Targeted sequencing of the short arm of chromosome 6V of a wheat relative *Haynaldia villosa* for marker development and gene mining

Jin Xiao

Nanjing Agricultural University

Wentao Wan

Nanjing Agricultural University

Mengli Li

Nanjing Agricultural University

Zhongyu Yu

Nanjing Agricultural University

Xu Zhang

Nanjing Agricultural University

Jia Liu

Nanjing Agricultural University

Kateřina Holuřov

Centre of the Region Han for Biotechnological and Agricultural Research

Jan Vrna

Centre of the Region Han for Biotechnological and Agricultural Research

Jaroslav Doleřel

Centre of the Region Han for Biotechnological and Agricultural Research

Yufeng Wu

Nanjing Agricultural University

Haiyan Wang

Nanjing Agricultural University

Xiue Wang (✉ xiuew@njau.edu.cn)

Nanjing Agricultural University <https://orcid.org/0000-0002-6312-3417>

Research article

Keywords: Comparative analysis, chromosome isolation, DNA marker development, flow cytometric sorting, *Haynaldia villosa*

Posted Date: January 29th, 2020

DOI: <https://doi.org/10.21203/rs.2.22109/v1>

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

[Read Full License](#)

Abstract

Background: Short arm of chromosome 6V (6VS) of *Haynaldia villosa* has been used in wheat breeding programs to introduce Pm21 resistance gene against powdery mildew and some other genes.

Results: In this work, 6VS was isolated from a wheat (*Triticum aestivum*) - 6VS telosome addition line by flow cytometric sorting and sequenced by illumina technology. The assembly length was 230.39 Mb with contig N50 of 9,788 bp. The sequence annotation identified 3,276 high confidence genes supported by RNA sequencing data, representing about 2.3% of the chromosome arm sequence; repetitive elements accounted for 74.91% of the arm sequence. Sequences homologous to 6VS genes were identified on short arms of chromosomes 6A of *T. urartu*, 6D of *Aegilops tauschii*, 6A and 6B of *T. dicoccoides*, 6A, 6B and 6D of *T. aestivum* and 6H of *Hordeum vulgare*, revealing synteny relationships among these chromosome arms. Based on differences in intron size between the homologous genes on 6VS and 6AS/6BS/6DS of *T. aestivum*, 222 primer pairs were designed. Out of them, 120 amplified 6VS-specific products and are suitable as intron-target (IT) markers to trace the 6VS chromatin introduced into wheat.

Conclusions: The results obtained and markers developed in this work will facilitate introduction of important genes to common wheat from its wild relative, while reducing the presence of unfavorable genes due to linkage drag.

Background

Haynaldia villosa L. ($2n = 14$, genome VV) is a wild relative of common wheat (*Triticum aestivum* L.) carrying resistance genes to numerous wheat diseases, including powdery mildew, wheat yellow mosaic virus, eyespot, take-all and rusts [1]. It has also been credited for improving tillering [2, 3], high grain protein content [4–6], and tolerances to frost and drought of wheat [7, 8]. These characters make *H. villosa* a highly attractive source of important genes and alleles for wheat improvement [1]. In the previous study, several useful genes were mapped on short arm of chromosome 6V, such as the *Pm21* locus, which provides immunity or high resistance to all powdery mildew isolates, and *NAM-V1*, which contributes to increased grain protein content (GPC) in the wheat-*H. villosa* 6AL/6VS translocation lines [9, 10]. However, the lack of the genome sequence hampered the efforts to mine other important genes from *H. villosa* and the use of molecular tools to introduce them to wheat, while avoiding unfavorable alien chromatin.

The progress in DNA sequencing technology now makes the production of whole genome sequence assemblies feasible by whole genome shotgun approaches and the number of sequenced genomes of wheat relatives keeps on increasing [11–13]. However, if chromosomal location of the loci of interest is known, the option is to sequence only the chromosome, or chromosome arm of interest. This approach significantly reduces the project costs and thus enables sequencing chromosomes from multiple lines of a species, if needed. It also simplifies bioinformatic analyses due to reduced volume of sequence data.

Targeted sequencing of a particular chromosome is possible after isolating a required number of chromosomes by flow cytometric sorting [14–21]. Next generation sequencing of flow-sorted chromosomes has been used to develop molecular markers in *Aegilops geniculata* and *H. villosa* [22–24]. Importantly, sequencing DNA from flow sorted chromosomes facilitated the production of draft genome assemblies of barley [25], rye [26] and common wheat [27] and to isolate genes in wheat and barley either by the MutChromSeq strategy [28] or the TACCA approach [29].

Purification of a particular chromosome by flow sorting may be hampered by the inability to discriminate the chromosome from other chromosomes in a karyotype if its size or relative DNA content is not different. Various strategies have been developed to overcome this difficulty and one of them is to sort translocation or deletion chromosomes with altered size [15–19]. Larger deletions are not viable in diploids, but they may be developed from wild type chromosomes after they are introduced to a polyploid species, such as wheat, which tolerate aneuploidy. Thus, Tiwari et al. sorted chromosome 5Mg from a wheat/*Ae. geniculata* disomic substitution line [20]. In a similar way, Xiao et al. used a wheat-alien ditelosomic addition line “NAU1201” to isolate chromosome arm 4VS of *H. villosa*. Thus, the line needs not to be prepared exclusively for flow cytometric sorting and may be already available [21].

In this work, a wheat-alien addition line containing a pair of short arms of chromosome 6V of *H. villosa* was used to isolate, sequence and assemble the sequence of 6VS. The draft sequence made it possible to characterize molecular composition of 6VS including DNA repeat content, identify genes and characterize syntenic relationships with the genomes of tribe Triticeae and other sequenced grasses. The 6VS sequences would also be used to develop PCR-based 6VS-specific markers, which will support alien introgression breeding of wheat and the cloning of favorable genes from 6VS.

Results

Flow sorting and sequencing of chromosome arm 6VS of *H. villosa*

Flow cytometric analysis of chromosomes isolated from *T. aestivum*-*H. villosa* 6VS ditelosomic addition line resulted in bivariate flow karyotypes FITC (log scale) vs. DAPI (linear scale) fluorescence on which a number of populations could be resolved (Figure 1). The population representing 6VS telosome was identified after screening all populations with lower DAPI fluorescence, which were expected to correspond to smaller chromosomes. Microscopic analysis of flow-sorted particles after FISH with probes for pSc119.2 and Afa family repeats enabled unambiguous identification of the population representing 6VS telosomes (Figure 1). A detailed microscopic analysis showed that 6VS telosome could be sorted at an average purity of 89.41%. The sorted DNA was amplified by multiple displacement amplification (MDA) reactions before illumina sequencing. Sequencing of DNA amplified from flow-sorted chromosome 6VS in illumina MiSeq system generated 47.7 Gb high-quality paired-end reads from two libraries, with insert sizes of 500 bp and 1,000 bp, respectively.

De novo assembly was performed using the software Hecate (<http://bgi-international.com/us/>, unpublished) with different k-mer sizes (41, 45, 49 and 63). The result of the 45-mer run provided the assembly with the best sequence coverage and N50 size, and therefore was used to generate the 6VS scaffolds. The sequencing data and detailed assembly for 6VS are summarized in Table 1. A total length of 230.39 Mb assembled sequences was obtained, comprising 153,177 scaffolds. The maximum and minimum lengths of the scaffold were 138,620 bp and 100 bp, respectively, with contig N50 length of 9,788 bp and mean length of 1,464 bp.

Identification of repetitive DNA elements

The overall repetitive DNA composition including transposable elements (TEs) and tandem repeats across the 6VS assembly was analyzed. When compared with two DNA repeat databases combined, the Repbase UPdate library [30] and the TREP library [31], a total of 74.91% of the 6VS assembly corresponded to repeat elements (Table. 2). Among the repeat elements, the most abundant were LTR retrotransposons comprising about 88%, out of which *Gypsy* superfamily repeats accounted for about 69.3%, followed by *Copia* superfamily, which comprised 14.43%. DNA transposons were mainly represented by *TIR* family, which made up about 8.72% of all repeats. After masking all repetitive DNA elements, the remaining non-repetitive sequence reads from 6VS equaled 49.1 Mb, which was used for the following gene prediction and sequence comparisons.

Gene content of chromosome 6VS

The repeat-masked sequences of 6VS were used for *ab initio* gene prediction by AUGUSTUS program which identified a total of 5,973 loci of predicted coding sequences. The unigenes from the RNA-seq data obtained from leaves and endosperm of *H. villosa* were used to support the identification of coding loci. Using both *ab initio* and evidence-based gene predictions, we finally identified a total of 3,276 high-confidence protein-coding loci on 2,871 scaffolds of 6VS. The gene length distribution is shown on Fig. S1A. The genic sequences represented a total length of 5,278,412 bp, which accounted for 2.3% of the 6VS assembly. A total of 1,672 genes were functionally assigned to one or more Gene Ontology (GO) terms (Fig. S1B). To summarize, 1,441, 1,510 and 1,432 genes were annotated with cellular component, molecular function and biological process, respectively.

In order to test the annotation quality of the 6VS, we used genes *NLR-V* [32], *STPK-V* [33] and *NAM-V1* [9] cloned from 6VS to perform BLASTn search. We found sequences homologous at 99.93%, 100.00% and 99.93%, respectively, implying a high quality of *H. villosa* 6VS annotation. Thus, the 6VS draft sequence obtained in this work will facilitate extensive mining of 6VS genes in wheat breeding.

Comparative analysis of 6VS sequence composition

We then used a set of “toplevel” wheat sequences consisting of molecule-level assemblies (<http://plants.ensembl.org/>) that were released by IWGSC [34] to identify 6VS syntenic regions on wheat chromosomes 6A, 6B and 6D. We also identified 6VS syntenic regions on chromosomes 6A and 6B of tetraploid *T. dicoccoides*, 6D of *Ae. tauschii*, 6A of *T. urartu* and 6H of *H. vulgare*. All predicted genes were used to identify possible syntenic regions in genomes of the related grass species. After filtering, 2,867 6VS predicted genes had 1,499, 1,577 and 1,430 blastn hits with homologous genes in wheat chromosomes 6A, 6B and 6D, respectively; the number of hits with homologous genes in *T. dicoccoides* chromosomes 6A and 6B was 1,323 and 1,374, respectively; in *Ae. tauschii* chromosome 6D it was 1,301, in *T. urartu* 6A it was 1,424 and in *H. vulgare* 6H it was 1,307. Moreover, 634 out of 2,867 identified homologous syntenic genes were shared among all eight genomes. The syntenic genes on *T. aestivum* chromosomes 6A, 6B and 6D, *T. dicoccoides* chromosomes 6A and 6B, *Ae. tauschii* chromosome 6D, *T. urartu* chromosome 6A and *H. vulgare* chromosome 6H were plotted according to the position of their respective chromosomes to highlight the syntenic regions (Fig. 2). Similarly, the syntenic regions with high gene density were observed on chromosomes 6AS, 6BS and 6DS of *T. aestivum*, 6AS and 6BS of *T. dicoccoides*, 6DS of *Ae. tauschii*, 6AS of *T. urartu* and 6HS of *H. vulgare*.

NB-ARC domain proteins are commonly known as disease resistance genes. In the 6VS assembly, a total of 45 genes were predicted to encode NB-ARC domain proteins using HMMER model [35]. In a separate project, we analyzed transcriptome of the wheat-*H. villosa* translocation line T6VS/6AL after the treatment with two *Blumeria graminum f.sp tritici* (*Bgt*) isolates E26 and E31 (data not shown). We found that 28 genes were expressed after inoculation of both isolates within 24 hours, with 15 genes up-regulated two-fold or more when compared to the control (Figure 3). As 6VS chromatin introduced to wheat showed the main contributor of the resistance to various *Bgt* isolates, 6VS genes might be involved in the innate immunity of *H. villosa* to powdery mildew.

Wheat cultivars with 6VS/6AL translocation have been used extensively in wheat production with more than four million hectares in China [32], not only due to broad-spectrum resistance to powdery mildew, but also due to their contributions to higher 1000-grain weight (TGW) [36]. In a previous study, *TaGW2-6A* was described as negative regulator of grain-width and grain-weight [37–39]. Four SNPs that occurred in the promotor region of *TaGW2-6A* were reported to be associated with TGW at positions –998bp, –739bp, –593bp and –494bp, in which SNP at –494 bp showing significant association with TGW and was located in the ‘CGCG’ motif [37]. SNP–494 has most effect on *TaGW2-6A* expression level and TGW, with haplotypes with A allele having significantly lower *TaGW2-6A* expression and higher TGW compared with those with G allele. To figure out if the increased TGW of 6VS/6AL translocations was due to the substitution of 6AS with 6VS, the *TaGW2-6A* gene homologue *HvGW2-6V* was identified in the 6VS assembly. The *HvGW2-6V* in *H. villosa* belongs to G allele at SNP–494, which was associated with low TGW (Figure 4). We speculate that higher TGW of 6VS/6AL translocations might be affected by other genes rather than *GW2-6V*, or that the expression of alien gene is suppressed due to genomic shock in wheat background although the genotype at position –494 was the same with low TGW.

Development of intron targeted (IT) markers

Zhang et al. and Wang et al. developed IT markers for all chromosomes except for 6VS of *H. villosa* [22, 23]. With the shotgun sequences of 6VS, IT markers for this short arm could now be developed. All 2,063 annotated genes along the pseudomolecule of *Ae. tauschii* 6DS were aligned by Blastn search against the wheat genome reference and 6VS assembly sequences to determine exon-exon junction lengths on chromosomes 6AS, 6BS, 6DS of *T. aestivum* as well as on 6VS. A total of 222 genes had the intron length in *H. villosa* differing by at least 10% as compared to those in wheat subgenomes A, B and D. Then, we designed PCR primers in the conserved exon regions which spanned the targeted introns using the online software Primer 3 V0.4.0 (<http://frodo.wi.mit.edu/primer3/>).

In order to test the 222 IT primers on 6VS of *H. villosa*, we perform PCR analysis for the DNA from *T. aestivum* cv. Chinese Spring (AABBDD), *H. villosa* (VV) and *T. aestivum*-*H. villosa* translocation line T6AL·6VS. If the amplification of a primer pair generated a distinct PCR product visualized as polymorphic band in *H. villosa*, and *T. aestivum*-*H. villosa* translocation line T6AL·6VS, but not in common wheat, the primer pair was considered 6VS-specific. In total, 120 markers were obtained with the success rate of developing 6VS chromosome arm-specific molecular markers as high as 54.05% (Table S1). All IT markers were tested on three different translocation lines, NAU418, NAU419 and NAU1203, involving 6VS with different introgressed segments. The chromosome arm could be dissected into four bins: bin1 to bin4 (Fig 5), which contained 34, 11, 46 and 29 markers, respectively. Given that all three translocation lines are resistant, the resistant gene *Pm21* was mapped within bin3. The 40 markers within this physical bin are suitable for marker-assisted breeding.

Discussion

Aneuploidy germplasm facilitate flow-sorting target chromosomes or its arms

In order to characterize short arm of *H. villosa* chromosome 6V (6VS) at DNA level, we combined flow cytometric chromosome sorting and next generation DNA sequencing. When compared to whole genome sequencing, this approach provided a massive and lossless reduction of DNA sample complexity and facilitated DNA sequence analysis. A chromosome can be purified by flow sorting if it differs in relative DNA content from other chromosomes in a karyotype, which is not the case of chromosome 6V in *H. villosa*. Thus, we have sorted 6VS chromosome arm from *T. aestivum*-*H. villosa* 6VS ditelosomic addition line, where the telocentric chromosome 6VS is smaller than other chromosomes. With the aim to achieve high resolution of 6VS, we employed bivariate analysis of DNA content (DAPI fluorescence) and the amount of GAA microsatellites labelled by FITC following the FISHS protocol [40]. This approach permitted sorting 6VS arm at almost 90% purity.

The 6VS sequences would accelerate breeding program

H. villosa has been an important donor of disease resistance in wheat breeding, and *Pm21* transferred from *H. villosa* to wheat remains the most effective powdery mildew resistance gene [10]. Although *Pm21* has been cloned, its introduction by genetic transformation may not be acceptable by the market [41]. On the other hand, *Pm21* transferred from wheat-*H. villosa* translocation line T6AL.6VS, has been successfully utilized in wheat breeding, and more than 20 wheat varieties have been released in China [42]. Thus, the introgression of alien chromatin harboring traits of interest by chromosome engineering remains a priority. However, due to linkage drag, this strategy often introduces favorable traits together with deleterious loci, compromising yield and quality [43]. Thus, advanced chromosome engineering is needed to minimize alien chromatin during alien introgression breeding. The main procedures for reducing alien chromatin in wheat is to induce chromatin break-rejoining by ionizing radiation, or induce meiotic recombination between the alien chromatin and its homoeologous common wheat counterpart. In order to preserve beneficial genes and remove deleterious loci, it is important to know the location of beneficial and deleterious genes and define the size of introgressed chromatin.

Development specific molecular markers using chromosome sorting strategy

Development of molecular markers is now much easier than before due to falling costs of next-generation sequencing. As shown in this work, this is true also in species without genome sequence, especially if a chromosome of interest can be purified by flow sorting. The sequences from alien chromosome could then be combined with available wheat genome sequence to develop molecular markers suitable for detecting alien chromatin. Tiwari et al. developed a total of 2,178 unique 5MgS-specific SNPs from *Ae. geniculata* by combining chromosome flow sorting and next generation sequencing and highlighted the power of this approach for mining markers specific for alien chromatin [24]. Zhang et al. developed 1624 intron targeting markers for all *H. villosa* chromosomes, except 4VS and 6VS arms, and out of them 841 (51.79%) markers were specific for tracing *H. villosa* chromatin in wheat background [22]. Wang et al. developed 359 intron targeting markers by combining chromosome sorting and next-generation sequencing, among which 232 (64.62%) markers were suitable for tracing the 4VS chromatin in the wheat background [23]. In this study, based on the DNA sequence of 6VS, a total of 222 IT primer pairs were designed and 120 (54.05%) proved to be specific for *H. villosa* 6VS. Apart from improving the knowledge of genome structure of an important donor of genes for wheat improvement and development of markers to support its use in alien introgression breeding of wheat, the results of this work confirm that chromosome sorting combined with next-generation sequencing is an efficient strategy for IT marker development.

Conclusions

Here, we report a draft DNA sequence of *H. villosa* chromosome arm 6VS and annotation of high-confidence of 3,276 genes. The coding genes showed a fine synteny with Triticeae group 6 chromosomes. A total of 120 IT markers specific to 6VS were developed and used to identify 6VS chromatin in three alien introgression lines. The results and resources developed will support further analysis of molecular organization of 6VS and accelerate its utilization in improvement of bread wheat.

Methods

Plant materials

H. villosa (VV, Accession No. 91C43, $2n = 14$) was obtained from Cambridge Botanical Garden, UK. *T. aestivum* - *H. villosa* ditelosomic addition line Dt6VS [$2n = 42(\text{AABBDD}) + 2t(6\text{VS})$] (Accession No. NAU1202), three *T. aestivum* - *H. villosa* small fragment translocation lines (Accession No. NAU418, NAU419 and NAU1203), and *T. aestivum* - *H. villosa* T6VS·6AL translocation line 92R137 (Accession No. NAU405) were developed at the Cytogenetics Institute, Nanjing Agricultural University (CINAU, hereafter). Common wheat (*T. aestivum*, AABBDD) cv. Chinese Spring maintained at CINAU was used as a control in this work.

Chromosome sorting and DNA sequencing

Suspensions of mitotic metaphase chromosomes were prepared from synchronized meristem root tips of young seedlings according to Vrána et al. and Kubaláková et al. [44, 45]. GAA microsatellite repeats on isolated chromosomes were fluorescently labelled by FISHIS [40] using GAA-fluorescein isothiocyanate (FITC) conjugate (Sigma, Saint Louis, USA) and counterstained by DAPI (4',6-diamidino 2-phenylindole) at 2 µg/ml. The samples were analyzed by FACSaria II SORP flow cytometer and sorter (Becton Dickinson Immunocytometry Systems, San José, USA) at rates of 2000–3000 particles per second and sort windows were set on bivariate flow karyotypes FITC vs. DAPI fluorescence. The identity of sorted particles and contamination of sorted fractions by other chromosomes were determined following Kubaláková et al. [46]. Briefly, one thousand particles were sorted from each sample into a 7 µl drop of P5 buffer on a microscope slide. After air-drying, the slides were used for FISH with probes for pSc119.2 and Afa family repetitive DNA sequences and evaluated by fluorescence microscopy.

Chromosomes were sorted at rates of 15 - 20 / sec into 40 µl sterile deionized water in 0.5 ml PCR tubes and two different 6VS DNA samples were prepared and sequenced. The first was produced by multiple displacement amplification (MDA) of DNA prepared from two batches of 100,000 copies of 6VS telosomes. The amplification was done using Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Piscataway, USA) as described by Šimková et al. [47] and the two MDA products were pooled into one sample to reduce amplification bias. Two micrograms of amplified DNA were used to prepare

sequencing library using TruSeq DNA PCR-Free Library Prep Kit (Illumina, San Diego, USA). The library was sequenced in one run on Illumina MiSeq System (1000 bp insert, 2 x 300 bp) yielding 14 Gb sequence data (~44x coverage of 6VS). The second type of 6VS DNA sample was not amplified and DNA from 100,000 copies of 6VS telosome was purified and directly used to prepare sequencing library using Nextera DNA Library Prep Kit (Illumina). The library was sequenced in one run on Illumina MiSeq System (500 bp insert, 2 x 300 bp) yielding 10.2 Gb sequence data (~32x coverage of 6VS). The sequence reads data were deposited in the (NCBI) Sequence Read Archive (SRA) and is available under accession number PRJNA590539. De novo assembly of the Illumina paired-end reads was performed using the software Hecate (unpublished, <http://bgi-international.com/us/>) using different k-mer sizes (41, 45, 49, and 63). The result of the 45-mer run provided the assembly with the best sequence coverage and N50 size, and therefore was determined to generate 6VS scaffolds.

Identification of DNA repeats

RepeatMasker software (version open-4.0.5, <http://www.repeatmasker.org/>) was used to identify DNA repeat regions and mask repetitive DNA across the 6VS assembly sequence with WU-blast algorithm. Repetitive sequences were searched by aligning 6VS sequence against the known repeat library Repbase Update (<http://www.girinst.org/repbase/>) as well as TREP database (<http://wheat.pw.usda.gov/ITMI/Repeats>), using default settings.

Scanning of coding sequences in the repeat-masked 6VS scaffolds

Ab initio gene prediction was performed by AUGUSTUS program using the repeat-masked sequences. The transcriptome data containing 204,258 unigenes that were obtained from leaves and endosperm of *H. villosa* (unpublished) were used to support the presence of the loci with coding genes. We blasted the predicted gene sequences against the transcriptome data of *H. villosa* with $e\text{-value} \leq 10^{-5}$. Predicted genes with more than 95.0% identity and a minimum alignment of 300 bp on a “unigene” of transcriptome were defined as ‘evidenced gene’. For GO analysis, we used Blast2GO program to get GO annotation and WEGO software for GO functional classification to understand the distribution of gene functions at the macro level.

Transcriptome data

6VS/6AL translocation line grown in a growth chamber with 20°C/16°C (day/night), 16 h/8 h (light/dark). The translocation line was inoculated with *Bgt* isolates E26 and E31 at two-leaf stage, respectively, and inoculated water as control. RNAs were isolated at 1 h, 8 h, 18 h and 24 h after *Bgt* and water inoculation, respectively, followed by freezing in liquid nitrogen for subsequent RNA extraction. The samples were

submitted to the BGI for sequencing using the Illumina Hiseq 4000 platform. After sequencing, Trinity was used to de novo assembly of clean reads. In the transcript three-level classification of Trinity results, taking the longest transcript of gene level as unigenes and then were Blast to NT, NR, COG, KEGG and SwissProt for annotation. Bowtie2 to compare clean reads to unigenes. Then, using the RSEM to calculate the expression level of each sample. *Bgt* isolates E26 and E31 were collected from Institute of Plant Protection, Chinese Academy of Agricultural Sciences.

Development of Intron Target markers

In the first step, we chose a set of genes to calculate exon-exon junction size in genome sequences of homologous arms 6AS, 6BS and 6DS of wheat as well as in 6VS assembly. This gene set contained gene sequence from the annotated *Ae. tauschii* chromosome 6DS and *T. aestivum* chromosome 6DS [13, 48]. We removed repeats in the two gene databases by Blastn analysis using the cut-off parameters e-value > 1e-5, coverage > 80%. In the second step, all obtained genes were compared with genomic sequences of wheat cv. Chinese Spring chromosome arms 6AS, 6BS, 6DS and *H. villosa* 6VS using a local Blastn program. All genes matching 6AS, 6BS, 6DS and 6VS assemblies and possessing at least one predicted exon-exon junction were selected. Intron sizes of corresponding genes were then determined and compared against each other. Genes whose intron sizes in 6VS differed for at least 10% from that of 6AS, 6BS and 6DS simultaneously in the same homologue allele were chosen for designing the primer pairs. The primers were designed in the exons flanking the targeted introns using the online software Primer 3 (Version 4.0, <http://frodo.wi.mit.edu/primer3/>).

Abbreviations

DAPI: 4', 6-diamidino-2-phenylindole; FITC: fluorescein isothiocyanate; MDA: multiple displacement amplification; GO: gene ontology; NR: non-redundant protein sequence database; NT: nucleotide sequence database; COG: clusters of orthologous groups; KEGG: kyoto encyclopedia of genes and genomes; FISH: fluorescence in situ hybridization; TREP: triticeae repeat sequence database; LTR: long terminal repeat; TGW: thousand grain weight; IT: intron targeting; SNP: single nucleotide polymorphism; SINE: short interspersed nuclear elements; TIR: terminal inverted repeat.

Declarations

Authors' Contributions

WXE and JD conceived, designed and coordinated the work; JV and JD flow-sorted telosome 6VS, determined the purity in flow-sorted fractions and amplified chromosomal DNA; KH sequenced amplified chromosomal DNA; XJ, WWT, JD, and WHY wrote the manuscript; WWT, LML, YZY and LJ performed experiments; XJ, WWT, ZX, and WYF analyzed the data. and all authors have read and approved the final manuscript.

Availability of data and materials

The sequence read data of 6VS chromosome was deposited in the (NCBI) Sequence Read Archive (SRA) and is available under accession number PRJNA590539.

Acknowledgements

We also thank Zdeňka Dubská, Marie Kubaláková, Romana Šperková and Jitka Weiserová for the assistance with chromosome sorting and DNA amplification.

Funding

This work was supported by the National Key Research and Development Program (2016YFD0101004, 2016YFD0102001–004), the National Natural Science Foundation of China (Nos. 31571653, 31771782 and 31201204), the International Cooperation and Exchange of the National Natural Science Foundation of China (No. 31661143005), the special fund of Jiangsu Province for the transformation of scientific and technological achievements (BA2017138), the Creation of Major New Agricultural Varieties in Jiangsu Province (PZCZ201706), the SAAS Program for Excellent Research Team, the Science and Technology Service Programs of the Chinese Academy of Sciences (KFJ-STS-ZDTP–002), the Jiangsu Agricultural Technology System (JATS) (No. 2019429), the Key Research and Development Major Project of Ningxia Hui Autonomous Region (No. 2019BBF02022–04), and the Plants as a tool for sustainable global development (No. CZ.02.1.01/0.0/0.0/16_019/0000827).

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.Grądzielewska A: *The genus Dasypyrum—part 2. Dasypyrum villosum—a wild species used in wheat improvement. Euphytica* 2006, 152(3):441–454.
- 2.Mohammad P, Hossain M, Khodaker N, Shiraishi M: *Study for morphological characters of species alien to wheat in Bangladesh. Sarhad Journal of Agriculture (Pakistan)* 1997.
- 3.Okocha P: *Peculiarities of nucleo-cytoplasmic interactions in allocytoplasmic forms of wheat. GLOBAL JOURNAL OF PURE AND APPLIED SCIENCES* 1999, 5:431–436.
- 4.Della Gatta C, Tanzarella O, Resta P, Blanco A: *Protein content in a population of Haynaldia villosa and electrophoretic pattern of the amphiploid T. durum × H. villosa. Breeding Methodologies in Durum Wheat and Triticale Univ Tuscia, Viterbo (Italy)* 1984:39–43.
- 5.Pace Cd, Paolini R, Scarascia M, Qualset C, Delre V: *Evaluation and utilization of Dasypyrum villosum as a genetic resource for wheat improvement. In: Wheat genetic resources: meeting diverse needs: 1990: John Wiley & Sons; 1990: 279–379.*
- 6.De Pace C, Snidaro D, Ciaffi M, Vittori D, Ciofo A, Cenci A, Tanzarella O, Qualset C, Mugnozza GS: *Introgression of Dasypyrum villosum chromatin into common wheat improves grain protein quality. Euphytica* 2001, 117(1):67–75.
- 7.Qualset C, De Pace C, Jan C, Scarascia Mugnozza G, Tanzarella O, Greco B: *Haynaldia villosa (L.) Schur: a species with potential use in wheat breeding. In: Am Soc Agron Abstr: 1981;1981.*
- 8.Zhong GY, Dvořák J: *Evidence for common genetic mechanisms controlling the tolerance of sudden salt stress in the tribe Triticeae. Plant Breeding* 1995, 114(4):297–302.
- 9.Zhao C, Lv X, Li Y, Li F, Geng M, Mi Y, Ni Z, Wang X, Xie C, Sun Q: *Haynaldia villosa NAM-V1 is linked with the powdery mildew resistance gene Pm21 and contributes to increasing grain protein content in wheat. BMC genetics* 2016, 17(1):82.
- 10.Xing L, Hu P, Liu J, Witek K, Zhou S, Xu J, Zhou W, Gao L, Huang Z, Zhang R et al: *Pm21 from Haynaldia villosa Encodes a CC-NBS-LRR that Confers Powdery Mildew Resistance in Wheat. Molecular plant* 2018.
- 11.Bauer E, Schmutzer T, Barilar I, Mascher M, Gundlach H, Martis MM, Twardziok SO, Hackauf B, Gordillo A, Wilde P: *Towards a whole-genome sequence for rye (Secale cereale L.). The Plant Journal* 2017, 89(5):853–869.
- 12.Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y et al: *Draft genome of the wheat A-genome progenitor Triticum urartu. Nature* 2013, 496(7443):87–90.
- 13.Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X et al: *Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. Nature* 2013, 496(7443):91–95.

14. Berkman PJ, Skarshewski A, Lorenc MT, Lai K, Duran C, Ling EY, Stiller J, Smits L, Imelfort M, Manoli S *et al*: *Sequencing and assembly of low copy and genic regions of isolated Triticum aestivum chromosome arm 7DS*. *Plant biotechnology journal* 2011, 9(7):768–775.
15. Vitulo N, Albiero A, Forcato C, Campagna D, Dal Pero F, Bagnaresi P, Colaiacovo M, Faccioli P, Lamontanara A, Simkova H *et al*: *First survey of the wheat chromosome 5A composition through a next generation sequencing approach*. *Plos One* 2011, 6(10):e26421.
16. Berkman PJ, Skarshewski A, Manoli S, Lorenc MT, Stiller J, Smits L, Lai K, Campbell E, Kubalakova M, Simkova H *et al*: *Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation*. *TAG Theoretical and applied genetics Theoretische und angewandte Genetik* 2012, 124(3):423–432.
17. Lucas SJ, Akpinar BA, Simkova H, Kubalakova M, Dolezel J, Budak H: *Next-generation sequencing of flow-sorted wheat chromosome 5D reveals lineage-specific translocations and widespread gene duplications*. *BMC genomics* 2014, 15:1080.
18. Tanaka T, Kobayashi F, Joshi GP, Onuki R, Sakai H, Kanamori H, Wu J, Simkova H, Nasuda S, Endo TR *et al*: *Next-generation survey sequencing and the molecular organization of wheat chromosome 6B*. *DNA research: an international journal for rapid publication of reports on genes and genomes* 2014, 21(2):103–114.
19. Helguera M, Rivarola M, Clavijo B, Martis MM, Vanzetti LS, Gonzalez S, Garbus I, Leroy P, Simkova H, Valarik M *et al*: *New insights into the wheat chromosome 4D structure and virtual gene order, revealed by survey pyrosequencing*. *Plant science: an international journal of experimental plant biology* 2015, 233:200–212.
20. Tiwari VK, Wang S, Danilova T, Koo DH, Vrana J, Kubalakova M, Hribova E, Rawat N, Kalia B, Singh N *et al*: *Exploring the tertiary gene pool of bread wheat: sequence assembly and analysis of chromosome 5M(g) of Aegilops geniculata*. *The Plant journal: for cell and molecular biology* 2015, 84(4):733–746.
21. Xiao J, Dai K, Fu L, Vrana J, Kubalakova M, Wan W, Sun H, Zhao J, Yu C, Wu Y *et al*: *Sequencing flow-sorted short arm of Haynaldia villosa chromosome 4V provides insights into its molecular structure and virtual gene order*. *BMC genomics* 2017, 18(1):791.
22. Zhang X, Wei X, Xiao J, Yuan C, Wu Y, Cao A, Xing L, Chen P, Zhang S, Wang X *et al*: *Whole genome development of intron targeting (IT) markers specific for Dasypyrum villosum chromosomes based on next-generation sequencing technology*. *Molecular Breeding* 2017, 37(9):115.
23. Wang H, Dai K, Xiao J, Yuan C, Zhao R, Dolezel J, Wu Y, Cao A, Chen P, Zhang S *et al*: *Development of intron targeting (IT) markers specific for chromosome arm 4VS of Haynaldia villosa by chromosome sorting and next-generation sequencing*. *BMC genomics* 2017, 18(1):167.

24. Tiwari VK, Wang S, Sehgal S, Vrana J, Friebe B, Kubalaková M, Chhuneja P, Doležel J, Akhunov E, Kalia B *et al*: *SNP Discovery for mapping alien introgressions in wheat. BMC genomics* 2014, **15**:273.
25. Mayer KF, Martis M, Hedley PE, Simková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H *et al*: *Unlocking the barley genome by chromosomal and comparative genomics. The Plant cell* 2011, **23**(4):1249–1263.
26. Martis MM, Zhou R, Haseneyer G, Schmutzer T, Vrana J, Kubalaková M, König S, Kugler KG, Scholz U, Hackauf B *et al*: *Reticulate evolution of the rye genome. Plant Cell* 2013, **25**(10):3685–3698.
27. International Wheat Genome Sequencing C: *A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science* 2014, **345**(6194):1251788.
28. Sanchez-Martin J, Steuernagel B, Ghosh S, Herren G, Hurni S, Adamski N, Vrana J, Kubalaková M, Krattinger SG, Wicker T *et al*: *Rapid gene isolation in barley and wheat by mutant chromosome sequencing. Genome biology* 2016, **17**(1):221.
29. Thind AK, Wicker T, Šimková H, Fossati D, Moullet O, Brabant C, Vrána J, Doležal J, Krattinger SG: *Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly. Nature biotechnology* 2017, **35**(8):793.
30. Bao W, Kojima KK, Kohany O: *Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA* 2015, **6**:11.
31. Wicker T, Matthews DE, Keller B: *TREP: a database for Triticeae repetitive elements*. In.: Elsevier Current Trends; 2002.
32. Xing L, Hu P, Liu J, Witek K, Zhou S, Xu J, Zhou W, Gao L, Huang Z, Zhang R: *Pm21 from Haynaldia villosa Encodes a CC-NBS-LRR Protein Conferring Powdery Mildew Resistance in Wheat. Molecular plant* 2018.
33. Cao A, Xing L, Wang X, Yang X, Wang W, Sun Y, Qian C, Ni J, Chen Y, Liu D: *Serine/threonine kinase gene Stpk-V, a key member of powdery mildew resistance gene Pm21, confers powdery mildew resistance in wheat. Proceedings of the National Academy of Sciences* 2011, **108**(19):7727–7732.
34. International Wheat Genome Sequencing C, investigators IRp, Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, investigators lw-gap, Pozniak CJ *et al*: *Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science* 2018, **361**(6403).
35. Eddy SR: *Accelerated Profile HMM Searches. PLoS computational biology* 2011, **7**(10):e1002195.
36. Li G, Chen P, Zhang S, Wang X, He Z, Zhang Y, Zhao H, Huang H, Zhou X: *Effects of the 6VS·6AL translocation on agronomic traits and dough properties of wheat. Euphytica* 2007, **155**(3):305–313.

37. Jaiswal V, Gahlaut V, Mathur S, Agarwal P, Khandelwal MK, Khurana JP, Tyagi AK, Balyan HS, Gupta PK: *Identification of Novel SNP in Promoter Sequence of TaGW2-6A Associated with Grain Weight and Other Agronomic Traits in Wheat (Triticum aestivum L.). PloS one* 2015, *10*(6):e0129400.
38. Yang Z, Bai Z, Li X, Wang P, Wu Q, Yang L, Li L, Li X: *SNP identification and allelic-specific PCR markers development for TaGW2, a gene linked to wheat kernel weight. TAG Theoretical and applied genetics Theoretische und angewandte Genetik* 2012, *125*(5):1057–1068.
39. Su Z, Hao C, Wang L, Dong Y, Zhang X: *Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (Triticum aestivum L.). TAG Theoretical and applied genetics Theoretische und angewandte Genetik* 2011, *122*(1):211–223.
40. Giorgi D, Farina A, Grosso V, Gennaro A, Ceoloni C, Lucretti S: *FISHIS: Fluorescence In Situ Hybridization in Suspension and Chromosome Flow Sorting Made Easy. PloS one* 2013, *8*(2):e57994.
41. Khan SJ, Muafia S, Nasreen Z, Salariya AM: *GENETICALLY MODIFIED ORGANISMS (GMOs): FOOD SECURITY OR THREAT TO FOOD SAFETY. Pakistan Journal of Science* 2012.
42. Li H, Chen X, Xin ZY, Ma YZ, Xu HJ, Chen XY, Jia X: *Development and identification of wheat-Haynaldia villosa T6DL·6VS chromosome translocation lines conferring resistance to powdery mildew. Plant Breeding* 2005, *124*(2):203–205.
43. Abrouk M, Balcarkova B, Simkova H, Kominkova E, Martis MM, Jakobson I, Timofejeva L, Rey E, Vrana J, Kilian A et al: *The in silico identification and characterization of a bread wheat/Triticum militinae introgression line. Plant biotechnology journal* 2017, *15*(2):249–256.
44. Kubalakova M, Vrana J, Cihalikova J, Simkova H, Dolezel J: *Flow karyotyping and chromosome sorting in bread wheat (Triticum aestivum L.). TAG Theoretical and applied genetics Theoretische und angewandte Genetik* 2002, *104*(8):1362–1372.
45. Vrana J, Kubalakova M, Simkova H, Cihalikova J, Lysak MA, Dolezel J: *Flow sorting of mitotic chromosomes in common wheat (Triticum aestivum L.). Genetics* 2000, *156*(4):2033–2041.
46. Kubalakova M, Valarik M, Barto J, Vrana J, Cihalikova J, Molnar-Lang M, Dolezel J: *Analysis and sorting of rye (Secale cereale L.) chromosomes using flow cytometry. Genome* 2003, *46*(5):893–905.
47. Simkova H, Svensson JT, Condamine P, Hribova E, Suchankova P, Bhat PR, Bartos J, Safar J, Close TJ, Dolezel J: *Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. BMC genomics* 2008, *9*:294.
48. Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E et al: *Structural and functional partitioning of bread wheat chromosome 3B. Science* 2014, *345*(6194).

Supplementary Files Legend

Additional file 1: Fig. S1 (A) Sequence length distribution of 3,276 genes of short arm of *H. villosa* 6V chromosome; (B) Percent distribution of the GO entries for *H. villosa* 6VS genes. The most represented entries within the three ontologies (Molecular function, Biological process and Cellular component) are indicated.

Additional file 2: Table S1 Markers specific for short arm of *H. villosa* 6V chromosome

Tables

Table 1 The statistics of assembly of flow-sorted short arm of *H. villosa* 6V chromosome

Total bases (Gbp)	47.7
Number of assembly scaffolds	153,177
Total assembly bases (bp)	230,388,792
Max. length of assembly scaffolds (bp)	138,620
Min. length of assembly scaffolds (bp)	100
N50 (bp)	9,788
Mean length (bp)	1,464
GC-content (%)	45.68

Table 2 Identification of repetitive DNA elements in short arm of *H. villosa* 6V chromosome

Type	Sub-type	Total length (bp)	% genome
DNA transposon	TIR	11,269,751	6.53
	Helitron	189,843	0.11
retrotransposon	LTR_Copia	18,656,357	10.81
	LTR_Gypsy	89,588,481	51.91
	LTR_Unknown	3,831,370	2.22
	SINE	1,691,326	0.98
	Unknown	4,694,291	2.72
tandem repeat		535,011	0.31
unknown		4,694,291	2.72

Figures

Figure 1

Flow sorting of *H. villosa* chromosome arm 6VS Bivariate flow karyotype obtained after the analysis chromosomes isolated from bread wheat-Haynaldia villosa 6VS ditelosomic addition line. Prior to analysis, GAA microsatellite repeats on chromosomes were labelled by fluorescein isothiocyanate (FITC) and counterstained by 4',6-diamidino 2-phenylindole (DAPI). The population representing 6VS telosome is marked by red rectangle. The inset shows examples of flow-sorted 6VS telosomes after fluorescence in situ hybridization (FISH) with probes for pSc119.2 (yellow-green) and Afa family repeats (red).

Figure 2

Comparative analysis of 6VS DNA sequence The distinguish of 6VS homologous genes in group 6 of Triticeae genome. The blue lines represent the homologous genes of 6VS located on the chromosomes of Triticeae genome. Ta, Td, At, Tu and Hv represent *T. aestivum*, *T. dicoccoides*, *Ae. tauschii*, *T. urartu* and *H. vulgare*, respectively. The blue arrows represent the position of centromere of common wheat.

Figure 3

Gene expression patterns in response to powdery mildew infection and gene structures of the resistance gene analogs (RGAs) from 6VS of *H. villosa*. Left: The heatmap for the expression of predicted RGAs that treated with two Bgt isolates, E26 and E31, respectively at 3 and 24 hours after inoculation. Right: The structure of resistance gene analogs which predicted from 6VS of *H. villosa*. Different colors represent different domains. The scale represents the length of amino acids of genes.

HvGW2-6V-promotor	AAG ⁻⁹⁸⁸ GCG.....CCCGTGCAG ⁻⁷³⁹ TAGCT.....TCG ⁻⁵⁹³ GAC.....CC ⁻⁴⁹⁴ GCGG High TGW
TaGW2-6A-promotor Hap_1	AAG ⁻⁹⁸⁸ GCG.....CCCGTACAG ⁻⁷³⁹ CAGCT.....TCG ⁻⁵⁹³ GAC.....CC ⁻⁴⁹⁴ GCGG Low TGW
TaGW2-6A-promotor Hap_2	AAG ⁻⁹⁸⁸ GCG.....CCCGTACAG ⁻⁷³⁹ CAGCT.....TCG ⁻⁵⁹³ AAC.....CC ⁻⁴⁹⁴ GCGG Low TGW
TaGW2-6A-promotor Hap_3	AAG ⁻⁹⁸⁸ GCG.....CCCGTACA ⁻⁷³⁹ ACAGCT.....TCG ⁻⁵⁹³ GAC.....CC ⁻⁴⁹⁴ GCGG Low TGW
TaGW2-6A-promotor Hap_4	AAG ⁻⁹⁸⁸ ACG.....CCCGTACAG ⁻⁷³⁹ CAGCT.....TCG ⁻⁵⁹³ AAC.....CC ⁻⁴⁹⁴ GCGG Low TGW
TaGW2-6A-promotor Hap_5	AAG ⁻⁹⁸⁸ GCG.....CCCGTACA ⁻⁷³⁹ ACAGCT.....TCG ⁻⁵⁹³ GAC.....CC ⁻⁴⁹⁴ ACGG High TGW

Figure 4

The SNPs at the promotor region of TaGW2-6A and HvGW2-6V Characters that highlighted with red color were the SNPs that reported association with thousands of grain weight (TGW) within the promotor region of TaGW2-6A and HvGW2-6V. 'CGCG' motif was underlined.

Figure 5

Development of IT markers and construction of cytogenetic physical map of *H. villosa* 6VS chromosome arm Left: Molecular markers in each bin Right: The amplification results of representative marker in each bin which was underlined in the left 1: *T. aestivum* (AABBDD); 2: *H. villosa* (VV); 3: *T. aestivum*-*H. villosa* translocation line (T6VS·6AL); 4-6: Three *T. aestivum*-*H. villosa* small fragment translocation lines (NAU418, NAU419 and NAU1203); 7: DNA Ladder

Supplementary Files

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

- [Additionalfile1FigureS1.jpg](#)
- [Additionalfile2TableS1.docx](#)