Using small RNA technology to efficiently identify tomato viruses and viroids in mixed-infected field samples

Kai-Shu Ling[1], Rugang Li[1], Patrick, Wechter[1], Alvaro Hernandez[2] and Zhangjun Fei[3].
[1] USDA-ARS, U.S. Vegetable Laboratory, Charleston, SC, [2] W.M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champaign, Urbana, IL, [3] Boyce Thompson Institute, USDA-ARS, Robert W. Holley Center for Agriculture and Health, Cornell University, Ithaca, NY. Kai.Ling@ars.usda.gov

Small interfering RNAs (siRNA) are produced in plants as a defense mechanism against virus or viroid infection. Analysis of a siRNA profile upon virus infection in plants may allow the *de novo* assembly of the viral genome. In the present study, we were interested in developing an efficient sequencing protocol for virus and viroid identification in leaf tissues collected from greenhouse-grown tomato plants with possible virus infection from Arizona (2 samples), California (1 sample) and Mexico (1 sample), respectively. Small RNA (sRNA) libraries were prepared and multiplexed for sequencing with an Illumina GAII system. Each library generated between 5-7 million of sRNA reads, with 2-3 million unique ones. Most of the sRNA reads were 21-24 nt long, with a majority being 24 nt in length. Approximately 90% of sRNAs were tomato-related. After computer-assisted subtraction of tomato sRNAs, the highly enriched viral siRNAs pools were assembled to scaffolds of previously identified *Pepino mosaic virus* (PepMV) and several pospiviroids genomes. For the small *Potato spindle tuber viroid* (PSTVd, 358 nt), a complete genome was assembled using siRNAs generated from the California isolate. siRNAs were also assembled to near completion to target strains of PepMV. Interestingly, two strains (EU and US1) of PepMV, which share only 80% of nucleotide sequence identity, could be clearly differentiated. Upon gap filling and validation through Sanger sequencing and random amplification of cDNA ends (RACE), six full genome sequences of PepMV (~6,400 nt) in three U.S. samples were assembled. Finally, we applied this technology to characterize an unknown etiology of a severe disease with necrotic lesions and chlorosis in infected tomatoes in Mexico. With *de novo* genome assembly and BLAST analyses of contigs against the NCBI protein database (nr), a novel virus with the highest sequence identity only 70% to a potyvirus were revealed. Upon gap filling and RACE, a complete virus genome was obtained. With the successful applications in the identification of various types of viruses and viroids in tomato samples from different origins, the small RNA technology will likely become an efficient and powerful generic tool for a complete virus and viroid analysis in plants.