

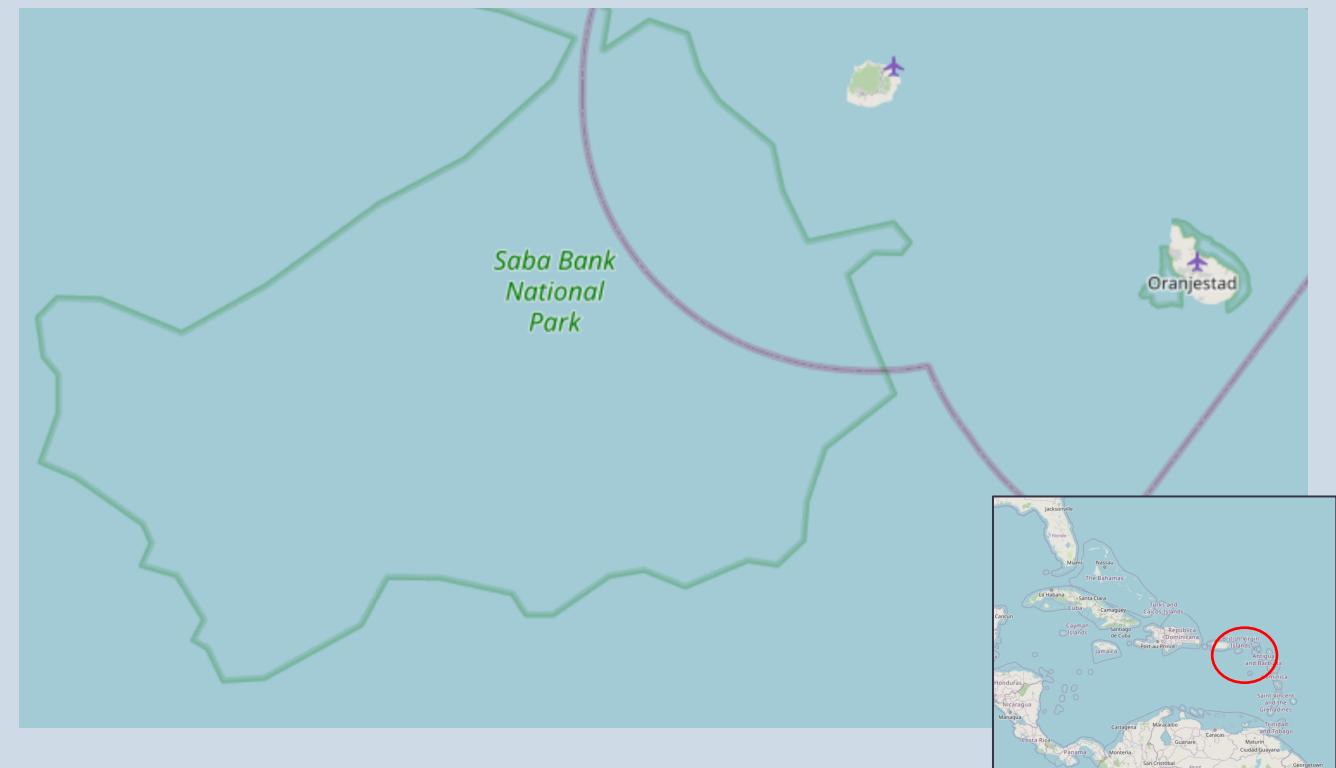


Metabarcoding with MinION for taxonomic identification

& ecological analysis of data from Saba and St. Eustatius expeditions

Background

- Internship as part of MSc Biology: Molecular Genetics & Biotechnology
- Project done with supervision/guidance by
 - Arjen Speksnijder
 - Rutger Vos
 - Elza Duijm
 - Dick Groenenberg



Research Question

Metabarcoding:

- How to demultiplex, cluster, and make consensus sequences for MinION pooled amplicon sequencing data, given the higher error rate and longer sequence length?

Environmental Analysis

- Does exchange between species take place across the Saba Bank and between the Saba Bank and St. Eustatius?

Metabarcoding Environmental DNA

- Metabarcoding: DNA based approach for taxonomic identification of mixed samples using high throughput sequencing.
- COI gene as barcode
 - Mitochondrial cytochrome c oxidase I
 - other studies may use ITS, 12S, 16S.
- Barcoding region requirements:
 - Similar enough so DNA can be amplified with a universal primer.
 - Diverged enough to allow species identification
- BOLD database as reference

Why Minion



Small and portable

Not constrained by a lab



Cheap (\$1000 startup cost)



Long reads

No merging forward and reverse read required

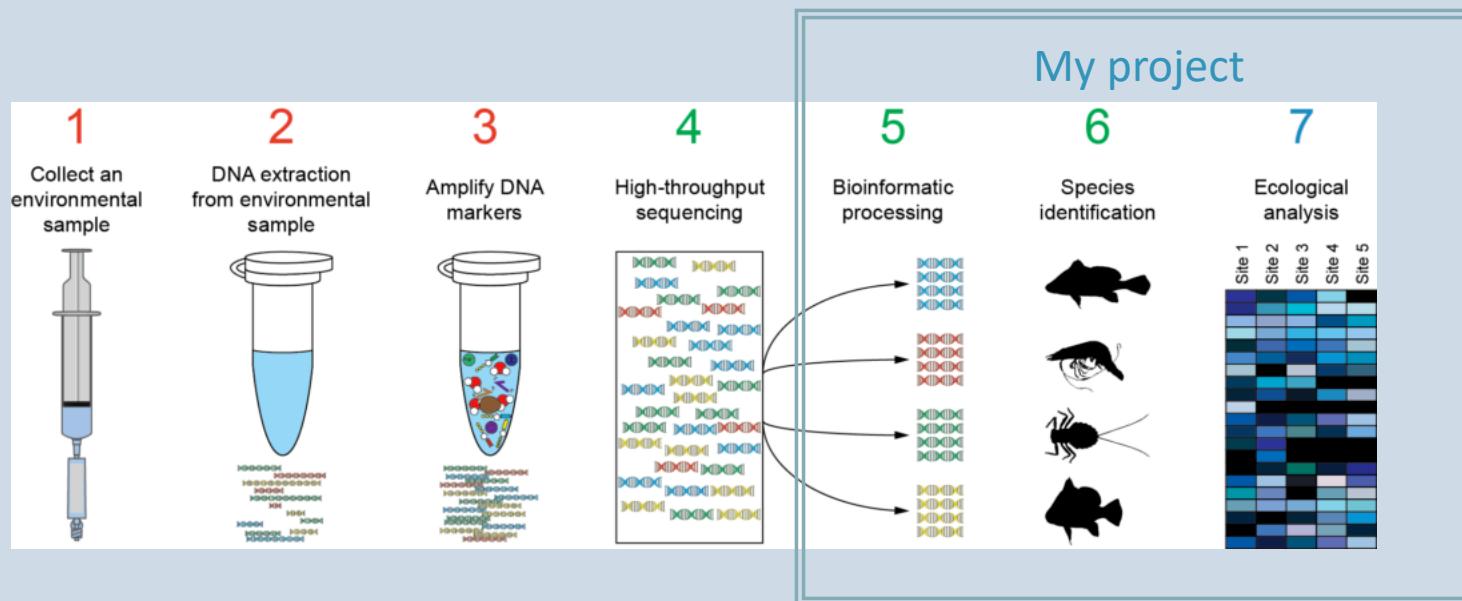


Emerging technology

rapidly improving

Sampling

- Water samples taken during expeditions to St Eustatius (2015) and Saba bank (2018)
- Water samples taken at several locations and depths.
- eDNA extraction, PCR and sequencing in Naturalis laboratory



Results

Basecallers & error rate

Demultiplexing

Clustering and consensus

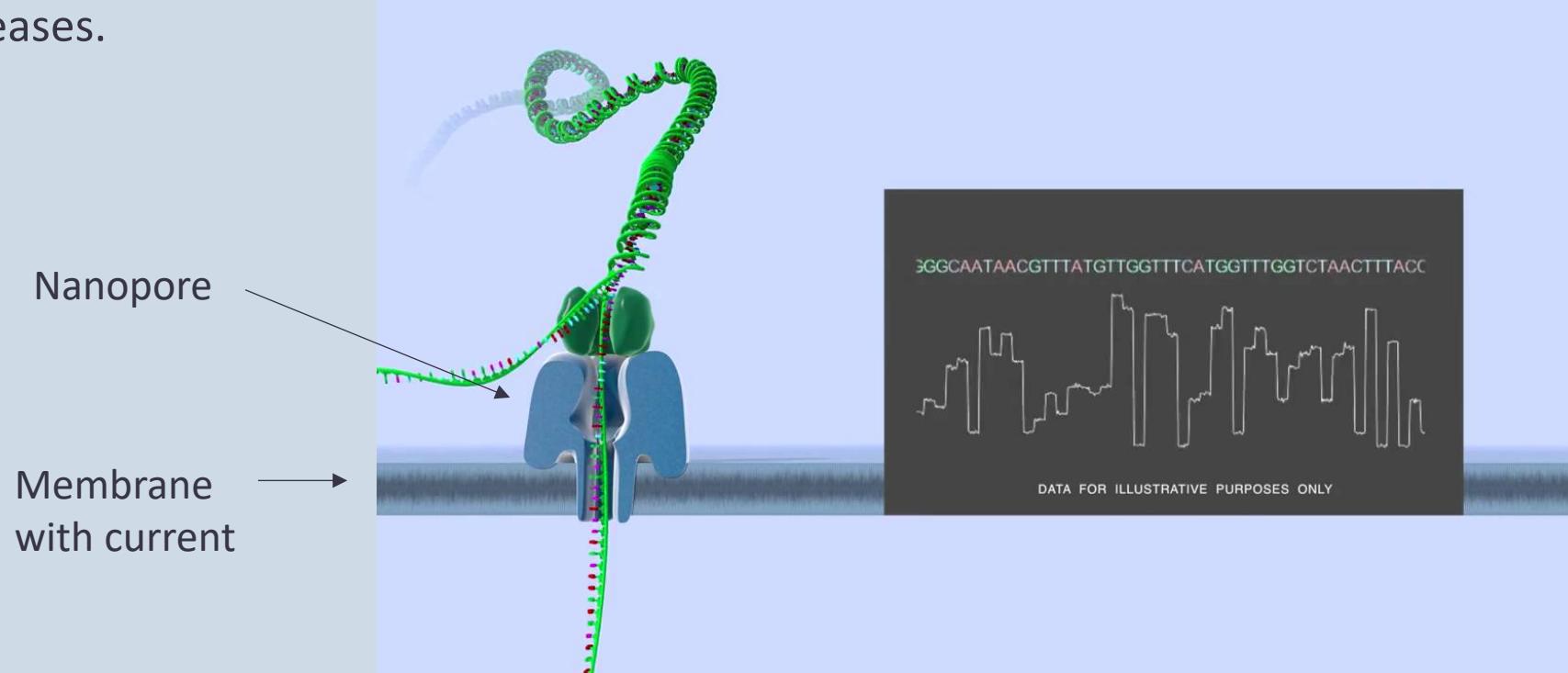
Taxonomic Identifications

Ecological analysis (with Illumina data)

Basecaller & Error Rate

Quick explanation:

- Basecalling = converting the current disruptions generated by the passing of the DNA strand through the nanopore to a sequence of nucleotides (ACTG).
- Oxford Nanopore Technologies periodically releases new and improved basecallers. And also provides research releases.



Basecallers & Error rate

- Basecallers tested

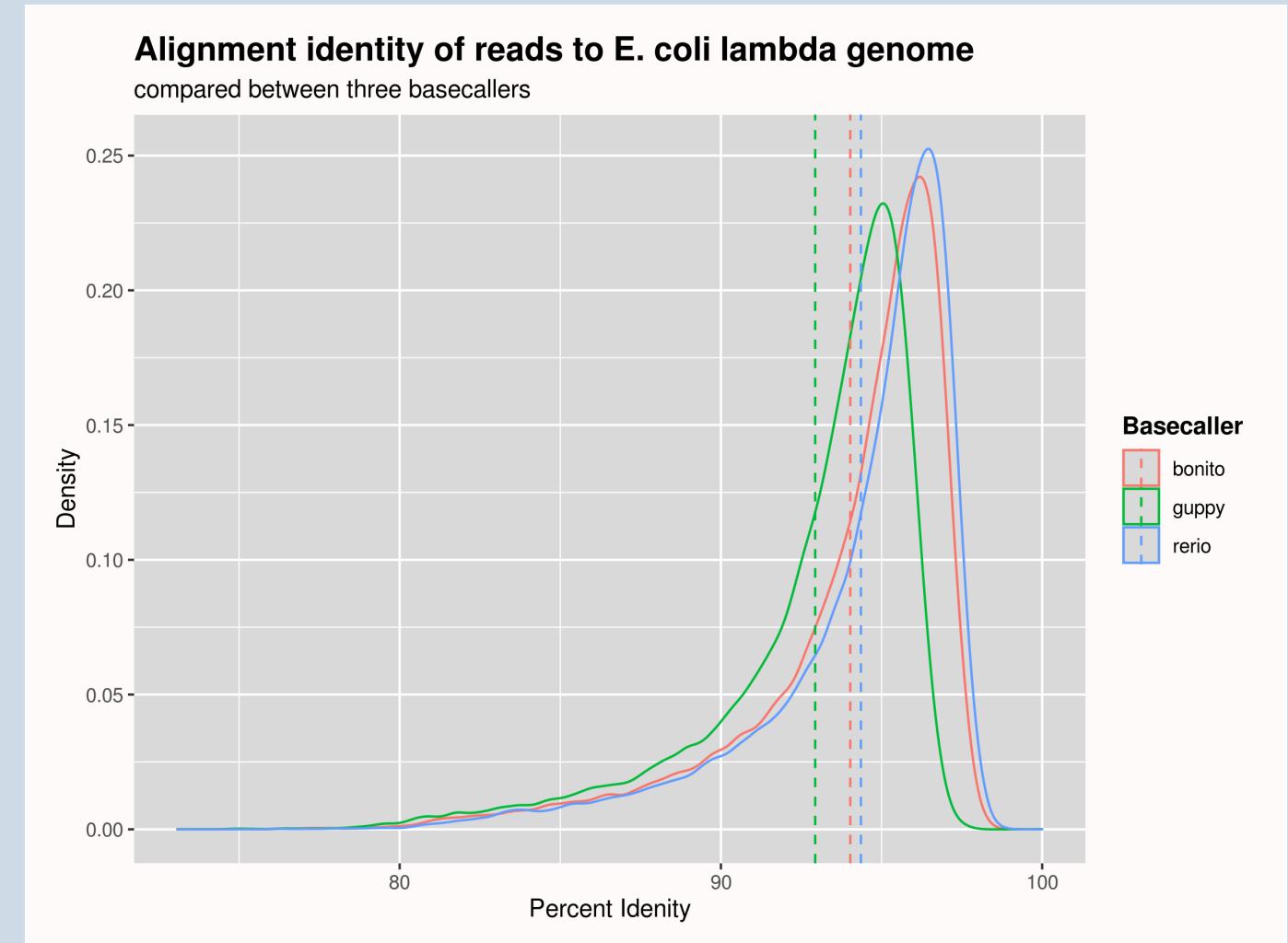
Guppy (Current version)

Bonito (Research Release,
March 2020)

Rerio (Research Release,
April 2020)

- Estimate error rates by aligning to E. coli lambda genome

```
Mode guppy:    94.83
Mode bonito:   96.26
Mode rerio:    96.89
```



Demultiplexing

With Minibar

Task:

Separating the samples based on dual Illumina tags

- 7bp tags were used
- Edit distance between tag as low as 2

Hypothesis:

Problems due to lower quality of nanopore data.

- Errors may cause mis- or no identification.

Result:

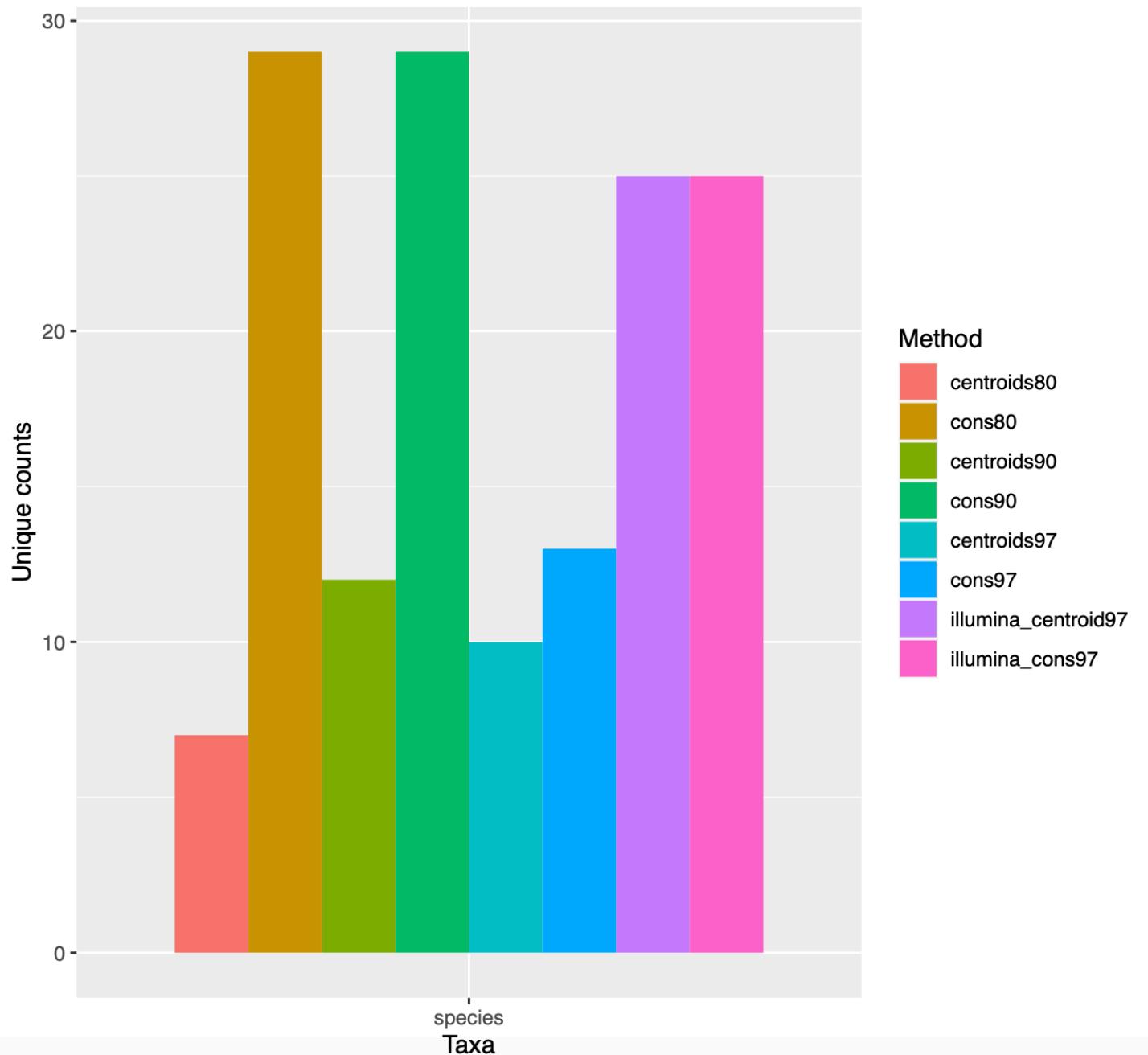
Not satisfactory. Distinguishes few samples with certainty.

- Solution would be to use Oxford Nanopore tags, or design longer tags.

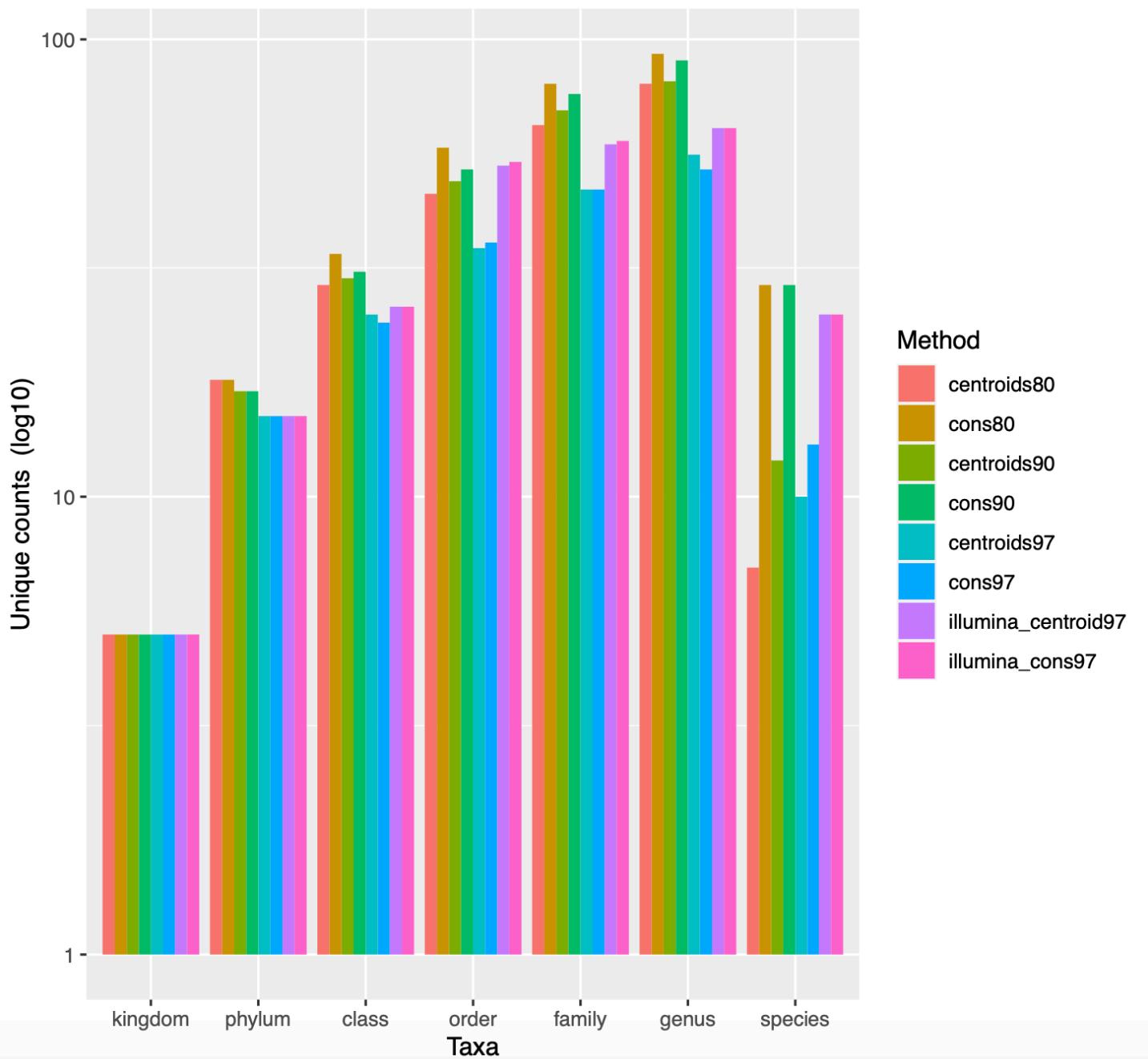
Clustering and consensus

- Clustering with vsearch
- Vsearch can create consensus from clusters
- Tested at different clustering percentages
- Count number of lca identifications and compare against centroid.

Number of lowest common ancestors per taxonomic level
excluding cluster sizes up to 1



Number of lowest common ancestors per taxonomic level
excluding cluster sizes up to 1



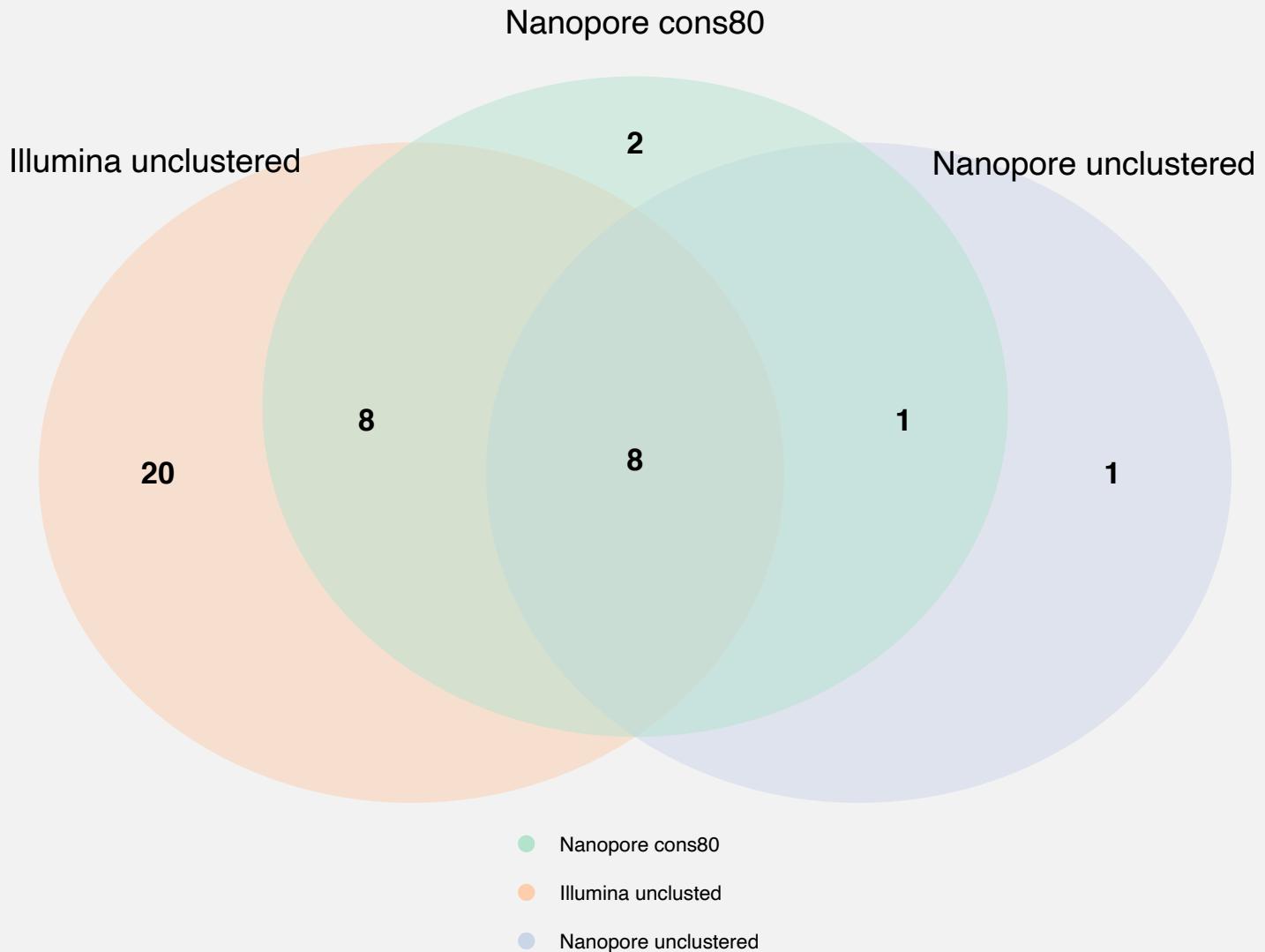
Taxonomic identifications

Consensus looks better, but does it also lead to better identification?

Venn diagram to look at overlap of identifications between methods

species

overlap between unique identifications at species level, not filtered on cluster size



Taxonomic identifications

Found on genus level in unclustered data

nano_cons80_lca90 (2):
Rhopalonema velatum, Centropages violaceus

rawilluminalca_90 (20):
Limacina inflata, Eurythenes gryllus, Diadema antillarum, Ptychodera flava, Asymmetron lucayanum, Coryphopterus venezuelae, Amphinome rostrata, Lensia campanella, Aglaura hemistoma, Meoma ventricosa, Lytechinus variegatus, Euchaeta marina, Pleuroamma abdominalis, Porroecia spinirostris, Gymnothorax moringa, Ophiocoma echinata, Cyclothona sp., Apodemus sylvaticus, Linopneustes longispinus, Nezumia aequalis

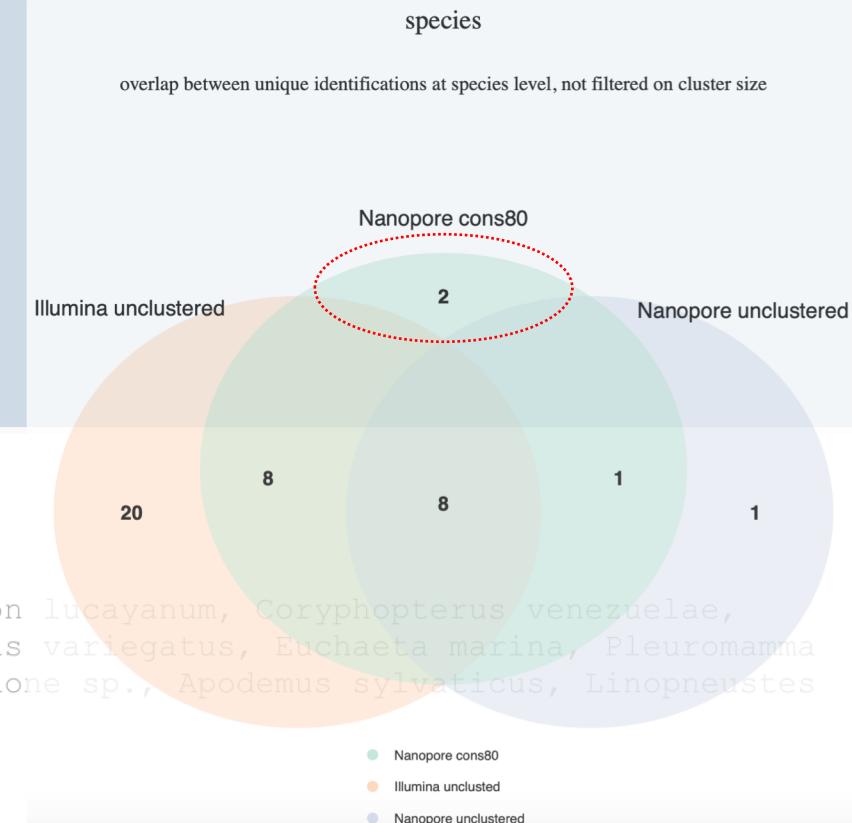
rawnanopore_lca90 (1):
Acanella eburnea

nano_cons80_lca90 - rawilluminalca_90 (8):
Homo sapiens, Nanomia bijuga, Chaunax pictus, Aspidosiphon laevis, Delibus nudus, Clausocalanus furcatus, Gammarus pulex, Cephalopholis fulva

nano_cons80_lca90 - rawnanopore_lca90 (1):
Rosacea sp.

rawilluminalca_90 - rawnanopore_lca90 (0):

nano_cons80_lca90 - rawilluminalca_90 - rawnanopore_lca90 (8):
Epinephelus guttatus, Lucicutia flavicornis, Cordagalma cordiforme, Euphausia tenera, Dasybranchus sp. DH1, Raphanus sativus, Athorybia rosacea, Rhizophysa filiformis



Taxonomic identifications

nano_cons80_lca90 (2):
Rhopalonema velatum, Centropages violaceus

rawilluminalca_90 (20):
Limacina inflata, Eurythenes gryllus, Diadema antillarum, Ptychodera flava, Asymmetron lucayanum, Coryphopterus venezuelae, Amphinome rostrata, Lensia campanella, Aglaura hemistoma, Meoma ventricosa, Lytechinus variegatus, Euchaeta marina, Pleuromamma abdominalis, Porroecia spinirostris, Gymnothorax moringa, Ophiocoma echinata, Cyclothona sp., Apodemus sylvaticus, Linopneustes longispinus, Nezumia aequalis

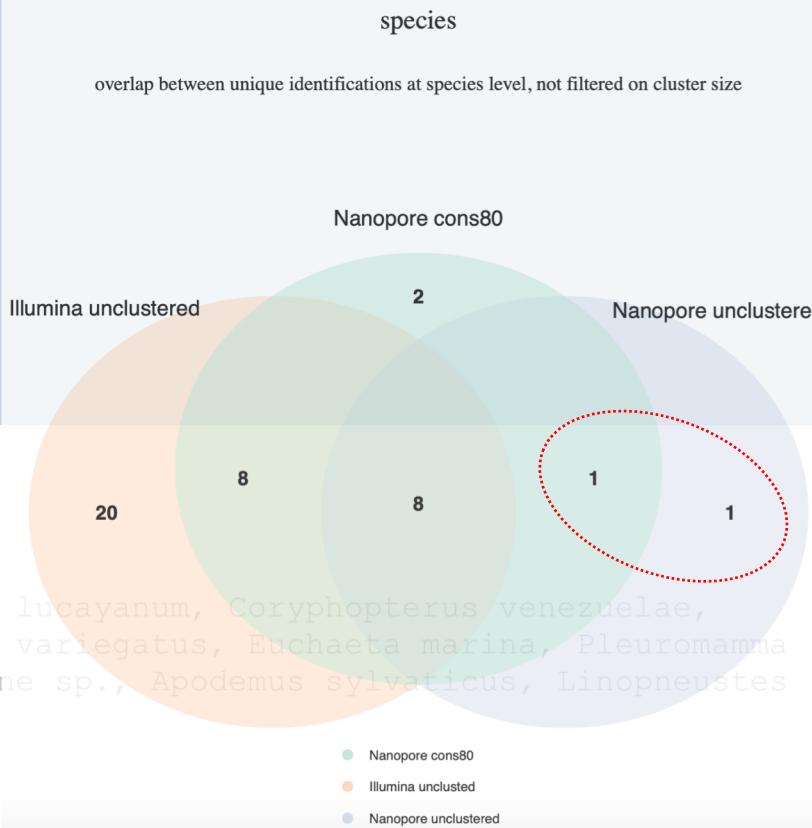
rawnanopore_lca90 (1): **Probably false positive:** read has other good blast hits,
Acanella eburnea ← and not found in illumina on genus or species level

nano_cons80_lca90 - rawilluminalca_90 (8):
Homo sapiens, Nanomia bijuga, Chaunax pictus, Aspidosiphon laevis, Delibus nudus, Clausocalanus furcatus, Gammarus pulex, Cephalopholis fulva

nano_cons80_lca90 - rawnanopore_lca90 (1):
Rosacea sp. ← **Likely a true hit:** high quality reads, not a chimera, several reads in data

rawilluminalca_90 - rawnanopore_lca90 (0):

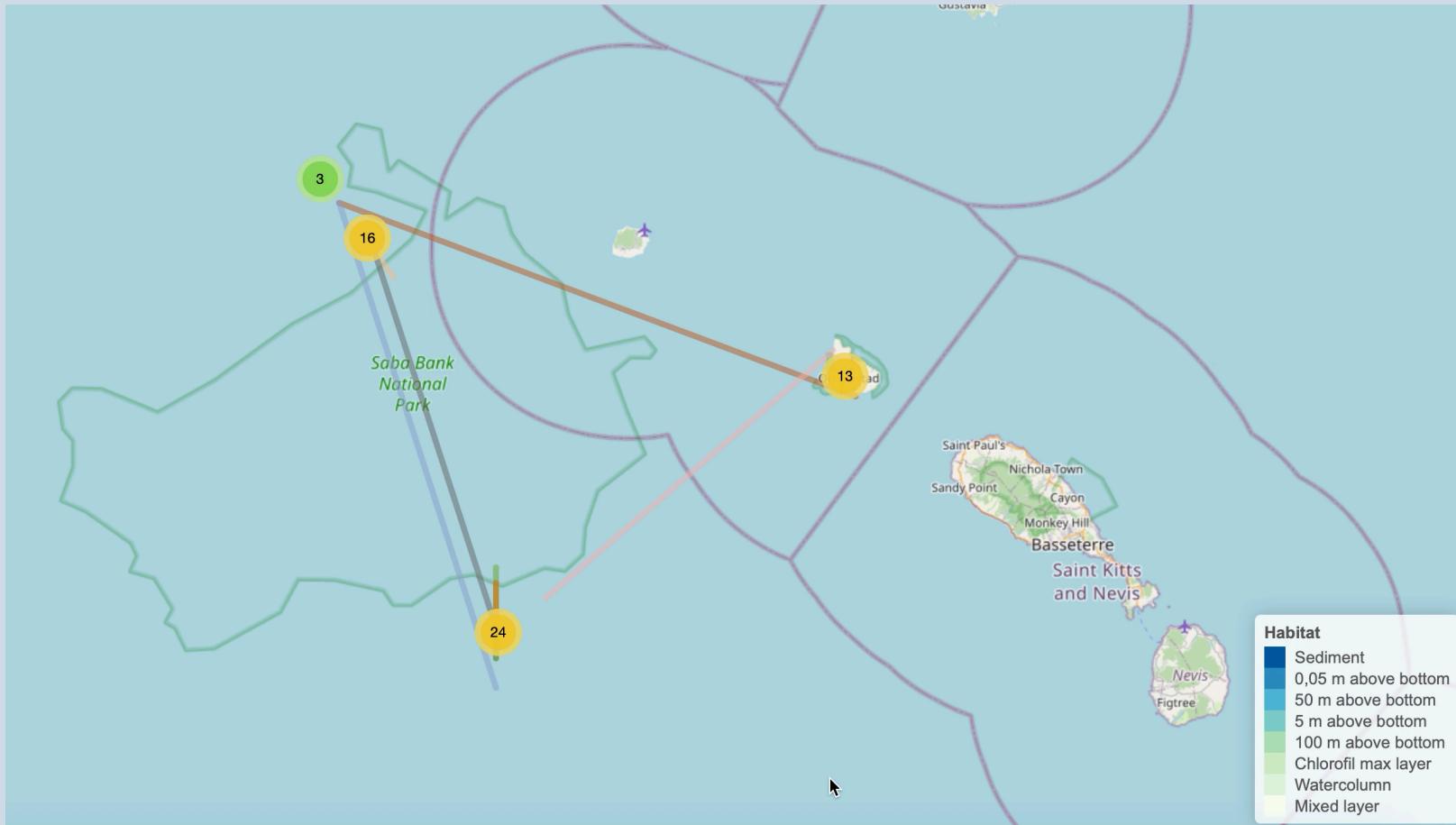
nano_cons80_lca90 - rawilluminalca_90 - rawnanopore_lca90 (8):
Epinephelus guttatus, Lucicutia flavicornis, Cordagalma cordiforme, Euphausia tenera, Dasybranchus sp. DH1, Raphanus sativus, Athorybia rosacea, Rhizophysa filiformis



Ecological analysis

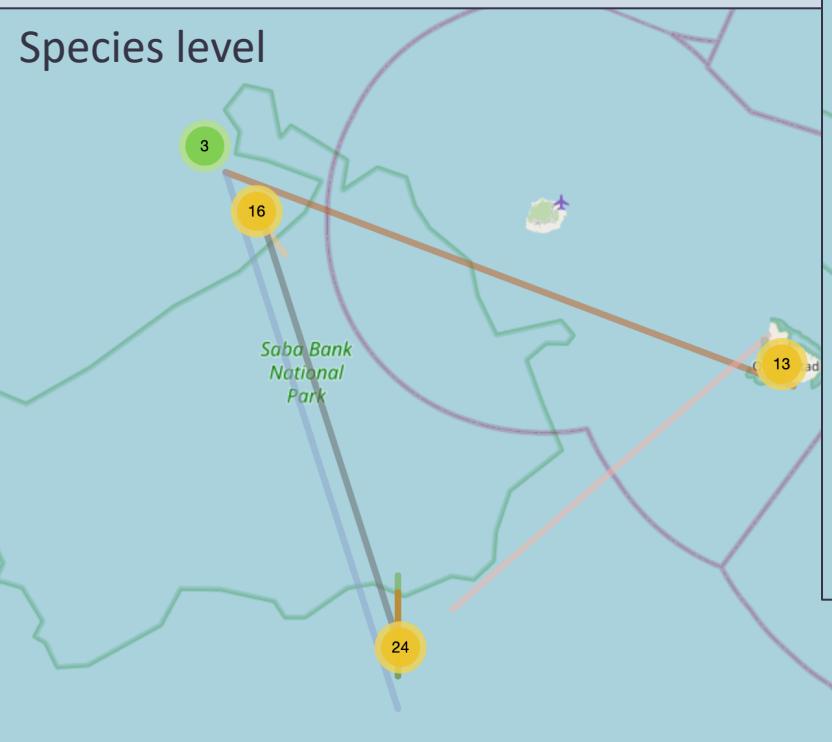
- Not enough data to do reliable statistical analysis
- Enough data to visually discover patterns
- Interactive html map
 - Samples plotted on map, with species and information
 - Created with leaflet (R package)

Ecological analysis



Ecological analysis

Species level



Genus level



Family level



Conclusion

- Metabarcoding with Nanopore data:
 - Data quality improves after making a consensus sequence from a cluster
 - Cannot be demultiplexed based on illumina tags
- Ecological Analysis
 - Visual patterns of exchange across Saba Bank

Questions?



Images harvested by Naturalis script