Progress Report

David Zaitlin, Ph.D.
November 4, 2014

Project Title: "The DNA Sequence of the Burley Tobacco Genome"

Investigators: David Zaitlin[1], Robert Miller[1,2,3], and Ling Yuan[1,2]

[1]Kentucky Tobacco Research and Development Center and [2]Department of Plant & Soil Sciences, University of Kentucky, and [3]KY-TN Tobacco Improvement Initiative (KTTII)

Funding Period: April 1, 2014 – March 31, 2015

Summary

The goal of this project is to produce a usable, high-quality draft genome sequence of an elite cultivar of burley tobacco. This sequence will be an invaluable resource for the development of molecular markers linked to important disease resistance genes, many of which originated in other species of *Nicotiana*. Burley inbred line TKF1112B, a product of the KTTII breeding program, is used as a parental line in high-yielding commercial burley hybrids. TKF1112B is resistant to blue mold (from *N. debneyi*), black shank Race 0 (from *N. plumbaginifolia*), wildfire (from *N. longiflora*), black root rot (from *N. debneyi*), TMV (from *N. glutinosa*), and PVY, TEV, *etc*. It is often difficult to develop molecular markers for genes that originated in other species of *Nicotiana*; the chromosomal regions transferred into the tobacco genome are of unknown size and location, and because marker loci within these regions tend to segregate together, there is no way to know how close any of them are to a given disease resistance gene. Having the genome sequence of a burley tobacco line that carries many of these "alien" disease resistance introgressions will allow KTRDC researchers to identify the exotic chromosomal segments and to rapidly develop markers that are closely linked to the specific genes of interest. These molecular markers can then be used in backcross breeding schemes to move the linked genes into other tobacco lines in the development of new varieties for growers. In addition, the genome sequence of an elite burley line can also be used for gene discovery and comparative genomics, and will be unique to the University of Kentucky.

In the six months since this project was funded, a paper describing the genome sequences of three distinct tobacco cultivars was published by a research group at Philip Morris International[1]. The three tobacco genomes (burley, flue-cured, and oriental) are now available on the Sol Genomics Network and can be searched and downloaded, apparently without restrictions.

Progress to Date

As explained in the original proposal, the total funding required for this project is ~$64,000. The money generously provided by the Council for Burley Tobacco will be used to generate Illumina genomic DNA sequence data at the Weill Cornell Medical College Genomics Resources Core Facility. We have chosen to seek additional funding from other sources before initiating the project. This decision was made in part because the plant bioinformatician, Dr. Aureliano Bombarely, recently took a faculty position at Virginia Tech and is in the process of establishing his laboratory and computer infrastructure. Fortunately, we received a verbal commitment this week for additional funding from another party, so we anticipate starting the DNA sequencing in early 2015.

The availability of the three PMI tobacco genomes will affect the design of the current project somewhat; rather than perform a *de novo* assembly of the TKF1112B genome, we can use the existing burley sequence (from TN90) as a template on which to align sequencing reads from TKF1112B. We will then take all of the reads that do not align to the template and assemble them into contigs, many of which will represent the chromosomal regions that were transferred into TKF1112B from species such as *N. debneyi* and *N. plumbaginifolia*. We can also take the assembled sequence contigs from TKF1112B and use them in comparisons with the other tobacco genomes to identify SNPs and other types of useful genomic polymorphisms. Having a tobacco genome to use as an assembly template will reduce the amount of DNA sequence data required for our project and will allow us to get a better assembly. We will also include GBS (genotyping-by-sequencing) to generate specific molecular markers and possibly a linkage map of the burley tobacco genome.

[1]Sierro, N., Battey, J.N.D., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., Goepfert, S., Peitsch, M., Ivanov, N.V. (2014) The tobacco genome sequence and its comparison with those of tomato and potato. Nature Communications 5: 3833. doi: 10.1038/ncomms4833