Transcriptomics, chromosome engineering and mapping identify a restorer-of-fertility region in the CMS wheat system msH1

Cristina Rodríguez-Suárez [1*], Paolo Bagnaresi [2], Luigi Cattivelli [2], Fernando Pistón [1], Almudena Castillo [1], Azahara C Martín [3], Sergio G Atienza [1], Carmen Ramírez [1], Antonio Martín [1].

[1] Instituto de Agricultura Sostenible, Consejo Superior de Investigaciones Científicas (IAS-CSIC), Avda. Menéndez Pidal s/n, 14080 Córdoba, Spain

[2] Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria. Research Centre for Genomics and Bioinformatics, via S. Protaso 302 I -29017 Fiorenzuola d'Arda PC, Italy

[3] Crop Genetics, John Innes Centre, Norwich Research Park NR4 7UH Norwich, UK

*corresponding author:
Cristina Rodríguez-Suárez https://orcid.org/0000-0002-5409-1839
+34 957499241 crodriguez@ias.csic.es

Paolo Bagnaresi https://orcid.org/0000-0001-8466-4805
Luigi Cattivelli https://orcid.org/0000-0002-6067-4600
Fernando Pistón https://orcid.org/0000-0002-5496-0205
Almudena Castillo
Azahara C Martín https://orcid.org/0000-0002-6087-7491
Sergio G Atienza https://orcid.org/0000-0001-5175-8303
Carmen Ramírez
Antonio Martín https://orcid.org/0000-0002-9755-9278

Acknowledgements

Research funded by grants AGL2013-43329-R and SOLAUT_00025062, from Ministerio de Economía y Competitividad, Spain (MINECO).

Conflict of interest statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Abstract

Cytoplasmic male sterility (CMS) is a valuable trait for hybrid seed production. The msH1 CMS system in common wheat results from the incompatibility between the nuclear genome of wheat and the cytoplasm of the wild barley *Hordeum chilense*. This work aims to identify *H. chilense* candidate genes for fertility restoration in the msH1 system with a multidisciplinary strategy based on chromosome engineering, differential expression analysis and genome mapping. Alloplasmic isogenic wheat lines differing for fertility, associated to the presence of an acrocentric chromosome H^{ch}ac resulting from the rearrangement of the short arms of *H. chilense* chromosomes $1H^{ch}$ and $6H^{ch}$, were used for transcriptome sequencing. Two novel RNA-seq mapping approaches were designed and compared to identify DEGs (Differentially Expressed Genes) of *H. chilense* associated to male fertility restoration. Minichromosomes (H^{ch}mi), new smaller reorganizations of the H^{ch}ac also restoring fertility, were obtained and used to validate the candidate genes. This strategy was successful identifying a putative restorer-of-fertility region on $6H^{ch}$S, with six candidate genes, including the ortholog of the barley restorer gene *Rfm1*. Additionally, transcriptomics gave preliminary insights on sterility and restoration networks showing the importance of energy supply, stress, protein metabolism and RNA processing.

Keywords

Cytoplasmic male sterility, common wheat, *Hordeum chilense*, minichromosomes, restorer-of-fertility, RNA-seq

Author contribution statement

AC, AM, CR-S and FP conceived and planned the experiments. AC and CR collected the plant material. PB performed the bioinformatic analysis. CR-S and PB analyzed RNA-seq results. CR-S performed primer design, mapping and results interpretation. ACM, AM, CR and CR-S participated in minichromosomes obtaining. ACR performed GISH. CR-S wrote the manuscript in consultation with ACM, LC, PB and SGA. All authors gave critical feedback.

Key message

An original RNA-seq mapping strategy, validated with chromosome engineering and physical mapping, identifies candidate genes for fertility restoration in the $6H^{ch}S$ chromosome of *H. chilense* in the wheat msH1 system.

Introduction

Cytoplasmic male sterility (CMS) is a natural phenomenon in plants, a maternally inherited trait that leads to the production of non-functional pollen due to the disruption of the cytonuclear communication. CMS is caused by mitochondrial genes whereas nuclear encoded genes, named restorer-of-fertility (*Rf*), can suppress male sterility and restore fertility (Hanson and Bentolila 2004). In most cases, *Rf* genes encode proteins that act directly on the mitochondrial encoded sterilizing transcripts by specifically binding and processing them (Chen and Liu 2014; Kim and Zhang 2018). The majority of known *Rf* belong to the pentatricopeptide repeat (PPR) protein family, although other families as glycine-rich proteins, aldehyde dehydrogenase, acyl-carrier proteins and a peptidase have been identified as *Rf* (Dahan and Mireau 2013; Gaborieau et al. 2016).

CMS represents a cost-effective system for hybrid seed production in self-pollinated crop species to exploit heterosis, and therefore, it has been successfully used in many cropping systems including rice, maize, sunflower and rye among others (Bohra et al. 2016). In experimental wheat hybrids, heterotic grain yields of more than 10% have been described (Longin et al. 2013) along with enhanced yield stability (Mühleisen et al. 2014). However, no stable system for hybrid wheat production has been successfully obtained (Whitford et al. 2013). A new CMS source in bread wheat was described by (Martín et al. 2008) and named msH1. This CMS system uses the cytoplasm of the wild barley *Hordeum chilense Roem. et* Schultz. accession H1 as a source of male sterility. Restoration of fertility was first observed associated with the addition of the short arm of chromosome $6H^{ch}$ from H1 (Martín et al. 2008). Other wheat lines harboring new reorganizations of $6H^{ch}S$ chromosome were obtained, as the double translocation $T6H^{ch}S{\cdot}6DL$, and its fertility restoration ability was also confirmed (Martín et al. 2009). Furthermore, a new acrocentric recombined chromosome ($H^{ch}ac$) including fragments from $1H^{ch}S$ and $6H^{ch}S$ was identified as a new source of fertility restoration in the msH1 system (Martín et al. 2010; Castillo et al. 2014). Recently, the translocation $T6H^{ch}S{\cdot}6DL$ has been transferred to durum wheat confirming the restoration ability also in alloplasmic durum wheat lines (Martín et al. 2018). Although there exists evidence of a putative enhancing effect of chromosome $1H^{ch}S$ in the restoration process (Castillo et al. 2015), the importance of the *Rf* on chromosome $6H^{ch}S$ has been recurrently observed.

To enable the utilization of msH1 system in hybrid wheat production, a deeper knowledge of the restoration process, including the identification of the genes involved, is needed. For that purpose, isogenic alloplasmic wheat lines with the addition of the restorer-of-fertility H^{ch}ac chromosome (fertile phenotype) and without the H^{ch}ac chromosome (sterile phenotype) were selected as an optimum material, as they only differ for the alien chromosome and share the same common wheat background and *H. chilense* cytoplasm. Differentially expressed genes (DEGs) analysis between these two wheat lines is a convenient starting point to identify *H. chilense* genes whose expression is associated to the fertility-restored phenotype, as well as to find markers to delimit the genomic region carried by the H^{ch}ac chromosome. We therefore performed the transcriptome sequencing of alloplasmic lines contrasting for fertility restoration in the search for potential candidate genes for *Rf* in msH1.

Material and Methods

Plant material

The plant material used for this study is shown in Table 1. Alloplasmic lines T528 and T749, harboring an acrocentric chromosome restoring fertility (H^{ch}ac), were developed and characterized in previous works (Martín et al. 2010; Castillo et al. 2014). Alloplasmic wheat lines T527 and T854 are described in this work.

In the self-progeny of the fertile alloplasmic line T749, a plant harboring a single H^{ch}ac chromosome (T749-84) was selected and selfed. Its progeny was cytologically screened by somatic chromosome counting and divided in two groups: plants with 42 chromosomes with the addition of the H^{ch}ac (fertile), and plants with 42 chromosomes without H^{ch}ac (sterile). Plants were grown in a climate chamber (light period: 22 ºC during 11 h + 16 ºC during 1 h; dark period: 10 ºC during 11 h + 16 ºC during 1 h). Anthers were cytologically examined for the stage of development and collected at the late uninucleate stage, as this is the moment when microspores collapse in the alloplasmic sterile genotypes or continue a normal development in the restored and euplasmic fertile lines (Martín et al. 2008; Martín et al. 2010). A single anther per floret was squashed in acetocarmine and mounted for microscopy. The remaining two anthers from the same floret were isolated, frozen in liquid nitrogen and stored at −80 °C.

RNA-seq library preparation and sequencing

Twenty developmentally equivalent anthers from at least five different plants were pooled for each of the two biological replicates of the two contrasting genotypes. Total RNA was extracted using TRIzol reagent (Life Technologies, Grand Island, NY). RNA quality was checked by electrophoresis in a 1.5% agarose gel and the concentration of total RNA was determined by NanoDrop (Thermo Scientific, Wilmington, DE, USA). The four cDNA libraries (Fert_3, Fert_5, Ster_4 and Ster_9) were prepared and sequenced by Illumina paired-end technology on the Illumina HiSeq 2000 instrument by BGI–Hong Kong Co., Ltd (Hong Kong, China).

Sequence data analysis and mapping strategies

Raw fastQ files (paired-end reads; 90 nt) were checked by means of FastQC application (version 11.1, downloaded from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to track contaminants and low-quality reads. Such identified poor quality reads/regions were filtered out by Cutadapt application (Martin 2011). Two different read mapping strategies were used, namely BarWheat and OnlyBarley. Regarding BarWheat approach, filtered reads for each biological replicate were mapped with Bowtie2 (Langmead and Salzberg 2012) and Tophat2 (Kim et al. 2013) to a virtual BarWheat genome consisting of the merged genome multifasta files for barley (*Hordeum vulgare* L.) and common wheat (*Triticum aestivum* L.) obtained from Ensembl Plants repository (http://plants.ensembl.org/info/website/ftp/index.html; releases Hv_IBSC_PGSB_v2 and Triticum_aestivum.IWGSC1.0+popseq.30, respectively). Care was taken to modify, as necessary, chromosome names in merged multifasta genome files by adding genome-specific prefixes to chromosome names to avoid ambiguities. Read counts were collected from the lexicographically sorted BAM alignment files with HTSeq version 0.6.1p1 in the ‘union’ mode using the custom BarWheat GTF file. This file was created by combining barley GTF files (Hv_IBSC_PGSB_v2; both low confidence and high confidence genes) and wheat GTF file as obtained from the Ensembl Plants repository. HTSeq was set to discard ambiguous (multimapper) reads from counts. If necessary, chromosome names in BarWheat GTF file were modified to match modified chromosome names in BarWheat genome file. For OnlyBarley approach, reads were mapped to barley genome (Hv_IBSC_PGSB_v2) and read counts were collected from BAM alignment files with HTSeq version 0.6.1p1 in the ‘union’ mode using the default

GTF file (Hv_IBSC_PGSB_v2) as available in Ensembl Plants repository and set in order to discard ambiguous (multimapper) reads as in the BarWheat approach.

DEG Calling

Differentially expressed genes (DEGs) were called implementing the Bioconductor DESeq2 package version 1.8 (Love et al. 2014) with local fit, betaPrior parameter set to False and enabling independent filtering. Thresholds for FDR (Benjamini-Hochberg false discovery rate) and fold change (FC) were set to 0.05 and 2, respectively.

Mapman

MapMan tool (Thimm et al. 2004; Lohse et al. 2014) images were generated by importing DESeq2-normalized expression data in MapMan application. BarWheat DEG (MM2) sequences were binned to MapMan "BINs" (i.e. Functional classes of genes) by means of the Mercator application (Lohse et al. 2014) with default parameters. In order to provide a snapshot of a larger set of genes, all called DEGs (no FC threshold applied in this case) were visualized.

IGV visualizations

BarWheat genome visualization with Integrative Genomics Viewer (IGV) was employed for help in primer design. As the large size of some chromosomes in BarWheat genome (and also some large barley chromosome) does not allow proper visualization in IGV in the BAM default format, BAM alignment files as resulting from tophat2/bowtie2 mapping were converted to the SAM decompressed format prior to loading on IGV.

Primer design and PCR

Primer pairs were designed with Primer-BLAST software tool (Ye et al. 2012) using *H. chilense* DEGs consensus sequences from *H. chilense* or from *H. vulgare* as templates, and the Triticeae sequence database to examine potentially undesired targets. Leaf tissue was harvested from plantlets for DNA isolation. Genomic DNA was extracted following the CTAB protocol with slight modifications (Murray and Thompson 1980). All PCR amplifications were carried out using MyTaq^{TM} DNA polymerase (Bioline, London, UK) following manufacturer's instructions. Amplification products were resolved in agarose gels and visualized with Safeview^{TM} Nucleid Acid Stain (NBS Biologicals Ltd, Cambridgeshire, England).

Primer pair amplification specificity was checked with parent lines H1, T21 and T26. The *H. chilense* addition lines T21A1H_1S and T21A6H_1S were used for DEGs physical mapping in *H. chilense*. Polymorphic pairs were used to characterize lines T749, T528, T854 and T527.

Genome in situ hybridization

Chromosome preparation and genome in situ hybridization (GISH) were carried out as described previously (Rey et al. 2018). *H. chilense* genomic DNA was directly labelled with tetramethyl-rhodamine-5-dUTP (Sigma) by nick translation. Images were taken using a Leica DM5500B microscope equipped with a Hamamatsu ORCA-FLASH4.0 camera and controlled by Leica LAS X software v2.0.

Results

Line T749 is an alloplasmic common wheat line with restored fertility due to the presence of the H^{ch}ac from *H. chilense*. The segregating self-progeny of one of these plants was classified into fertile (with H^{ch}ac) and sterile (without H^{ch}ac) to build the contrasting libraries for RNA-seq. These isogenic lines shared the same wheat background and the sterilizing *H. chilense* H1 cytoplasm, and differed only for the presence of the restorer H^{ch}ac chromosome. Transcriptomic analysis was carried out in anthers at the late uninucleate stage when the *Rf* gene was expected to be active.

Once samples were collected, plants were grown to maturity. Fertility was evaluated in absolute terms, confirming that all plants with the H^{ch}ac chromosome were fertile and all plants without H^{ch}ac chromosome were completely sterile.

Mapping strategies and DEG calling

Two different mapping strategies were assayed and evaluated: (1) 'OnlyBarley': using *H. vulgare* as reference genome and (2) 'BarWheat': using a virtual genome constructed by merging barley and wheat genomes. In both strategies, different tolerance levels were assayed by changing the number of mismatches allowed between reads and reference genomes. The number of reads obtained in each library (around 20 million), the number of mapped reads, indicating the number of reads with multiple alignments, for each strategy and with each tolerance level is shown in Supplemental Table 1 (EMS_1).

(1) OnlyBarley (OB) mapping was assessed using *H. vulgare* as reference genome. Maximum allowed mismatches between the 90-nt long reads and the reference genome were set from zero to three (MM0-MM3). Table 2 shows the number of DEGs identified by this strategy considering these three conditions. In the most stringent one (MM0), 11 DEGs were identified. This number increased, when mismatches were allowed, to 18 (in MM1), 35 (in MM2) and 42 (in MM3). The complete list and descriptions of DEGs identified by this strategy is shown in Supplemental Tables 2.1 to 2.4 (EMS_2).

(2) BarWheat (BW) mapping, the novelty of this approach is that a virtual genome constructed by merging barley and wheat genomes was used as reference genome for mapping. Allowed mismatches between reads and the virtual genome were set in a range from 0-2 (MM0-MM2). Regarding barley genes, in the MM0 condition 35 DEGs were identified, 83 in MM1 and 136 in MM2. The number of DEGs identified per chromosome is shown in Table 2. The complete list and descriptions of DEGs identified by this strategy are shown in Supplemental Tables 3.1 to 3.3 (EMS_3).

BW mapping strategy in the MM2 condition was selected as the best option and used for downstream analyses (see discussion section). In addition to the 136 DEGs from barley, 106 genes from wheat were identified as differentially expressed (Supplemental Table 3.3). All 106 DEGs assigned to chromosome 1H are located consecutively in a region from 261,782 to 235,074,660 Mb of the *H. vulgare* genome. Regarding DEGs in 6H, 19 genes were differentially expressed being located in the distal part of the short arm, in a region between 1,578,516 Mb (delimited by HORVU6Hr1G000640) and 19,420,779 Mb (HORVU6Hr1G011020). Gene HORVU6Hr1G083600 is located out of this region, in the long arm of chromosome 6H of *H. vulgare*, but may be probably wrongly assigned, being instead the paralog HORVU6Hr1G006840 within the above indicated region (14,301,793-14,307,089 Mb). A schematic representation of DEGs location in 1H and 6H chromosomes is shown in Figure 1. Genes on chromosome 2H (HORVU2Hr1G097220 and HORVU2Hr1G110120), 3H (HORVU3Hr1G084830), 5H (HORVU5Hr1G037580 and HORVU5Hr1G051310) and 7H (HORVU7Hr1G056770 and HORVU7Hr1G117000) may be wheat transcripts wrongly mapped to barley. This is consistent with the fact that five of these seven genes show negative $\log_2$fold values (Supplemental Table 3.3). Also, four DEGs were identified with an unknown position in

barley genome: HORVU0Hr1G000210, HORVU0Hr1G000390, HORVU0Hr1G000410 and HORVU0Hr1G014960 (Table 3).

Male-sterility and restoration networks

The MapMan analysis was performed to define the functional annotation and categorization of the DEGs identified in the MM2 condition in BW mapping. All DEGs, with no fold change filtering, were considered (Supplemental Table 3.4 in EMS_3). In total, 307 DEGs were mapped to 315 bins as some genes were assigned to more than one category or to more than one subcategory within the same category. Considering each gene only once in each functional group, the number of genes assigned to the functional category *protein* was highest (47), followed by *RNA* (27), *miscellanea* (20), *stress* (18), *signaling* (16), *cell* (11), *transport* (10), *TCA* and *DNA* (9), *development* and *secondary metabolisms* (8), *lipid metabolism* (6), *nucleotide metabolism* (5) and *cell wall*, *hormone metabolism* and *amino acid metabolisms* (4), being the rest of the classes with less than 4 genes (Supplemental Table 4 in ESM_4). The metabolism overview displayed by Mapman is shown in Figure 2.

Validation of DEGs

Validation of BW mapping strategy

The subset of DEGs assigned to 6H chromosome in *H. vulgare* and those of unknown position was selected for validation (Table 3). To validate the mapping strategy and DEG calling, the presence of the genes identified by using *H. vulgare* genome had to be confirmed in the H^{ch}ac of *H. chilense*. For that purpose, consensus sequences of *H. chilense* DEGs were retrieved from IGV and used as template for specific primer design. For genes HORVU6Hr1G000640, HORVU6Hr1G001860, HORVU6Hr1G083600, HORVU0Hr1G000210 and HORVU0Hr1G000410 it was not possible to design successful primers after several attempts. For the remaining 18 DEGs, specific amplification of *H. chilense* genes in wheat genomic background was obtained (primer pairs are listed in Supplemental Table S5, EMS_1). An example of the amplification pattern of some of these primers is shown in Figure 3. Primer pair designed for HORVU6Hr1G003890 amplification was valid in T26 wheat background but not in T21, and the opposite happened with primer pair designed for HORVU6Hr1G003900, which was only useful in T21 wheat background.

DEGs location was first confirmed on chromosome 6H^{ch}S by using the *H. chilense*-wheat addition lines. Two DEGs of unknown position, HORVU0Hr1G014960 and HORVU0Hr1G000390, were also physically mapped to 6H^{ch} and to 1H^{ch}, respectively, by this means. Then, the presence of DEGs was tested in the wheat lines T749 and T528 harboring the H^{ch}ac. All primer pairs showed a positive amplification in both lines (Table 4), confirming that DEGs were actually located in the H^{ch}ac chromosome. Therefore, BW mapping strategy was successful in the identification of *H. chilense* genes.

Defining a Rf region using minichromosomes, new H^{ch}ac reorganizations

The H^{ch}ac chromosome in line T528 shows a stable transmission; however, spontaneous reorganizations of the H^{ch}ac were observed in the progeny of some of the lines T528 carrying a single H^{ch}ac chromosome copy. These reorganized chromosomes were smaller than the original acrocentric and thus, were referred to as minichromosomes (H^{ch}mi) when first observed. Based on their size under microscope observation, at least two different H^{ch}mi were obtained. Molecular markers, previously described for characterization of H^{ch}ac (Castillo et al. 2014), were used to confirm that both H^{ch}mi were shorter than the H^{ch}ac and different between them (Supplemental Table S6, EMS_1). Genomic in situ hybridization (GISH) was carried out to confirm their *H. chilense* origin (Figure 4). Lines harboring these new reorganizations were named T527 (H^{ch}mi1) and T854 (H^{ch}mi2).

The resulting minichromosomes are an excellent tool to examine their fertility restoration ability and their association with the presence / absence of DEGs. For that purpose, self-progenies from T854 and T527 lines were obtained. Segregation for the presence/absence of the H^{ch}mi was observed by chromosome counting in both progenies. Fertility restoration ability of both minichromosomes was evaluated in absolute terms: all plants harboring H^{ch}mi1 or H^{ch}mi2 were fertile, whereas all plants without minichromosomes were sterile. Plants from both progenies were characterized for the presence of the validated DEGs. Plants without H^{ch}mi1 and H^{ch}mi2 did not show positive amplification for any DEGs as expected. All plants harboring H^{ch}mi2 showed the same amplification pattern for DEGs on 6H^{ch}S than line T528, but were negative for HORVU0Hr1G000390 mapped to 1H^{ch}. Conversely, plants harboring H^{ch}mi1 were positive for HORVU0Hr1G000390 and showed positive amplification for only five of the sixteen DEGs located in 6H^{ch}S: HORVU6Hr1G003170, HORVU6Hr1G003470,

HORVU6Hr1G003770, HORVU6Hr1G003890 and HORVU6Hr1G004350 (Table 4, Figure 3).

These genes are common to all three H^{ch}ac, H^{ch}mi1 and H^{ch}mi2 fertility restorer chromosomic reorganizations and therefore, they are delimiting a target chromosome region for fertility restoration. As most of the *Rf* identified in other CMS systems belong to the PPR class (Gaborieau et al. 2016), a search of PPR-domain containing genes in the vicinity of this region in barley was carried out, which allowed the identification of candidate genes HORVU6Hr1G002890, HORVU6Hr1G004120, HORVU6Hr1G005240 and HORVU6Hr1G05380. Specific primer pairs for the four genes were designed (Supplemental Table S5, EMS_1), validated in H^{ch}ac lines and tested in H^{ch}mi1 and H^{ch}mi2 (Table 4). From the four PPR genes, only HORVU6Hr1G004120 was detected in the acrocentric and in the two minichromosomes (Table 4, Figure 3).

Discussion

RNA-seq bioinformatic data processing was a technical challenge due to several limitations. First, *H. chilense* genome is not sequenced and it could not be used as reference for mapping. Instead, *H. vulgare* genome was the closest available. Additionally, the H^{ch}ac is in common wheat genomic background, and then, transcripts of A, B and D genomes coexist with those from *H. chilense*. The challenge therefore was the correct identification of *H. chilense* transcripts avoiding the interference with wheat transcripts, while overcoming the expected polymorphisms between *H. vulgare* and *H. chilense* genomes. For this purpose, the two mapping strategies OB and BW were assayed. The results obtained can be compared in terms of number of DEGs identified and chromosome assignment (Table 2). In OB mapping, DEGs are identified in all chromosomes and they increase in number as long as more mismatches are allowed. For some of these genes, negative $\log_2$fold change values were observed (up-regulated in sterile plants compared to the fertile ones). This expression profile may not correspond to *H. chilense* genes, as sterile plants do not harbor the H^{ch}ac chromosome, and may rather be explained by reads from wheat transcripts artifactually mapped to barley genes. Also, DEGs are identified in all seven barley chromosomes and not only on chromosomes 1H and 6H as expected due to the H^{ch}ac chromosome origin (Martín et al. 2010; Castillo et al. 2014). To overcome these deficiencies, a second mapping

strategy was assayed. In BW mapping, a virtual barley and wheat merged genome was created and used as reference. In this original approach, the risk of wrong alignment is reduced as wheat reads preferentially map to wheat genome and they do not interfere to barley mapping. Consistently, DEGs are mainly localized on chromosomes 1H and 6H, which show a good degree of collinearity with chromosomes $1H^{ch}$ and $6H^{ch}$ from *H. chilense* (Avila et al. 2019), and when more mismatches are allowed, there are almost no DEGs detectable on other chromosomes. The location of DEGs (Figure 1) is coincident with the expected origin of the acrocentric chromosome, which was characterized with molecular markers in a previous work as a reorganization of $1H^{ch}S$ and $6H^{ch}S$ chromosomes (Castillo et al. 2014). Also, $\log_2$fold values were higher in general terms with BW. Therefore, BW mapping strategy in the MM2 condition was finally used for downstream analyses.

The functional categorization performed by MapMan, provides an overview of the metabolic pathways involved in the sterility and restored fertility conditions. There exists evidence of an activation of energy metabolism in restored plants via: (1) up-regulation of glucose-6-phosphate isomerase (HORVU1Hr1g006860) and a hexokinase (HORVU1Hr1g028200) in glycolysis for pyruvate production; (2) pyruvate transport from cytoplasm to mitochondria by up-regulation of mitochondria pyruvate carrier (Traes_1AS_B1DE9E844, Traes_1BS_AC2BB761D and Traes_7AS_887C5E8C5) (3) conversion of pyruvate to mitochondrial acetyl-coA to enter the tricarboxylic acid cycle (TCA), by the up-regulation of the pyruvate dehydrogenase complex (*PDC*) subunits (HORVU1Hr1g023060, HORVU1Hr1g023630 and HORVU6Hr1g003770); (4) activation of the TCA by up-regulating the conversion of isocitrate to α-oxoglutarate (Traes_3AL_3d5c860fd and Traes_3B_63651ecff) and the conversion of malate to pyruvate (HORVU1Hr1g028030). Conversion of aspartate to oxaloacetate and alanine to 2-oxoglutarate are activated by up-regulation of aspartate and alanine aminotransferases, respectively (HORVU6Hr1g003470 and HORVU1Hr1g028030) providing substrates to TCA. Increases in alanine aminotransferase, aspartate aminotransferase and glutamate dehydrogenase have been associated with alternative route to 2-oxoglutarate under conditions in which flux through the usual TCA cycle pathway is inhibited (Sweetlove et al. 2010). These two enzymes are connected to amino acid metabolism, which is up-regulated in fertile plants, and has been demonstrated to be involved in other CMS systems (Fang et al. 2016). This activation of

TCA cycle is consistent with the observations of the high energy requirements for anther and pollen development in wheat normal conditions (Tang et al. 2018). Mitochondria proteomic studies have also shown the down-regulation of the mitochondrial electron transport chain (mtETC) and TCA proteins in CMS wheat (Wang et al. 2015). In CMS maize, genes involved in energy metabolism have been also associated with male fertility restoration (Liu et al. 2018a). On the other hand, conversion of cytosolic citrate to acetyl-coA by ATP-citrate lyase (ACL) is upregulated in the sterile condition (Traes_3AS_27b1d22dd and Traes_7BL_3ed49605b). ACL generates cytosolic acetyl-CoA for the synthesis of a variety of phytochemicals including waxes, isoprenoids, stilbenes, and flavonoids (Fatland et al. 2002). Accordingly, secondary metabolism related to wax, flavonoids and phenylpropanoids is up-regulated in male sterile anthers. Regarding phenylpropanoids and lignin synthesis, Traes_1DL_93e80c14a (ferulate 5-hydroxylase; F5H) and Traes_6DS_211935e65 (cinnamyl alcohol dehydrogenase) are upregulated in the sterile condition. It has been shown that overexpression of FH5 in *Arabidopsis* leads to disruption of pollen wall formation, revealing a link between the biosynthetic pathways of lignin and sporopollenin (Weng et al. 2010). Genes related to wax metabolism (Traes_6AS_a40580bca and Traes_7AS_c71fe9684) and to cell wall (Traes_2AL_d3715cc49, Traes_2BS_d44e4b43a, Traes_5BL_2136d403e and Traes_4AS_ff5c27e0a) are also up-regulated in sterile anthers that, along with secondary metabolisms activation, is compatible with a response to oxidative stress. Excessive accumulation of reactive oxygen species (ROS) has been related to aberrant tapetal programmed cell death (PCD) progression, affecting the development of microspores and causing male sterility in alloplasmic wheat (Liu et al. 2018b). Accordingly, the functional category *stress* is overrepresented including modulated genes related to redox state, peroxidation, signaling, and heat shock proteins, among others.

The most abundant functional classifications correspond to *protein* and *RNA* (Supplemental Table S4), especially to protein synthesis and degradation and RNA regulation of transcription. The same phenomenon has been observed at the proteomic level in male-sterile wheat floret mitochondria (Wang et al. 2015) and complete wheat anthers (Zhang et al. 2018). Several genes related to ubiquitination are modulated belonging to the E2 and E3 sub-family genes (RING/U-box, SKP and cullins). Down-

regulation of ubiquitin/ proteasome pathway has been observed in male-sterile anthers of pepper, being related to cell cycle control response to damage DNA and PCD (Qiu et al. 2018). Involvement of protein degradation via proteasomes has been also associated to stress as described in rice post-meiosis panicle under heat stress, as a way of eliminating denatured and dysfunctional proteins (Zhang et al. 2014). Recently, it has been demonstrated that ubiquitination occurs in the mitochondria and that it is related to energy metabolism regulation (Lavie et al. 2018). The most direct relation with restoration of fertility has been observed in Lead rice-type CMS, where *RIF2* encodes a ubiquitin domain-containing protein, necessary for the restoration complex (Fujii et al. 2014).

Regarding the validation of the mapping strategy, the subset of DEGs assigned to 6H chromosome and those of unknown position was selected for that purpose, as fertility restoration in the msH1 system has been associated with the short arm of chromosome $6H^{ch}$ (Martín et al. 2008; Martín et al. 2009). The presence of all tested genes was verified in the H^{ch}ac (Table 4), validating BW mapping strategy. A second validation round was possible by using the new H^{ch}ac reorganizations named minichromosomes. The loss of integrity and stability of the H^{ch}ac is an intriguing event that will be deeply studied and discussed in an independent work. Due to their fertility restoration ability, H^{ch}mi1 and H^{ch}mi2 served to reduce the number of candidate genes to five: HORVU6Hr1G003170, HORVU6Hr1G003470, HORVU6Hr1G003770, HORVU6Hr1G003890 and HORVU6Hr1G004350. These five genes may be considered as candidates for the restoration gene of *H. chilense* as they are expressed in a crucial moment of pollen development and they are located in the three different restoring-of-fertility *H. chilense* chromosomal reorganizations. HORVU6Hr1G003890 is a polyadenylate binding-interacting protein 4-like, similar to ataxin-2, involved in regulation of cytoplasmic mRNA processing body assembly (Jiménez-López and Guzmán 2014). HORVU6Hr1G004350 encodes a protein similar to DExH-box ATP-dependent RNA helicases that plays essential roles in pre-mRNA splicing (Liu and Cheng 2015). HORVU6Hr1G003170 encodes for an ubiquitin-conjugating enzyme E2 protein, implicated in proteasome-mediated protein degradation via ubiquitination (Hershko et al. 1983). HORVU6Hr1G003470 gene is an aspartate aminotransferase, involved in amino acids metabolism and also related to metabolic flux in TCA cycle (Sweetlove et al. 2010; Fang et al. 2016). HORVU6Hr1G003770 encodes for an

acetyltransferase component PDC, which catalyzes the conversion of pyruvate to acetyl-CoA. Acetyl-CoA is indeed the molecule through which pyruvate enters the TCA in the mitochondria, and it is directly related to energy supply.

Along with their putative role as candidates for *Rf*, these five genes also tag a candidate restorer region on chromosome 6H^{ch}S. Taking part of this region, common in the three fertility restorer chromosomic reorganizations, it is also located the PPR gene HORVU6Hr1G004120. Noteworthy, this gene corresponds to the restorer gene *Rfm1* described in *H. vulgare* (Ui et al. 2015; Rizzolatti et al. 2017), although authors use the nomenclature Hv_Bradi3g00900A and B due to discrepancies with the assembly and annotation of HORVU6Hr1G004120. Although no significant differential expression was detected for this gene, a minimal expression was detected in BW MM2: 1.9 and 1.01 normalized counts in Fert_3 and Fert_5 libraries, respectively, and 0 in both sterile libraries (data not shown). These values, due to the few reads mapped and consequently the very poor statistical power, did not translate to a confidence DEG call of FDR < 0.05. Furthermore, a sequence homology up to 70% has been described in some restorer-of-fertility-like (RFL) PPR genes in barley (Melonek et al. 2019), and for shorter sequence lengths up to 96% (http://plants.ensembl.org/Hordeum_vulgare/Tools/Blast). Due to this similarity, reads from HORVU6Hr1G004120 transcripts could have been discarded since they mapped to more than one gene, giving an explanation to the low expression values observed. Besides, it cannot be ruled out that this gene is expressed earlier during anther development. Even though it was not reported as a DEG, HORVU6Hr1G004120 can be considered a suitable candidate for *Rf* in the msH1 system due to its position (between HORVU6Hr1G003890 and HORVU6Hr1G004350), its presence in the H^{ch}ac and in the two minichromosomes and its nature of PPR gene.

Final remarks

Comparative transcriptomics is revealed as a powerful method for candidate gene identification in *H. chilense* despite the limitations of a wild species with scarce genomic information. These limitations were compensated with an original RNA-seq mapping strategy, which was validated by genomic mapping of *H. chilense* DEGs and used to delimit a candidate genomic region. Additionally, minichromosomes were

successfully used for a second validation round to shorten this region and the number of candidate genes to six, being one of them the orthologue of the restorer gene *Rfm1* in *H. vulgare*.

The novel BW mapping approach was very effective discerning *H. chilense* transcripts in a wheat background. This strategy of constructing virtual genomes for mapping by merging genomes from different species could be also useful when investigating tritordeum (or other amphiploids) or introgression lines from different species.

Finally, this RNA-seq analysis gave some preliminary insights of restoration process, suggesting the importance of stress, energy supply, protein metabolism and RNA processing in the msH1 system.

Conflict of interests

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Avila CM, Mattera MG, Rodríguez-Suárez C, Palomino C, Ramírez MC, Martín A, Kilian A, Hornero-Méndez D, Atienza SG (2019) Diversification of seed carotenoid content and profile in wild barley (*Hordeum chilense* Roem. et Schultz.) and *Hordeum vulgare* L.–*H. chilense* synteny as revealed by DArTSeq markers. Euphytica 215:45 doi:10.1007/s10681-019-2369-6

Bohra A, Jha UC, Adhimoolam P, Bisht D, Singh NP (2016) Cytoplasmic male sterility (CMS) in hybrid breeding in field crops. Plant Cell Rep 35:967-993 doi:10.1007/s00299-016-1949-3

Castillo A, Atienza SG, Martín AC (2014) Fertility of CMS wheat is restored by two *Rf* loci located on a recombined acrocentric chromosome. J Exp Bot 65:6667-6677 doi:10.1093/jxb/eru388

Castillo A, Rodríguez-Suárez C, Martín AC, Pistón F (2015) Contribution of chromosomes $1H^{ch}S$ and $6H^{ch}S$ to fertility restoration in the wheat msH1 CMS system under different environmental conditions. PLOS ONE 10:e0121479 doi:10.1371/journal.pone.0121479

Chen L, Liu YG (2014) Male sterility and fertility restoration in crops. Annu Rev Plant Biol 65:579-606 doi:10.1146/annurev-arplant-050213-040119

Dahan J, Mireau H (2013) The Rf and Rf-like PPR in higher plants, a fast-evolving subclass of PPR genes. RNA Biol 10:1469-1476 doi:10.4161/rna.25568

Fang X, Fu HF, Gong ZH, Chai WG (2016) Involvement of a universal amino acid synthesis impediment in cytoplasmic male sterility in pepper. Sci Rep 6:23357 doi:10.1038/srep23357

Fatland B, Ke J, Anderson MD, Mentzen WI, Cui LW, Allred CC, Johnston JL, Nikolau BJ, Wurtele ES (2002) Molecular characterization of a heteromeric ATP-citrate Lyase that generates cytosolic acetyl-coenzyme A in *Arabidopsis*. Plant Physiol 130 (2):740-56 doi:10.1104/pp.008110

Fujii S, Kazama T, Ito Y, Kojima S, Toriyama K (2014) A candidate factor that interacts with RF2, a restorer of fertility of Lead rice-type cytoplasmic male sterility in rice. Rice (NY) 7:21-21 doi:10.1186/s12284-014-0021-6

Gaborieau L, Brown GG, Mireau H (2016) The propensity of pentatricopeptide repeat genes to evolve into restorers of cytoplasmic male sterility. Front Plant Sci 7:1816 doi:10.3389/fpls.2016.01816

Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. The Plant Cell 16:S154-S169 doi:10.1105/tpc.015966

Hershko A, Heller H, Elias S, Ciechanover A (1983) Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. J Biol Chem 258:8206-8214

Jiménez-López D, Guzmán P (2014) Insights into the evolution and domain structure of ataxin-2 proteins across eukaryotes. BMC Res Notes 7:453 doi:10.1186/1756-0500-7-453

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36 doi:10.1186/gb-2013-14-4-r36

Kim YJ, Zhang D (2018) Molecular control of male fertility for crop hybrid breeding. Trends Plant Sci 23:53-65 doi:https://doi.org/10.1016/j.tplants.2017.10.001

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357 doi:10.1038/nmeth.1923

Lavie J, De Belvalet H, Sonon S, Ion AM, Dumon E, Melser S, Lacombe D, Dupuy JW, Lalou C, Bénard G (2018) Ubiquitin-dependent degradation of mitochondrial proteins regulates energy metabolism. Cell Rep 23:2852-2863 doi:10.1016/j.celrep.2018.05.013

Liu YC, Cheng SC (2015) Functional roles of DExD/H-box RNA helicases in pre-mRNA splicing. J Biomed Sci 22:54 doi:10.1186/s12929-015-0161-z

Liu Y, Wei G, Xia Y, Liu X, Tang J, Lu Y, Lan H, Zhang S, Li C, Cao M (2018a) Comparative transcriptome analysis reveals that tricarboxylic acid cycle-related genes are associated with maize CMS-C fertility restoration. BMC Plant Biol 18:190 doi:10.1186/s12870-018-1409-z

Liu Z, Shi X, Li S, Zhang L, Song X (2018b) Oxidative stress and aberrant programmed cell death are associated with pollen abortion in isonuclear alloplasmic male-sterile wheat. Front Plant Sci 9:595 doi:10.3389/fpls.2018.00595

Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, Tohge T, Fernie AR, Stitt M, Usadel B (2014) Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. Plant, Cell Environ 37:1250-1258 doi:10.1111/pce.12231

Longin CF, Gowda M, Mühleisen J, Ebmeyer E, Kazman E, Schachschneider R, Schacht J, Kirchhoff M, Zhao Y, Reif JC (2013) Hybrid wheat: quantitative genetic parameters and consequences for the design of breeding programs. Theor Appl Genet 126:2791-2801 doi:10.1007/s00122-013-2172-z

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550 doi:10.1186/s13059-014-0550-8

Martín AC, Atienza SG, Ramírez MC, Barro F, Martín A (2008) Male fertility restoration of wheat in Hordeum chilense cytoplasm is associated with $6H^{ch}S$ chromosome addition. Aust J Agric Res 59:206-213 doi:https://doi.org/10.1071/AR07239

Martín AC, Atienza SG, Ramírez MC, Barro F, Martín A (2009) Chromosome engineering in wheat to restore male fertility in the msH1 CMS system. Mol Breed 24:397-408 doi:10.1007/s11032-009-9301-z

Martín AC, Atienza SG, Ramírez MC, Barro F, Martín A (2010) Molecular and cytological characterization of an extra acrocentric chromosome that restores

male fertility of wheat in the msH1 CMS system. Theor Appl Genet 121:1093-1101 doi:10.1007/s00122-010-1374-x

Martín AC, Castillo A, Atienza SG, Rodríguez-Suárez C (2018) A cytoplasmic male sterility (CMS) system in durum wheat. Mol Breed 38:90 doi:10.1007/s11032-018-0848-4

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:3 doi:10.14806/ej.17.1.200

Melonek J, Zhou R, Bayer PE, Edwards D, Stein N, Small I (2019) High intraspecific diversity of restorer-of-fertility-like genes in barley. The Plant J 97:281-295 doi:10.1111/tpj.14115

Mühleisen J, Piepho H-P, Maurer HP, Longin CFH, Reif JC (2014) Yield stability of hybrids versus lines in wheat, barley, and triticale. Theor Appl Genet 127:309-316 doi:10.1007/s00122-013-2219-1

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321-4325

Qiu Y, Liao L, Jin X, Mao D, Liu R (2018) Analysis of the meiotic transcriptome reveals the genes related to the regulation of pollen abortion in cytoplasmic male-sterile pepper (*Capsicum annuum* L.). Gene 641:8-17 doi: 10.1016/j.gene.2017.10.022

Raupp W, Friebe B, Gill B (1995) Suggested guidelines for the nomenclature and abbreviation of the genetic stock of wheat and its relatives. Wheat Information Service 81:51-54

Rey MD, Moore G, Martín AC (2018) Identification and comparison of individual chromosomes of three accessions of *Hordeum chilense*, *Hordeum vulgare*, and *Triticum aestivum* by FISH. Genome 61:387-396 doi:10.1139/gen-2018-0016

Rizzolatti C, Bury P, Tatara E, Pin PA, Rodde N, Bergès H, Budar F, Mireau H, Gielen JJL (2017) Map-based cloning of the fertility restoration locus *Rfm1* in cultivated barley (*Hordeum vulgare*). Euphytica 213:276 doi:10.1007/s10681-017-2056-4

Sweetlove LJ, Beard KFM, Nunes-Nesi A, Fernie AR, Ratcliffe RG (2010) Not just a circle: flux modes in the plant TCA cycle. Trends Plant Sci 15:462-470 doi:10.1016/j.tplants.2010.05.006

Tang H, Song Y, Guo J, Wang J, Zhang L, Niu N, Ma Sm Zhang G, Zhao H (2018) Physiological and metabolome changes during anther development in wheat

(*Triticum aestivum* L.). Plant Physiol Biochem 132:18-32 doi: 10.1016/j.plaphy.2018.08.024

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. The Plant J 37:914-939 doi:10.1111/j.1365-313X.2004.02016.x

Ui H, Sameri M, Pourkheirandish M, Chang MC, Shimada H, Stein N, Komatsuda T, Handa H (2015) High-resolution genetic mapping and physical map construction for the fertility restorer *Rfm1* locus in barley. Theor Appl Genet 128:283-290 doi:10.1007/s00122-014-2428-2

Wang S, Zhang G, Zhang Y, Song Q, Chen Z, Wang J, Guo J, Niu N, Wang J, Ma S (2015) Comparative studies of mitochondrial proteomics reveal an intimate protein network of male sterility in wheat (*Triticum aestivum* L.). J Exp Bot 66:6191-6203 doi:10.1093/jxb/erv322

Weng JK, Mo H, Chapple C (2010) Over-expression of F5H in COMT-deficient *Arabidopsis* leads to enrichment of an unusual lignin and disruption of pollen wall formation. The Plant J 64:898-911 doi:10.1111/j.1365-313X.2010.04391.x

Whitford R, Fleury D, Reif JC, Garcia M, Okada T, Langridge P, Korzun V (2013) Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. J Exp Bot 64:5411-5428 doi:10.1093/jxb/ert333

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134-134 doi:10.1186/1471-2105-13-134

Zhang G, Ye J, Jia Y, Zhang L, Song X (2018) iTRAQ-based proteomics analyses of sterile/fertile anthers from a thermo-sensitive cytoplasmic male-sterile wheat with *Aegilops kotschyi* cytoplasm. Int J Mol Sci 19(5):1344 doi:10.3390/ijms19051344

Zhang X, Xiong H, Liu A, Zhou X, Peng Y, Li Z, Luo G, Tian X, Chen X (2014) Microarray data uncover the genome-wide gene expression patterns in response to heat stress in rice post-meiosis panicle. J Plant Biol 57(6):327-336 doi:10.1007/s12374-014-0177-z

Tables

Table 1. Plant material used in this work. Nomenclature for the genetic stocks of wheat and its relatives suggested by (Raupp et al. 1995) is used

Line	Germplasm	Chromosome	
		Number	Configuration
H1	*H. chilense* Roem. et Schultz. accession H1	14	7''
T26	*T. aestivum* cv. T26	42	21''
T21	*T. aestivum* cv. Chinese Spring	42	21''
T21A1H_1S	*T. aestivum* cv. CS–*H. chilense* ditelosomic addition of 1H^{ch}S	42+t''	21'' + t''1H^{ch}S
T21A6H_1S	*T. aestivum* cv. CS–*H. chilense* ditelosomic addition of 6H^{ch}S	42+t''	21'' + t''6H^{ch}S
T749	*T. aestivum* cv. CS-*H. chilense* disomic addition acrocentric chromosome in H1 cytoplasm	42+ac''	21''+1''H^{ch}ac
T528	*T. aestivum* cv. T26-*H. chilense* disomic addition acrocentric chromosome in H1 cytoplasm	42+ac''	21''+1''H^{ch}ac
T854	*T. aestivum* cv. T26-*H. chilense* monosomic addition minichromosome 2 in H1 cytoplasm	42+mi'	21''+1' H^{ch}mi2
T527	*T. aestivum* cv. T26-*H. chilense* monosomic addition minichromosome 1 in H1 cytoplasm	42+mi'	21''+1' H^{ch}mi1

Table 2. Number of differentially expressed genes (DEGs) identified with the two mapping strategies. DEGs identified in barley chromosomes 1H-7H and of unknown position (Unk), with OnlyBarley (OB) and BarWheat (BW) mapping approaches considering the different mismatching conditions (MM)

	Number of DEGs						
	OB				BW		
Chromosome	MM0	MM1	MM2	MM3	MM0	MM1	MM2
1H	8	6	8	15	29	68	106
2H	1	2	3	3	-	2	2
3H	-	4	8	5	-	-	1
4H	1	1	2	3	-	-	-
5H	-	1	3	6	-	-	2
6H	-	2	3	4	4	9	19
7H	-	2	7	5	-	-	2
Unk	1	0	1	1	2	4	4

Table 3. DEGs identified on chromosome 6H and of unknown position in *H. vulgare*. Log_2fold and p values and a short description of each gene are shown

H. vulgare ID	log_2fold	p value	Short description
HORVU6Hr1G000640	7.73	5.77E-07	SKP1 1B
HORVU6Hr1G001670	1.65	5.97E-06	Hsc70-interacting/ TPR-like superfamily protein
HORVU6Hr1G001860	3.18	3.04E-16	Homeobox engrailed-1 / Pollen Ole e 1 allergen and extension fam
HORVU6Hr1G002130	6.31	0.000191	Topless-related 2 isoform X2
HORVU6Hr1G003170	6.10	8.37E-15	NEDD8-conjugating enzyme Ubc12 /RUB1 conjugating enzyme 1
HORVU6Hr1G003470	6.27	1.01E-08	Aspartate mitoch / aspartate aminotransferase 1
HORVU6Hr1G003770	5.64	3.28E-09	Dihydrolipoamide S-acetyltransferase /component of PDC
HORVU6Hr1G003890	3.25	9.71E-05	Polyadenylate-bind-interacting 4-like /CTC-interacting domain 4
HORVU6Hr1G003900	6.84	2.44E-05	Probable methyltransferase PMT17 isoform X2
HORVU6Hr1G004350	4.31	1.12E-06	DExH-box ATP-dependent RNA helicase DExH12
HORVU6Hr1G005370	4.67	3.98E-11	Pre-mRNA-processing 40A-like
HORVU6Hr1G005390	2.31	2.61E-05	60S ribosomal L37
HORVU6Hr1G005730	7.06	9.86E-06	Pre-mRNA-processing factor 39-like
HORVU6Hr1G006050	7.32	9.67E-07	ATP-dependent helicase BRM
HORVU6Hr1G006560	6.80	2.86E-05	2-oxoisovalerate dehydrogenase subunit beta 1
HORVU6Hr1G006880	4.34	2.92E-05	Serine carboxypeptidase-like/ Carboxypeptidase Y homolog A
HORVU6Hr1G007030	6.91	1.81E-05	F-box kelch-repeat SKIP4
HORVU6Hr1G011020	2.88	2.31E-06	Histone H4
HORVU6Hr1G083600	7.57	1.2E-06	-
HORVU0Hr1G000210	7.69	6.64E-07	-
HORVU0Hr1G000390	7.94	2.41E-07	Glycine-rich protein
HORVU0Hr1G000410	8.17	8.06E-08	Zinc transporter ZupT
HORVU0Hr1G014960	10.69	1.96E-40	GDSL esterase lipase

Table 4. DEGs and PPR genes validation and evaluation by using *H. chilense* (H1), wheat lines (T21 and T26), wheat lines with the H^{ch}ac addition (T749 and T528) and wheat lines with the minichromosomes H^{ch}mi1 or H^{ch}mi2 addition (T527 and T854, respectively)

H. vulgare ID	*H. vulgare* position	H1	T21	T26	T21+ H^{ch}ac T749	T26+ H^{ch}ac T528	T26+ H^{ch}mi2 T854	T26+ H^{ch}mi1 T527
HORVU6Hr1G001670	chr6h:5062417-5067619	+	-	-	+	+	+	-
HORVU6Hr1G002130	chr6h:6017177-6023449	+	-	-	+	+	+	-
HORVU6Hr1G002890 [P]	chr6h:7028863-7034152	+	-	-	+	+	+	-
HORVU6Hr1G003170	chr6h:7538700-7542614	+	-	-	+	+	+	+
HORVU6Hr1G003470	chr6h:7898535-7902987	+	-	-	+	+	+	+
HORVU6Hr1G003770	chr6h:8213254-8219841	+	-	-	+	+	+	+
HORVU6Hr1G003890	chr6h:8449061-8457843	+	+	-	*	+	+	+
HORVU6Hr1G003900	chr6h:8535557-8541335	+	-	+	+	*	*	*
HORVU6Hr1G004120 [P]	chr6h:9155451-9171768	+	-	-	+	+	+	+
HORVU6Hr1G004350	chr6h:9845758-9853851	+	-	-	+	+	+	+
HORVU6Hr1G005240 [P]	chr6h:11509347-11515094	+	-	-	+	+	+	-
HORVU6Hr1G005370	chr6h:11928429-11944450	+	-	-	+	+	+	-
HORVU6Hr1G005380 [P]	chr6h:11980250-11981113	+	-	-	+	+	+	-
HORVU6Hr1G005390	chr6h:12012122-12015583	+	-	-	+	+	+	-
HORVU6Hr1G005730	chr6h:12657549-12663784	+	-	-	+	+	+	-
HORVU6Hr1G006050	chr6h:13265913-13277591	+	-	-	+	+	+	-
HORVU6Hr1G006560	chr6h:14070752-14075041	+	-	-	+	+	+	-
HORVU6Hr1G006880	chr6h:14328002-14332255	+	-	-	+	+	+	-
HORVU6Hr1G007030	chr6h:14502121-14506662	+	-	-	+	+	+	-
HORVU6Hr1G011020	chr6h:19418348-19420721	+	-	-	+	+	+	-
HORVU0Hr1G014960	chrUn:82980303-82982385	+	-	-	+	+	+	-
HORVU0Hr1G000390	chrUn:2848447-2853394	+	-	-	+	+	-	+

(*) Not valid in this wheat background

([P]) PPR genes, not DEGs

Figure captions

Fig. 1 Localization of *H. chilense* DEGs based on *H. vulgare* orthologues position on chromosomes 1H and 6H

Fig. 2 Overview of the metabolism by MapMan. Positive fold change values (red) and negative (blue) correspond to up-regulated genes in fertile and in sterile libraries, respectively

Fig. 3 Validation of DEGs and the PPR-like genes. Genes HORVU6Hr1G005730, HORVU6Hr1G004350, HORVU0Hr1G000390 and the PPR-like gene HORVU6Hr1G004120 were validated by amplification in *H. chilense* (H1), common wheat (T21 and T26), addition lines T21A1H_1S and T21A6H_1S (ad1H^{ch} and ad6H^{ch}, respectively), lines harboring the H^{ch}ac (T749 and T528), T527 descents with and without H^{ch}mi1 (T527$^+$ and T527$^-$ respectively) and T854 descents with and without H^{ch}mi2 (T854$^+$ and T854$^-$ respectively)

Fig. 4 In situ hybridization to root-tip metaphase cells from restored lines T527 (A) and T854 (B). GISH signal using *H. chilense* genomic DNA is shown in magenta. Blue DAPI staining shows wheat chromosomes. Boxed regions show enlarged views of minichromosomes H^{ch}mi1 (A) and H^{ch}mi2 (B), which display magenta colour indicating its *H. chilense* origin. DAPI staining of minichromosomes is also shown in blue inside boxed regions.