

# Marine Molecular Biodiversity Assessment in the Dutch Caribbean

Heleen Bouwer, MSc Internship

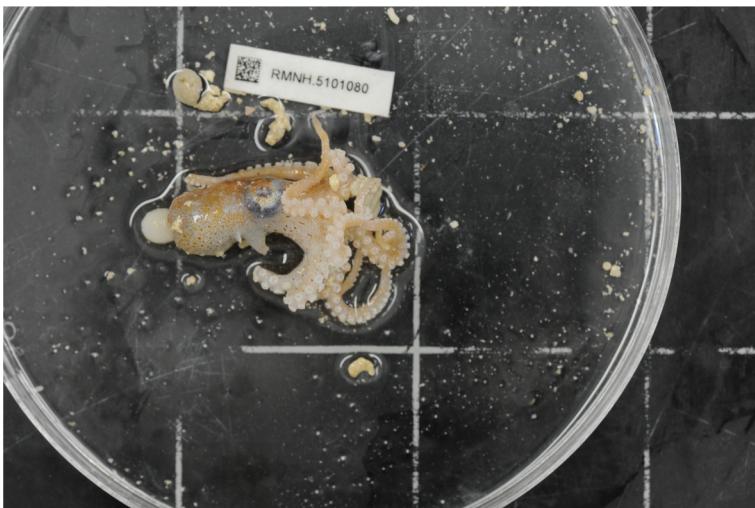


# Background

- Internship for MSc Biology: Molecular Genetics and Biotechnology
- Supervisor: Rutger Vos
  - Project led by Arjen Speksnijder
  - Labwork by Elza Duijm
  - Bioinformatics supervision Rutger Vos

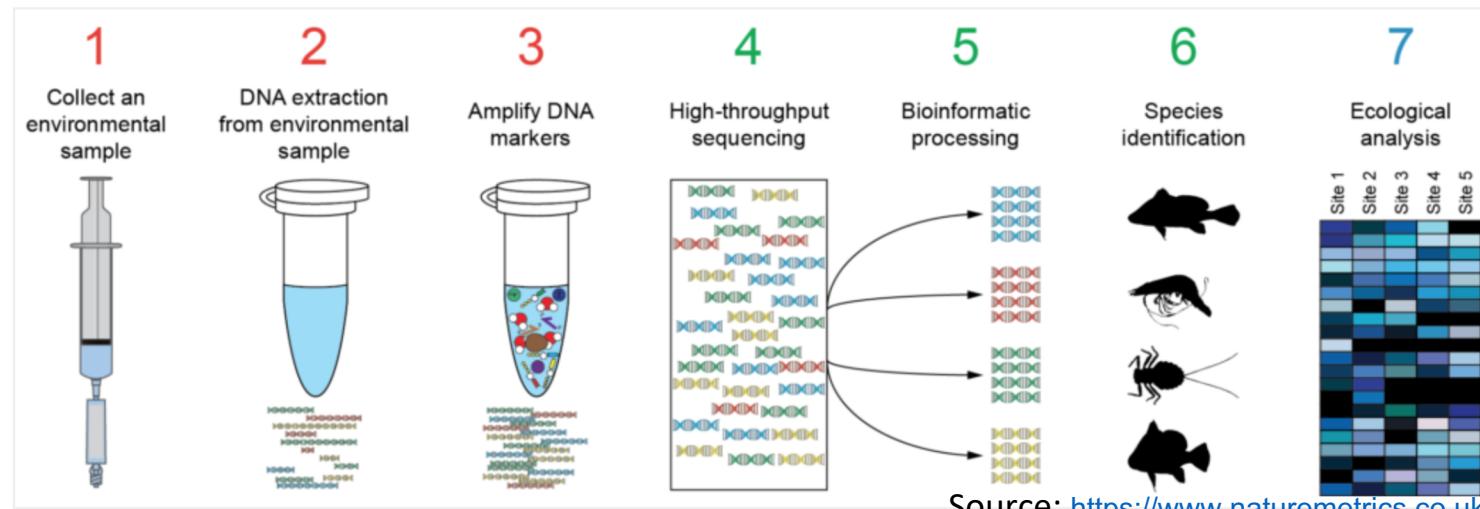
# Background

- Naturalis went on St. Eustatius (2015) and Saba expeditions (2018)
  - Joint expedition with NIOZ and IMARES
  - Overarching goal: establishing baseline marine biodiversity in Dutch Caribbean
  - Lots of specimen caught, recorded and barcoded.
  - Water samples taken at several locations and depths for eDNA biodiversity analysis



# Molecular Biodiversity Assessment

- Metabarcoding: DNA based approach for taxonomic identification of environmental samples using high throughput sequencing.
  - COI gene used as marker
- Using database of barcodes as reference
  - Naturalis local database
  - International Barcode of Life (IBOL)



# MinION as sequencer

- MinION start up pack costs \$1000
  - No capital expenditure required
  - flow cells can be washed and reused
- Small and portable device
  - Not constrained by a lab
- Long reads
- Scalable



**For MinION / GridION  
Flongle**

Adapter to enable small, rapid nanopore sequencing tests, for mobile or desktop sequencers



**MinION Mk1B**

Your personal nanopore sequencer, putting you in control



**MinION Mk1C**

Your personal nanopore sequencer including compute and screen, putting you in control



**GridION Mk1**

Higher-throughput, on demand nanopore sequencing at the desktop, for you or as a service



**PromethION 24/48**

Ultra-high throughput, on-demand nanopore sequencing, for you or as a service

Source: Oxford Nanopore,

# How it works

- Membrane that holds nanopores
- Passing of nucleotides causes characteristic current disruptions
- Raw signals can be basecalled



Source: Oxford Nanopore,

Problem: higher error rates than Illumina (5-15%)

# Research Questions

## Bioinformatics:

- How to demultiplex, cluster and create consensus sequences for MinION pooled sequencing data for metabarcoding, given the higher error rate and long sequences that the data contain?

## eDNA dynamics

- How does the gradient of eDNA change geographically and along the water column?

# What was already done

- Samples taken
- eDNA extracted
- Tag per sample added
  - So the location can be traced back
- Samples pooled for PCR
- Sequencing of all pools with Illumina MiSeq (by BaseClear)
  - Established approach: This serves as control data for the current project
- **Sequencing of one pool with Oxford Nanopore MinION (by Naturalis)**

# Approach

- MinION sequencing done mid 2019
  - Run time 20 hours
  - 23.4 GB raw data
  - 5,314,990 reads generated
- Output format
  - Fast5 (raw data, not basecalled)
- First step: basecall data using most recent basecaller (Guppy)

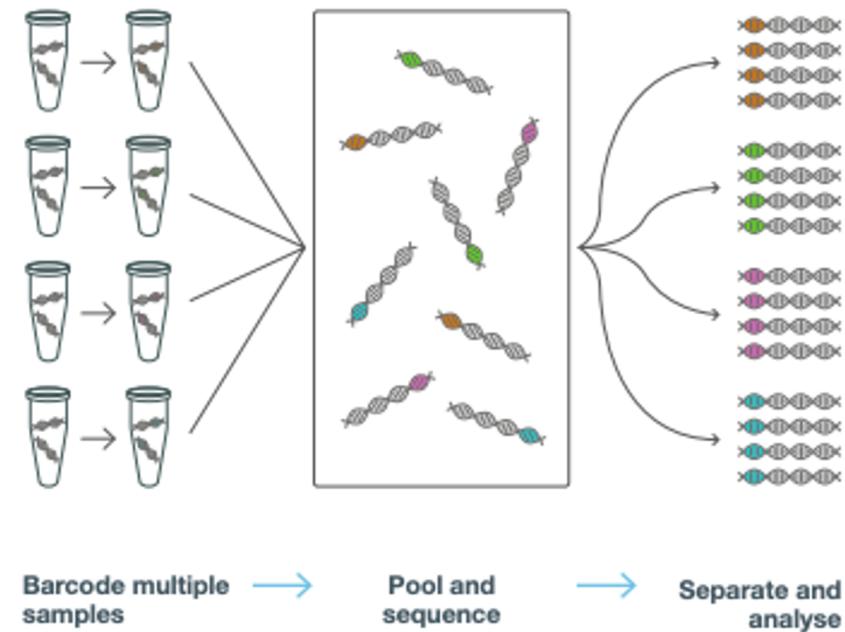


# Trimming and demultiplexing

What does a read look like?

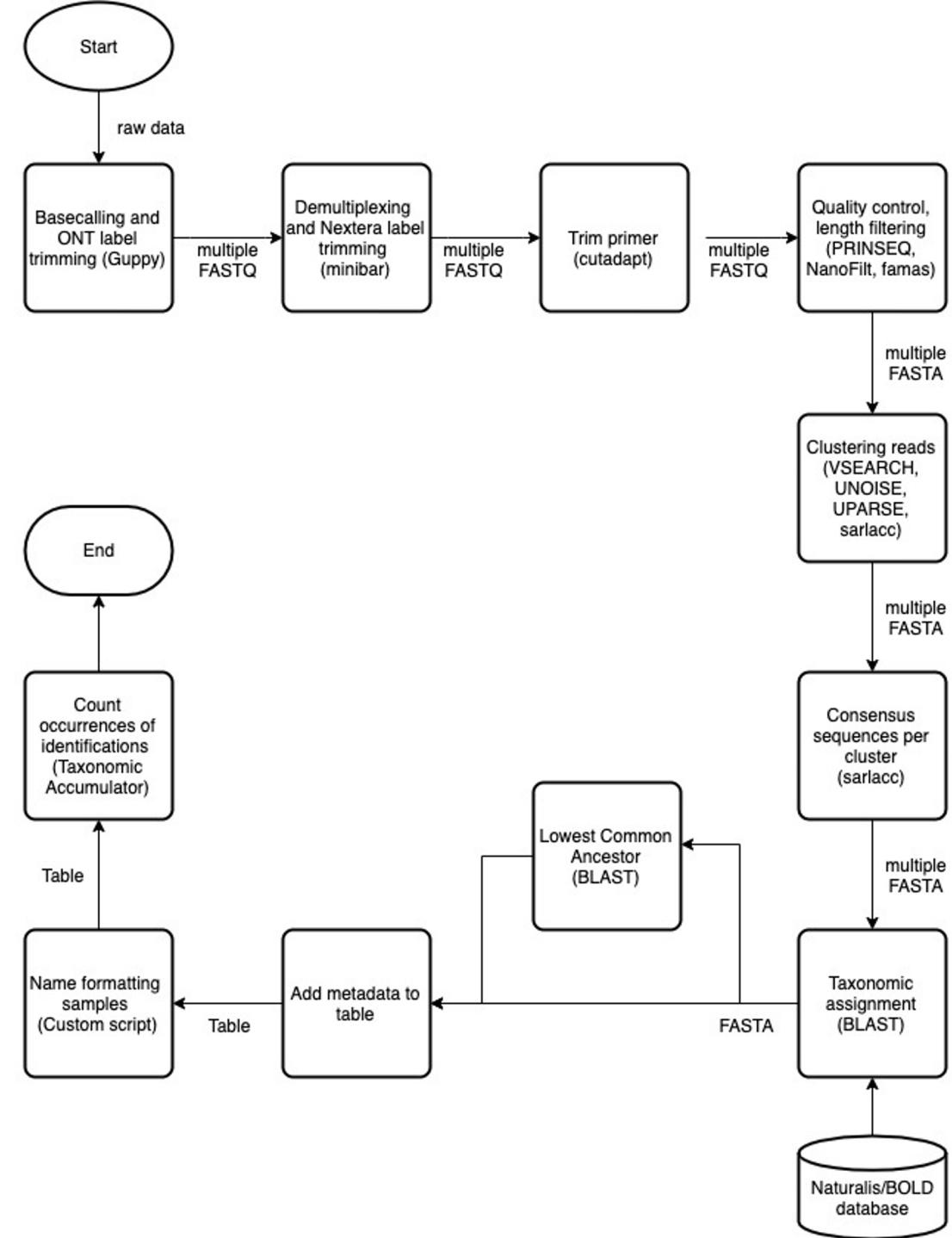
ONT adapter	Illumina chip adapter	Forward sample labels	Forward primer	Region of interest (COI gene)	Reverse primer	Reverse sample labels	Illumina chip adapter	ONT adapter
19 bp	8 bp	24 bp	424 bp		29 bp	8 bp	19 bp	

Demultiplexing: determining what sample a read belongs to

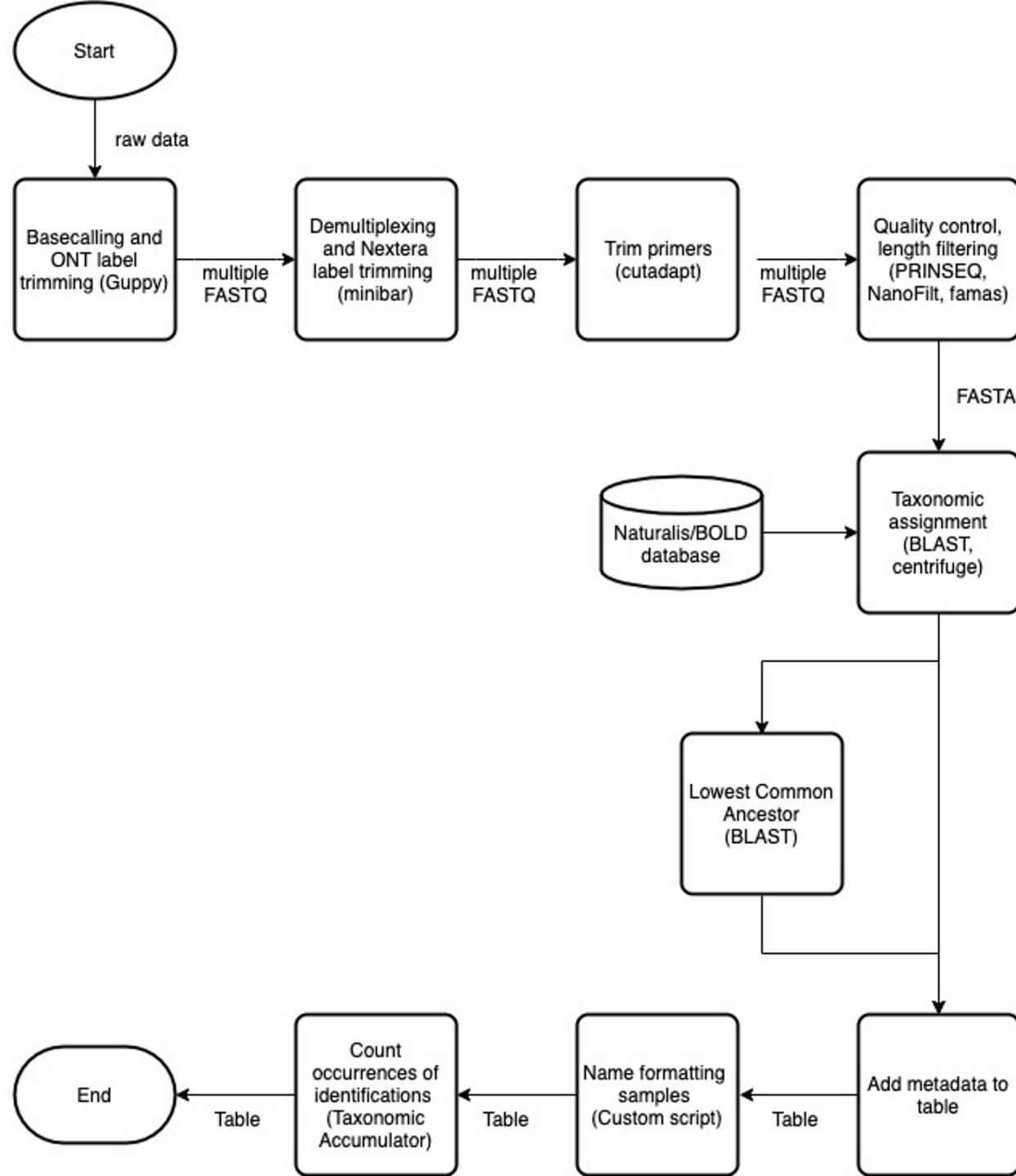


# Conceptual workflow

- Test and adapt workflow
- Compare outcomes to the MiSeq workflow
- Assemble workflow in Galaxy for future use



# Alternative



# Challenges

- Consensus sequence from clusters
  - instead of picking a (random) representative sequence
  - establish parameters to prevent clustering several species in one cluster
- Database matching
  - establish thresholds

read 1	A	G	C	A
read 2	A	G	C	T
read 3	A	T	C	T
read 4	A	G	C	T
consensus	A	G	C	T

# Final output

- Taxonomy: Blast hits on species level
- Reverse taxonomy (Lowest Common Ancestor)
  - Although not on species level, still useful for ecological analysis

# Planning and timeframe

- November:
  - Literature research and project proposal
- December – January:
  - Meeting with Wageningen University researchers
    - They successfully used metabarcoding for biodiversity assessment in North Sea (using 12s and 16s region)
  - Learn how to use Guppy & basecall data with up to date version
  - Test demultiplexing & trimming scripts
- February – March:
  - Assemble workflow in Galaxy
  - Possibly sequence more samples for more data
- April-May
  - Analyze and compare composition of eDNA signals between watercolums
- May-June
  - Write up project report and journal article

# Questions/Suggestions?