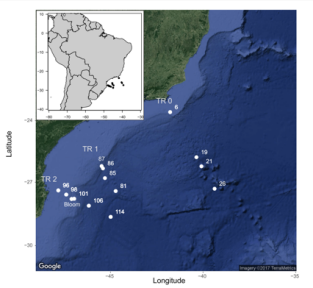


Dada2 tutorial

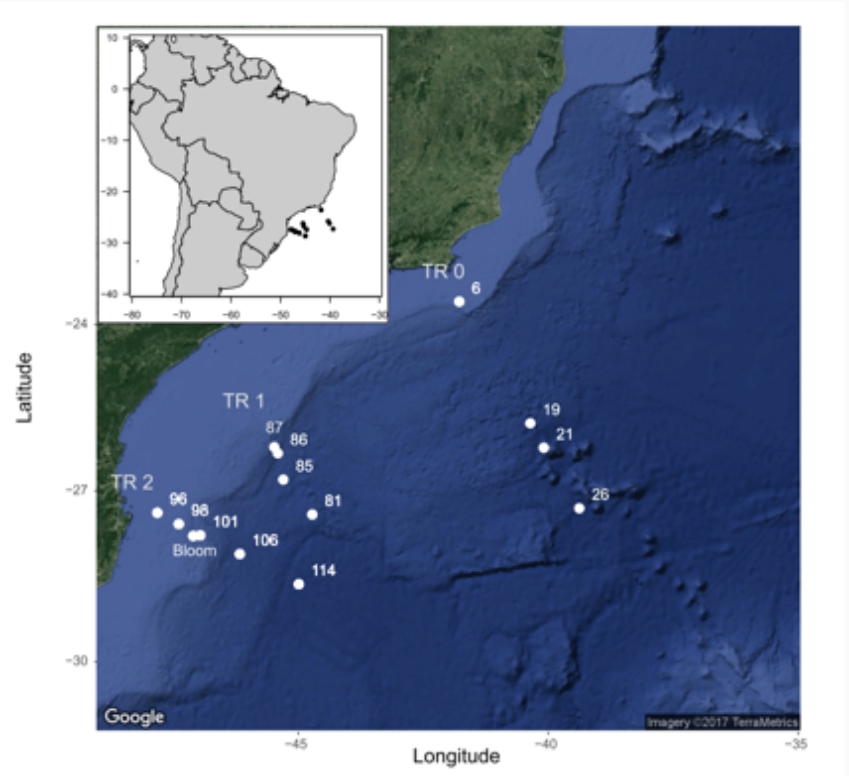
Daniel Vaulot

23 11 2018

Station Biologique de Roscoff, CNRS-Sorbonne Université

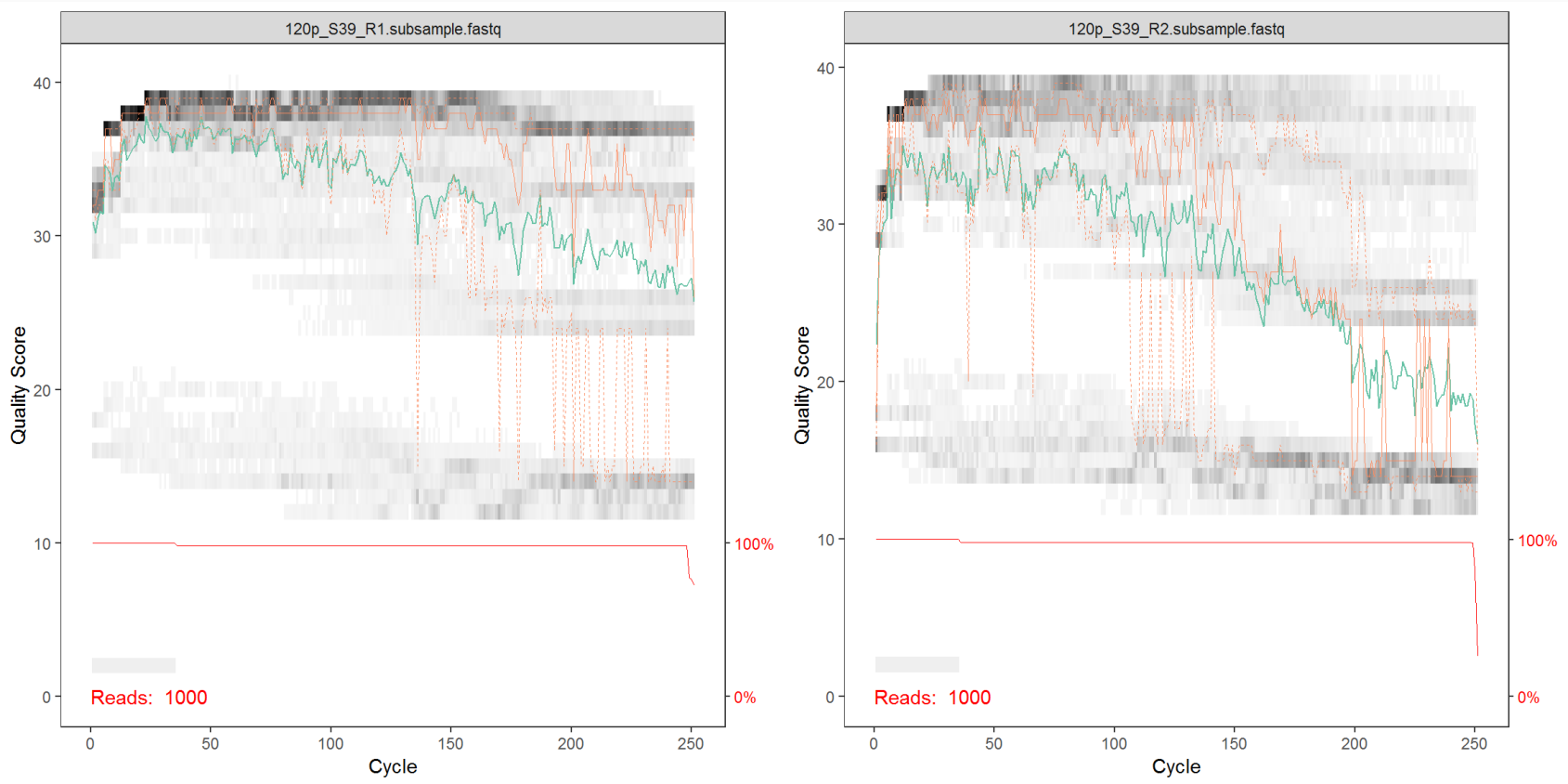
Data used

- The samples originate from the CARBOM cruise (2013) off Brazil (Ribeiro et al. 2018).
- Samples have been sorted by flow cytometry and 3 genes have been PCR amplified :
  - 18S rRNA - V4 region
  - 16S rNA with plastid
  - nifH
- The PCR products have been sequenced by 1 run of Illumina 2*250 bp.
- The data consist of the picoplankton samples from one transect and fastq files have been subsampled with 1000 sequences per sample.



Inspect and process raw sequences

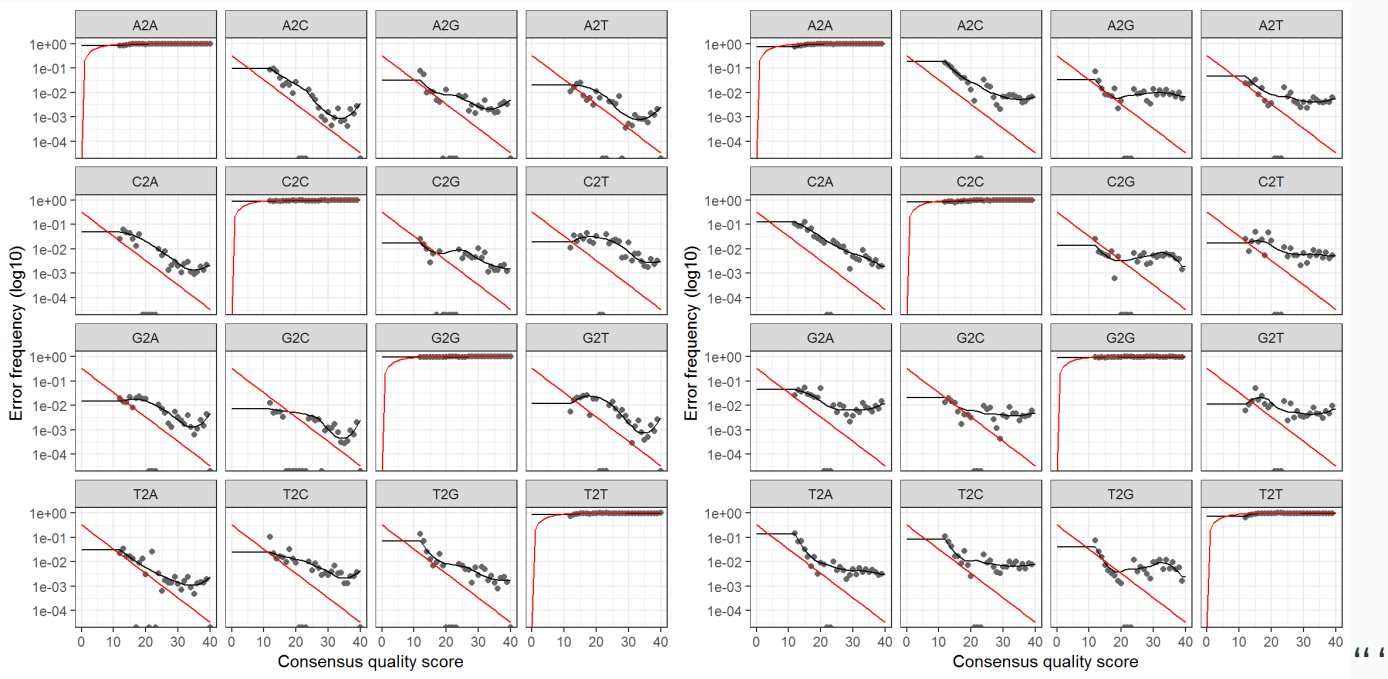
- Construct a list of the fastq files
- Get sequences per file
- Plot quality



- Remove primers
  - dada2 by length
  - cutadapt by sequence

Dada2

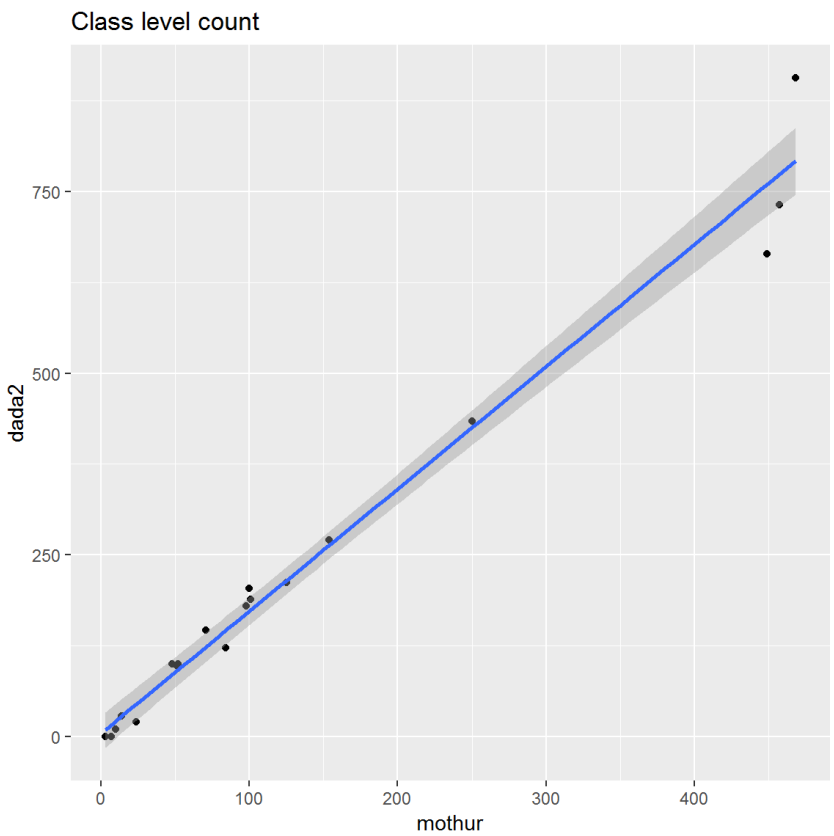
- Learn error rates

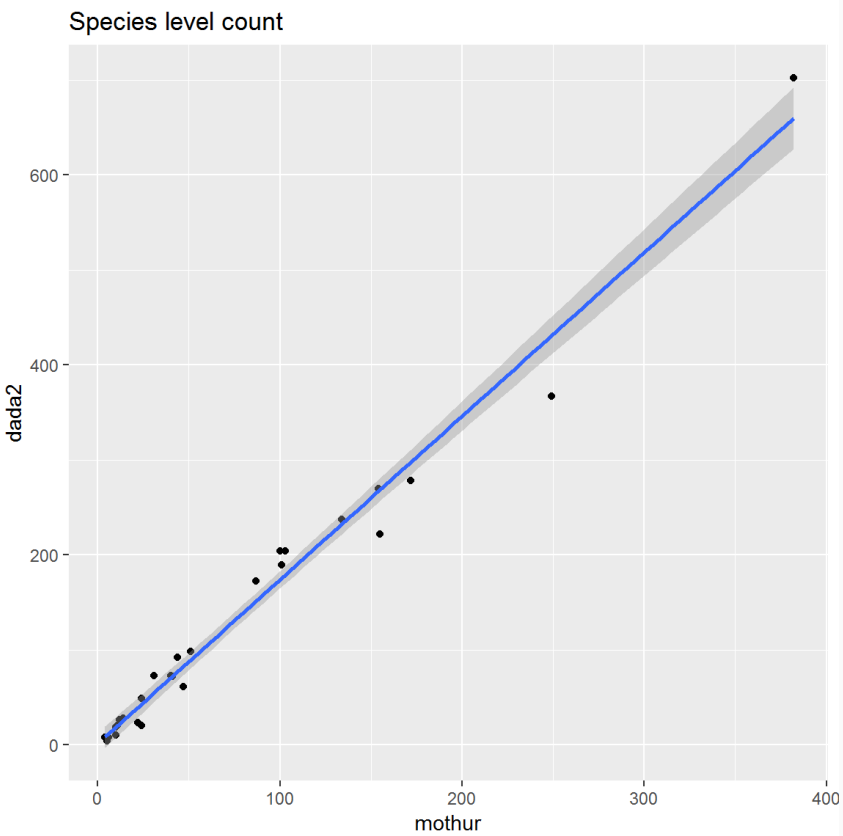


- Dereplicate reads
- Sequence-variant inference algorithm to the dereplicated data (the heart of dada2)
- Merge pairs
- Remove chimeras de-novo
- Assign taxonomy with Wang classifier
  - For euks use PR2 database
  - Provide taxonomy and bootstrap
- Filter out sequences that have bootstrap < 80 at higher taxonomic level

Compare mothur vs dada2

- Class level
- Species level





Now

Let us do it !