Agriculture and Agri-Food Canada | Agriculture et Agroalimentaire Canada

Genomic Tool for Fine Mapping and Marker Development of Wheat Resistance Genes

Frank M. You[1], Pingchuan Li[1], Jin Xiao[1,2], Mulualem T. Kassa[1,3], Curt A. McCartney[1], Colin W. Hiebert[1], Mark C. Jordan[1], Sylvie Cloutier[4], Curtis Pozniak[5], Pierre Fobert[3], Xiue Wang[2], Ming-Cheng Luo[6] and Jan Dvorak[6]

[1]Agriculture and Agri-Food Canada, Morden, MB, Canada; [2]Nanjing Agricultural University, Nanjing, China; [3]National Research Council, Saskatoon, SK, Canada; [4]Agriculture and Agri-Food Canada, Ottawa, ON, Canada; [5]University of Saskatchewan, Saskatoon, SK, Canada; [6]University of California Davis, Davis, CA, United States

Introduction

Bread wheat (2n = 6x = 42, AABBDD) evolved from hybridization of tetraploid emmer wheat (AABB) and diploid *Aegilops tauschii* (2n=14, DD) approximately 8,000 years ago (Nesbitt, 1995; Petersen et al., 2006; Jia et al., 2013). The bread wheat chromosome survey sequence (CSS) is completed and physical mapping and reference sequencing is on going and yet to be fully sorted for three sub-genomes (http://www.wheatgenome.org/Projects/).

Over 100 diseases caused by different pathogens have been reported in wheat, which threatened wheat production around the globe (Chelkowski, 2001). Resistance gene analogs (RGAs) are a class of potential resistance (*R*) genes, including three major types, NBS-encoding proteins, receptor-like protein kinases (RLKs) and receptor-like proteins (RLPs) (Kumar et al. 2015). To date, most of cloned *R*-genes can be grouped into one of them. Analyses of NBS encoding genes from various crop plants have suggested that only a small part of the *R*-genes may be functional (Chin et al., 2001). Major approach to face this challenge is to discover new disease resistance genes and develop new RGAs associated high density SNP markers. Currently the map-based identification and isolation of disease resistance genes are the most efficient strategy and have been used in cloning of all currently available wheat genes (Sekhwal, 2015).

In wheat, major challenge for map-based cloning is to efficiently find the target disease resistance gene associated flanking markers because of large size genome, polyploid, and highly repeated DNA sequences (Paux et al., 2006; International Wheat Genome Sequencing, 2014).The primary goals of this study were to (1) identify all RGA candidates in hexaploid genome *Triticum aestivum*, and its diploid ancestors including *Triticum urartu* and *Aegilops tauschii*, the A and D genome donor respectively; (2) identify the RGA associated SNP markers from dozens of wheat cultivar exome sequence data; and (3) apply GenomeZipper and 90k SNP markers to establish a linear RGA gene order model of wheat genome on the basis of conserved synteny to rice, *Brachypodium* and sorghum. These data will provide useful genomic tool for fine mapping and marker development of wheat resistant genes.

Materials and Methods

RGA identification: The protein sequences of *T. aestivum*, *T. uratu*, *A. tauschii* and several other cereal crop genomes were used for RGA identification with a comprehensive genome-wide RGA identification package, RGAugury (Li et al. 2016).

SNP calling: Illumina exome sequences of 100 bp pair-ends of 89 hexaploid/tetraploid cultivars were used for SNP discovery. The raw FASTQ reads were directly aligned with genomic gene sequences by using Burrows-Wheeler Aligner (BWA, v0.7.15), and a custom SNP calling pipeline was used for SNP calling.

RGA homolog analysis: The protein sequences of RGAs identified from *T. aestivum*, *T. uratu*, *A. tauschii*, *Brachypodium* and *O. sativa* were aligned reciprocally by BLASTP at 1e-5. Top one hits above the E-value cutoff were used for homolog analysis. An ortholog between two species was defined as two genes from two different species which are reciprocally top one hits in BLAST. Similar approached was applied to homologs in subgenomes of bread wheat.

Sorting of RGAs onto chromosomes: *GenomeZipper* (GZ) data of wheat and other species were obtained from PGSB (formerly MIPS). Wheat RGAs existed in GZ were kept to build a GZ linkage map based on their collinearity with other species. All 90k SNP probe sequences (Wang et al., 2014) were aligned with CSS genome sequence and only the best hits with the correct chromosome and arms assignment were retained. The SNP markers associated contigs were thus sorted by their centimorgan distance for each chromosome. A comparative map was created using CMap.

Results and Discussion

Using RGAugury, 4,848 (2.2% of total genes), 1,591 (4.6%), 1,803 (2.2%) RGAs from hexaploid wheat (AABBDD), and its diploid progenitors *Triticum uratu* (AA) and *Aegilops tauschii* (DD), respectively, were identified. Of these in the bread wheat genome, 1,424 (29.3% of total RGAs), 1,722 (35.5%) and 1,430 (29.5%) RGAs were distributed in A, B and D sub-genomes, respectively, and the remaining 272 (5.6%) were not aligned to chromosomes. RGAs were unevenly distributed and most of them were clustered together and spread across chromosomes by collinearity studies. The homolog analysis of RGAs between wheat and other cereal species indicated that wheat RGA genes were subjected to substantial duplication events after the differentiation of wheat and other cereal species, especially the progenitors of A, B and D genomes. A total of 14,299 genome-specific SNPs on 1,727 out of 4848 RGAs (35.6%) were identified from exome sequencing data of 89 wheat cultivars and breeding lines in Canada and other countries. Of the 8 types of RGAs, CNL and NL are the most abundant SNP associated NBS encoding genes.

Table 1. RGAs identified from wheat, rice and *Brachypodium*

	NBS-encoding							RLP	RLK	TM-CC	Total
	NBS	CNL	TNL	CN	TN	NL	TX				
T. aestivum (ABD)	658	396	0	323	10	866	4	385	1900	306	4848
T. urartu (A)	36	229	0	56	3	258	1	110	791	107	1591
A. tauschii (D)	68	214	0	41	3	272	1	146	925	132	1803
Brachypodium	26	115	0	21	3	94	1	53	612	96	1021
O. sativa	47	220	0	38	3	191	3	117	855	134	1608

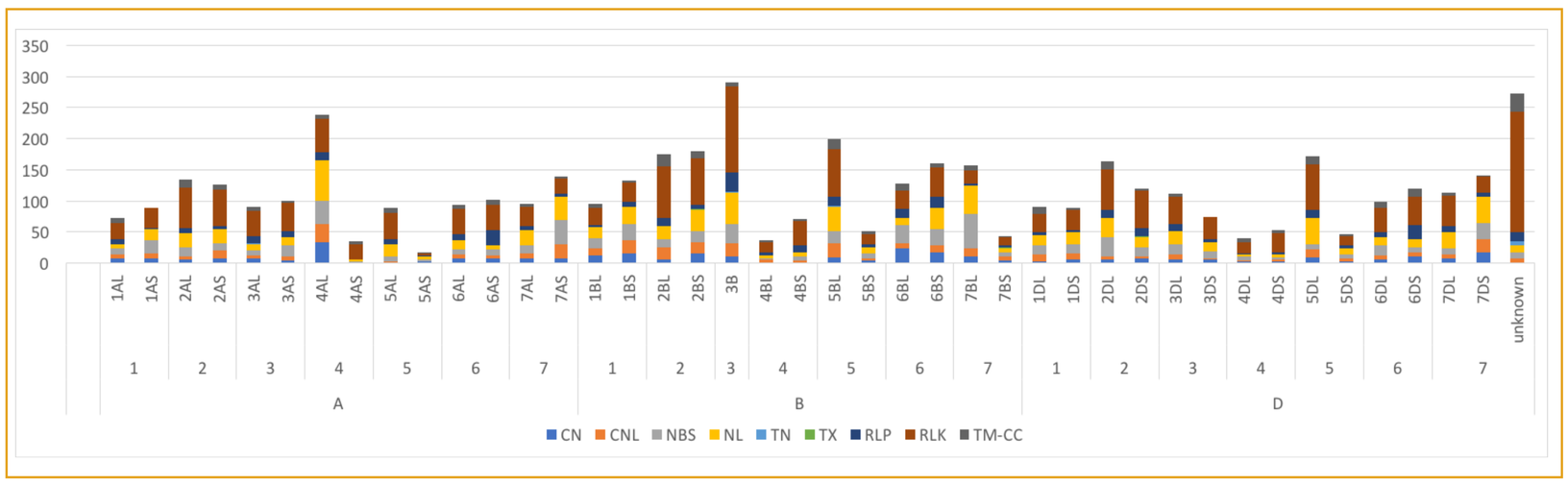

Fig 1. RGA distribution across arms of hexaploid wheat sub-genomes. Chromosome 3B has no further separation by arms. The unsorted RGAs were treated as the unknown group.

Fig 2. The RGA syntenic relationships between bread wheat and *Brachypodium*, rice and sorghum. Gene density of orthologous RGAs of wheat in *Brachypodium*, rice and sorghum were calculated using a sliding window of 500kb. Top two hits were analyzed.

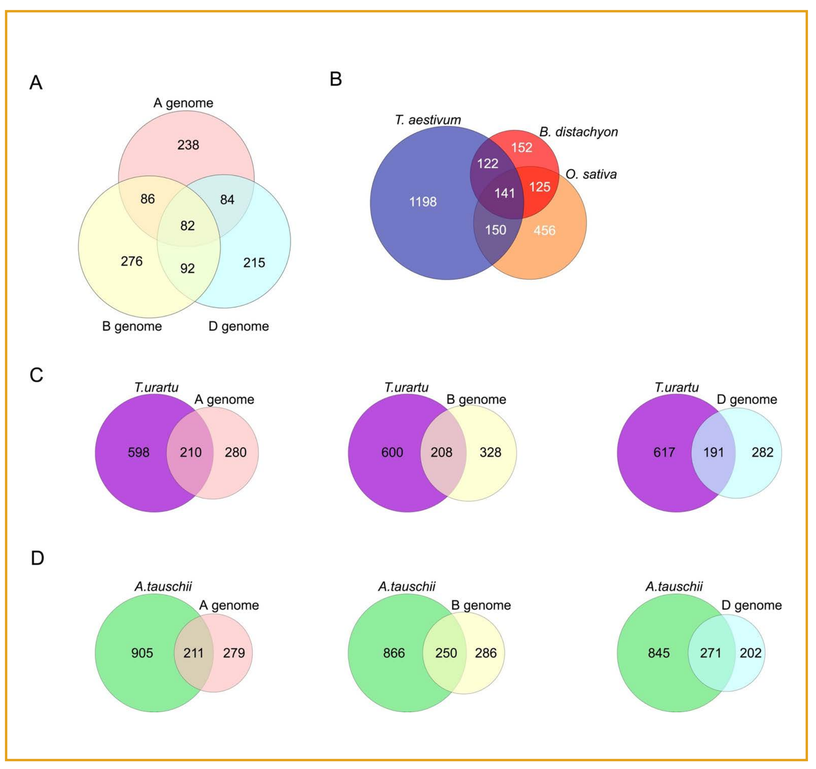

Fig 3. Venn diagram of RGAs identified from wheat, *brachypodium* and rice. (A) RGAs identified from three subgenomes of bread wheat. (B) RGAs identified from bread wheat, *Brachypodium* and rice. (C) RGAs identified from *T. urartu* and subgenome of bread wheat. (D) RGAs identified from *A. tauschii* and subgenomes of bread wheat. The diagram were drawn using Venn Diagram Plotter (goo.gl/JY0juv).

These RGAs and RGA-specific SNPs have been successfully used for fine mapping high impact wheat resistance genes such as Ug99 stem rust resistance gene (*SrCad*), leaf rust resistance gene (*Lr16*), and orange wheat blossom midge resistance gene (*Sm1*). Cost effective, breeder friendly and diagnostic SNP markers that are suitable for high throughput Marker Assisted Selection (MAS) have been developed for *Lr16* (Kassa et al. 2017) and *SrCad* (Kassa et al. 2016). Moreover, marker haplotypes were identified to determine the presence and absence of wheat midge resistance gene *Sm1* (Kassa et al. 2016). An RGA database integrating the developed genomic data has been developed. Partial of RGAs have been sorted across the arms or chromosome of bread wheat by using the 90K SNP markers. All these data provides genomic resources to locate specific *R* gene candidates and find gene-specific markers for fine-mapping and cloning of wheat *R* genes, and resistance breeding.

Fig 5. CMap of the RGA linkage map on bread wheat 1B between a 90k SNP chip based genetic map and a GZ based synteny map. The lines between the two maps are corresponding to the locations of RGAs (highlighted in red) on each map. The unit of genetic distance is cM.

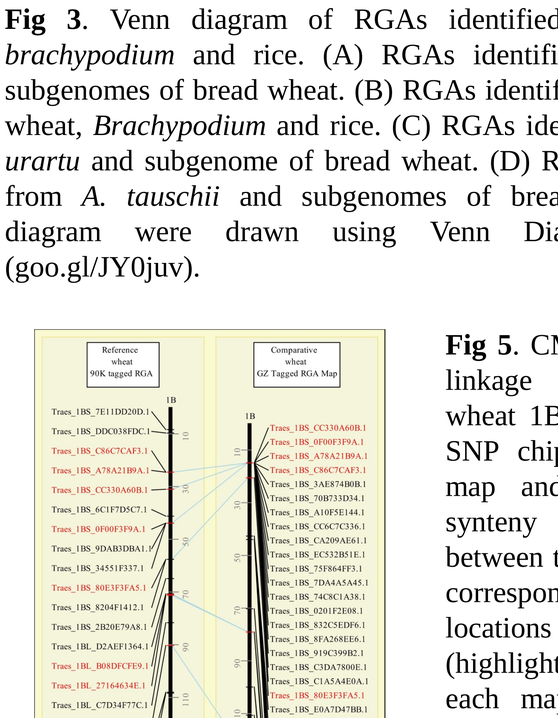

Fig 4. Box-plots for different types of SNP associated RGAs. *y*-axis represents the number of SNPs on RGAs.

Acknowledgements

This research is supported by the Canada GRDI project.

Agriculture and Agri-Food Canada, Morden Research and Development Center, 101 Route 100, Morden, Manitoba, Canada R6M 1Y5 Email : frank.you@agr.gc.ca

Canada