Next generation *de novo* sequencing of the *L. raphanica* genome

Alyson Askew, Justin Roberts, Michael Nguyen, Bhoomika Vakharia, Brittany Wallace, Juan Mata, Glen Borchert

[1]Department of Biology, University of South Alabama, Mobile, AL

Abstract

Recently, the commercial Shiitake mushroom, Lentinula edodes, has been sequenced. However, the genome of the Gulf Coast's American Shiitake mushroom, or Lentinula raphanica, remained undefined. We undertook and have successfully sequenced the American shiitake mushroom and are in the process of comparing it to the already sequenced Gymnopus luxurians. So far we have obtained 4.8 billion base pairs of genomic sequence, which considering that Lentinula edodes genome is known to be ~40.8 million base pairs, we have now sequenced the genome ~120 times. We have found the genome to be ~45% GC rich and to have a 99.97% rate of base call accuracy with only a 0.1% rate of unidentified bases. We have recently begun assembling the scaffolds into larger pieces to have a more continuous genome. We currently have ~80,000 scaffolds and hope to narrow this down to ~60 in the next few weeks by blasting them against the genome of Gymnopus luxurians. The genome is also currently being blasted against known protein coding genes and other known fungal genes and will be annotated with the findings in the near future. We also will define the differences between the Lentinula raphanica genome and the genomes of other closely related species. We will also focus on finding genome characteristics specific to Lentinula raphanica. Ultimately, we plan to publish our findings in the online journal PlosOne.

Methods

L. Raphanica culture DNA extraction and verification

American Shitake mushrooms were collected in the wild and cultured in the lab. The stock cultures were given approximately three weeks for the mycelium to exhibit growth. Each Lentinula strain was inoculated onto three plates of each medium, resulting in 15 plates per strain (3 replicates X 5 media). After all of the agar plates were prepared, they were inoculated with mycelial discs. After the inoculations were complete, all of the plates were placed in an incubator set at 25˚C and grown until the mycelium reached the edge of the plate or after 40 days.

DNA identity was confirmed by performing PCR based enzyme digest analyses (Fig 1) (Dunning et al, 2007). A pure culture was grown on nutrient agar for each of the strains and inoculated on liquid potato broth for two weeks. Once there was enough mycelium present, it was filtered and weighed for DNA extraction. DNA extraction was then performed on each of the samples using protocols modified from the guidelines in White et al. (1990). After DNA extraction was successful, a PCR was performed to amplify a marker sequence capable of differentiating fungal species. Restriction enzyme digests with HINFI and HINDIII (data not shown) were performed to differentiate between the different species in Lentinula. Products were run on a standard 1% agarose gel and imaged.

Next generation sequencing

The results here inscribed are derived from the use of Otogenetics, a commercial sequencing company based in Norcross, GA. American Shiitake DNA was quantified using the Qubit (Life Technologies, Bethesda, MD) fluorometer. Next, purified DNA was fragmented and primed using the NEBNext Fast DNA Fragmentation and Library Prep Set (New England Biolabs, Ipswich, MA) per standard manufacturer protocol. Following this, DNA was re-quantified using the Qubit and diluted to the specifications of Otogenetics' commercial sequencing protocol. The sample was sent to Otogenetics for sequencing. Once received, the raw data was assembled into contigs on DNAStar's Lasergene program.

Genome assembly

The contigs were assembled into scaffolds using SOAPdenovo2. The scaffolds were also run through the Genetic Information Research Institute's CENSOR masking software. The scaffolds were then blasted against protein coding genes, themselves, and other closely related fungal species' genomes.

Results

L. Raphanica DNA confirmation

The American Shitake mushrooms that were collected in the wild were confirmed by PCR based enzyme digest using HINFI (Figure 1) to be *Lentinula raphanica*.(Dunning et al, 2007).

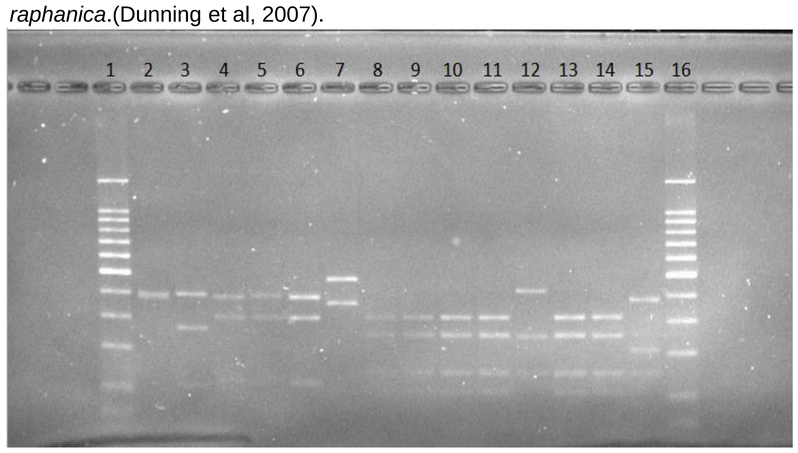

Figure 1. Fungal Genome HINFI digestion. To confirm species identity, ribosomal DNA was amplified from isolated fungal genomes using ITS 4 and ITS 5 primers then the product was digested with HINFI and run on a 2% agarose gel. Composition of each lane is as follows: 1, DNA Ladder; 2, TFB 2306 *L. edodes*; 3, TFB 7563 *L. novozealandica;* 4, TFB 8467 *L. boryana*; 5, TFB 7829 *L. boryana*; 6, TFB 8421 *L. boryana*; 7, TFB 8436 *L. boryana*; 8, JLM 1620 *L. raphanica;* 9, JLM 1647 *L. raphanica*; 10, JLM 1636 *L. raphanica*; 11, JLM 1623 *L. raphanica*; 12, TFB 9929 *L. raphanica*; 13, JLM 1587 *L. raphanica*; 14, TFB 9564 *L. raphanica*; 15, TFB 9447 *L. aciculospora*; 16, DNA Ladder.

Genome sequencing

The sample sent to Otogenetics resulted in 48 million paired reads with reads running approximately 100 base pairs in length, resulting in the data obtained equaling 4.8 billion base pairs. Sequence analysis found that Lentinula raphanica's genome contains ~45% GC base content. (Figure 2)

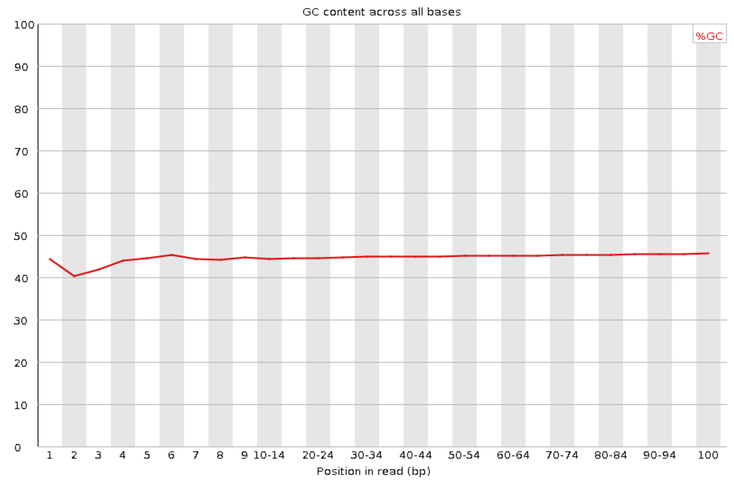

Figure 2. Genome GC content determined from 4.8 billion base pairs of sequencing reads. 45% on average GC content shown here translates to approximately 2.16 billion G and C bases throughout all of the reads.

Genome sequencing continued

Approximately 37% of sequences were found to be duplicates. Sequence integrity was particularly high with n or unidentifiable reads at less than 0.1% and the Phred score was found to be approximately 37. (Figure 3) Assembling the genome in SeqMan (Lasergene) produced 16572 Contigs between 223 and 31379 bp long.

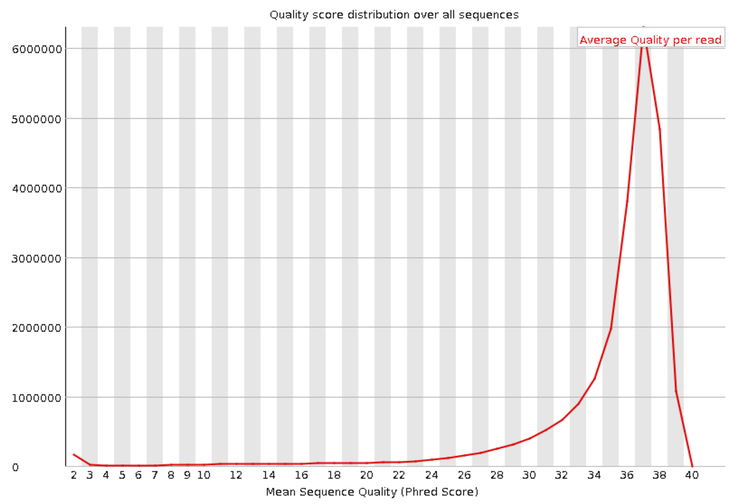

Fig. 3 Representation of the Phred score of the raw sequence reads. This chart shows a Phred score of 37, which translates to an estimated 99.98% base call accuracy reading. This was calculated by deriving the formula of e=10^(-Q/10) from the formula given for Q in the Illumina protocol, where e is the probability of error and Q is the Phred score. The probability of error multiplied by 100 gave the percent error, and subtracting that from 100% gave the percent accuracy depicted here.

Sequence Assembly

We currently have ~80,000 scaffolds. We also have a database of known protein coding genes and other fungal genes pulled from Ensembl that the genome is currently being blasted against.

Conclusions

- In all, we sequenced 4.8 billion base pairs. The *Lentinula edodes* (Shiitake mushroom) genome is 40.2 Mb (Kwan et al, 2011). Based on this, we estimate that we have sequenced the *Lentinula raphanica* genome approximately 119 times.
- We found the genome consists of ~45% GC content.
- We found ~37% of the returned sequences were duplicates.
- Unidentifiable reads were found at less than 0.1%.
- The Phred score of 37 translates into a 99.98% base call accuracy.
- We currently have ~80,000 scaffolds.

Future Directions

- The scaffolds will be blasted into larger scaffolds using *Gymnopus luxurians* as a reference genome.
- These scaffolds will be analyzed for unique genes to *L. raphanica* as well as for similar genes to the reference genome.
- Our final genome will be submitted to ensemble.org and we will compile our final findings in a comprehensive report to be submitted for publication in either PlosOne or Genome Biology.

ACKNOWLEDGEMENTS

The authors would like to thank UCUR and the College of Arts and Sciences of the University of South Alabama for their support in this work.

Literature Cited

Alexopoulos, C. J., C. W. Mims, et al. (1996). Introductory Mycology. New York, USA, J Wiley & Sons.
Dunning, K., A. Morris, and J.L. Mata, *RFLP Analyses of the Mushroom Lentinula*, 2007, University of South Alabama.
Grigoriev, I. V., H. Nordberg, et al. (2012). "The genome portal of the Department of Energy Joint Genome Institute." Nucleic Acids Res 40(Database issue): D26-32.
Kwan, H. S., C. H. Au, et al. (2011). Genome sequence and genetic linkage analysis of Shiitake mushroom Lentinula edodes. Fungal Genetics Conference.
Lee, C. C., D. W. Wong, et al. (2005). "Cloning and characterization of the xyn11A gene from Lentinula edodes." Protein J 24(1): 21-26.
Quail, M. A., M. Smith, et al. (2012). "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers." BMC Genomics 13: 341.
Scheffers, B. R., L. N. Joppa, et al. (2012). "What we know and don't know about Earth's missing biodiversity." Trends Ecol Evol 27(9): 501-510.
Taylor, J. W., J. Spatafora, et al. (2004). Assembling the Tree of Life. USA, Oxford University Press.
Wasser, S. P. (2005). Shiitake (Lentinus edodes). Encyclopedia of Dietary Supplements. New York, New York, Marcel Dekker, Inc.: 653-664.