Developing and Mapping Single Nucleotide Polymorphisms (SNPs) from the Red Raspberry Genome Sequence

Progress Report submitted to the North American Raspberry and Blackberry Association Research Committee regarding 2010 Funding

Prepared and submitted by: Dr. Courtney Weber, Associate Professor, and Judson Ward, Ph.D. Candidate, Department of Horticulture, Cornell University, 630 W. North St.-NYSAES, Geneva, NY 14456

The objective of the project is to develop reliable sequence based DNA markers based on single nucleotide polymorphisms (SNPs or "snips") between different raspberry genotypes. These markers have utility in breeding new varieties, disease resistance screening and functional genomics as well as other genetic studies in crop species. Potential SNPs markers have been identified throughout the red raspberry genome and specific primers developed for a subset of these SNPs to demonstrate their potential in genetic mapping of red raspberry. A set of reference SNP markers distributed throughout the raspberry genome is being developed. They will act as reference markers for other researchers to compare linkage maps and as a beginning towards more detailed mapping studies of specific areas of the genome associated with traits of interest such as disease resistance, fruit quality parameters and yield components. The specific objectives of the project are listed below.

Objectives:

1) Analyze the complete raspberry genome sequence and transcriptome sequences from BYU and Cornell for potential SNPs throughout the genome.

2) Design specific primers for thirty-five, approximately evenly spaced, SNPs (5 on each of the seven chromosomes/linkage groups) for use as reference SNPs for the raspberry genome.

3) Test and map the new markers on a population segregating for resistance to raspberry bushy dwarf virus that is currently being analyzed at Cornell for markers associated with resistance.

Over the past year we have sequences significant portions of the genome of the red raspberry varieties 'Titan', 'Latham' and 'Heritage'. We sequenced the genomic DNA of the 'Heritage' variety using two different sequencing techniques (Illumina and 454) to a coverage depth of approximately 150 times. This sequence data is currently being assembled into a complete genome. In addition, we have sequenced the transcriptomes (the sum of expressed genes from a specific tissue at a specific time point) from the roots of the 'Latham' and 'Titan' varieties in order to study the genes expressed in the roots in response to attack by *Phytophthora fragariae* var. *rubi*.

We compared the DNA sequence from the three genotypes to identify possible SNPs throughout the entire genome. Utilizing the transcriptome data from

'Titan' and 'Latham' ensures that the majority of these SNPs originated within actual genes making them very useful for future studies on gene action and genome structure. The initial comparison yielded 137,167 potential SNPs between 'Titan' and 'Heritage' and 157,148 SNPs between 'Latham' and 'Heritage'.

The 'Heritage' genomic sequence was aligned to unpublished genome data from the strawberry *Fragaria vesca.* This provided a scaffold of the genome of raspberry so that markers could be placed relatively uniformly across the genome. We filtered the potential markers to identify a subset of the sequences with the highest probability for being true SNPs (as opposed to an artifact of sequencing) and spaced throughout the raspberry genome. This yielded 208 potential markers to be analyzed for primer design and marker testing.

The primary method being used for developing SNP markers into specific PCR based markers in this project is by using restriction enzymes that recognize the specific sequence containing the SNP to cut the DNA fragment into 2 smaller pieces. These cleaved amplified polymorphic sequence (CAPS) markers can be visualized on an agarose gel thus allowing the researcher to identify individuals containing the SNP sequence. This conversion of the SNP into a marker is very useful for use across different populations and germplasm sources because they are very sequence specific and co-dominant in nature thus providing data on heterozygosity.

Initially, we designed 50 primer sets for single or paired SNPs that also are located within a restriction site. The 50 primer sets are being tested for amplification and transference to the test population WSU1499 x 'Newburgh'. Once fully screened, the markers will be compared to data on resistance to raspberry bushy dwarf virus (RBDV) to further isolate the region of the genome where the resistance gene is located. This process in ongoing and will be completed by the term date of the project, February 28, 2011.