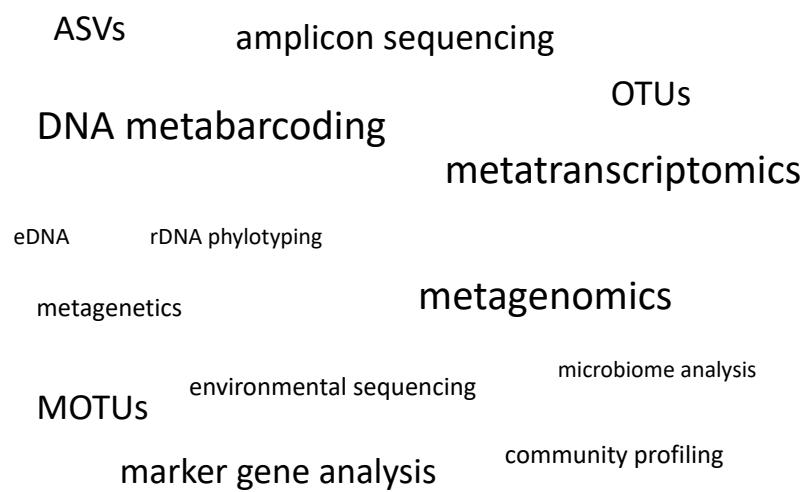


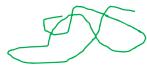
## Introduction to DNA metabarcoding

- Explain some terms
- Introduce key steps
- Introduce some literature

## Some confusing / important terms

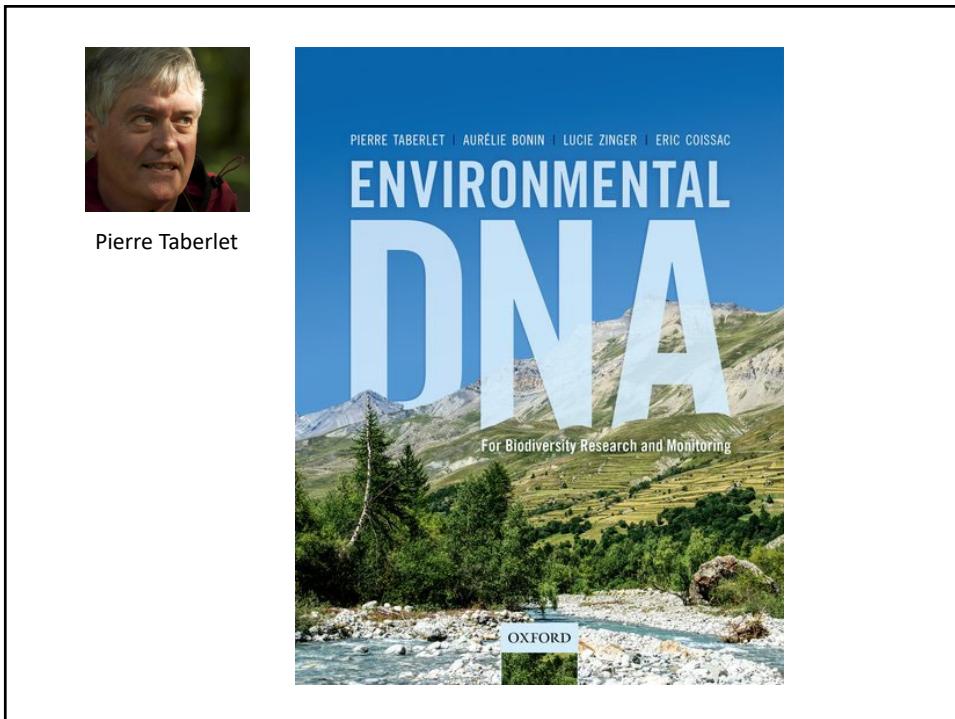
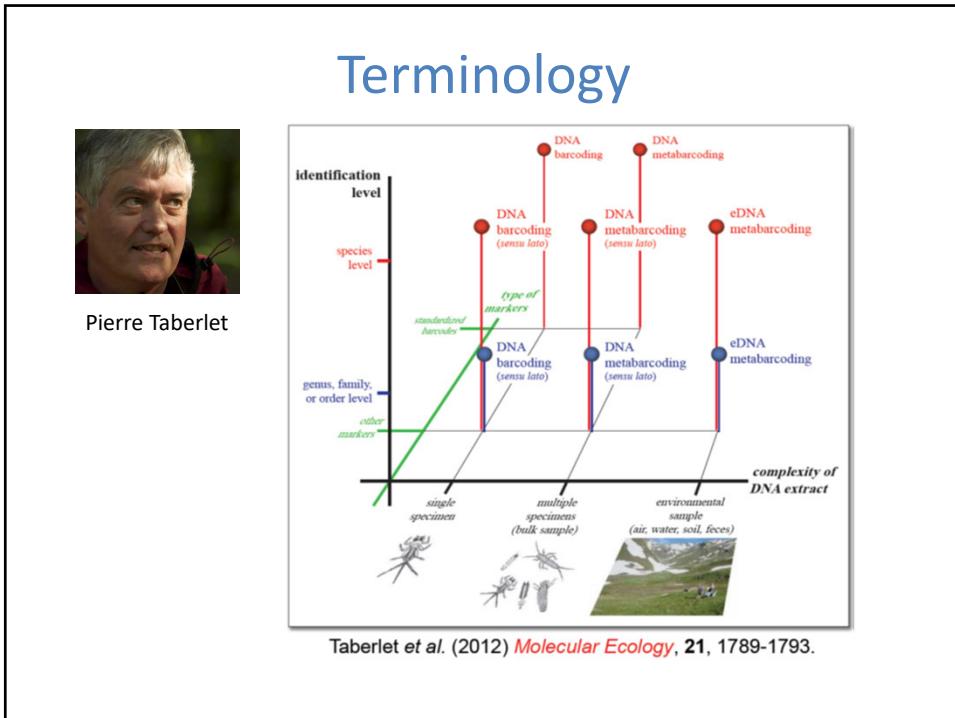


## Some important terms

- DNA barcoding  → Sequence variation in a single locus (e.g. ITS) in a single specimen
- Metabarcoding  → Sequence variation in a single locus (e.g. ITS) in a community  
Mixed DNA templates
- Metagenomics  → Genome wide sequence variation in a community  
Mixed DNA templates
- Metatranscriptomics  → cDNA sequence variation in a community  
Mixed RNA

## Some important terms

- Metabarcoding  → Sequence variation in a single locus (e.g. 16S)  
Mixed DNA templates Who are there?
- Metagenomics  → Genome wide sequence variation  
Mixed DNA templates Which genes (and who) are there?
- Metatranscriptomics  → cDNA sequence variation  
Mixed RNA Who are active and doing what?



## DNA metabarcoding

- Research questions:
  - Who are there?
  - Richness: How many taxa / operational taxonomic units (alpha/gamma diversity)?
  - Compositional differences (beta diversity)?
  - Which processes and drivers are shaping the communities?
  - Co-occurrence patterns → Interactions
- Qualitative data

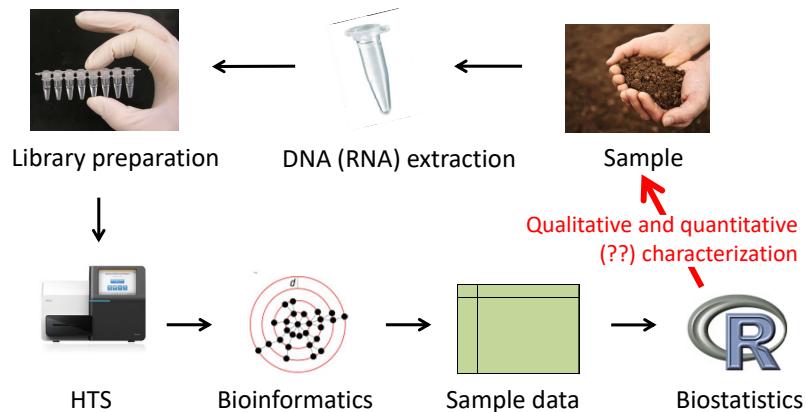


## DNA metabarcoding

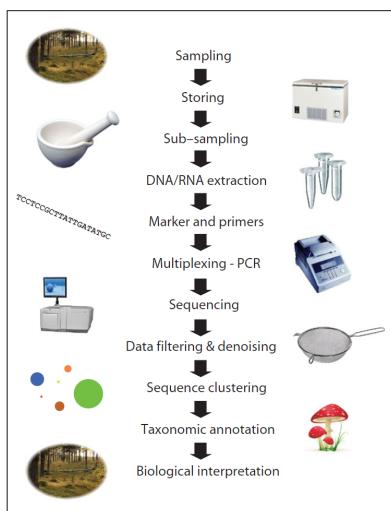
- Research questions:
  - Who are there?
  - Richness: How many taxa / operational taxonomic units (alpha/gamma diversity)?
  - Compositional differences (beta diversity)?
  - Which processes and drivers are shaping the communities?
  - Co-occurrence patterns → Interactions
- Qualitative data
- Quantitative data
  - Relative abundances only...



## General workflow in DNA metabarcoding studies

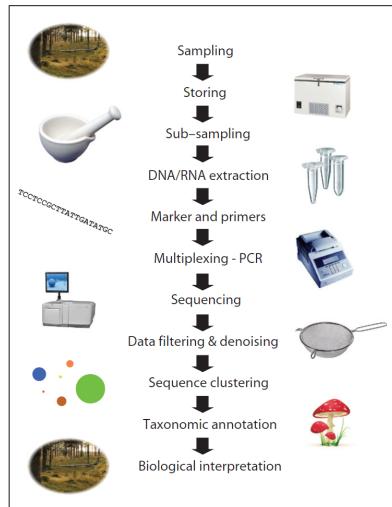


## DNA metabarcoding - many steps



Lindahl et al. 2013

## DNA metabarcoding - many steps



Lindahl et al. 2013

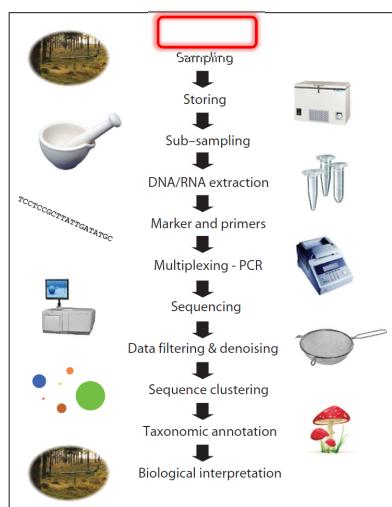


Many steps...



... to go wrong

## DNA metabarcoding - many steps



Lindahl et al. 2013



Many steps...



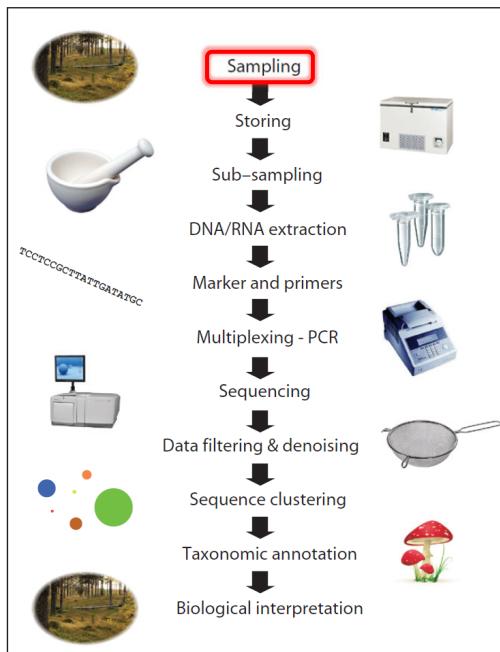
... to go wrong

## If new study system – conduct a pilot?

- Which sampling scheme?
- How many replicates?
- Which extraction protocol
- Which primers?
- Which sequencing depth?
- Which sequencing technique?
- Etc.



→ Depends on the alpha, beta and gamma diversity  
(which you might not know anything about..).



## Representativeness (in space)

- Many communities highly heterogenous
- Should obtain samples that are representative
- If you are not interested in the small scale variation in itself → pool sub-samples?



Fig. 5. What is the optimal relationship between primary sample size and the analytical sample volume (insert) and how can it come about? When sample size increases one can intuitively understand that the sample becomes more representative. But at the same time, today's analytical volumes continue to decrease (insert) as the analytical instruments become more and more precise. For all heterogeneous materials, there is consequently an intrinsic contradiction between primary sampling representativity and the instrumental analytical volume requirements. This is the root cause of all sampling and representativity issues.

## Representativeness (in time)

- Many communities often display high temporal variation! Repeated temporal sampling?
- When comparing across different areas: May use phenological markers to obtain the 'same' temporal phase.



**«Replicate or lie»**

**Opinion**  
**Replicate or lie**

James P. Prosser<sup>1</sup>  
Institute of Biological and Environmental Sciences,  
University of Aberdeen, Crookston Building, St.  
Mary's Street, Aberdeen, AB22 2PU, UK.

**Introduction**  
Andren and colleagues (2005) recently published a paper in this journal (Environmental Microbiology) that I have cited and quoted during my early years of research, but I am afraid that it has now become one of the most highly cited papers in the majority of microbial ecology literature. This is despite the fact that the basic premises of statistical analysis and its significance and relevance to microbial ecology have been well known to this problem, and my authority of the consequences, since the 1980s. The paper in question concerns the relationship between measured diversity and the number of species, and the relationship between measured diversity and the number of individuals. The authors conclude that measured diversity is best explained by the number of individuals, and not species richness. The paper also claims that the relationship between measured diversity and species richness is non-significant. The paper is based on a comparison of two methods, beyond these techniques and beyond studies of microorganisms.

**Why replicate?**  
The main point of the journal could be that we articles describing and discussing applications of statistical analysis in environmental microbiology should have a clear base and fundamental aspect – the need for replicates. In general, replicates are required to estimate the variance observed, without to concern biological abundance. In this case, the authors used a sample size of 100 samples, which they say was enough, giving values of  $4 \times 10^3$  and  $3.2 \times 10^3$  individuals. One might think that the first value is greater than one like the other. However, the authors say that they had to remove the outliers from their basic statistics because of overdispersion among the data.

Received 1 December 2004, accepted 2 January 2005. "Replicate or lie" is available online at [www.blackwell-science.com](http://www.blackwell-science.com). © 2005 Blackwell Publishing Ltd

**Clone library analysis and pyrosequencing**

	Number of articles	% with replicates
<i>Appl Environ Microbiol</i>	60	23
<i>Environ Microbiol</i>	47	15
<i>FEMS Microbiol Ecol</i>	29	24
<i>ISME J</i>	23	13
<i>Microbial Ecol</i>	22	9
Total	181	18

It doesn't help that you  
are dealing with HTS data if  
you don't replicate properly!

Prosser JI. 2010, Environmental Microbiology

**Biological replicates**

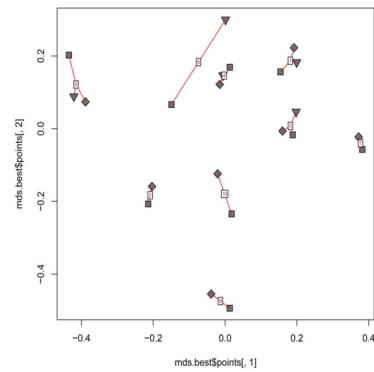
**Fungal communities associated with mosses in different forest management types**

Davey et al. 2014. FEMS Microbial Ecology

## Technical replicates

- Some samples could/should be analyzed multiple times
- Reveal the variability (experimental error) of the analysis technique → allows to set limits for what is meaningful and significant data

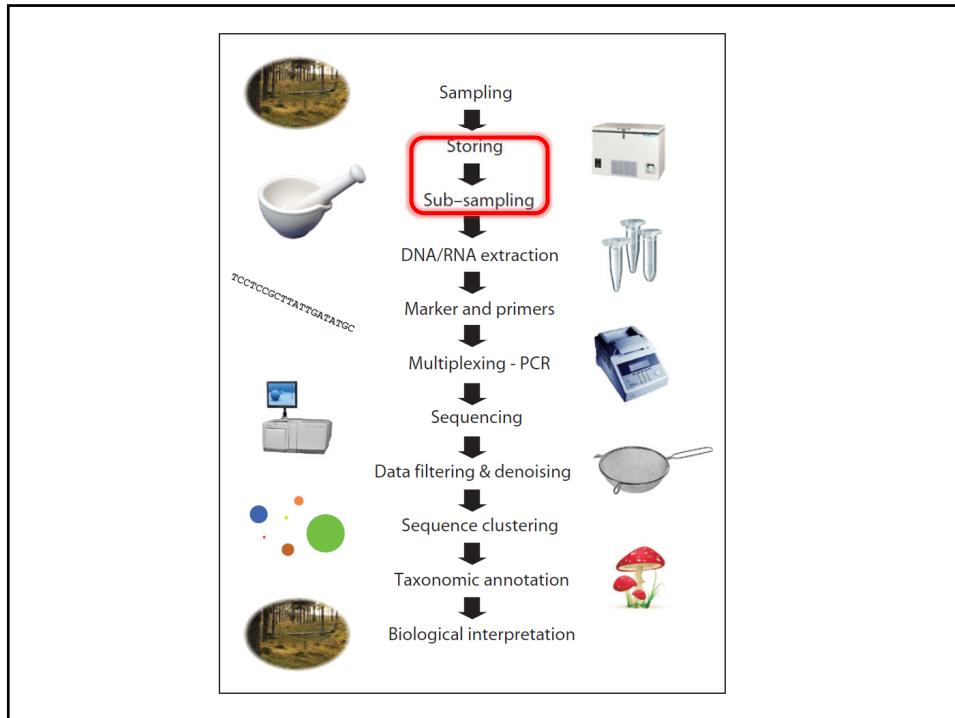
Communities with low DNA content!!



Davey *et al.* 2014. FEMS Microbial Ecology

## Sample types

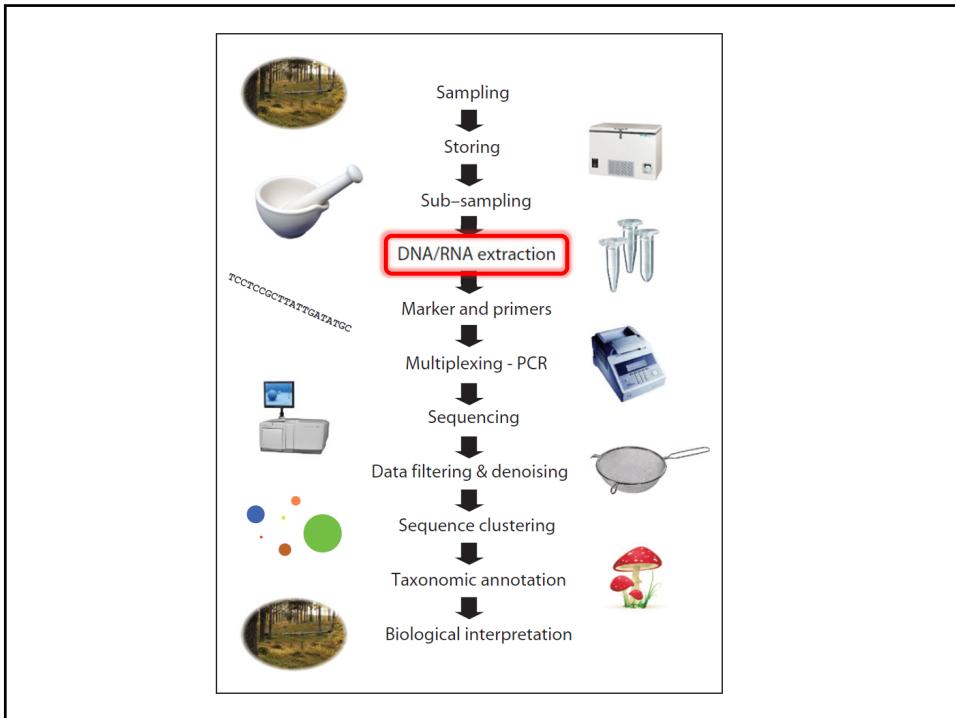
1. Biological replicates
2. Technical replicates



## Storing

- Unappropriate storage may introduce severe biases!
- Community members can respond quickly to altered conditions
- 'Arrest' the communities!
- Process the samples asap. If needed, long time storage at -80C often suggested





## DNA extraction

**An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil**  
 S.M. Dineen<sup>1,2</sup>, R. Aranda IV<sup>1,2</sup>, D.L. Anders<sup>3</sup> and J.M. Robertson<sup>2</sup>

<sup>1</sup> Visiting Scientist, Federal Bureau of Investigation Laboratory, Quantico, VA, USA  
<sup>2</sup> Counterterrorism and Foreign Science Research Unit, Federal Bureau of Investigation Laboratory, Quantico, VA, USA  
<sup>3</sup> Hazardous Materials Science Response Unit, Federal Bureau of Investigation Laboratory, Quantico, VA, USA

**Influence of DNA extraction and PCR amplification on studies of soil fungal communities based on amplicon sequencing**  
 Lihui Xu, Sabine Ravanska, John Larsson, and Magnus Nilssen\*

**Molecular biology, genetics and biotechnology**  
**Effect of DNA extraction and sample preservation method on rumen bacterial population**  
 Katerina Fliegerova<sup>a,b,\*</sup>, Ilma Tapiö<sup>b</sup>, Aurelie Bonin<sup>c</sup>, Jakub Mrazek<sup>a</sup>, Maria Luisa Callegari<sup>c</sup>, Paolo Ratti<sup>c</sup>, Alireza Bayat<sup>d</sup>, Johanna Vilkkki<sup>c</sup>, Jan Kopečný<sup>a</sup>, Kevin J. Shilling<sup>c,d</sup>, Frederic Boyer<sup>c</sup>, Eric Coissac<sup>c</sup>, Pierre Taberlet<sup>c</sup>, R. John Wallace<sup>c</sup>

**Effect of DNA Extraction Methods and Sampling Techniques on the Apparent Structure of Cow and Sheep Rumen Microbial Communities**  
 Gemma Henderson<sup>1</sup>, Faith Cox<sup>1</sup>, Sanna Kuitiainen<sup>1</sup>, Vahideh Heidarian Miri<sup>2</sup>, Michael Zehfouj<sup>2</sup>, Samantha J. Noel<sup>3</sup>, Garry C. Waghorn<sup>2</sup>, Peter H. Jansson<sup>1</sup>

**The Impact of Different DNA Extraction Kits and Laboratories upon the Assessment of Human Gut Microbiota Composition by 16S rRNA Gene Sequencing**  
 Nicholas A. Kennedy<sup>1</sup>, Alan W. Walker<sup>2</sup>, Susan H. Berry<sup>3</sup>, Sylvia H. Duncan<sup>4</sup>, Freda M. Farquharson<sup>4</sup>, Petra Louis<sup>5</sup>, John M. Thomson<sup>5</sup>, UK IBD Genetics Consortium<sup>6</sup>, Jack Satsangi<sup>1</sup>, Harry J. Flint<sup>6</sup>, Julian Parkhill<sup>2</sup>, Charlie W. Lee<sup>1</sup>, Georgina L. Hold<sup>1</sup>

\* indicates author for correspondence

## DNA extraction

- Should yield high and uniform amounts of DNA
- Concentration of PCR inhibitors minimized
- Same protocol for all samples!
- If no proper literature are available on your study system → conduct a pilot?!
- Extraction negatives!



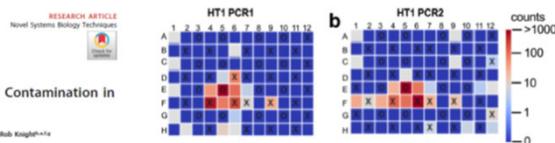
- MoBio Power Soil?
- FastDNA kit for Soil?
- EZNA Soil kit?
- CTAB + cleanup kit?

## DNA extraction



### Quantifying and Understanding Well-to-Well Contamination in Microbiome Research

Jeremiah J. Minich,<sup>a</sup> Jon G. Sanders,<sup>b</sup> Amnon Amir,<sup>b</sup> Greg Humphrey,<sup>b</sup> Jack A. Gilbert,<sup>a,f</sup> Rob Knight<sup>a,b,e</sup>

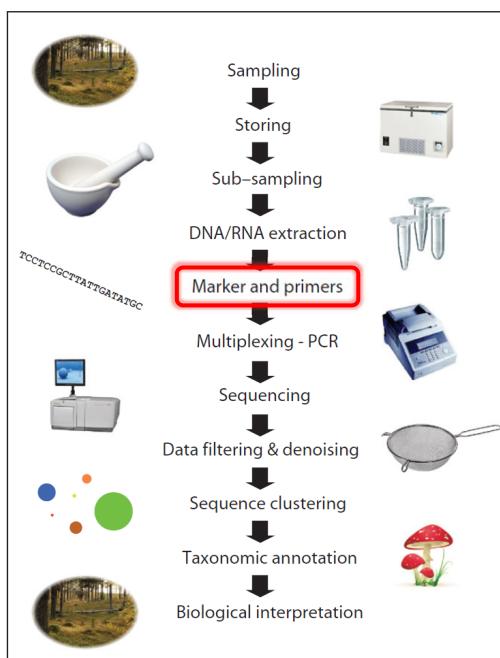


**ABSTRACT** Microbial sequences inferred to belong to one sample may not have originated from that sample. Such contamination may arise from laboratory reagent sources or from physical exchange between samples. This study seeks to rigorously assess the behavior of this often-neglected between-sample contamination. Using unique bacteria, each assigned a particular well in a plate, we assess the frequency at which sequences from each source appear in other wells. We evaluate the effects of different DNA extraction methods performed in two laboratories using a consistent plate layout, including blanks and low-biomass and high-biomass samples. Well-to-well contamination occurred primarily during DNA extraction and, to a lesser extent, in library preparation, while barcode leakage was negligible. Laboratories differed in the levels of contamination. Extraction methods differed in their occurrences and levels of well-to-well contamination, with plate methods having more well-to-well contamination and single-tube methods having higher levels of background contaminants. Well-to-well contamination occurred primarily in neighboring samples, with rare events up to 10 wells apart. This effect was greatest in samples with lower biomass and negatively impacted metrics of alpha and beta diversity. Our work emphasizes that sample contamination is a combination of cross talk from nearby wells and background contaminants. To reduce well-to-well effects, samples should be randomized across plates, samples of similar biomasses should be processed together, and manual single-tube extractions or hybrid plate-based cleanups should be employed. Researchers should avoid simplistic removals of taxa or operational taxonomic units (OTUs) appearing in negative controls, as many will be microbes from other samples rather than reagent contaminants.

Minich et al. 2019, mSystems

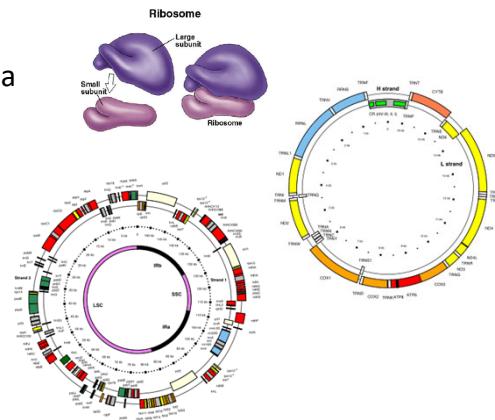
## Sample types

1. Biological replicates
2. Technical replicates
3. Extraction negatives



## Markers used in DNA metabarcoding

- Standard markers (>500 bp):
  - 18S: Eukaryotes
  - 16S: Bacteria/archaea
  - ITS: Fungi & plants
  - COI: Metazoa
  - *RbcL*: Plants
  - *trnL*: Plants



## Markers in DNA metabarcoding

- The ideal marker should:
  - Have primer sites that are shared by all target organisms
  - Be easy to amplify
  - Be of appropriate length for efficient amplification and sequencing
  - Be of similar length
  - No intragenomic variation (i.e. no paralogs)
  - Be possible to align
  - Have high interspecific variation
  - Have low intraspecific variation
- No known markers meet all these requirements!

## Markers in DNA metabarcoding

- The ideal marker should:
  - Have primer sites that are shared by all target organisms
  - Be easy to amplify
  - Be of appropriate length for efficient amplification and sequencing
  - Be of similar length
  - No intragenomic variation (i.e. no paralogs)
  - Be possible to align
  - Have high interspecific variation
  - Have low intraspecific variation
- No known markers meet all these requirements!

## ITS and primer bias

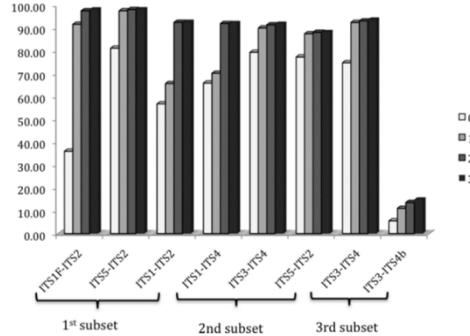
Bellemain et al. BMC Microbiology 2010, **10**:189  
<http://www.biomedcentral.com/1471-2160/10/189>



Open Access

### ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases

Eva Bellemain<sup>1</sup>, Tor Carlsen<sup>2</sup>, Christian Brochmann<sup>1</sup>, Eric Coissac<sup>3</sup>, Pierre Taberlet<sup>3</sup> and Håvard Kauserud<sup>2</sup>



Bellemain et al. 2010, BMC Microbiology

## Markers in DNA metabarcoding

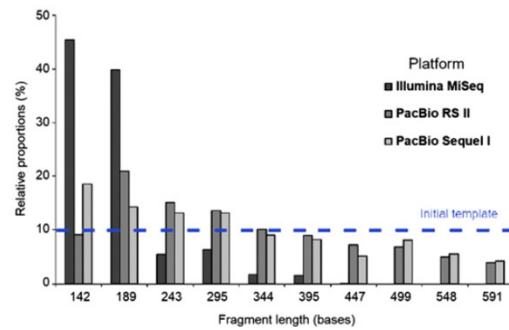
- The ideal marker should:
  - Have primer sites that are shared by all target organisms
  - Be easy to amplify
  - Be of appropriate length for efficient amplification and sequencing
  - Be of similar length
  - No intragenomic variation (i.e. no paralogs)
  - Be possible to align
  - Have high interspecific variation
  - Have low intraspecific variation
- No known markers meet all these requirements!



### Methods

Optimized metabarcoding with Pacific biosciences enables semi-quantitative analysis of fungal communities

Carles Castaño<sup>1</sup> , Anna Berlin<sup>1</sup> , Mikael Brandström Durling<sup>1</sup> , Katharina Ihrmark<sup>1</sup>, Björn D. Lindahl<sup>2</sup> , Jan Stenlid<sup>1</sup> , Karina E. Clemmensen<sup>1\*</sup> and Ake Olson<sup>1\*</sup>



## Markers in DNA metabarcoding

- The ideal marker should:
  - Have primer sites that are shared by all target organisms
  - Be easy to amplify
  - Be of appropriate length for efficient amplification and sequencing
  - Be of similar length
  - **No intragenomic variation (i.e. no paralogs)**
  - Be possible to align
  - Have high interspecific variation
  - Have low intraspecific variation
- No known markers meet all these requirements!

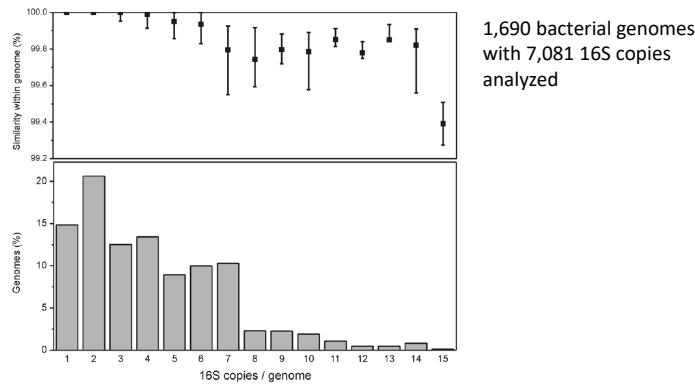
## (Intra)genomic variability in 16S

OPEN  ACCESS Freely available online PLOS ONE

### The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses

Tomáš Větrovský, Petr Baldrian\*

Laboratory of Environmental Microbiology, Institute of Microbiology of the Academy of Sciences of the Czech Republic, Praha, Czech Republic

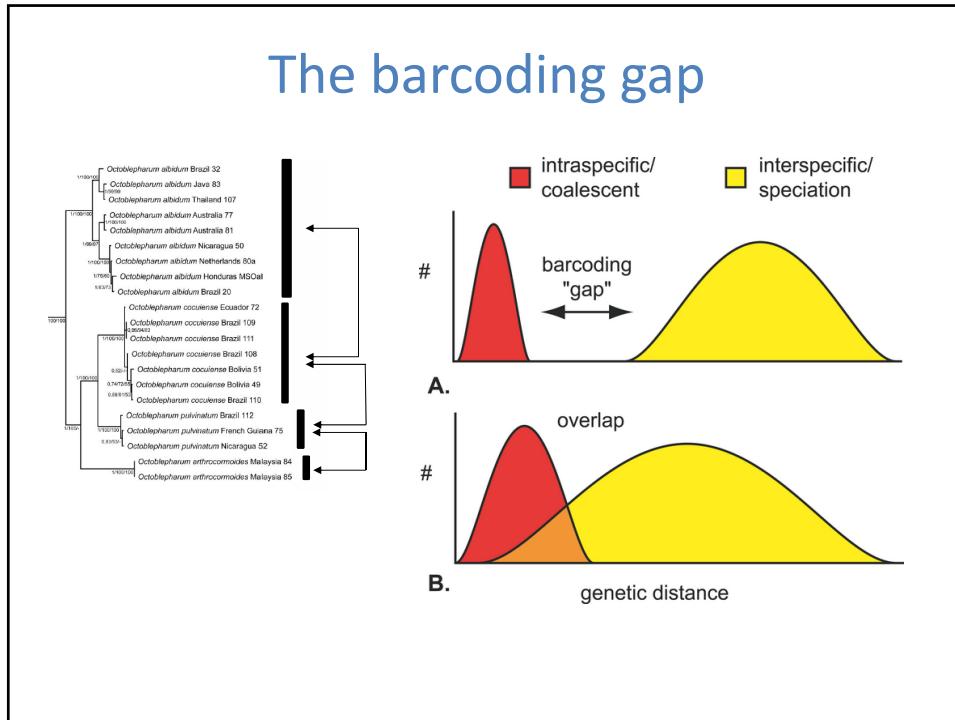


## Markers in DNA metabarcoding

- The ideal marker should:
  - Have primer sites that are shared by all target organisms
  - Be easy to amplify
  - Be of appropriate length for efficient amplification and sequencing
  - Be of similar length
  - No intragenomic variation (i.e. no paralogs)
  - Be possible to align
  - Have high interspecific variation
  - Have low intraspecific variation
- No known markers meet all these requirements!

## Markers in DNA metabarcoding

- The ideal marker should:
  - Have primer sites that are shared by all target organisms
  - Be easy to amplify
  - Be of appropriate length for efficient amplification and sequencing
  - Be of similar length
  - No intragenomic variation (i.e. no paralogs)
  - Be possible to align
  - Have high interspecific variation
  - Have low intraspecific variation
- No known markers meet all these requirements!

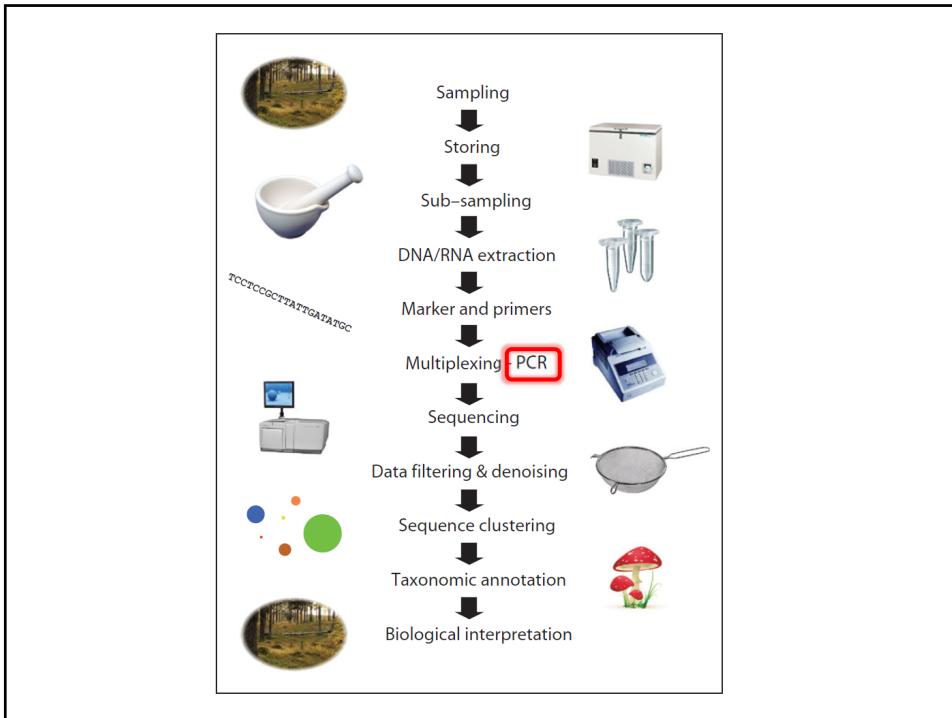


## How conserved/variable are the marker?

- 18S and 16S: Low variability, low intraspecific variation, low interspecific variation
- ITS: High variability, high intraspecific variation, high 'interspecific' variation

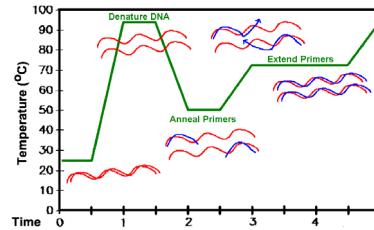


- Affects how some of the bioinformatics analyses should be conducted!! (i.e. no single way)



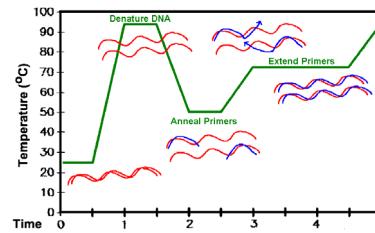
## PCR

- Different relevant factors during PCR:
  - Which polymerase enzyme (proofreading or not)?
  - Which RAMP speed?
  - How many cycles?
  - Which annealing temperature?
  - Multiple/replicate PCR reactions?
  - PCR negatives!



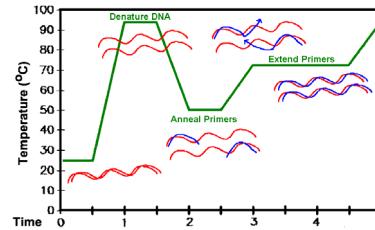
## PCR

- Different relevant factors during PCR:
  - Which polymerase enzyme (proofreading or not)?
  - Which RAMP speed?
  - How many cycles?
  - Which annealing temperature?
  - Multiple/replicate PCR reactions?
  - PCR negatives!



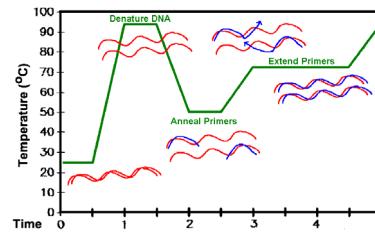
## PCR

- Different relevant factors during PCR:
  - Which polymerase enzyme (proofreading or not)?
  - Which RAMP speed?
  - **How many cycles?**
  - Which annealing temperature?
  - Multiple/replicate PCR reactions?
  - PCR negatives!



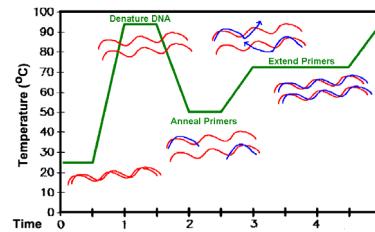
## PCR

- Different relevant factors during PCR:
  - Which polymerase enzyme (proofreading or not)?
  - Which RAMP speed?
  - How many cycles?
  - Which annealing temperature?
  - **Multiple/replicate PCR reactions?**
  - PCR negatives!



## PCR

- Different relevant factors during PCR:
  - Which polymerase enzyme (proofreading or not)?
  - Which RAMP speed?
  - How many cycles?
  - Which annealing temperature?
  - Multiple/replicate PCR reactions?
  - **PCR negatives!**

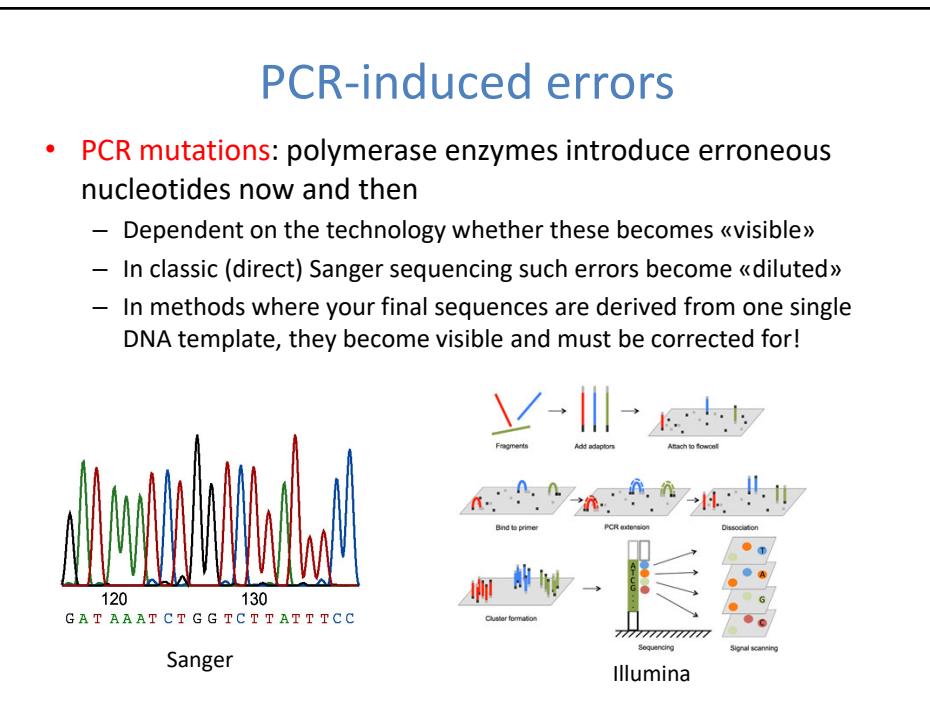
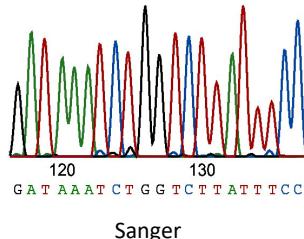


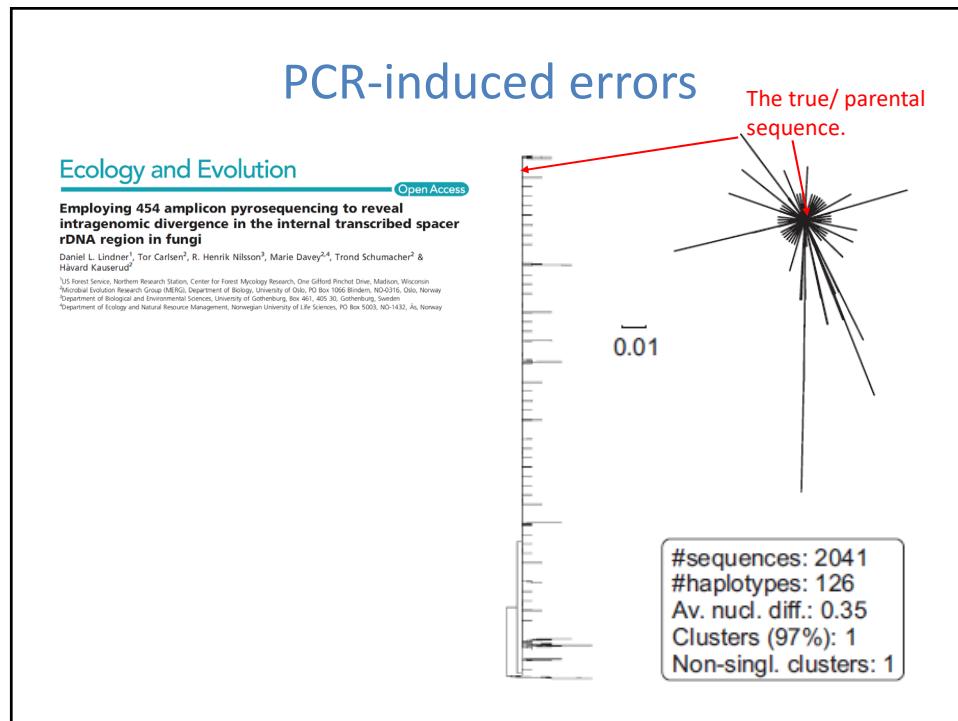
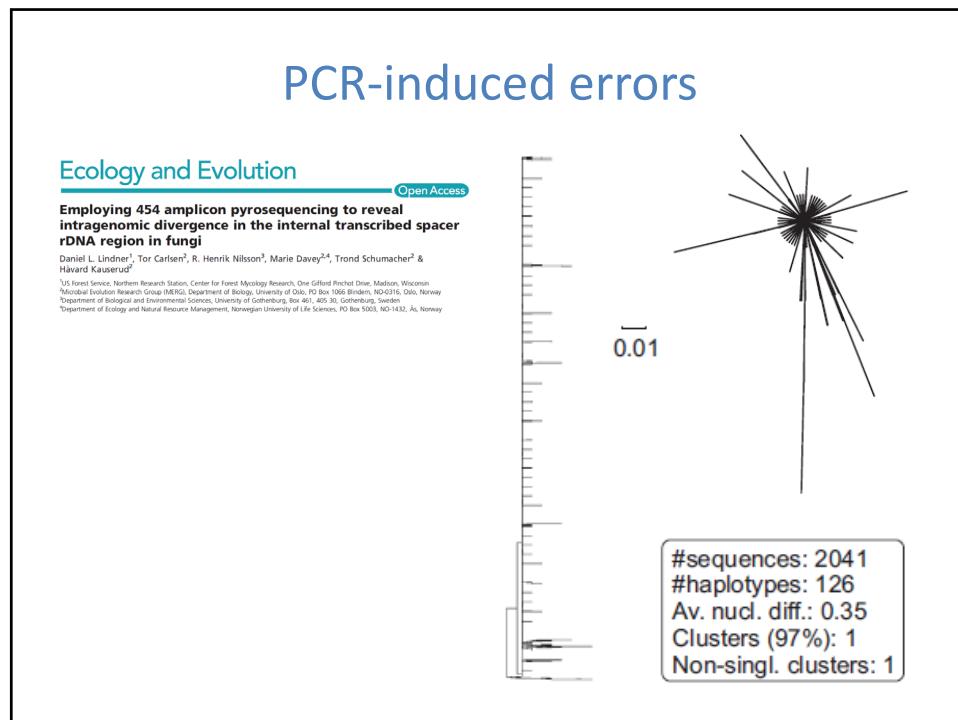
## Sample types

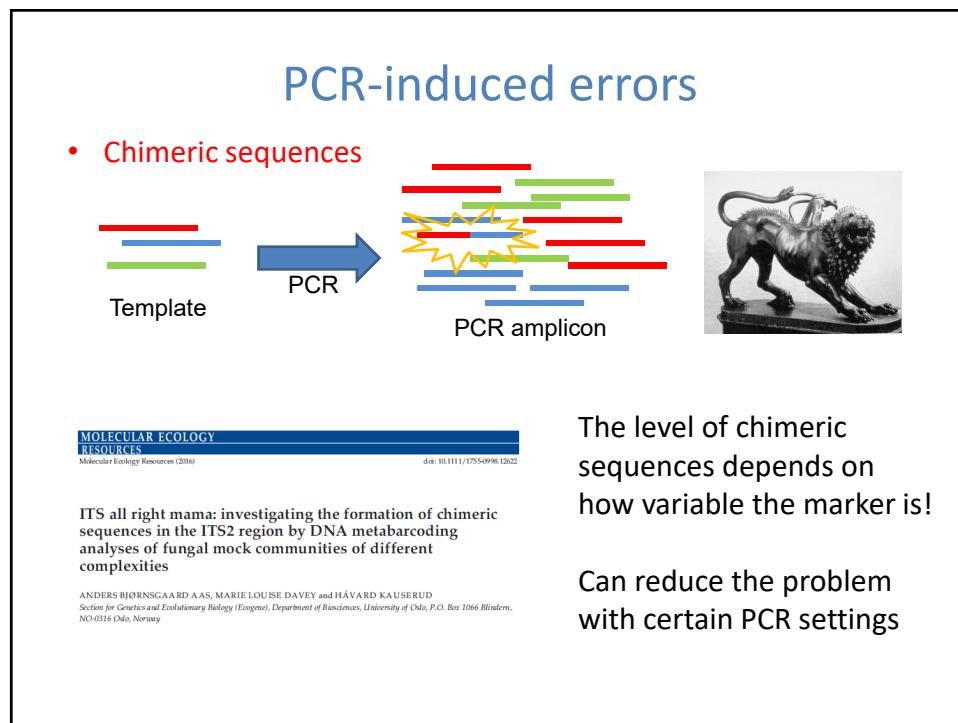
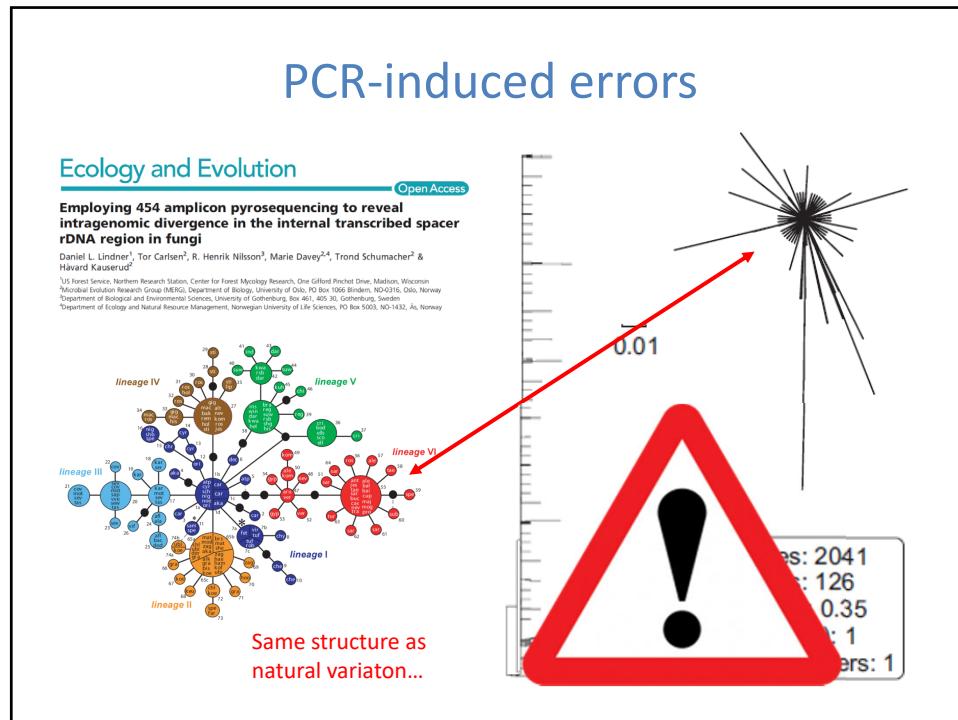
1. Biological replicates
2. Technical replicates
3. Extraction negatives
4. PCR negatives

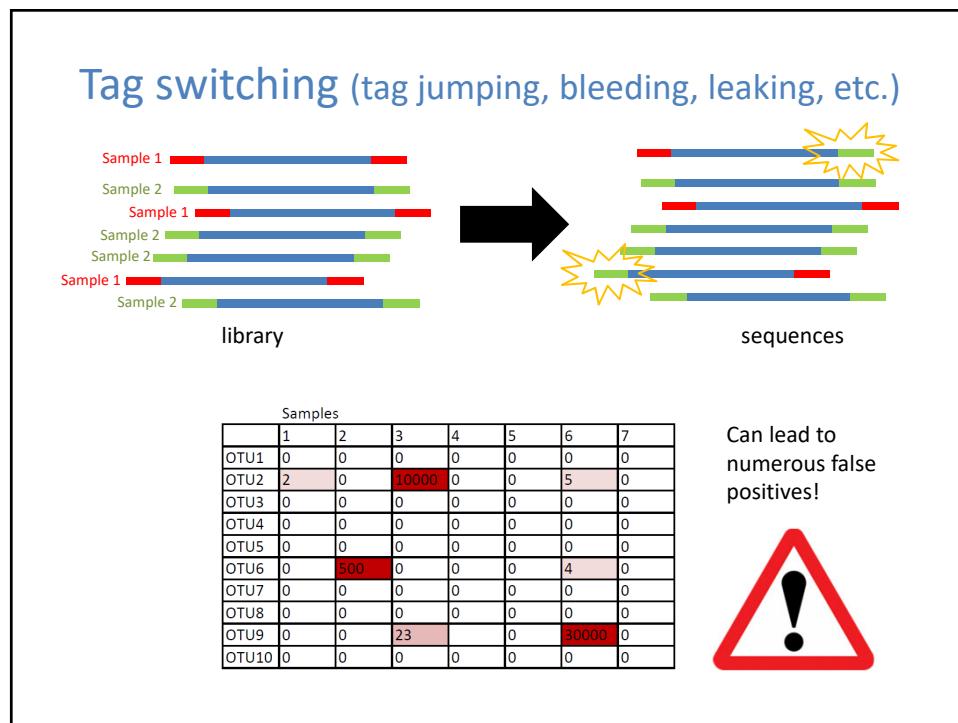
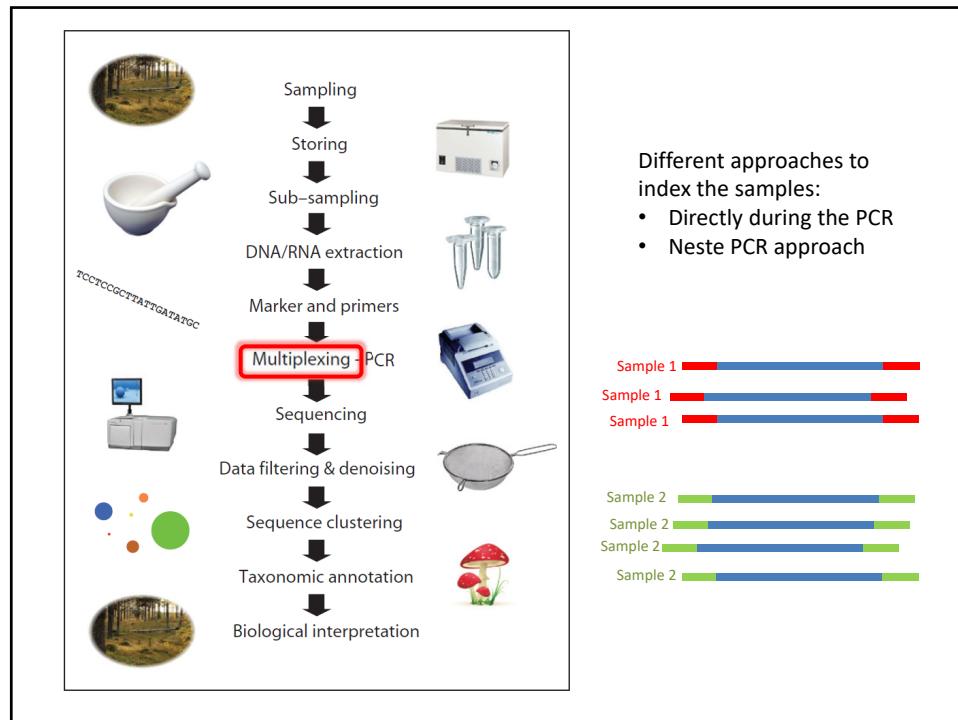
## PCR-induced errors

- **PCR mutations:** polymerase enzymes introduce erroneous nucleotides now and then
  - Dependent on the technology whether these becomes «visible»
  - In classic (direct) Sanger sequencing such errors become «diluted»
  - In methods where your final sequences are derived from one single DNA template, they become visible and must be corrected for!









# Tag switching

MOLECULAR ECOLOGY  
RESOURCES

Molecular Ecology Resources (2015)

doi: 10.1111/1755-0998.12402

## Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies

IDA BÆRHOLM SCHNELL,\*† KRISTINE BOHMANN\*‡§ and M. THOMAS P. GILBERT\*§  
 \*Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen K, Denmark, †Center for Zoo and Wild Animal Health, Copenhagen Zoo, 2000 Frederiksberg, Denmark, §School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK, §Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia 6102, Australia

“We found that an average of 2.6% and 2.1% of sequences had tag combinations, which could be explained by tag jumping...”

# Tag switching

- The problem can be reduced or controlled for by:
  - Tagging in both ends with unique tag combinations
  - Rinse the PCR amplicons thoroughly
  - Avoid PCR step during the final library preparations steps before sequencing (i.e. when adaptors are introduced)
  - Include positive controls during PCR (mock community) → can better identify the level of switching/leakage

## Sample types

1. Biological replicates
2. Technical replicates
3. Extraction negatives
4. PCR negatives
5. Positive control (mock community)

EDITORIAL

MOLECULAR ECOLOGY WILEY

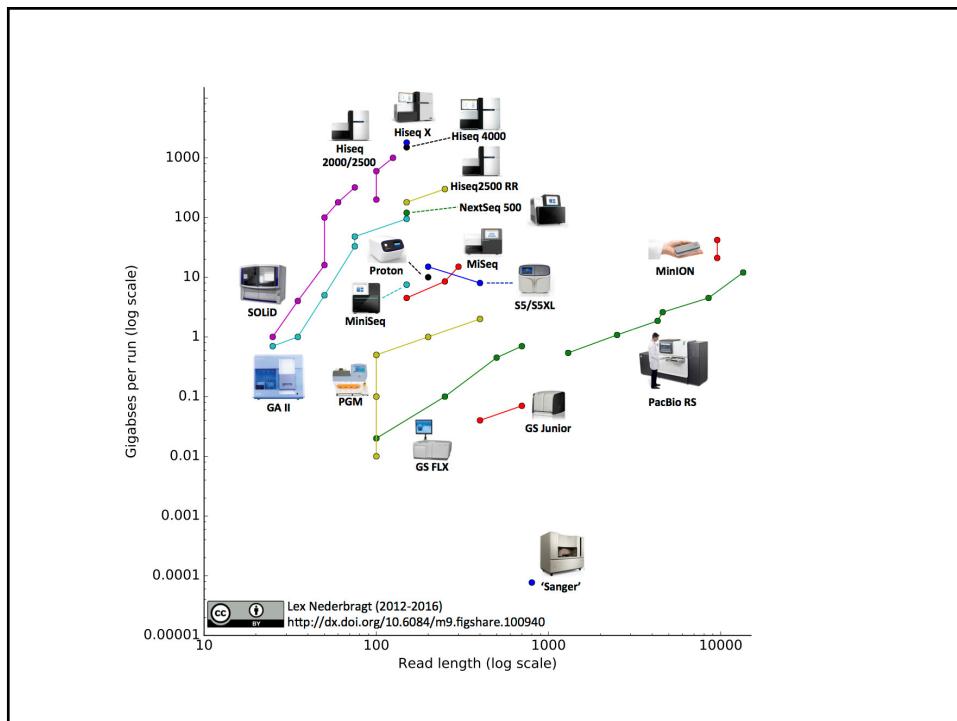
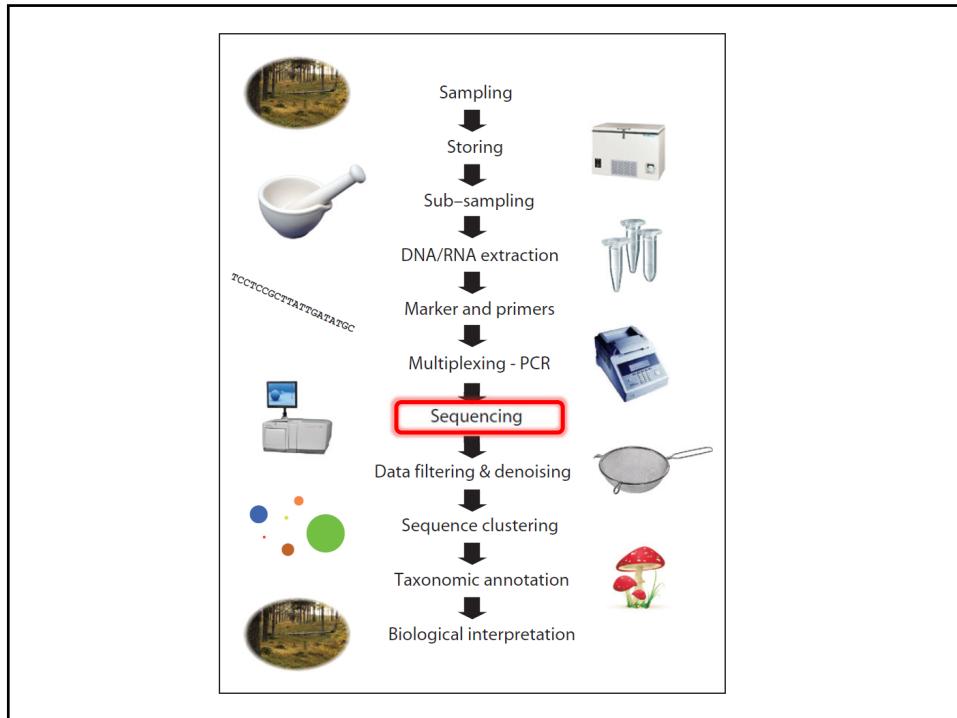
DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions

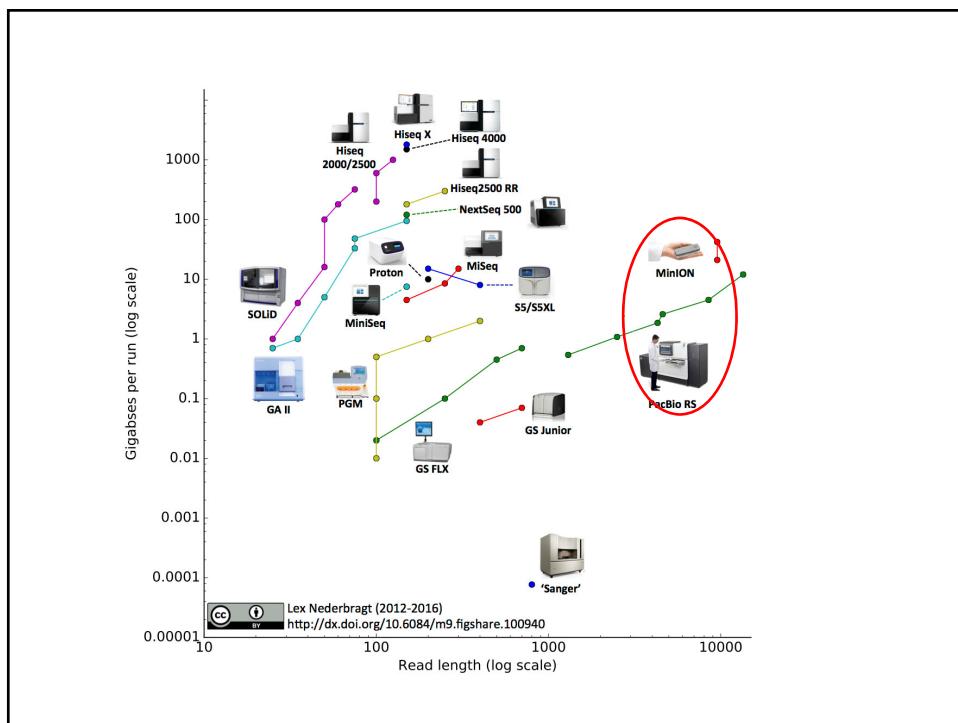
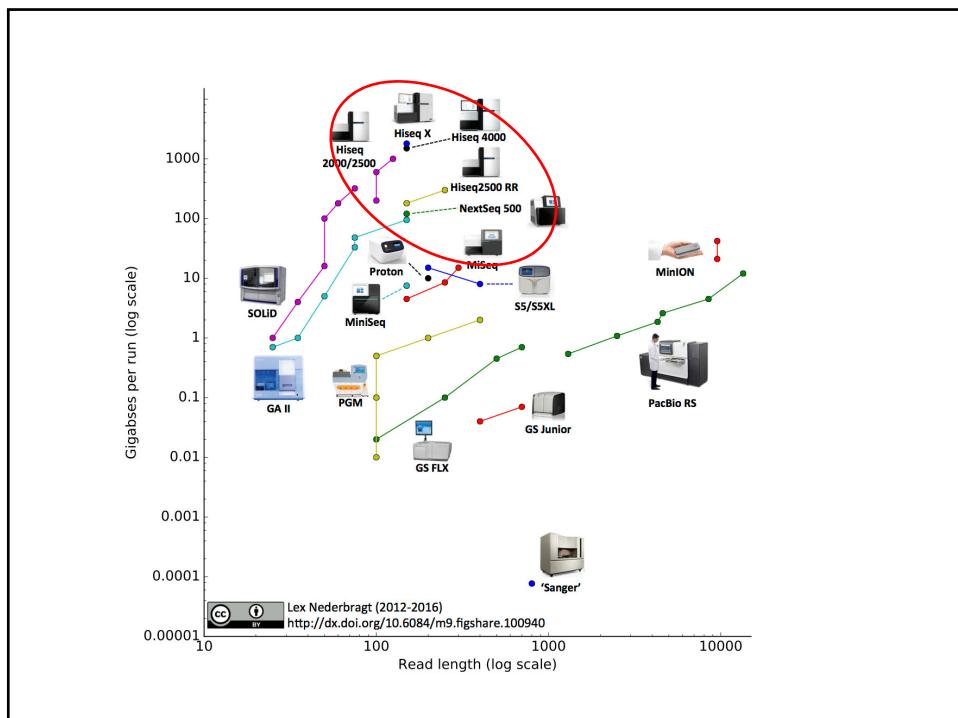
Zinger et al. 2019. Molecular Ecology Resources

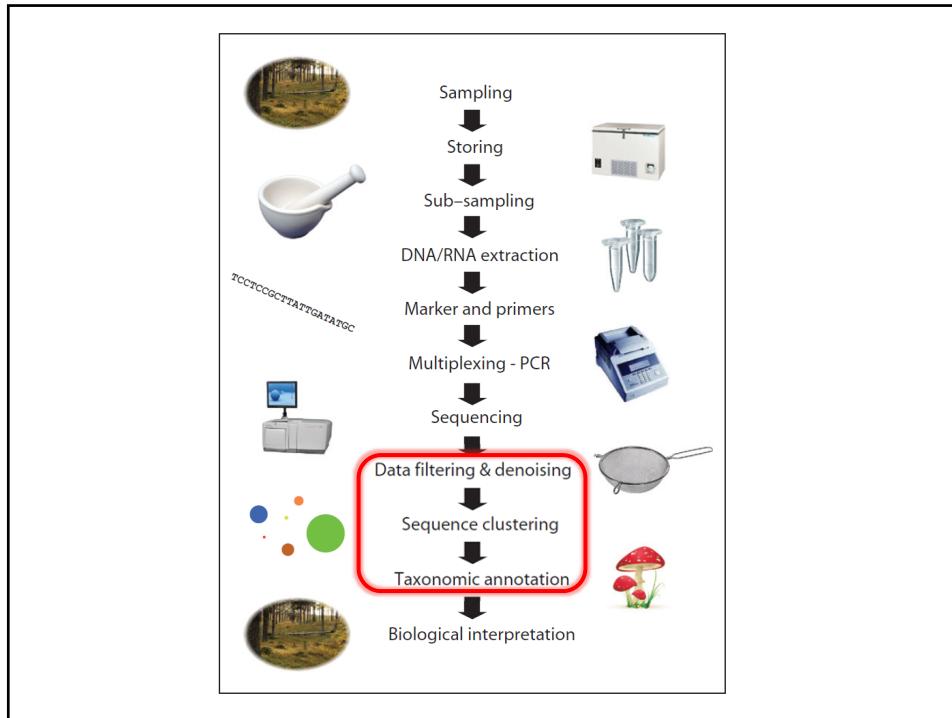
## Tag switching

- The problem can be reduced or controlled for by:
  - Tagging in both ends with unique tag combinations
  - Rinse the PCR amplicons thoroughly
  - Avoid PCR step during the final library preparations steps before sequencing (i.e. when adaptors are introduced)
  - Include positive controls during PCR (mock community) → can better identify the level of switching/leakage
  - (Drastic) *ad hoc* solution: Process your OTU/sample matrix by removing low frequency occurrences (proportion-wise)

	Samples						
	1	2	3	4	5	6	7
OTU1	0	0	0	0	0	0	0
OTU2	2	0	0	0	0	5	0
OTU3	0	0	0	0	0	0	0
OTU4	0	0	0	0	0	0	0
OTU5	0	0	0	0	0	0	0
OTU6	0	500	0	0	0	4	0
OTU7	0	0	0	0	0	0	0
OTU8	0	0	0	0	0	0	0
OTU9	0	0	23	0	0	0	0
OTU10	0	0	0	0	0	0	0



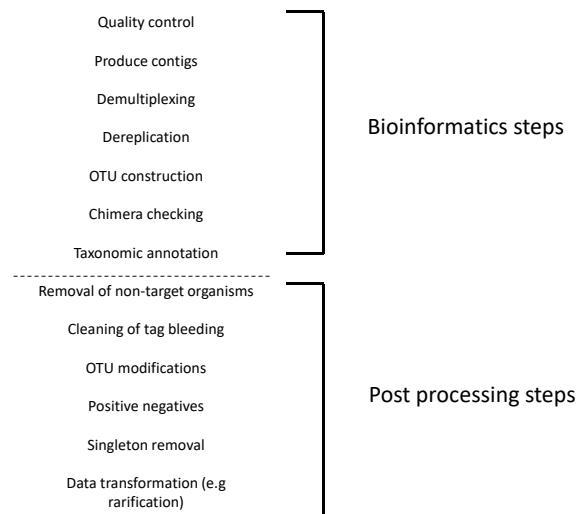




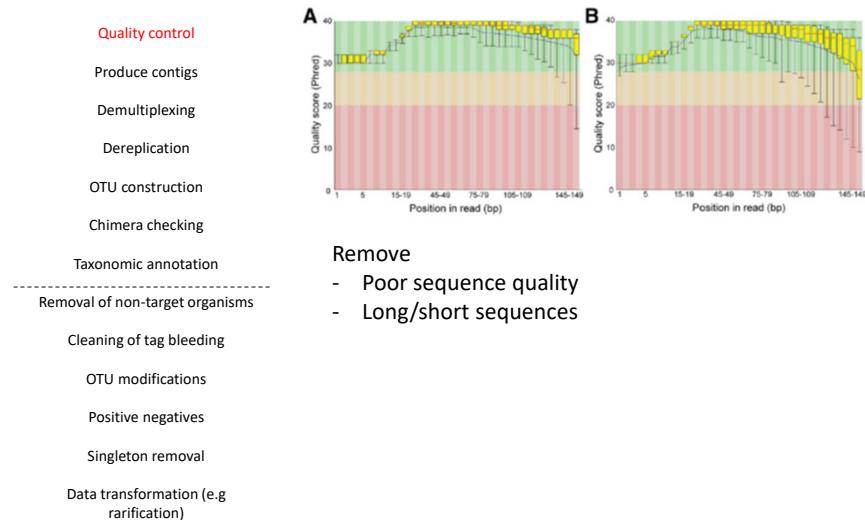
## Bioinformatics – main steps

- (The order of steps depends somewhat on the pipeline/programs)
- Quality control
  - Produce contigs
  - Demultiplexing
  - Dereplication
  - OTU construction
  - Chimera checking
  - Taxonomic annotation
- 
- Removal of non-target organisms
- Cleaning of tag bleeding
  - OTU modifications
  - Positive negatives
  - Singleton removal
  - Data transformation (e.g. rarification)

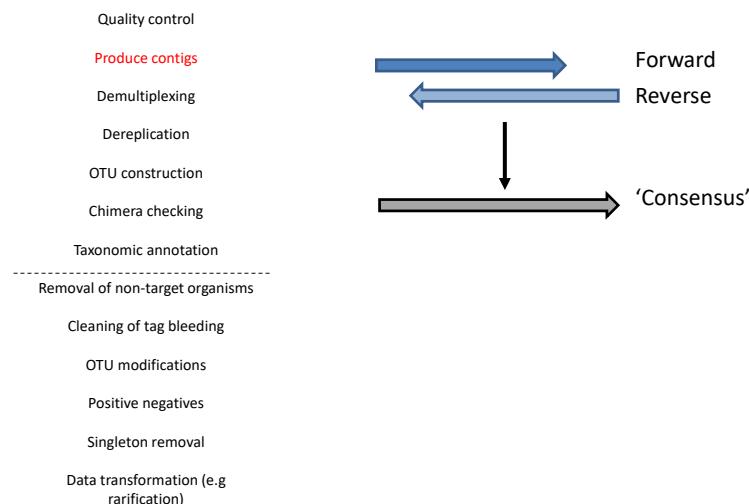
## Bioinformatics – main steps



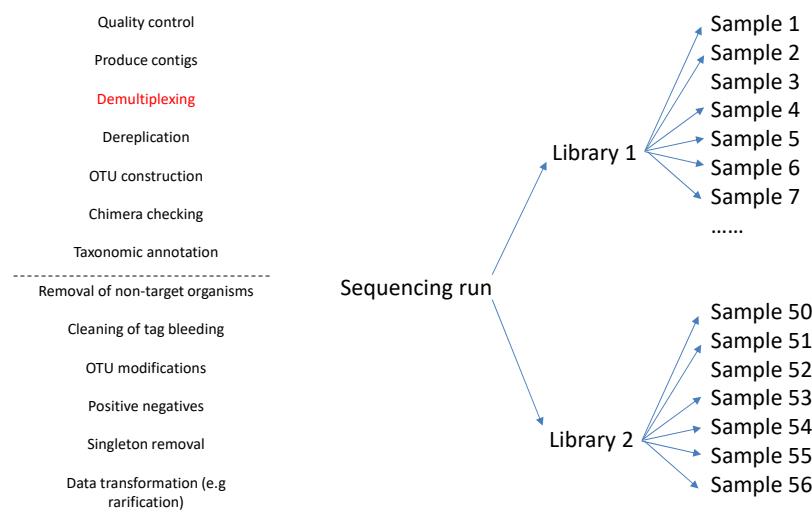
## Bioinformatics – main steps



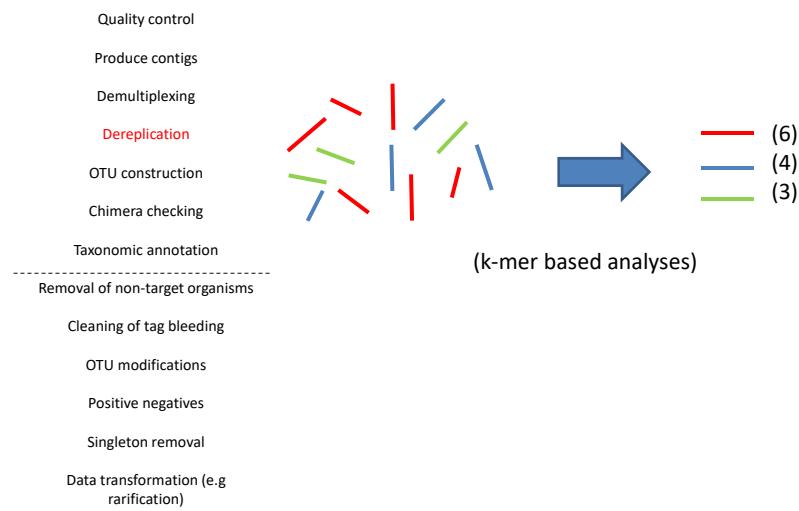
## Bioinformatics – main steps



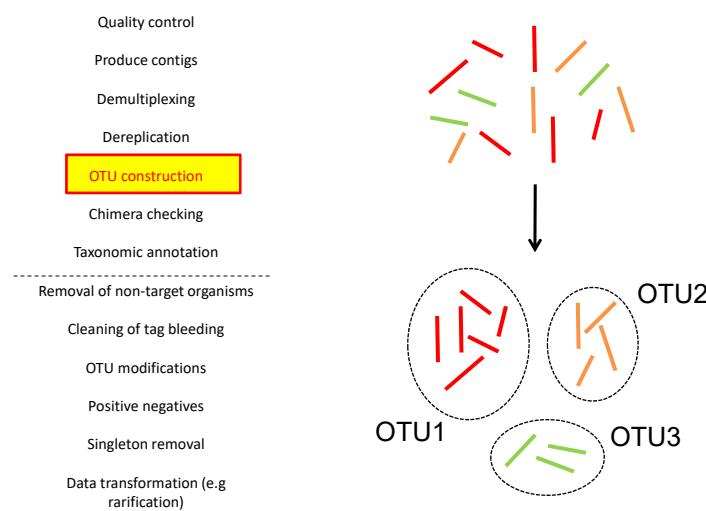
## Bioinformatics – main steps

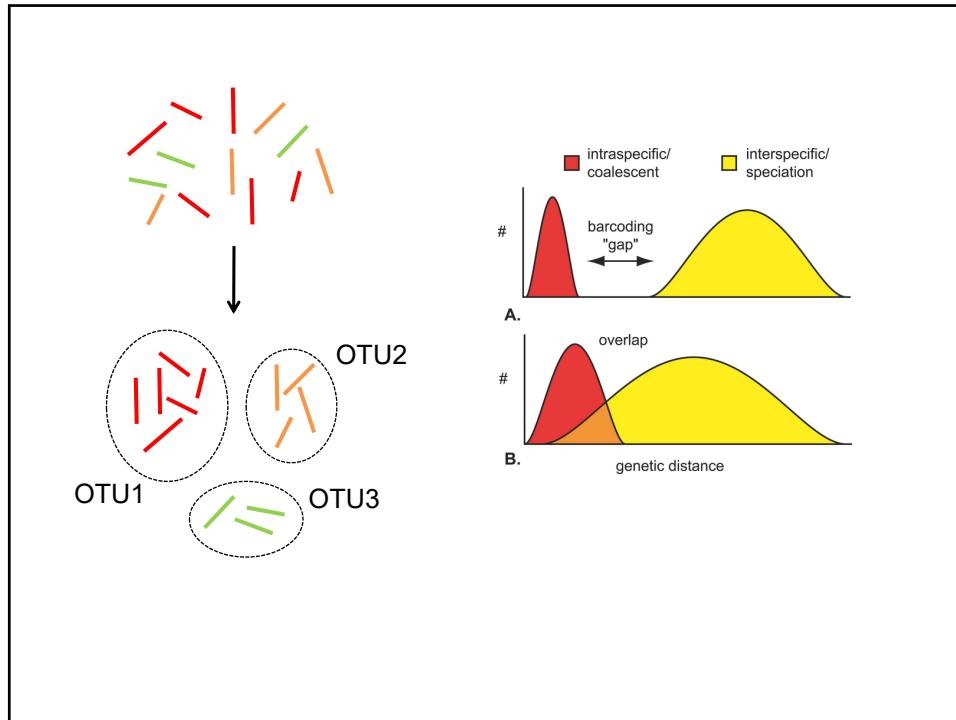


## Bioinformatics – main steps



## Bioinformatics – main steps





## Bioinformatics – main steps

- Quality control
  - Produce contigs
  - Demultiplexing
  - Dereplication
  - OTU construction**
  - Chimera checking
  - Taxonomic annotation
  - 
  - Removal of non-target organisms
  - Cleaning of tag bleeding
  - OTU modifications
  - Positive negatives
  - Singleton removal
  - Data transformation (e.g. rarification)
- de novo versus closed (reference based) OTU construction?*

## Bioinformatics – main steps

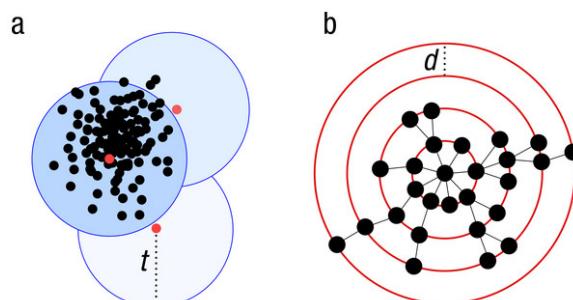
Quality control  
 Produce contigs  
 Demultiplexing  
 Dereplication  
**OTU construction**  
 Chimera checking  
 Taxonomic annotation  
 Removal of non-target organisms  
 Cleaning of tag bleeding  
 OTU modifications  
 Positive negatives  
 Singleton removal  
 Data transformation (e.g. rarification)

*de novo* versus closed (reference based) OTU construction

## Bioinformatics – main steps

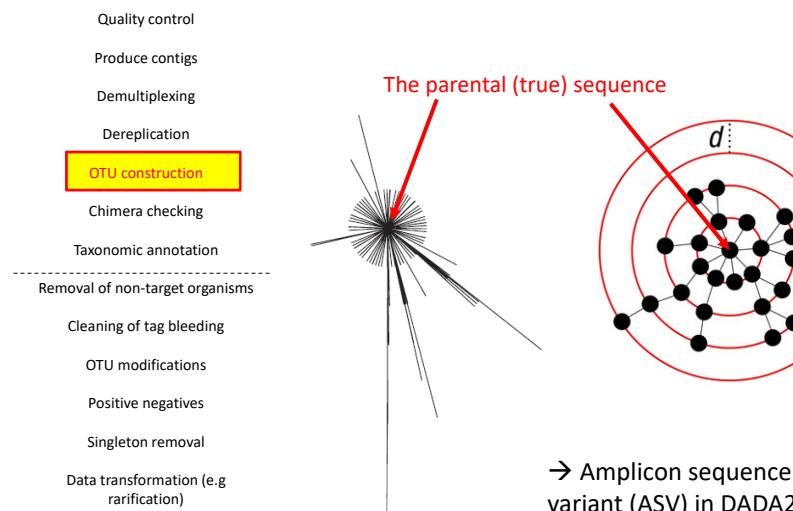
Quality control  
 Produce contigs  
 Demultiplexing  
 Dereplication  
**OTU construction**  
 Chimera checking  
 Taxonomic annotation  
 Removal of non-target organisms  
 Cleaning of tag bleeding  
 OTU modifications  
 Positive negatives  
 Singleton removal  
 Data transformation (e.g. rarification)

Many different clustering approaches

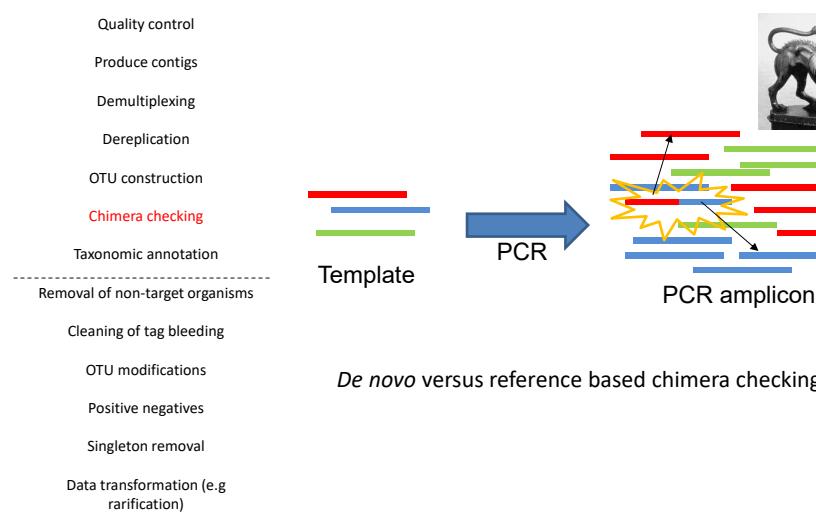


Mahé et al. 2014

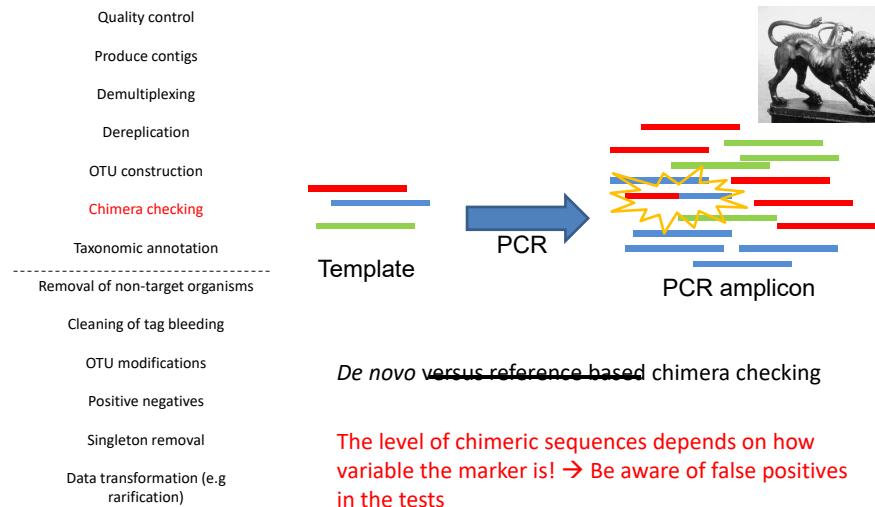
## Bioinformatics – main steps



## Bioinformatics – main steps



## Bioinformatics – main steps



## Bioinformatics – main steps



## Bioinformatics – main steps

Quality control

Produce contigs

Demultiplexing

Dereplication

OTU construction

Chimera checking

Taxonomic annotation

-----  
Removal of non-target organisms

Cleaning of tag bleeding

OTU modifications

Positive negatives

Singleton removal

Data transformation (e.g.  
rariification)

## Bioinformatics – main steps

Quality control



Produce contigs



Demultiplexing



Dereplication



OTU construction



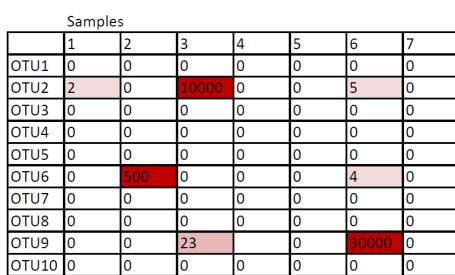
Chimera checking



Taxonomic annotation

-----  
Removal of non-target organisms

Cleaning of tag bleeding

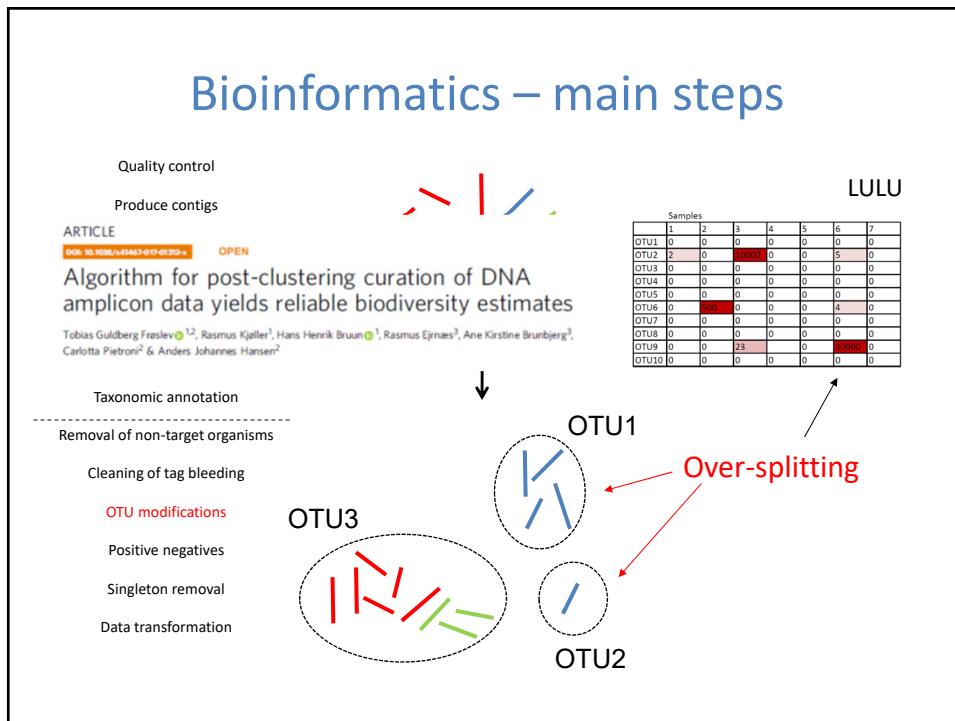
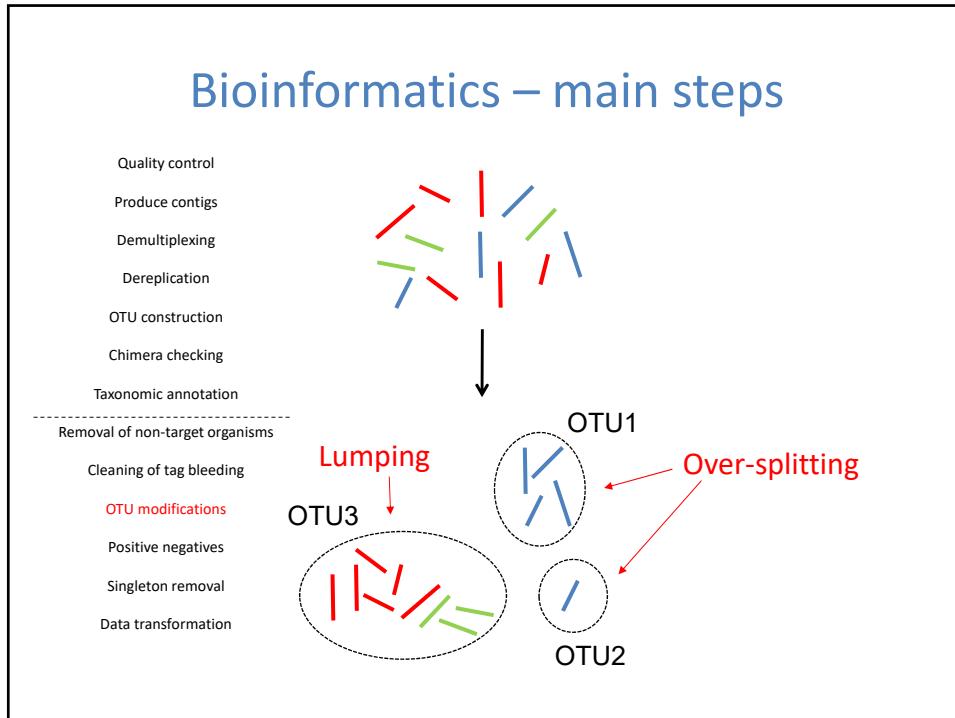


OTU modifications

Positive negatives

Singleton removal

Data transformation



## Bioinformatics – main steps

Quality control

Produce contigs

Demultiplexing

Dereplication

OTU construction

Chimera checking

Taxonomic annotation

-----  
Removal of non-target organisms

Cleaning of tag bleeding

OTU modifications

**Positive negatives**

Singleton removal

Data transformation

External or internal contaminants?



Quantifying and Understanding Well-to-Well Contamination in Microbiome Research

Jeremiah J. Minich,<sup>a</sup> Jon G. Sanders,<sup>b</sup> Amnon Amir,<sup>b</sup> Greg Humphrey,<sup>b</sup> Jack A. Gilbert,<sup>c,d</sup> Rob Knight<sup>b,c,e,f</sup>

Should be very careful how you treat positive negatives!!

## Bioinformatics – main steps

Quality control

Produce contigs

Demultiplexing

Dereplication

OTU construction

Chimera checking

Taxonomic annotation

-----  
Removal of non-target organisms

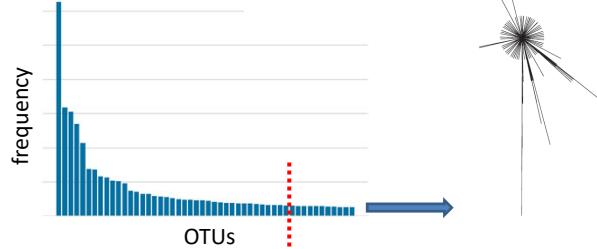
Cleaning of tag bleeding

OTU modifications

**Positive negatives**

**Singleton removal**

Data transformation

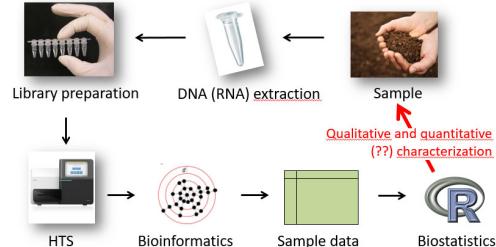


What is a ‘singleton’?  
Depends on your sequencing depth..

## Bioinformatics – main steps

Quality control  
 Produce contigs  
 Demultiplexing  
 Dereplication  
 OTU construction  
 Chimera checking  
 Taxonomic annotation  
 Removal of non-target organisms  
 Cleaning of tag bleeding  
 OTU modifications  
 Positive negatives  
 Singleton removal  
 Data transformation

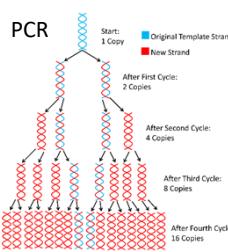
### General workflow in DNA metabarcoding studies



## Bioinformatics – main steps

Quality control  
 Produce contigs  
 Demultiplexing  
 Dereplication  
 OTU construction  
 Chimera checking  
 Taxonomic annotation  
 Removal of non-target organisms  
 Cleaning of tag bleeding  
 OTU modifications  
 Positive negatives  
 Singleton removal  
 Data transformation

	Samples						
	1	2	3	4	5	6	7
OTU1	0	0	0	0	0	0	0
OTU2	2	0	0	10000	0	0	0
OTU3	0	0	0	0	0	0	0
OTU4	0	0	0	0	0	0	0
OTU5	0	0	0	0	0	0	0
OTU6	0	500	0	0	0	4	0
OTU7	0	0	0	0	0	0	0
OTU8	0	0	0	0	0	0	0
OTU9	0	0	23	0	0	0	0
OTU10	0	0	0	0	0	0	0



Depends on the study aims!

Be careful with resampling and transformations!

Check the effect from various data treatments options on the results!

## The importance of controls

1. Biological replicates
  2. Technical replicates
  3. Extraction negatives
  4. PCR negatives
  5. Positive control (mock community)
- Different purposes

## Conclusions

- Which methods to use? → No general answer – it's context dependent. You must argue for your choices!

INVITED REVIEWS AND SYNTHESSES

WILEY MOLECULAR ECOLOGY

**Environmental DNA metabarcoding: Transforming how we survey animal and plant communities**

Kristy Deiner<sup>1</sup> | Holly M. Bik<sup>2</sup> | Elvira Mächler<sup>3,4</sup> | Mathew Seymour<sup>5</sup> | Anais Lacoursière-Roussel<sup>6</sup> | Florian Altermatt<sup>7,8</sup> | Simon Creer<sup>5</sup> | Ilana Bista<sup>5,7</sup> | David M. Lodge<sup>1</sup> | Natasha de Vere<sup>8,9</sup> | Michael E. Pfrender<sup>10</sup> | Louis Bernatchez<sup>6</sup>

EDITORIAL

MOLECULAR ECOLOGY WILEY

**DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions**

Zinger et al. 2019. Molecular Ecology Resources



Pierre Taberlet

