E

BMT-TWA/Wheat/1/2
ORIGINAL: English
DATE: February 18, 2001

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

AD HOC CROP SUBROUP ON MOLECULAR TECHNIQUES FOR WHEAT

First Session
Cambridge, United Kingdom, February 26 to 28, 2001

DISCOVERY AND IMPLEMENTATION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN WHEAT VARIETY IDENTIFICATION

Document prepared by experts from Canada

Discovery and implementation of single nucleotide polymorphisms in wheat variety identification.

DARYL J. SOMERS

Agriculture and Agri-Food Canada, Cereal Research Centre, 195 Dafoe Rd., Winnipeg, MB, Canada.

INTRODUCTION

The following discussion paper is focused on the development, implementation and impact of a novel DNA fingerprinting strategy for variety identification. The paper considers the questions raised in the work program of the crop subgroups on molecular techniques.

DNA fingerprinting of many species of plants and cultivated varieties has a long scientific history. When DNA profiling technology first came into use, restriction fragment length polymorphisms (RFLPs) were considered state-of-the-art. RFLP technology was followed by random amplification of polymorphic DNA (RAPD), followed by amplified length polymorphisms (AFLP) and most recently we use microsatellite markers. Each of these technologies is very useful and there have been numerous alternative methods of DNA fingerprinting where these technologies form the foundation. Today, DNA fingerprinting generally relies on 1) differences in restriction enzyme digestion sites 2) random differences in primer annealing sites or 3) differences in numbers of tandem repeats of DNA. In all cases, the molecular polymorphism can be described as a difference in DNA sequence, and researchers use different methods (restriction digestion, and PCR) to visualize the sequence differences.

In the late 1990s and now in the new millenium, the age of genomics research has moved to the foreground of many plant molecular biology research programs, and this meeting in Cambridge is timely in that we can consider using genomics-based approaches for variety identification. The primary genomics resource available to most researches is sequences of ‘expressed sequence tags’ (ESTs). These sequences represent the expressed portion of the genome and are derived by first extracting messenger RNA and converting this into DNA strands that are easily sequenced. The most prominent source of publically available EST sequences is from the NSF-funded, USDA project in Albany CA, USA which proposes greater than 100,000 EST sequences. In addition, the International Triticea EST Constortium (ITEC) was formed in July of 1998 which proposed approximately 40,000 EST sequences. By the end of year 2001, there will be an estimated 100,000 wheat EST sequences available in the public domain and many more will be available shortly after that. As I stated above, the foundation of all molecular DNA fingerprinting is in DNA sequences, so the availability of so many DNA sequences in the public domain presents a remarkable opportunity to extract DNA fingerprinting technology.

The EST sequences represent ‘exon’ sequences from functional, expressed wheat genes, introns have been removed by cellular processing. Since these sequences are from coding DNA, there are naturally no gross polymorphisms in the sequences, between varieties, but there is an abundance of single base substitutions and deletions, collectively referred to as single nucleotide polymorphisms (SNPs). The SNPs can act as molecular polymorphisms and be used to distinguish varieties from each other. Although there are only four possible alleles

at any SNP locus (A, G, C, T), the shear abundance of SNPs provides enough useful molecular markers to distinguish varieties of wheat.

Wheat is a polyploid species with three different genomes (A, B, D) and DNA fingerprinting of varieties with any of the conventional marker technologies has always compared polymorphisms from identical loci between varieties. Previous marker technology accommodated the polyploid nature of wheat by virtue of the fact that homoeologous loci (ie. from different genomes) have differences in molecular length which is resolved by RFLP and PCR-based assays. In contrast, homoeologous EST sequences do not differ in molecular length, but rather differ largely by SNPs. Therefore, inorder to distinguish varieties with SNP technology, we must first derive a method to identify loci between varieties without detecting highly similar loci/sequences from homoeologous genomes. This is accomplished by 'mining' SNPs from EST databases and developing PCR-based strategies whereby, SNPs between homoeologous genomes and SNPs between varieties are distinguishable.

Microarray technology is the next generation of detection platform for DNA polymorphism. A microarray is a matrix of small DNA molecules, commonly on a silica chip. Each DNA molecule is capable of distinguishing allelic differences between identical loci in wheat varieties. This detection platform is amenable to high throughput, reproducibility and repeatability. Microarray technology is developing very quickly and we should consider that detection platforms will be substantially different at year end, than they are now. But, the fundamental, primary biological information that is used by microarray technology is the SNP. Therefore, a research program toward SNP discovery and demonstration of its use in varietal identification is valid today and will always be valid as long as microarray detection platforms develop and are in use.

ADDRESSING THE WORK PROGRAM

Molecular methods available and suitability for use

1a There are only 4 possible alleles at each SNP locus. This may seem low, but the level of polymorphism is enhanced by the opportunity to interrogate large numbers of SNP loci simultaneously. As with any marker technology, allele frequency will determine the usefulness and informativeness of the marker.

1b The availability of the public domain markers is high, and will increase as more EST sequences become available from the USDA and ITEC. Also, 'inhouse' EST sequencing projects in various research labs is becoming very common, which further enhances the availability of SNP markers.

1c The accessibility of current SNPs is from public domain databases, where there are not any restrictions to the use of the database and its contents. Detection platforms are largely patented and it is suggested that a detection platform be selected, and purchase of detection regents or services will include license fees. All detection platforms I am aware of, do not use radioisotope, typically it is fluorescence. I would estimate detection hardware, readers will cost $50,000 to $200,000 USD. As microarray and SNP technology is adopted by more plant and animal research programs, the costs of SNP technology will be reduced substantially due to competition in the marketplace to deliver reliable, innovative detection platforms.

1d Repeatability is largely untested at this time. SNP detection begins with genomic DNA and uses a very high stringency PCR-based assay. Therefore, presumably, the repeatability should be quite good.

1e Agreement of results between labs is untested at this time.

1f As SNP detection and detection platforms are further developed, the technology will undoubtedly become more repeatable and consistent. In addition, SNP discovery will eventually, over time, identify a very useful set of SNPs to be used in variety identification. The initial panel of SNPs used may not be the best.

DUS issues

Distinctness

2a This is untested at this time. The use of SNPs will be no different from existing molecular diversity calculations. SNPs will have certain allele frequencies among a panel of wheat varieties which will be used to calculate variability between varieties and genetic distance. We fully anticipate, there will be little difficulty in distinguishing between closely related varieties, again based on the large number of SNP identifiers available.

Uniformity

2b SNP technology is expected to have a very high level of resolution, in that it will be capable of detecting intra-varietal genetic variation. Thus uniformity can be determined in certain seed lots. The SNP technology is flexible in that data derived from SNP detection could be used in multivariate analysis, which will permit changes in uniformity and still be used for distinctness calculations.

Stability

2c DNA stability is very important. Since the SNPs are derived from coding DNA, this DNA is far less likely to undergo changes over time. Also, since only single base differences are assayed, the chance of natural DNA sequence errors arising from one generation to the next are very remote as opposed to changes in microsatellite lengths or changes in RFLP allele sizes.

Influence of different methods on levels of DUS

2d I cannot comment on this issue, the SNP technology has not progressed to the point of detailed examination of marker sets. Although, no impact on DUS of marker sets is anticipated.

Relationships to phenotype

3a The SNPs acts as a typical molecular marker and thus can be used to construct conventional genetic maps, which links SNP alleles to phenotypic characteristics.

3b SNP technology has not progressed far enough to draw relationships between traditional characteristics and SNP data.

3c The potential to link SNPs to phenotypic data is great. The shear number of SNPs will enable researchers to assay the entire genome, particularly in regions with genes since the SNP are gene sequences themselves. In addition, SNP technology enables one to compare allelic sequences and look for differences in function. Thus, the potential to link allelic differences to phenotype is possible.

3d Comparing molecular and phenotypic distances is untested at this time using SNP technology.

Potential applications

4a SNPs, as will most molecular markers, will give an accurate description of the genotype of a wheat variety and will also measure the intra-variety variation. Thus, SNPs are a good fit for describing DUS.

4b SNPs can assess reference collections. As with any molecular marker technology and its use in describing DUS, a reference collection may be essential to be able to calibrate the molecular system periodically.

4c We expect the SNP technology to provide very high resolution of variety identification and thus if guidelines are set that describe the minimum molecular genetic distance to be 'essentially derived', then SNPs should be able to detect this genetic distance.

4d The sequences of SNPs within any variety can be determined and documented and thus easily incorporated into variety descriptions if the variety registration policies required this sort of data.

Possible impacts of the introduction and unsolved problems

5a The development of SNP technology in wheat, in fact plants, is very new. There are many areas of research to be completed, many of the issues described above were un-answerable since certain experiments have simply not been completed. So at this time, the implementation of SNP technology in variety identification is not advised. We expect to have completed a full analysis of this technology in Canadian labs by late 2003.

5b If the technology were fully developed, variety identification will be available in a high throughput format and easily accessible to many research labs and grain handling industries. Experts in statistical sciences, population genetics, and molecular genetics will have to clearly outline the minimum genetic distances permitted and proceed with caution as the molecular technology is calibrated and tested for a particular system.

[End of document]