Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Extensive structural variation in the Bowman-Birk inhibitor family in common wheat (*Triticum aestivum L.*)

Yucong Xie
Colorado State University
Karl Ravet
Colorado State University
Stephen Pearce (Stephen.Pearce@colostate.edu)
Colorado State University https://orcid.org/0000-0002-1794-7618

Research article

Keywords: Protease inhibitor, Bowman-Birk inhibitor, Tandem duplication, Biotic stress, Wheat.

Posted Date: November 16th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-108242/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at BMC Genomics on March 25th, 2021. See the published version at https://doi.org/10.1186/s12864-021-07475-8.

Abstract

Background Bowman-Birk inhibitors (BBI) are a family of serine-type protease inhibitors that modulate endogenous plant proteolytic activities during different phases of development. They also inhibit exogenous proteases as a component of plant defense mechanisms, and their overexpression can confer resistance to phytophagous herbivores and multiple fungal and bacterial pathogens. Dicot BBIs are multifunctional, with a “double-headed” structure containing two separate inhibitory loops that can bind and inhibit trypsin and chymotrypsin proteases simultaneously. By contrast, monocot BBIs have a non-functional chymotrypsin inhibitory loop, although they have undergone internal duplication events giving rise to proteins with multiple BBI domains.

Results We used a Hidden Markov Model (HMM) profile-based search to identify 57 BBI genes in the common wheat (*Triticum aestivum* L.) genome. The BBI genes are unevenly distributed, with large gene clusters in the telomeric regions of homeologous group 1 and 3 chromosomes that likely arose through a series of tandem gene duplication events. The genomes of wheat progenitors also contain contiguous clusters of BBI genes, suggesting this family underwent expansion before the domestication of common wheat. However, the BBI gene family varied in size among different cultivars, showing this family remains dynamic. Because of these expansions, the BBI gene family is larger in wheat than other monocots such as maize, rice and *Brachypodium*.

We found BBI proteins in common wheat with intragenic homologous duplications of cysteine-rich functional domains, including one protein with four functional BBI domains. This diversification may expand the spectrum of target substrates. Expression profiling suggests that some wheat BBI proteins may be involved in regulating endogenous proteases during grain development, while others were induced in response to biotic and abiotic stresses, suggesting a role in plant defense.

Conclusions Genome-wide characterization reveals that the BBI gene family in wheat is subject to a high rate of homologous tandem duplication and deletion events, giving rise to a diverse set of encoded proteins. This information will facilitate the functional characterization of individual wheat BBI genes to determine their role in wheat development and stress responses and their potential application in breeding.

Background

Plant proteases play vital roles in diverse biological processes by modulating programmed cell death, nutrient remobilization and defense responses [1]. Their activity is regulated by different classes of protease inhibitors (PIs) which bind to their protease substrates either through an irreversible trapping reaction or a tight-binding reaction [2–4]. In plants, PIs regulate the activity of endogenous proteases to prevent proteolytic degradation, for example, by controlling the mobilization of storage proteins in seeds and kernels, and regulating senescence [5, 6]. They also play important roles in plant defense by regulating the activity of exogenous proteases from different types of pests and pathogens to prevent

cellular damage [7]. In response to insect feeding, plant PIs are released into the insect's guts and inhibit digestive protease enzymes, which can prevent nutrient absorption, retarding their growth and development [8]. Plant PIs are also induced by effector triggered immunity in response to bacterial and fungal pathogens to inhibit their proteolytic enzymes [9–11]. PIs are categorized into four broad classes according to their target protease specificity: serine PI (serpins), cysteine PI (cystatins), aspartic acid PI (pepstatins), and metallo-carboxy PI [2]. PIs are further classified into types, families and clans to reflect their evolutionary relationships based on sequence homology, structural variation and biochemical function [12–14]. The latest PI classifications are maintained in the MEROPS database [15].

Bowman-Birk inhibitors (BBIs) are a family of serine-type PIs in MEROPS family I12, clan IF, that inhibit trypsin and chymotrypsin protease activity via the tight-binding reaction mechanism [16, 17]. Members of the BBI family are best known for their role in plant defense against phytophagous insects, and have been used to engineer insect-resistant transgenic crops [18]. Overexpression of a cowpea trypsin inhibitor gene, which encodes a BBI protein, confers resistance to insects in the orders Coleoptera and Lepidoptera in tobacco [19], rice [20], and wheat [21]. Several BBI proteins also exhibit trypsin-like protease inhibition against fungal pathogens including *Mycosphaerella arachidicola, Fusarium oxysporum*, and *Botrytis cinerea* [22, 23], *Fusarium culmorum* [24] and *Pyricularia oryzae* [25], as well as bacterial pathogens such as *Xanthomonas oryzae* pv. *Oryzae* [26]. One rice BBI, APIP4, interacts at the protein level with both a fungal effector and host NLR receptors as part of the innate immune response, and plants carrying loss-of-function mutations in this gene exhibit increased susceptibility to *Magnaporthe oryzae* [27]. In wheat, genetic mapping studies identified putative BBI genes as candidates for seedling resistance to tan spot [28] and Fusarium head blight [29]. There is also evidence that BBIs play roles in more diverse processes, such as tolerance to salinity [30], oxidative [31], and drought stress [32, 33], and regulating Fe uptake via an unknown mechanism [34].

First discovered in soybean in 1946 [35], BBIs had until recently only been described in the Fabaceae and Poaceae families [36]. The BBIs are now known to be widely distributed in angiosperms [36–38], and evolutionary and phylogenetic analyses suggest they share a common ancestral sequence [38]. The characterization of five BBIs in *Selaginella moellendorffii,* the oldest known extant vascular plant, show that this ancestral protein has a characteristic "double-headed" structure with two homologous and spatially separated inhibitory loops within one BBI domain [38]. Conserved inhibitory loops form reactive motifs providing dual specificity [36]. BBI domains are also characterized by a series of conserved Cysteine (Cys) residues, which form disulfide bridges to provide structural stability required to maintain inhibitory loop conformation [36, 37]. The mutation of a single conserved Cys residue forming a disulfide bridge is sufficient to abolish the activity of either inhibitory loop [39]. The Cys-formed inhibitory loops contain reactive domains composed of variable amino acids responsible for binding to trypsin and to chymotrypsin, including two conserved residues P1 and P1' that are particularly important in determining protease substrate specificity [36]. BBI proteins also commonly have a hydrophobic signal peptide (SP) at their N-terminus, with high sequence diversity among different BBIs [40, 41]. The SP is required for BBI protein translocation and secretion into the extracellular space, although it is not necessary for protease inhibition since the inhibitory loops can function independently of the rest of the BBI protein [42]. There is

also evidence that BBI proteins can act in the nucleus [34]. All characterized BBI proteins in dicotyledonous plants have a conserved “double-headed” structure with a consistent molecular weight of approximately 8 kDa [36–38, 43].

By contrast, almost all BBIs in monocotyledonous plants lack conserved Cys residues in the second inhibitory loop that are required to inhibit chymotrypsin, leading to a “single-headed” structure so that each BBI domain consists of only one functional reactive loop to inhibit trypsin activity [36]. The only known exceptions are three “double-headed” BBIs in the banana (*Musa acuminate*) genome, indicating that the “single-headed” BBI structure originated since the monocot and dicot lineages diverged [38]. Evolutionary models indicate that monocot BBIs underwent internal domain duplications within a single protein that resulted in multiple inhibitory loops [25, 36, 44]. Previous studies divided monocot BBI proteins into six groups (MI-I to MI-VI) on the basis of their functional domain number and the number and position of conserved Cys residues [10, 36, 45]. To simplify, these six BBI models in monocots can be grouped into three broad classes; one comprised of 8 kDa proteins with a single functional domain (groups MI-I, MI-II, and MI-III), a second class with a molecular weight of approximately 16 kDa and a duplicated single-inhibitory loop (groups MI-IV and MI-V) and a final category of larger proteins with three tandemly duplicated BBI domains. While the first two classes are widespread in monocots, only three rice BBIs have been described which fall into the final class [25].

Genome-wide studies of the BBI gene family have been performed in rice [25], common bean [46] and other angiosperms [38]. However, to date, only three BBIs have been characterized in common wheat (*Triticum aestivum* L.), a crop which provides approximately 20% of the calories and proteins consumed by the human population [47]. Of the three BBI proteins isolated from wheat germ, IBB1 has two homologous functional domains, each with one functional inhibitory loop [48, 49], whereas IBB2 and IBB3 have only one functional domain [48, 50]. These three BBIs inhibit protease activity, control protein metabolism during wheat kernel development and germination, and inhibit fungal trypsin-like activity and hyphal growth [51]. Three other putative genes with sequence homology to BBIs (*wali3, wali5,* and *wali6*) were isolated as cDNAs from wheat root tips [30, 52, 53]. These putative BBI genes are transcriptionally induced by wounding or by the imposition of toxic metal stress, but their function against protease was not tested [30, 53].

The identification of wheat BBI genes is complicated by the high frequency of residue substitution and sequence variability among encoded proteins, and the complexity of the wheat genome. Common wheat is an allopolyploid (genomes AABBDD) produced from two separate hybridization events. The first occurred approximately 0.5 to 0.9 million years ago between *T. urartu* (AA) and an unknown species related to *Aegilops speltoides* to form the tetraploid wild emmer wheat *T. turgidum* ssp. *diccocoides* (AABB). A second hybridization event between *T. turgidum* ssp. durum and *Ae. tauschii* (DD) gave rise to common wheat, approximately 10,000 years ago [54].

In the current study, we used a Hidden Markov Model (HMM)-based approach to describe the BBI gene family in common wheat, revealing it to be larger than in other monocot species. We found evidence of

extensive gene duplications throughout wheat's evolutionary history, as well as internal duplications that further diversified the functional BBI domains of individual proteins. The findings from our study highlight the extent of variation in the BBI gene family in the Triticeae lineage and will facilitate their functional characterization to explore how this diversity impacts wheat development and plant defense.

Results

Bowman-Birk inhibitor genes are unevenly distributed in the common wheat genome

We identified 57 BBI genes in the hexaploid common wheat genome using a three-step HMM-based approach outlined in Fig. 1. We first used the HMM profile for BBI (Pfam: PF00228, downloaded from the Pfam database) to search the IWGSC RefSeq v1.1 protein database and identified 39 BBI proteins. We generated a new HMM profile based on the alignment of these 39 sequences and used this in a second search against the same protein database to identify 62 BBI proteins, including 23 that were not found in the first step. We performed HMMscan on each protein and excluded five sequences that lacked a BBI Pfam domain (Additional file 2, Table S1). A final search using an HMM profile built from an alignment of the remaining 57 BBIs did not yield any additional proteins, confirming this is a comprehensive list of annotated BBI proteins in the wheat landrace 'Chinese Spring' (Additional file 2, Table S1).

We manually adjusted the start codon position for five BBIs to match homologous sequences (Additional file 2, Table S2). After manual curation, 50 full-length BBIs are predicted to have an N-terminal SP domain, with cleavage positions ranging from 15 to 30 amino acids. Seven N-terminally truncated BBIs are predicted to lack a functional SP domain (Additional file 2, Table S1).

The 57 BBIs include three genes (TraesCS3A02G046000, TraesCS3B02G036400, and TraesCS1B02G025900) that encode previously characterized BBI proteins - IBB1, IBB2, and IBB3 (Additional file 2, Table S3) [48, 50]. Three other previously described putative BBI genes (*wali3, wali5* and *wali6* [52, 53]) were not found among the 57 BBIs. An HMMscan analysis of the corresponding full-length proteins (TraesCS1D02G265900, TraesCS1D02G265800 and TraesCS1B02G276900) revealed that they did not contain a BBI domain, indicating these genes do not encode functional BBI proteins (Additional file 2, Table S3).

Wheat BBI genes are unevenly distributed across the genome with two gene triads on chromosomes 4 and 5 and large clusters on homeologous group 3 (36 BBIs) and group 1 chromosomes (15 BBIs) (Fig. 2a, b). The BBI genes in these clusters are separated by short physical distances and in several instances include adjacent BBIs, suggesting they arose through tandem gene duplication events (Fig. 2a, b). For example, the ten BBIs on chromosome 3A span a region of just 270 kb and include four adjacent BBIs (Fig. 2b). All wheat BBIs were located in the telomeric regions (R1 and R3) of their respective chromosomes (Fig. 2a).

This pattern of gene duplication is consistent with homology analysis that divided the 57 BBIs into six homeologous categories (Table 1). Overall, 21 BBI genes (36.8% of the total) formed seven complete

triads (1:1:1 for A:B:D genome), close to the 35.8% for all wheat genes in the genome [55]. By contrast, 14% of BBI genes form groups characterized by gene duplication (n:1:1/1:n:1/1:1:n) compared to 5.7% of all wheat genes [55] (Table 1). In addition, one group of genes consisted of four tandemly duplicated genes on chromosome 1B (0:4:0), while on chromosome 3, one group exhibited duplications of both the A and B homeologs (2:2:1) (Table 1; Additional file 2, Table S4).

Table 1 - Homeologous group identification and categorization of the BBI gene family in wheat

Category number	Homeologous group (A:B:D)	Number of groups	Number of genes	% of genes
1	1:1:1	7	21	36.8
2	2:1:1 and 1:2:1	2	8	14
3	1:1:0 and 0:1:1	2	4	7
4	0:4:0	1	4	7
5	2:2:1 and 2:0:2	2	9	15.8
6	Singletons	11	11	19.4
-	Total	25	57	100

To determine whether these duplication events affected the selective pressure on BBI genes, we performed a Ka/Ks ratio analysis to calculate the sequence divergence rate for the clusters of BBIs on individual homeologous group 1 and 3 chromosomes. A ratio of non-synonymous (Ka) to synonymous (Ks) nucleotide changes greater than one indicates divergent function of two genes, whereas a Ka/Ks ratio of less than one indicates purifying selection and conserved function. The Ka/Ks ratios for pairwise comparisons of BBI genes on homeologous group 1 chromosomes were all less than one, except for one branch on chromosome 1D between TraesCS1D02G020600 and TraesCS1D02G018700LC that had a value of 1.17 (Additional file 1, Fig. S1). By contrast, eight branches on homeologous group 3 chromosomes had Ka/Ks values greater than one, including four branches on 3A, two branches on 3B, and two branches on 3D (Additional file 1, Fig. S1).

Overall, our analysis shows that the BBI family in wheat is unevenly distributed across the genome and includes large gene clusters in the telomeric regions of homeologous group 1 and group 3 chromosomes. The distribution of the genes in these clusters suggest they originated from paralogous expansion through tandem duplication events.

BBI genes underwent extensive tandem duplications in the Triticeae

We next compared the BBI family in wheat with other monocot species. Using the same approach and criteria (Fig. 1), we identified six BBIs from *Brachypodium* (*B. distachyon*), seven from maize (*Z. mays*),

eleven from rice (*O. sativa*), and sixteen from barley (*H. vulgare*) (Fig. 3a). A full list of BBIs from each species is provided in Additional file 2, Table S5. Considering its hexaploid genome, common wheat has an average of 19 BBI genes per diploid genome, 3.2-fold more than *Brachypodium*, 2.7-fold more than maize, 1.7-fold more than rice, but just 1.2-fold more than barley (Fig. 3b).

To explore the genetic relationships between BBIs in these species, we constructed a phylogenetic tree from all identified proteins. The tree separated wheat BBIs into three broad clades, each of which also contained BBIs from other species, except clade A that does not contain maize BBIs (Fig. 3c). Clade A clustered all wheat BBIs located on homeologous group 1 and 5 chromosomes. Clade B included the majority of wheat BBIs located on homeologous group 3 chromosomes, with the remainder clustered in clade C together with the BBI gene triad from chromosome 4 (Fig. 3c).

Consistent with their relatively recent divergence and the similarity in size of the BBI gene family, most barley BBIs co-located with wheat BBIs (Fig. 3c). However, one cluster of contiguous BBIs on barley chromosome 3H suggests that gene duplication events also occurred independently in this species (Clade C, Fig. 3c). Maize and rice BBIs formed two distinct clusters in clade B and clade C, which included several adjacent BBIs in their respective genome assemblies, suggesting that BBI gene duplication also occurred independently in both these species (Fig. 3c).

BBI proteins were also separated according to the type of reactive site and the number of active domains they contained, as defined by Mello *et al.* [36]. Every BBI from all species in clade A contains a single active BBI domain and all fall into the MI-I group except for one barley BBI (HORVU5Hr1G068510) that does not match any previously characterized BBI group (Fig. 3c). The wheat BBIs clustered in clades B and C are all multi-domain proteins, and fall into either the MI-II or MI-IV groups except for three wheat BBIs with more than two domains that are most similar to the MI-IV group (Fig. 3c). The cluster of rice, maize and *Brachypodium* BBIs in clade C were most similar to the wheat BBIs on homeologous group 3 chromosomes, and were also all multi-domain proteins, represented by groups MI-IV, MI-V and MI-VI (Fig. 3c).

This phylogeny reveals that the BBI gene family in monocots is subject to a complex pattern of internal and external gene duplication events, resulting in multi-domain BBIs and gene copy number variation in each species. In wheat, extensive gene duplication on homeologous group 1 and especially group 3 chromosomes, that also occurred in barley, account for the greater numbers of BBI genes in the Triticeae lineage compared to other grasses.

The BBI gene family underwent gene duplication and deletion events both before and after common wheat's domestication

To gauge the approximate timing of the BBI gene family expansion in wheat, we identified BBI proteins from common wheat's ancestors. We found 12 BBIs from *T. urartu*, and 17 from *Ae. tauschii*, the diploid progenitors of the A and D genomes of common wheat, respectively (Fig. 4a). Because the diploid wheat B genome progenitor is unknown, we identified 24 BBIs in *T. turgidum* ssp. *dicoccoides*, an allotetraploid

progenitor with genomes AABB. We excluded three of these genes from our analysis (TRIDC0UG001830, TRIDC0UG020640, and TRIDC0UG000460) because they were not assembled into chromosomes, leaving 10 BBIs on the A genome and 11 on the B genome (Fig. 4a). Compared to each diploid progenitor genome, the corresponding genome in *T. aestivum* contained a greater number of BBIs (Fig. 4b). There were 1.3-fold more BBIs on the A genome of *T. aestivum* than in *T. urartu* and 1.5-fold more genes than in the A genome of *T. turgidum* ssp. *dicoccoides* (Fig. 4b). The difference in BBI gene number was greatest for the B genome, where *T. aestivum* had 1.9-fold more BBIs than *T. turgidum*. By contrast, the *T. aestivum* D genome contains only 1.2-fold more BBI genes than *Ae. tauschii* (Fig. 4b).

Phylogeny showed that most genes from wheat ancestors were clustered into orthologous groups with their corresponding genes in common wheat (Fig. 4c, Additional file 2, Table S6). The BBI genes on chromosomes 4 and 5 showed no evidence of duplication in any wheat species and both triads in *T. aestivum* grouped with their ancestral genes in *T. turgidum* (A genome) and *Ae. tauschii* (D genome) (Fig. 4c). We did not find the B homeologous copies of these genes in the *T. turgidum* assembly.

By contrast, the similarity and genomic position of BBI gene clusters on homeologous group 1 and 3 chromosomes in progenitor wheat species suggests that many BBI gene duplication events occurred before common wheat's domestication. On *Ae. tauschii* chromosome 1D, six contiguous BBI genes are clustered within 800 kb, while on chromosome 3D, eight BBI genes are clustered within 500 kb, suggesting they arose through tandem duplication (Additional file 2, Table S5). In *T. turgidum*, the eight BBI genes on chromosome 3A are within 200 kb and the ten BBI genes on chromosome 3B are contiguous within 700 kb (Additional file 2, Table S5).

This phylogeny also revealed several instances of gene duplications in tetraploid *T. turgidum* and hexaploid *T. aestivum* that were absent in the diploid progenitors, suggesting that they occurred since polyploidization (Additional file 2, Table S6). Furthermore, we also found a cluster of four adjacent paralogous BBIs on chromosome 1B of *T. aestivum* present only as a single copy in *T. turgidum*, suggesting that additional tandem duplication events occurred after common wheat's domestication (Fig. 4c).

To analyze the most recent evolution of the BBI gene family in common wheat, we identified BBIs in the genome assemblies of four common wheat cultivars (Additional file 2, Table S7). The total number of BBI genes in these cultivars ranged from 55 in 'Mace' to 60 in 'Jagger' (Table 2). While the BBI gene triads on chromosomes 4 and 5 were conserved in all cultivars, phylogenetic analysis indicated several instances of gene loss and gain on homeologous group 1 and 3 chromosomes (Additional file 1, Fig. S2). Although the BBI gene number varied between cultivars on each of these chromosomes, this variation was greatest on chromosomes 1B, 1D and 3B (Fig. 5a, b, Table 2). Strikingly, none of the five analyzed cultivars shared an identical complement of BBI genes.

Table 2 – BBI genes in five common wheat varieties, separated by chromosome.

Chromosome	Chinese Spring	Jagger	Mace	Julius	Landmark	$\Delta_{max-min}$
1A	3	3	3	3	2	1
1B	5	5	4	6	3	3
1D	7	9	9	9	10	3
3A	10	11	10	10	11	1
3B	14	13	11	12	13	3
3D	12	13	12	13	13	1
4A	1	1	1	1	1	0
4B	1	1	1	1	1	0
4D	1	1	1	1	1	0
5A	1	1	1	1	1	0
5B	1	1	1	1	1	0
5D	1	1	1	1	1	0
Total	57	60	55	59	58	5

$\Delta_{max-min}$ shows the inter-varietal variation in BBI gene number for each chromosome.

Taken together, analysis of the BBI gene family in different wheat germplasm reveals that while the gene triads on chromosomes 4 and 5 did not undergo expansion throughout wheat evolution, the gene clusters on homeologous group 1 and 3 chromosomes are more variable. Many gene duplication events occurred before domestication, but the increase in gene number in common wheat and variation among modern wheat cultivars shows that the BBI family remains dynamic.

Wheat BBI genes on homeologous group 3 chromosomes encode proteins with duplicated active domains

We next studied in greater detail the functional domains in the 57 BBIs from ‘Chinese Spring’. The majority of wheat BBIs (36 proteins, 63%) had one functional BBI domain, including all 15 BBIs located on homeologous group 1 chromosomes, the gene triads on chromosomes 4 and 5 and 15 BBIs on homeologous group 3 chromosomes (Fig. 6a, b). Of the remaining BBIs on group 3 chromosomes, 18 had two functional BBI domains (Fig. 6c), two proteins (TraesCS3D02G036400 and TraesCS3D02G035700) had three domains and one protein (TraesCS3B02G038300) had four domains (Fig. 6d). The gene structure of wheat BBIs reveals that while the majority have either one (6 BBIs, 10%) or two exons (45 BBIs, 79%), five genes including all three-domain proteins had three exons, while the gene (TraesCS3B02G038300) encoding the four-domain protein had four exons (Additional file 1, Fig. S3). This suggests that the genes encoding three- or four-domain BBI proteins may have evolved either from complete or partial gene duplication followed by fusion of tandem-duplicated genes.

We characterized the number and positions of conserved Cys residues within the reactive motifs of wheat BBIs according to the evolutionary scheme of Mello *et al*. [36] and with respect to substrate specificity. The first reactive inhibitory motif for trypsin was predicted to be conserved and functional in all wheat

BBIs except for three proteins: TraesCS3A02G049400LC, which has a truncated motif in a functional BBI domain (Fig. 6b), TraesCS3D02G033700, which has a two amino-acid deletion within the reactive motif (Fig. 6b), and TraesCS3B02G037200, which carries a Cys to Tyrosine (Y) amino acid substitution in the final residue of the reactive motif (Fig. 6c). The vast majority of BBIs carried K/R and S amino acids at the P1 and P1' positions, respectively (Fig. 6), the consensus motif determining trypsin inhibition in monocots [36]. All MI-I type BBIs had a K/R-S motif except one protein (TraesCS1D02G019800LC) that has a Glycine (G) in the P1 residue (Fig. 6a). Among the MI-II type BBIs, the P1-P1' motif was more diverse. Notably, four homeologous BBIs on group 3 chromosomes each exhibited Serine (S) to Valine (V) substitutions at position P1', while each protein in the triad on chromosome 4 had a Glutamate (E) residue at position P1 (Fig. 6b). Most MI-IV type BBIs also had a K/R-S motif in both reactive sites except a homeologous triad of BBIs with Serine to Tyrosine (T) substitutions in the P1' residue (Fig. 6c). In all 57 wheat BBIs the disulfide bridge (C_{10} and C_{11}) supporting the second inhibitory motif for chymotrypsin was lost (Fig. 6a-d).

The gene triad on chromosome 5 and all 15 BBIs on homeologous group 1 chromosomes each encode BBI proteins with functional domains comprised of 12 Cys residues that form six disulfide bridges, except for TraesCS1A02G022000 and TraesCS1A02G019800LC which carry amino acid substitutions at Cys residues in positions C_6 and C_{14}, respectively (Fig. 6a). The homeologous triad of BBIs on chromosome 4 fall into the MI-II group (Fig. 6b). BBIs on group 3 chromosomes were the most divergent. There were 15 BBIs categorized into the MI-II group that each contain ten Cys residues, except for TraesCS3A02G049400LC which has a deletion encompassing four Cys residues, and TraesCS3A02G045700, TraesCS3B02G042700LC and TraesCS3D02G034100 which each carry a single Cys amino acid substitution (Fig. 6b). Another 18 BBIs encode two-domain proteins categorized in the MI-IV group although three (TraesCS3D02G035300, TraesCS3B02G037300 and TraesCS3B02G037200) had a truncated second domain (Fig. 6c). The three wheat BBIs with more than two domains could not be categorized into any previously described MI evolutionary group (Fig. 6d). The three-domain proteins TraesCS3B02G036400 and TraesCS3D02G035700 are most similar to the MI-IV group but each underwent internal duplication of one domain resulting in three adjacent BBI domains that have distinct Cys positions from the previously proposed MI-VI three-domain group [36]. TraesCS3B02G036400 has a truncated second domain and a deletion of five Cys residues while the Cys positions in TraesCS3D02G035700 are also divergent from existing models (Fig. 6d). Each of the four domains in TraesCS3B02G038300 are full length and contain ten conserved Cys residues, suggesting all four may be functional (Fig. 6d). A summary of the different types of wheat BBI proteins in common wheat is shown in Fig.7.

Wheat BBI genes exhibit diverse expression profiles during development and in response to biotic and abiotic stress

We next used public RNA-Seq datasets to characterize transcript levels of the 57 BBI genes in common wheat. Genes were clustered into four main groups based on their expression profile in different wheat tissues and at different stages of development (Fig. 8a). Genes in group I showed relatively high

transcript levels in most plant tissues during development. BBIs in group II were predominantly expressed in root tissues, while BBIs in group III were expressed most highly during the early stages of leaf, stem and spike development. Finally, genes in group IV showed low levels of expression in most tissues and included ten genes with no detectable transcripts in any assayed tissue (Fig. 8a).

We also identified a subset of wheat BBIs that exhibit stress-responsive changes in expression (Fig. 8b). The majority of the highly expressed BBIs in group I are induced in response to stripe rust and *Septoria tritici* blotch infection and are suppressed by heat stress (Fig. 8b). Several BBIs in other groups were induced by multiple biotic stresses, including some genes in group IV that were only expressed in response to stress, indicating they may play a role in general immunity. Many of the BBI genes with no detectable expression in any of the reported conditions encode proteins lacking a SP or with truncations and amino acid changes in critical domains, suggesting they may be non-functional (Fig. 8).

Discussion

Diverse wheat genomic resources facilitate gene family characterization studies

In this study, we identified and characterized the BBI gene family in the common wheat landrace 'Chinese Spring', four modern cultivars, and their extant progenitors, using HMM-based homology searches (Fig. 1). This approach incorporates position-specific alignment scores and ensemble algorithms to evaluate all possible alignments. By weighting the relative likelihood of each alignment to identify orthologous proteins against a Pfam protein database, HMM may provide greater sensitivity than other sequence-based searches to identify all members of a gene family [56]. For each species, a single HMMsearch using a profile downloaded from the Pfam database was insufficient to identify all BBI proteins, likely because this general profile does not reflect species-specific diversity in this protein family [57]. A second search using a custom HMM profile built from an alignment of BBIs from the first screen yielded additional BBIs in every species analyzed, and for wheat, included 13 BBI proteins that were not annotated as such in the IWGSC RefSeq v1.0 gene model annotations [55]. We confirmed that each protein contained at least one BBI Pfam domain using HMMscan, although it is important to note that these sequences represent *in silico* predictions and the inhibitory function of each protein should be validated using biochemical assays, especially for those lacking conserved Cys residues.

Access to a greater diversity of high-quality genome assemblies for wheat will allow for more detailed gene characterization studies in this species. For example, the recent assembly of a more contiguous 'Chinese Spring' genome using both short and long read sequencing resolved 5,799 gene duplications that were not annotated in IWGSC RefSeq v1.1 [58]. These include two BBI genes (a paralog of TraesCS3A02G046300 located 6 Mb downstream on the same chromosome, and a paralog of TraesCS5B02G498100 located on chromosome 3B) that were not present in the IWGSC RefSeq v1.1 assembly. Beyond 'Chinese Spring', an international wheat pan-genome project aims to sequence and assemble multiple common wheat genomes (http://www.10wheatgenomes.com/). Among the five varieties we analyzed in this study, no two had the same complement of BBI genes (Fig. 5), highlighting

the value of a broader set of genomic resources to characterize the full extent of natural genetic variation in wheat.

The wheat BBI gene family underwent extensive duplication resulting in copy number variation and multi-domain proteins

Consistent with previous studies, our phylogenetic analysis shows that the BBI family is subject to widespread gene duplication events that likely occurred independently in each monocot species since they last shared a common ancestor [25, 38]. In rice, ten BBI genes are located in a 430 kb region of chromosome 1 [25], in maize, four BBIs are 200 kb apart on chromosome 3 and in barley, eleven BBIs are located within a 450 kb region of chromosome 3H (Fig. 3c, Additional file, Table S5). Each of these regions is syntenic with the distal region of wheat homeologous group 3 chromosomes [59, 60], suggesting that a common mechanism associated with this region of the genome, likely conserved in all crop species, triggers gene duplication at these loci. However, we found no evidence of systematic duplication of other genes in the region surrounding BBI clusters (+/- 200 kb, data not shown), suggesting that the high rate of gene duplications is limited to the BBI genes in these chromosomal regions.

One possible factor contributing to the high rate of duplications in the BBI family may be the location of gene clusters in distal telomeric regions of each chromosome (Fig. 2), which are hotspots for evolution, recombination events [61] and, in polyploid species, homeologous exchange [62]. Characterization of the MADS-box transcription factor family in wheat revealed a positive correlation between the number of genes in a subfamily and their proximity to the telomere [63]. In barley, large segmental duplications occurred more frequently in the telomeres, and were associated with increased gene copy number variation, potentially because of higher rates of non-allelic homologous recombination in these regions [64]. However, their position alone cannot account for the extent of BBI duplication, because the genes on homeologous group 4 and 5 chromosomes are similarly located in the telomere but did not undergo duplication in any barley or wheat genome analyzed in our study (Fig. 3c).

Although we found evidence of BBI gene duplication in all analyzed monocot genomes, this family was larger in wheat and barley due to more extensive tandem duplication events on wheat homeologous group 1 and 3 chromosomes and barley chromosome 3H (Fig. 3). Although many of these gene duplication events had already occurred in wheat's diploid and tetraploid progenitors, we also identified several duplication events that occurred since common wheat's domestication (Fig. 4), demonstrating that the process driving BBI family expansion in wheat remains active. In polyploid wheat species, relaxed selection pressure arising from gene redundancy may partially account for the greater expansion of the BBI gene family [65]. However, the similar size of the BBI gene family in diploid barley and wheat progenitors shows that gene duplication occurs to a similar degree in different Triticeae species, demonstrating that polyploidy is not necessary for BBI duplication. Further studies will be required to determine the mechanism or factors driving BBI gene family expansion in the Triticeae.

Our study also revealed that BBI domain duplication, possibly originating from incomplete gene duplication followed by gene fusion or internal duplication, resulted in further diversification of encoded

wheat BBI proteins, potentially enlarging the spectrum of their protease substrates (Fig. 6). Domain duplication is a common feature of BBI evolution in different plant species, including an ancient event that gave rise to the “double-headed” BBI structure conserved in dicots [25, 36, 38]. Our *in silico* analysis predicted that all wheat BBI proteins lack a functional second reactive motif to inhibit chymotrypsin activity (Fig. 6), consistent with analyses of other monocot BBIs [36, 37]. However, previous studies have detected chymotrypsin inhibition in protein extracts from the wheat endosperm, so it is likely that this activity is performed by a distinct family of protease inhibitors, potentially members of the cereal trypsin/a-amylase inhibitor family [66, 67].

Multi-domain monocot BBIs were previously isolated and characterized in other monocot species [25, 36, 68, 69]. The separation of all single-domain BBIs and all multi-domain BBIs in our phylogenetic tree suggests that these multi-domain BBIs were already present in the common ancestor of these grasses (Fig. 3c). In wheat, all multi-domain BBIs are located on homeologous group 3 chromosomes (Fig. 6c-d). Our finding that BBIs on both group 1 and group 3 chromosomes underwent complete gene duplication but only the BBIs on group 3 chromosomes underwent domain duplication (Fig. 3c and Fig. 6), suggests that the mechanism of gene duplication differs between group 1 and group 3 chromosomes. Alternatively, the reduced selective pressure on BBI genes on homeologous group 3 chromosomes (Additional file 1, Fig. S1) may result in a higher magnitude of gene expansion and an increased frequency of internal duplications giving rise to multi-domain proteins. These include three- and four-domain BBI proteins distinct in structure from any previously proposed BBI protein model (Fig. 6d). In order to determine the impact of this variation, it will be critical to identify the endogenous and exogenous interacting substrates of the BBI family, which remain poorly understood.

Functional characterization of wheat BBI genes

Gene duplication events can impact molecular evolution in different ways [70]. These include: (i) loss of protein function resulting from excessive mutation accumulation (ii) gain of protein function as a result of gene over-expression, (iii) neo- or sub-functionalization , and (iv) modulation of protein activity by duplicating and diversifying reactive sites. Our analyses indicate that the wheat BBI family potentially contains members exhibiting each of these features. Several wheat BBI genes exhibited truncations, mutations in active sites and undetectable transcript levels in all assayed tissues (Fig. 8a), suggesting they may be non-functional pseudogenes. Conversely, we also identified homeologous BBI genes that exhibit divergent expression profiles, suggesting they may have taken on new functional roles during wheat development (Fig. 8a). Several wheat BBIs exhibit high transcript levels in the grain, suggesting they may regulate endogenous protease activity during grain development (Fig. 8a). We also identified a subset of BBIs that are transcriptionally induced in response to fungal and bacterial pathogens, consistent with previous studies in other plants [25, 51, 71], which may indicate these genes contribute to plant defense responses (Fig. 8b). Some pathogens secrete proteases as part of their infection cycle, and in response, plants have co-evolved different classes of PIs to inhibit their activity [25]. In wheat, a greater number of BBI proteins with more numerous and diverse reactive sites may allow the wheat plant to inhibit a wider range of pathogenic protease substrate variants as part of an effective response against

fungal and bacterial pathogens [25]. Interestingly, the distal area of chromosome arm 3BS underlying a large cluster of BBI genes, coincides with the wheat streak mosaic virus resistance locus *Wsm2* [72, 73]. Although there is no evidence that BBI proteins act as an R gene for virus resistance, they might function as antagonistic interacting proteins with other R proteins to trigger defense responses [33].

Methods

Identification of Bowman-Birk inhibitors in plant genomes

High and low confidence wheat protein annotations from IWGSC RefSeq v1.1 [55] were downloaded from the IWGSC sequence repository hosted by URGI (https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Annotations/v1.1/) and concatenated into a single FASTA file consisting of 298,774 protein sequences. Protein sequences were obtained from the reference assemblies of *Hordeum vulgare* (IBSC_v2, 236,301 protein sequences), *Brachypodium distachyon* (v3.0, 52,972 protein sequences), *Aegilops tauschii* (Aet_v4.0, 258,680 protein sequences) [75], and *Triticum urartu* (ASM34745v1, 33,483 protein sequences) [76] from EnsemblPlants (https://plants.ensembl.org/info/website/ftp/index.html). *Oryza sativa* proteins were downloaded from the Rice Genome Annotation Project (Oryza_japonica.MSUv7, 55,986 protein sequences) (RGAP, http://rice.plantbiology.msu.edu) and *Zea mays* proteins (Zea_mays.B73_RefGen_v4, 131,585 protein sequences) were downloaded from MaizeGDB (https://www.maizegdb.org). *Triticum turgidum* ssp. *dicoccoides* wild emmer wheat 'Zavitan' WEWseq v1 proteins were downloaded from https://wewseq.wixsite.com/consortium with 110,314 protein sequences [77].

The identification of BBI proteins in each species was performed with HMMER analysis [56] against the local protein annotation database using a three-step approach outlined in Fig. 1. First, we performed an HMMsearch using the HMM profile for the Bowman-Birk protease inhibitor family (Pfam: PF00228) which was downloaded from Pfam 32.0 [57] using an E-value threshold of $1e^{-5}$. We next aligned the BBI protein sequences identified from the first step using HMMalign and built a new HMM profile based on the multiple alignment using HMMbuild. We used the new generated HMM profile to conduct a second HMMsearch against the same species-specific protein databases. Finally, we examined the list of BBI proteins for the presence of a BBI Pfam domain (PF00228) using HMMscan with an E-value threshold of 0.05. Proteins that contained the Pfam domain were classified as BBI. We then performed alignment of the identified BBI protein sequences from all species with MAFFT [78] and noticed that several BBIs were predicted to lack a signal peptide due to misannotation of the methionine start codon. We manually curated the position of the N-terminal start codon of several BBIs from *T. aestivum*, *Ae. tauchii*, *T. urartu*, *T. turgidum* and *H. vulgare* to match homologous sequences. Full curation details are provided in Additional file 2, Table S2, and includes details of BBI proteins with N-terminal truncations likely caused by point mutations. The curated sequences were used in all subsequent analyses.

Chromosomal locations and homology identification

All identified wheat BBIs were mapped to the IWGSC Refseq v1.1 genome assembly to identify their chromosomal location [55]. To determine homologous relationships between genes, we performed all-to-all BLAST using the 57 proteins as queries and applied an E-value threshold of $1e^{-10}$. Putative paralogs or homeologs were defined as homologous BBIs with a BLASTP e-value < $1e^{-10}$ and identity > 75% on the same or homeologous group chromosome, respectively. This approach was also used to identify orthologous relationships between BBIs in common wheat and progenitor genomes. The synteny and homologous relationship of wheat BBI genes were visualized with Circos plot using R shinyCircos [79]. Each chromosome was divided into telomere (R1/R3), centromere (C) and R2 segments according to information from the IWGSC RefSeq v1.1 genome assembly [55]. The distance of wheat BBIs and other high- and low-confidence gene models were mapped to individual chromosomes using the R Sushi package plotBed function [80].

We calculated Ka/Ks ratios using an online tool hosted by the computational biology unit (CBU http://services.cbu.uib.no/tools/kaks) using the coding sequence of each common wheat BBI gene. We excluded one BBI (TraesCS3D02G035800) from the Ka/Ks ratio analysis due to a premature stop codon in its coding sequence. The remaining 56 BBIs were grouped according to their chromosome and used to construct phylogenetic trees and calculate the pairwise Ka/Ks ratio for each branch.

Alignment and phylogenetic analysis

We performed multiple sequence alignments using Clustal Omega using full-length BBI protein sequences identified in all species. Model selection was conducted with IQ-TREE using the lowest Bayesian information criterion (BIC) as WAG+G4 model [81]. We constructed the phylogenetic tree using the selected model with 1000 ultrafast bootstrap replicates UFBoot2 [82, 83]. The resulting tree was visualized and annotated with the R package ggtree v2.0.4 [84]. The domain model type for BBIs in grasses were determined manually by comparing the number and position of Cys residues to the model proposed by Mello *et al.* [36].

Identification of BBI on homeologous group 3 chromosomes in different wheat varieties

The draft genome assembly for four common wheat varieties ‘Jagger’ (U.S.A, winter growth habit), ‘Julius’ (Germany, winter), ‘Landmark’ (Canada, spring), and ‘Mace’ (Australia, spring) were downloaded from the 10+ Wheat Genomes Project (https://wheat.ipk-gatersleben.de/downloads/) and used to build local BLAST databases. We then used the full-length protein-coding sequences of each BBI gene from ‘Chinese Spring’ as queries and performed BLAST against the genomes of each wheat variety to identify their chromosomal position [85]. The position of each BBI genes in these varieties was cross-referenced with GFF files to identify the corresponding gene ID. To identify BBIs present in these varieties but absent from the ‘Chinese Spring’ assembly, the corresponding genomic region spanning all BBI genes on homeologous group 1 and group 3 chromosomes from each wheat variety were extracted locally using the bedtools getfasta command for *ab initio* gene prediction [86]. The open reading frame (ORF) and putative gene model for each extracted DNA fragment was predicted with OrfM-0.7.1 [87]. To determine

whether predicted gene models contain a functional BBI domain, all predicted ORFs were scanned with HMMscan using an e-value cutoff as 0.05 and those BBIs with PF00228 domains were retained. After exclusion of common orthologs in other varieties, the unique BBIs in each variety were named using the first two letters of the cultivar name, followed by chromosome number, and ordered by their relative position on that chromosome. For example, JA_3A-1 represents the first unique BBI on 'Jagger' chromosome 3A. The BBI protein sequences from all varieties were then used to construct a phylogenetic tree with IQ-TREE using the WAG+G4 model with 1000 ultrafast bootstrap replicates UFBoot2 [82, 83]. The resulting tree was visualized and annotated with the R package ggtree v2.0.4 [84].

Gene structure analysis of the functional domains and motifs

The complete genomic, CDS and amino acid sequences, as well as gene feature information of all BBIs identified were downloaded from IWGSC RefSeq v1.1 [55]. Schematic representation of the exon-intron organization of wheat BBIs was conducted by comparing the CDS and the corresponding genomic sequences using Gene Structure Display Server 2.0 [88]. To find conserved Cys-rich domains, the amino acid sequence for the functional domains of all identified BBIs in wheat, by aligning amino acid sequences between the first and last conserved Cys residue in each domain using MAFFT v7 for multiple sequence alignment [78]. All sequences were analyzed using Signal P v5.0 [89] to predict the presence of N-terminal SP and for potential cleavage sites.

Gene expression analysis

The expression data for wheat BBI genes in five tissues (spike, root, leaf, grain and stem) at three different developmental stages from hexaploid wheat var. 'Chinese Spring' [90] and under abiotic stress (heat and drought) condition at the one-week-old seedling stage [91] were mapped to the IWGSC RefSeq v1.1 genome and processed into TPM values as described previously [92]. Separately, we downloaded several biotic stress expression datasets as TPM from the online wheat expression browser expVIP [93], including studies on fusarium head blight [94, 95], stripe rust [96, 97], powdery mildew [97], fusarium crown rot [98], *Septoria tritici* blotch [99, 100] and PAMP elicitors [101]. For each pathogen, we calculated the log2 fold change of the transcript abundance for each treated sample compared to mock controls or samples at time zero at each time point and averaged the values of all time points. Heatmaps for tissue specific time course expression were constructed using log2 transformed TPM values with the R package pheatmap v1.0.12. Genes were clustered according to their expression level (metric, Euclidian; method, complete) and grouped by their chromosome type.

List Of Abbreviations

BBI, Bowman-Birk Inhibitor; BLAST, Basic Local Alignment Search Tool; cDNA, complementary Deoxyribose Nucleic Acid; CDS, Coding Sequence; Cys, Cysteine; GFF, General Feature Format; HMM, Hidden Markov Model; IWGSC, International Wheat Genome Sequencing Consortium; kb, kilobase pair; kDa, kilo Dalton; Mb, Megabase pair; ORF, Open Reading Frame; PAMP, Pathogen-Associated Molecular Patterns; Pfam, Protein family; PI, Protease Inhibitor; SP, Signal Peptide; TPM, Transcripts Per Million.

Declarations

Acknowledgements

We are grateful to James Hamilton for assistance in generating expression data.

Author contributions

YX performed data analysis and wrote the first draft of the manuscript. KR performed data analysis and wrote the manuscript. SP wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was partially funded by the Colorado Wheat Research Foundation and Colorado Wheat Administrative Committee.

Availability of data and raw materials

Gene sequence data from this project are all available in public databases. Full details of all data analysis steps and outputs are provided in supplementary files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Clemente M, Corigliano MG, Pariani SA, Sánchez-López EF, Sander VA, Ramos-Duarte VA. Plant serine protease inhibitors: Biotechnology application in agriculture and molecular farming. Int J Mol Sci. 2019;20:1345. doi:10.3390/ijms20061345.
2. Laskowski M, Kato I. Protein inhibitors of proteinases. Annu Rev Biochem. 1980;49:593–626. doi:10.1146/annurev.bi.49.070180.003113.
3. S. Bateman K, N.G. James M. Plant protein proteinase inhibitors: structure and mechanism of inhibition. Curr Protein Pept Sci. 2011;12:341–7. doi:10.2174/138920311796391124.
4. Laskowski M, Qasim MA. What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? Biochim Biophys Acta - Protein Struct Mol Enzymol.

2000;1477:324–37. doi:10.1016/S0167-4838(99)00284-8.

5. Volpicella M, Leoni C, Costanza A, De Leo F, Gallerani R, R. Ceci L. Cystatins, serpins and other families of protease inhibitors in plants. Current Protein & Peptide Science. 2011;12:386–98. doi:10.2174/138920311796391098.
6. Pak C, Van Doorn WG. Delay of Iris flower senescence by protease inhibitors. New Phytol. 2005;165:473–80. doi:10.1111/j.1469-8137.2004.01226.x.
7. Haq SK, Atif SM, Khan RH. Protein proteinase inhibitor genes in combat against insects, pests, and pathogens: natural and engineered phytoprotection. Arch Biochem Biophys. 2004;431:145–59. doi:https://doi.org/10.1016/j.abb.2004.07.022.
8. Chen MS. Inducible direct plant defense against insect herbivores: A review. Insect Sci. 2008;15:101–14. doi:10.1111/j.1744-7917.2008.00190.x.
9. Hellinger R, Gruber CW. Peptide-based protease inhibitors from plants. Drug Discov Today. 2019;24:1877–89.
10. Lawrence PK, Koundal KR. Plant protease inhibitors in control of phytophagous insects. Electron J Biotechnol. 2002;5:93–109.
11. Jashni MK, Mehrabi R, Collemare J, Mesarich CH, de Wit PJGM. The battle in the apoplast: Further insights into the roles of proteases and their inhibitors in plant–pathogen interactions. Front Plant Sci. 2015;6:584. doi:10.3389/fpls.2015.00584.
12. Rawlings ND, Barrett AJ. Evolutionary families of peptidases. Biochem J. 1993;290:205–18. doi:10.1042/bj2900205.
13. Rawlings ND, Tolle DP, Barrett AJ. Evolutionary families of peptidase inhibitors. Biochem J. 2004;378:705–16. doi:10.1042/BJ20031825.
14. Birk Y. Plant protease inhibitors. 1st edition. Springer-Verlag Berlin Heidelberg; 2003.
15. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 2018;46:D624–32. doi:10.1093/nar/gkx1134.
16. Birk Y, Gertler A, Khalef S. A pure trypsin inhibitor from soybeans. Biochem J. 1963;87:281–4. doi:10.1042/bj0870281.
17. Bowman DE. Differentiation of soybean antitryptic factors. Proc Soc Exp Biol Med. 1946;63:547–50. doi:10.3181/00379727-63-15668.
18. Singh S, Singh A, Kumar S, Mittal P, Singh IK. Protease inhibitors: recent advancement in its usage as a potential biocontrol agent for insect pest management. Insect Sci. 2020;27:186–201.
19. Hilder VA, Gatehouse AMR, Sheerman SE, Barker RF, Boulter D. A novel mechanism of insect resistance engineered into tobacco. Nature. 1987;330:160–3. doi:10.1038/330160a0.
20. Xu D, Xue Q, McElroy D, Mawal Y, Hilder VA, Wu R. Constitutive expression of a cowpea trypsin inhibitor gene, *CpTi*, in transgenic rice plants confers resistance to two major rice insect pests. Mol Breed. 1996;2:167–73. doi:10.1007/BF00441431.

21. Bi RM, Jia HY, Feng DS, Wang HG. Production and analysis of transgenic wheat (*Triticum aestivum* L.) with improved insect resistance by the introduction of cowpea trypsin inhibitor gene. Euphytica. 2006;151:351–60. doi:10.1007/s10681-006-9157-9.
22. Ye XY, Ng TB, Rao PF. A Bowman-Birk-type trypsin-chymotrypsin inhibitor from broad beans. Biochem Biophys Res Commun. 2001;289:91–6. doi:10.1006/bbrc.2001.5965.
23. Komarnytsky S, Borisjuk N, Yakoby N, Garvey A, Raskin I. Cosecretion of protease inhibitor stabilizes antibodies produced by plant roots. Plant Physiol. 2006;141:1185–93. doi:10.1104/pp.105.074419.
24. Pekkarinen AI, Longstaff C, Jones BL. Kinetics of the inhibition of Fusarium serine proteinases by barley (*Hordeum vulgare* L.) inhibitors. J Agric Food Chem. 2007;55:2736–42. doi:10.1021/jf0631777.
25. Qu LJ, Chen J, Liu M, Pan N, Okamoto H, Lin Z, et al. Molecular cloning and functional analysis of a novel type of Bowman-Birk Inhibitor gene family in rice. Plant Physiol. 2003;133:560–70.
26. Pang Z, Zhou Z, Yin D, Lv Q, Wang L, Xu X, et al. Transgenic rice plants overexpressing *BBTI4* confer partial but broad-spectrum bacterial blight resistance. J Plant Biol. 2013;56:383–90. doi:10.1007/s12374-013-0277-1.
27. Zhang C, Fang H, Shi X, He F, Wang R, Fan J, et al. A fungal effector and a rice NLR protein have antagonistic effects on a Bowman–Birk trypsin inhibitor. Plant Biotechnol J. 2020;n/a n/a. doi:10.1111/pbi.13400.
28. Juliana P, Singh RP, Singh PK, Poland JA, Bergstrom GC, Huerta-Espino J, et al. Genome-wide association mapping for resistance to leaf rust, stripe rust and tan spot in wheat reveals potential candidate genes. Theor Appl Genet. 2018;131:1405–22. doi:10.1007/s00122-018-3086-6.
29. Sari E, Cabral AL, Polley B, Tan Y, Hsueh E, Konkin DJ, et al. Weighted gene co-expression network analysis unveils gene networks associated with the *Fusarium head blight* resistance in tetraploid wheat. BMC Genom. 2019;20:925. doi:10.1186/s12864-019-6161-8.
30. Shan L, Li C, Chen F, Zhao S, Xia G. A Bowman-Birk type protease inhibitor is involved in the tolerance to salt stress in wheat. Plant, Cell Environ. 2008;31:1128–37. doi:10.1111/j.1365-3040.2008.01825.x.
31. Dramé KN, Passaquet C, Repellin A, Zuily-Fodil Y. Cloning, characterization and differential expression of a Bowman-Birk inhibitor during progressive water deficit and subsequent recovery in peanut (*Arachis hypogaea*) leaves. J Plant Physiol. 2013;170:225–9.
32. Yan KM, Chang T, Soon SA, Huang FY. Purification and characterization of Bowman-Birk protease inhibitor from rice coleoptiles. J Chinese Chem Soc. 2009;56:949–60. doi:10.1002/jccs.200900139.
33. Malefo MB, Mathibela EO, Crampton BG, Makgopa ME. Investigating the role of Bowman-Birk serine protease inhibitor in *Arabidopsis* plants under drought stress. Plant Physiol Biochem. 2020;149:286–93. doi:10.1016/j.plaphy.2020.02.007.
34. Zhang L, Nakanishi Itai R, Yamakawa T, Nakanishi H, Nishizawa NK, Kobayashi T. The Bowman-Birk trypsin inhibitor IBP1 interacts with and prevents degradation of IDEF1 in rice. Plant Mol Biol Report. 2014;32:841–51. doi:10.1007/s11105-013-0695-8.

35. Bowman DE. Fractions derived from soybeans and navy beans which retard tryptic digestion of casein. Proc Soc Exp Biol Med. 1944;57:139–40. doi:10.3181/00379727-57-14731P.
36. Mello MO, Tanaka AS, Silva F MC. Molecular evolution of Bowman-Birk type proteinase inhibitors in flowering plants. Mol Phylogenet Evol. 2003;27:103–12.
37. Qi RF, Song ZW, Chi CW. Structural features and molecular evolution of Bowman-Birk protease inhibitors and their potential application. Acta Biochim Biophys Sin (Shanghai). 2005;37:283–92.
38. James AM, Jayasena AS, Zhang J, Berkowitz O, Secco D, Knott GJ, et al. Evidence for ancient origins of Bowman-Birk inhibitors from *Selaginella moellendorffii*. Plant Cell. 2017;29:461–73.
39. Clemente A, Arques MC, Dalmais M, Le Signor C, Chinoy C, Olias R, et al. Eliminating anti-nutritional plant food proteins: the case of seed protease inhibitors in pea. PLoS One. 2015;10:e0134634–e0134634. doi:10.1371/journal.pone.0134634.
40. Baek JM, Song JC, Choi Y Do, Kim S Il. Nucleotide sequence homology of cDNAs encoding soybean Bowman-Birk type proteinase inhibitor and its isoinhibitors. Biosci Biotechnol Biochem. 1994;58:843–6. doi:10.1271/bbb.58.843.
41. Baek JM, Kim SI. Nucleotide sequence of a cDNA encoding soybean Bowman-Birk proteinase inhibitor. Plant Physiol. 1993;102:687. doi:10.1104/pp.102.2.687.
42. Nishino N, Aoyagi H, Kato T, Izumiya N. Studies on the synthesis of proteinase inhibitors: Synthesis and activity of nonapeptide fragments of soybean Bowman-Birk inhibitor. J Biochem. 1977;82:901–9. doi:10.1093/oxfordjournals.jbchem.a131767.
43. Birk Y. The Bowman-Birk inhibitor. Trypsin- and chymotrypsin-inhibitor from soybeans. Int. J. Pept. Protein Res. 1985;25:113–31.
44. Prakash B, Murthy MRN, Sreerama YN, Rao DR, Gowda LR. Studies on simultaneous inhibition of trypsin and chymotrypsin by horsegram Bowman-Birk inhibitor. J Biosci. 1997;22:545–54. doi:10.1007/BF02703392.
45. Habib H, Fazili KM. Plant protease inhibitors : a defense strategy in plants. Biotechnol Mol Biol Rev. 2007;2:68–85.
46. Galasso I, Piergiovanni AR, Lioi L, Campion B, Bollini R, Sparvoli F. Genome organization of Bowman–Birk inhibitor in common bean (*Phaseolus vulgaris* L.). Mol Breed. 2009;23:617–24. doi:10.1007/s11032-009-9260-4.
47. Food and Agriculture Organization of the United Nations, Database, FAOSTAT statistics, Crops. 2018. http://www.fao.org/faostat/en/.
48. Odani S, Koide T, Ono T. Wheat germ trypsin inhibitors. isolation and structural characterization of single-headed and double-headed inhibitors of the Bowman-Birk type. J Biochem. 1986;100:975–83. doi:10.1093/oxfordjournals.jbchem.a121810.
49. Raj SSS, Kibushi E, Kurasawa T, Suzuki A, Yamane T, Odani S, et al. Crystal structure of bovine trypsin and wheat germ trypsin inhibitor (I-2b) complex (2:1) at 2.3 Å resolution. J Biochem. 2002;132:927–33. doi:10.1093/oxfordjournals.jbchem.a003306.

50. Poerio E, Caporale C, Carrano L, Caruso C, Vacca F, Buonocore V. The amino acid sequence and reactive site of a single-headed trypsin inhibitor from wheat endosperm. J Protein Chem. 1994;13:187–94. doi:10.1007/BF01891977.
51. Chilosi G, Caruso C, Caporale C, Leonardi L, Bertini L, Buzi A, et al. Antifungal activity of a Bowman-Birk-type trypsin inhibitor from wheat kernel. J Phytopathol. 2000;148:477–81. doi:10.1046/j.1439-0434.2000.00527.x.
52. Richards KD, Snowden KC, Gardner RC. *wali6* and *wali7*. Genes induced by aluminum in wheat (*Triticum aestivum* L.) roots. Plant Physiol. 1994;105:1455–6. doi:10.1104/pp.105.4.1455.
53. Snowden KC, Richards KD, Gardner RC. Aluminum-induced genes. Induction by toxic metals, low calcium, and wounding and pattern of expression in root tips. Plant Physiol. 1995;107:341–8. doi:10.1104/pp.107.2.341.
54. El Baidouri M, Murat F, Veyssiere M, Molinier M, Flores R, Burlot L, et al. Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). New Phytol. 2017;213:1477–86. doi:10.1111/nph.14113.
55. Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science. 2018;361:7191. doi:10.1126/science.aar7191.
56. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Res. 2018;46:W200–4. doi:10.1093/nar/gky448.
57. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2018;47:D427–32. doi:10.1093/nar/gky995.
58. Alonge M, Shumate A, Puiu D, Zimin A, Salzberg SL. Chromosome-scale assembly of the bread wheat genome reveals thousands of additional gene copies. Genetics. 2020;216:599–608. doi:10.1534/genetics.120.303501.
59. Munkvold JD, Greene RA, Bermudez-Kandianis CE, La Rota CM, Edwards H, Sorrells SF, et al. Group 3 chromosome bin maps of wheat and their relationship to rice chromosome 1. Genetics. 2004;168:639–50.
60. La Rota M, Sorrells ME. Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. Funct Integr Genomics. 2004;4:34–46.
61. Glover NM, Daron J, Pingault L, Vandepoele K, Paux E, Feuillet C, et al. Small-scale gene duplications played a major role in the recent evolution of wheat chromosome 3B. Genome Biol. 2015;16:188. doi:10.1186/s13059-015-0754-6.
62. Zhang Z, Zhang Z, Gou X, Gou X, Xun H, Xun H, et al. Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. Proc Natl Acad Sci U S A. 2020;117:14561–71. doi:10.1073/pnas.2003505117.
63. Schilling S, Kennedy A, Pan S, Jermiin LS, Melzer R. Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. New Phytol. 2020;225:511–29. doi:10.1111/nph.16122.

64. Bretani G, Rossini L, Ferrandi C, Russell J, Waugh R, Kilian B, et al. Segmental duplications are hot spots of copy number variants affecting barley gene content. Plant J. 2020;103:1073–88. doi:10.1111/tpj.14784.
65. Comai L. The advantages and disadvantages of being polyploid. Nat Rev Genet. 2005;6:836–46. doi:10.1038/nrg1711.
66. Di Maro A, Farisei F, Panichi D, Severino V, Bruni N, Ficca AG, et al. WCI, a novel wheat chymotrypsin inhibitor: purification, primary structure, inhibitory properties and heterologous expression. Planta. 2011;234:723–35. doi:10.1007/s00425-011-1437-5.
67. Tedeschi F, Di Maro A, Facchiano A, Costantini S, Chambery A, Bruni N, et al. Wheat Subtilisin/Chymotrypsin Inhibitor (WSCI) as a scaffold for novel serine protease inhibitors with a given specificity. Mol Biosyst. 2012;8:3335–43. doi:10.1039/C2MB25320H.
68. Nagasue A, Fukamachi H, Ikenaga H, Funatsu G. The amino acid sequence of barley rootlet trypsin inhibitor. Agric Biol Chem. 1988;52:1505–14. doi:10.1080/00021369.1988.10868867.
69. Tashiro M, Asao T, Hirata C, Takahashi K, Kanamori M. The complete amino acid sequence of a major trypsin inhibitor from seeds of foxtail millet (*Setaria italica*). J Biochem. 1990;108:669–72. doi:10.1093/oxfordjournals.jbchem.a123260.
70. Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution. J Genet. 2013;92:155–61.
71. Othman T, Abu Bakar N, Zainal Abidin R, Mahmood M, Saidi N, Shaharuddin N. Potential of plant's Bowman-Birk protease inhibitor in combating abiotic stresses: a mini review. Bioremediation Sci Technol Res. 2014;2:53–61.
72. Dhakal S, Tan CT, Anderson V, Yu H, Fuentealba MP, Rudd JC, et al. Mapping and KASP marker development for wheat curl mite resistance in "TAM 112" wheat using linkage and association analysis. Mol Breed. 2018;38:119.
73. Tan CT, Assanga S, Zhang G, Rudd JC, Haley SD, Xue Q, et al. Development and validation of KASP markers for wheat streak mosaic virus resistance gene *Wsm2*. Crop Sci. 2017;57:340–9.
74. Uauy C, Wulff BBH, Dubcovsky J. Combining traditional mutagenesis with new high-throughput sequencing and genome editing to reveal hidden variation in polyploid wheat. Annu Rev Genet. 2017;51:435–54. doi:10.1146/annurev-genet-120116-024533.
75. Luo MC, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, et al. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. Nature. 2017;551:498–502. doi:10.1038/nature24486.
76. Ling HQ, Ma B, Shi X, Liu H, Dong L, Sun H, et al. Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. Nature. 2018;557:424–8. doi:10.1038/s41586-018-0108-0.
77. Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science. 2017;357:93–7. doi:10.1126/science.aan0032.
78. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2017;20:1160–6. doi:10.1093/bib/bbx108.

79. Yu Y, Ouyang Y, Yao W. shinyCircos: an R/Shiny application for interactive creation of Circos plot. Bioinformatics. 2017;34:1229–31. doi:10.1093/bioinformatics/btx763.
80. Phanstiel DH, Boyle AP, Araya CL, Snyder MP. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. Bioinformatics. 2014;30:2808–10. doi:10.1093/bioinformatics/btu379.
81. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9. doi:10.1038/nmeth.4285.
82. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–22. doi:10.1093/molbev/msx281.
83. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. doi:10.1093/molbev/msu300.
84. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. GGTREE: an R package for visualization and annotation of phylogenetic trees With their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36. doi:10.1111/2041-210X.12628.
85. Deng W, Nickle DC, Learn GH, Maust B, Mullins JI. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. Bioinformatics. 2007;23:2334–6. doi:10.1093/bioinformatics/btm331.
86. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2. doi:10.1093/bioinformatics/btq033.
87. Woodcroft BJ, Boyd JA, Tyson GW. OrfM: a fast open reading frame predictor for metagenomic data. Bioinformatics. 2016;32:2702–3. doi:10.1093/bioinformatics/btw241.
88. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2014;31:1296–7. doi:10.1093/bioinformatics/btu817.
89. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37:420–3. doi:10.1038/s41587-019-0036-z.
90. Lukaszewski AJ, Alberti A, Sharpe A, Kilian A, Stanca AM, Keller B, et al. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science. 2014;345:1251788. doi:10.1126/science.1251788.
91. Liu Z, Xin M, Qin J, Peng H, Ni Z, Yao Y, et al. Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). BMC Plant Biol. 2015;15:152. doi:10.1186/s12870-015-0511-8.
92. Pearce S, Vazquez-Gross H, Herin SY, Hane D, Wang Y, Gu YQ, et al. WheatExp: an RNA-seq expression database for polyploid wheat. BMC Plant Biol. 2015;15:299. doi:10.1186/s12870-015-0692-1.

93. Borrill P, Ramirez-Gonzalez R, Uauy C. expVIP: A customizable RNA-seq data analysis and visualization platform. Plant Physiol. 2016;170:2172–86. doi:10.1104/pp.15.01667.
94. Kugler KG, Siegwart G, Nussbaumer T, Ametz C, Spannagl M, Steiner B, et al. Quantitative trait loci-dependent analysis of a gene co-expression network associated with Fusarium head blight resistance in bread wheat (*Triticum aestivum* L.). BMC Genom. 2013;14:728. doi:10.1186/1471-2164-14-728.
95. Schweiger W, Steiner B, Vautrin S, Nussbaumer T, Siegwart G, Zamini M, et al. Suppressed recombination and unique candidate genes in the divergent haplotype encoding *Fhb1*, a major Fusarium head blight resistance locus in wheat. Theor Appl Genet. 2016;129:1607–23. doi:10.1007/s00122-016-2727-x.
96. Cantu D, Segovia V, MacLean D, Bayles R, Chen X, Kamoun S, et al. Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. BMC Genom. 2013;14:270. doi:10.1186/1471-2164-14-270.
97. Zhang H, Yang Y, Wang C, Liu M, Li H, Fu Y, et al. Large-scale transcriptome comparison reveals distinct gene activations in wheat responding to stripe rust and powdery mildew. BMC Genom. 2014;15:898. doi:10.1186/1471-2164-15-898.
98. Powell JJ, Carere J, Fitzgerald TL, Stiller J, Covarelli L, Xu Q, et al. The Fusarium crown rot pathogen *Fusarium pseudograminearum* triggers a suite of transcriptional and metabolic changes in bread wheat (*Triticum aestivum* L.). Ann Bot. 2017;119:853–67. doi:10.1093/aob/mcw207.
99. Yang F, Li W, Jørgensen HJL. Transcriptional reprogramming of wheat and the hemibiotrophic pathogen *Septoria tritici* during two phases of the compatible interaction. PLoS One. 2013;8:81606. doi:10.1371/journal.pone.0081606.
100. Rudd JJ, Kanyuka K, Hassani-Pak K, Derbyshire M, Andongabo A, Devonshire J, et al. Transcriptome and metabolite profiling of the infection cycle of *Zymoseptoria tritici* on wheat reveals a biphasic interaction with plant immunity involving differential pathogen chromosomal contributions and a variation on the hemibiotrophic lifestyle definition. Plant Physiol. 2015;167:1158–85. doi:10.1104/pp.114.255927.
101. Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, et al. The transcriptional landscape of polyploid wheat. Science. 2018;361:eaar6089. doi:10.1126/science.aar6089.

Figures

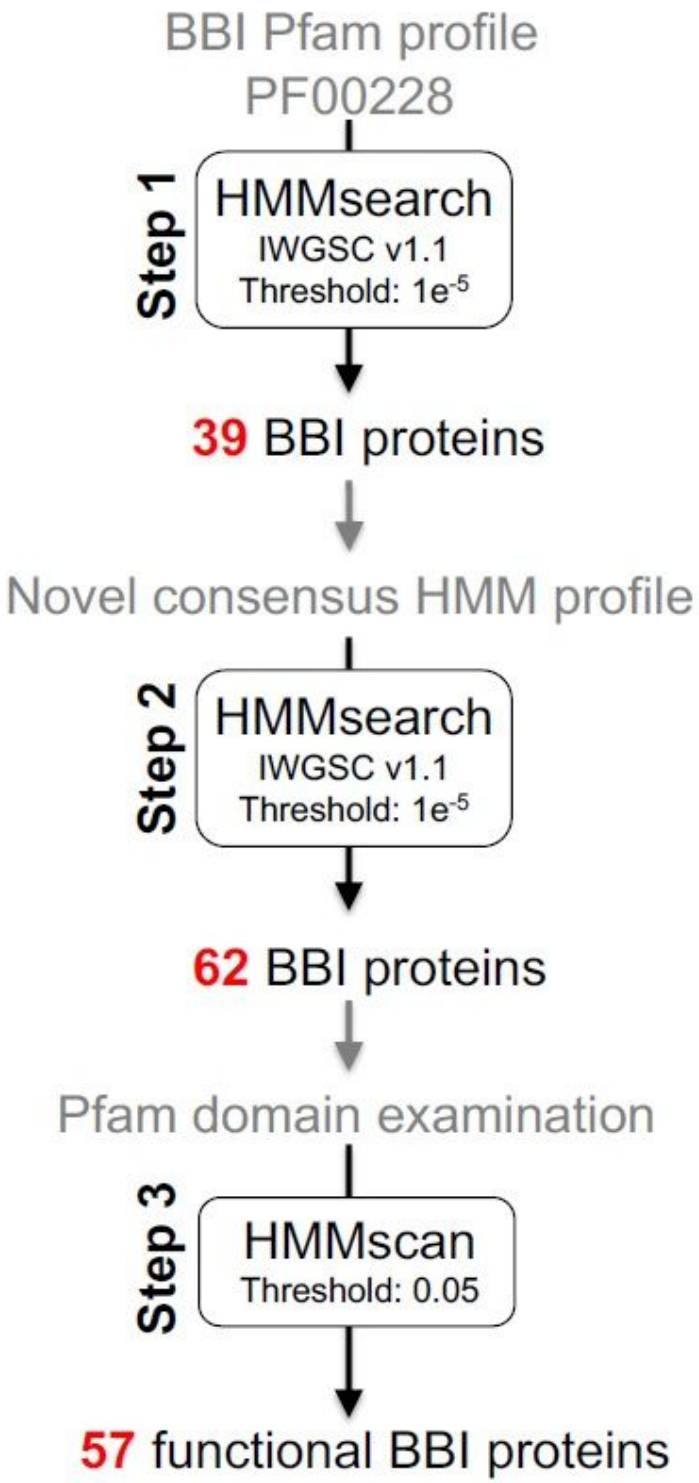


Figure 1

Pipeline for Bowman-Birk inhibitor (BBI) gene family identification in plant genomes. The identification of BBIs in the T. aestivum genome is presented as an example, including key steps and criteria for each step. The number of proteins identified at each stage are highlighted in red.

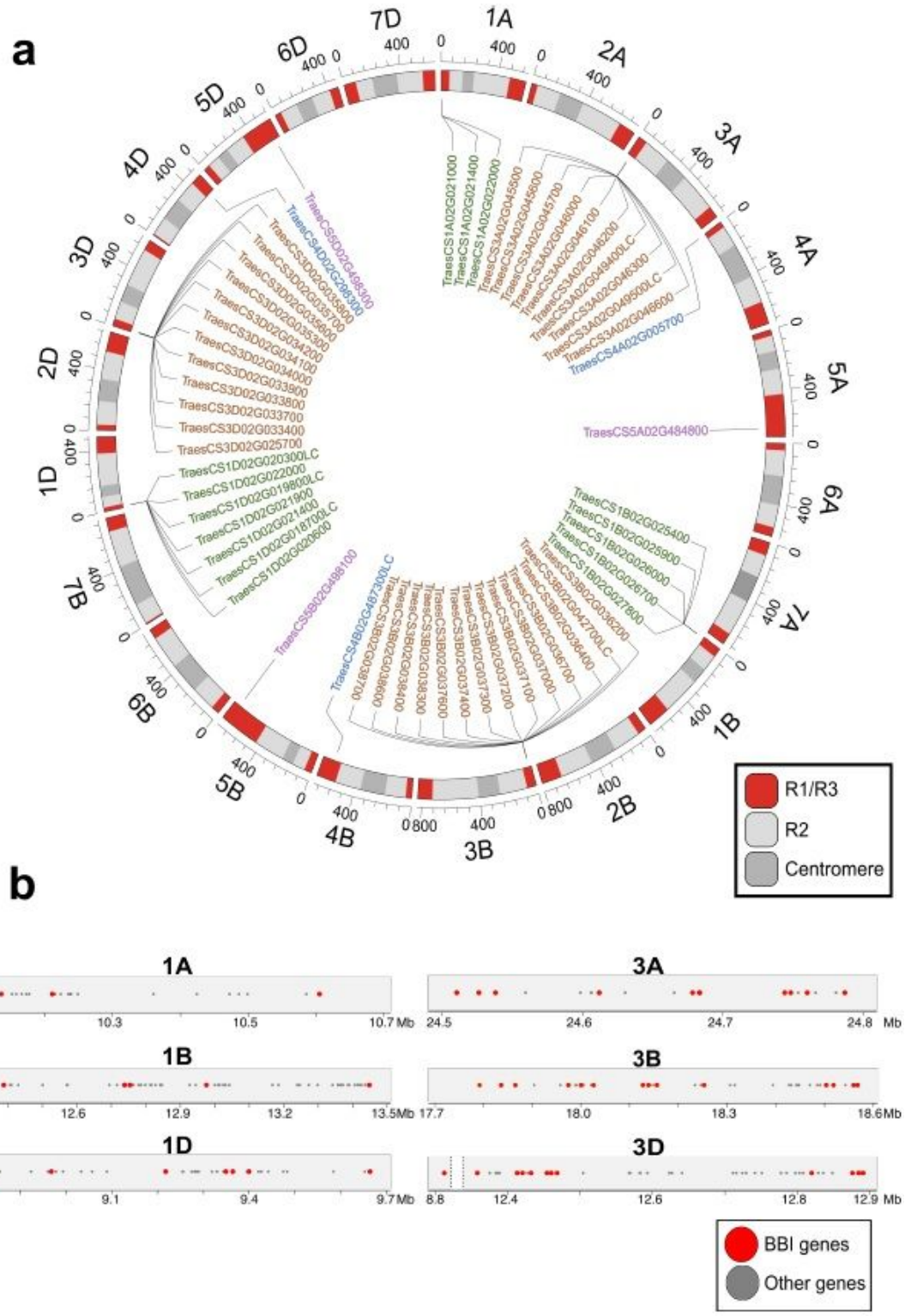


Figure 2

Distribution of 57 BBI in the T. aestivum genome. a Chromosomal positions of wheat BBIs. Gene names are colored according to their homeologous group. Chromosomal segments are indicated by different colors - distal regions of the chromosome R1 and R3 in red, centromeric region C in dark grey, and region R2 in light grey. b Distribution of genes within BBI clusters on homeologous group 1 and group 3 chromosomes. Red dots represent BBI genes, whereas grey dot represent other annotated genes in the

region, positioned according to their physical location in the IWGSC Refseq v1.1 genome assembly. All high confidence (HC) and low confidence (LC) gene models are presented.

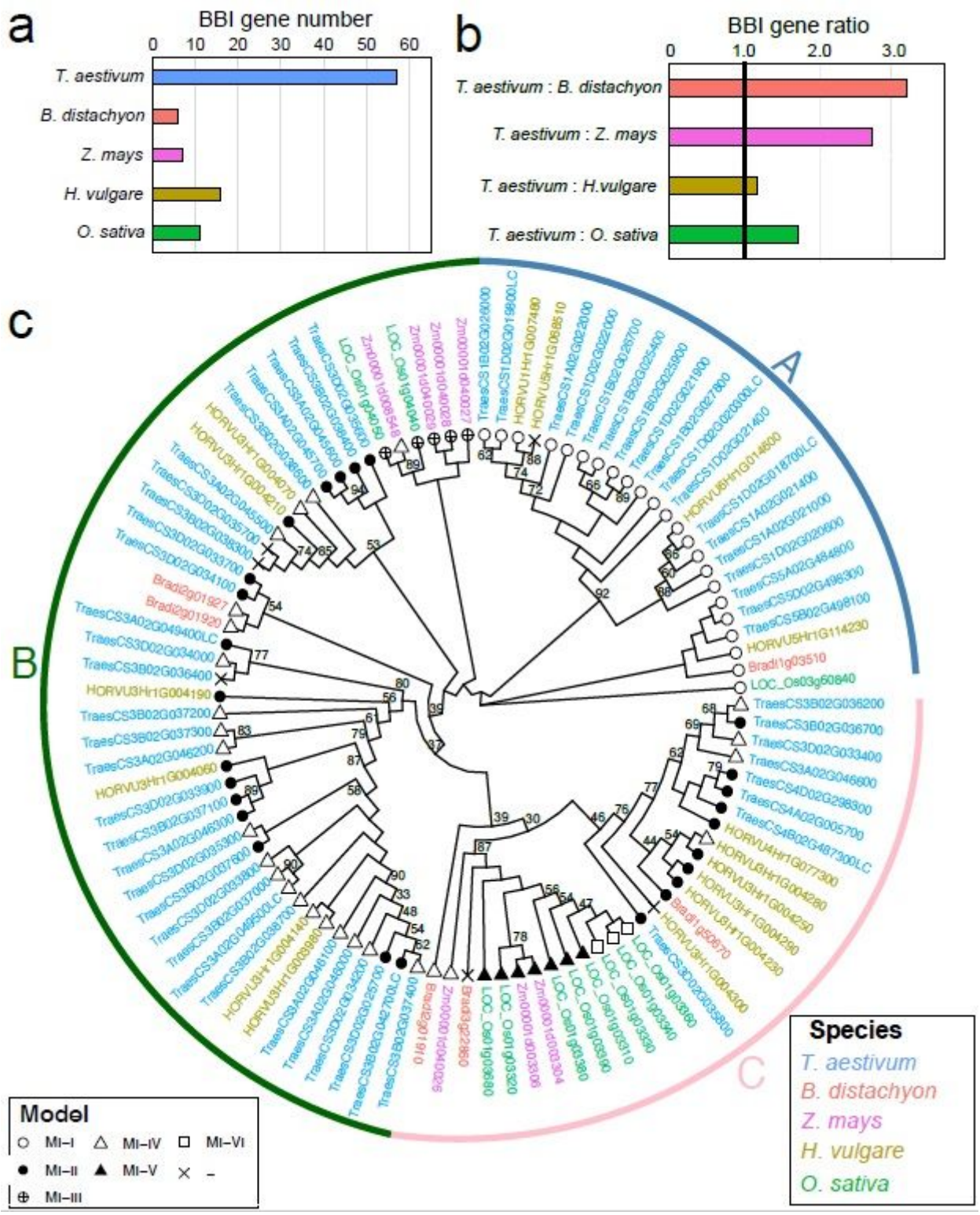


Figure 3

Comparison of the wheat BBI gene family with other monocots. a Total number of BBI genes in monocot genomes. Bars are color-coded based on species. b Ratios of total BBI gene numbers in common wheat compared to other monocot species, adjusted for wheat's hexaploid genome. The 1:1 ratio is indicated by

a bold line. c Circular phylogenetic tree of all BBI proteins from rice, maize, barley, Brachypodium and common wheat. Only bootstrap support values below 95 are indicated on the tree. Gene labels are color-coded by species and includes the BBI group based on the classification of Mello et al. [36].

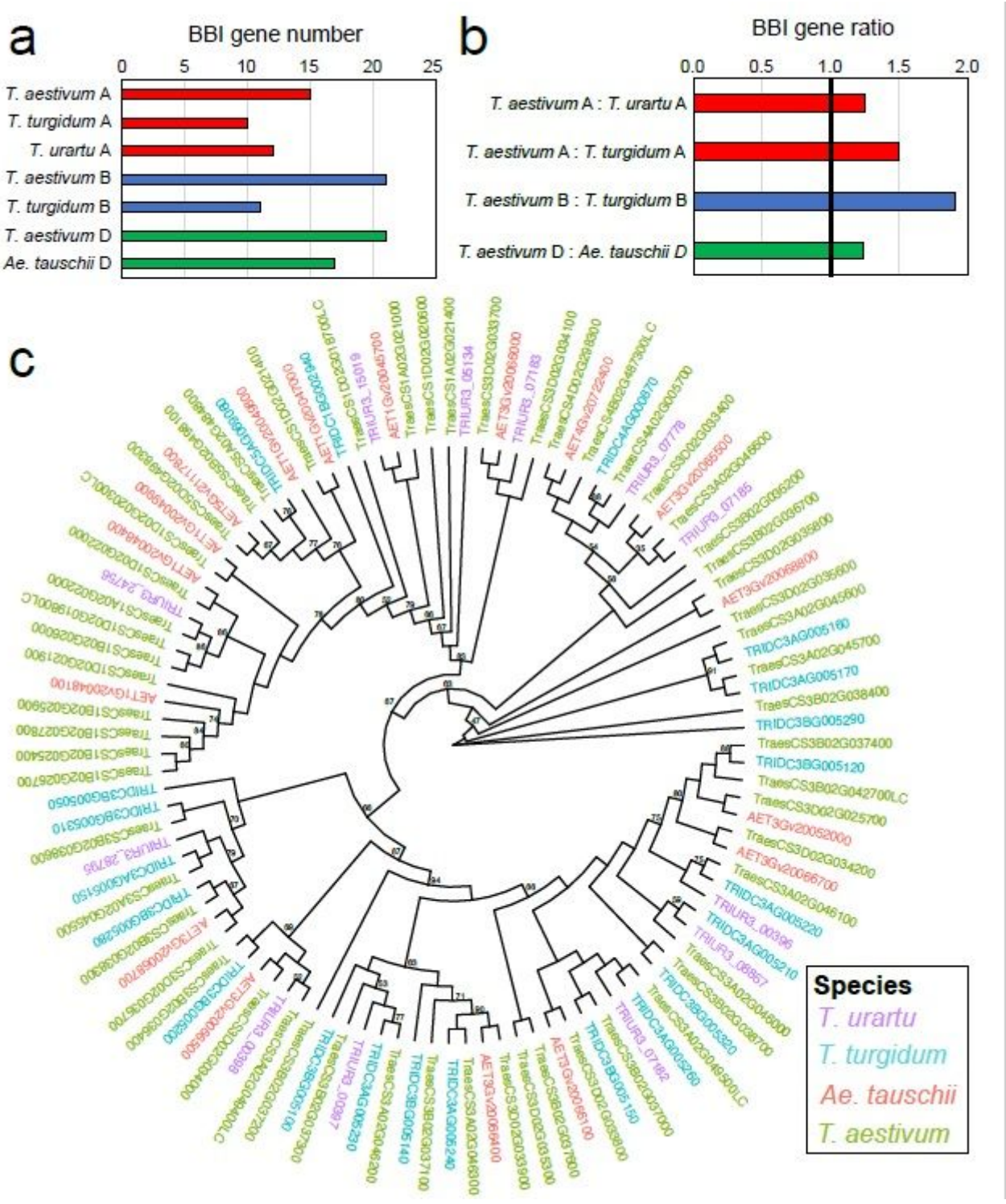


Figure 4

Comparison of the BBI gene family in different wheat germplasm. a The number of BBI genes in the genomes of different wheat species. Bars are color coded by species. b Ratio of total BBI gene numbers in common wheat compared to progenitor species. The 1:1 ratio is indicated by a bold line. c

Phylogenetic tree constructed from all BBI proteins from each wheat species. Only bootstrap support values below 95 are indicated on the tree. Genes are color-coded based on species.

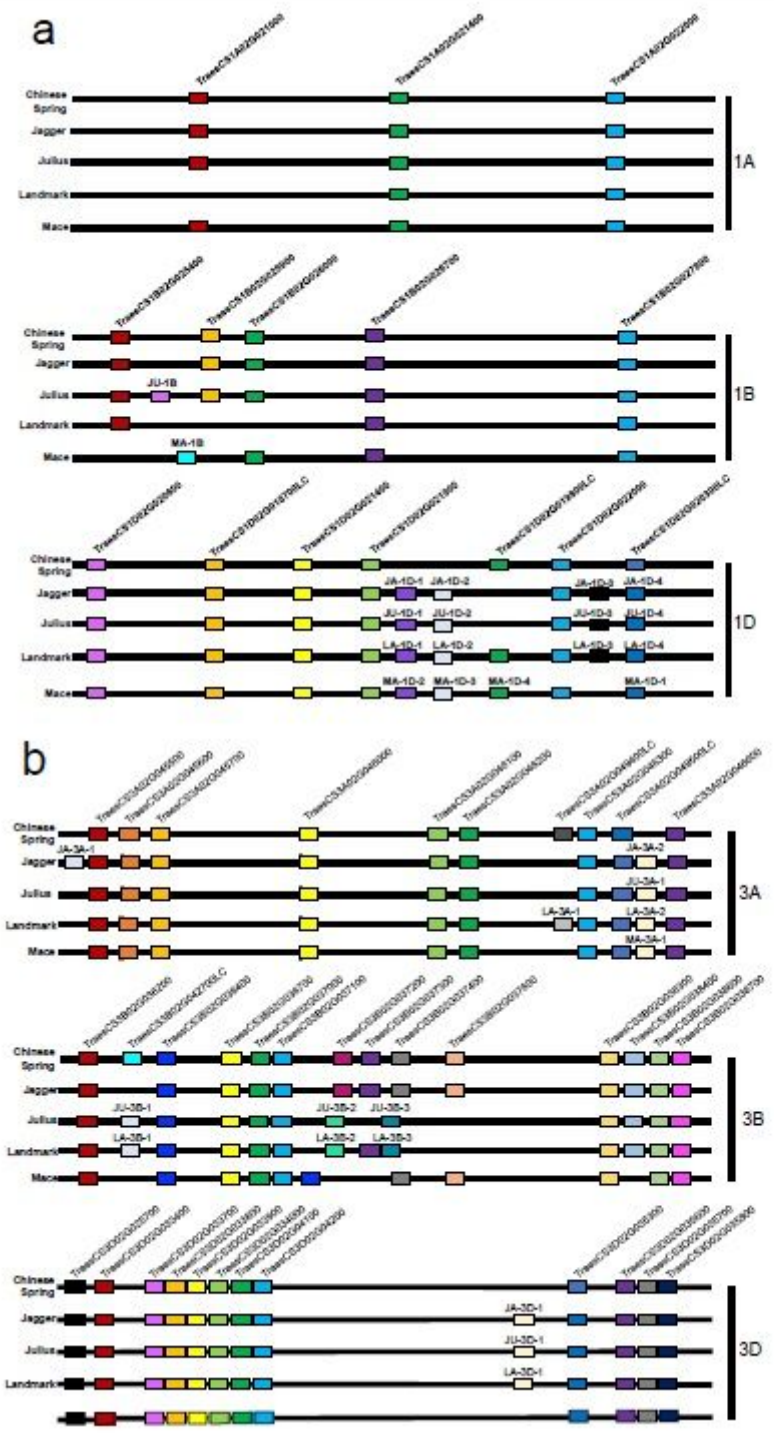


Figure 5

Distribution of BBI genes on homeologous group 1 and 3 chromosomes in different common wheat varieties. a. Homeologous group 1 chromosomes. b. Homeologous group 3 chromosomes. The ‘Chinese Spring’ BBIs are ordered according to their physical position in the IWGSC RefSeq v.1.1 genome assembly, but not to scale. Genes are colored according to their homology so that genes in the same color are

orthologous in different varieties. The BBIs present in other varieties but absent in ‘Chinese Spring’ are labeled such that JA_3A-1 indicates the first unique BBI on ‘Jagger’ chromosome 3A.

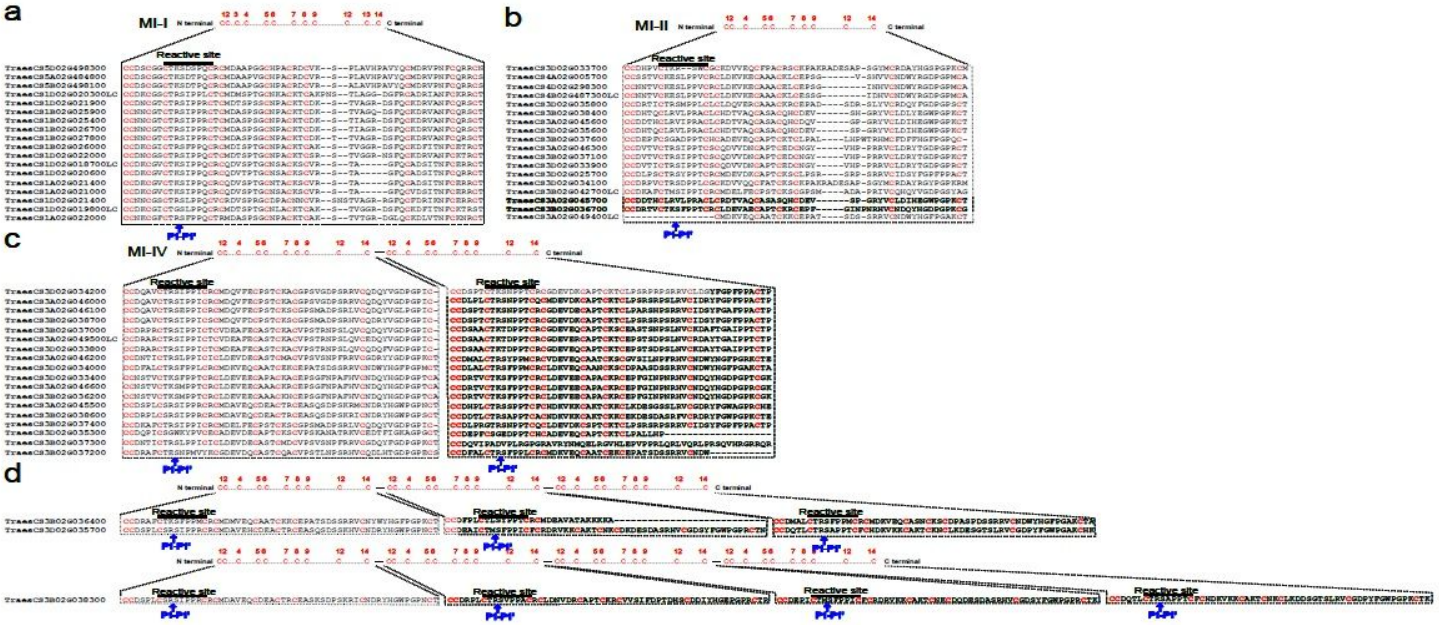


Figure 6

Alignment of the conserved Cys-rich domains of common wheat BBI proteins. a Alignment of common wheat BBI proteins falling into group MI-I; b MI-II group; c MI-IV group; and d BBI proteins that cannot be classified into an existing group. The Cys residues are highlighted in red with their corresponding position indicated above each alignment. The blue arrow underneath the domain sequences highlight the P1 and P1’ positions.

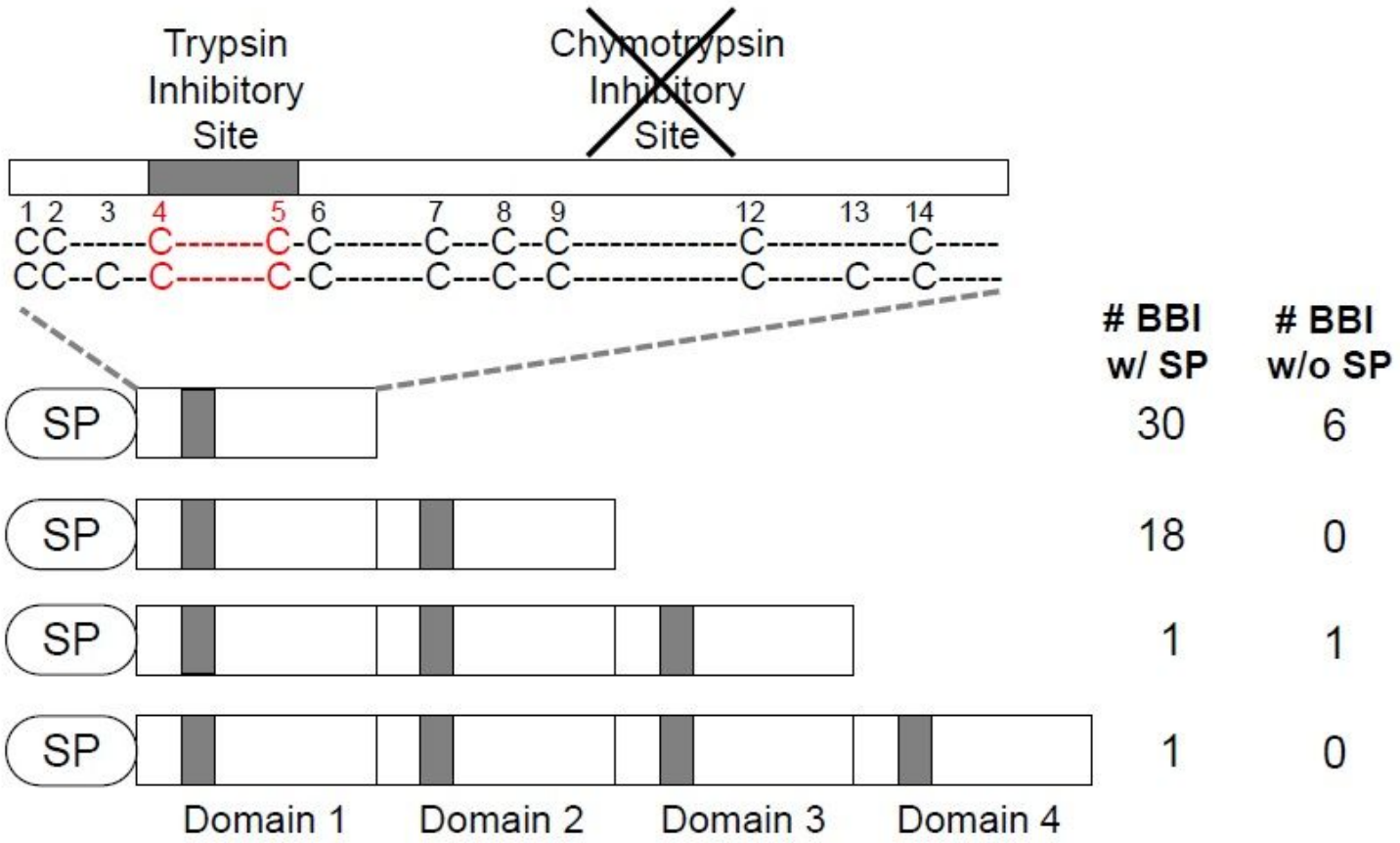


Figure 7

Summary of the proposed structural composition of the BBI gene family in common wheat. Most wheat BBI proteins contain an N-terminal signal peptide, and between one and four reactive loop domains at the C-terminus. The conserved Cys residues in the inhibitory domain are listed as C, and other amino acid residues indicated as dashes. The first reactive site is highlighted in red. The numbers of wheat BBIs from 'Chinese Spring' falling into each category are separated on the basis of presence or absence of complete signal peptides.

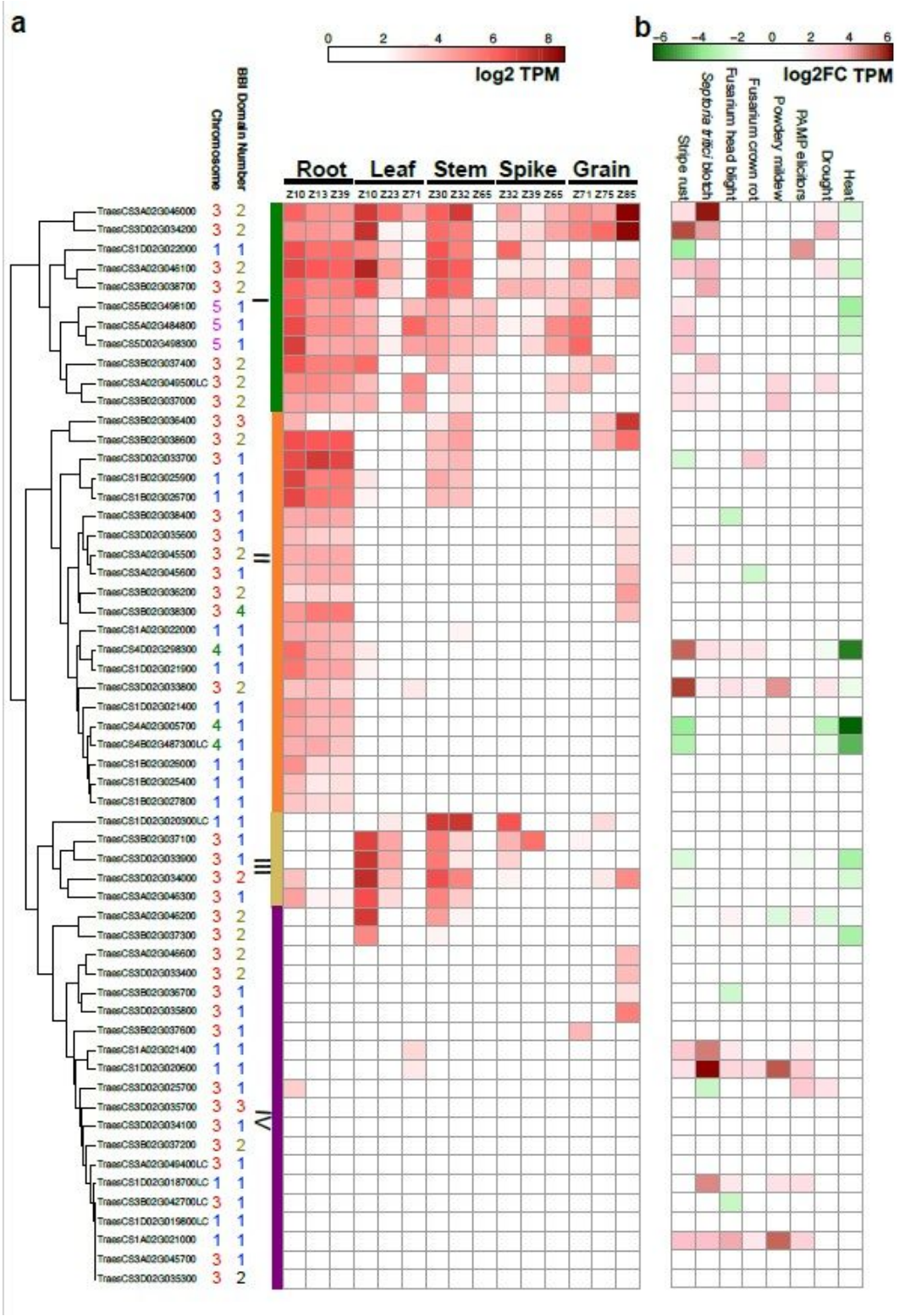


Figure 8

Expression profiles of common wheat BBI genes. a Transcript levels of BBI genes in five tissue types (root, leaf, stem, spike and grain) each at three different developmental stages based on the Zadoks scale. Genes were clustered based on their expression profile and all expression data is presented as log2 TPM. b Transcript levels of BBI genes in response to biotic and abiotic stress. Expression levels of each BBI gene in wheat plants infected with fusarium head blight, stripe rust, powdery mildew, fusarium crown

rot, Septoria tritici blotch, and bacterium chitin and flg22 (PAMP) were compared to the corresponding mock treatment. Expression is presented as log2 fold change of the TPM between pathogen treatment and mock control. Drought and heat stress data is taken from one-week-old seedlings and presented as log2 fold change of the TPM between stress and control seedlings. Where multiple timepoints were in each dataset, mean fold-changes are presented. Genes are in the same orders as in the tissue specific developmental conditions in panel a. The number of BBI domains and chromosome for each gene are color-coded.

Supplementary Files

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

- Additionalfile1.pdf
- Additionalfile2.xlsx