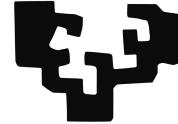




eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea



NATURAL HISTORY MUSEUM OF DENMARK  
UNIVERSITY OF COPENHAGEN

# Genomics of Marine Resources Metagenomics

2018 May 14-17

Antton Alberdi - [antton.alberdi@smn.ku.dk](mailto:antton.alberdi@smn.ku.dk)  
Assistant Professor - University of Copenhagen

## 2 Metabarcoding



Analytical procedure whereby **minibarcode primers** are used to **PCR amplify** taxa within a particular taxonomic group after which samples are sequenced **in parallel on a high-throughput sequencing platform** followed by **taxonomical identification** of sequences through comparison to DNA reference databases.

***Commonly employed synonyms:***

Metabarcoding

Amplicon sequencing

Targeted sequence metagenomics

...

## 2 Metabarcoding



## 2 Metabarcoding



Received: 18 May 2017 | Accepted: 26 June 2017

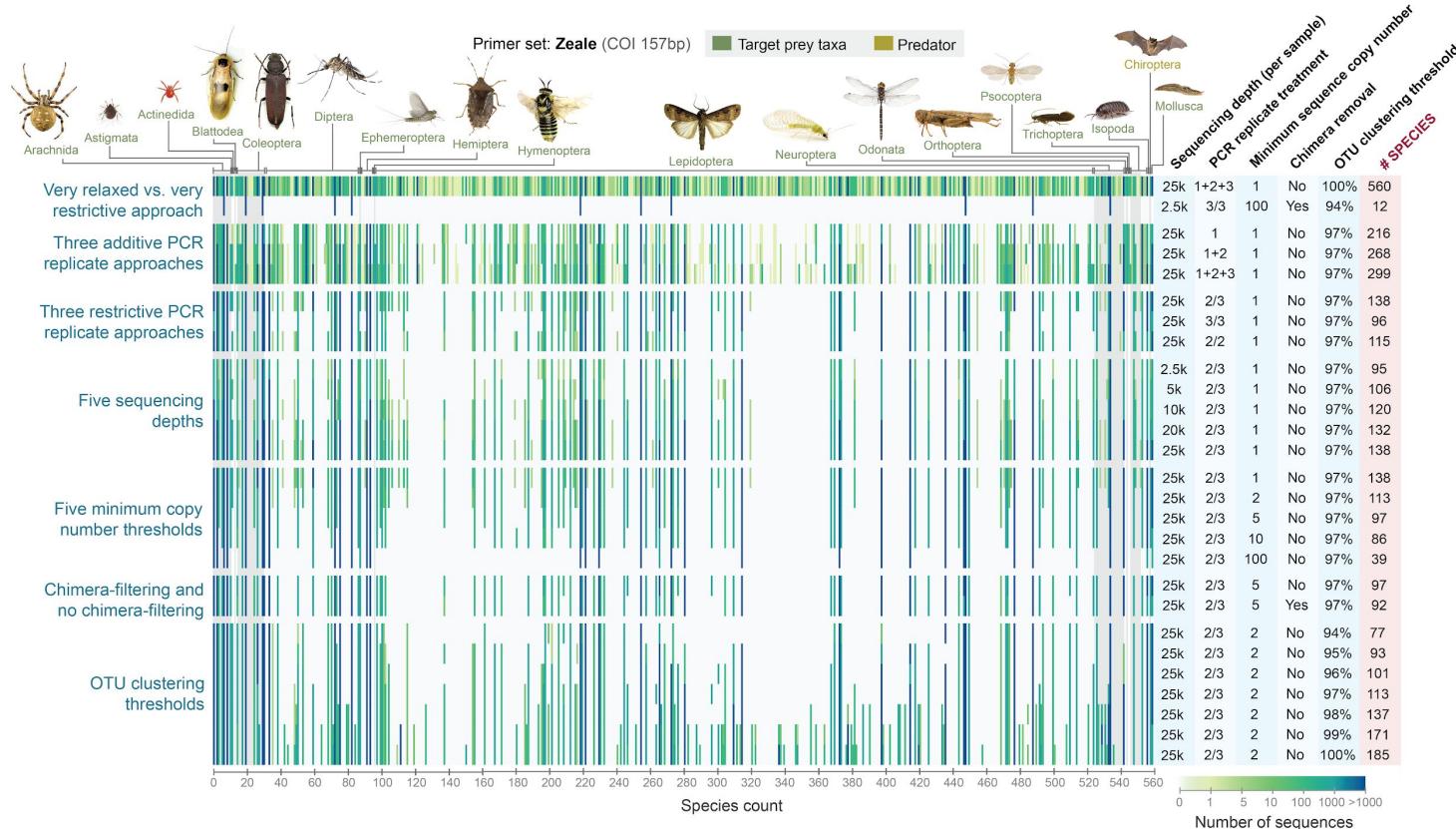
DOI: 10.1111/2041-210X.12849

**RESEARCH ARTICLE**

Methods in Ecology and Evolution 

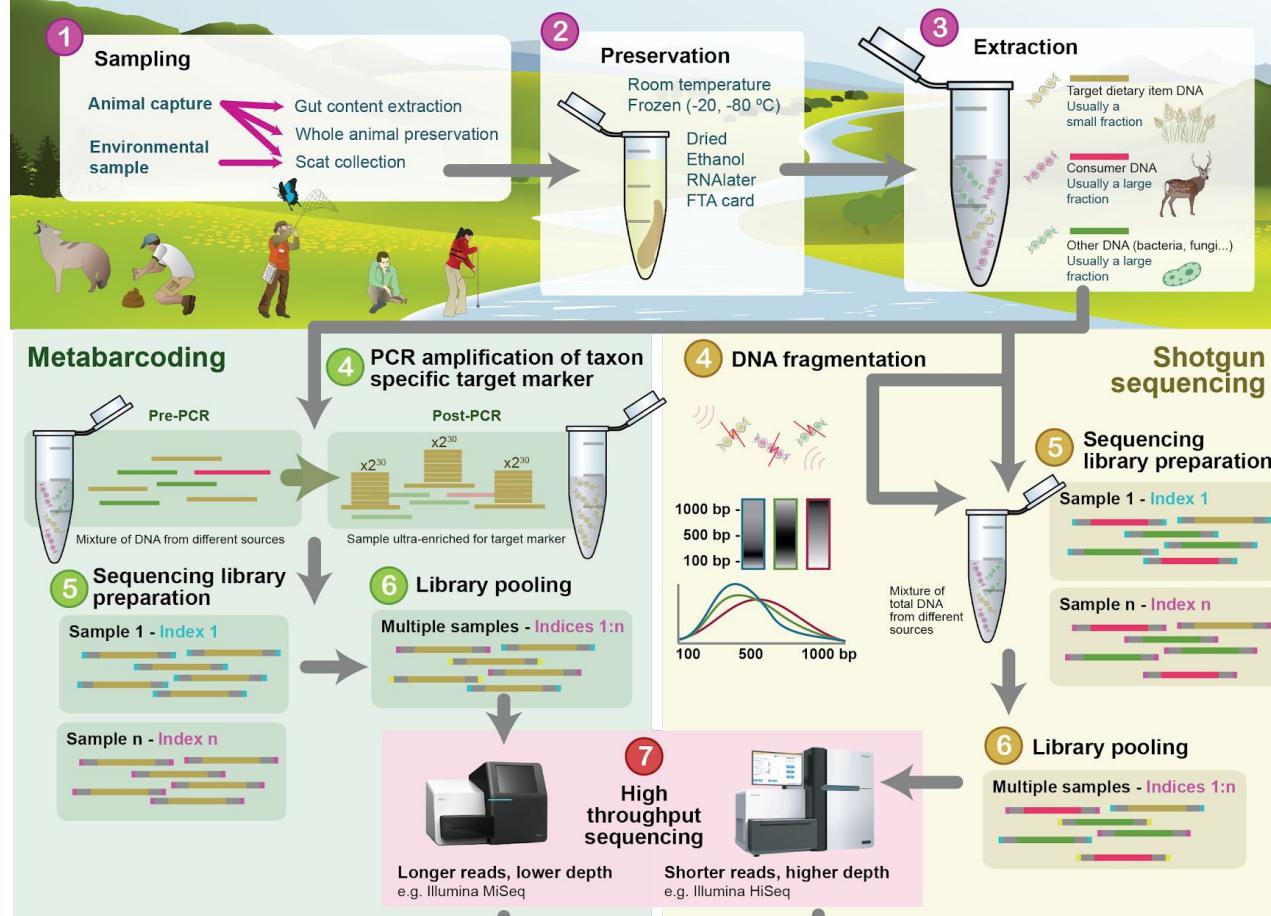
# Scrutinizing key steps for reliable metabarcoding of environmental samples

Antton Alberdi<sup>1</sup>  | Ostaizka Aizpurua<sup>1</sup> | M. Thomas P. Gilbert<sup>1,2,3</sup> | Kristine Bohmann<sup>1,4</sup>



- Collection biases
- Fragmented DNA
- Interspecific variability across the marker
- Intraspecific variability across the marker
- Targeted approach
- Inhibition
- Stochasticity
- Primer bias
- Numts
- PCR errors
- Sequencing errors
- Chimera formation in tagging PCR
- Chimera formation in index PCR
- False positives
- False negatives
- False assignment of sequences to samples
- No quantification within samples – equal weight to scarce and abundant species
- Contamination
- Predator and collector DNA
- Secondary predation
- Parasites
- Reference database
- etc?

Things to consider in the lab,  
behind the screen and in the  
interpretation of results



## 2 Metabarcoding

# Sampling

Avoid contamination



## 2 Metabarcoding

# Sampling Techniques

- **Water filtering**
- Soil sampling
- Air sampling
- Core sampling: soil, ice, guano...
- Dropping or gut collections
- ...



## 2 Metabarcoding

# Sampling Techniques

- **Water filtering**
- Soil sampling
- Air sampling
- Core sampling: soil, ice, guano...
- Dropping or gut collections
- ...



## 2 Metabarcoding

# Sampling

- Water filtering
- **Soil sampling**
- Air sampling
- Core sampling: soil, ice, guano...
- Dropping or gut collections
- ...



## 2 Metabarcoding

# Sampling

- Water filtering
- Soil sampling
- Air sampling
- **Core sampling: soil, ice, guano...**
- Dropping or gut collections
- ...



## 2 Metabarcoding

# Sampling

- Water filtering
- Soil sampling
- Air sampling
- Core sampling: soil, ice, guano...
- **Dropping or gut collections**
- ...



## 2 Metabarcoding

# Sampling

- Water filtering
- Soil sampling
- Air sampling
- Core sampling: soil, ice, guano...
- **Dropping or gut collections**
- ...



## 2 Metabarcoding

# Preservation

Snap freezing



Air dried, silica beads



Preservatives: RNAlater,  
ethanol, OMNIgene Gut (DNA  
Genotek), FTA cards  
(Whatman)



## 2 Metabarcoding

# Preservation



RESEARCH ARTICLE  
Methods for Improving Human Gut Microbiome Data by Reducing Variability through Sample Processing and Storage of Stool

Monika A. Gorzelak, Sandeep K. Gill, Nishat Tasnim, Zahra Ahmadi-Vand, Michael Jay, Deanna L. Gibson\*

Contents lists available at ScienceDirect

Journal of Microbiological Methods



journal homepage: [www.elsevier.com/locate/jmicmeth](http://www.elsevier.com/locate/jmicmeth)



Effect of preservation method on spider monkey (*Ateles geoffroyi*) fecal microbiota over 8 weeks

Vanessa L. Hale <sup>a,\*</sup>, Chia L. Tan <sup>b</sup>, Rob Knight <sup>c,d</sup>, Katherine R. Amato <sup>e</sup>



RESEARCH ARTICLE  
Applied and Environmental Science



Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies

Se Jin Song,<sup>a,b</sup> Amnon Amir,<sup>a</sup> Jessica L. Metcalf,<sup>a,b</sup> Katherine R. Amato,<sup>c</sup> Zhenjiang Zech Xu,<sup>a</sup> Greg Humphrey,<sup>a</sup> Rob Knight<sup>a,c</sup>



eman ta zabal zazu  
Universidad del País Vasco  
Euskal Herriko Unibertsitatea

NATIONAL HISTORY MUSEUM OF DENMARK  
UNIVERSITY OF COPENHAGEN



Open Access

METHODOLOGY ARTICLE

## Comparison of methods for fecal microbiome biospecimen collection

Christine Dominianni<sup>1†</sup>, Jing Wu<sup>1†</sup>, Richard B Hayes<sup>1,2</sup> and Jiyoung Ahn<sup>1,2\*</sup>

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: [www.elsevier.com/locate/meegid](http://www.elsevier.com/locate/meegid)

ELSEVIER

Optimal sample storage and extraction protocols for reliable multilocus genotyping of the human parasite *Schistosoma mansoni*

F. Van den Broeck<sup>a,b</sup>, S. Geldof<sup>a</sup>, K. Polman<sup>b</sup>, F.A.M. Volckaert<sup>a</sup>, T. Huyse<sup>a,b,\*</sup>

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: [www.elsevier.com/locate/jmicmeth](http://www.elsevier.com/locate/jmicmeth)

ELSEVIER

Effects of field conditions on fecal microbiota

Vanessa L. Hale <sup>a,\*</sup>, Chia L. Tan <sup>b</sup>, Kefeng Niu <sup>c,d</sup>, Yeqin Yang <sup>c</sup>, Duoying Cui <sup>e</sup>, Hongxia Zhao <sup>f</sup>, Rob Knight <sup>g,h</sup>, Katherine R. Amato<sup>i</sup>

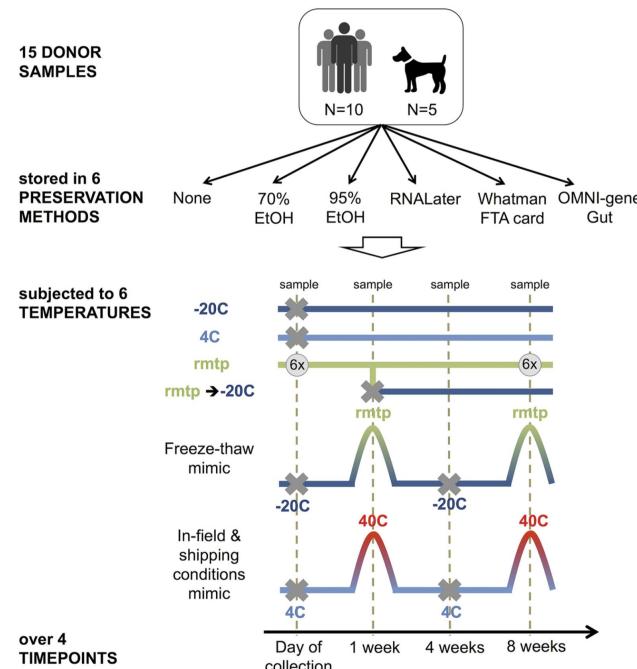


## 2 Metabarcoding

# Preservation

### Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies

Se Jin Song,<sup>a,b</sup> Amnon Amir,<sup>a</sup> Jessica L. Metcalf,<sup>a,b</sup> Katherine R. Amato,<sup>c</sup> Zhenjiang Zech Xu,<sup>a</sup> Greg Humphrey,<sup>a</sup> Rob Knight<sup>a,d</sup>



**FIG 1** Diagram of experimental design. Bulk feces were collected from humans and dogs, homogenized, treated with a range of common preservatives, and subjected to a range of likely temperature scenarios. Conditions not sampled are marked with a gray X. Additional specific combinations of conditions not sampled are described in Materials and Methods and listed in Table S2 in the supplemental material. For freshly collected samples in no preservative, six replicate subsamples were taken from the same primary aliquot. At the 8-week time point, six replicate subsamples were taken from samples with no preservative and samples stored in 95% ethanol. rmt, room temperature.

## 2 Metabarcoding

# DNA extraction



Resource Article

### Measuring the gut microbiome in birds: comparison of faecal and cloacal sampling

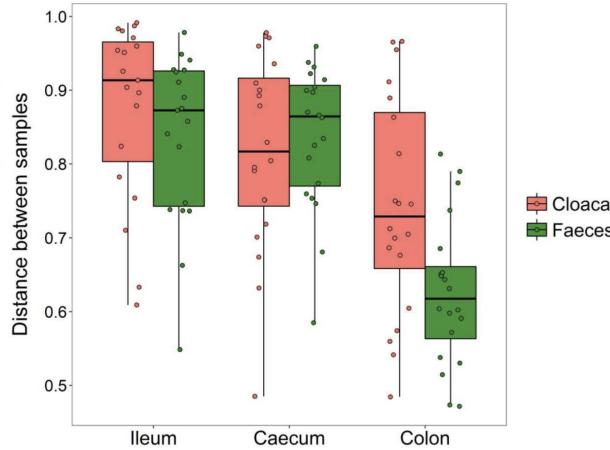
Elin Videvall , Maria Strandh, Anel Engelbrecht, Schalk Cloete, Charlie K. Cornwallis

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, July 2002, p. 3401–3407  
0099-2240/02/\$04.00+0 DOI: 10.1128/AEM.68.7.3401–3407.2002  
Copyright © 2002, American Society for Microbiology. All Rights Reserved.

Vol. 68, No. 7

### Mucosa-Associated Bacteria in the Human Gastrointestinal Tract Are Uniformly Distributed along the Colon and Differ from the Community Recovered from Feces

Erwin G. Zoetendal,<sup>1,2\*</sup> Atte von Wright,<sup>3</sup> Terttu Vilpponen-Salmela,<sup>4</sup> Kaouthar Ben-Amor,<sup>2,5</sup> Antoon D. L. Akkermans,<sup>2</sup> and Willem M. de Vos<sup>1,2</sup>



## 2 Metabarcoding

# DNA extraction

Custom protocols vs. commercial kits



- Qiagen:
  - DNeasy PowerSoil
  - DNeasy PowerWater
- FastDNA SPIN Kit
- ThermoFisher

## 2 Metabarcoding

# DNA extraction

## SOIL

SORT BY: Position ↑



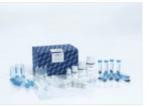
### DNeasy PowerSoil Kit

Isolate microbial genomic DNA from all soil types.



### DNeasy PowerLyzer PowerSoil Kit

Isolate DNA from tough soil microbes, optimized for use with bead-based homogenizers.



### DNeasy PowerMax Soil Kit

Isolate microbial DNA from large quantities of soil - great for samples with low microbial load.



### MagAttract PowerSoil DNA Kit

Hands-free isolation of DNA from soil using automated processing and liquid handling systems.

[VIEW DETAILS](#)

[VIEW DETAILS](#)

[VIEW DETAILS](#)

[VIEW DETAILS](#)

## WATER/AIR

SORT BY: Position ↑



### DNeasy PowerWater Kit

For the isolation of genomic DNA from filtered water samples, including turbid water.



### DNeasy PowerWater Sterivex Kit

Isolate genomic DNA from water samples collected with Sterivex Filter Units.



### MagAttract PowerWater DNA/RNA Kit

Automated isolation of nucleic acids from filtered air and water samples.



### AllPrep PowerViral DNA/RNA Kit

Isolate viral or bacterial total nucleic acids from waste water and stool samples.

## PLANTS/SEEDS



### DNeasy PowerSoil HTP 96 Kit

For high-throughput isolation of DNA from up to 384 soil samples in less than one day.



### RNeasy PowerSoil DNA Elution Kit

For co-isolation of DNA and RNA from soil using the RNeasy PowerSoil Total RNA Kit.

[VIEW DETAILS](#)

[VIEW DETAILS](#)

## FECAL/MICROBIOME

SORT BY: Position ↑



### QIAamp PowerFecal DNA Kit

Isolate DNA from stool, gut material and biosolids.



### QIAamp 96 PowerFecal QIAcube HT Kit

For automated high-throughput purification of genomic DNA from fresh or frozen stool samples that are high in PCR inhibitors.



### MagAttract PowerMicrobiome DNA/RNA Kit

Hands-free isolation of nucleic acids from stool and gut material using an automated processing or liquid handling system.

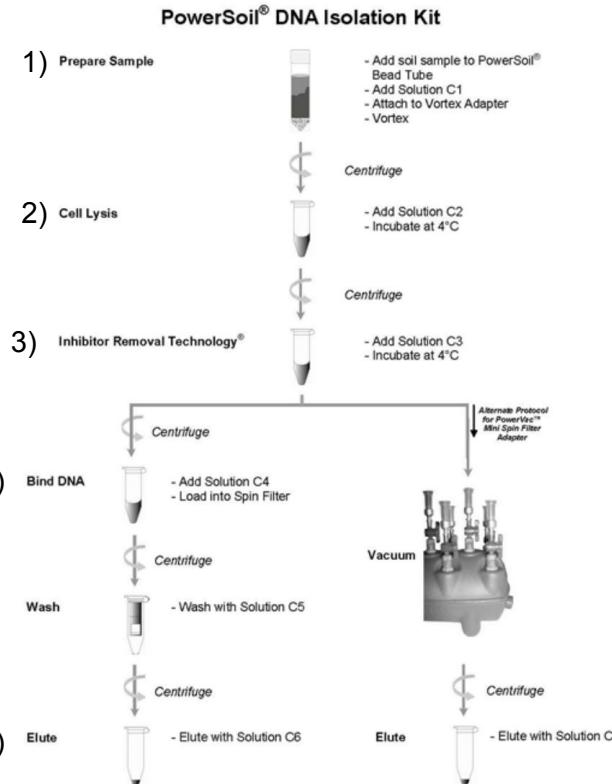


### AllPrep PowerViral DNA/RNA Kit

Isolate viral or bacterial total nucleic acids from waste water and stool samples.

## 2 Metabarcoding

# DNA extraction



## 2 Metabarcoding

OPEN  ACCESS Freely available online

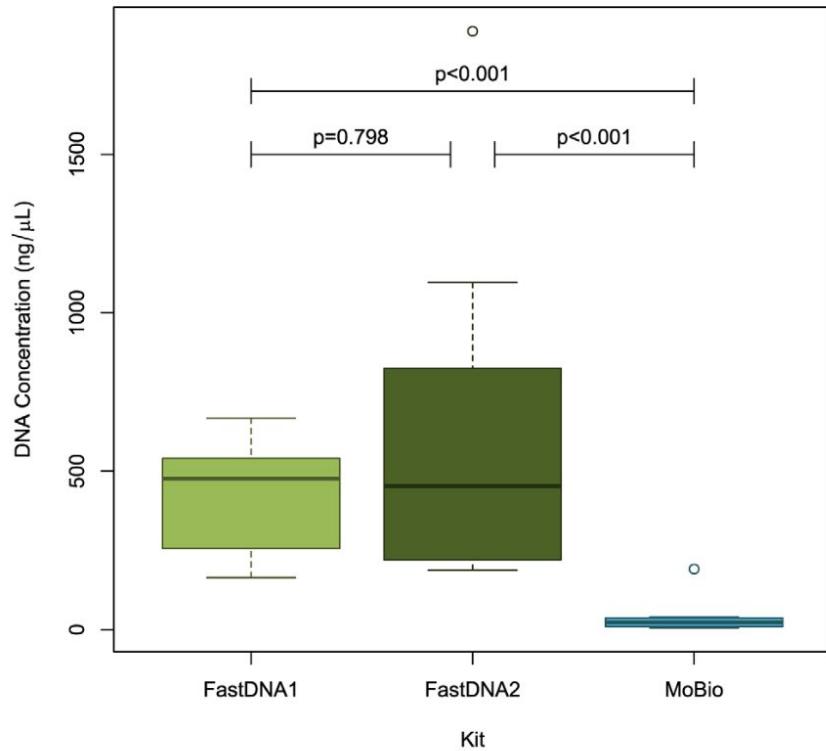


### The Impact of Different DNA Extraction Kits and Laboratories upon the Assessment of Human Gut Microbiota Composition by 16S rRNA Gene Sequencing

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>4</sup>, John M. Thomson<sup>5</sup>, UK IBD Genetics Consortium<sup>†</sup>, Jack Satsangi<sup>1</sup>, Harry J. Flint<sup>4</sup>, Julian Parkhill<sup>2</sup>, Charlie W. Lees<sup>1,9</sup>, Georgina L. Hold<sup>3,\*</sup>

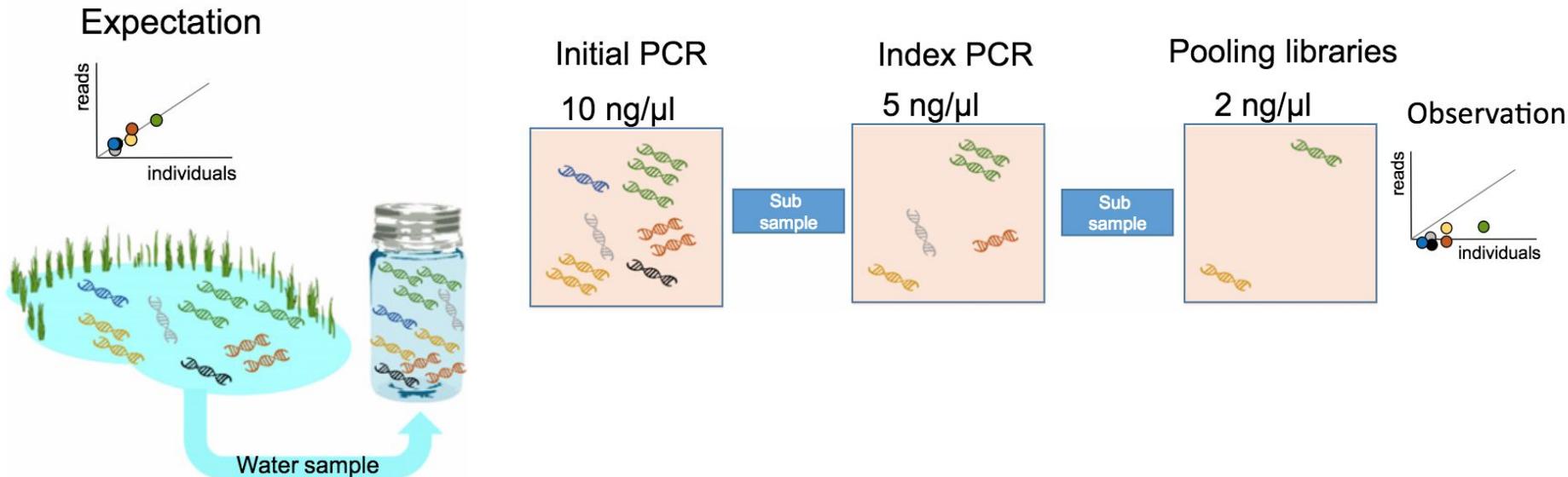


# DNA extraction



## 2 Metabarcoding

# DNA extraction



## 2 Metabarcoding

# DNA extraction



FEMS Microbiology Ecology, 93, 2017, fw240  
doi: 10.1093/femsec/fw240  
Advance Access Publication Date: 3 December 2016  
Research Article

Wesolowska-Andersen et al. *Microbiome* 2014, 2:19  
<http://www.microbiomejournal.com/content/2/1/19>

### RESEARCH ARTICLE

#### DNA extraction protocols may influence bacterial community composition detected in the intestinal microbiome: a metagenomic analysis from wild Prussian carp, *Carassius gibelio*

Elena N. Kashinskaya<sup>1</sup>, Karl B. Andree<sup>2</sup>, Evgeniy P. Simonov<sup>3</sup>, and Mikhail M. Solovyev<sup>1,3,\*</sup>

OPEN ACCESS Freely available online

#### The Impact of Different DNA Extraction Kits and Laboratories upon the Assessment of Human Gut Microbiota Composition I

Nicholas A. Kennedy<sup>1</sup>, Alan W. Walker<sup>2</sup>, Susan Petra Louis<sup>4</sup>, John M. Thomson<sup>5</sup>, UK IBD Gene Julian Parkhill<sup>2</sup>, Charlie W. Lees<sup>1,9</sup>, Georgina L.

PLOS ONE

OPEN ACCESS Freely available online

#### Evaluation of Methods for the Extraction and Purification of DNA from the Human Microbiome

Sanqing Yuan<sup>1,4</sup>, Dora B. Cohen<sup>1,4</sup>, Jacques Ravel<sup>5</sup>, Zaid Abdo<sup>2,3,4</sup>, Larry J. Forney<sup>1,4\*</sup>



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)  
ScienceDirect

Journal of Microbiological Methods 72 (2008) 124–132

Journal of Microbiological Methods

[www.elsevier.com/locate/jmicmeth](http://www.elsevier.com/locate/jmicmeth)

Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces



Jfrey L. Ram <sup>a,\*</sup>, Marc D. Basson <sup>b,c,d</sup>, Phanramphoei Namprachan <sup>a</sup>, R. Niec <sup>a</sup>, Kawsar Z. Baisha <sup>e</sup>, Larry H. Matherly <sup>d,f</sup>, Adhip P.N. Majumdar <sup>h</sup>, Ikuko Kato <sup>d,g</sup>

Open Access

RESEARCH

#### Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis

Agata Wesolowska-Andersen<sup>1</sup>, Martin Jain Bah<sup>2</sup>, Vera Carvalho<sup>2</sup>, Karsten Kristiansen<sup>3</sup>, Thomas Sicheritz-Pontén<sup>1</sup>, Ramneek Gupta<sup>1</sup> and Tine Rask Licht<sup>2\*</sup>

ELSEVIER

Contents lists available at SciVerse ScienceDirect

Journal of Microbiological Methods

journal homepage: [www.elsevier.com/locate/jmicmeth](http://www.elsevier.com/locate/jmicmeth)



A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples

Shantelle Claassen <sup>a,\*</sup>, Elloise du Toit <sup>a</sup>, Mamadou Kaba <sup>a,d</sup>, Clinton Moodley <sup>b</sup>, Heather J. Zar <sup>c</sup>, Mark P. Nicol <sup>a,d</sup>

PLOS ONE



## 2 Metabarcoding



# Contamination

JOURNAL OF CLINICAL MICROBIOLOGY, July 2003, p. 3452–3453  
0095-1137/03/\$8.00+0 DOI: 10.1128/JCM.41.7.3452-3453.2003  
Copyright © 2003, American Society for Microbiology. All Rights Reserved.

Vol. 41, No. 7

### Contamination of Qiagen DNA Extraction Kits with *Legionella* DNA

van der Zee and Crielaard recently reported the contamination of Qiagen DNA extraction columns with *Legionella* species (6). We have independently made this discovery and have attempted to further characterize the contaminating DNA.

extraction control, sequenced in the 23S-5S rRNA region (4), and submitted for identification in the same manner. Of the 312 bp of sequence submitted, only 246 bp shared any similarity with any reference species that had been submitted to the National Center for Biotechnology Information BLAST data-

Salter et al. *BMC Biology* 2014, **12**:87  
<http://www.biomedcentral.com/1741-7007/12/87>

RESEARCH ARTICLE

Open Access

### Reagent and laboratory contamination can critically impact sequence-based microbiome analyses

Susannah J Salter<sup>1\*</sup>, Michael J Cox<sup>2</sup>, Elena M Turek<sup>2</sup>, Szymon T Calus<sup>3</sup>, William O Cookson<sup>2</sup>, Miriam F Moffatt<sup>2</sup>, Paul Turner<sup>4,5</sup>, Julian Parkhill<sup>1</sup>, Nicholas J Loman<sup>3</sup> and Alan W Walker<sup>1,6\*</sup>



## 2 Metabarcoding

# Contamination



## 2

# Metabarcoding

GENOMIC  
RESOURCES



eman ta zabal zazu

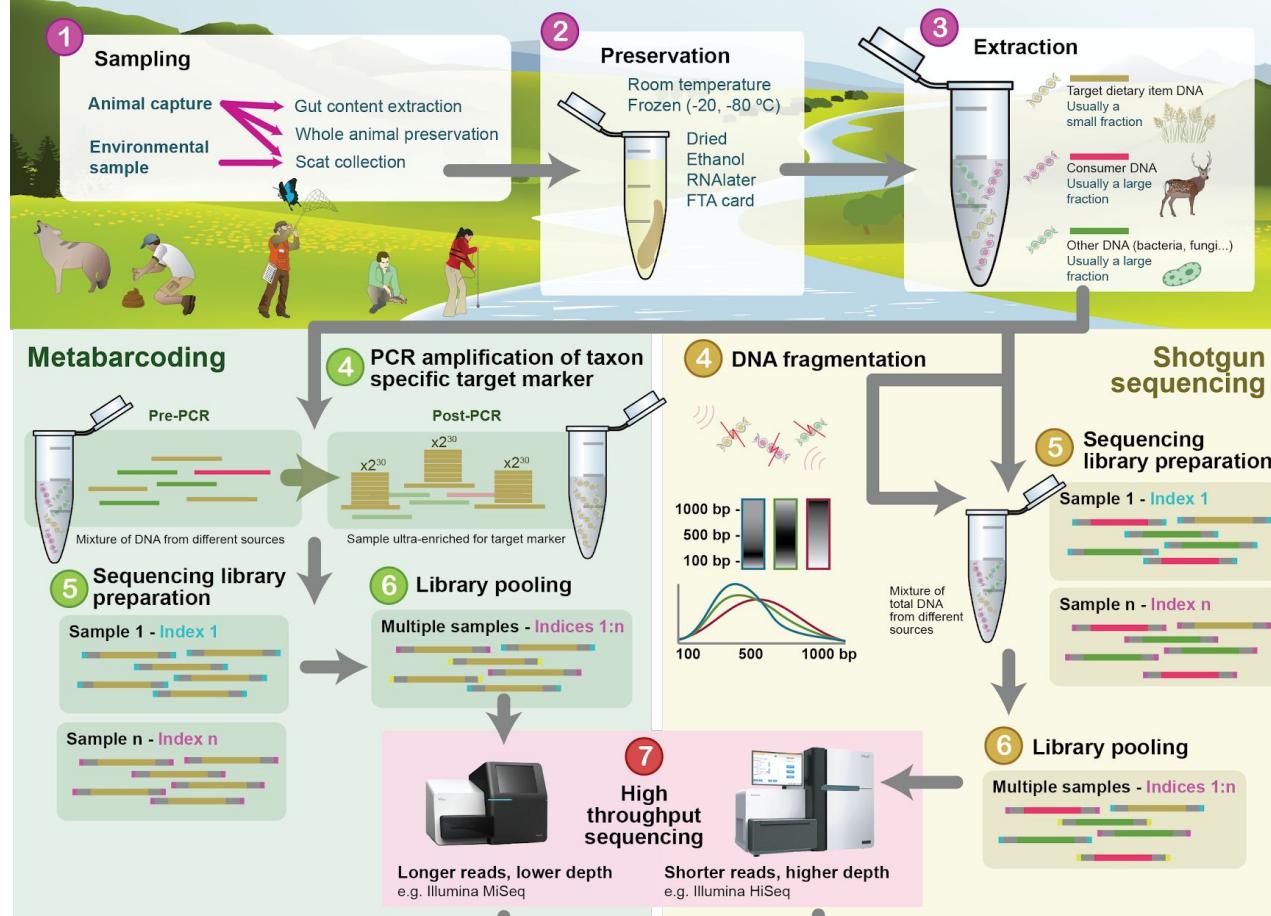


Universidad  
del País Vasco  
Euskal Herriko  
Unibertsitatea



NATURAL HISTORY MUSEUM OF DENMARK  
UNIVERSITY OF COPENHAGEN





## 2 Metabarcoding



Analytical procedure whereby **minibarcode primers** are used to **PCR amplify** taxa within a particular taxonomic group after which samples are sequenced **in parallel on a high-throughput sequencing platform** followed by **taxonomic identification** of sequences through comparison to DNA reference databases.

***Commonly employed synonyms:***

Metabarcoding

Amplicon sequencing

Targeted sequence metagenomics

...

## 2 Metabarcoding



# Reference databases

BOLD SYSTEMS

DATABASES

IDENTIFICATION

TAXONOMY

WORKBENCH

RESOURCES

LOGIN



BARCODE OF LIFE DATA SYSTEM <sup>v4</sup>

Advancing biodiversity science through DNA-based species identification.

EXPLORE THE DATA

<http://www.boldsystems.org>

## 2 Metabarcoding

# Reference databases



Home SILVAngs Browser Search ACT Download Documentation Projects FISH & Probes Contact

### SILVA

#### Welcome to the SILVA rRNA database project

A comprehensive on-line resource for quality checked and aligned ribosomal RNA sequence data.

SILVA provides comprehensive, quality checked and regularly updated datasets of aligned small (16S/18S, SSU) and large subunit (23S/28S, LSU) ribosomal RNA (rRNA) sequences for all three domains of life (*Bacteria*, *Archaea* and *Eukarya*).

SILVA are the official databases of the software package ARB.

For more background information → [Click here](#)

### SILVAngs



Check out our service for Next Generation Amplicon data

### News

11.05.2018

#### SILVA ACT Tool released



The new SILVA ACT service combines alignment, search and classify as well as reconstruction of trees in a single web application.

23.04.2018

#### Change of SILVA Terms of Use/License Model



With the next full database release which is expected for Fall 2018, the SILVA project will resign its current dual licensing model and the SILVA data sets will become free also for commercial/non-academic users.

10.04.2018

#### QIIME files for SILVA release 132 have arrived



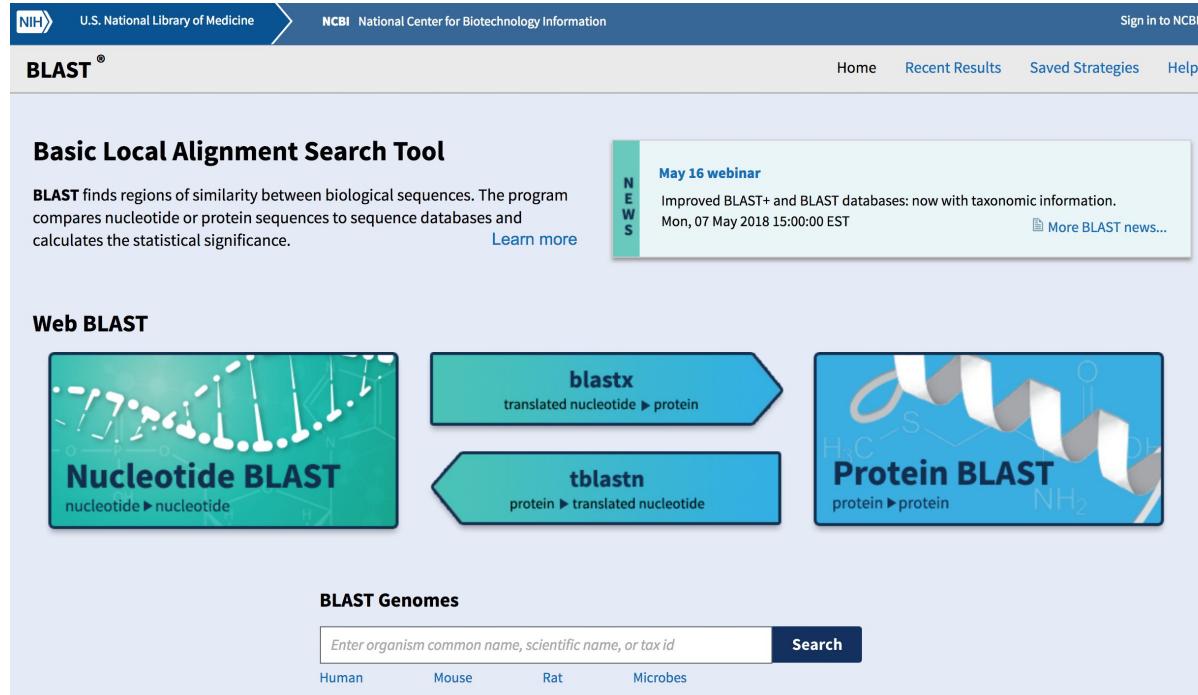
Many thanks to the QIIME Team for providing us with the QIIME 132 release files.

15.03.2018

#### de.NBI Summer School: Riding the Data Life Cycle - registration now open!

## 2 Metabarcoding

# Reference databases



The screenshot shows the NCBI BLAST homepage. At the top, there's a navigation bar with the NIH logo, U.S. National Library of Medicine, NCBI National Center for Biotechnology Information, and a Sign in to NCBI link. Below the navigation, the word "BLAST" is prominently displayed. A teal sidebar on the left contains the letters "N E W S" vertically. The main content area features a section titled "Basic Local Alignment Search Tool" which describes what BLAST does. To the right of this, there's a box for a "May 16 webinar" about improved BLAST databases. Below these sections, there's a "Web BLAST" section with three boxes: "Nucleotide BLAST" (nucleotide to nucleotide), "blastx" (translated nucleotide to protein), and "tblastn" (protein to translated nucleotide). To the right of these is a "Protein BLAST" section (protein to protein). At the bottom, there's a "BLAST Genomes" search bar with a "Search" button and links for Human, Mouse, Rat, and Microbes.

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

**BLAST®**

**Basic Local Alignment Search Tool**

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

Learn more

**NEWS**

**May 16 webinar**

Improved BLAST+ and BLAST databases: now with taxonomic information.  
Mon, 07 May 2018 15:00:00 EST

[More BLAST news...](#)

**Web BLAST**

**Nucleotide BLAST**  
nucleotide ► nucleotide

**blastx**  
translated nucleotide ► protein

**tblastn**  
protein ► translated nucleotide

**Protein BLAST**  
protein ► protein

**BLAST Genomes**

Enter organism common name, scientific name, or tax id

Search

Human    Mouse    Rat    Microbes

# Reference databases

- Different standardised markers (genomic) regions depending on the target taxa.
  - Genbank: “Mammalia COI” 38,509 entries
  - Genbank: “Mammalia NADH2” 1,453 entries

## 2 Metabarcoding

GENOMIC  
RESOURCES 



## Plants

**rBcL**

RUBISCO gene  
Plastid

**matK**

Maturase K  
Plastid

## 2 Metabarcoding



# Reference databases

## A DNA barcode for land plants

CBOL Plant Working Group<sup>1</sup>

Communicated by Daniel H. Janzen, University of Pennsylvania, Philadelphia, PA, May 27, 2009 (received for review March 18, 2009)

DNA barcoding involves sequencing a standard region of DNA as a tool for species identification. However, there has been no agreement on which region(s) should be used for barcoding land plants. To provide a community recommendation on a standard plant barcode, we have compared the performance of 7 leading candidate plastid DNA regions (*atpF-atpH* spacer, *matK* gene, *rbcL* gene, *rpoB* gene, *rpoC1* gene, *psbK-psbI* spacer, and *trnH-psbA* spacer). Based on assessments of recoverability, sequence quality, and levels of species discrimination, we recommend the 2-locus combination of *rbcL+matK* as the plant barcode. This core 2-locus barcode will provide a universal framework for the routine use of DNA sequence data to identify specimens and contribute toward the discovery of overlooked species of land plants.

*matK* | *rbcL* | species identification

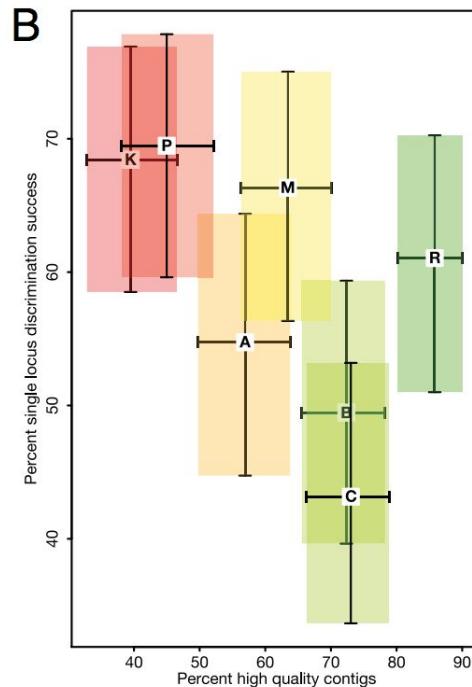
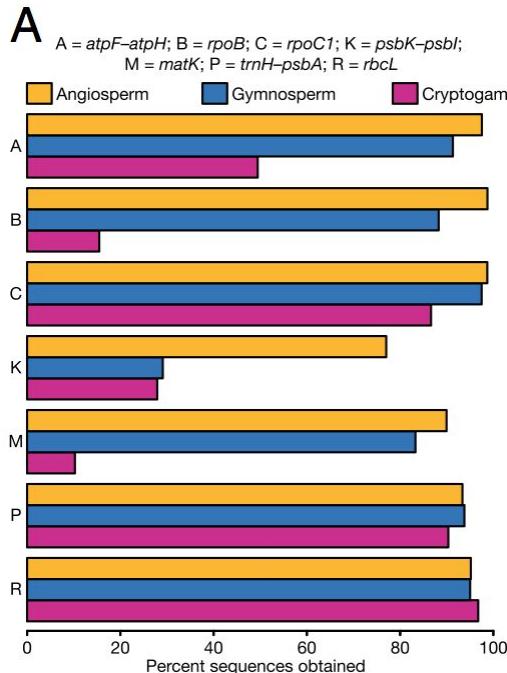
intergenic spacers *trnH-psbA* and *psbK-psbI*, in part attributable to a high frequency of mononucleotide repeats disrupting individual sequencing reads.

**Species Discrimination.** Among 397 samples successfully sequenced for all 7 loci, species discrimination for single-locus barcodes ranged from 43% (*rpoC1*) to 68%–69% (*psbK-psbI* and *trnH-psbA*), with *rbcL* and *matK* providing 61% and 66% discrimination respectively (rank order: *rpoC1*<*rpoB*<*atpF-*

Author contributions: P.M.H., L.L.F., J.L.S., M.H., S.R., M.v.d.B., M.W.C., R.S.C., D.L.E., A.J.F., S.W.G., K.E.J., K.-J.K., W.J.K., H.S., S.C.H.B., C.v.d.B., M.C., T.A.J.H., B.C.H., G.P., J.E.R., G.A.S., V.S., O.S., M.J.W., and D.P.L. designed research; D.L.E., A.J.F., K.E.J., J.v.A.S., D.B., K.S.B., K.M.C., J.C., A.C., J.J.C., F.C., D.S.D., C.S.F., M.L.H., L.J.K., P.R.K., J.S.K., Y.D.K., R.L., H.-L.L., D.G.L., S.M., O.M., I.M., S.G.N., C.-W.P., D.M.P., and D.-K.Y. performed research; L.L.F., J.L.S., M.H., S.R., and D.P.L. analyzed data; and P.M.H., S.W.G., S.C.H.B., and D.P.L. wrote the

## 2 Metabarcoding

# Reference databases



CBOL Plant Working  
Group (2009) PNAS

## 2 Metabarcoding

GENOMIC  
RESOURCES



Universidad  
del País Vasco  
Euskal Herriko  
Unibertsitatea

eman ta zabal zazu  
NATURAL HISTORY MUSEUM OF DENMARK  
UNIVERSITY OF COPENHAGEN

# Reference databases



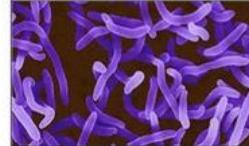
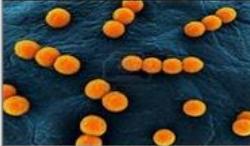
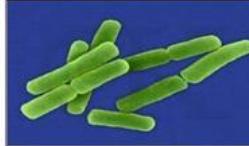
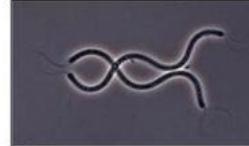
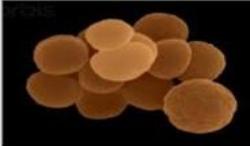
Fungi

ITS

Internal transcribed spacer

## 2 Metabarcoding

# Reference databases

Circular (Coccus)	Rod-shaped (Bacillus)	Curved Forms
		
Diplo- (in pairs)	Coccobacilli (oval)	Vibrio (curved rod)
		
Strepto- (in chains)	Streptobacilli	Spirilla (coil)
		
Staphylo- (in clusters)	Mycobacteria	Spirochete (spiral)

## Bacteria

### 16S rRNA

Small subunit of a prokaryotic ribosome

## 2 Metabarcoding

# Reference databases



Animals

COI

Cytochrome Oxidase Subunit I  
Mitochondrion

## 2 Metabarcoding

# Reference databases

**BOLDSYSTEMS**

DATABASES IDENTIFICATION TAXONOMY WORKBENCH RESOURCES LOGIN 

ANIMAL IDENTIFICATION [COI]

FUNGAL IDENTIFICATION [ITS]

PLANT IDENTIFICATION [RBCL & MATK]

The BOLD Identification System (IDS) for COI accepts sequences from the 5' region of the mitochondrial Cytochrome c oxidase subunit I gene and returns a species-level identification when one is possible. Further validation with independent genetic markers will be desirable in some forensic applications.

**Historical Databases:** [Current](#) Jul-2015 Jul-2014 Jul-2013 Jul-2012 Jul-2011 Jul-2010 Jul-2009

Search Databases:

**All Barcode Records on BOLD (5,359,409 Sequences)**

Every COI barcode record on BOLD with a minimum sequence length of 500bp (warning: unvalidated library and includes records without species level identification). This includes many species represented by only one or two specimens as well as all species with interim taxonomy. This search only returns a list of the nearest matches and does not provide a probability of placement to a taxon.

**Species Level Barcode Records (3,040,613 Sequences/187,242 Species/75,345 Interim Species)**

Every COI barcode record with a species level identification and a minimum sequence length of 500bp. This includes many species represented by only one or two specimens as well as all species with interim taxonomy.

**Public Record Barcode Database (1,179,472 Sequences/96,801 Species/26,003 Interim Species)**

All published COI records from BOLD and GenBank with a minimum sequence length of 500bp. This library is a collection of records from the published projects section of BOLD.

**Full Length Record Barcode Database (1,901,740 Sequences/168,553 Species/62,433 Interim Species)**

Inded for short sequence identification as it provides

<http://www.boldsystems.org>

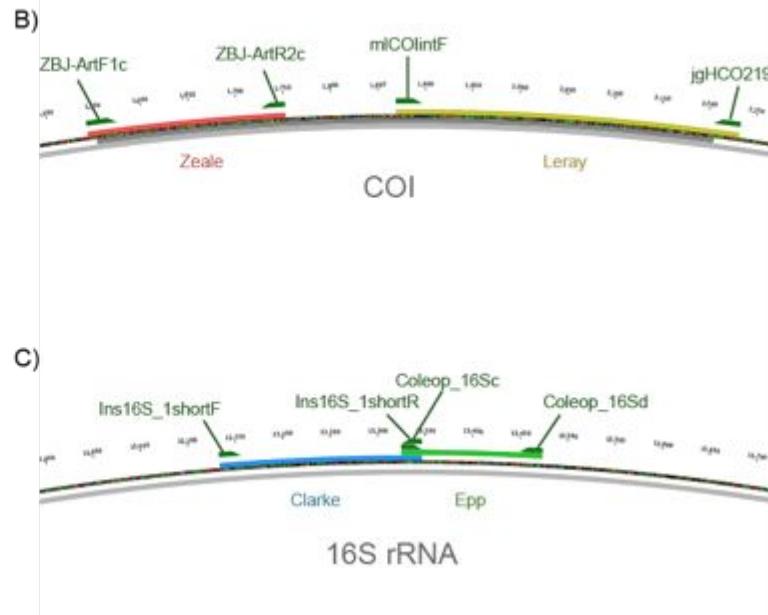
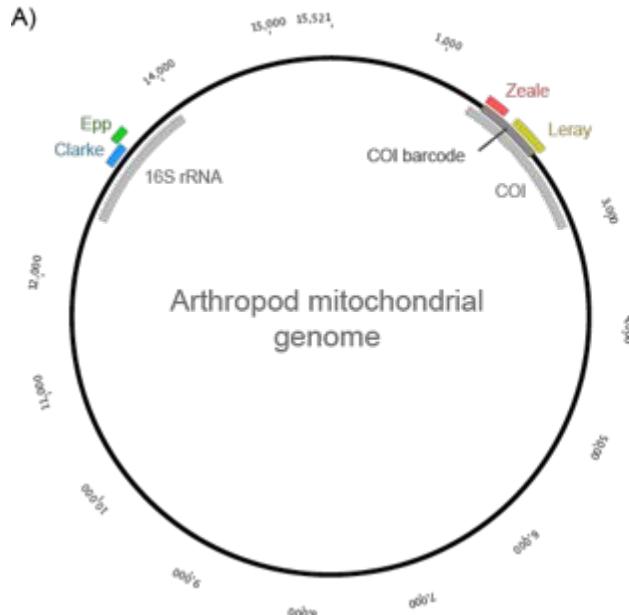
## 2 Metabarcoding

GENOMIC  
RESOURCES

Universidad  
del País Vasco  
Euskal Herriko  
Unibertsitatea

eman ta zabal zazu  
NATURAL HISTORY MUSEUM OF DENMARK  
UNIVERSITY OF COPENHAGEN

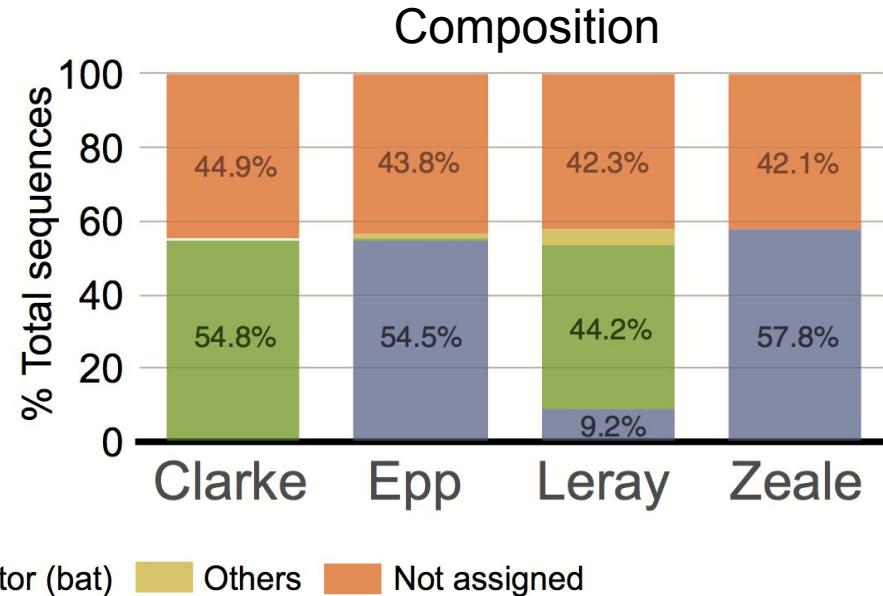
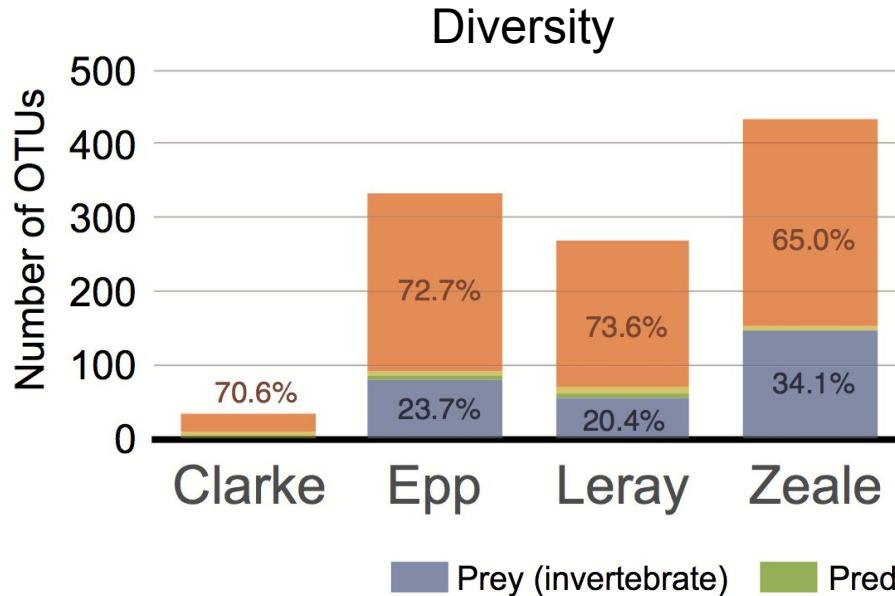
# Primers



## 2 Metabarcoding

# Primers

Different primers = Different results



## 2 Metabarcoding

# Primers

**PRIMER SEQUENCE:** ATGCGAACGGCTACCAC

**SPECIES 1** ATGCGAACGGCTACCAC  
ATGCGAACGGCTACCAACTAATCGGACTACAGCTAACAGACT

**SPECIES 2** ATGCGAACGGCTACCAC  
ATGCCAACGGCTTCCACTAATCGGACTACAGCTAACAGACT

**SPECIES 3** ATGCGAACGGCTACCAC  
ATGCCAACCGGCTTGGACTAATCGGACTACAGCTAACAGACT

**SPECIES 4** ATGC~~G~~AACGGCTACCAC  
ATGCC~~CCCC~~CCTGCTTGGACTAATCGGACTACAGCTAACAGACT

Mismatches

Different amplification success

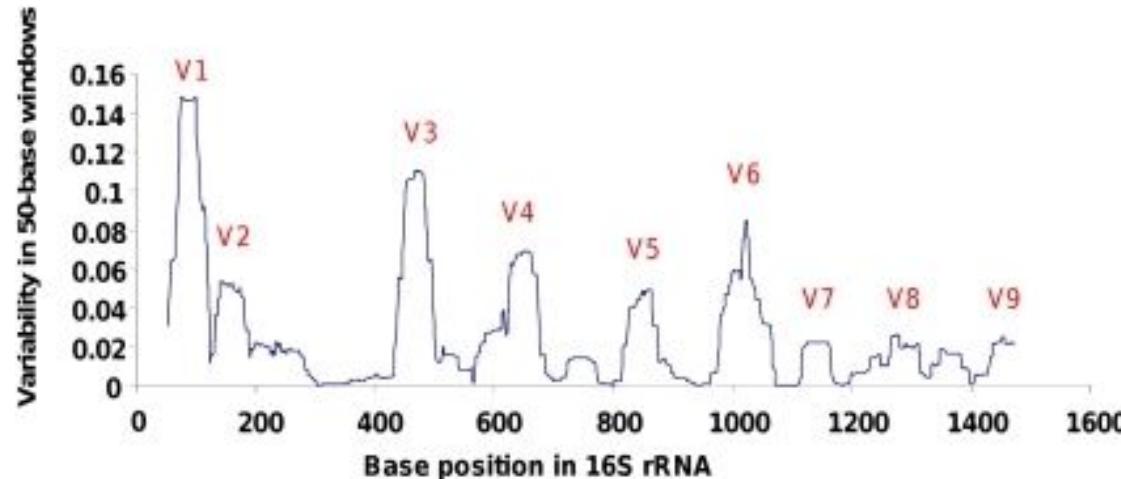
## 2 Metabarcoding

# Primers

C V Conserved Variable Conserved

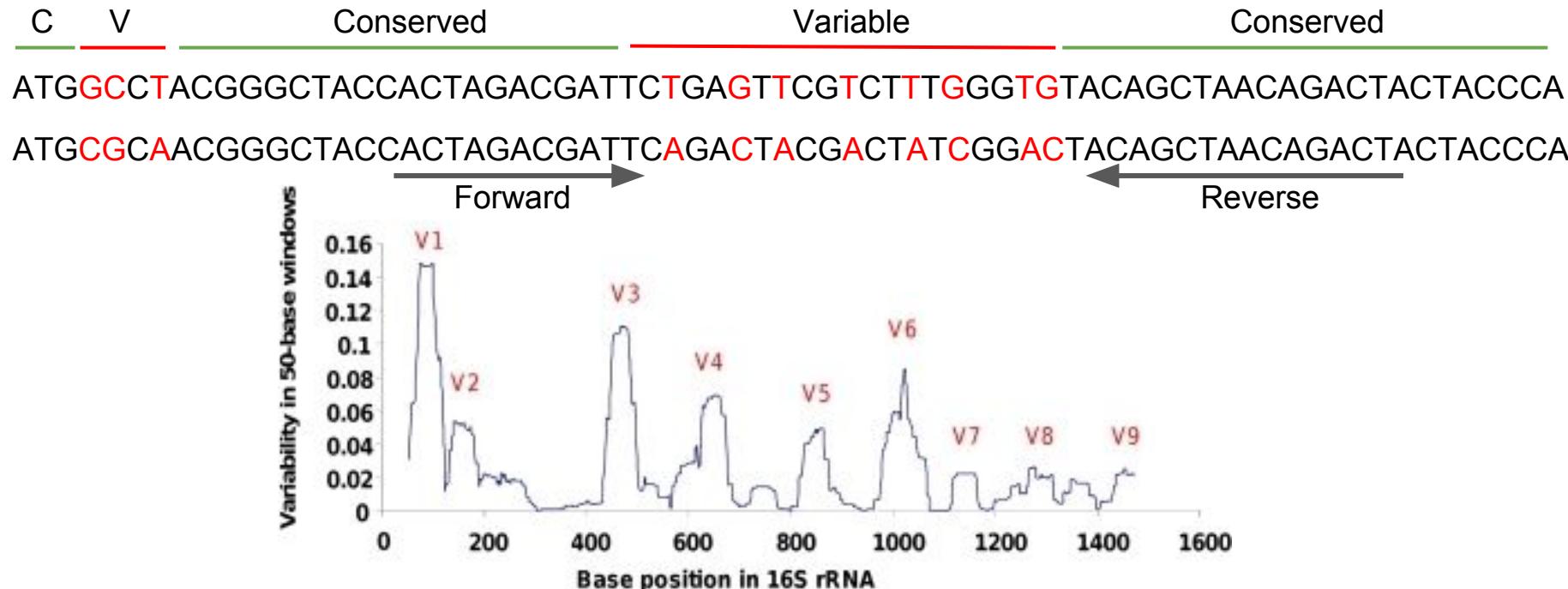
ATGG**CCTACGGGCTACC**ACTAGACGATT**TGAGTTCGTCTTG**GGGTGTACAGCTAACAGACTACTACCCA

ATG**CGCAACGGGCTACC**ACTAGACGATT**CAGACTACGACTATCGGACTAC**AGCTAACAGACTACTACCCA



## 2 Metabarcoding

# Primers



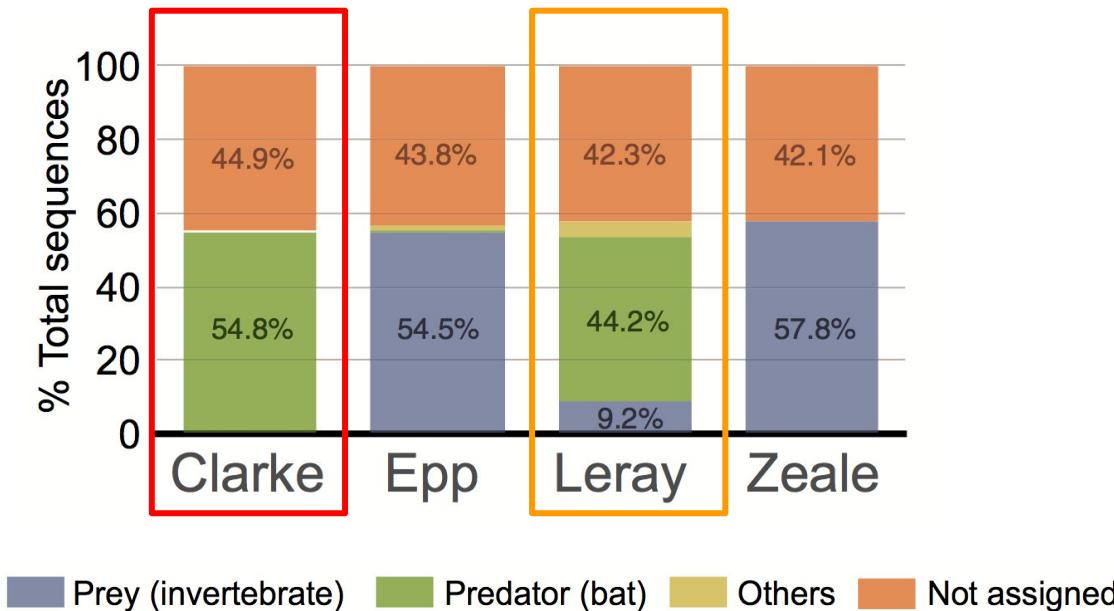
## 2 Metabarcoding

# Primers

		Target taxa
SPECIES 1	ATGCGCAACGGGCTACCAC ATGCGCAACGGGCTACCACTAATCGGACTACAGCTAACAGACT	
SPECIES 2	ATGC <span style="color:red">G</span> CAACGGGCT <span style="color:red">A</span> CCAC ATGCCCAACGGGCT <span style="color:red">T</span> CCACTAATCGGACTACAGCTAACAGACT	
SPECIES 3	ATGC <span style="color:red">G</span> CAAC <span style="color:red">G</span> GGCT <span style="color:red">A</span> CCAC ATGCCCAAC <span style="color:red">C</span> GGCT <span style="color:red">T</span> GGACTAATCGGACTACAGCTAACAGACT	
SPECIES 4	ATGC <span style="color:red">G</span> CAAC <span style="color:red">G</span> GGCT <span style="color:red">A</span> CCAC ATGCC <span style="color:red">CCCCC</span> T <span style="color:red">G</span> CT <span style="color:red">T</span> GGACTAATCGGACTACAGCTAACAGACT	Non-target taxa

## 2 Metabarcoding

# Primers



## 2 Metabarcoding

# Primers

SPECIES 1	ATGCGAACGGGCTACCA ATGCGAACGGGCTACCACTAATCGGACTACAGCTAACAGACT
SPECIES 2	ATGCGAACGGGCTACCAC ATGCCAACGGGCTTCCACTAATCGGACTACAGCTAACAGACT
SPECIES 3	ATGCGAACGGGCTACCAC ATGCCAACCGGCTTGGACTAATCGGACTACAGCTAACAGACT

Target  
taxa

### Actual proportions

Species 1: 33%

Species 2: 33%

Species 3: 33%

### Post-amplification proportions

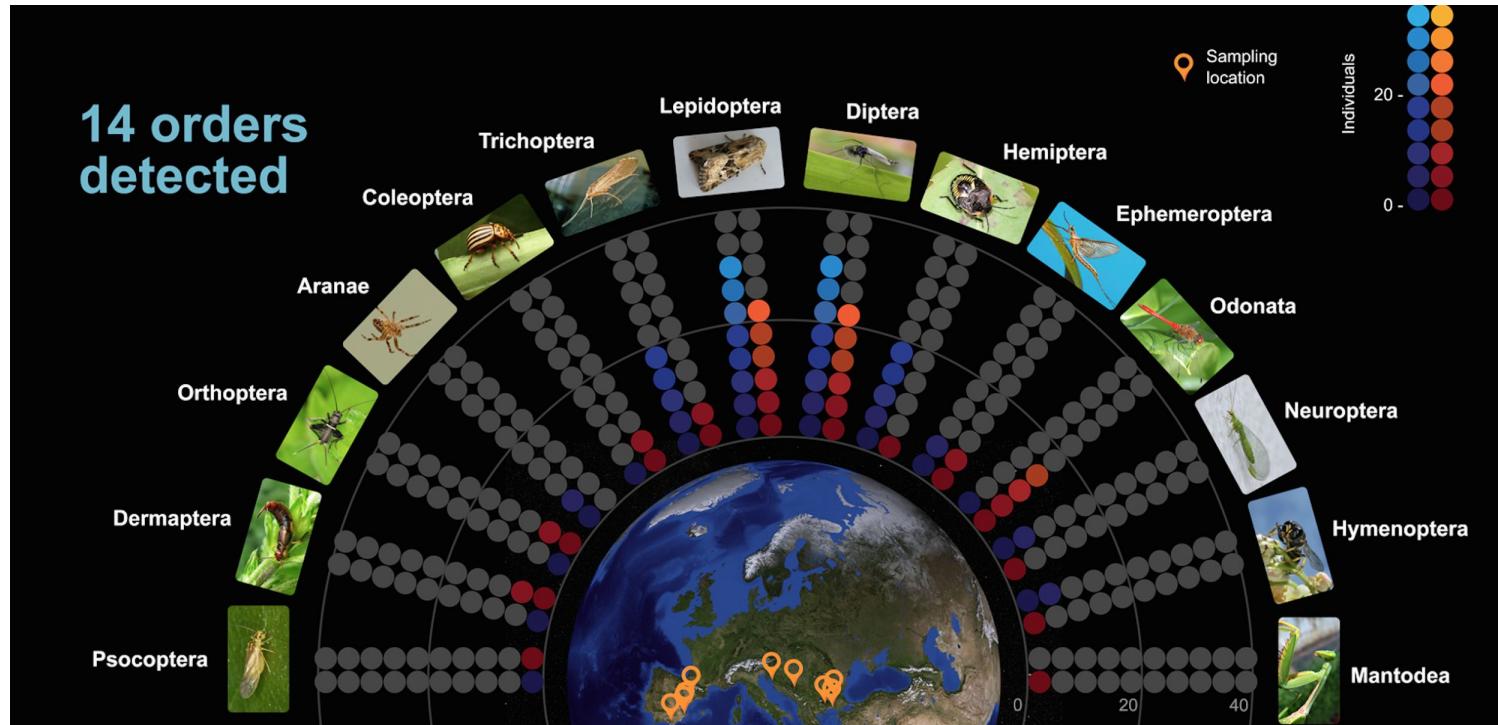
Species 1: 60%

Species 2: 35%

Species 3: 5%

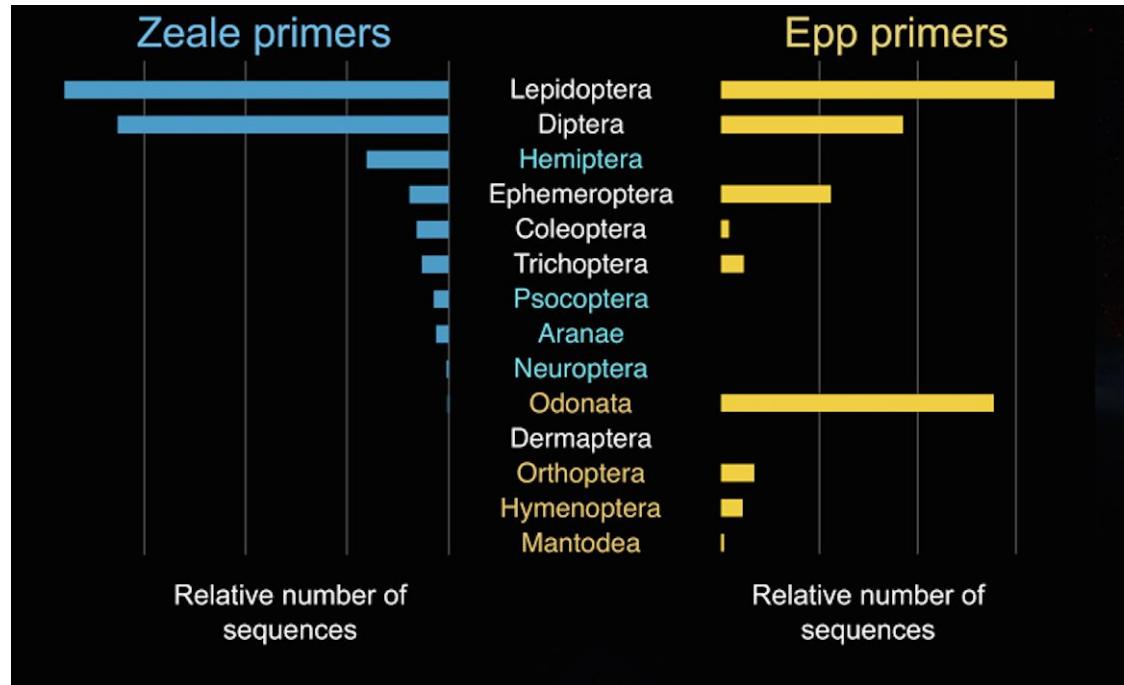
## 2 Metabarcoding

# Primers



## 2 Metabarcoding

# Primers



## 2 Metabarcoding

# Primers

**DEGENERATED PRIMER SEQUENCE:** ATGC~~S~~CAACG~~N~~GCT~~W~~SBAC

### IUPAC nucleotide code

Symbol	Bases	Description	
A	A	Adenine	ATGCGCAACGAGCTACCAC
C	C	Cytosine	ATGCCCAACGTGCTTGGAC
G	G	Guanine	ATGCCCAACGTGCTAGTAC
T (or U)	T (or U)	Thymine (or Uracil)	ATGCGCAACGGGCTTCGAC
W	A or T	Weak	ATGCGCAACGCGCTTCTAC
S	C or G	Strong	ATGCCCAACGAGCTAGCAC
M	A or C	aMino	ATGCGCAACGCGCTTCTAC
K	G or T	Keto	ATGCCCAACGAGCTAGCAC
R	A or G	puRine	ATGCCCAACGAGCTAGCAC
Y	C or T	pYrimidine	ATGCCCAACGAGCTAGCAC
B	C or G or T	not A (B comes after A)	
D	A or G or T	not C (D comes after C)	

## 2 Metabarcoding

# Primers

**DEGENERATED PRIMER SEQUENCE:** ATGC~~S~~CAACG~~N~~GCT~~W~~SBAC

ATGC~~S~~CAACG~~N~~GCT~~W~~SBAC

**SPECIES 1** ATGCGAACGGGCTACCAACTAATCGGACTACAGCTAACAGACT

ATGC~~S~~CAACG~~N~~GCT~~W~~SBAC

**SPECIES 2** ATGCCAACGGGCTTCCACTAATCGGACTACAGCTAACAGACT

ATGC~~S~~CAACG~~N~~GCT~~W~~SBAC

**SPECIES 3** ATGCCAACCGGCTTGGACTAATCGGACTACAGCTAACAGACT

## 2 Metabarcoding

# Primers

**MARKER:** 16S rRNA

**341F** CCTAYGGGRBGCASCAG  
**806R** GGGACTACNNGGGTATCTAAT

**TARGET:** prokaryotes

**MARKER:** 16S rRNA

**Meta01F** YYCRMCTGTTTAYCAAAAACA  
**Meta01R** CCGGTYTGAACTCARATC

**TARGET:** metazooa

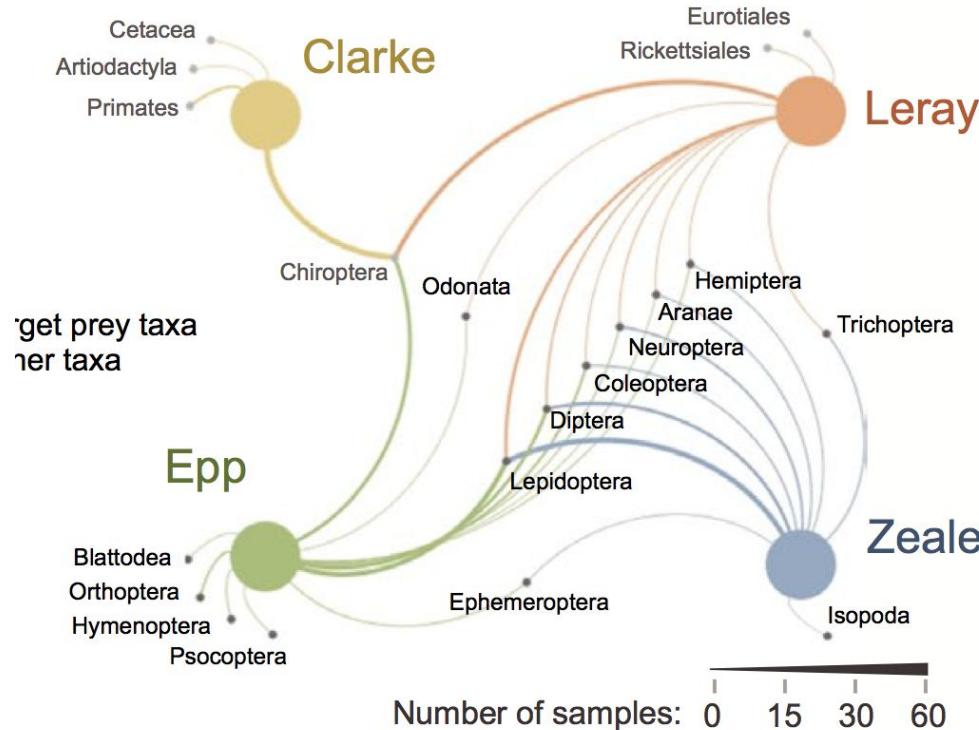
**MARKER:** 12S rRNA

**Aves01F** GATTAGATACCCCACTATGC  
**Aves01R** GTTTAAGCGTTGTGCTCG

**TARGET:** birds

## 2 Metabarcoding

# Primers



## 2 Metabarcoding



# Primers

- Ecoprimers
- PrimerMiner
- Primer3
- <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>
- <http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>
- ...

## 2 Metabarcoding

GENOMIC  
RESOURCES 

eman ta zabal zazu  
Universidad  
del País Vasco  
Euskal Herriko  
Unibertsitatea

NATURAL HISTORY MUSEUM OF DENMARK  
UNIVERSITY OF COPENHAGEN

# Primers



1) In Silico



2) Mock communities



3) Real samples

## 2 Metabarcoding

Leray et al. *Frontiers in Zoology* 2013, 10:34  
http://www.frontiersinzoology.com/content/10/1/34



Open Access

A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents

Matthieu Leray<sup>1,2\*</sup>, Joy Y Yang<sup>3</sup>, Christopher P Meyer<sup>2</sup>, Suzanne C Mills<sup>1</sup>, Natalia Agudelo<sup>2</sup>, Vincent Ranwez<sup>4</sup>, Joel T Boehm<sup>5,6</sup> and Ryuji J Machida<sup>7</sup>

### New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems

LAURA S. EPP,<sup>\*1</sup> SANNE BOESSENKOOL,<sup>\*1</sup> EVA P. BELLEMAIN,<sup>\*</sup> JAMES HAILE,<sup>†‡</sup> ALFONSO ESPOSITO,<sup>\*¶¶</sup> TIAYYBA RIAZ,<sup>§</sup> CHRISTER ERSÉUS,<sup>¶</sup> VLADIMIR I. GUSAROV,<sup>\*</sup> MARY E. EDWARDS,<sup>\*\*</sup> ARILD JOHNSEN,<sup>\*</sup> HANS K. STENØIEN,<sup>†‡</sup> KRISTIAN HASSEL,<sup>†‡</sup> HÅVARD KAUSERUD,<sup>†‡</sup> NIGEL G. YOCCOZ,<sup>§§</sup> KARI ANNE BRÄTHEN,<sup>§§</sup> ESKE WILLERSLEV,<sup>†</sup> PIERRE TABERLET,<sup>§</sup> ERIC COISSAC<sup>§</sup> and CHRISTIAN BROCHMANN<sup>\*2</sup>

<sup>\*</sup>National Centre for Biosystematics, Natural History Museum, University of Oslo, PO Box 1172, Blindern, NO-0318 Oslo, Norway, <sup>†</sup>The Centre of Excellence for GeoGenetics, Natural History Museum of Denmark, Øster Voldgade 5-7, 1350 Copenhagen K, Denmark, <sup>‡</sup>Ancient DNA Research Laboratory, Murdoch University, South Street, Perth, WA 6150 Australia,

<sup>§</sup>Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France, <sup>¶</sup>Department of Biological and Environmental Sciences, University of Gothenburg, Box 463, SE-405 30 Göteborg, Sweden,

<sup>\*\*</sup>Geography and Environment, University of Southampton, University Road, Southampton SO17 1BJ, UK, <sup>†‡</sup>Systematics and Evolution Group, Museum of Natural History and Archeology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway, <sup>†‡</sup>Microbial Evolution Research Group (MERG), Department of Biology, University of Oslo, PO Box 1066 Blindern, N-0316, Oslo, Norway, <sup>§§</sup>Department of Arctic and Marine Biology, University of Tromsø, NO-9037 Tromsø, Norway



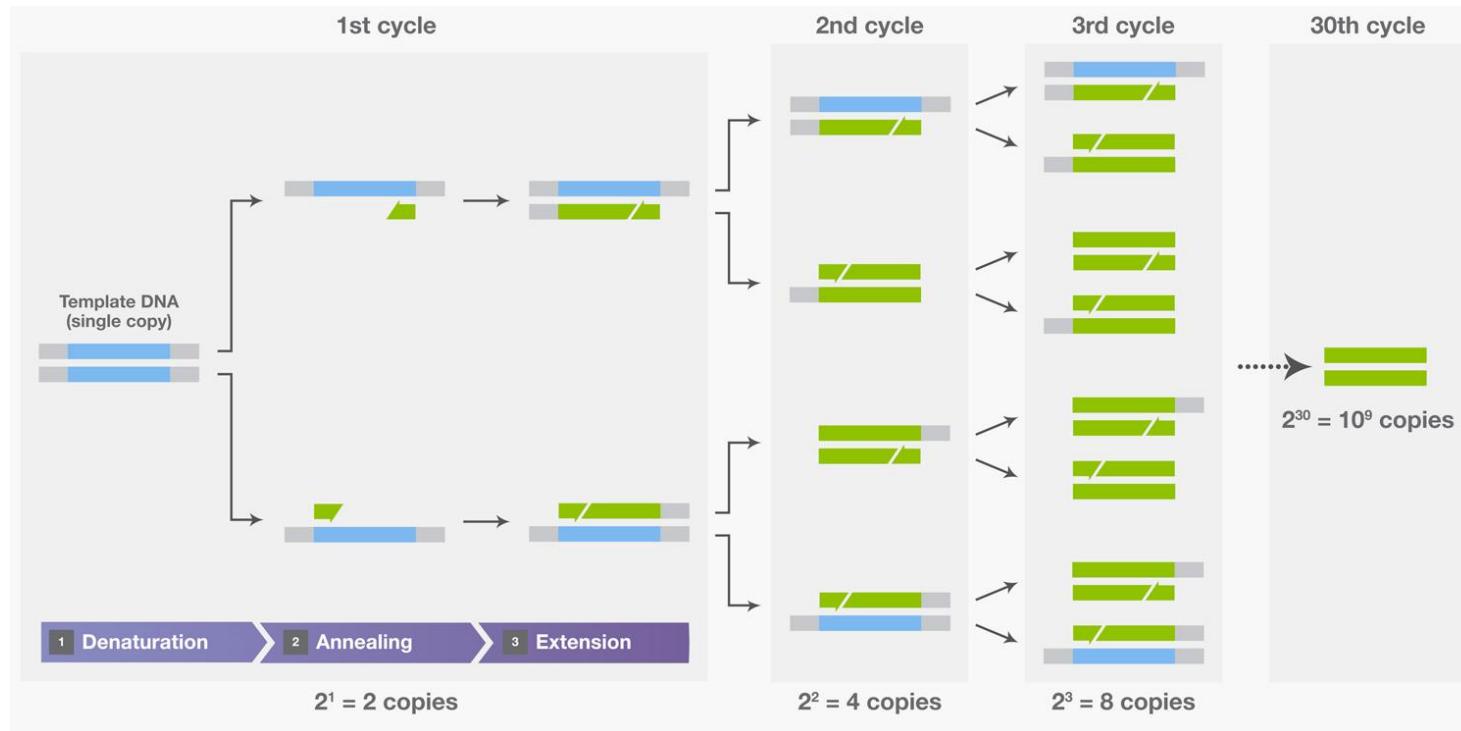
## Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment

Vasco Elbrecht<sup>1\*</sup> and Florian Leese<sup>1,2</sup>

<sup>1</sup>Aquatic Ecosystem Research, Faculty of Biology, University of Duisburg-Essen, Essen, Germany, <sup>2</sup>Centre for Water and Environmental Research, University of Duisburg-Essen, Essen, Germany

## 2 Metabarcoding

# PCR amplification



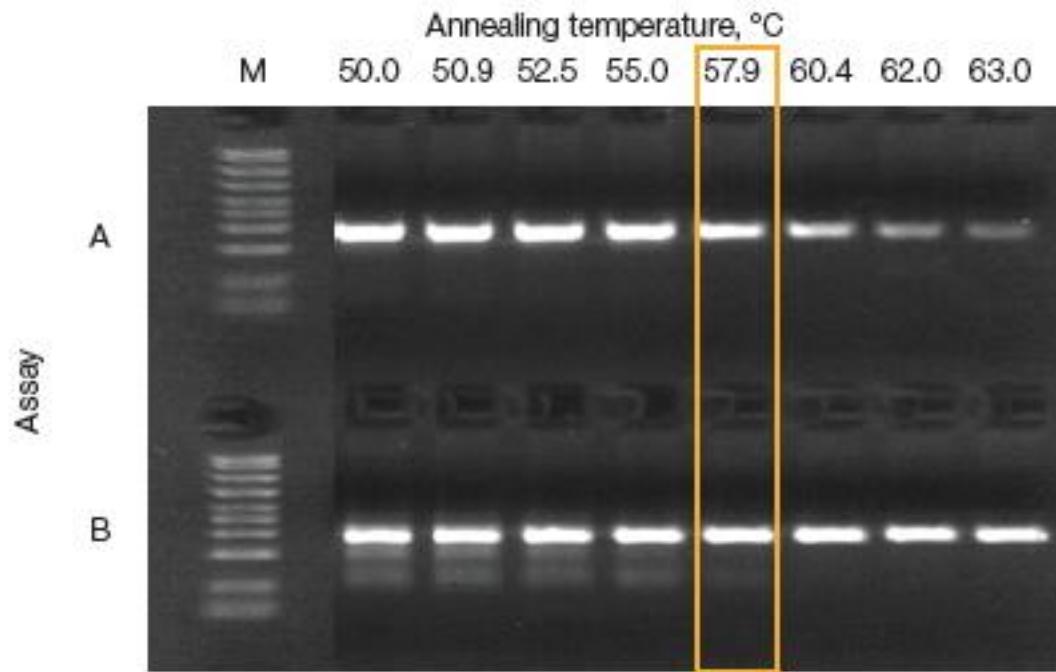
## 2 Metabarcoding

# PCR amplification

Zeale	Leray	Epp	Clarke
<b>1x</b> 95 °C 10 min	<b>1x</b> 95 °C 10 min	<b>1x</b> 95 °C 10 min	<b>1x</b> 95 °C 10 min
<b>40x</b> 95 °C 20 sec 55 °C 30 sec 72 °C 1 min	<b>36x</b> 95 °C 20 sec 51 °C 30 sec 72 °C 1 min	<b>37x</b> 95 °C 20 sec 55 °C 30 sec 72 °C 30 sec	<b>34x</b> 95 °C 20 sec 54 °C 30 sec 72 °C 30 sec
<b>1x</b> 72 °C 7 min	<b>1x</b> 72 °C 7 min	<b>1x</b> 72 °C 7 min	<b>1x</b> 72 °C 7 min
<b>1x</b> 4 °C hold	<b>1x</b> 4 °C hold	<b>1x</b> 4 °C hold	<b>1x</b> 4 °C hold

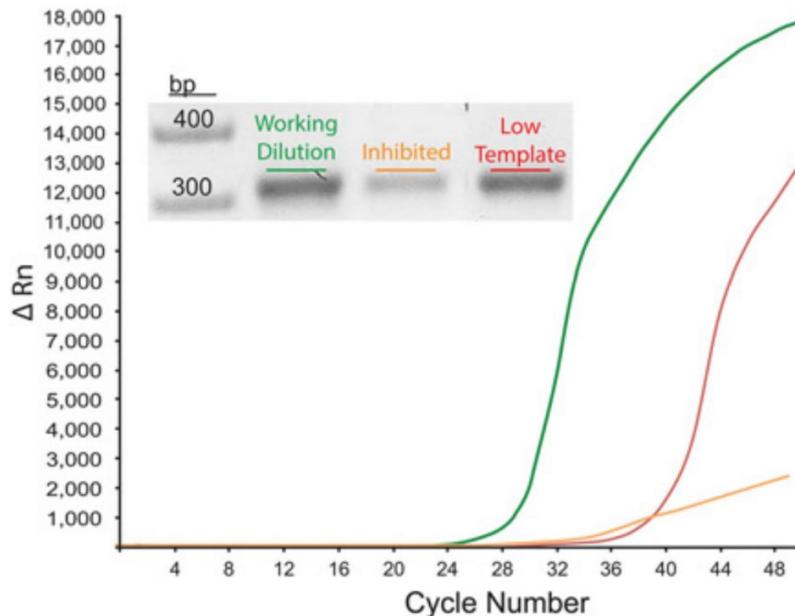
## 2 Metabarcoding

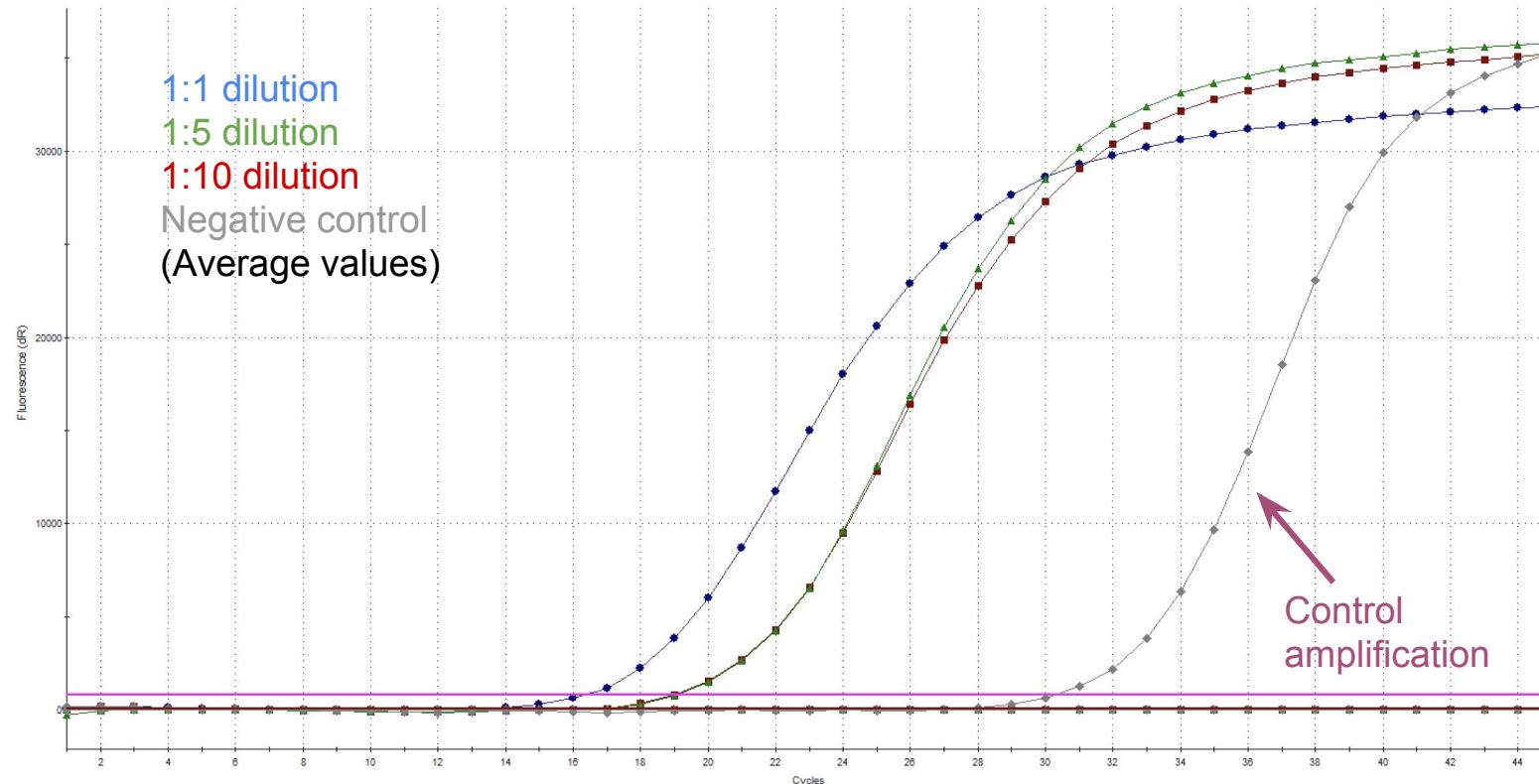
# PCR amplification

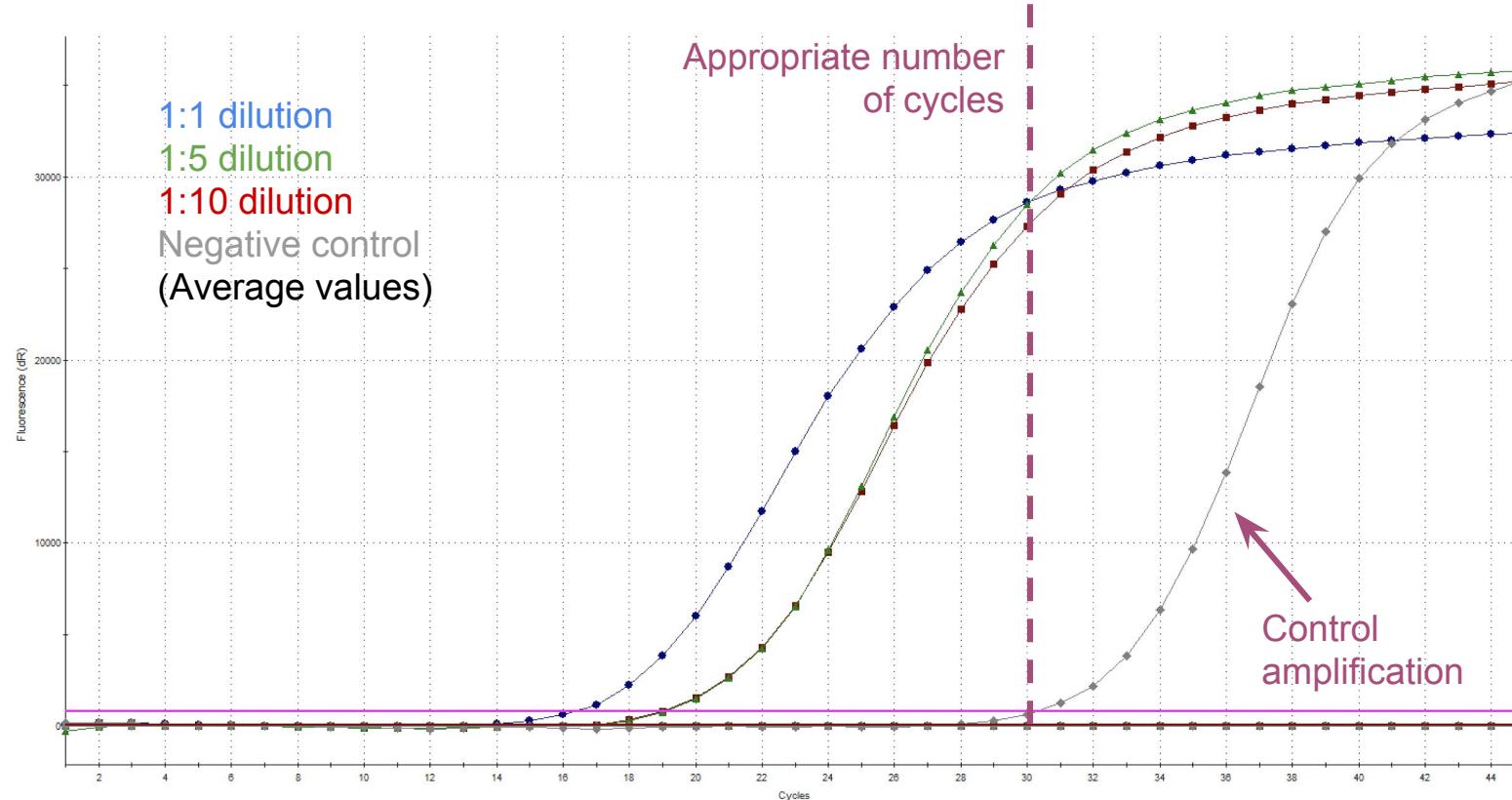


## 2 Metabarcoding

# PCR amplification

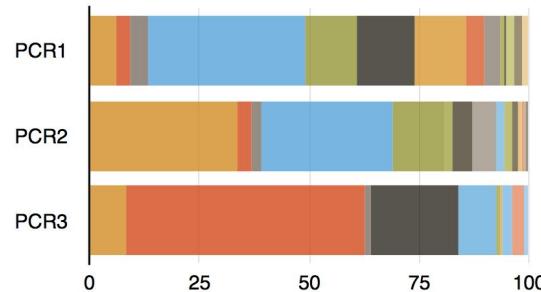
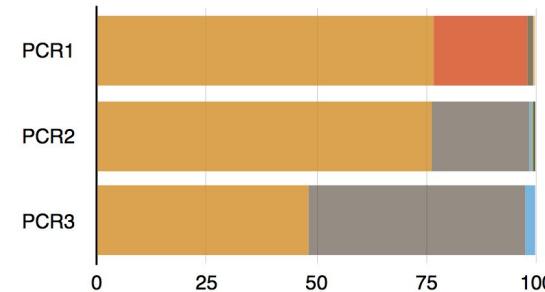
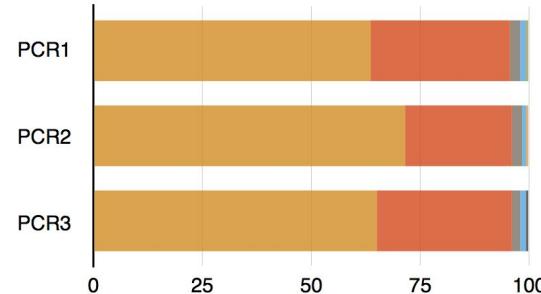
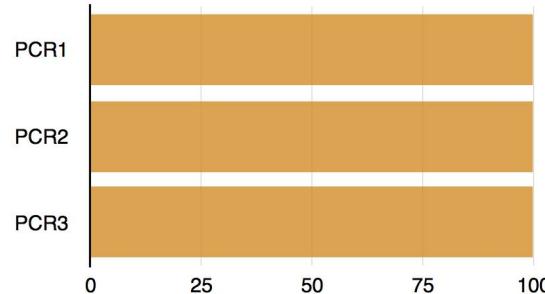






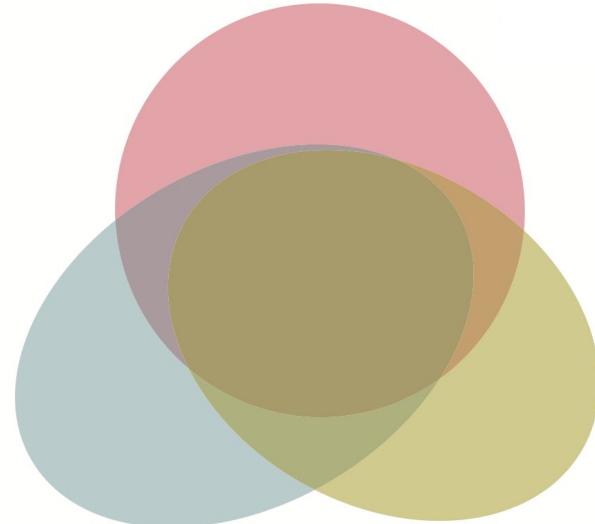
## 2 Metabarcoding

# PCR replicates



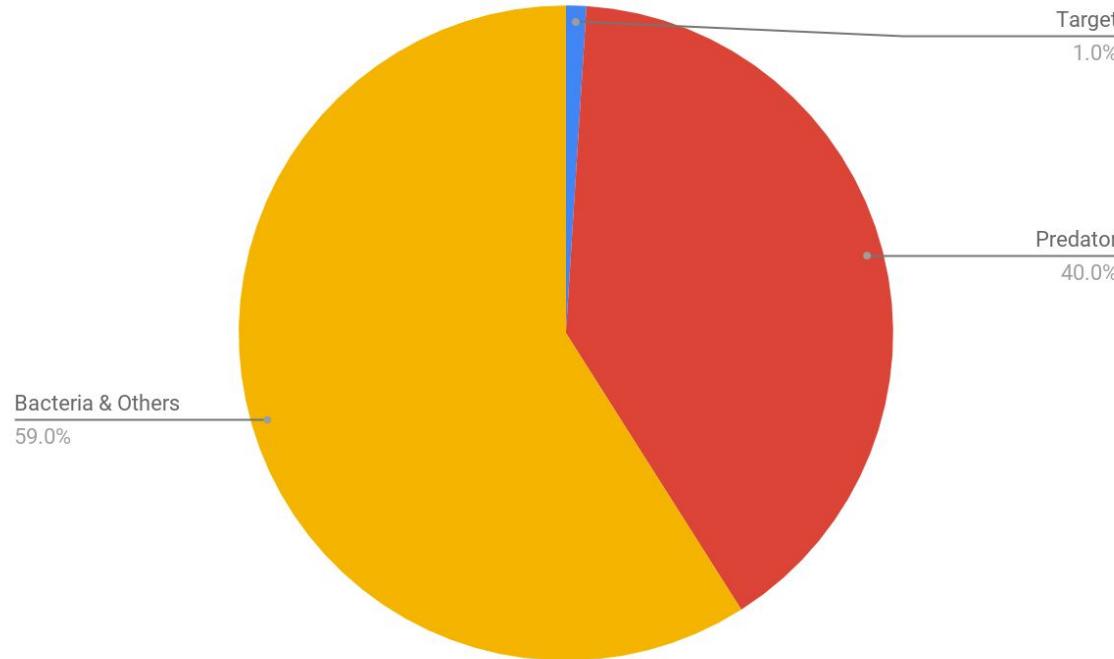
## 2 Metabarcoding

# PCR replicates



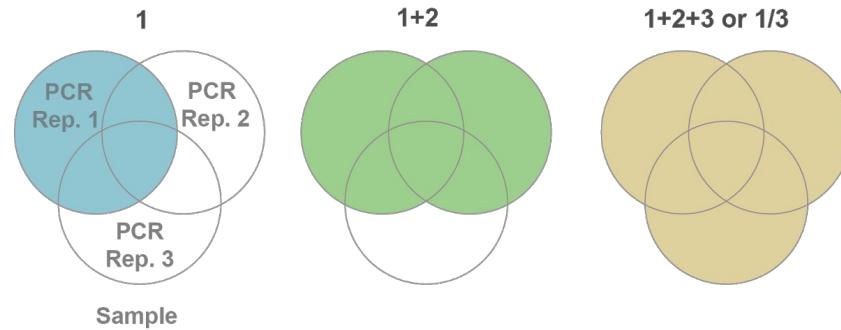
## 2 Metabarcoding

# PCR replicates



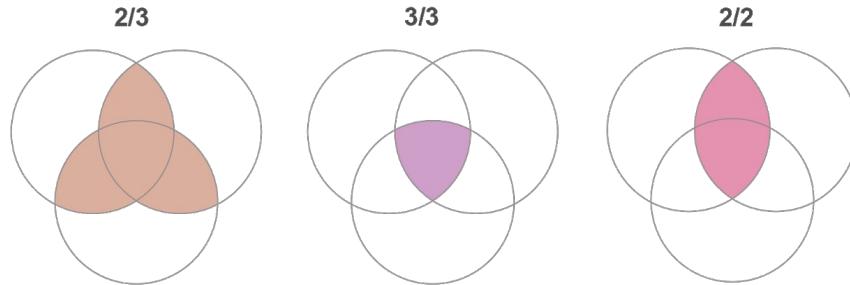
# PCR replicates

## Replicate filtering strategies



# PCR replicates

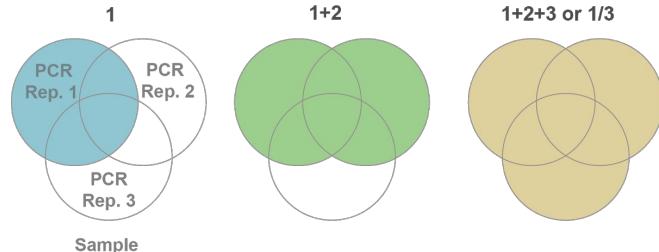
## Replicate filtering strategies



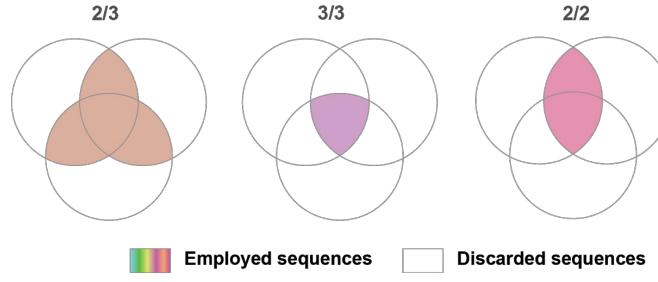
## 2 Metabarcoding

# PCR replicates

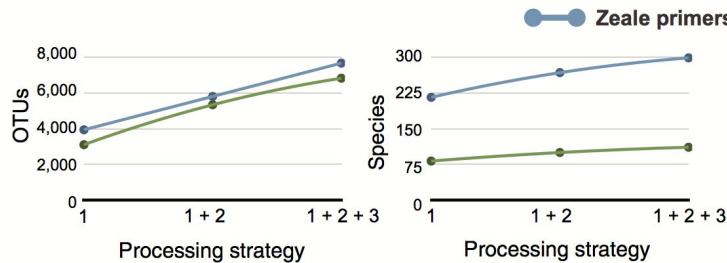
A) Additive PCR replicate filtering approaches



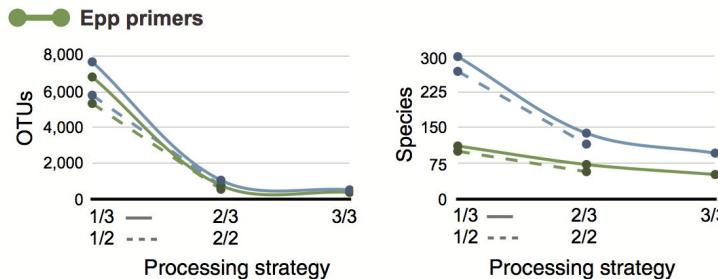
B) Restrictive PCR replicate filtering approaches



(a) Additive PCR replicate processing strategies

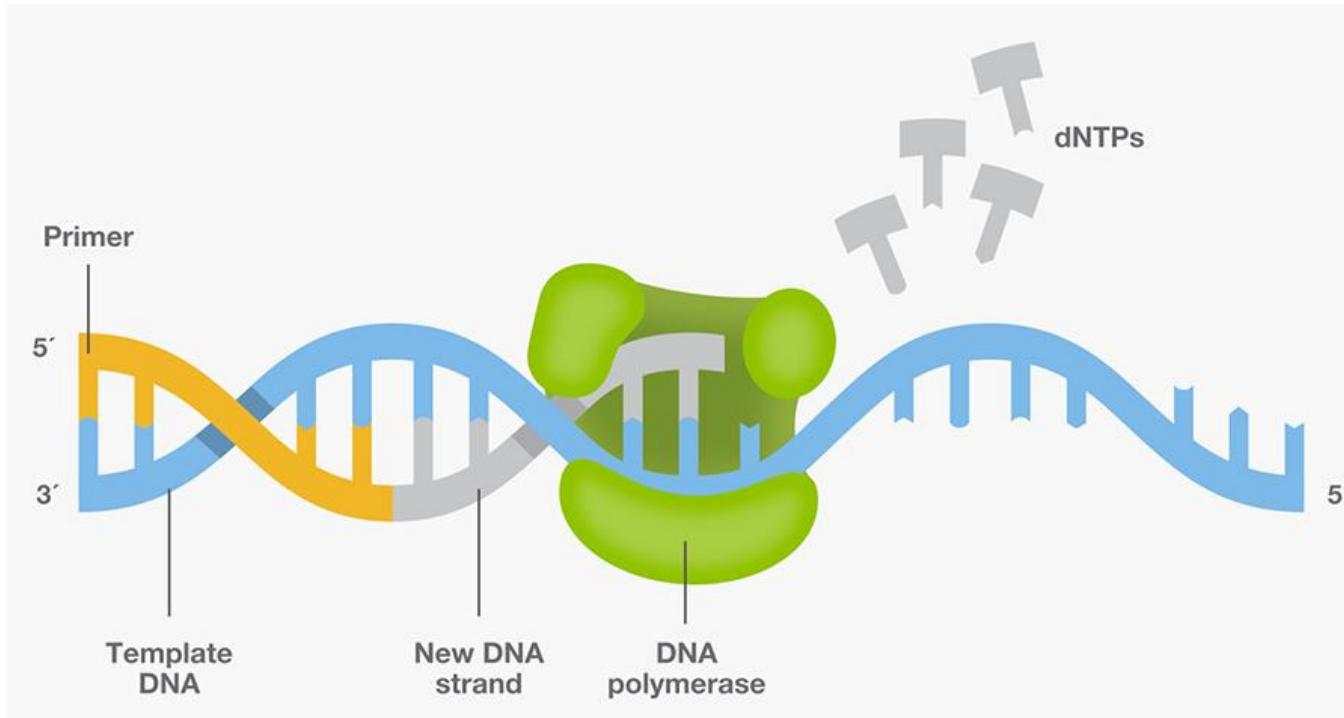


(b) Restrictive PCR replicate processing strategies



## 2 Metabarcoding

# PCR errors



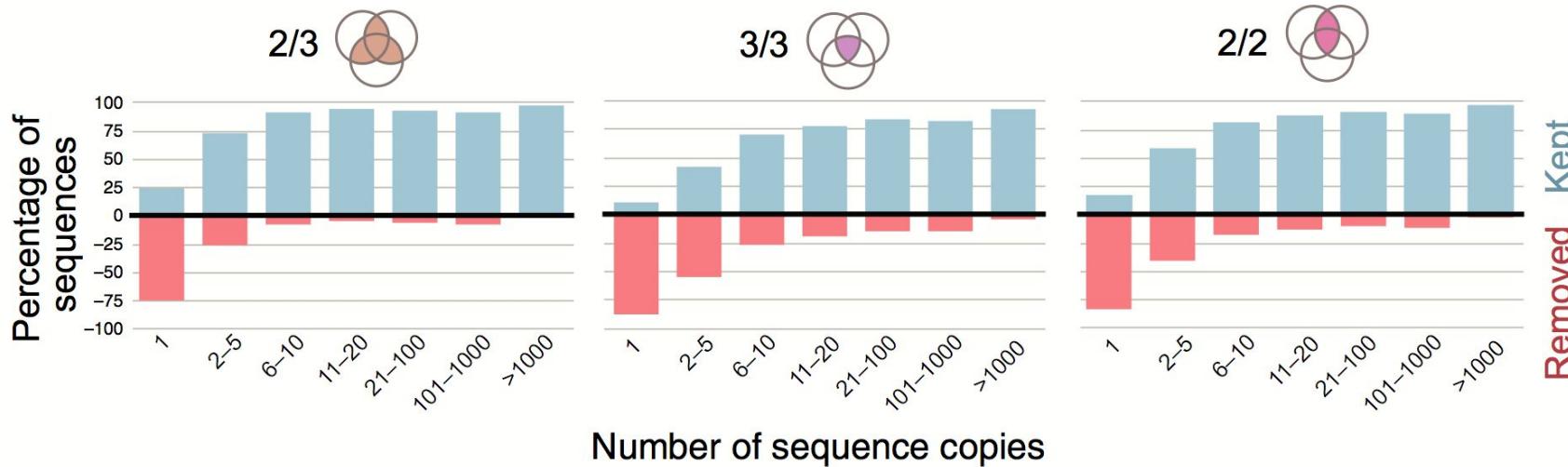
## 2 Metabarcoding

# PCR replicates

Phred Quality Score	Probability of Incorrect Base Call	Base Call Accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%

## 2 Metabarcoding

# PCR replicates

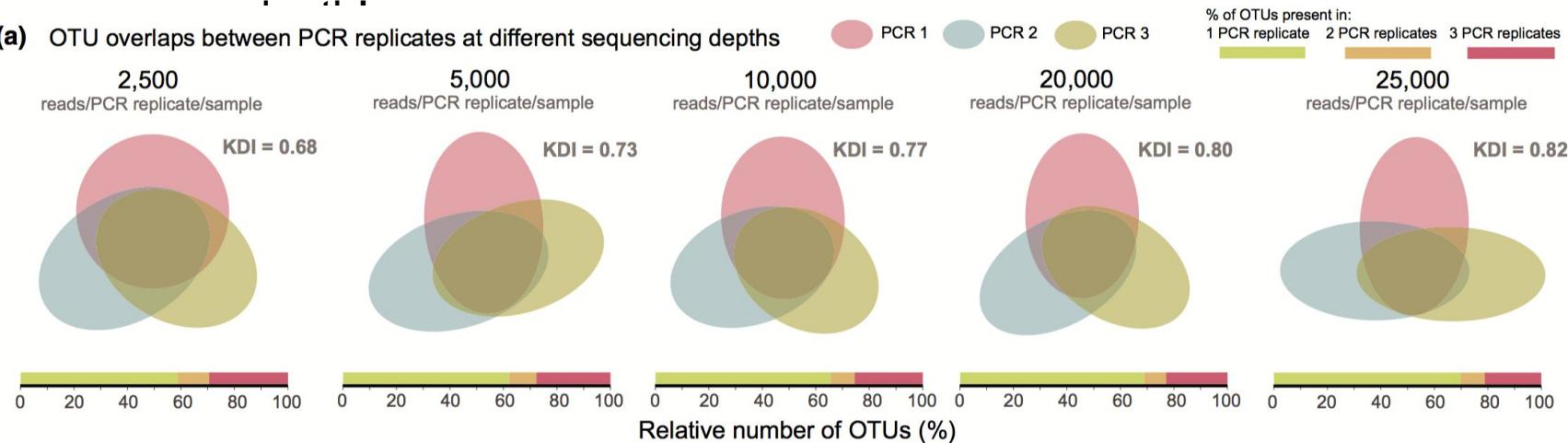


## 2 Metabarcoding

# PCR replicates

- Extremely popular: singleton removal
- Copy number of artifactual sequences depend on sequencing

(a) OTU overlaps between PCR replicates at different sequencing depths

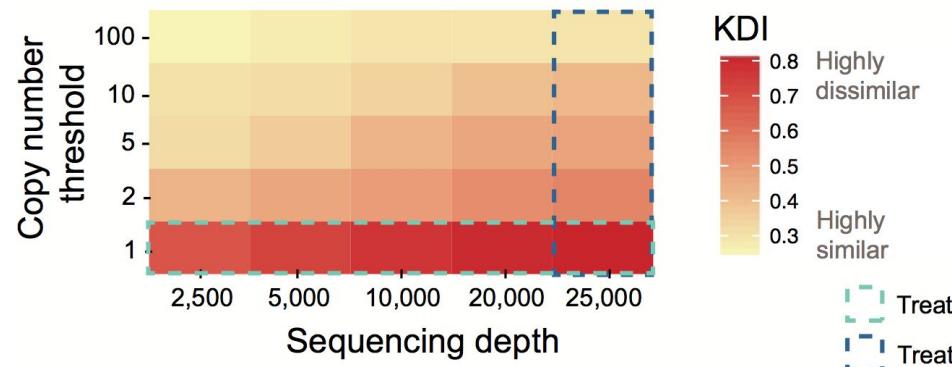


## 2 Metabarcoding

# PCR replicates

- Extremely popular: singleton removal
- Copy number of artifactual sequences depend on sequencing depth!

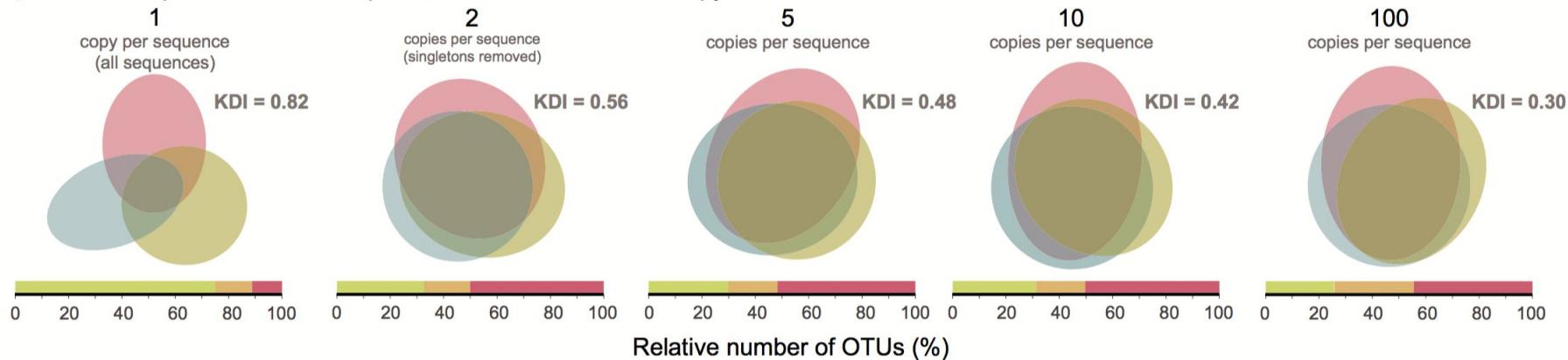
(c) Effect of copy number threshold and sequencing depth on dissimilarity between PCR replicates



## 2 Metabarcoding

# PCR replicates

(b) OTU overlaps between PCR replicates at different minimum copy number thresholds



## 2 Metabarcoding

# Multiplexing



Illumina  
MiSeq

**Output:**  
20 million sequences

**Diet:** 10.000-20.000 sequences/sample



1000 samples / sequencing run

**Microorganisms:** 20-50.000 sequences/replicate



400 samples / sequencing run

## 2 Metabarcoding

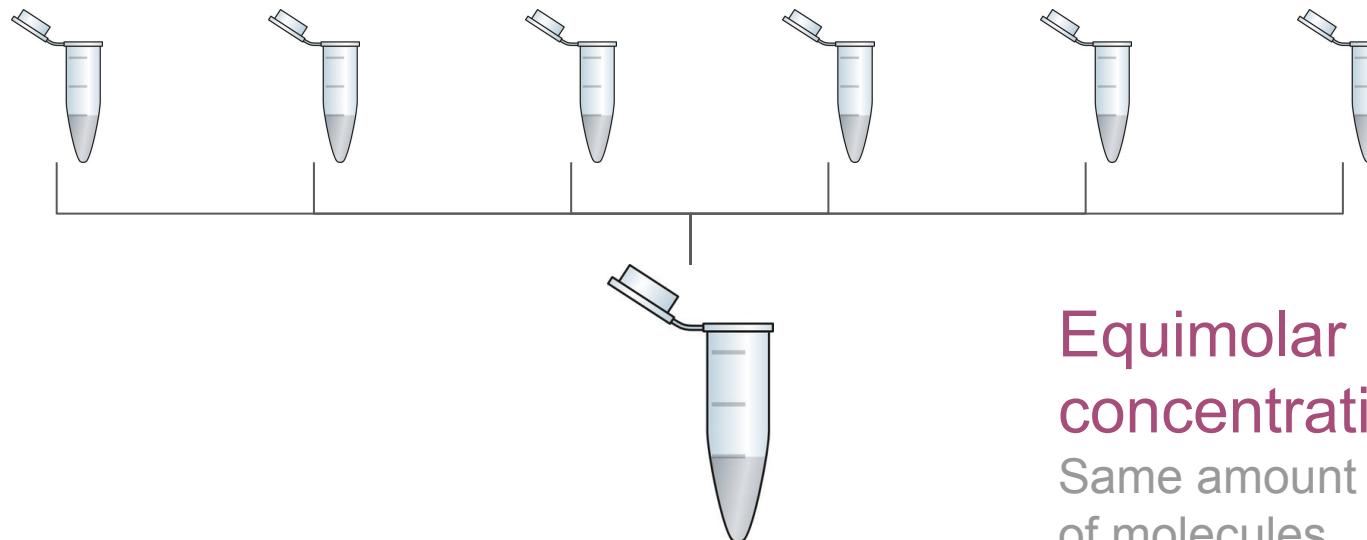
# Multiplexing

Tag 1 **ATGCGCA**ACGGGCTACCACTAATCGGACTACAGCTAACAGACT  
Tag 2 **AGGCCCA**ACGGGCTACCACTAATCGGACTACAGCTAACAGACT  
Tag 3 **TAGCCCT**ACGGGCTACCACTAATCGGACTACAGCTAACAGACT  
Tag 4 **GATCTGA**ACGGGCTACCACTAATCGGACTACAGCTAACAGACT  
Tag 5 **ATGCCGT**ACGGGCTACCACTAATCGGACTACAGCTAACAGACT

## 2 Metabarcoding

# Pooling

Sample 1      Sample 2      Sample 3      Sample 4      Sample 5      Sample 6

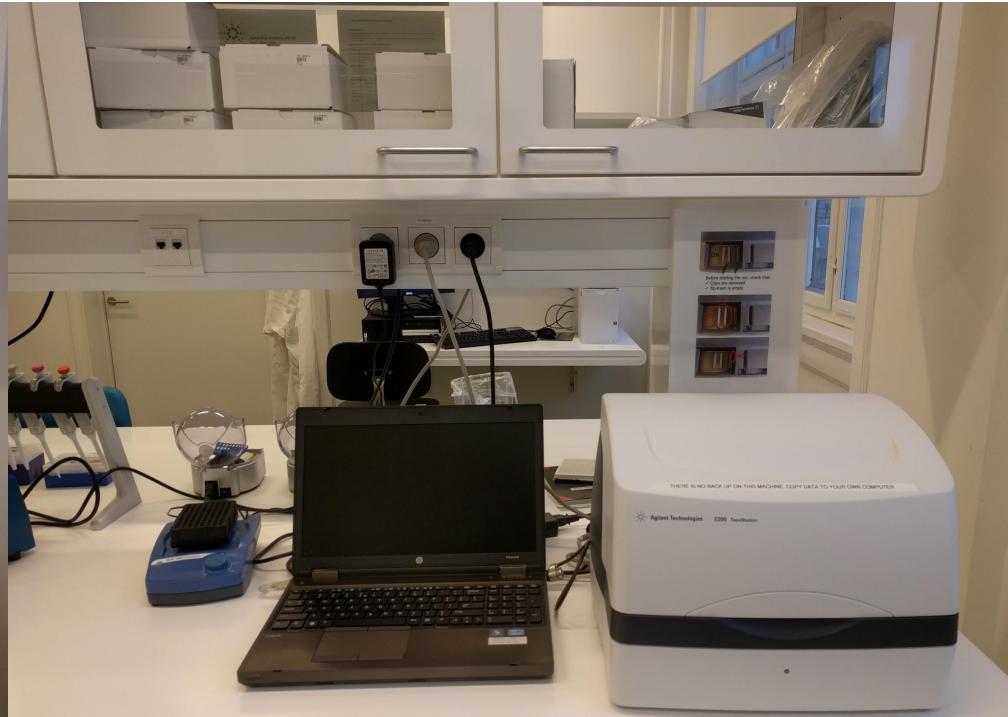


Equimolar  
concentrations  
Same amount  
of molecules

Amplicon pool/Library pool

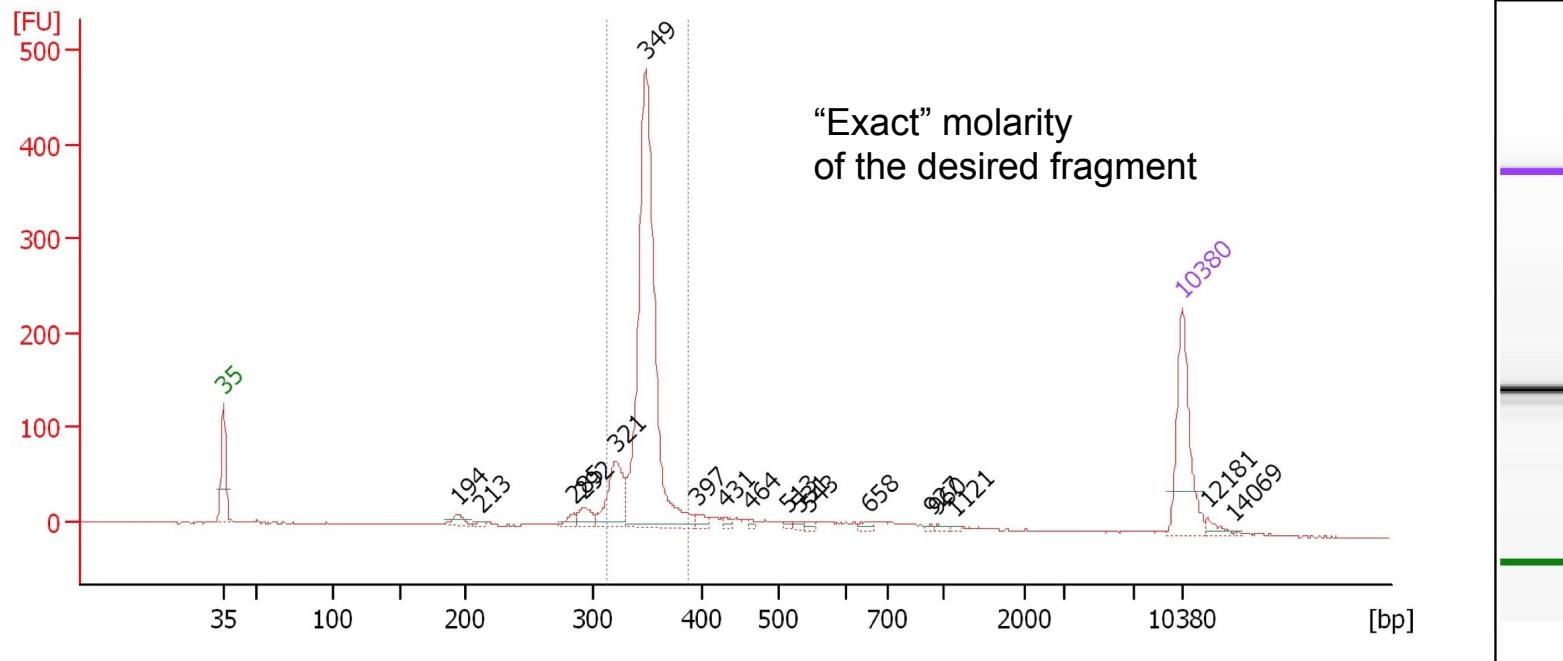
## 2 Metabarcoding

# Agilent Bioanalyzer / TapeStation



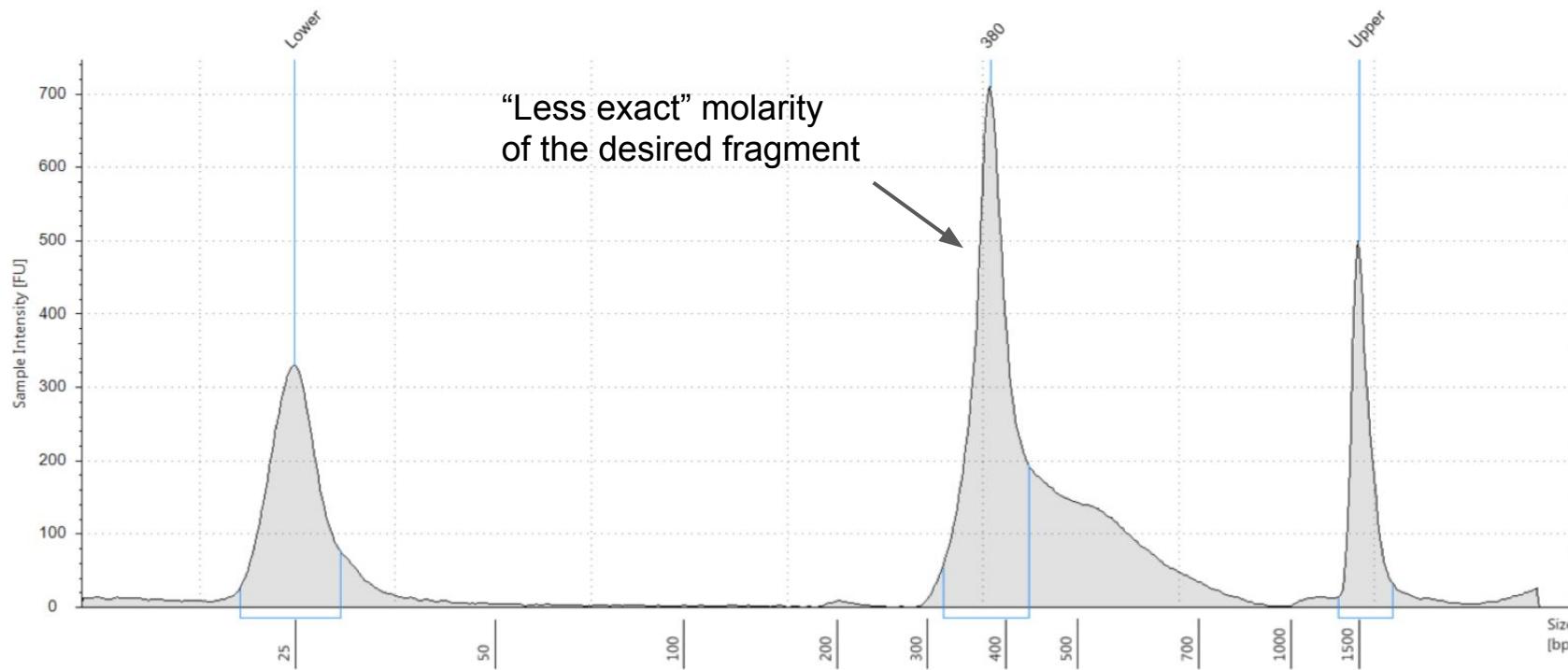
## 2 Metabarcoding

# Agilent Bioanalyzer



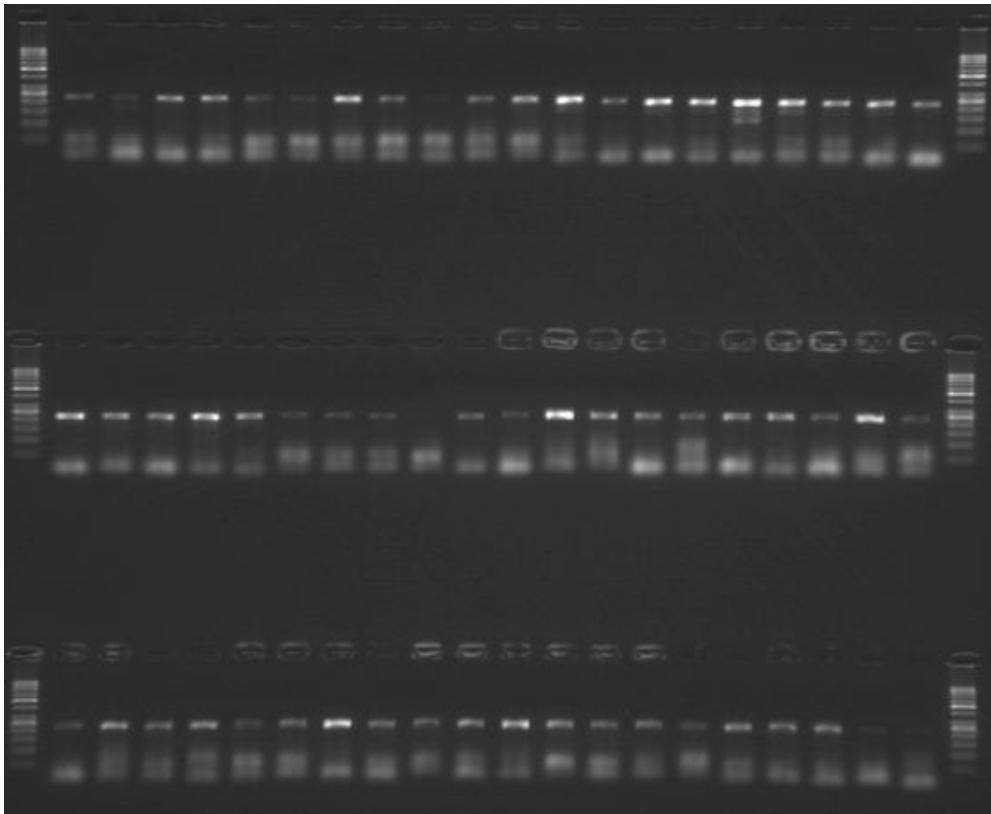
## 2 Metabarcoding

# Agilent Tapestation



## 2

# Metabarcoding

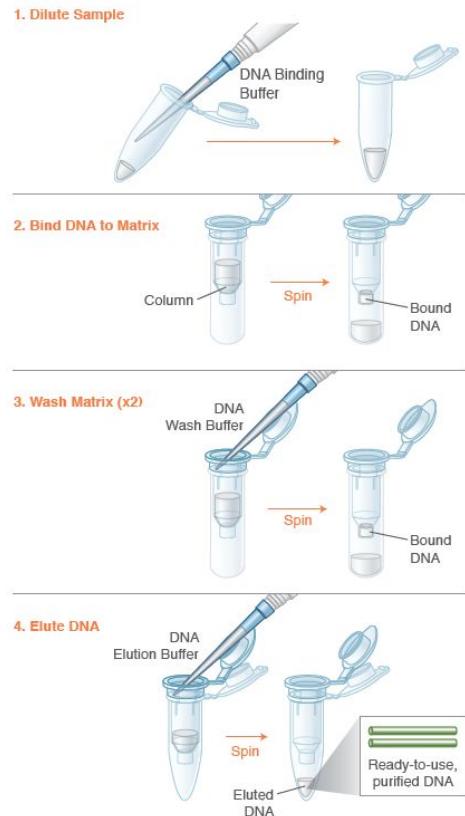


**2% agarose gel**

Cheaper  
Far less accurate

## 2 Metabarcoding

# Purification

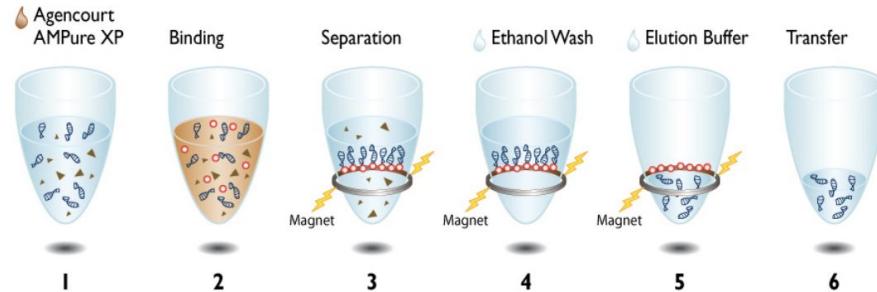


## Silica columns

- Qiagen Qiaquick
- Qiagen Minelute
- NEB Monarch
- ...

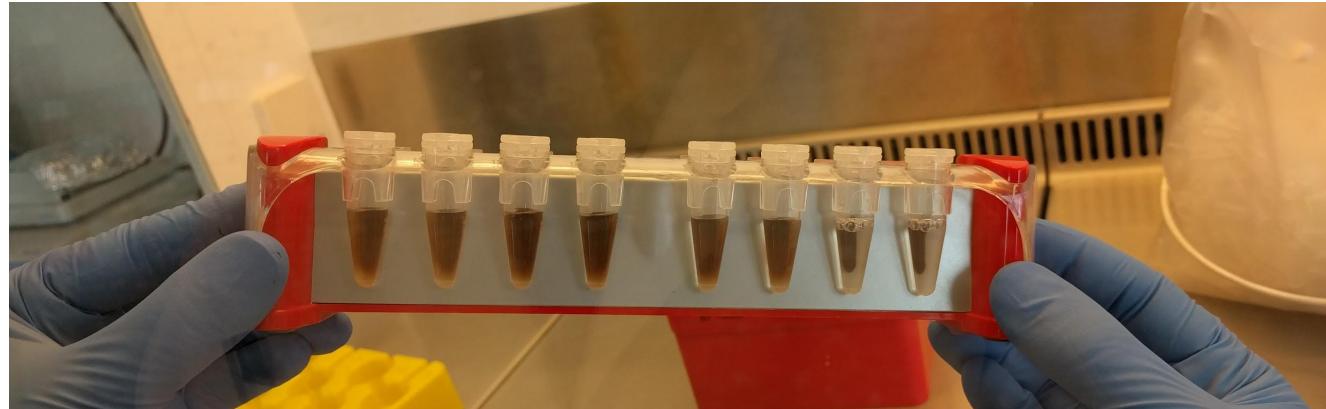
## 2 Metabarcoding

# Purification



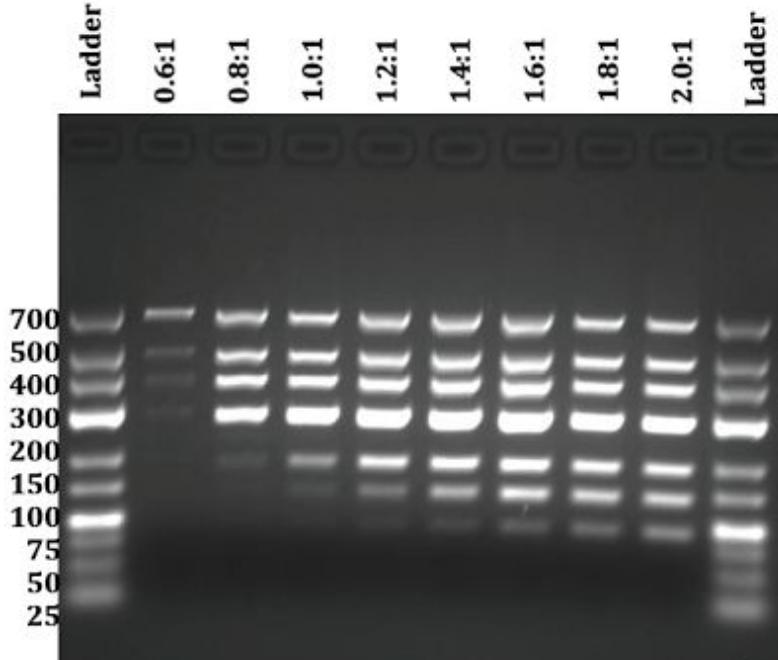
# Magnetic beads

- Agencourt AMPure XP
- ThermoFisher MagJET
- ...



## 2 Metabarcoding

### Purification



keatslab.org

### Magnetic beads

- Agencourt AMPure XP
- ThermoFisher MagJET
- ...

Bead:DNA volume ratio  
Size-selection

## 2 Metabarcoding

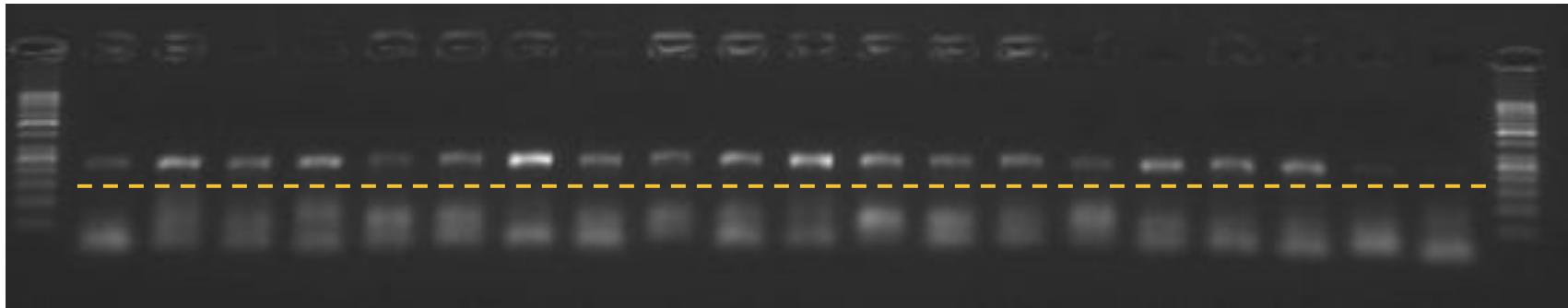
# Purification

## Magnetic beads

Beads 0.8 - 1 DNA

Beads 80 ul - 100 ul DNA

Cutoff around 300bp



## 2 Metabarcoding



# Library preparation

DNA sequence



## 2 Metabarcoding



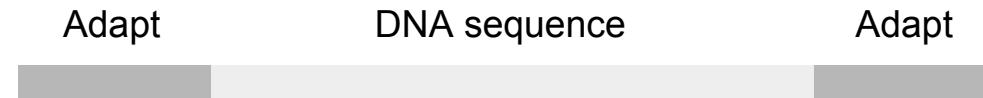
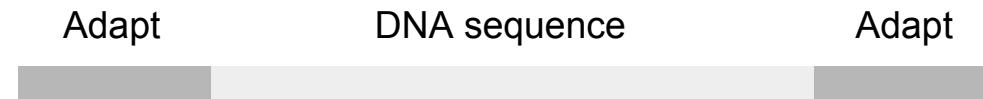
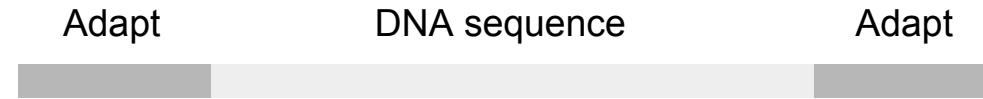
# Library preparation

Adapt                  DNA sequence                  Adapt



## 2 Metabarcoding

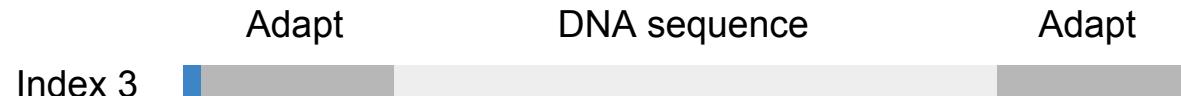
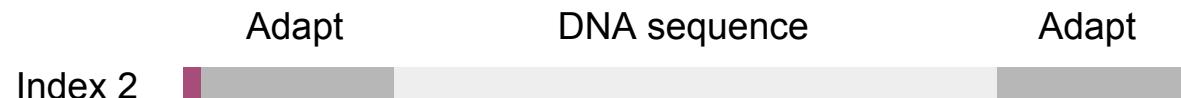
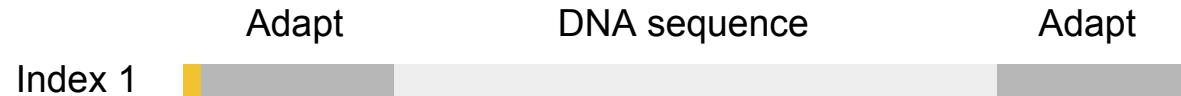
# Library preparation



Multiple samples sequenced together

## 2 Metabarcoding

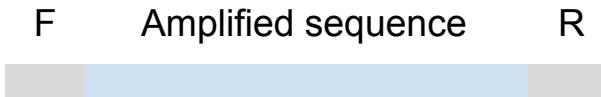
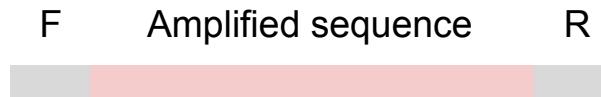
# Library preparation



Multiple samples sequenced together  
Indexed adaptors

## 2 Metabarcoding

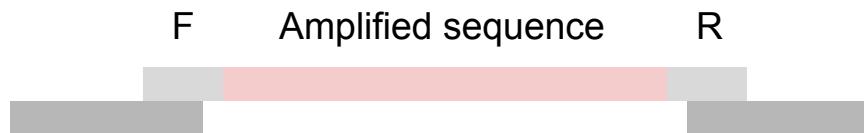
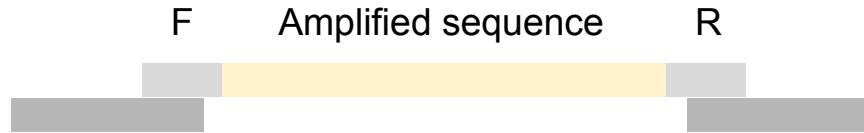
# Library preparation



Primers: common sequences

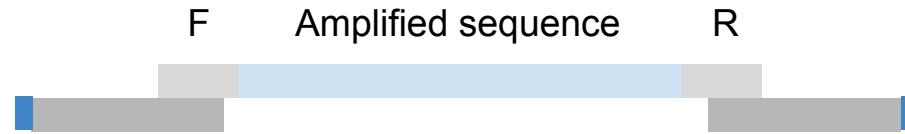
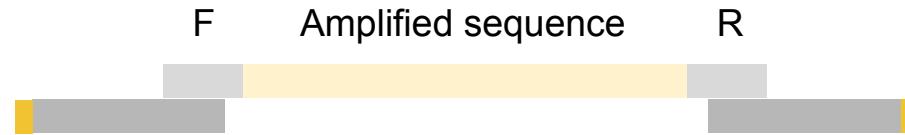
## 2 Metabarcoding

# Library preparation



Adaptors added by PCR  
Risk of chimera formation

# Library preparation

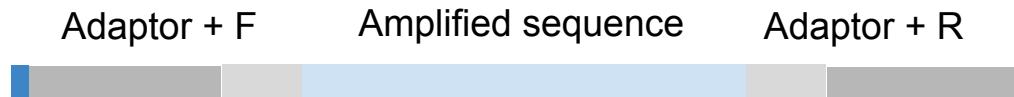
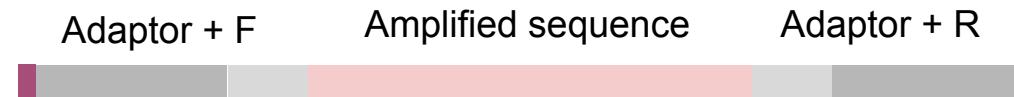


Adaptors added by PCR

**Dual indexing:** inter-sample chimeras detected

Intra-sample chimeras undetected

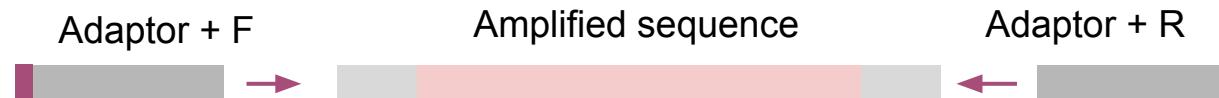
# Library preparation



## Fusion primers

Low amplification efficiency and risk of bias

# Library preparation



Adaptors added by ligation  
Loss of DNA templates

## 2 Metabarcoding



# Sequencing

## 2 Metabarcoding



# Quality check

## 2 Metabarcoding



# OTU clustering