**GenBank genome submission**

1. Now go to new genome submission:  
   <https://submit.ncbi.nlm.nih.gov/subs/genome/SUB5261114/submission_type>
2. Select ‘single genome’ (unless you have lots, I think this is easier)
3. Fill in your details (it doesn’t matter if you put in a group or not)
4. Click ‘No’ to whether you already registered a BioProject or BioSample for this project
5. You can then either click on ‘release on a specified date’, or release immediately (I think release immediately is easier, as it ensures that it will be available whenever you publish.
6. For the assembly data: if you got the genome from MicrobesNG, then this was assembled using SPAdes, and I have put both dates as the month that you received the data from MicrobesNG. The genome coverage is listed on the MicrobesNG project page (Mean Coverage), and MicrobesNG use either MiSeq of HiSeq, so I have put MiSeq for the ‘Sequencing Technology’ (I think the .sqn file may actually have the assembly data, and it also gives the date that this was done in the automatic file name)
7. ‘Yes’ your sample includes the full genome, and ‘yes’ this is the final version. MicrobesNG do not use *de novo* assembly (they map the reads to a similar existing genome based on the taxonomy you give – so I just gave the genus name as the reference assembly), so ‘No’ to this and ‘no’ to being an update of an existing genome
8. I gave the submission title as the name that I gave my bacterium, in this case *Mycobacterium* sp. DBP42
9. For the public description, I gave a brief overview of what I aimed to do when I obtained this isolate, *i.e.* This study aimed to isolate bacteria that could grow using six common plasticizers (dibutyl phthalate, bis(2-ethyl hexyl phthalate), diisononyl phthalate, diisodecyl phthalate, acetyl tributyl citrate and trioctyl trimellitate) as a sole carbon source. This was narrowed down to two bacteria that appeared to be able to grow on all six plasticizers and their growth was characterised using proteomics.
10. I put that this is of Environmental relevance
11. I didn’t add a link or consortium name, but did put in my MIBTP grant details
12. For ‘Sample Type’, I selected: Genome, metagenome or marker sequences (MIxS compliant) > Cultured Bacterial/Archaeal Genomic Sequences MIGS > Water
13. For sample name, organism and strain, I put: DBP42; *Mycobacterium* sp.; *Mycobacterium* sp. DBP42  
    Sometimes it doesn’t seem to be happy with the sample name, even though it tells you that you can put whatever you like, but it lets you go forward if you just click submit again, and so far I haven’t had problems further down the line from doing this!
14. Now input the data about where you collected the original samples that this was isolated from (hopefully you have this in your lab book! – Plymouth ones are July 2016) – if you hover over the question marks then they will give you examples, so I put for mine: Seawater biome; Nearshore; Seawater (these samples were from Plymouth Sound). For the location, you need the full country name, *i.e.* United Kingdom, not UK. Plymouth latitude and longitude: 50.3355 N 4.1527 W
15. For nucleic acid sequence source, I put ‘Enrichment’ for isolation and growth conditions, and ‘NA’ for number of replicons and reference for biomaterial (unless this is already published for you!)
16. Depth: 0m
17. I have put that the Bacteria and/or source DNA is available from Joseph, but I think it may also be available from MicrobesNG, if you have used them for the sequencing
18. I have then said ‘yes’ to annotating the file (as trying to add a file that you have annotated is ridiculously painful, and might even be impossible, but there are some instructions on that below, that may or may not work…)
19. On Files, click option 2 – chromosome is in one or more pieces (unless you are totally sure that you have a complete genome), and then choose Fasta
20. This should be a small file, so is fine to ‘upload now’. If you used Prokka, then this should be the .fsa file.
21. No, you don’t have the AGP files. When you click continue after this, it should check the file for you.
22. If you now get an error message because there are contigs of below 200 nucleotides, you can just delete these from the end of the file. The contigs in the .fsa file are sorted by length anyway, and they say how long they are in the >contig name part. So then just repeat with your new file. (I haven’t had any other error messages).
23. Assignment: ‘No’ it’s not a complete chromosome, and ‘No’ no sequences belong to plasmid (unless you know they do, and can provide information on where the plasmids start and stop)
24. I then put in my name as the sequence author, gave the bacteria species name as the reference title, and said it was unpublished, and said reference authors were the same as sequence authors.
25. Now, submit! It will give you the project number fairly quickly (10 minutes or so), but will take longer for it to be finalized (especially if it needs curation). Now you can add something like: All sequences have been deposited in the NCBI Short Read Archive (SRA) database under Bioproject PRJNA499076.
26. Mine also came back with errors after submission, and I had to remove some of the shorter contigs, and a small sequence in one of the contigs that it said it thought were contamination.
27. It will give you the project number fairly quickly (10 minutes or so), but will take longer for it to be finalized (especially if it needs curation). Now you can add something like: All sequences have been deposited in the NCBI Short Read Archive (SRA) database under Bioproject PRJNA499076.

**Genbank amplicon sequencing submission**

1. Go to the sequence read archive submission portal: <https://submit.ncbi.nlm.nih.gov/subs/sra/>
2. Click ‘new submission’ (you can also leave the submissions half completed and come back to them here)
3. Fill in your details (it doesn’t matter if you put in a group or not)
4. Click ‘No’ to whether you already registered a BioProject or BioSample for this project
5. You can then either click on ‘release on a specified date’, or release immediately (I think release immediately is easier, as it ensures that it will be available whenever you publish.
6. Enter a title, e.g. “Microbial community succession on PET”
7. Provide a short description, e.g. “This project aimed to characterise microbial community succession on different types of poly(ethylene terephthalate) (PET), and the PET oligomer BHET, in laboratory incubation experiments.”
8. I have then clicked that it is of environmental relevance
9. I then put in my grant details, but leave the rest of this page blank.
10. I then clicked “Metagenome or environmental sample”
11. It is then easier to upload a file using excel, as I think it would take FOREVER to fill in online, but you do then need to be careful about how you fill in this form as it is very picky. I can provide an example of how this is filled in if necessary.  
    Your sample name can be anything at this point – I stuck with the file name that the reads were from, so they are S1-S140. You can then give any descriptor (mine are e.g. RD01W1, which means it is day 1 of the ‘water’ or no carbon control). You don’t have a bioproject accession until you’ve completed this process, so leave that blank, the organism is ‘Community’ or ‘DNA extraction and sequencing control’ or something like this, N/A for host, isolation source seawater, collection date (should be in the e.g. Jul-2018 format), geographic location needs to be in the format: ‘United Kingdom: Porthcawl’, and then get the coordinates from google. Aerobe for relation to oxygen, DNA extraction method used in samp\_mat\_proc,, and then 1.5 ml for sample size. I put some more sample details in description, e.g. ‘PET succession - day 1 - no carbon control - replicate 1’. I also made a note here if this sample contains reads from another project, i.e. Mira’s and Craig’s 18S samples. Ref biomaterial won’t allow you to have the same for every sample, so I just put the same as is in description.
12. Now you can upload this file, and hopefully it will let you! ☺ It may give you a warning that it will need to be curated, but this shouldn’t be a problem.
13. Next you have to do the metadata. This will be similar to the previous file, and will need to have the same names, as well as the additional info on sequencing instruments etc. Again, I have example files that I can send.  
    The sample names should be the same as in the previous file, and the library ID can be the sample title. The title can then be the same as the description from the previous file.  
    Library strategy – amplicon  
    Library source – metagenomic  
    Library selection – PCR  
    Library layout – paired  
    Platform – Illumina  
    Instrument model – Illumina MiSeq  
    Design description - Taxa PCR, magnetic bead cleanup, index PCR, normalisation, pooling   
    Filetype – fastq  
    Filename and filename2 should then be the forward (R1) and reverse (R2) reads (I have a script that will give these in a new excel document if you want to use it – let me know)
14. Now upload this!
15. Now you will need to upload the files. They will probably be too big to just upload, so you’ll need to use the ‘FTP or Aspera Command Line file preload’. You’ll need to install Aspera command line for this (https://downloads.asperasoft.com/en/downloads/62).
16. First, you need to make this script executable, so navigate to the folder where it is (in terminal), and type chmod +x file\_name
17. Now run it using ./file\_name
18. This gives you the line to run to add this to your path, for me it was: export PATH=/Users/u1560915/Applications/Aspera\ CLI/bin:$PATH
19. Click on the Aspera command line upload instructions. Download the key file, put this in a sensible folder, and remove the .txt from the file name
20. Now use the link it gives you, e.g. ascp -i <path/to/key\_file> -QT -l100m -k1 -d <path/to/folder/containing files>[subasp@upload.ncbi.nlm.nih.gov:uploads/robynw371@gmail.com\_qRm5Cf4h](mailto:subasp@upload.ncbi.nlm.nih.gov:uploads/robynw371@gmail.com_qRm5Cf4h)  
    For me, this is (after navigating to the right folder): ascp -i SRA\ submission/aspera.openssh -QT -l100m -k1 -d raw\_files/ [subasp@upload.ncbi.nlm.nih.gov:uploads/robynw371@gmail.com\_qRm5Cf4h](mailto:subasp@upload.ncbi.nlm.nih.gov:uploads/robynw371@gmail.com_qRm5Cf4h)
21. Now you should see the files starting to upload, obviously this will probably take a while (especially if you use uncompressed files, like me)… and you will probably need to wait a little while again before they are visible online.