## 1. SCOPE

This protocol is designed for the extraction of gDNA from unidentified bacterial isolates, amplification of the whole 16S rRNA gene, sequencing and identification by comparison to 16S databases. This protocol replaces my old "HDA_and_extraction.pdf" protocol from 2012 which had some minor errors in it. Additionally it is advised to use the pA and pH primers as they will amplify and sequence the whole 16S gene as opposed to a limited variable region. Recently updated for Turnbaugh Lab

## 2. PRINCIPLE

Genomic DNA is extracted using Instagene for a quick and dirty DNA preparation. You could use other kits such as the Purelink Genomic DNA kit but for the purposes of this protocol, it would be overkill. A simple PCR is performed and checked via gel electrophoresis. Then the PCR products are cleaned up and prepared for sequencing. Cleaning and trimming the sequencing results and interpreting the database comparisons is also included.

## 3. MATERIAL REQUIREMENTS

Instagene Matrix (Biorad #732-6030)
Wide bore 200 $\mu$L pipette tips
1.5 mL Sterile Microcentrifuge Tubes
Sterile $H_2O$ or PBS
200 $\mu$l PCR Tubes
Water bath or heat block set to 56˚C
Water bath or heat block set to 100˚C
Vortex

2x Mastermix (Ex. Amplitaq Gold 360 Life Tech/Thermofisher 4398881)

2 µM PA primer (8F) 5-AGAGTTTGATCCTGGCTCAG-3
2 µM PH primer (1543R) 5- AAGGAGGTGATCCAGCCGCA-3

Purelink PCR purification kit (Life Technologies #K3100-01; or Qiagen)
High-sensitivity dsDNA kit for the Qubit (Life Technologies #Q32854)

## 4. EXTRACTION OF gDNA FROM ISOLATE

4.1 Pick a single colony from a freshly prepared culture plate and resuspend in 1mL of PBS or sterile H2O in a microcentrifuge tube
4.2 Centrifuge at ~10,000 xg for 1 minute.
4.3 Remove the supernatant and resuspend in 200 µL of Instagene matrix*
** Note: matrix should be well mixed and use wide bore tips*
4.4 Mix thoroughly by vortexing
4.5 Incubate at 56˚C for 30 minutes
4.6 Mix thoroughly by vortexing
4.7 Incubate at 100˚C for 8 minutes
4.8 Mix thoroughly by vortexing
4.9 Centrifuge for 3 minutes at max (~16,000 xg)
4.10 Transfer 100 µL of the supernatant containing gDNA to a new microcentrifuge tube. Discard the pellet*.
**Note: If you are lazy you do not have to separate the supernatant you can just centrifuge before every time you use it.*
4.11 Store gDNA at -20˚C for long term storage or 4˚C for short term.
4.12 If you wish you may quantify the gDNA using the high-sensitivity dsDNA kit for the Qubit (Life Technologies #Q32854).

## 5. PCR

5.1 Prepare aa master mix for PCR as outlined in Table 5.1.
5.2 Add 5 µL of gDNA for a total reaction volume of 50 µL. Be sure to include a reaction with 5 µL of H2O as no template control (NTC).
5.2 Run the PCR using the PCR cycle parameters indicated in Table 5.2

| Component | Stock Concentration | Final concentration | Volume per 50 µL Rxn |
|---|---|---|---|
| pA | 2 µM | 400 nM | 10 µL |
| pH | 2 µM | 400 nM | 10 µL |
| 2X Amplitaq | 2x | 1x | 25 µL |
| | | Total Volume: | 45 µL |

Table 5.1. Master mix for pApH amplifications. Add components in the order indicated.

| Step | Time | Temperature |
|---|---|---|
| Initial Denaturation | 10 minutes | 95.0 ˚C |
| 30 Cycles of: | | |
| Denaturation | 30 seconds | 95.0 ˚C |
| Annealing | 30 seconds | 55.0 ˚C |
| Extension | 1 minute* | 72.0 ˚C |
| Final Extension | 10 minutes | 72.0 ˚C |
| Storage | Hold | 4.0 ˚C |

Table 5.2. PCR cycles for amplification with pA and pH. **Note: this could probably be increased to 1.5 minutes as the general rule of thumb is 1min/1kb*

**6. PCR CLEANUP**

Using the Purelink PCR purification kit:

6.1 Add 200 µL of buffer B2 to 50 µL of the reaction (4 volumes).

6.2 Transfer 250 µL of the mixture to a spin column and centrifuge at maximum speed for 1 minute.

6.3 Discard flow through and add 650 µL of wash buffer W1. Centrifuge at maximum speed for 1 minute.

6.4 Discard flow through and recentrifuge for 3 minutes.

6.5 Transfer spin column to fresh nuclease free microcentrifuge tube and add 50 µL of Elution buffer (Tris-HCl, TE, $H_2O$ will all work as well).

6.6 Incubate at room temperature for 1 minute and then centrifuge for 2 minutes at maximum speed.

6.7 Dispose of spin column and keep eluted DNA.

**7. CONFIRMING PRODUCT BY GEL ELECTROPHORESIS**

Prepare and run a 1% agarose gel using TAE or TBE buffer using 10 µL of PCR reaction. Confirm presence of ~1,500 bp band and the absence of product in the NTC.

**8. CAPILLARY SEQUENCING**

**For genewiz sequencing*

8.1 Quantify PCR product using the Qubit HS dsDNA

8.2 Combined 10µL of pA with 5 µL PCR product (diluted to 10ng/µL)
**Note: the primer will be used to sequence from. You can expect up to 900 bp of usable sequence from either pA or pH. If you wish for the full sequence you must prepare two tubes for sequencing, one with pA and one with pH. DO NOT ADD BOTH PRIMERS INTO THE SAME SEQUENCING REACTION!!!*

8.3 Using genewiz account create a sequencing request for and leave reaction in the dropoff outside the labs on the 10$^{th}$ floor.

## 9. CLEANING AND INTERPRETTING SEQUENCE

Returned sequence will need to be trimmed and corrected before comparisons can be made to a database. This will require a program capable of viewing the electropherograms (.ab1 files). I suggest the following programs:

4 Peaks (http://nucleobytes.com/index.php/4peaks) for Mac
FinchTV (http://www.geospiza.com/Products/finchtv.shtml) for Windows

The very beginning and end of every sequencing run is poor and yields many Ns (Figure 9.1). This has to be removed (highlight and hit delete). Also, you want to ensure that you have no unknown bases in your sequence (Ns). You can either crop out these areas, or in the case of base 24 in the example in figure 9.1, study the trace to see if you can manually change the base. In the example we can see the correct base is truly an A but the signal didn't rise above the background from the Cs. When you are content, copy and paste the corrected new sequence to a text file or use File>Export to create a new file.

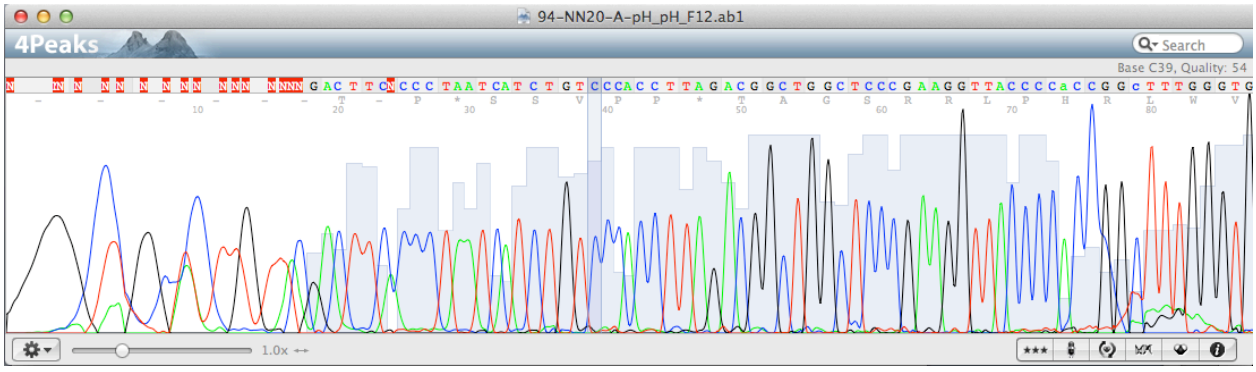

Figure 9.1. Example of beginning of trace shows poor quality.

If you have sequenced from both ends of the gene trim and correct both files independently and overlap the sequences. I have used Cap3 for this with good results (http://doua.prabi.fr/software/cap3). Use the resulting overlapped contig for futher analysis.

Using NCBI BLAST (other good databases include NCBI 16S database, Green Genes, RDP, HOMD) compare your sequence to the non redundant (nr/nt) database but be sure to exclude uncultured/environmental sample sequences (otherwise you will see the OTUs from Metagenomic surveys). See Figure 9.2.

To evaluate the hits you want to find a high % ID over 100% of the gene. Ideally to call a species you want >97%. In Figure 9.3 you can see that NN20 is identical to *W. cibaria* and *W. confusa.* This means you can not distinguish which species it is. The correct identification would be *Weissella* sp. Review of the literature suggests these are two highly similar species so it is not surprising we cannot distinguish them by 16S.

To be confident in our identification we must also look at the other hits. **IF YOU ARE LOOKING AT LACTOBACILLUS CASEI GROUP OR ENTEROBACTERACEAE BE SURE TO DO THIS!!!!!!!** To do this, go back to the BLAST interface (Figure 9.1) and enter the best hit into the Organism box and hit exclude. In this case I have excluded *Weissella*. In this case we can see the nearest non-*Weissella* hit is L. curvatus with 92% ID over 95% coverage indicating that we can be confident in our identification. In the case of *E. coli* many other organisms would show up with high % IDs. To definitively make the *E. coli* identification biochemical tests would be recommended.

If no hits are found try other databases:

http://greengenes.lbl.gov/cgi-bin/nph-index.cgi
http://rdp.cme.msu.edu/seqmatch/seqmatch_intro.jsp
http://www.homd.org/index.php?name=RNAblast&link=upload

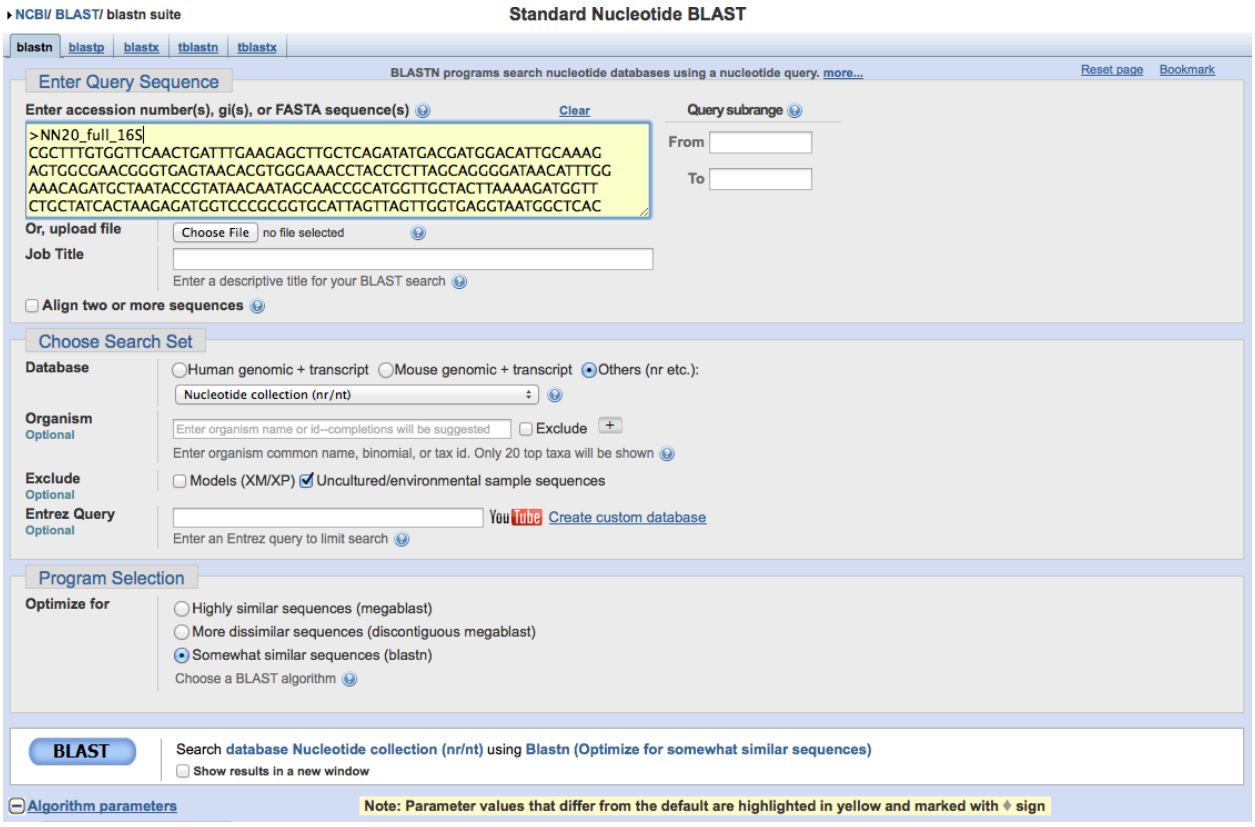


Figure 9.2. Blast parameters for comparison to NCBI NR database.

| Description | Max score | Total score | Query cover | E value | Ident | Accession |
|---|---|---|---|---|---|---|
| Weissella cibaria gene for 16S rRNA, partial sequence, strain: JCM 7777 | 2904 | 2904 | 100% | 0.0 | 100% | AB911503.1 |
| Weissella cibaria gene for 16S rRNA, partial sequence, strain: NRIC 0136 | 2904 | 2904 | 100% | 0.0 | 100% | AB362617.1 |
| Weissella cibaria gene for 16S rRNA, partial sequence, strain: NRIC 0133 | 2904 | 2904 | 100% | 0.0 | 100% | AB362614.1 |
| Weissella cibaria 16S rRNA gene, strain ACA-DC 3411t2 | 2904 | 2904 | 100% | 0.0 | 100% | AJ422031.1 |
| Weissella cibaria strain Uga49-1 16S rRNA gene, complete sequence | 2904 | 2904 | 100% | 0.0 | 100% | DQ294961.1 |
| Weissella cibaria gene for 16S rRNA, partial sequence, strain: JCM 7755 | 2900 | 2900 | 100% | 0.0 | 99% | AB911490.1 |

Figure 9.3. Top BLAST results for strain NN20.

## 10. INFORMATIVE REFERENCES

Falsen et al. 1999. International Journal of Systematic Bacteriology. 49, 217-221.