

---

# A Genomic Sequencing Approach to Study Wood Decay and Copper Tolerance in the Brown Rot Fungus, *Antrodia radiculosa*

**Juliet D. Tang**<sup>1</sup>, Andy Perkins<sup>2</sup>, Tad Sonstegard<sup>3</sup>,  
Shane Burgess<sup>4</sup>, and Susan Diehl<sup>1</sup>

<sup>1</sup> Forest Products, Mississippi State University, MS

<sup>2</sup> Computer Science & Engineering, Mississippi State University, MS

<sup>3</sup> USDA ARS Bovine Functional Genomics Laboratory, Beltsville, MD

<sup>4</sup> Life Sciences & Biotechnology Institute, Mississippi State University, MS

## What is a genome?

- All the DNA in a haploid set of chromosomes

## Why sequence genomes?

- Genes orchestrate how organisms look, grow, reproduce, and respond to their environment.
- Known gene sequence
- Determine molecular function
- Understand biological processes

# *Antrodia radiculosa*

---

- Brown rot fungus - worldwide distribution
- Rapidly breaks down cellulose and hemicelluloses
- Grows in the presence of high copper

What genes control these biological traits?

# Pipeline

---

## Steps

Sequence → Filter → Assemble → Predict Genes

## Tools

ILLUMINA  
NGS  
GAII

→

Custom  
Perl  
Scripts

→

Velvet  
0.7.55

→

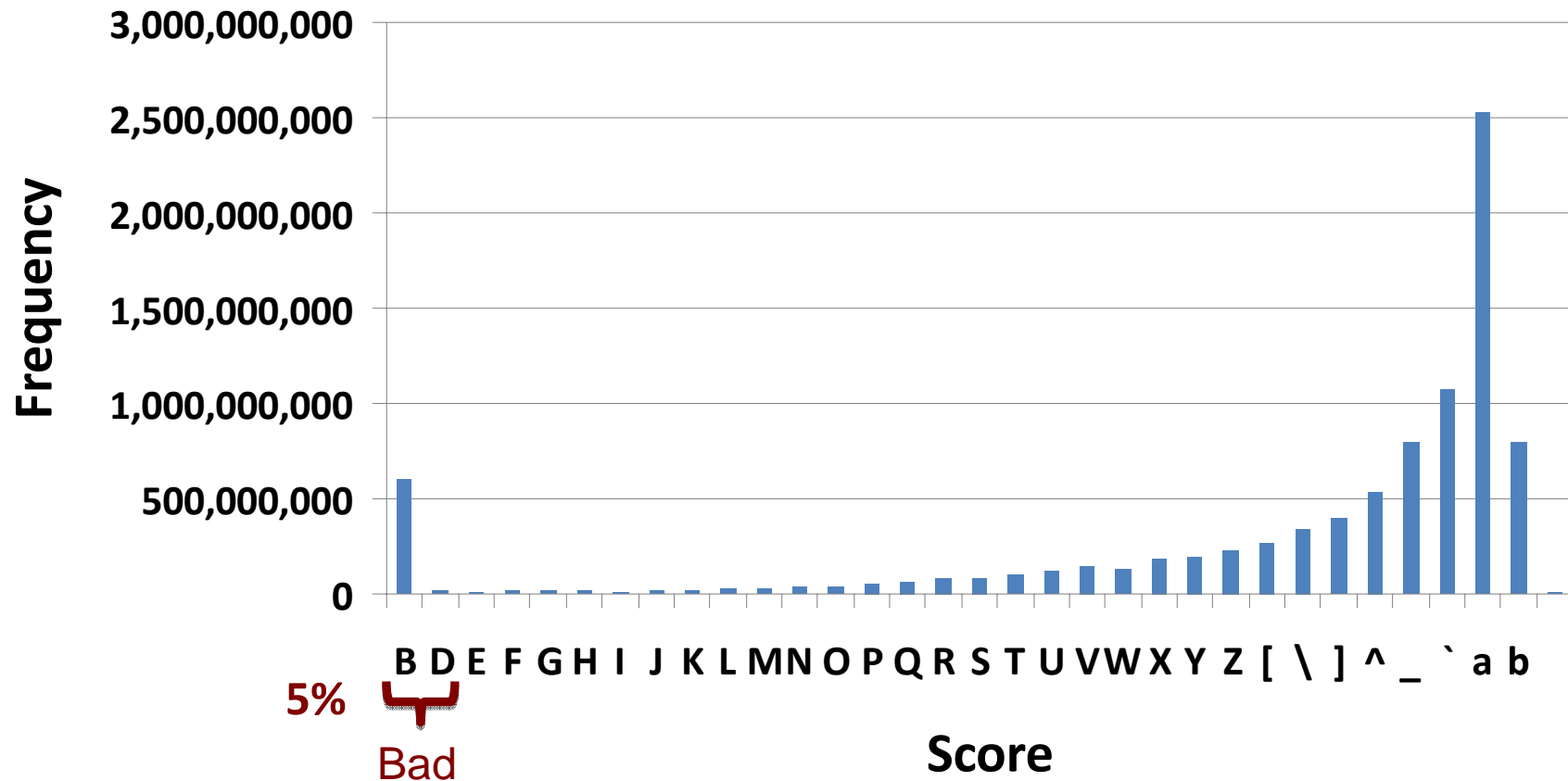
Genemark  
ES v2

# Short Read Sequencing Strategy

---

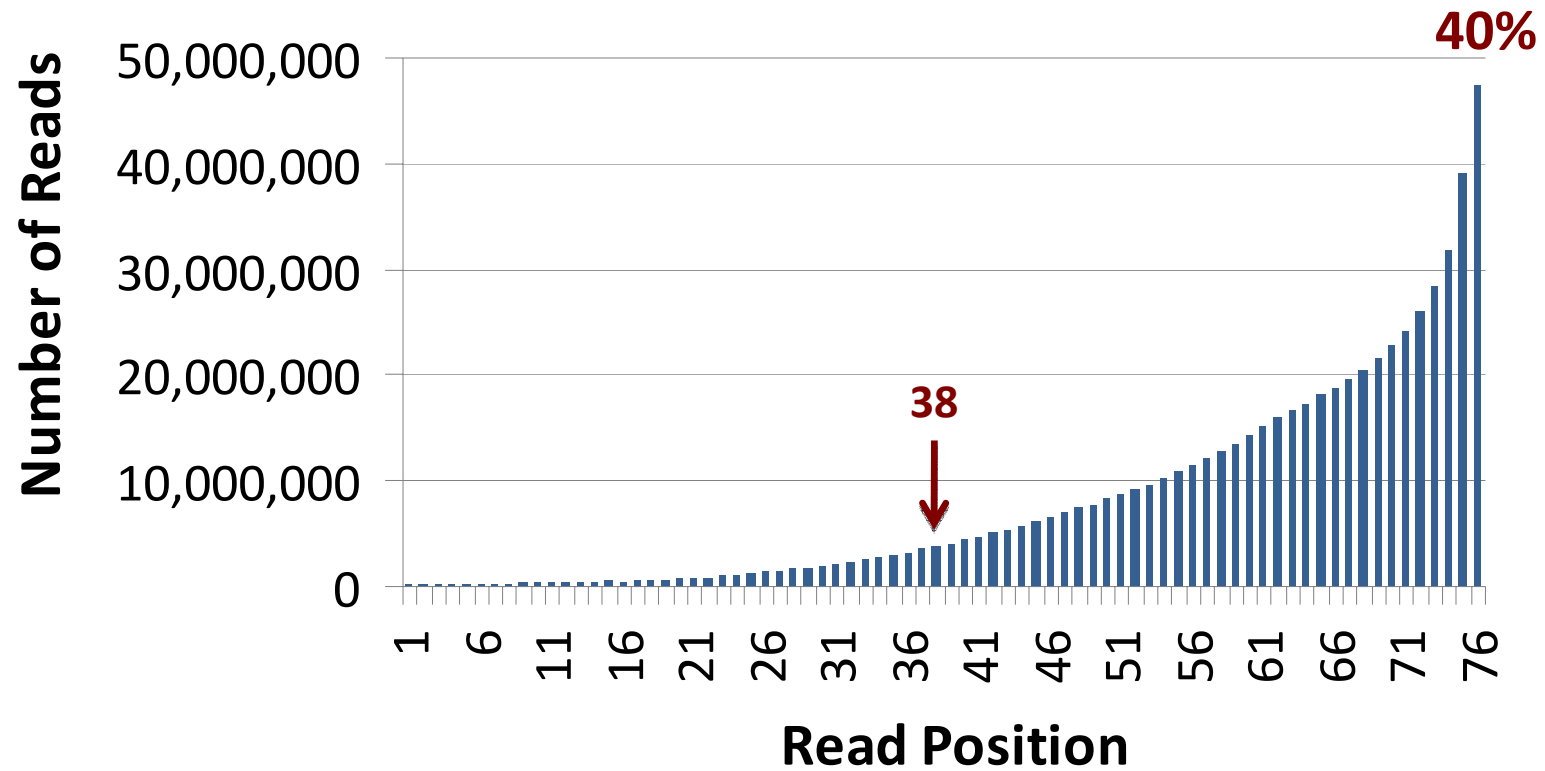
- Paired end PE
- Read length 76 nt
- Fragment size 300 bp
- Estimated genome size 30 Mb
- Number of reads for 1 flow cell  $100 \times 10^6$
- Depth of coverage  $\frac{(\# \text{ reads} \times \text{read length})}{\text{genome\_size}}$  250x
- Sequence minus mito

# Base Quality Score Frequency (9 Gb)



# Bad Scores by Read Position

---

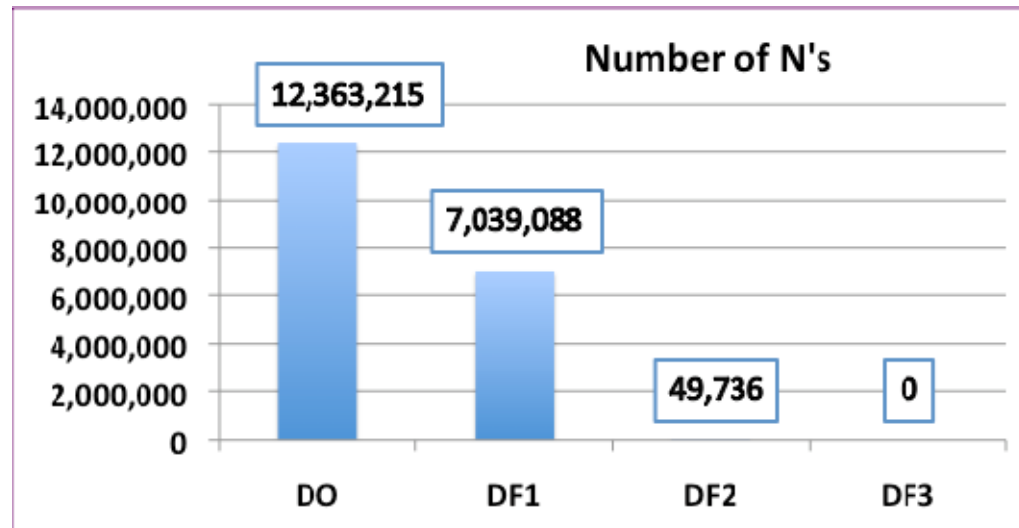


# Quality Filters

---

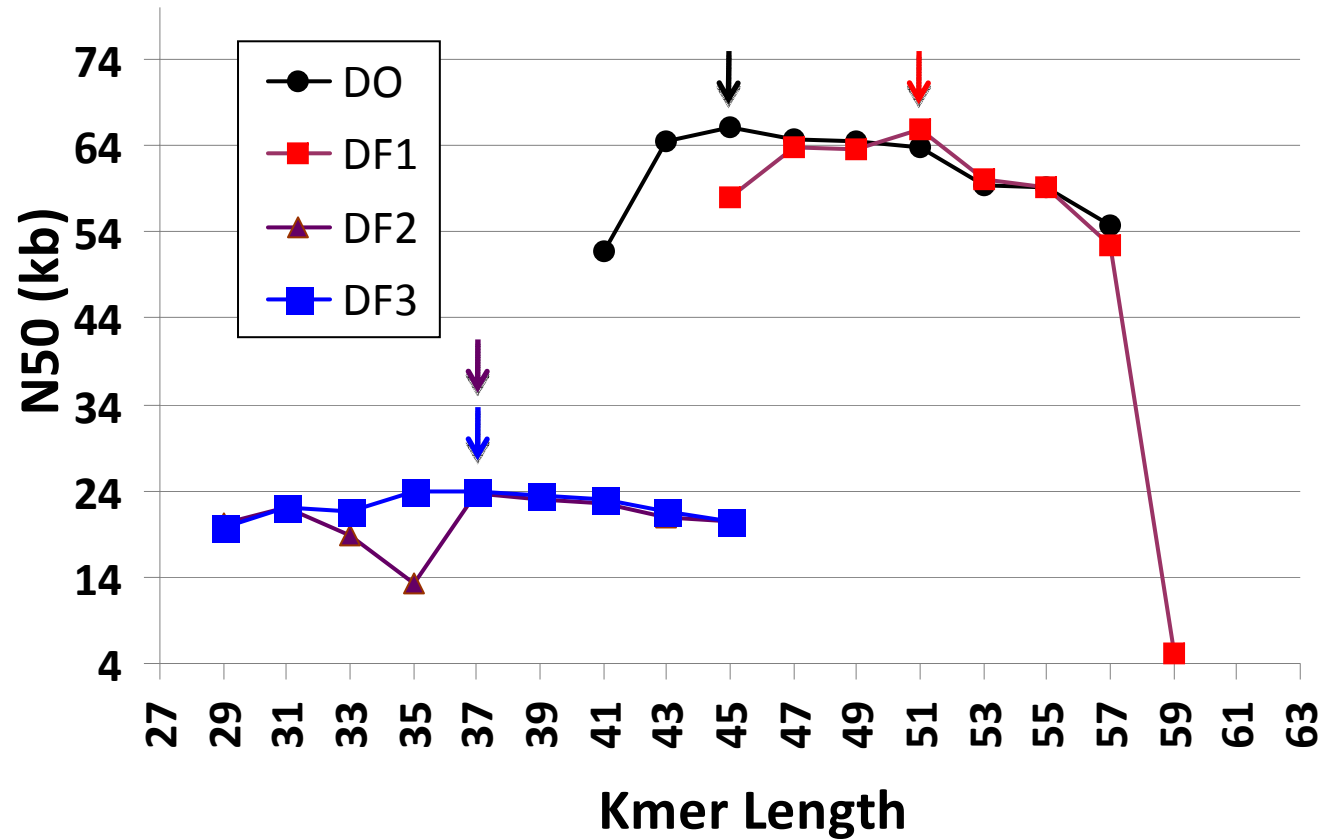
- F1: 38 or more bad scores per read
- F2: 1 or more bad scores per read
- F3: 1 or more bad scores or N's per read

<b>Dataset</b>	<b>Number of Reads</b>
<b>DO</b>	<b>117,745,354</b>
<b>DF1</b>	<b>113,780,324</b>
<b>DF2</b>	<b>62,927,297</b>
<b>DF3</b>	<b>62,878,344</b>





# Velvet Assemblies



N50 =  $\frac{1}{2}$   
assembly is  
covered by  
contigs of size  
N50 or larger

# Max N50 - Velvet Assemblies

---

Assembly	k	N50 (kb)	Coverage*	Used Reads	Unused Reads
VO	45	66.2	61.6	86,599,818	31,145,536
VF1	51	65.8	57.0	85,129,618	28,650,706
VF2	37	23.7	50.1	46,078,009	16,849,288
VF3	37	24.0	52.5	46,141,809	16,737,164

---

\* Ave contig coverage

---

# Max N50 - Velvet Assemblies

---

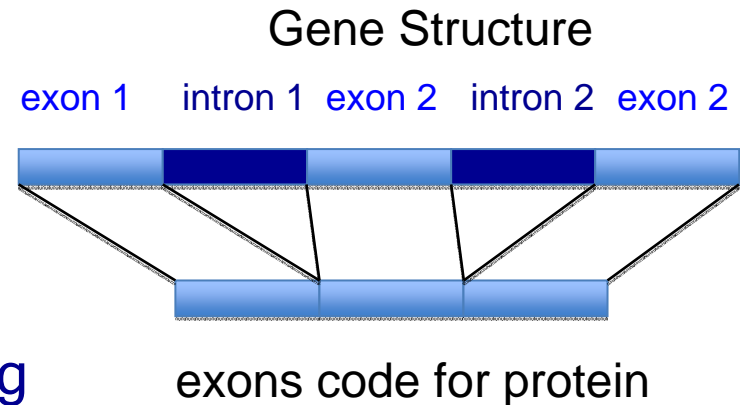
Assembly	k	N50 (kb)	Coverage*	Used Reads	Unused Reads
VO	45	66.2	61.6	86,599,818	31,145,536
VF1	51	65.8	57.0	85,129,618	28,650,706
VF2	37	23.7	50.1	46,078,009	16,849,288
VF3	37	24.0	52.5	46,141,809	16,737,164

---

# Genemark-ES v2\*

---

- HMM for gene prediction
- Specifically for fungi
  - greater sensitivity and accuracy
- Does not require dataset for training
- Train on assembly
  - contigs  $\geq 20$  kb
  - sum  $>10$  Mb



# Gene Descriptive Stats

---

	Ar (BRF)	Pc (WRF)	Pp (BRF)
genome size	33.7 Mb	30 Mb	33 Mb
gene models	8137	12,000	17,000
% GC in exons	54%	59%	.
ave exon lgth	220 bp	232 bp	.
ave intron lgth	70 bp	117 bp	.

Gene predictions  
from the optimal assembly look  
**PROMISING!**

# Summary

---

- NGS is a viable option
- Requires some filtering
- Evaluate optimal assembly
- Got 50% of genes from 1 run (< \$1.50 per gene)



80 glycoside hydrolase-like genes  
3 cellulase-like genes  
2 glyoxalate oxidase-like genes  
13 laccase-like genes.

# Acknowledgments

---

Dr. Chuan Yu Hsu  
Brandon Malone  
Linda Sites

Cathy Gresham  
Dr. Susan Bridges  
Dr. Yogi Dandass

Dr. Darrel Nicholas  
Dr. Carol Clausen  
Dr. Tor Schultz