**Kennedy Lab Molecular Ecology Protocol v1.3**

***Illumina sequencing of fungal ITS1 amplicon libraries***

**Nhu Nguyen & Peter Kennedy**

*Updated April 22, 2015*

**I. Materials**

1. Good DNA from samples
2. General PCR reagents
3. Multiplex barcoded and MiSeq Illumina sequencing primers

**II. Library preparation**

1. Raw sample preparations
2. DNA extraction and clean-up
3. PCR primer plate set-up
4. PCR parameters
5. PCR clean-up
6. Library preparation
7. Organization of your samples
8. Library submission

**III. Sequence library processing and analysis**

1. Demultiplexing samples
2. Quality filtering
3. Downstream processing and analyses
4. OTU identification for fungal studies
5. Parsing out fungal guilds

**IV. References**

**I. Materials**

1. *Good DNA from samples*

This is crucial to producing good PCR products and saving downstream troubleshooting time during PCR and library preparation. If you’re working from a MoBio soil kit, run your samples through a NanoDrop or plate reader that will show absorbance at 260/280 and 260/230. A DNA sample is considered pure if the 260/280 ratio is ~1.8 and 260/230 ratio is between 1.8-2.2. If the ratio of 260/280 is appreciably lower, it may indicate the presence of protein, phenol or other contaminants. If your samples do not show a good 260/280 ratio, run them through any other DNA purification kit. If an absorbance spectrophotometer is not available, it’s a good idea to run all your extracted DNA samples through a purification kit.

2. *General PCR reagents*

Regular Taq polymerase introduces too many mistakes during polymerization. Therefore, use a high fidelity polymerase with proof-reading function and low error rates. We recommend a high fidelity polymerase such as KAPA HiFi polymerase. Extra MgCl2 (0.5M) and/or BSA (10mg/ml) may be required to minimize the effects of PCR inhibitors in some DNA samples.

3. *Multiplexed barcoded primers & MiSeq Illumina sequencing primers*

There are two sets of sequencing primers – the PCR primers and the MiSeq Illumina sequencing primers. The figure below from Smith & Peay (2014) shows the primer set-up and graphically shows them in relation to the ITS1, our targeted gene region. Note that the primers were designed for the MiSeq platform.

The barcoded primer list can be found at <http://www.cbs.umn.edu/labs/kennedy/protocols>



**Figure from:** Smith, D.P., Peay, K.G., 2014. PLoS One 9, e90234. doi:10.1371/journal.pone.0090234

PCR primers

The forward PCR primer contain an Illumina adapter, linker, and gene primer

The reverse PCR primer contains an Illumina adapter, barcode, linker, and gene primer

All samples will receive the same forward primer but each sample will receive a different barcoded reverse primer. The barcodes are GOLAY error correcting 12bp barcodes.

These primers should be ordered in separate tubes from any primer company such as IDT. There is a significant amount of cross contamination with a plate order so it’s best to avoid ordering whole plates. For barcoded primers (reverse), order at 25nmol scale. For non-barcoded primer (forward), order at 100nM scale so that you will have enough to use for many reactions. Standard desalting should be good enough for PCR. Once received, resuspend the primers in TE buffer to 100μM.

Illumina sequencing primers

The user must provide the sequencing center with sequencing primers. There are three sequencing primers: Read 1, Read 2, and Index. Read 1 primer will sequence one direction, Read 2 primer will sequence the reverse direction, and Index primer will sequence the barcode. Provide the sequencing center with the requested amount of each in a separate tube. The BMGC will require 5μl of 100μM sequencing primers.

These primers should be ordered in separate tubes from any primer company such as IDT. Request HPLC purification. HPLC purification requires a large amount of product to begin with and the yield is minimal. Order the primers at 250nmol scale so that you will have enough for multiple runs in the future. Resuspend the primers in PCR grade water to 100μM.

**II. Library preparation**

1. ***Raw sample preparations***

Field samples should be processed according to the requirement of the project and the type of organisms involved. For ectomycorrhizal work, we typically process three types of samples: soil cores, in-growth bag, and whole root system.

* For soil cores – keep the soil cool and process them as quickly as possible from the field. Soil should be mixed well in the sampling bag, then sieved by using a disposable sieve set following the Kennedy Lab protocol. A designated amount of sieved soils should be weighted out and used for extraction. If extraction cannot be performed during the same day, freeze the sample until they can be extracted. Avoid freeze thaw cycles.
* For in-growth bags – keep the bags cool and process as quickly as possible. It is not recommended that the bags be frozen before processing.
* For whole root systems – follow the method by Nguyen & Bruns (in prep). In short, wash the root system free from soil as much as possible. Then in a small tray of water, remove all the mycorrhizal root-tips by gently rubbing them between your fingers. Filter out the mycorrhizal root-tips, dry them at 40°C for an hour or until well dehydrated, crush them gently to mix the root-tips together.
1. ***DNA extraction and clean-up***
* For soil cores – follow the methods described by the UMFuN sequencing initiative ([Song et al. 2015. PLOS ONE 10: e0127234](http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127234)). The MOBIO soil extraction kits generally produce good DNA for PCR amplification.
* For in-growth bags – follow the hyphal floating method ([Branco et al. 2013. PLOS ONE 8:e78295](http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078295)). In short, pour the contents of the ingrowth bags into 10 ml of ddH2O, then vortex. Use 2 ml of the supernatant, pellet the fungi, and run the pellet through any tissue DNA extraction kit. The Sigma extraction protocol is one cheap and easy approach.
* For whole root systems – follow the method by Nguyen & Bruns (in prep). In short, weigh out 20 mg of the dried and mixed root-tips, bead beat the tissue into powder, resuspend in 2% CTAB buffer with 2% PVP-40, and extract using chloroform followed by isopropanol precipitation and ethanol dehydration.

As recommended above, the DNA should be as clean as possible to avoid trouble downstream. Quantify and check the quality of the DNA. If the quality is not good, run the DNA again through any clean-up kit. Soil samples often have PCR inhibitors and should be cleaned and quantified.

To minimize PCR bias, all samples should contain about the same amount of DNA. 25ng/μl is very good for PCR. Dilute or concentrate your DNA to get an even volume for all samples.

1. ***Organization of your samples***

It’s crucial to keep good records of which sample was amplified with which barcoded primer. The best way to keep track of this is to use the “mapping file” that is also used later in downstream QIIME analyses. See the documentation on how to produce this file at (<http://qiime.org/documentation/file_formats.html>).

1. ***PCR primer set-up***

The PCR primers that you ordered will most-likely come in dehydrated form. If you have not done so already, resuspend each tube to 100uM with TE buffer. Prepare the primer mix by first diluting each primer set (barcoded and non-barcoded), then mixing them together. Strip tubes with individual caps are best for this. Even though plates are more convenient, they are not recommended because the possibility of cross contamination is too high during primer and PCR preparations.

Generally, you should only mix enough primers for several PCR reactions because long-term storage causes evaporation and will skew primer concentration. Too much primers in a PCR reaction causes primer dimers and will get sequenced, wasting chip space that could otherwise be used for actual sequences. Because of the size of these dimers (due to longer primers used), they are difficult to get rid of without a precise method of band selection such as gel excision.

The standard mix for 10 μl of 10μM primers:

1 μl, 100 μM barcoded primer

1 μl, 100 μM non-barcoded primer

18 μl TE buffer

1. Make a master mix of your non-barcoded primer in a 2 ml eppi tube. If you are preparing a 96 primer samples, and want 10ul of primers per tube, mix: 1746 μl of TE and 97 μl (+1 extra) of 100 μM non-barcoded primer. Mix well. Dispense 10ul of the dilute primer into each tube.

To each of the tubes, add 1 μl of 100 μM barcoded primer. This should give you a total of 20 μl of 10μM primers, which is enough for 28 PCR reactions.

1. Make a primer map (see “*3. Organization of your samples*”) so that you know which barcoded primer went into which tube. This step is essential because any mistakes here will affect the identity of your sequenced sample later on. Make sure there are NO mistakes. If you think there is even a slight possibility that a mistake happened, start over. You have plenty of primers to play with. The other thing to watch out for is cross contamination. If for any reason cross contamination occurred, start over with a fresh batch!
2. Primer storage

Store all your primers at -20°C. Remember that the dilute primers should not be used for too long. A fresh batch should be made every once in a while.

1. ***PCR parameters***

Amplicon PCR is full of biases so we have considered a range of ways to minimize these biases. Smith & Peay (2014) suggested that single amplification using a low annealing temperature versus pooled samples have very similar richness, evenness values. Based on these results and our own studies, we recommend using a low annealing temperature for a single PCR per sample. You will also notice that we’re running a lower number of cycles (30 vs. 35). With more cycles, there is a higher chance of amplifying contaminants.

At least two extra control samples should be included with your library. One sample is the mock community of fungi (positive control) and the other is the PCR control (negative control). Ideally some negative controls at the extraction stage can also be included – see [Nguyen et al. 2015](http://onlinelibrary.wiley.com/doi/10.1111/nph.12923/full) for a more in-depth discussion on mock communities.

The mock community serves as an anchoring point in which we know the identity of all fungi that have been put into them. This allows us to assess and normalize run-to-run variations from each library and determine the best clustering value for OTU determination. The mock community is built modularly, so species could be added or removed depending on the version. The Kennedy Lab keeps track of these versions and improve each as necessary.

The negative control sample allows us to determine which OTUs contaminated our samples and they could be dealt with prior to statistical analyses. Each PCR reaction you run should include a negative control.

PCR for a 20 μl reaction:

ddH2O 11.8 μl

10X Buffer w/MgCl2 2.0 μl

10X dNTP mix 2.0 μl

MgCl2 (50mM) 0.64 μl

Forward primer (10μM) 0.35 μl

Reverse primer (10μM) 0.35 μl

HotStarTaq polymerase (5U/ml) 0.16 μl

Template (25ng) 2.0 μl

PCR cycle:

1) 95 °C, 10 min

2) 95 °C, 30s

3) 52 °C, 20s

4) 72 °C, 30s

5) go to step 2 (29x)

6) 72 °C, 8 min

7) 12 °C, forever

1. ***PCR clean-up***

Once you have determined that your PCRs are good through gel electrophoresis, the next step is to clean them up. If you have run replicate PCRS of each sample, combine those replicates into one.

PCRs may be cleaned up using any preferred method, but one of the quickest methods is using magnetic beads, such as the AMPure XP beads from Beckman Coulter. The magnetic beads work with any magnetic stand, but one of the easier ones to use is the two strip-tube DynaMag - PCR stand from Invitrogen/Life Technologies (part no. 49-2015). This stand allows you to clean 16 PCR reactions at a time, and you can actually see what’s going on in the tube vs. the plate magnetic stands.

For good and strong bands on the gel, purify 10 μl. For weaker reactions, you may need a little bit more. For the negative controls, combine them all into a single sample, then clean 30 μl. Follow the manufacturer’s instructions. Resuspend your PCR product in 35 μl PCR quality water, but pipe out only 30 μl, leaving the extra 5 μl at the bottom of the tube that is often mixed with magnetic beads.

Quantify the amount of each cleaned sample you have by using a Qubit or a plate reader with fluorescent capability. Note that a NanoDrop does not give an accurate quantification of PCR products so don’t use it here.

\***Important note:** if you are running samples with the same barcodes as used in previous studies in the same lab, be careful about contamination at this stage. We recommend that you take precaution (wiping pipettes down with nucleases, using filtered tips) to avoid that possibility. See Nguyen et al. (2015) for more details.

1. ***Library preparation***

Once you have all the samples quantified, fill out the concentration in the excel spreadsheet called [PCR\_dilutions.xlsx](http://www.cbs.umn.edu/labs/kennedy/protocols). This spreadsheet will give you the amount you need to add from each sample so that they are combined in equimolar concentration. The spreadsheet give you several columns and final concentrations, but for Illumina, it’s only necessary to use the 25 ng/μl column. Combine all products according to the values given in the PCR dilution spreadsheet. For the negative control, add the amount that is the average volume of all of the PCR products you’ve added.

Once you have combined all the products together into one tube, quantify the amount again to make sure that you have about 10 ng/μl. This is the minimum amount needed for the BMGC. If it’s lower, you must concentrate the sample by either running part of it again through a column or using the AMPure kit. If your sample is higher than 10 ng/μl, don’t worry about it. Before running any samples, the BMGC staff will do a much more sensitive quantification, dilute the product to the proper concentration, and run CAPA analysis to make sure your library works before running it on the Illumina chip.

Once you have the appropriate concentration, aliquot 20 μl into a clean tube and keep it frozen until you’re ready to submit. Keep the rest of your samples frozen for backup.

1. ***Preparing sequencing primers***

Once you have received the three Illumina sequencing primers, dilute them to 100 μM in PCR water.

Prepare one tube for each primer, and aliquot 5 μl of primer into each. The sequencing facility will use 3.5 μl for the reaction, but it’s good to give them a little extra.

1. ***Library submission***

Before submitting your sample at the University of Minnesota, you need to create an account with BMGC, and from there you can submit a sample submission form.

To create an account, go here:
<http://secure.ahc.umn.edu/ahc/bmgc_new/>

If you already have an account simply log in, download the appropriate submission form, in this case the “Libraries only” form under “Illumina Next Generation Sequencing Request”. Fill out the form according to the instructions and submit through email.

You will also need to submit the set of barcodes you’ve used in your library so that your samples can be demultiplexed. The barcodes must be submitted in reverse complement. This can be found in the primer spreadsheet.

Once you have done this, print out the first page of the order form and bring that along with your library and 3 sequencing primers to the BMCG St. Paul Campus facility in 28 Snyder Hall.

**III. Sequence library processing**

1. ***Demultiplexing samples***

Often samples will already be demultiplexed. You will receive two files for each of the samples in your library. One file is for the forward and the other is the reverse direction. You will need to provide the sequencing facility with the sample name and associated reverse complement of the barcodes for demultiplexing.

1. ***Quality filtering***

We have found that using the forward read (R1) gives much better results than the reverse read (R2) or paired sequences. Therefore, we suggest using only your forward reads, although you may want to try pairing the reads as well. There are many approaches to quality filter your data, we use the workflow below, which gives basically the equivalent results to the QIIME workflow “split\_libraries\_fastq.py”.

For our workflow, you will need three pieces of software:

cutadapt (<https://code.google.com/p/cutadapt>)

Trimmomatic (<http://www.usadellab.org/cms/?page=trimmomatic>)

mothur (<http://www.mothur.org/>) - included with QIIME

Download and install cutadapt and Trimmomatic. Put them into your bin (/usr/local/bin) and make sure that you give execute permission (chmod a+x) to the binaries.

**a. make working directories**

mkdir processing/

mkdir 2\_201/

**b. remove distal priming/adapter sites from ends of reads**

cutadapt -a GATCTCTTGGNTCTNGCATCGATGAAGAACG -q 20 -e 0.2 raw\_data/65\_S65\_L001\_R1\_001.fastq -o processing/65\_R1.fastq #Stops forward read at the ITS2 priming site

cutadapt -a GGAAACCTTGTTACGACTTTTACTTCCTCTAAATGACCAA -q 20 -e 0.3 raw\_data/65\_S65\_L001\_R2\_001.fastq -o processing/65\_R2.fastq #Stops reverse read at the ITS1f priming site

**c. trim remaining un-trimmed low-quality regions from ends of reads**

java -jar /usr/local/bin/trimmomatic-0.32.jar SE -phred33 processing/65\_R1.fastq processing/65\_R1\_trimmed.fastq LEADING:20 TRAILING:20 MINLEN:125

java -jar /usr/local/bin/trimmomatic-0.32.jar SE -phred33 processing/65\_R2.fastq processing/65\_R2\_trimmed.fastq LEADING:20 TRAILING:20 MINLEN:125

**d. convert fastq files to fasta**

mothur

fastq.info(fastq=/Users/nhu/Desktop/Research/IDENT/general\_processed/processing/65\_R1\_trimmed.fastq)

fastq.info(fastq=/Users/nhu/Desktop/Research/IDENT/general\_processed/processing/65\_R2\_trimmed.fastq)

**e. filter out short seqs and ones with ambiguous bases**

mothur

screen.seqs(fasta=/Users/nhu/Desktop/Research/IDENT/general\_processed/processing/65\_R1\_trimmed.fasta, maxambig=0, minlength=125, maxhomop=9, processors=2)

screen.seqs(fasta=/Users/nhu/Desktop/Research/IDENT/general\_processed/processing/65\_R2\_trimmed.fasta, maxambig=0, minlength=125, maxhomop=9, processors=2)

**f. Adding QIIME labels and combining all sequences into 1 file**

This code will process all files within a directory and concatenate them into a single file.

Put all of your filtered fasta files into 1 directory before running the script

-c is a column in the mapping file which tells QIIME to connect that file to the sample ID

add\_qiime\_labels.py -i processed/ -m alnus\_mapping.txt -c InputFileName -o processed/

1. ***Downstream processing and analyses***

We prefer to use QIIME for the rest of the steps such as clustering OTUs, identifying the sequences, creating alpha and beta diversity plots, statistical contrasts, etc…

Learn QIIME by starting with the standard tutorial, then applying it to fungal sequences.

<http://qiime.org/tutorials/>

We suggest that prior to the OTU picking, you perform a prefilter to get rid of non-target sequences. This is done through the script “parallel\_pick\_otus\_uclust\_ref.py”, clustering at 0.60, and using the latest UNITE database (with singletons) as a reference.

OTU picking is a crucial step to producing a good OTU table that is more representative of your community. There is much debate about the best method but there are pros and cons to each method. We have chosen to use a two-step clustering method first using USEARCH, then passing the results from the USEARCH clustering to UCLUST to cluster OTUs that did not cluster in the first USEARCH step (see [Nguyen et al. 2015](http://onlinelibrary.wiley.com/doi/10.1111/nph.12923/full)). The second step helps to collapse the OTUs that could not be collapsed in the first pass. UPARSE may also be use and VSEARCH (the open source version of USEARCH) are also other pieces of software to consider. We DO NOT recommend using just UCLUST alone because many papers now document OTU inflation

1. ***OTU identification for fungal studies***

We use local BLAST as the standard for sequence identification. The UNITE fungal ITS database is currently the best-curated database for high throughput studies of fungi. It is updated several times a year, so it is best to update your local BLAST database using the latest UNITE release.

Local BLAST allows enough flexibility to remove OTUs that do not match to anything in the database (probably a by-product of PCR) and allows cleaning of sequences that have only a tiny portion of actual fungal sequences. We typically remove any sequences that have the length/query length of < 0.85 (85% of the query sequence matches to a database sequence). Some studies consider as high of a ratio as 0.90.

1. ***Parsing out fungal guilds***

We created the FUNGuild database along with a python script that will parse processed high-throughput sequencing OTUs into ecological (trophic) guilds. The script and database, along with an instruction manual can be found on GitHub (https://github.com/UMNFuN/FUNGuild). Our manuscript is currently in review.

**IV. References (how to cite this protocol)**

Nguyen, N. H., Smith, D. P., Peay, K. G., & Kennedy, P. G. (2015). Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytologist* 205:1389-1393. DOI: 10.1111/nph.12923

**Other papers to look at:**

Branco, S., Bruns, T.D., Singleton, I., 2013. Fungi at a small scale: spatial zonation of fungal assemblages around single trees. PLoS One 8, e78295. doi:10.1371/journal.pone.0078295

Smith, D. P., & Peay, K. G. (2014). Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PloS One*, *9*(2), e90234. doi:10.1371/journal.pone.0090234

Nguyen, N. H., Song Z., Bates S. T., Branco S., Tedersoo L., Menke J, Schilling J., Kennedy P. FUNGuild: an open annotation tool for parsing high-throughput fungal molecular data by ecological guild. Submitted to *Fungal Ecology*.

 Song Z, Schlatter D, Kennedy P, Kinkel LL, Kistler HC, et al. (2015) Effort versus Reward: Preparing Samples for Fungal Community Characterization in High-Throughput Sequencing Surveys of Soils. PLoS ONE 10(5): e0127234. doi: 10.1371/journal.pone.0127234