127-P

In silico simulation of massively parallel sequencing as a diagnostic tool for bacterial phytopathogens

J. Daniels[1], T. Stobbe[1], A. Espindola[1], W. Schneider[2], J. Fletcher[1] and F.M. Ochoa-Corona[1]

1. Oklahoma State University, Department of Entomology & Plant Pathology, Stillwater, OK 74078
2. USDA ARS, Foreign Disease-Weed Science Research Unit, Fort Detrick, MD 21702

E-mail: *jon.daniels@okstate.edu* **and** *francisco.ochoa_corona@okstate.edu*

INTRODUCTION

- With increasing imports of commodities from countries abroad, the possible introduction of exotic plant pathogens has greatly risen over the past few decades.
- The aim of this study was to develop bioinformatic pipelines to generate mock sample databases (MSDs) and design electronic probes (e-probes) for the purpose of validating massively parallel sequencing (MPS) as a means to detect and identify high risk plant pathogenic microbes in a metagenomic sample.
- Massively parallel sequencing provides the capacity to process millions of sequence reads in parallel from a non-clonal fragmented library (1).
- Metagenomics is the study of genomic material of the entire complex community of microorganisms in an environmental sample (2,3).

OBJECTIVES

1. This study addresses the current need for simultaneously detecting all classes of plant pathogenic microbes, including bacterial, fungal, and viral pathogens, in a single assay.
2. The MPS system will discriminate strain-type and identify plant pathogens that underwent genetic modification to increase virulence.

MATERIALS AND METHODS

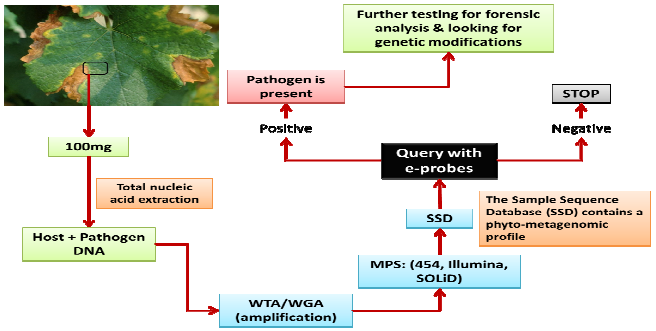

Figure 1: Flowchart showing extraction of infected plant tissue, to sequencing and querying with designed e-probes. (**WTA**= whole transcriptone amplification, **WGA**= whole genome amplification)

Definitions:

- Mock sample database (MSD): a simulated 454-sequencing database containing reads of +/- 400bp. MetaSim (4) was used to develop MSDs containing host & bacterial pathogen genomes at various abundances (3 replicates each).
- E-probe: a digital nucleotide sequence designed for a target pathogen. This was done by using a modified program, Tools for Oligonucleotide Fingerprint Identification (TOFI)(5) for use in unique sequence identification.

Table 1: List of pathogens & near neighbors used in e-probe design.

Target Pathogen	Near Neighbor
Xylella fastidiosa 9a5c	*Xanthomonas oryzae* pv. oryzae
Ralstonia solanacearum race 3 biovar 2	*Polynucleobacter necessarius*
Xanthomonas oryzae pv. *oryzae*	*Xylella fastidiosa* 9a5c
Candidatus Liberibacter asiaticus	*Agrobacterium tumefaciens*

RESULTS

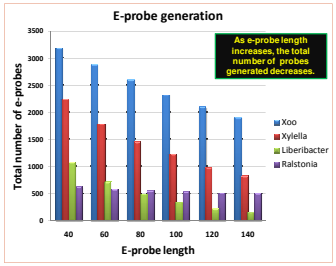

Figure 2: Comparison of the total number of e-probes generated for various lengths (40 – 140 nt).

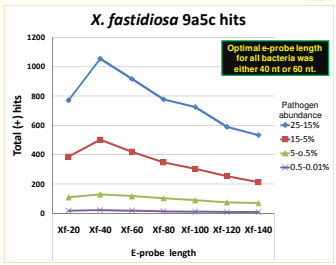

Figure 3: Comparison of the total hits in MSDs at varying pathogen abundances at varying e-probe length.

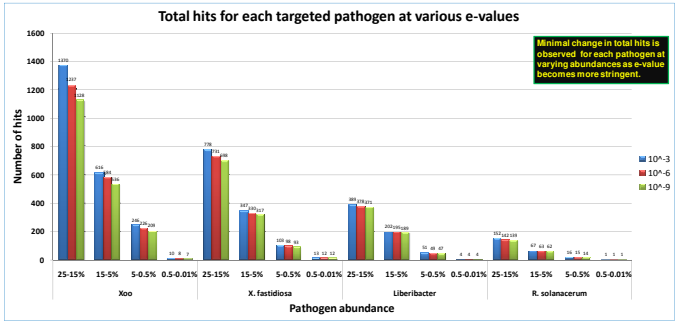

Figure 4: Comparison of the total hits of each target pathogen at varying e-values.

CONCLUSIONS

1. Bioinformatic pipelines were developed for quick (<48 hours) e-probe generation of targeted pathogens having a sequenced genome.
2. MetaSim MSD generation is streamlined to accept a multitude of modifications depending on the host and current biosecurity needs for use in expediting BLAST searches.
3. Optimal MSD blasting parameters were established, including query length and e-value.
4. The use and capability of MPS as a means to detect plant pathogenic microbes in complex host samples was demonstrated *in silico*.

LITERATURE CITED

1. Mardis, Elaine R. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 24 (3):133-141.
2. Understanding Our Microbial Planet: The new Science of Metagenomics. 2007. edited by T. N. A. o. Sciences: The National Academies Press.
3. Handelsman, Jo, Michelle R. Rondon, Sean F. Brady, Jon Clardy, and Robert M. Goodman. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology* 5 (10):R245-R249.
4. Richter DC, Ott F, Auch AF, Schmid R, Huson DH (2008): MetaSim—A Sequencing Simulator for Genomics andMetagenomics.PLoSONE3(10):e3373.doi:10.1371/journal.pone.0003373http://www.plosone.org/article/info:doi/10.1371/journal.pone.0003373
5. Vijaya Satya, Ravi, Nela Zavaljevski, Kamal Kumar, and Jaques Reifman. 2008. A high-throughput pipeline for designing microarray-based pathogen diagnostic assays. *BMC Bioinformatics* 9 (1):185.

ACKNOWLEDGEMENTS

This project was supported by the Agriculture and Food Research Initiative Competitive Grants Program Grant No. 2010-85605-20542 from the National Institute of Food and Agriculture. Thank you to the National Institute for Microbial Forensics & Food and Agricultural Biosecurity (NIMFFAB) for their review, comments, and guidance on this project.