BIOAg Progress Report, Snyder-Elling-Fu

TITLE: Sequencing the genomes of two critically-important biological control agents

PRINCIPAL INVESTIGATORS AND COOPERATORS: PI: Bill Snyder; wesnyder@wsu.edu; co-PIs: Axel Elling (Plant Pathology); Zhen "Daisy" Fu (postdoc in Snyder lab). Our project is in coordination with Andy Jensen, Research Director of the ID, WA and OR Potato Commissions; nematodes are collected on-farm in cooperation with potato growers across the Columbia Basin.

KEY WORDS: Biodiversity; Biological Control; Potato; Nematode; Genomics

ABSTRACT: Potatoes are a valuable Washington crop that is threatened by devastating insect pests. We have found (1) that insect-killing "entomopathogenic" nematodes (EPNs) are key natural enemies of these insects, and (2) that organic farming greatly increases EPN genetic diversity. Indeed, genetically-diverse mixes of EPN strains are more lethal to insects than any single worm strain. We would like to identify the specific genes that allow different EPN strains to "complement" one another in killing pests. Unfortunately, our efforts have been limited by the lack of well-constructed reference genomes for the two most common species of EPNs in Washington, *Steinernema feltiae* and *Heterorhabditis bacteriophora*. We are using the third-generation advanced sequencers now available at WSU (e.g., the PacBio RS) to construct high-coverage genomes of these two species. Insect-killing nematodes can be used as bio-pesticides. Understanding the traits that make worms lethal to pests will allow us to design effective bio-pesticide blends, and to conserve and enhance beneficial nematode biodiversity on farms. Thus, our project is developing *novel approaches to pest management* that *increase the sustainability of farming systems*.

PROJECT DESCRIPTION

Recent advances in genomics hold the potential to revolutionize plant breeding. Sequencing technology is getting both cheaper and more powerful, such that complete genomes exist for many important crops (e.g., Valesco et al. 2010). When genes associated with agronomically-important traits can be identified, targeted breeding programs can be designed to promote these genes (Perez de Castro et al. 2012). However, modern genomics approaches have not been used for the targeted development of biological control agents that possess desirable traits. If genes associated with biocontrol effectiveness could be identified, strategies could then be developed to enhance the frequency of these traits during natural enemy conservation or when biological control agents are deployed as bio-pesticides. Linking a natural enemy's pest-control effectiveness to underlying genes would also enhance our fundamental understanding of biological control.

The Colorado potato beetle, *Leptinotarsa decemlineata*, is a major pest of potato crops worldwide. The beetles quickly develop resistance to new insecticides, making biological control an attractive (and sometimes the only) control option. Potato beetles are attacked by

insect-killing, “entomopathogenic” nematodes (EPNs). Working in Washington potato fields, we have found that two EPN species, *Steinernema feltiae* and *Heterhabditis bacteriophora*, are most common. EPNs are exceptionally abundant on organic potato farms, exerting natural pest control that could explain how organic potato farmers produce high yields without much insecticide use (Ramirez and Snyder 2009, Crowder et al. 2010). The nematodes can be sprayed as a “bio-pesticide”to improve beetle control (Grewal et al. 2005).

Most recently, we have found that organic farming promotes greater genetic diversity within each EPN species, and that greater intraspecific genetic diversity improves pest control. For example, pairs of two genetically-different isolates of *S. feltiae* kill far more insects than any single strain (Fig. 1). Thus, EPN strains “complement” one another. BIOAg funding in 2012 allowed us to conduct RAD-TAG sequencing to delineate genetic differences among complementary nematode strains. This sequencing approach involves digesting the genome with a restriction endonuclease, which, after shearing, is then sequenced. The short fragments (or “tags”) of DNA that flank each digestion site are screened for genetic variation (i.e., single nucleotide polymorphisms or "SNPs"). Interpreting these sequences requires alignment with high-quality reference genomes. Unfortunately, we found that the *S. feltiae* and *H. bacteriophora* genomes provided to us by other researchers were of exceptionally low quality, with poor genome coverage and thus weak alignment with our RAD-TAG sequences. This made it difficult to annotate our SNPs and correlate them with specific genes, such that we could not generate the convincing preliminary data that we need to pursue federal grants.

In our 2012 BIOAg submission, we proposed to make use of the third-generation sequencing technology now available at the genomics core facility at WSU, to construct high-quality reference genomes for *S. feltiae* and *H. bacteriophora*. This will allow us to use our RAD-TAG data to search for genes that underlie nematode complementarity. Additionally, we proposed to use new sequencing technology at U Idaho to compare gene expression profiles among pairs of complementary nematode strains, which will provide an additional, and complementary, tool to search for complementarity-related EPN genes. Our ultimate goal is to engineer EPN bio-pesticides that combine strains with complementary modes of activity, and to gain a greater fundamental understanding of how EPN strains complement one another to kill more pests.

Objectives

Our project has 3 inter-related objectives:

(1) Sequence the genomes of the two most common insect-attacking nematode species in WA potato fields, *S. feltiae* and *H. bacteriophora*.

(2) Compare gene-activity patterns among pairs of different strains of *S. feltiae* that complement one another in order to kill more beetle pests than could any single strain alone.

(3) Use these data to identify genes that lead nematodes to complement one another.

OUTPUTS

Initially, we have focused on developing a complete genome for *S. feltiae*, the nematode species for which we have (1) collected the largest number of strains and (2) that most clearly shows intraspecific complementarity. To accomplish this, genomic DNA was sheared to generate large fragments for library construction with the Pacific Biosciences DNA template prep kit (3-10Kb). The sample did not shear predictably and generated a library whose peak size was smaller than anticipated. When attempting to sequence the resulting library, it proved to be un-sequenceable. Additional purification of the starting DNA was performed, and the shearing and library construction was then repeated. Again the resulting library produced very little sequence. Pacific Biosciences provided an unpublished affinity purification that has been used to successfully make sequenceable libraries from challenged samples. The small amount of library recovered did in fact sequence. Since the average size of the library was smaller than desired, the longest read enzyme available at the time, XL, was used in conjunction with the longest read chemistry available at the time, Sequencing Reagents XL version 1.0. This combination was used to generate 5 GB of data from the sample, which represents approximately 50x coverage (Table 1).

This level of coverage was considered an acceptable level of coverage for the assemblers to use in early 2013, when this work was initiated. Specifically, we envisioned using the alloraEC protocol of SMRTpipe Version 1.3.3. Over two million long reads (Table 1, average read length = 2622 bp) were actually generated from 24 SMRT cells; this number of the reads provided decent coverage of the genome (59x coverage). However, because of the technical issues described below we currently have 77.8% of the raw reads that did not pass the filtering settings; as a result, the assembly of the data has been problematic. The initial assembly that we conducted covered only 40% (coverage 0.4) of the genome. The XL enzyme, while providing the longest reads at the time, is also the lowest quality enzyme available. We envisioned that high depth would overcome these deficiencies. It did not. The XL enzyme is now being phased out, as it does not provide the high consensus accuracy needed for *de novo* assembly. In addition, since this project was initiated, both the enzymes and the assembly strategies have changed. The P4 polymerase provides just as long – if not longer – reads at much higher quality, and the best *de novo* assembler available now at the WSU Core Facility is HGAP 2.0. According to Core Facility personnel, HGAP has proven to be a superior assembler compared to earlier technologies, assembling some of the most challenging regions of very difficult genomes.

HGAP 2.0 works best when it is provided with 60-100x coverage of reads >5kb. In addition, HGAP 2.0 uses the full range of quality values and kinetics provided by the PacBio RSII instrument to polish the assembly using the Quiver algorithm. The XL enzyme we initially used for this project is not supported by the quiver program. Assembly parameter screens with the existing data have not generated a decent assembly to date. The Core Facility has now revisited library construction using newer methods, and personnel there are attempting to generate sequence with the P4 enzyme so it can be assembled with the latest, most effective assembler.

The existing XL data may still be useful for scaffolding high confidence contigs generated by the P4 enzyme and HGAP 2.0. Recent upgrades to the RS instrument have dramatically increased yield, and the Core Facility predicts being able to generate enough data, without much additional cost, that should assemble better than the XL data already in our possession.

Because the Genomics Core of Washington State University was having technical difficulties with DNA library preparation, as we have just described, we tried an alternative route to analyze the restriction enzyme associated tag (RAD-tag) data without aligning to a reference genome. Briefly, we aligned all of the reads to the two nematode strains that generated the largest number of RAD-tag reads (32-4D and 38-6 2P) using the Bowtie aligner (Langmead et al., 2009). Sequence Alignment/Map (SAM) files were generated. Samtools (Li et al., 2009) was used to process the SAM files and generate the Variant Call Format (VCF) files. VCF is a text file format, containing meta-genotype information of each strain for each locus. Allele frequency of each locus was calculated using in-house script. Afterwards, genetic divergence between strains was calculated from the sum of difference of allele frequency for each locus using in-house script. The genetic distance matrices, which include all pairwise genetic divergences between strains, were constructed. The resulting genetic distance matrices were imported to the PHYLIP program (Felsenstein, 2005) to construct phylogenetic trees with various algorithms. Finally the phylogenic trees were visualized in the program Integrative Tree of Life (Letunic and Bork, 2006, Figure 1).

No matter whether we use nematode strain 32-4D, or 38-6 2P, as our reference, we conclude that the phylogenic relationship between Pat NW 12-6 and Pat NW 6-2 is further than between Pat NW6-2 and Pat NW12-1, and that Pat NW6-2 and Pat NW12-1 are more closely related (Figure 1). The bioassay we conducted earlier showed synergistic effect of strain pairs NW 12-1 and NW12-6, and of NW 6-2 and NW12-6, on mortality of waxworm (*Galleria mellonella*, Figure 2). The results of the phylogenetic tree, together with the complementarity bioassay, illustrate that it is more likely that two nematode strains are complementary if they are more-distantly related.

IMPACTS

· Short-Term: We now have identified the species of EPN that occur in Washington organic potato fields, and have determined that significant genetic diversity exists within these populations. Furthermore, combining genetically-different EPN strains leads to higher mortality of pest insects. Thus, the genetic diversity among EPNs that organic farming promotes, seems likely to improve natural pest control.

· Intermediate-Term: Our key intermediate-term goal is to successfully compete for federal grant funding to further investigate genetic differences among EPN species, identify genes that correspond to important traits tied to a worm-strain's ability to kill hosts, and develop the means to search for these traits in nematodes in potato fields. We next hope to begin the design and testing of bio-pesticides that combine beneficial and complementary nematode traits.

· Long-Term: Our ultimate goal is to provide new commercial bio-pesticides that effectively control potato beetles, and to provide a model approach for understanding why natural enemies complement one another that can be applied to other pests or cropping systems.

ADDITIONAL FUNDING APPLIED FOR / SECURED

This project has allowed PI Snyder to develop molecular biology expertise in his laboratory, and in turn this has allowed him to successfully compete for a large federal grant in this area. **We learned this fall that our USDA Organic Transitions (USDA-ORG) proposal, which includes a molecular-biology component led by BIO-Ag postdoc Fu and PI Snyder, was funded at the full requested amount of $747, 955.** Snyder is PI of the new USDA-ORG grant. Also, BIO-Ag co-PIs Snyder, Elling and Fu will be submitting a proposal to USDA-NIFA's Foundational Program this winter, with preliminary data from the BIOAg project forming the core of this new proposal.

GRADUATE STUDENTS FUNDED

This project supported the research of 2 PhD students in PI Snyder's laboratory, Christine Lynch and Karol Krey. Christine successfully defended her dissertation in spring 2013 and is currently a postdoc at the University of Hawaii; Karol continues to make excellent progress toward her degree with graduation expected in 2015. This BIOAg project also provided data being used by Elliott Moon to complete his MS degree in Statistics. Elliott also graduated this past spring, and recently accepted a job at a statistical consulting firm in Seattle.

RECOMMENDATIONS FOR FUTURE RESEARCH

Our next step is making use of the third-generation sequencing technology now available at the genomics core facility at WSU, to construct high-quality reference genomes for *S. feltiae* and *H. bacteriophora*. This will allow us to use our RAD-TAG data to search for genes that underlie nematode complementarity. Additionally, new sequencing technology at U Idaho will be used to compare gene expression profiles among pairs of complementary nematode strains, providing an additional, and complementary, tool to search for complementarity-related EPN genes. Our ultimate goal is to engineer EPN bio-pesticides that combine strains with complementary modes of activity, and to gain a greater fundamental understanding of how EPN strains complement one another to kill more pests.

LITERATURE CITED

Conesa, A. et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676.

Crowder, DW, TD Northfield, MR Strand and WE Snyder. 2010. Organic agriculture promotes evenness and natural pest control. Nature 466:109-112.

Eid, J. et al. 2009. Real-time DNA sequencing from single polymerase molecules. Science 5910:133–138.

Grewal, P.S., R-U Ehlers and DI Shapiro-Ilan. 2005. Nematodes as Biocontrol Agents. CABI Pub., Wallingford, UK; Cambridge, MA.

Koren, S, et al. 2012. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. Nat. Biotechnol. 30:693-700

Myers, EW, et al. 2000. A whole-genome assembly of Drosophila. Science 5461:2196-2204.

Perez deCastro, AM, et al. 2012. Application of genomic tools in plant breeding. Current Genomics 132:179-195.

Ramirez, RA and WE Snyder. 2009. Scared sick? Predator-pathogen facilitation enhances the exploitation of a shared resource. *Ecology* 90:2832-2839.

Velasco, R, et al. 2010. The genome of the domesticated apple. Nature Genetics 42:833.

Table 1. Summary of raw reads and assembled contigs for the genome sequences of *Steinernema feltiae*.

Filtering

	Pre-Filter	Post-Filter
Polymerase Read Bases (bp)	5910335093	3722486870
Polymerase Reads	2254506	515963
Polymerase Read Length (bp)	2622	7215
Polymerase Read Quality	0.323	0.83
Coverage (fold)	59.1	37.22

Post-Preassembler

Length cutoff (bp)	1077
Pre-assembled bases (bp)	654172809
Pre-assembled reads	492662
Pre-assembled read length (bp)	1309
Pre-assembled N50 (bp)	1343
Coverage (fold)	6.54

Assembly (polishing step failed)

scaffolds	16507
Max Contig Length (bp)	24551
N50 Contig Length (bp)	2764
Sum of Contig Lengths (bp)	41343576
Coverage (fold)	0.41

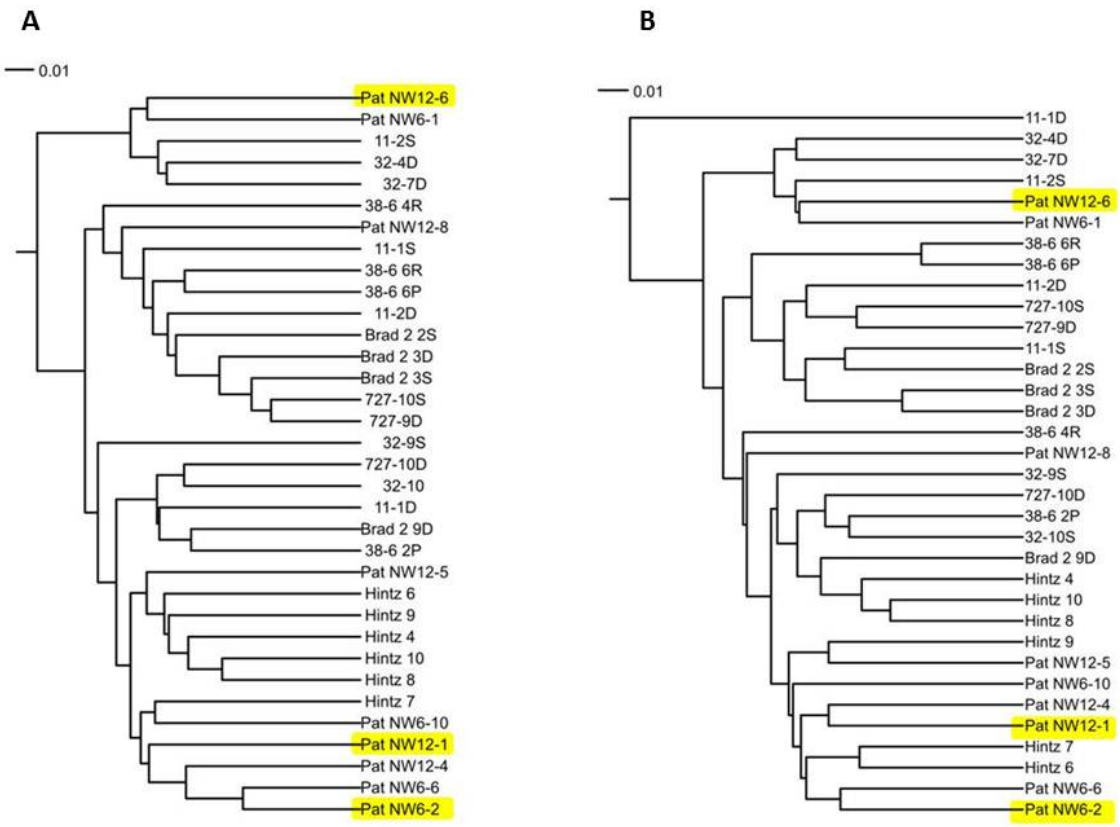

Figure 1. Phylogenetic relationship of 34 strains of entomopathogenic nematodes collected from potato fields in Washington. The phylogenic tree (A) was constructed based on alignment of all of the reads to strain 32-4D; tree (B) was constructed based on alignment of all of the strains to 38-6 2P. Both phylogenetic trees were constructed with the algorithm Kitsch in the PHYLIP program. The strains we tested in the bioassay (Figure 2) are highlighted.

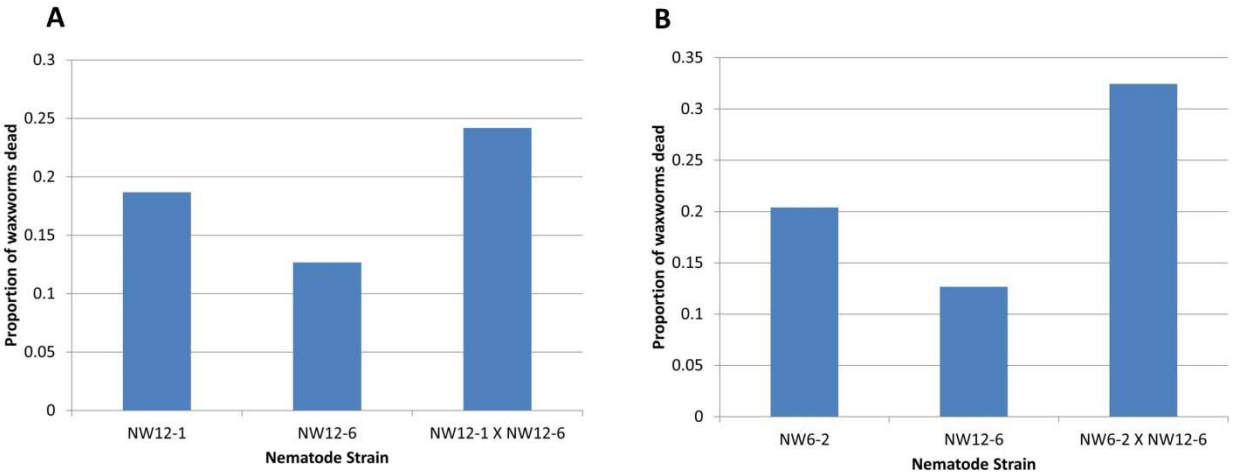

Figure 2. Synergistic relationship of entomopathogenic nematode strains on mortality of waxworms (*Galleria mellonella*). (A) With the same total number of nematodes, the combination of NW 12-1 and NW 12-6 caused higher mortality on waxworms than either single strain exerted on its own. (B) Likewise, with the same total number of nematodes, the combination of NW 6-2 and NW 12-6 caused higher mortality on waxworms than either single strain exerted on its own.