Resequencing of the *Phytophthora ramorum* Genome to Characterize Genetic Variation and Population Dynamics of the Invasive Pathogen[1]

Jennifer Yuzon,[2] David Rizzo,[2] Mathu Malar C,[2] Sucheta Tripathy,[3] and Takao Kasuga[4]

Abstract

Phytophthora ramorum has spread and diversified throughout California's northwestern coast since its introduction in the 1990s. Tracking the spread of *P. ramorum* and the functional response of the pathogen to the environment is of particular interest to managing the epidemic. Using genetic tools such as microsatellite markers, researchers have learned much of the pathogen's epidemiology by identifying migrational pathways and new introductions. However, higher resolution markers may reveal previously undetected substructure.

Work at the genetic and genomic level is underway to identify markers in the form of single nucleotide polymorphisms (SNPs), not only for a higher resolution of population structure, but to identify genetic responses of the pathogen to its adopted environment. Because SNPs are more common throughout the genome and have a lower likelihood of violating the infinite sites assumption, a genome-wide SNP dataset will help reconstruct the evolutionary past of *P. ramorum*; identify fitness and selection for genes; and possibly associate population subdivision based on geography, climate, or other influences on population dynamics. Structural variants and chromosomal abnormalities are also of interest because such genetic variation can add another layer of evolutionary potential for the rapidly expanding population.

Characterizing genetic variation has been challenging given the current reference genome. As much as 30% of the sequence is missing in the first draft of the *P. ramorum* genome published in 2006. Copy number variation (CNV) analysis and flow cytometry (FCM) revealed that the genome of the reference strain Pr102 was aberrated and unstable. Using such a reference genome can lead to biases for characterizing important genetic variation in extant lineages.

We have chosen the NORS-DUC standard isolate CDFA1418886 and started to revise the reference genome of *P. ramorum*. CDFA1418886 was cultured from the foliar host camellia and is more likely to depict the type of genetic variation seen in the transmissive population of *P. ramorum* in California. Phenotypically, CDFA1418886 is stable and more aggressive compared to Pr102. Genome analyses and FCM confirm that the genome lacks aberration and is more stable.

Using PacBio sequencing and an in-house pipeline for assembly, the genome is larger (~75-80MB) than initially reported (65MB) and reflects the genome size seen in FCM. Further refinement of the reference utilizes the conserved synteny between *P. ramorum* and closely related species.

---

[1] A version of this paper was presented at the Sixth Sudden Oak Death Science Symposium, June 20-23, 2016, San Francisco, California.

[2] University of California, Davis, CA.
[3] CSIR Indian Institute of Chemical Biology, Kolkata, India.
[4] USDA-ARS, Davis, CA.
Corresponding author: jdyuzon@gmail.com.

Work on identifying genetic variants by resequencing nursery and forest isolates from foliar hosts in California is underway. The software BIC-seq is used to call structural variants such as CNV and LOH. SNP caller software are also being tested.