The *Rotylenchulus reniformis* Genome: Sequencing, Assembly, and Annotation

William S. Sanders[1,*], Kurt Showmaker[1,2], Ben Bartlett[1,3], Hui Wang[1], Satish Ganji[1], Martin Wubben[1,4], Fiona McCarthy[1,5], Zenaida Magbunga[1,3], Daniel Peterson[1,3]

[1]Institute for Genomics, Biocomputing, and Biotechnology, Mississippi State University, USA; [2]Department of Biochemistry & Molecular Biology, Mississippi State University, USA; [3]Department of Plant and Soil Sciences, Mississippi State University, USA; [4]USDA-ARS, Mississippi State, USA; [5]Department of Basic Sciences, College of Veterinary Medicine, Mississippi State University, USA

*-wss2@igbb.msstate.edu

Abstract

Rotylenchulus reniformis (the reniform nematode – Linford & Oliveria 1940) is a plant parasitic nematode with a primarily tropical and subtropical geographic distribution. The reniform nematode is an emerging problem in US cotton production, causing an estimated annual $130,000,000 USD crop loss. Adult females establish feeding sites in the plant root, where the head portion of the nematode is positioned inside the plant root and the external tail portion begins to swell, developing a kidney-like shape. Understanding of the genome structure of this organism can help provide insight into the mechanisms of its parasitism and aid in the development of reniform resistant cotton lines. Genomic reniform DNA was isolated and sequences were generated using two platforms, combining both Illumina Sequencing-by-Synthesis and Roche 454 Pyrosequencing technologies. Sequences were assembled using both Roche's GS *De Novo* assembly algorithm (Newbler) and the ABySS assembly algorithm. Here we present the current status of the *Rotylenchulus reniformis* genome sequencing project and discuss the present state of our draft assembly and the associated annotations.

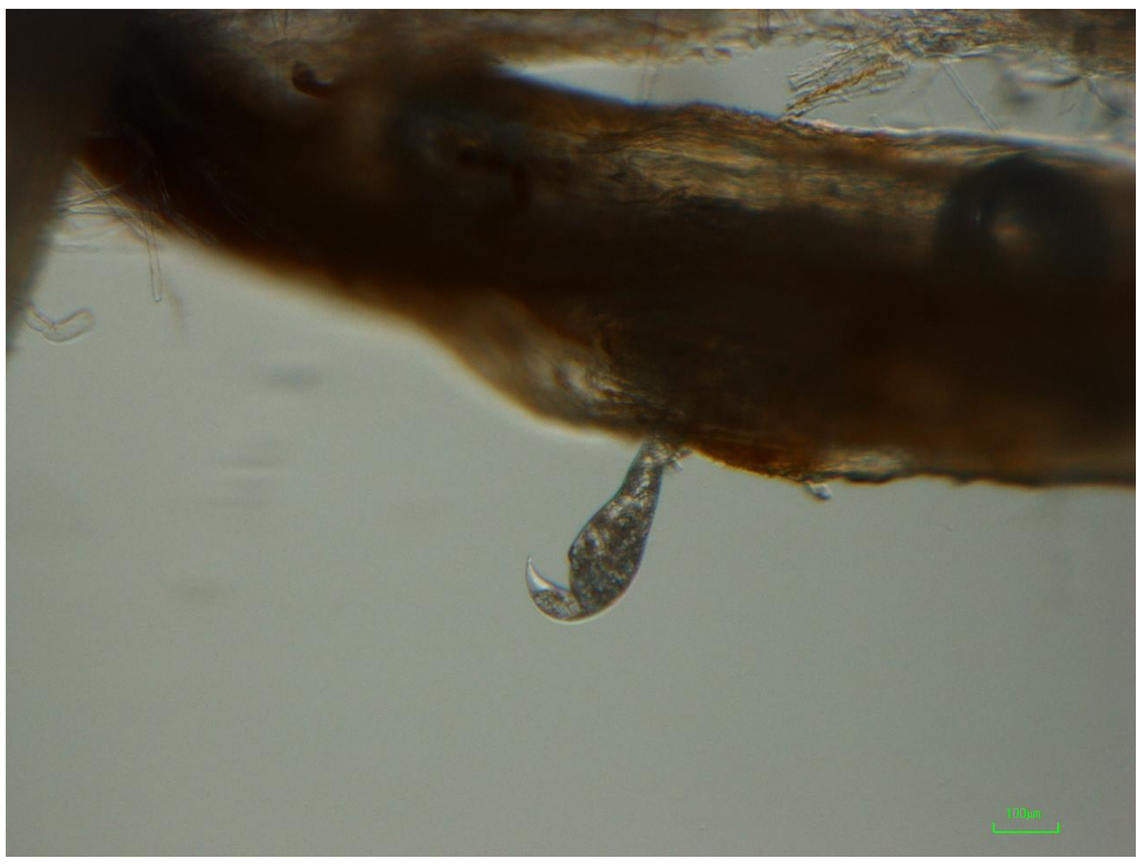

Figure 1. *R. reniformis* sedentary female with an established feed site on the root of a cotton plant.

Infection by *R. reniformis* can result in reduction of plant weight, a reduction in the number of pods per plant, a reduction in the chlorophyll content of leaves, and a decrease in the bulk density of stem parts.

Challenges When Working with *R. reniformis*

- Lack of a close reference genome (Fig. 2).
- DNA isolation from a pooled sample of nematodes
- Parasitic relationship requires culturing the nemaode with its associated host plant and soil
- Genome size not currently known – based on other sequenced nematodes, the size is currently estimated to be between 58 Mb and 144 Mb.

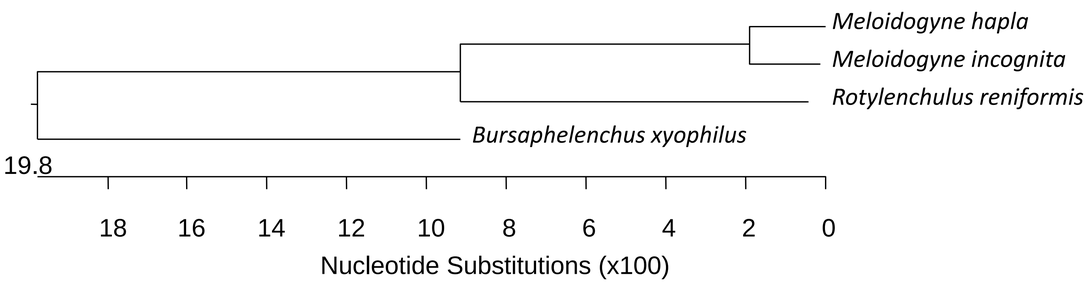

Figure 2. *R. reniformis* 18S rRNA (EU306342) aligned using CLUSTALW with other sequenced plant parasitic nematodes, *M. hapla* (AY593892), *M. incognita* (AY284621), and *B. xylophilus* (FJ235886).

Materials & Methods

A pooled population of of *R. reniformis* was sampled and DNA isolated using a QIAGEN Dneasy Blood & Tissue Kit. The pooled sample contained a number of nematodes, resulting in increased SNP sampling.

Illumina Sequencing:
Single reads –
7 lanes of 1x75 bp reads
Paired End reads w/ inserts –
1 lane of 2x100 bp reads with 250 bp inserts
1 lane of 2x100 bp reads with 350 bp inserts

Roche 454 Sequencing:
Single reads –
½ run of whole genome shotgun reads
Paired End reads w/ inserts
½ run of paired end with 8,000 bp inserts

Figure 3. *R. reniformis* sedentary female with an established feed site on the root of a cotton plant.

Assembly was performed using both the ABySS *de novo* assembly algorithm (version 1.3.2) and the Roche 454 GS *De Novo* Assembler (version 2.6).

Separate assemblies were generated using the Illumina sequences with ABySS and the Roche 454 sequences with the Roche 454 GS De Novo Assembler. A combined assembly was generated using the Roche 454 assembly and ABySS.

Sequencing Statistics

A total of 11,155,121,426 bp have been sequenced using combined Illumina and Roche 454 technologies.

Based on an estimated minimum genome size of 58 Mb and an estimated maximum genome size of 144 Mb, we have sequenced in the range of 78X – 191X genome coverage.

In order to obtain a more accurate coverage estimate, we are currently performing real-time PCR analysis in order to get an estimate of genome size of *R. reniformis*.

Table 1. Sequencing Statistics for *R. reniformis*.

	# Sequences	Total Sequence Length (bp)
Illumina 1x75	56,182,791	4,262,542,542
Illumina 1x100	15,613,588	1,561,358,800
Illumina 2x100 (250 bp insert)	28,920,210	2,892,021,000
Illumina 2x100 (350 bp insert)	21,594,142	2,159,414,200
	122,310,731	10,875,336,542
Roche 454 SE	462,594	161,287,004
Roche 454 PE (8kb insert)*	679,515	118,497,880
	1,142,109	279,784,884

Assembly Statistics

	# Contigs	N50	Largest Contig (bp)	Total Sequence
Illumina Sequences	213,248	365	34,739	78 Mb
Roche 454 Sequences	19,075	808	9,444	15 Mb
Illumina + Roche 454 Sequences	225,980	457	39,345	96 Mb

The combined assembly contained 1,571 contigs > 2,000 bp in length (N50 = 2,812) covering a total of 4.7 Mb of genomic sequence.

Microsatellite Analysis

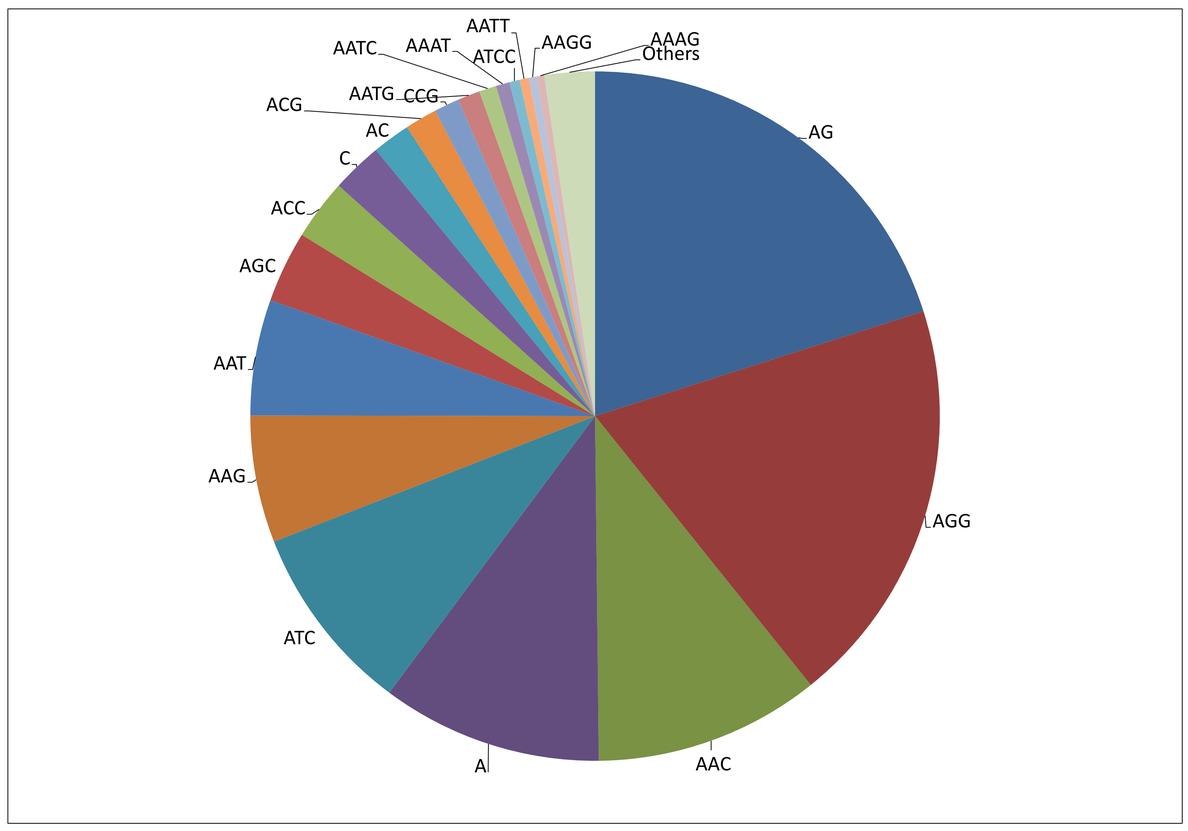

Figure 4. Frequency of microsatellite classes in the genome of *R. reniformis*. Among the 58 classes identified, the 20 most frequent are shown in individual divisions. The remaining 38 microsatellites are considered in a single division defined as others.

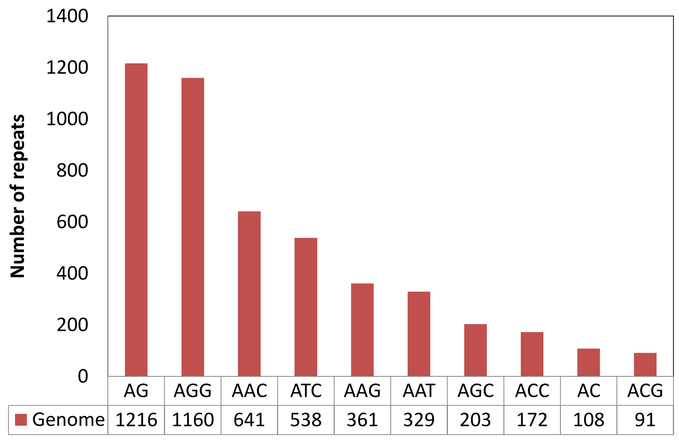

Figure 5. Frequency of the 10 most abundant repeats from our *R. reniformis* genome assembly.

Microsatellite analysis was performed using PHOBOS 3.3.11 and only examined perfect microsatellites from 1 to 6 bp in length, with detection thresholds of 12 repeats (for 1 bp microsatellites), 8 repeats (for 2 bp microsatellites), and 5 repeats (for 3,4,5, and 6 bp microsatellites).

The frequencies of the 10 most abundant microsatellite repeats were then counted and compared to previously identified microsatellites from an SSR-enriched library of *R. reniformis* (Arias, et al. J of Nematology. 41(2):146-156. 2009).

Annotation Progress

The 1,571 contigs > 2,000 bp in length were used in conjunction with GeneMark.hmm to predict 2,440 protein coding genes, which are being subjected to further annotation.
74.94% of RNA-Seq reads of the R. reniformis transcriptome (See Poster P0067) mapped to the genome assembly with at least 1 reported alignment.

Future Work

- More sequencing incorporating mitochondrial DNA removal and non-specific whole genome amplification of R. reniformis DNA
- Further refinement of the assembly using alternative assembly algorithms and visualization tools
- Further structural and functional annotation – identification of ncRNAs, repeat elements, GO annotation
- Submission of curated sequences to repositories
- Incorporation of transcriptome sequences (P0067) to help further refine our predicted gene models
- Using our R. reniformis BAC-end library to further improve the assembly
- Proteomics (Proteogenomic Mapping) to help refine gene models during structural genome annotation

Funding
USDA-ARS Award: 5864027241-08050415