High-Throughput Single Nucleotide Polymorphism (SNP) Discovery and Marker Validation in *Brassica napus*

Michelle Wiggins[1], Shunxue Tang[1], Yonghe Bai[1], Fang Lu[1], Carol Powers[1], Fabiano Pita[1], Lasantha Ubayasena[2], Zoe Ehlert[2], Tom Kubik[2], Gregory Gingera[2], Christof Stoll[3], Van Ripley[2], Thomas Greene[1], Steve Thompson[1], Siva Kumpatla[1]

[1] Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268, USA

[2] Dow AgroSciences LLC, 101-421 Downey Road, Saskatoon, SK S7N-4LB, Canada

[3] Dow AgroSciences LLC, Herberner Straße 43a, D-48268, Greven, Germany

Introduction

Single nucleotide polymorphisms (SNPs) are the most useful markers for genetic diversity study, fine mapping, association mapping and molecular breeding because of their abundance in the genome and amenability for high throughput genotyping. Oilseed rape (*Brassica napus* L., AACC, 2n = 4x = 38), an allotetraploid formed from the diploid *B*. *rapa* (AA, 2n = 2x = 20) and *B*. *oleracea* (CC, 2n = 2x = 18), is one of the most important vegetable oilseed crops in the world. However, development of SNP markers, genomics tools and resources for molecular breeding in oilseed rape is very challenging because of the highly duplicated nature of its genome.

Through the joint efforts of Agriculture & Agri-Food Canada (AAFC), DNA LandMarks (DLM), Dow AgroSciences (DAS) and 11 other industrial partners participating in the *Brassic*a SNP Discovery Consortium, a large number of SNPs have been identified via *in silico* SNP discovery using existing *B*. *napus* expressed sequence tags (ESTs) (Chatel et al. 2008) or 454 sequencing of *B*. *napus* transcripts and gene families (Sidebottom et al. 2010), and released to partners in the Consortium. In this poster, we describe the selection of 1,536 SNPs for the creation of an Illumina OPA, validation of these SNPs through the screening of 235 winter and spring oilseed rape lines using Illumina GoldenGate SNP assay, and the use of informative SNPs to investigate the genetic relationships among DAS oilseed rape lines.

Materials and Methods

More than 21,000 *in silico* SNPs were identified from nearly 500,000 ESTs from DLM and AAFC, and 3,000 putative SNPs with high confidence were selected to design a 1,536-plex Illumina OPA. The 1,536 SNPs in the OPA were validated with 235 winter and spring oilseed rape lines from DAS using the GoldenGate SNP assay on the Illumina BeadStation.

Heterozygosities or polymorphic information content scores of SNP markers were calculated according to Ott (1991). Pairwise genetic distances among the 235 lines were estimated using the 'proportion of shared alleles' estimator in MicroSat. The UPGM trees were constructed using command 'neighbor.exe' in Phylip 3.69 (Felsenstein 2005) and visualized with Dendroscope (Huson et al. 2007).

Results and Conclusions

• Illumina GoldenGate chemistry is an efficient technology for high-throughput SNP validation even in polyploid species like *B*. *napus*. More than 90% of the SNP markers produced good SNP clustering patterns. Based on the clustering patterns of the genotypes of inbred or double haploid lines, four types of SNPs were identified, which included 3 types of intragenomic SNPs caused by allelic variations and the intergenomic SNPs caused by polymorphisms between paralogs (Fig. 1).

• The attrition rate in *B*. *napus* SNP marker development was high. Only 42% of the 1,536 SNPs were intragenomic SNPs, which are useful for DNA fingerprinting or genetic mapping (Fig. 2).

• The heterozygosities of the 569 polymorphic SNP markers ranged from 0.02 to 0.50 with an average of 0.36 (Fig. 3).

• The genetic distances ranged from 0.01 to 0.49 with an average of 0.31 between spring lines, and ranged from 0.01 to 0.50 with an average of 0.28 between winter lines (Fig. 4).

• The UPGMA (unweighted pair group method with arithmetic mean) trees showed 7 clades and 5 clades in the spring and winter lines, respectively (Fig. 4). The genetic relationships among the inbred lines can help in the selection of cross combinations in the hybrid breeding program in oilseed rape.

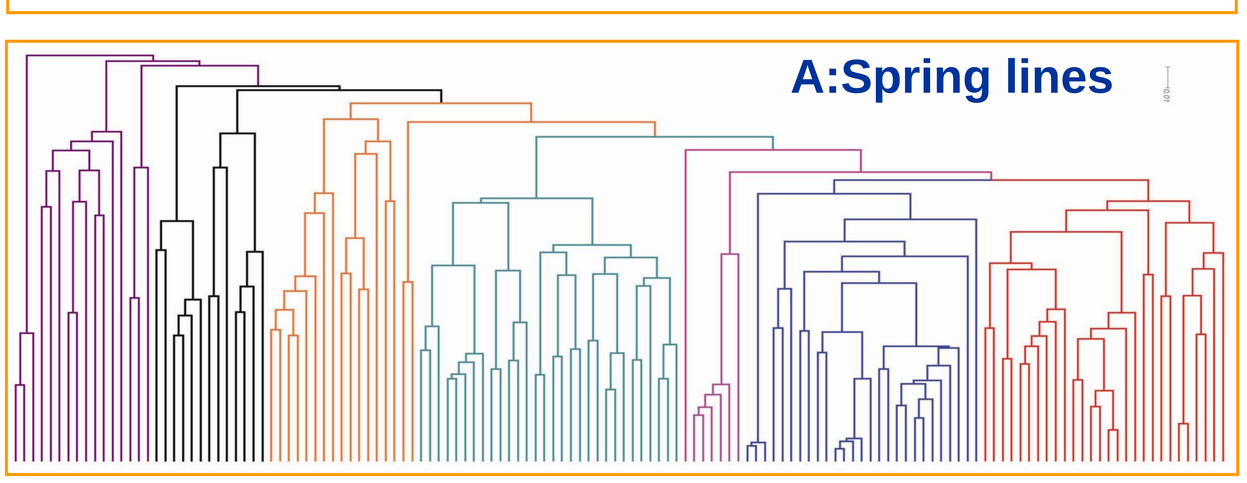


Figure 4. UPGMA trees produced by genetic distances estimated from 569 polymorphic SNP markers. Seven clades were observed in the spring lines (A) and 5 clades were observed in the winter lines (B), respectively. Different colors were used to highlight individual clades.

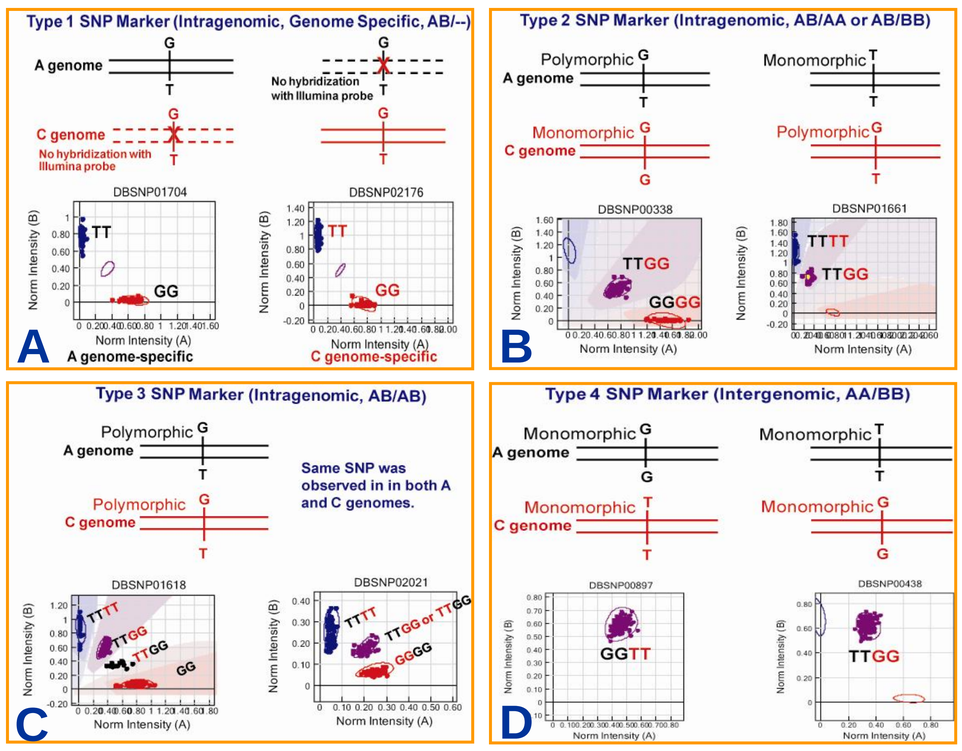
Figure 1. Categories of SNP markers.

A: Type 1 SNPs (intragenomic, genome-specific)

B: Type 2 SNP (intragenomic, hemi-SNP)

C: Type 3 SNP (intragenomic, AB/AB)

D: Type 4 (intergenomic, non-allelic)



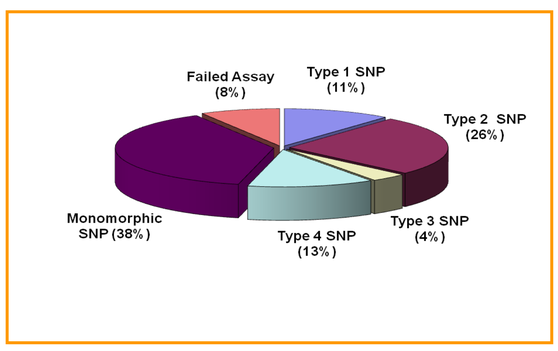


Figure 2. The validation and quality of the 1,536 SNP markers selected for Illumina GoldenGate assay.

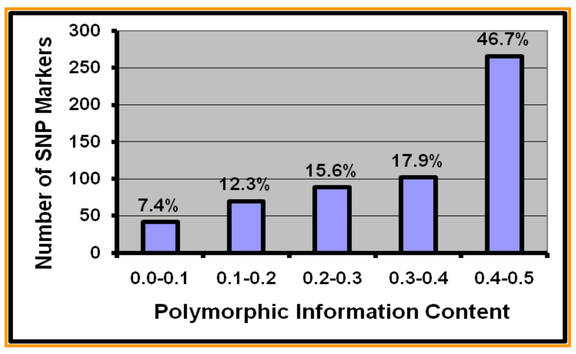


Figure 3. The PIC scores for 569 polymorphic SNP markers genotyped on 235 winter and spring oilseed rape lines.

References

Chatel G, Guigou G, Kalogerakis M, Dayton L, Guo Y, Cheung WY (2008) *Brassica* SNP Discovery And Validation. Plant & Animal Genome Conference XVI, San Diego, CA, USA. Jan 12-16, 2008

Sidebottom C, Clarke W, Li R, Higgins E, Links M, Parkin I, and Sharpe A (2010) SNP Discovery In *Brassica napus* Using 454 Titanium Sequencing. Plant & Animal Genome Conference XVIII, San Diego, CA, USA. Jan 9-13, 2010.

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M and Rupp R (2007) Dendroscope- An interactive viewer for large phylogenetic trees, BMC Bioinformatics 8:460

Felsenstein J (2005) PHYLIP (http://evolution.genetics.washington.edu/phylip.html).

Microsat: http://hpgl.stanford.edu/projects/microsat/microsat.html.

Ott J (1991). Analysis of human genetic linkage. Johns Hopkins University Press. Baltimore, Maryland