The Genome of Reniform Nematode, *Rotylenchulus reniformis*

Satish Ganji[1*], William S. Sanders[1], John V. Stokes[2], Kurt Showmaker[1,3], Ben Bartlett[1], Hui Wang[1], Martin Wubben[3,5], Fiona McCarthy[1,6], Zenaida Magbanua[1], Daniel Peterson[1,4]

[1]Institute for Genomics, Biocomputing, and Biotechnology, Mississippi State University, USA; [2]College of Veterinary Medicine, Mississippi State University, USA; [3]Department of Biochemistry & Molecular Biology, Mississippi State University, USA; [4]Department of Plant and Soil Sciences, Mississippi State University, USA; [5]USDA-ARS, Mississippi State, USA; [6]Department of Veterinary Science and Microbiology, University of Arizona.
*satishg@igbb.msstate.edu

Abstract

The reniform nematode (RN; *Rotylenchulus reniformis*) is a pest that causes considerable damage to cotton. For example, in 2011 yield losses of approximately 279,000 bales (total estimated value > $90 million) were attributed to RN damage. Ostensibly, sequencing the genome of *R. reniformis* represents a key step in identifying genes underlying RN's ability to infect host plants. Ultimately, knowledge of the RN genome may suggest means of minimizing *R. reniformis* damage through targeted disruption of RN-specific gene pathways. Towards this end, we determined the RN genome size and initiated whole genome sequencing of the nematode. Our flow cytometric analysis indicates that the genome size of *R.reniformis* to be ~190 Mb, almost twice the size of the genome of *C. elegans* (~100 Mb) and 3-4 times the size of the root-knot nematode (*Meloidogyne incognita*) genome (~50 Mb). Reniform nematode genome sequencing was performed using Illumina and Roche 454 technologies, and sequence reads were assembled using Newbler and ABySS. Here we present the current status of the *R.reniformis* genome sequencing project and discuss the present state of our draft assembly and its annotation.

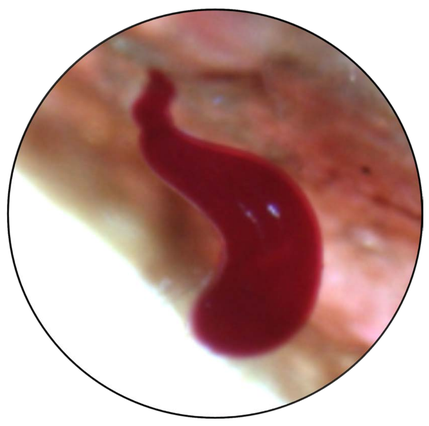

Figure 1. *R. reniformis* sedentary female (stained with Acid Fuschin) showing characteristic reniform shape, with an established feeding site on the root of cotton host plant (background).

R. reniformis adult female stage infects and feeds on the host plant root from a specialized feeding structure called syncytium, formed by dissolution of the cell walls of cells adjacent to initial feeding cell. This results in the loss of nutrients to the sink .

Materials & Methods

A pooled population of *R. reniformis* mixed-stage vermiform tissue was collected from host plant cotton cultures and surface-sterilized with 0.01% $HgCl_2$ and 0.001% hibitane.

Flow Cytometry

To estimate the genome size of *R.reniformis* by flow cytometry, nuclei were isolated from vermiform tissue. The tissue was homogenized in 2 mL Dounce tissue grinder with ice cold Galbraith buffer (45 mM $MgCl_2$, 30 mM sodium citrate, 20 mM MOPS, 1mg/mL Triton X-100, 1mg/L boiled RNase, pH 7.2). The homogenate was passed through 10 µm nylon membrane filter and the filtrate containing nuclei was stained with propidium iodide (50 µg/mL). *C.elegans* was used as known control with calculated genome size of ~100 Mb. N2 strain of *C.elegans* was cultured on NGM plates for six days. The vermiform was collected by washing the plates with water followed by surface sterilization, isolation and staining of nuclei with PI. The stained nuclei were analyzed with FACSCalibur flow cytometer (BD Biosciences)using CellQuest Pro software.

Genome Sequencing

Genomic DNA for sequencing was 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

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.

Genome Size Estimation of *R.reniformis*

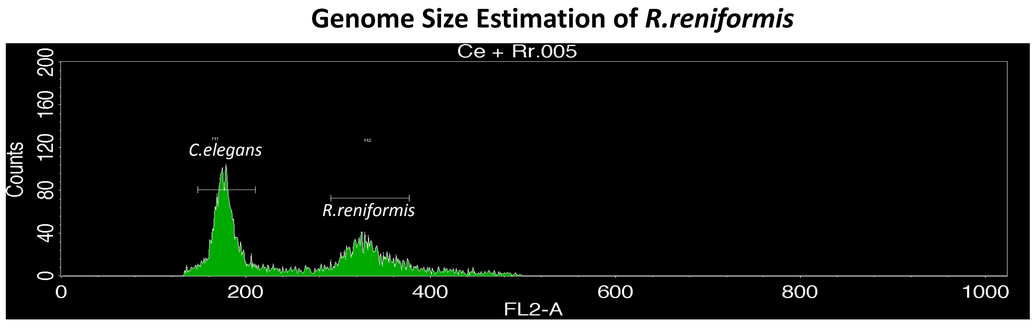

Figure 2. Histogram of fluorescence intensity of Propidium Iodide (PI) stained nuclei of *C.elegans* and *R.reniformis*. The cytometer argon laser was adjusted at 488 nm at 20mW in potential and the measurement was realized at 585 nm.

By comparing the flouresence intensities of PI stained nuclei, the genome size of *R.reniformis* was estimated to be ~190 Mb, based on the calculated genome size of *C.elegans* (~100 Mb).

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

	# Sequences	Total Sequence Length (bp)
Illumina 1x75	56182791	4262542542
Illumina 1x100	15613588	1561358800
Illumina 2x100 (250 bp insert)	28920210	2892021000
Illumina 2x100 (350 bp insert)	21594142	2159414200
Total	**122310731**	**10875336542**
Roche 454 SE	462594	161287004
Roche 454 PE (8kb insert)*	679515	118497880
Total	**1142109**	**279784884**

Table 2. Assembly Statistics for *R.reniformis* Genome Sequence

	# 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 of >2,000 bp in length (N50 = 2,812) covering a total of 4.7 Mb of genomic sequence.

Microsatellite Analysis

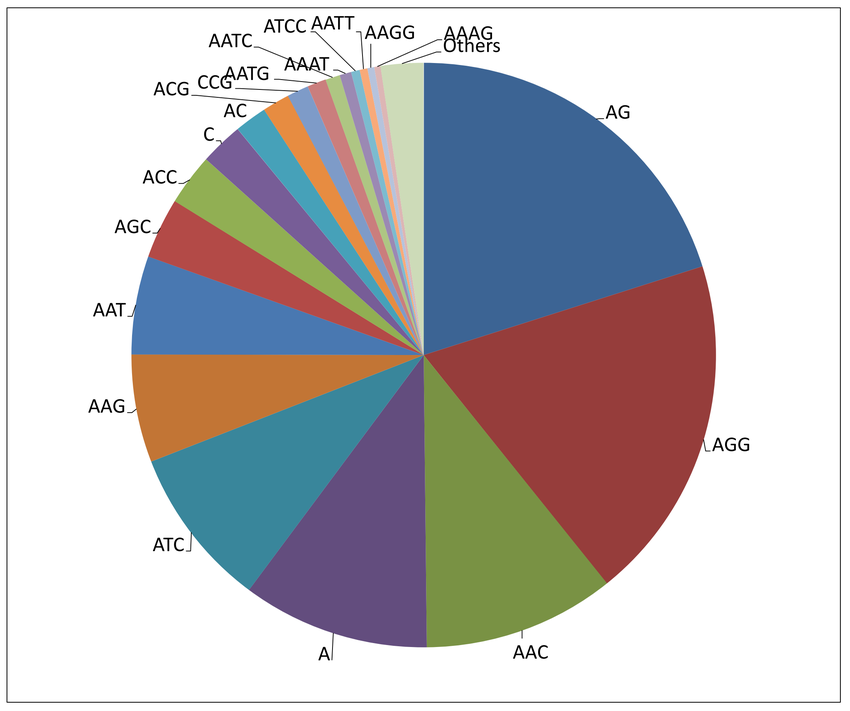

Figure 3. Frequency of microsatellite classes in the genome of *R. reniformis*. Among the 58 classes identified, the 20 most frequent are shown in individual divisions. The remaining 38 microsatellites are considered in a single division defined as others.

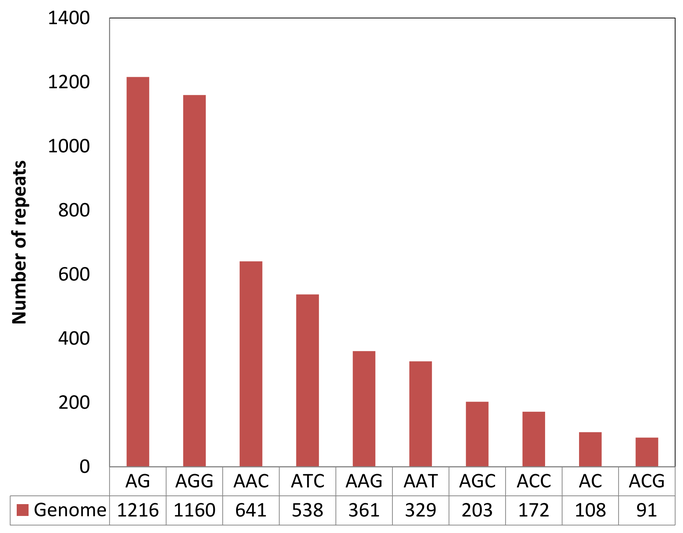

Figure 4. Frequency of the 10 most abundant repeats from our *R. reniformis* genome assembly.

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

- Cross-check the genome size estimate with other known control organisms
- More sequencing of *R. reniformis* DNA extracted and amplified from single egg
- Further refinement of the assembly using alternative assembly algorithms
- Further structural and functional annotation - ncRNAs, repeat elements, GO annotation
- Incorporation of transcriptome sequences 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 structural genome annotation

Funding
USDA-ARS Award: 5864027241-08050415