Improving *Globodera rostochiensis* genome with ultra-long Nanopore reads.

The potato cyst nematode *Globodera rostochiensis* is a devastating plant parasite that is present on both the USDA and EPPO quarantine organism lists. This biotrophic pathogen possesses an arsenal of effector proteins in order to infect its host. Some of these are adapted from endogenous genes while others have been acquired by horizontal gene transfer. Eves-van den Akker's lab aims to understand the molecular basis of plant parasitism and virulence by exploiting recent advances in genomics. Through a recent community wide effort, the draft genome of *G. rostochiensis* was completed. It was found that both effectors clustered in islands, and genes acquired by horizontal gene transfer, were in close proximity to transposons: indicating a role for transposon repeats in genome re-organisation.

The current draft *G. rostochiensis* genome was assembled from short Illumina sequencing reads, typically of 100-200 bases in length. Given that the short-read assemblers will fail when a sequence repeat (i.e. a transposon) is longer that the read itself, these regions of critical important in the evolution and execution of parasitism are, by the very nature of how they are formed, incompletely assembled in the draft *G. rostochiensis* genome. The aim of my project was thus to explore the efficacy of second generation, ultra-long read, Oxford Nanopore sequencing technology for plant-parasitic nematode genome improvement. Ultra-long reads have the potential to generate sequences far longer than the longest repeat in the *G. rostochiensis* genome, with clear promise to improve the existing genome assembly. Exploiting this technology to generate a more complete assembly will facilitate analyses of these key genomic regions and their implication for plant parasitism.

Nanopore sequencing takes advantage of protein nanopores inserted into a synthetic polymer membrane with a high electrical resistance. As the DNA passes through the nanopore, it creates a characteristic disruption in the ionic current, generated by applied potential. The disruption patterns are used to distinguish between the four standard DNA bases. With current Nanopore protocols, it is possible to obtain reads of several hundred thousand bases in length. However, with the increased length of sequenced DNA the overall coverage is low.

My project was divided in two phases. First, the DNA extraction of ultra-long DNA molecules had to be optimised for *G. rostochiensis*. This was critical for the overall success of the project. Given the fast development of tools for long read sequencing, it would also provide valuable transferable information to a wide range of projects. Through a very gentle extraction of DNA from approximately 1 million individual nematodes, molecules of sufficient length and quality were obtained. The second stage of the project focused on Nanopore sequencing and assembly of ultra-long molecules. We adapted the original protocols and the previously tested protocols by the nanopore community, to optimise the sequencing for *G. rostochiensis*. In theory, a small number of long reads would likely resolve ambiguously assembled regions in the *G. rostochiensis* genome.

We were able to successful sequence ultra-long reads (max. ~450 kb), although as expected, sequencing error rates were worse than relatively shorter reads, and low overall. Nevertheless, those shorter reads of (1,000 - 8,000 bp) with relatively higher quality could also potentially improve the assembly. As this technology is in its infancy, there is no 'best practice' pipeline for read correction and assembly. We are therefore in the process of exploring the efficacy of existing assembly tools that permit the inclusion of such long-read data (e.g. SPAdes), for genome improvement in plant-parasitic nematodes. The flexibility of this assembler lies in its ability to integrate short read, and long read data, in combination with an existing draft assembly. In parallel, we are exploring a range of different and new pipelines for long-read based genome assembly. If successful, we have the potential to provide a more faithful representation of a Globodera genome, and with it, facilitate the analyses of key genomic regions involved in parasitism, and their implications for virulence.

Simon Bajew

University of Dundee

The British Society for Plant Pathology
Bursary Report