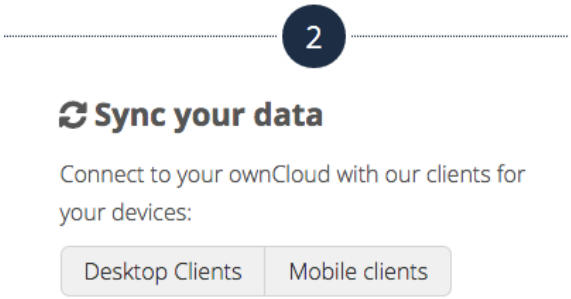
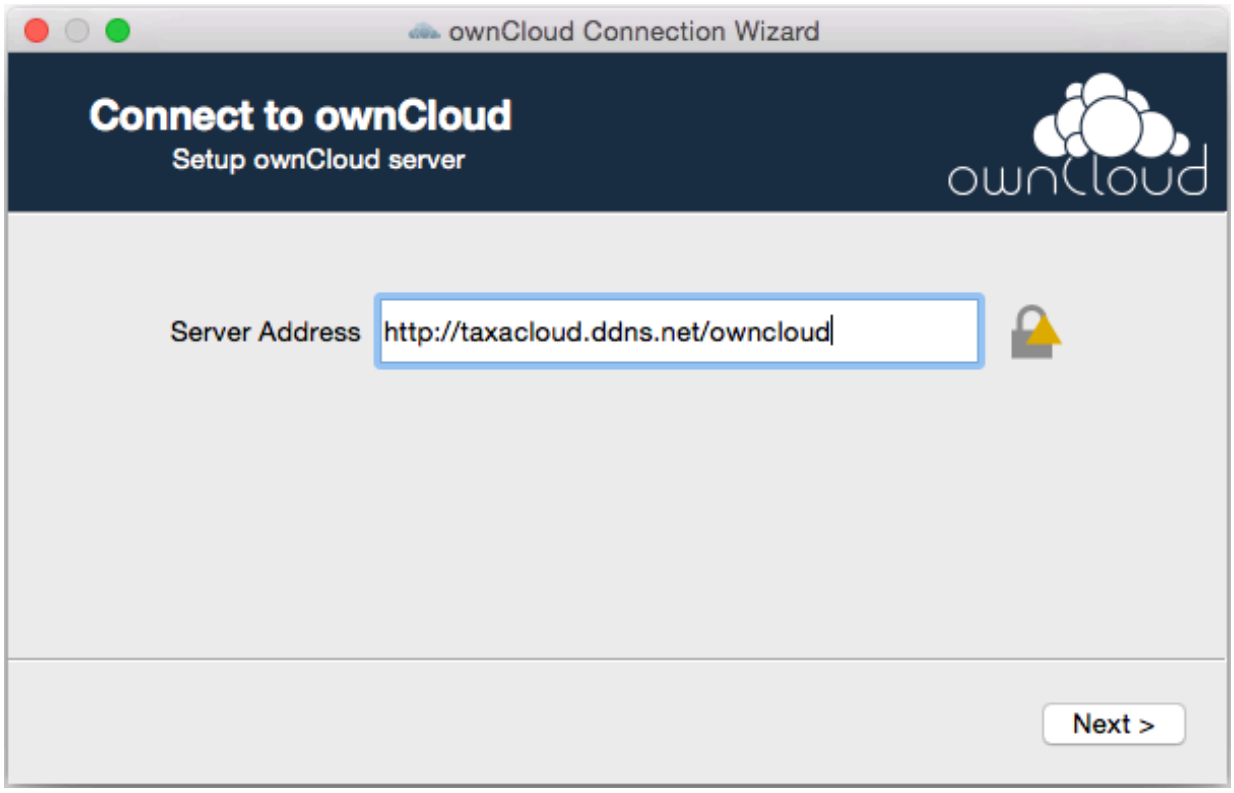
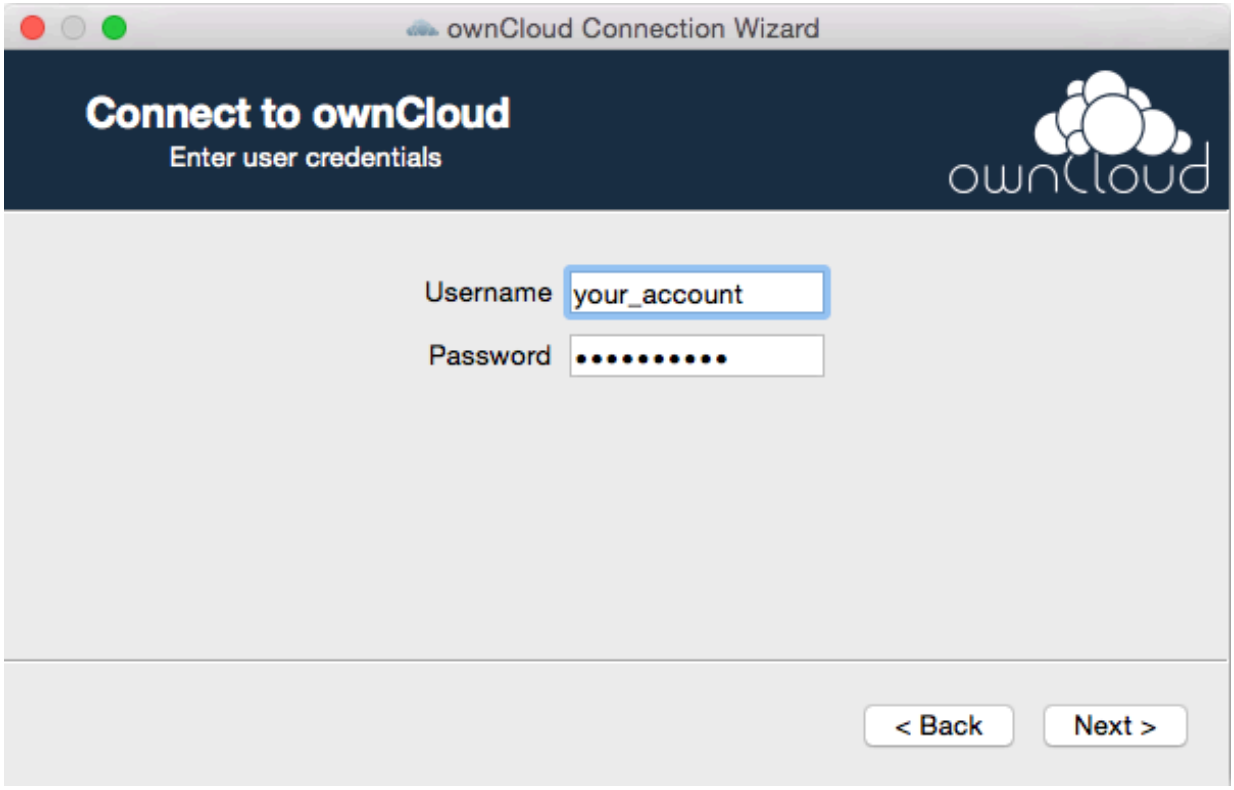
TaxaCloud User Guide

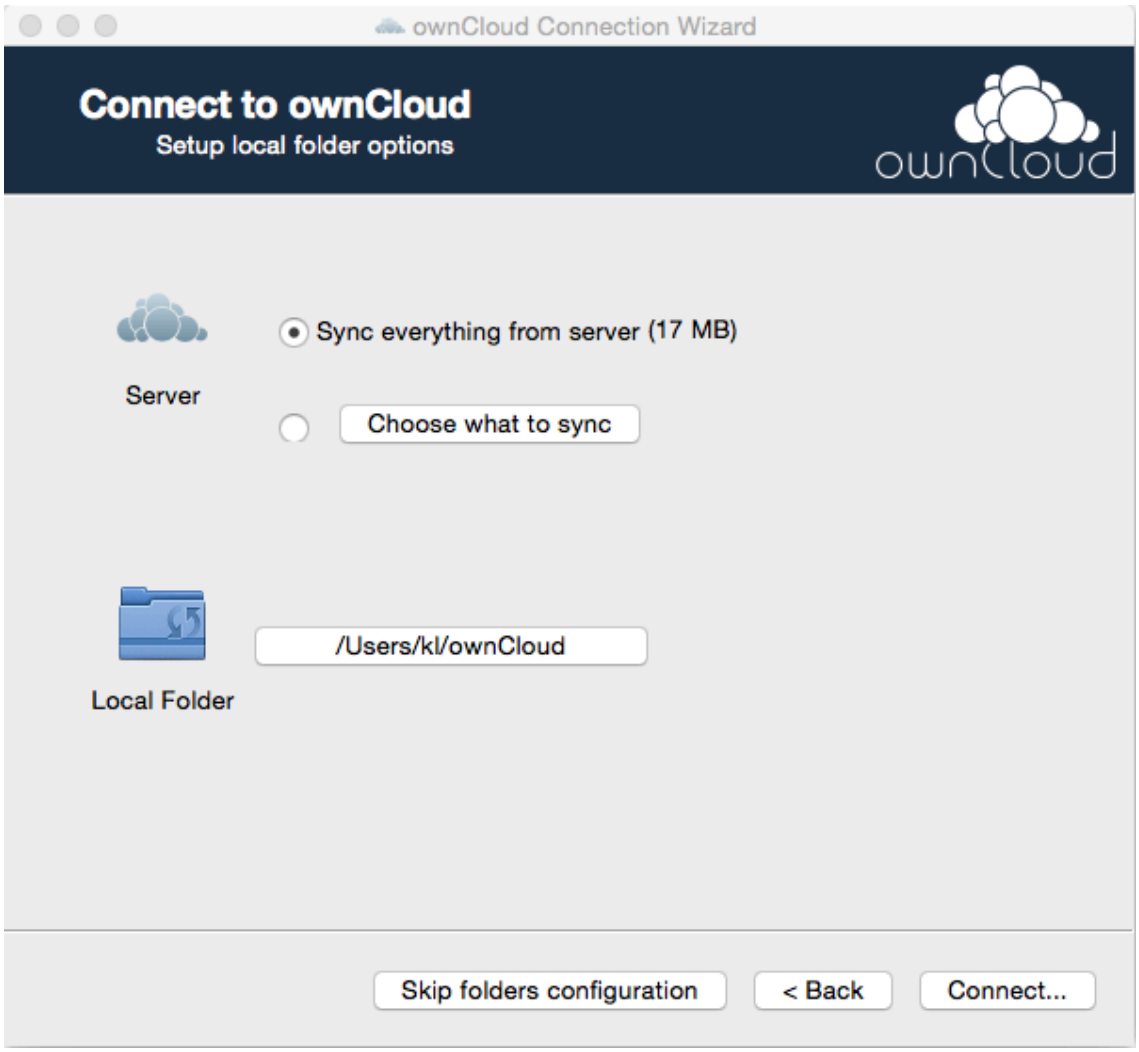
1. To get your personal service access account, please contact the main author Kaitao Lai (Kaitao.Lai@westernsydney.edu.au) with your contact details and your project description for requesting this service.
2. To upload your sequence data, download the owncloud desktop client on your desktop by browsing https://owncloud.org/install/, then click 'Desktop Clients' in Step 2: Sync your data. You must click the suitable version for your system (Windows 7, 8.x and 10, Mac OSX 10.7+, 64 bit, or Linux Multiple distributions).

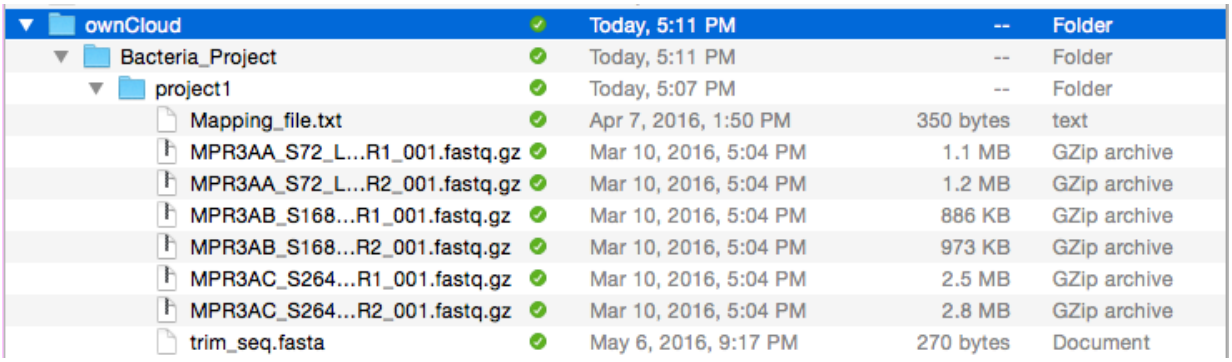


3. After installation of Owncloud desktop client, setup an account on Owncloud desktop client, input 'http://taxacloud.ddns.net/owncloud' as Server Address, then input your account name and password, and specify your local storage folder. After done the setup, copy your sequence data to local storage folder for synchronising your sequence data to ownCloud server.

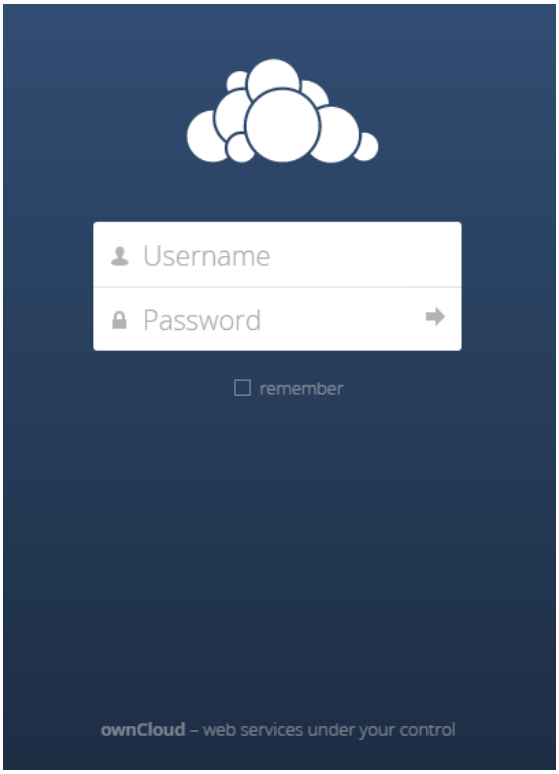



ownCloud Connection Wizard
Connect to ownCloud
Enter user credentials
ownCloud
Username
your_account
Password
< Back
Next >


ownCloud Connection Wizard
Connect to ownCloud
Setup local folder options
ownCloud
Server
Sync everything from server (17 MB)
Choose what to sync
/Users/kl/ownCloud
Local Folder
Skip folders configuration
< Back
Connect...

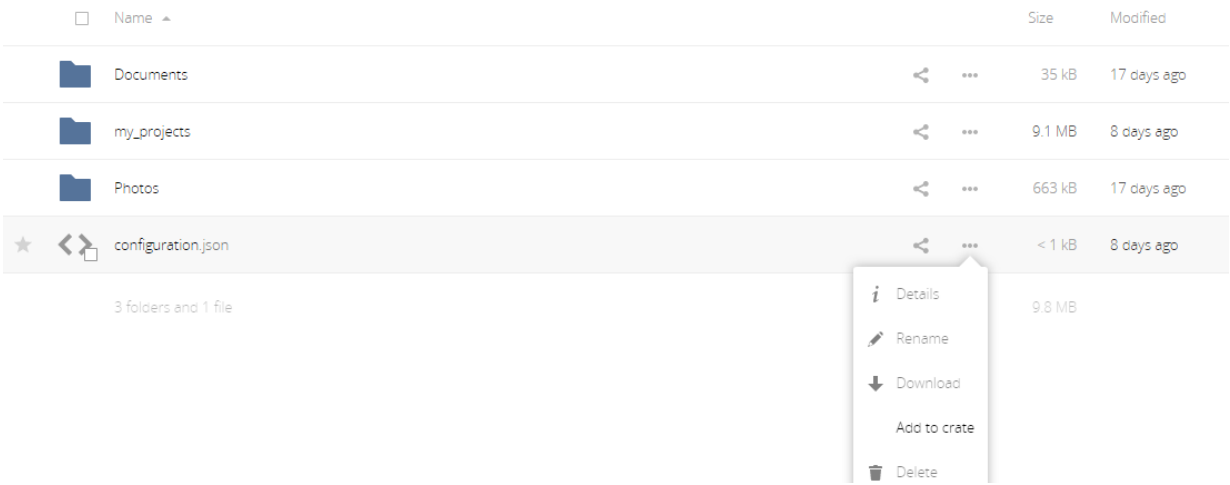


4. To access the TaxaCloud system, click the link below the first subheading on the TaxaCloud webpage. You will then be asked to input a username and password



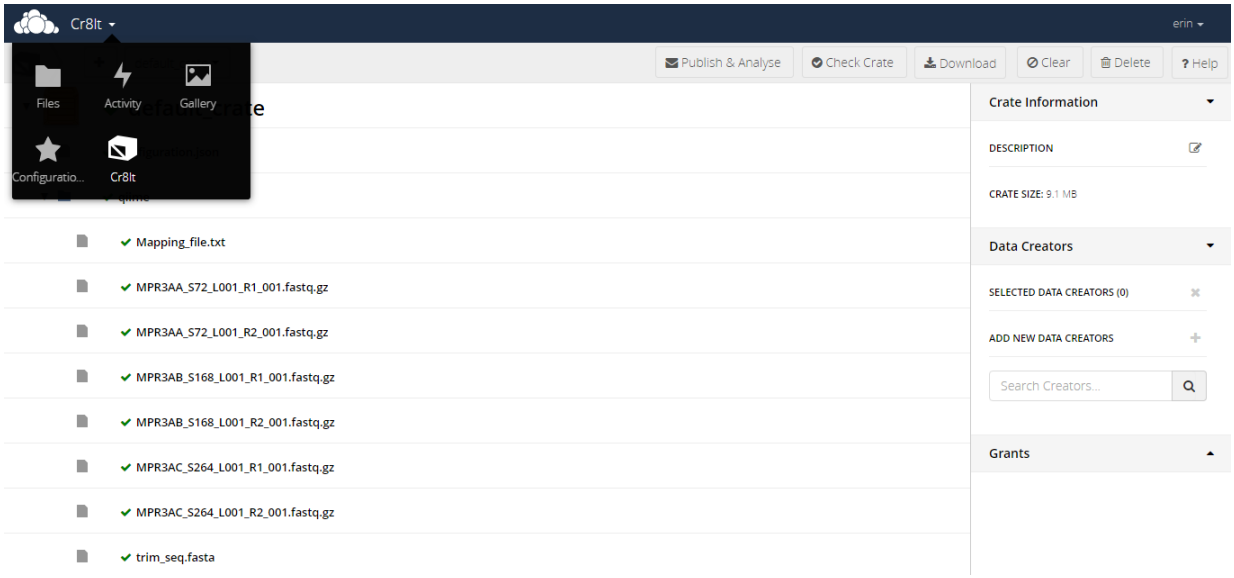
provided to you by the Author.

5. This will take you to the main page, where you can upload your projects and

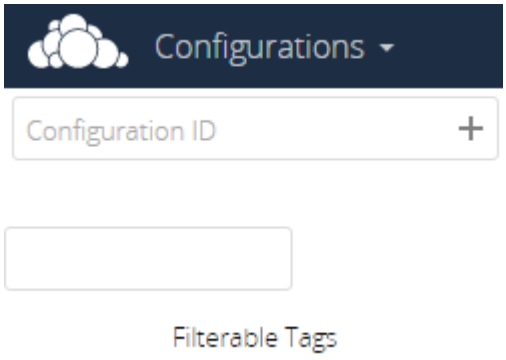


sequences for analysing. To start the configuration process, you must first add the configuration file ('configuration.json') by clicking on the three dots to the left of the size column, then clicking on 'Add to crate'. If you haven't do the configuration (creating configuration.json file) for this analysis, please go to step 7 – step 32 for doing configuration before this step.

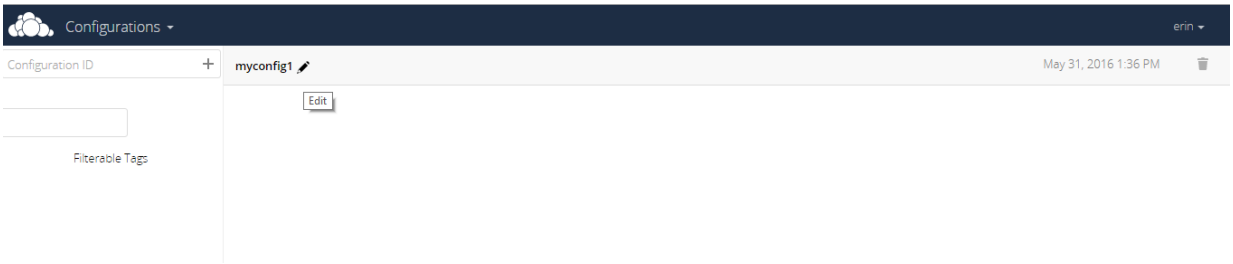
6. Do the same for your QIIME sequence in the 'my_projects' folder. Now your files have been added for configuration. You can view your added files by clicking the arrow next to 'Files' in the top left-hand corner and clicking 'Cr8lt'. This will take you to a page that shows you your added files in a 'default_crate'.



7. To begin configuration, click the arrow next to 'Files' in the top left-hand corner and click 'Configuration'.
8. You must first add a configuration ID by typing a name into the box in the top left-hand corner under 'Files' and clicking the '+' to add the configuration ID. The ID will then appear on the main page.



9. Hover your mouse over your newly created ID and an image of a pencil will appear next to the ID name. Click this pencil image to edit the configuration.



10. The first step is to select Data Type. Click the grey box next to 'Data Type' and a drop-down box will appear with several options. Click '16s' for bacteria, or 'its2' for fungus.
11. The next step is the FastQC Process. FastQC is a quality control tool for high throughput sequence data that generates a map to gauge the quality of the sequence. Click the grey box true and select 'true' if you wish to run the process, or 'false' if you do not wish to run the process.
12. Next, add a file name for contaminant oligonucleotide sequences such as primers and adaptors. If no file name has been specified, a default file from the system for the most common primers and adaptors will be used.
13. Next, is the trimming of primers and adaptors. If you wish to run this process, click the grey box true and select 'true' if you wish to run the process, or 'false' if you do not wish to run the process.
14. The next step is to determine which software for trimming sequences and quality control you wish to use. Next to 'Trimming Method', click the grey box and choose 'seqtk' or 'trimmomatic'. 'Seqtk' is a fast and lightweight tool for processing sequences in the FASTA or FASTQ format which seamlessly parses both FASTA and FASTQ files. The TaxaCloud system automatically uses the 'seqtk' formula to process the sequence, and is most commonly preferred by users. 'Trimmomatic' is a fast, multithreaded command line tool that can be used to trim and crop Illumina (FASTQ) data as well as to remove adapters however is slightly more difficult to use.
15. The SeqTK Tool Parameters are set at 'auto', where the system automatically detects the suitable bp to trim from but can also be calculated manually to double check.
16. The Trimmomatic Tool Parameters can also be set to 'auto' for the Phred Score Scale where the system will automatically make a judgement between 'phred33' and 'phred64', however these options can also be chosen. 'Phred33' and 'phred64' are different encoding formats with different standards.
17. Under 'The Trimmomatic Tool Parameters', 'Trimmomatic Trailing', 'Trimmomatic SlidingWindow', and 'Trimmomatic MinLen' will be automatically generated based on

which option was chosen for ‘Phred Score Scale’ but can also be manually calculated by inputting your chosen numbers into the box.

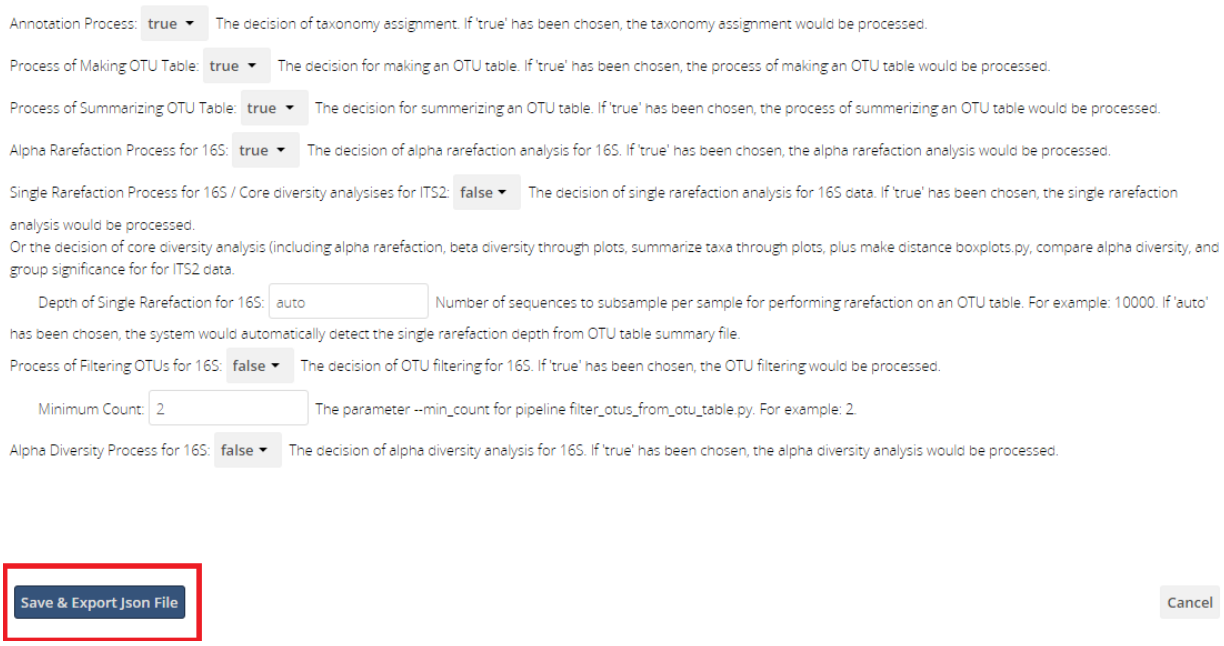
18. The next step is to determine the Mapping File, which provides information on the sequence such as sample time, soil temperature data, how to identify the sample etc. If you are providing a mapping file, you must choose ‘true’ from the grey drop-down box. If you are not providing a mapping file, you must choose ‘false’ and the system will automatically create a dummy mapping file.
19. The next step is to choose whether to validate the mapping file. If ‘true’ is chosen from the grey drop-down box, then the system will analyse the mapping file for errors. If ‘false’ is chosen, the process will not run.
20. Next, the process of joining paired ends will run if ‘true’

is selected or ‘false’ if you do not wish to run the process.
21. Next is the process of splitting libraries, which will produce a file that merges all of the sequences but adds a line to each FASTA header, telling the system what sample each sequence belongs to. If you wish to run this process, click the grey box

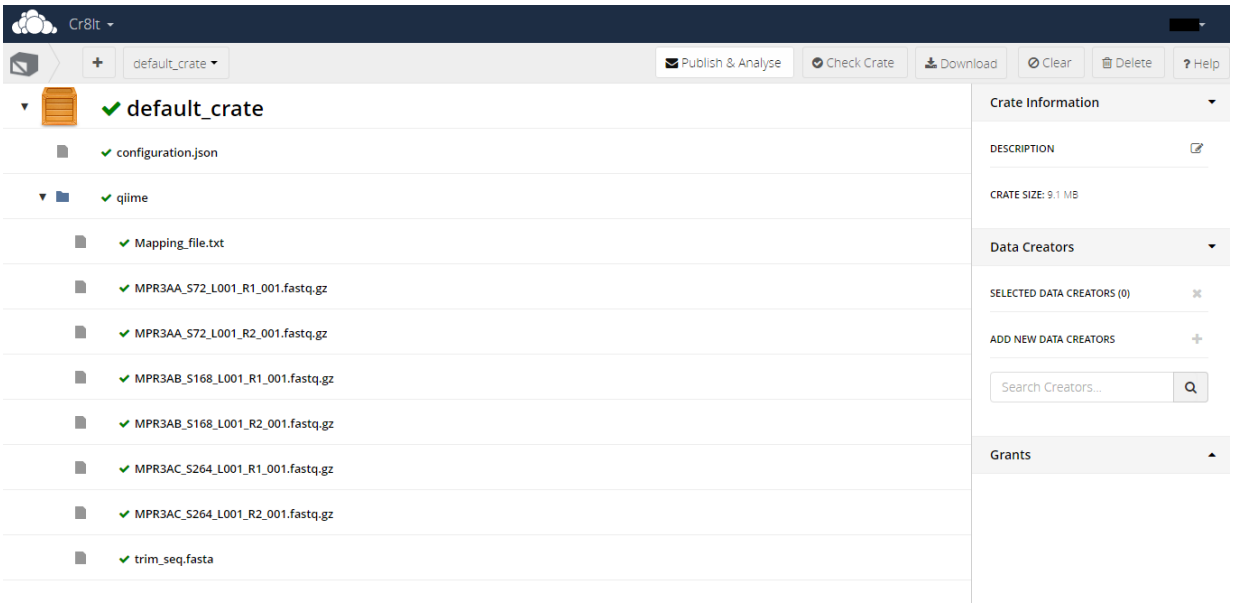
and select ‘true’ or ‘false’ if you do not wish to run the process.
22. Chimeras are artificial recombinants between two or more parental sequences, and they are normally formed when prematurely terminated fragments reanneal to other template DNA during PCR amplification. Once quality control processes are completed, chimeric sequences should be removed from the dataset. The process of removing chimeras will run using a ‘vsearch trimmer’ if ‘true’ is chosen from the grey drop-down box and will not run if ‘false’ is chosen.
23. Next, ‘OTU Picking’ will commence if ‘true’ is chosen from the grey drop-down box and will not run if ‘false’ is chosen. OTU Picking is a process that picks operational taxonomic units ‘de novo’ (i.e. clusters of sequences which share 97% identity).
24. Next is the ‘Process of Picking Representative Set of Sequences’ which picks a representative sequence for each OTU in order to compare to online databases. This process will run if ‘true’ is chosen from the grey drop-down box and will not run if ‘false’ is chosen.
25. Next, the ‘Annotation process’ assigns taxonomy to each representative sequence (i.e. annotating). This process will run if ‘true’ is chosen from the grey drop-down box and will not run if ‘false’ is chosen.
26. Next, an OTU table will be generated if ‘true’ is chosen from the grey drop-down box and will not run if ‘false’ is chosen.
27. Then, the next step summarizes the OTU table if ‘true’ is chosen from the grey drop-down box. The summarized OTU table helps to compare your sequence and taxonomy information to online databases.
28. Next, ‘Alpha Rarefaction Process for 16S’ and ‘Single Rarefaction Process for 16S / Core diversity analysis for ITS2’ can be run if ‘true’ is chosen from the grey drop-down boxes and will not run if ‘false’ is chosen.
29. Then, Depth of Single Rarefaction for 16S, a process which follows a number of sequences to subsample per sample for performing rarefaction on an OTU table. If 'auto' has been chosen, the system would automatically detect the single rarefaction

depth from OTU table summary file, however, you can manually put in a number of your choice (e.g. 10000) by writing it in the box next to 'Depth of Single Rarefaction for 16S'.

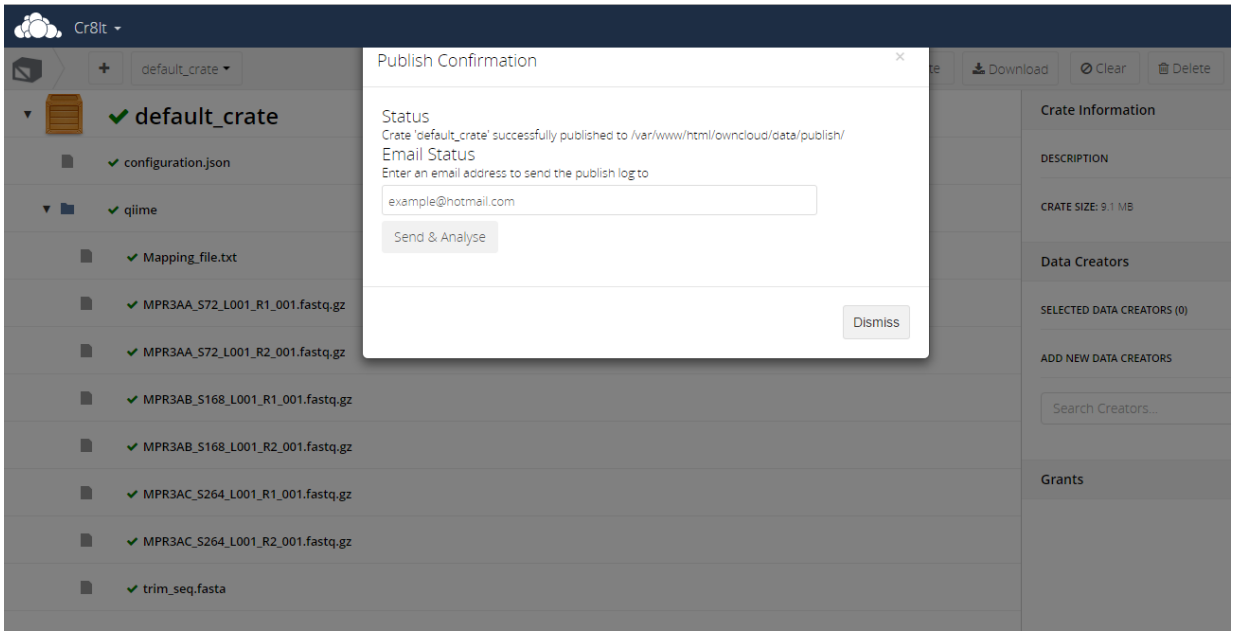
30. The 'Process of Filtering OTUs for 16S' filters the alignment and will be run if 'true' is chosen from the grey drop-down box and will not run if 'false' is chosen. Then you can specify the minimum parameter count for pipeline filtering (e.g. 2) by writing it in the box next to 'Minimum Count'.
31. The final step is 'Alpha Diversity Process for 16S' which measures the alpha diversity and will run if 'true' is chosen from the grey drop-down box and will not run if 'false' is chosen.
32. Finally, click the blue 'Save and Export Json File' at the bottom of the configuration page to save and export your file. Now you have finished the configuration.



33. Now you want to go back to the 'Cr8lt' page and click the 'Publish and Analyse' button at the top of the page, and click the 'Publish and Analyse' button again in the pop-up box after selecting the folder.

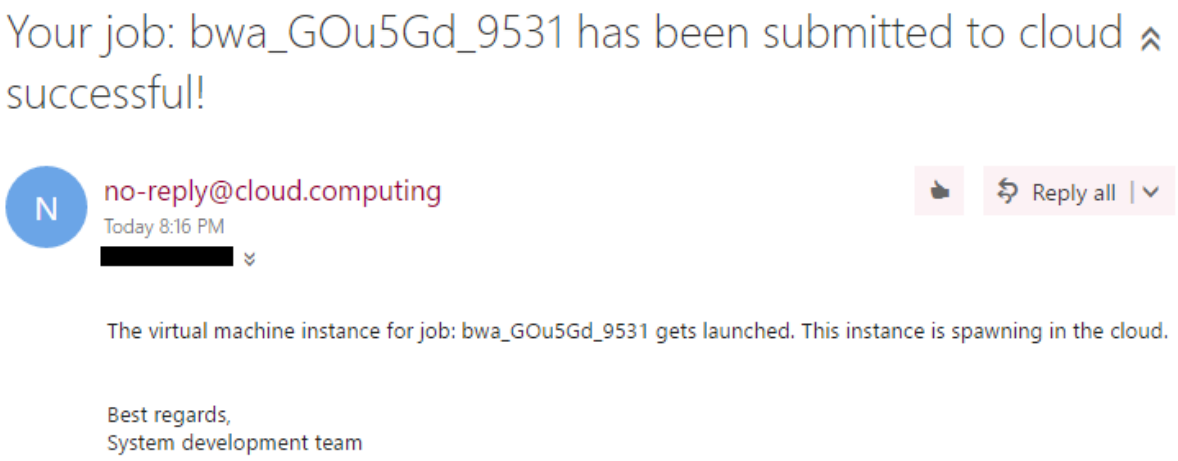


34. A ‘Publish Confirmation’ box will pop up and will ask you to input your email address. Type your email address into the box and click the ‘Send and Analyse’ button when finished. Be very careful typing in your email address or you will not receive the analysis. The system will automatically pack all the files into a single file for analysis

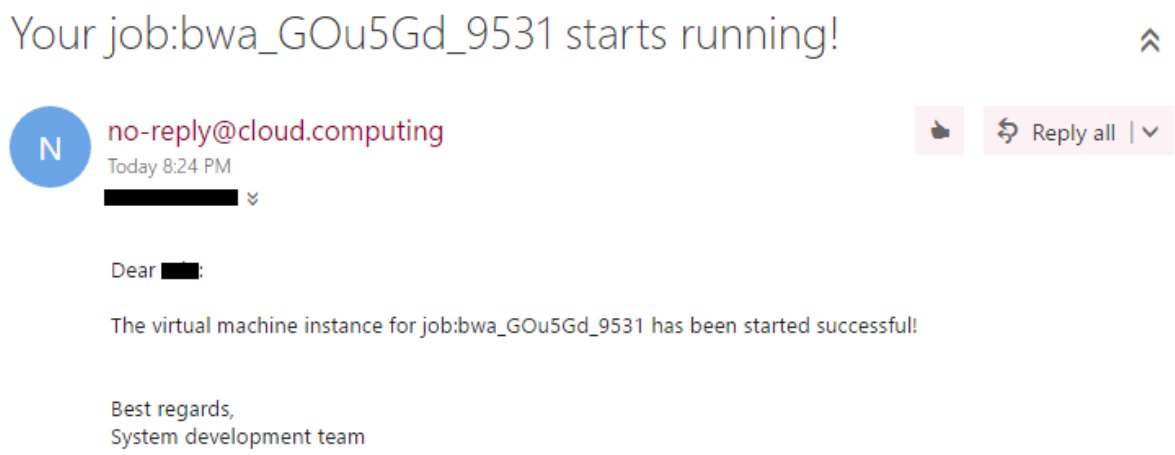


and then send them to your inputted email address. Now click the ‘Dismiss’ button.

35. You should then receive an email from TaxaCloud with your Job ID (e.g. Job: bwa_GOu5Gd) that has been submitted successfully.



36. You should then receive a second email from TaxaCloud to state that the analysis is currently taking place.



37. And finally you will receive a third and final email from TaxaCloud with your completed sequence analysis. The completed analysis is sent back to the project folder in TaxaCloud as an ‘Output Zip’ file.

Your job:bwa_GOu5Gd_9531 has been completed successful!

N

no-reply@cloud.computing

Fri 8:36 PM

[redacted]

Reply all

Dear [redacted]

Your job:bwa_GOu5Gd_9531 has been completed successful! The log details are listed below. Please look through the attached details. Thanks!

Best regards,
System development team

Analysing data type: 16s
Command:
cd /workspace/qiime-analysis && mkdir -p analysis/fastqc
Output:

Command:
cd /workspace/qiime-analysis && /workspace/software/FastQC/fastqc MPR3AC_S264_L001_R1_001.fastq.gz -o analysis/fastqc
Output:
Analysis complete for MPR3AC_S264_L001_R1_001.fastq.gz