Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars

Colin R. Cavanagh[a], Shiaoman Chao[b], Shichen Wang[c], Bevan Emma Huang[d], Stuart Stephen[a], Seifollah Kiani[c], Kerrie Forrest[e], Cyrille Saintenac[c], Gina L. Brown-Guedira[f], Alina Akhunova[c], Deven See[g], Guihua Bai[h], Michael Pumphrey[i], Luxmi Tomar[j], Debbie Wong[e], Stephan Kong[e], Matthew Reynolds[k], Marta Lopez da Silva[k], Harold Bockelman[l], Luther Talbert[m], James A. Anderson[n], Susanne Dreisigacker[k], Stephen Baenziger[o], Arron Carter[i], Viktor Korzun[p], Peter Laurent Morrell[n], Jorge Dubcovsky[q,r], Matthew K. Morell[a], Mark E. Sorrells[s], Matthew J. Hayden[e,1], and Eduard Akhunov[c,1]

[a]Commonwealth Scientific and Industrial Research Organization, Plant Industry and Food Futures National Research Flagship, Canberra, ACT 2601, Australia; [b]US Department of Agriculture–Agricultural Research Service Biosciences Research Laboratory, Fargo, ND 58102; [c]Department of Plant Pathology, Kansas State University, Manhattan, KS 66506; [d]Food Futures National Research Flagship and Mathematics, Informatics and Statistics Division, Commonwealth Scientific and Industrial Research Organization, Dutton Park, QLD 4102, Australia; [e]Department of Primary Industries Victoria, Victorian AgriBiosciences Center, Bundoora, VIC 3083, Australia; [f]US Department of Agriculture–Agricultural Research Service Eastern Regional Small Grains Genotyping Laboratory, Raleigh, NC 27695; [g]US Department of Agriculture Western Regional Small Grains Genotyping Laboratory, Pullman, WA 99164; [h]US Department of Agriculture–Agricultural Research Service, Manhattan, KS 66506; [i]Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164; [j]Department of Molecular Biology and Biotechnology, Chaudhary Charan Singh Haryana Agricultural University, Hisar 125004, India; [k]Genetic Resources and Enhancement Unit, Centro Internacional de Mejoramiento de Maíz y Trigo, Mexico City, 06600, Mexico; [l]US Department of Agriculture–Agricultural Research Service, Aberdeen, ID 83210; [m]Department of Plant Sciences, Montana State University, Bozeman, MT 59717; [n]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108; [o]Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68588; [p]KWS Lochow GmbH, 37555 Einbeck, Germany; [q]Department of Plant Sciences, University of California, Davis, CA 95616; [r]Howard Hughes Medical Institute, Chevy Chase, MD 20815; and [s]Plant Breeding and Genetics, Cornell University, Ithaca, NY 14853

Domesticated crops experience strong human-mediated selection aimed at developing high-yielding varieties adapted to local conditions. To detect regions of the wheat genome subject to selection during improvement, we developed a high-through-put array to interrogate 9,000 gene-associated single-nucleotide polymorphisms (SNP) in a worldwide sample of 2,994 accessions of hexaploid wheat including landraces and modern cultivars. Using a SNP-based diversity map we characterized the impact of crop improvement on genomic and geographic patterns of genetic diversity. We found evidence of a small population bottleneck and extensive use of ancestral variation often trace-able to founders of cultivars from diverse geographic regions. Analyzing genetic differentiation among populations and the extent of haplotype sharing, we identified allelic variants subjected to selection during improvement. Selective sweeps were found around genes involved in the regulation of flowering time and phenology. An introgression of a wild relative-derived gene conferring resistance to a fungal pathogen was detected by haplotype-based analysis. Comparing selective sweeps identified in different populations, we show that selection likely acts on distinct targets or multiple functionally equivalent alleles in different portions of the geographic range of wheat. The majority of the selected alleles were present at low frequency in local populations, suggesting either weak selection pressure or temporal variation in the targets of directional selection during breeding probably associated with changing agricultural practices or environmental conditions. The developed SNP chip and map of genetic variation provide a resource for advancing wheat breeding and supporting future population genomic and genome-wide association studies in wheat.

SNP genotyping | polyploid wheat | selection scans | wheat improvement | breeding history

Since its origin around 8,000 BC (1), hexaploid bread wheat (*Triticum aestivum* ssp. *aestivum*) has been subject to intense selection aimed at developing improved, high-yielding varieties that are adapted to diverse environmental conditions and agricultural practices (2). This process has included modification of vernalization and photoperiod requirements (3), adaptation to water-limiting conditions (4), low temperature (5), salt (6), and soil toxicity (7, 8). Development of semidwarf photoperiod-insensitive wheat varieties in the 1940s formed the basis of the "Green Revolution" (9).

The detection of loci under selection during crop improvement can contribute to more targeted breeding efforts and the opportunity to improve genomic selection models (10). Although traditional phenotype-to-genotype approaches using mapping or association mapping populations are useful for identifying large-effect trait loci (11–13), they are limited to phenotypes that are readily measured and may fail to detect a large portion of the genetic changes associated with plant domestication and improvement (10). Because selection has localized effects on molecular variation in populations (14), methods based on detecting genomic regions with patterns of variation that differ from the genome-wide average can identify loci subject to selection (15, 16). The advantage of these population genetics approaches is that they do not require identifying adaptive phenotypes a priori. Recently, genome-wide scans based on patterns of linkage disequilibrium (LD) as well as genetic differentiation between populations have been successfully applied to detect targets of selection in plants species by surveying both natural populations (16) and cultivated species (17–19).

Although multiple studies have reported the genetic basis of individual phenotypes associated with wheat improvement and

AGRICULTURAL SCIENCES

adaptation (9, 12, 20–24) or sought to characterize the structure of genetic variation in regional populations (25–28), the genome-scale impact of selection for improvement on the patterns of genetic variation in wheat remains largely unknown. Here we performed single-nucleotide polymorphism (SNP) discovery in the wheat transcriptome and developed a high-throughput SNP genotyping array. The array was used to construct a high-density wheat SNP map and to assess genetic variation in the coding region of 2,994 wheat accessions composing a broad geographical sample of landraces and wheat cultivars adapted to spring and winter growth conditions. Cultivars were selected to represent the major breeding programs of Australia, eastern and western Asia, Europe, and North America. The SNP distribution among the populations of cultivars and landraces was investigated to understand the impact of crop improvement on the structure of genetic diversity in wheat and to identify likely targets of selection.

Results

SNP Discovery. To maximize the utility of the genotyping assay and reduce the effect of ascertainment bias, SNP discovery was performed in a diverse sample of cultivars (Dataset S1; Fig. 1; Fig. S1). Roche 454 sequence reads from nine wheat accessions originating from Australia, the United States, and Mexico were assembled into 477,291 reference transcripts (RTs) and used to identify 25,454 SNPs with a validation rate of 85–90% (Dataset S1 and SI Methods). The distribution of alleles at the SNP sites was assessed using deep-coverage Illumina sequence data generated from a sample of 20 diverse wheat cultivars from Australia, China, Mexico, and the United States, 3 of which overlapped with the panel of nine accessions used for 454 sequencing (Dataset S1 and Fig. S2). In addition, 655 SNPs discovered in a diverse panel of wheat landraces (27) and SNPs identified from a sequence capture assay of 3,500 genes in the parents of the SynOp mapping population (29) were selected for assay design. These datasets were used to select 9,000 SNPs for Illumina iSelect SNP assay design (SI Methods).

Genotyping with 9K iSelect Beadchip Assay. Of the 9,000 attempted SNP assays on the iSelect genotyping array, 8,632 were functional, representing a 96% conversion rate (SI Methods). A total of 7,733 SNPs were genotyped across a diverse panel of 2,994 hexaploid wheat accessions. The remaining loci did not produce clear clustering patterns or were monomorphic. Manual data curation was required for accurate genotype calling of 65% of the SNPs (Dataset S2 and SI Methods). Of the 7,733 SNPs, 49% permitted genotype calls in heterozygous hexaploid individuals, whereas the remaining SNPs produced closely spaced clusters suitable for genotyping homozygous plants only. The majority of

SNPs behaved as biallelic markers; 338 (4.3%) of the SNP assays revealed segregation at more than one locus (Fig. S3).

Consensus Wheat SNP Map. The consensus map was built using the genotypic data from seven mapping populations including six biparental populations and one four-parent Multiparent Advanced Generation InterCross (MAGIC) population (20) (Dataset S3). In total, 7,504 polymorphic loci were positioned in the consensus map. The average SNP density across chromosomes was 1.9 ± 1.0 SNP/cM. The polymorphic loci were genotyped by 7,160 assays; 6,822 loci mapped to a single chromosome, 332 mapped to two positions, and 6 mapped to three chromosomal positions across the mapping populations. The number of SNPs mapped was similar for the A and B genomes (3,469 and 3,415, respectively) and lowest (by about fivefold) for the D genome (620 SNPs). Among 3,588 markers on the consensus map for which the linkage group was unambiguous among the contributing individual maps, 436 markers showed evidence for segregation distortion. The majority (222) of these markers were on chromosome 2B, where distortion is known to occur in the region around the *Sr36* locus, introgression from wild relative (20). Of the remaining 214 markers, 145 were mapped to chromosome 1A, with only small regions showing distortion on chromosomes 4B, 5A, 5B, 6A, 6B, and 7A (Fig. 2B).

Distribution of SNP Diversity Among Populations. A total of 6,305 (74%) SNPs were included in the analysis of genetic diversity and population structure in a worldwide sample of 2,994 hexaploid wheat accessions (Dataset S4). Despite their broad geographic distribution, from 86% to 100% of the SNPs were polymorphic within the individual populations, suggesting that the assay was enriched for common SNPs. This result is consistent with the shift of minor allele frequency (MAF) in the landraces and cultivars toward alleles with MAF >0.2 (Fig. 3A). The observed MAF is the consequence of intentional bias in SNP selection, where common alleles were favored by choosing more broadly distributed SNPs (SI Methods).

The joint allele frequency distribution between landraces and cultivars showed a strong correlation for both the complete set of 6,305 SNPs ($r^2 = 0.78$) (Fig. 3B) and the subset of 655 SNPs ($r^2 = 0.74$) discovered in landraces, suggesting a small effect of the SNP discovery procedures on the relative estimates of allele frequencies in these populations. The proportion of rare SNPs (MAF <0.05) at nonsynonymous sites in landraces (0.127) was higher than that at synonymous sites (0.079, Fisher's exact test, $P = 8.2 \times 10^{-5}$). No fixed differences were found between the cultivars and landraces, and only 1.3% of variants were private to cultivars. The sample of 134 landraces used in our study captured nearly 99% of the alleles present in the analyzed set of 2,860
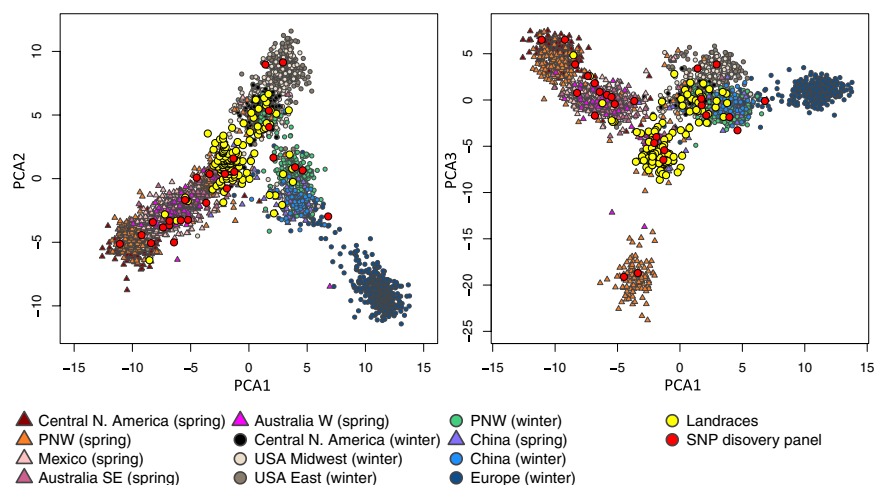

Fig. 1. Population structure of a worldwide collection of wheat accessions. The DAPC of populations of spring (triangles) and winter (circles) wheat from the Pacific Northwest of the United States (PNW), eastern United States, midwestern United States, central North America, China, Europe, and southeastern and western Australia. The SNP discovery panel and landraces are shown by red and yellow circles, respectively.

▲ Central N. America (spring)
▲ PNW (spring)
△ Mexico (spring)
▲ Australia SE (spring)
▲ Australia W (spring)
● Central N. America (winter)
○ USA Midwest (winter)
● USA East (winter)
● PNW (winter)
▲ China (spring)
● China (winter)
● Europe (winter)
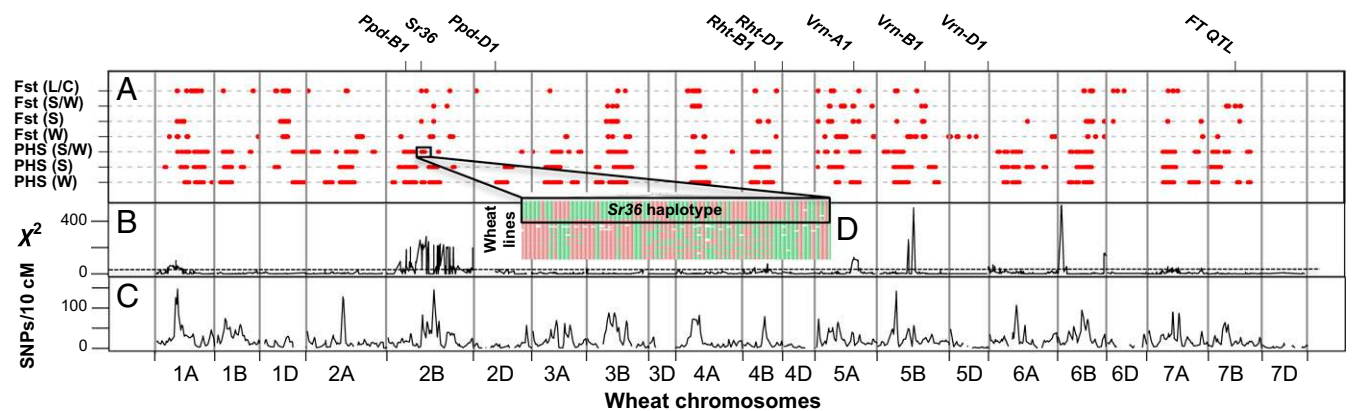● Landraces
● SNP discovery panel

Cavanagh et al.

Fig. 2. Wheat-genome selection scans. (*A*) Distribution of extreme F_{ST} and PHS values across the wheat genome. The map locations of *Ppd-B1*, *Ppd-D1*, *Sr36*, *Rht-B1*, *Rht-D1*, *Vrn-A1*, *Vrn-B1*, *Vrn-D1*, and *FT QTL* genes are shown on top. Chromosome boundaries are shown by vertical gray lines. F_{ST} was estimated between landraces and cultivars [F_{ST}(L/C)], between spring and winter wheat [F_{ST}(S/W)], and between populations within spring [F_{ST}(S)] and winter wheat [F_{ST}(W)]. (*B*) Distribution of single-locus χ^2 tests of segregation distortion across the wheat genome. The horizontal dashed line corresponds to the significance threshold $P \leq 1e^{-10}$. (*C*) Distribution of SNP density of across the wheat genome calculated for 10-cM sliding windows with 2-cM overlaps. (*D*) The graphical genotype of the *Sr36* genomic region with SNP alleles shown in red and green.

wheat cultivars. The overall genetic diversity of the cultivars (π = 0.36) was comparable to that of the landraces (π = 0.33) (Dataset S4). These results are consistent with an estimated limited (6%) reduction in population size accompanying the transition from landraces to cultivars (*SI Methods* and Fig. 4C). Taken together, our data suggest that most of the diversity present in the modern cultivars was also present in landraces.

Wheat Population Structure and Linkage Disequilibrium. The relationships among the wheat accessions inferred using discriminant analysis of principal components (DAPC) (30) were mostly similar for both the complete set of 6,305 SNPs (Fig. 1 and Fig. S1) and a subset of 655 SNPs (Fig. S4 *A and C*) discovered in a panel of landraces (27). The total amount of genetic variation explained by the first five principal components was 93.8%. Most populations were clearly separated into spring and winter wheat, which could be further subdivided into groups largely coinciding with the accessions' geographic origin (Fig. 1 and Fig. S4 *A and C*). Interestingly, despite the absence of European accessions in the SNP discovery panel, the European winter wheat population showed the strongest degree of genetic differentiation from the remaining populations, likely reflecting the use of genetically diverged founders (26). Compared with winter wheat, a higher level of admixture was detected in spring wheat populations. For example, clustering of spring wheat from the Pacific Northwest (PNW) of the United States and from Australia with varieties from Mexico or of spring and winter wheat accessions from China with the PNW winter wheat suggests the extensive use of lines sharing common ancestry in the development of these

varieties. The results of the DAPC were consistent with the results of model-based clustering analyses (Fig. S4A).

Although the significant proportion of landraces clustered separately from cultivars (Fig. 1 and Fig. S4B), we found landrace accessions grouping with populations from different geographic regions. For example, coclustering of Chinese cultivars and landraces can result from broad use of landraces in breeding programs. The hierarchical *F* statistics, estimated from all SNPs as the ratio of sums of variance components (31), showed that the proportion of genetic differentiation explained by geographic location (13.4%) was higher than that explained by growth habit (7.8%) or improvement level (2.9%). Similar estimates of variance components were obtained using the smaller subset of 655 SNPs discovered in landraces (*SI Methods*). These results suggest that the genetic composition of local populations can be shaped by alleles contributed by founding landraces, as well as by the divergence of these populations from the ancestral population of landraces.

Our data revealed a high level of heterogeneity in the extent of LD across the wheat genome with blocks of high-LD SNPs (r^2 > 0.75) separated by regions with high historic recombination rates (Fig. 4 *A* and *B*; *SI Methods*). LD between neighboring SNPs was higher in cultivars than in landraces (Wilcoxon signed-rank test, $P < 0.05$) (Fig. 4 *A* and *B*). Likewise, the estimate of shared haplotype length around every SNP (32) was lower in the 134 landraces (4.9 cM) than in a comparable random sample of cultivars (5.5 cM; *t* test, $P < 2.2 \times 10^{-16}$). These changes in LD are possibly caused by population bottleneck and selection during wheat improvement.

Evidence for Postdomestication Selection in the Wheat Genome. We used a comparison of genetic differentiation (F_{ST}) between populations and pair-wise haplotype sharing (PHS) in populations (16, 33) to identify genomic regions subject to selection. Because selection scan approaches based on F_{ST} and haplotype sharing are not strongly affected by ascertainment bias (34, 35), they are better suited for analyzing data generated using SNP chips.

The extent of genetic differentiation in a five-SNP window among the nine spring wheat populations [$F_{ST} \sim 0.15 \pm 0.02$ (SD)] and seven winter wheat populations [$F_{ST} \sim 0.15 \pm 0.02$ (SD)] were similar. We identified 21 regions in the spring wheat and 39 regions in the winter wheat that exceed the 0.173 and 0.175 F_{ST} thresholds ($P \leq 0.05$), respectively (Fig. 2A). Only two regions on the homeologous group 1 chromosomes were shared between these two scans.

With the worldwide sample of wheat accessions grouped according to their growth habit, the overall extent of genetic differentiation between spring and winter wheat [F_{ST}(S/W)] was small, with a mean
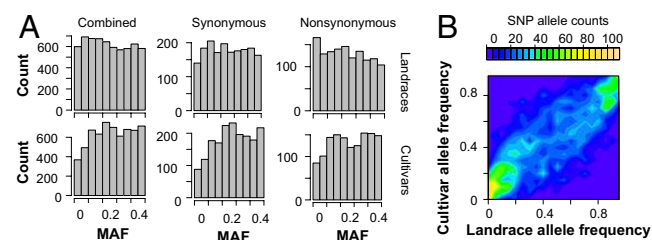

Fig. 3. Comparison of allele frequencies between cultivars and landraces. (*A*) Distribution of minor allele-frequency (MAF) counts in landraces and cultivars at combined, synonymous, and nonsynonymous SNP loci. (*B*) Distribution of joint allele-frequency density between cultivars and landraces.
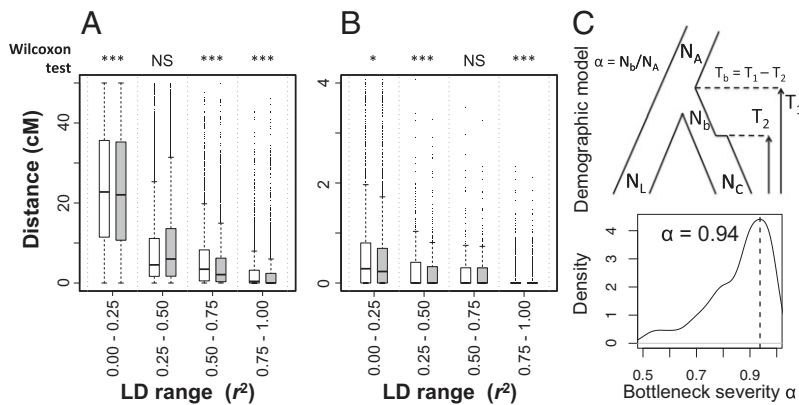
Fig. 4. Impact of improvement on LD and effective population size. (*A* and *B*) Boxplots showing the interquartile range of genetic distances for all pair-wise comparisons of either linked (*A*) or neighboring (*B*) SNP pairs grouped on the basis of the extent of LD (r^2). Open and gray boxplots correspond to cultivars and landraces, respectively. Wilcoxon signed-rank test: NS, nonsignificant; *$P \leq 0.05$; ***$P \leq 0.001$. (*C*) Demographic model used to assess the size of the improvement bottleneck (α). N_A, N_L, and N_C are population sizes of ancestral, landrace, and cultivar populations. The population of cultivars underwent a population bottleneck T_1 generations ago. The population remains at $\alpha * N_A$ size for $T_b = T_1 - T_2$ generations. The density plot shows the posterior distribution of α with the peak at 0.94. For details see SI Methods.

F_{ST} in a five-SNP sliding window of 0.091 ± 0.008 (SD). We identified 15 genomic regions that differentiated the spring and winter wheat (mean $F_{ST} > 0.097$) (Fig. 2*A*). Included in these genomic regions were SNPs flanking previously identified flowering-time QTL on chromosomes 5A and 5B (*Vrn-A1*, *Vrn-B1*) (21, 24) and 7B (22) (Fig. 2*A* and Dataset S5). No genetic differentiation was detected around the photoperiod regulation genes *Ppd-B1* (23), *Vrn-2* (36), and *Vrn-3* (37). However, the region around the *Ppd-B1* contained SNP wsnp_Ex_c66052_64232430 with $F_{ST} = 0.099$, which was above the specified window-based threshold.

To identify genomic regions selected during postdomestication wheat improvement, we assessed F_{ST} between landraces and cultivars (Fig. 2*A*) [F_{ST}(L/C)]. The mean F_{ST} in a sliding window was 0.079 ± 0.007 (SD) with a total of 32 regions showing extreme F_{ST} (>0.082). Among these regions we find evidence of strong differentiation around a major "green revolution" gene *Rht-B1* (12).

The extent of haplotype sharing as measured by PHS was assessed in the populations of 1,192 spring and 1,802 winter wheat accessions and in a mixed population including winter and spring wheat (Fig. 2*A*). Because the PHS test has increased power to detect selection of variants that have not reached fixation (17), only partial overlap was found among SNP variants identified in the 2.5% tail of the PHS statistic calculated for spring (136 variants), winter (136 variants), and the mixed spring/winter (133 variants) populations (Dataset S5). Among a total of 308 SNPs showing evidence of selection, 34 (11%) were shared between spring and winter wheat, 52 (17%) were shared between the winter wheat and mixed spring/winter wheat populations, and 27 (9%) were shared between the spring and mixed spring/winter wheat populations. We identified only 16 SNPs (5%) that were shared between all three populations. Selected alleles were found on 16 and 17 of the 21 wheat chromosomes in spring and winter wheat, respectively (Fig. 2*A*). Twenty percent (61/308) of the SNPs identified as PHS outliers were also F_{ST} outliers (Dataset S5). These 61 SNPs were mapped to 23 genomic regions distributed across 10 wheat chromosomes.

The strength and duration of selection can impact the frequency and distribution of selected alleles among individual populations. The SNP variants in the 2.5% tail of the PHS distribution were present at relatively low frequencies in individual wheat populations (Fig. 5). In the spring and winter wheat populations, from 40% to 80% and from 70% to 80%, respectively, of the SNP variants identified by the PHS scan had a MAF < 0.2. In the spring and winter wheat populations, 53 and 35 SNP variants from the PHS scan were present at a frequency >0.5 in at least one population, respectively. Some of these SNP alleles showed limited geographic distribution. For example, one of the alleles of SNP wsnp_BE499016B_Ta_2_1 located on chromosome 6B reached high frequency in the spring wheat cultivars from Mexico (Fig. 5). An allele of SNP wsnp_Ku_c28756_38667953 located near the *Rht-B1* gene responsible for dwarfism showed high frequency in winter wheat populations from North America (Fig. 5). These distinct patterns of geographic distribution of alleles

subjected to selection can potentially be linked with adaptation to local environmental conditions.

Genomic regions showing evidence of selection include a number of genes contributing to agronomically important phenotypes in wheat or other plants. The regions identified by the PHS scan included the *Rht-B1* locus, associated with wheat dwarfing phenotypes (12); *Ppd-B1* and *Vrn1*, variants that are associated with day-length insensitivity and flowering time in both wheat and barley (21–24); and the *Sr36* locus, associated with resistance to a fungal pathogen (20) (Fig. 2 *A* and *D*; Dataset S5). Three of these genes, *Rht-B1*, *Vrn-A1*, and *Vrn-B1*, also fell in genomic regions identified in the F_{ST} genetic differentiation scan.

Discussion

Patterns of Genetic Diversity and Population Structure. The high-throughput SNP genotyping array and a high-density SNP map developed in our study provided us with an unprecedented opportunity to gain insights into the impact of crop improvement on the genome-wide patterns of genetic variation and identify putative targets of selection in the wheat genome.

Relatively small differences in diversity observed between modern cultivars and landraces are consistent with a minor bottleneck during wheat improvement that resulted in only a 6% reduction in population size. This observation is similar to findings in maize,
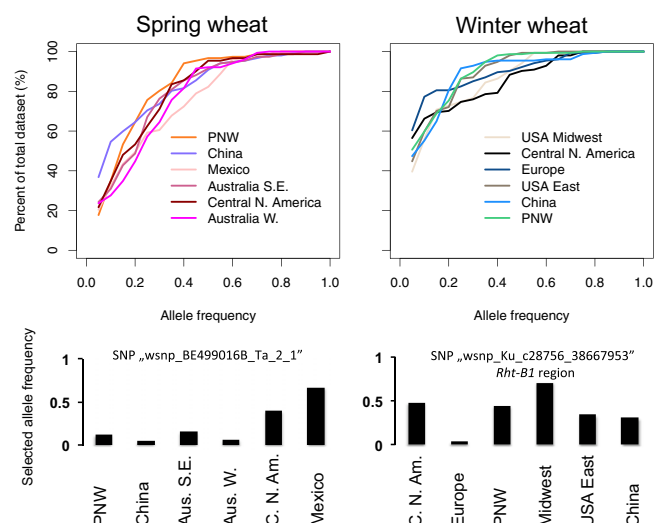

Fig. 5. Cumulative frequency of SNP alleles subjected to selection in wheat populations. Cumulative allele frequency distribution in the spring (*Left*) and winter (*Right*) wheat from different geographic regions was calculated for SNP alleles identified in selection scans. (*Lower*) Population-specific frequencies of the two SNP alleles showing evidence of selection.

Cavanagh et al.

which also show a minor effect of crop improvement on diversity (17, 38) and suggests the extensive use of landraces in the development of crop varieties. This is in contrast to domestication bottlenecks that result in significant changes in population size due to selection for alleles contributing to the domesticated phenotype (17) and the demographic effect linked with sampling from limited geographic areas (39, 40).

Low genetic differentiation between landraces and modern cultivars suggests that selection during wheat breeding, when we consider the total population, has not dramatically altered allele frequency genome-wide, but may have been accomplished by selection on a relatively limited number of loci. At the same time strong geographic differentiation among wheat populations found in the current and previously published studies (25–28) and the relatedness of landraces and cultivars suggest that the use of distinct founders as well as allele-frequency divergence from the ancestral population of landraces could have contributed to the development of regional breeding populations. Relatively few alleles were exclusive to to wheat cultivars. Although the introgression of favorable traits from wild relatives has been proposed as a potential path to wheat improvement (41), the rarity of exclusive alleles suggests that these efforts to date have not notably altered the genetic composition of elite cultivars.

Growth habit is one of the primary mechanisms driving local adaptation. Although the measured impact of growth habit on genetic differentiation is limited, spring and winter wheat could be distinguished by both genetic assignment and DAPC analyses. The relative genetic similarity between growth habits likely reflects the common practice of using lines from both groups in breeding spring and winter wheat cultivars and the complex architecture of flowering-time regulation in wheat, where spring growth habit can result from independent mutations in multiple genes (42). These conclusions are consistent with the limited overlap between strongly differentiated SNPs and the candidate flowering-time loci.

Selection Scans. We have identified a number of candidate selection targets associated with wheat improvement including regions containing genes involved in the regulation of flowering (21–24), development (12), and stress response (20). Although the biological function of many selection targets is unknown, the genomic resources developed in our study provide an opportunity for the identification of genes underlying wheat adaptation to diverse climatic conditions.

Because F_{ST} and PHS scans tend to identify loci at different stages of selection (15), partial (20%) but nonetheless substantial overlap in the loci identified by the two approaches was consistent with previous studies in *Arabidopsis* (15) and humans (43), which have demonstrated the dependence of genome-scan results on the approach applied to detect selection (44). Fewer loci were identified as outliers based on F_{ST} than on PHS, potentially reflecting limited differentiation in allele frequency in response to selection. Recent introgression of favorable alleles into breeding programs can generate admixture LD around the introgressed locus (20), which is readily detected by the PHS scan. In contrast to the results obtained in natural populations (15), human-driven selection in crops may have strong effects on both LD and genetic differentiation at a significant number of selection targets (17–19).

The limited overlap (25% for PHS and 5–10% for F_{ST}) between the targets of selection identified in the spring and winter wheat populations suggests that selection may occur on distinct loci. There are two plausible explanations for this observation, both of which may play a role in the differences in targets of selection identified in these populations. First, selection pressures are likely to vary temporally and within a heterogeneous environment. This is consistent with the low MAF for many apparent targets of selection and may reflect changing selective pressures related to pathogen pressure or climatic regime. Second, selection may be acting on multiple functionally equivalent mutations in different portions of the broad geographic range of wheat. Functionally equivalent mutations are most plausible under strong selective pressures with limited migration to promote the spread of favorable variants (45).

Conclusion. The targets of selection, identified as extended haplotypes with low MAF, have made only a small contribution to genetic differentiation among geographic regions. Our results suggest that regional adaptation likely resulted from the selection of multilocus genotypes by sampling among common variants rather than among few favorable variants (39) and support studies that suggest the quantitative nature of major agronomic traits (46). Although the potential of using existing common alleles for crop improvement still requires further investigation, crop breeding will likely benefit from the introduction of new allelic variation from distant relatives. The high-throughput SNP array and diversity map will provide a resource for the accelerated analysis of wheat genetic diversity, identification of genes targeted by selection, designing of high-power genome-wide association studies experiments and marker-assisted breeding and genomic selection.

Methods

Plant Material. For SNP discovery, the transcriptomes of 26 accessions of hexaploid wheat (Dataset S1) were sequenced using Roche 454 and Illumina (GAII$_x$ and HiSeq2000) next-generation sequencing. The consensus genetic map was developed using seven experimental populations: four-parent MAGIC population (20) and six biparental populations (Dataset S3). Ditelosomic lines for Chinese Spring wheat (47) were used to orient and assign the consensus genetic map linkage groups to wheat chromosomes.

RT Assembly and SNP Discovery. The strategy outlined in SI Methods was applied to assemble transcripts produced from homeologous and paralogous copies of genes. The assembly of RTs was performed using MIRA v.3.0 (48). SNP discovery was performed by aligning reads against the RT (SI Methods) followed by validation in a set of an additional 20 wheat cultivars (Dataset S1). SNP discovery was also performed by sequence capture of 3,500 genes (29) in the parents of the SynOp population (49). A set of 655 SNPs discovered in landraces was also included into our study (27). A total of 9,000 SNPs were selected based on their distribution across genome and frequency in the discovery population (SI Methods).

SNP Genotyping. Infinium iSelect SNP genotyping was performed on the BeadStation and iScan instruments according to the manufacturer's protocols (Illumina). SNP clustering and genotype calling were performed using GenomeStudio v2011.1 software (Illumina). A genotype calling algorithm was generated for bread wheat using an iterative process to account for observed shifts in SNP clusters caused by differences in the number of duplicated (homeologous and paralogous) gene copies detected between assays (SI Methods).

Consensus Map Construction. The MAGIC map was constructed with the R package mpMap (20). The linkage maps for the six biparental populations (Dataset S3) were created using the program MSTmap (50). Linkage groups were assigned to chromosomes based on the MAGIC map and the results of SNP genotyping of wheat ditelosomic lines. MapMerge was used to integrate the MAGIC map with maps from each of the biparental populations (51). The consensus map positions were scaled according to the average slope of the MAGIC and SynOp maps relative to the consensus map (SI Methods).

SNP Data Analysis. Basic summary statistics for each SNP (MAF, average pairwise diversity π) were calculated in the R package genetics and the libsequence C++ library (52). Population structure was inferred using the program Structure (53) and the R package adegenet V1.3–4 (30). To asses the impact of wheat improvement on genetic diversity, we simulated data using coalescent simulations implemented in the program ms (54). The data were simulated under a demographic model that was previously used to investigate the effect of a domestication bottleneck on crop diversity (39, 40). Simulation details are provided in SI Methods.

Pair-wise estimates of LD were obtained SNPs with a MAF ≥ 0.05 by measuring r^2 according to Weir (55). The average length of pair-wise shared haplotypes in populations (32) was calculated around each SNP for 134 accessions of landraces and random samples of 134 wheat cultivars.

F_{ST} and PHS Scans. Subpopulations used in the hierarchical F_{ST} analysis were partitioned on the basis of geographic location, crop improvement status, and growth habit. The variance components were estimated for each SNP and used to calculate the F statistics (31). For F_{ST} scan, the estimates of single-locus F_{ST} were calculated for each SNP locus using the BayeScan program (33). F_{ST} was estimated in populations grouped according to their geographic origin {within winter [F_{ST} (W)] and spring [F_{ST} (S)] wheat}, growth habit [between spring and winter wheat F_{ST} (S/W)], or improvement status [between landraces and cultivars F_{ST} (L/C)]. The Markov chain was run for 100,000 steps with a burn-in of 50,000 steps and a thinning interval of 10. For each SNP, population-specific F_{ST} estimates were averaged. Comparisons were based on mean F_{ST} in a sliding window of five SNPs (two-SNP overlap). A genomic region was considered to be an outlier if two consecutive windows showed mean F_{ST} above the 95th percentile generated through bootstrap resampling (1,000 times). The PHS statistic was calculated using

the approach of Toomajian et al. (16), which assesses the average length of shared haplotype blocks around a SNP position using pair-wise comparison of individuals within a population. SNP variants showing extreme PHS values were defined as those falling above the 97.5 percentile of the PHS distribution in a sliding window (from 0 to 1, step 0.01).

1. Willcox G (1997) Archaeobotanical evidence for the beginnings of agriculture in Southwest Asia. *The Origins of Agriculture and Crop Domestication*, eds Damania AB, Valkoun J, Willcox G, Qualset CO, [International Center for Agricultural Research in the Dry Areas, Aleppo (Syria); International Plant Genetic Resources Institute, Rome (Italy); Food and Agricultural Organization, Rome (Italy); Genetic Resources Action International, Barcelona (Spain)], pp 25–38.
2. Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316(5833):1862–1866.
3. Worland T, Snape JW (2001) Genetic basis of worldwide varietal improvement. *The World Wheat Book: A History of Wheat Breeding*, eds Bonjean AP, Angus WJ (Lavoisier Publishing, Paris) pp 59–100.
4. Reynolds M, Dreccer F, Trethowan R (2007) Drought-adaptive traits derived from wheat wild relatives and landraces. *J Exp Bot* 58(2):177–186.
5. Snape JW, et al. (2001) Mapping genes for flowering time and frost tolerance in cereals using precise genetic stocks. *Euphytica* 120(3):309–315.
6. Dubcovsky J, Santa-Maria GE, Epstein E, Luo MC, Dvorak J (1996) Mapping of the K+/Na+ discrimination locus Kna1 in wheat. *Theor Appl Genet* 92(3–4):448–454.
7. Sasaki T, et al. (2004) A wheat gene encoding an aluminum-activated malate transporter. *Plant J* 37(5):645–653.
8. Jefferies SP, et al. (2000) Mapping and validation of chromosome regions conferring boron toxicity tolerance in wheat (*Triticum aestivum*). *Theor Appl Genet* 101(5–6):767–777.
9. Hedden P (2003) The genes of the Green Revolution. *Trends Genet* 19(1):5–9.
10. Morrell PL, Buckler ES, Ross-Ibarra J (2011) Crop genomics: Advances and applications. *Nat Rev Genet* 13(2):85–96.
11. Doebley J, Stec A, Gustus C (1995) teosinte branched1 and the origin of maize: Evidence for epistasis and the evolution of dominance. *Genetics* 141(1):333–346.
12. Peng J, et al. (1999) 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 400(6741):256–261.
13. Simons KJ, et al. (2006) Molecular characterization of the major wheat domestication gene Q. *Genetics* 172(1):547–555.
14. Cavalli-Sforza LL (1966) Population structure and human evolution. *Proc R Soc Lond B Biol Sci* 164(995):362–379.
15. Horton MW, et al. (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat Genet* 44(2):212–216.
16. Toomajian C, et al. (2006) A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. *PLoS Biol* 4(5):e137.
17. Hufford MB, et al. (2012) Comparative population genomics of maize domestication and improvement. *Nat Genet* 44(7):808–811.
18. Xu X, et al. (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat Biotechnol* 30(1):105–111.
19. Morris GP, et al. (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc Natl Acad Sci USA* 110(2):453–458.
20. Huang BE, et al. (2012) A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol J* 10(7):826–839.
21. Yan L, et al. (2003) Positional cloning of the wheat vernalization gene VRN1. *Proc Natl Acad Sci USA* 100(10):6263–6268.
22. Sourdille P, et al. (2000) Detection of QTLs for heading time and photoperiod response in wheat using a doubled-haploid population. *Genome* 43(3):487–494.
23. Börner A, Korzun V, Worland AJ (1998) Comparative genetic mapping of loci affecting plant height and development in cereals. *Euphytica* 100(1):245–248.
24. Szucs P, et al. (2006) Positional relationships between photoperiod response QTL and photoreceptor and vernalization genes in barley. *Theor Appl Genet* 112(7):1277–1285.
25. Zhang L, et al. (2011) Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers. *BMC Genet* 12:42.
26. White J, et al. (2008) The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. *Theor Appl Genet* 116(3):439–453.
27. Akhunov ED, et al. (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics* 11:702.
28. Chao S, et al. (2010) Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genomics* 11:727.
29. Saintenac C, Jiang D, Akhunov ED (2011) Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biol* 12(9):R88.
30. Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet* 11:94.
31. Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes* 5(1):184–186.
32. Mathews DJ, Kashuk C, Brightwell G, Eichler EE, Chakravarti A (2001) Sequence variation within the fragile X locus. *Genome Res* 11(8):1382–1391.
33. Foll M, Gaggiotti OE (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 180(2):977–993.
34. Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG (2007) Recent and ongoing selection in the human genome. *Nat Rev Genet* 8(11):857–868.
35. Albrechtsen A, Nielsen FC, Nielsen R (2010) Ascertainment biases in SNP chips affect measures of population divergence. *Mol Biol Evol* 27(11):2534–2547.
36. Yan L, et al. (2004) The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. *Science* 303(5664):1640–1644.
37. Yan L, et al. (2006) The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proc Natl Acad Sci USA* 103(51):19581–19586.
38. van Heerwaarden J, Hufford MB, Ross-Ibarra J (2012) Historical genomics of North American maize. *Proc Natl Acad Sci USA* 109(31):12420–12425.
39. Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998) Investigation of the bottleneck leading to the domestication of maize. *Proc Natl Acad Sci USA* 95(8):4441–4446.
40. Haudry A, et al. (2007) Grinding up wheat: A massive loss of nucleotide diversity since domestication. *Mol Biol Evol* 24(7):1506–1517.
41. Ortiz R, et al. (2007) High yield potential, shuttle breeding, genetic diversity, and a new international wheat improvement strategy. *Euphytica* 157(3):365–384.
42. Zhang XK, et al. (2008) Allelic variation at the vernalization genes *Vrn-A1*, *Vrn-B1*, *Vrn-D1*, and *Vrn-B3* in Chinese wheat cultivars and their association with growth habit. *Crop Sci* 48(2):458–470.
43. Pickrell JK, et al. (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* 19(5):826–837.
44. Akey JM (2009) Constructing genomic maps of positive selection in humans: Where do we go from here? *Genome Res* 19(5):711–722.
45. Ralph P, Coop G (2010) Parallel adaptation: One or many waves of advance of an advantageous allele? *Genetics* 186(2):647–668.
46. Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol* 147(3):969–977.
47. Kimber G, Sears ER (1968) Nomenclature for the description of aneuploids in the Triticinae. *Proceedings of Third International Wheat Genetics Symposium*, eds Findlay KW, Shepherd KW (Canberra, Australia), pp 468–473.
48. Chevreux B, et al. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res* 14(6):1147–1159.
49. Sorrells ME, et al. (2011) Reconstruction of the synthetic W7984 × Opata M85 wheat reference population. *Genome* 54(11):875–882.
50. Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet* 4(10):e1000212.
51. Wu Y, Close TJ, Lonardi S (2011) Accurate construction of consensus genetic maps via integer linear programming. *IEEE/ACM Trans Comput Biol Bioinform* 8(2):381–394.
52. Thornton K (2003) Libsequence: A C++ class library for evolutionary genetic analysis. *Bioinformatics* 19(17):2325–2327.
53. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure: Extensions to linked loci and correlated allele frequencies. *Genetics* 164(4):1567–1587.
54. Hudson RR (1990) Gene genealogies and the coalescent process, *Oxford Surveys in Evolutionary Biology*, eds Futuyma D, Antonovics J (Oxford Univ. Press, Oxford), vol. 7, pp. 1–43.
55. Weir BS (1996) *Genetic Data Analysis II* (Sinauer, Sunderland, MA).