Identification of the *VERNALIZATION 4* gene reveals the origin of spring growth habit in ancient wheats from South Asia

Nestor Kippes[a], Juan M. Debernardi[a], Hans A. Vasquez-Gross[a], Bala A. Akpinar[b], Hikment Budak[b], Kenji Kato[c], Shiaoman Chao[d], Eduard Akhunov[e], and Jorge Dubcovsky[a,f,1]

[a]Department of Plant Sciences, University of California, Davis, CA 95616; [b]Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli, Tuzla-Istanbul, 34956, Turkey; [c]Graduate School of Environmental and Life Science, Okayama University, Okayama, 700-8530, Japan; [d]Biosciences Research Lab, US Department of Agriculture–Agricultural Research Service, Fargo, ND 58102; [e]Department of Plant Pathology, Kansas State University, Manhattan, KS 66506; and [f]Howard Hughes Medical Institute, Chevy Chase, MD 20815

Contributed by Jorge Dubcovsky, August 3, 2015 (sent for review May 25, 2015; reviewed by Richard M. Amasino and Simon Griffiths)

Wheat varieties with a winter growth habit require long exposures to low temperatures (vernalization) to accelerate flowering. Natural variation in four vernalization genes regulating this requirement has favored wheat adaptation to different environments. The first three genes (*VRN1*–*VRN3*) have been cloned and characterized before. Here we show that the fourth gene, *VRN-D4*, originated by the insertion of a ∼290-kb region from chromosome arm 5AL into the proximal region of chromosome arm 5DS. The inserted 5AL region includes a copy of *VRN-A1* that carries distinctive mutations in its coding and regulatory regions. Three lines of evidence confirmed that this gene is *VRN-D4*: it cosegregated with *VRN-D4* in a high-density mapping population; it was expressed earlier than other *VRN1* genes in the absence of vernalization; and induced mutations in this gene resulted in delayed flowering. *VRN-D4* was found in most accessions of the ancient subspecies *Triticum aestivum* ssp. *sphaerococcum* from South Asia. This subspecies showed a significant reduction of genetic diversity and increased genetic differentiation in the centromeric region of chromosome 5D, suggesting that *VRN-D4* likely contributed to local adaptation and was favored by positive selection. Three adjacent SNPs in a regulatory region of the *VRN-D4* first intron disrupt the binding of *GLYCINE-RICH RNA-BINDING PROTEIN 2* (*TaGRP2*), a known repressor of *VRN1* expression. The same SNPs were identified in *VRN-A1* alleles previously associated with reduced vernalization requirement. These alleles can be used to modulate vernalization requirements and to develop wheat varieties better adapted to different or changing environments.

wheat | flowering | vernalization | VRN1 | *Triticum aestivum* ssp. *sphaerococcum*

Hexaploid wheat (*Triticum aestivum* L., genomes AABBDD) originated in the coastal area of the Caspian Sea ∼8,000 y ago from the hybridization of a tetraploid wheat (*Triticum turgidum* L., genomes AABB) and diploid *Aegilops tauschii* (genome DD) (1, 2). Its polyploid nature and rapidly changing genome facilitated wheat adaptation to a wide range of environments as it spread westward into Europe and eastward into Asia (3). Since then, wheat has become a major global crop contributing approximately one-fifth of the calories and proteins consumed by humans. Although wheat production in 2013 exceeded 700 million tons (Food and Agriculture Organization, faostat.fao.org), further increases are required to satisfy the food demands of a fast-growing human population. We hypothesize that a better understanding of the adaptative changes that occurred in wheat during the early expansion of agriculture can contribute to the development of new strategies to increase wheat productivity in changing environments.

Wheat adaptation to different latitudes and planting dates is mainly associated with natural variation in the photoperiod gene *PPD1* that promotes flowering under long days (4–6) and in the vernalization gene *VRN1* that modulates the requirement of long exposures to cold temperatures (vernalization) to induce flowering (7–9). The photoperiod and vernalization pathways converge at the regulation of *FT1* (= *VRN3* in wheat) (10), which encodes a mobile protein that travels from leaves to the shoot apical meristem (SAM) (11, 12). Once in the SAM, FT1 becomes part of a florigen activation complex that binds to the *VRN1* promoter, inducing its transcription (13–15). In common wheat, most genes exist in three copies (homeologs), which are designated using their respective genomes (e.g., *VRN-A1*, *VRN-B1*, and *VRN-D1*).

VRN1 is a MADS-box (MCM1/AGAMOUS/DEFICIENS/SRF) transcription factor homologous to the *Arabidopsis* meristem identity gene *APETALA 1* (7–9), and its activation by vernalization (16) or by FT1 (14) accelerates the transition of the SAM from the vegetative to the reproductive phase. Wheat varieties carrying the ancestral *VRN1* allele have a strong vernalization requirement and are classified as "winter wheats." Mutation in the *VRN1* promoter (17–19) or large deletions in the first intron (20) greatly reduce or eliminate the vernalization requirement, and wheat varieties carrying these alleles are classified as "spring wheats." A critical regulatory region in the *VRN1* first intron is transcribed and contains binding sites for the RNA-binding protein TaGRP2 (21). TaGRP2 is homologous to the *Arabidopsis GLYCINE-RICH RNA-BINDING PROTEIN7* (GRP7), which is a single-stranded

Significance

A precise regulation of flowering time is critical for plant reproductive success and for cereal crops to maximize grain production. In wheat, barley, and other temperate cereals, vernalization genes play an important role in the acceleration of reproductive development after long periods of low temperatures during the winter (vernalization). In this study, we identified *VERNALIZATION 4* (*VRN-D4*), a vernalization gene that was critical for the development of spring growth habit in the ancient wheats from South Asia. We show that mutations in regulatory regions of *VRN-D4* are shared with other *VRN-A1* alleles and can be used to modulate the vernalization response. These previously unknown alleles provide breeders new tools to engineer wheat varieties better adapted to different or changing environments.

Author contributions: N.K. and J.D. designed research; N.K., J.M.D., H.A.V.-G., B.A.A., H.B., K.K., S.C., and E.A. performed research; N.K., E.A., and J.D. analyzed data; and N.K., J.M.D., and J.D. wrote the paper.

The authors declare no conflict of interest.

Reviewers: R.M.A., University of Wisconsin-Madison; and S.G., John Innes Centre.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. KR422423, KR422424, KR119059–KR119063, and AC270085).

See Commentary on page 11991.

[1]To whom correspondence should be addressed. Email: jdubcovsky@ucdavis.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1514883112/-/DCSupplemental.

RNA-binding protein involved in the regulation of flowering (22). In the absence of cold, TaGRP2 binds to *VRN1* pre-mRNA and inhibits *VRN1* expression. During vernalization, TaGRP2 is gradually modified (O-GlcNAcylation), allowing the interaction with a carbohydrate-binding protein called VER2, which reduces TaGRP2-mediated repression of *VRN1* expression (21).

In wheat and other temperate grasses, *VRN1* is also expressed in the leaves where it acts as a repressor of *VRN2* (23). *VRN2* encodes a protein containing a putative zinc finger and a CCT protein–protein interaction domain (24) that acts as a long-day repressor of *FT1* to prevent flowering during the fall (25, 26). The induction of *VRN1* during winter prevents the up-regulation of *VRN2* when daylength increases during spring (23). In the absence of *VRN2*, *FT1* is up-regulated by long days, further increasing *VRN1* expression, and closing a positive feedback loop that leads to an irreversible acceleration of flowering (27).

A fourth vernalization locus, *VRN-D4*, has been mapped on the proximal region of the short arm of chromosome 5D (henceforth, 5DS) (28, 29), but the gene underlying this locus has not been identified yet. *VRN-D4* was first described in the Australian wheat variety Gabo (30), which likely inherited it from the Indian cultivar Muzaffarnagar (31). *VRN-D4* is found at higher frequency among wheat accessions from South Asia than from other parts of the world (32–34). *VRN-D4* was transferred from Gabo into Triple Dirk to develop Triple Dirk "F" (TDF), which is part of a set of isogenic lines carrying different vernalization genes (35, 36). Segregating populations generated from these isogenic lines revealed that *VRN-D4* has strong epistatic interactions with the other vernalization genes and is part of the same vernalization pathway (29).

The *VRN-D4* gene is located in a proximal region of 5DS that shows no recombination (29), limiting the use of map-based cloning to identify this gene. In this study we used alternative approaches to show that the *VRN-D4* locus originated from the insertion of a large 5AL chromosome segment including the *VRN-A1* gene into the proximal region of chromosome arm 5DS. We also show that this insertion is almost fixed in the ancient *T. aestivum* ssp. *sphaerococcum* (Perc.) Mac Key from South Asia and is associated with a reduction in diversity in the neighboring chromosome regions. Finally, we characterize natural variation in the regulatory regions of the first intron of *VRN-D4* and related *VRN-A1* alleles and demonstrate that they affect the binding of a repressor of *VRN1* expression.

Results

***VRN-D4* Is Linked to Early Expression of *VRN1*.** In a previous study, we showed that the presence of the *Vrn-D4* allele was associated with the up-regulation of *VRN1* and *FT1* and the down-regulation of *VRN2* transcript levels in the leaves (28). Based on this result and on the epistatic interactions among these genes, we suggested that *VRN-D4* likely operates upstream or as a part of the positive regulatory feedback loop connecting these three genes. To determine which of these genes is affected first by *VRN-D4*, we conducted a developmental time course analysis of gene expression under nonvernalizing conditions in isogenic lines with and without the dominant *Vm-D4* allele.

Triple Dirk "C" plants (TDC, winter growth habit, *vrn-D4*) exhibited low transcript levels of *VRN1* and *FT1* and high levels of

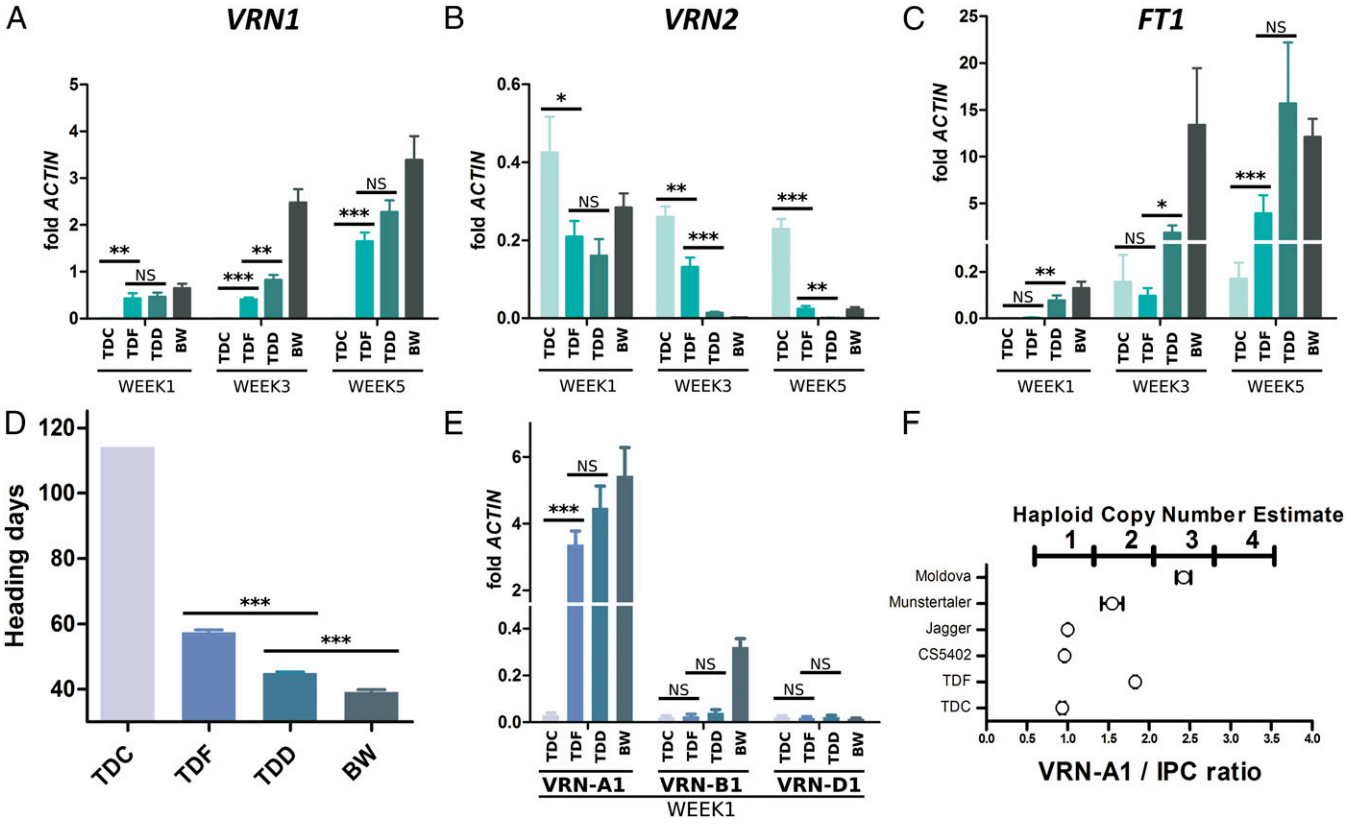

Fig. 1. Characterization of the TDF (*Vrn-D4*) line. Transcript levels of (*A*) *VRN-1* (all homeologs combined), (*B*) *VRN2*, and (*C*) *FT1*. (*D*) *VRN-D4* effect on heading time. The experiment was stopped 114 d after sowing. (*E*) Transcript levels of the three different *VRN1* homeologs. (*F*) *VRN-A1* haploid copy number. Jagger, Munstertaler, and Moldova were included as controls for *VRN-A1* copy number equal to 1, 2, and 3, respectively (37). (*A–E*) Means were calculated from eight biological replications. (*F*) Means were calculated from four biological replications. Error bars in all graphs represent SEM. TDC and CS5402: winter, no allele for spring growth habit. TDF: spring, *Vrn-D4*. TDD: spring, *Vrn-A1a*. BW (Bobwhite): spring, *Vrn-A1*, *Vrn-B1*, *Vrn-D1*. NS, nonsignificant; $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$.

VRN2 at the three tested sampling points (Fig. 1 *A*–*C*) confirming previous results (28). By contrast, Triple Dirk F plants (TDF, spring growth habit, *Vrn-D4*) showed increased expression of *VRN1* and *FT1* and decreasing levels of *VRN2* at week 5 (Fig. 1 *A*–*C*). At the earliest sampling point (week 1), the differences in *VRN1* between TDF and TDC were already highly significant ($P = 0.002$, 38-fold change, Fig. 1*A*), whereas the differences in *VRN2* were only marginally significant ($P = 0.05$, 2-fold change, Fig. 1*B*). Differences in *FT1* transcript levels between TDC and TDF were detected only in the latest sampling point (week 5, Fig. 1*C*). These results suggest that *VRN1* is the earliest target (direct or indirect) of *VRN-D4* among these three genes.

In the absence of vernalization, TDF flowered much earlier than TDC but still 12.6 d later than Triple Dirk "D" (TDD, strong *Vrn-A1a* allele) (17) and 18.3 d later than Bobwhite (BW, alleles for spring growth habit at the three *VRN1* homeologs) (Fig. 1*D*). The later flowering of TDF relative to TDD and Bobwhite was associated with lower transcript levels of *VRN1* and *FT1* and higher transcript levels of *VRN2* (Fig. 1 *A*–*C*, week 3).

Surprisingly, we found that *VRN-A1* was the only *VRN1* homeolog significantly up-regulated in TDF relative to TDC. *VRN-A1* transcript levels were higher from week 1, and the levels were similar in TDF and TDD (Fig. 1*E*). By contrast, neither *VRN-B1* nor *VRN-D1* transcripts were detected at the first (Fig. 1*E*) or third week (*SI Appendix*, Fig. S1*A*) in TDF or TDD. *VRN-B1* and *VRN-D1* were detected in Bobwhite (dominant *Vrn-A1 Vrn-B1 Vrn-D1*) after the first week (*SI Appendix*, Fig. S1). These results suggested that either the *VRN-A1* homeolog is a specific target of *VRN-D4* or that an additional copy of *VRN-A1* is present in TDF.

TDF Carries a Second Copy of *VRN-A1* Completely Linked to *VRN-D4*. Using a *VRN1* copy number Taqman assay described in previous studies (6, 37), we observed one copy of *VRN-A1* in winter lines TDC and CS5402 (recessive *vrn-D4* allele), but two copies of *VRN-A1* in TDF (Fig. 1*F*). When we sequenced the coding region of *VRN-A1* in TDF and Gabo (dominant *Vrn-D4* allele), we discovered a fixed A/C polymorphism at position 367 (henceforth, A367C, *SI Appendix*, Fig. S2*A*) that was not present in any of the *VRN1* sequences deposited in GenBank. Lines carrying the recessive *vrn-D4* allele showed only the canonical "A" at position 367, present in all available *VRN-A1* sequences (*SI Appendix*, Fig. S2*A*). The A367C SNP was completely linked with spring growth habit in a high-density mapping population (3,182 gametes) segregating for *VRN-D4* on chromosome arm 5DS (29). The six plants with the *VRN-D4* closest recombination events are shown in *SI Appendix*, Fig. S2*B*. Based on this result, we hypothesized that the duplicated copy of *VRN-A1* on chromosome arm 5DS was a good candidate for *VRN-D4*.

This hypothesis was further supported by the earlier expression of the *VRN-A1* gene inserted on chromosome arm 5DS transcript. In the leaf samples collected from 1- and 3-wk-old TDF plants, all of the transcripts carry the C variant at position 367 (*SI Appendix*, Fig. S3). Only by the fifth week, low expression of the A variant (*vrn-A1* allele) was detected in TDF together with the stronger expression of the C variant. The control lines showed the expected results: the winter TDC showed no *VRN1* expression, whereas the spring TDD and BW plants showed early expression of the A variant (*Vrn-A1a* allele from 5AL, *SI Appendix*, Fig. S3). Taken together, these results support the hypothesis that the *VRN-A1* copy in chromosome 5DS is *VRN-D4*.

Validation of the *VRN-D4* Candidate Gene Using Induced Mutations. To test the previous hypothesis, we mutagenized a population of 1,153 TDF lines with ethyl methane sulphonate (EMS) and screened them using a *Cel*I assay as described before (38). We identified four nonsynonymous mutations and one splicing site mutant in the *VRN1* copy in 5DS (*SI Appendix*, Method S1). Based on their predicted effect on protein function, we selected two mutations for phenotypic characterization (*SI Appendix*, Table S1). The first mutation generated a change from glutamic acid to lysine at position 158 (E158K, Fig. 2*A*). Although this is usually a tolerated amino acid change (BLOSUM 62 score = 1), it is located in a conserved position that suggests some evolutionary constraints (*SI Appendix*, Figs. S4 and S5). In a population segregating for the E158K mutation, plants heterozygous and homozygous for the mutation flowered 5.1 d ($P = 0.0004$) and 9.1 d ($P = 2.15e^{-5}$) later than plants carrying the wild-type allele, respectively. Plants carrying the wild-type allele flowered at the same time as the TDF control ($P = 0.76$, Fig. 2*B*).

The second mutation, found in homozygous state, eliminated the donor splicing site at the fourth exon (Fig. 2*A*). Translation of the adjacent intron sequence predicts a premature stop codon 8 bp downstream of the induced mutation that eliminates 40% of the VRN-D4 protein. The presence of this premature stop codon was confirmed in the transcripts from the TDF mutant (*SI Appendix*, Method S1). TDF control plants flowered on average 48 d after sowing, whereas plants homozygous for the splicing site mutation flowered on average 146 d later than TDF ($P < 5.04e^{-14}$) and at the same time of the winter TDC control ($P = 0.910$, Fig. 2*C*). These two independent mutations confirmed that the duplicated *VRN-A1* copy on the proximal region of 5DS is *VRN-D4*.

Physical Map of the *VRN-D4* Region. A BLAST search of the International Wheat Genome Sequencing Consortium (IWGSC) draft sequence of Chinese Spring (CS) chromosome arms (39) using *VRN-A1* as query revealed the presence of *VRN-D4* (carrying the diagnostic A367C SNP) in the 5DS database. Although CS does not carry *VRN-D4*, the CS ditelosomic 5DS line (henceforth, CS-dt5DS) used to sequence the 5DS arm was generated using the variety Gabo (40), which does carry *VRN-D4* (36, 41).

The physical map of the 5DS arm from CS-dt5DS, developed as part of the IWGSC, was used to determine the size of the inserted 5AL segment. A screen of the CS-dt5DS BAC library minimum tiling path (MTP) clones (*SI Appendix*, Method S2) (42) yielded BAC TaeCsp5DShA_0038_M14, which is part of overlapping contigs CTG87-CTG61 (*SI Appendix*, Table S2). We sequenced 12 BACs from these two contigs (*SI Appendix*, Method S2) and deposited the assembled ~680-kb sequence in GenBank (AC270085). The

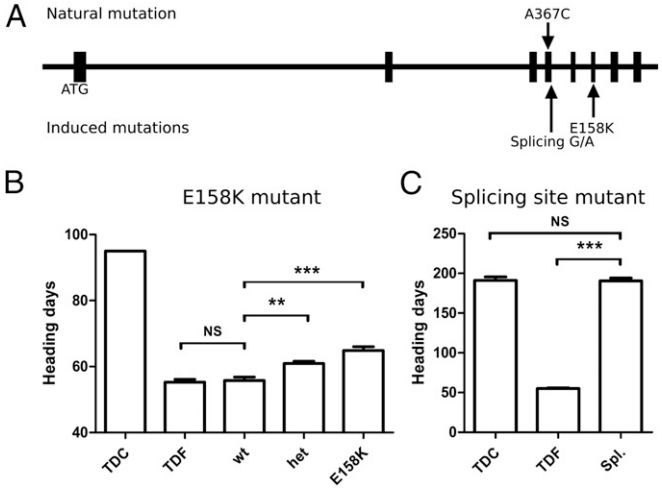

Fig. 2. Effect of *VRN-D4* TILLING mutants on flowering time. (*A*) *VRN-D4* gene structure; arrows indicate the location of natural SNP A367C, and TILLING mutants E158K and splicing site mutant (*SI Appendix*, Table S1). (*B*) Effect of the E158K mutation on heading time. WT, homozygous wild-type allele; het, heterozygous; E158K, homozygous mutant allele. (*C*) Effect of the splicing site mutation on heading time. Spl, homozygous splicing site mutation. Bars indicate means of at least six biological replications. Error bars indicate SEM. NS, nonsignificant; $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$.

comparison of the assembled sequence with databases for CS-5AL (39), *Triticum urartu* (diploid donor of the A genome) (43), and three databases for *Ae. tauschii* (D genome) (*SI Appendix*, Method S2) identified a ~290-kb region with high identity (>99%) to the A genome databases, flanked by sequences with high identity (>99%) to *Ae. tauschii* (Fig. 3*A*). This result indicates that the *VRN-D4* locus originated by the insertion of a large segment from chromosome arm 5AL into chromosome arm 5DS.

Annotation of the assembled region of CS-dt-5DS revealed the presence of three genes (AC270085). *VRN-D4* was the only gene in the 290-kb 5AL region. None of the genes known to flank *VRN-A1* in chromosome arm 5AL (7) were detected in the sequence of the inserted 5AL segment or by PCR in the DNA extracted from sorted 5DS chromosome arms from CS-dt-5DS (*SI Appendix*, Fig. S6). The two other genes were located in the D-genome region downstream of *VRN-D4* (*CYCLIN-A2-1* and hypothetical gene TRIUR3_08792). No genes were detected in the 225-kb sequence upstream of the 5AL insertion.

The first border of the 5DS/5AL insertion is located 88.3 kb upstream of the start codon of *VRN-D4* (henceforth "upstream insertion site," Fig. 3*B*) and the second border is located between 201 and 203 kb downstream of *VRN-D4* (henceforth "downstream insertion site," Fig. 3*B*). We developed diagnostic PCR markers for each border (Fig. 3*C* and *SI Appendix*, Method S3), that were perfectly correlated with the A367C SNP diagnostic for *VRN-D4* in the critical recombinants of the TDF × CS5402 fine mapping population (*SI Appendix*, Fig. S2*C*) and in all of the accessions described in *SI Appendix*, Tables S4 and S5. These results suggest that *VRN-D4* originated from the same 5DS/5AL insertion event.

We could not find a single *Ae. tauschii* contig matching both D-genome borders of the CS-dt-5DS sequence. *Ae. tauschii* contig 5265.1 showed >99% identity with a 104-kb region of CS-dt-5DS upstream of the 5AL insertion site (Fig. 3*B*). However, the additional 200 kb of this *Ae. tauschii* contig showed no similarity with the CS-dt-5DS D-genome sequence on the other side of the 5AL insertion. This result suggests that one of the original borders of the 5DS/5AL insertion was likely modified during or after the insertion event. This hypothesis was further supported by high levels of identity between different *Ae. tauschii* contigs and the CS-dt-5DS region downstream of the 290-kb insertion (Fig. 3*B*).

To identify the closest source of the 5AL segment inserted on chromosome arm 5DS, we compared the genomic sequence of

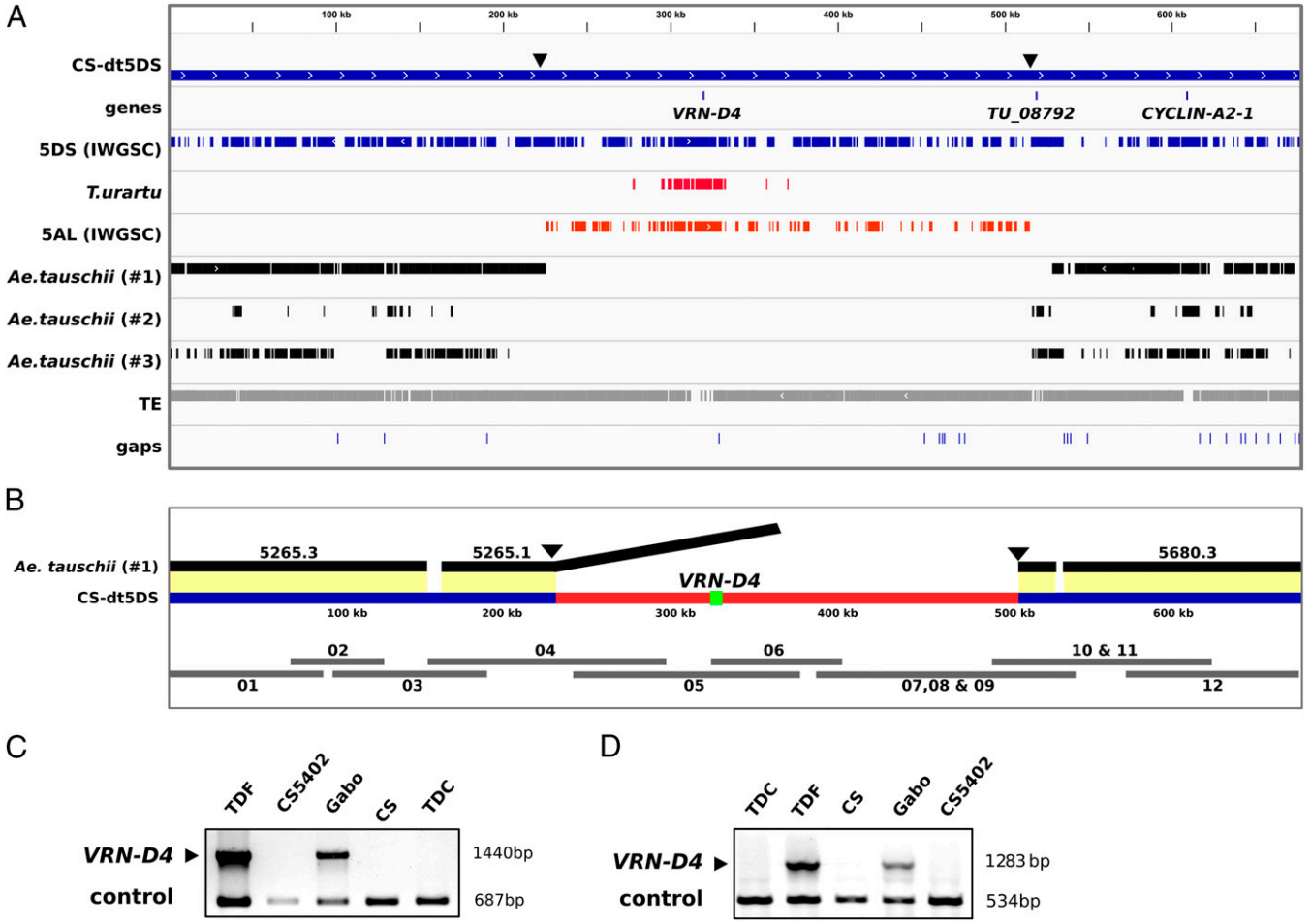

Fig. 3. Graphical representation of the *VRN-D4* region in chromosome 5DS. (*A*) CS-dt-5DS sequence assembly (~680 kb, AC270085) compared with 5DS and 5AL chromosome arm databases (IWGSC), *T. urartu*, and three different *Ae. tauschii* databases (*SI Appendix*, Method S2). Two genes [*VRN-D4* and *Cyclin A2-1* (TRIUR3_24995)] and one hypothetical gene TRIUR3_08792 were annotated. (*B*) Schematic representation of CS-dt-5DS sequence. Dark blue, CS-dt-5DS sequence; red, >99% identical to CS 5AL; black, *Ae. tauschii* contigs aegilops.wheat.ucdavis.edu/ATGSP/; gray rectangles, CS-dt-5DS BACs; yellow, regions >99% identical; and triangles, 5DS/5AL insertion point upstream and downstream of *VRN-D4*. (*C* and *D*) Molecular marker for the 5DS/5AL insertion sites upstream and downstream of *VRN-D4*, respectively. The 1,440-bp product indicates the presence of the 5DS/5AL upstream insertion and the 687-bp band is a PCR amplification control from BJ315664 (*SI Appendix*, Method S3). The 1,283-bp product indicates the presence of the 5DS/5AL downstream insertion and the 534-bp band is a PCR amplification control from BE606654 (*SI Appendix*, Method S3).

VRN-D4 with seven *VRN-A1* alleles from polyploid wheat (*SI Appendix*, Fig. S7). *VRN1* sequences form *Triticum monococcum* and *T. urartu* were included as outgroups. A neighbor joining tree showed that the closest sequence to *VRN-D4* was the *VRN-A1* allele from CS. The two sequences differed only in one SNP (A367C) in 14,912 bp. The *VRN-A1* sequences from Jagger and Claire were also very similar, showing only one and three additional SNPs in the analyzed region, respectively (*SI Appendix*, Fig. S7).

Characterization of the *VRN-D4* Allele for Spring Growth Habit. Although *VRN-D4* is expressed early in development in the absence of vernalization, its sequence lacks previously known mutations associated with spring growth habit (7, 17–20). Instead, *VRN-D4* has three close SNPs in the RIP-3 region of the first intron (*SI Appendix*, Fig. S8), which acts as a binding site for TaGRP2, a known negative regulator of *VRN1* (21) (Fig. 4*A*).

We found the canonical RIP-3 sequence in *VRN-B1* and *VRN-D1* in polyploid wheat and in *VRN-A^{m}1* in diploid wheat *T. monococcum* (*SI Appendix*, Fig. S8). By contrast, we detected two different RIP-3 haplotypes in the *VRN-A1* sequence (*SI Appendix*, Fig. S8). The first one includes a T-to-C mutation at position 2,783 from the start codon (henceforth, T2783C or RIP-3$^{VRN\text{-}A1}$) and was found in all 33 *T. urartu*, 13 *T. turgidum* ssp. *dicoccoides*, 13 *T. turgidum* ssp. *dicoccum*, and 12 *T. aestivum* sequenced accessions (*SI Appendix*, Table S3). The second haplotype, found in *VRN-D4* and the related *VRN-A1* sequences in CS, Jagger, and Claire (henceforth RIP-3$^{VRN\text{-}D4}$), included the previous T2783C SNP and two additional SNPs at positions 2,780 (G2780C) and 2,784 (C2784T). Interestingly, the *VRN-A1* alleles from Jagger and Claire have been previously linked to a weak vernalization requirement (6, 44).

To test if these SNPs at the RIP-3 region affect the binding of the TaGRP2 protein (*SI Appendix*, Fig. S9*A*), we performed RNA-electrophoretic mobility shift assays (REMSA, *SI Appendix*, Method S4). First, we tested a 27-nt RNA probe (*SI Appendix*, Fig. S9*B*) identical to the previously published RIP-3 fragment (21) and compared it to two 27-nt RNA probes containing the SNPs present in RIP-3$^{VRN\text{-}A1}$ and RIP-3$^{VRN\text{-}D4}$ (*SI Appendix*, Fig. S9*C*). TaGRP2 bound to the RIP-3 fragment in a dose-dependent manner (*SI Appendix*, Fig. S9*B*). The strength of the binding decreased in the presence of the T2783C mutation in RIP-3$^{VRN\text{-}A1}$ and was almost eliminated in the presence of the three mutations characteristic of the RIP-3$^{VRN\text{-}D4}$ allele (*SI Appendix*, Fig. S9*C*).

The predicted secondary structure of the 27-nt RIP-3 sequence was different from the one generated using longer RIP-3 flanking sequences (200 nt) and was affected by the SNPs in RIP-3$^{VRN\text{-}A1}$ and RIP-3$^{VRN\text{-}D4}$ (Fig. 4*B* and *SI Appendix*, Fig. S9*D*). To test if these differences have an effect on the differential binding of

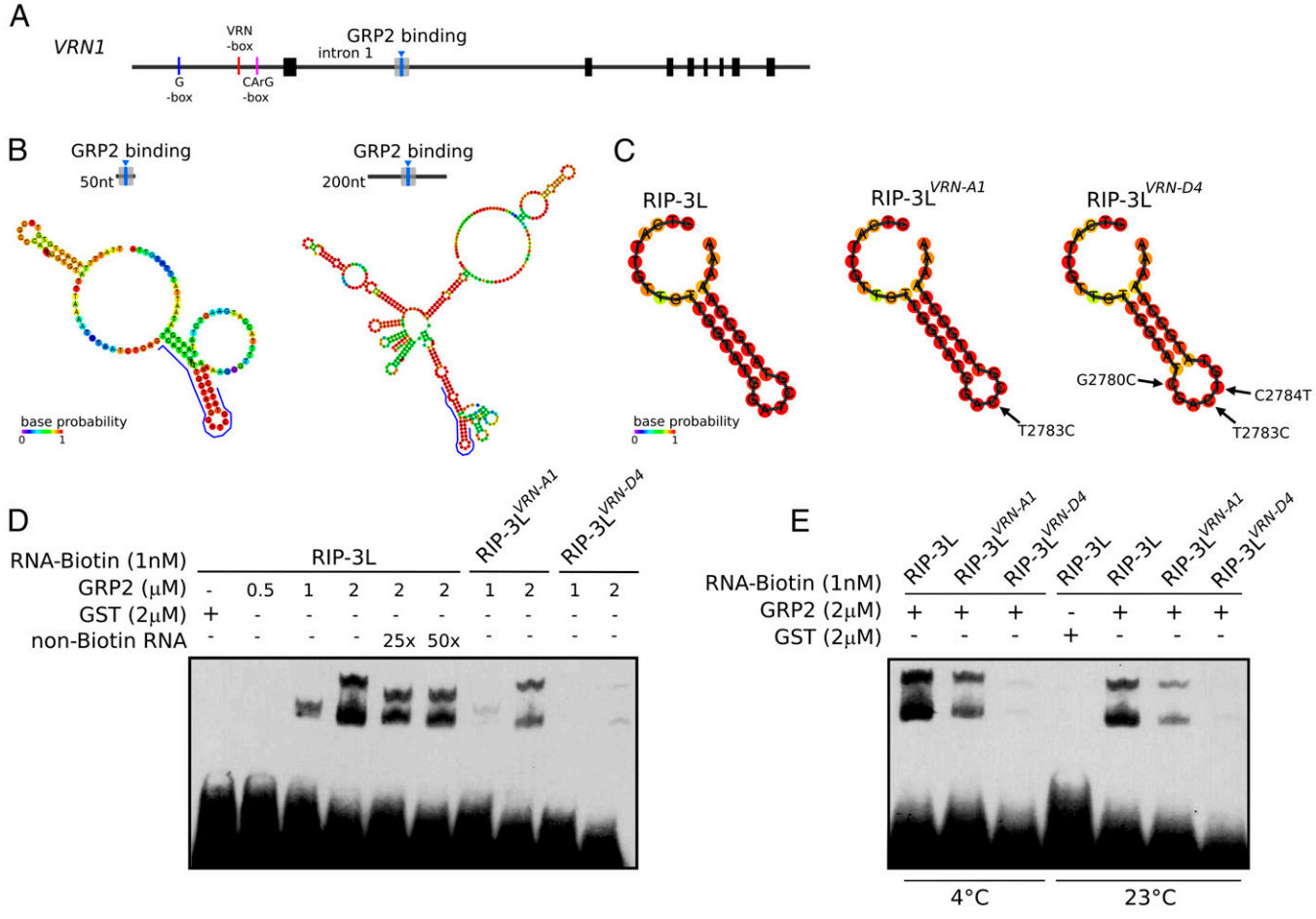

Fig. 4. Binding of GRP2 to RNA probes including the RIP-3 site. (*A*) GRP2 binding region in *VRN1* first intron. (*B*) Predicted RNA structure of intron one region including 50 nt or 200 nt at each side of the RIP-3 binding site (highlighted with a blue line). (*C*) Predicted RNA structure of the three 34-nt probes (L). RIP-3L, canonical sequence; RIP-3L$^{VRN\text{-}A1}$, SNP T2783C; and RIP-3L$^{VRN\text{-}D4}$, SNPs G2780C, T2783C, and C2784T. Arrows indicate the position of the SNPs. Nucleotide colors represent base pair probability for the predicted RNA secondary structure, with red representing the highest probability. (*D*) RNA-electrophoretic mobility shift assay (REMSA) showing the interaction of TaGRP2 with the three RNA probes. Interactions were tested with two different protein amounts (1: 1 μM and 2: 2 μM). Nonbiotinylated RIP-3L was used as competitor. (*E*) Same probes as in *D* were tested at two different binding temperatures (lines 1–3: 4 °C and lines 4–7: 23 °C).

TaGRP2, we developed longer RIP-3 RNA probes (34 nt, henceforth RIP-3L) that have similar structures to that predicted for RIP-3 with longer flanking regions (Fig. 4*B*). The three polymorphic SNPs are all located within a single-stranded RNA loop of the predicted RIP-3L structure and their presence does not affect the predicted structure of the probes (Fig. 4*C*). TaGRP2 showed a strong binding to the RIP-3L probe with the canonical sequence, a weaker binding to the RIP-3L$^{VRN\text{-}A1}$ probe with one SNP, and almost no binding to the RIP-3L$^{VRN\text{-}D4}$ probe with 3 SNPs (Fig. 4 *D* and *E*). A similar result was found when the experiment was repeated under low binding temperature (4 °C, Fig. 4*E*). In summary, the two probe lengths and two binding temperatures tested in the in vitro assays showed similar relative binding intensities for RIP-3, RIP-3$^{VRN\text{-}A1}$, and RIP-3$^{VRN\text{-}D4}$.

Geographical Distribution of *VRN-D4*. The two markers for the 5DS/5AL insertion and for the A367C SNP are diagnostic for the presence of *VRN-D4* and were used to study the distribution of *VRN-D4* in hexaploid wheat. The RIP-3$^{VRN\text{-}D4}$ haplotype is also informative, but its presence should be interpreted with caution because this haplotype is also present in the *VRN-A1* locus in some accessions.

These four markers were all present in 22 of the 26 *T. aestivum* ssp. *aestivum* accessions (*SI Appendix*, Table S4) previously proposed to carry *VRN-D4* based on genetic analyses (32–34). However, markers for the 5DS/5AL insertion and for the A367C SNP were not detected in accessions IL193, CL035, IL346, and CL030 (*SI Appendix*, Table S4). Copy number assays showed that these four accessions have no additional copies of *VRN-A1* outside the ones in chromosome arm 5AL. Accessions IL193 and CL035 have only one copy of *VRN-A1* carrying the RIP-3$^{VRN\text{-}D4}$ haplotype (same as Jagger, Claire, and CS), whereas accessions IL346 and CL030 have two linked copies of *VRN-A1*. The last two lines have a C349T fixed SNP (*SI Appendix*, Table S4) that is diagnostic for a tandem duplication of *VRN-A1* on chromosome 5AL (6). These results confirmed the absence of *VRN-D4* in these four accessions, suggesting that the original crosses and genetic analyses should be repeated.

The collection sites of the *T. aestivum* ssp. *aestivum* accessions carrying *VRN-D4* (*SI Appendix*, Table S4) partially overlap with the distribution of *T. aestivum* ssp. *sphaerococcum* (Punjab region, current Northeast Pakistan and Northwest of India, *SI Appendix*, Fig. S10). This last subspecies has been previously proposed as a putative source of *VRN-D4* (41), a suggestion that is supported by the high frequency of *VRN-D4* (94%) among the 33 *T. aestivum* ssp. *sphaerococcum* accessions characterized in this study (*SI Appendix*, Table S5). The only two accessions classified as *T. aestivum* ssp. *sphaerococcum* lacking *VRN-D4* were collected in China (CItr 8610 and CItr 10911). These two lines also differed from the other *T. aestivum* ssp. *sphaerococcum* accessions by the presence of the *Vrn-D1a* allele (*SI Appendix*, Table S5), which carries a large intron deletion that is associated with a spring growth habit and that is present at high frequency in China (45).

Analyses of additional molecular markers for six *VRN1* and *VRN3* alleles showed no additional alleles for spring growth habit among the 31 *T. aestivum* ssp. *sphaerococcum* accessions that carry *VRN-D4* (*SI Appendix*, Table S5). For each accession, we also extracted RNA from leaves of 1-wk-old unvernalized plants and performed RT-PCR with primers that amplify both *VRN-D4* and *VRN-A1*. Sequencing of the amplification products showed only the *VRN-D4* allele (A367C SNP), confirming that this gene is expressed early in development in these *T. aestivum* ssp. *sphaerococcum* accessions (*SI Appendix*, Table S5). To determine if the absence of *VRN-D4* from these two lines from China was the result of admixture with other *T. aestivum* subspecies, we compared them with accessions from five different *T. aestivum* subspecies (*SI Appendix*, Method S5 and Table S6) using the 90K iSelect SNP chip

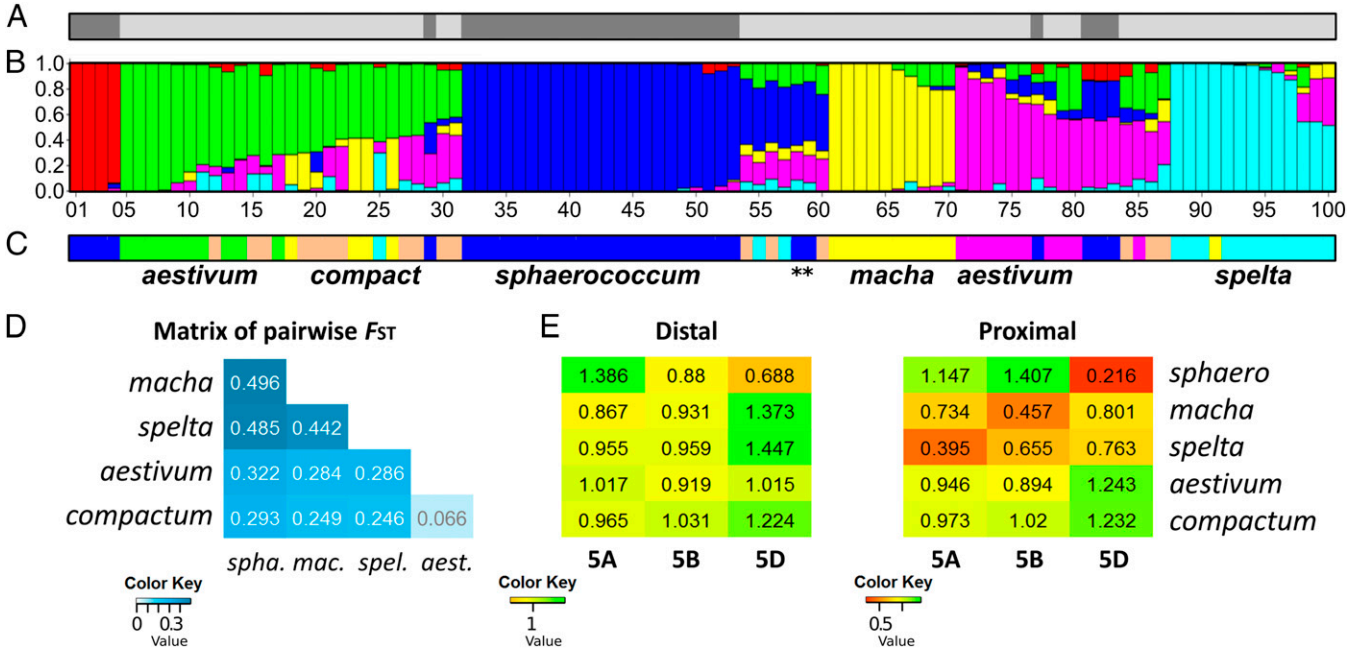

Fig. 5. Genetic relationships among *T. aestivum* subspecies. (*A*) Dark gray indicate presence and light gray absence of *VRN-D4* based on the markers developed in this study. (*B*) Population structure based on the analysis of 100 lines with 16,371 SNP markers (six groups). Each bar represents one genotype, and the identification numbers below correspond to the same numbers in *SI Appendix*, Table S6 describing each accession. (*C*) Classification of *T. aestivum* subspecies based on GRIN. Green, *aestivum* (spring growth habit); orange, *compactum*; blue, *sphaerococcum*; yellow, *macha*; pink, *aestivum* (winter growth habit); and cyan, *spelta*. Asterisks indicate two Chinese lines classified as ssp. *sphaerococcum* that show extensive admixture and lack *VRN-D4*. (*D*) Level of genetic differentiation between all pairs of subspecies determined using the fixation index (F_{ST}). All comparisons are significant at $P < 0.001$. (*E*) Standardized genetic diversity values for the distal and proximal regions in chromosomes from homeologous group 5. Note the high reduction in diversity in chromosome 5D proximal region of *T. aestivum* ssp. *sphaerococcum* relative to the other subspecies.

(46). The five subspecies were clustered into six groups, including two different groups for *T. aestivum* ssp. *sphaerococcum* and one group combining the spring accessions of *T. aestivum* ssp. *aestivum* with the accessions from *T. aestivum* ssp. *compactum* (Fig. 5*B*). Chinese accessions CItr 8610 and CItr 10911 exhibited evidence of extensive admixture, which may explain the absence of *VRN-D4* in these two lines (asterisks in Fig. 5*C* and *SI Appendix*, Fig. S11).

The *VRN-D4* gene was found in all of the *T. aestivum* ssp. *sphaerococcum* accessions from the Punjab region, suggesting that it likely contributed to local adaptation and was favored by positive selection. To test this possibility, we estimated the average genetic diversity (47) across chromosomes and chromosome regions in the different *T. aestivum* subspecies using 14,236 SNPs with known chromosome locations. To reduce the possible effect of ascertainment bias, diversity values per chromosome were divided by the average diversity of their respective genome and subspecies (*SI Appendix*, Method S6 and Table S7). These values, referred to hereafter as "standardized genetic diversity," are summarized in *SI Appendix*, Fig. S12. For chromosome 5D, the standardized genetic diversity of the *T. aestivum* ssp. *sphaerococcum* accessions carrying *VRN-D4* is less than half the average of the values from the other four *T. aestivum* subspecies. None of the other 20 chromosomes showed such a large difference in standardized genetic diversity among subspecies (*SI Appendix*, Fig. S12*A*).

To determine which region of chromosome 5D was responsible for the reduced diversity, we divided each chromosome arm into one proximal (1/3 of the genetic length including *VRN-D4*) and one distal region (2/3 of the genetic length). For comparison, a similar analysis was performed for homeologous chromosomes 5A and 5B (Fig. 5*E*) and for all of the chromosomes of the D genome (*SI Appendix*, Fig. S12*B*). Chromosome 5D of *T. aestivum* ssp. *sphaerococcum* showed a 1.8-fold reduction in standardized genetic diversity in the distal region and a 4.7-fold reduction in the proximal region compared with the average of the other four subspecies (Fig. 5*E*). Such a drastic relative reduction in diversity in the proximal region was not observed in chromosomes 5A and 5B (Fig. 5*E*) or in any of the other chromosomes of the D genome in *T. aestivum* ssp. *sphaerococcum* (*SI Appendix*, Fig. S12*B*). The centromeric region of chromosome 5D, where *VRN-D4* is located, also showed the highest level of genetic differentiation (F_{ST}) between *T. aestivum* ssp. *sphaerococcum* and the pooled data from the other four subspecies (*SI Appendix*, Fig. S13). In summary, these results indicate that the proximal region of chromosome 5D in *T. aestivum* ssp. *sphaerococcum* accounts for most of the observed reduction in genetic diversity in chromosome 5D in this subspecies and contributes to its genetic differentiation of *T. aestivum* ssp. *sphaerococcum* from the other subspecies of *T. aestivum*.

Discussion

The Dynamic Wheat Genome. This study shows that the *VRN-D4* gene originated from the insertion of a large region of chromosome arm 5AL (including *VRN-A1*) into chromosome arm 5DS. The presence of genes and gene fragments in noncolinear locations is relatively frequent in the large genomes of the temperate grasses, which originated from massive amplifications of repetitive elements (39, 48–50). Some DNA transposons in wheat (e.g., CACTA) are able to move genes and gene fragments to noncolinear locations (51), as reported for rice 'Pack-MULEs' (52) and maize helitrons (53). However, this mechanism is associated with mobilization of small DNA fragments (<3 kb) (54) and is unlikely to explain the 290-kb insertion that originated *VRN-D4*.

An alternative mechanism proposed to explain noncolinear DNA sequences involves the insertion of nonhomologous DNA segments to repair double-strand breaks generated by the insertion of repetitive elements (54). A hallmark of this mechanism is the presence of a retrotransposon insertion in one of the borders of the acceptor site (54). This mechanism is associated with a wide range of insertion sizes and therefore can explain the 290-kb 5AL insertion in chromosome arm 5DS. Unfortunately, one of the original downstream borders of the 5AL insertion has been altered in the CS-dt-5DS sequence, limiting our ability to determine the molecular mechanism that originated this large 5AL insertion.

The insertion of the 5AL chromosome segment into the 5DS chromosome arm provides a well-documented example of the ability of large wheat chromosome regions to move to different chromosome locations within historical times. This result suggests that contigs in a chromosome arm database that match sequences in different chromosomes may represent real insertion events and should not be automatically discarded as DNA contaminations. In addition, our results demonstrate that some of the sequences deposited in the wheat survey sequences of CS (39) actually belong to different wheat varieties. Differences between CS and CS-derived cytogenetic stocks have been reported before (40) and should be considered when analyzing the CS survey sequences.

The Origin of *VRN-D4*. Previous genetic studies identified the presence of *VRN-D4* in a few accessions of *T. aestivum* ssp. *sphaerococcum* and suggested that this subspecies could be the source of *VRN-D4* (32, 41, 55). This subspecies, first described in 1921 (56), has been found mainly in the Punjab region, including vast areas of eastern Pakistan and northern India. The characteristic round grains are abundant in archeological sites from the Indus Valley Civilization between 4,600 and 3,900 y ago (57).

The identification and characterization of *VRN-D4* allowed us to study the distribution of this gene in a larger germplasm collection (*SI Appendix*, Tables S4 and S5). We found *VRN-D4* in 31 of the 33 accessions of *T. aestivum* ssp. *sphaerococcum* (*SI Appendix*, Table S5) but in none of the accessions from the other subspecies described in *SI Appendix*, Table S6. These results, together with the higher frequency of *VRN-D4* in *T. aestivum* ssp. *aestivum* accessions from South Asia (*SI Appendix*, Table S4), support the hypothesis that *VRN-D4* originated in *T. aestivum* ssp. *sphaerococcum*. The only two accessions classified as *T. aestivum* ssp. *sphaerococcum* that lacked the *VRN-D4* were collected in China (outside the central area of distribution for this subspecies) and carry the alternative *VRN-D1a* allele for spring growth habit (*SI Appendix*, Table S5). These two accessions exhibit high levels of admixture (Fig. 5), which may explain the absence of *VRN-D4* and the presence of a different allele for spring growth habit.

Except for the previous two accessions from China, none of the *T. aestivum* ssp. *sphaerococcum* accessions described in *SI Appendix*, Table S5 carry known *VRN1* or *VRN3* alleles for spring growth habit (*SI Appendix*, Table S5). This result suggests that *VRN-D4* may be the only allele for spring growth habit in this subspecies. This was confirmed for accessions K-5498 and K-23822 (*SI Appendix*, Table S5) in a previously published genetic study (32). The importance of *VRN-D4* in the induction of spring growth habit in this subspecies was also supported by the early expression of this gene during development in the absence of vernalization. Early expression was confirmed for all of the accessions of *T. aestivum* ssp. *sphaerococcum* carrying this gene (*SI Appendix*, Table S5).

Patterns of Genetic Diversity in the *VRN-D4* Region. The proximal region of chromosome 5D in *T. aestivum* ssp. *sphaerococcum* showed a four- to fivefold reduction in standardized genetic diversity and a local increase in genetic differentiation relative to the other four subspecies. We did not observe reductions in genetic diversity of this magnitude in the proximal regions of homeologous chromosomes 5A and 5B (Fig. 5*E*) or in the proximal regions of the other six chromosomes from the D genome (*SI Appendix*, Fig. S12*B*). These patterns of genetic diversity are likely associated with adaptive selection for *Vrn-D4* or a tightly linked gene in the centromeric region of chromosome 5D.

Indirect support for *Vrn-D4*'s role in local adaptation is provided by previous studies showing evidence of selection for

different *VRN1* loci (58). Among modern wheat varieties, the strong *Vrn-A1a* allele for early flowering is predominant in regions of very cold winters, where spring wheats are planted in the spring. By contrast, the weaker *Vrn-B1a* and *VrnD1a* alleles are present at higher frequencies in the Mediterranean climates where spring wheats are planted during the fall (45, 59, 60). These results provide an example of the selective forces that could have also operated on *VRN-D4*. Adaptive selection for *VRN-D4* might have been favored by the absence of other alleles for spring growth habit in the *T. aestivum* ssp. *sphaerococcum* accessions from South Asia (*SI Appendix*, Table S5) and by the partial dominant nature of the *Vrn-D4* allele for early flowering (29).

We currently do not know when the *VRN-D4* gene originated in *T. aestivum* ssp. *sphaerococcum* and how fast it was adopted. However, if DNAs can be obtained from ancient archeological samples of *T. aestivum* ssp. *sphaerococcum*, the diagnostic markers developed in this study for *VRN-D4* can be used to answer these questions.

Genetic Differentiation of *T. aestivum* ssp. *sphaerococcum*. In addition to the characterization of the *VRN-D4* region, this study provides an overall view of *T. aestivum* ssp. *sphaerococcum* genetic diversity that is very different from the one suggested in earlier studies. Early studies based on few morphological traits proposed that *T. aestivum* ssp. *sphaerococcum* and ssp. *aestivum* differed only in a few loci (61, 62). However, our study based on a genome-wide sample of SNPs, suggests a very different picture. Both the F_{ST} and cluster analyses showed that *T. aestivum* ssp. *sphaerococcum* is one of the most differentiated subspecies of *T. aestivum* (Fig. 5*D* and *SI Appendix*, Fig. S11). A genome-wide scan of F_{ST} values across the 21 chromosomes revealed multiple genomic regions with strong differentiation, including *VRN-D4* (*SI Appendix*, Fig. S14). It would be interesting to investigate if these differentiated regions are also targets of adaptive selection.

We hypothesize that as common wheat moved eastward and westward from its original area on the south coast of the Caspian Sea, it differentiated and adapted to new environments. Limited exchanges between the western and Asian human populations during the early stages of agriculture expansion likely favored the simultaneous differentiation of crops, cultures, and languages. The divergence of *T. aestivum* ssp. *sphaerococcum* from the western *T. aestivum* subspecies parallels the differentiation of the Indo-European languages, where the Indo-Iranian subfamilies are more distinct from the western Celtic, Germanic, Italic, and Balto-Slavic families (63).

***VRN-D4* Provides Insights into the Regulation of Vernalization Requirement.** None of the polymorphisms previously known to be associated with spring growth habit in *VRN1* regulatory regions (17–20) were found in *VRN-D4*. Therefore, alternative mechanisms are required to explain the dominant spring growth habit associated with the *Vrn-D4* allele. We discuss below three nonmutually exclusive hypotheses.

Hypothesis 1: "Chromosome location." It is possible that the new chromosome location of the *VRN-D4* gene affects its regulation. The original *VRN-A1* gene is located in the middle of the 5AL arm in a region of normal recombination, whereas *VRN-D4* is located in a more proximal location where no recombination has been detected so far (29). It is well established that chromosome translocations and rearrangements can have a deep impact on gene expression. Neighboring genes can have highly correlated expression profiles (64, 65) and can impact the expression of relocated genes. However, no genes were identified close to *VRN-D4* limiting their potential effect on *VRN-D4* expression.

Position effects on gene expression can also be associated with changes in the proximity to euchromatic or heterochromatic regions. In *Drosophila*, these effects can spread over considerable distances, but the genes located closest to the breakpoint are commonly the most severely affected (66). However, in the *VRN-D4* locus, the duplicated *VRN1* gene is still flanked by large segments of the original 5AL chromatin (~80 kb upstream and ~200 kb downstream of *VRN-D4*), which may buffer potential effects of the different chromosome location on *VRN-D4* expression.

Hypothesis 2: "Effect of the K123Q polymorphism." The A367C SNP in *VRN-D4* results in an amino acid change (K123Q) that has not been detected so far in any published VRN1 allele (or in related MADS box proteins, *SI Appendix*, Figs. S4 and S5). Therefore, we currently have no information to determine if this polymorphism can contribute to the differences in flowering time. Any proposed mechanism for this polymorphism will have to explain why the transcript levels of *VRN-D4* are higher than those from *VRN-A1* during early developmental stages (Fig. 1*E* and *SI Appendix*, Fig. S3), when both genes have identical promoters. Hypothesis 3 provides a potential explanation for these differences.

Hypothesis 3: "SNP mutations in the RIP-3 region." The three linked SNPs in the transcribed RIP-3 region in the *VRN-D4* first intron provide an attractive explanation for the early expression of *VRN-D4*. These three linked SNPs are located within a predicted single-stranded RNA loop targeted by TaGRP2 (Fig. 4*C*). The RIP-3 region in wheat is similar to the binding site of the *Arabidopsis* GRP7 homolog (*SI Appendix*, Fig. S8*A*), and one of the three SNPs found in RIP-$3^{VRN\text{-}D4}$ is in a position that has been shown to be critical for the binding of GRP7 in *Arabidopsis* (67). Variation in AtGRP7 affects *Arabidopsis* flowering time by regulating the expression of the MADS-box *FLOWERING LOCUS C* (*FLC*) (22).

Previous studies in transgenic wheats have shown that overexpression of *TaGRP2* delays flowering, whereas its down-regulation by RNAi accelerates flowering relative to the wild-type control. These results confirmed that TaGRP2 functions as a flowering repressor (21). Our REMSA results show that the presence of three SNPs in the RIP-$3^{VRN\text{-}D4}$ region is sufficient to disrupt the binding between TaGRP2 and its target site (Fig. 4). TaGRP2 binding to the *VRN1* pre-mRNA is required for the inhibition of *VRN1* expression (21), so the disruption of this interaction can explain the earlier expression of *VRN-D4*. This mechanism is also consistent with the spring growth habit associated with large deletions in the first intron of *VRN1*, including the RIP-3 region (20, 68).

We found that the same three SNPs present in the RIP-3 region in *VRN-D4* are also present in the RIP-3 regions of the related *VRN-A1* alleles in common wheat varieties Jagger, Claire, and CS (*SI Appendix*, Fig. S8). The *VRN-A1* allele from CS has not been characterized in detail, but the winter wheat varieties Jagger and Claire have a reduced vernalization requirement (6, 44) that maps to the *VRN-A1* locus. The cause of the reduced vernalization requirement is still controversial and has been attributed to copy number variation in Claire (6) and to a polymorphism (A180V) in the VRN-A1 protein in Jagger (44). The discovery that both Claire and Jagger carry the RIP-$3^{VRN\text{-}D4}$ allele provides an alternative explanation for their reduced vernalization requirement. We propose that the limited binding of the TaGRP2 repressor to the RIP-$3^{VRN\text{-}D4}$ allele in these two varieties results in higher *VRN-A1* transcript levels and reduced vernalization requirement. This last interpretation is consistent with the expression reported for these two varieties. *VRN-A1* transcript levels increased more rapidly in Jagger than in the winter line 2174 (RIP-$3^{VRN\text{-}A1}$) (44) and also more rapidly in Claire than in the winter varieties Malacca and Hereward (6).

Despite carrying the same RIP-$3^{VRN\text{-}D4}$ allele, the effects of *VRN-D4* and the *VRN-A1* allele present in Jagger and Claire are not identical. A short vernalization treatment (3 wk) accelerates flowering in Claire and Jagger more than 1 mo (6, 44), but has no significant effects in TDF (*VRN-D4*) (28). This result suggests that additional mechanisms contribute to the earlier flowering associated with the *VRN-D4* allele. The different chromosome location or the additional K123Q polymorphism in *VRN-D4* can contribute to these differences, but it is also possible that the

different genetic backgrounds in these varieties can also modulate the effect of the RIP-3$^{VRN\text{-}D4}$ allele.

An interesting observation was that the *VRN-A1* alleles characterized so far in diploid *T. urartu* and in wild and cultivated polyploid wheat accessions (*SI Appendix*, Table S3) all have the T2783C SNP in the RIP-3 region (RIP-3$^{VRN\text{-}A1}$ allele). Our REMSA experiments showed that this mutation is sufficient to weaken the interaction between TaGRP2 and its target RNA sequence, although the disruption is not as strong as the one observed when the three SNPs are present in the RIP-3$^{VRN\text{-}D4}$ allele (Fig. 4). It is tempting to speculate that the T2783C SNP in RIP-3$^{VRN\text{-}A1}$ can contribute to the higher transcript levels in *VRN-A1* relative to *VRN-B1* and *VRN-D1* observed in winter wheat plants after 6 wk of vernalization (69). We are currently generating the genetic stocks required to compare the expression levels of the different *RIP-3* alleles at the *VRN-A1* locus.

In summary, the cloning of a vernalization gene originated in the ancient Indus Valley Civilization has improved our understanding of the natural variation of the main wheat vernalization gene in modern winter wheat varieties. Breeders can use these alleles to engineer wheat varieties better adapted to different or changing environments. One more time, "What's past is prologue."

Materials and Methods

Plant Materials and Growth Conditions. Triple Dirk F (TDF, spring, *Vrn-D4*) and Triple Dirk C (TDC, winter, *vrn-D4*), and Triple Dirk D (TDD, *Vrn-A1a*) are part of a series of isogenic lines carrying different alleles for spring growth habit in the genetic background of Triple Dirk (35, 36). The Australian variety Gabo is the source of the *VRN-D4* allele in TDF. Lines with the closest recombination events to *VRN-D4* were used to validate the *VRN-D4* candidate gene. These lines were identified in a high density map of *VRN-D4* generated from the cross between TDF and a substitution line of chromosome 5D from synthetic wheat 5402 in CS (CS5402) (28, 29).

We characterized two common wheat varieties with reduced vernalization requirement linked to the *VRN-A1* locus (6, 44). Jagger (KS-82-W-418/STEPHENS), a hard red winter wheat, was the most widely planted variety in Kansas and Oklahoma for 12 y. Its early maturity helps Jagger to partially avoid wheat rust and heat stress. Claire (WASP/FLAME) is a variety from the United Kingdom released in 1999 that remains popular for its suitability for early sowing.

To determine the geographic distribution of *VRN-D4*, we characterized a collection of 26 *T. aestivum* accessions with known vernalization genes (32–34) (*SI Appendix*, Table S4) and a collection of 33 *T. aestivum* ssp. *sphaerococcum* accessions obtained from the National Small Grains Collection (NSGC) and the Vavilov Institute of Plant Industry (*SI Appendix*, Table S5). For progeny tests and expression experiments, plants were grown in PGR15 growth chambers (Conviron) adjusted to 16 h of light (22 °C) and 8 h of darkness (18 °C).

Expression Profiles and Copy Number Variation. RNA samples were extracted from leaves using the Spectrum Plant Total RNA Kit (Sigma-Aldrich). *VRN1*, *VRN2*, *FT* and *ACTIN* primers for SYBR GREEN quantitative PCR were developed in previous studies (10, 70, 71). *VRN-A1*, *VRN-B1* and *VRN-D1* expression was tested using genome specific qRT-PCR primers described before (69). Quantitative RT-PCR was performed using SYBR Green and a 7500 Fast Real-Time PCR system (Applied Biosystems). Amplified products were confirmed by sequencing. *VRN-A1* and *VRN-D4* transcripts were differentiated by the presence of an additional BstNI restriction site in *VRN-D4* (A367C polymorphism, *SI Appendix*, Fig. S3). *VRN-A1* copy number estimation was performed using the Taqman assay described before (6).

Physical Map. The MTP clones of the CS-dt-5DS BAC library (42) were screened at Sabanci University (Turkey) with three pairs of primers for *VRN-D4* listed in *SI Appendix*, Table S8. The selected BAC and 11 overlapping BACs belonging to two overlapping contigs (*SI Appendix*, Table S2) were sequenced using a combination of Illumina and 454 sequencing (*SI Appendix*, Method S2). Annotation was conducted using "Triannot" (72). The assembled sequence was compared with the 5AL arm database of the IWGSC (39), *T. urartu* (v1.26 plants.ensembl.org) (43), and three *Ae. tauschii* (D genome) databases (plants.ensembl.org/Aegilops_tauschii/, GCA_000347335.1 (73), aegilops.wheat.ucdavis.edu/ATGSP/ (74), and https://urgi.versailles.inra.fr/download/iwgsc/TGAC_WGS_assemblies_of_other_wheat_species, TGAC_WGS_tauschii_v1). We used BLAT v. 35 filtered for matches that were longer than 500 bp and >99% identical (*SI Appendix*, Method S2).

"Targeting Local Lesions in Genomes" Mutants. A targeting local lesions in genomes (TILLING) population of 1,153 lines of hexaploid wheat line TDF (*VRN-D4*) was mutagenized with 0.9% ethyl methane sulphonate (EMS) and screened using a *CelI* assay as described before (38). DNA pools (four lines per pool) were screened with three sets of primers (*SI Appendix*, Method S1). The potential effects of the different mutations on protein function were predicted using programs PROVEAN, SIFT, and PolyPhen-2 (*SI Appendix*, Method S1).

***T. aestivum* ssp. *sphaerococcum* Population Structure.** A total of 100 lines including 33 *T. aestivum* ssp. *sphaerococcum*, 15 *T. aestivum* ssp. *macha*, 17 *T. aestivum* ssp. *compactum*, 15 *T. aestivum* ssp. *spelta*, and 20 *T. aestivum* ssp. *aestivum* (10 spring and 10 winter) accessions obtained from the NSGC (*SI Appendix*, Table S6) were genotyped with the iSelect 90k SNP array (46). The population structure analysis was conducted using the program STRUCTURE version 2.3.4 (*SI Appendix*, Method S5). The fixation index (F_{ST}) between all possible pairwise combinations of the five subspecies was calculated using the program Arlequin (75). Mean F_{ST} values were calculated using a sliding window of seven mapped SNPs (46) with steps of four SNPs to generate smoother curves. F_{ST} values for each SNP were calculated by comparing *T. aestivum* ssp. *sphaerococcum* with pooled data from the other four subspecies.

ACKNOWLEDGMENTS. We thank Francine Paraiso, Tyson Howell, Stephen Pearce, Chengxia Li, and Eligio Bossolini for valuable suggestions; Dr. Nikolay Goncharov (Institute of Cytology and Genetics), Dr. Jan Dvorak (University of California, Davis), and Harold Bockelman [US Department of Agriculture (USDA) National Small Grain Collection] for the seeds of the different germplasm used in this study; and Dr. Jan Dvorak and Dr. M.-C. Luo for early access to the *Ae. tauschii* genome database. This project was supported by the National Research Initiative (Grants 2011-67013-30077 and 2011-68002-30029) from the USDA National Institute of Food and Agriculture, by the Howard Hughes Medical Institute, and the Gordon and Betty Moore Foundation Grant GBMF3031. J.M.D. is supported by a fellowship of the Human Frontier Science Program.

1. Wang J, et al. (2013) *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytol* 198(3):925–937.
2. Huang S, et al. (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc Natl Acad Sci USA* 99(12):8133–8138.
3. Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316(5833):1862–1866.
4. Beales J, Turner A, Griffiths S, Snape JW, Laurie DA (2007) A *pseudo-response regulator* is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). *Theor Appl Genet* 115(5):721–733.
5. Wilhelm EP, Turner AS, Laurie DA (2009) Photoperiod insensitive *Ppd-A1a* mutations in tetraploid wheat (*Triticum durum* Desf.). *Theor Appl Genet* 118(2):285–294.
6. Díaz A, Zikhali M, Turner AS, Isaac P, Laurie DA (2012) Copy number variation affecting the *Photoperiod-B1* and *Vernalization-A1* genes is associated with altered flowering time in wheat (*Triticum aestivum*). *PLoS One* 7(3):e33234.
7. Yan L, et al. (2003) Positional cloning of the wheat vernalization gene VRN1. *Proc Natl Acad Sci USA* 100(10):6263–6268.
8. Trevaskis B, Bagnall DJ, Ellis MH, Peacock WJ, Dennis ES (2003) MADS box genes control vernalization-induced flowering in cereals. *Proc Natl Acad Sci USA* 100(22):13099–13104.
9. Danyluk J, et al. (2003) *TaVRT-1*, a putative transcription factor associated with vegetative to reproductive transition in cereals. *Plant Physiol* 132(4):1849–1860.
10. 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.
11. Corbesier L, et al. (2007) FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science* 316(5827):1030–1033.
12. Tamaki S, Matsuo S, Wong HL, Yokoi S, Shimamoto K (2007) Hd3a protein is a mobile flowering signal in rice. *Science* 316(5827):1033–1036.
13. Taoka K, et al. (2011) 14-3-3 proteins act as intracellular receptors for rice *Hd3a* florigen. *Nature* 476(7360):332–335.
14. Li C, Dubcovsky J (2008) Wheat FT protein regulates *VRN1* transcription through interactions with FDL2. *Plant J* 55(4):543–554.
15. Li C, Lin H, Dubcovsky J (2015) Factorial combinations of protein interactions generate a multiplicity of florigen activation complexes in wheat and barley. *Plant J*, 10.1111/tpj.12960.
16. Oliver SN, Finnegan EJ, Dennis ES, Peacock WJ, Trevaskis B (2009) Vernalization-induced flowering in cereals is associated with changes in histone methylation at the *VERNALIZATION1* gene. *Proc Natl Acad Sci USA* 106(20):8386–8391.
17. Yan L, et al. (2004) Allelic variation at the *VRN-1* promoter region in polyploid wheat. *Theor Appl Genet* 109(8):1677–1686.

18. Zhang J, et al. (2012) A single nucleotide polymorphism at the *Vrn-D1* promoter region in common wheat is associated with vernalization response. *Theor Appl Genet* 125(8):1697–1704.
19. Chu CG, et al. (2011) A novel retrotransposon inserted in the dominant *Vrn-B1* allele confers spring growth habit in tetraploid wheat (*Triticum turgidum* L.). *G3 (Bethesda)* 1(7):637–645.
20. Fu D, et al. (2005) Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. *Mol Genet Genomics* 273(1):54–65.
21. Xiao J, et al. (2014) O-GlcNAc-mediated interaction between VER2 and TaGRP2 elicits *TaVRN1* mRNA accumulation during vernalization in winter wheat. *Nat Commun* 5:4572.
22. Streitner C, et al. (2008) The small glycine-rich RNA binding protein AtGRP7 promotes floral transition in Arabidopsis thaliana. *Plant J* 56(2):239–250.
23. Chen A, Dubcovsky J (2012) Wheat TILLING mutants show that the vernalization gene *VRN1* down-regulates the flowering repressor *VRN2* in leaves but is not essential for flowering. *PLoS Genet* 8(12):e1003134.
24. Yan L, et al. (2004) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* 303(5664):1640–1644.
25. Dubcovsky J, et al. (2006) Effect of photoperiod on the regulation of wheat vernalization genes *VRN1* and *VRN2*. *Plant Mol Biol* 60(4):469–480.
26. Hemming MN, Peacock WJ, Dennis ES, Trevaskis B (2008) Low-temperature and daylength cues are integrated to regulate *FLOWERING LOCUS T* in barley. *Plant Physiol* 147(1):355–366.
27. Distelfeld A, Li C, Dubcovsky J (2009) Regulation of flowering in temperate cereals. *Curr Opin Plant Biol* 12(2):178–184.
28. Yoshida T, et al. (2010) *Vrn-D4* is a vernalization gene located on the centromeric region of chromosome 5D in hexaploid wheat. *Theor Appl Genet* 120(3):543–552.
29. Kippes N, et al. (2014) Fine mapping and epistatic interactions of the vernalization gene *VRN-D4* in hexaploid wheat. *Mol Genet Genomics* 289(1):47–62.
30. Knott DR (1959) The inheritance of rust resistance. IV. Monosomic analysis of rust resistance and some other characters in six varieties of wheat including Gabo and Kenya Farmer. *Can J Plant Sci* 39(2):215–228.
31. O'Brien L, Morell M, Wrigley C, Appels R (2001) Genetic pool of Australian wheats. *The World Wheat Book*, eds Bonjean AP, Angus WJ (Lavoisier Publishing, Paris), pp 611–648.
32. Goncharov NP, Shitova IP (1999) The inheritance of growth habit in old local varieties and landraces of hexaploid wheats. *Russ J Genet* 35(4):386–392.
33. Iwaki K, Haruna S, Niwa T, Kato K (2001) Adaptation and ecological differentiation in wheat with special reference to geographical variation of growth habit and *Vrn* genotype. *Plant Breed* 120(2):107–114.
34. Iwaki K, Nakagawa K, Kuno H, Kato K (2000) Ecogeographical differentiation in east Asian wheat, revealed from the geographical variation of growth habit and *Vrn* genotype. *Euphytica* 111(2):137–143.
35. Pugsley AT (1971) A genetic analysis of the spring-winter habit of growth in wheat. *Aust J Agric Res* 22:21–31.
36. Pugsley AT (1972) Additional genes inhibiting winter habit in wheat. *Euphytica* 21:547–552.
37. Zhu J, et al. (2014) Copy number and haplotype variation at the *VRN-A1* and central *FR-A2* loci are associated with frost tolerance in hexaploid wheat. *Theor Appl Genet* 127(5):1183–1197.
38. Uauy C, et al. (2009) A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. *BMC Plant Biol* 9:115–128.
39. Mayer KFX, et al. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* 345(6194):1251788.
40. Devos KM, et al. (1999) Chromosome aberrations in wheat nullisomic-tetrasomic and ditelosomic lines. *Cereal Res Commun* 27(3):231–239.
41. Goncharov NP (2003) Genetics of growth habit (spring vs winter) in common wheat: Confirmation of the existence of dominant gene *Vrn4*. *Theor Appl Genet* 107(4): 768–772.
42. Akpinar BA, et al. (2015) The physical map of wheat chromosome 5DS revealed gene duplications and small rearrangements. *BMC Genomics* 16:453.
43. Ling HQ, et al. (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* 496(7443):87–90.
44. Li G, et al. (2013) Vernalization requirement duration in winter wheat is controlled by TaVRN-A1 at the protein level. *Plant J* 76(5):742–753.
45. Zhang XK, Xia XC, Xiao YG, Dubcovsky J, He ZH (2008) Allelic variation at the vernalization genes *Vrn-A1, Vrn-B1, Vrn-D1* and *Vrn-B3* in Chinese common wheat cultivars and their association with growth habit. *Crop Sci* 48(2):458–470.
46. Wang S, et al.; International Wheat Genome Sequencing Consortium (2014) Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol J* 12(6):787–796.
47. Weir BS (1996) *Genetic Data Analyisis II* (Sinauer, Sunderland, MA).
48. SanMiguel PJ, Ramakrishna W, Bennetzen JL, Busso CS, Dubcovsky J (2002) Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m). *Funct Integr Genomics* 2(1-2):70–80.
49. Wicker T, et al. (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. *Plant Cell* 23(5):1706–1718.
50. Ma J, et al. (2015) Putative interchromosomal rearrangements in the hexaploid wheat (*Triticum aestivum* L.) genotype 'Chinese Spring' revealed by gene locations on homoeologous chromosomes. *BMC Evol Biol* 15:37.
51. Wicker T, Guyot R, Yahiaoui N, Keller B (2003) CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. *Plant Physiol* 132(1):52–63.
52. Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431(7008):569–573.
53. Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence nonhomologies among maize inbreds. *Plant Cell* 17(2):343–360.
54. Wicker T, Buchmann JP, Keller B (2010) Patching gaps in plant genomes results in gene movement and erosion of colinearity. *Genome Res* 20(9):1229–1237.
55. Stelmakh AF, Avsenin VI (1996) Alien introgression of spring habit dominant genes into bread wheat. *Euphytica* 89(1):65–68.
56. Percival J (1921) *The Wheat Plant, a Monograph* (Duckworth and Co, London).
57. Weber SA (1999) Seeds of urbanism: Paleoethnobotany and the Indus civilization. *Antiquity* 73:813–826.
58. Cavanagh CR, et al. (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc Natl Acad Sci USA* 110(20):8057–8062.
59. Stelmakh AF (1998) Genetic systems regulating flowering response in wheat. *Euphytica* 100(1-3):359–369.
60. Stelmakh AF (1990) Geographic-distribution of *Vrn*-genes in landraces and improved varieties of spring bread wheat. *Euphytica* 45(2):113–118.
61. Swaminathan MS, Chopra VL, Jagathes D (1963) Induced sphaerococcoid mutations in *Triticum aestivum* and their phylogenetic and breeding significance. *Curr Sci India* 32(12):539–540.
62. Gupta N, Swaminathan MS (1967) An induced sphaerococcoid mutant in *Triticum dicoccum*. *Curr Sci India* 36(1):19.
63. Bouckaert R, et al. (2012) Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097):957–960.
64. Williams EJB, Bowles DJ (2004) Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. *Genome Res* 14(6):1060–1067.
65. Zhan S, Horrocks J, Lukens LN (2006) Islands of co-expressed neighbouring genes in *Arabidopsis thaliana* suggest higher-order chromosome domains. *Plant J* 45(3): 347–357.
66. Lima-de-Faria A (1983) *Molecular Evolution and Organization of the Chromosome* (Elsevier, Amsterdam).
67. Leder V, et al. (2014) Mutational definition of binding requirements of an hnRNP-like protein in Arabidopsis using fluorescence correlation spectroscopy. *Biochem Biophys Res Commun* 453(1):69–74.
68. Hemming MN, Fieg S, Peacock WJ, Dennis ES, Trevaskis B (2009) Regions associated with repression of the barley (*Hordeum vulgare*) *VERNALIZATION1* gene are not required for cold induction. *Mol Genet Genomics* 282(2):107–117.
69. Loukoianov A, Yan L, Blechl A, Sanchez A, Dubcovsky J (2005) Regulation of *VRN-1* vernalization genes in normal and transgenic polyploid wheat. *Plant Physiol* 138(4): 2364–2373.
70. Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J (2006) A NAC Gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* 314(5803): 1298–1301.
71. Distelfeld A, Tranquilli G, Li C, Yan L, Dubcovsky J (2009) Genetic and molecular characterization of the *VRN2* loci in tetraploid wheat. *Plant Physiol* 149(1):245–257.
72. Leroy P, et al. (2012) TriAnnot: A versatile and high performance pipeline for the automated annotation of plant genomes. *Front Plant Sci* 3:5.
73. Jia J, et al.; International Wheat Genome Sequencing Consortium (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496(7443):91–95.
74. Luo MC, et al. (2013) A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. *Proc Natl Acad Sci USA* 110(19):7940–7945.
75. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.