May 21, 2021

Fungal isolate identification using ITS-LSU regions

Cassie Ettinger[1]

[1]University of California, Riverside

1 Works for me Share dx.doi.org/10.17504/protocols.io.butvnwn6

EisenLab Early Career Ocean Scientists 2 more workspaces

Cassie Ettinger
University of California, Riverside

ABSTRACT

This protocol was written as a companion to the 'Swabs to Genomes' workflow to enable undergraduate interns to take Sanger sequences for the ITS-LSU region from unidentified fungal isolates associated with seagrasses and identify preliminary taxonomy via databases searches and then build phylogenies to further assess putative identities.

ATTACHMENTS

Fungal ITS-LSU Identification.docx

DOI

dx.doi.org/10.17504/protocols.io.butvnwn6

PROTOCOL CITATION

Cassie Ettinger 2021. Fungal isolate identification using ITS-LSU regions. **protocols.io** https://dx.doi.org/10.17504/protocols.io.butvnwn6

KEYWORDS

Fungal isolate identification, ITS-LSU regions, Sanger sequences

LICENSE

CREATED

May 07, 2021

LAST MODIFIED

May 21, 2021

OWNERSHIP HISTORY

May 07, 2021 Urmilas

May 21, 2021 Cassie Ettinger University of California, Riverside

PROTOCOL INTEGER ID

49749

Process Sanger sequences following the ‘Swabs to Genome’ workflow using Seqtrace

1

Follow directions here on how to install and use Seqtrace: **https://peerj.com/articles/960/#sanger-sequence-processing**.

If on undergraduate computer in the Eisen lab, you can type in 'seqtrace' into 'terminal' to start. Otherwise you may need to type 'python seqtrace.py'

2

Create a new project (File > New Project).

2.1 Choose the directory your Sanger sequences are in.

2.2 If your files have the primer names in them, you can enter their names under 'search strings…'.

2.3 Add your forward and reverse primer names (sequences not necessary).

1. Possible forward primers: ITS5, ITS1F.
2. Possible reverse primers: LR3, ITS4.

2.4 Now click on the 'Sequencing Processing' tab.
1. Under "Sequence trimming", unclick the 'Automatically trim sequence ends' button.
2. Under "Consensus" decrease the min. confidence score to 15.

3

To add files, go to "Traces" and click on "Add trace files", then select the reads (.abi files) you want to work with.

4

Group your forward and reverse reads by highlighting both of them and clicking "Group selected forward/reverse files" (under "Traces").

5

Under "Sequences" go to "Generate Finished Sequences" and click on "for all trace files".

6

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

To view your consensus sequence, click on the read pair group and then click on the magnifying glass at the top of the page.

> The "Trace View" shows the quality scores, the chromatogram (trace) display, and the raw base calls from both the forward and reverse reads, as well as the consensus sequence. The consensus sequence is the middle list of nucleotides.

7 🖥

To export the consensus from the trace view, go to "Sequence", hover on "Export Sequences", and select "Export Sequences from Selected Trace Files".

> This will create a file containing the consensus sequence.

Identify fungal sequences using existing databases

8 🖥

Go to NCBI's BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

8.1 Click nucleotide blast.

8.2 Upload the fasta file containing your consensus sequences or copy / paste them in the box.

8.3 For each sequence write down the top hit & its % identity (this is how much the top hit matches your sequence). Note that the top match is database dependent, and for our purposes here is just a hypothesis. We will check our sequences against several databases and build a tree to determine how confident we are in this hypothesis.

9 🖥

Go to Ribosomal Database Project (RDP) website (https://rdp.cme.msu.edu/).

9.1 Make your way to the "Classifier" (https://rdp.cme.msu.edu/classifier/classifier.jsp).

9.2 Upload/paste your sequences.

> For "choose a gene", we are selecting what database to use for ITS1-ITS4 this is UNITE; for ITS5-LR3 you will want to do this process three times, once using UNITE and once using "Fungal LSU training set", and one using "WARCUP".

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

9.3 Click on “Show assignment detail for Root”.

1. Write down the assignment for each sequence and the %.

> **Stop here and decide on next steps!** There are now two options to build fungal phylogenies, the first is to try to build a tree using SILVA (described below), but this is not very effective for fungi that SILVA does not have lots of data for (e.g. works well for Penicillium, but not for more novel or less studied taxa). The second way is to build a phylogeny manually - which involves (1) downloading fungal sequences to include in the tree from BLAST and the literature (2) generating an alignment (3) trimming the alignment and checking it manually (4) picking an evolutionary model for tree building and (5) building a phylogeny. Steps (2), (4), and (5) can be performed on CIPRES (https://www.phylo.org/portal2/home.action).

Build a tree for the LSU region (if available)

10

Go to SILVA ACT (https://www.arb-silva.de/aligner/).

11

You will need at **least 4 sequences** to build a tree - so wait until you have 4 sequences or add in sequences you already have to get to 4.

11.1

> We want to add in a bacteria as an outgroup so you actually only need **3 fungal sequences** to build your tree. We can download an outgroup from https://www.arb-silva.de/browser/.

Pick a **named** bacteria, click its name to add it to your cart and then click “Download” to get the sequence as a fasta file.

11.2 Paste at the end of your fasta file or add to box if copy/pasting.

11.3 Alternatively use the bacterial sequence below:
>OUTGROUP Bacteria;Chloroflexi;Chloroflexia;Chloroflexales;Chloroflexaceae;Roseiflexus;Roseiflexus castenholzii DSM 13941
AAGACAGAAGGAGCACACGGAGGAUGCCUCGCCGCUGCGCGCCGAUGAAGGACGCGGCAACGCUGCG
AAAAGCGUGGGGGAGCCGCGCGCAGGCUGUGAUCCCACGAUGUCCGAAUGGGGCAACCCGUCCGUGA
UACGGACACCGCGCAAUGCGGAGGGCACCGGGGGAACUGAAACAUCUUAGUACCCCGAGGAACAGACA
GCAUCCCCGGAGUAGUGGCGAGCGAAACGGGGUGAGCCCAAACCACGACCGUGGCAAGGCGGCAGCC
GUUGCGGUCGUGGGGUUGAAGGGAGCGCUGCGGGGGGACUGCCGACCCCCCCGCGCGCGUCGCACCC
GGGCGACGAAGCCGACUGGAAAGCGGCGCCACAGAGCGUGACAGCCGCGUAGUCGGGCGGGUGCGGC
GGCGCGAGGCGCUUGCCUGAGUAGCGCCGGACACGAGCAACCCGGCGUGAAACUGGGGCGACCACGC
UCCAAGGCUAAAUACGCGCAGCGAGCGAUAGCGAACGAGUACCGUGAGGGAACGGUGAAAAGCACCCC

GGCGAGGGGAGUGAAAGAGAACCUGAAACCGUGUGCUUCCAUGCAGUCGGAGCCCCGUGAGGGGU
GACGGCGUGCCUUUUGGAGUAUGAUCCGGCGAGUUACCCUCGGUGGCCAGGUUAAGGCAGUGACAGC
CGGAGCCGGAGCGAAAGCGAGUCUGAAGAGGGCGCGCGGUCGCCGGGGGUAGACCCGAAACCGCUUG
AGCUACCCAUGGGCAGGGUGAAGCGCGCGUAACAGCGCGUGGAGGCCCGCACCGGUGUGGAUUGCAA
ACCGCUCGGAUGACCUGUGGGUAGGGGUGAAAUGCCAAACGAAAGCGGAGAUAGCUGGUUCUCCCCGA
AAUGCAUUGAGGUGCAGCCUGCGGAAUGCCGCGCGCGGAGGUAGAGCGCUGGUGUGGUGCGGGGGCU
UCACCGCCUACCAAACGACGCCAAACUGCGAAUGCCGCGCGGGGAAGCGCAGAGUGAGACGUGCGGCG
CAAACGUUGUACGUCGAGAGGGAAACAACCCAGACCUGCAGCUAAGGUCCCGAAAUCGACCCUGAGUG
GGAAAGGAUGUGUCGUUGCGGAGACAACCAGGAGGUUGGCUUAGAAGCAGCCACCCUUGAAAGAGUG
CGUAACAGCUCACUGGUCGAGUGACGAUGCGCCGACAAUCCAGCGGGGCAAAGGGUCGUACCGAAGCU
CAGGACCGGAAGGGUGGUAGGGGAGCGUCGUGGACGCGGCGAAGCGGCGGCGGAAGCCGGUCGUGG
AGCGUCGACGAGUGCGAAUGCCGGAACGAGUAACAGCGAAGCGGGGUGCGAACCCCCGCCACCGGAA
GCCUCAGGGUUCCGCCGCAAGGGUGAUCCGCGGCGGGUUAGUCGGGCCUAAGGGGAAGCGCGAGCGA
CACCCGAUGGACAACAGGUAGACAGUCCUGUACCGGUUAUGCGCCGUUUGACGGACGCGGGACGCUGG
GGGGUAGGCGCUGCGGUCAGACGAGGACCGUCCAAGCUGGUAGGGAGACGGCGGAGGGAAGUCCGCG
CCGUCGGCGACCGAGAAGCGAGAGCGAGCUGGCGCGCGAGCGCACGGCAAGAGCGUUGAACCCGCAC
AGCCGAGAAAGGCCUGCUACGGAGGCGCAUGACCGCCCGUACCGCAAACCGACCCAGGUAGGCUGGCU
GAGGAAGCUAAGGUGGACGAGUGAUCCCUGGUUAAGGAACUCGGCAAAUUGACCCCGUACCUUCGGAA
GAAGGGGUGCUCGGUCGGUGAACUCCGGCGCGGAGGGAGCCGAUGCGAGUCGCAGCAAAGCGGCCCA
GGCGACUGGAUACCAGAACCACAGGUCCGUGCUGAAGCCGAAAGGCGACGUAUACGGGCUGAUGCCUG
CCCAGUGCCGGAAGGUUAAGGGGAGGGCUGUAAGGUCUGAACCGAAGCCCCGGUGAACGGCGGCUGU
AACUAUAACAGUCCUAAGGUAGCGAAAUUCCUUGUCGGGUAAGUUCCGACCCGCACGAAAGGCAUCAC
GAUCUGGGCAGUGUCUCGACCAGGGGCUCGGUGAAACUGCGCUGGCCGUUAGGACGCGGCCAACCCG
UAGCAGGACAAAAAGACCCCGUGGAGCUUUACUACAGCUUGCCAUUGUGCGCGCUCCGGGCUUGCGUA
GGAUAGGUGGGAGCCGGAGAACGGACCCUUGCGGGGGUGCGGGAGGCGACGGUGAAAUACCACUCUG
GCACGGCGUGCGCACUCACCCGCGAGAGCGGGGACCGUGGCUGGCGGGUAGUUUGUCUGGGGCGGAC
GCCUCCUAAAAGGUAACGGAGGCGCGCAACGGCGCCCUCAGGCGGGAUGGCAAUCCGCCGGUGAGUG
CAAAGGCAGAAGGGCGCUUGACUGCAAGGCAGACAGGCCGCGCAGAGAUGAAAGUCGGCCUUAGUGA
UCCUACGGUUCCGCGUGGAAGGGCCGUAGCAUAACGGAUAAAAGCUACCCCGGGGAUAACAGGCUGAU
CGUGCCCAAGAGUUCACAUCGACGGCACGGUUUGGCACCUCGAUGUCGGCUCGUCGCAUCCUGGGGCU
GGAGUCGGUCCCAAGGGUUGGGCUGUUCGCCCAUUAAAGCGGCACGCGAGCUGGGUUCAGAACGUCG
UGAGACAGUUCGGUCUCUAUCCGCUACGGGCGGUUAGUGGCGUGCGGGGAGCUGCUCUUAGUACGAG
AGGACCGGAGUGGACGGACCGCUGGUGGAGGAGUUGUGGGGCGGCCUGCAGGCUCCGUAGCCAUGUC
CGGCGGCAAGAACCGCUGAAAGCAUCUAAGUGGGAAAUGCCCCCCAAGAUGACGCCACUCACGACGGA
AGGUCGGUAAGUCCCCCAGAAGACGACUGGGAAGGCGGCGCCAGCUGGAGGCCCCGUGAGGGGUAGA
GGCGAGGCGUGCGCAUGGACGAGGGCUU.

12

Under “Basic alignment parameters”.

12.1 Choose “LSU”.

12.2 Click “removed”.

13

Click next to “Search and Classify”.

13.1 Change “Min. Identity with query” to 0.7.

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

14

Click next to “Compute tree”.

15

Under “Output settings”.

15.1 Change “reject sequences below…” to 50.

16

Under “advanced tree building parameters”.

16.1 Click “include neighbors from…”.

17

Give the job a name and click run, the job will now show up under “Alignment Taskmanager”.

18

When the job is finished, click on it and click “show results”.

18.1 An “Alignment Results Table” will now appear.

18.2 Click on “Export to CSV” to save this information.

19

Go back to the “Alignment Taskmanager” and click on your job.

19.1 Now click on “Download file” and then “zip” to download your tree.

Viewing tree using Dendroscope

20

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

See here for additional details and how to download Dendroscope: https://peerj.com/articles/960/#building-a-16s-rdna-tree
FigTree (another tree viewing software) also works!

Open the tree in Dendroscope.

21

Re-root the tree to the outgroup.

22

Expand the tree by clicking the expansion button, then scroll through the tree to locate the outgroup.

23

Click on the beginning of the taxon name, to select it, and reroot the tree by going to edit and selecting “re-root”.

24

Change the view to “phylogram" to make it easier to look at.

25

Write down who is the closest sister taxa/group and check to see who that taxa/group is. Now assess your phylogeny.

Where do your sequences fall on the tree? Does this make sense with your BLAST/RDP assignments?

Build phylogeny manually

26

Download fungal sequences to include in the tree from BLAST and the literature.

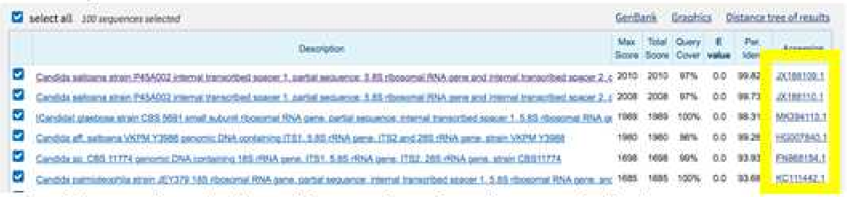
26.1 Based on the fungal taxonomy of your sequences, decide which of **your** fungi to group into a single tree.

e.g. all the fungi that blast to Penicillium or all fungi in the same “Genus” or “Family” based on RDP.

26.2 Next write down the accession numbers for fungal sequences from BLAST that are close matches.

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

e.g. the top hits on BLAST - maybe take the top 3 hits for each taxa. As you make your list for all the top hits to your taxa, you may have duplicates if many of your fungal taxa are similar, but we don't want duplicate sequences in our tree so you can remove the duplicate accession numbers from your list now.



☑ select all 100 sequences selected GenBank Graphics Distance tree of results

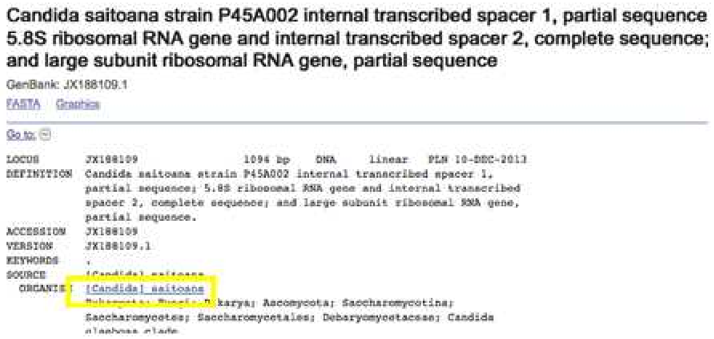
☑	Description	Max Score	Total Score	Query Cover	E value	Per. Iden	Accession
☑	Candida saitoana strain P45A002 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, c	2010	2010	97%	0.0	99.82	JX188109.1
☑	Candida saitoana strain P45A002 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, c	2008	2008	97%	0.0	99.73	JX188110.1
☑	[Candida] glaebosa strain CBS 5691 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA g	1989	1989	100%	0.0	98.31	MK394113.1
☑	Candida aff. saitoana VKPM Y3988 genomic DNA containing ITS1, 5.8S rRNA gene, ITS2 and 28S rRNA gene, strain VKPM Y3988	1980	1980	96%	0.0	99.28	HG007840.1
☑	Candida sp. CBS 11774 genomic DNA containing 18S rRNA gene, ITS1, 5.8S rRNA gene, ITS2, 28S rRNA gene, strain CBS11774	1698	1698	90%	0.0	93.93	FN868154.1
☑	Candida palmioleophila strain JEY379 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, an	1685	1685	100%	0.0	93.68	KC111442.1

26.3

Identify an outgroup and also some close relatives from other closely related genera (**this is the 'hardest' part**).

You can do this through literature searches (e.g. Google Scholar, "Fusarium phylogeny") and writing down accession numbers used to build trees previously. Or you can use BLAST/GenBank Taxonomy Viewer to pick some fungal taxa. For example:

1. Click on the Accession number of a top hit for one of your sequences.
2. Click on the "Organism".



Candida saitoana strain P45A002 internal transcribed spacer 1, partial sequence 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence

GenBank: JX188109.1

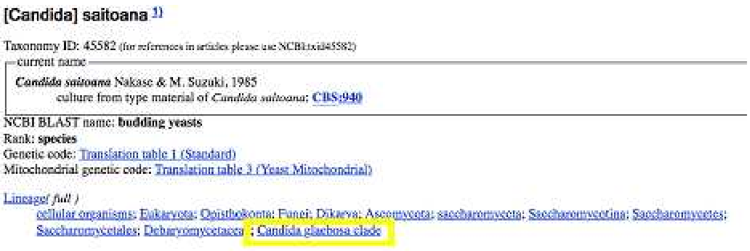
FASTA Graphics

Go to:

```
LOCUS       JX188109                1094 bp    DNA     linear   PLN 10-DEC-2013
DEFINITION  Candida saitoana strain P45A002 internal transcribed spacer 1,
            partial sequence; 5.8S ribosomal RNA gene and internal transcribed
            spacer 2, complete sequence; and large subunit ribosomal RNA gene,
            partial sequence.
ACCESSION   JX188109
VERSION     JX188109.1
KEYWORDS    .
SOURCE      [Candida] saitoana
  ORGANISM  [Candida] saitoana
            Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina;
            Saccharomycetes; Saccharomycetales; Debaryomycetaceae; Candida
            glaebosa clade
```

3. This will take you the GenBank Taxonomy Browser where you can browse for fungal data by taxonomy.

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6



[Candida] saitoana 1)

Taxonomy ID: 45582 (for references in articles please use NCBI:txid45582)

current name

Candida saitoana Nakase & M. Suzuki, 1985
culture from type material of *Candida saitoana*: CBS:940

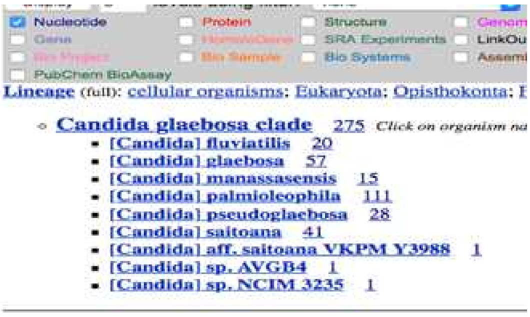
NCBI BLAST name: **budding yeasts**
Rank: **species**
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 3 (Yeast Mitochondrial)

Lineage(*full*)
cellular organisms; Eukaryota; Opisthokonta; Fungi; Dikarya; Ascomycota; saccharomyceta; Saccharomycotina; Saccharomycetes; Saccharomycetales; Debaryomycetaceae ; Candida glaebosa clade

Notes:

4. Check “nucleotide”.

These are all fungi where there is available DNA data in the database (the number next to the taxa is the number of DNA sequences available).



☑ Nucleotide ☐ Protein ☐ Structure ☐ Genom
☐ Gene ☐ HomoloGene ☐ SRA Experiments ☐ LinkOu
☐ Bio Project ☐ Bio Sample ☐ Bio Systems ☐ Assem
☐ PubChem BioAssay

Lineage (full): cellular organisms; Eukaryota; Opisthokonta; F

- **Candida glaebosa clade** 275 *Click on organism na*
  - [Candida] fluviatilis 20
  - [Candida] glaebosa 57
  - [Candida] manassasensis 15
  - [Candida] palmioleophila 111
  - [Candida] pseudoglaebosa 28
  - [Candida] saitoana 41
  - [Candida] aff. saitoana VKPM Y3988 1
  - [Candida] sp. AVGB4 1
  - [Candida] sp. NCIM 3235 1

5. You can also click on other levels in the “Lineage”. e.g. “Family” (Debaryomyces, etc) to obtain close relatives outside of the group of interest.

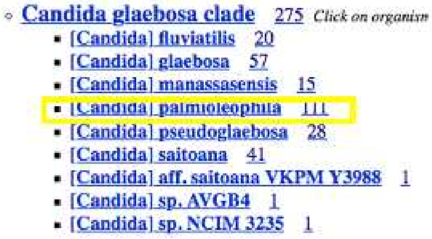
When building a tree you want to have:

i. 2-3 very close relatives (obtained previously based on ‘top hits’).
ii. And then 2-3 sequences from the same Genus (e.g. Candida) that are not the closest hits (so not from the same species (so in our example, not *Candida saitoana*)).
iii. 2-3 sequences that are outside the Genus (so from other Genera in the Debaryomyces).
iv. 1-2 sequences that are even more distant (so from another Order of fungi, etc).

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

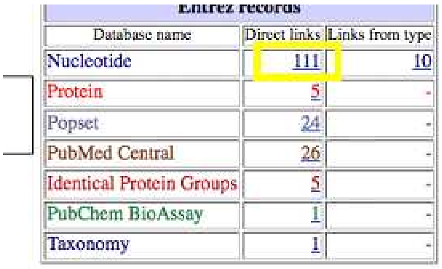
6. If you now want to obtain the Accession numbers of these fungal taxa to use in your tree, (in this example we are going for a sequence in the same Genus (Candida), but that is not *Candida saitona*).

a. Click on the name of the Genus you are interested in.



- Candida glaebosa clade 275 *Click on organism*
  - [Candida] fluviatilis 20
  - [Candida] glaebosa 57
  - [Candida] manassasensis 15
  - [Candida] palmioleophila 111
  - [Candida] pseudoglaebosa 28
  - [Candida] saitoana 41
  - [Candida] aff. saitoana VKPM Y3988 1
  - [Candida] sp. AVGB4 1
  - [Candida] sp. NCIM 3235 1

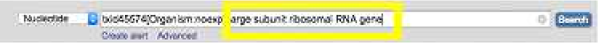
b. Then on the right, click on the number by "Nucleotide", this will take us to a list of all the DNA sequences for this organism.



Entrez records

Database name	Direct links	Links from type
Nucleotide	111	10
Protein	5	-
Popset	24	-
PubMed Central	26	-
Identical Protein Groups	5	-
PubChem BioAssay	1	-
Taxonomy	1	-

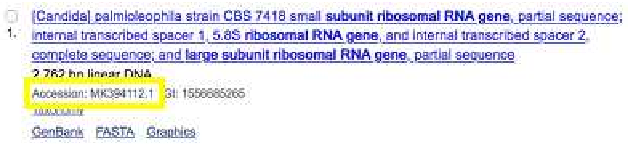
c. At this point, you may see a sequence already that says "large subunit ribosomal RNA gene", this what we want, if not, you can search for this - just paste "large subunit ribosomal RNA gene" after what is already included in the search bar.



Nucleotide txid45674[Organism:noexp large subunit ribosomal RNA gene

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

d. Ideally you should then see a hit that looks like this!! Write down that Accession Number.



□ 1. [Candida] palmioleophila strain CBS 7418 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence

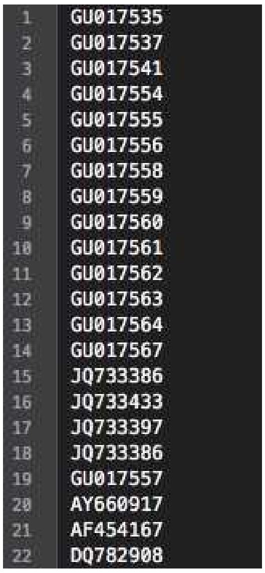
2,762 bp linear DNA

Accession: MK394112.1 GI: 1556686265

Taxonomy

GenBank FASTA Graphics

26.4 You should now have a list of Accession numbers that looks like this - with a new Accession number on each line - save this as a text file (e.g. .txt). We now want to obtain the sequences for all of these.

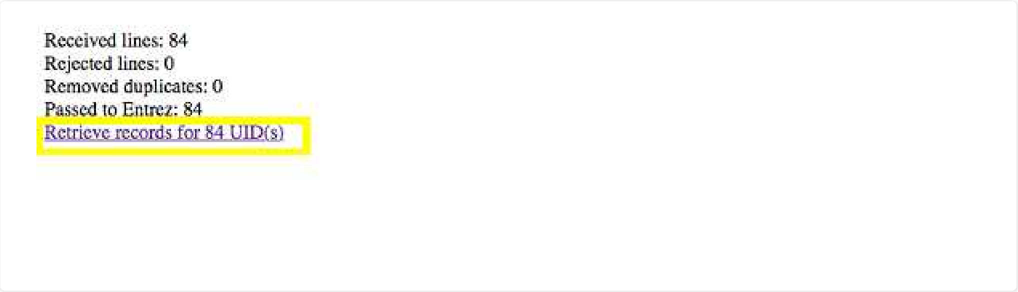


```
GU017535
GU017537
GU017541
GU017554
GU017555
GU017556
GU017558
GU017559
GU017560
GU017561
GU017562
GU017563
GU017564
GU017567
JQ733386
JQ733433
JQ733397
JQ733386
GU017557
AY660917
AF454167
DQ782908
```

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

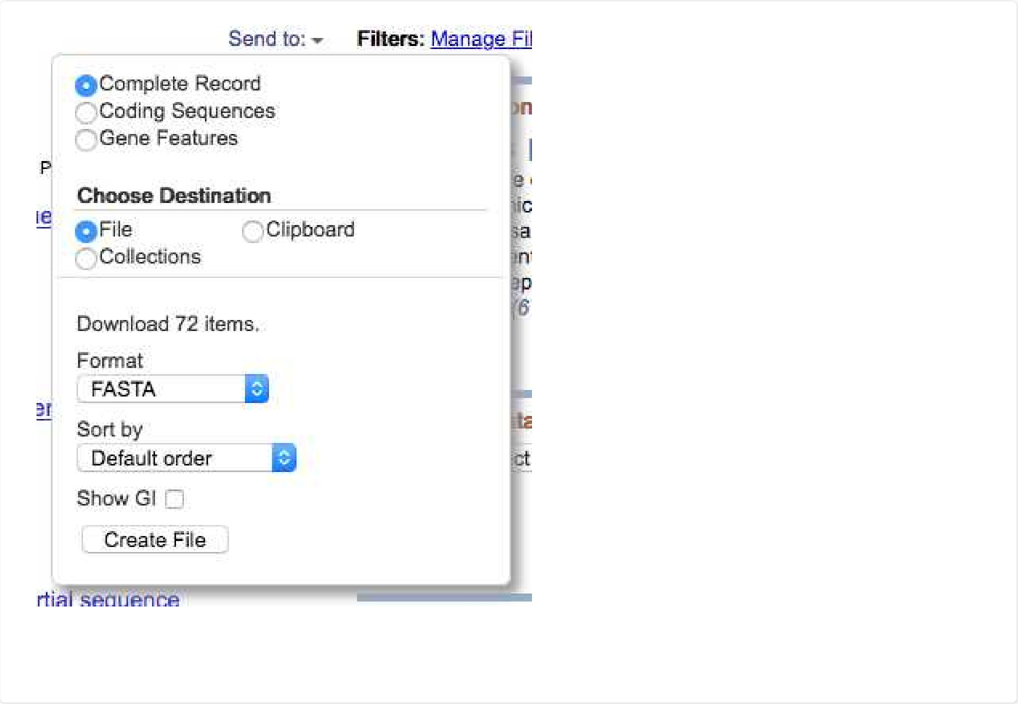
Go to https://www.ncbi.nlm.nih.gov/sites/batchentrez.

1. “Database” should be “Nucleotide”.
2. Select your file of Accession numbers.
3. Click “Retrieve”.
4. It will then take you to a screen where it tells you if it had any issues with your Accession numbers (e.g. duplicates, etc), and here you want to click on “Retrieve records”.



Received lines: 84
Rejected lines: 0
Removed duplicates: 0
Passed to Entrez: 84
Retrieve records for 84 UID(s)

5. This will take you back to NCBI which will now list all the sequences you want, you now want to find “Send to” on the right of the screen and pick “File”, Format of “Fasta”.



● Complete Record
○ Coding Sequences
○ Gene Features

Choose Destination

● File ○ Clipboard
○ Collections

Download 72 items.

Format: FASTA

Sort by: Default order

Show GI □

6. Yay, you’ve just downloaded all the sequences! Now you just need to combine these sequences into one file with your sequences (so copy + paste into one file). You should now have one “fasta” file that contains all your fungal sequences and all of the close relatives and outgroup sequences obtained from the above methods.

27

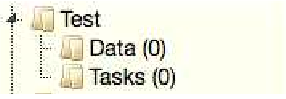
Generating an alignment.

27.1 We now need to align these sequences so that we know how each of the nucleotides in each sequence lines up.
Go to CIPRES (https://www.phylo.org/portal2/home.action) and create an account.

27.2 Log in to CIPRES. (https://www.phylo.org/portal2/home.action) and create an account.

27.3 Make a new project folder (e.g. Test).

This folder has both "Data" and a "Tasks" folders inside.

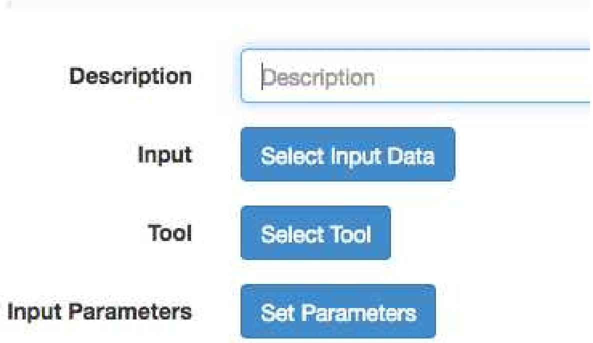


27.4 Click on "Data".

27.5 Upload your fasta file with all your fungal sequences.

27.6 Now click on "Tasks".

27.7 You want to create a new "Task".

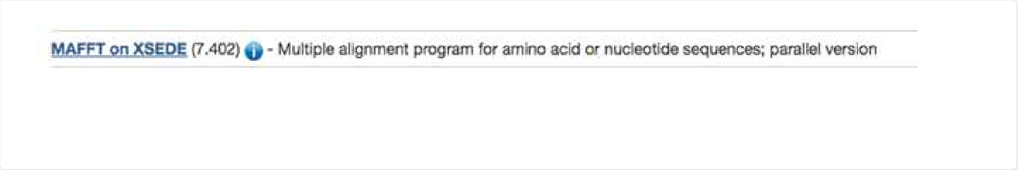


Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

27.8 Enter something that makes sense in the “Description” to name the task, eg. Fungal alignment with MAFFT.

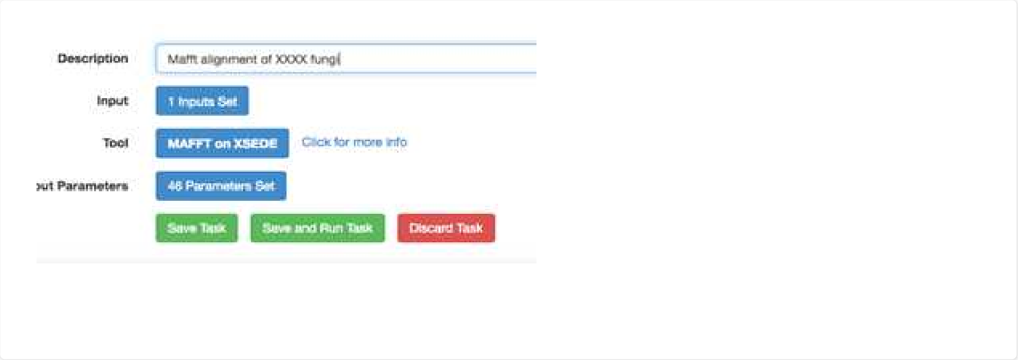
27.9 Next click “Select Input Data” and pick your uploaded fasta file.

27.10 Next “Select tool” and pick MAFFT (alignment software).



MAFFT on XSEDE (7.402) ⓘ - Multiple alignment program for amino acid or nucleotide sequences; parallel version

You should now have something that looks like this:



Description: Mafft alignment of XXXX fungi

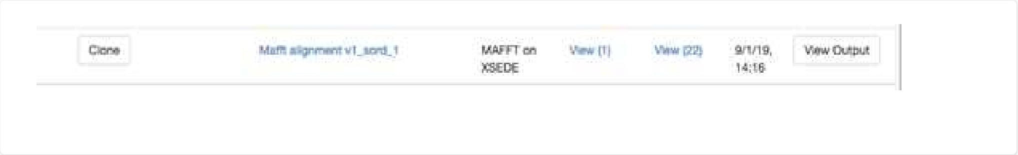
Input: 1 Inputs Set

Tool: MAFFT on XSEDE — Click for more info

›ut Parameters: 46 Parameters Set

27.11 Click “Save and Run Task”.

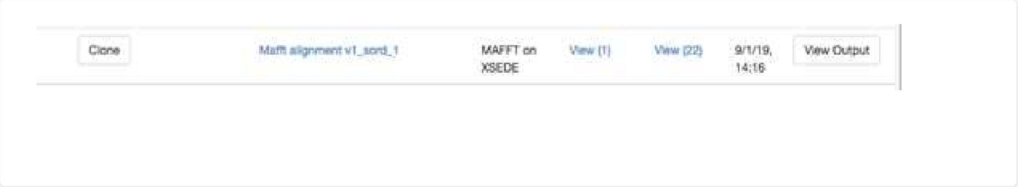
27.12 You will get an email when it is finished running and then you can go back to CIPRES, to “Tasks” and now click on “View Output".



Clone	Mafft alignment v1_sord_1	MAFFT on XSEDE	View (1)	View (22)	9/1/19, 14:16	View Output

27.13 You want to click “Download” next to the “output.mafft” file.

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6



Clone	Mafft alignment v1_sord_1	MAFFT on XSEDE	View (1)	View (22)	9/1/19, 14:16	View Output

You now have your aligned sequences!

28

Trim the alignment (remove parts that don't overlap).

28.1 Go to http://phylemon2.bioinfo.cipf.es/.

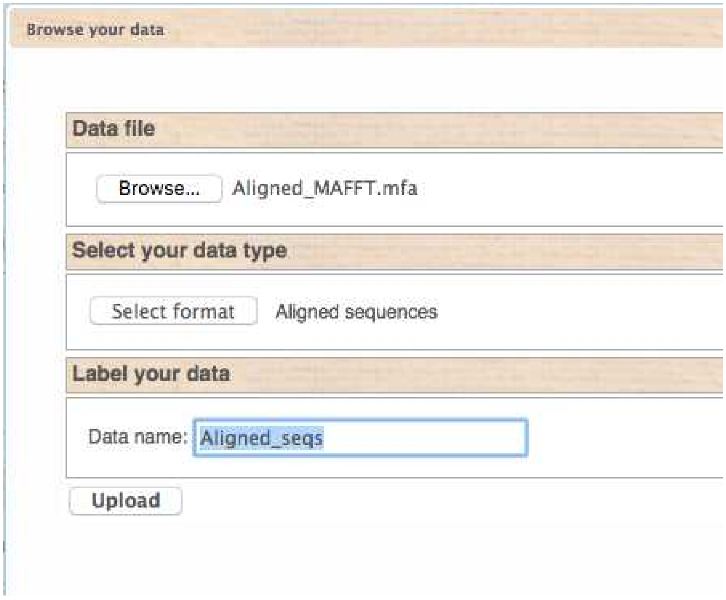
28.2 Sign in / make an account or work as an "anonymous user".

28.3 Go to "Utilities".

28.4 Click on "Alignment Utilities".

28.5 Click on "TrimAl".

i. Upload your aligned sequence file from CIPRES.



Browse your data

Data file

Browse... Aligned_MAFFT.mfa

Select your data type

Select format Aligned sequences

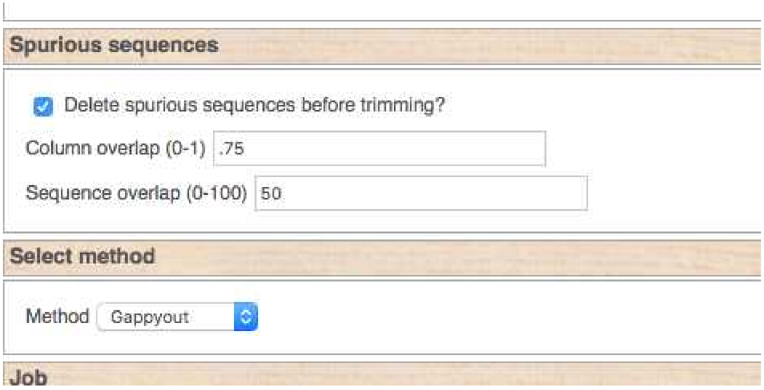
Label your data

Data name: Aligned_seqs

Upload

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

ii. Now edit the parameters for TrimAl, click on the checkmark by “delete spurious ....” and add the values 0.75 and 50 respectively (or for advanced users, your values of choice) and then also select “gappyout”.



Spurious sequences

☑ Delete spurious sequences before trimming?

Column overlap (0-1) .75

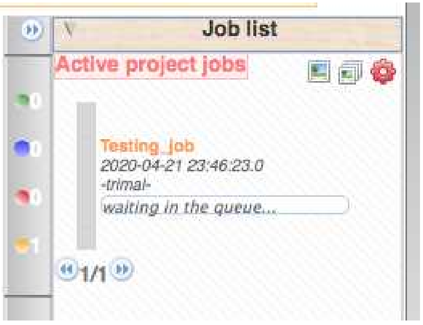
Sequence overlap (0-100) 50

Select method

Method Gappyout

Job

iii. Now run the job!.



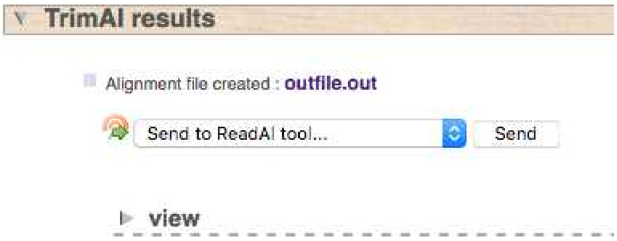
Job list

Active project jobs

Testing_job
2020-04-21 23:46:23.0
-trimal-
waiting in the queue...

If its running, it should show up in the job list!

iv. When its finished running you want to download the new fasta file, just click on “outfile.out” to download.



TrimAl results

Alignment file created : outfile.out

29

Now we want to look at our alignment! (optional, but recommended).

29.1 Download JalView (http://www.jalview.org/).

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

29.2 Open JalView and drag your trimmed alignment file into JalView and it should open it up.

> For beginners - you just want to make sure that it looks like things “line up” and that nothing looks too “weird”.

30 🖥

Next, we can either make a slower, potentially more accurate phylogeny or a faster, potentially less accurate phylogeny. To make a slow phylogeny skip to steps 32 and 33. To make a fast phylogeny go to step 31.

31 🖥

Build a phylogeny using FastTree2.

31.1 Go back to CIPRES.

31.2 Click on "Data".

31.3 Upload your fasta file (outfile.out) with your aligned trimmed sequences.

31.4 Now click on "Tasks".

31.5 Create a new "Task".

31.6 Enter something in the description (e.g. fasttree of Candida).

31.7 Next "Select input data" and select your uploaded file.

31.8 Next "select tool" and pick FastTreeMP on XSEDE.

31.9 Click on “Parameters”.

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

31.10 Change “Maximum Hours to Run (up to 168 hours)” to 2.

31.11 Save parameters and submit task.

31.12 When finished “View output”.

31.13 Download the file that ends in ".tre".

> Yay we have a tree!. You can stop here unless you want to make a more accurate tree in steps 31 and 32 below.

32 🖥

Pick an evolutionary model!.

32.1 Back to CIPRES!.

32.2 Click on “Data”.

32.3 Upload your fasta file (outfile.out) with all your trimmed and aligned sequences.

32.4 Now click on “Tasks”.

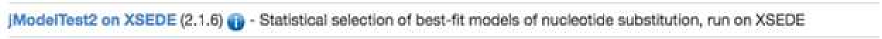
32.5 You want to create a new “Task”.

32.6 Enter something that makes sense in the “Description” to name the task, eg. Fungal model test.

32.7 Next click “Select Input Data” and pick your uploaded file.

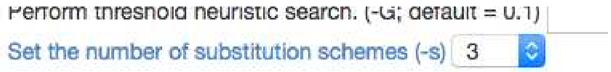
Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

32.8 Next “Select tool” and pick JmodelTest2.



jModelTest2 on XSEDE (2.1.6) - Statistical selection of best-fit models of nucleotide substitution, run on XSEDE

32.9 Click on “Parameters”.
i. Change “Maximum Hours to Run (up to 168 hours)” to 2.
ii. Under “Configure Model Search” set the number of substitution schemes to 3.

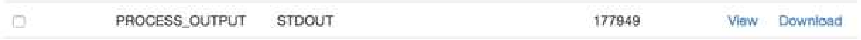


Perform threshold heuristic search. (-G; default = 0.1)
Set the number of substitution schemes (-s) 3

32.10 Save parameters and submit task.

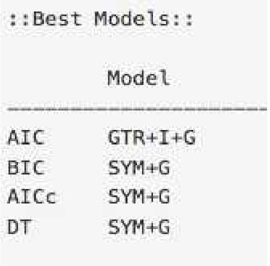
32.11 When finished “View output”.

32.12 Next click on “View” next to STDOUT.



□ PROCESS_OUTPUT	STDOUT	177949	View	Download

32.13 Scroll to the bottom and find this information:



```
::Best Models::

        Model
---------------------
AIC     GTR+I+G
BIC     SYM+G
AICc    SYM+G
DT      SYM+G
```

This tells us which evolutionary model works best, hopefully they all match! Very likely it will say GTR + I + G or GTR.

33

Build a phylogeny !! Finally!!

33.1 Also on CIPRES!.

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6

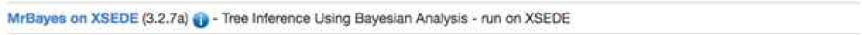
33.2 Now click on “Tasks”.

33.3 You want to create a new “Task”.

33.4 Enter something that makes sense in the “Description” to name the task, eg. Fungal phylogeny.

33.5 Next click “Select Input Data” and pick the same file you just used for the model test above (the trimmed and aligned file).

33.6 Next “Select tool” and pick MrBayes. *Note for advanced users: CIPRES also supports RAxML*



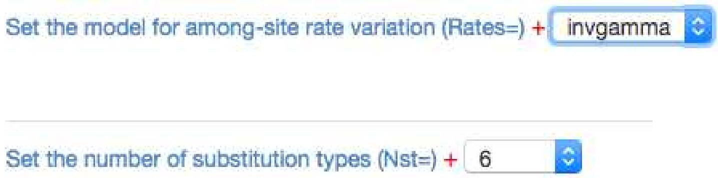
MrBayes on XSEDE (3.2.7a) ⓘ - Tree Inference Using Bayesian Analysis - run on XSEDE

33.7 Click on “Parameters”.

33.8 Pick “MrBayes 3.2.6”.

Now we want to set up the evolutionary model we decided on above (see: https://gist.github.com/brantfaircloth/895282).

E.g. for GTR + I + G



Set the model for among-site rate variation (Rates=) + invgamma

Set the number of substitution types (Nst=) + 6

33.9 Save parameters and submit task.

33.10 You will get an email when it is finished running and then you can go back to CIPRES, to “Tasks” and now click on “View Output”.

33.11 Download the “infile.nex.con.tre” file and open this file in a tree viewing software to see your tree!.

□ infile.nex.con.tre 47314 View Download

YAY! We finished making our tree!.

Citation: Cassie Ettinger (05/21/2021). Fungal isolate identification using ITS-LSU regions. https://dx.doi.org/10.17504/protocols.io.butvnwn6