

# Submitting RCC sequences to Genbank with Geneious

*Daniel Vaulot & Adriana Lopes dos Santos*

*Version 2.0 - 11 08 2018*

## Contents

<b>1 Aim of document</b>	<b>1</b>
<b>2 Assemble and clean sequences</b>	<b>2</b>
<b>3 Add informations to sequences</b>	<b>9</b>
3.1 Taxonomy . . . . .	9
3.2 Gene annotation . . . . .	13
3.3 Metadata . . . . .	15
3.4 Primers information . . . . .	17
<b>4 GenBank submission - General case (not for 16S, 18S or ITS, see next part)</b>	<b>19</b>
<b>5 GenBank submission - 16S, 18S or ITS</b>	<b>25</b>
5.1 Prepare files . . . . .	26
5.2 Submit to NCBI web portal . . . . .	34
<b>6 Appendixes</b>	<b>45</b>
6.1 Retrieve sequences from Genbank using Geneious . . . . .	45

## 1 Aim of document

This document explains how to use Geneious to :

- assemble and clean final sequences from several traces using different internal primers
- annotate the sequences
- submit to Genbank using Bankit through the Genious plug-in
- submit to Genbank for 18S, ITS and 16S that cannot be any more be submitted using Bankit

## Notes

- Look at legends **below** screen captures for directions.
- Changes from previous versions have been labelled with 

## 2 Assemble and clean sequences

- Import the ab1 trace
  - Drag and Drop



Figure 1: Import trace sequences

- Trim the sequences
  - Annotate & Predict → Trim Ends
  - Use error probability limit from 0.01 to 0.02 (increase to 0.05 if cannot assemble correctly, the trimming will be less drastic). For single reads (e.g. 528F) use a maximum of 0.02.

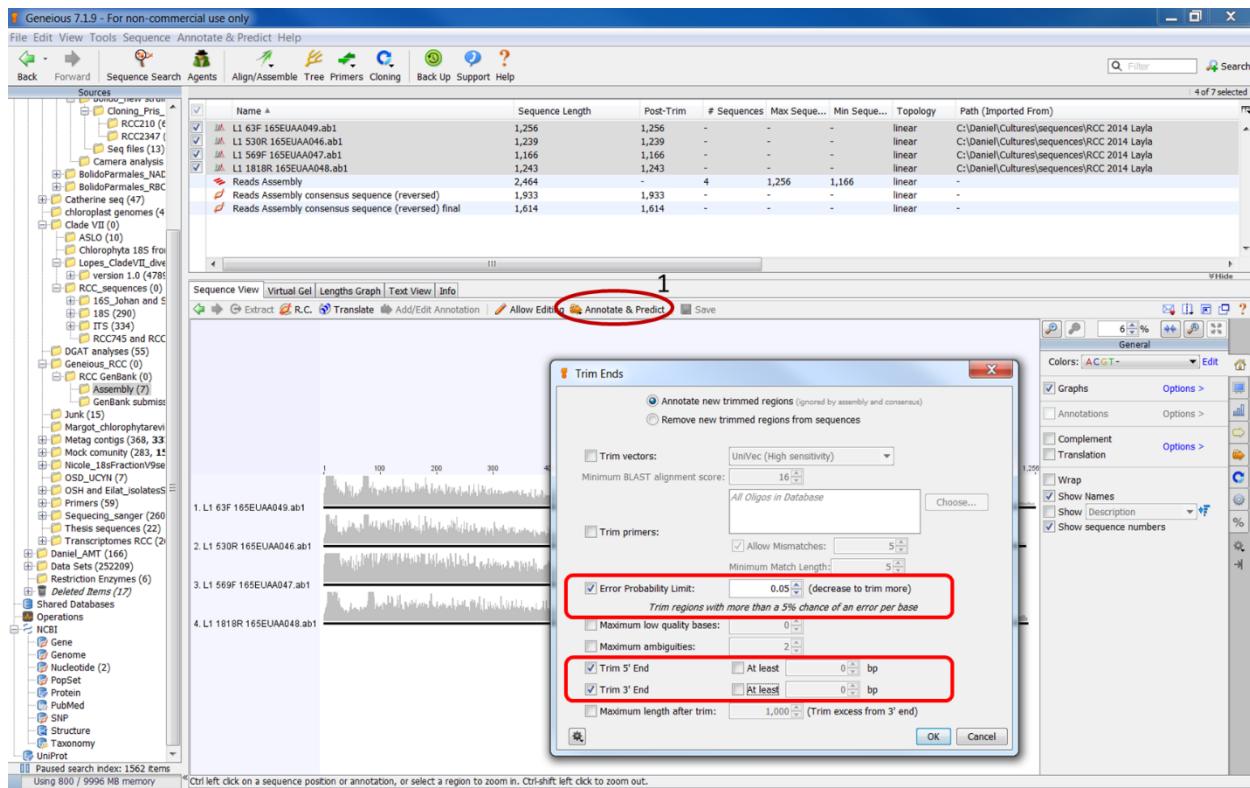


Figure 2: Trim sequences

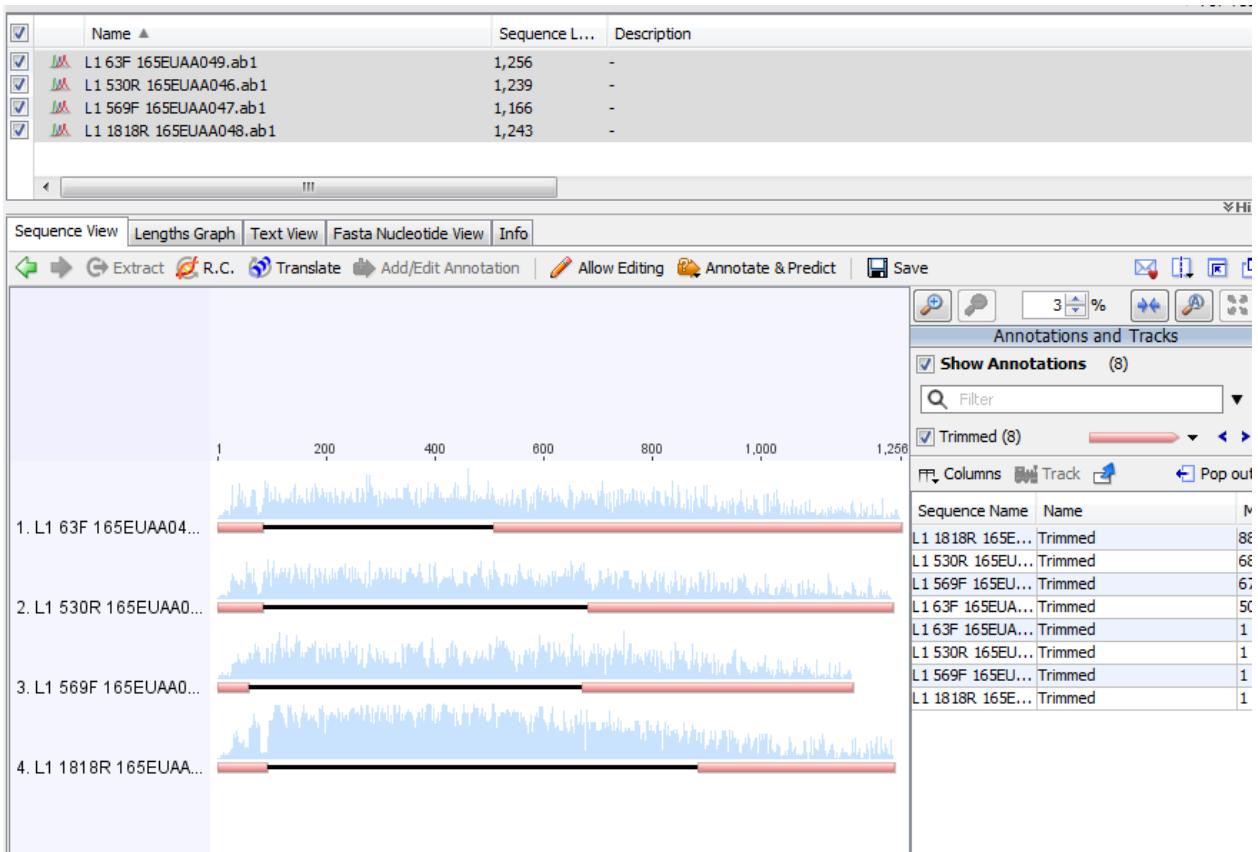


Figure 3: Visualize the trimmed sequences

- Assemble if several primers have been used
    - Align/Assemble → De Novo Assemble
    - Use for assembly name: RCC####\_gene-name\_your-initials\_date
      - \* e.g. RCC2497\_18S\_PG\_2018\_02\_15
-  The name should not contain any space
- Select “save the consensus”
  - Select “save contigs”.
  - You may have to change the trimming level (increase probability level - see above) if traces cannot be assembled

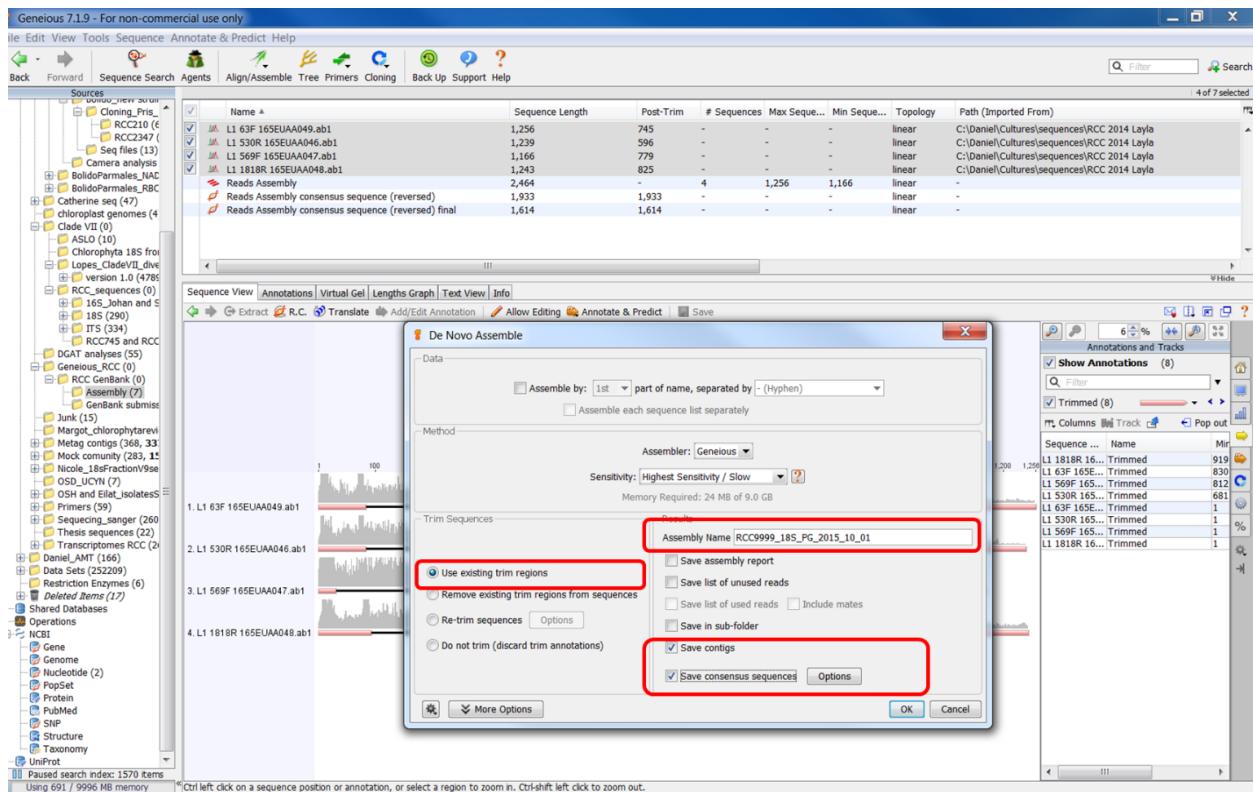


Figure 4: Assemble sequences

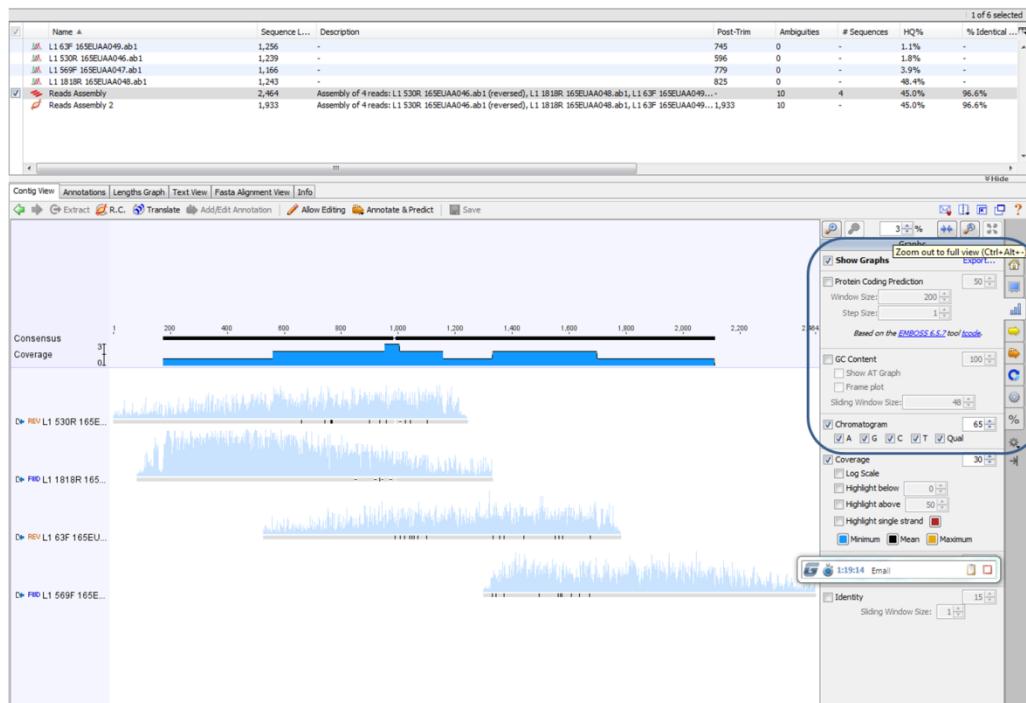


Figure 5: Visualize the assembled sequences

- Check the assembly and edit the consensus if necessary.



This is very important to make sure that your sequence is clean.  
 – Allow editing  
 – Edit bases that maybe wrongly assigned in one the trace.

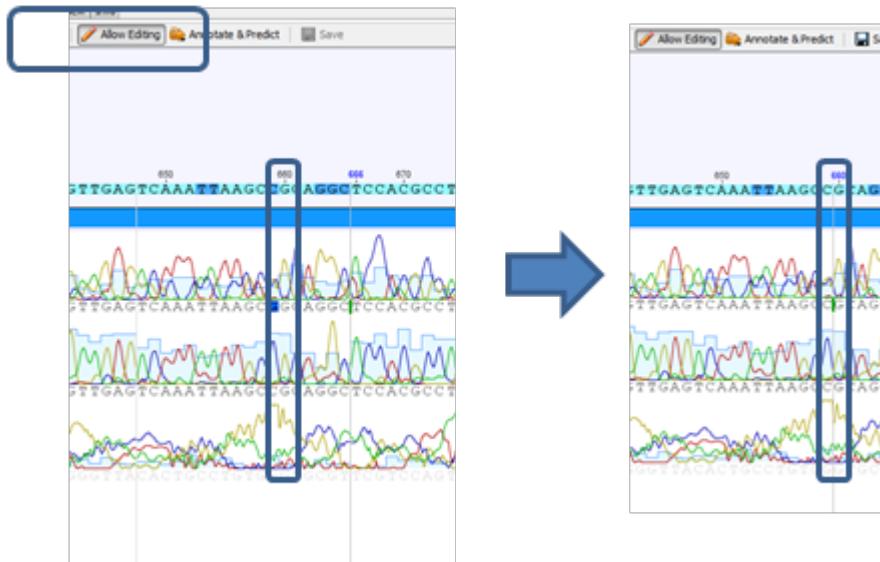


Figure 6: Check and correct assembly

- Select and extract consensus

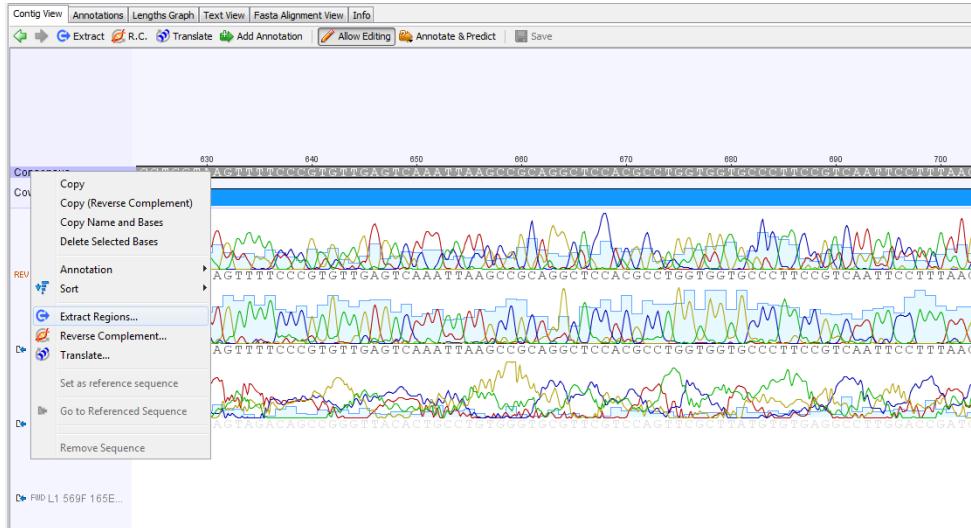


Figure 7: Extract consensus

- Reverse complement if necessary (if the sequence was assembled the other way around).
- Locate primers, test forward and reverse separately.  
   – Tools -> Primers > Test with saved primers

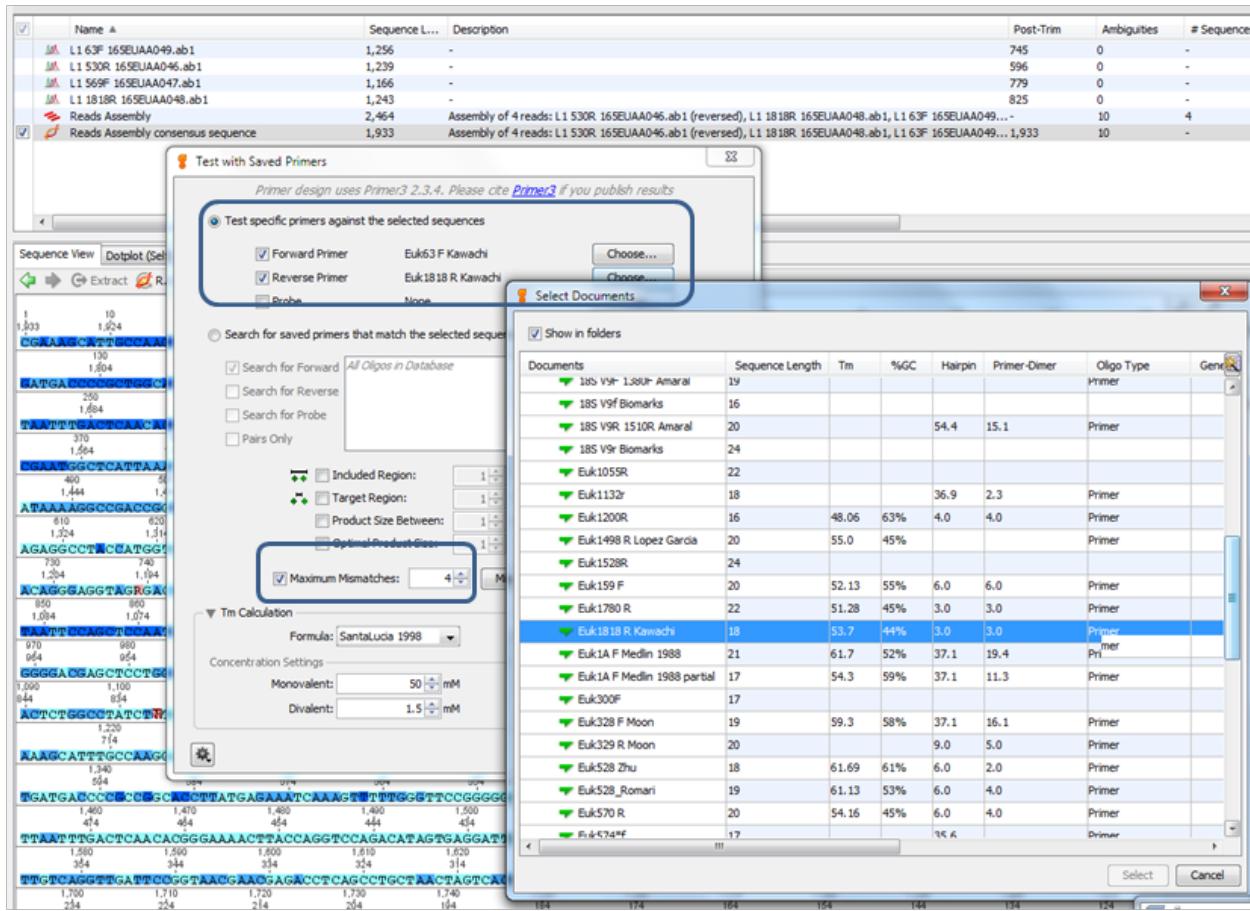


Figure 8: Test with saved primers

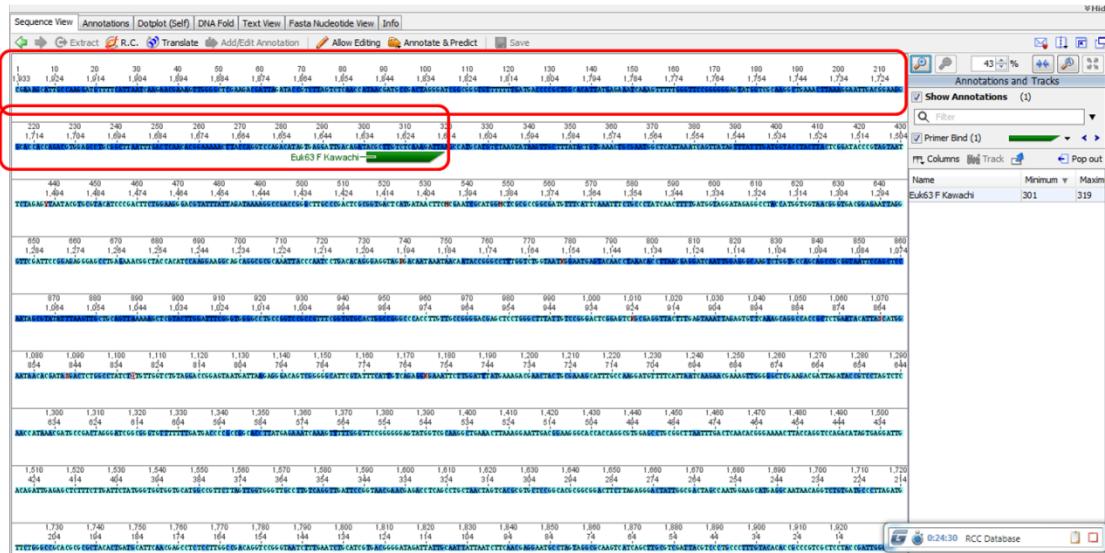


Figure 9: Locate primers

- Remove everything which is outside of primers **including the primers**.
  - allow editing
  - pressing the left button of the mouse, mark the region to be delete, It will show in green
  - press delete



Figure 10: Remove everything outside primers

- Et voilà, you have a clean sequence
  - The coloring corresponds to sequence quality based on the traces and assembly.

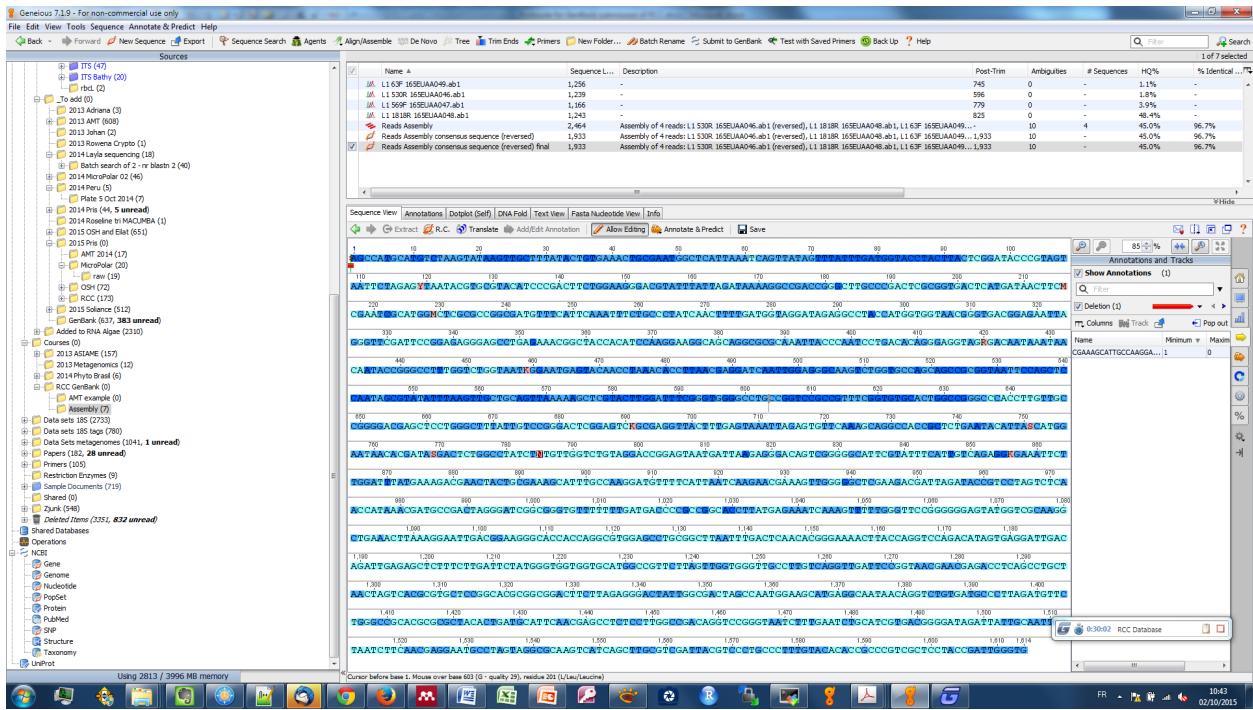


Figure 11: Final sequence

### 3 Add informations to sequences

#### 3.1 Taxonomy

- Do a batch BLAST search
  - Select the files -> "sequence search" or "Blast search"
  - Sometimes it does not work so you can do with the NCBI BLAST server
  - Pay attention to the following parameters:
    - \* **database** : nr genbank
    - \* **program** : blastn (for protein coding gene like rbcL the blastX can be also use to confirmed)
    - \* **results** : hit table
    - \* **maximum hits** : 25 at least
  - Parameters can be saved, recalled and deleted by clicking at the bottom of dialog box
    - \* save current settings -> name -> save.

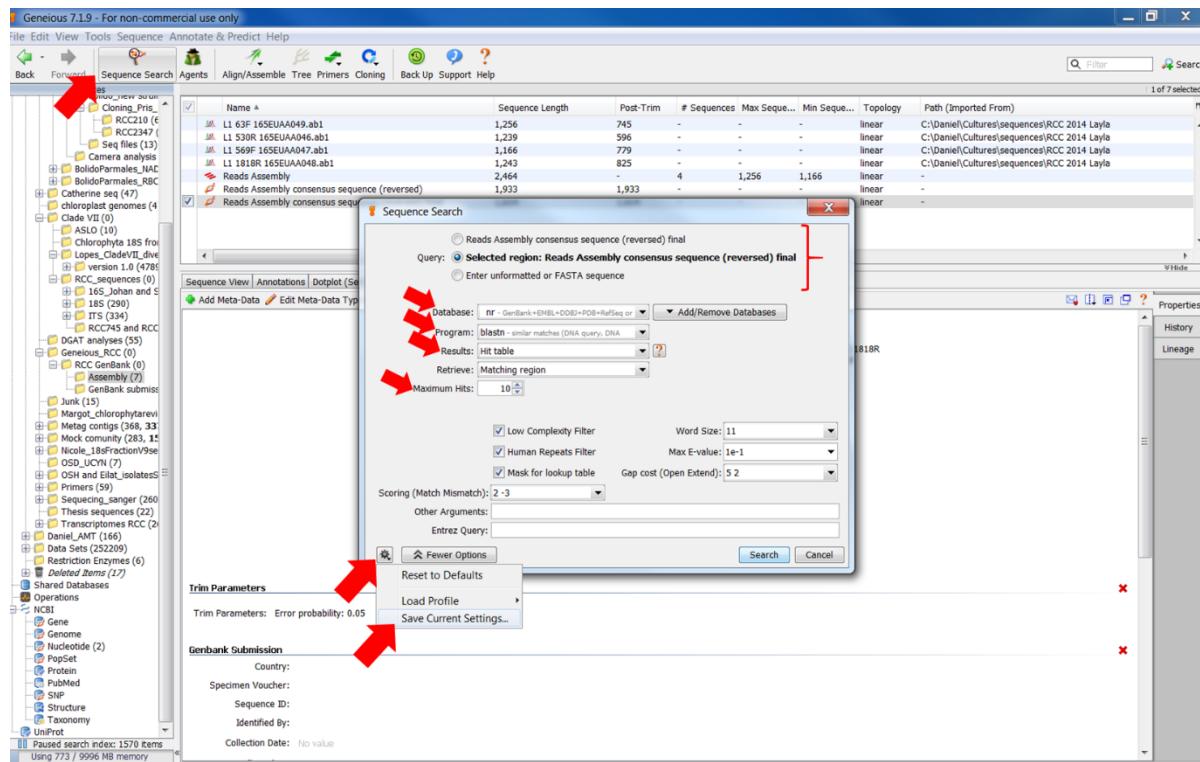


Figure 12: BLASTN

- Retrieve the closest sequence from GenBank (*Optional*)
  - From Geneious folder with the Blast results, select the closest result, drag the file into your folder in your local database if you wish to retain the file and/or modify it.
  - From Genbank: Copy the accession number -> go to NCBI -> nucleotide -> paste the accession number (look the figure bellow). You drag the file into your folder in your local database if you wish to retain the file and/or modify it.

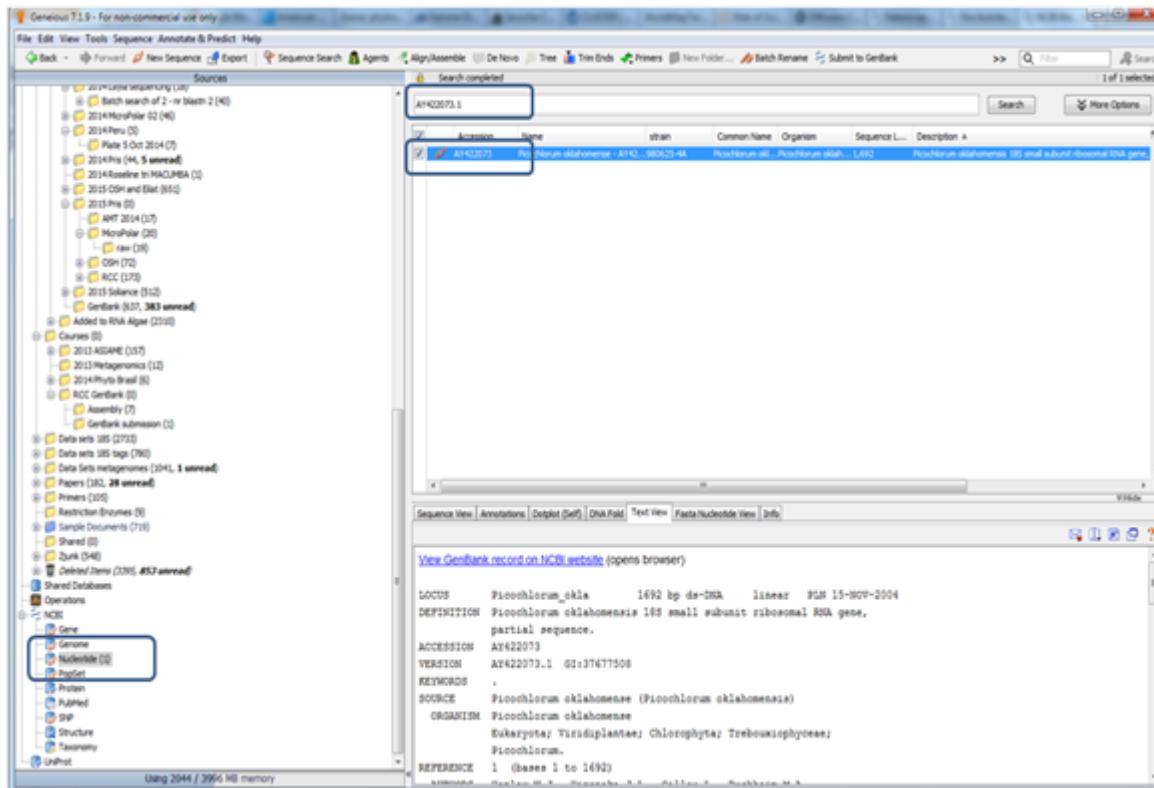


Figure 13: Retrieve closely related sequences from Genbank

- Do a manual alignment (*Optional*)
  - This is very useful to detect introns, for ITS sequences, combine gene sequencing partial 18S + 28S for example.
  - Align/Assemble -> Pairwise Align – MAFFT using the default parameters

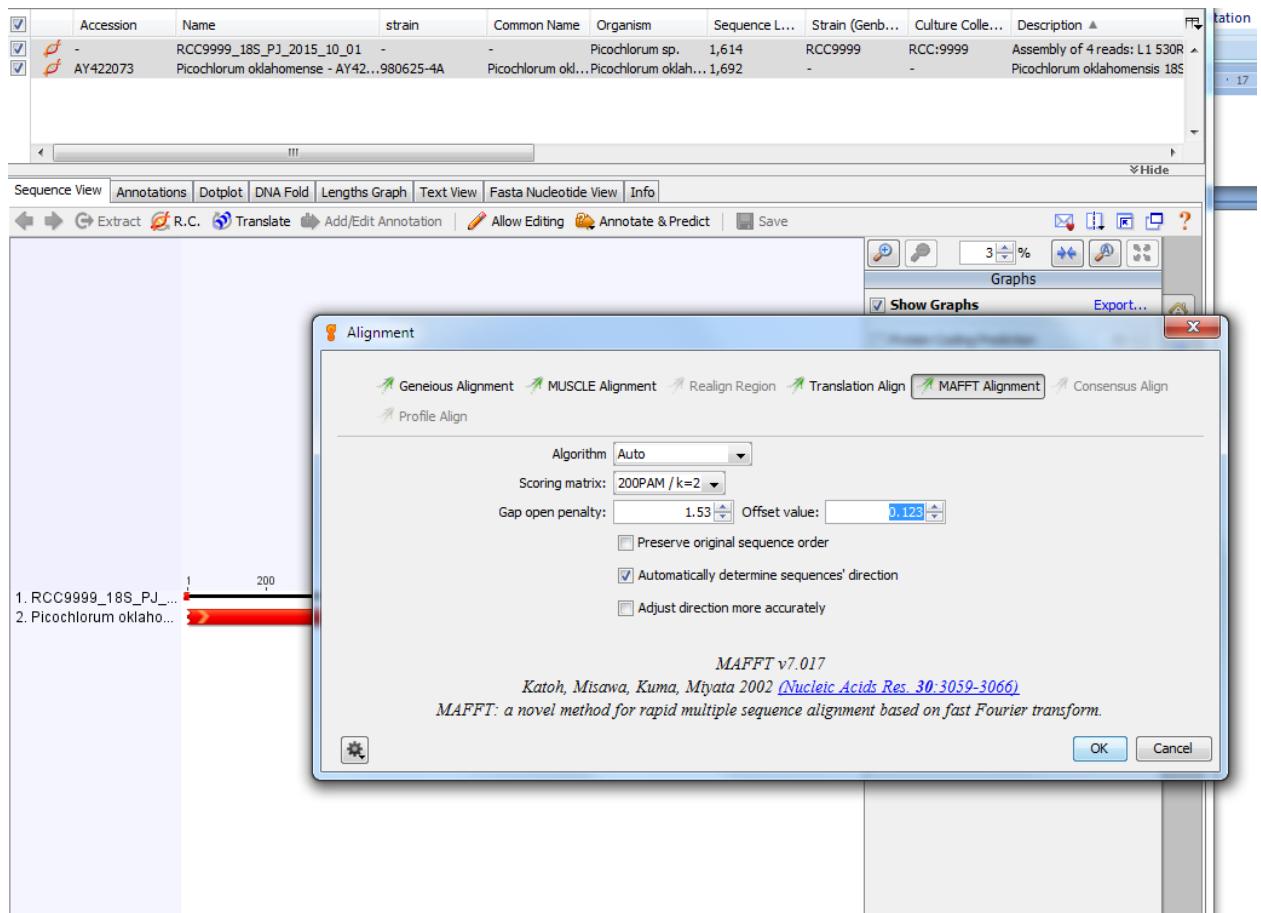


Figure 14: Alignement parameters

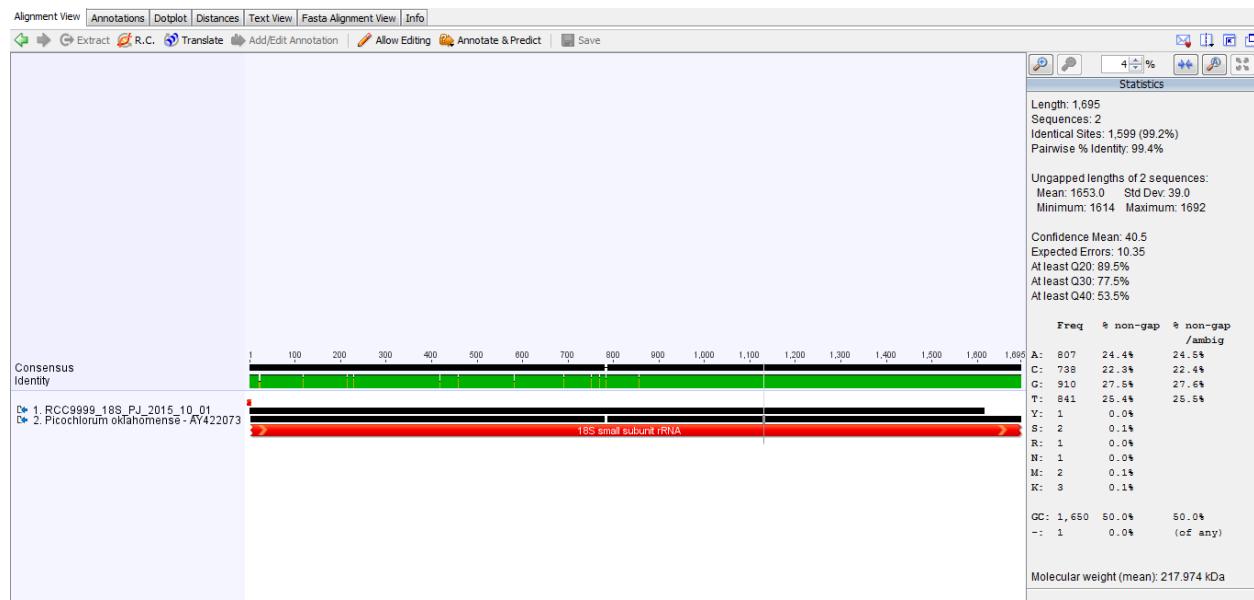


Figure 15: Alignement results

### 3.2 Gene annotation



This step is NOT necessary for 16S, 18S, ITS

- With the mouse, select your sequence, add notation
- Parameters to be changed (look the picture below)
  - Name:** name of the gene
  - Type:**
    - \* select rRNA for 18S, ITS, 16S platidial and 28S
    - \* CDS or gene coding sequence for example rbcl
- Add property using the 1st ADD: name = product, value = name of gene, for example 18S rRNA.
- Add annotation using the 2nd ADD (click in INTERVALS to see it): click in “truncated left end” and “truncated right end”. This is to tell that the sequence is not complete. For example, the 18S in this tutorial had the extremities before the primer removed, so is incomplete.



Make sure you do not have two annotations for the same gene !

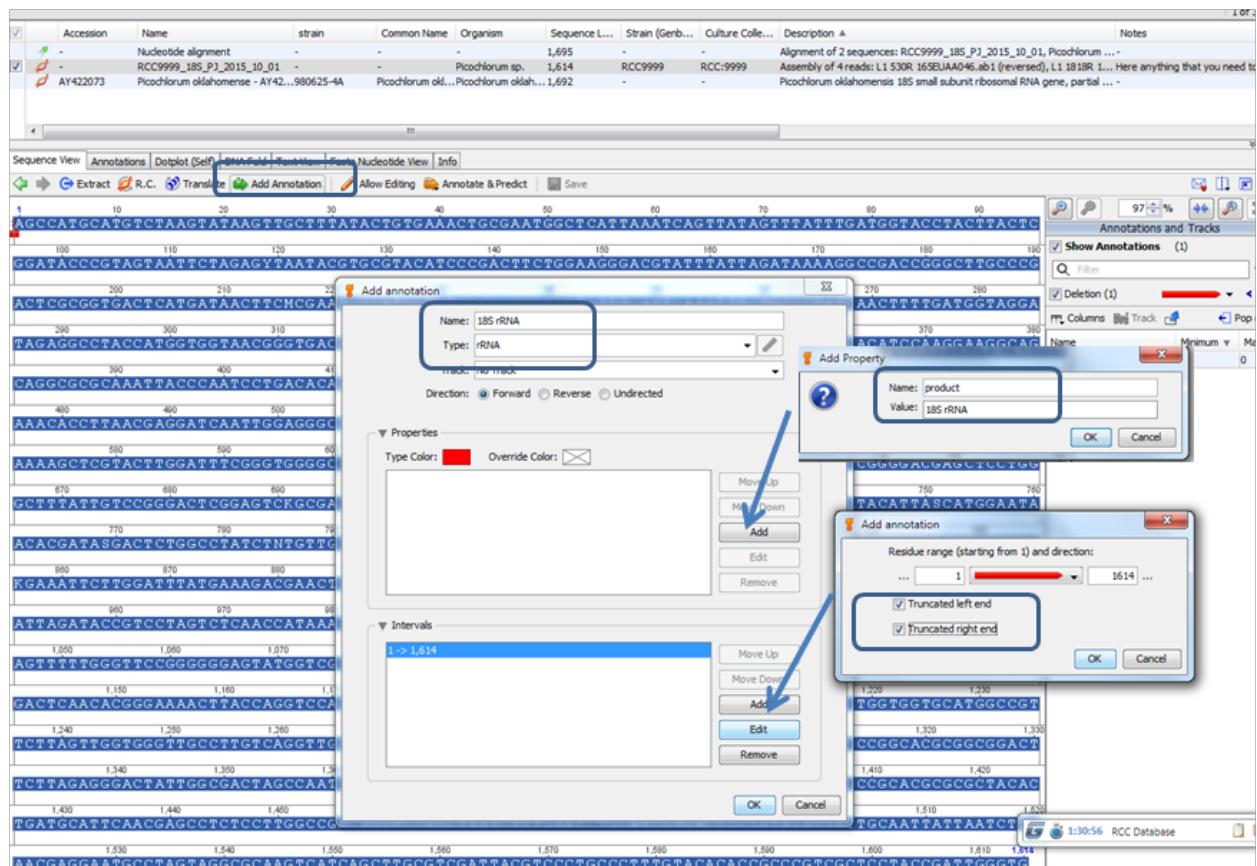


Figure 16: Annotate genes

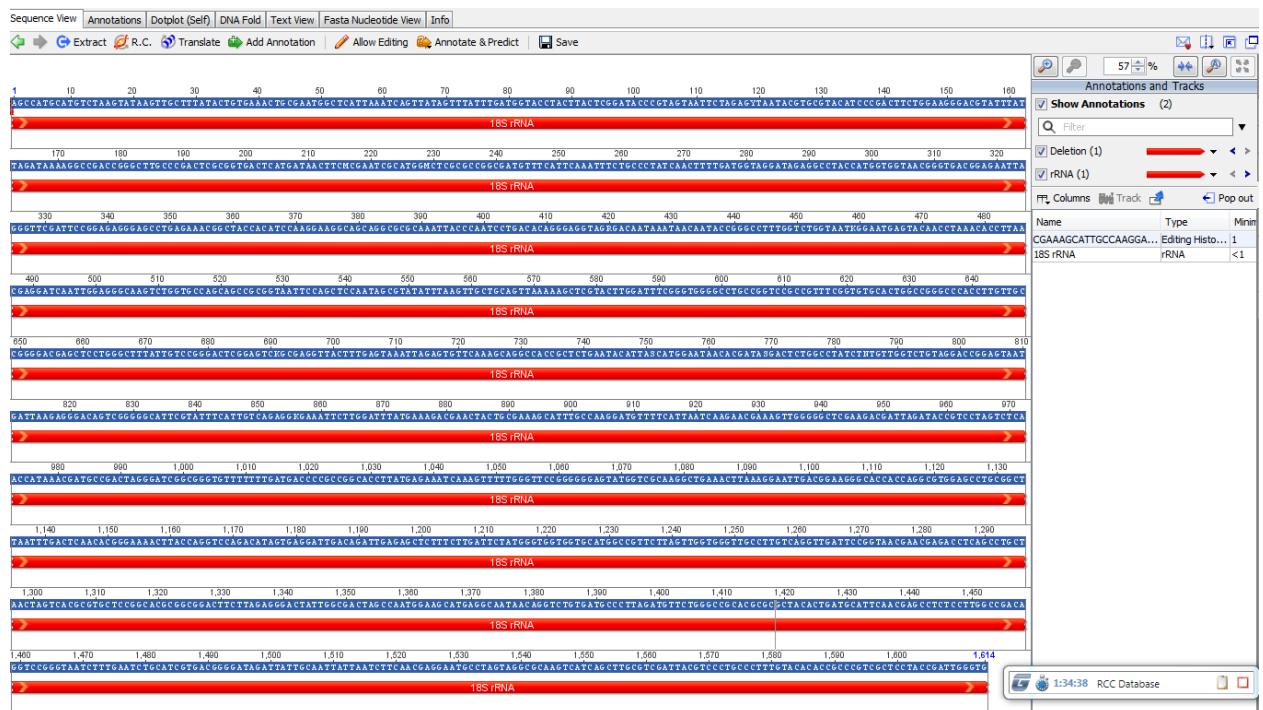


Figure 17: Annotated genes

### 3.3 Metadata

- Add three new type of metadata (it has to be done only once) in the GenBank submission category:
  - Sequence\_ID
  - Strain
  - Culture\_collection   Edit Meta data Types -> Genbank Submission -> click on the + on the right side ->  
write Culture Collection on the new field - > ok

**⚠** Make sure that these new fields are in the Genbank Submission category. Do not recreate a new category.

**⚠** Use exactly the orthograph for names especially with underscores “Culture\_collection” and not as before “Culture Collection”. There is already a field called “Sequence ID” but you must create a new field “Sequence\_ID” with an underscore.

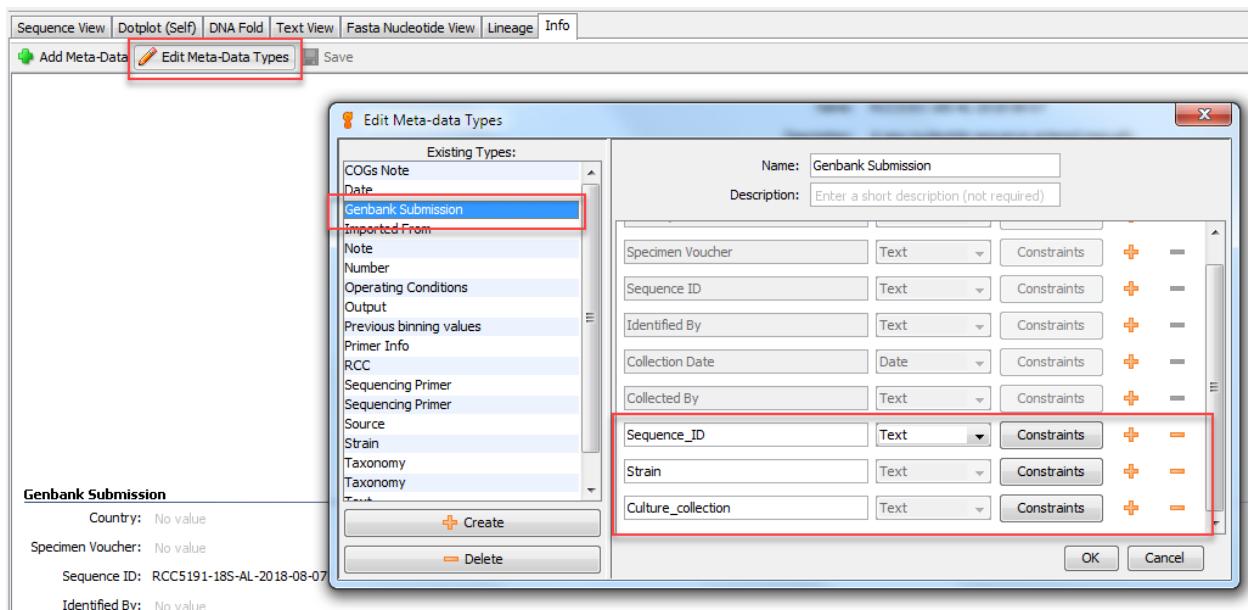


Figure 18: Add new meta-date type: Culture Collection

- Click on the final sequence, go to info and change or correct the following fields.
  - **Name** : RCC####\_gene-name\_your-initials\_date, e.g. RCC9999\_18S\_PG\_2015\_10\_01 (change if it is not in this format at this point).
    - \* This will be the ID of the sequence submitted to GenBank.
    - \* This name must not contain any space
    - \* This name must be unique. For example if you submit 2 sequences for the same strain and same gene you must use different names e.g.RCC9999\_18S\_PG\_2015\_10\_01\_A and RCC9999\_18S\_PG\_2015\_10\_01\_B
  - **Sequence\_ID** : It should be the same as the Name (see batch edit below to copy the Name to this field).
  - **Organism** : Picochlorum sp. or Trebouxiophyceae.
    - \* Enter the genus name or, if not known, the lowest taxonomic level known.
    - ⚠** Only use the species name if **you are absolutely sure** of the species as determined by

microscopy or ITS. Do not rely on BLAST!!

- \* DO NOT add the RCC number at the end of the organism name.
- \* For levels above the genus, do not use sp. For example use **Trebouxiophyceae** and not **Trebouxiophyceae sp.** or **Chlorophyta** and not **Chlorophyta sp.**
- **Strain** : This is the RCC code as **RCCxxxx** without space between RCC and number e.g. **RCC1236**.
- **Culture\_collection** : This is the RCC number as **RCC:xxxx** with “:” between RCC and number e.g. **RCC:1236**.

Name	Organism	Sequence L...	Strain (Genb...)	Culture Colle...	Description	Notes
<input checked="" type="checkbox"/> RCC9999_18S_P1_2015_10_01	Picohlorum sp.	1,614	RCC9999	RCC:9999	Assembly of 4 reads: L1 530R 16SEUAA046.ab1 (reversed), L1 1818R 16SEUAA048.ab1, L1 63F 16SEUAA049.ab1 (reversed), L1 569F 16SEUAA047.ab1	Post-Trim 1,614

Sequence View | Annotations | Dotplot (Self) | DNA Fold | Text View | Fasta Nucleotide View | Info

Add Meta-Data | Edit Meta-Data Types | Save

**Name:** RCC9999\_18S\_P1\_2015\_10\_01  
**Description:** Assembly of 4 reads: L1 530R 16SEUAA046.ab1 (reversed), L1 1818R 16SEUAA048.ab1, L1 63F 16SEUAA049.ab1 (reversed), L1 569F 16SEUAA047.ab1

**Common Name:** No value  
**Genetic Code:** No value  
**Free end gaps:** true  
**Modified:** 02 Oct 2015  
**Molecule Type:** DNA  
**Organism:** Picohlorum sp.  
**Taxonomy:** No value  
**Topology:** linear  
**Notes:** Here anything that you need to remember....

**Trim Parameters**  
Trim Parameters: Error probability: 0.05

**Genbank Submission**  
Country: No value  
Specimen Voucher ID: No value  
Sequence ID: No value  
Identified By: No value  
Collection Date: No value  
Collected By: No value  
Specimen Voucher: No value  
Strain: RCC9999  
Culture Collection: RCC:9999

Figure 19: Update the different meta-data fields

- It is possible to quickly change metadata for a set of sequences using the Batch edit mode. For example you can :
  1. Copy the Strain field to the Culture Collection field
  2. Add the “.” automatically for all sequences by replacing “RCC” by “RCC:.”

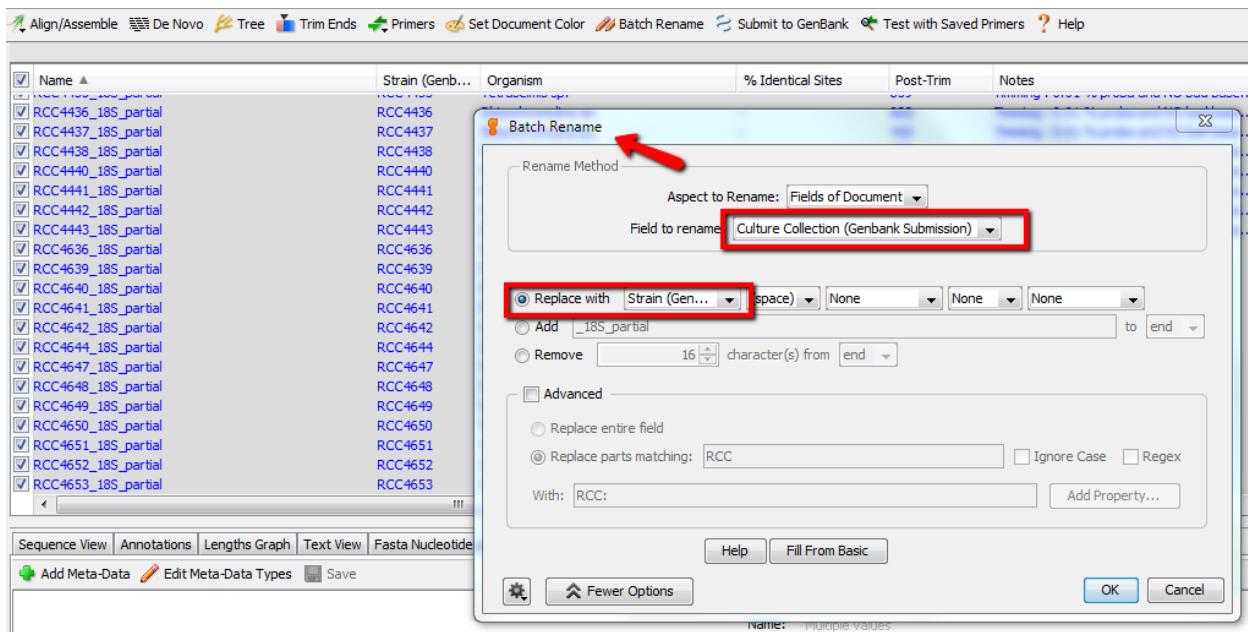


Figure 20: Batch edit - simple

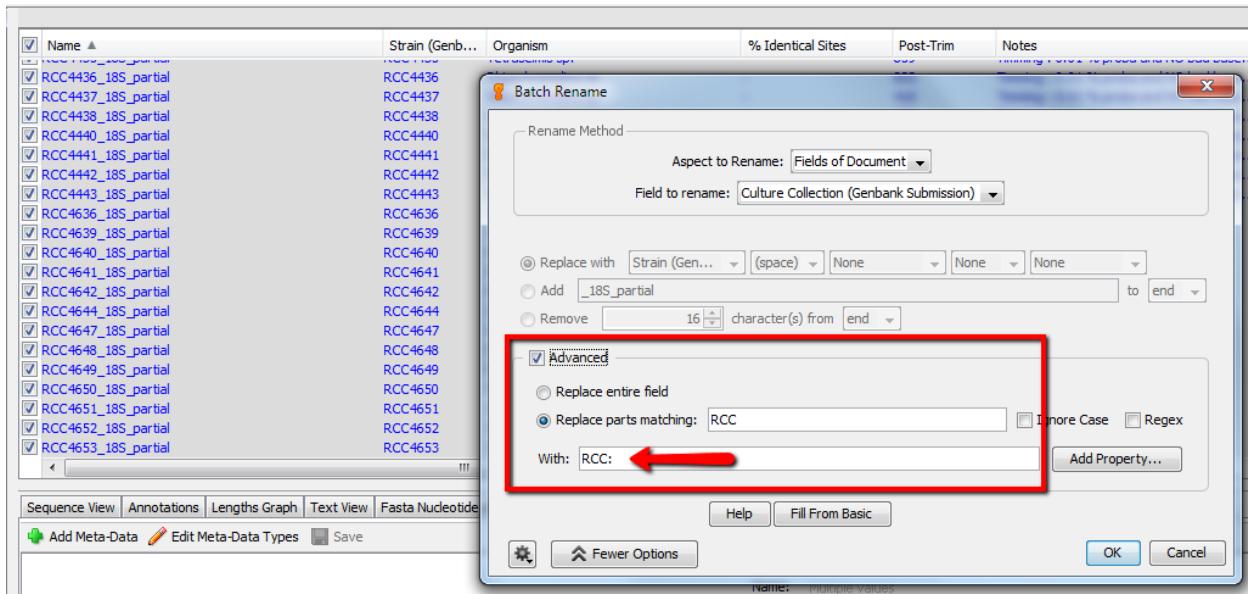


Figure 21: Batch edit - advanced

### 3.4 Primers information

**⚠ This step is Optional for 18S, but must be added for ITS, 28S and other genes**

Edit Meta data Types -> Sequencing Primer -> OK

You can also use **Batch edit** to go faster

**Sequencing Primer**

---

Forward PCR Primer Name: 18S 63F

Forward PCR Primer Sequence: ACGCTTGTCCTCAAAGATTA

Reverse PCR Primer Name: 18S 1818R

Reverse PCR Primer Sequence: ACGGAAACCTTGTACGA

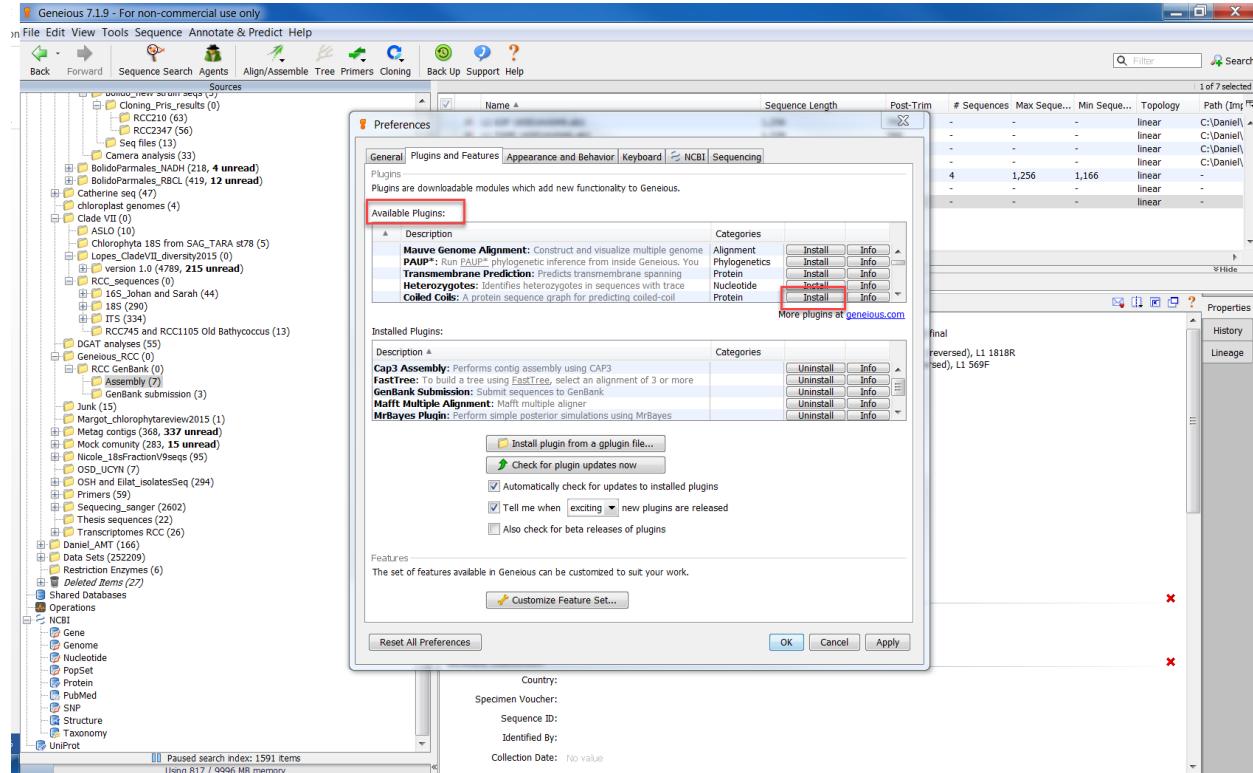
Forward Sequencing Primer Name: No value

Forward Sequencing Primer Sequence: No value

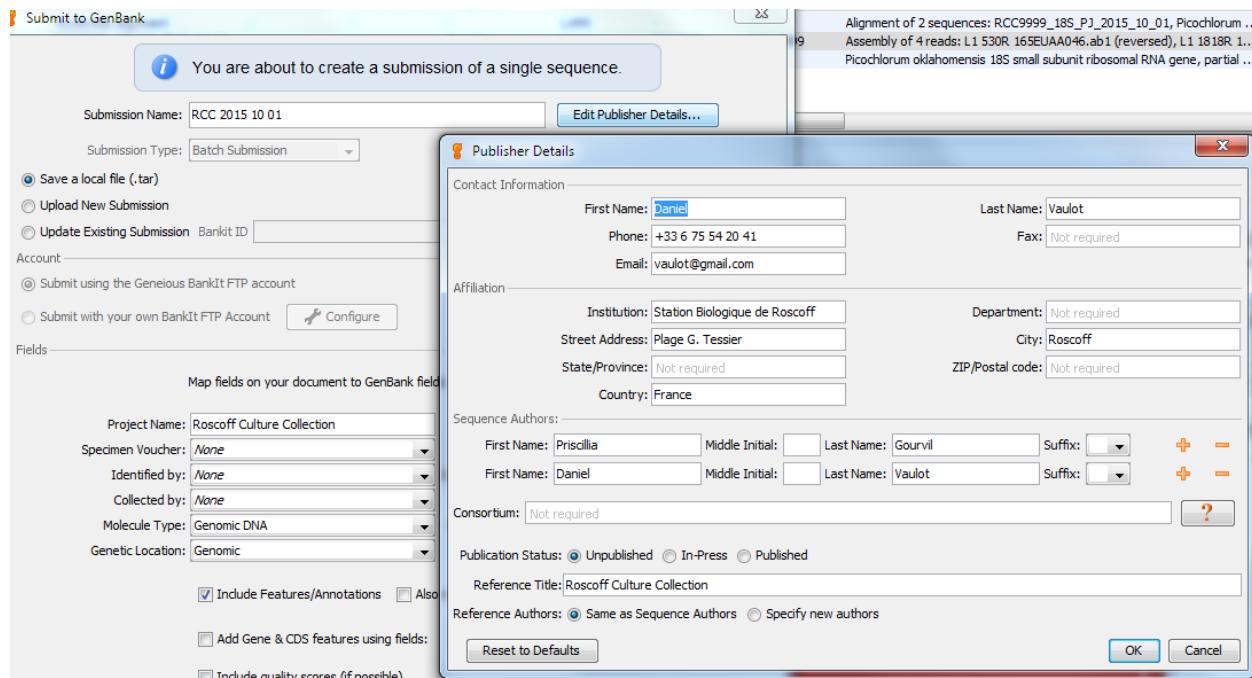
Figure 22: Add primer information to meta-data

## 4 GenBank submission - General case (not for 16S, 18S or ITS, see next part)

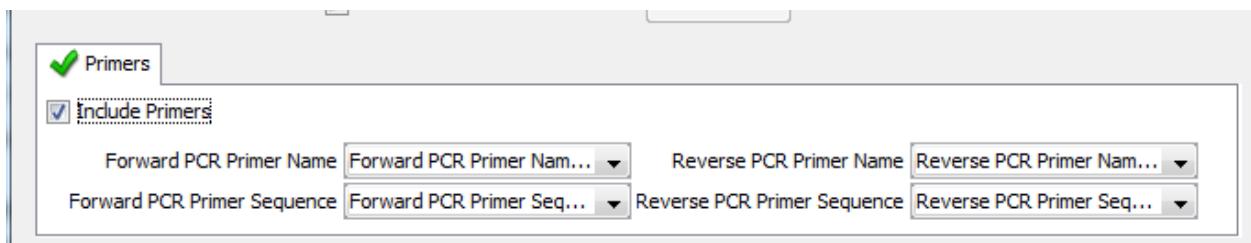
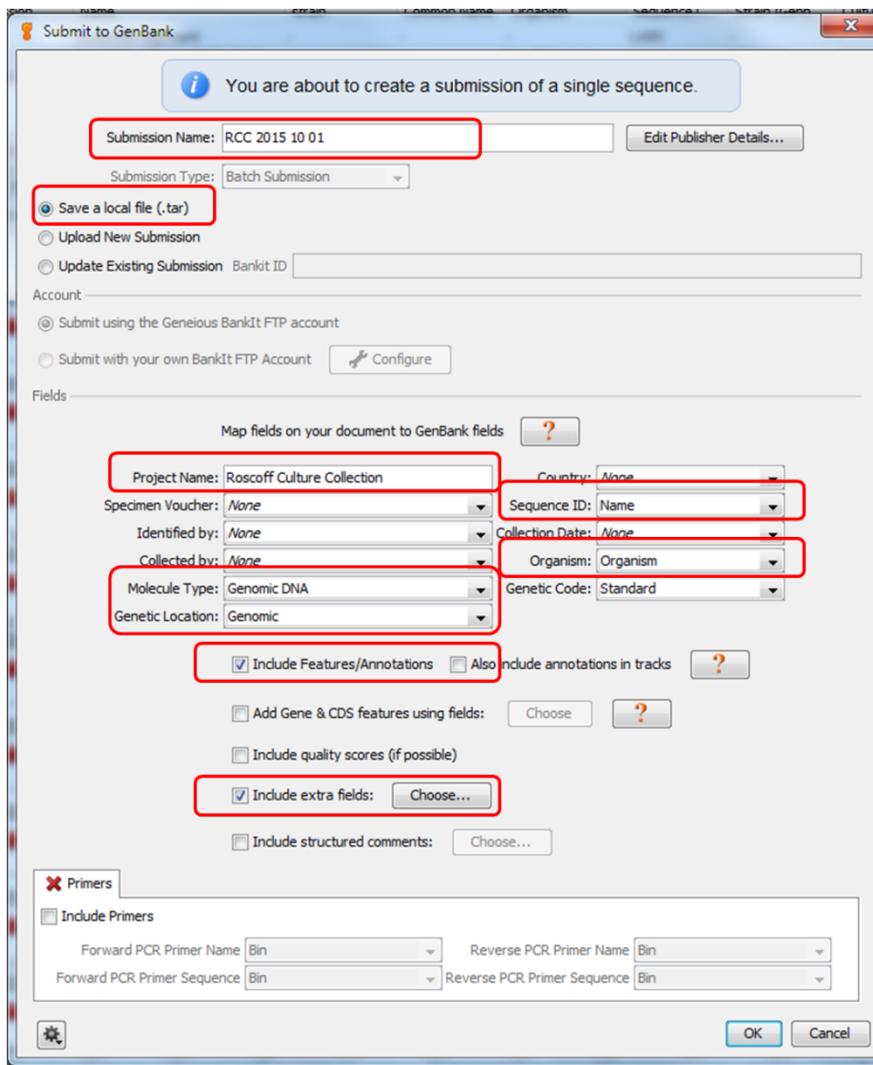
- Note that since August 2018 16S, 18S, 28S and ITS cannot be submitted by BankIt and must submitted through a web interface.
- Install plugin GenBank submission
  - Tools → plugin → choose the plugin and click in install

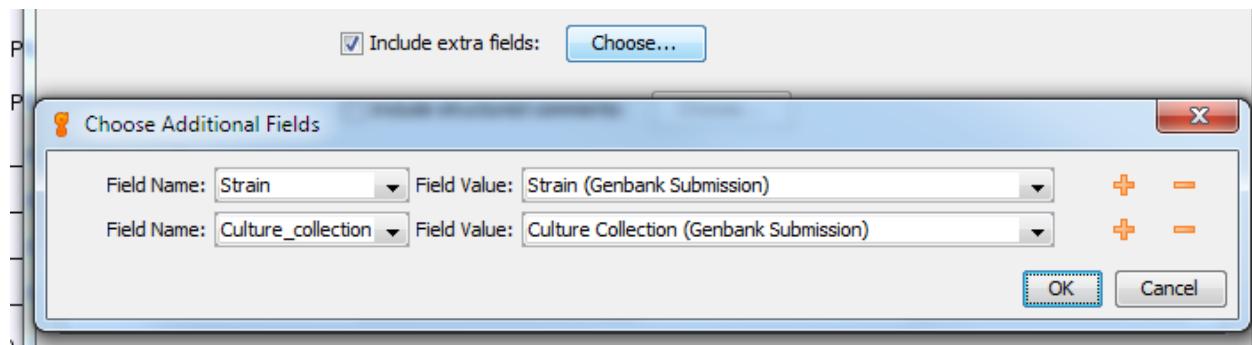


- Select the sequences you want to submit
- Select GenBank submission
- Enter first the **Publisher details** (add the info like the picture bellow, except that the sequence authors is Daniel Vaulot + who did the sequence)
  1. Name
  2. email
  3. Adress
  4. Sequence authors
  5. Select **Unpublished**
  6. Reference should be “Roscoff Culture Collection”



- Check very carefully all the fields
  - **Submission name** : the name of the file to be saved (this should be kept on the Databases computer)
  - **Save a local file** (only upload when everything is OK)
  - **Project name** : Roscoff Culture Collection
  - **Molecule type** : Genomic DNA
  - **Genetic location** : in general Genomic but can also be Plastid or Mitochondrion or Nucleomorph for Cryptophytes
  - **Sequence ID** : Name
  - **Organism** : Organism
  - **Include features/annotation** : Yes
  - **Include other fields** : Yes
    - \* **Culture\_collection** : Culture\_collection (GenBank submission)
    - \* **Strain** : Strain (GenBank submission).
  - **Primers** : You can put the primers if necessary but they need to be entered Sequencing primers





- Check submission in the Preview mode
  - If Errors you need to correct
  - Ignore warning about “Organism not found” and “Collection”

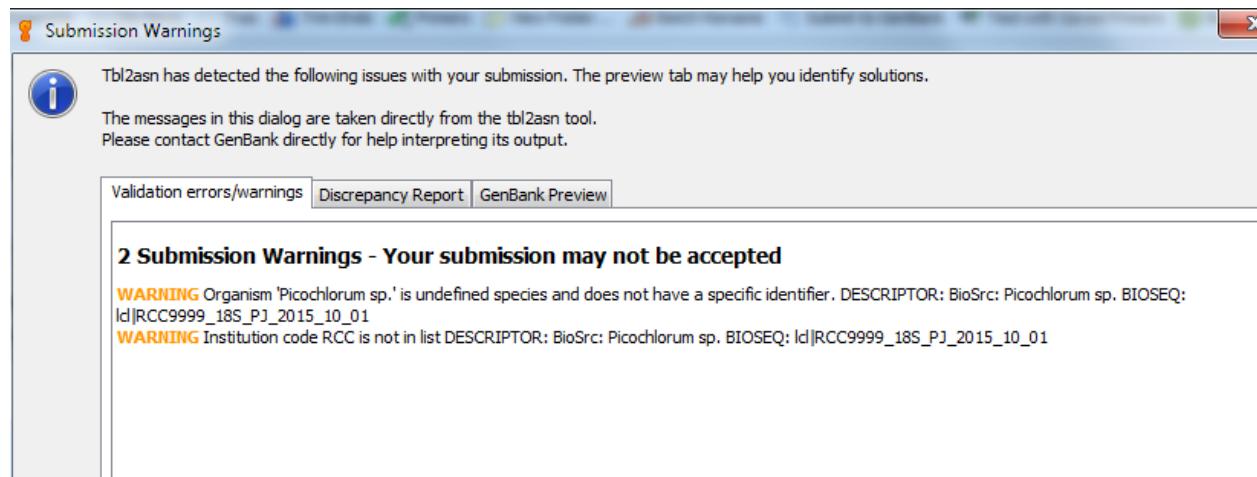


Figure 23: Warnings - Ignore

Submission Warnings

Tbl2asn has detected the following issues with your submission. The preview tab may help you identify solutions.

The messages in this dialog are taken directly from the tbl2asn tool.  
Please contact GenBank directly for help interpreting its output.

	Validation errors/warnings	Discrepancy Report	GenBank Preview	
LOCUS	RCC9999_18S_PJ_2015_10_011614 bp	DNA	linear	02-OCT-2015
DEFINITION	Picochlorum sp. strain RCC9999.			
ACCESSION				
VERSION				
KEYWORDS	.			
SOURCE	Picochlorum sp.			
ORGANISM	Picochlorum sp.			
	Unclassified.			
REFERENCE	1 (bases 1 to 1614)			
AUTHORS	Gourvil, P. and Vaulot, D.			
TITLE	Roscoff Culture Collection			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 1614)			
AUTHORS	Gourvil, P. and Vaulot, D.			
TITLE	Direct Submission			
JOURNAL	Submitted (02-OCT-2015) Station Biologique de Roscoff, Plage G. Tessier, Roscoff, France			
FEATURES	Location/Qualifiers			
source	1..1614 /organism="Picochlorum sp." /mol_type="genomic DNA" /strain="RCC9999" /isolate="RCC9999" /culture_collection="RCC:9999" /PCR_primers="fwd_name: 18S 63F, fwd_seq: acgcttgtctcaaagatta, rev_name: 18S 1818R, rev_seq: acggaaaccttgttacga"			
rRNA	<1..>1614 /product="18S ribosomal RNA"			
ORIGIN	1 agccatgcat gtctaagtat aagttgcattt atactgtgaa actgcgaatg gtcattaaaa 61 tcagtttag tttatttgat ggtacctact tactcgata cccgtatcaa ttcttagagyt 121 aatacgtgcg tacatcccga cttctgaaag ggacgtatTTT attagataaa aggccgaccg 181 ggcttggcccg actcgcgggtg actcatgata acttcmcgaa tcgcattggmc tcgcggccggc 241 gatgtttcat tcaaatttct gccttatcaa cttttatgg taggatagag gcctaccatg 301 gtggtaacgg gtgacggaga attagggttc gattccggag agggagcctg agaaacggct 361 accacatcca aggaaggcag caggcgcga aattacccaa tcctgacaca gggaggttagr 421 gacaataaat aacaataccg ggcctttgtt ctggtaatkgaatgagttac aacctaaaca 481 ccttaacgg gatcaattgg agggcaagtc tggtgccagc agccgcggta attccagctc 541 caatagcgtt tatttaagtt gctgcgttaaaatgcgtt acttggattt cgggtggggc 601 ctggccgtcc ggcgttccgg tggacttgg ccggccccac ctttgtccgg gggacggact 661 cctgggtttt attgtccggg actcggagtc kgcgaggta ctttgatgg attagagtgt 721 tcaaaggcagg ccaccgctctt gaatacattt scatgaaata acacgatag actctggcct 781 atctntgttg gtctgttagga ccggatgtat gattaaggagg gacagtccgg ggcattcgta 841 .....			

Figure 24: Genbank record preview

- Save as tar file.
  - The submission has to be done before processing a new one starts because Geneious keep at the memory the info from the last .tar file you saved.

- The tar file can be uncompressed to an .asn file which can be opened with Sequin which can be downloaded from NCBI.
- Finally submit using the Geneious BankIt account and record the BankIt number

The screenshot shows a web-based submission interface for GenBank. At the top, there's a blue header bar with the text "Submit to GenBank". Below it is a light blue information box containing an info icon and the text "You are about to create a submission". The main form area has several sections:

- Submission Name:** RCC 2015 10 01
- Submission Type:** Batch Submission
- File Options:** Radio buttons for "Save a local file (.tar)" (unchecked), "Upload New Submission" (checked), and "Update Existing Submission" (unchecked). A "BankIt ID" input field is next to the update option.
- Account Options:** Radio buttons for "Submit using the Geneious BankIt FTP account" (checked) and "Submit with your own BankIt FTP Account". A "Configure" button with a wrench icon is next to the second option.
- Fields:** A section labeled "Fields" at the bottom.

Figure 25: Submit to GenBank

## 5 GenBank submission - 16S, 18S or ITS

Submission must now be done at <https://submit.ncbi.nlm.nih.gov/subs/genbank/>. If you do not have a login you must create one.

Information about the NCBI submission portal is here. We recommend to read these instructions very carefully before submitting the sequences.

The main steps are :

1. Create fasta file with unique name for each sequence.
  - Sequence Name (Sequence\_ID) cannot contain spaces. The Sequence\_ID identifies the same specimen in all the steps of a submission. We use a convention of the following type RCC9999\_18S\_PG\_2015\_10\_01 (see above)
  - Sequence Name must be unique within the set and may not contain spaces.
  - Sequence Name may contain only the following characters - letters, digits, hyphens (-), underscores (\_), periods (.), colons (:), asterisks, and number signs(#).
2. Create a tabulated file as Text (tsv - tab-delimited) containing all the information about the sequence. See this link for the description of all the modifiers. This file can be easily exported from Geneious and finalized with Excel. For the RCC, the following columns are necessary (fields in bold are **mandatory**):
  - **Sequence\_ID** - This field must be set to be equal to the Name
  - **Organism** - Picochlorum sp. or Trebouxiophyceae sp.
  - **Strain** - e.g. RCC1236.
  - **Culture\_collection** - e.g. RCC:1236.
  - Fwd\_primer\_name - name of forward PCR primer
  - Fwd\_primer\_seq - nucleotide sequence of forward PCR primer
  - Rev\_primer\_name - name of reverse PCR primer
  - Rev\_primer\_seq - nucleotide sequence of reverse PCR primer

 The Sequence\_ID must rigorously identical to the name field because the FASTA file will be saved using the Name field while the tsv file contains the Sequence\_ID. If you do changes to the Name, you must copy it to the Sequence\_ID using batch rename.

## 5.1 Prepare files

We will do a simple case but you can add more columns (see list of modifiers.

- Fasta file
  - Select sequences
  - Export as fasta

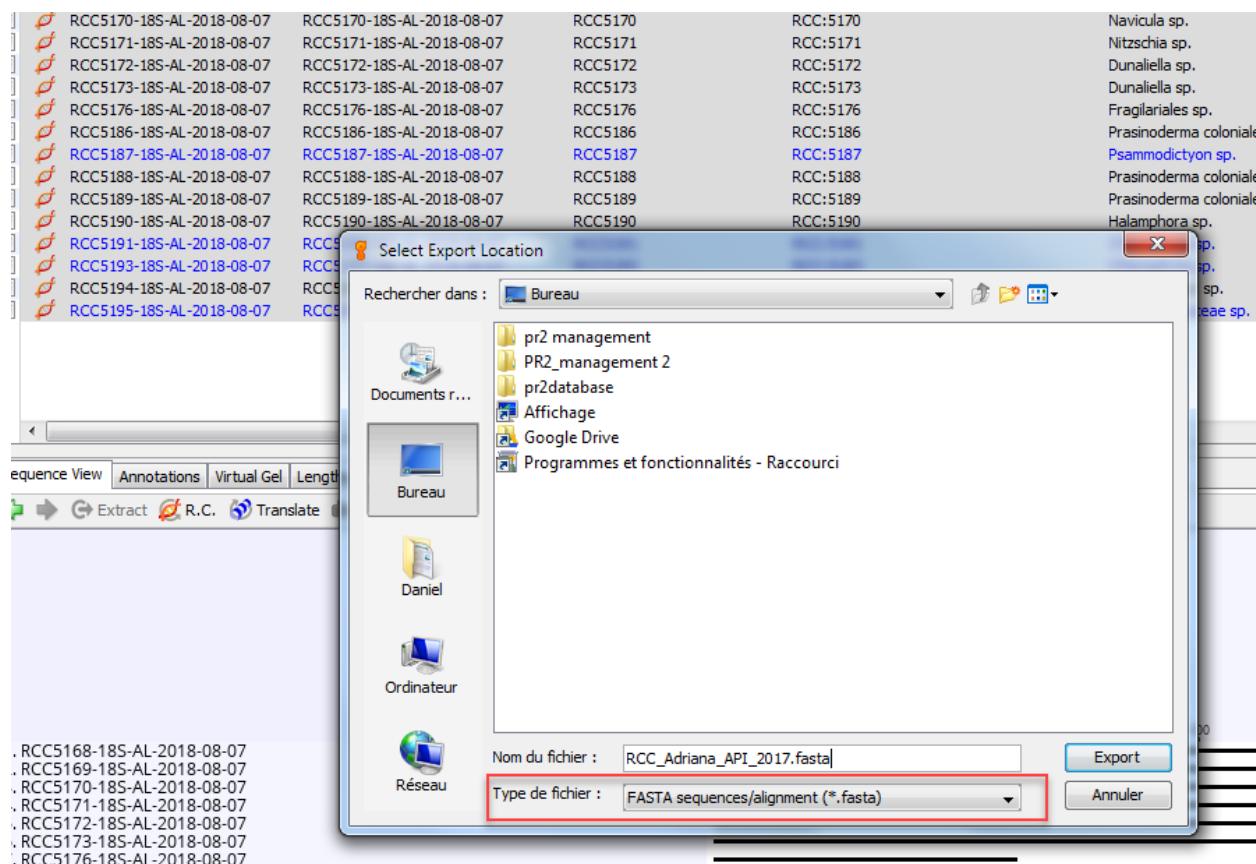


Figure 26: Export to Fasta

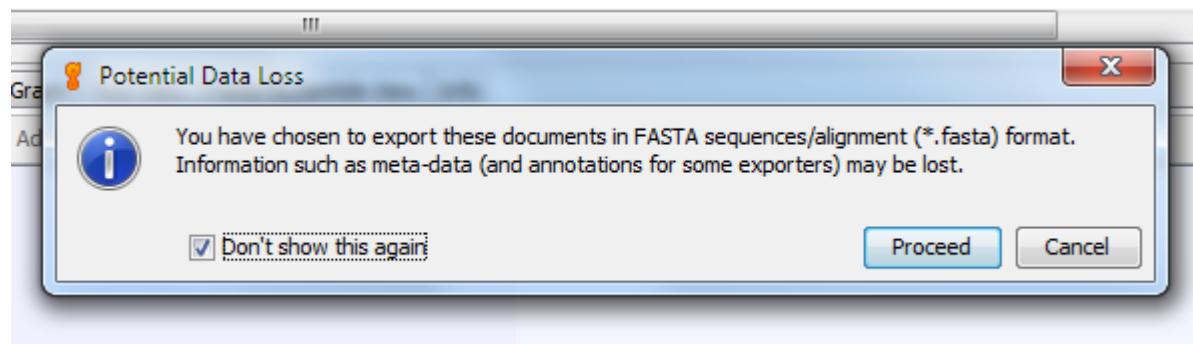


Figure 27: Ignore this warning

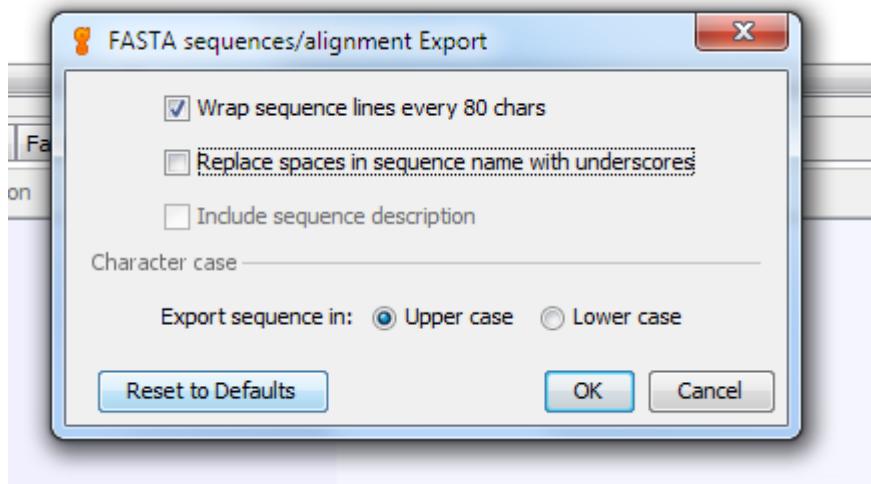


Figure 28: Wrap sequences to 80 characters

```

1 >RCC5168-18S-AL-2018-08-07
2 CTCATTATATCAGTTATAGTTCTTGATAGTCCCTCACTACTTGATAACCGTAGTAATTCTAGAGCTAATACATGCGT
3 CGATACCCCTTGGGTAGTATTATTAGATGGAAACCAACCCCTCGGGGTGATGTGGTAATCATAATAAGCTTGCAGGA
4 TCGCCGGTGGCGATGGATCATCAAGTTCTGCCCTATCAGTTGGACGTTGGTATTGGCAAACGTGGCTTAACG
5 GGTAAACGGGAATTAGGGTCTGATTCCGGAGAGGGAGGCTGAGAGACGGCTACACATCCAAGGAAGGCAGCAGGGCGGT
6 AAATTACCCAATCTTGACACAAGGAGGTAGTGCACAAATAAAATAACAAIGCCGGCCTTGTAGGTCTGGCAATTGGAATGA
7 GAACAATTAAACCCCTATCAGGGATCAATTGGAGGGCAAGTCTGGTGCAGCAGCCGCGTAATTCCAGCTCCAATAG
8 CGTATATTAAAGTGTGCAAGTTAAAAGCTCGTAGTTGGACTTGTGGTGGTCTGAGGTCTGTTGGTACTTTGG
9 GGACTGCCATCCTGGGTGGATCCTGTGTCGATTAGTTGTGTCGTCAGGGATGCCATCGTTACTGTGAAAAAATTA
10 GAGTGTCAAAGCAGGCTTATGCCGTGAATATATTAGCATGAAATAATGAGATAGGACCTGGTACTATTTGGTT
11 TGCACCGAGGTAATGATTAATAGGGACAGTTGTGGTATTGTAITCCATTGTCAGAGGTGAAATTCTGGATTCTG
12 GAAGACGAACGAATGCGAAAGCATTACCAAGGGATTTCTTACAAAGGGATTGCGGGGTTCTGTTACGTCCTCGTCAGCACCTTATG
13 GATACCACCGTAGTCTAACCATAAAATATGCCGACAAGGGATTGGCGGGGTTCTGTTACGTCCTCGTCAGCACCTTATG
14 AGAAATCACAAGTTGGTCCGGGGGAGTATGGTCGCAAGGCTGAAACTAAAGAAATTGACGGAAGGGCACCACC
15 AGGAGTGGAGCCTGGGCTTAATTGACTCAACACGGAAAACCTACCGGTCAGACATAGTGAGGATTGACAGATTG
16 GAGCTCTTCTTGATTCTATGGTGGTGCATGGCGTTCTAGTTGGGGAGTGATTGCTGGTTAATTCCGTTAA
17 CGAACGAGACCCCTGCGCTGCTAAATAGTCTTGTGAGTATTGGGCTTCTAGAGGGACGTGCATTCTAT
18 TAGATGCAGGAAGATAAGGGCAATAACAGGTCTGTGATGCCCTAGATGTTCTGGGCCGACGCCGCTACACTGATGCA
19 TTCAACGAGTTCTTGGCCGAGAGGCCTGGCAATCTTGGAACGTGCAICGTGATAGGGATAGATTATTGCAATTATT
20 AATCTGAACGAGGAATTCTAGTAAACGCAAGATCATCAATCTGATTACGTCCTGCCCTTGTACACACCGCCC
21 GTCGCACCTACCGATTGGATGGTCCGGTGAAGCCTCGGGATTGTGACCGGAGCCTTACGGG
22 >RCC5169-18S-AL-2018-08-07
23 ACTGCGAACGGCTCATTATATCAGTTATAGTTATTGATAGTCCCTACTATTGGATAACCGTAGTAATTCTAGAGCT
24 AATACATGCGTCAATACCCCTGGGTAGTATTATTAGATAGAAACCAACCCCTCGGGGTGATGTGGTAATCATAA
25 TAAGCTTGGGATCGATGGCTCGGCGACGGATCATTCAAGTTCTGCCCTATCAGTTGGATGGTAGGGTATT
26 GGCCTACCATGGCTTAACGGTAACGGGAAATTAGGGTTGATTCCGGAGAGGGGCTGAGAGATGGCACCCACATCC
27 AAGGAAGGCAGCAGGCGCTAAATTACCAATCCTGACACAGGGAGGTAGTGCACAAATAAAACAATGCCGGCTTAT
28 AGGTCTGGCAATTGGATGAGAACAAATTAAATCCCTATCAGGGATCAATTGGAGGGCAAGTCTGGTGCAGCCGC
29 GGTAAATTCCAGCTCAAAGCGTATATTAAAGTTGTGCAAGTAAAAAGCTCGTAGTTGGATTGTGGTTACGGCGTGT
30 ACCAGGCACTTGTGCTGAGTTCATGCCGTGCCATCCTGGTGGAACCTGTGTCGATTAGGTGTCGTCAGGGG
31 ATGCCCATCGTTACTGTGAAAAAAATTAGAGTGTCAAAGCAGGCTTATGCCGTGAATATATTAGCATGGAATAATAAG
32 ATAGGACTTTCTGCTATTGTTGGTTGCGCGAAGAGGTAAATGATTAATAGGGACAGTTGGGGTATTGTATTCCAT
33 TGTCAAGGGTGAATTCTGGATTGGAAAGACGAACACTGCGAACAGCATTACCAAGGATTTTCAATTCAAGA
34 ACGAAAGTTAGGGGATCGAAGATGATTAGATACCATCGTAGTCTAACCATAAACTATGCCGACAAGGGATTGGTGGGG
35 CTCGTTACGTCTCCATCAGCACCTTATGAGAAATCACAAGTCTGGGTTCCGGGGGAGTATGGTCGAAGGCTGAAAC
36 TTAAAGAAATTGACGGAAGGGCACCACCAAGGAGTGGAGCCTGCGGCTTAATTGACTCAACACGGGAAACTTACCGAGT
37 CCAGACATAGTGAGGATTGACAGATTGAGAGCTTCTGATTCTATGGTGGTGTGCACTGGCGTTCTAGTTGGT
38 GAGTGTGATTGCTGGTAACTCGTTAACGAACGAGACCAACTGCCGTCTAAATAGCCAGTGTGAGTGAATCTCACTGACT
39 GCTGGCTTCTAGAGGGACGTGCATTCTATCAGATGCAGGAGGTAGTGGCAATAACAGGTCTGTGATGCCCTTAGATGT
40 CCTGGCCGACCGCGCTACACTGATGCATTCAACGAGTTTACCTGGCGAGAGGCCCTGGCAATCTTGAACGTG
41 CATCGTGTAGGGATAGATTATTGCAATTATTAAATCTGAAACGAGGAATTCTAGTAAACGCAATCATCAATTGCAATT
42 GATTACGTCCTGCCCTTGTACACACCAGGCGTCGACCTACCGATTGAAATGGTCCGGTGAAGCCTCGGGATTGTGATC
43 TGTTCTTATTGGAGATG
44 >RCC5170-18S-AL-2018-08-07
45 CGGCTCATTATATCAGTTATAGTTATTGATAGTCCCTACTATTGGATAACCGTAGTAATTCTAGAGCTAATACATG
46 CGTCAATACCCCTGGGTAGTATTATTAGATAGAAACCAACCCCTCGGGGTGATGTGGTAATCATAATAAGCTTGC
47 CGGATCGCATGGCTCGGCCGCGACGGATCATTCAAGTTCTGCCCTATCAGTTGGATGGTAGGGTATTGGCCTTAC
48 ATGGCTTAACGGTAACGGGAAATTAGGGTTGATTCCGGAGAGGGGCTGAGAGATGGCACCACATCCAAGGAAGG
49 CAGCAGGCGCTAAATTACCAATCCTGACACAGGGAGGTAGTGCACAAATAAAACAATGCCGGCCTTATAGGTCTGG
50 CAATTGGAATGAGAACAAATTAAATCCCTATCAGGGATCAATTGGAGGGCAAGTCTGGTGCAGCAGCCGCGTAATTC
51 CAGCTCCAATAGCGTATATTAAAGTTGTGCAAGTAAAAGCTCGTAGTTGGATTGTGGTTACGGCGTGTACCGAGGCA

```

Figure 29: Final fasta file

- Source information file tab-delimited (tsv file)

Name	Sequence_ID (Genbank Submission)	Strain (Genbank Submission)	Culture_collection (Genbank Submission)	Organism
RCC5168-18S-AL-2018-08-07	RCC5168-18S-AL-2018-08-07	RCC5168	RCC:5168	Halimphora sp.
RCC5169-18S-AL-2018-08-07	RCC5169-18S-AL-2018-08-07	RCC5169	RCC:5169	Navicula sp.
RCC5170-18S-AL-2018-08-07	RCC5170-18S-AL-2018-08-07	RCC5170	RCC:5170	Navicula sp.
RCC5171-18S-AL-2018-08-07	RCC5171-18S-AL-2018-08-07	RCC5171	RCC:5171	Nitzschia sp.
RCC5172-18S-AL-2018-08-07	RCC5172-18S-AL-2018-08-07	RCC5172	RCC:5172	Dunaliella sp.
RCC5173-18S-AL-2018-08-07	RCC5173-18S-AL-2018-08-07	RCC5173	RCC:5173	Dunaliella sp.
RCC5176-18S-AL-2018-08-07	RCC5176-18S-AL-2018-08-07	RCC5176	RCC:5176	Fragilariales sp.
RCC5186-18S-AL-2018-08-07	RCC5186-18S-AL-2018-08-07	RCC5186	RCC:5186	Prasinoderma coloniale
<b>RCC5187-18S-AL-2018-08-07</b>	<b>RCC5187-18S-AL-2018-08-07</b>	<b>RCC5187</b>	<b>RCC:5187</b>	<b>Psammodictyon sp.</b>
RCC5188-18S-AL-2018-08-07	RCC5188-18S-AL-2018-08-07	RCC5188	RCC:5188	Prasinoderma coloniale
RCC5189-18S-AL-2018-08-07	RCC5189-18S-AL-2018-08-07	RCC5189	RCC:5189	Prasinoderma coloniale
RCC5190-18S-AL-2018-08-07	RCC5190-18S-AL-2018-08-07	RCC5190	RCC:5190	Halimphora sp.
<b>RCC5191-18S-AL-2018-08-07</b>	<b>RCC5191-18S-AL-2018-08-07</b>	<b>RCC5191</b>	<b>RCC:5191</b>	<b>Chlorophyta sp.</b>
<b>RCC5193-18S-AL-2018-08-07</b>	<b>RCC5193-18S-AL-2018-08-07</b>	<b>RCC5193</b>	<b>RCC:5193</b>	<b>Chlorophyta sp.</b>
RCC5194-18S-AL-2018-08-07	RCC5194-18S-AL-2018-08-07	RCC5194	RCC:5194	Prorocentrum sp.
<b>RCC5195-18S-AL-2018-08-07</b>	<b>RCC5195-18S-AL-2018-08-07</b>	<b>RCC5195</b>	<b>RCC:5195</b>	<b>Heterocapsaceae sp.</b>

Figure 30: Select sequences. Check the 4 fields (Sequence\_ID, Strain, Culture\_collection and Organism) are correct.

Name	Sequence_ID (Genbank Submission)	Strain (Genbank Submission)	Culture_collection (Genbank Submission)	Organism
RCC5168-18S-AL-2018-08-07	RCC5168-18S-AL-2018-08-07	RCC5168	RCC:5168	Halimphora sp.
RCC5169-18S-AL-2018-08-07	RCC5169-18S-AL-2018-08-07	RCC5169	RCC:5169	Navicula sp.
RCC5170-18S-AL-2018-08-07	RCC5170-18S-AL-2018-08-07	RCC5170	RCC:5170	Navicula sp.
RCC5171-18S-AL-2018-08-07	RCC5171-18S-AL-2018-08-07	RCC5171	RCC:5171	Nitzschia sp.
RCC5172-18S-AL-2018-08-07	RCC5172-18S-AL-2018-08-07	RCC5172	RCC:5172	Dunaliella sp.
RCC5173-18S-AL-2018-08-07	RCC5173-18S-AL-2018-08-07	RCC5173	RCC:5173	Dunaliella sp.
RCC5176-18S-AL-2018-08-07	RCC5176-18S-AL-2018-08-07	RCC5176	RCC:5176	Fragilariales sp.
RCC5186-18S-AL-2018-08-07	RCC5186-18S-AL-2018-08-07	RCC5186	RCC:5186	Prasinoderma coloniale
<b>RCC5187-18S-AL-2018-08-07</b>	<b>RCC5187-18S-AL-2018-08-07</b>	<b>RCC5187</b>	<b>RCC:5187</b>	<b>Psammodictyon sp.</b>
RCC5188-18S-AL-2018-08-07	RCC5188-18S-AL-2018-08-07	RCC5188	RCC:5188	Prasinoderma coloniale
RCC5189-18S-AL-2018-08-07	RCC5189-18S-AL-2018-08-07	RCC5189	RCC:5189	Prasinoderma coloniale
RCC5190-18S-AL-2018-08-07	RCC5190-18S-AL-2018-08-07	RCC5190	RCC:5190	Halimphora sp.
<b>RCC5191-18S-AL-2018-08-07</b>	<b>RCC5191-18S-AL-2018-08-07</b>	<b>RCC5191</b>	<b>RCC:5191</b>	<b>Chlorophyta sp.</b>
<b>RCC5193-18S-AL-2018-08-07</b>	<b>RCC5193-18S-AL-2018-08-07</b>	<b>RCC5193</b>	<b>RCC:5193</b>	<b>Chlorophyta sp.</b>
RCC5194-18S-AL-2018-08-07	RCC5194-18S-AL-2018-08-07	RCC5194	RCC:5194	Prorocentrum sp.
<b>RCC5195-18S-AL-2018-08-07</b>	<b>RCC5195-18S-AL-2018-08-07</b>	<b>RCC5195</b>	<b>RCC:5195</b>	<b>Heterocapsaceae sp.</b>

Figure 31: Export selected documents

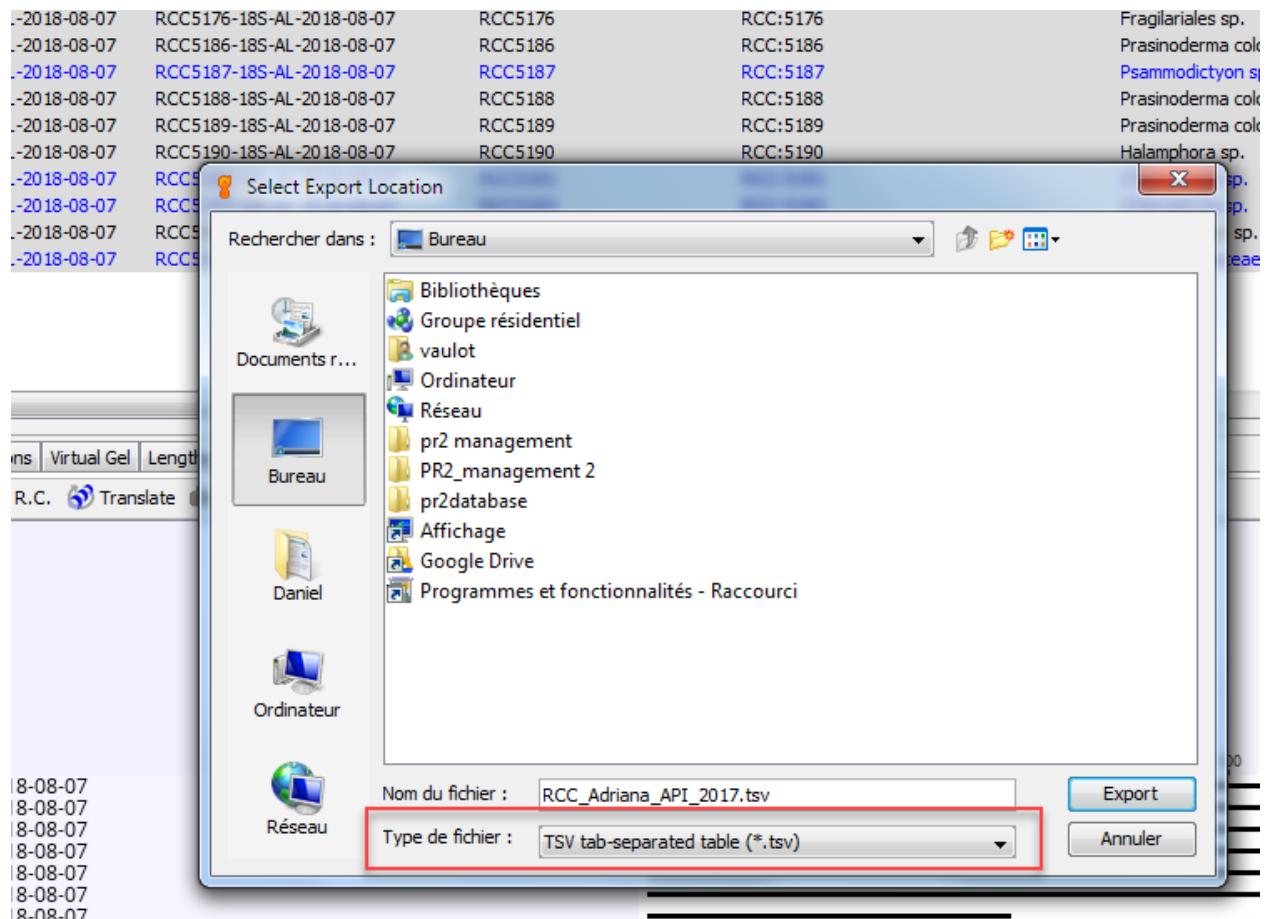


Figure 32: Export as tsv

Name	Sequence_ID (Genbank Submission)	Strain (Genbank Submission)	Culture_collection (Genbank Submission)
RCC5168-18S-AL-2018-08-07	RCC5168-18S-AL-2018-08-07	RCC5168	RCC:5168
RCC5169-18S-AL-2018-08-07	RCC5169-18S-AL-2018-08-07	RCC5169	RCC:5169
RCC5170-18S-AL-2018-08-07	RCC5170-18S-AL-2018-08-07	RCC5170	RCC:5170
RCC5171-18S-AL-2018-08-07	RCC5171-18S-AL-2018-08-07	RCC5171	RCC:5171
RCC5172-18S-AL-2018-08-07	RCC5172-18S-AL-2018-08-07	RCC5172	RCC:5172
RCC5173-18S-AL-2018-08-07	RCC5173-18S-AL-2018-08-07	RCC5173	RCC:5173
RCC5176-18S-AL-2018-08-07	RCC5176-18S-AL-2018-08-07	RCC5176	RCC:5176
RCC5186-18S-AL-2018-08-07	RCC5186-18S-AL-2018-08-07	RCC5186	RCC:5186
RCC5187-18S-AL-2018-08-07	RCC5187-18S-AL-2018-08-07	RCC5187	RCC:5187
RCC5188-18S-AL-2018-08-07	RCC5188-18S-AL-2018-08-07	RCC5188	RCC:5188
RCC5189-18S-AL-2018-08-07	RCC5189-18S-AL-2018-08-07		
RCC5190-18S-AL-2018-08-07	RCC5190-18S-AL-2018-08-07		
RCC5191-18S-AL-2018-08-07	RCC5191-18S-AL-2018-08-07		
RCC5193-18S-AL-2018-08-07	RCC5193-18S-AL-2018-08-07		
RCC5194-18S-AL-2018-08-07	RCC5194-18S-AL-2018-08-07		
RCC5195-18S-AL-2018-08-07	RCC5195-18S-AL-2018-08-07		

Sequence View Annotations Virtual Gel Lengths Graph Text View Fasta Nuc

Extract R.C. Translate Add/Edit Annotation

RCC5168-18S-AL-2018-08-07  
RCC5169-18S-AL-2018-08-07  
RCC5170-18S-AL-2018-08-07  
RCC5171-18S-AL-2018-08-07  
RCC5172-18S-AL-2018-08-07  
RCC5173-18S-AL-2018-08-07  
RCC5176-18S-AL-2018-08-07  
RCC5186-18S-AL-2018-08-07  
RCC5187-18S-AL-2018-08-07

TSV tab-separated table Export X

Select fields to export:

- # Source Sequences
- % Identical Sites
- % Pairwise Identity
- %GC
- Ambiguities
- Approx. Consensus Length
- Bin
- Color
- Created Date
- Description
- Free end gaps
- Genbank Submission: Culture\_collection
- Genbank Submission: Sequence ID
- Genbank Submission: Sequence\_ID
- Genbank Submission: Strain
- HQ%
- Imported From: Filename
- Imported From: Path
- LQ%
- MQ%
- Mean Coverage
- Modified
- Molecule Type
- Organism
- Post-Trim
- Sample
- Sequence
- Sequence Length

Reset to Defaults OK Cancel

Figure 33: Select the columns to be exported (be careful to export Sequence\_ID and not Sequence ID without underscore)

Genbank Submission:	Culture_collection	Genbank Submission:	Sequence_ID	Genbank Submission:	Strain	Organism
RCC:5168	RCC5168-18S-AL-2018-08-07	RCC5168	<i>Halimphora</i> sp.			
RCC:5169	RCC5169-18S-AL-2018-08-07	RCC5169	<i>Navicula</i> sp.			
RCC:5170	RCC5170-18S-AL-2018-08-07	RCC5170	<i>Navicula</i> sp.			
RCC:5171	RCC5171-18S-AL-2018-08-07	RCC5171	<i>Nitzschia</i> sp.			
RCC:5172	RCC5172-18S-AL-2018-08-07	RCC5172	<i>Dunaliella</i> sp.			
RCC:5173	RCC5173-18S-AL-2018-08-07	RCC5173	<i>Dunaliella</i> sp.			
RCC:5176	RCC5176-18S-AL-2018-08-07	RCC5176	<i>Fragilariales</i> sp.			
RCC:5186	RCC5186-18S-AL-2018-08-07	RCC5186	<i>Prasinoderma coloniale</i>			
RCC:5187	RCC5187-18S-AL-2018-08-07	RCC5187	<i>Psammodictyon</i> sp.			
RCC:5188	RCC5188-18S-AL-2018-08-07	RCC5188	<i>Prasinoderma coloniale</i>			
RCC:5189	RCC5189-18S-AL-2018-08-07	RCC5189	<i>Prasinoderma coloniale</i>			
RCC:5190	RCC5190-18S-AL-2018-08-07	RCC5190	<i>Halimphora</i> sp.			
RCC:5191	RCC5191-18S-AL-2018-08-07	RCC5191	<i>Chlorophyta</i> sp.			
RCC:5193	RCC5193-18S-AL-2018-08-07	RCC5193	<i>Chlorophyta</i> sp.			
RCC:5194	RCC5194-18S-AL-2018-08-07	RCC5194	<i>Prorocentrum</i> sp.			
RCC:5195	RCC5195-18S-AL-2018-08-07	RCC5195	<i>Heterocapsaceae</i> sp.			

Figure 34: Edit the tsv fil to remove ‘GenBank Submission:‘ in the titles of the columns. This is best done with an editor such as Notepad++ or with Excel. In the latter case the file must be saved as a text tabulated file.

Note: if include the primers, make sure that you rename the column, not forgetting the underscores: \* Fwd\_primer\_name - name of forward PCR primer \* Fwd\_primer\_seq - nucleotide sequence of forward PCR primer \* Rev\_primer\_name - name of reverse PCR primer \* Rev\_primer\_seq - nucleotide sequence of reverse PCR primer

Example of header : Sequence\_ID Culture Collection Strain Organism Fwd\_primer\_name  
Fwd\_primer\_seq Rev\_primer\_name Rev\_primer\_seq

Culture collection → Sequence ID → Strain → Organism	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5168 → RCC5168-18S-AL-2018-08-07 → RCC5168 > <i>Halimphora</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5169 → RCC5169-18S-AL-2018-08-07 → RCC5169 > <i>Navicula</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5170 → RCC5170-18S-AL-2018-08-07 → RCC5170 > <i>Navicula</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5171 → RCC5171-18S-AL-2018-08-07 → RCC5171 > <i>Nitzschia</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5172 → RCC5172-18S-AL-2018-08-07 → RCC5172 > <i>Dunaliella</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5173 → RCC5173-18S-AL-2018-08-07 → RCC5173 > <i>Dunaliella</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5176 → RCC5176-18S-AL-2018-08-07 → RCC5176 > <i>Fragilariales</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5186 → RCC5186-18S-AL-2018-08-07 → RCC5186 > <i>Prasinoderma coloniale</i> CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5187 → RCC5187-18S-AL-2018-08-07 → RCC5187 > <i>Psammodictyon</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5188 → RCC5188-18S-AL-2018-08-07 → RCC5188 > <i>Prasinoderma coloniale</i> CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5189 → RCC5189-18S-AL-2018-08-07 → RCC5189 > <i>Prasinoderma coloniale</i> CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5190 → RCC5190-18S-AL-2018-08-07 → RCC5190 > <i>Halimphora</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5191 → RCC5191-18S-AL-2018-08-07 → RCC5191 > <i>Chlorophyta</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5193 → RCC5193-18S-AL-2018-08-07 → RCC5193 > <i>Chlorophyta</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5194 → RCC5194-18S-AL-2018-08-07 → RCC5194 > <i>Prorocentrum</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq
RCC:5195 → RCC5195-18S-AL-2018-08-07 → RCC5195 > <i>Heterocapsaceae</i> sp. CRLF	→ Fwd_primer_name → Fwd_primer_seq → Rev_primer_name → Rev_primer_seq

Figure 35: After editing and removing GenBank Submission:

The screenshot shows a Microsoft Excel spreadsheet with the ribbon menu at the top. The active tab is "Accueil". The toolbar includes standard functions like "Coller" (Paste), "Police" (Font), and "Alignement" (Text Alignment). The formula bar shows the cell reference "A4" and the value "RCC:5170". The main area displays a table with the following data:

	A	B	C	D	E
1	Culture_collection	Sequence_ID	Strain	Organism	
2	RCC:5168	RCC5168-18S-AL-2018-08-07	RCC5168	Halaphora sp.	
3	RCC:5169	RCC5169-18S-AL-2018-08-07	RCC5169	Navicula sp.	
4	RCC:5170	RCC5170-18S-AL-2018-08-07	RCC5170	Navicula sp.	
5	RCC:5171	RCC5171-18S-AL-2018-08-07	RCC5171	Nitzschia sp.	
6	RCC:5172	RCC5172-18S-AL-2018-08-07	RCC5172	Dunaliella sp.	
7	RCC:5173	RCC5173-18S-AL-2018-08-07	RCC5173	Dunaliella sp.	
8	RCC:5176	RCC5176-18S-AL-2018-08-07	RCC5176	Fragilariales sp.	
9	RCC:5186	RCC5186-18S-AL-2018-08-07	RCC5186	Prasinoderma coloniale	
10	RCC:5187	RCC5187-18S-AL-2018-08-07	RCC5187	Psammodictyon sp.	
11	RCC:5188	RCC5188-18S-AL-2018-08-07	RCC5188	Prasinoderma coloniale	
12	RCC:5189	RCC5189-18S-AL-2018-08-07	RCC5189	Prasinoderma coloniale	
13	RCC:5190	RCC5190-18S-AL-2018-08-07	RCC5190	Halaphora sp.	
14	RCC:5191	RCC5191-18S-AL-2018-08-07	RCC5191	Chlorophyta sp.	
15	RCC:5193	RCC5193-18S-AL-2018-08-07	RCC5193	Chlorophyta sp.	
16	RCC:5194	RCC5194-18S-AL-2018-08-07	RCC5194	Prorocentrum sp.	
17	RCC:5195	RCC5195-18S-AL-2018-08-07	RCC5195	Heterocapsaceae sp.	
18					
19					

Figure 36: Editing with Excel (save file as tab-delimited tsv)

## 5.2 Submit to NCBI web portal

- Go to web portal : <https://submit.ncbi.nlm.nih.gov/subs/genbank/>

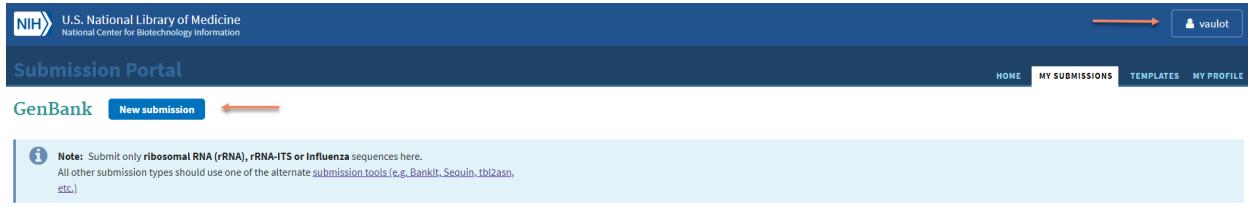


Figure 37: Web portal. Register or login if you have already an ID

# GenBank submission: SUB4388090

New

1 SUBMISSION TYPE    2 SUBMITTER    3 SEQUENCING TECHNOLOGY    4 SEQUENCING

## Submission Type

### Submission type

\* What do your sequences contain?

- Prokaryotic rRNA/IGS
- Eukaryotic Nuclear rRNA/ITS
- Eukaryotic Organelle rRNA
- Influenza virus



\* What do these Prokaryotic rRNA/IGS sequences contain?

- small subunit rRNA only (16S rRNA)
- large subunit rRNA only (23S rRNA)
- intergenic spacer (16S-23S rRNA IGS)



**i** If none of the options above describe your sequence

Optional. You can provide a title for your submission in the Submission Portal. This title will not be displayed in the final records. Limit your title to 60 characters.

### Submission title (Optional, not displayed in final records) [?](#)

RCC Adriana 2018 Prokaryotes



**Continue**

Figure 38: Enter the type of sequence



## Submitter

### Affiliation

**i** The information you give here will be displayed in the final sequence records.  
For address details, provide the primary address where work was done to generate the data in this submission.

**\* Submitting organization**

Sorbonne Université, CNRS, Station Bi

**\* Department**

UMR7144

**\* Street**

Place Georges Teissier

**\* City**

Roscoff

**State/Province**

**\* Postal code**

29680

**\* Country**

France



### Contact information

**i** GenBank may use this information to contact you about your submission, it will not be displayed in the final sequence records.

**\* E-mail (primary)**

vaulot@sb-roscff.fr

**\* E-mail (secondary)**

vaulot@gmail.com

**i** Please provide an alternate email address to ensure that messages are received

**\* First (given) name**

Daniel

**Middle name**

**\* Last (family) name**

Vaulot

**Phone**

**Fax**

**Continue**

Update my contact information in profile

Figure 39: Enter submitter information

# Submission Portal

## GenBank submission: SUB4388090

Prokaryotic 16S rRNA / RCC Adriana 2018 Prokaryotes

1 SUBMISSION TYPE    2 SUBMITTER    3 SEQUENCING TECHNOLOGY    4 SEQUENCES    5 SEQUENCE PROCE

### Sequencing Technology

#### Method

\* What methods were used to obtain these sequences? 

- Sanger dideoxy sequencing
- 454
- Helicos
- Illumina
- IonTorrent
- Pacific Biosciences
- SOLiD
- Other



#### Assembly State

These sequences are:

- Unassembled sequence reads
- Assembled sequences (each sequence was assembled from two or more overlapping sequence reads)



**Continue**

Figure 40: Enter the sequence technology. In almost all cases choose Sanger and Assembly

## Sequences

**Release date**

**Note:** Release of BioProject or BioSample is also triggered by the release of linked data.

\* When should this submission be released to the public?

Release immediately following processing

Release on specified date or upon publication, whichever is first

**Chimera check**

\* Did you check and remove low-quality and chimeric sequences from your FASTA file prior to preparing this submission?

Yes

No

**Only for prokaryotes**

**Note:** Please provide the name and version of the chimera checking program. BLAST alone is not sufficient as a chimera checking program.

Program Name      Version

Geneious      10

**Cultured or Uncultured**

Select whether your sequences were obtained from cultured or uncultured samples.

**Only for prokaryotes**

\* Bacterial/archaeal Sequences: How were they obtained? [?](#)

Pure-cultured strains (axenic cultures containing only one microbial species each)

Uncultured, bulk environmental DNA (PCR-amplified directly from environmental sample or host; samples were not grown in culture)

**Sequences**

\* Upload a nucleotide [FASTA](#) formatted file.

Choisir un fichier Aucun fichier choisi

**Note:** If you have multiple sequences, all of your sequences need to be in one file. [Help on FASTA file](#).

Example FASTA nucleotide format:

```
>Seq1
aaccgatatacgatgtgatccgatatacgagagagga
>Seq2
gtacgataaaagatgtgatccgatatacgagagagga
```

**Note:** Use the latest version of the [Aspera Connect plugin](#) for faster file uploads. If a pop-up box about 'fasp protocol' is displayed, click 'Allow' or 'Open' to let [Aspera Connect](#) handle file uploads more efficiently.

Figure 41: Sequences. Release date: Choose immediate release in most cases, there is really no need to delay release.- The chimera question is only for Prokaryotes. - Chose pure cultures for cyanos. - Upload the fasta sequence file

- Source information. Since it will be loaded in the text file, choose “NONE of these”

## GenBank submission: SUB4388090

cultured Prokaryotic 16S rRNA / RCC Adriana 2018 Prokaryotes

1 SUBMISSION TYPE 2 SUBMITTER 3 SEQUENCING TECHNOLOGY 4 SEQUENCES 5 SOURCE INFO 6 SOURCE MODIFIERS 7 REFERENCES 8 OVERVIEW

### Source Information

The first few sequence IDs that we found are:

RCC5180-18S-AL-2018-08-07  
RCC5183-18S-AL-2018-08-07  
RCC5185-18S-AL-2018-08-07

\* Do your sequence IDs represent one of these?

Strain

NONE of these

Values for these are typically alpha-numeric sample codes used in your laboratory to track individual samples. Select 'NONE of these' if it does not describe your sequence IDs or the sequence IDs contain more information than the [description/scope](#) of that field.

**Continue**

Figure 42: Source info

## GenBank submission: SUB4388090

cultured Prokaryotic 16S rRNA / RCC Adriana 2018 Prokaryotes

1 SUBMISSION TYPE 2 SUBMITTER 3 SEQUENCING TECHNOLOGY 4 SEQUENCES 5 SOURCE INFO 6 SOURCE MODIFIERS 7 REFERENCES 8 OVERVIEW

### Source Modifiers

For each sequence, GenBank requires the following source information:

- scientific name of the **Organism** that was sequenced and
- strain** (NOT from a top BLAST hit, NOT the species name).

If you have already provided all the required information, you can press Continue to proceed.

More help: [how to provide source modifiers](#), [description of each modifier](#), [what is a source modifier?](#)

\* How do you want to apply source modifiers?

Use an editable table

Upload a tab-delimited file (use our provided template)

**Continue**

Figure 43: Upload tsv file saved from Geneious

## GenBank submission: SUB4388090

cultured Prokaryotic 16S rRNA / RCC Adriana 2018 Prokaryotes

1 SUBMISSION TYPE > 2 SUBMITTER > 3 SEQUENCING TECHNOLOGY > 4 SEQUENCES > 5 SOURCE INFO > 6 SOURCE MODIFIERS > 7 REFERENCES > 8 OVERVIEW

### Source Modifiers

▼ For each sequence, GenBank requires the following source information:

- scientific name of the **Organism** that was sequenced and
- **strain** (NOT from a top BLAST hit, NOT the species name).

If you have already provided all the required information, you can press Continue to proceed.

ⓘ More help: [how to provide source modifiers](#), [description of each modifier](#), [what is a source modifier?](#)

▶ Current Source Modifiers - what you have provided so far

▶ \* How do you want to apply source modifiers?

1. [Download source modifier template](#) with any source information provided so far.

2. Edit the [downloaded table](#) in Microsoft Excel or any other editor.

↳ [See an example Source Modifiers table](#)

3. Save the table as a tab-delimited text file.

4. Upload Source Modifiers file.

RCC\_Adriana\_API\_2017\_Prok.tsv ✓



Choisir un fichier Aucun fichier choisi ⓘ

5. Click Continue to validate the information and follow the instructions.

Continue

Figure 44: After uploading tsv file

## GenBank submission: SUB4388231

Eukaryotic Nuclear rRNA/ITS / RCC Adriana 2018 Eukaryotes



### Source Modifiers

**⚠ Warning:** One or more of the organism names listed below are not in the NCBI Taxonomy database. Confirm that the spelling of the listed organism name(s) is correct. If any are not correct, provide corrected names. If they are correct, click the Continue button.

Organism Name
Trebouxiophyceae sp.
Chlorophyceae sp.
Fragilariales sp.
Chlorophyta sp.
Heterocapsaceae sp.

Figure 45: Taxonomy error - This error is due to the addition of sp. to taxa at the rank above the genus. You need to correct and remove sp. to the tsv file. If the error comes from a new taxon not yet described you can ignore and GenBank will contact you probably to add this taxon to their database.

## GenBank submission: SUB4388090

cultured Prokaryotic 16S rRNA / RCC Adriana 2018 Prokaryotes

1 SUBMISSION TYPE 2 SUBMITTER 3 SEQUENCING TECHNOLOGY 4 SEQUENCES 5 SOURCE INFO 6 SOURCE MODIFIERS 7 REFERENCES 8 OVERVIEW

### References

#### Sequence authors

Who should be publicly credited as the submitter of this sequence data?

\* First (given) name MI  \* Last (family) name   
Adriana  Lopes dos Santo   
Daniel  Vaulot

Delete

Delete

Names will appear in your records as:

Lopes dos Santos,A. and Vaulot,D.



[Add another sequence author](#)

#### Reference

Please provide the title and relevant publication details of **your paper** that discusses **this submission**.

\* Publication status

Unpublished  In-press  Published

Reference title

Roscoff Culture Collection



Select Reference Authors

Same as sequence authors  Specify new authors

[Continue](#)

Figure 46: Add the reference. For the Roscoff Culture Collection just fill as indicated with the name of the person who produced the sequence first.

**GenBank submission: SUB4388090**  
cultured Prokaryotic 16S rRNA / RCC Adriana 2018 Prokaryotes

[Delete submission](#)

**1 SUBMISSION TYPE** **2 SUBMITTER** **3 SEQUENCING TECHNOLOGY** **4 SEQUENCES** **5 SOURCE INFO** **6 SOURCE MODIFIERS** **7 REFERENCES** **8 OVERVIEW**

### Overview

**Submit**

You have requested that your sequence data be released **immediately following processing**.

**Submitter**

Submitter Daniel Vaulot  
[vaulot@sub-moscov.fr](mailto:vaulot@sub-moscov.fr)  
[vaulot@gmail.com](mailto:vaulot@gmail.com)

Institution Sorbonne Université, CNRS, Station Biologique

Department UMR7144

Street Place Georges Teissier

City Roscoff

Postal code 29680

Country France

**Sequence authors**

- Adriana Lopes dos Santos
- Daniel Vaulot

**References**

Publication status unpublished

Reference title Rosecif Culture Collection

Authors same as sequence authors

**Sequencing Technology**

Methods Sanger dideoxy sequencing

Assembly state assembled

Chimera tool used? Geneious 10

**Uploaded files**

Figure 47: Final check

**Submission Portal**

**GenBank** [New submission](#)

**Note:** Submit only ribosomal RNA (rRNA), rRNA-ITS or Influenza sequences here.  
All other submission types should use one of the alternate [submission tools](#) (e.g. BankIt, Sequin, [tbl2asn](#), etc.).

**Attention:** If you have corrections to an existing submission with status:  

- Processed-error: use the FIX button to correct a submission.
- Queued or Processing: email your request with the Submission ID.
- Processed: follow these directions and email your request with the Accession numbers.

**Do not create a new submission to fix or update an existing submission whose status is Queued, Processed-error, Processing, or Processed!**

**Short description and brief instructions**

1 submission		Group	Status	Updated
Submission	Title		<input checked="" type="checkbox"/> Submitted	01/51
SUB4388090	cultured Prokaryotic 16S rRNA / RCC Adriana 2018 Prokaryotes		Awaiting processing.	

**Filter / Search**

From date To date Status Sort by

Data archives [Show](#)

Query  [Search](#) [Clear](#)

Figure 48: Submission status. When you press submit you should arrive at the final screen showing your submission.

## GenBank MH732916-MH732918



De [gb-admin@ncbi.nlm.nih.gov](mailto:gb-admin@ncbi.nlm.nih.gov)  
à [vaulot@gmail.com](mailto:vaulot@gmail.com), [vaulot@sb-roscoff.fr](mailto:vaulot@sb-roscoff.fr)

Dear GenBank Submitter:

Thank you for your direct submission of sequence data to GenBank. We have provided GenBank accession number(s) for your nucleotide sequence(s):

SUB4388090 RCC5180-18S-AL-2018-08-07 MH732916  
SUB4388090 RCC5183-18S-AL-2018-08-07 MH732917  
SUB4388090 RCC5185-18S-AL-2018-08-07 MH732918

GenBank accession numbers should appear in any publication that reports or discusses the data, as it gives the community a unique label with which they may retrieve your data from our online servers.

Based on the data submitted to us, the scheduled release date for your submission is:

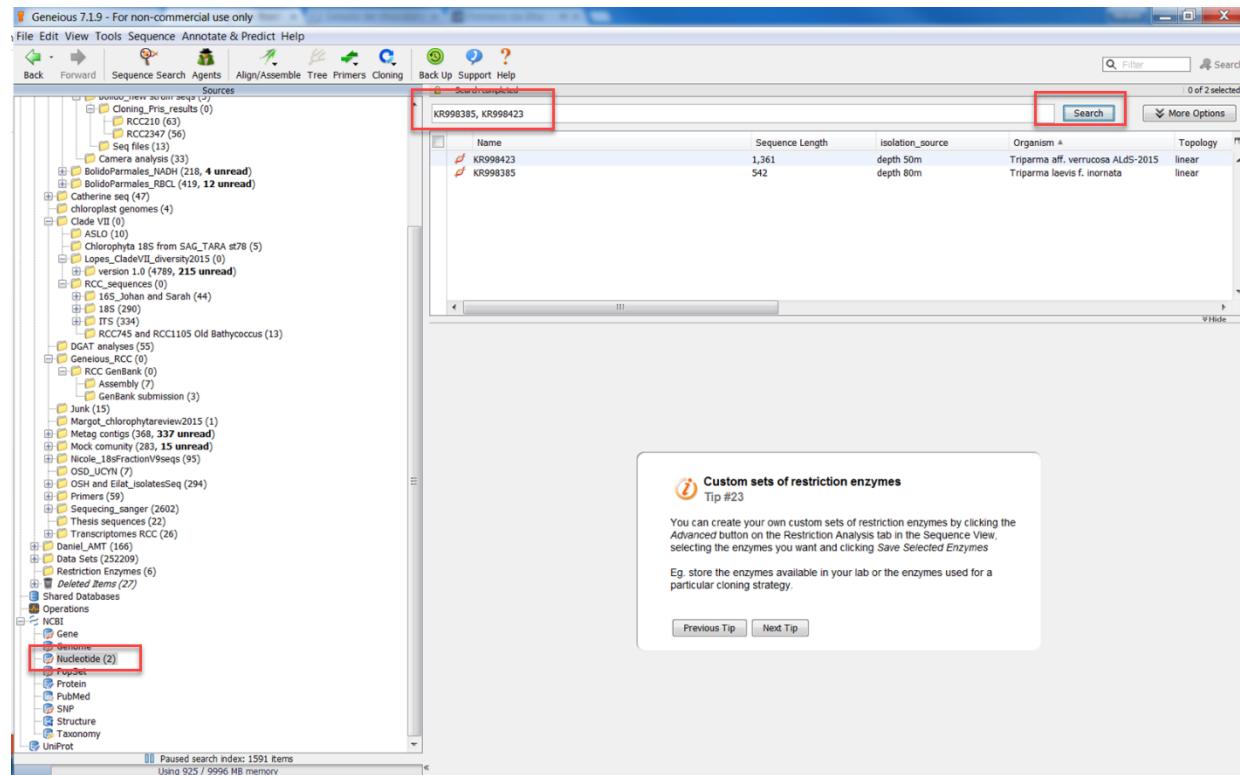
Aug 15, 2018

Figure 49: A few moments later, you should receive an email with the accession numbers. Please forward to [rcc@sb-roscoff.fr](mailto:rcc@sb-roscoff.fr).

## 6 Appendixes

### 6.1 Retrieve sequences from Genbank using Geneious

- **For a list:** Go to nucleotide, type the numbers separate by coma and click search. The results will appear in the bottom panel. You must drag the file into a folder in your local database if you wish to retain the file and/or modify it.



\*

**For consecutive accessions numbers:** type the first and last numbers separated by :, click in more options, change All fields to Accession

The screenshot shows the Genious 7.1.9 software interface. The main window displays a search results table for accessions KR998385:KR998423. The table includes columns for Name, Sequence Length, Isolation Source, Organism, and Topology. The results show various entries from Bolidomonas spp., including KR998394, KR998397, KR998414, KR998389, KR998390, KR998393, KR998398, KR998399, KR998409, KR998416, KR998417, KR998418, KR998391, KR998392, KR998396, KR998400, KR998408, KR998410, KR998411, KR998412, KR998413, KR998421, KR998422, KR998395, KR998411, KR998420, KR998388, KR998403, KR998407, and KR998423. The search bar at the top has 'Accession' selected and 'is' set. A red box highlights the 'Search' button. Below the table, a tip box titled 'Custom sets of restriction enzymes' provides information on creating custom restriction enzyme sets.

Name	Sequence Length	Isolation Source	Organism	Topology
KR998394	531	depth 20m	Bolidomonas mediterranea	linear
KR998397	1,665	depth 20m	Bolidomonas mediterranea	linear
KR998414	1,234	depth 20m	Bolidomonas mediterranea	linear
KR998389	574	depth 15m	Bolidomonas pacifica	linear
KR998390	575	depth 15m	Bolidomonas pacifica	linear
KR998393	601	depth 5m	Bolidomonas pacifica	linear
KR998398	1,698	depth 15m	Bolidomonas pacifica	linear
KR998399	1,686	depth 5m	Bolidomonas pacifica	linear
KR998409	458	depth 15m	Bolidomonas pacifica	linear
KR998416	1,358	depth 15m	Bolidomonas pacifica	linear
KR998417	1,200	depth 15m	Bolidomonas pacifica	linear
KR998418	1,200	depth 5m	Bolidomonas pacifica	linear
KR998391	513	depth 5m	Bolidomonas pacifica var. eleuthera	linear
KR998392	746	depth 15m	Bolidomonas pacifica var. eleuthera	linear
KR998396	536	depth 5m	Bolidomonas pacifica var. eleuthera	linear
KR998400	1,640	depth 5m	Bolidomonas pacifica var. eleuthera	linear
KR998408	584	depth 5m	Bolidomonas pacifica var. eleuthera	linear
KR998410	462	depth 5m	Bolidomonas pacifica var. eleuthera	linear
KR998411	458	depth 5m	Bolidomonas pacifica var. eleuthera	linear
KR998412	607	depth 25m	Bolidomonas pacifica var. eleuthera	linear
KR998413	1,200	depth 5m	Bolidomonas pacifica var. eleuthera	linear
KR998421	1,015	depth 25m	Bolidomonas pacifica var. eleuthera	linear
KR998422	560	English Channel depth 10m	Bolidomonas sp. ALd5-2015	linear
KR998395	692	English Channel depth 10m	Bolidomonas sp. ALd5-2015	linear
KR998411	1,386	English Channel depth 10m	Bolidomonas sp. ALd5-2015	linear
KR998420	546	depth 50m	Triparma aff. verrucosa ALd5-2015	linear
KR998388	1,739	depth 50m	Triparma aff. verrucosa ALd5-2015	linear
KR998403	691	depth 50m	Triparma aff. verrucosa ALd5-2015	linear
KR998407	1,361	depth 50m	Triparma aff. verrucosa ALd5-2015	linear
KR998423	510			

**Custom sets of restriction enzymes**  
Tip #23  
You can create your own custom sets of restriction enzymes by clicking the Advanced button on the Restriction Analysis tab in the Sequence