Preadapted to adapt: underpinnings of adaptive plasticity revealed by the downy brome genome

Samuel R Revolinski
Washington State University https://orcid.org/0000-0001-7868-0585
Peter J Maughan
Brigham Young University https://orcid.org/0000-0003-3714-3411
Craig E Coleman
Brigham Young University https://orcid.org/0000-0002-7244-2532
Ian C Burke (icburke@wsu.edu)
Washington State University https://orcid.org/0000-0002-4384-2684

Research Article

Keywords: Bromus, GWAS, invasive species, flowering time, phenology, reference genome, genome annotation, cheatgrass, downy brome, Bromus tectorum

Posted Date: September 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2050485/v1

Version of Record: A version of this preprint was published at Communications Biology on March 27th, 2023. See the published version at https://doi.org/10.1038/s42003-023-04620-9.

Abstract

Bromus tectorum L. is arguably the most successful invasive weed in the world. It has fundamentally altered arid ecosystems of the western United States, where it now found on an excess of 20 million hectares and costs land managers and growers through lost yield, land utility, and increased incidence of fire. Invasion success is often related to avoidance of abiotic stress and human management. Early flowering is a complex but heritable trait utilized by *B. tectorum* that enables the species to temporally monopolize limited resources and thus outcompete native plant community. Thus, understanding the genetic underpinning of flowering time is critical for the design of integrated management strategies – regardless of the invaded ecosystem. To study flowering time traits in *B. tectorum*, we assembled the first chromosome scale reference genome using PacBio long reads, assembled using the Canu assembler, and scaffolded using Omni-C chromatin contact mapping technology. The final assembly spanned 2.482 Gb in length and has an N50 and L50 of 357 Mb and 4, respectively. To assess the utility of the assembled genome for trait discovery, 121 diverse *B. tectorum* accessions were phenotyped in replicated greenhouse trials, genotyped by sequencing and subjected to a genome wide association study (GWAS). Significantly ($q < 0.05$) associated QTLs were identified for height, days to first joint (J1), days to first visible panicle (VPN), and days to first ripe seed (FRS). Overlap between significant QTLs was present between traits, suggesting pleiotropy or closely linked QTLs for life cycle related traits. Candidate genes, representing homologs of an array of genes that have been previously associated with plant height or flowering phenology traits in related species, were located near significant QTLs. The GWAS, combined with a well annotated genome, is a viable method for identifying candidate genes associated with weedy characteristics in invasive weeds. This is the first study using high-resolution GWAS to identify phenology related genes in a weedy species and represents a significant step forward in our understanding of the mechanisms underlying genetic plasticity in one of the most successful invasive weed species in the world.

Main

Bromus tectorum is the most abundant invasive weed in North America. In the western United States, it is estimated to infest 31.4% (210,000 km^2) of the Great Basin[1]. It is most notorious for and especially problematic in non-cropland and rangelands of the intermountain west where *B. tectorum* invasion has altered intervals between fires from at least 60 years to less than five years[2,3], exacerbating the degradation of ecosystems caused by climate change[1,4]. *Bromus tectorum* is also a damaging weed in agricultural crops, causing significant yield loss in winter wheat (*Triticum aestivum L.*) across a large proportion of the western North American wheat producing region[5].

The history of *Bromus tectorum* (L.) in North America is proposed to have begun as a genetic bottleneck, resulting from the arrival of a small number of founder genotypes from Eurasia[6]. The seed likely arrived in animal bedding, grain contaminants, or it was imported intentionally as a potential forage[7]. *Bromus tectorum*, known as cheatgrass or downy brome, is predominantly a self-fertilizing, cleistogamous

species, adapted to multiple ecosystems in its native range[8]. Although the species is distributed across North American ecosystems, the population has retained genetic signatures that trace back to ancestral populations and ecosystems[6]. Thus, the success of *B. tectorum* invasiveness is due in large part to the diversity of these original populations that exhibit significant levels of plasticity in phenological traits. Plasticity allows individual *B. tectorum* plants to adapt quickly to local changes in the availability of in-season resources[9], while populations of *B. tectorum* are composed of an assemblage of diverse genotypes, facilitating success in response to long term and local variation in climate[10,11,12]. Such adaptive variation in life cycle traits[13,14] hinders management efforts in both natural and agricultural ecosystems. In short, *B. tectorum* individuals and populations express adaptive plasticity and, when expressed as earlier and variable reproductive phenology, allow *B. tectorum* to outcompete for limited resources at the expense of the native plant community[15]. Where it successfully invades, *B. tectorum* germinates and flowers early, facilitating access to limited resources, usually moisture, well before the native vegetation or crops can compete successfully for these resources[16,17].

Phenotypic variation in *B. tectorum* for adaptive traits like the aforementioned flowering time, but also seed dormancy and vernalization are undoubtedly major drivers of *B. tectorum's* highly successful invasion across a wide variety of North America ecosystems[13,16,17,18]. In model and crop *Poacea* species, growth and flowering phenology is controlled by an array of light and temperature sensitive gene pathways. Unfortunately, little is currently known about the genetic basis of adaptive variation in *B. tectorum*. Orthologous genes, such as *FT* (flowering time), *VRN1* (VERNALIZATION1), and *VRN2* (VERNALIZATION2) are likely contributing factors[19]. As a member of the Pooideae subfamily in the Poaceae family, *B. tectorum* is closely related to the Triticeae subfamily, which includes the agriculturally important species barley (*Hordeum vulgare L.*), wheat (*Triticum aestivum L.*) and rye (*Secale cereale L.*)[20,21]. Thus, extensive genomic resources, including well annotated genomes, in these sister taxa are available that will facilitate comparative genomics and gene discovery efforts in *B. tectorum.*

Here we report the first high-quality, annotated reference genome for *B. tectorum*. We utilize the reference genome in a genome wide association study (GWAS) to identify candidate genes for reproductive phenology (days to first joint, days to first visible panicle, days to first ripe seed and number of tillers) and plant height – traits that are known to directly influence the success of *B. tectorum* as an invasive weed species. Genome wide association studies (GWAS) are particularly useful for dissecting complex traits in species where controlled crossing is not practical or possible [22,23,24,25,26]. Our study demonstrates the first successful application of GWAS for the identification of QTL controlling heritable reproductive phenology traits in a weedy, highly invasive species. The identification of candidate gene targets controlling important climate and management adaptive characteristics underlying these QTL suggests how *B. tectorum* might respond to climate change, thus enabling the development enhanced and more reliable management practices for this highly invasive and problematic weed species.

Results

Genome Assembly. The Omni-C chromosomal assembly produced a near complete assembly for *B. tectorum.* The assembly was 2,482 megabases (Mb) in total length (Supplementary Table S1). The contig N50 was 19.4 Mb, the scaffold N50 was 357.4, and the resulting assembly contained 92.1% of the BUSCO genes with 259 gaps. The final assembly had a L50 of four and a L90 of seven corresponding to the seven chromosomes ($x = 7$) expected for members of the Pooideae in the the Poeae, Aveneae, Bromeae and Triticeae tribes[27].

Based on CDS comparisons using MCScanX[28], the seven largest scaffolds of the *B. tectorum* reference genome were found to have a one-to-one syntenic relationship with the seven chromosomes of the barley genome – not surprising given their phylogenetic proximity. As expected, synteny between *B. tectorum* and *H. vulgare* was highest in the chromosome arms where gene density is known to be high and lowest though the centromeric region where gene density is substantially reduced (Fig. 2). A translocation is present between chromosomes two and five where the first half of *Bt5* has synteny with chromosome two in barley and the second half of *Bt2* has synteny with chromosome 5 in barley (Fig. 2).

Phenotypic Analysis. Broad Sense Heritability (Reliability), the mean, standard deviation, min, max and correlation Best Linear Unbiased Estimates (BLUEs) were calculated to characterize the variation in traits and obtain a measure of the effect of each genotype. The distributions of BLUEs for reproductive phenology traits indicated a bimodal distribution (Fig. 1), where more rapidly flowering and taller plants consisted of accessions from Washington and GRIN collections. In contrast, Montana accessions flowered later and were shorter (Supplementary Table S3). Broad sense heritability was high for all the traits and ranged from 0.94 for tiller number to 0.99 for days to first visible panicle (VPN) (Table 1). Genotypes had a wide range of BLUEs for all traits measured: plant height (PH) ranged from 36.5 to 89 cm, number of tillers ranged from 5.5 to 38.5, and days to first ripe seed (FRS) ranged from 43.2 to 112.5 d (Table 1). Spearman correlations between BLUEs of different reproductive phenology traits were all above 0.95 (Fig. 1). Reproductive phenology and PH traits were moderately negatively correlated, with Spearman correlations ranging from the − 0.44 (PH and FRS) to -0.51 (PH and J1).

Genome-wide Association Mapping for Height, Tiller Number and Phenology Traits. To identify regions of the genome associated with variation in adaptive traits, a GWAS was performed on 121 genotypes to find QTL for PH, VPN, days to first visible joint (J1), FRS, days to 50% ripe seed (AWN_{50}) and tiller number using BLINK with 3 principal components and significance threshold of 0.05 after multiple testing correction. Nineteen QTLs were significantly ($q < 0.05$) associated with PH and reproductive phenology, with one QTL for PH (1), nine QTL for VPN (9), three QTL for J1 (3), nine QTL for FRS (9), and three QTL for AWN_{50} (3). A QTL on *Bt*6 (Bt6:8628087) was significant ($q < 0.05$) for all the reproductive phenology related traits except J1, where a second significant ($q = 2.070E-07$) QTL for J1 was at a nearby position (Bt6:8417488) on *Bt6* (Table 2). The QTL on *Bt1* (Bt1:276755111) was significant ($q < 0.05$) for J1, FRS, and AWN_{50} (Table 2). QTLs on *Bt2* (Bt2:9403921) and *Bt7* (Bt7:347158582) were significantly ($q < 0.05$) associated with VPN and FRS (Table 2).

Plant Height. We identified a QTL that was significantly (q = 1.360E-05, Table 2) associated with PH on Bt6:301800092. The QTL explained 16.9% (Table 2) of the phenotypic variation. The MAF at the PH QTL on *Bt6* was 0.44, indicating both allelic states are common in our panel. Searching the area flanking the QTL associated with PH at Bt6:301800092 on both sides by up to 500 Kb revealed a homolog of Xanthine Dehydrogenase (XDH) 29 Kb from the QTL and a homolog of Indole-3-pyruvate monooxygenase YUCCA6 (YUC6) 242 Kb from the QTL (Table 3). XDH and YUC6 are promising candidate genes for the QTL associated with PH we identified on *Bt6* as they both are well document to be associated with changes in PH, senescence and response to drought[29,30]. In rice (*Oryza sativa* L.) overexpression of the XDH homolog led to increased PH while under-expression of the XDH homolog resulted in reduced PH[29], indicating that homologs of XDH would be fitting candidate genes for PH. Indeed, in rice, a GWAS identified XDH as a candidate gene for coleoptile length in response to flooding[31] indicating XDH homologs may be involved with stem elongation in grasses and thus PH. Furthermore, in Arabidopsis knock-out mutations of the XDH gene led to reduced PH[32]. Knock-out mutations of the YUC6 gene in Arabidopsis was found to increase auxin production in the shoots leading to reduced PH[33]. Further investigation, including gene expression experiments and/or knockout of the XDH gene homolog in *B. tectorum*, is needed to validate and further understand the underlying genetics controlling the large effect QTL for PH.

Unexpectedly, PH and reproductive phenology timing were negatively correlated, contradicting previous findings that PH and reproductive phenology were positively correlated (i.e., earlier flowering plants do not have as much time to grow)[34,35]. *Lolium perenne* (L.) was also found to have a negative genetic correlation between PH and flowering time[36]. The negative correlation was thought to be the result of selection imposed by grazing, where plants biological fit plants remained short until they were ready to flower at which point they elongated and flowered quickly[37]. *Bromus tectorum* may also be under similar grazing selection pressure[38]. Competition with crops could also select for taller, fast-growing phenotypes to facilitate competition for space.

Phenology Traits. Flowering time is an important adaptive trait for ensuring the survival of plants in a broad range of climates, often driving local adaptation[39,40,41]. In predominantly self-fertilizing species, large effect QTLs control the variation in flowering time[42,43], in contrast to maize where flowering time is controlled by many QTL of small effect[44]. Here, we phenotyped four traits (VPN, J1, FRS, AWN_{50}) associated with flowering time to facilitate a GWAS to identify candidate reproductive phenology genes. Days to first visible panicle (VPN) was the first reproductive phenology stage observed and had the highest heritability at 0.99 (Table 1). The largest effect estimated for a significant (q = 3.630E-07) QTL for VPN was found at Bt7:70795764, explaining 16.3% of the variation in VPN with a MAF of 0.26. The eight-remaining significant (q < 0.05) QTL on *Bt3* (Bt3:2025814; q = 8.210E-07), *Bt4* (Bt4:382636922; q = 1.920E-05), *Bt6* (Bt6:8628087; q = 4.830E-03), *Bt7* (Bt7:347158582; q = 1.046E-02), *Bt2* (Bt2:9403921; q = 3.485E-02), *Bt3* (Bt3:383345273; q = 3.797E-02), *Bt3* (Bt3:324738958; q = 3.797E-02) and *Bt3*

(Bt3:324739133; q = 3.797E-02) explained 4.3, 4.3, 2.9, 4.7, 3.3, 3.8, 3.4 and 3.4 percent of the phenotypic variation, respectively, with MAF ranging from 0.2 to 0.49 (Table 2).

Developmentally, days to first joint (J1) was the second reproductive phenology trait to occur. Three significant (q < 0.05) QTL were associated with J1. The significant QTL at *Bt6* (Bt6:8417488; q = 2.070E-07) was 12 Kb from a homolog of HDR1 and explained 6.7% of the phenotypic variation with an MAF of 0.35 (Table 2 and Table 3). The QTL at *Bt1* (Bt1:276755111; q = 3.870E-05) explained 8.1% of the phenotypic variation with a MAF of 0.2. The other QTL on *Bt1* (Bt1:169814345; q = 1.909E-03) explained 4.2% of the phenotypic variation and had an MAF of 0.35.

The next developmental stage was first ripe seed (FRS). Days to first ripe seed was associated with nine significantly (q < 0.05) associated QTL. The leading QTL located on *Bt7* (Bt7:347158582; q = 1.360E-05) explained 7.2% of the phenotypic variation of FRS and was common (MAF = 0.29) (Table 2). The remaining QTLs on *Bt7* (Bt7:347158582; q = 5.280E-09), *Bt1* (Bt1:173453655; q = 1.582E-04), *Bt1*(Bt1:39144935; q = 1.786E-04), *Bt3* (Bt3:48964562; q = 1.786E-04), *Bt1* (Bt1:276755111; q = 2.823E-04), *Bt2* (Bt2:9403921; q = 4.166E-04), *Bt4* (Bt4:382201414; q = 1.026E-03), *Bt3* (Bt3:4974562; q = 1.836E-03) and *Bt6* (Bt6:8628087; q = 4.110E-02) explained 1.4, 1.8, 1.2, 4.1, 1.9, 0.6, 4.0 and 2.3 percent of the variation, respectively (Table 2). The MAF of QTL significantly associated with FRS were lower than those identified for the other reproductive phenology traits, varying from 0.058 to 0.29.

The final developmental reproductive phenology trait to occur was AWN_{50}, resulting in three significant associations (q < 0.05). The leading QTL on *Bt1* (Bt1:276755111; q = 6.760E-05) explained 6.5% of the phenotypic variation and had a MAF of 0.2 (Table 2). The other two QTL were located on *Bt5* (Bt5:348443767; q = 1.637E-04) and *Bt6* (Bt6:8628087; q = 5.973E-03) explaining 5.6% and 4.8% of the phenotypic variation with MAFs of 0.37 and 0.2, respectively.

Survey of Candidate Genes Associated with Phenology. Seventeen genes were identified as punitive candidates for reproductive phenology traits, based on their proximity to QTL (within a 500Kb interval on either side of the significant SNP) and previous functional characterization related to specific phenologies. The most promising of these genes is a homolog of *Heading Date Repression 1* (*HDR1*) The proximity of a homolog of HDR1 with QTL from all the maturity related trait analyzed suggests that HDR1 significantly influences maturation in *B. tectorum*. In *O. sativa*, *HDR1* promotes *Heading date 1* (*Hd1*) and represses *Early heading date 1* (*Ehd1*) delaying flowering[45]. Knockout mutations or RNA interference of *HDR1* resulted in rice plants that flowered 30 days earlier in long day light conditions, making *HDR1* a promising candidate gene for maturity traits[45].

Multiple candidate genes were identified for several reproductive phenology associated QTL, indicating multiple genes may underlie reproductive phenology in *B. tectorum*. The QTL on *Bt2* (9403921) associated with VPN and FRS was 13Kb, 29Kb and 133Kb from homologs of ABC transporter B family member 19 (*ABCB19*), *Cullin-3A* (*CUL3A*), and *Gibberellin 20 oxidase 2* (*GA20OX2*), respectively. Loss of function mutations in *ABCB19*, *CUL3A*, and *GA20OX2* led to longer flowering times, indicating all three

promote advancement of reproductive phenology[46,47,48]. *FRS5* is a punitive transcription factor regulating far-red light control of development[49], associated with adaptation to photoperiod[50]. Dehydration-responsive element-binding protein 1F (*DREB1F*) is a putative transcription factor, when upregulated, represses GA biosynthesis catalyzed by *GA20OX* genes, causing shorter plants with delayed flowering[51]. A likelihood ratio test of the interaction between the FRS QTLs on *Bt2* (9403921) and *Bt1* (173453655) revealed epistasis is likely (X^2 = 3.71 df = 1, p = 0.054), indicating *DREB1F* and *GA20OX* are interacting to control FRS in an epistatic manner.

The QTL associated with VPN on *Bt3* (2025814) was 120Kb, 177Kb and 240Kb from homologues of BTB/POZ and MATH domain-containing protein 1 (*BPM1*), BTB/POZ and MATH domain-containing protein 2 (*BPM2*) and PHYTOCHROME-DEPENDENT LATE-FLOWERING (*PHL*), respectively. *PHL* triggers flowering under long-day conditions by repressing the phytochrome b (*PHYB*) and the constans (*CO*) genes[52]. *BPM1* and *BPM2* are transcription factors of the BTB/POZ and MATH domain-containing protein (*BPM*) gene family and are involved in regulating flowering time by making proteins that are part of the Cullin E3 ubiquitin ligase complexes that include *CUL3A*[53]. A likelihood ratio test did not detect epistasis (X^2 = 0.02, df = 1, p = 0.901) between the QTL on *Bt2* (9403921) near *CUL3A* and the QTL on *Bt3* (2025814) near *BPM1* and *BPM2*.

A QTL on *Bt3* (324738958) associated with *VPN* was located 64Kb, 113Kb and 189Kb from homologs of *UPSTREAM of FLC (UFC)*, *Ultraviolet-B receptor 8 (UVR8)* and *Early Flowering 7 (VIP2)*, respectively. *UVR8, VIP2* and *UFC* are involved with the regulation of *Flowering Locus C* (FLC) in *A. thaliana*[54,55,56], suggesting one mechanism maintaining genetic variation in *B. tectorum* flowering time is regulation of FLC expression. Homologs of four genes in the *FHY3/FAR1* gene family were near QTL found associated with maturity traits indicating the *FHY3/FAR1* gene family is part of a mechanism maintaining variation in maturity traits in *B. tectorum*. Homologues of *FHY3* and *FRS6* were found 147 and 149 Kb from a QTL on *Bt7* (Table 3), respectively. A homologue of *FRS7* was near the QTL on *Bt3* associated with VPN (Table 3) *FRS5* discussed earlier is also member of the *FHY3/FAR1* gene family. The *FHY3/FAR1* gene family is comprised of 14 homologous genes, regulating transcription as a response to far red light (Lin and Wang, 2004). *FHY3*, *FRS6*, *FRS7* have been demonstrated to regulate flowering time in *A. thaliana*[48,57,58] making members of the *FHY3/FAR1* good candidate genes for maturity traits.

Additionally, a *Cytokinin dehydrogenase 2* (*CKX2*) homolog was identified as a candidate gene for the QTL at Bt3:48964562 associated with FRS. *CKX2* catalyzes the oxidation of cytokinins[59]. Cytokinins have been shown to regulate flowering time via transcriptional activation of Twin Sister of FT (*TST*)[60] and overexpression in members of the Cytokinin dehydrogenase (*CKX*) gene family have shown to delay flowering time in long day conditions[59] indicating that the *CKX2* homolog could be controlling maturity traits through the regulation of cytokinins in *B. tectorum*.

Discussion

The reference quality genome we have assembled in an invaluable resource for understanding the fundamental genetic controls that have facilitate one of the most successful invasive weeds in North America. Understanding the genetic basis of adaptive traits in *B. tectorum* will lead to improved management strategies. Using the genome we explored the genetic underpinnings of reproductive phenology traits in *B. tectorum*, revealing pathways and mechanisms contributing to adaptive plasticity which directly contributes to the species invasive spread across a wide range of environments. Indeed, our GWAS uncovered genetic mechanisms contributing to plasticity within the species. We identified significantly associated QTL for maturity traits that were near candidate genes responsible for controlling the photoperiod pathway, plant hormone regulation, and transcription factors triggered by far-red or UV light.

Our study indicates that not only is the genome very similar to barley (Fig. 2), but the adaptive control of reproductive phenology also closely mirrors barley. Both domesticated and wild barley are adapted to latitudinal clines, where reproductive phenology is controlled by genes responding to environmental cues[61]. Our identification of a *GA20OX* ortholog as a candidate gene for reproductive phenology corroborates previous work, where *GA20OX* loss of function was found to reduce PH and delay flowering[62]. Interestingly, *GA20OX* orthologs are semi-dwarfing genes implicated in the green revolution[63]. The semi-dwarfing gene we identified as a candidate gene could also explain the negative correlation between reproductive phenology traits and PH. Additionally, wild barley was found to adapt using variation in photoperiod response genes[61], supporting our findings where the FHY3/FAR1 photoperiod receptor gene family was implicated in controlling reproductive phenology.

Although *GA20OX* is a known controller of PH, its homolog was not identified by the GWAS as significant for PH. Plant height is an adaptive trait with large effect QTL conserved for maintaining phenotypic variation within in and between species of Poaceae[64]. Genome wide association studies in predominantly self-fertilizing wheat and rice identified QTLs associated with PH that explained 2.4% and 23.9%, respectively[65,66], indicating self-fertilizing species have QTL with relatively large effects. Our study also indicates that self-fertilizing species use a smaller number of large effect loci to control PH as we only identified a single loci explaining 16.9% of the phenotypic variation. The homologs of *XDH* and *YUC6* as candidate genes reflected the involvement of stress response and hormonal regulator genes controlling PH as found in other phylogenetically proximal grasses, such as wheat[67], barley[68] and oats[69].

Our study indicates *B. tectorum* is armed with a complex array of genetic mechanisms to create adaptive variation underlying reproductive phenology and PH which has facilitated its invasion into N. America and suggests that it is likely to continue to spread north into western Canada as climate change facilitates range expansion[70]. Identifying candidate *FHY3*, *FRS5*, *FRS6*, *FRS7*, *ABCB19*, *UVR8* and *PHL* genes that all act in response to light stimulus indicate that photoreceptor genes are critical for controlling variation in flowering time in *B. tectorum*. Photoreceptors controlling reproductive phenology will result in phenotypic plasticity because light signals of the environment will influence the underlying

pathways[71]Genetic control of traits was quite high, ranging from 0.99 to 0.94 for VPN and tiller number, respectively.

Furthermore, our results indicated the presence of small to moderate effect QTL controlling reproductive phenology traits, rather than singular large effect loci. The high heritability and moderate effect QTL detected for adaptive traits indicate that *B. tectorum* has already been adapting and will continue to adapt to a wide range of environments utilizing moderate sized QTL– as genetic recombination is very rare. Further investigation into the plasticity and regulation of flowering time in *B. tectorum* is needed to understand how local populations respond to stress or climate variation, and how the genetic variation we have discovered confers success in the arid, dry southern reaches of the American southwest and northern Mexico north to western Canada.

Materials And Methods

Plant Material and DNA extraction for Whole Genome Sequencing. For whole genome assembly, a single plant from the *B. tectorum* accession (FMH10) was grown hydroponically in an isolated, disease free growth chamber under a 12-h photoperiod. Growing temperatures ranged from 18°C (night) to 20°C (day). The hydroponic growth solution was based on MaxiBloom® Hydroponics Plant Food (General Hydroponics, Sevastopol, CA, United States) at a concentration of 1.7 g/L. FMH10 is a common clade accession (US Forest Service, Shrub Sciences Laboratory, Provo, UT, USA; Meyer et al., 2016). In preparation for PacBio CLR sequencing, high molecular weight DNA was extracted from 72-h dark-treated leaf samples using a CTAB-Qiagen Genomic-tip protocol as described by Vaillancourt and Buell[72].

Whole Genome Sequencing. For whole-genome sequencing, large-insert SMRTBell libraries (> 20 kb), selected using a SageElf (Sage Science, Inc., Beverly, MA, USA), were prepared according to standard manufacture protocols and sequenced at the BYU DNA Sequencing Center (Provo, UT, USA) using P6-C4 chemistry on a Sequel II instrument (Pacific BioSciences, Menlo Park, CA, USA). For whole genome polishing, DNA was sent to for Illumina HiSeq (2 X 150 bp) sequencing from standard 500-bp insert libraries. Trimmomatic v0.35 (ref. 73 was used to remove adapter sequences and leading and trailing bases with a quality score < 20 or with an average per-base quality of 20 over a four-nucleotide sliding window. After trimming, any reads shorter than 75 nucleotides in length were removed. Raw PacBio and Illumina reads have been deposited in GenBank.

Genome assembly, polishing and Hi-C scaffolding. A primary assembly of *B. tectorum* accession FMH10 was constructed using Canu v1.9[74] with default parameters (corMhapSensitivity = normal and corOutCoverage = 40). The primary assembly was polished twice with Illumina short reads using Arrow from the GenomicConsensus package in the Pacific BioSciences SMRT portal v5.1.0 followed by a single round of insertion/deletion correction using PILON v0.22[75]. Fresh leaf tissue from a single 3-week-old FMH10 plant was sent to Dovetail Genomics LLC (Santa Cruz, CA, USA) in preparation for construction of an Omni-C™ proximity-guided final chromosome-scale assembly. The Omni-C™ technology uses a novel approach to Hi-C library preparation via DNA digestion with a non-specific endonuclease to increase

uniformity and genomic coverage (https://dovetailgenomics.com/omni-c/). The libraries were prepared using a standard Illumina library prep followed by sequencing on an Illumina HiSeq X in rapid run mode. The HiRiSE™ scaffolder and the Omni-C™ library-based read pairs were used to produce a likelihood model for genomic distance between read pairs, which was used to break putative miss-joins and to identify and make prospective joins in primary contig assembly to produce the final chromosome scale reference assembly.

GWAS Plant Materials Collection. *Bromus tectorum* genotypes were obtained from Genome Resources Information Network (GRIN), field samples in eastern Washington from Lawrence 2018 et al.[76], and samples from natural areas in Montana (contributed by Lisa Rew, Montana State University) with 11, 64, and 46 samples, respectively. Each genotype was grown for one generation in the greenhouse to increase seed and verify purity. Six replicates of each line were vernalized at 4° Celsius with 10 h light per day for 53 d, then planted in 1.4-liter square pots. Three replicates, or blocks, were placed in one of two greenhouses. The plants were arranged in a completely randomized block design, with each block grown on a separate greenhouse bench. Supplemental lighting was used to keep day lengths at least 15 h per d.

Phenotyping. Plants in the greenhouse were observed daily to record reproductive phenology associated phenotypes. The phenotypes measured included days until first panicle visible (VPN, Feekes 10.1), days to first joint (J1, Feekes 6), days until first mature seed (FRS), days until 50% of seeds dry with awns angled outward (AWN_{50}), number of tillers, and height of the tallest panicle (PH). Number of tillers were counted for each plant at the end of the experiment before harvest. The height of the tallest panicle was measured in centimeters, measuring from the base of the plant to the tip of the longest panicle.

GWAS DNA Extraction and Resequencing. DNA was extracted using a bromide CTAB protocol as previously described[77]. Samples were diluted to a 50 ng/µl concentration. Genotyping by sequencing libraries were prepared for each sample by LGC Genomics (Berlin, Germany following Elshire et al. 2011[78] using the *MslI* restriction enzyme. Barcode adapters were ligated to each sample and the samples were put into 48-plex library plates. The polymerase chain reaction was used to amplify samples on the plates which were then sequenced using a single lane of Illumina NextSeq 500 V2. Approximately 1.5 million (2X150 bp) reads were generated per sample.

After sequencing, all the library groups were de-multiplexed with bcl2fastq v2.17.1.14 (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html) software allowing for up to two mismatches on the barcodes. Library groups were de-multiplexed further into separate samples according to the inline barcodes, where no mismatches were allowed. The adapter barcodes were clipped and reads less than 20 bases in length were discarded as were any reads where the 5’ end did not match the restriction enzyme cutting motif. Reads were quality trimmed from the 3’ end so that the average Phred quality score across ten neighboring bases > 20.

GWAS SNP Calling. De-multiplexed filtered reads for each sample were aligned to the de novo reference genome using BWA mem[79] with default settings for paired end reads. SAM files generated from the

alignments were converted to BAM files and sorted using SAMTOOLS[80]. The mpileup and call functions from Bcftools[81] were used to call SNPs. The vcf file generated from Bcftools was filtered so only SNP variants were kept, minor allele frequency (MAF) > 0.05, Missing Alleles < 25% and a QUAL score of at least 30 for each SNP, using Bcftools[80]. Scripts in R statistical programming language were used to read the vcf file into an allelic dosage table and filter out markers with more than two alleles or more than 5 heterozygous calls (*B. tectorum* is an autogamous species). Missing calls were imputed with a k^{th} nearest neighbor imputation using the "impute" package[82] from the Bioconductor project[83] in the R statistical programming language[84].

Linkage-disequilibrium. Linkage-disequilibrium (LD) was calculated on a pairwise basis between SNP on the same chromosome within 3 Mb using 1500 randomly sampled SNP from each chromosome. LD, measured as r^2_{sv}, was calculated using the method described by Mangin et al., 2012[85] that calculates the Pearson correlation between SNP but corrects for population structure and kinship in LD calculations implemented in the "LDcorSV" R package[86]. A genomic kinship matrix was estimated from the SNP data using the method from VanRaden 2008[87] implemented in GAPIT[88]. Landscape and Ecological Analysis (LEA)[89] was used in R to determine the optimal number of ancestral groups, and then calculate the admixture of each of these ancestral groups. Values of K, ranging from 1 to 20, were evaluated using the snmf function in LEA for cross entropy in ten replications and the lowest K near the lowest cross entropy was selected as the optimal K. The kinship matrix and population components from snmf described above were used with LDcorSV to correct for population structure, with distances above 3 Mb being discarded. The data from all the chromosomes was pooled together after filtering SNP by distance. The nlrq function of the quantreg R package[90] was used estimate an asymptotic decay by physical genetic distance for the 90th percent quantile of LD values. The LD decay was defined as the distance required for the LD (90th percent quantile of corrected r^2) to drop from its initial starting point to half-way between the starting point and the lower asymptotic limit ($LD_{90,1/2}$). $LD_{90,1/2}$ as a measure was found by simulation to be a more accurate estimate then $r^2 = 0.1$ as a measure of LD decay[91].

GWAS Analysis. GWAS was performed using the Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) algorithm[92] with the BLUEs for each phenotype as the response variables and the numeric SNP matrix as the genetic data. BLINK was ran using the GAPIT package[88] in R. Three principal components were chosen to be included as a covariate in the BLINK which represented the smallest number of principal components that controlled inflation on the P-P diagnostic plots generated by GAPIT. Potential candidate genes within a 1 Mb window, based on LD (Fig. 3) for each of the significant genomic regions identified in the GWAS analysis were manually identified from the previously described annotated gene set with functional annotations indicating association with maturity traits.

Declarations

Ethics approval and consent to participate

All relevant permissions were obtained for the collection of *B. tectorum* genotypes. The methods carried out in our manuscript were in accordance with the local, national and international guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author Contributions

SRR wrote the manuscript, designed experiments and analyzed data. PJM assembled and annotated the reference genome, analyzed data and edited the manuscript. CEC contributed to the assembly and annotation the reference genome, and edited the manuscript. ICB contributed to writing the manuscript, designed experiments and edited the manuscript.

Funding

This work was funded in part by the Washington Grain Commission and by U.S. Department of Agriculture National Institute of Food and Agriculture grant No. 2017-68002-26819.

Availability of Data and Materials

The raw sequences used for the *B. tectorum* genome assembly are deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive database under the BioProject PRJNA728981 with the following accession numbers: SRR14498212–SRR14498217 (PacBio reads), SRR14578284–SRR14578290 (Hi-C reads), SRR14578282 (Transcriptome) and SRR14498209–SRR14498211 (Polishing short reads). The raw reads for the resequencing panel are found in BioProject PRJNA728981 with the following NCBI accession numbers: SRR15308470–SRR15308851 (resequencing panel). Genome browsing and bulk data downloads, including annotations and BLAST analysis of the final proximity-guided assemblies are available at CoGe (https://genomevolution.org/coge/) with genome ID: id56596.

Code Availability

The main custom scripts and SNP data have been deposited in figshare (10.6084/m9.figshare.c.6116910)

Acknowledgments

We would like to acknowledge Amber Hauvermale for providing sequence data on *B. tectorum* that was used to conceive the experiment but was not included in analyses present in the manuscript. We would also like acknowledge Lisa Rew for providing seed for the *B. tectorum* genotypes from Montana that we re-sequenced for the GWAS.

Corresponding Author

Ian C. Burke (icburke@wsu.edu)

References

1. Bradley, B. A., Curtis, C. A., Fusco, E. J., Abatzoglou, J. T., Balch, J. K., Dadashi, S. & Tuanmu, N. Cheatgrass (*Bromus tectorum*) distribution in the intermountain western United States and its relationship to fire frequency, seasonality, and ignitions. *Biol Invasions* **20**, 1493–1506 (2018).
2. Balch, J. K., Bradley, B. A., D'Antonio, C. M. & Gomez-Dans, J. Introduced annual grass increases regional fire activity across the arid western USA (1980–2009). *Glob. Change Biol.* **19**, 173–183 (2012).
3. Pilliod, D. S., Welty, J. L. & Arkle, R. S. Refining the cheatgrass–fire cycle in the Great Basin: Precipitation timing and fine fuel composition predict wildfire trends. *Ecol. Evol.* **7**, 8126–8151 (2017).
4. Zimmer, S. N., Grosklos, G. J., Belmont, P. & Adler, P. B. Agreement and uncertainty among climate change impact models: a synthesis of sagebrush steppe vegetation projections. *Rangel Ecol Manag* **75**, 119–129 (2021).
5. Blackshaw, R. E. Downy brome (*Bromus tectorum*) density and relative time of emergence affects interference in winter wheat (*Triticum aestivum*). *Weed Sci.* **41**, 551–556 (1993).
6. Novak, S. J. & Mack, R. N. Genetic variation in *Bromus tectorum* (Poacea): comparison between native and introduced populations. Heredity 71, 167–176 (1993).
7. Mack, R. N. Invasion of *Bromus tectorum* L. into western North America: an ecological chronicle. *Agro-Ecosystems* **7**, 145–165 (1981).
8. Merrill, K. R., Meyer, S. E. & Coleman, C. E. Population genetic analysis of *Bromus tectorum* (Poaceae) indicates recent range expansion may be facilitated by specialist genotypes. *Am. J. Bot.* **99**, 529–537 (2012).
9. Wolkovich, E. M. & Cleland, E. E. The phenology of plant invasions: a community ecology perspective. *Front. Ecol. Environ.* **9**, 287–294 (2010).
10. Arnesen, S., Coleman C. E. & Meyer, S. E. Population genetic structure of *Bromus tectorum* in the mountains of western North America. *Am. J. Bot.* **104**, 879–890 (2017).
11. Merrill, K. R., Coleman, C. E., Meyer, S. E., Leger, A. L. & Collins, K. A. Development of single-nucleotide polymorphism markers for *Bromus tectorum* (Poaceae) from a partially sequenced transcriptome. Appl. Plant Sci. **4**, 1600068 (2016).

12. Meyer, S. E., Leger, E. A., Eldon, D. R. & Coleman, C. E. Strong genetic differentiation in the invasive annual grass *Bromus tectorum* across the Mojave– Great Basin ecological transition zone. Biol. Invasions **18**, 1611–1628 (2016).
13. Lawrence, N. C., Hauvermale, A. L. & Burke, I. C. Downy brome (*Bromus tectorum*) vernalization: variation and genetic controls. *Weed Sci.* **66**, 310–316 (2018).
14. Mack, R. N. & Pyke, D. A. The demography of *Bromus tectorum*: variation in time and space. *J. Ecol.* **71**, 69–93 (1983).
15. Wolkovich, E. M. & Cleland, E. E. The phenology of plant invasions: a community ecology perspective. *Front. Ecol. Environ.* **9**, 287–294 (2010).
16. Rice, K. J. & Mack, R. N. Ecological genetics of *Bromus tectorum*. I. A hierarchical analysis of phenotypic variation. *Oecologia* **88**, 77–83 (1991).
17. Meyer, S. E. Ecological genetics of seed germination regulation in *Bromus tectorum* L.. I. phenotypic variance among and within populations. *Oecologia* **120**, 27–34 (1999).
18. Meyer, S. E., Nelson, D. L. & Carlson, S. L. Ecological genetics of vernalization response in *Bromus tectorum* L. (Poaceae). *Ann. Bot.* 2004, <bvertical-align:super;>93</bvertical-align:super;>, 653–663 (2003).
19. Fernández-Calleja, M., Casas, A.M. & Igartua, E. Major flowering time genes of barley: allelic diversity, effects, and comparison with wheat. *Theor. Appl. Genet.* **134**, 1867–1897 (2021).
20. Mathews, S., Tsai, R. C., Kellogg, E. A. Phylogenetic structure in the grass family (Poaceae): evidence from the nuclear gene phytochrome b. *Am. J. Bot.* 2000, <bvertical-align:super;>87</bvertical-align:super;>, 96–107 (2000).
21. Fortune, P. M., Pourtau, N., Viron, N., Ainouche, M. L. Molecular phylogeny and reticulate origins of the polyploid *Bromus* species from Section Genea (Poaceae). *Am. J. Bot.* **95**, 454–464 (2008).
22. Pais, A. L., Whetten, R. W. & Xiang, Q. Y. Population structure, landscape genomics, and genetic signatures of adaptation to exotic disease pressure in *Cornus florida* L.— Insights from GWAS and GBS data. *J. Syst. Evol.* **58**, 546–570 (2020).
23. Kosch, T. A., Silva, C. N. S, Brannelly, L. A., Roberts, A. A., Lau, Q., Marantelli, G., Berger, L., Skerratt, L. F. Genetic potential for disease resistance in critically endangered amphibians decimated by chytridiomycosis. *Anim. Conserv.* **22**, 238–250 (2019).
24. Nichols, K. M., Kozfkay, C. C. & Narum, S. R. Genomic signatures among Oncorhynchus nerka ecotypes to inform conservation and management of endangered Sockeye Salmon. *Evol. Appl.* **9**, 1285–1300 (2016).
25. Wright, B. R. et al. A demonstration of conservation genomics for threatened species management. *Mol. Ecol. Resour.* **20**, 1526–1541 (2020).
26. Duntsch, L. et al. Polygenic basis for adaptive morphological variation in a threatened Aotearoa New Zealand bird, the hihi (*Notiomystis cincta*). *Proc. R. Soc. Lond. B Biol. Sci.* <bvertical-align:super;>287</bvertical-align:super;>, 20200948 (2020).

27. Hsiao, C., Chatterton, N. J., Asay, K. H. & Jensen, K. B. Molecular phylogeny of the Pooideae (Poaceae) based on nuclear RDNA (ITS) sequences. *Theoret. Appl.* Genetics **90**, 389–398 (1995).
28. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
29. Han, R. et al. Enhancing xanthine dehydrogenase activity is an effective way to delay leaf senescence and increase rice yield. *Rice* **13**, 16 (2020).
30. Kim, J. I. et al. Overexpression of Arabidopsis YUCCA6 in potato results in high- auxin developmental phenotypes and enhanced resistance to water deficit. *Mol. Plant* **6**, 337–349 (2013).
31. Rohilla M. et al. Genome-wide association studies using 50 K rice genic SNP chip unveil genetic architecture for anaerobic germination of deep-water rice population of Assam, India. *Mol. Genet. Genome.* **295**, 1211–1226 (2020).
32. Brookbank, B. P., Patel, J., Gazzarrini, S. & Nambara, E. Role of basal ABA in plant growth and development. Genes **12**, 1936 (2021).
33. Kim, J. I. et al. yucca6, a dominant mutation in Arabidopsis, affects auxin accumulation and auxin-related phenotypes. Plant Physiol. **45**, 722–735 (2007).
34. Jia, P., Bayaerta, T., Li, X. & Du G. Relationships between flowering phenology and functional traits in eastern Tibet alpine meadow. *Arct. Antarct. Alp. Res.* **43**, 585–592 (2011).
35. Bolmgren, K. & Cowan, P. D. Time - size tradeoffs: a phylogenetic comparative study of flowering time, plant height and seed mass in a north-temperate flora. *Oikos* **117**, 424–429 (2008).
36. Hazard, L., Betin, M. & Molinari, N. Correlated response in plant height and heading date to selection in perennial ryegrass populations. *J. Agron.* **98**, 1384–1391 (2006).
37. Brown, R. F. Tiller development as a possible factor in the survival of the two grasses, *Aristida armata* and *Thyridolepis mitchelliana*. *Rangel.* **4**, 34–38 (1982).
38. Williamson, M. A. et al. Fire, livestock grazing, topography, and precipitation affect occurrence and prevalence of cheatgrass (Bromus tectorum) in the central Great Basin, USA. *Biol. Invasions* **22**, 663–680 (2020).
39. Totland, O. Effects of temperature and date of snowmelt on growth, reproduction, and flowering phenology in the arctic/alpine herb, *Ranunculus glacialis*. *Oecologia* **133**, 168–175 (2002).
40. Hall, M. C. & Willis, J. H. Divergent selection on flowering time contributes to local adaptation in mimulus guttatus populations. *Evolution* **60**, 2466–2477 (2007).
41. Ågren, J. & Schemske, D. W. Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. *New Phytol.* **194**, 1112–1122 (2012).
42. Salomé, P. A. et al. Genetic Architecture of Flowering-Time Variation in *Arabidopsis thaliana*. *Genetics* **188**, 421–433 (2011).
43. Xu, Y. et al. Quantitative trait locus mapping and identification of candidate genes controlling flowering time in *Brassica napus* L. *Front. Plant Sci.* **11**, 626205 (2021).
44. Buckler, E. S. et al. The genetic architecture of maize flowering time. *Science* **325**, 5941 (2009).

45. Sun, X. et al. The *Oryza sativa* regulator HDR1 associates with the kinase OsK4 to control photoperiodic flowering. *PLoS Genet* **12**, e1005927 (2016).
46. Dieterle, M. et al. Molecular and functional characterization of Arabidopsis Cullin 3A. *Plant J.* **41**, 386–399 (2005).
47. Rieu, I. et al. The gibberellin biosynthetic genes AtGA20ox1 and AtGA20ox2 act, partially redundantly, to promote growth and development throughout the Arabidopsis life cycle. *Plant J.* **53**, 488–504 (2008).
48. Zhao, H. *et al.* The ATP-binding cassette transporter ABCB19 regulates postembryonic organ separation in *Arabidopsis*. *PLoS One* **8**, e60809 (2013).
49. Lin, R. & Wang, H. Arabidopsis FHY3/FAR1 Gene family and distinct roles of its members in light control of Arabidopsis development. *Plant Physiol.* **136**, 4010–4022 (2004).
50. McKown, A. D. et al. Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *Populus trichocarpa*. *New Phytol.* **201**, 1263–1276 (2014).
51. Magome, H., Yamaguchi, S., Hanada, A., Kamiya, Y. & Oda, K. dwarf and delayed- flowering 1, a novel Arabidopsis mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. *Plant J.* **37**, 720–729 (2004).
52. Endo, M., Tanigawa, Y., Murakami, T., Araki, T. & Nagatani, A. PHYTOCHROME- DEPENDENT LATE-FLOWERING accelerates flowering through physical interactions with phytochrome B and CONSTANS. *Proc. Natl. Acad. Sci. U.S.A.* <bvertical-align:super;>110</bvertical-align:super;>, 18017–18022 (2013).
53. Škiljaica, A. et al. The protein turnover of Arabidopsis BPM1 is involved in regulation of flowering time and abiotic stress response. *Plant. Mol. Biol.* **102**, 359–372 (2020).
54. Dotto, M., Gómez, M. S., Soto, M. S. & Casati, P. UV-B radiation delays flowering time through changes in the PRC2 complex activity and miR156 levels in *Arabidopsis thaliana*. *Plant Cell Environ.* **41**, 1394–1406 (2018).
55. He, Y., Doyle, M. R. & Amasino, R. M. PAF1-complex-mediated histone methylation of FLOWERING LOCUS C chromatin is required for the vernalization-responsive, winter-annual habit in Arabidopsis. *Genes Dev.* **18**, 2774–2784 (2004).
56. Finnegan, E. J., Sheldon, C. C., Jardinaud, F., Peacock, W. J. & Dennis, E. S. A cluster of Arabidopsis genes with a coordinate response to an environmental stimulus. *Curr. Biol.* **14**, 911–916 (2004).
57. Ritter, A. et al. The transcriptional repressor complex FRS7-FRS12 regulates flowering time and growth in Arabidopsis. *Nat Commun.* **8**, 15235 (2017).
58. Xie, Y. et al. FHY3 and FAR1 Integrate light signals with the miR156-SPL module- mediated aging pathway to regulate Arabidopsis flowering. *Mole. Plant* **13**, 483–498 (2020).
59. Werner, T. et al. Cytokinin-deficient transgenic *Arabidopsis* plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. *Plant Cell* **15**, 2532–2550 (2003).

60. D'Aloai, M. et al. Cytokinin promotes flowering of Arabidopsis via transcriptional activation of the *FT* paralogue *TSF*. *Plant J.* **65**, 972–979 (2011).
61. Wiegmann, M. et al. Barley yield formation under abiotic stress depends on the interplay between flowering time genes and environmental cues. *Sci. Rep.* **9**, 6397 (2019).
62. Yan, H. et al. Position validation of the dwarfing gene *Dw6* in oat (*Avena sativa L.*) and its correlated effects on agronomic traits. Front. Plant Sci. **12**, 668847 (2021).
63. Hedden, P. The genes of the Green Revolution. *Trends Genet.* **19**, 5–9 (2003).
64. Lin, Y. R., Schertz, K. F. & Paterson, A. H. Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. *Genetics* **141**, 391–411 (1995).
65. Tessmann, E. W. & Sanford, D. A. V. GWAS for fusarium head blight related traits in winter wheat (*Triticum Aestivum* L.) in an artificially warmed treatment. *Agronomy* **8**, 68 (2018).
66. Biscarini, F. et al. Genome-wide association study for traits related to plant and grain morphology, and root architecture in temperate rice accessions. *PLoS One* **11**, e0155425 (2016).
67. Han, R. et al. Enhancing xanthine dehydrogenase activity is an effective way to delay leaf senescence and increase rice yield. *Rice* **13**, 16 (2020).
68. Kim, J. I. et al. Overexpression of Arabidopsis YUCCA6 in potato results in high- auxin developmental phenotypes and enhanced resistance to water deficit. *Mol. Plant* **6**, 337–349 (2013).
69. Rohilla M. et al. Genome-wide association studies using 50 K rice genic SNP chip unveil genetic architecture for anaerobic germination of deep-water rice population of Assam, India. *Mol. Genet. Genome*. **295**, 1211–1226 (2020).
70. DeBeer, C. M., Wheater, H. S., Carey, S. K. & Chun, K. P. Recent climatic, cryospheric, and hydrological changes over the interior of western Canada: a review and synthesis. Hydrol. Earth Syst. Sci. **20**, 1573–1598 (2016).
71. Casal, J. J., Fankhauser, C., Coupland, G. & Blázquez, M. A. Signalling for developmental plasticity. Trends Plant Sci. **9**, 309–314 (2004).
72. Vaillancourt, B. & Buell, C. R. High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing. Preprint at https://www.biorxiv.org/content/10.1101/783159v2.full.pdf (2019).
73. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
74. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
75. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* **9**, e112963 (2014).
76. Lawrence, N. C., Hauvermale, A. L., Dhingra, A. & Burke, I. C. Population structure and genetic diversity of *Bromus tectorum* within the small grain production region of the Pacific Northwest. *Ecol. Evol.* **7**,

8316–8328 (2018).

77. Porebski, S. L., Bailey, G. & Baum, B. R. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol Components. *Plant Mole. Biol. Rep.***15**, 8–15 (1997).
78. Elshire, R. J. et al. A Robust, Simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS One* **6**, e19379 (2011).
79. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA- MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).
80. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
81. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
82. Hastie, T., Tibshirani, R., Narasimhan, B. & Chu, G. Impute: imputation for microarray data. R package version 1.58.0. (2019).
83. Gentleman, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
84. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/ (2019).
85. Mangin, B. et al. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* 2012, <bvertical-align:super;>108</bvertical-align:super;>, 285–291 (2012).
86. Desrousseaux, D., Sandron, F., Siberchicot, A., Cierco-Ayrolles, C. & Mangin, B. LDcorSV: linkage disequilibrium corrected by the structure and the relatedness. R package version 1.3.2. (2017).
87. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 2008, <bvertical-align:super;>91</bvertical-align:super;>, 4414–4423 (2008).
88. Lipka, et al. Genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399 (2012).
89. Frichot, E. & Francois, O. LEA: an R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929 (2015).
90. Koenker, R. Qauntreg: quantile regression. R package version 5.83. (2021).
91. Vos, P. G. et al. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theor. Appl. Genet.* **130**, 123–135 (2017).
92. Huang, M., Liu, X., Zhou, Y., Summers, R. M., Zhang, Z. BLINK: A Package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* **8**, giy154 (2019).
93. Todd, J. J. & Vodkin, L. O. Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. *Plant Cell* **8**, 687–699 (1996).

94. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
95. Page, J. T., Liechty Z. S., Huynh, M. D. & Udall, J. A. BamBam: genome sequence analysis tools for biologists. *BMC Res. Notes* **7**, 829 (2014).
96. Lee, T., Guo, H., Wang, X., Kim, C. & Paterson, A. H. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genet.* **15**, 162 (2014).
97. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
98. Kim, D. Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
99. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
100. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* **12**, 491 (2011).
101. Smit, A. F. A. & Arian, F. A. RepeatModeler Open-1.0. (2008).
102. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-4.0. (2015).
103. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. 3210–3212 (2015).
104. Cox, A. V. et al. Comparison of plant telomere locations using a PCR-generated synthetic probe. *Annal. Bot.* **72**, 239–247 (1993).
105. Krzywinski, M. et al. Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
106. Bates, D., Mächler, M., Bolker B. & Walker, S. Fitting linear mixed-effects models using lme4. Preprint at https://arxiv.org/abs/1406.5823 (2014).
107. Cullis B. R., Smith, A. B. On the design of early generation variety trials with correlated data. *J. Agric. Biol. Environ. Stat.* **11**, 381 (2006).
108. Ziyatdinov, A. et al. lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinform.* **19**, 68 (2018).

Tables

Table 1. Description of the variation and consistency of phenotypic data across lines in greenhouse trials. Mean ($BLUE_{\mu}$), Standard Deviation ($BLUE_{sd}$), Minimum ($BLUE_{min}$), and Maximum ($BLUE_{max}$) of Best Linear Unbiased Estimates (BLUE) and Reliability (H^2) of the phenotypic traits measured in the greenhouse.

Trait	H^2	$BLUE_{\mu}$	$BLUE_{sd}$	$BLUE_{min}$	$BLUE_{max}$
Height	0.96	65.57	9.82	36.5	89
Tillers	0.94	2.88 (19.05)	0.37	1.70(5.5)	3.65(38.5)
VPN[1]	0.99	3.22 (27.89)	0.49	1.48(4.41)	4.28(72.00)
J1[2]	0.98	3.58(38.70)	0.38	0.38(18.50)	4.43(83.77)
FRS[3]	0.96	4.12(62.94)	0.2	3.76(43.15)	4.72(112.5)
AWN_{50}[4]	0.97	4.23(70.46)	0.21	3.89(70.46)	4.77(118.1)

[1] Days to first visible panicle
[2] Days to first visible joint
[3] Days to first ripe seed
[4] Days to 50% ripe seed

Table 2. Significant SNP detected by BLINK, where trait includes plant height (PH) (FRS), (VPN), or (AWN_{50}); the chromosome the trait appears to be associated with; the minor allele frequency (MAF) the percent of variation explained by the QTL, and the P-value of the SNP after FDR correction for multiple testing.

Trait	Chromosome	Minor allele frequency	Variation Explained (%)	P-value
PH[1]	6	0.44	16.9	1.360E-05
FRS[2]	7	0.29	7.2	5.280E-09
FRS	1	0.16	1.4	1.582E-04
FRS	1	0.22	1.8	1.786E-04
FRS	3	0.21	1.2	1.786E-04
FRS	1	0.2	4.1	2.823E-04
FRS	2	0.25	1.9	4.166E-04
FRS	4	0.058	0.6	1.026E-03
FRS	3	0.15	4	1.836E-03
FRS	6	0.2	2.3	4.110E-02
VPN[3]	7	0.26	16.3	3.630E-07
VPN	3	0.31	4.3	8.210E-07
VPN	4	0.42	4.3	1.920E-05
VPN	6	0.2	2.9	4.830E-03
VPN	7	0.29	4.7	1.046E-02
VPN	2	0.25	3.3	3.485E-02
VPN	3	0.45	3.8	3.797E-02
VPN	3	0.49	3.4	3.797E-02
VPN	3	0.49	3.4	3.797E-02
AWN_{50}[4]	1	0.2	6.5	6.760E-05
AWN_{50}	5	0.37	5.6	1.637E-04
AWN_{50}	6	0.2	4.8	5.973E-03
J1[5]	6	0.35	6.7	2.070E-07
J1	1	0.2	8.1	3.870E-05
J1	1	0.35	4.2	1.909E-03

[1] Plant Height

[2] Days to first ripe seed
[3] Days to first visible panicle
[4] Days to 50% ripe seed
[5] Days to first visible joint

Table 3. Candidate genes within 250 kbp of significant SNP identified by GWAS, with two ortholog species *Arabidopsis thaliana* and *Oryza sativa*.

Trait(s)	Gene	Ortholog Species	Associated SNP[1]	Distance From SNP (Kb)
PH[2]	XDH	*O. sativa*	Bt6_301800092	29
PH	YUC6	*A. thaliana*	Bt6_301800092	242
FRS[3], VPN[4]	FHY3	*A. thaliana*	Bt7_347158582	149
VPN	FRS7	*A. thaliana*	Bt3_383280959	64
FRS, VPN	FRS6	*A. thaliana*	Bt7_347158582	147
FRS	FRS5	*A. thaliana*	Bt1_173453655	78
FRS	DREB1F	*O. sativa*	Bt1_173453655	236
FRS	CKX2	*O. sativa*	Bt3_48964562	55
FRS, VPN	CUL3A	*A. thaliana*	Bt2_9403921	29
FRS, VPN	ABCB19	*A. thaliana*	Bt2_9403921	13
FRS, VPN	GA20OX2	*O. sativa*	Bt2_9403921	133
FRS, VPN, AWN_{50}[5]	HDR1	*O. sativa*	Bt6_8628087	212
J1[6]	HDR1	*O. sativa*	Bt6_8417488	12
VPN	PHL	*A. thaliana*	Bt3_2025814	240
VPN	BPM1	*A. thaliana*	Bt3_2025814	120
VPN	BPM2	*A. thaliana*	Bt3_2025814	177
VPN	UVR8	*A. thaliana*	Bt3_324738958	113
VPN	UFC	*A. thaliana*	Bt3_324738958	64
VPN	VIP2	*A. thaliana*	Bt3_324738958	189

[1] Single nucleotide polymorphism

[2] Plant height
[3] Days to first ripe seed
[4] Days to first visible panicle
[5] Days to 50% ripe seed
[6] Days to first visible joint

Figures

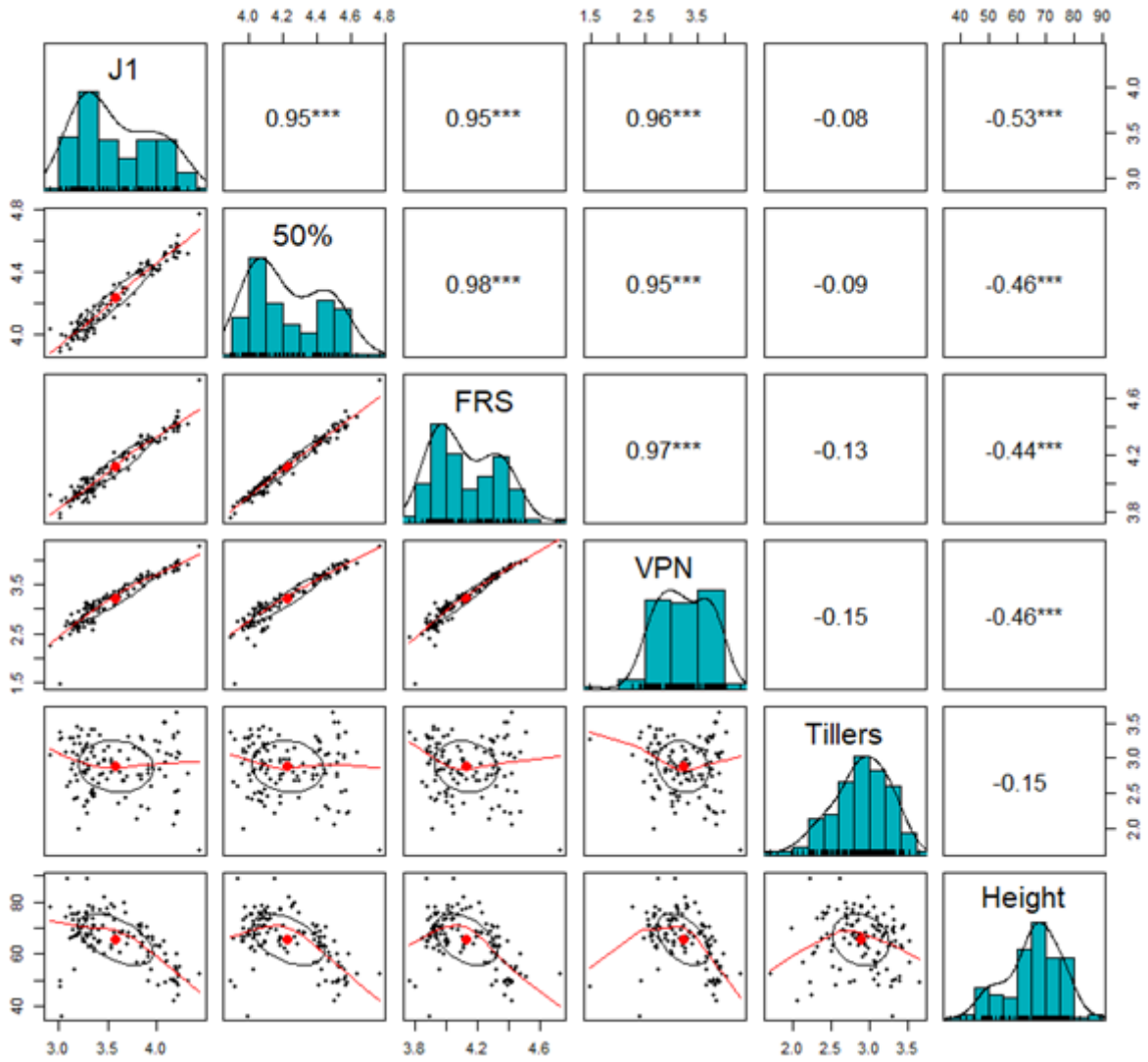

Figure 1

Distributions, Spearman correlations, and scatter plots of the genotype BLUEs for height (cm), number of tillers, days until first panicle (VPN), days until first joint (J1), days until first ripe seed (FRS) used in the GWAS analysis. Histograms with density ticks and a smoothing line are on the diagonals. The upper diagonal contains the Spearman correlation between BLUEs of traits with "***" denoting a significant level ($p<0.0001$) of correlation between traits. The lower diagonal contains scatter plots between traits

with center dot, a centroid with a standard deviation of 1 and locally weighted smoothing (LOESS) smoothing curve.

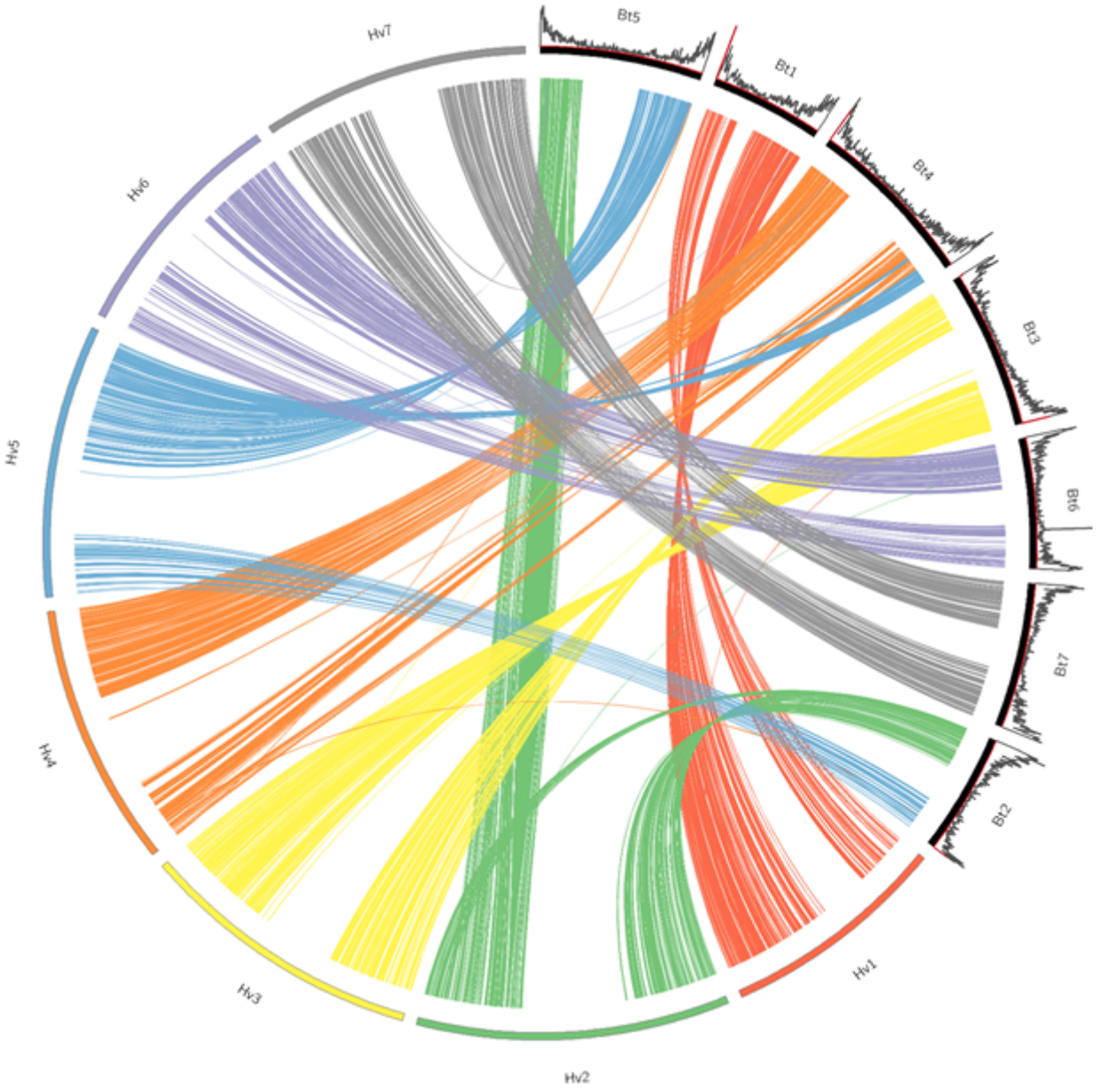

Figure 2

Circos synteny plot depicting the synteny between the *Bromus tectorum* draft reference genome and a previously published barley genome. Strips and barley chromosomes colored in reference to the corresponding barley chromosomes. Histograms above *B. tectorum* chromosomes are coded black for gene density and red for telomere repeats for each 1 mb window of the chromosome.

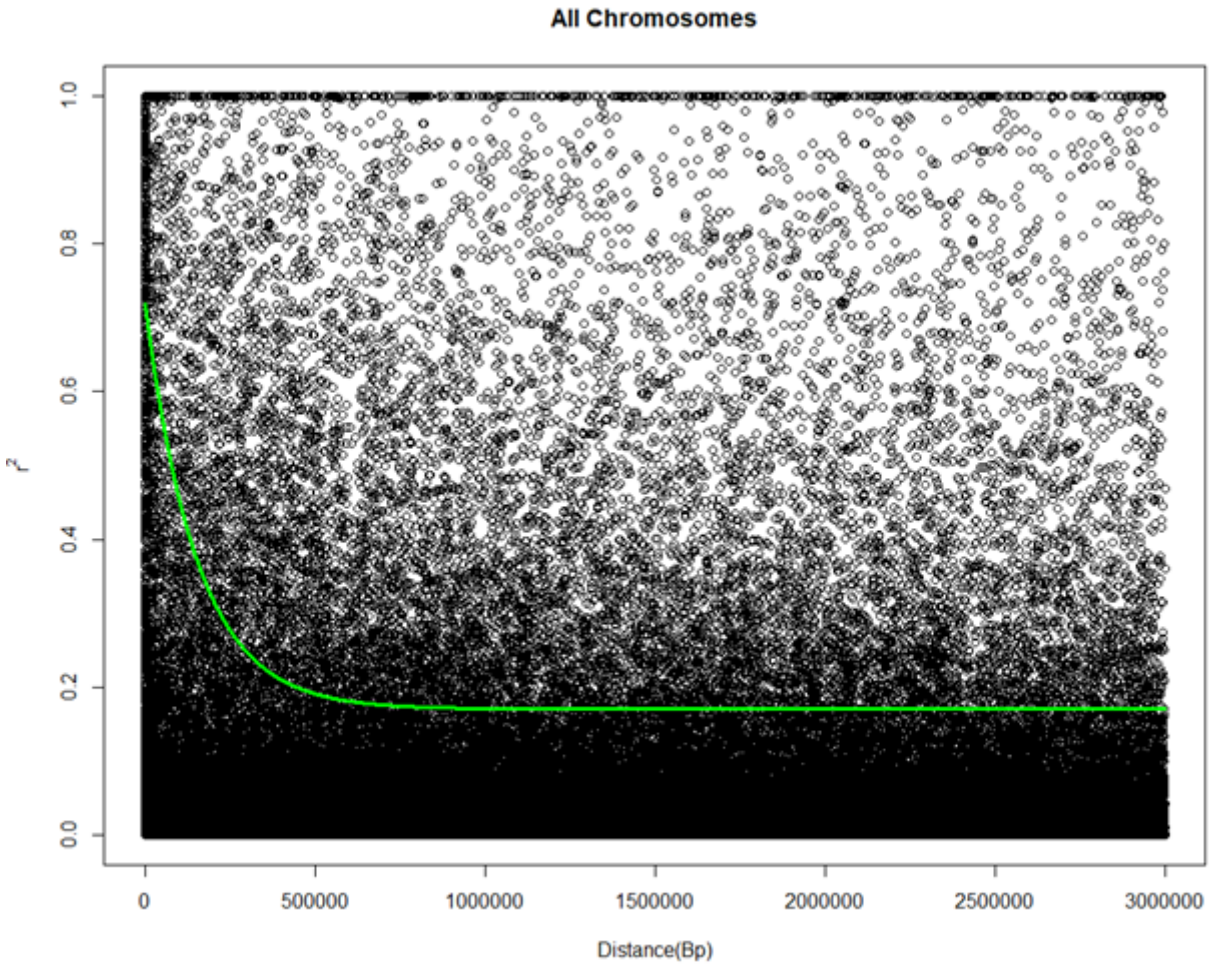

Figure 3

Asymptotic decay of linkage disequilibrium (LD) as a function of distance in *Bromus tectorum*. Green line is the asymptotic decay curve of best fit for the 90thpercentile of LD values.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformationandtablesready4SubmissionFIXED.docx