Background: Animal and plant pathogens are emerging at accelerated rates and on a global scale. New emergences of fungal and fungal-like pathogens in both natural and managed landscapes are often intensified by human activity and result in the most severe population declines and unprecedented extinctions ever witnessed in wild species[1]. Understanding how pathogens of global concern evolve during an outbreak will allow us to better control outbreaks. Tracing an outbreak to its origin is challenging but is routinely done for human pathogens, such as the Ebola virus outbreak of 2014, allowing implementation of quarantines[2]. Famously, the fungal-like pathogen implicated in the Irish Potato Famine, *Phytophthora infestans*, originated in Central Mexico and a series of migrations brought it to Europe in the 1800s[3]. However, the center of origin for *P. ramorum*, an emerging sister species causing Sudden Oak Death (SOD) in US and European forests, is unknown. Since first detection in 1995, SOD has killed millions of oaks and tanoaks and resulted in dramatic financial losses, including bankruptcy, to nursery owners from mandated destruction of infected stock[4]. Further, the disease is expected to cost $135 million before 2020 in California alone[5]. *P. ramorum* and other emerging and evolving pathogens are increasingly important to understand as trade becomes more global and the impact of disease more severe. *I propose to use P. ramorum as a model system to characterize evolution of genome structure and virulence associated with pathogen emergence.*

Aims: **1) *Construct a new P. ramorum reference genome.*** *(In 2019).* The current reference genome of *P. ramorum* assembly has low read coverage (7x) and is fragmented, with half of the ~66Mb genome in scaffolds larger than or equal to 308,042 (N50) and 2500 scaffolds in total. I will improve this reference by assembly with long reads (Pacific BioSciences: >10,000 base pairs), refining resolution in repeat-rich regions and allowing greater structural variant detection[6].

2) *Infer the center of origin and a center of diversity.* *(2020).* All *P. ramorum* outbreaks to date have come from unknown origins and are highly clonal, caused by one of four clonal lineages (Fig. 1). Despite clonal epidemic populations, its genome shows signatures of recent outcrossing, in contrast to evidence of extensive inbreeding in other published *Phytophthora* genomes. This indicates that the original, unknown pathogen source is likely a recombinant center of diversity. Inferences will rely on 77 Eastern Asian *P. ramorum* samples isolated by Dr. Thomas Jung (Mendel University).

3) *Analyze population genomics of natural and epidemic populations.* *(2021).* The natural population collected in Asia may contain private alleles not observed in existing data and ancestral to emergent, clonal populations. Genome sequencing of Asian isolates, in addition to the improved assembly, will allow for population analysis of structural variation (SV) in copy number, ploidy, or indels contributing to pathogen diversity and fitness.

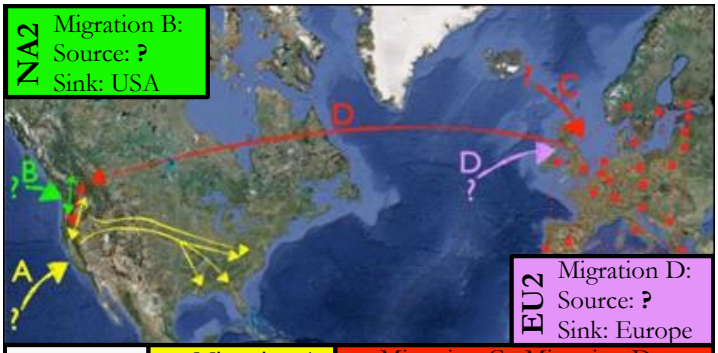

Fig 1. P. ramorum *has emerged repeatedly[4]. Four strains, diverging into four clonal lineages, were defined by SSRs and named by continent and order in which they were discovered.*

Methods: **1)** I will use long-read sequencing (PacBio: Oregon State University) to improve the available genome for the reference strain Pr-102. For genome annotation, I will cross-reference the existing annotations and subsequently improve the prediction of gene models using additional RNA-seq data (Sequence Read Archive; Dr. Richard Hamelin, University of British Columbia).

2) All Asian isolates will be sequenced using HiSeq (OSU). These short reads will be aligned to the new reference assembled in step 1 to identify single nucleotide polymorphisms (SNPs). Phylogeographic analyses will be performed on SNPs in a core set of orthologous genes with high read depth determined to be neutrally evolving by the ratio of synonymous to non-synonymous mutations (dN/dS). I will test the hypothesis that the most recent common ancestor of *P. ramorum* is located in Asia as supported by root state probabilities based on a coalescent analysis[7].

3) I will compare the Asian population with a previously-sequenced clonal population implicated in a Southwest Oregon SOD epidemic (data from UBC). Alignments against the improved reference genome will be used to analyze gene duplications, large insertions/deletions, and ploidy changes. Using SNPs, I will characterize genetic diversity in the two populations with a focus on virulence genes found in *P. ramorum* functional genetics literature. I will reconstruct the ancestral states of these genes at the coalescence of *P. ramorum* and its sister *P. lateralis*[8].

Intellectual Merit: Increasing the N50 and lowering the number of unknown bases (18% Ns) found in the original *P. ramorum* reference genome will result in the discovery of new genes, and more data with modern gene annotation algorithms will improve understanding of existing genes; these are the foundation for future work on genotypic and phenotypic variation and evolution. Long reads will allow characterization of SV, which has never before been studied in *P. ramorum*. Knowledge of gene duplication or deletion, especially in virulence genes which tend to dominate repetitive and gene-sparse regions[9], would be invaluable in understanding genomic architecture of pathogen emergence. My proposed research will be the second *Phytophthora* genome assembled using long reads but the first to investigate associations between SV and pathogen emergence using its center of origin. Determining the genetic diversity of *P. ramorum* in its center of origin will provide new information on the evolutionary dynamics that influence diversification and selection in natural environments (native forests) versus non-native forests or agricultural production (nurseries). In addition, the prediction of ancestral states of adaptive genes (e.g., genes encoding virulence factors) will improve on our knowledge of the molecular mechanisms linked with the evolutionary forces associated with pathogen emergence and increased virulence.

Broader Impacts: Discovery of the center of origin will impact policy, showing the importance of careful screening before shipment of live plants. Further, it will improve the search for natural resistance to use in plant breeding; in a wild population, the pathogen and host have long engaged in an evolutionary arms race. Identification of novel and complex virulence genes in the new reference genome and information about possible allele combinations from a sexual population will assist molecular breeding of resistance. Breeding SOD-resistant varieties of hosts like oaks and Rhododendron would economically benefit the US, saving on management costs.

To widen my impact, I will develop an open-source web application to explore my data in genetic and geographic space on an interactive map. This map, packaged with genomic analyses including phylogenetic trees, principal component analyses, will allow scientists, professional or amateur, to explore spatial and genetic relationships of pathogen emergence for this project as well as others. To broaden participation in STEM, I will work with local, at-risk high school students in our minorities in agriculture club. They will explore my app and provide feedback to improve its accessibility. We will visit SOD-impacted sites with the US Forest Service, who our organization regularly works with, to emphasize the importance of natural studies and connect underrepresented groups with STEM employers.

[1] Fisher et al. 2012. *Nature*. [2] Gire et al. 2014. *Science*. [3] Goss et al. 2014. *PNAS*. [4] Grunwald et al. 2012. *TIM*. [5] Kovacs et al. 2011. *J. Environ. Manage.* [6] Pollard et al. 2018. *Hum. Mol. Genet.* [7] Lemey et al. 2009. *PLoS Comput. Biol.* [8] Quinn et al. 2013. *FEMS Microbiol. Lett.* [9] Raffaele & Kamoun. 2012. *Nat. Rev. Microbiol.*