Genotyping-by-Sequencing Reveals Genetic Diversity of Crested Wheatgrass [*Agropyron cristatum* (L.) Gaert.]

Kiran Baral[1], Yong-Bi Fu[2], Bill Biligetu[1], Bruce Coulman[1]

[1]Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, S7N 5A8, Canada. [2]Plant Gene Resources of Canada, Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, Saskatoon, SK, S9H 3X2, Canada.

INTRODUCTION

Crested wheatgrass [*Agropyron cristatum* (L.) Gaertn.], occurs in diploid, tetraploid and hexaploid forms (Dewey, 1984). Characteristics such as drought tolerance, winter hardiness, high quality early spring forage, ease of establishment, strong competitive ability, tolerance to insect predation, and high forage yield makes it an important commercial species in Canadian grasslands (Looman & Heinrichs, 1973). Genotyping-by-sequencing (GBS) is capable of producing high-density, low-cost genotypic information without the requirement for a reference genome sequence. This study represents the first use of GBS technology to sample genome-wide variants of crested wheatgrass and assess the genetic diversity.

OBJECTIVES

- Identify genome-wide SNP markers in 12 accessions of *A. cristatum* from Canada, Kazakhstan, and Siberia and,
- Evaluate the utility of the GBS in the genetic diversity analysis of complex polyploid plants.

MATERIALS AND METHODS

Plant material:

- Sixteen genotypes each from 12 tetraploid crested wheatgrass accessions comprising 192 genotypes in total (Table 1).

Genotyping-by-Sequencing:

- Diversity-focused (gd-GBS) using two restriction enzyme (*Pst*I/*Msp*I) digestion as described by Peterson et al., (2014).

Bioinformatic Analysis:

- Raw sequence (FASTQ) data were cleaned using Trimmomatic v0.36 (Bolger et al., 2014).
- FASTQ files of 250 bp was split into three fragment sets with a custom Perl script with an intact 5-base *Pst*I residue (TGCAG) at the beginning for UNEAK-GBS pipeline (Lu et al., 2013).
- Each fragment was analyzed with UNEAK and the Haplotag pipelines and generated a matrix of samples by SNP loci.
- Character by Taxa (CbyT) program generated a filtered SNP file.

Genetic Diversity Analysis:

- A model-based Bayesian method available in the program STRUCTURE version 2.2.3 (Pritchard et al., 2000; Falush et al., 2007).
- A neighbor-joining (NJ) analysis using PAUP* (Swofford, 2000).
- A principal coordinate analysis (PCoA) using the R routine *AveDissR* (Yang & Fu, 2017; R Development Core Team, 2016).
- Genetic variation were computed through analysis of molecular variance (AMOVA) using Arlequin version 3.5 (Excoffier & Lischer, 2010).
- An unweighted pair group method with arithmetic mean (UPGMA) dendrogram based on pairwise genetic distances.

MATERIALS AND METHODS (continued)

Table 1. List of the 12 crested wheatgrass (*A. cristatum*) accessions used in the study

Accession	CN number	Alternative identification	Origin	
AC Goliath	CN108673	-	Canada	Cultivar
AC Newkirk	-	FOR552	Canada	Cultivar
Vysokij 9	CN30995	PI 370654	Siberia	Cultivar
Karabalykskij 202	CN31068	PI 326204	Kazakhstan	Cultivar
PGR 16830	CN43478	-	Kazakhstan	Collection
Kirk	CN108662	PI 536010	Canada	Cultivar
S8959E	-	FOR917	Canada	Breeding line
S9491	-	S9491	Canada	Breeding line
S9514	-	S9514	Canada	Breeding line
S9516	-	S9516	Canada	Breeding line
S9544	-	S9544	Canada	Breeding line
S9556	-	S9556	Canada	Breeding line

RESULTS

- There were approximately 87.8 million raw forward (R1) sequence reads.
- The number of raw forward sequence reads per sample ranged from 190,606 to 775,160 with an average of 457,279.
- Combined UNEAK and Haplotag analysis at the 50% level of missing data generated 45,507 SNPs across 192 genotypes.

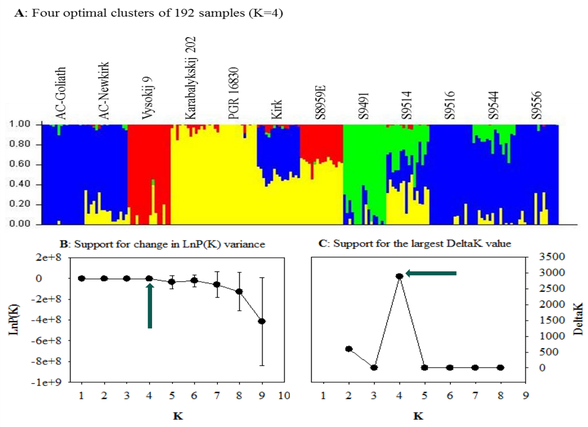

Figure 1. Four genetic clusters of 192 *Agropyron cristatum* genotypes inferred by STRUCTURE based on 45,507 SNP markers. **A**: The mixture coefficients of 192 plants with K=4, presented in the original order of genopytpes from 12 accessions (see Table 1 for accession label). **B**: Support from the LnP(K) estimation. **C**: Support from the estimation of the largest value of the delta K= mean (|Ln"(K)|) / sd (LnP(K)).

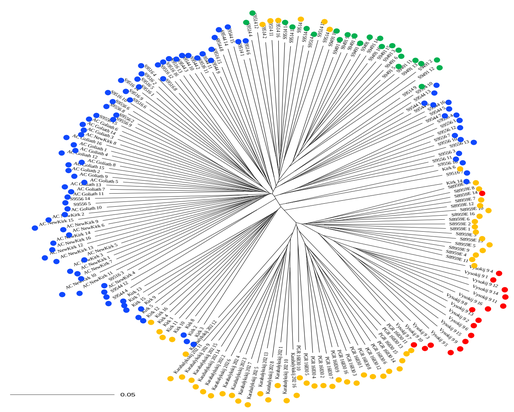

Figure 2. Genetic relationship of 192 genotypes of the 12 crested wheatgrass accessions as revealed by neighbor-joining clustering with the 45,507 SNP markers. Each genotype is numbered after its accession label. Each node for a genotype is represented with colored circle followed genotype name. Red, green, blue and yellow represents plants in Clusters 1, 2, 3 and 4 inferred from the STRUCTURE analysis (Figure 2A), respectively.

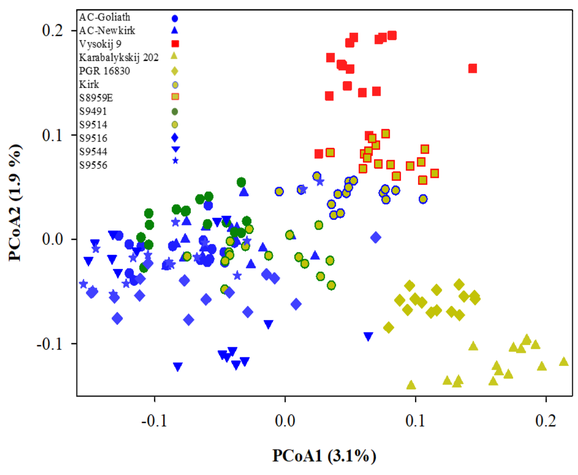

Figure 3. Genetic relationship of 192 *Agropyron cristatum* genotypes is labelled with shape and color of the shape represents the clusters obtained from the STRUCTURE analysis. Red, green, blue and yellow represents Clusters 1, 2, 3 and 4 inferred from the STRUCTURE analysis (Figure 2A), respectively.

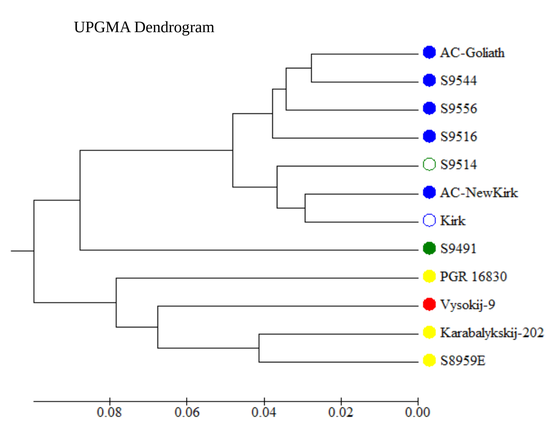

Figure 4. Genetic diversity and genetic relationships of the 12 *Agropyron cristatum* accessions shown by an UPGMA dendrogram based on the Phi statistics obtained from the AMOVA.

RESULTS

Table 2. Results of the analysis of molecular variance for two models of genetic structure (12 accessions and four clusters from the structure analysis) based on 45,507 SNP markers

Source of variation	df	Sum of squares	Variance explained	Variance (%)
12 accessions				
Among accessions	11	101048.8	246.0	15.8
Within accessions	372	488598.0	1313.4	84.2
Four clusters from STRUCTURE				
Among groups	3	54736.5	193.3	12.07
Within groups	380	534910.3	1407.7	87.93

CONCLUSIONS

These results are valuable in crested wheatgrass breeding and encourage the GBS application in non-model plants with complex genetic structure towards characterization of genome-wide genetic variability.

REFERENCES

1. Bolger et al. (2014). *Bioinformatics*, *30*(15), 2114–2120.
2. Dewey, (1984). Plenum Publishing Corp, New York, 209–279.
3. Excoffier & Lischer, (2010). *Ecology Resources*, *10*(3), 564–567.
4. Falush et al. (2007). *Molecular Ecology Notes*, *7*(4), 574–578.
5. Looman & Heinrichs (1973). *Canadian Journal of Plant Science*, *53*, 501–506.
6. Peterson et al. (2014). *Diversity*, *6*(4), 665–680.
7. Pritchard et al. (2000). *Genetics*, *155*, 945–959.
8. R Development Core Team, (2016).Available online: http://ww.r-project.org/
9. Yang & Fu, (2017). *Applications in Plant Sciences*, *5*(7), 1700018.

ACKNOWLEDGEMENTS

We are thankful to Mr. Gregory Peterson and Ms. Carolee Horbach for their technical assistance, and ComputeCanada and Westgrid for providing the high-performance computing service as well as of their technical support. The work was financially supported from Beef Cattle Research Council of Canada and Agriculture and Agri-Food Canada (AAFC) Growing Forward 2 Funds (FGR.08.13).

Contact:
kiran.baral@usask.ca
bruce.coulman@usask.ca