AGRICULTURAL RESEARCH FOUNDATION
INTERIM REPORT
FUNDING CYCLE 2021– 2023

TITLE: Genomic Epidemiology: Revolutionizing Plant Disease Diagnostics for Oregon's Growers

RESEARCH LEADER: Jeff Chang

COOPERATORS: Melodie Putnam

SUMMARY/ABSTRACT:

Plant diseases, along with pests and weeds, present significant impediments to agricultural economic prosperity. The global movement of plant materials has dramatically increased risks of introducing and spreading diseases. The rapidity with which diseases can spread, both naturally and artificially through trade, means plant disease diagnostic laboratories have a critical need to increase their capacity to help stakeholders. Whole genome sequencing (WGS) holds promise in meeting this need and revolutionizing disease diagnostics. WGS reveals all the genetic information of an organism, which can be analyzed in powerful ways to answer questions that cannot be addressed with traditional disease diagnostics. However, implementation is not trivial. Most diagnostic labs lack sufficient computational infrastructure, biocomputing expertise, and instruments for WGS. We are a microbial genomics research group with specialties in genomics of bacterial pathogens of plants and have been collaborating long-term with the Oregon State University (OSU) Plant Clinic. We developed a standard computational workflow and applied this to study hundreds of samples received from the OSU Plant Clinic. With this framework, we are positioned to take the next significant steps towards implementing use of WGS as a standard method in disease diagnostics. This proposal will test novel applications of WGS to greatly accelerate and inform disease diagnostics. Results will have significant benefits to Oregon's agriculture industry.

OBJECTIVES:

Develop culture-independent methods for plant disease diagnostics. **Rationale:** Culturing introduces a bias into diagnostics. Bacteria isolated from diseased plant samples are grown on a limited number of media, which neither support growth of all bacteria nor all plant pathogenic bacteria. **Aim 1** is a method for extracting as much bacterial DNA with as little contamination from plant DNA possible. **Aim 2** is a computational method to assembly genomes from a mixed population and identify features that can be used to identify *Agrobacterium.*

PROCEDURES:

Aim 1: Inoculated samples will be used to develop and evaluate methods. We will use tomato as a surrogate herbaceous host. Plants will be inoculated to produce galls. The galls will be excised from plants and sonicated, which uses sound energy to agitate samples. Total DNA will be prepared from samples. We will vary parameters to identify steps that improve DNA extraction. Parameters include amount of tissue, power and duration of sonication, and different DNA preparation kits and protocols. Rather than using WGS to assess for enrichment and purity, we will use quantitative PCR

(qPCR). This is more cost effective and allows us to process approximately 100 samples in parallel. Briefly, we target very short, but informative, regions of the pathogen and host genome. We then use qPCR to determine the relative amounts of each (compared to pure pathogen and host DNA). This relative comparison will be used to compare effectiveness of DNA extraction methods and their variants.

Aim 2: We will also use inoculated samples as a first test of computational methods. We will use DNA samples prepared in the most optimal way resulting from activities described for Aim 1. Total DNA will be sequenced on the Illumina iSeq100. The short sequencing reads will be processed to remove portions that were added on and remove those that are of poor quality (high probably that they have a large number of errors). We will use existing programs to assemble the sequencing information into larger fragments. This is the most important variable of this aim. We will then use different types of information to determine which assembly program is most effective. First, pathogens, by definition, gain an advantage in causing disease to plants and because of that, replicate faster than non-pathogenic strains present in the same location. Hence, the amount of genetic information that corresponds to the inoculated strain *Agrobacterium* should exceed that of any other bacteria that was present on or in the plant. Thus, we will count the number of sequencing reads that were generated per fragment. Secondly, we can expect that the most highly represented fragments should correspond to the inoculated strain of *Agrobacterium.* We will therefore align fragment sequences to the genome sequence of the inoculated strain of *Agrobacterium* to determine if it is indeed the case. From these tests, we will be able to identify the best computational approach to assemble genome sequences from a mixed population.

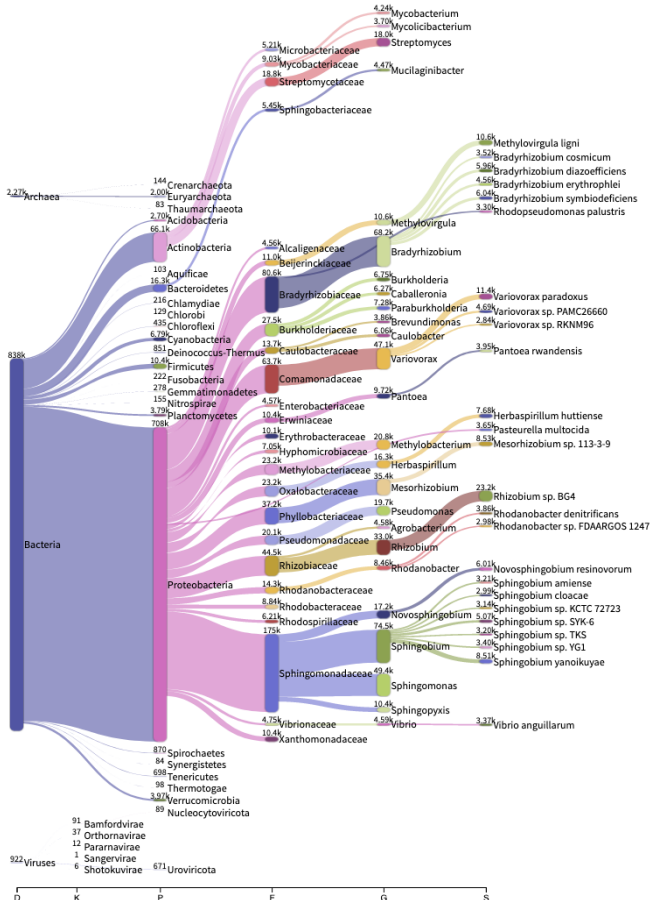
As a final test, we will apply our methods to samples received by the OSU Plant Clinic. Traditional and approaches described herein will be used in parallel, which allows us to crosscheck and validate the efficacy of our methods. In cases where a traditional approach fails to identify *Agrobacterium,* we will mine the genome sequences for marker genes that are indicative of the pathogen (5). We will also mine genome sequences for marker genes that are indicative of susceptibility to the biocontrol product.

SIGNIFICANT ACCOMPLISHMENTS TO DATE:

We purchased and installed an Illumina miniSeq, which is one model up from the iSeq100 we had originally intended to purchase. The miniSeq provides greater flexibility and greater depth of sequencing. Five postdocs/staff have been trained to operate the machine.

We have successfully sequenced 38 bacterial genomes and two metagenome sequences as pilot experiments. For the latter, samples of putative crown galls on blackberry were used to optimize methods for metagenome sequencing and diagnostics. These samples were PCR positive for the *Agrobacterium* virulence marker *virD2*, however the Plant Clinic was unable to culture pathogenic agrobacteria from galls. Total microbial DNA was isolated from the surface of galls by applying a sonicator and beadbeater to the samples, followed by purification using a MoBio Powersoil DNA purification kit. Two samples were included on a sequencing run on the MiniSeq, and the sequencing reads were analyzed for the presence of agrobacteria and key virulence genes/plasmids. Sequencing reads were aligned to reference genomes, and 30 reads were identified that have 100% identity to an Agrobacterium Ti plasmid reference. Additional reads were identified as most similar to genomospecies 4 *Agrobacterium*. The program Kraken 2

was also used to perform a metagenomics community analysis. This pipeline uses k-mer signatures to classify the taxonomy of all sequenced microbial members of the crown gall metagenome. The program Pavian was then used to visualize results (see attached figure). Metagenome community analysis revealed that agrobacteria represent only 0.5% of the sequenced microbial community (~4.5k sequencing reads out of 838k microbial reads). This suggests that *Agrobacterium* may have been present to initiate the gall, but was subsequently outcompeted by other bacteria and is only present at low levels.



We are currently inoculating tomato with a known isolate of pathogenic agrobacterium and will isolate bacteria from galls of various ages and expect to carry out more metagenome sequence analyses of samples received by the Plant Clinic. We will also test if sequencing bacteria en masse isolated from non-selective plates will help identify pathogenic microorganisms.

ADDITIONAL FUNDING RECEIVED DURING PROJECT TERM:

No additional funding has been received since project initiation; but prior and current to this proposal, we had received funding from the ODA to purchase the sequencing machine and from USDA to develop computational methods.

FUTURE FUNDING POSSIBILITIES:

We have begun applying our approaches to more than plant-associated pathogens. In collaboration with colleagues at OSU, we have submitted a pre-proposal to USDA-NIFA-SCRI to study pathogens of honeybees. We are developing a proposal that studies forest microbiome

health; to be submitted to NSF. We are in the early stages of proposal development with colleagues in Food Science and Technology to submit two proposals to USDA-NIFA to track foodborne pathogens. Last, in 2023, we plan on submitting a proposal to NSF to develop a center that predicts and tracks emergence of pathogens.