National Center for Biotechnology Information (NCBI) contains multiple publicly available databases, including genomic sequences.

NCBI also has the BLAST tool. Basic Local Alignment Search Tool (BLAST) is a sequence similarity search program. BLAST can take sequences of nucleotides or proteins and search them agains the entire publicly available database. Try taking some of your nucleotide sequences (from your genomes) to BLAST them.

Each of your genomes comes in a ".fasta" file. This is a type of FASTA file. FASTA format is a text-based file for nucleotide (genomic) or animo acid (protein) sequences. You have nucleotide sequences. A FASTA file can have multiple sequences in it. Your sequences are in the ZIPed folder.

A sequence in a FASTA file begins with a greater-than character (">") followed by a description of the sequence (all in a single line). This is called the FASTA header. The lines immediately following the description line is the nucleotide (or amino acid) sequence.

Each of your barcodeX.fasta files contains multiple organisms 16S RNA that we were able to recover with the sequencing. To figure out YOUR barcode please see the methodology document "matériel et méthode 16S nanopore".

Each isolate will look something like this:

> cluster_0_consensus
CCGCTGACCCATCTGGTGAGTTCCTGGAGCCTGGAGCAGGCCGTCGAGCGCCTTCAGGTTCTCGGCGACGGCGC
CGGGCCGTTGGGTGGGCGAGGAAGTCGTCGTACGCGTAGGGGAAGTCC

In your cases the FASTA header is not very informative, but this is normal for output from sequencing.

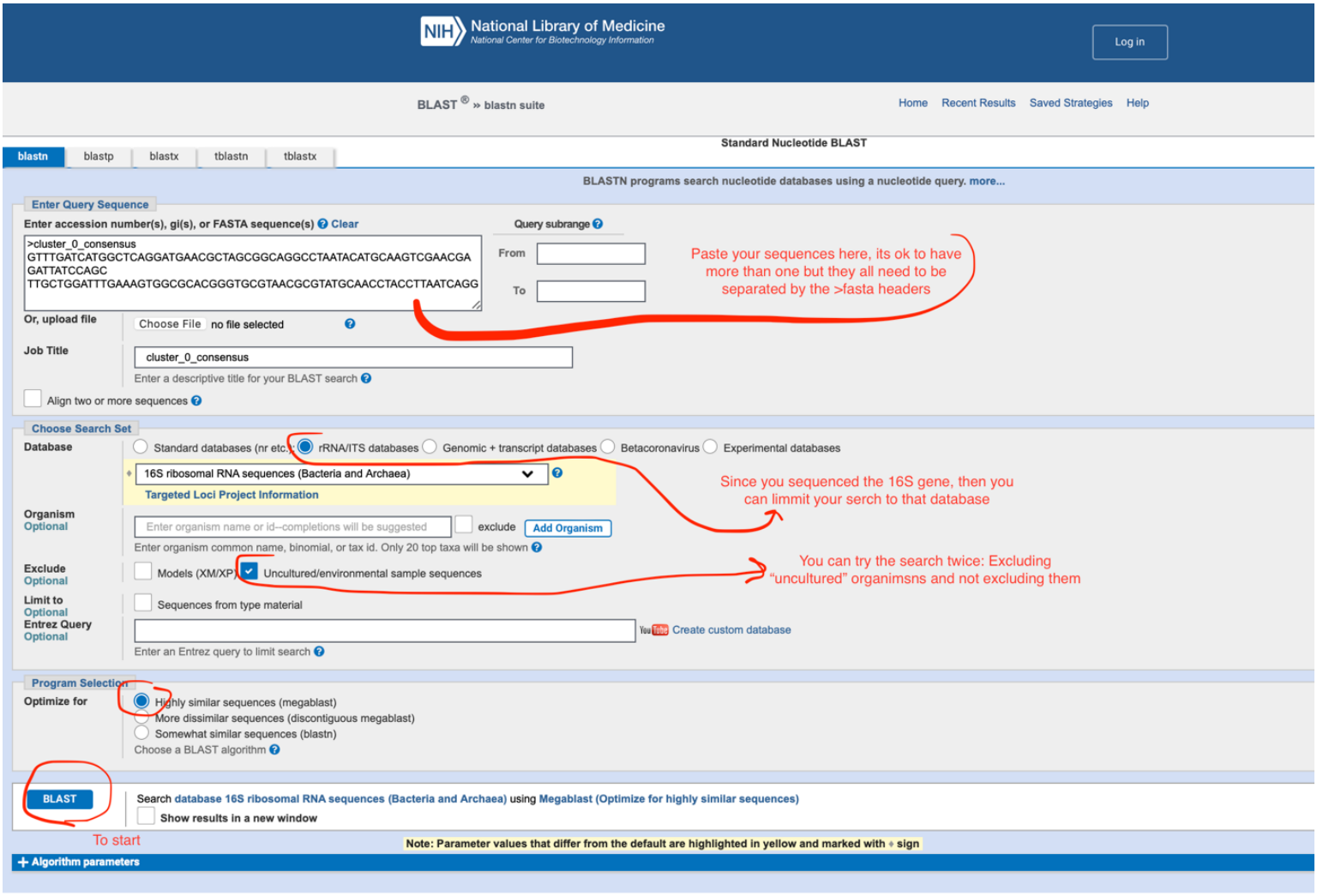
Try the BLAST tool on NCBI with your sequences. Make sure to use "nucleotide > nucleotide" (also known as blastn) as you are searching your nucleotide sequences agains the entire nucleotide sequence database. The tool starts off with default values, you can keep these (you have to hit the BLAST button at the bottom, after copy/pasting your sequence. Do only one sequence at a time (remember the sequences in the FASTA file are separated by a ">" sign. You don't need to BLAST all of your sequences (there are too many) but check a few to give yourself an idea what organism you have. Explore the outputs and try to understand where else your isolates have been found.

You may not have pure cultures. Knowing the relative amount of each organism you extracted the DNA from, may be useful. To figure this out, you can look at the relative abundance of # of sequences recovered for each organism in a barcode using this file "read_n.xlsx".

https://blast.ncbi.nlm.nih.gov/Blast.cgi

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

Example of the BLAST interface:


National Library of Medicine
National Center for Biotechnology Information
Log in
BLAST ® » blastn suite
Home Recent Results Saved Strategies Help
blastn blastp blastx tblastn tblastx
Standard Nucleotide BLAST
BLASTN programs search nucleotide databases using a nucleotide query. more...
Enter Query Sequence
Enter accession number(s), gi(s), or FASTA sequence(s) Clear
>cluster_0_consensus
GTTTGATCATGGCTCAGGATGAACGCTAGCGGCAGGCCTAATACATGCAAGTCGAACGA
GATTATCCAGC
TTGCTGGATTTGAAAGTGGCGCACGGGTGCGTAACGCGTATGCAACCTACCTTAATCAGG
Query subrange
From
To
Paste your sequences here, its ok to have more than one but they all need to be separated by the >fasta headers
Or, upload file
Choose File no file selected
Job Title
cluster_0_consensus
Enter a descriptive title for your BLAST search
Align two or more sequences
Choose Search Set
Database
Standard databases (nr etc.) rRNA/ITS databases Genomic + transcript databases Betacoronavirus Experimental databases
16S ribosomal RNA sequences (Bacteria and Archaea)
Targeted Loci Project Information
Since you sequenced the 16S gene, then you can limmit your serch to that database
Organism
Optional
Enter organism name or id--completions will be suggested
exclude
Add Organism
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown
Exclude
Optional
Models (XM/XP)
Uncultured/environmental sample sequences
You can try the search twice: Excluding "uncultured" organimsns and not excluding them
Limit to
Optional
Sequences from type material
Entrez Query
Optional
Create custom database
Enter an Entrez query to limit search
Program Selection
Optimize for
Highly similar sequences (megablast)
More dissimilar sequences (discontiguous megablast)
Somewhat similar sequences (blastn)
Choose a BLAST algorithm
BLAST
Search database 16S ribosomal RNA sequences (Bacteria and Archaea) using Megablast (Optimize for highly similar sequences)
Show results in a new window
To start
Note: Parameter values that differ from the default are highlighted in yellow and marked with + sign
+ Algorithm parameters