Journal of Cereal Science 65 (2015) 267–276

Contents lists available at ScienceDirect

Journal of Cereal Science

journal homepage: www.elsevier.com/locate/jcs

Identification of Canadian wheat varieties using OpenArray genotyping technology☆

Daniel J. Perry*, Sung-Jong Lee

Canadian Grain Commission, Grain Research Laboratory, 1404-303 Main Street, Winnipeg, Manitoba, R3C 3G8, Canada

ARTICLE INFO

Article history:
Received 27 March 2015
Received in revised form
4 August 2015
Accepted 11 August 2015
Available online 14 August 2015

Keywords:
Wheat
Single nucleotide polymorphism
Variety identification
Cultivar identification

ABSTRACT

Accurate and efficient means to determine the variety composition of wheat composites are required to provide assurances in the grain handling system. We developed a variety identification system for hexaploid wheat (*Triticum aestivum* L.) that operates with a simple workflow and can sustain a high level of throughput. A set of 32 SNP genotyping assays was developed for use with the TaqMan OpenArray genotyping system. A reference profile database was constructed based on a survey of these markers in 24 kernels of each of 128 wheat varieties representing four market classes. Custom software was written to generate allele profiles from exported OpenArray fluorescence data and query those profiles against the reference database to produce variety calls. We demonstrated the effectiveness of the system in blind analyses of 192 kernels comprising 119 varieties. Each kernel was correctly identified with the exception of two that were called as unidentified and were later confirmed to have uncommon variant profiles not previously encountered.

1. Introduction

Canadian wheat is marketed by class; varieties designated within each class share specific functional characteristics such that class is a reliable indicator of the processing and end-use quality. Tolerances for the presence of wheats of other classes, including nonregistered varieties, are specified for each class and grade. Historically, in western Canada the varieties within each wheat class were required to have a characteristic kernel size, shape and appearance, such that kernel visual distinguishability (KVD) could be used to identify and segregate the classes in the grain handling system. However, in 2008 KVD was removed as a requirement for variety registration and in its place a system of producer declarations, supported by monitoring and testing was implemented. New varieties can now be registered regardless of kernel appearance, making it possible to develop a variety intended for one specific class based on end-use functionality, but have kernel characteristics similar to varieties in another. Concurrent with the removal of KVD, a new class, Canada Western General Purpose (CWGP), was introduced to accommodate high-yielding, non-milling wheat varieties intended for feed or ethanol production. The purpose of the CWGP class has evolved somewhat since its inception; recently several varieties were moved into it from the Canada Western Red Winter (CWRW) wheat class. Other changes to the class structure of Canadian wheat have been proposed that could, among other things, see the varieties currently designated as Canada Western Red Spring (CWRS) wheat, the most popular Canadian wheat class, be divided into two separate classes based on gluten strength. In light of developments such as these, accurate and efficient methods to determine variety composition are becoming increasingly essential to support Canada's grain quality assurance system. Such tools may also be useful for other purposes like assessment of seed purity and the protection of plant breeder rights.

Variations of polyacrylamide gel electrophoresis of seed storage proteins have been used for wheat variety identification (VID) for several decades and such methods have been developed as international standards (International Association for Cereal Science and Technology, 1995; International Organization for Standardization, 1993). However, seed proteins have limited variability and, although higher resolution techniques such as capillary electrophoresis and high performance liquid chromatography have also

Abbreviations: CPSR, Canada Prairie Spring Red; CWGP, Canada Western General Purpose; CWRS, Canada Western Red Spring; CWRW, Canada Western Red Winter; KVD, kernel visual distinguishability; SNP, single nucleotide polymorphism; UPGMA, unweighted pair group method with arithmetic mean; VID, variety identification.

☆ Paper no. 1092 of the Grain Research Laboratory, Canadian Grain Commission.

* Corresponding author.
E-mail address: daniel.perry@grainscanada.gc.ca (D.J. Perry).

http://dx.doi.org/10.1016/j.jcs.2015.08.002
0733-5210/

been applied, it is generally recognized that seed storage proteins cannot sufficiently distinguish among large numbers of wheat varieties. This has led others to explore potentially higher resolution DNA-based approaches. An assortment of DNA marker types have been investigated, but microsatellites (also known as simple-sequence repeats or SSR) have been the most widely embraced for wheat VID (Bányai et al., 2006; Fujita et al., 2009; Leigh et al., 2003; Manifesto et al., 2000; Prasad et al., 2000; Röder et al., 2002; Tran-Dinh et al., 2009; Zhen et al., 2014) because they are abundant and highly polymorphic. Microsatellites may provide complete or nearly complete discrimination among varieties; however, owing to the presence of multiple size-differentiated alleles they can require considerable investment of time for data interpretation, even when allele analysis software is used. This limitation is of particular concern in the context of grain handling due to the complexity of samples that may be experienced. For example, there are more than eighty varieties eligible for delivery as CWRS wheat and a CWRS sample may contain any number and combination of those varieties. Clearly, for assessment of variety composition and class eligibility, automation of data analysis should be a primary objective. For this reason our interest turned to SNP-based approaches; although SNP loci generally have lower diversity than microsatellites, their bi-allelic nature can make them much more amenable to automated allele calling.

Many platforms are available for SNP gentoyping and they vary considerably in throughput, both in terms of numbers of samples and multiplexing levels. Some techniques that can be used to genotype a large number of individuals at a moderate number of SNP loci have been investigated in wheat; for example, (Bérard et al., 2009) used SNPlex (Applied Biosystems) to examine a set of 47 SNPs in 1314 lines and Akhunov et al. (2009) demonstrated the effectiveness of the Illumina GoldenGate assay, a genotyping technology now quite widely adopted by the wheat genomics community, by genotyping 91 tetraploid and hexaploid lines of wheat at 96 SNP loci. However, the rather complex workflow of these approaches could be prohibitive for implementation in a routine testing environment where a continuous throughput of large numbers of samples and an expectation of prompt turnaround is likely. Our goal was to develop a VID system for hexaploid wheat (*Triticum aestivum* L.) that operates with a simple workflow and can sustain a high level of throughput. We achieved this objective using TaqMan OpenArray genotyping technology (Applied Biosystems, Life Technologies) in combination with custom data analysis software.

2. Materials and methods

2.1. OpenArray procedures and instrumentation

TaqMan OpenArray genotyping technology utilizes a microscope slide-sized stainless-steel plate that has 3072 through-holes arranged in 48 subarrays of 64 through-holes each. Each through-hole acts as a reaction chamber for a TaqMan genotyping assay. Plates are purchased with assays preloaded and dried down, enabling a simple genotyping workflow that consists of loading DNA extracts combined with a universal master mix onto OpenArray plates, thermal cycling and imaging. Instrumentation available for OpenArray analyses evolved over the course of the work presented here. Initially we used OpenArray NT instrumentation (Applied Biosystems, Life Technologies), for which 2 μL of each DNA extract was mixed with 2 μL of TaqMan OpenArray Genotyping Master Mix in a 384-well plate and loaded onto Format 32 TaqMan OpenArray Genotyping Plates using an OpenArray NT Autoloader (Format 32 plates contain one set of 32 assays replicated 96 times). Thermal cycling was performed in a GeneAmp 9700 thermal cycler equipped with a dual flat block, followed by endpoint fluorescence detection in an OpenArray NT Imager.

Later we transitioned to QuantStudio 12K Flex instrumentation (Applied Biosystems, Life Technologies). With this system, the combined DNA and master mix was loaded onto Format 32 OpenArray plates using the QuantStudio 12K Flex OpenArray AccuFill System. Thermal cycling and fluorescence detection occurred in a QuantStudio 12K Flex Real-Time PCR instrument fitted with an OpenArray block. To simplify run initiation, an application was developed in Visual Basic (Microsoft Visual Studio 2010) to accept entry of sample identifiers and the numbers of kernels per sample and automatically map the kernels to the appropriate through-holes for integration with plate information files by the QuantStudio 12K Flex software.

2.2. SNP discovery and marker selection

Primer pairs for genome-specific amplification of sequencing templates were selected from the wheat SNP database (http://probes.pw.usda.gov:8080/snpworld/Search) described by Akhunov et al. (2010). Preference was given to primers targeting 500 to 1000 bp products with multiple SNPs reported. A small number of primer pairs for other loci of interest were also designed based upon published sequences. Sequencing templates were amplified from DNA extracted from single kernels using an SDS procedure (Perry, 2004). Aliquots were examined via agarose gel electrophoresis and those appearing to consist of single amplicons were column purified (Qiagen, MinElute 96 UF PCR Purification Kit or QIAquick PCR Purification Kit) and submitted for Sanger sequencing at the National Research Council of Canada (NRC) Plant Biotechnology Institute (Saskatoon, Canada).

In a first round of sequencing, 155 targets, selected almost exclusively from the A and B wheat genomes, were examined in a diverse panel of eight varieties, comprising one representative from each western Canadian milling wheat class. Sequences were aligned using ClustalX 2.1 (Larkin et al. 2007) and TaqMan SNP genotyping assays were designed for one or two SNPs or small indels per polymorphic alignment using Primer Express 2.0 (Applied Biosystems), making an effort to include the original genome-specific primer when practical. The assays were evaluated through analysis of eight individual kernels from each of 132 varieties from various sources using NT instrumentation; fluorescence plots were examined using OpenArray SNP Genotyping Analysis Software Version 1.0.3 (Biotrove, Woburn, MA). A first set of OpenArray plates contained assays targeting 32 polymorphisms; second and third sets of plates contained assays targeting a total of 55 additional polymorphisms as well as a small number of redesigned assays targeting polymorphisms represented on earlier plates.

A second round of sequencing was carried out in search of additional polymorphism; 133 additional target sequences were examined in a panel that included several combinations of varieties that were not well resolved by assays developed from first round sequences. Unlike the first round of sequencing, this second round had a strong representation of D-genome targets. A set of 32 assays designed to target 28 of these second round polymorphisms were examined in OpenArray analyses in a manner similar to above.

Following assessment of all assays, 32 were chosen to be included in a wheat VID marker set. Selection was based upon individual genotyping performance (strength of fluorescence signal and separation of VIC and FAM groups in scatter plots) and upon combined ability to differentiate among varieties.

2.3. Characterization of polymorphism within and among varieties

OpenArray analysis of the selected markers was performed on