Research Statement

Jie Wang

My research interest is using computational approaches that analyze large data set to answer biological questions with a particular emphasis on evolution. The field of biology has expanded greatly due to the recent development of large-scale high-throughput methodologies, such as next generation sequencing and high-throughput automated phenotyping tools. While the technologies provided keys to a greater understanding of the biology, the data produced by these technologies remained challenging or underutilized by many biologists. Facing these challenges, I am interested in utilizing bioinformatics methods, in combination with my knowledge in disease biology and evolution, to generate biological interpretations from large next generation sequencing data set and answer important biological questions.

Research experiences

My interests in bioinformatics started at the second year of my graduate school. One of my research projects was to characterize the genetic diversity and population structure for a fungal plant pathogen, *Fusarium virguliforme*, which causes sudden death syndrome (SDS) on soybean. The Sanger sequencing data archived in GenBank database demonstrated that there was no genetic diversity among *F. virguliforme* isolates. Therefore, I decided to develop a new set of microsatellite genetic markers with higher mutation rate for evaluation of genetic diversity for *F. virguliforme* using a 454 genome shotgun genome sequencing data [1]. When I was working on this project, I have learned a lot of skills on next generation sequencing data analysis, genome sequence analyses, file parsing using Perl, and data visualizations. This experience let me understand the wide application of bioinformatics and sparked my research interest in this area. In the meantime, I have utilized the genetic markers to develop a population genetics project to characterize genetic diversity and determine population structure of *F. virguliforme* for my dissertation research. Contrary to the long-believed asexual clonal lineage of *F. virguliforme* in the United States, I have detected high genetic diversity and strong population structure for *F. virguliforme* isolates collected in the United States using the microsatellite genetic markers I developed. By expanding the population size with the isolates collected from South America, I have detected high genetic diversity and clustered *F. virguliforme* populations into four ancestor clusters with a total of 539 isolates. Using Bayesian coalescence analysis and approximate Bayesian computation methods, I proposed evidences to support the Southern U.S. as the center of origin model for *F. virguliforme*. My project is the first study to elucidate the demographic history of this fungal pathogen using both North and South American isolates [3].

Besides the population genetics project, I have also worked on three other projects focusing on molecular diagnostics, epidemiology, and fungicide sensitivity. *Fusarium virguliforme* is a newly emerged pathogen in the U.S. and the diagnosis of this pathogen was cumbersome due to the lack of sensitive and specific diagnostic assay. I started my graduate research by developing an improved molecular diagnostic qPCR assay for detection and quantification of *F. virguliforme* from plant and soil samples [2]. Using the developed qPCR assay, I was able to conducted field experiments to determine the temporal dynamics of *F. virguliforme* colonization in soybean roots throughout the growing season. One important finding was that the quantity of *F. virguliforme* colonized in soybean was not correlated with SDS foliar symptom severity, which has suggested the possibility of distinct resistance mechanisms responsible for SDS foliar symptoms and root infection in soybean genome. This study not only showed the different sources of SDS resistance in soybean, but also provided a fast and accurate method to collect phenotypic data for characterizing root colonization levels among soybean cultivars [4]. The third project that I worked on was to determine the baseline fungicide sensitivity of the *F. virguliforme* isolates collected in the population genetics study to the succinate dehydrogenase inhibitor fungicide using image analysis and dose-response models to calculate EC_{50} for those isolates [5].

In addition to my dissertation research projects, I have also involved in several collaborative projects by applying my bioinformatics skills on different research topics and pathosystems. This includes bacterial comparative genomics and plant-pathogen interaction transcriptomics. The original objective for the bacterial comparative genomics project was to mine 23 bacterial genomes for a pair of primer and probe set for quick diagnostic of the bacterial pathogen from turf grass using real-time PCR. To fulfill this objective, I have worked on a Python pipeline to design, test, and validate primer sets for diagnostic purposes [6]. While I was working on this project, I have also performed a series of comparative genomics analyses on the same genomic sequencing data to study the genomic elements involved in host adaptation and pathogenicity. I have used the sliding window genome scanning combined with gene-tree-species-tree method, and identified genes and gene clusters that are associated with host adaptation in the turf pathogen genomes [7]. Furthermore, I have also worked on the pathogen transcriptomics of *Sclerotinia sclerotiorum* - *Pisum sativum* interactions by using a data set generated by a former postdoc, who was interested in the host response. I have developed my own pipeline to analyze the gene expression profile of the fungal pathogen on the interaction, and identify genes that contributed to pathogenicity/virulence and pathogen lifestyle transition during infection [8].

Future research interests

In the immediate future, my research interests include: *microbial pathogens host adaptations*, *pathogen lifestyle transition during infection,* and *demographic history of emerging plant pathogens*.

Microbial pathogen host adaptations

Plant pathogens contain species ranging from specialists with very narrow host ranges to generalists that attack a wide range of host plants. The driving force for the pathogen diversification and specification is the interaction between hosts and pathogens. However, the genomic basis for broad or narrow host range remains unclear for most of the plant pathogens. With continuous growing number of pathogen genomes sequenced and available to public, a collective effort is needed to combine the available genomic data into a comprehensive study to answer the question: "how does plant pathogen adapt to its host?". I had the experience to work on a comparative genomics project of a group of turf bacterial pathogens, and I have found that negative selection on virulence genes present in a pathogenicity island are associated with host range specification. This approach combined genomic analysis method and bench experiments and elucidated the genomic signatures for the pathogen host adapt process.

Pathogen lifestyle transition during infection

The lifestyle transition of the plant fungal pathogen occurred several times during the infection within the plant tissues. Understanding the genetic mechanisms behind this transition will deep our understanding of the host-pathogen interactions. However, thousands of genes in pathogen transcriptome could be involved in the interactions with host plants, and to facilitate the success of infection. Many of these genes were differentially expressed at different lifestyle stages during infection, which makes it very difficult to identify the key genes that determine the pathogenicity of the pathogen. Furthermore, the top differentially expressed genes are not always the switch genes that determine the gene expression profile. To learn the infection biology from the gene expression pattern using the transcriptomics data requires in-depth bioinformatics analysis to cluster and extract the key genes determining the infection process, and thus to narrow down the candidate gene list for wet lab validation. I would like to use the approach that takes advantage of gene co-expression network and multivariate statistical methods to determine the gene function modules contribute to the infection trait. Using this approach, we will not only identify genes important for host interaction, but also will determine the "hubs" that regulate the transitions between different infection stages and pathogen lifestyle transition.

Demographic history of emerging plant pathogens

In the past decades, there are many notorious cases of exotic plant pathogens that caused severe yield losses worldwide. To achieve a full control of emerging plant pathogens to additional geographical locations, collaborative efforts from basic biology to extension outreach are urgently needed. Understanding the demographic history and pathogen migration route serves as the first step toward developing effective disease mitigation strategies and set up proper quarantine zones to prevent spreading of the pathogens. I have utilized the microsatellite genetic markers to study the demographic history for the SDS pathogen, and proposed possible center of origin and migration route for this pathogen. The different genotypes of SDS pathogens identified in this study were also used for the screening for SDS resistant soybean cultivars. Currently, the more affordable sequencing technology provides tremendous amount of genome sequencing data for the population genomics study, which enables stronger statistical power to make more accurate inferences. Meanwhile, the genomic sequencing data can also be used for evolutionary genomics study, which determines the gene or genomic fragments that are under strong selection to adapt to local environment using the "reverse ecology" method proposed by Ellison et al. [9].

Taken together, I am interested in using system biology methods in combination with omics data to dissect the pathogenicity and host adaptation mechanisms of plant pathogens. With the experience gained in graduate school, I'd like to further deep and expand these research areas.

Short-term and long-term career goals

My short-term goal is to obtain a postdoctoral position in the area of bioinformatics and to further build up my experience and tool set in this area while using it to make novel biological discoveries. My long-term goal is to obtain an academia position and utilize bioinformatics tools to make scientific discoveries and train graduate students and postdocs in the area of agriculture.

Reference:

[1] **Wang, J.**, and Chilvers, M. I. 2016 Development and characterization of microsatellite markers for *Fusarium virguliforme* and their utility within clade 2 of the *Fusarium solani* species complex. *Fungal Ecology* 20:7-14

[2] **Wang, J**., Jacobs, J. L., Byrne, J. M., and Chilvers, M. I. 2015 Improved diagnoses and quantification of *Fusarium virguliforme*, causal agent of soybean sudden death syndrome. *Phytopathology* 105(3):378-87

[3] **Wang, J**., Jacobs, J. L., Rojas, A. J., and Chilvers, I. M. 2016 Population genetic analysis infers migration pathways of *Fusarium virguliforme*. *In preparation*

[4] **Wang, J**., Jacobs, J. L., Roth, M. G., and Chilvers, M. I. 2016 Temporal dynamics of *Fusarium virguliforme* colonization in soybean root. *In preparation*

[5] **Wang, J**., Stenzel, O. and Chilvers, M. I. Baseline sensitivity of *Fusarium virguliforme* to the succinate dehydrogenase fungicide, fluopyram. *In preparation*

[6] Giordano, P. R., **Wang, J**., et al. Genetic relationships among plant pathogenic members of the *Acidovorax* genus: Towards specific molecular diagnosis of the turfgrass pathogen. *In preparation*

[7] Zeng, Q., **Wang, J**., Giordano, P. R., et al. Genomic analysis of the turfgrass pathogen Acidovorax avenae subsp. avenae identifies evidence for directional selection during pathogen evolution. *In preparation*

[8] **Wang, J**., Rojas, A. J., Santos, P., and Chilvers, I. M. Fungal transcriptome analysis of the Sclerotinia sclerotiorum-Pisum sativum interaction. *In preparation*

[9] Ellison, C.E., Hall, C., Kowbel, D., Welch, J., Brem, R.B., Glass, N.L., and Taylor, J.W. 2011. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. P Natl Acad Sci USA 108:2831-2836.

Last update: March 24, 2016