

The logo consists of the word "SKILL" in white, bold, sans-serif capital letters inside a red rounded rectangle, followed by the word "PILLS" in red, bold, sans-serif capital letters inside a white rounded rectangle.

SKILL PILLS

Skill Pill+: Evolutionary Genomics

Bourguignon Unit

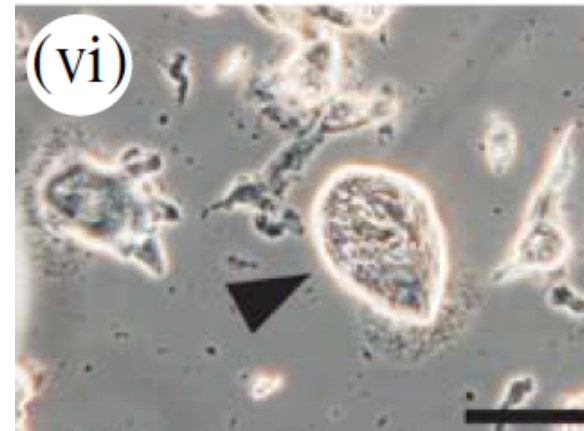
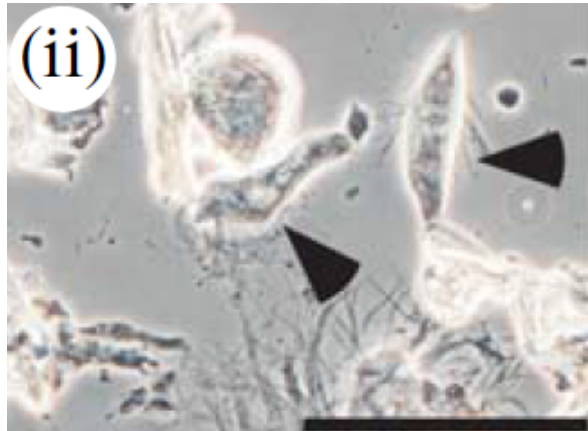
December 2018



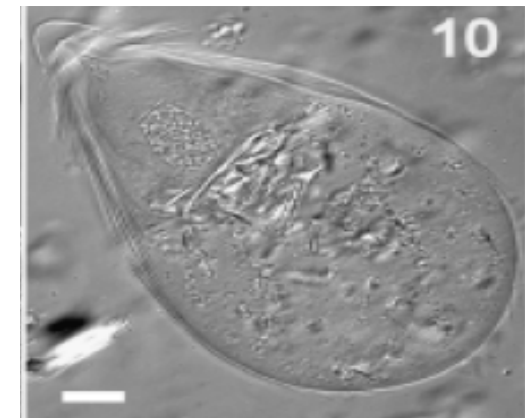
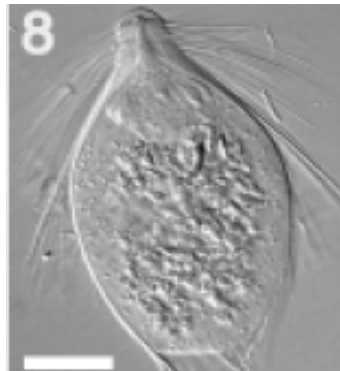
OIST

The workflow

- Clean the bench with 70% ethanol and tissue
- Clean the microscope lens and platform with 70% ethanol and kim-wipe (as the lens is sensitive).
- Pour 500 ul of 70% ethanol into a petri plate.
- Add two drops of Solution U to the slide using 100 ul pipette.
- Transfer the termite to the petri plate containing ethanol for a few seconds to clean the termite.
- Remove the termite from petri plate, transfer it to the cover of the petri plate (or any clean surface).
- Dissection: Using two tweezers, hold the head of the termite with one tweezer, and pull the gut out of the termite from the rear side [abdomen side] using the other tweezer.
- Transfer the gut to the slide, mix the gut content in one of the drops of Solution U on the slide.
- Transfer a few drops of the gut content from one drop to the other drop, using pipette. This is called as dilution.
- Check under the microscope (80X magnification).



Ohkuma *et al.* 2008



Carpenter *et al.* 2009

**Lets learn about the
microbes in the
termite gut!**



- What are protists?
- What are the function of protists in the termite gut?
- Types of protists in the gut
- How to identify the protists in the termite gut?

- Protists are eukaryotes, which are neither plant, animal, bacteria or fungi.
- They can have long tails called flagellates on their surface, to help them move.
- They can eat by photosynthesis, enzyme secretions outside the body, or by eating pre-digested food. Termite gut protists are the latter ones.

<https://www.ducksters.com/science/biology/protists.php>

Functions of protists

- They can be parasites, causing disease or symbiont, helping the host organism.
- The termite gut protists are symbionts. They help digest cellulose eaten up by termites.

- There are more than 400 different species of protists in the termite gut.
- They can be classified into two groups:

Parabasalids - They are larger in size, they have parabasal filaments.

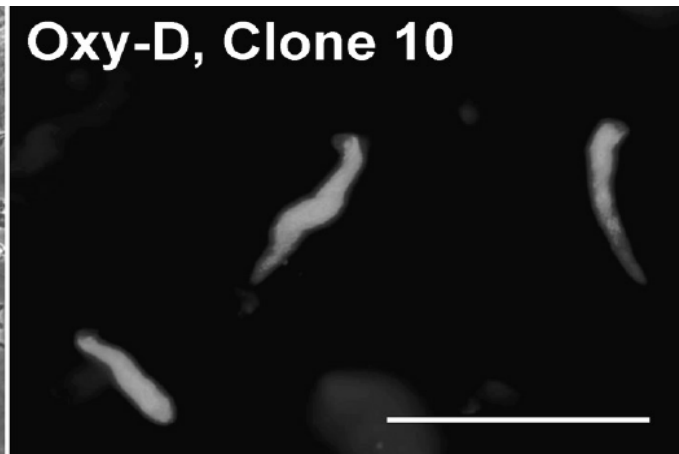
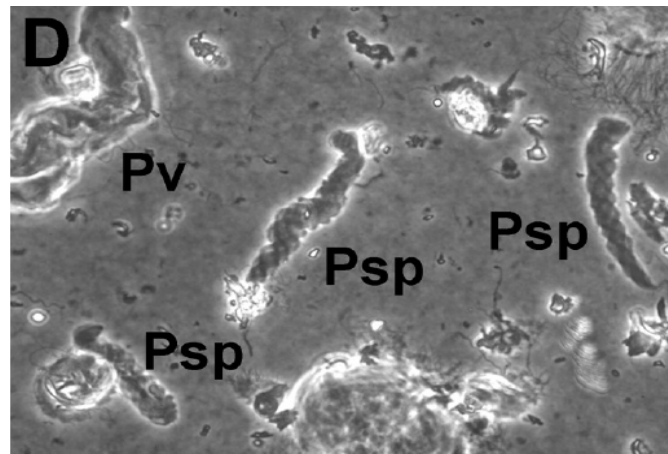
Oxymonads - They are smaller in size, they lack these filaments.

Parabasalids

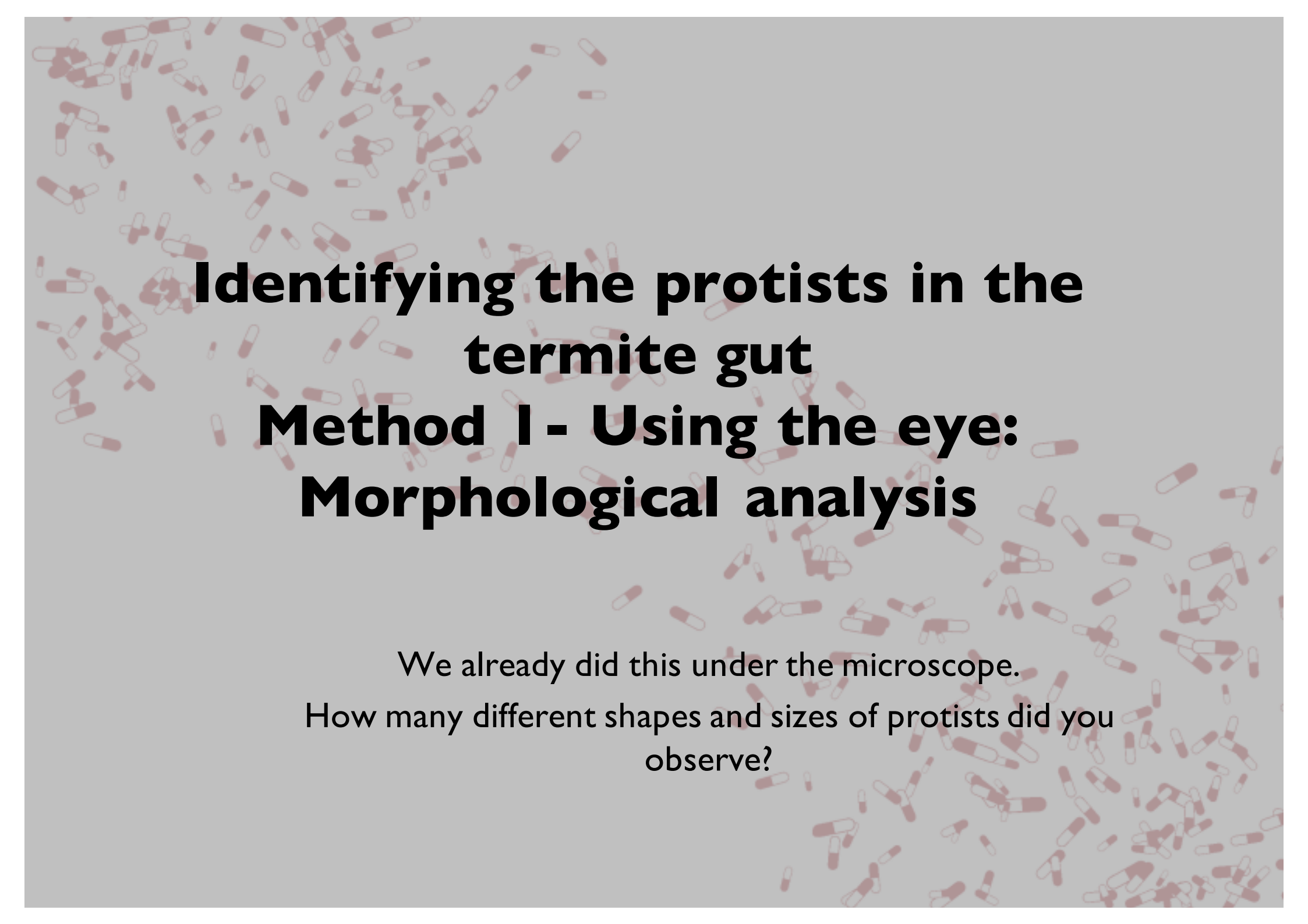


Carpenter *et al*, 2009

Oxymonads



Stingl and Brune, 2003

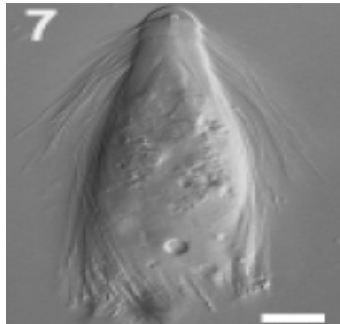
The background of the slide is a light gray color with a pattern of numerous small, red and white capsules scattered across it. The capsules are oriented in various directions, some horizontally, some vertically, and some at angles. They are semi-transparent, allowing the gray background to show through.

Identifying the protists in the termite gut

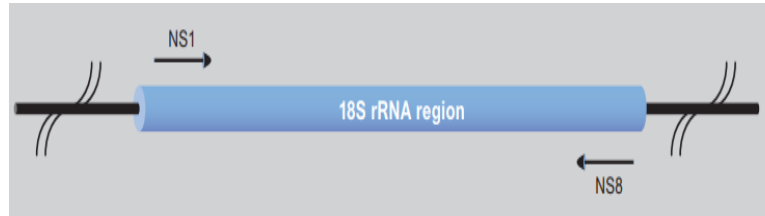
Method I - Using the eye: Morphological analysis

We already did this under the microscope.
How many different shapes and sizes of protists did you
observe?

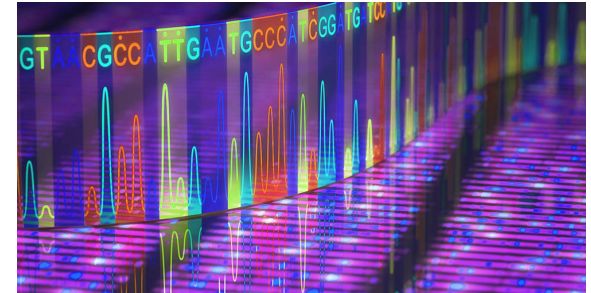
Method 2 - 18S sequence analysis



**Isolate protists
from the gut**



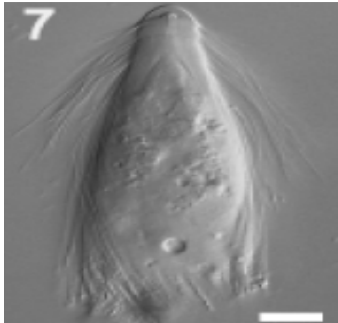
**Amplify the 18S region
using primers**



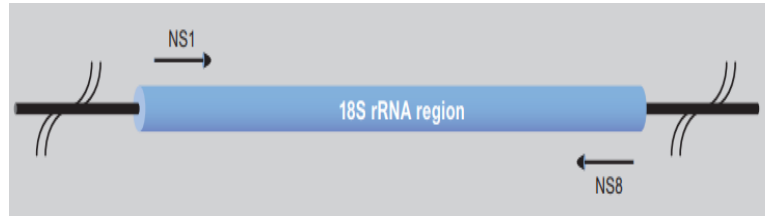
**Next generation
sequencing**

- Amplify the 18S ribosomal RNA sequence of the protists observed under the microscope.
- 18S ribosomal RNA is ~13,105 base pairs long. The primers are selected to represent the diversity of each eukaryotic order. For termite gut Parabaslid species, Ohkuma *et al.* (1998) primers were used. These primers target regions ~19-42 using forward primer and ~1772-1795 using reverse primer.

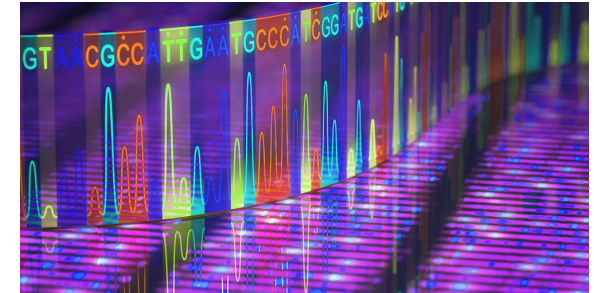
Method 2 - 18S sequence analysis



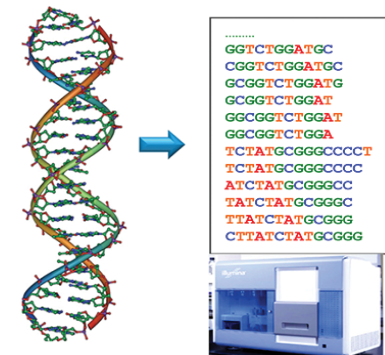
**Isolate protists
from the gut**



**Amplify the 18S region
using primers**



**Next generation
sequencing**



Data analysis

Today we will be making sense of the data we usually get from the sequencing machine.

- The sequenced data is in fastq format. This is not in human readable format.
- To check the quality of the reads, we will use Fastqc software.

<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

NOTE-

(Taxonomy analysis can be done with fastq or fasta files. I have already converted the fastq files to fasta files for ease of analysis. The overall steps are the same for both file formats.)

Diagram illustrating the Fastq format structure:

```
@FORJUSP02AJWD1
CCGTCAATTCATTTAAGTTTAACTTGCAGCGTACTCCCCAGGCGGT
+
AAAAAAAAAAAA:99@:::??@::FFAAAAACAA:::BB@@?A?
```

Labels:

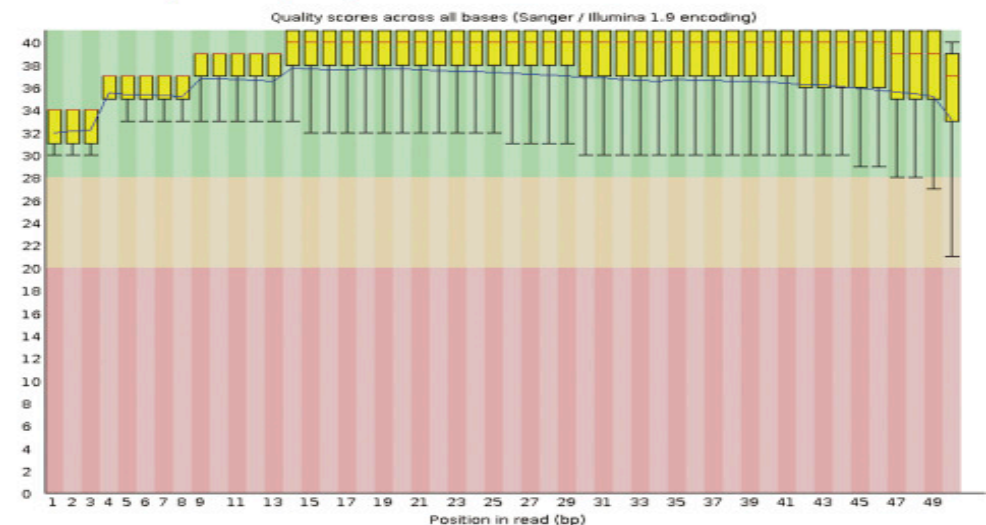
- Label:** @FORJUSP02AJWD1
- Sequence:** CCGTCAATTCATTTAAGTTTAACTTGCAGCGTACTCCCCAGGCGGT
- Q scores (as ASCII chars):** :99@:::??@::FFAAAAACAA:::BB@@?A?
- Base=T, Q=':'=25**

Fastq format

A

```
@HWI-ST193:397:D16B3ACXX:2:1101:1091:2467 1:N:0:CGATGT
ATCACAGACAGAAGAGGATTGTACAGAGGAGCTCTTGACTTCCTGCATC
+
:=:ABBD AFFDDFHHIGCEEB:CEBF<+A??F3?D*?D*?B*:?B<?)?#
```

B Per base sequence quality



Quality check using Fastqc software

- Blast search can be done on the web or on command line.

The web interface

The screenshot displays the NCBI BLAST web interface. At the top, there is a navigation bar with the NIH logo, "U.S. National Library of Medicine", and the NCBI logo, "National Center for Biotechnology Information". A "Sign in to NCBI" link is also present. Below this, a banner shows "BLAST® >> blastn suite" and navigation links for "Home", "Recent Results", "Saved Strategies", and "Help".

The main section is titled "Standard Nucleotide BLAST". It features a tabbed interface with "blastn" selected. A description states: "BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)". There are links for "Reset page" and "Bookmark".

The "Enter Query Sequence" section includes a large text area for "Enter accession number(s), gi(s), or FASTA sequence(s)", a "Clear" link, and a "Query subrange" section with "From" and "To" input fields. Below this is a file upload section with "Or, upload file", a "Choose File" button, and "No file chosen" text. A "Job Title" field is also present with the prompt "Enter a descriptive title for your BLAST search". A checkbox for "Align two or more sequences" is also visible.

The "Choose Search Set" section includes a "Database" section with radio buttons for "Human genomic + transcript", "Mouse genomic + transcript", and "Others (nr etc.):". A dropdown menu is set to "Nucleotide collection (nr/nt)". The "Organism" section is optional and includes a text field for "Enter organism name or id—completions will be suggested", an "Exclude" checkbox, and a "+" button. A note states: "Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown".

What are the advantages and disadvantages of web based blast search?

- Align DNA sequences with MEGA7 or with MAFFT

```
mkdir MAFFT
```

```
cd MAFFT
```

- MAFFT command#copy fas file to MAFFT folder

```
cp <path to phylogeny_parabasalid.fasta> .
```

```
module load mafft/7.305
```

```
mafft --maxiterate 1000 --globalpair  
      phylogeny_parabasalid.fasta >  
      phylogeny_parabasalid_aligned.fasta
```



```
l) scp -r <path to
alignment> <local
directory>
```

2) view alignment and trim if needed

3) save trimmed fasta file if possible

- Determine best model, and partitioning strategy with PartitionFinder2

- #partitionfinder requires .phy format and .cfg format files

1) make .phy file. go to geneious >> import .fasta >> export .phy file

Also see example in

```
/work/spp_dec2018/Menglin/FILE/phylogeny_parabasalid_alignedTrim.phy
```

2) make .cfg file see example in /work/spp_dec2018/Menglin/FILE/partition_finder.cfg

3) copy all two files in one directory

```
mkdir Partitionfinder
```

```
cp <path to .phy file> Partitionfinder/
```

```
cp <path to .cfg file> Partitinfinder/
```

4) Partitionfinder command (slurm script)

```
module load partitionfinder/2.1.1
```

```
PartitionFinder.py Partitionfinder/
```

5) check result

```
cd analysis
```

```
less best_scheme.txt
```



- Identify the best tree, with RAxML. That tree will be used as the tree.

```
module load raxml/8.2.11
raxmlHPC -m GTRGAMMAI -p 12345 -# 20 -k -q partitions.txt
        -s phylogeny_parabasalid_alignedTrim.phy -n theTree
```

- Carry out bootstrapping. This command makes 1000 trees with bootstrap method

```
raxmlHPC -f d -# 1000 -p 1000 -x 1000 -k -q partitions.txt -s
        phylogeny_parabasalid_alignedTrimm.phy -n RAxML_bootstrap.trees
        -m GTRGAMMAI
```

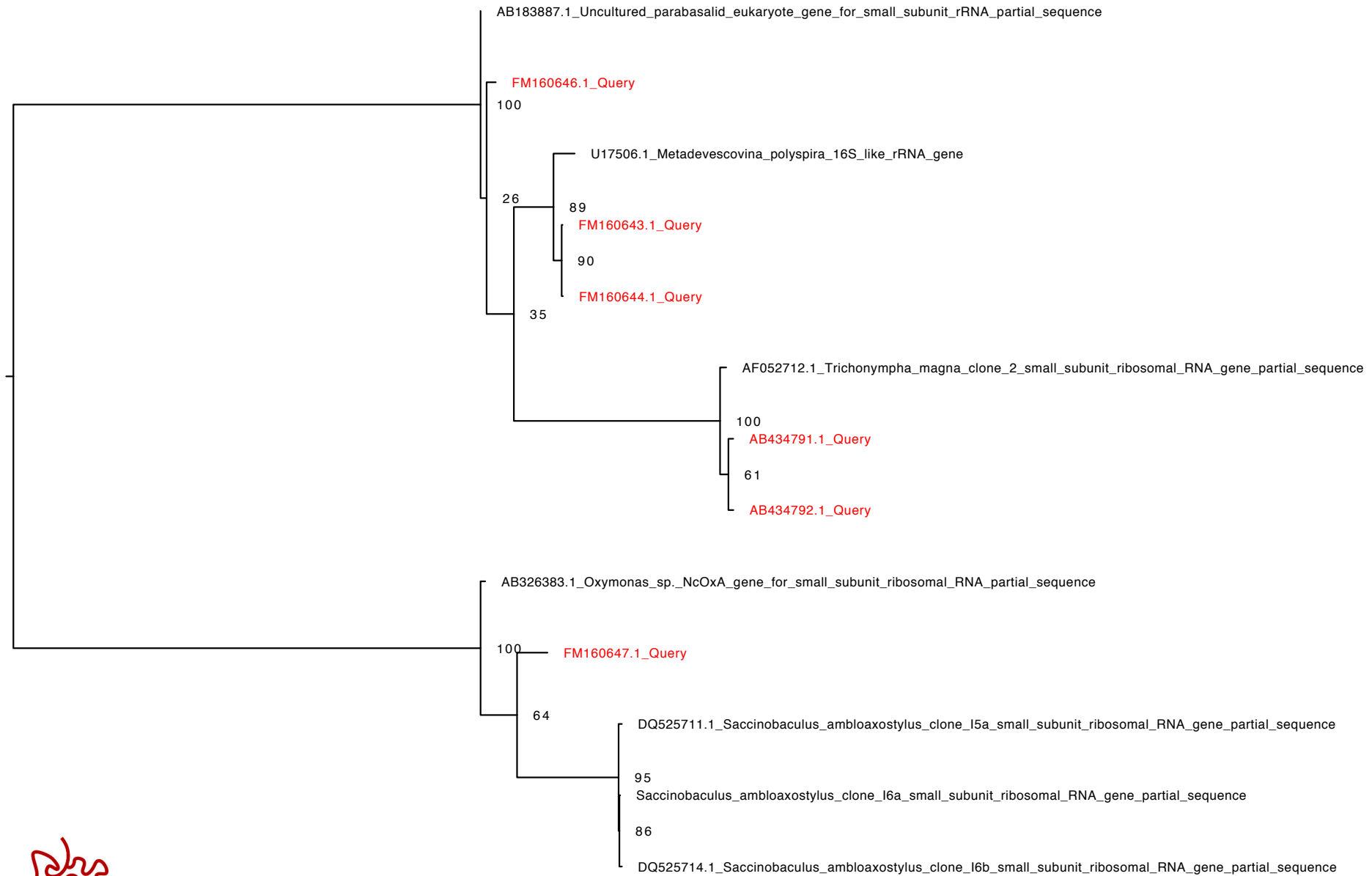
- Create bipartition tree that can be read with Figuretree

```
raxmlHPC -m GTRGAMMAI -p 12345 -f b -t RAxML_bestTree.theTree
        -z RAxML_bootstrap.trees -n tree
```

- ALL IN ONE

```
raxmlHPC -f a -m GTRGAMMA -q partitions.txt
        -s mt_for_Menglin.phy -n tree -p 12345 -x 12345 -# 1000
```

Phylogenetic tree



- Tracer-

<https://github.com/beast-dev/tracer/releases/tag/v1.7.1>

- BEAST-

<http://beast.community/installing>