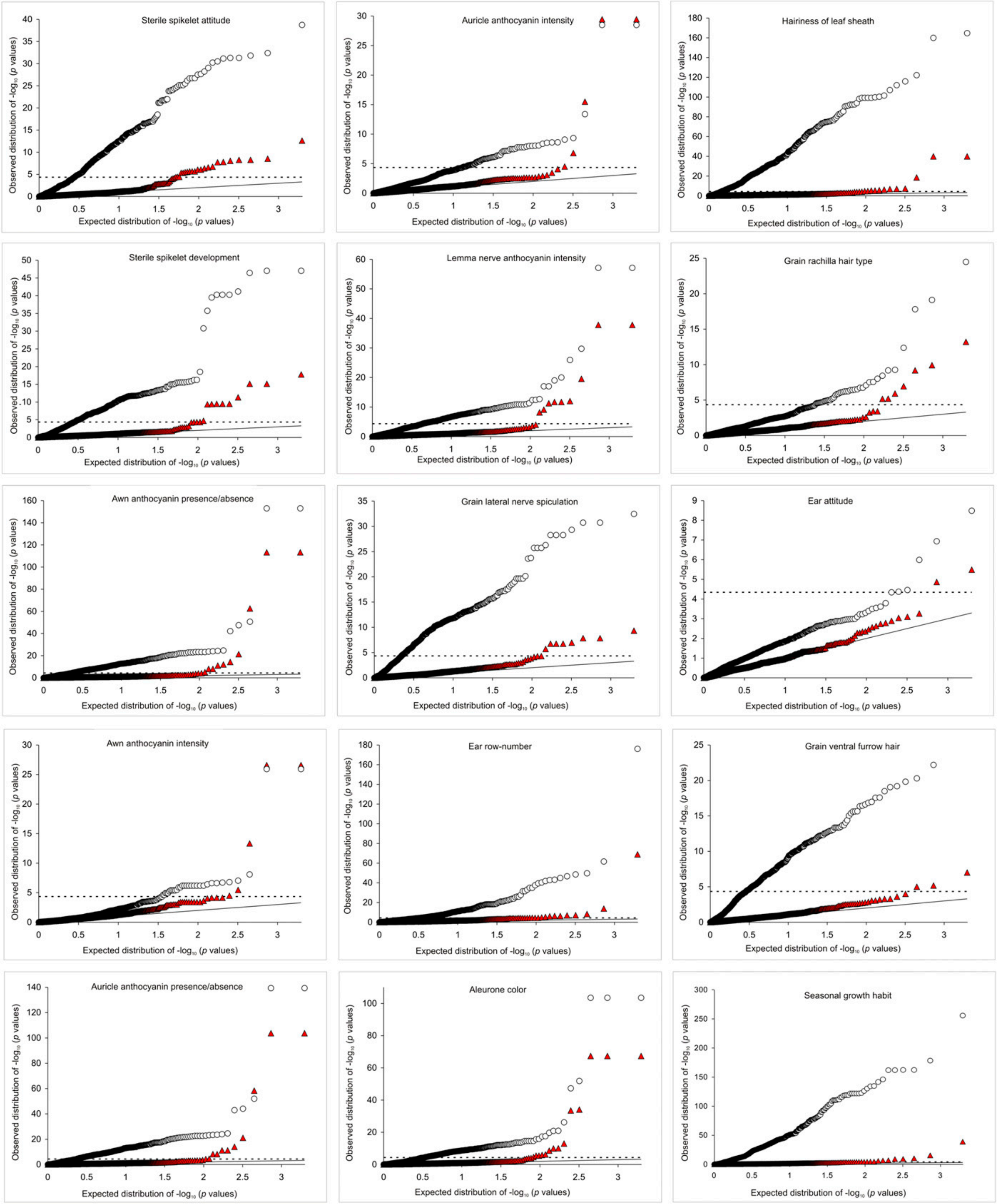

Fig. S4. Quantile–quantile plots for traits returning significant associations after GWA analysis. Expected vs. observed *P* values are plotted for naive (circles) and mixed model corrected (triangles) analyses. The $x = y$ line (solid) and Bonferroni corrected $P = 0.05$ significance thresholds (dashed line) are indicated.

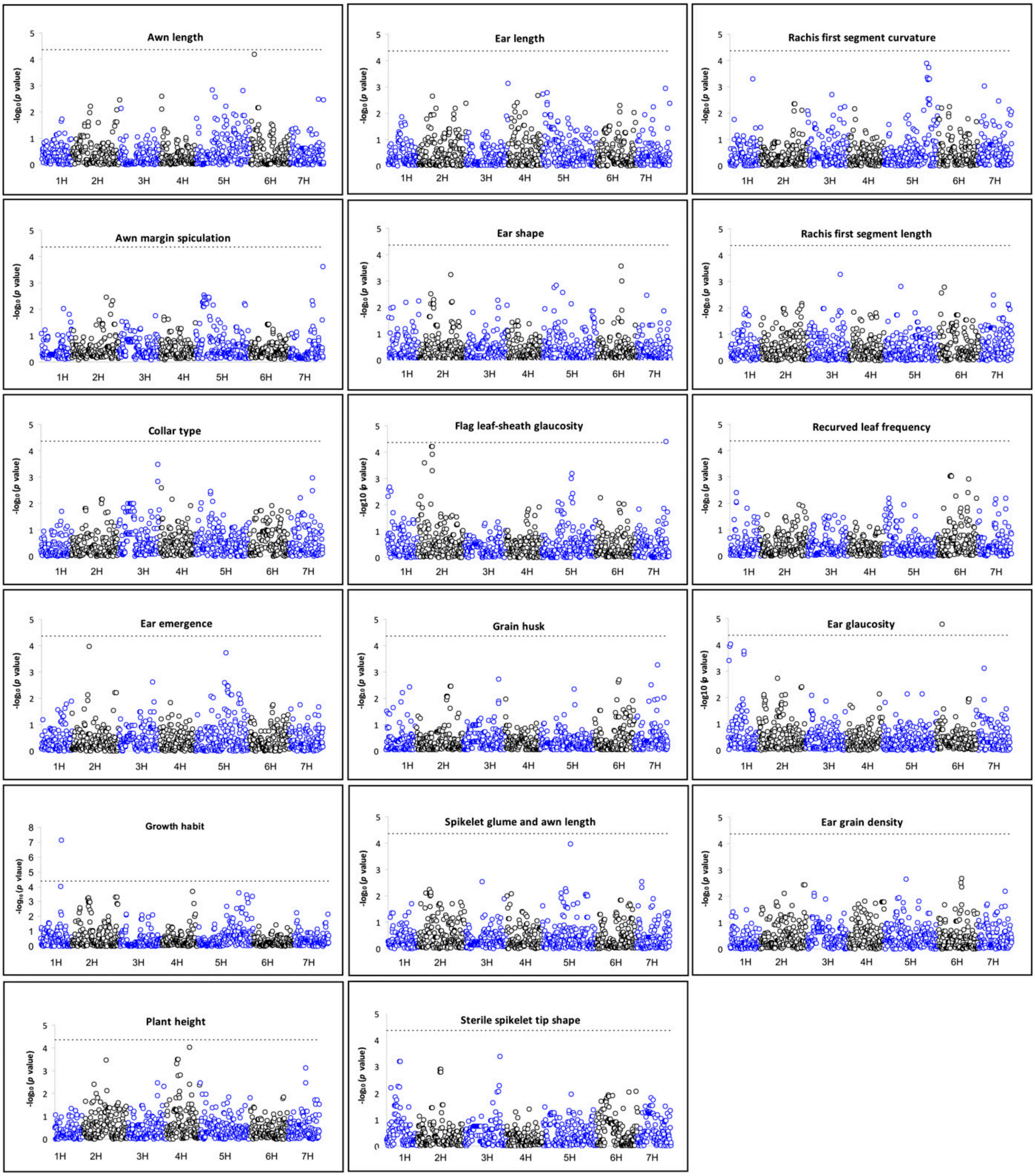

Fig. S5. GWA scans for the 18 traits that did not detect significant associations according to the discovery criteria used in this study ($-\log_{10} P \geq 4.35$; more than or equal to two significant markers within a 4-cM window).

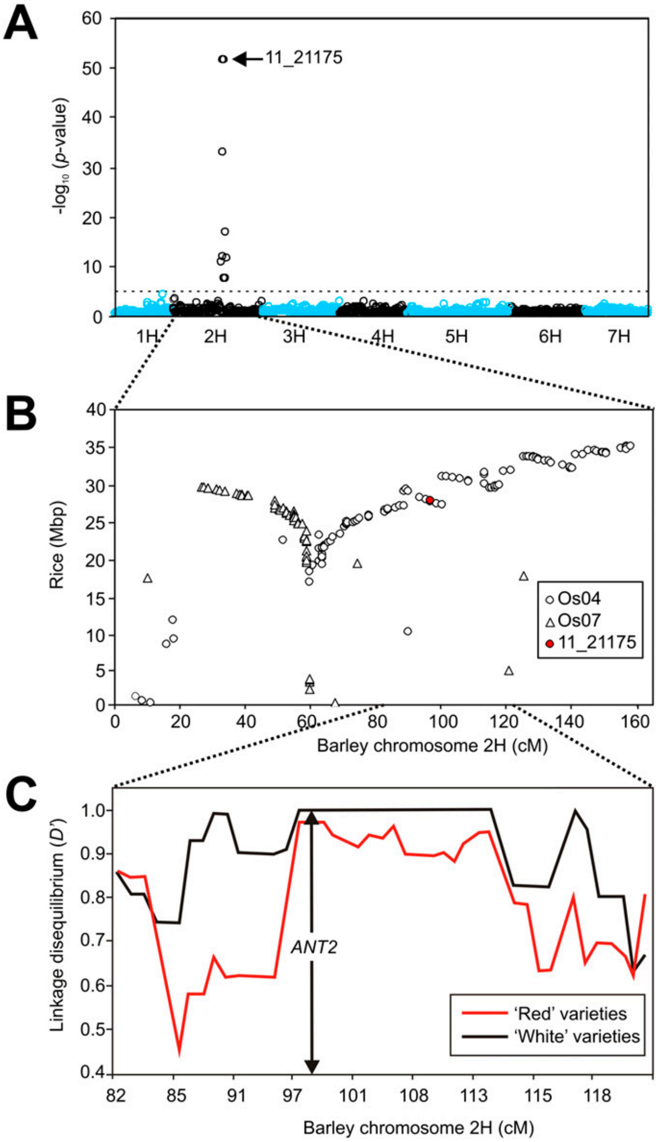

Fig. S6. Region of chromosome 2H associated with anthocyanin expression per se. (*A*) GWA scan. The Bonferroni corrected *P* = 0.05 significance threshold is indicated. (*B*) Colinearity between the genetic map of barley chromosome 2H and the physical maps of rice chromosomes 4 and 7. (*C*) Sliding window analysis of LD (measured as *D*′) around the *ANT2* locus in white and red cultivars.

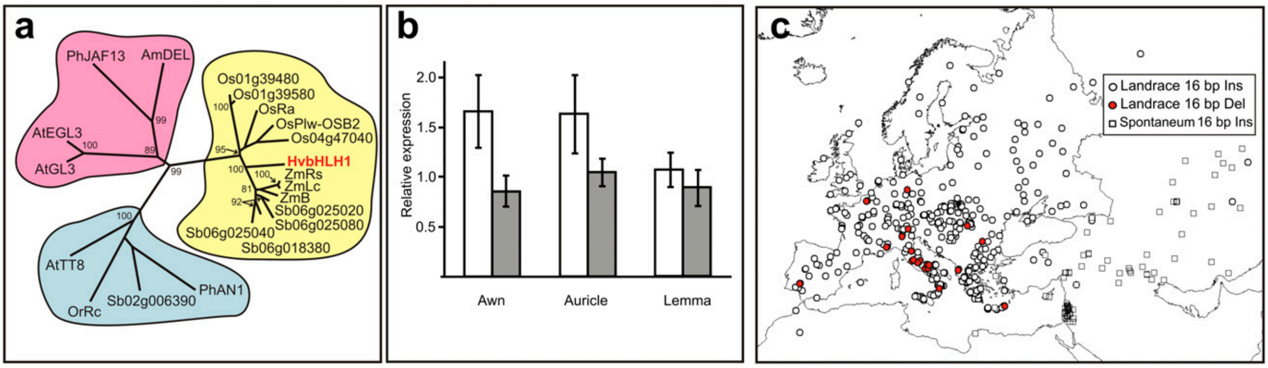

Fig. S7. *HvbHLH1* phylogeny, expression, and geography. (*A*) Unrooted phylogenetic analysis of HvbHLH1 and additional bHLH proteins from rice, sorghum, and maize as well as proteins known to act within the anthocyanin pathways of dicotenous species. Bootstrap frequencies (1,000 replicates) above 80% are indicated. Am, *Antirrhinum major*; At, *Arabidopsis thaliana*; Bd, *B. distachyon*; Hv, *H. vulgare* spp. *vulgare*; Or, *O. rufipogon*; Os, *O. sativa*; Ph, *Petunia* x *hybrid*; Zm, *Zea mays*. (*B*) Semiquantitative RT-PCR analysis of *HvbHLH1* expression in tissues from the white variety Saffron and the red variety Retriever normalized against *ACTIN.* Error bars indicate ±1 SD. (*C*) Geographic distribution of the *HvbHLH1* 16-bp InDel in 647 landrace and *H. vulgare* ssp. *spontaneum* accessions.

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)
Table S3 (DOCX)
Table S4 (DOCX)
Table S5 (DOCX)
Table S6 (DOCX)