GenoBaits®WheatplusEE: a targeted capture sequencing panel for quick and accurate identification of wheat-Thinopyrum derivatives

Pingchuan Deng
Northwest A&F University
Xin Du
Northwest A&F University
Yanzhen Wang
Shanxi Agricultural University
Xiaoying Yang
Northwest A&F University
Xiaofang Cheng
Northwest A&F University
Chenxi Huang
Northwest A&F University
Tingting Li
Yangling Vocational and Technical College
Tingdong Li
Northwest A&F University
Chunhuan Chen
Northwest A&F University
Jixin Zhao
Northwest A&F University
Changyou Wang
Northwest A&F University
Xinlun Liu
Northwest A&F University
Zengrong Tian
Northwest A&F University
Wanquan Ji (jiwq1319@nwafu.edu.cn)
Northwest A&F University https://orcid.org/0009-0001-3767-2172

Research Article

Keywords: wheat, wild relatives, SNP array, genotyping by target sequencing, GBTS, sequence capture in solution (liquid chip), derivatives

Posted Date: June 30th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3083249/v1

Version of Record: A version of this preprint was published at Theoretical and Applied Genetics on January 31st, 2024. See the published version at https://doi.org/10.1007/s00122-023-04538-0.

Abstract

Thinopyrum species are an important source for new genetic variation for wheat improvement because they carry many agronomically important traits, such as abiotic/biotic resistances. Accurate identification of exogenous chromosome(s) or chromosome segments or genes is very important after alien genetic material has been successfully introduced into wheat, but remains challenging. Here, we report the development of a high-resolution wheat-Thinopyrum elongatum array, named GenoBaits®WheatplusEE, to trace alien genetic information through genotyping with a target sequencing system and a liquid chip. A total of 90000 capture probes derived from two species were integrated into one chip, including 10000 and 80000 within wheat and Th. elongatum, respectively. The capture probes were designed to be preferentially located in genes and evenly distributed across the genome, supporting the development of a roadmap guiding each alien genes. The array was applied to high-throughput identification of the alien chromosomes or segments in Thinopyrum and distantly related species, as well as derivatives. Our results validated that the GenoBaits®WheatplusEE array could direct identification of the breakpoint for the alien segment, the alien chromosome copy number, and the wheat chromosome variation by target sequencing the sample only once. Additionally, we can efficiently and cost-effectively genotype, supporting the exploration of subgenome composition, phylogenetic relationships, and important gene (e.g., Fhb7 gene) polymorphisms among Thinopyrum species and derivates. We expect GenoBaits®WheatplusEE to become a popular tool or model that empowers future exploration of the wild germplasm for wheat improvement.

Key message

A total of 90000 capture probes derived from wheat and *Th. elongatum* were integrated into one chip, serving as a cost-effectively genotype for the exploration of *Thinopyrum* species and derivates.

Introduction

Bread wheat (*Triticum aestivum* L., 2n = 6x = 42) is one of the most widely grown crops and supplies one fifth of the total protein and calorific requirements of mankind. Like all cultivated plant species, domestication and improvement have resulted in the wheat that feeds the world today, but at the cost of reducing its genetic variation, including resistance/tolerances to biotic/abiotic stresses (Dubcovsky and Dvorak 2007; Haudry et al. 2007). Growing evidence suggests that wheat yields are now starting to plateau around the world, and this is believed to be a direct result of a lack of genetic variation (Brisson et al. 2010; Ray et al. 2013). Luckily, many of these traits are still preserved in wild relatives of crops, which have been shown to be important sources for improving important agronomic traits in wheat, especially in dynamic abiotic and biotic threats.

Th. elongatum (syn. *Agropyron elongatum* or *Lophopyrum elongatum*), a grass of the Triticeae family, is of great interest to a diverse community of breeders and researchers as a tertiary gene pool of cultivated wheat. The *Thinopyrum* taxon is composed of a complex of polyploidy series, including diploid (2*n* = 2*x* =

14), tetraploid (2*n* = 4*x* = 28), and decaploid (2*n* = 10*x* = 70). Over the past few decades, alien genes of the *Thinopyrum* genus have been successfully transferred to wheat by distant hybridization and chromosome engineering for a set of traits that include abiotic/biotic resistances (e.g., salinity, *Sr24*, *Sr25*, *Sr26*, *Pm51*, *FHB7*) as well as morphological characteristics (e.g., blue grain) (Akhtar et al. 1994; Li et al. 2018; Tong et al. 2022; Wang et al. 2020; Zeven 1991; Zhan et al. 2014; Zheng et al. 2014a). The most recent and successful example is the transfer of a segment of *Th. elongatum* and *Th. ponticum* to wheat conferring resistance to *Fusarium head blight* (FHB) (Dvořák and Ross 1986; Guo et al. 2015; Wang et al. 2020; Zhang et al. 2022).

Traditionally, the identification of *Thinopyrum*-wheat derivatives has mainly relied on genomic *in situ* hybridization (GISH), fluorescence *in situ* hybridization (FISH), non-denaturing FISH (ND-FISH), and multicolor FISH (mcFISH). On the basis of these cytological methods, the karyotype of diploid *Th. elongatum* has been established and served as a useful resource for the rapid identification of potential donor chromosomes in wheat improvement programs (Linc et al. 2012). Recently, a chromosome painting system was developed using seven synthesized group-specific bulked oligonucleotide pools derived from single-copy sequences on chromosomes 1 to 7 of barley and the corresponding collinear regions of wheat, which could achieve high-throughput karyotyping of the chromosomes of *Thinopyrum* species (Li et al. 2021). In general, cytological-based methods can provide a whole image of the chromosomal composition and alterations in both target region and background simultaneously, but their application has been hampered by low-throughput and laborious. The development of next-generation sequencing technologies and high-throughput marker systems is an important complement to cytological detection.

Genotyping by Target Sequencing (GBTS) is an emerging technology that is developed based on targeted sequence capture and high-throughput sequencing (Guo et al. 2019). According to their capture strategy, the GBTS technology can be further classified as multiplex PCR-based target sequencing (markers number: several to 5K) and probe-in-solution-based target sequencing (markers number: 1K-20K)(Guo et al. 2019; Zhang et al. 2020). Until, multiple GBTS platforms have been developed, including SureSelect (Jiang et al. 2014), AmpliSeq (Li et al. 2017), NimbleGen (Krasileva et al. 2017), GenoBaits, and GenoPlexs (Guo et al. 2019). Due to high stability, reliability, flexibility, resolution, and cost-effectiveness, GBTS have been designed for more than 100 organisms and widely used in numerous areas such as genetic resource evaluation, genetic map construction, gene mapping, and cloning (Guo et al. 2021; Lee et al. 2022; Liu et al. 2022; Si et al. 2022).

Recently, chromosome-level assemblies have become available for both wheat ((IWGSC) 2018) and diploid *Th. elongatum* (Wang et al. 2020), which open up exciting prospects for harnessing the potential of *Thinopyrun* species. Taking advantage of the released genome and resequencing data, we identified candidate genomic regions specific for *Th. elongatum* and wheat, which were further used to design a GBTS-based integrated wheat-*Th. elongatum* liquid chip, named GenoBaits®WheatplusEE. This chip has been successfully applied in detecting and characterizing introgressions from *Thinopyrum* species

containing the E or a closely related genome such as *Pseudoroegneria spicata* (*Ps. spicata*), *Th. bessarabicum*, *Th. intermedium*, and *Th. ponticum*.

Materials and methods

GenoBaits probe design, evaluation, and synthesis

Genome sequences and annotation files for wheat (IWGSC RefSeq v1.1) and *Th. elongatum* (Thinopyrum_elongatum-REFERENCE-SDAU-1.0) were collected from the EnsemblPlants and National Genomics Data Center (NGDC) of the China National Center for Bioinformation (CNCB) database (https://ngdc.cncb.ac.cn). Whole genome sequencing data from *Th. elongatum* and the wheat cultivars (Chinese Spring and Abbondanza) were downloaded from the NCBI database under the BioProject accession code PRJNA540081 and PRJNA597250. The raw reads were trimmed using Trimmomatic (Bolger et al. 2014). To identify *Th. elongatum* specific regions, we aligned clean resequencing data from CS and Abbondanza with *Th. elongatum* reference genomes (Wang et al. 2020) using BWA-mem 0.7.17 with default parameters (Li and Richard 2010). The genomic regions that could not be covered by both CS and Abbondanza were identified as *Th. elongatum*-specific genomic regions, which were involved in 11830821 windows with total length of 1.85 Gb (40.77%). Then, *Th. elongatum*-specific genomic regions with length ≥ 100 nt were retained as liquid array candidate loci, which was further evaluated as follows: (1) All loci that contained the ambiguous base N were discarded. (2) There should be no significant sequence homology between the candidate loci and the CS reference genome sequence. (3) Loci with sequences that mapped to multiple locations on the *Th. elongatum* reference genome sequence were removed. (4) Loci overlapping with genes and their flanking regions were preferentially retained. (5) Loci were selected with the goal of optimizing coverage and even spacing throughout the genome, notably gene-poor regions. (6) Candidate regions were further filtered based on several criteria to choose the best loci for the GBTS system, which have been described in detail previously (Guo et al. 2021). CS candidate regions were identified in the same way.

Finally, 45135 target regions (5035 loci in the wheat genome, 40100 loci in the *Th. elongatum* genome) were selected as candidate target loci. These loci were further used for GenoBaits probe design and evaluation by MolBreeding Biotechnology Company (Shijiazhuang, China). As a result, 90000 liquid-based hybrid capture arrays were selected for GenoBaits®WheatplusEE, which constituted of 10000 wheat and 80000 *Th. elongatum* probes. All probes were 110 nt in size and were synthesized by Shanghai Invitrogen Biotechnology Co. Ltd. (Shanghai, China).

Plant materials and genotyping

A set of 39 diverse materials was used for the initial evaluation of the GenoBaits®WheatplusEE panel, including 4 wheat relatives, 3 wheat variates, 14 wheat-*Th. elongatum* addition lines, 16 wheat-*Th. elongatum* substitution lines, and 2 wheat-*Th. ponticum* partial amphidiploids. More detailed information can be found in Table S2. All materials were planted in the glasshouse of Northwest A & F University, Yangling, China. Genomic DNA was extracted from freeze-dried young leaves using the

cetyltrimethylammonium bromide (CTAB) method (Porebski et al. 1997). DNA concentration and quality were estimated using a NanoDrop 2000 instrument. DNA library construction, probe hybridization, and high-throughput sequencing was performed by MolBreeding Biotechnology Company (Shijiazhuang, China) as described previously (Guo et al. 2021).

In-silico analysis of sequence data

The raw reads were trimmed using Trimmomatic (Bolger et al. 2014), and high-quality clean reads were assigned to the mixed genome sequences of wheat and *Th. elongatum* by BWA-MEM (Li and Richard 2010). Potential PCR duplicate read filtering and genomic variation calling were performed using GATK v4.1.9 (McKenna et al. 2010). The allelic data genotyped by the GenoBaits®WheatplusEE panel were used to calculate the pairwise genetic distance matrix. The phylogenetic tree was constructed using the neighbor joining method within PHYLIP-3.697 software and visualized using Evolview (He et al. 2016). In addition, aligned reads were converted into a depth map for the overall targeted region by bamdst v.1.0.6 (https://github.com/shiquan/bamdst). The depth for each locus was then normalized by depth per million (DPM) in each accession.

FISH–GISH and mc-GISH

The drop method of chromosomal preparations (Han et al. 2004) was used for sequential FISH–GISH and mc-GISH analyzes. Total genomic DNA was extracted from all plant materials with a modified CTAB method. The genomic DNA from *Th. ponticum* was used for GISH probes, while sheared CS DNA was used for blocking during GISH or mc-GISH. The oligonucleotide probes Oligo-pSc119.2 (green) and Oligo-pTa535 (red) were used for sequential FISH–GISH as previously described (Yang et al. 2014).

Results

Development of high-resolution GenoBaits probes for wheat- Th. elongatum

Chinese Spring (CS) and Abbondanza are the wheat cultivars that have been used the most for distant hybridization. To design the GenoBaits array for *Th. elongatum* target capture-based sequencing, the whole genome resequencing data from CS and Abbondanza were aligned with *Th. elongatum* reference genomes (Wang et al. 2020). The genomic regions that could not be covered by both CS and Abbondanza were identified as *Th. elongatum* specific genomic regions, which participated in 11830821 windows with total length of 1.85 Gb (40.77%) (Fig. S1).

The design of the GenoBaits probes used in this study was described in the Experimental Procedures. For *Th. elongatum*, the goal of our probe design was to maximize the usability of the probes to trace each gene. After a selection pipeline, 40100 regions were ultimately selected as targets. Subsequently, we developed 80000 *Th. elongatum* bait probes to capture these targets. Each chromosome was covered by 7880 to 14315 probes (Fig. 1a). The interval length between the genes and adjacent probes was distributed mainly from 0 to 3 kb, representing 87.0% (38391/44144) of the genes on the seven assembled pseudochromosomes (Wang et al. 2020) (Fig. 1b). We observed that 42.2% of the probes were

located in the gene body region, involving 18646 different genes (Fig. 1c). Only 4.4% (1933/44144) of the gene-probe pairs were more than 100 kb apart. Additionally, a total of 7027 probes were collocated with 458 genes that are related to seed quality and disease resistance (e.g., high molecular gluten, *FHB7*, R genes) to increase resolution for their haplotyping, with an average of 15.3 probes per gene. Overall, these results indicate that we basically realize the goal of developing a roadmap that guids each gene in the *Th. elongatum* genome.

For wheat, we primarily intended to select species-specific loci that would uniformly cover all the chromosomes to ensure an even distribution of detection signals along each chromosome. The final array contained 10000 probes, targeting 5035 loci in the wheat genome. The lowest number of probes was for chromosome 4D at 332, and the highest number was for chromosome 3B at 711 probes, with a variation of 2.1 times (Fig. S2a). However, the probe density on each chromosome was similar, ranging from 0.61 to 0.88 probes per megabase (Mb) of DNA sequences. Of these probes, 82.7% were located in the genic region, with 10.2% distributed from 0 to 100 kb, and 7.1% had a distance greater than 100 kb between the gene-probe pairs (Fig. S2b).

Finally, a total of 90000 capture probes (each 110 nt in length) derived from two species were integrated into one GBTS liquid chip, named GenoBaits®WheatplusEE. The density of the probes decreased along the chromosome from the distal regions to the centromere (Fig. 1a and 1d), which was consistent with the fact that the distal regions are rich in genes ((IWGSC) 2018; Wang et al. 2020). Taken together, the probes in the GenoBaits®WheatplusEE array were well distributed throughout the genome for both wheat and *Th. elongatum*.

Sensitivity and reproducibility of GenoBaits®WheatplusEE array

To evaluate the efficiency of target capture, we tested the array using wheat variates CS and Aikang58 in DNA duplicates, as well as *Th. elongatum*. Following targeted capture and sequencing using the MGISEQ-2000 platform (PE150), approximately 4.9×10^{-7} pair-end reads were generated, leading to an expected average sequencing depth of 107.9x. More than 96.2% of the sequences aligned with the original reference genome sequence (Table S1). The highest mapping rate was obtained for CS (98.5%) and the lowest for *Th. elongatum* (89.1%). Meanwhile, about 78.1% of reads could be mapped back to the longer references from which probes were originally designed (110-bp capture probe sequence + 1000-bp flanking sequences at both sides of probe), with number ranged between 76.0% (CS repeat2) and 83.0% (*Th. elongatum*).

To display the data in a form that is independent of the absolute number of sequences, the depths for each targeted location were normalized by depths per million reads (DPM) in each sample. For CS, 99.5% of the wheat-targeted loci received a detection signal (DPM 0), 97.5% received a detection signal with DPM > 1, 0.5% was not covered in this experiment (Fig. 2a-b). Similar results were also observed for AK58, except for chromosome 1B (Fig. 2b). We observed that weak detection signals occurred in the 1BL chromosome of AK58 (Fig. 2c), suggesting the potential of using this array to confirm wheat-rye 1BL.1RS

translocation line. Furthermore, more than 78.6% of the wheat-targeted loci were not covered in *Th. elongatum*. Consequently, *Th. elongatum* achieved a high detection rate level for the target loci derived from *Th. elongatum*, 67.9% compared to 1.7% (AK58) and 0.1% (CS). For the two DNA replicates, the normalized depth distribution plot for the targeted regional sequencing was considerably consistent (Fig. 2d and Fig. S3-S4), indicating high reproducibility. In summary, the GenoBaits®WheatplusEE array generated unambiguous high-reproducibility signal patterns for each of the chromosomes, allowing the identification of individual chromosomes for wheat and *Th. elongatum*.

Characterization of chromosome constitution in wheat- Th. elongatum derivatives

The designed array was applied to direct the capture and targeted sequencing of various *wheat-Th. elongatum* derivatives to evaluate its specificity and efficiency. We first tested in 14 disomic addition lines of wheat (CS)-*Th. elongatum*. In the 7E addition line, 7E and 21 wheat chromosomes showed densely and evenly detected signals along the chromosomes, while the successful captured loci in the remaining 6 *Th. elongatum* chromosomes were sparsely and randomly distributed (Fig. 3a). Similar results were observed for the addition lines CS-1E, CS-2E, CS-3E, CS-4E, CS-5E and CS-6E (Fig. S5). Furthermore, the designed array could clearly visualize the deletion of the chromosome and the corresponding break point (Fig. 3b). According to the distribution pattern of the successfully captured loci among a set of *Th. elongatum* telosomic addition lines, the arm ratio expressed as the ratio of the length of the short arm to that of the long arm (Gill et al. 1996) was calculated for each of the *Th. elongatum* chromosomes (Fig. 3c). For example, the longest chromosome for *Th. elongatum* was 7E (~ 744.1 Mb)(Wang et al. 2020). The overlap of successfully captured loci between 7EL and 7ES was the centromere region, with a length of 19.8 Mb. Based on this, the fraction length in 7ES was 0.51 and that in 7EL was 0.49, with an arm ratio of 1.02 for 7E chromosome. However, we did not observe robust capture signals in the centromere region for CS-7ES when comparing it with CS-7E and CS-7EL, indicating that 7ES might lack a natural centromere.

We also collected 16 *wheat-Th. elongatum* substitution lines for further evaluation. Referring to the successfully captured loci, the array could determine the chromosome constitution for each material as expected (Fig. S6). As shown in Fig. 3d, most of the probes only captured their corresponding chromosomes, while very few detection signals were randomly distributed in the absence of the chromosome. For example, three disomic substitution lines each with one of the three homoeologous wheat chromosomes of group 1 (1A, 1B or 1D) were replaced by chromosome 1E, designated as CS-1E(1A), CS-1E(1B), and CS-1E(1D), respectively. The sequencing results showed that the array could not only reflect the composition of the foreign chromosome, but could also succeed in distinguishing the three homoeologous chromosomes of wheat. These results demonstrated the high specificity and efficiency of this array in capturing the wheat and *Th. elongatum* chromosomes.

Validation of array for chromosome identification in closely species

Elytrigia Desv. and *Pseudoroegneria* (Nevski) A. Löve are important resources for breeding distant hybrids of wheat. To examine the potential application of this array in distantly related species, three representative species (*Ps. spicata*, *Th. bessarabicum*, and *Th. intermedium*) and 2 wheat-*Th. ponticum* derivatives were used for targeted capture and sequencing analyzes. We found that a large number of probes designed based on the *Th. elongatum* reference genome could capture sequences in these three closely related species (Fig. 4a-c, Fig. S7a-d). The successfully captured loci appeared evenly distributed across the chromosomes, indicating good conservation of homoeologous groups among them. Using a DPM > 1 as a threshold, the highest capture rate was obtained for *Th. bessarabicum* (37.0%, 14829/40100), followed by *Th. intermedium* (26.7%, 14829/40100) and *Ps. spicata* (23.9%, 9602/40100) (Fig. S7e-f). This result indicated that *Th. elongatum* was more closely related to *Th. bessarabicum*.

A previous study has shown that it was difficult to identify the individual chromosomes of *Th. ponticum* (Host) D.R. Dewey (2n = 10x = 70) using current tandem repeat-based probes (Li and Wang 2009; Xi et al. 2019). CH18058 was a stable line derived from a cross between the wheat(7182)-*Th. ponticum* partial amphiploid line Xiaoyan784 (2n = 8x = 56). GISH using *Th. ponticum* genomic DNA as probes showed that CH18058 contained three *Th. ponticum* chromosomes (Fig. 4d). The result of FISH analysis revealed that a pair of 1D chromosomes were absent and replaced by two alien chromosomes, which was further confirmed by the result of GenoBaits®WheatplusEE array detection. Furthermore, the array data analysis also showed that another alien chromosome was a short arm of the *Th. ponticum* group 5 chromosome (Fig. 4e and Fig. S8a). Therefore, CH18058 carried 40 wheat chromosomes consisting of 14 A-genome, 14 B-genome, 12 D-genome, 2 *Th. ponticum* chromosome, and 1 *Th. ponticum* chromosome arm.

Analysis of chromosome compositions of the partial amphiploid

XiaoYan693 is a wheat-*Th. ponticum* partial amphiploid (2n = 8x = 56). Using GISH and FISH, we observed that XiaoYan693 possessed 16 *Th. ponticum* chromosomes and 40 wheat chromosomes composed of 14 A chromosomes, 12 B chromosomes (missing a pair of 6B), and 14 D chromosomes (Fig. 5a-b). This was consistent with the result of previous studies (Li et al. 2021; Zheng et al. 2014b). Based on the results of the matrix, the 1B chromosome displayed a significantly lower capture rate than other wheat chromosomes, 17.1% compared to 84.3% (average)(Fig. 5c). A comparable number of successfully captured loci were observed for the 7 *Th. elongatum* chromosomes, ranging from 775 (4E) to 1533 (2E). However, 6E has a higher sequencing depth than other chromosomes, with a 1.44-fold variation (Fig. 5d). A similar result was observed for CH18058, in which the sequence depth of the single chromosome arm could be reduced by 23.1% compared to the other pair of *Th. ponticum* chromosomes (Fig. S8b). These results indicated that we can use sequence depth to estimate the chromosome copy number.

We also compared the alien chromosomes of XiaoYan693 with other distantly related species and derivates to infer its subgenome constitution. Based on the mapped sequence reads, we remained with 644180 high-quality biallelic SNPs for constructing phylogenetic trees. The highest frequency of SNPs

was identified in chromosome 2E (17.9%, 115547), followed by 7E (17.9%, 115483), 6E (15.0%, 96590), 5E (14.0%, 90133), 3E (13.0%, 83853), 1E (12.4%, 79562) and 4E (9.8%, 63012). As expected, *Th. elongatum* and its derivatives were preferentially clustered together. For XiaoYan693, the five alien chromosome groups were more closely related to the E genome of *Th. elongatum* (E^e) and *Th. bessarabicum* (E^b), while the remaining alien chromosomes were preferentially clustered together with the St genome of *Ps. Spicata*. Based on these findings, we speculate that six chromosomes of the St genome and one of the E genome (group 5) were present in its alien genomes for XiaoYan693. Similarly, the alien genome of CH18058 may be composed of one pair of St chromosomes (1St) and one short arm of E^b chromosome ($5E^bS$) (Fig. 5e and Fig. S8).

An investigation of polymorphism in Fhb7 homologs from Thinopyrum and its derivatives

Fhb7 is a gene horizontally transferred from fungus and can confer broad resistance of wheat to head blight and crown rot without a yield penalty (Wang et al. 2020). We designed 26 probes to target capture of the gene body and flanking sequences (± 2Kb) of *Fhb7* in GenoBaits®WheatplusEE (Fig. 6a). Target sequencing of *Th. elongatum* demonstrated that *Fhb7* was successfully captured for both the coding and flanking regions using GenoBaits®WheatplusEE (Fig. 6b). Further results supported that *Fhb7* homologs were indeed detected in the genera *Thinopyrum*, including *Th. bessarabicum*, *Th. intermedium*, and *Th. ponticum* (Guo et al. 2022). XiNong 511, a leading cultivar in China, was derived from the common wheat– *Th. ponticum* line and has been shown to exhibit good resistances to FHB (Fig. 6c). However, we observed that *Fhb7* was present in XiaoYan693 but absent in XiNong511 (Fig. 6b), indicating that resistance was probably contributed by another gene than *Fhb7*. In comparison, only 12 high-quality biallelic SNPs occurred in the coding region of *Fhb7* among *Thinopyrum* species, but significantly higher levels of nucleotide variation (154 SNPs vs 12 SNPs) distributed in the flanking region. Furthermore, large fragment deletions were observed in the flanking region of *Fhb7* for *Th. intermedium* and XiaoYan693. In summary, in contrast to solid chip-based technology, the GBTS platform is highly convenient due to capture-in-solution (liquid chip), which will better serve *Thinopyrum* species functional studies and wheat improvement.

Discussion

Since the initial discovery of the *Ph1* chromosome pairing control locus in wheat, the utilization of wild relatives based on distant hybridization is the widely used and very valuable approach to improving wheat characters (Riley and Chapman 1958). However, the lack of high-throughput screening technology to quickly identify and characterize exogenous chromatin remains the first block to the large-scale exploitation of genetic variation in wild relatives. Traditional cytogenetic methods, such as GISH and FISH, are highly specialized, low throughput, and difficult to clearly distinguish between wheat and alien chromatin, especially for small alien fragment. Nevertheless, in contradiction with the situation, breeders need to minimize the alien genomic regions to eliminate the issue of linkage drag, examine the targeted genes, map their locations, and final clone.

Due to the lack of a species-specific *Thinopyrum* genotyping array, using high-density wheat SNP genotyping arrays is currently one of the most approaches to accelerate the exploitation of exotic genes of *Thinopyrum* species in breeding (Grewal et al. 2018; Yu et al. 2019; Zhang et al. 2017). Initial wheat SNP genotyping arrays were designed based on the genetic variation available in various wheat germplasms, such as the Illumina Wheat 9K iSelect SNP array (Cavanagh et al. 2013), and the Illumina Wheat 90K iSelect SNP genotyping array (Wang et al. 2014). Although these arrays have facilitated the identification of introgressions in *Thinopyrum*/wheat derivatives (Yu et al. 2019; Zhang et al. 2017), both are ineffective in detecting polymorphisms in germplasms derived from secondary and tertiary gene pools (Rasheed et al. 2018). Subsequently, several attempts have been made to design arrays by integrating high-variability loci from both wheat and wild relatives to overcome this limitation, promoting the development of the Axiom® Wheat 660K SNP array and the Axiom® HD Wheat genotyping (820K) array (Winfield et al. 2016). These genotyping arrays facilitates high-throughput screening of introgressions, including but not limited to *Thinopyrum* species (Grewal et al. 2018; Jia et al. 2022; Zhou et al. 2018). However, the application of wheat genotyping arrays in tracing introgressions is dependent on prior knowledge for the pedigree and cannot display the details on alien fragment/genes. In this work, we present the development of a high-throughput system for genome-wide screening for introgressed segments of *Thinopyrum* species in wheat background. In comparison, the GenoBaits®WheatplusEE array contained 80000 bait probes that target the capture of 40100 loci in the *Th. elongatum* genome. We have significantly reduced the probe density to reduce the initial cost of developing the probe pools while maintaining a reasonable level of resolution to basically realize a roadmap guiding each gene in the *Th. elongatum* genome. Our results validated that the GenoBaits®WheatplusEE array could not only detect alien chromosomes in wheat-*Thinopyrum* substitution and addition lines, but also direct identification of the break point for telosomic addition lines. This capability shows great potential in marker-assisted reduction of the size of introgressed segments to eliminate the issue of linkage drag.

Because the three progenitor genomes of A, B, and D are very similar to each other, it is challenging to perform specific genotyping to distinguish homoeologous chromosomes in polyploid wheat (Makhoul et al. 2020; Mao et al. 2010), which will become more difficult when adding another closely related foreign genome in wheat-*Thinopyrum* derivatives (Wang et al. 2020). Therefore, the development of the GenoBaits®WheatplusEE array is focusing more attention on species-specific capture. This is different from traditional genotyping arrays, which have a much greater appreciation of the genotype calling rate and the locus polymorphic to study genomic patterns of diversity and investigate the genetic basis of trait variation (Cavanagh et al. 2013; Wang et al. 2014; Winfield et al. 2016). To achieve this goal, the resequencing data of *Th. elongatum* and wheat were searched against each other's reference genome, then the candidate species-specific sequences were extracted and further analysis to ensure that GenoBaits probes could specifically capture the targeted sequences at the inter-genomes level, as well as within intra-subgenomes. Based on the above works, GenoBaits probes from two species were integrated into one array, providing opportunities to better utilize the assay that reveals chromosome variation for both wheat and distant species by target sequencing of the derivative lines only once. Capture sequencing analysis of a set of wheat-*Thinopyrum* derivates validated that the array could not only

reflect the composition of foreign chromosomes but could also succeed in distinguishing the three homoeologous chromosomes from wheat. Additionally, nearly 78.1% of the reads could be mapped back to the sequences on the array, which was higher than for the Axiom® HD wheat genotyping (820K) array (38%)(Winfield et al. 2016). These results proved the high specificity and efficiency of this array in capturing wheat and *Th. elongatum* chromosomes.

Instead of using array-based SNPs, previous studies have demonstrated that introgression can be identified by analyzing the coverage of data from whole-genome-sequencing (Causse et al. 2013; Keilwagen et al. 2022; Walkowiak et al. 2020) and Genotyping-by-Sequencing (GBS)(Keilwagen et al. 2019). The GBTS platform is a sequencing-based genotyping technology (Guo et al. 2019). In this study, normalized depth analysis was used to infer introgressed regions and wheat genomic variation because chromosome intervals for sequencing reads are known for the GBTS platform, which is different from the former coverage analysis (Keilwagen et al. 2019). In comparison, the normalized depth analysis enables the researcher to identify the introgressed regions, as well as infer the chromosome copy number. Furthermore, the GBTS platform is an efficient genotyping technology, which could generate multiple marker panels with different numbers of mSNPs (from 1K to up to 40K) by sequencing at different depths for only one mother mSNP panel (from 25x to 100x) (Guo et al. 2019). Based on the sequencing data derived from GenoBaits®WheatplusEE, nearly 600K high-quality biallelic SNPs were obtained for XiaoYan693, other distantly related species and derivates, supporting the exploration of the compositions and structures of the introgressed *Th. Ponticum* chromosomes in partial amphiploids. Moreover, the array will be conveniently improved with the addition or deletion of targeted capture loci as the knowledge of *Thinopyrum* species increases. There are about 50 species of *Thinopyrum* and are represented with species of various levels of ploidy including diploids, tetraploids, hexaploids, octaploids and decaploids (Lu et al. 2007). The GenoBaits®WheatplusEE array will serve as a useful platform for efficiently and cost-effectively exploring the subgenome composition and phylo-genetic relationships between *Thinopyrum* species and derivates.

Declarations

Statements & Declarations

Contributions

P.C. Deng, W.Q. Ji, and C.H. Chen conceived the project and designed the experiments. X. Du, Y.Z. Wang, C.X. Huang and X.F. Chen performed the experiments. X.Y. Yang performed an FHB evaluation for XiNong511. P.C. Deng and T.T. Li analyzed and interpreted the data. P.C. Deng and W.Q. Ji draft the manuscript, whereas J.X. Zhao, C.Y. Wang, X.L. Liu and Z.R. Tian revised the manuscript. All authors discussed the results and commented on the manuscript. All authors read and approved the manuscript.

Funding

This research is supported by grants from the basic Research Plan for Yangling Seed Industry Innovation Center (Ylzy-xm-03), the basic Research Plan for Two Chain Fusion in ShaanXi Province of China and the China National Key Research and Development Program (2021YFD1200601-04).

Acknowledgements

We thank the High-Performance Computing (HPC) of the State Key Laboratory of Crop Stress Biology in Arid Areas and Northwest A&F University for providing computing resources.

Data availability

The sequencing data reported in this article have been deposited in the National Genomics Data Center (NGDC) of the China National Center for Bioinformation (CNCB) database (accession number: CRA010636).

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. (IWGSC) IWGSC (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361:661-+
2. Akhtar J, Gorham J, Qureshi RH (1994) Combined Effect of Salinity and Hypoxia in Wheat (Triticum-Aestivum L) and Wheat Thinopyrum Amphiploids. Plant and Soil 166:47-54
3. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120
4. Brisson N, Gate P, Gouache D, Charmet G, Oury FX, Huard F (2010) Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. Field Crop Res 119:201-212
5. Causse M, Desplat N, Pascual L, Le Paslier MC, Sauvage C, Bauchet G, Berard A, Bounon R, Tchoumakov M, Brunel D, Bouchet JP (2013) Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. BMC genomics 14:791
6. Cavanagh CR, Chao SM, Wang SC, Huang BE, Stephen S, Kiani S, Forrest K, Saintenac C, Brown-Guedira GL, Akhunova A, See D, Bai GH, Pumphrey M, Tomar L, Wong DB, Kong S, Reynolds M, da Silva ML, Bockelman H, Talbert L, Anderson JA, Dreisigacker S, Baenziger S, Carter A, Korzun V, Morrell PL, Dubcovsky J, Morell MK, Sorrells ME, Hayden MJ, Akhunov E (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proceedings of the National Academy of Sciences of the United States of America 110:8057-8062
7. Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. Science 316:1862-1866

8. Dvořák J, Ross K (1986) Expression of Tolerance of Na+, K+, Mg2+, Cl− and SO2− 4 Ions and Sea Water in the Amphiploid of Triticum aestivum Elytrigia elongata. Crop Sci 26:658-660
9. Gill KS, Gill BS, Endo TR, Taylor T (1996) Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. Genetics 144:1883-1891
10. Grewal S, Yang CY, Edwards SH, Scholefield D, Ashling S, Burridge AJ, King IP, King J (2018) Characterisation of Thinopyrum bessarabicum chromosomes through genome-wide introgressions into wheat. Theoretical and Applied Genetics 131:389-406
11. Guo J, Zhang X, Hou Y, Cai J, Shen X, Zhou T, Xu H, Ohm HW, Wang H, Li A, Han F, Wang H, Kong L (2015) High-density mapping of the major FHB resistance gene Fhb7 derived from Thinopyrum ponticum and its pyramiding with Fhb1 by marker-assisted selection. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 128:2301-2316
12. Guo X, Wang M, Kang H, Zhou Y, Han F (2022) Distribution, Polymorphism and Function Characteristics of the GST-Encoding Fhb7 in Triticeae. Plants 11
13. Guo Z, Yang Q, Huang F, Zheng H, Sang Z, Xu Y, Zhang C, Wu K, Tao J, Prasanna BM (2021) Development of high-resolution multiple-SNP arrays for genetic analyses and molecular breeding through genotyping by target sequencing and liquid chip. Plant Commun 2:100230
14. Guo ZF, Wang HW, Tao JJ, Ren YH, Xu C, Wu KS, Zou C, Zhang JN, Xu YB (2019) Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. Mol Breeding 39
15. Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glemin S, David J (2007) Grinding up wheat: A massive loss of nucleotide diversity since domestication. Molecular biology and evolution 24:1506-1517
16. He ZL, Zhang HK, Gao SH, Lercher MJ, Chen WH, Hu SN (2016) Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. Nucleic acids research 44:W236-W241
17. Jia HW, Feng H, Yang GT, Li HW, Fu SL, Li B, Li ZS, Zheng Q (2022) Establishment and identification of six wheat-Thinopyrum ponticum disomic addition lines derived from partial amphiploid Xiaoyan 7430. Theoretical and Applied Genetics 135:3277-3291
18. Jiang L, Liu X, Yang J, Wang HF, Jiang JC, Liu LL, He S, Ding XD, Liu JF, Zhang Q (2014) Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. BMC genomics 15
19. Keilwagen J, Lehnert H, Berner T, Badaeva E, Himmelbach A, Borner A, Kilian B (2022) Detecting major introgressions in wheat and their putative origins using coverage analysis. Scientific reports 12
20. Keilwagen J, Lehnert H, Berner T, Beier S, Scholz U, Himmelbach A, Stein N, Badaeva ED, Lang D, Kilian B, Hackauf B, Perovic D (2019) Detecting Large Chromosomal Modifications Using Short Read Data From Genotyping-by-Sequencing. Frontiers in plant science 10
21. Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, Simmonds J, Ramirez-Gonzalez RH, Wang XD, Borrill P, Fosker C, Ayling S, Phillips AL, Uauy C, Dubcovsky J (2017)

Uncovering hidden variation in polyploid wheat. Proceedings of the National Academy of Sciences of the United States of America 114:E913-E921

22. Lee C, Cheon KS, Shin Y, Oh H, Jeong YM, Jang H, Park YC, Kim KY, Cho HC, Won YJ, Baek J, Cha YS, Kim SL, Kim KH, Ji H (2022) Development and Application of a Target Capture Sequencing SNP-Genotyping Platform in Rice. Genes 13
23. Li DY, Long D, Li TH, Wu YL, Wang Y, Zeng J, Xu LL, Fan X, Sha LN, Zhang HQ, Zhou YH, Kang HY (2018) Cytogenetics and stripe rust resistance of wheat-Thinopyrum elongatum hybrid derivatives. Mol Cytogenet 11
24. Li GR, Zhang T, Yu ZH, Wang HJ, Yang EN, Yang ZJ (2021) An efficient Oligo-FISH painting system for revealing chromosome rearrangements and polyploidization in Triticeae. Plant Journal 105:978-993
25. Li H, Richard D (2010) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics:14
26. Li HJ, Wang XM (2009) Thinopyrum ponticum and Th. intermedium: the promising source of resistance to fungal and viral diseases of wheat. J Genet Genomics 36:557-565
27. Li L, Fang ZW, Zhou JF, Chen H, Hu ZF, Gao LF, Chen LH, Ren S, Ma HY, Lu L, Zhang WX, Peng H (2017) An accurate and efficient method for large-scale SSR genotyping and applications. Nucleic acids research 45
28. Linc G, Sepsi A, Molnar-Lang M (2012) A FISH Karyotype to Study Chromosome Polymorphisms for the Elytrigia elongata E Genome. Cytogenetic and genome research 136:138-144
29. Liu YC, Liu SL, Zhang ZF, Ni LB, Chen XM, Ge YX, Zhou GA, Tian ZX (2022) GenoBaits Soy40K: a highly flexible and low-cost SNP array for soybean studies. Science China-Life Sciences 65:1898-1901
30. Lu WD, Xu PB, Xun P (2007) Summary of the situation for applying genetic resources from Elytrigia in Triticum aestivum breeding. Acta Prataculturae Sinica 16:36-140
31. Makhoul M, Rambla C, Voss-Fels K, Hickey L, Snowdon R, Obermeier C (2020) Overcoming polyploidy pitfalls: a user guide for effective SNP conversion into KASP markers in wheat. Theoretical and Applied Genetics 133:2413-2430
32. Mao P, Huang Y, Wang X, Lin M, Mao P, Zhang G (2010) Cytological evaluation and karyotype analysis in plant germplasms of Elytrigia Desv. Agricultural sciences in China 9:1553-1560
33. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 20:1297-1303
34. Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol Biol Rep 15:8-15
35. Rasheed A, Mujeeb-Kazi A, Ogbonnaya FC, He Z, Rajaram S (2018) Wheat genetic resources in the post-genomics era: promise and challenges. Ann Bot 121:603-616

36. Ray DK, Mueller ND, West PC, Foley JA (2013) Yield Trends Are Insufficient to Double Global Crop Production by 2050. PloS one 8:e66428
37. Riley R, Chapman V (1958) Genetic control of the cytologically diploid behaviour of hexaploid wheat. Nature 182:713-715
38. Si ZF, Jin SK, Li JY, Han ZG, Li YQ, Wu X, Ge YX, Fang L, Zhang TZ, Hu Y (2022) The design, validation, and utility of the "ZJU CottonSNP40K" liquid chip through genotyping by target sequencing. Ind Crop Prod 188
39. Tong CY, Yang GT, AoenBolige, Terigen, Li HW, Li B, Li ZS, Zheng Q (2022) Screening of Salt-Tolerant Thinopyrum ponticum Under Two Coastal Region Salinity Stress Levels. Frontiers in genetics 13
40. Walkowiak S, Gao LL, Monat C, Haberer G, Kassa MT, Brinton J, Ramirez-Gonzalez RH, Kolodziej MC, Delorean E, Thambugala D, Klymiuk V, Byrns B, Gundlach H, Bandi V, Siri JN, Nilsen K, Aquino C, Himmelbach A, Copetti D, Ban T, Venturini L, Bevan M, Clavijo B, Koo DH, Ens J, Wiebe K, N'Diaye A, Fritz AK, Gutwin C, Fiebig A, Fosker C, Fu BX, Accinelli GG, Gardner KA, Fradgley N, Gutierrez-Gonzalez J, Halstead-Nussloch G, Hatakeyama M, Koh CS, Deek J, Costamagna AC, Fobert P, Heavens D, Kanamori H, Kawaura K, Kobayashi F, Krasileva K, Kuo T, McKenzie N, Murata K, Nabeka Y, Paape T, Padmarasu S, Percival-Alwyn L, Kagale S, Scholz U, Sese J, Juliana P, Singh R, Shimizu-Inatsugi R, Swarbreck D, Cockram J, Budak H, Tameshige T, Tanaka T, Tsuji H, Wright J, Wu JZ, Steuernagel B, Small I, Cloutier S, Keeble-Gagnere G, Muehlbauer G, Tibbets J, Nasuda S, Melonek J, Hucl PJ, Sharpe AG, Clark M, Legg E, Bharti A, Langridge P, Hall A, Uauy C, Mascher M, Krattinger SG, Handa H, Shimizu KK, Distelfeld A, Chalmers K, Keller B, Mayer KFX, Poland J, Stein N, McCartney CA, Spannagl M, Wicker T, Pozniak CJ (2020) Multiple wheat genomes reveal global variation in modern breeding. Nature 588
41. Wang HW, Sun SL, Ge WY, Zhao LF, Hou BQ, Wang K, Lyu ZF, Chen LY, Xu SS, Guo J, Li M, Su PS, Li XF, Wang GP, Bo CY, Fang XJ, Zhuang WW, Cheng XX, Wu JW, Dong LH, Chen WY, Li W, Xiao GL, Zhao JX, Hao YC, Xu Y, Gao Y, Liu WJ, Liu YH, Yin HY, Li JZ, Li X, Zhao Y, Wang XQ, Ni F, Ma X, Li AF, Xu SS, Bai GH, Nevo E, Gao CX, Ohm H, Kong LR (2020) Horizontal gene transfer of Fhb7 from fungus underlies Fusarium head blight resistance in wheat. Science 368:844-+
42. Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant biotechnology journal 12:787-796
43. Winfield MO, Allen AM, Burridge AJ, Barker GLA, Benbow HR, Wilkinson PA, Coghill J, Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, King J, West C, Griffiths S, King I, Bentley AR, Edwards KJ (2016) High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. Plant biotechnology journal 14:1195-1206
44. Xi W, Tang ZX, Tang SY, Yang ZJ, Luo J, Fu SL (2019) New ND-FISH-Positive Oligo Probes for Identifying Thinopyrum Chromosomes in Wheat Backgrounds. International journal of molecular sciences 20

45. Yang XF, Wang CY, Chen CH, Zhang H, Tian ZR, Li X, Wang YJ, Ji WQ (2014) Chromosome constitution and origin analysis in three derivatives of Triticum aestivum - Leymus mollis by molecular cytogenetic identification. Genome / National Research Council Canada = Genome / Conseil national de recherches Canada 57:583-591
46. Yu Z, Wang H, Xu Y, Li Y, Lang T, Yang Z, Li G (2019) Characterization of chromosomal rearrangement in new wheat—Thinopyrum intermedium addition lines carrying Thinopyrum—specific grain hardness genes. Agronomy 9:18
47. Zeven A (1991) Wheats with purple and blue grains: a review. Euphytica 56:243-258
48. Zhan HX, Li GR, Zhang XJ, Li X, Guo HJ, Gong WP, Jia JQ, Qiao LY, Ren YK, Yang ZJ, Chang ZJ (2014) Chromosomal Location and Comparative Genomics Analysis of Powdery Mildew Resistance Gene Pm51 in a Putative Wheat-Thinopyrum ponticum Introgression Line. PloS one 9
49. Zhang J, Yang JJ, Zhang LK, Luo J, Zhao H, Zhang JA, Wen CL (2020) A new SNP genotyping technology Target SNP-seq and its application in genetic analysis of cucumber varieties. Scientific reports 10
50. Zhang W, Cao YP, Zhang MY, Zhu XW, Ren SF, Long YM, Gyawali Y, Chao SM, Xu S, Cai XW (2017) Meiotic Homoeologous Recombination-Based Alien Gene Introgression in the Genomics Era of Wheat. Crop Sci 57:1189-1198
51. Zhang W, Danilova T, Zhang MY, Ren SF, Zhu XW, Zhang QJ, Zhong SB, Dykes L, Fiedler J, Xu S, Frels K, Wegulo S, Boehm J, Cai XW (2022) Cytogenetic and genomic characterization of a novel tall wheatgrass-derived Fhb7 allele integrated into wheat B genome. Theoretical and Applied Genetics 135:4409-4419
52. Zheng Q, Klindworth DL, Friesen TL, Liu AF, Li ZS, Zhong SB, Jin Y, Xu SS (2014a) Characterization of Thinopyrum Species for Wheat Stem Rust Resistance and Ploidy Level. Crop Sci 54:2663-2672
53. Zheng Q, Lv Z, Niu Z, Li B, Li H, Xu SS, Han F, Li Z (2014b) Molecular cytogenetic characterization and stem rust resistance of five wheat– Thinopyrum ponticum partial amphiploids. J Genet Genomics 41:591-599
54. Zhou SH, Zhang JP, Che YH, Liu WH, Lu YQ, Yang XM, Li XQ, Jia JZ, Liu X, Li LH (2018) Construction of Agropyron Gaertn. genetic linkage maps using a wheat 660K SNP array reveals a homoeologous relationship with the wheat genome. Plant biotechnology journal 16:818-827

Figures

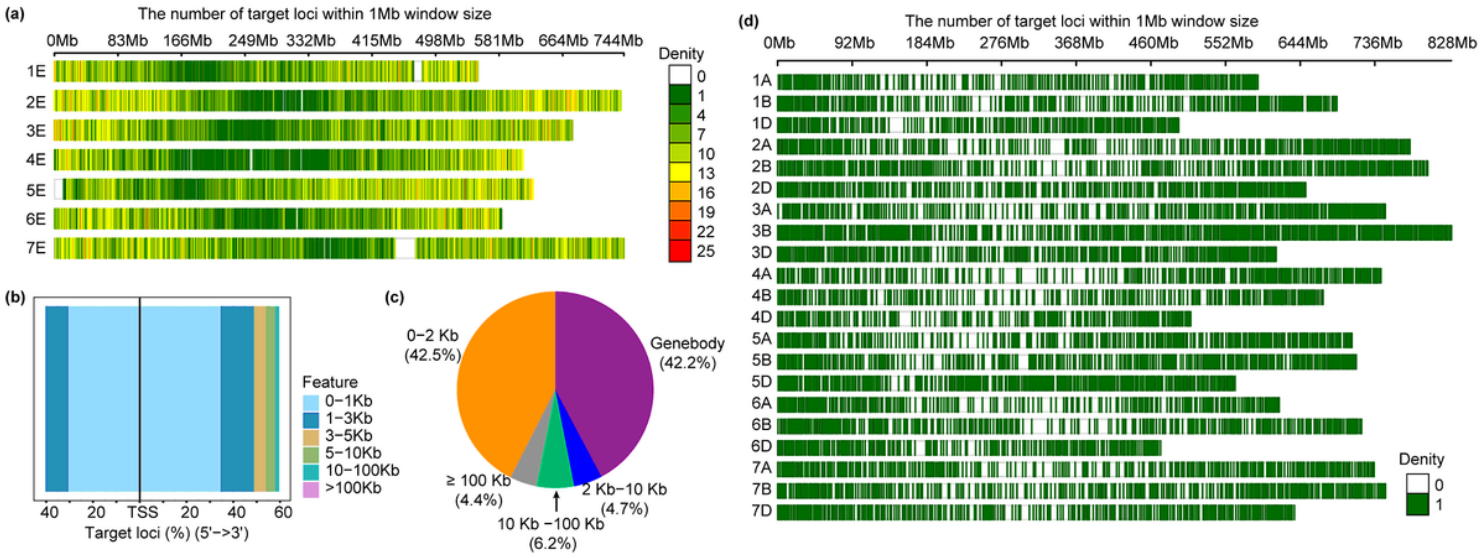

Figure 1

Characteristics of GenoBaits probes located on *Th. elongatum* chromosomes for the GenoBaits®WheatplusEE array. (a) Distribution of GenoBaits probes on seven *Th. elongatum* chromosomes. GenoBaits probes are indicated by different bar colors, and each bar represents a window size of 1 Mb. **(b)**Proportions of the distance between each GenoBaits probe and its closest gene transcription start site (TSS). **(c)**Proportions of the distance between each gene and its nearest GenoBaits probe. A total of 44,144 high-confidence protein-encoding genes were annotated and anchored on the seven chromosomes of *Th. elongatum*. **(d)** Distribution of bait probes on 21 wheat chromosomes. The bait probes are indicated by different bar colors and each bar represents a 1-Mb window size.

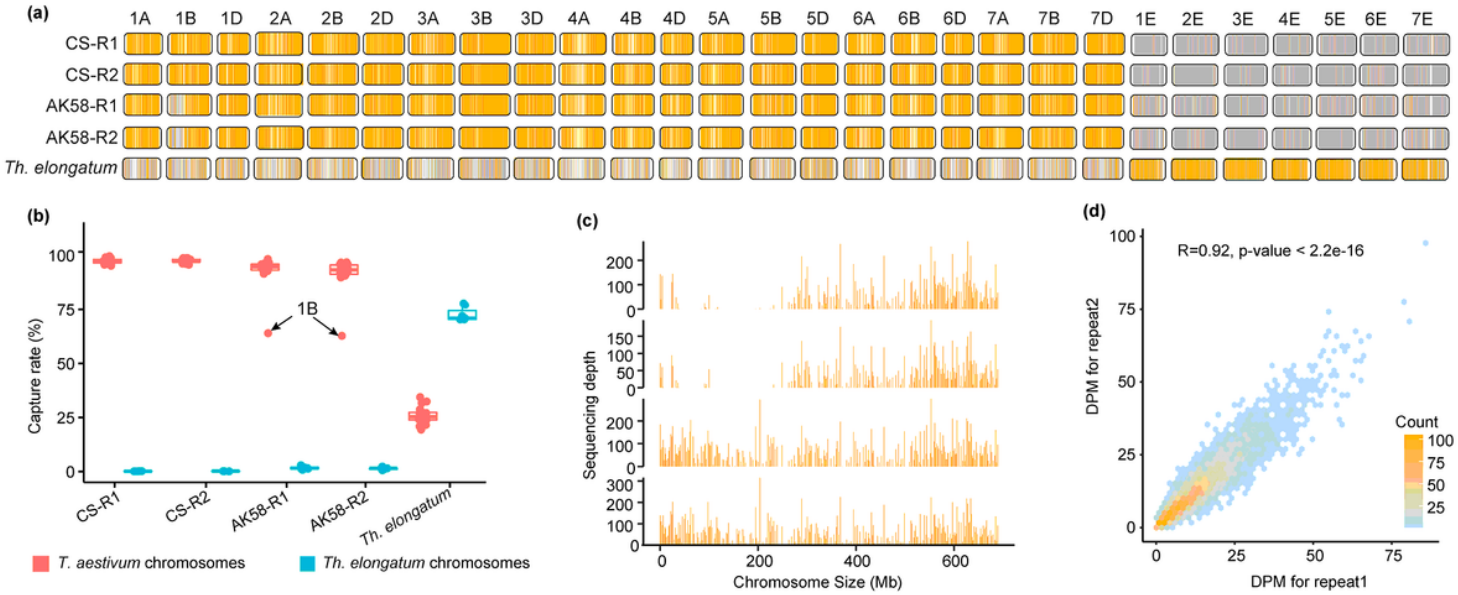

Figure 2

Assessment of reproducibility and sensitivity. (a) The distribution of the successful capture of target loci along the chromosomes. Orange represents the successful capture of the target locus (DPM≥1), while grey represents the failure locus. Two biological repeats were performed for both AK58 and CS, which corresponded to a wheat variety and landrace, respectively. **(b)** Boxplots showing the distribution of the successful capture rate for each chromosome. **(c)**Normalized depth for each target locus along the wheat chromosome 1B. **(d)** Correlation coefficients between the sequencing depth of the targeted loci from two biological duplicates of CS. The colors are proportional to the numbers of targeted loci.

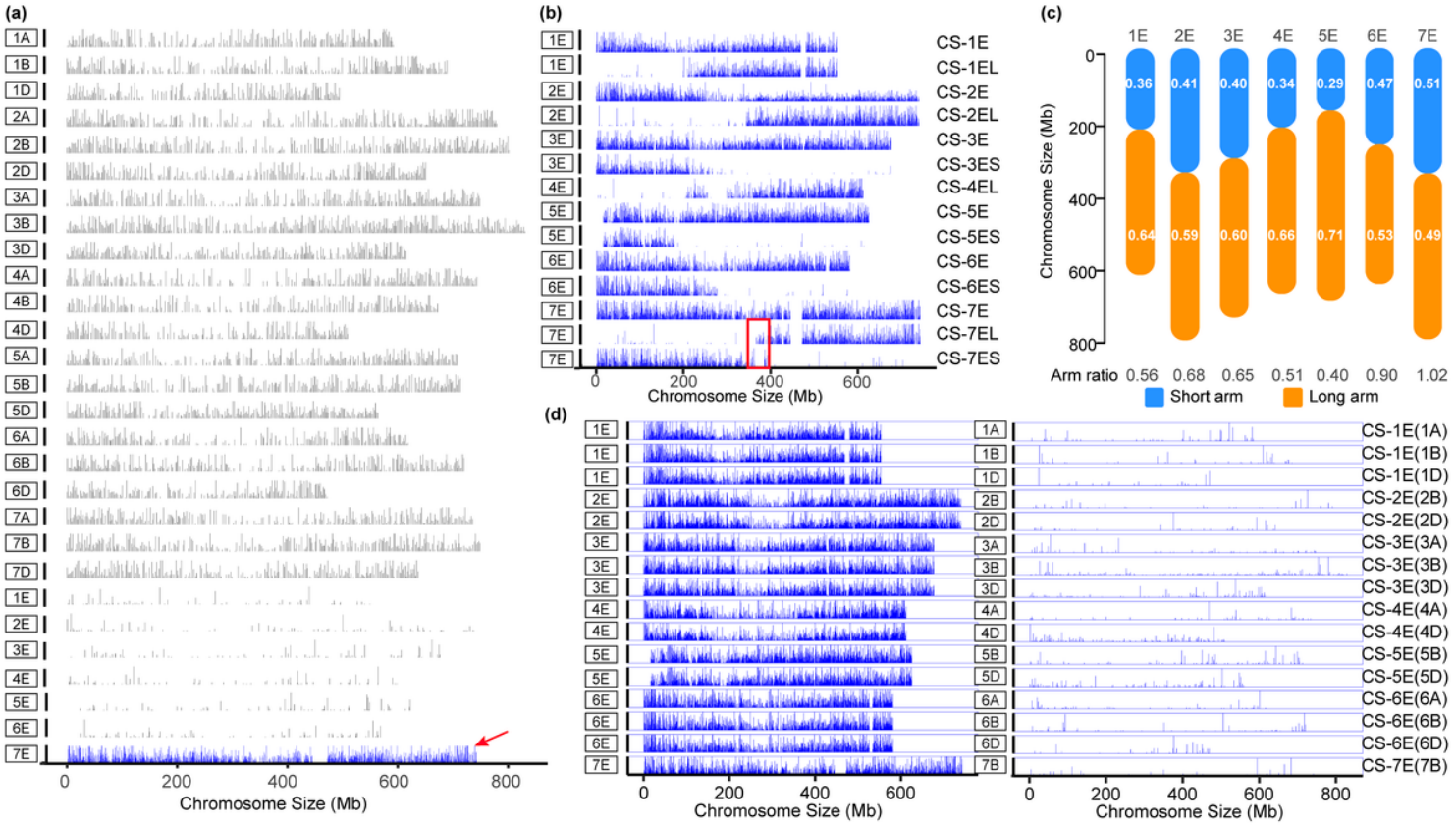

Figure 3

Application of the GenoBaits®WheatplusEE array in the identification of alien chromosomes in wheat-*Th. elongatum* derivatives. (a) Identification of alien chromosome in the disomic wheat (CS)-*Th. elongatum* addition line 4E (CS-4E). **(b)** Identification of *Th. elongatum* chromosomes in a set of wheat-*Th. elongatum*addition lines. **(c)** Evaluation of the candidate centromere region for *Th. elongatum* chromosomes based on capture sequencing analysis for a set of wheat-*Th. elongatum* ditelosomic additions. **(d)** Characterization of the chromosome composition in a set of wheat-*Th. elongatum* substitution lines. The x-axis indicates the physical position in Mb, while the y-axis indicates the distribution and abundance of sequencing depth (DPM) for each targeted capture locus along the chromosomes.

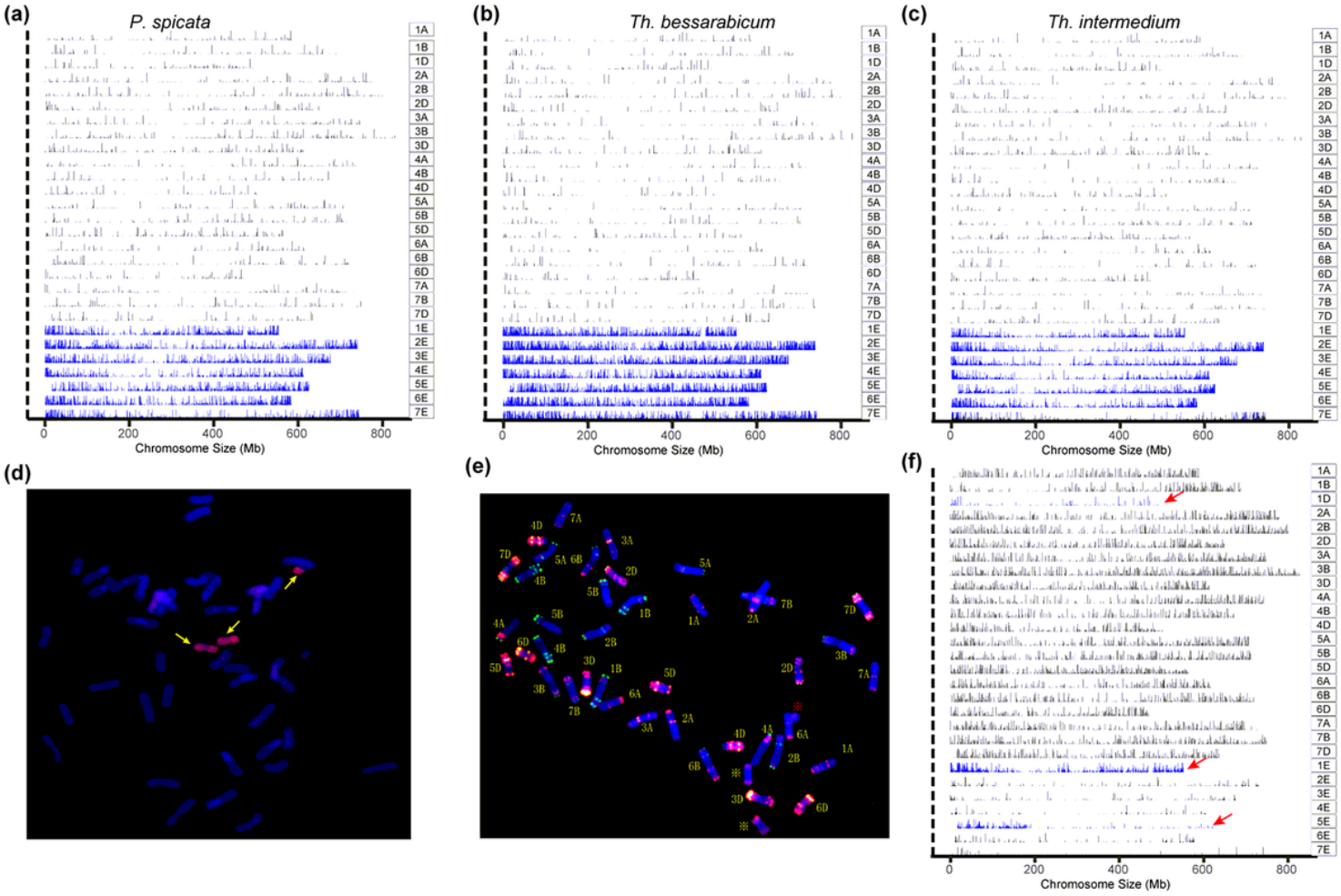

Figure 4

Capture sequencing analysis of *Thinopyrum* and closely related species. (a) to (c) The distribution and abundance of capture sequencing depth for *Ps. spicata* (**a**), *Th. bessarabicum* (**b**) and *Th. intermedium* (**c**). **(d)** GISH analysis of the wheat-*Th. ponticum* addition line CH18058 using genomic DNA from *Th. ponticum* as a probe (red). Chromosomes were counterstained with DAPI (blue). The alien chromosomes are indicated by yellow arrows. **(e)** FISH analysis of CH18058 using Oligo-pSc119.2 (green) and Oligo-pTa535 (red) probes. Chromosomes were counterstained with 4',6-diamidino-2-phenylindole. Green and red signals are from FAM-green and digoxigenin-red probes, respectively. **(f)** Distribution and abundance of the capture sequencing depth for CH18058.

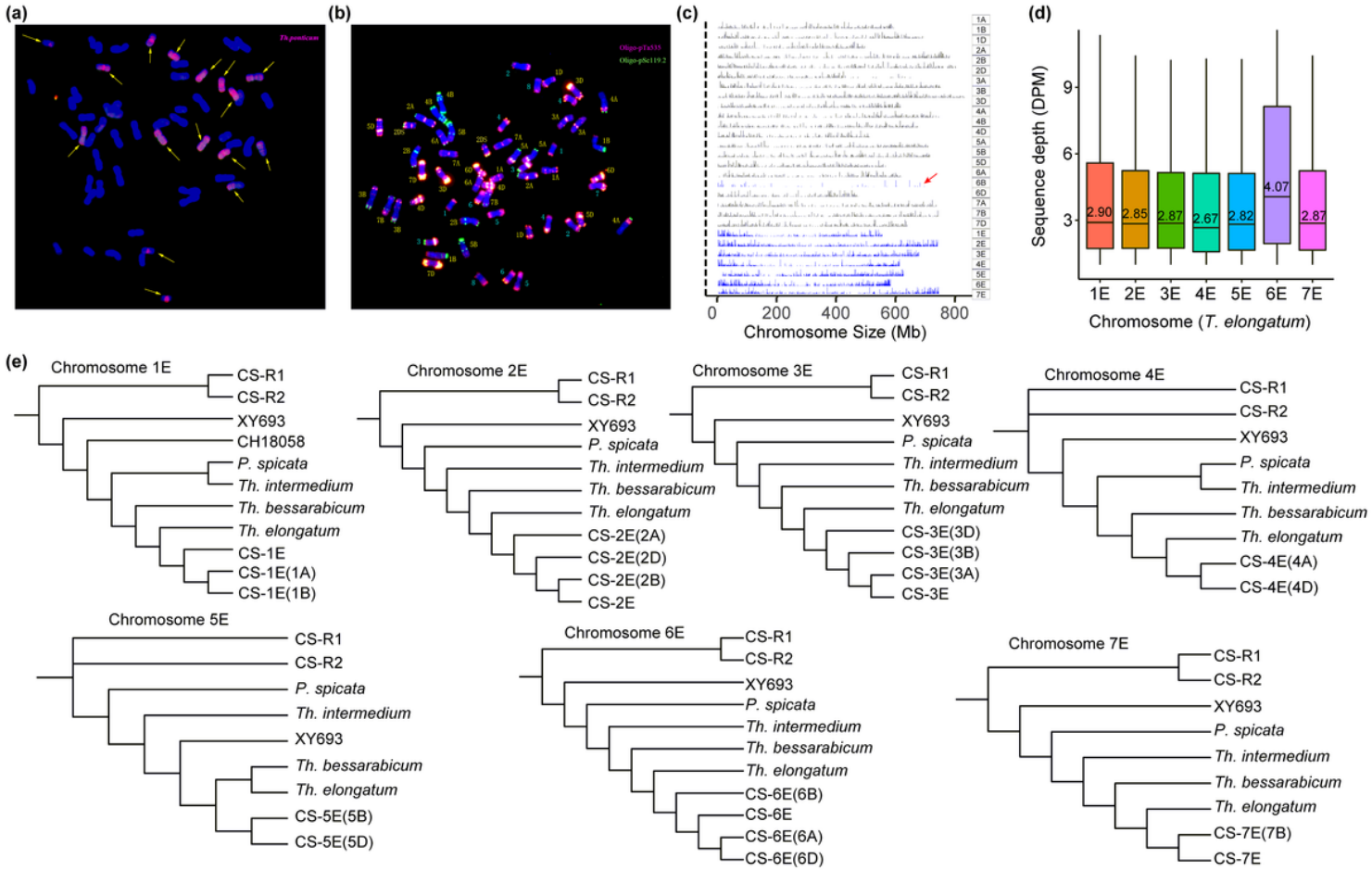

Figure 5

Capture sequencing analysis of a wheat-*Th. ponticum* partial amphiploid XY693by the GenoBaits®WheatplusEE array. (a) Sequential GISH analysis of XY693 using *Th. ponticum* genomic DNA as a probe (red). Chromosomes were counterstained with DAPI (blue). The alien chromosomes are indicated by yellow arrows. **(b)** FISH analysis of XY693 using Oligo-pSc119.2 (green) and Oligo-pTa535 (red) probes. **(c)** The distribution and abundance of capture sequencing depth for XY693. **(d)** Boxplot showing the distribution of sequencing depth for each chromosome. **(e)** Phylogenetic relationship between XY693 and other derivatives of wheat-*Th. elongatum*.

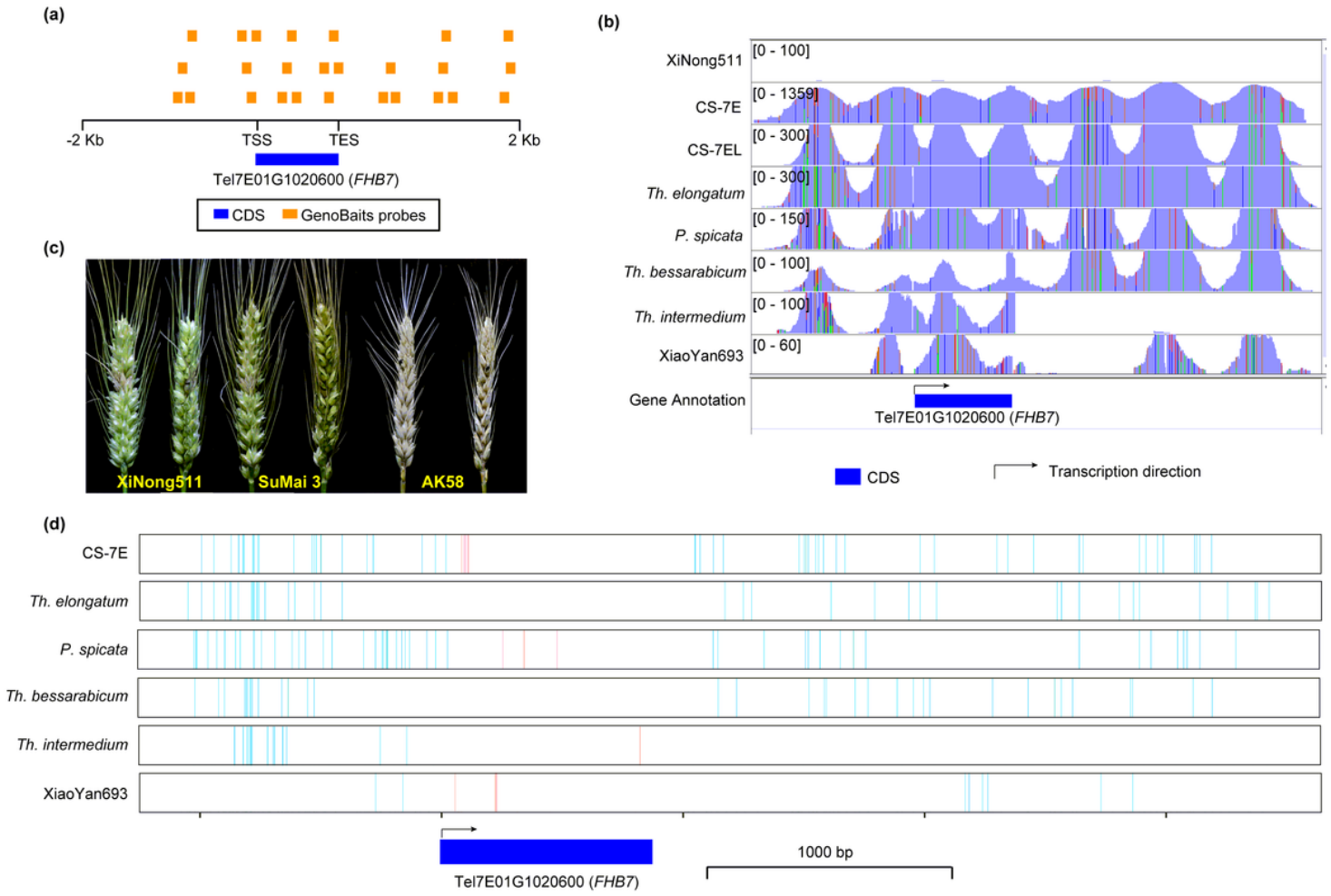

Figure 6

Application of the GenoBaits®WheatplusEE array for the genotyping of *Fhb7* gene. (a) Probe design to capture the genic and nearby region (± 2Kb) of the *FHB7* gene**.** TSS represents the transcription start site, while TES represents the transcription end site. **(b)** Genome browser view of capture sequencing for the *FHB7* gene**. (c)** Infection of wheat spikes with *FHB*(F.g. PH1). **(d)** Overview of SNPs distributed in *FHB7* of *Thinopyrum*and its derivatives.

Supplementary Files

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

- Supplementalfigures.docx
- Supplementaltables.xlsx