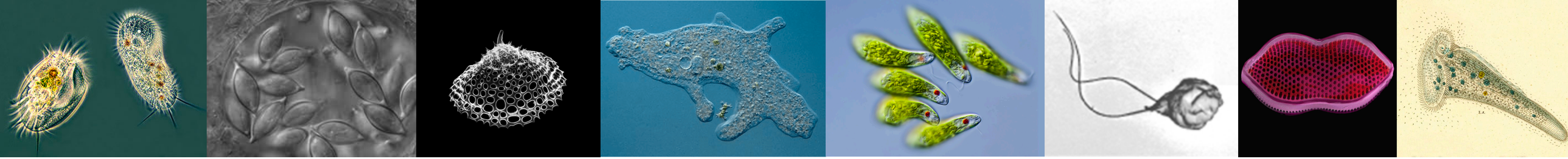


Long-read metabarcoding

Bioinformatics for Environmental Sequencing (DNA metabarcoding)

20th April 2023

Mahwash Jamy

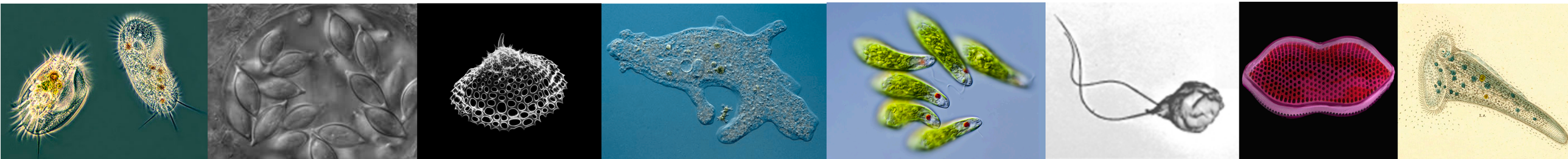
@JamyMahwash

Your lecturer for the next hour or so

- Postdoc at SLU and NHM Oslo
- Before that, PhD at Uppsala University

My research:

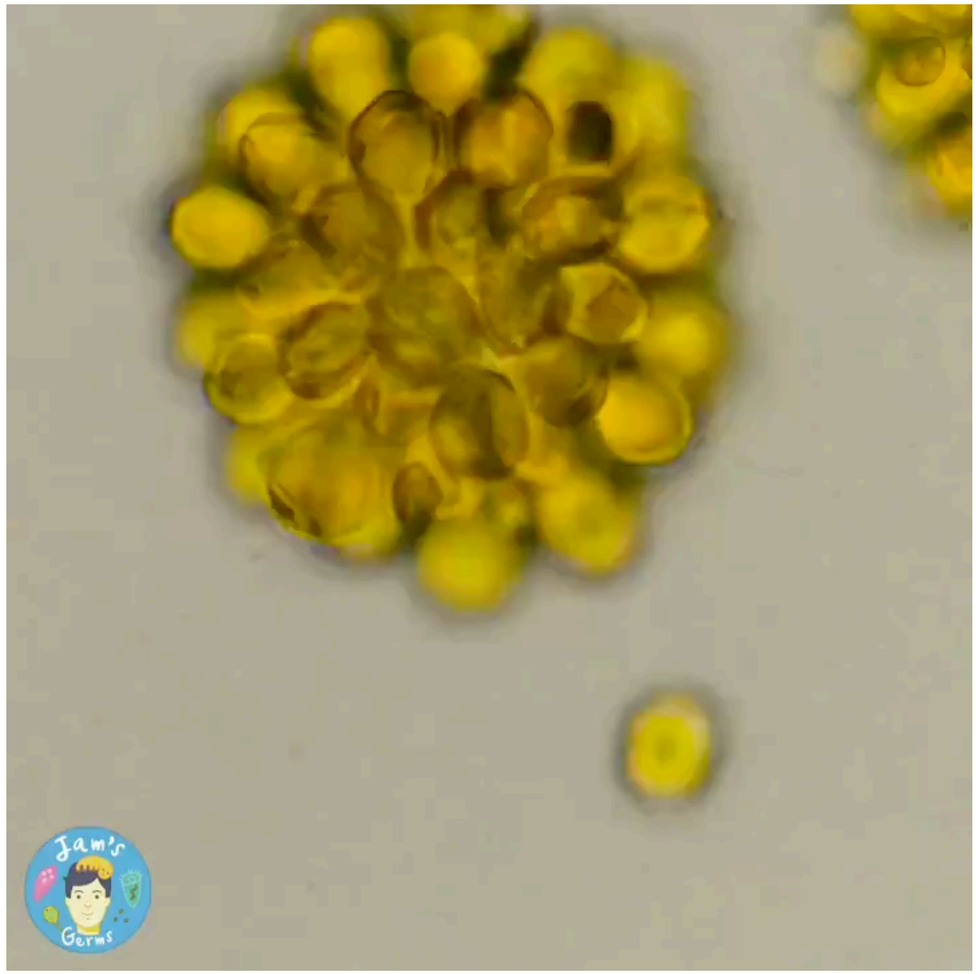
- Metabarcoding (particularly long-read)
- Phylogenetics
- Protist ecology and evolution



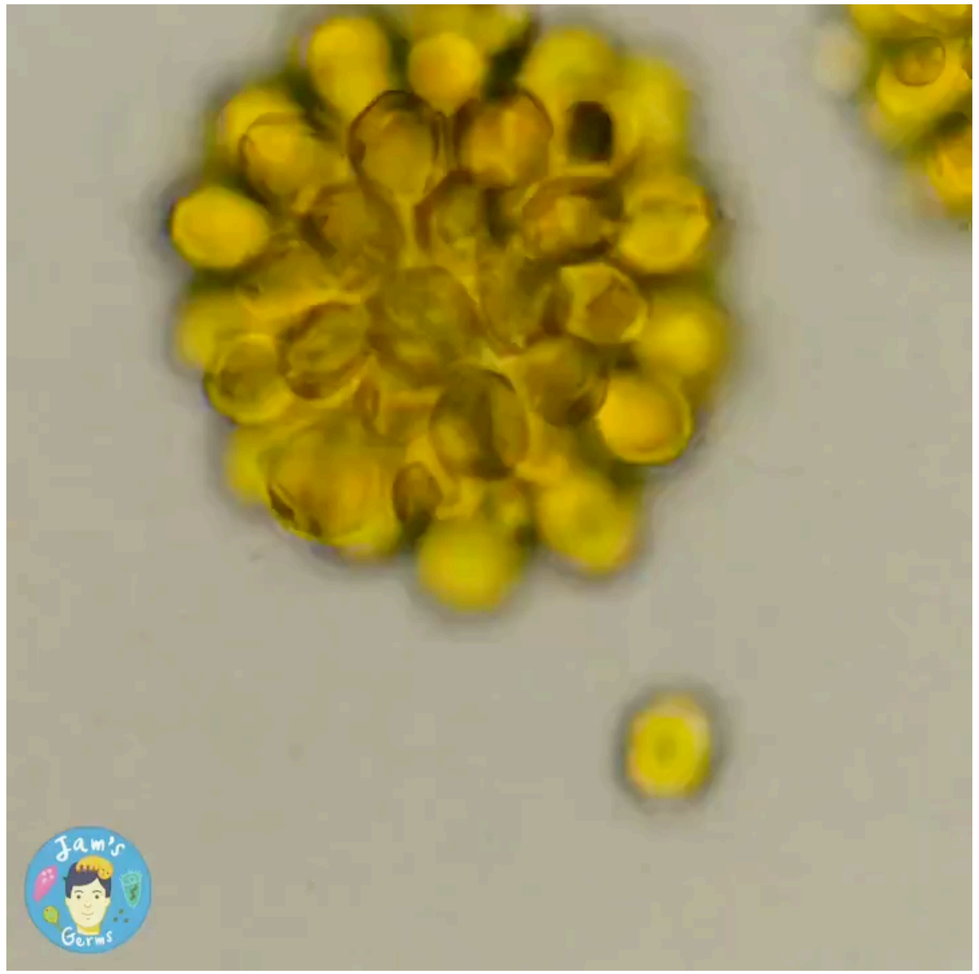
Outline

- What is long-read metabarcoding?
- Why do long-read metabarcoding? Some advantages and limitations.
- Generating long-read amplicons
- Which sequencing platform should I use?
- Process long-read data

Any questions?



Any questions?

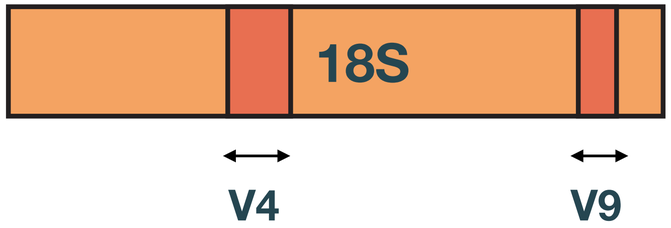


What is long-read metabarcoding?

Metabarcoding

Environmental sample

(Or fragment of 16S, ITS1 region,
rbcl, COI, or other)

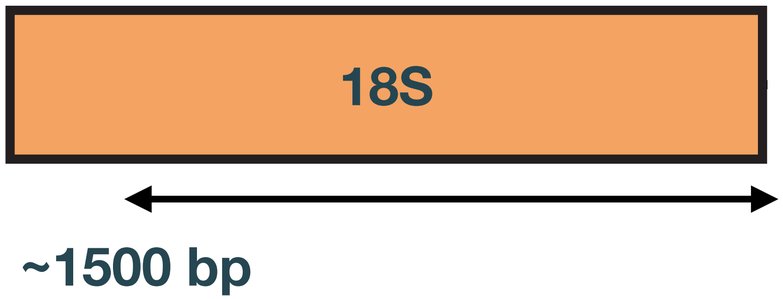


Sequence > 500 bp
fragment with
Illumina

Bioinformatic analyses

DADA2, VSEARCH, Swarm

Long-read metabarcoding

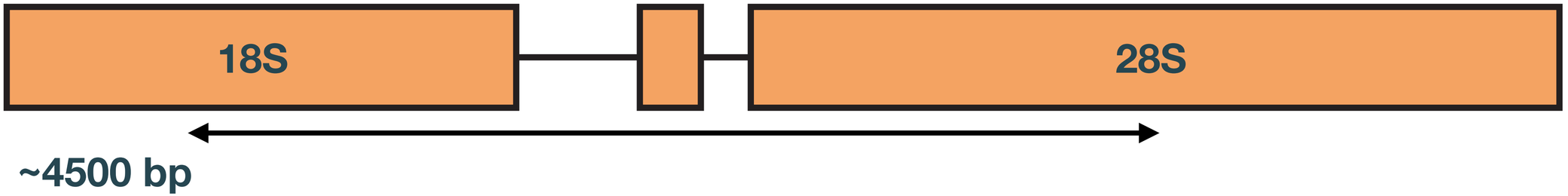


Other single gene markers:

- 16S
- ITS region
- COI

Other multi-gene markers:

- 16S-23S
- SSU-ITS region
- ITS region-LSU
- 12S-16S



Generally lower-throughput than Illumina sequencing

PACBIO®

Oxford NANOPORE Technologies

Increasing in popularity

- Error rates have decreased a lot and continue to decrease: from 25% error rate in early days to close to 100% accuracy (in some cases) now.

- Costs are decreasing rapidly (though currently more expensive and lower throughput than Illumina)

Sequencing technology (specifications)	Instrument cost	Library prep cost	Cost per lane/cell	Cost per million reads	Cost per billion bases	Runtime (h)	Throughput (k reads at maximum runtime)
PacBio Sequel I (maximum 8 cells)	300,000	150–400	1,000	2,000	100	10–20	500
PacBio Sequel II (maximum 16 cells)	650,000	75–400	2,000	500	17	10–30	4,000

Tedersoo et al, 2021. *Applied and Environmental Microbiology*

Why do long-read metabarcoding?

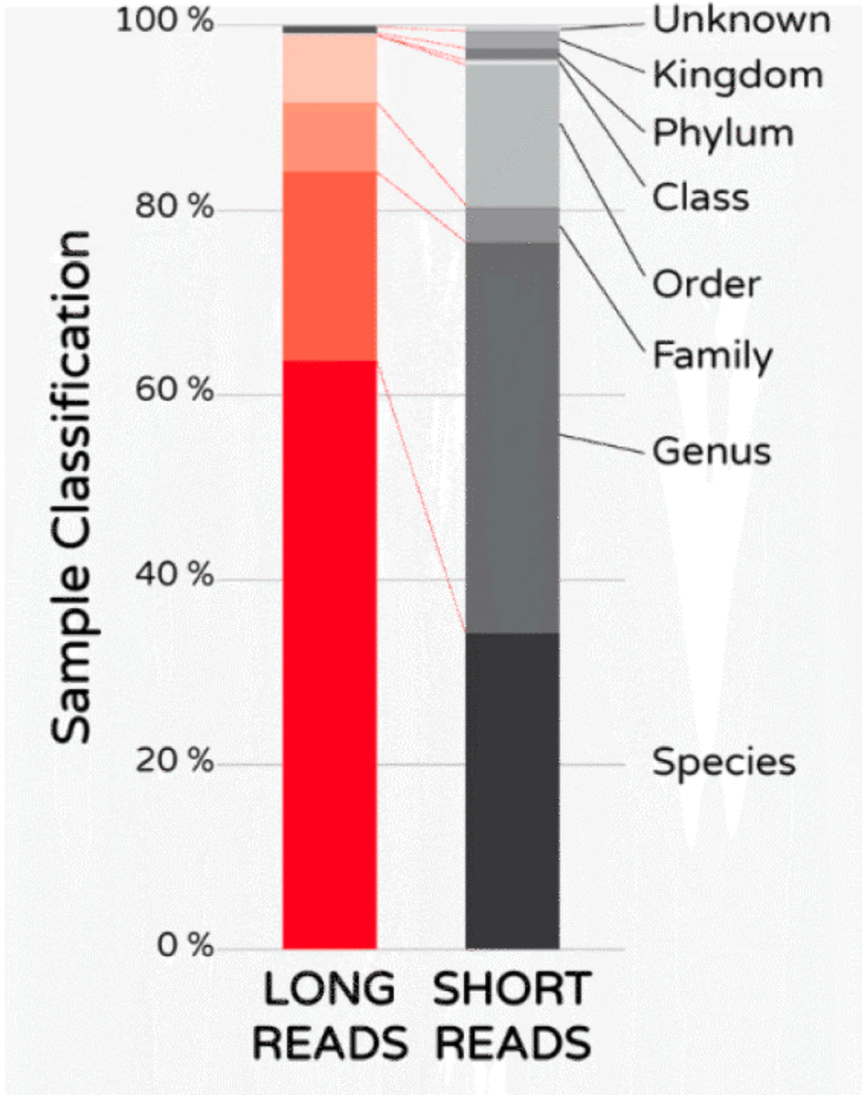
Long reads = increased phylogenetic and taxonomic information

Some advantages of long-read metabarcoding

Some advantages of long-read metabarcoding

- **Better taxonomic classification (“who is there?”)**

Better taxonomic classification

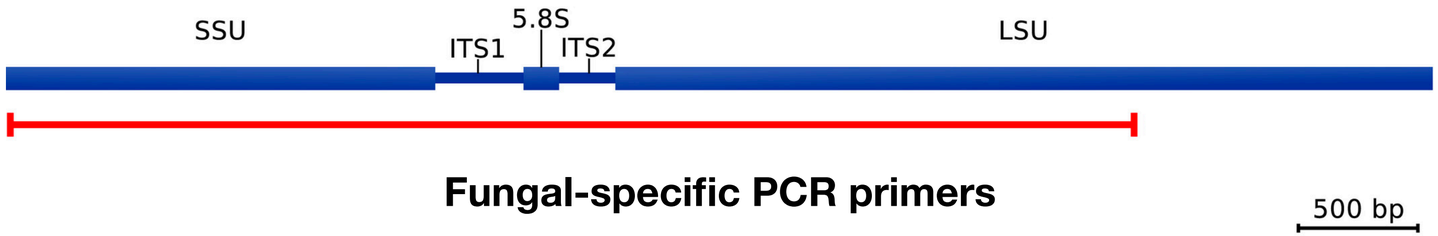


16S metabarcoding of prokaryotic community

More nucleotide data to tease apart closely related organisms

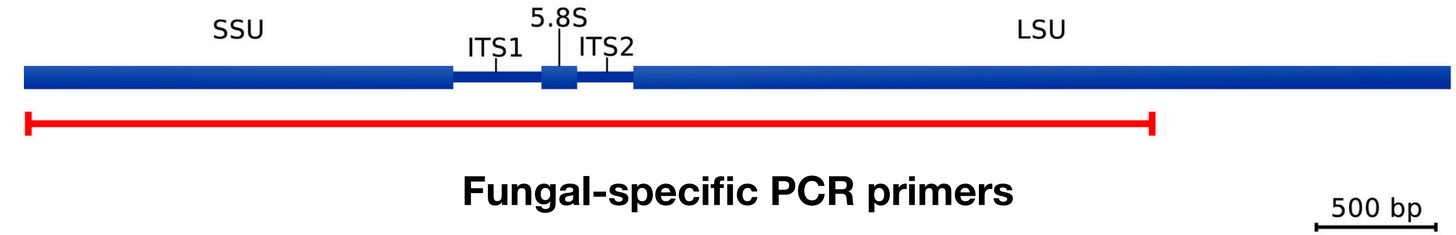
https://www.elementbiosciences.com/applications/full-length-16s-microbiome-classification

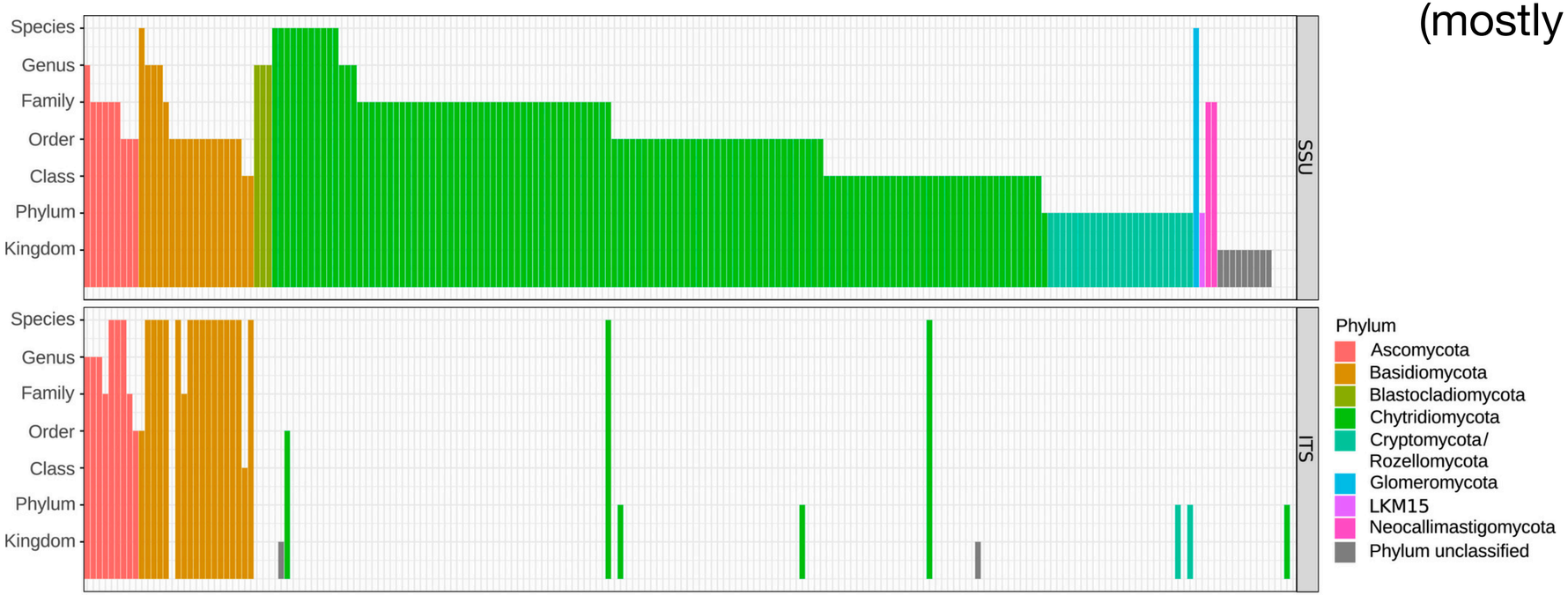
Better taxonomic classification



Heeger et al. 2018, Molecular Ecology Resources

Better taxonomic classification





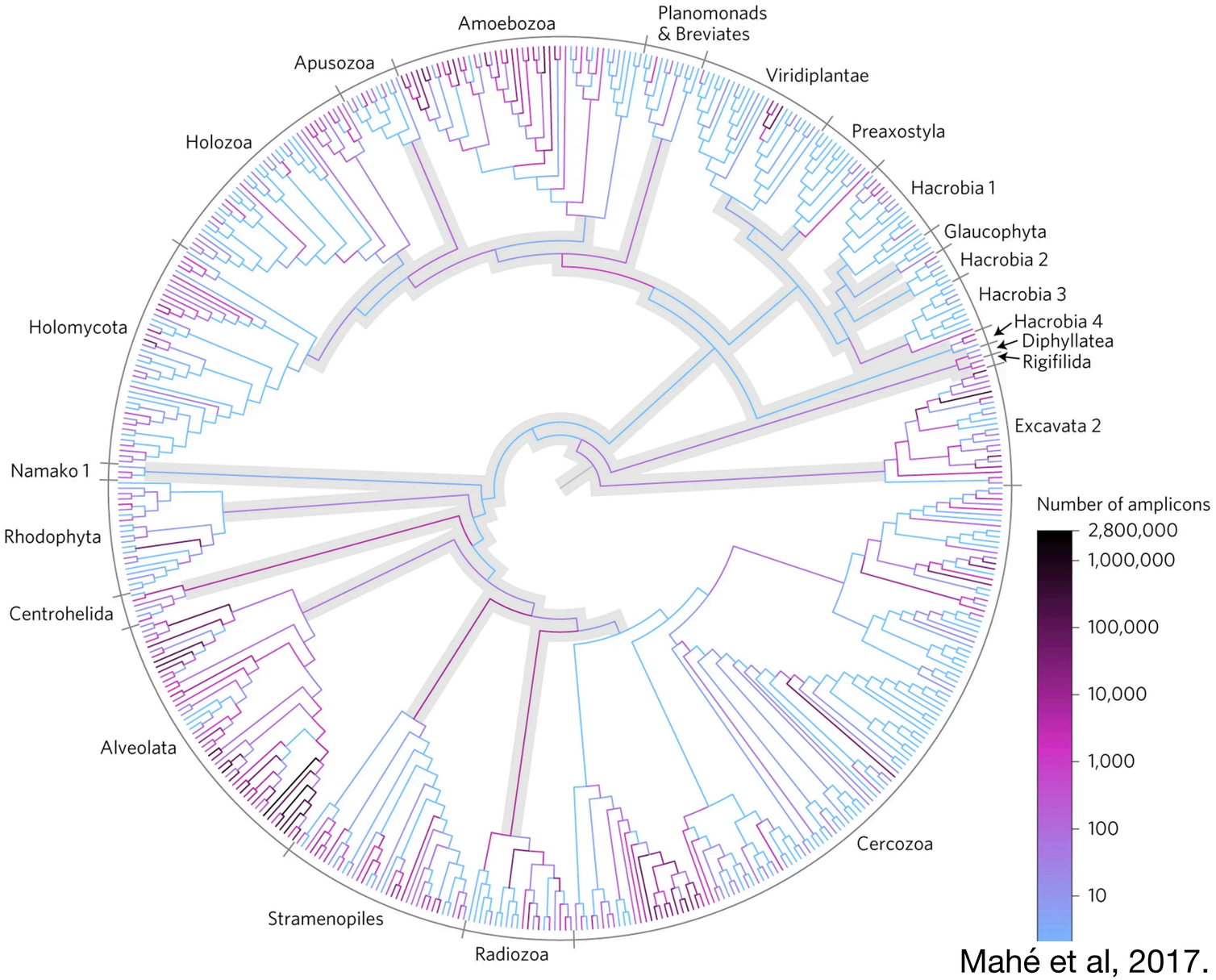
- Able to use different reference databases (UNITE, SILVA, RDP LSU)
- Able to classify OTUs that are not included in ITS and LSU databases (mostly non-Dikarya)

Heeger et al. 2018, Molecular Ecology Resources

Some advantages of long-read metabarcoding

- Better taxonomic classification (“who is there?”)
- **Infer more robust phylogenies to answer: “who is there?”**

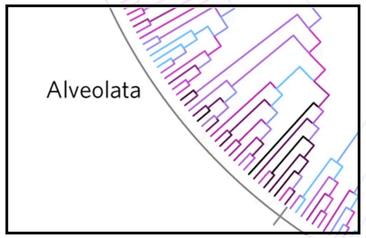
Infer more robust phylogenies to answer “who is there?”

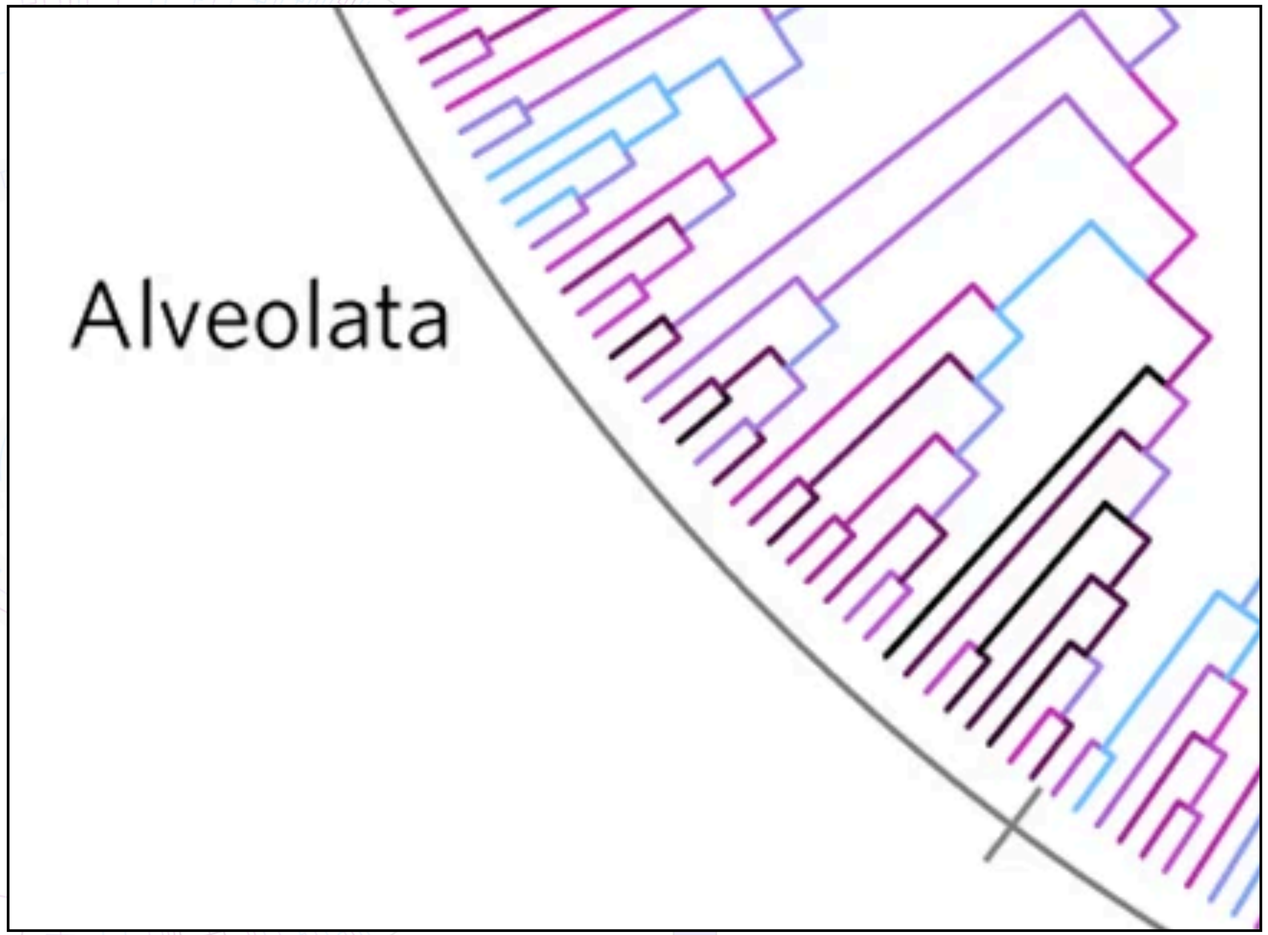


Mahé et al, 2017. *Nature Ecology & Evolution*

Infer more robust phylogenies to answer “who is there?”

- Lot of novel diversity in Apicomplexa
- All from one clade? Several clades?





Mahé et al, 2017. *Nature Ecology & Evolution*

Infer more robust phylogenies to answer “who is there?”

- Longer sequences = increased phylogenetic signal
- Can build multi-gene phylogenies (18S-28S)
- Particularly useful for discovering novel lineages, and investigating their position in the tree of life.

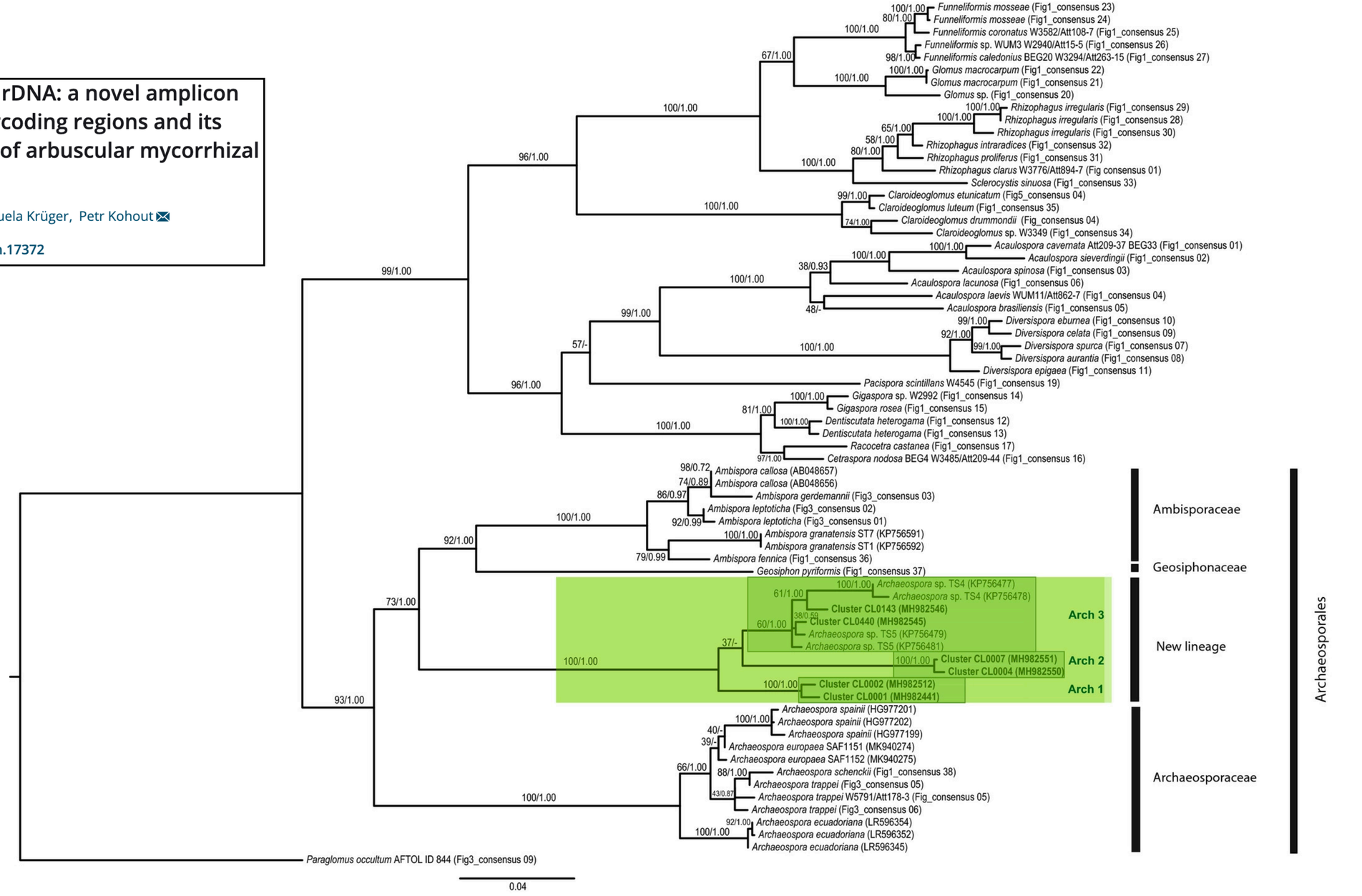
Infer more robust phylogenies to answer “who is there?”

PacBio sequencing of Glomeromycota rDNA: a novel amplicon covering all widely used ribosomal barcoding regions and its applicability in taxonomy and ecology of arbuscular mycorrhizal fungi

Zuzana Kolaříková, Renata Slavíková, Claudia Krüger, Manuela Krüger, Petr Kohout

First published: 29 March 2021 | https://doi.org/10.1111/nph.17372

18S-5.8S-28S tree

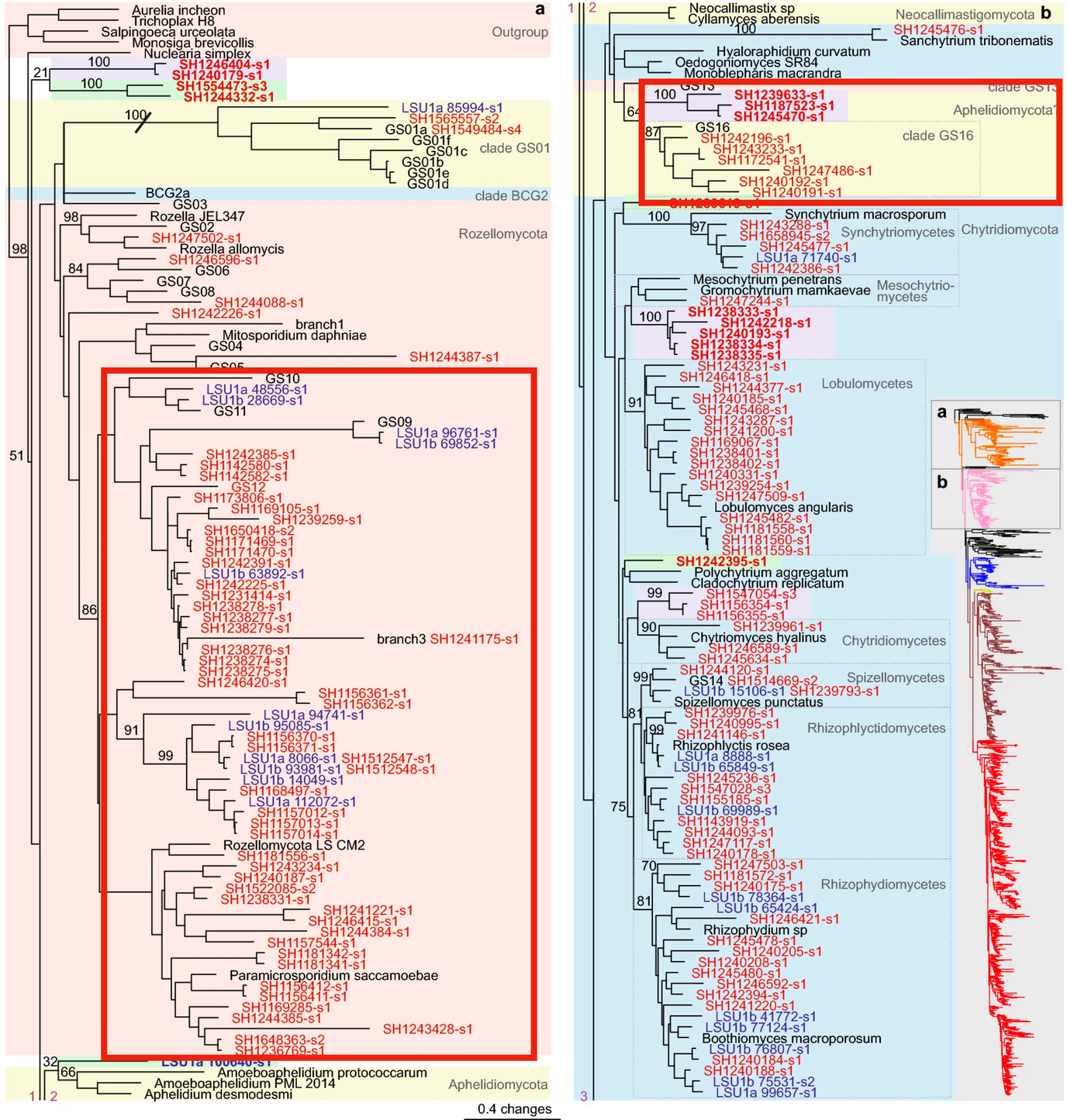


Published: 08 August 2020

Identifying the ‘unidentified’ fungi: a global-scale long-read third-generation sequencing approach

Leho Tedersoo ✉, Sten Anslan, Mohammad Bahram, Urmas Kõljalg & Kessy Abarenkov

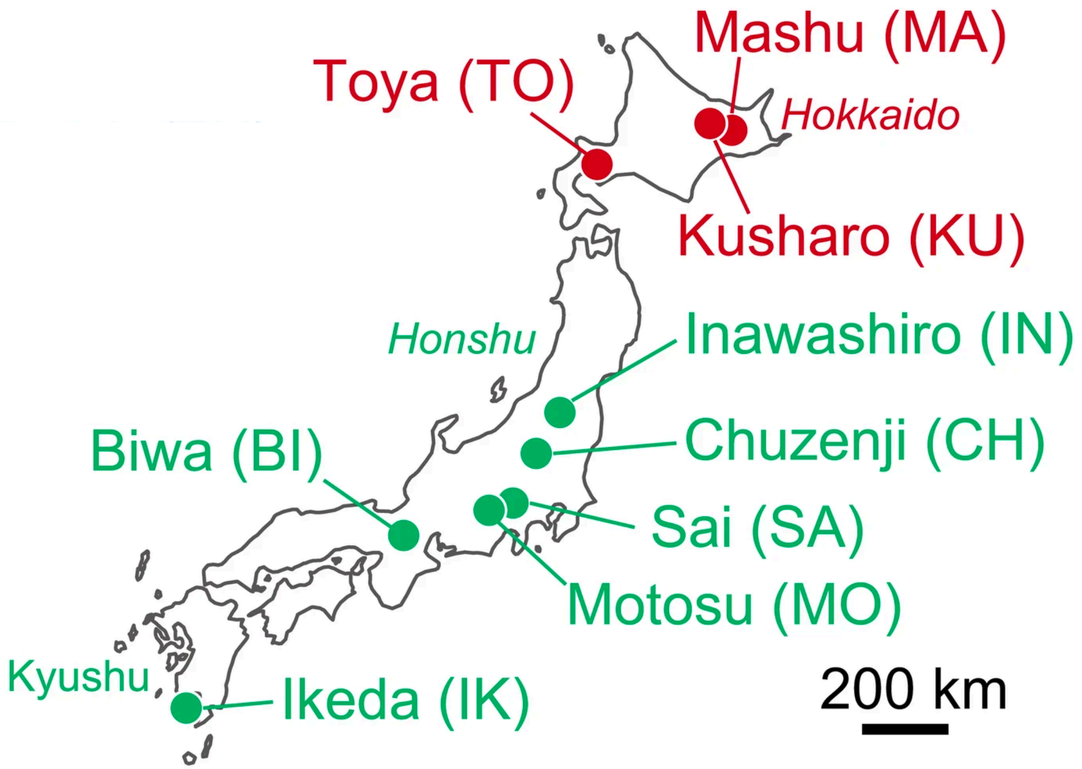
Fungal Diversity **103**, 273–293 (2020) | Cite this article



Some advantages of long-read metabarcoding

- Better taxonomic classification (“who is there?”)
- Infer more robust phylogenies to answer: “who is there?”
- **Higher resolution community profiling**

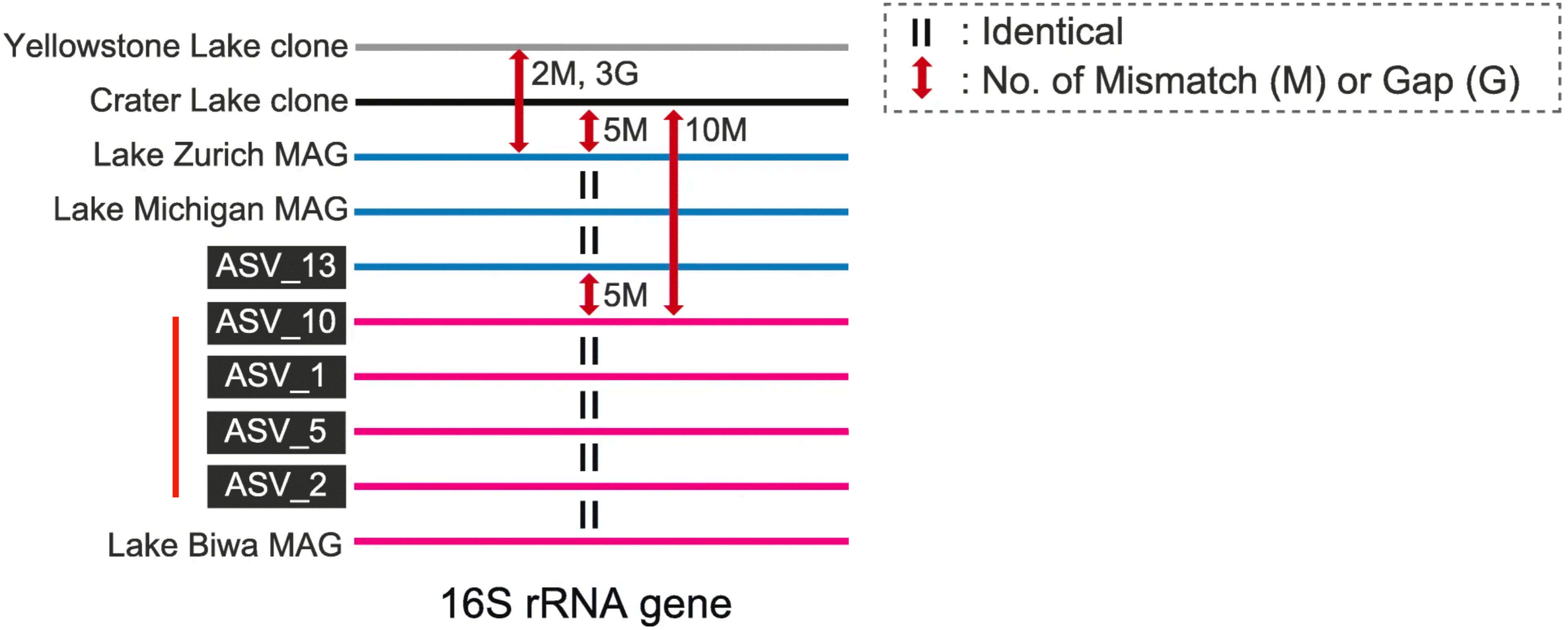
Higher resolution community profiling



Aim:
Compare the bacterial communities of lakes in Japan.
Sequenced 16S and ITS region.

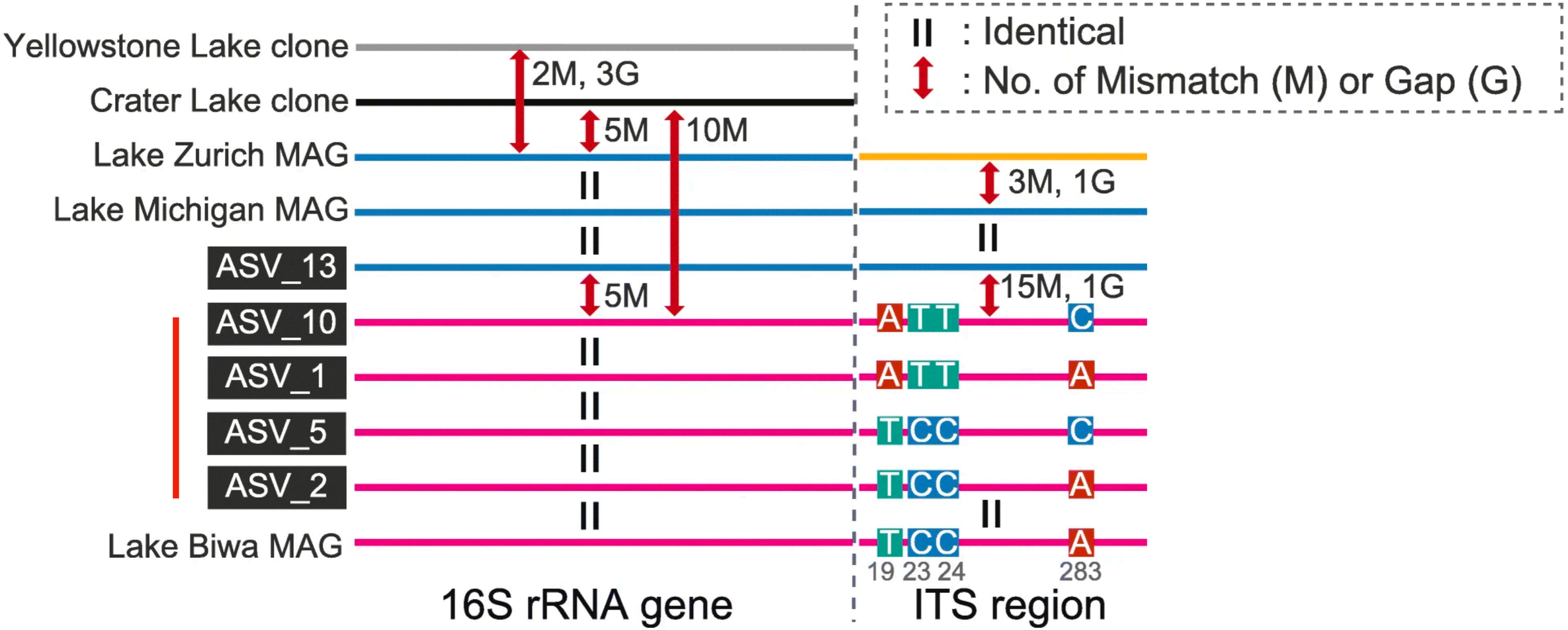
Okazaki et al, 2021. *Microbiome*

Higher resolution community profiling



Okazaki et al, 2021. *Microbiome*

Higher resolution community profiling



Okazaki et al, 2021. *Microbiome*

Some advantages of long-read metabarcoding

- Better taxonomic classification (“who is there?”)
- Infer more robust phylogenies to answer: “who is there?”
- Higher resolution community profiling
- **Enables macroevolutionary analyses**

Enables macroevolutionary analyses

- Model how traits evolve along a phylogeny
- Model speciation-extinction dynamics along a phylogeny

Requires a robust, taxon-rich, phylogeny

Article | Published: 22 October 2018

Clade-specific diversification dynamics of marine diatoms since the Jurassic

Eric Lewitus, Lucie Bittner, Shruti Malviya, Chris Bowler & Hélène Morlon

Nature Ecology & Evolution **2**, 1715–1723 (2018) | Cite this article

1557 Accesses | **28** Citations | **41** Altmetric | Metrics

Article | Open Access | Published: 04 August 2022

Global patterns and rates of habitat transitions across the eukaryotic tree of life

Mahwash Jamy, Charlie Biwer, Daniel Vaulot, Aleix Obiol, Hongmei Jing, Sari Peura, Ramon Massana & Fabien Burki

Nature Ecology & Evolution **6**, 1458–1470 (2022) | Cite this article

8806 Accesses | **3** Citations | **185** Altmetric | Metrics

ORIGINAL ARTICLE | Open Access

Analysing diversification dynamics using barcoding data: The case of an obligate mycorrhizal symbiont

Benoît Perez-Lamarque, Maarja Öpik, Odile Maliet, Ana C. Afonso Silva, Marc-André Selosse, Florent Martos, Hélène Morlon

First published: 22 April 2022 | **https://doi.org/10.1111/mec.16478**

Enables macroevolutionary analyses

- How often do microbial eukaryotes transition between habitats?
- Focus on transitions between marine and non-habitats
- Are certain groups better at transitioning across habitats?

Article | Open Access | Published: 04 August 2022

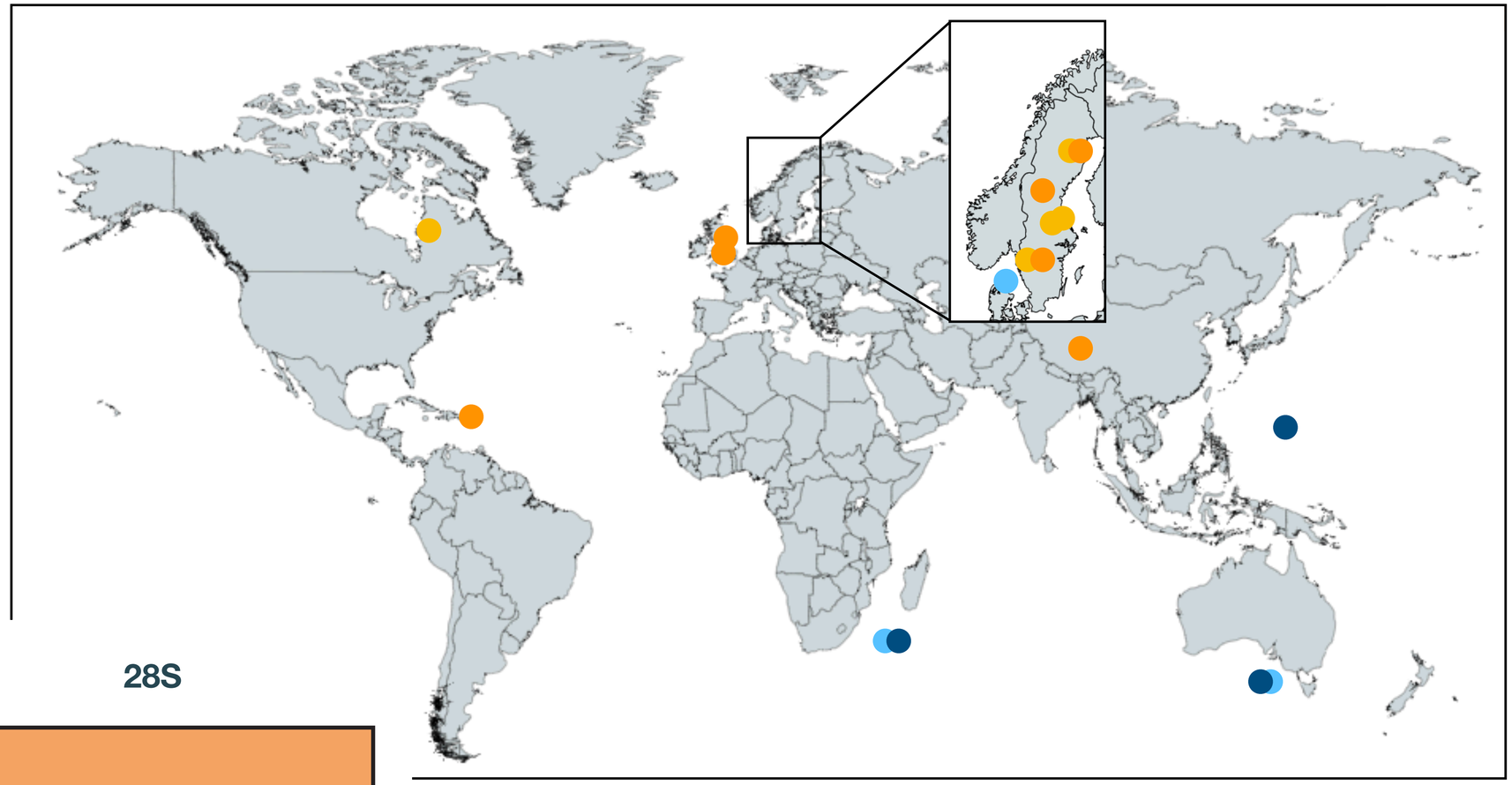
Global patterns and rates of habitat transitions across the eukaryotic tree of life

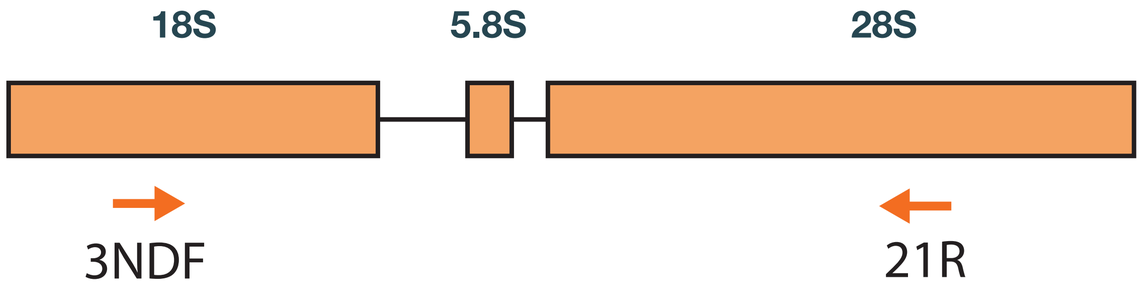
Mahwash Jamy, Charlie Biwer, Daniel Vaulot, Aleix Obiol, Hongmei Jing, Sari Peura, Ramon Massana & Fabien Burki

Nature Ecology & Evolution **6**, 1458–1470 (2022) | Cite this article

Generating a dense 18S-28S eukaryotic dataset

- Soil samples (x7)
- Freshwater samples (x5)
- Marine euphotic (x4)
- Marine aphotic (x5)





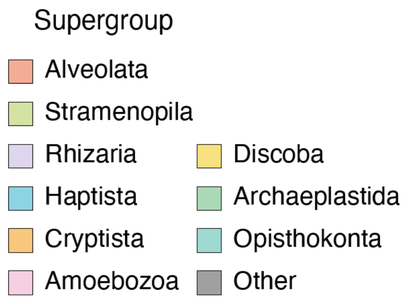
Sequenced on PacBio Sequel II
> 17,000 OTUs!

18S+28S tree

- **Contains environmental sequences only**
- **Contains all major eukaryotic lineages**

16,821 OTUs
7,160 sites

RAxML
GTR+G

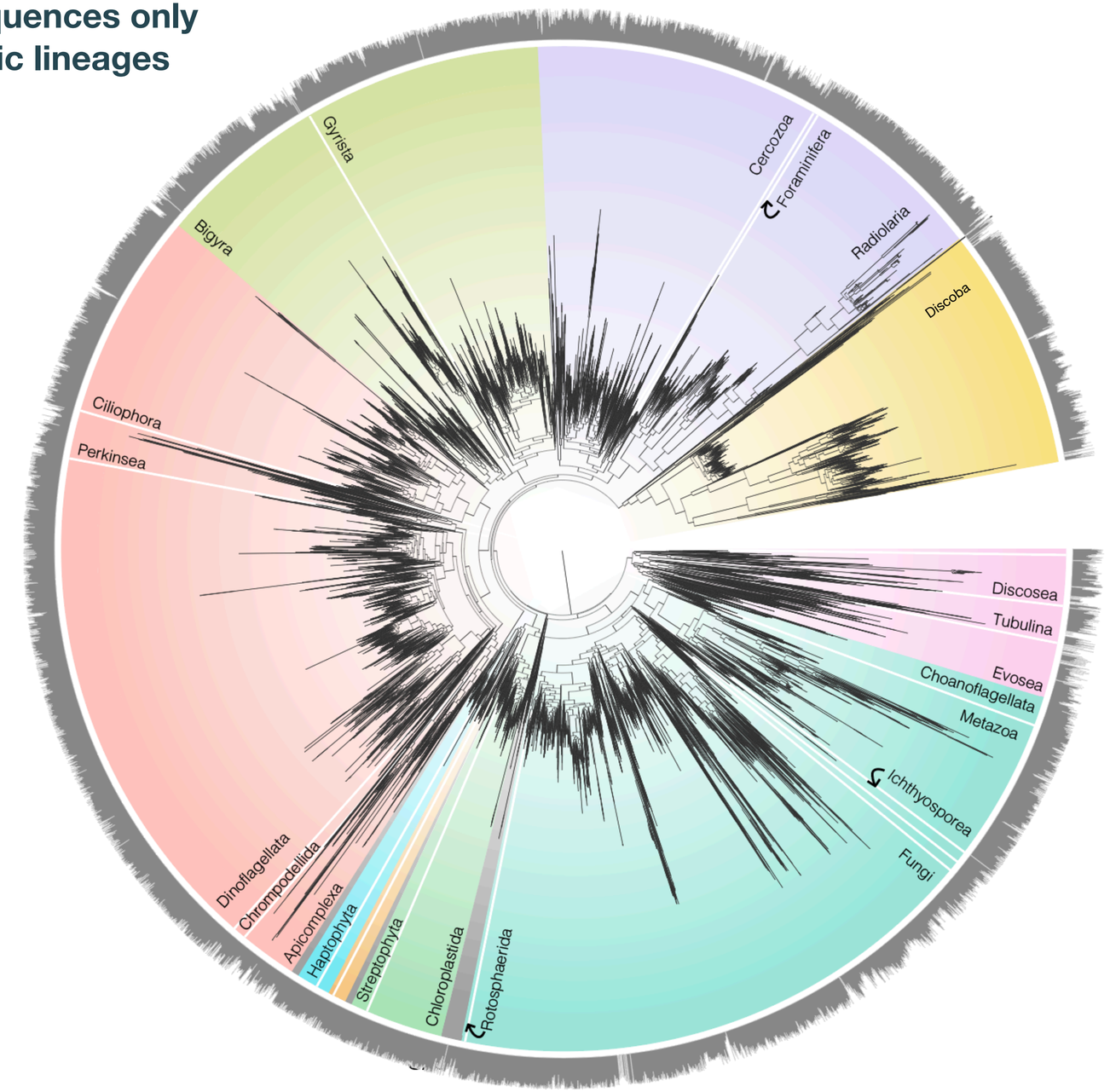


- **Contains environmental sequences only**
- **Contains all major eukaryotic lineages**

18S+28S tree

16,821 OTUs
7,160 sites

RAxML
GTR+G

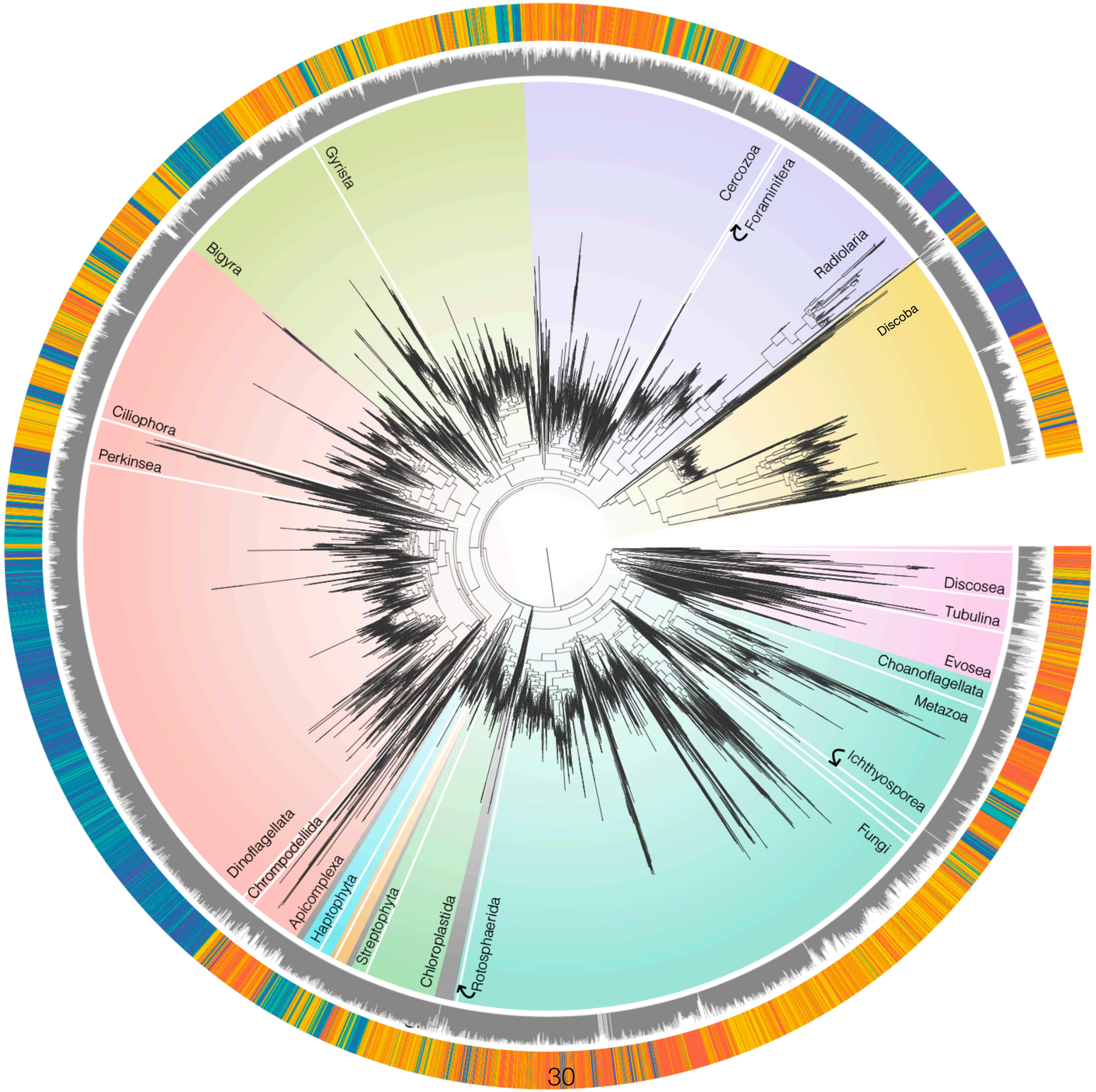


18S+28S tree

16,821 OTUs
7,160 sites

Habitat

- freshwater
- soil
- marine photic
- marine aphotic

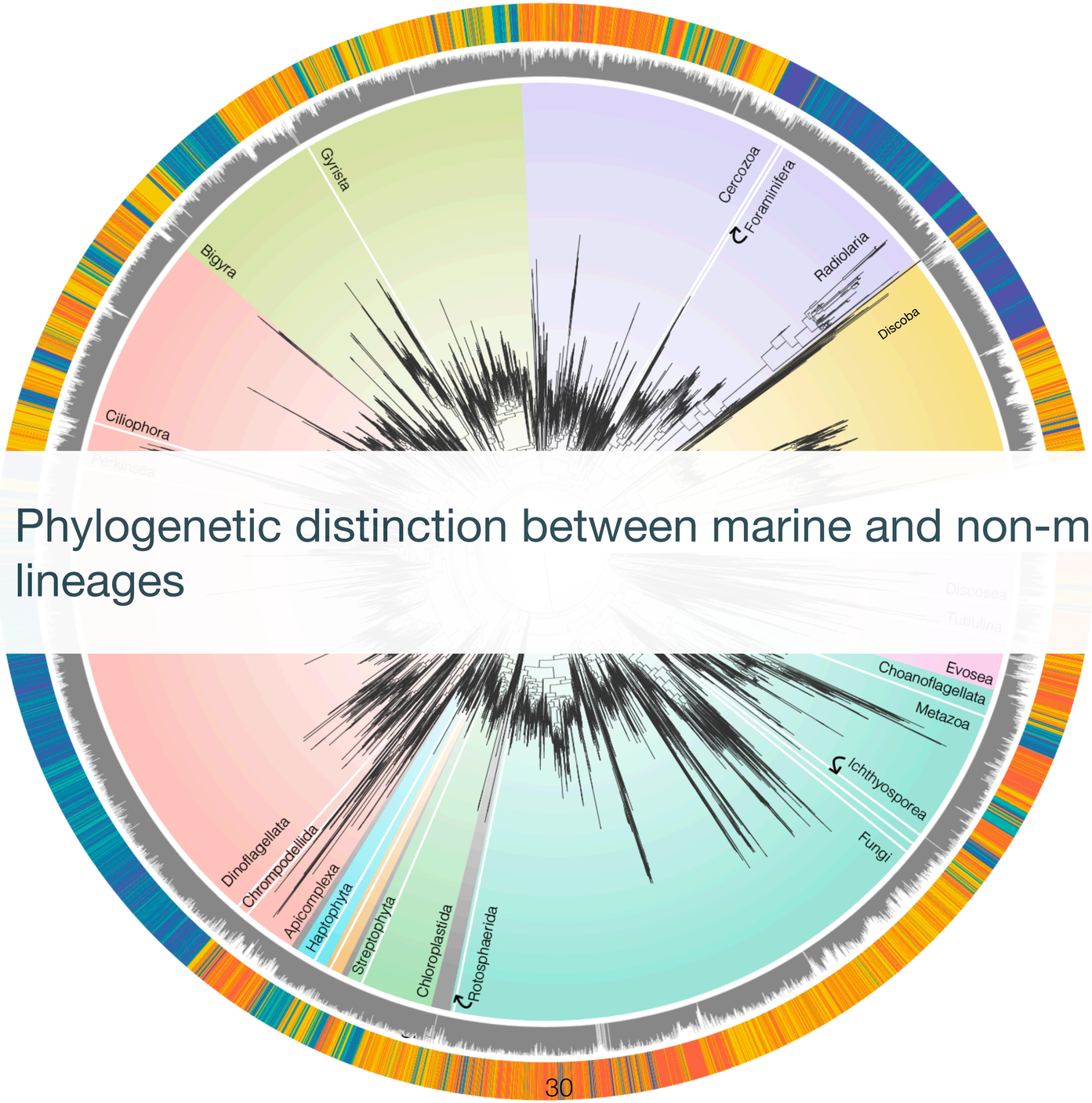


18S+28S tree

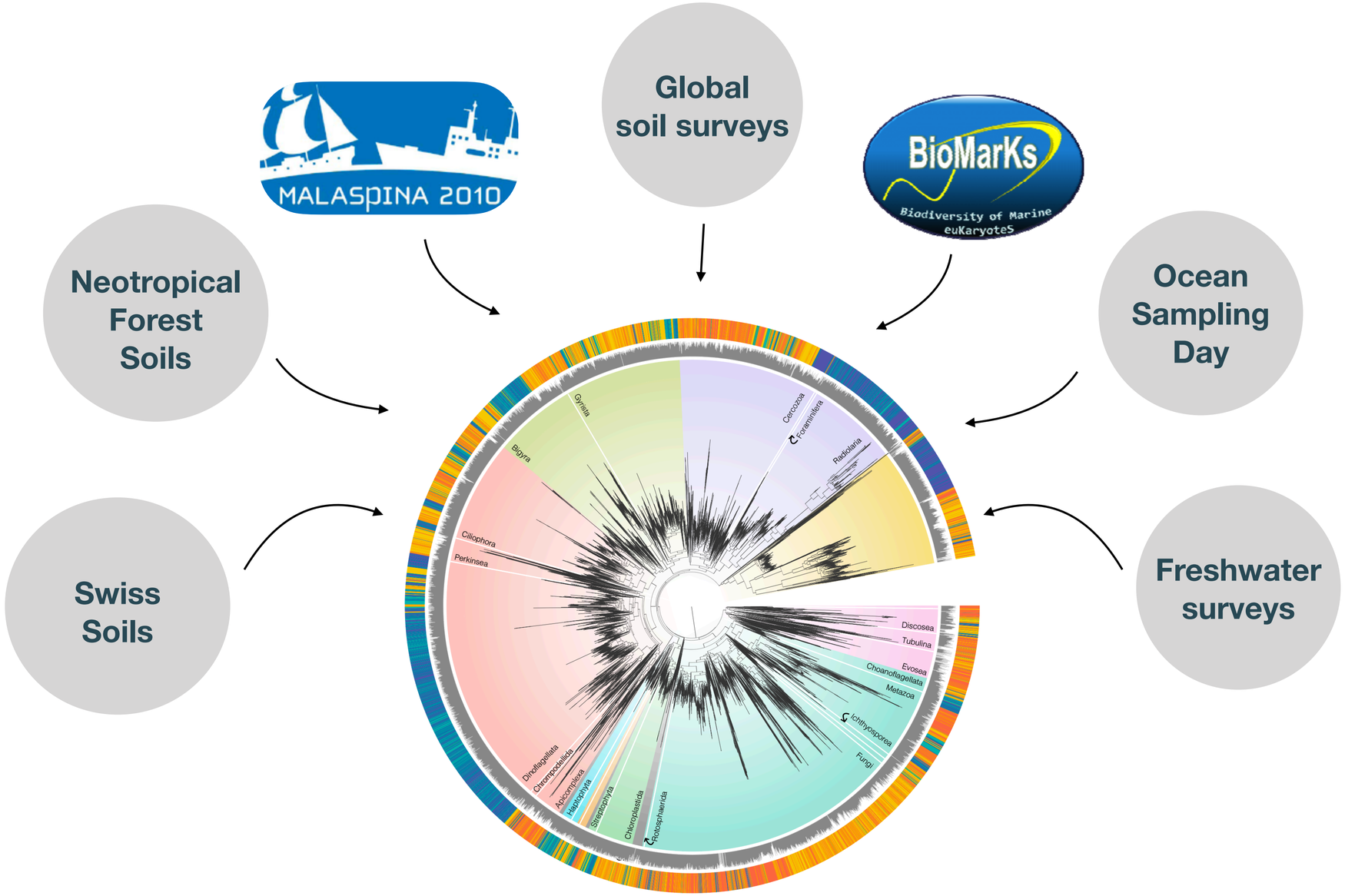
16,821 OTUs
7,160 sites

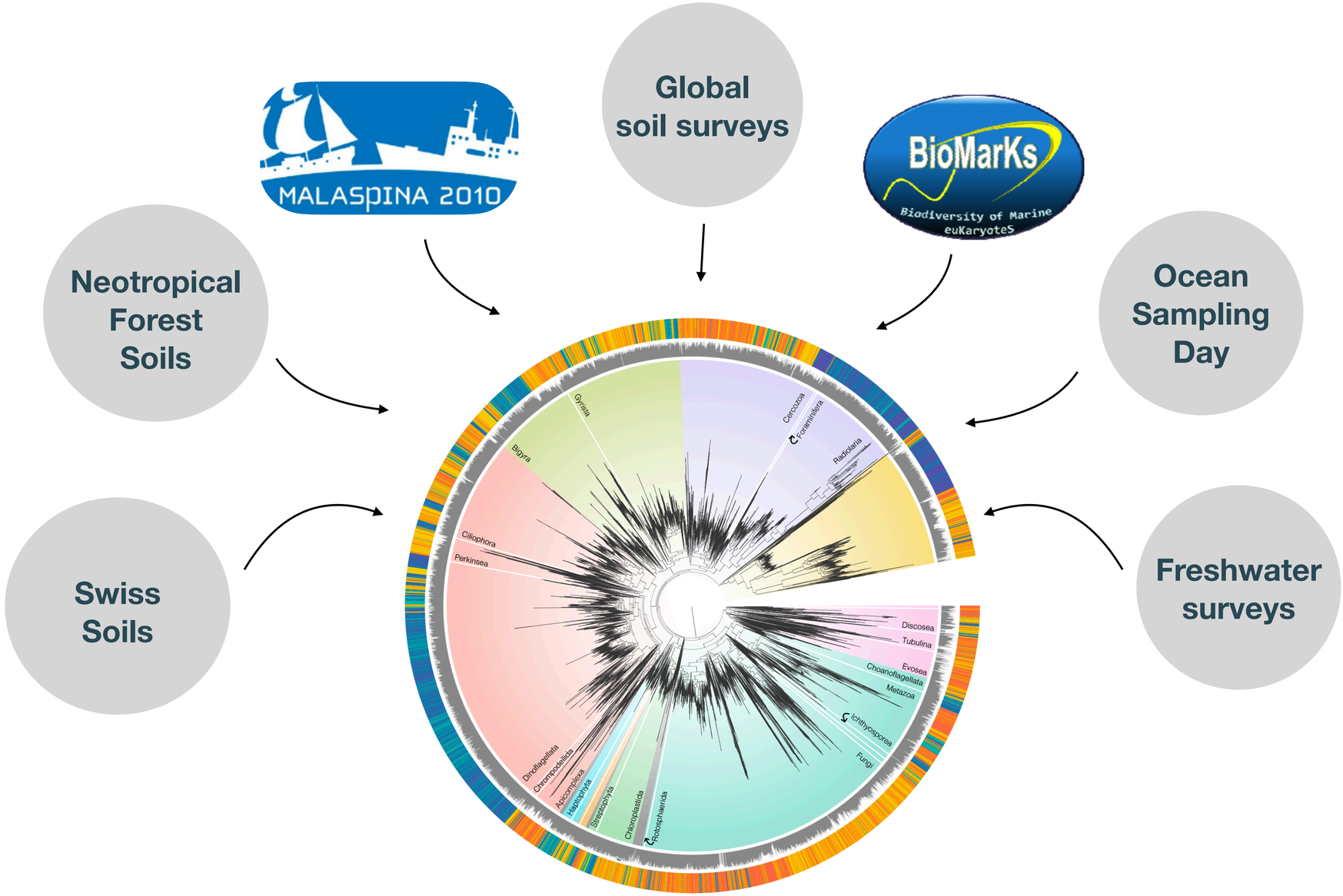
Habitat

- freshwater
- soil
- marine photic
- marine aphotic



- Phylogenetic distinction between marine and non-marine lineages


MALASPINA 2010
Global soil surveys
BioMarKs
Biodiversity of Marine euKaryoteS
Neotropical Forest Soils
Ocean Sampling Day
Swiss Soils
Freshwater surveys
Gyrista
Bigyra
Cercozoa
Foraminifera
Radiolaria
Ciliophora
Perkinsea
Discosea
Tubulina
Evosea
Choanoflagellata
Metazoa
Ichthyosporea
Fungi
Dinoflagellata
Chromerodavidida
Apicomplexa
Haptophyta
Stramenopiles
Chloroplastida
Rotosphaerida



Modelled rates of habitat transition across the eukaryotic tree of life.

- Detected hundreds of habitat transition events.

- Fungi, diatoms, golden algae have transitioned across marine/non-marine boundary more frequently than other clades.

Some advantages of long-read metabarcoding

- Better taxonomic classification (“who is there?”)
- Infer more robust phylogenies to answer: “who is there?”
- Higher resolution community profiling
- Enables macroevolutionary analyses
- **Rapidly generate sequences to populate reference databases**

Generate sequences to populate reference databases

BOLD SYSTEMS

- Reference sequences are very useful for taxonomic annotations, phylogenetic analyses, designing primers and probes, and more.

- Most sequences generated by Sanger sequencing which is low-throughput and expensive.

- Long-read sequencing can generate these reference sequences rapidly and in a cost effective way.

Mahé et al. 2017, Nature Ecology and Evolution

Rapid metabarcoding of herbarium specimen

MycoKeys. 2022; 86: 195–212.
PMCID: PMC8828592
Published online 2022 Feb 2. doi: 10.3897/mycokeys.86.77431
PMID: 35153530

A long-read amplicon approach to scaling up the metabarcoding of lichen herbarium specimens

Cécile Gueidan, Conceptualization, Formal analysis, Methodology, Writing - original draft[1] and Lan Li, Methodology[1]

Ecosystem-specific databases

- Sequenced full-length 16S rRNA genes from two soils commonly used for plant microbiome research.
- 18,042 ASVs
- Unidentified ASVs given placeholder names.
- Useful for future short-read metabarcoding studies

Askov soil. Used in plant and soil studies for more than 125 years.

Overgaard et al, 2022. *Frontiers in Microbiology*

Getting linked reference databases

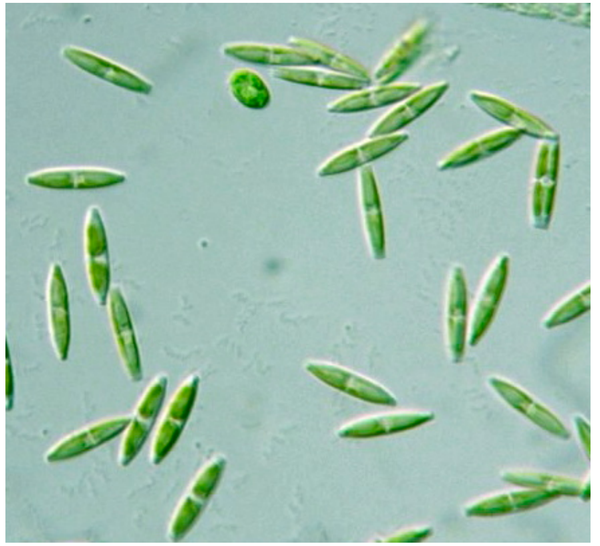
- Existing reference databases not linked

accession number	organism name	sequence length	sequence quality	alignment quality
AJ867001	*Nitzschia palea*	1548		

18S

accession number	organism name	sequence length	sequence quality	alignment quality
AM183226	*Nitzschia palea*	570		

28S



Getting linked reference databases

- Existing reference databases not linked

accession number	organism name	sequence length	sequence quality	alignment quality
AJ867001	*Nitzschia palea*	1548		

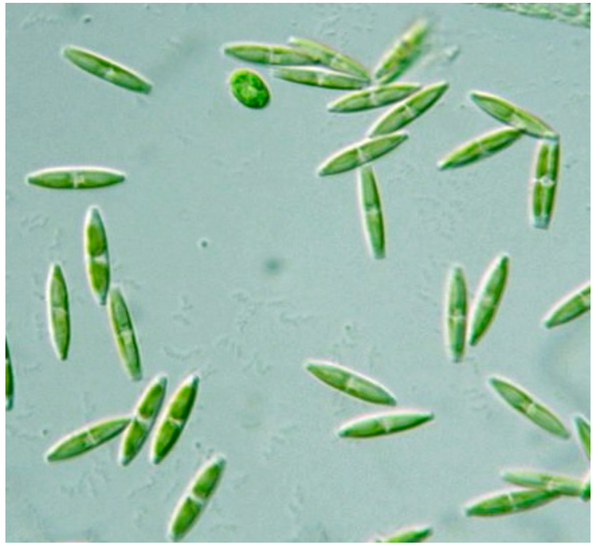
18S

accession number	organism name	sequence length	sequence quality	alignment quality
AM183226	*Nitzschia palea*	570		

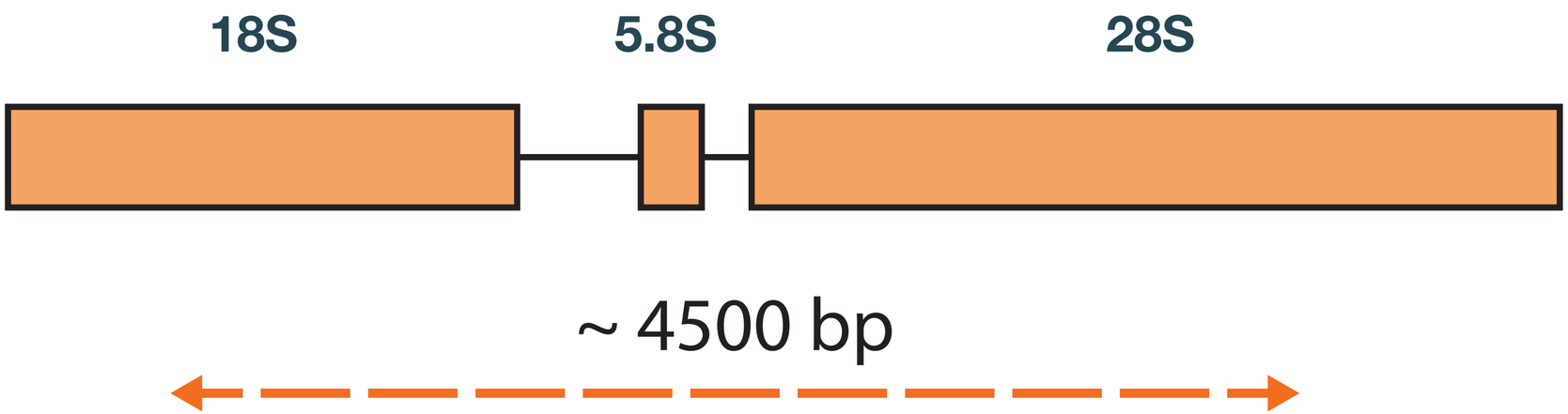
28S

- Did the two sequences come from the same strain? (Or even the same species?)

- ITS sequence?



Getting linked reference databases

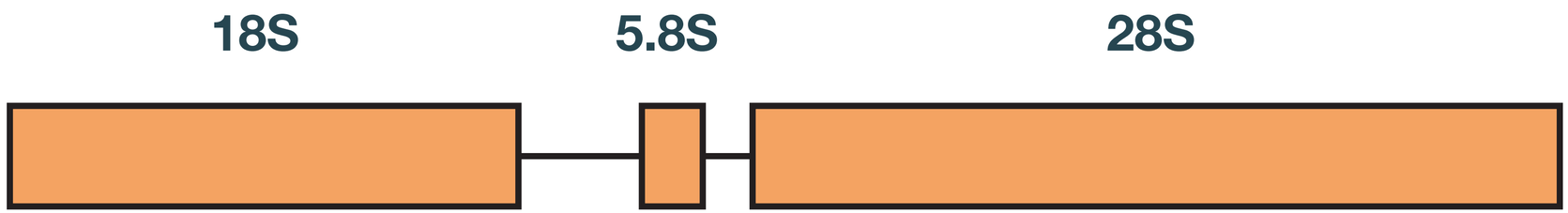


- Soil, freshwater, marine environmental samples
- General eukaryotic primers
- Close to 17,000 OTUs after processing data

Jamy et al, 2020. *Molecular Ecology Resources*

Jamy et al, 2022. *Nature Ecology & Evolution*

Getting linked reference databases

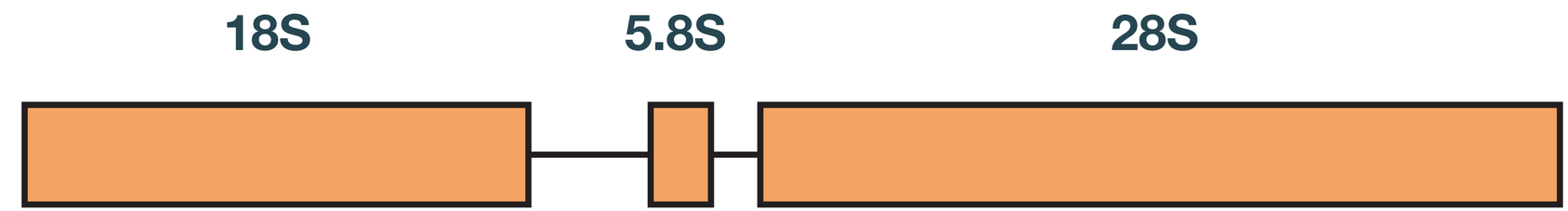


Euk; SAR; Alveolata; Ciliophora

Jamy et al, 2020. *Molecular Ecology Resources*

Jamy et al, 2022. *Nature Ecology & Evolution*

Getting linked reference databases

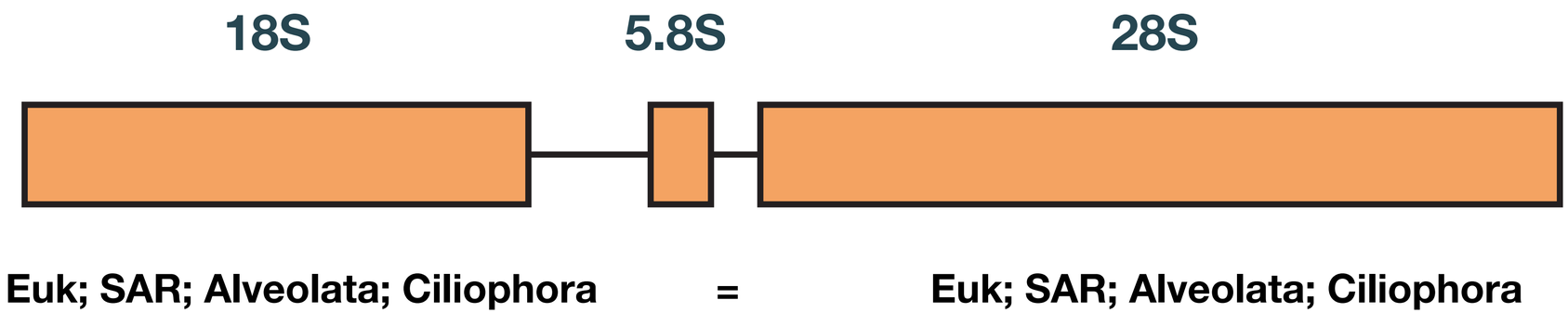


Euk; SAR; Alveolata; Ciliophora = **Euk; SAR; Alveolata; Ciliophora**

Jamy et al, 2020. *Molecular Ecology Resources*

Jamy et al, 2022. *Nature Ecology & Evolution*

Getting linked reference databases



- Transfer taxonomy to 28S (and ITS region)
- Linked database (including other studies) to be released this year (Tedersoo et al. *in prep*)

Jamy et al, 2020. *Molecular Ecology Resources*

Jamy et al, 2022. *Nature Ecology & Evolution*

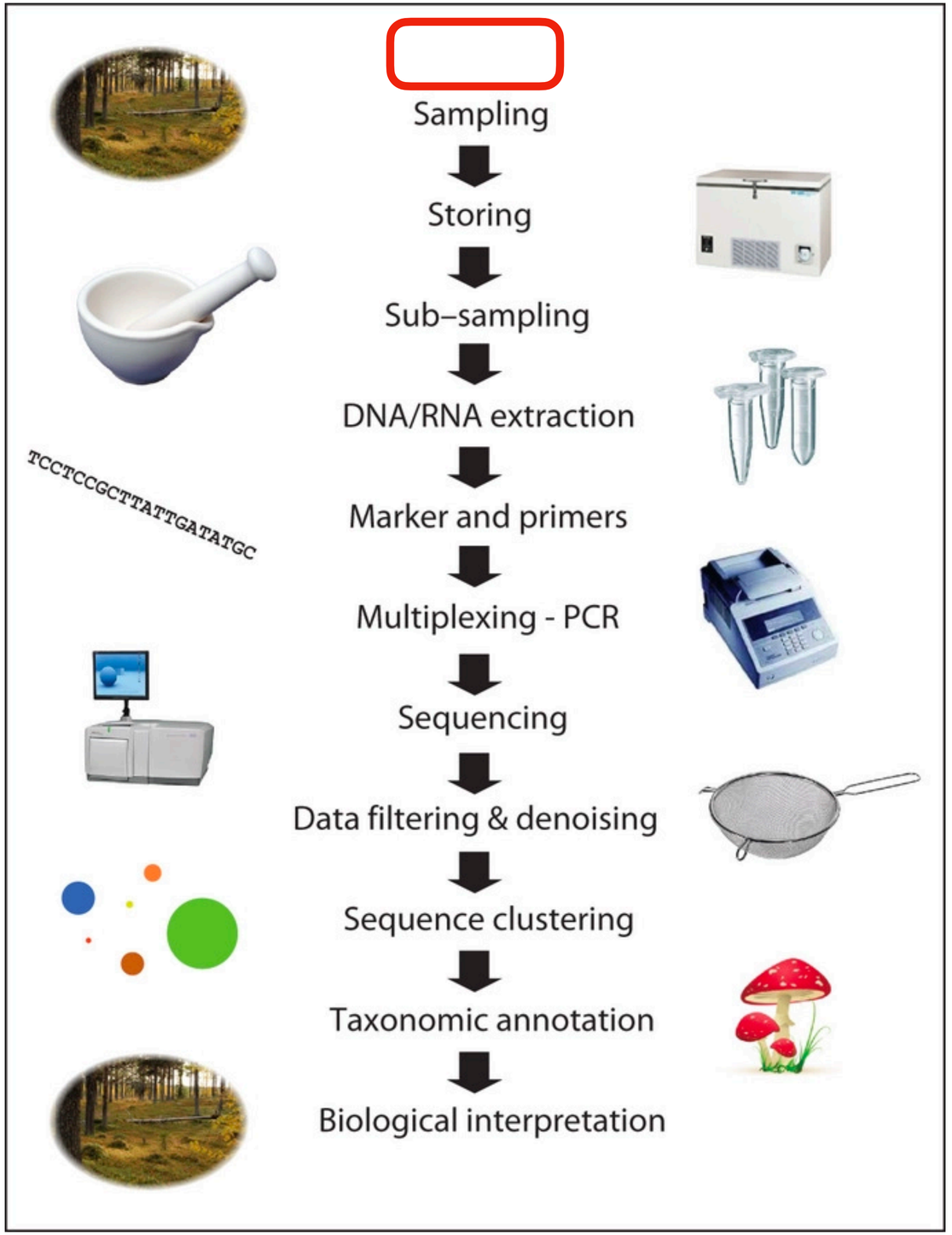
Limitations of long-read metabarcoding

- Higher cost and lower sequencing throughput than regular metabarcoding (this gap is becoming smaller and smaller).
- Usually requires greater amount and quality of input DNA
- Higher rate of chimera formation (but can be circumvented by using unique molecular identifiers, and bioinformatic tools).
- May not be worth it to study well-studied groups with good reference databases.
- Still relatively new!! Bioinformatic approaches are less well-established and are continue to evolve.

Any questions?

Any questions?

Generating long-read amplicons



Lindahl et al, 2013. *New Phytologist*

Things to consider

- Which sampling scheme? How many replicates? Which extraction protocol?

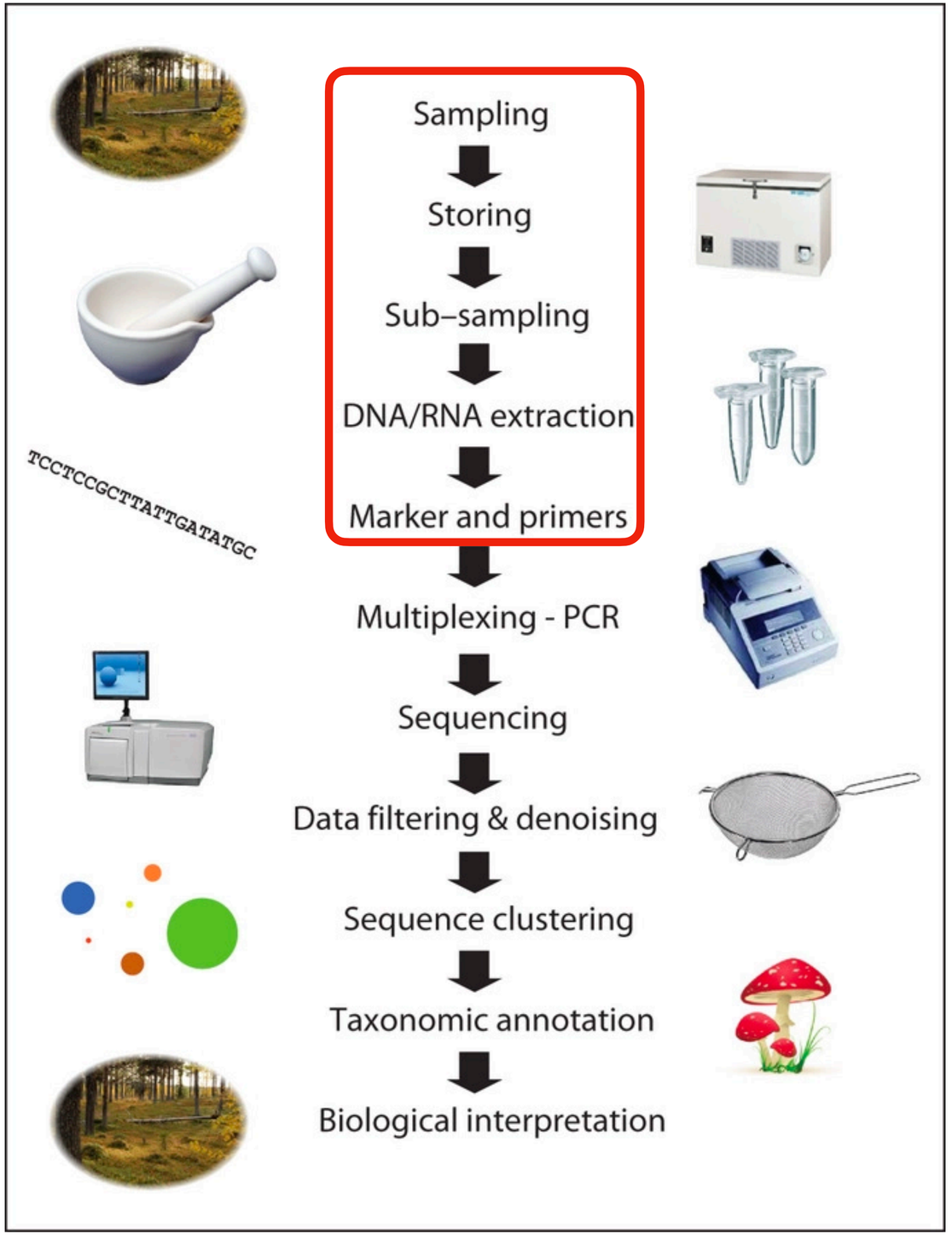
- Which marker(s)? Which primers?

- Which sequencing depth?

- Which sequencing technique? Unique molecular identifiers (UMIs) or not? Which sequencing platform?

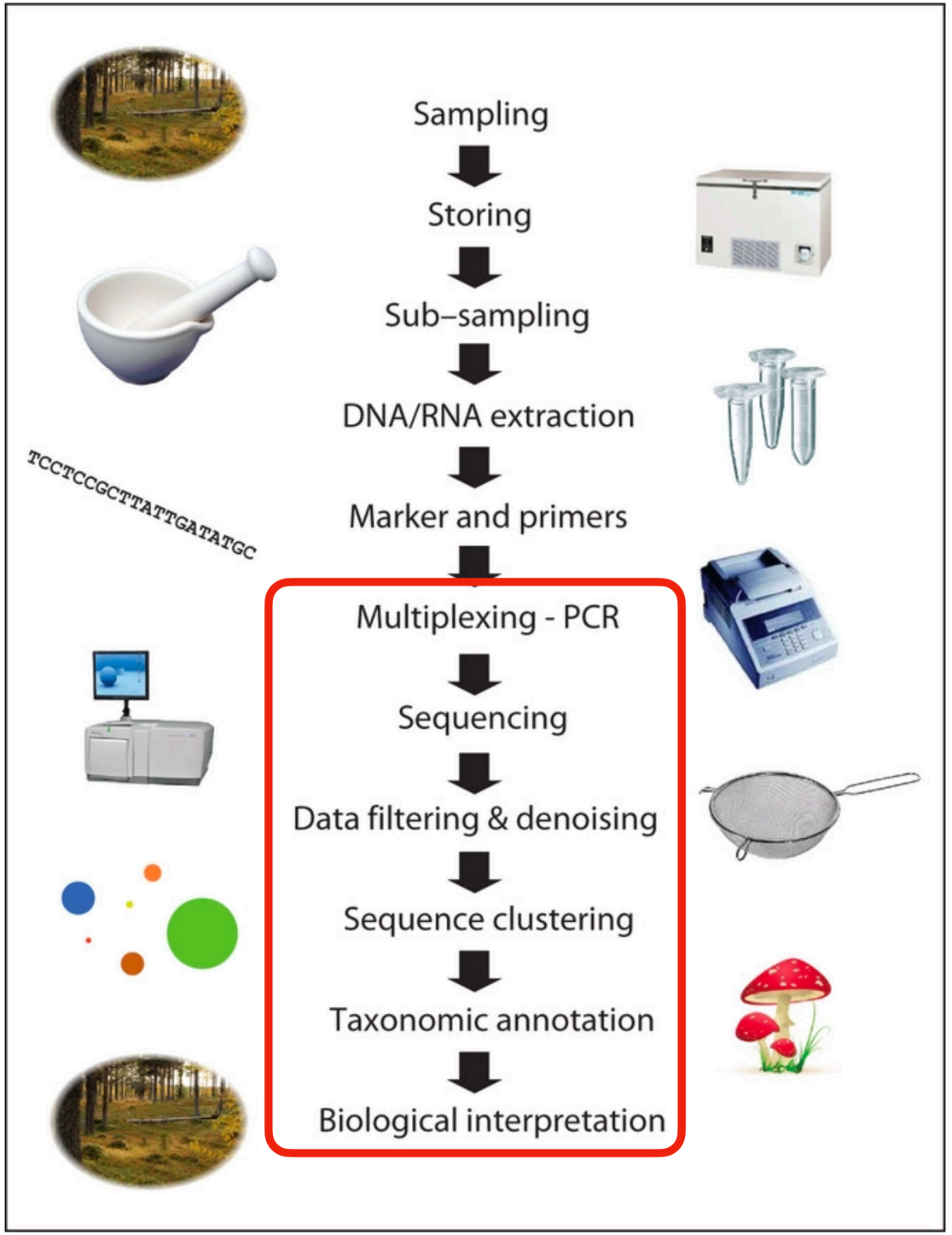
Sequencing techniques

- “Regular” long-read metabarcoding - more common!
  - PacBio
  - Nanopore
- Long-read metabarcoding with unique molecular identifiers (UMIs) - less common
  - PacBio
  - Nanopore
  - Illumina



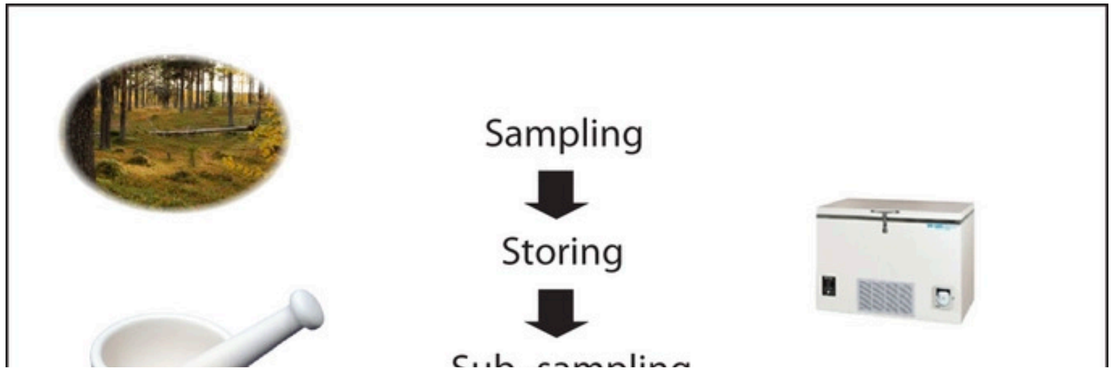
More or less the same as in regular metabarcoding!

Lindahl et al, 2013. *New Phytologist*

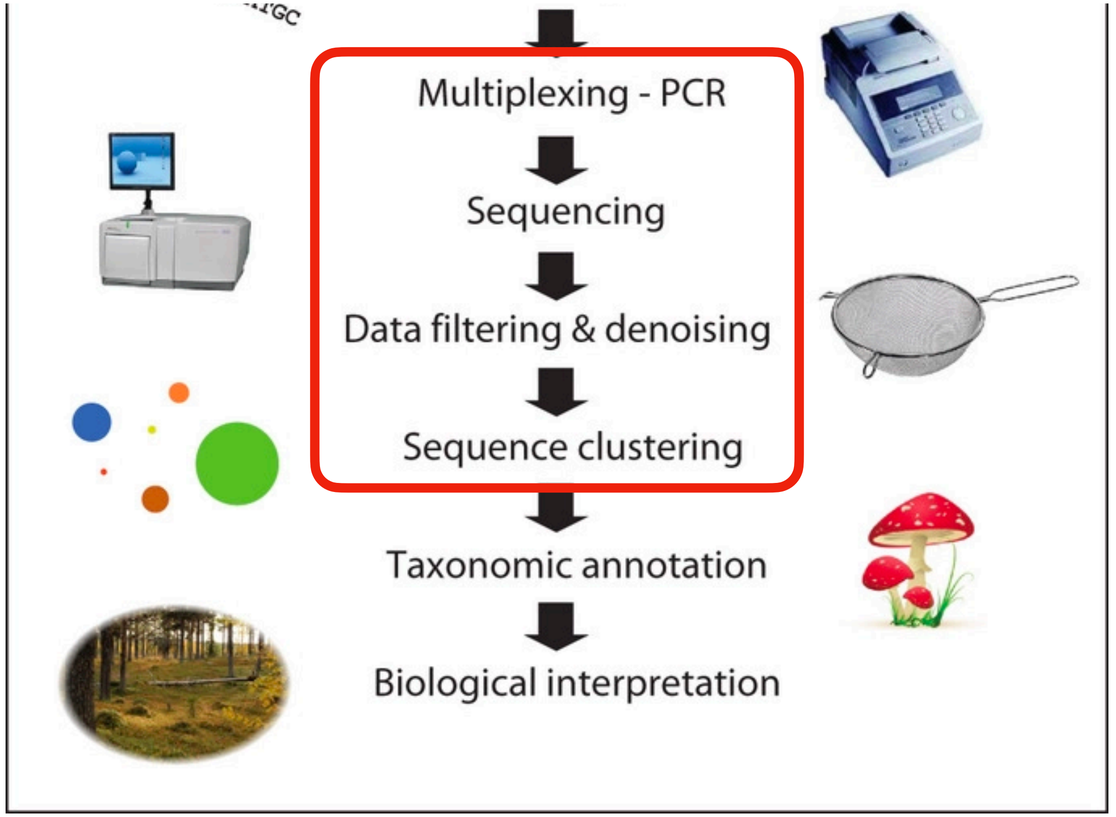


These steps differ, and will depend on the sequencing technique used.

Lindahl et al, 2013. *New Phytologist*

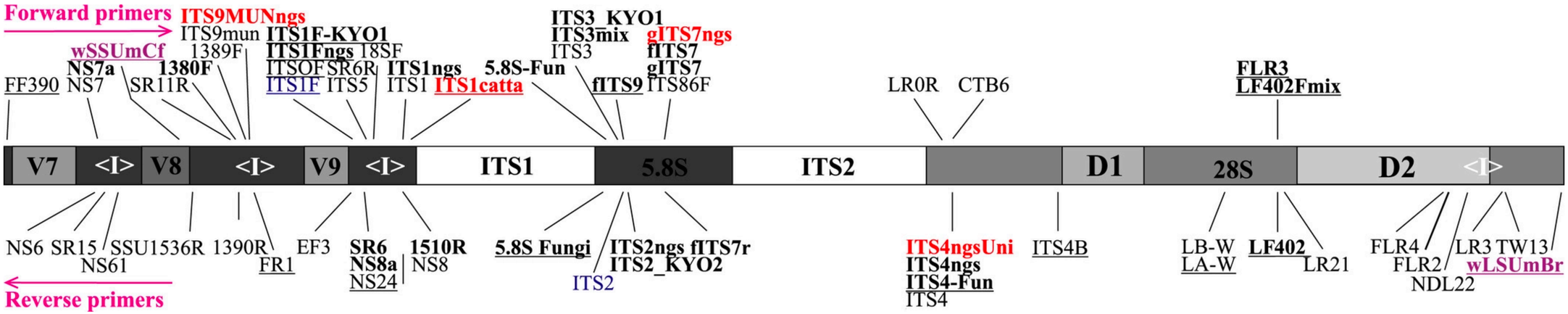


Long-read metabarcoding
without Unique Molecular Identifiers (UMIs)



Lindahl et al, 2013. *New Phytologist*

Primers



Tedersoo et al, 2022. *Molecular Ecology*

MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE | Open Access

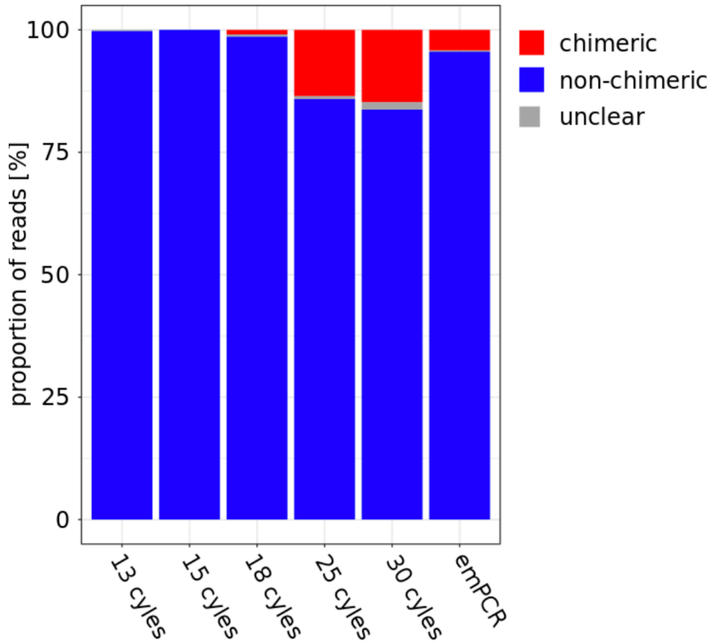
Short- and long-read metabarcoding of the eukaryotic rRNA operon: Evaluation of primers and comparison to shotgun metagenomics sequencing

Meike A. C. Latz, Vesna Grujcic, Sonia Brugel, Jenny Lycken, Uwe John, Bengt Karlson, Agneta Andersson, Anders F. Andersson

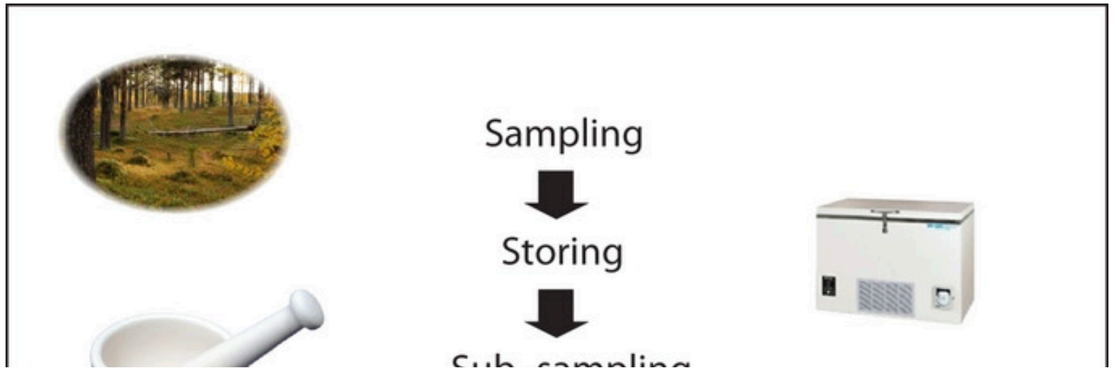
First published: 19 April 2022 | https://doi.org/10.1111/1755-0998.13623 | Citations: 1

PCR

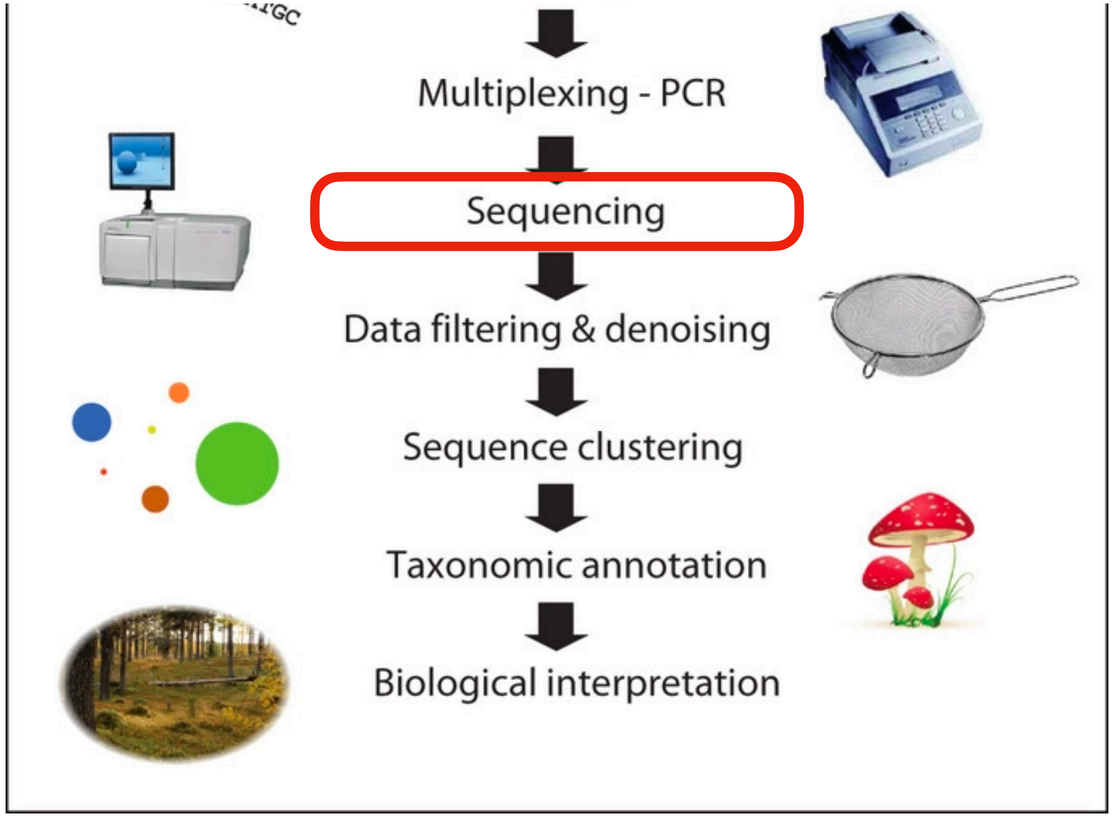
- Beware of chimeras!!
- Longer amplicons are more prone to chimeras
- Use fewer PCR cycles if possible



Heeger et al. 2018, *Molecular Ecology Resources*



Long-read metabarcoding
without Unique Molecular Identifiers (UMIs)



PacBio or
Nanopore?

Lindahl et al, 2013. *New Phytologist*

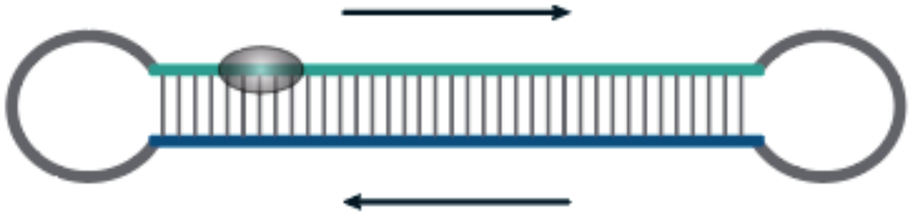
PacBio or Nanopore?

- PacBio Sequel and Sequel II
  (and now the new PacBio Revio!!!!!)
- Nanopore

- Error rate
- Portability
- The type of community you want to sequence
- Sequence length/marker

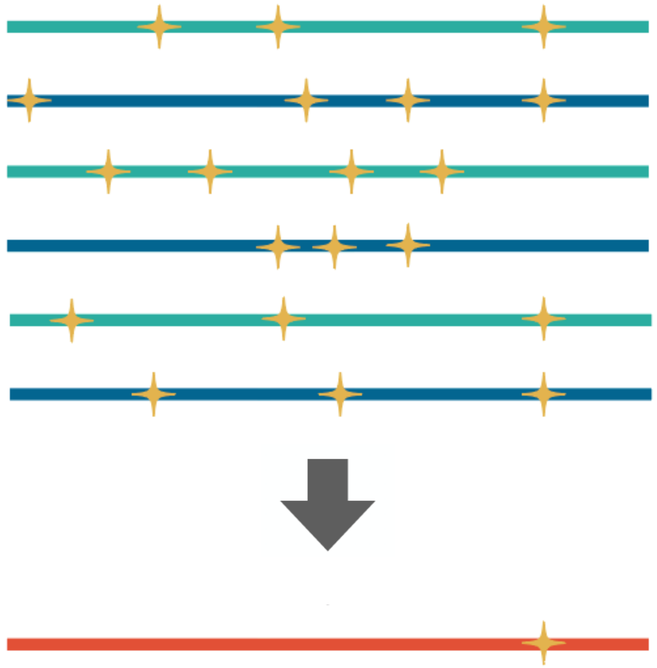
PacBio

SMRTbell template



Hairpin adaptors

Multi-pass sequencing



Subreads = passes
Error rate = 14-15%
BUT randomly distributed

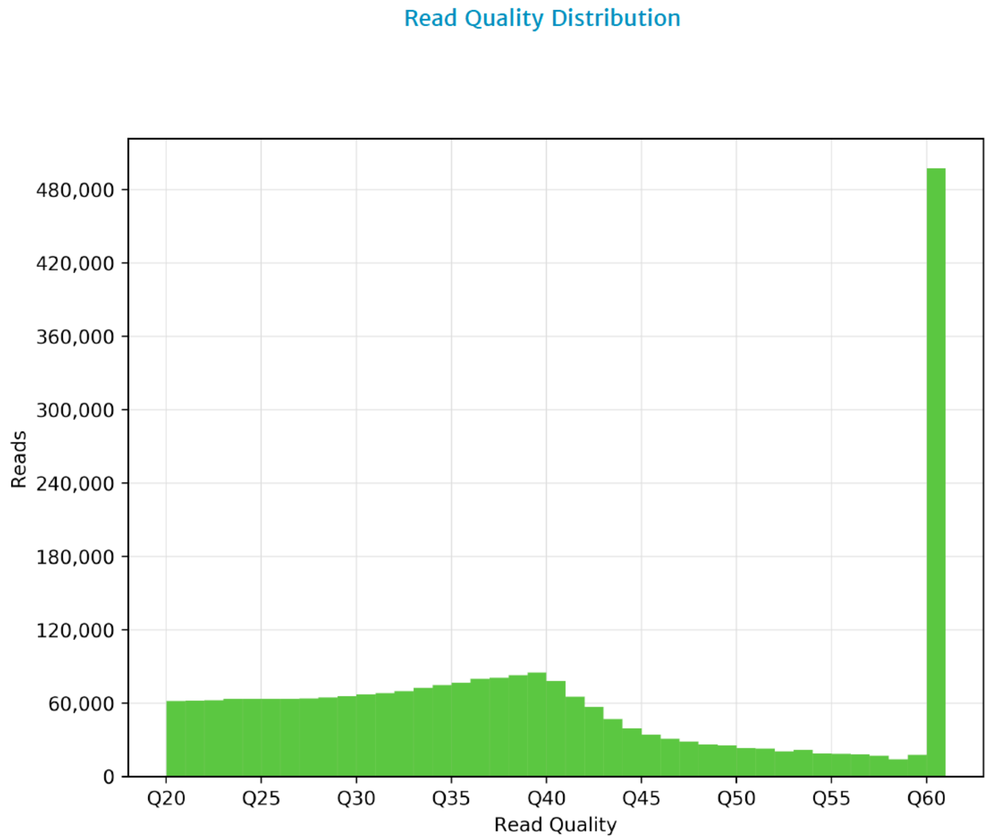
CCS = HiFi

Circular Consensus Sequence (CCS)
Much lower error rate

PacBio

- Use Sequel II if possible (cheaper and more high-throughput than Sequel) ~35,000 CCS per cell in Sequel vs ~3,000,000 CCS per cell for Sequel II

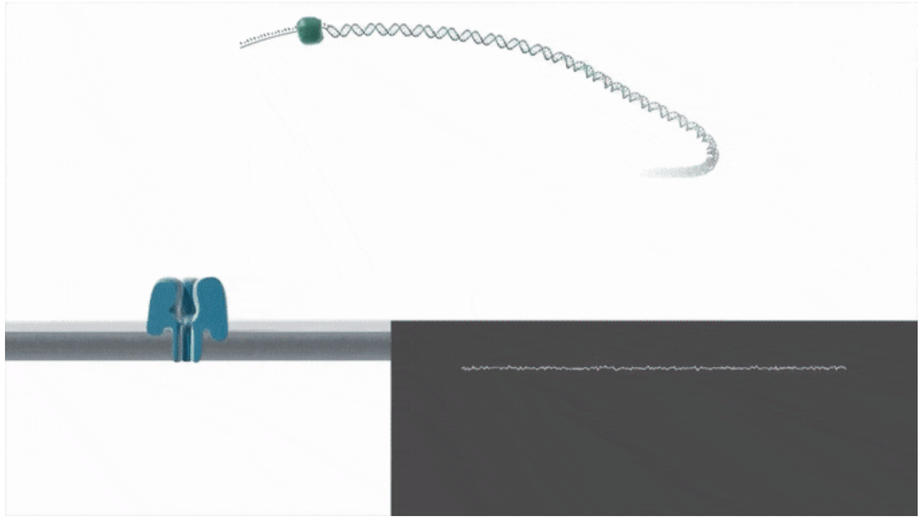
- Works well for complex communities



PacBio

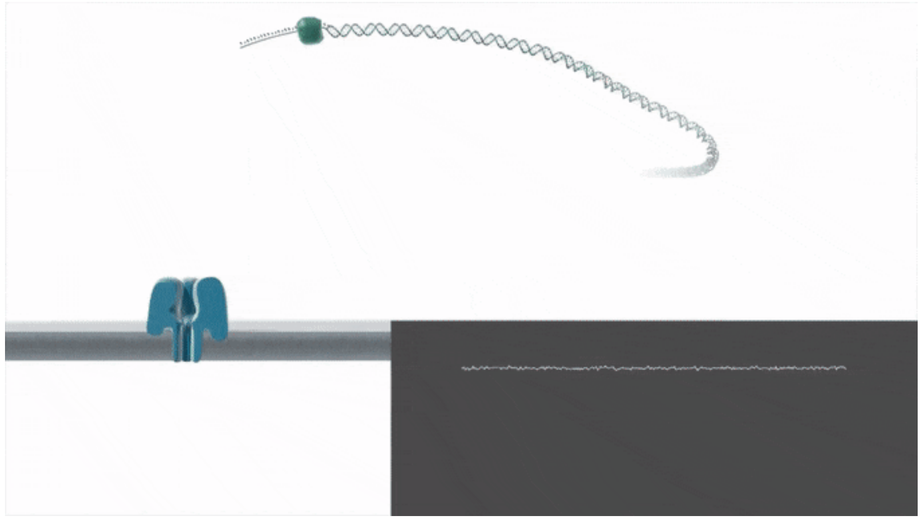
- Error rate. Raw error rate is high. Error rate of CCS is very low.
- Portability. Low
- The type of community you want to sequence. Complex community
- Sequence length/marker. Up to 20 kb

Nanopore



- Avg error rate = 10-15%
- No CCS technology

Nanopore



- Avg error rate = 10-15%
- No CCS technology

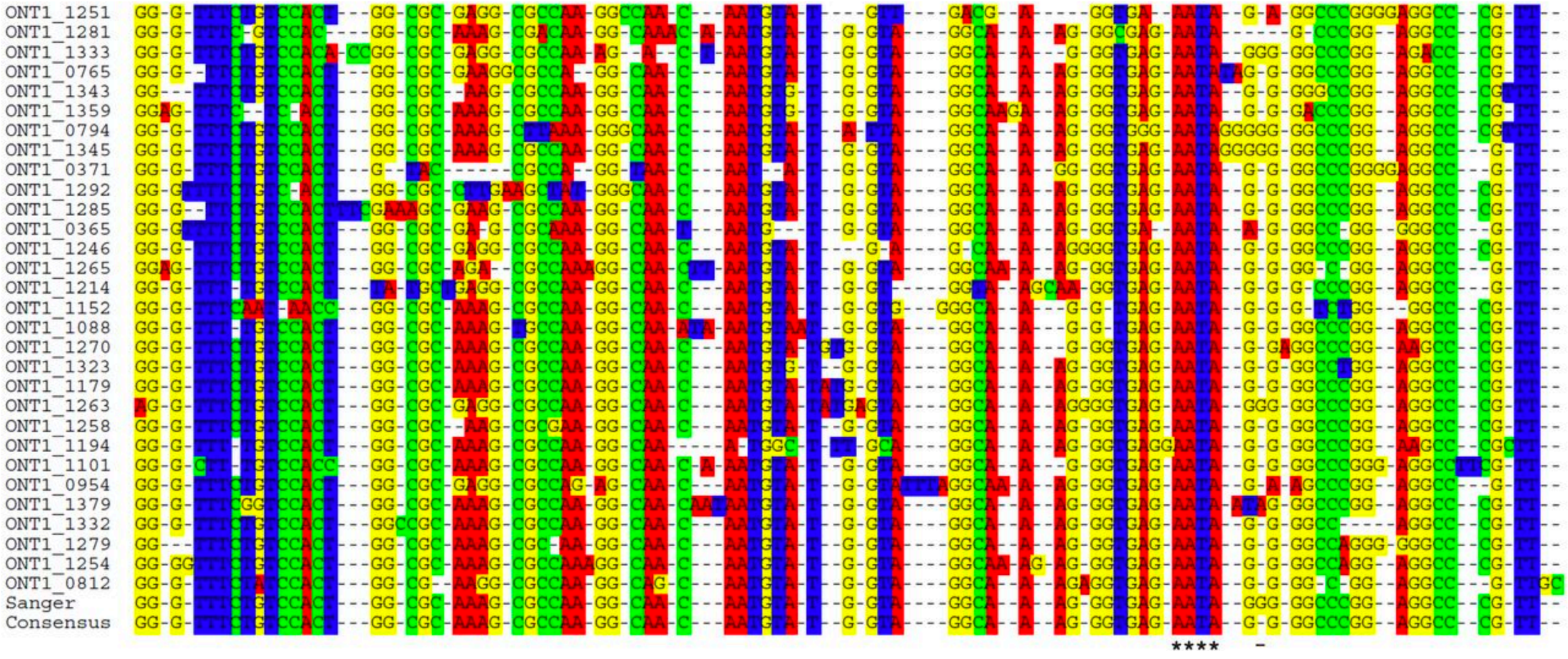
Nanopore

Relative Performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) Third-Generation Sequencing Instruments in Identification of Agricultural and Forest Fungal Pathogens

Kaire Loit, Kalev Adamson, Mohammad Bahram, Rasmus Puusepp, Sten Anslan, Riinu Kiiker, Rein Drenkhan, Leho Tedersoo

Irina S. Druzhinina, Editor

Cluster reads at 85%-92% seq id, and generate consensus to reduce error rate.



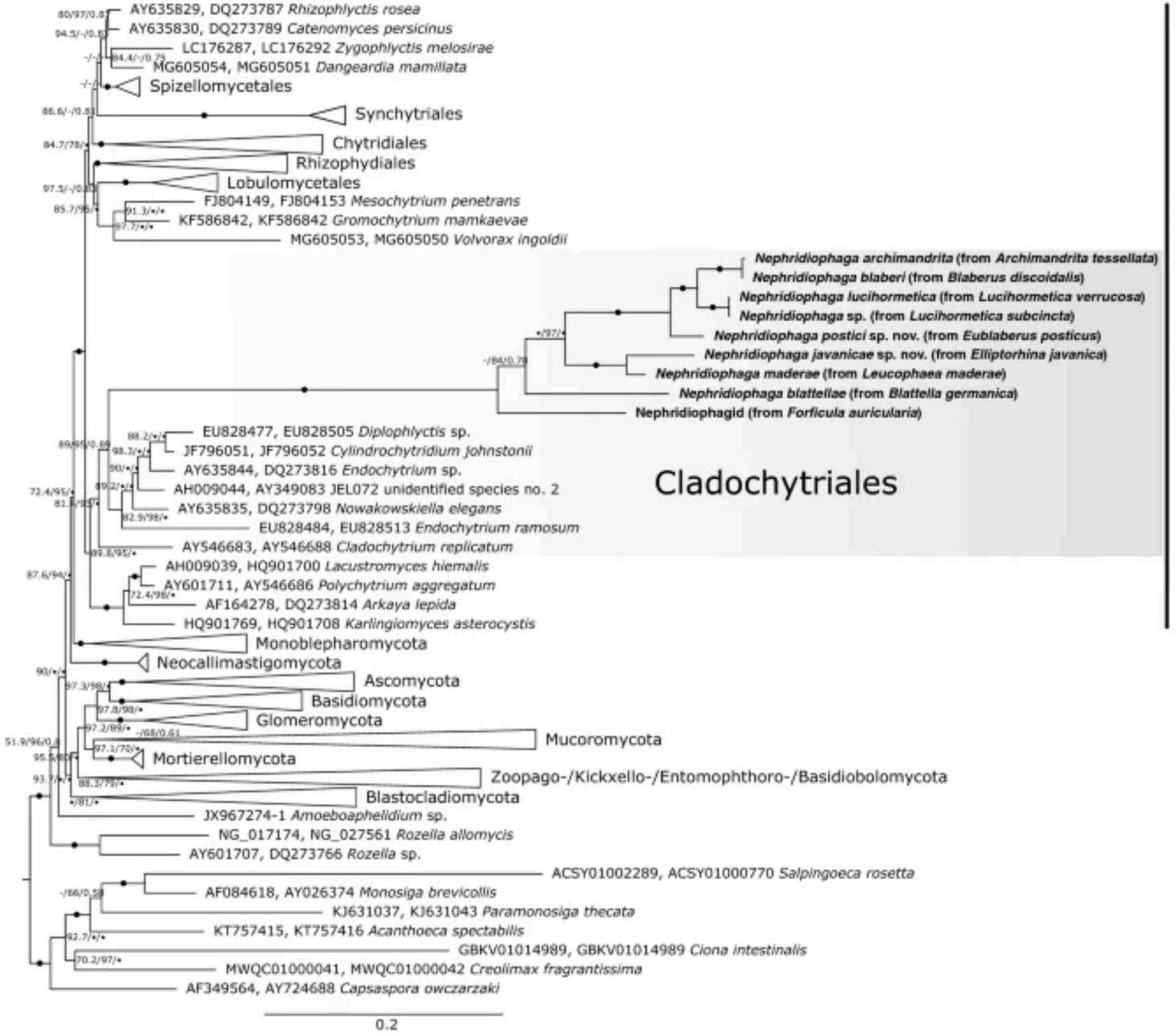
Nanopore

- Error rate. Raw error rate is high. Error rate of processed reads is still too high (compared to Illumina and PacBio).
- Portability. High!! Also very rapid.
- The type of community you want to sequence. Simple community
- Sequence length/marker. Up to 20 kb

Article | Open Access | Published: 11 January 2021

Long rDNA amplicon sequencing of insect-infecting nephridiophagids reveals their affiliation to the Chytridiomycota and a potential to switch between hosts

Jürgen F. H. Strassert ✉, Christian Wurzbacher, Vincent Hervé, Taraha Antany, Andreas Brune & Renate Radek ✉



Home › Nanopore Sequencing › Protocol

Full-Length 16S rRNA Gene Analysis Using Long-Read Nanopore Sequencing for Rapid Identification of Bacteria from Clinical Specimens

Yoshiyuki Matsuo

Protocol | First Online: 14 February 2023

489 Accesses | 2 Altmetric

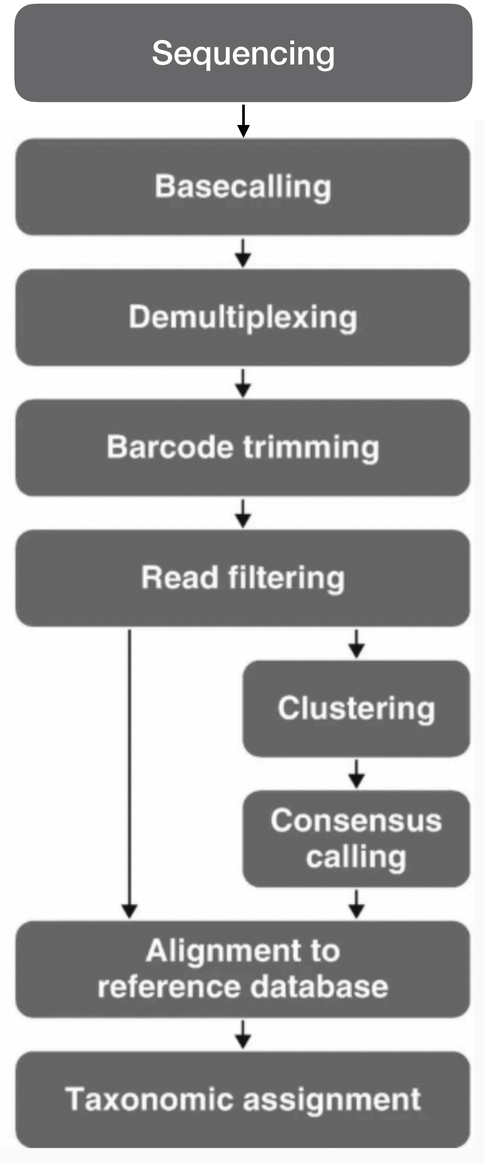
Part of the Methods in Molecular Biology book series (MIMB,volume 2632)

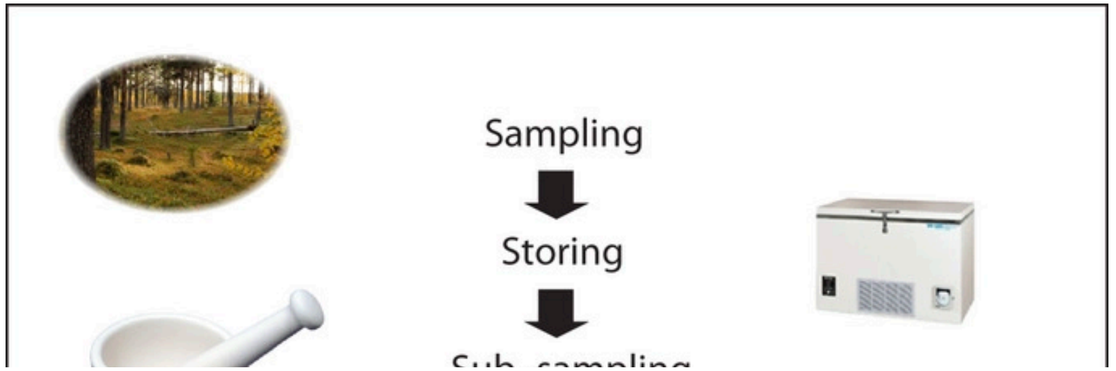
Speeding up the detection of invasive bivalve species using environmental DNA: A Nanopore and Illumina sequencing comparison

Bastian Egeter, Joana Veríssimo, Manuel Lopes-Lima, Cátia Chaves, Joana Pinto, Nicoletta Riccardi, Pedro Beja, Nuno A. Fonseca

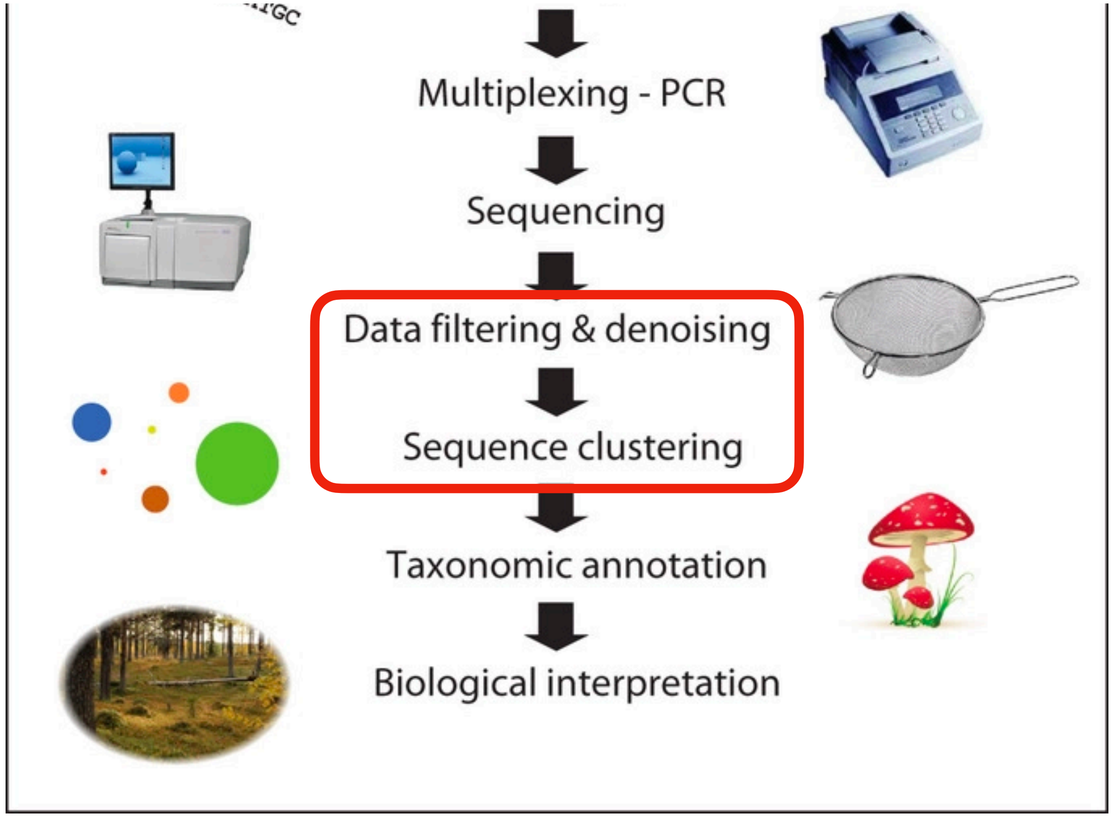
Turnaround time

1-2 days





Long-read metabarcoding
without Unique Molecular Identifiers (UMIs)



Lindahl et al, 2013. *New Phytologist*

Bioinformatic analyses of long-read amplicon data

- Fewer dedicated pipelines for long-read data.

Bioinformatic analyses of long-read amplicon data

- Fewer dedicated pipelines for long-read data.
- DADA2 (PacBio data only)
  - Great if you are using 16S/18S/ITS (or even 16S-23S….test on your dataset!)
  - Cannot handle longer reads yet (memory issues)

JOURNAL ARTICLE

High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution

Benjamin J Callahan, Joan Wong, Cheryl Heiner, Steve Oh, Casey M Theriot, Ajay S Gulati, Sarah K McGill, Michael K Dougherty

Bioinformatic analyses of long-read amplicon data

- Fewer dedicated pipelines for long-read data.
- DADA2 (PacBio data only)
  - Great if you are using 16S/18S/ITS (or even 16S-23S….test on your dataset!)
  - Cannot handle longer reads yet (memory issues)
- Custom pipelines (using VSEARCH, DADA2, etc.)
  - Depends on taxonomic group, sequencing technology, sequencing depth, length and marker(s) selected
  - Using the same programmes (VSEARCH, DADA2 etc. but adapt parameters for long reads)

Custom pipeline for processing PacBio data (18S-28S)

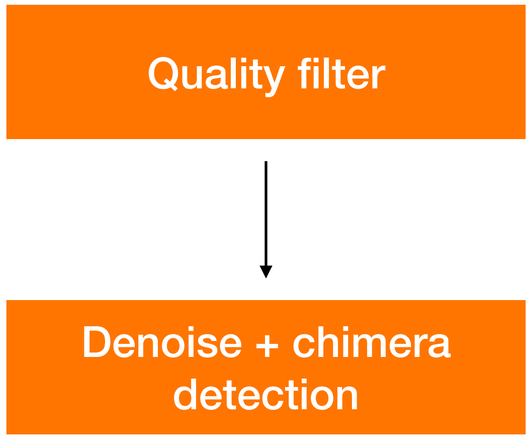
Quality filter

- Both primers present
- Max no. of expected errors = 4
- Remove too short and too long sequences

Tools used: command line tools, mafft, VSEARCH, DADA2, BLAST+, barrnap, custom perl script, seqkit

https://github.com/burki-lab/Transitions

Custom pipeline for processing PacBio data (18S-28S)



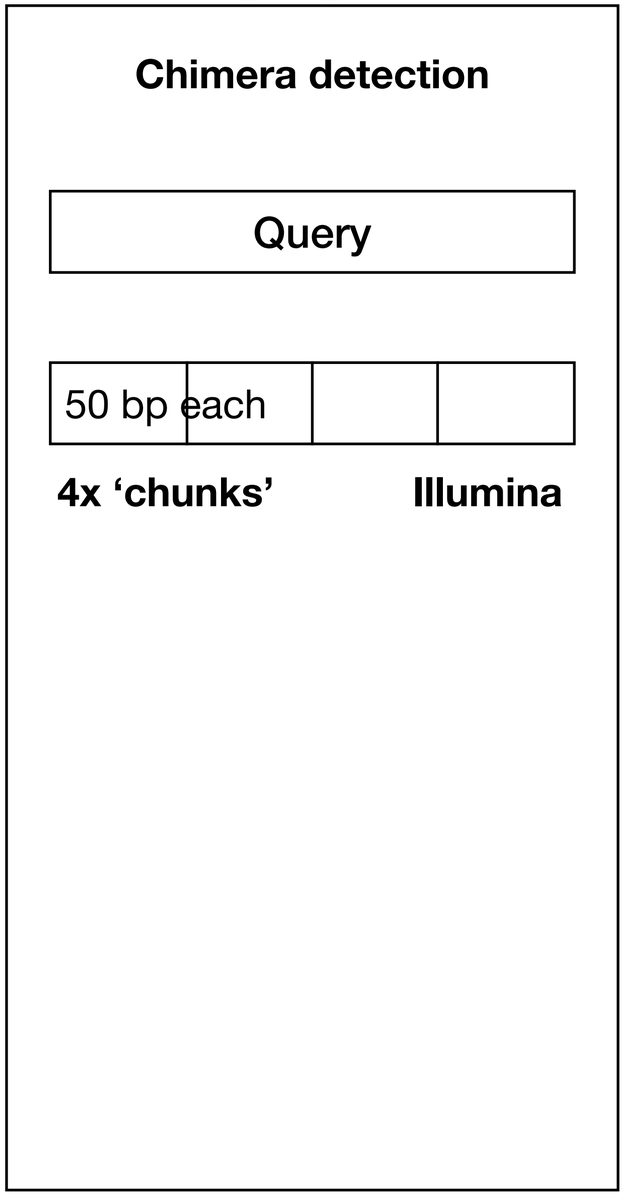
- Both primers present
- Max no. of expected errors = 4
- Remove too short and too long sequences

- Denoise by clustering at 99% similarity
- Remove prokaryotes
- *Denovo* chimera detection

Tools used: command line tools, mafft, VSEARCH, DADA2, BLAST+, barrnap, custom perl script, seqkit

https://github.com/burki-lab/Transitions

Custom pipeline for processing PacBio data (18S-28S)



Quality filter

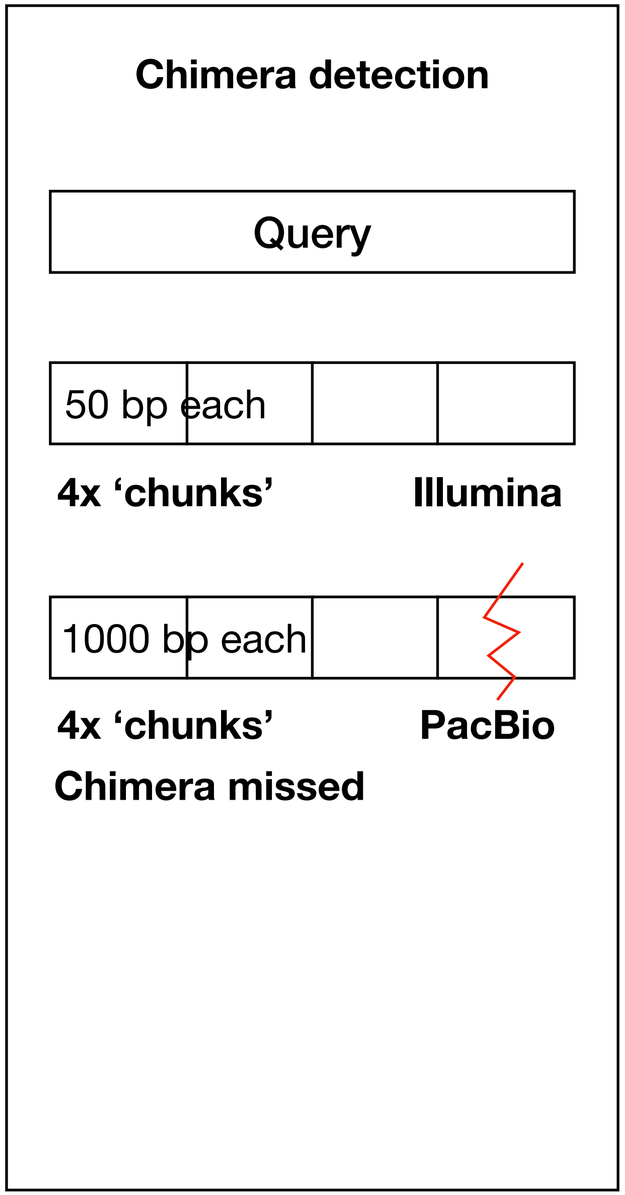
- Both primers present
- Max no. of expected errors = 4
- Remove too short and too long sequences

Denoise + chimera detection

- Denoise by clustering at 99% similarity
- Remove prokaryotes
- *Denovo* chimera detection

https://github.com/burki-lab/Transitions

Custom pipeline for processing PacBio data (18S-28S)



Quality filter

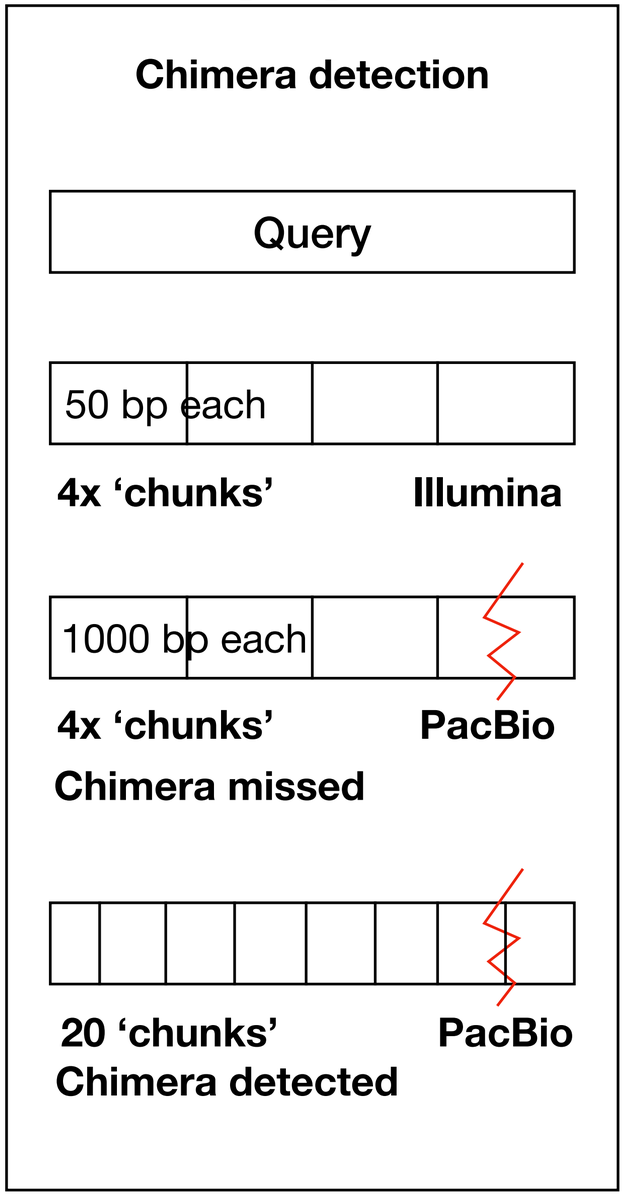
- Both primers present
- Max no. of expected errors = 4
- Remove too short and too long sequences

Denoise + chimera detection

- Denoise by clustering at 99% similarity
- Remove prokaryotes
- *Denovo* chimera detection

https://github.com/burki-lab/Transitions

Custom pipeline for processing PacBio data (18S-28S)



Quality filter

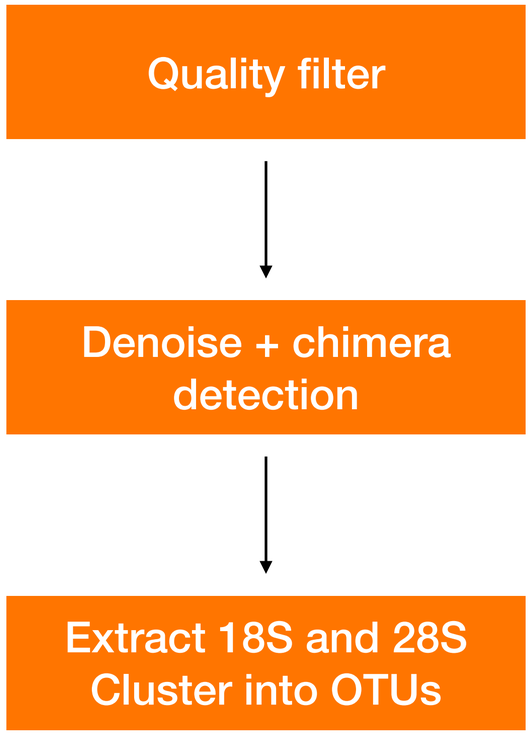
- Both primers present
- Max no. of expected errors = 4
- Remove too short and too long sequences

Denoise + chimera detection

- Denoise by clustering at 99% similarity
- Remove prokaryotes
- *Denovo* chimera detection

https://github.com/burki-lab/Transitions

Custom pipeline for processing PacBio data (18S-28S)



- Both primers present
- Max no. of expected errors = 4
- Remove too short and too long sequences

- Denoise by clustering at 99% similarity
- Remove prokaryotes
- *Denovo* chimera detection

- Cluster at 97% similarity based on 18S gene
- Remove singletons
- *Denovo* chimera detection (again!)

Tools used: command line tools, mafft, VSEARCH, DADA2, BLAST+, barrnap, custom perl script, seqkit

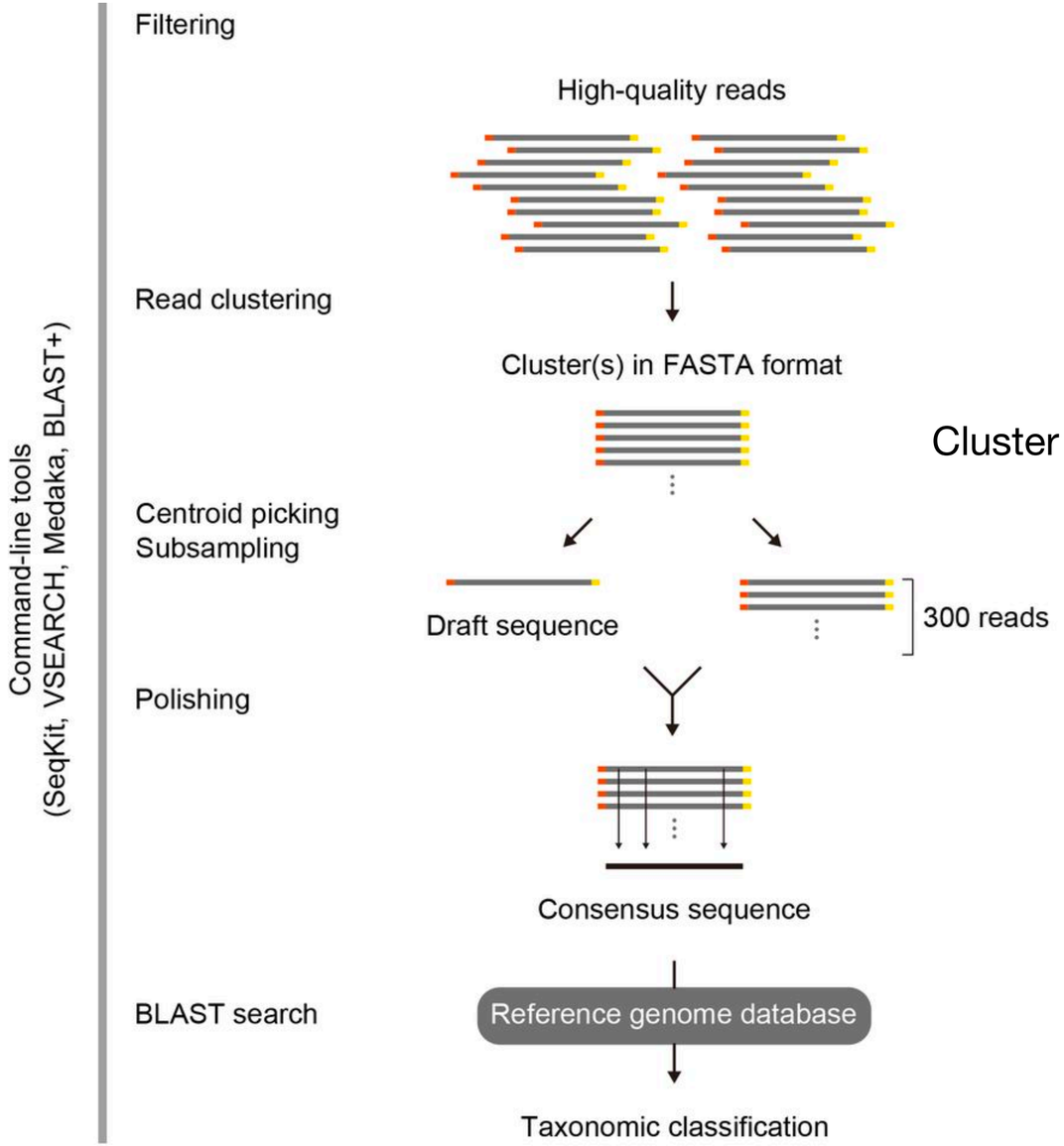
https://github.com/burki-lab/Transitions

Pipelines to process Nanopore data

DNA metabarcoding workflow utilizing nanopore long-read sequencing and consensus generation for rapid identification of fungal taxa with high phylogenetic resolution

Atsufumi Ohta, Kenichiro Nishi, Kiichi Hirota, Yoshiyuki Matsuo

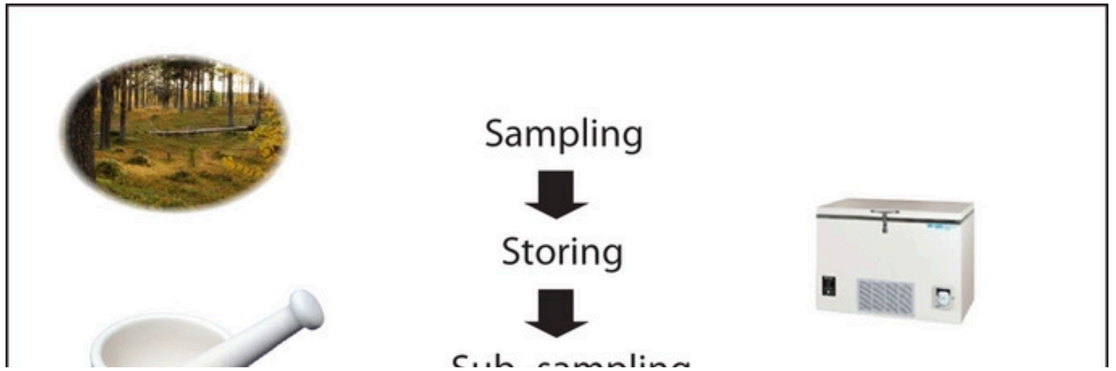
doi: https://doi.org/10.1101/2023.04.14.536971



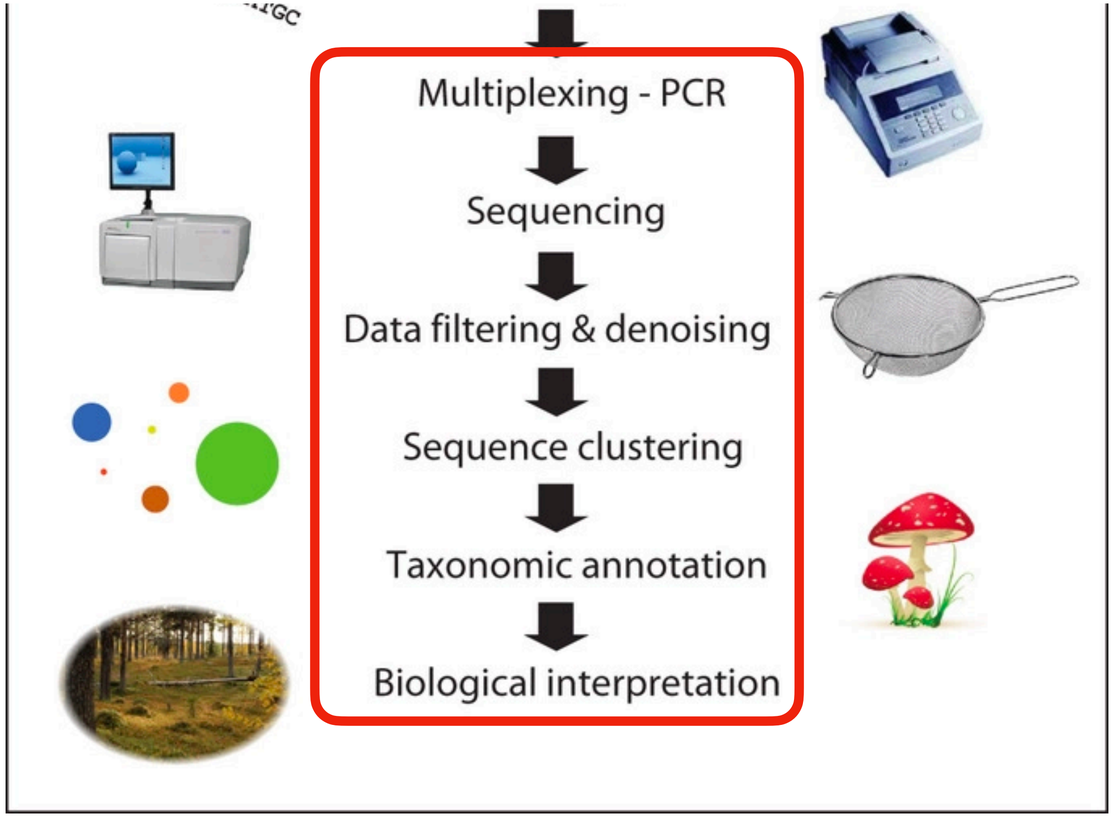
Cluster at 85-89% similarity

Any questions?

Any questions?



Long-read metabarcoding
with Unique Molecular Identifiers (UMIs)



Lindahl et al, 2013. *New Phytologist*

Unique molecular identifiers (UMIs)

- Illumina
  - Synthetic long-read sequencing
  - Commercial option (LoopSeq)
- Nanopore/PacBio
  - No commercial options available

Synthetic long-reads

Synthetic long reads

LoopSeq™ 16S Long Read Kit

$1,600.00

Quantity

1

Add to Cart

SHIPPING INFO −

$100 to any US desitination

PRODUCT INFO +

Quote +

Research | Open Access | Published: 05 June 2021

Ultra-accurate microbial amplicon sequencing with synthetic long reads

Benjamin J. Callahan, Dmitry Grinevich, Siddhartha Thakur, Michael A. Balamotis & Tuval Ben Yehezkel

Microbiome **9**, Article number: 130 (2021) | Cite this article

9540 Accesses | **21** Citations | **45** Altmetric | Metrics

LoopSeq

Attach.

Every sample is exposed to millions of unique barcodes, but only one barcode attaches per strand of DNA at the 16S site.

Each molecule is tagged with a unique barcode (UMI)

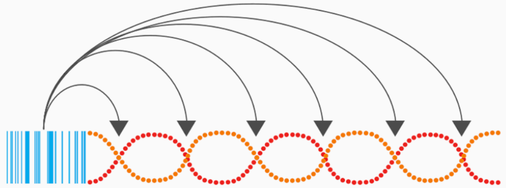
Amplify.

Every molecule, along with its unique barcode, is amplified using PCR.

It’s amplified

Distribute.

For each molecule copy, the barcode is randomly distributed within the molecule.



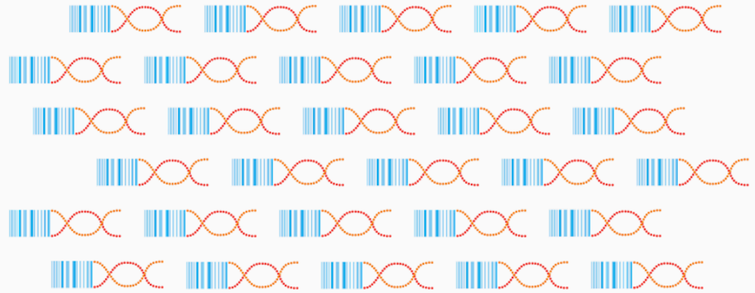
The barcode is inserted at various points in the SAME molecule

LoopSeq

Regular Illumina sequencing

Sequence.

Sequence the segment next to each barcode.



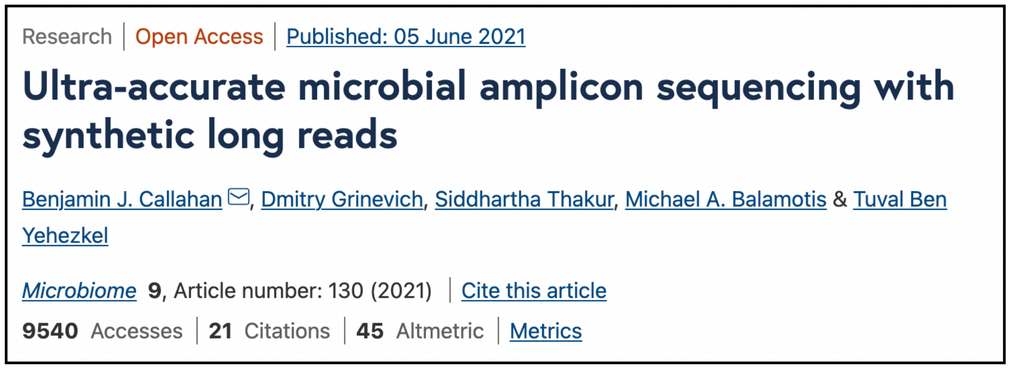
Assemble into long-reads using barcodes

Assemble.

Short reads that share the same barcode are combined algorithmically into a full-length molecule using linked-read de novo assembly.

Processing LoopSeq data

- Dedicated DADA2 workflow



Research | Open Access | Published: 05 June 2021

Ultra-accurate microbial amplicon sequencing with synthetic long reads

Benjamin J. Callahan, Dmitry Grinevich, Siddhartha Thakur, Michael A. Balamotis & Tuval Ben Yehezkel

Microbiome **9**, Article number: 130 (2021) | Cite this article

9540 Accesses | **21** Citations | **45** Altmetric | Metrics

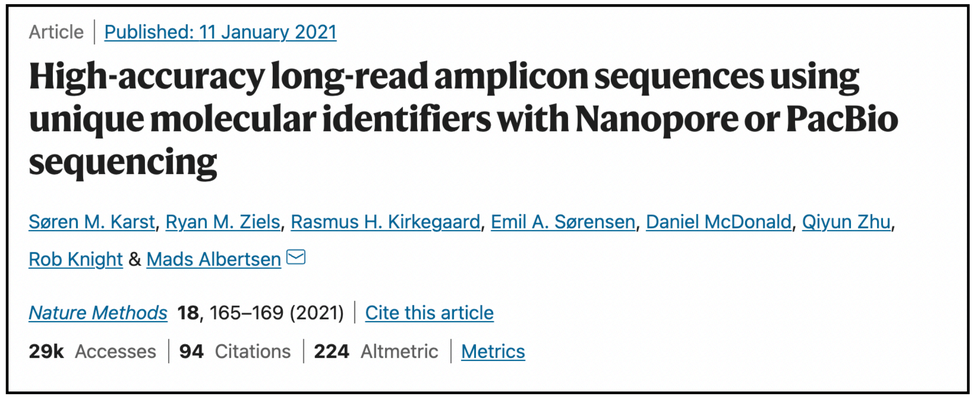
- Low rate of chimera formation since each molecule is being amplified with unique barcoded primers.

LoopSeq

- Error rate. Very low. 0.005% error rate per nucleotide
- Chimera rate. Low
- Cost. Kit for 1600 dollars. Plus Illumina sequencing.
- The type of community you want to sequence. Complex community
- Sequence length/marker. 16S (prok). 18S-ITS1-ITS2 (fungal)

Combining UMIs with Nanopore/PacBio

- Generates highly accurate single-molecule consensus sequences (0.0007-0.004% error rates)
- Low chimera rate (0.02%)
- Promising technique but rather difficult to implement at the moment (unless you’re prepared to spend a lot of time optimising the different lab work steps)



Article | Published: 11 January 2021

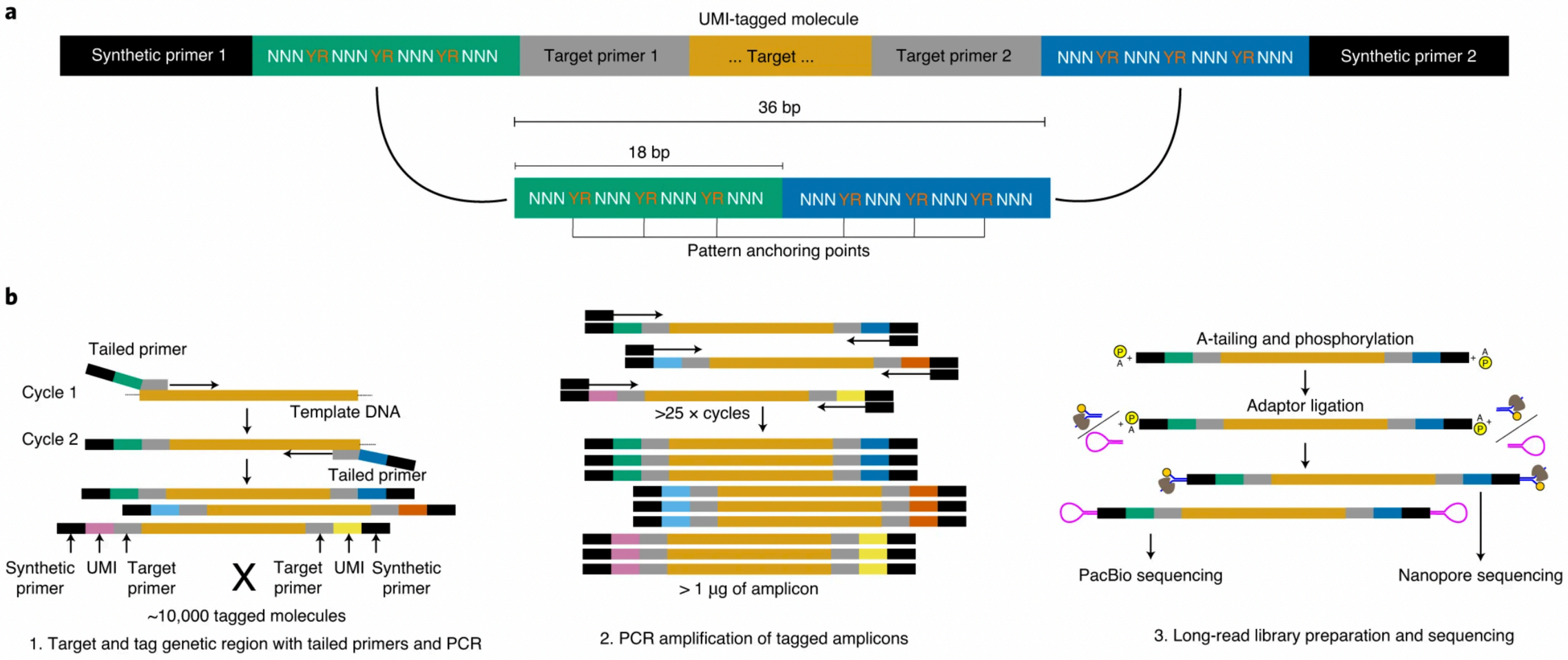
High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing

Søren M. Karst, Ryan M. Ziels, Rasmus H. Kirkegaard, Emil A. Sørensen, Daniel McDonald, Qiyun Zhu, Rob Knight & Mads Albertsen

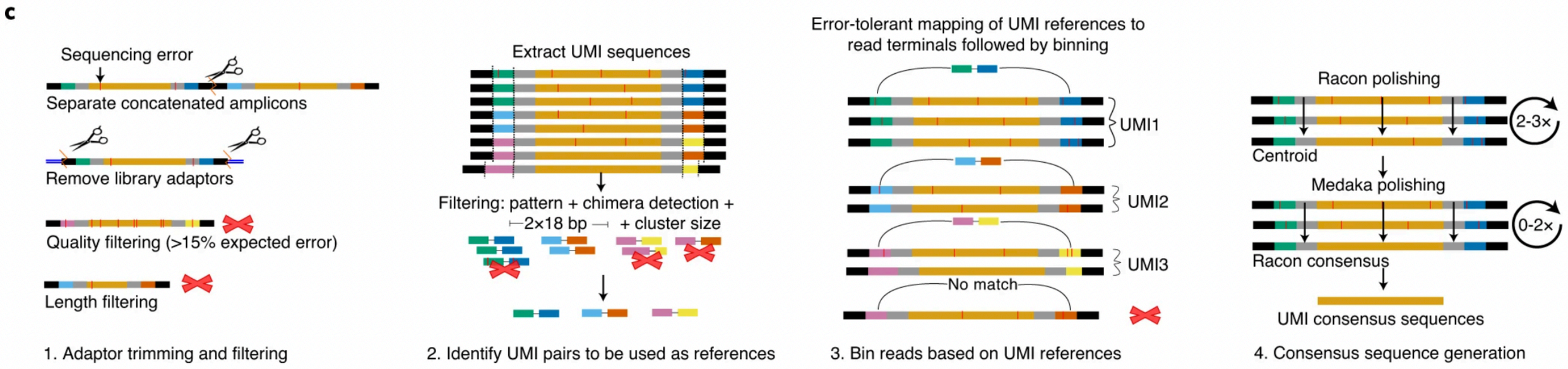
Nature Methods **18**, 165–169 (2021) | Cite this article

29k Accesses | **94** Citations | **224** Altmetric | Metrics

Combining UMIs with Nanopore/PacBio



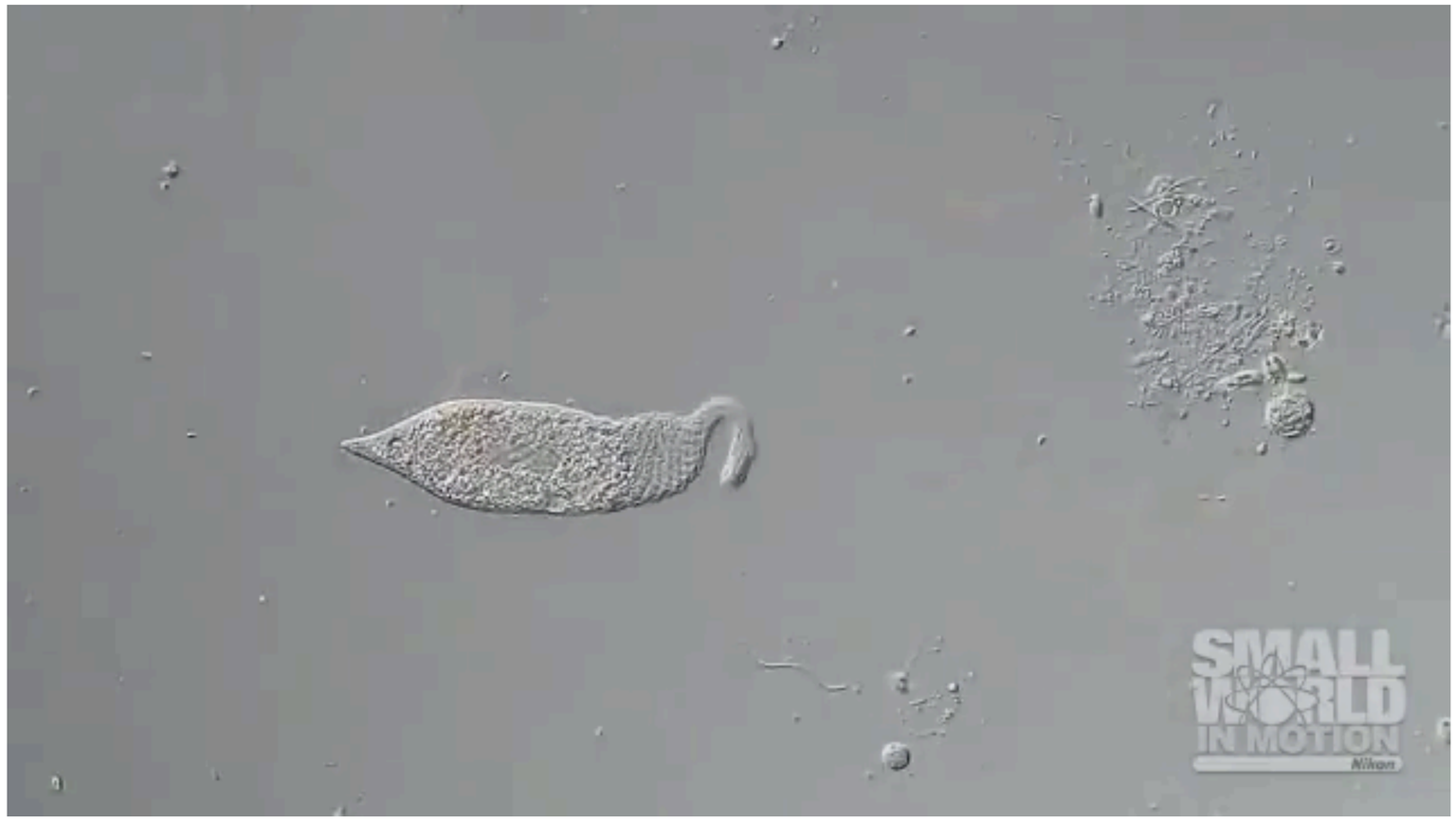
Processing data



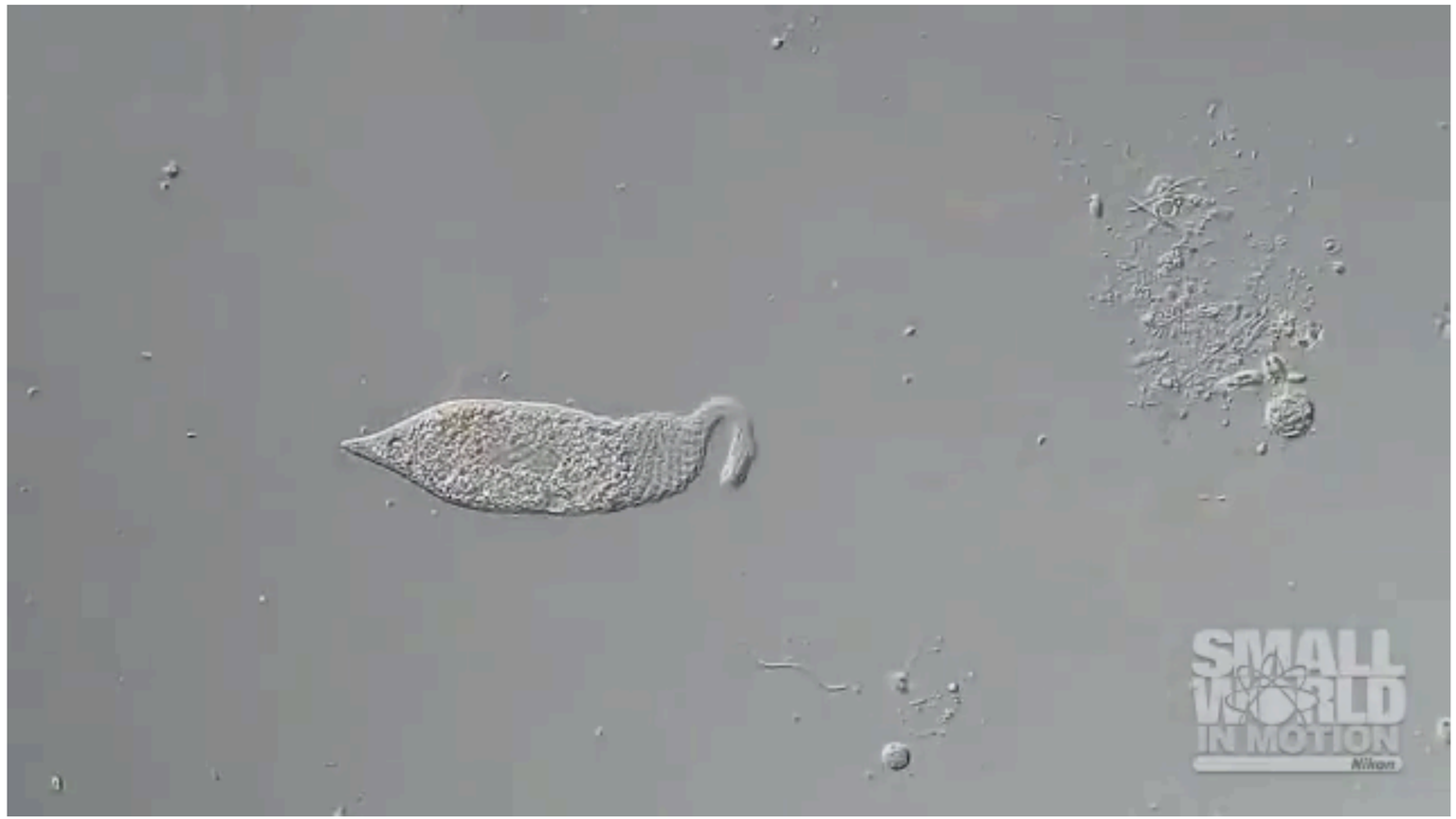
Combining UMIs with Nanopore/PacBio

- Error rate. Very low. 0.0007-0.004% error rate
- Chimera rate. Low
- Effort to set up method for first time. High
- Cost. Expensive
- The type of community you want to sequence. Complex community
- Sequence length/marker. Any!

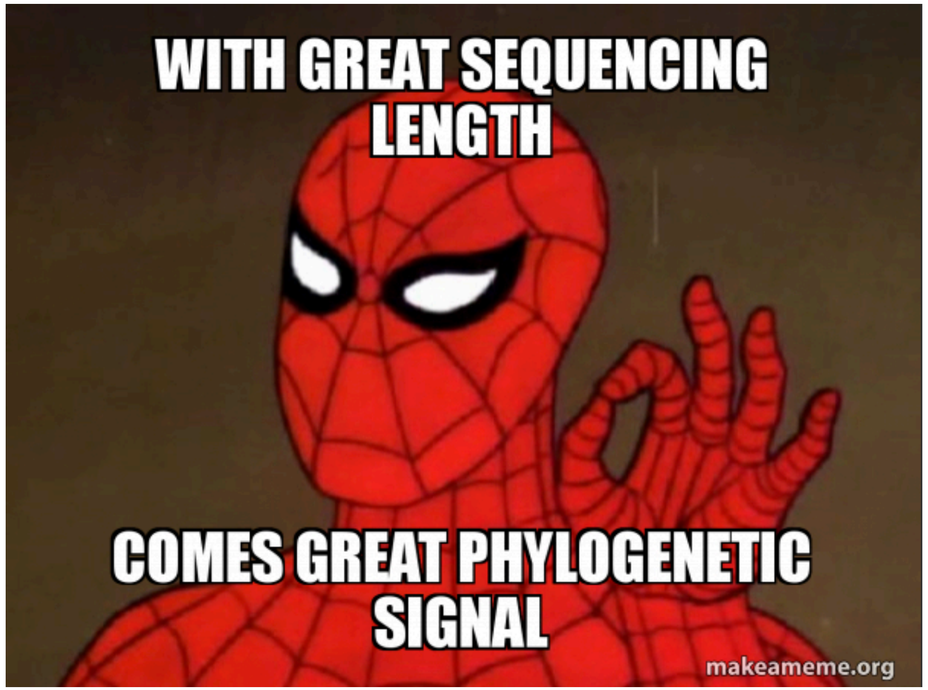
Any questions?



Any questions?



Take-home



Up and coming method. Expect lots of development in the field!!