Zolton Bair
PhD Student
Dept. of Botany and Plant Pathology
Oregon State University
Corvallis, OR 97330

Research Progress Update: IDENTIFICATION OF BLISTER RUST RESISTANCE GENES IN WHITEBARK PINE TO FACILITATE BREEDING AND RESTORATION

(2013 WPEF Student Research Grant)

Given the worsening outlook for the imperiled whitebark pine and associated ecosystems, it has become imperative that we employ any available means to combat the endangerment of this foundational tree. While the mountain pine beetle epidemic may appear to be on the decline, the whitebark pine must still overcome the potential consequences of climate change and white pine blister rust (WPBR). *Cronartium ribicola*, the pathogen that causes WPBR, is an exotic rust species that infects all five-needle white pines. Since indigenous species of pines have not been co-evolving with this pathogen, many pine populations are especially susceptible to the disease. Although traditional methods of resistance breeding have been employed to discover major genes that confer blister rust resistance in several related pine species, no such major gene has been identified in whitebark pine. However, partial genetic resistance does exist in low frequencies within naturally occurring whitebark pine populations. Experimental results indicate that this resistance is likely quantitative, meaning it involves the additive effects of several distinct genes. Though conventional means have proven inadequate to characterize the nature of blister rust resistance in whitebark pine, great promise lies in the rapidly advancing field of bioinformatics. With the right experimental design, bioinformatic approaches should be able to reveal particular genes that are differentially expressed in resistant trees. After identifying these candidate resistance genes, we can develop genetic markers that breeders can use to quickly and efficiently target resistant individuals for restoration efforts. Over the past two years as a graduate student Oregon State University, I have been working towards this goal. With the support of many agencies, organizations, and individuals, our research group has designed and executed an experiment that should reveal the genes underlying WPBR resistance in whitebark pine within the next year.

Initially, whitebark pine cones were collected from various populations in the Pacific Northwest and sown at Dorena Genetic Resource Center in 2010. In September 2012, we selected over 600 of these seedlings for our study of blister rust resistance. While half were inoculated with *C. ribicola*, the other half served as controls. Three days after inoculation, needles were collected from each experimental treatment and flash frozen in liquid nitrogen to preserve the tissue. Control seedlings were also inoculated shortly after their needles had been

sampled, so we could still evaluate resistance phenotypes from trees whose needles were sampled before inoculation. This experimental design allows us to distinguish between four groups of seedlings, based on their treatment (inoculated or control) and phenotype (resistant or susceptible).

While both resistant and susceptible trees may have very similar genomes, their phenotypes might differ significantly with variations in gene expression. Although each tree's potential genetic information is stored as DNA, its cellular activity is actually driven by expressing genes in the form of RNA. The whole collection of RNA molecules, known as the transcriptome, reveals the degree to which each gene is expressed. By analyzing the differences in gene expression between resistant and susceptible seedlings, we can more accurately target candidate genes for blister rust resistance.

Throughout the past two years, we have assessed each of the seedlings for symptoms indicating either susceptibility or resistance to blister rust. Given the results of our phenotypic observations, we have focused our research on one particularly promising family of seedlings, the half-sibling progeny of a single resistant mother tree from Mt. Rainier. Though these resistance phenotypes were only exhibited months or years after inoculation, we had extracted RNA from needles that were cryogenically preserved during the initial stages of infection. As a result, our transcriptomic data will represent the gene expression profiles of the seedlings three days post inoculation, as they were actively undergoing and potentially responding to infection. This approach should allow us to reveal any active resistance responses mediated by changes in transcription.

Once the RNA was extracted from viable needle samples, we employed a two-pronged sequencing strategy. First, we assembled a reference transcriptome, an important molecular resource and prerequisite for gene expression studies. Over 100 million individual sequences were assembled into approximately 45,000 transcripts, representing about 35,000 genes. We used a newly-developed sequencing technology called MiSeq to generate long paired-end reads, which facilitated the best possible quality for our reference transcriptome assembly. Since this transcriptome was generated from the combined read data of individuals representing every pairing of treatment and phenotype, we have developed a comprehensive atlas of potentially expressed genes in the needle tissue of this whitebark pine seedling family.

With the reference transcriptome assembled, we have begun using differential expression analyses to identify candidate blister rust resistance genes. For this study, we utilized a cheaper sequencing technology called HiSeq to generate huge read quantities from the RNA of 24 seedlings, with six individuals representing each experimental category. Altogether, we generated over 700 million reads for this phase of the study. Despite being relatively short and only single ended, we can easily map these reads against the reference transcriptome to assign their identities with greater certainty. Therefore, we can compare the transcriptomes of susceptible and resistant individuals, revealing

which specific transcripts are unique to resistant phenotypes. The experimental design and sample size will permit us to produce statistically significant results when comparing genes that are differential expressed between treatments and phenotypes.

While annotating these genes may help us to understand the mechanism underlying WPBR resistance in whitebark pine, our primary objective is to discover genetic markers within or proximal to the putative resistance genes. First, we must identify a particular single nucleotide polymorphism (SNP), single sequence repeat (SSR), or relative insertion/deletion (InDel) that is differentially expressed between resistant and susceptible individuals. While many of these SSR, SNPs, and InDels may be completely unrelated to blister rust resistance phenotypes, a small percentage of these markers will almost certainly be borne within or proximal to putative resistance genes. Once we have validated these resistance markers, we can develop cheap bioassays that allow researchers to quickly test for WPBR resistance in the field.

With this goal in mind, we are currently developing a pool of putative markers from the massive quantities of RNA sequencing data that were generated. After an exploratory stage where we are discovering and characterizing any markers with potential utility, we will begin to validate their efficacy by screening for their presence in trees that are known to be resistant. If several markers are found to be effective at discriminating between susceptible and resistant individuals, we might even develop a panel with several markers. In this way, we could hypothetically reveal which (if any) quantitative resistance genes are present in any given whitebark pine population or individual tree. With any luck, this protocol will serve to demonstrate how genomic markers can be developed from transcriptome data and ultimately used in breeding programs to target desirable traits like disease resistance.

We are fortunate to have this research supported by an outstanding group of collaborators, including a team led by Dr. David Neale at the University of California, Davis. For over a decade, this research group has been at the cutting edge of bioinformatic science, specifically the sequencing and genetic analysis of the genomes and transcriptomes of various *Pinus* species*.* Until recently, researchers have encountered great difficulty when attempting to assemble the *Pinus* genomes, which are notoriously large and repetitive as a result of their evolutionary history. However, a combination of advances in sequencing technology, faster computers with better algorithms, and significant cost reductions have allowed us to begin cracking the genetic codes of these ecologically and commercially important tree species. In March 2014, the group published the completed assembly of the loblolly pine (*Pinus taeda*) genome in the journals *Genetics* and *Genome Biology*. At over 22 billion base pairs, this assembly represents the largest genome ever sequenced to date. Moreover, the same working group is making excellent progress toward the completion of the sugar pine (*Pinus lambertiana*) genome, which---at an estimated 35 billion bases-

--will surpass their previous record by leaps and bounds. In the process, they have developed and confirmed the efficacy of several computing pipelines with broad applications in both genomics and transcriptomics. The same working group has also created several immense databases for *Pinus* bioinformatics data. Not only do these accomplishments exemplify the quality of our collaborators' research efforts, the annotated genomes of these related species provide a valuable reference resource for our study of whitebark pine transcriptomics.

In conjunction with these new bioinformatic resources, our successful sequencing efforts collectively bolster our prospects of identifying transcriptomic signatures of resistance. In addition to our primary funding provided by the Forest Service's Special Technology Development Program, we have received a great deal of support the Whitebark Pine Ecosystem Foundation (WPEF). Last year, the WPEF generously provided additional financial support for my graduate research on blister rust resistance in whitebark pine. The WPEF Student Research Grant has enabled me to sequence the transcriptomes of a greater number of individuals, providing a greater sample size for our differential gene expression study. In doing so, the WPEF has granted us the ability to assess a larger gene pool and yield more statistically robust results. I commend the members of this organization for spearheading the effort to halt and reverse the decline of the whitebark pine. Through our research, we aim to foster a basic scientific understanding of blister rust resistance in whitebark pine and to facilitate marker-assisted selection for resistance breeding programs. The student research grant funded by your WPEF membership has certainly supported our efforts.