**FINAL PROJECT REPORT**

**Project Title:** Apple genome project

**PI:** Amit Dhingra
**Organization:** WSU
**Telephone/email:** 509-335-3625
adhingra@wsu.edu
**Address:** PO Box 646414.
**Address 2:** 149 Johnson Hall
**City:** Pullman
**State/Zip:** WA 99164

**Cooperators:** Ananth Kalyanaraman, WSU; Charles-Eric_Durel, INRA at Angers, France, Riccardo Velasco, IASMA, Italy; Jasper Rees, University of Western Cape, South Africa; Gale Bozzo, University of Guelph, Canada; Tim Harkins, Roche Inc; Dan Rokhsar, Joint Genome Institute, Kate Evans, WSU.

**Other funding sources**

**Agency Name**: USDA - NRI
**Amount awarded**: $ 224,000
**Notes**: Supplemental funding provided by USDA for scaffold sequencing complementing the objectives of this project.

**Agency Name**: IASMA, Italy
**Amount awarded**: $ 30,000
**Notes**: These funds were dedicated towards generating sequence from DH apple at IASMA using the BAC library constructed with funds from USDA-NRI project. The data generated will also help in generating the scaffold.

**Total Project Funding**: 37,750

**Budget History:**

| **Item** | **2009** | | |
|---|---|---|---|
| **Salaries** | | | |
| **Benefits** | | | |
| **Wages** | 5000 | | |
| **Benefits** | 485 | | |
| **Equipment** | | | |
| **Supplies** | 30,265 | | |
| **Travel** | 2000 | | |
| | | | |
| | | | |
| | | | |
| **Miscellaneous** | | | |
| **Total** | **$37,750** | | |

**Note: Funding was approved for one year enabling us to focus on one objective.**

**ORIGINAL OBJECTIVE**

1. Generate additional genome sequence information to increase the depth of coverage of the apple genome.

**SIGNIFICANT FINDINGS**

The basic purpose of this project was to ensure efficient and complete assembly of the Double Haploid Apple genome sequence. Prior to the start of this project our program along with the collaborators had acquired sequence data randomly from the DH apple genome. At the same time USDA funding allowed us to construct a library of DNA fragments representing the DH apple genome. We were granted supplemental funding from USDA-NRI program to modify our approach of sequencing the ends of the DNA fragments to establish a new method of scaffold sequencing. The funds provided by USDA and this project was utilized for acquiring sequence information using our novel method. The objective of generating the information is 80% complete. We are currently in the process of refining the computational methods to integrate random and scaffold sequencing data for building a complete assembly of the apple genome. Thereafter this dataset will be compared with the apple genome sequence completed at IASMA with whom we continue to actively collaborate. We have provided the DH Apple genome DNA library to IASMA for generating sequences from the ends of DNA fragments. All this data will aid in our final goal of assembling the DH apple genome. They have committed their own funds to generate this information as part of our ongoing collaboration.

**RESULTS & DISCUSSION**

Sequence information can be rapidly utilized for developing molecular markers for the apple improvement program. It can also provide complete sequence information for genes where we only have partial information. Over the last year we have utilized the preliminary assemblies for mining such information for various colleagues at WSU. Most importantly this information has been the basis of identifying the complete coordinates and sequence for the putative bitter pit-controlling gene in apple that we identified in another project funded by WTFRC. Knowledge of genes underlying important traits can also serve as targets for improving existing varieties using controlled sports induction (CSI) using non-transgenic approaches. We have a continued emphasis on refining the CSI approach in our program to improve existing varieties thereby circumventing the marketing and retail shelf space issues.

The significance of this information will far outlive the duration of this project. Each economically important trait or desirable quality in the fruit tree is controlled at some level by genes. Availability of the genomic blueprint of apple enables us to pin point what gene or group of genes are responsible for

such traits. This information will guide apple improvement year after year from now on. Another testimony to this fact is that scientists have now discovered the gene underlying skin and lung cancer in humans utilizing the human genome information. As in case of humans, the potential economic benefits to the industry are apparent. With the apple genome sequence in hand, we can develop unique varieties for the PNW combining all priority traits that can create unique economic opportunities ranging from production to post-harvest stages.

BROADER IMPACTS

*Presentations*: The apple genome information has been highlighted at several forums over the last year including WSHA meetings. In 2009, the PI was invited to speak at the Hort Show about Enabling Economic Resilience through Genomics Research. Besides that the work has been shown as poster presentations at annual international meetings like American Society of Plant Biology and Plant and Animal Genome Meeting.

*Publications*: The data generated from WTFRC, WSU and USDA-supported DH apple genome has been integrated with the sequence information generated at IASMA and the seminal paper describing the results has been provisionally accepted at Nature Genetics.

*Research*: The apple genome sequencing project has enabled us to now sequence pear and cherry genomes. We are also a part of the strawberry and peach genome project consortia.

*Training opportunities*: This project has been steered by graduate student Scott Schaeffer who is independently supported by an NIH fellowship. We have graduated a computer science student Vandhana Krishnan who utilized the apple genome data for her MS thesis. A high school senior utilized the apple genome data for her senior project and has been accepted at MIT for higher studies.

**EXECUTIVE SUMMARY**

Significant progress: The objective of generating additional sequence information has been accomplished. We have devised a new method of generating far more useful information using a scaffold-sequence approach. At present we continue to refine the computational methods for creating complete and efficient apple genome assembly. It is a reiterative process owing to the computational constraints that involves testing different parameters to arrive at the best possible assembly.

Outcomes and summary of finding: Preliminary DH apple genome assemblies are available that are being used by our program to identify coordinates and sequence information of important genes linked to desirable traits. In summary this is just the start of the most efficient way of connecting traits to genes, an emphasis of our genomics program.

Future directions: We have two proposals under review at NSF and others at various stages of writing to build upon this foundational information. Our programmatic approach is to connect traits with genes using function information and the future projects are aimed at doing just that.