

Genome sequence of *Stachybotrys chartarum* Strain 51-11

D.A. Betancourt^{1*}, T. R. Dean¹, Jean Kim², and Josh Levy²

¹ U.S. Environmental Protection Agency,
National Risk Management Research Laboratory,
Air Pollution Prevention and Control Division,
Research Triangle Park, NC 27711, USA,
phone 919-541-9446, fax 919-541-2157,
email betancourt.doris@epa.gov

²RTI International
3040 E Cornwallis Rd, Durham, NC 27709

Abstract: *Stachybotrys chartarum* strain 51-11 genome was sequenced by shotgun sequencing utilizing Illumina Hiseq 2000 and PacBio long read technology. Since *Stachybotrys chartarum* has been implicated in health impacts within water-damaged buildings, any information extracted from the genomic sequence data relating to toxin and allergen production or the fungi's metabolism could be potentially useful.

Stachybotrys chartarum is a saprophytic fungus with world-wide distribution. In nature, it is isolated from soil, seeds and decaying organic matter. Indoors, *Stachybotrys chartarum* grows on wet cellulose-containing building material such as gypsum wallboard, cardboard and ceiling tiles [1].

S. chartarum is among the indoor fungal population frequently identified in water-damaged buildings and it has been linked to damp building-related illnesses (DBRIs) [2]. This fungus is capable of producing mycotoxins of which the macrocyclic trichothecenes are among the most toxic [3, 4]. Likewise, several studies have shown the possible association of *S. chartarum* proteins, e.g. proteinases and hemolysins, with inflammation of the respiratory system [5, 6]. Recently, Semeiks et al reported the genomes of four *S. chartarum* strains in an effort to determine toxin chemotypes within the toxigenic strains [7].

S. chartarum toxigenic strain 51-11, an environmental isolate obtained from the cluster of idiopathic pulmonary hemorrhage cases (Cleveland, Ohio) [8], is part of the RTI International (Research Triangle Park, NC) microbial collection assigned with a unique identification number and stored at -80°C for long term preservation. Propagation and maintenance of this fungus was on potato dextrose agar (PDA) (Becton, Dickinson & Company, Sparks, MD) at 25°C. Colony morphology and microscopic observations were made to confirm the phenotype that is characteristic of this species. A single spore isolate was used for genomic sequencing. The fungal spores were grown in potato dextrose broth (PDB) (Becton, Dickinson & Company, Sparks, MD) and were shaken at 25°C. The resulting fungal mycelia were harvested using sterile mira cloth (EMD Millipore, Darmstadt, Germany) and washed with sterile distilled water. The mycelia were flash frozen in liquid nitrogen and stored at -20°C until the genomic DNA was isolated using a protocol described by Kohler et al [9].

Two genomic DNA libraries for Illumina sequencing: one paired-end library (insert size 350-390bp), and one mate-paired library (insert size 100-5000 bp) were produced. These two

libraries were sequenced on an Illumina Hiseq 2000 (Illumina, San Diego, CA) at Argonne National Laboratory (Lemont, IL). The machine produced 460x coverage, and approximately 100x of that coverage was actually used for the assembly, though the whole dataset was exploited for error correction. Genome size was assessed using Kmer-counting tools, which pointed to a genome size of approximately 40 Mbases. The resulting 40 Mb error-corrected reads were assembled using MIRA [10] on a large-memory machine. The resulting contigs were hand-edited for quality. SSPACE (Leiden, The Netherlands) [11] basic 2.0 was run on the mate pair data to try to close gaps and produce better scaffolds. The resulting assembly contained 2843 scaffolds. Higher molecular weight DNA for PacBio long-read sequencing was also produced to close gaps in the Illumina assembly. Nine sequencing runs were performed, and after standard Quality Assurance/Quality Control cutoffs were applied, there were 284,625 reads, with ~736 million bases total (~2,600 bases per read), or ~18x coverage. Two hybrid assembly methods were compared, SSPACE-LongRead and PBJelly [12]. Multiple combinations using both programs were performed on a supercomputing cluster, and quality of assembly was assessed using standard metrics (N50, N90). Three iterations of SSPACE-LongRead alone (not in combination with PBJelly) showed the highest quality assembly, with the final number of scaffolds at 1,435, or approximately half of the Illumina assembly alone.

Nucleotide sequence accession numbers: submitted to GenBank. (Accession numbers pending)

Acknowledgements: We are grateful to Grace Byfield and Anthony Devine (RTI International), for their effort and advice during the first stages of this project.

Disclaimers: The views expressed in this article are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency."

"Any mention of trade names, products, or services does not imply an endorsement by the US Government or the United States Environmental Protection Agency. EPA does not endorse any commercial products, services, or enterprises."

References

1. **Chapter 1 : Identification of the common food and airborne fungi.** In: *Introduction to food and airborne fungi*. Edited by Samson R.A., Hoekstra E.S.; Frisvad, J.C.; Filtenborg, O., 6th edn. The Netherlands: Centraal Bureau Voor Schimmelcultures; 2002: 1-282.
2. **Institute of Medicine (IOM): Damp Indoor Spaces and Health:** National Academies Press , Washington DC; 2004.
3. Jarvis BB: *Stachybotrys chartarum: a fungus for our time. Phytochemistry* 2003, **64**(1):53-60.
4. Pestka JJ, Yike I, Dearborn DG, Ward MD, Harkema JR: *Stachybotrys chartarum, trichothecene mycotoxins, and damp building-related illness: new insights into a public health enigma. Toxicological Sciences* 2008, **104**(1):4-26.
5. Gregory L, Rand TG, Dearborn D, Yike I, Vesper S: **Immunocytochemical localization of stachylysin in *Stachybotrys chartarum* spores and spore-impacted mouse and rat lung tissue.** *Mycopathologia* 2002, **156**(2):109-117.

6. Yike I: **Fungal proteases and their pathophysiological effects.** *Mycopathologia* 2011, **171**(5):299-323.
7. Semeiks J, Borek D, Otwinowski Z, Grishin NV: **Comparative genome sequencing reveals chemotype-specific gene clusters in the toxigenic black mold *Stachybotrys*.** *BMC Genomics* 2014, **15**:590-605
8. Vesper SJ, Dearborn DG, Yike I, Sorenson WG, Haugland RA: **Hemolysis, toxicity, and randomly amplified polymorphic DNA analysis of *Stachybotrys chartarum* strains.** *Applied Environmental Microbiology* 1999, **65**(7):3175-3181.
9. Kohler, A, et al: **Convergent losses of decay mechanism and rapid turnover of symbiosis genes in mycorrhizal mutualists.** *Nature Genetics* 2015, **47**:410-415.
10. MIRA: Open source: <http://genome.cshlp.org/content/14/6/1147.full>
11. SSPACE: <http://www.baseclear.com/genomics/bioinformatics/basetools/SSPACE>
12. PBJelly: Open source: <http://www.biomedcentral.com/1471-2105/15/180/abstract>