A Bioinformatic Approach to Identification of Flowering-related Genes in Wheat and Barley

Rong-Cai Yang[1,2] , Fred Yefang Peng[1] and Zhiqiu Hu[2]

[1]*Feed Crops Branch, Alberta Agriculture and Forestry, Edmonton, AB T6H 5T6, Canada*

[2]*Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB T6G 2P5, Canada*

Abstract

Early flowering is an important trait influencing grain yield and quality in wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.) in short-season cropping regions such as western Canada. However, the genetics of this trait in wheat and barley is complex as a large number of genes are known or postulated to be involved in functions or pathways related to the timing of flowering. In addition, molecular characterization of the flowering-related genes remains challenging for large and complex genomes of wheat and barley. The purpose of this presentation is to investigate the usefulness of a bioinformatic approach to identification of flowering-related genes in these crop species. Using a two-stage bioinformatic approach, we combined reciprocal BLAST searches and OrthoMCL clustering to identify 900 and 275 putative orthologs in wheat and barley, respectively, from 190 known Arabidopsis [*Arabidopsis thaliana* (L.) Heynh.] genes related to flowering functions or pathways. The annotated flowering-related genes were clustered into 144 orthologous groups with identification of one-to-one, one-to-many, many-to-one and many-to-many orthology relationships. Our approach was further validated by domain and phylogenetic analyses of flowering-related proteins of flowering-related genes in Arabidopsis and the two crop species. The putative orthologous sequences inferred from known Arabidopsis genes can be confirmed and incorporated into molecular breeding for early flowering and maturing wheat and barley in western Canada and other high-latitude regions.

Introduction

Early flowering and maturity is an important agronomic trait for crop species such as barley and wheat grown in high-latitude regions with short growing seasons. The genetic mechanism behind flowering time control has been well characterized in Arabidopsis, and many genes implicated underlying this process have been identified [1, for review]. In contrast, only a few genes related to flowering have been studied in barley and wheat [2, for review]. The presence of a dominant *Ppd-h1* allele has been shown to cause early heading in barley under long-day conditions [3,4], and a 2kb INDEL in the upstream of the wheat *Ppd-D1* gene (*Ppd-D1a* allele) was shown to explain 56% of phenotypic variation of early flowering in wheat [5,6]. In this presentation, we used a bioinformatic approach to identify all putative orthologs of the Arabidopsis flowering genes in barley and wheat, and analyzed their gene structures, domains, and expression in different tissues. The identified candidate flowering genes will be useful for further studies on molecular breeding for early flowering barley and wheat.

Acknowledgments

This research is funded by the *Growing Forward 2* Research Opportunities and Innovation Internal Initiatives of Alberta Agriculture and Forestry.

Materials and Methods

Identification of flowering-related genes in barley and wheat. The Arabidopsis flowering genes were found in TAIR [7] and literature. Sequences of flowering-related genes and encoded proteins in wheat and barley were identified using reciprocal BLAST. First, the Arabidopsis protein sequences were used to search the barley and wheat protein databases from Ensembl with BLAST [8]. Then all unique hits were used to BLAST back against the Arabidopsis proteome, and reciprocal top three hit pairs were considered putative orthologs. Finally, the protein sequences of all putative orthologs were subject to OrthoMCL [9], forming ortholog groups (OG).

Structural analysis of orthologous flowering genes. The structural features of these flowering genes in Arabidopsis, wheat and barley were summarized based on the genome annotations in the Ensembl Plants database (Release 28) [10]. To calculate exon and intron sizes, their sequences within each gene were retrieved from Ensembl using its Perl API (Application Program Interface).

Results

Using the 190 flowering genes in Arabidopsis as the starting query, we identified 274 and 900 putative flowering-related genes in the genomes of barley and wheat, respectively. Most known flowering genes in barley and wheat have been found in our analysis. For example, the photoperiod (*PPD*) genes in barley (*Ppd-H1)* and wheat (*Ppd-D1*) are Ensembl sequences MLOC_81154 and Traes_2DS_2A961F39D respectively, clustered with AtPRR7 (AT5G02810) in OG5_139246.

Table 1. The average percentage of protein sequence similarity of flowering genes in the seven functional groups in Arabidopsis, wheat and barley. AT vs TA, Arabidopsis - wheat comparison; AT vs HV, Arabidopsis – barley comparison. Numbers in parenthesis denote standard deviation.

Functional group	AT vs TA	AT vs HV
Autonomous	62.7 (16.9)	53.4 (14.5)
Flower development	68.8 (11.8)	63.2 (12.4)
Gibberellin	61.3 (15.0)	54.7 (11.0)
Pathway integration	74.8 (13.4)	59.0 (17.0)
Photoperiod	61.4 (18.0)	57.4 (16.3)
Regulation	69.8 (13.0)	62.0 (13.5)
Vernalization	58.9 (8.3)	44.6 (13.0)

The overall sequence similarity of flowering genes in different functional groups is shown in Table 1.This result indicates that genes in different flowering pathways show lower sequence conservation between Arabidopsis and temperate cereals barley and wheat.

Table 2. Structural characteristics of flowering-related genes in Arabidopsis (AT), wheat (TA), and barley (HV). The numbers of flowering genes used for the summary statistics are shown in parentheses. Single-exon genes (no introns) were excluded for intron calculation. bp, base pair; aa, amino acid.

	AT (n = 190)		TA (n = 525)		HV (n = 265)	
	Mean	Range	Mean	Range	Mean	Range
Transcripts per gene	1.4	1-5	1.0	1-1	2.8	1-27
Gene length (bp)	3161	182-16871	3815	240-20952	4328	404-15512
Exons per gene	6.5	1-48	5.7	1-42	4.5	1-20
Exon size (bp)	466	79-4165	565	42-5550	878	87-5211
Intron size (bp)	468	78-2316	924	58-7291	856	44-5912
Protein length (aa)	529	77-3529	444	52-3250	500	50-2056

The structural characteristics of these flowering genes differ in wheat and barley compared with Arabidopsis (Table 2). On average, the Arabidopsis flowering genes are smaller yet encode longer proteins. Interestingly, the introns are significantly larger in wheat and barley than in Arabidopsis. The first large intron of *VRN-H1* has been shown to affect the vernalization sensitivity in barley [11, 12]. It is expected that the flowering genes with extremely large introns in wheat and barley may be further characterized in future studies.

In summary, we used a bioinformatics approach that combines both reciprocal BLAST searches and OrthoMCL clustering to predict a large number of flowering-related genes in large and complex barley and wheat genomes whose direct characterization would hardly be possible. The annotated orthologous genes can be further studied and used in molecular breeding for early flowering in barley and wheat.

References

1. J. Exp. Bot. doi: 10.1093/jxb/eru441 (2014)
2. Front. Plant Sci. 5: 431 (2014)
3. Science 310: 1031-1034 (2005)
4. Russian J. Genetics: Appl Res 3: 254-264 (2013)
5. Theor. Appl. Genet. 115:721-733 (2007)
6. Genetics 199: 609–623 (2015)
7. Nucleic Acids Res. 40: D1202-D1210 (2012)
8. Nucleic Acids Res. 25: 3389-3402 (1997)
9. Genome Res. 13: 2178-2189 (2003)
10. Nucleic Acids Res. 43: D662-D669 (2015)
11. Mol. Genet. Genomics 277:249-261 (2007)
12. Curr. Opin. Plant Biol. 12:178-184 (2009)

Growing Forward 2
A federal-provincial-territorial initiative

UNIVERSITY OF ALBERTA

Alberta Government

Canada