

Nanopore ITS Barcoding: 10x more throughput for 10x less cost

By: Stephen Russell – steve@hoosiermushrooms.org

The Hoosier Mushroom Society has been working with Sanger sequencing since 2015 in order to achieve a reasonably comprehensive understanding of all of the macrofungi that exist within the state. To date, the effort has produced over 10,000 specimens with ITS barcodes, representing well over 2,200 unique species. In recent months, however, we have transitioned away from Sanger and have been working exclusively with Oxford Nanopore Technologies (ONT) [MinION](#) and [Flongle](#) devices for all our DNA barcoding efforts.

The protocols for ONT nanopore sequencing are similar to Illumina in many ways. DNA barcoding with nanopore requires DNA extraction and amplification of individual target loci, just as traditional Sanger sequencing would. The resulting amplicons are ultimately pooled (mixed together) for sequencing. This means that each amplicon needs to be individually tagged with dual primer indexes (13bp unique barcodes on the forward and reverse primer for each sample), so the results can be disaggregated (demultiplexed) for analysis. Sequencing adapters are ligated to amplicons within the pool and the sequencing can commence on a flowcell within a small, portable device connected to a laptop computer. A typical sequencing run with a Flongle device may produce 750,000 – 1,500,000 raw ITS reads in 24 hours. These reads are then processed through a bioinformatics pipeline that calls the bases, filters for quality, and joins reads into a final “consensus” sequence.

When we started pursuing this methodology, there were very few protocols available for incorporating hundreds of specimens on a single sequencing run. There were even fewer protocols available with a focus on fungal amplicons. At the link below, you will find a complete set of working protocols (Version 1) for DNA barcoding fungal amplicons with ONT MinION and Flongle devices. It covers computer/program/dependency setup, DNA extraction, amplification, library prep, sequencing, and primary data analysis, including a full cost breakdown.

dx.doi.org/10.17504/protocols.io.36wgq7qykvk5/v1

Utilizing the nanopore methodology outlined here has several benefits over traditional Sanger sequencing. The ultimate cost per sample has dropped to less than \$0.60 - more than a 10x cost decrease from our previous Sanger protocols. This cost per sample could be further decreased by creating your own Extract-N-Amp-style reagent and ordering master-mix in bulk. Extraction and amplification reagents are the costliest aspect of the protocol (per sample). Another key benefit to nanopore vs. Sanger is the time savings. This protocol does away with amplicon validation through gel electrophoresis. The resulting barcodes are cheap enough that successful PCR is validated by examining whether a high-quality result is obtained.

Time savings are also significant post-sequencing. Manual sequence editing is not required with this protocol. Final results are presented for each specimen through a bioinformatics pipeline, rather than manually editing bases. The error rate for individual base calls from MinION devices has been a matter of concern for this type of sequencing since the introduction of the devices in 2014. That was also the primary concern that delayed our adoption of this technology. However, after examining hundreds of results from the ONT protocol, compared with Sanger sequences of the same species, more than 94% of the ONT consensus sequences from the pipeline matched Sanger sequences with 99.5% or greater BLAST identity similarity. A majority (60%+) of the results match one of our Sanger reads at 100%

identity (with 80%+ query coverage). The error rate from MinION devices is not a significant concern for our ongoing efforts, when building consensus sequences from depth of at least 100 reads. 100% matches have been acquired with as few as eight reads.

It should also be noted that the ultimate sequencing depth limits for MinION and Flongle devices have yet to be established for DNA barcoding. We are now using 960 specimens on Flongle 9.4.1 cells. This typically produces 200-450 quality-filtered reads that form the ultimate consensus sequence. It may still be possible to double the number of specimens per flowcell, with the 9.4.1 Flongle cells. Fitting more specimens on a flowcell decreases the overall cost per sample. Flongle 9.4.1 cells are currently \$100 each. Further, ONT will soon be releasing a Flongle 10.4.1 flowcell, which is likely to reduce the error rate for basecalling. This will allow for a further reduction in the number of reads that are necessary to form a high-quality consensus sequence.

Another interesting benefit to nanopore sequencing is that previous Sanger reads that may have failed due to contaminants are now often successful through nanopore. Two (or more) consensus sequences are obtained in the results for the specimen, rather than one muddled Sanger read. There are a number of situations where this has been particularly useful. As an example, when examining several *Entoloma abortivum* specimens, we received results validating the target, but also a second sequence validating the host – in this case *Armillaria gallica*. As another example, specimens that are beginning to succumb to mold can still be successfully sequenced. *Tylopilus rubrobrunneus* returned results with the target, as well as *Hypomyces chlorinigenus*.

The ultimate limiting factor for our Indiana barcoding efforts is now the number of modern specimens available – not cost or time. As this protocol uses a “quick” extraction method, it is most well suited for recent collections, rather than older herbarium specimens. The primary aspect that limits whether a sequencing result is obtained is often the quality of the extraction protocol employed. A trial of 2 plates (192 specimens) utilizing a modified Promega Wizard extraction kit returned sequences for 100% of the specimens. The “quick” extraction protocol we primarily use has 80-90% success rates. The most common points of failure are specimens with recalcitrant fresh material – polypores and crusts. For best results, specimens of this type should be moved into a secondary pipeline which utilizes an extraction protocol that involves grinding the material. A secondary factor for success of an individual specimen is primer selection. Ex – the ITS1F-ITS4 primer combination remains unlikely to amplify most *Cantharellus* specimens, despite the sequencing technology being employed.

In order to overcome the limitation on quantity of specimens, we have been focusing on efforts to further integrate citizen scientists into our collection efforts. By far, the most effective methodology has been online forays. Twice a year – summer and fall – we hold week-long virtual forays on iNaturalist.org. Individuals from across the state will hunt in their own localities, dry specimens they find, and mail them to me for triage. Recent events typically bring in over 1,000 specimens each, with our [Online Summer Foray 2022](#) (OSF2022) bringing in over 1,500 specimens. Combining ONT sequencing with a citizen science pipeline for modern collections is critical for making full use of the potential of the technology for DNA barcoding. A preprint of the methods we use for our online forays can be found here: <https://doi.org/10.1101/2022.05.24.493314>

In a short timeframe, ONT sequencing - combined with citizen science integration - has fundamentally changed our capacity to document the biodiversity of Indiana. We are on a pace to barcode 5,000 new

collections in six months time. We are looking forward to continual utilization of ONT protocols and hope that other groups will join us on continued protocol development and improvement.