

Contamination in DNA metabarcoding studies

Kristine Bohmann

Associate Professor
Environmental DNA Group
Section for Evolutionary Genomics
Globe Institute, Faculty of Health and Medical Sciences
University of Copenhagen



European Research Council
Established by the European Commission

Grant agreement No 856488

THE VELUX FOUNDATIONS
VILLUM FONDEN ✕ VELUX FONDEN



Please have your video on

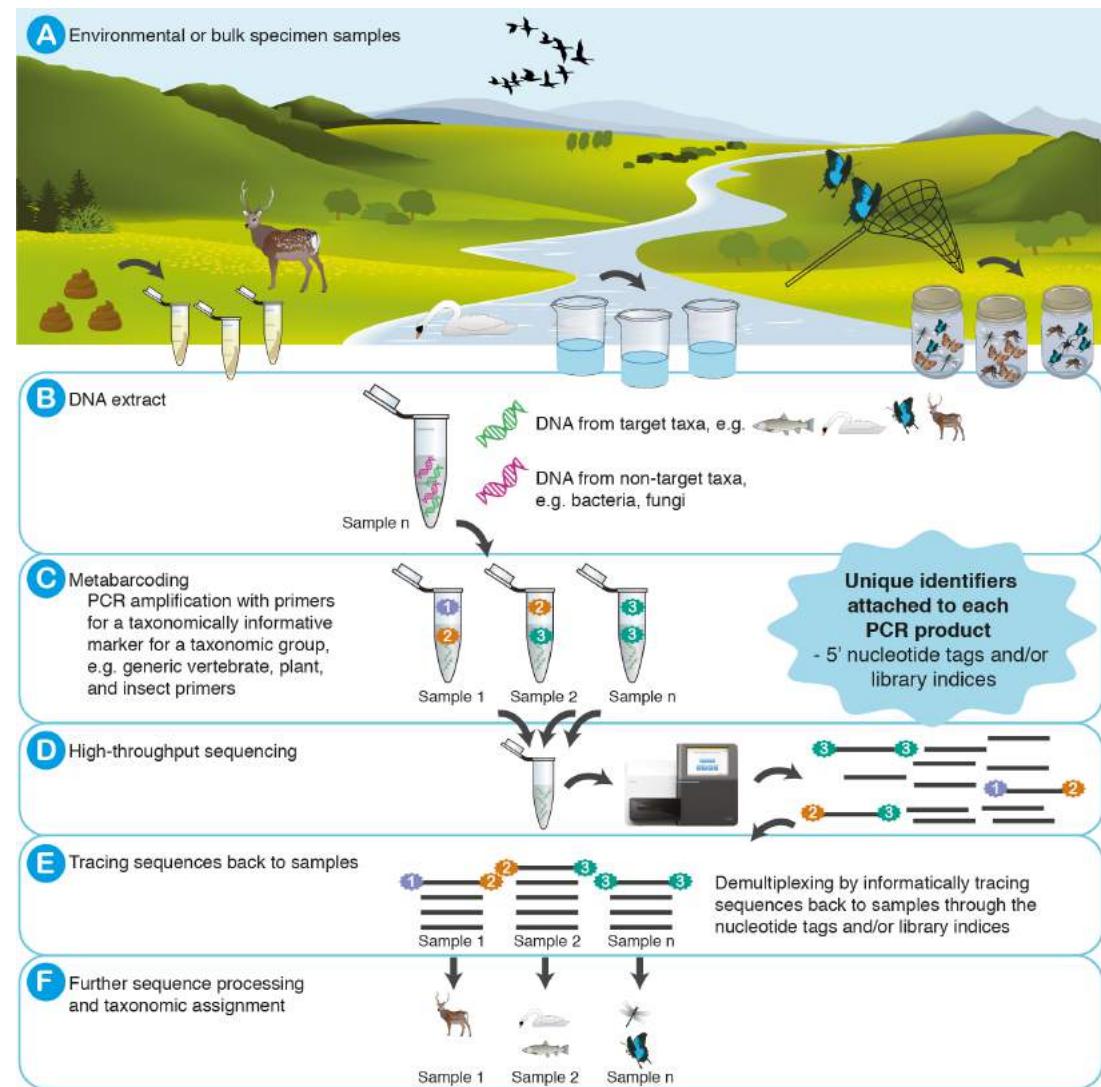


What scares you about contamination in your metabarcoding study?

Go to www.menti.com - I'll paste code in chat

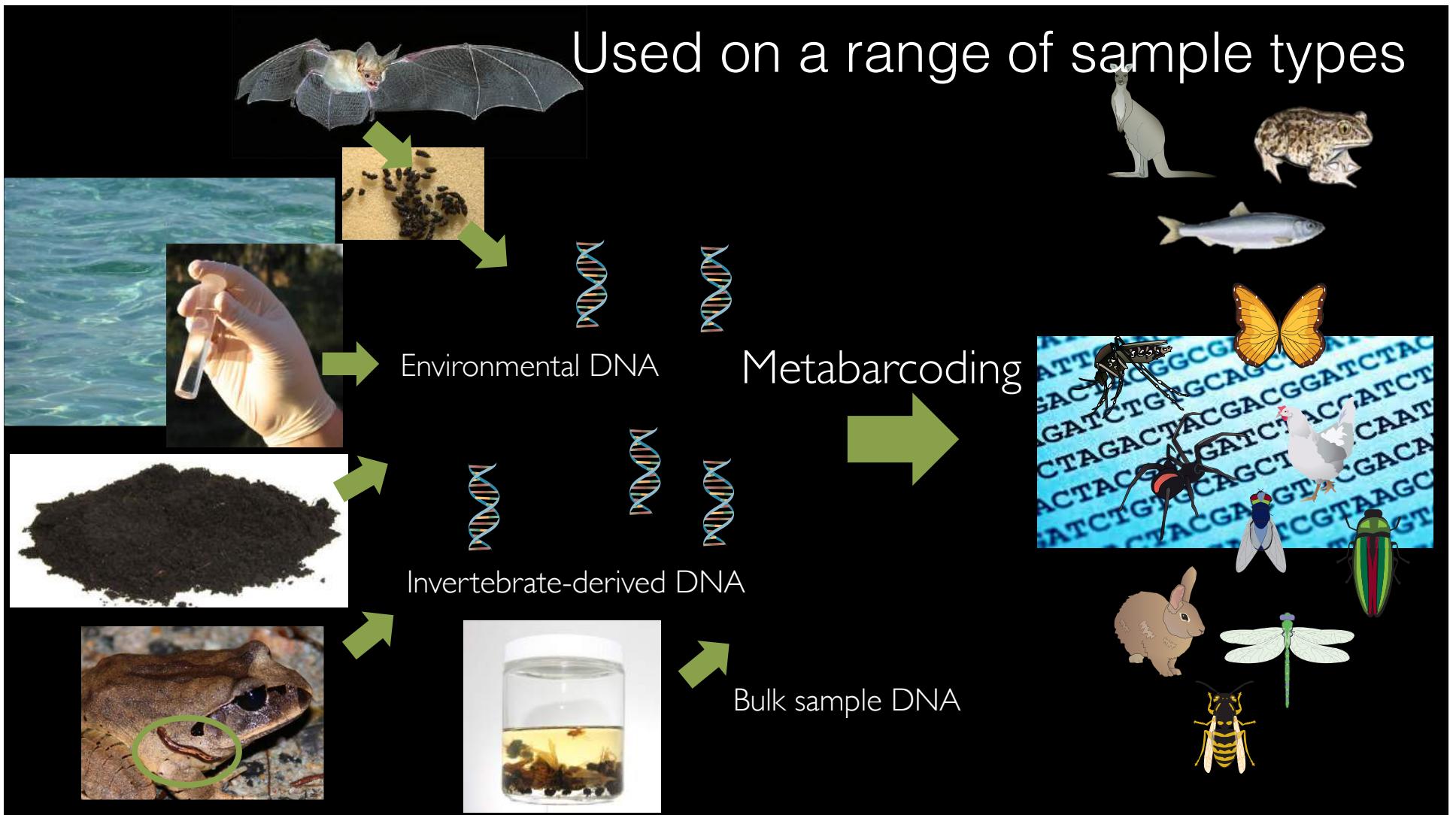
Please don't waste time
and money by cutting
corners in the lab set-up
– good investment to
invest a bit more in the
lab part

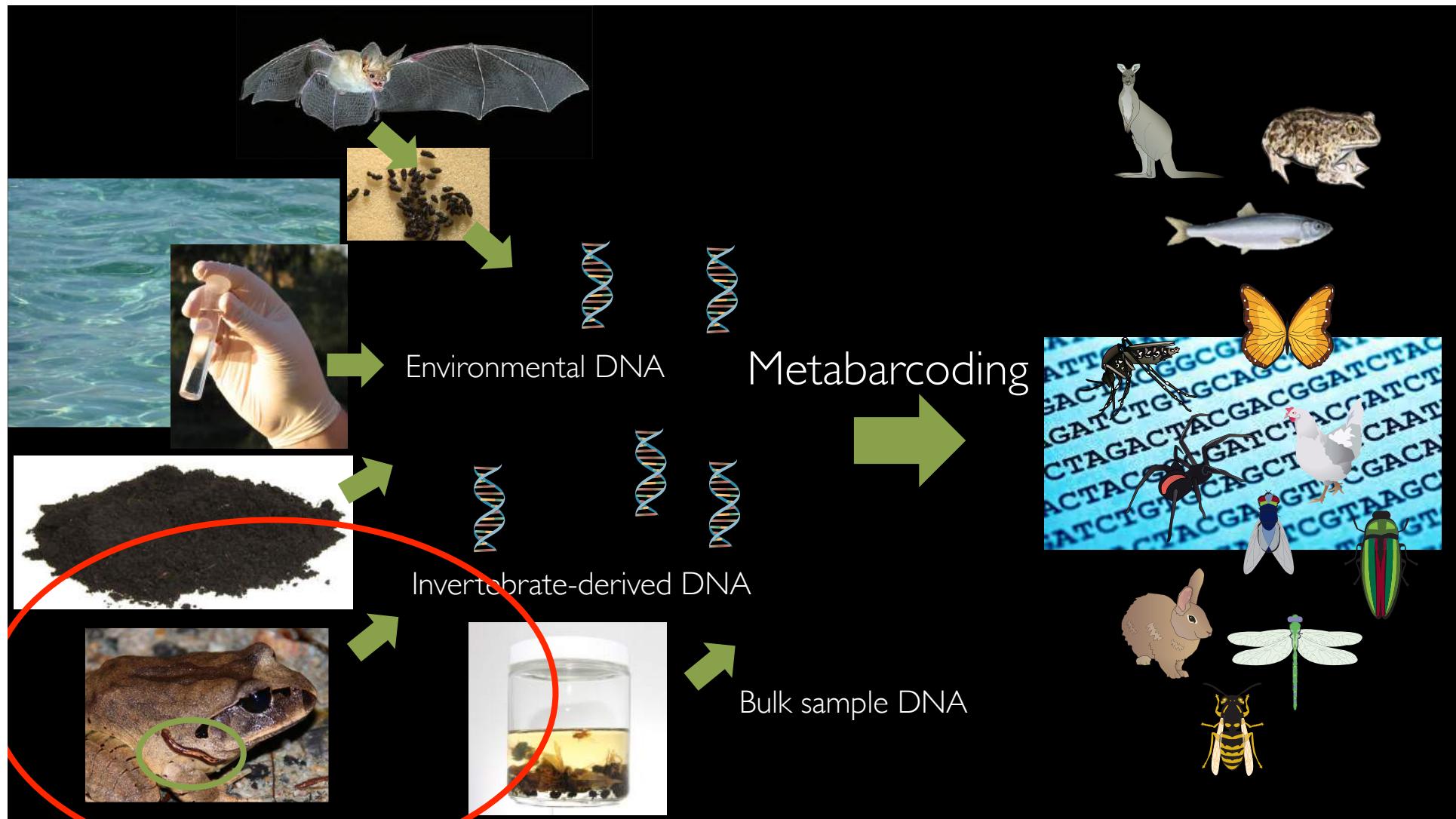
Outline of the metabarcoding workflow



From Bohmann, Elbrecht, Leese, Bunce, ... Creer, et al, submitted to Mol Ecol Res

Used on a range of sample types







Loss of biodiversity – to inform conservation efforts it's important with monitoring tools

All species shown here are protected by CITES

Monitoring can be difficult – solution: blood-feeding terrestrial leeches

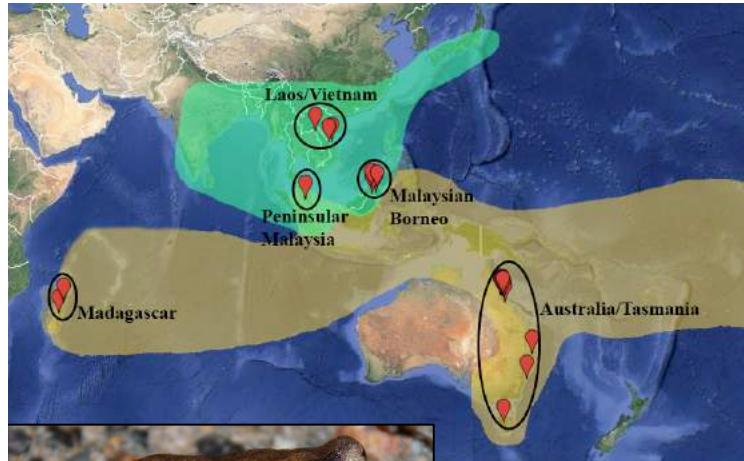




Blood-feeding leeches are found in rainforests with rich biodiversity and high numbers of endemic or threatened species

They find the collector





RESOURCE ARTICLE

WILEY MOLECULAR ECOLOGY RESOURCES

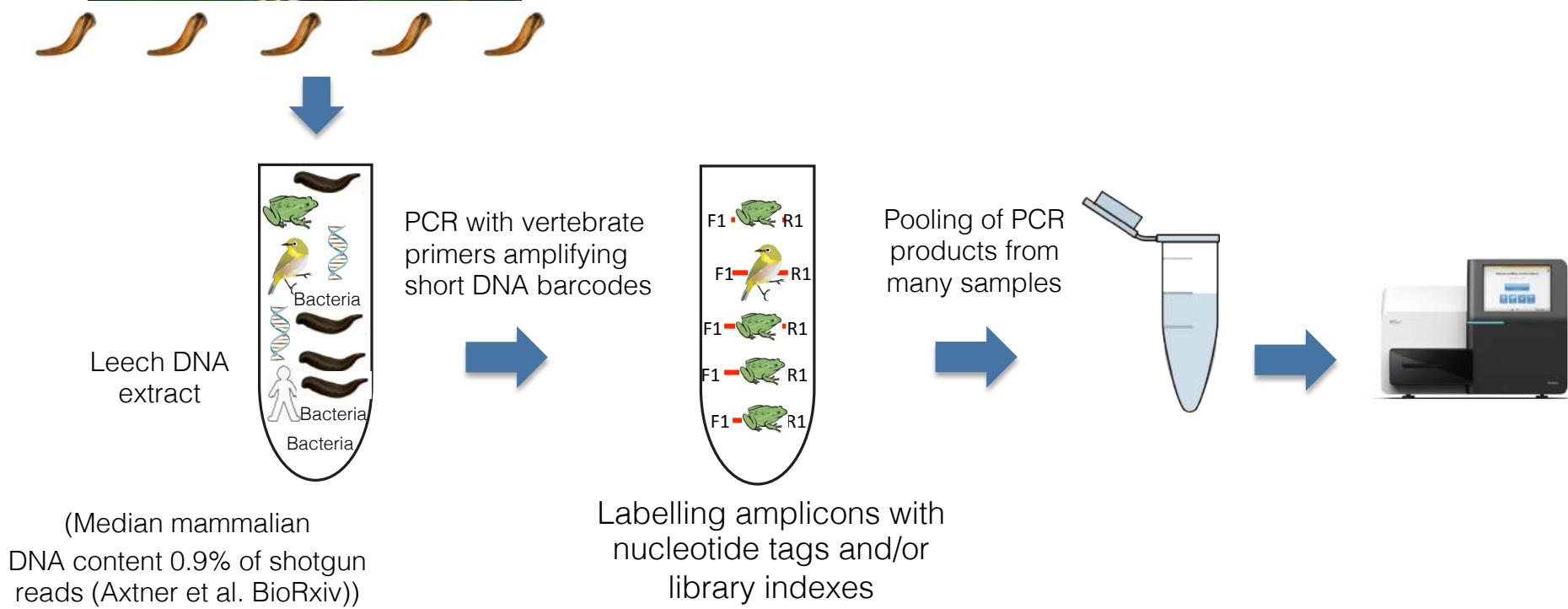
Debugging diversity – a pan-continental exploration of the potential of terrestrial blood-feeding leeches as a vertebrate monitoring tool

Ida Bærholm Schnell^{1,2} | Kristine Bohmann^{1,3} | Sebastian E. Schultze¹ | Stine R. Richter¹ | Dáithí C. Murray⁴ | Mikkel-Holger S. Sinding^{1,5} | David Bass^{6,7} | John E. Cadle⁸ | Mason J. Campbell⁹ | Rainer Dolch¹⁰ | David P. Edwards^{9,11} | Thomas N. E. Gray¹² | Teis Hansen¹ | Anh Nguyen Quang Hoa¹³ | Christina Lehmkühl Noer^{1,2} | Sigrid Heise-Pavlov¹⁴ | Adam F. Sander Pedersen¹⁵ | Juliet Carl Ramamonjisoa¹⁶ | Mark E. Siddall¹⁷ | Andrew Tilker^{18,19} | Carl Traeholt² | Nicholas Wilkinson²⁰ | Paul Woodcock²¹ | Douglas W. Yu^{3,22} | Mads Frost Bertelsen² | Michael Bunce⁴ | M. Thomas P. Gilbert^{1,4,23}



Metabarcoding

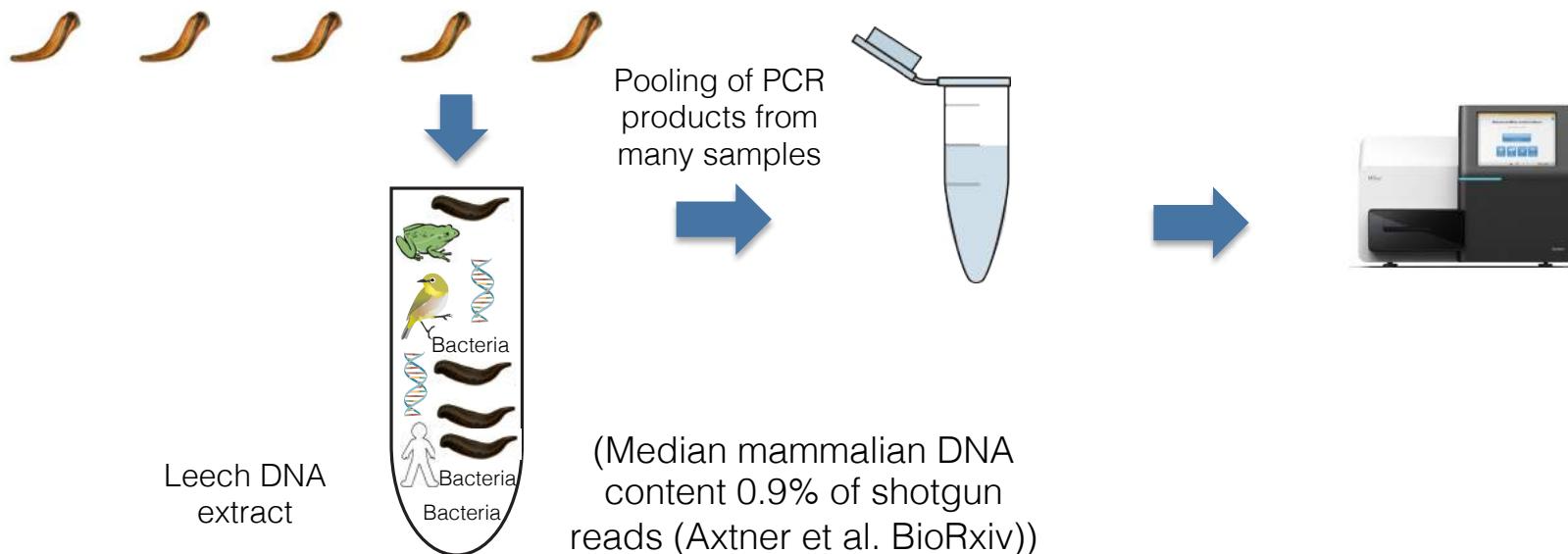
Targeted, sensitive approach with parallel sequencing on NGS platform



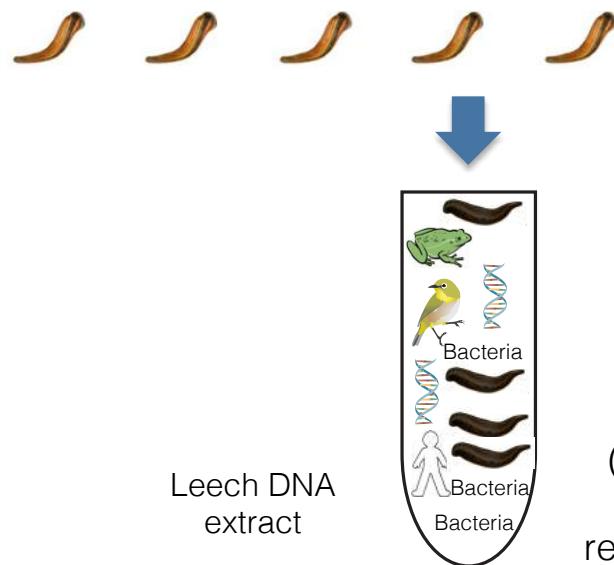
The good thing



Targeted, sensitive approach with parallel sequencing on NGS platform

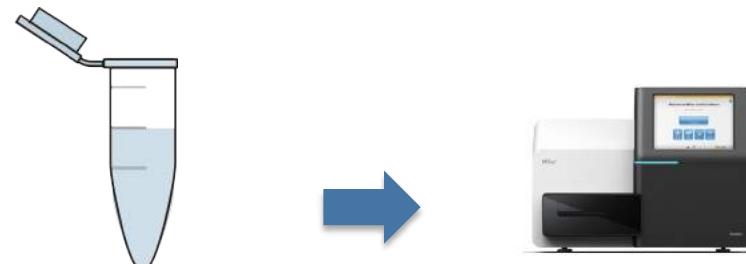


...are also where problems arise



(Median mammalian DNA
content 0.9% of shotgun
reads (Axtner et al. BioRxiv))

Targeted, sensitive approach with parallel sequencing on NGS platform



Today we will focus on
sensitivity and issues with
pooling

Contamination & issues with pooling
prior to sequencing

Intended learning outcomes

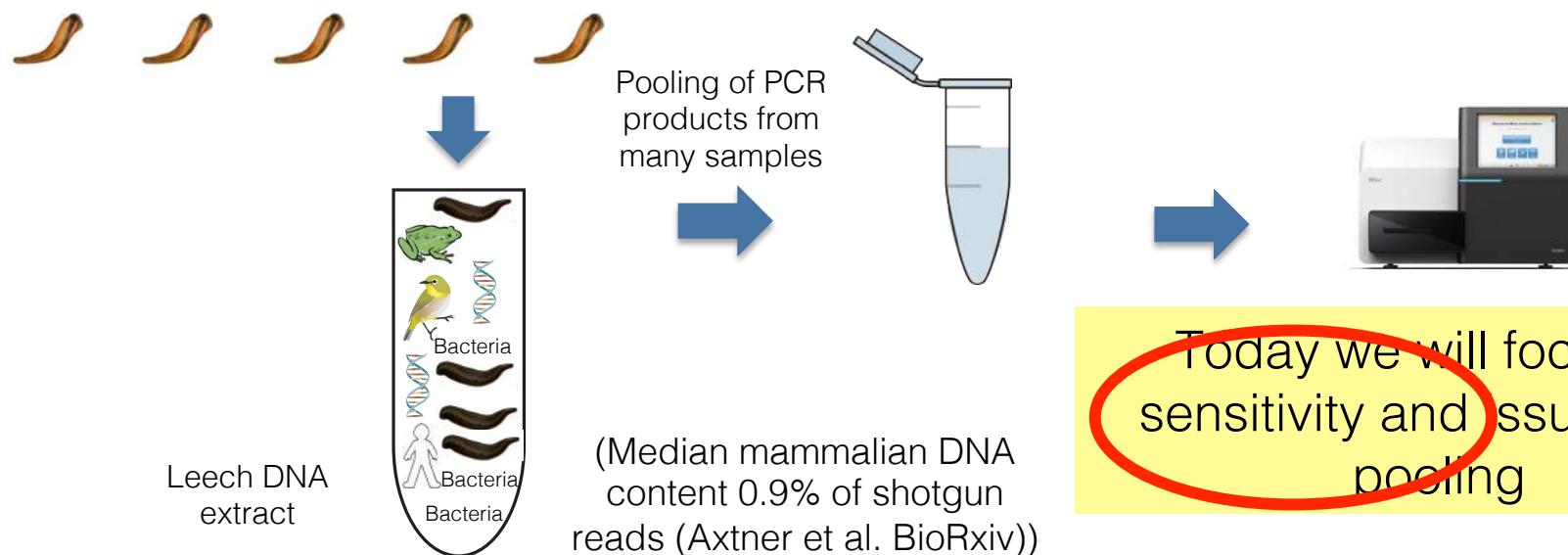
After this session you will be able to:

- Identify different sources of contamination – where it can arise, how to detect it and how to avoid it
- Outline the three main metabarcoding strategies for sample labelling
- Describe how PCR replicates can be used during data processing to balance error removal with detection of diversity
- Demonstrate an understanding of tag-jumps - how they can be accounted for and/or avoided and how they can be identified during data processing
- Formulate a strategy for the lab set-up of your metabarcoding study to account for or avoid different sources of contamination and facilitate the ability to balance error removal with diversity detection during data processing

...are also where problems arise



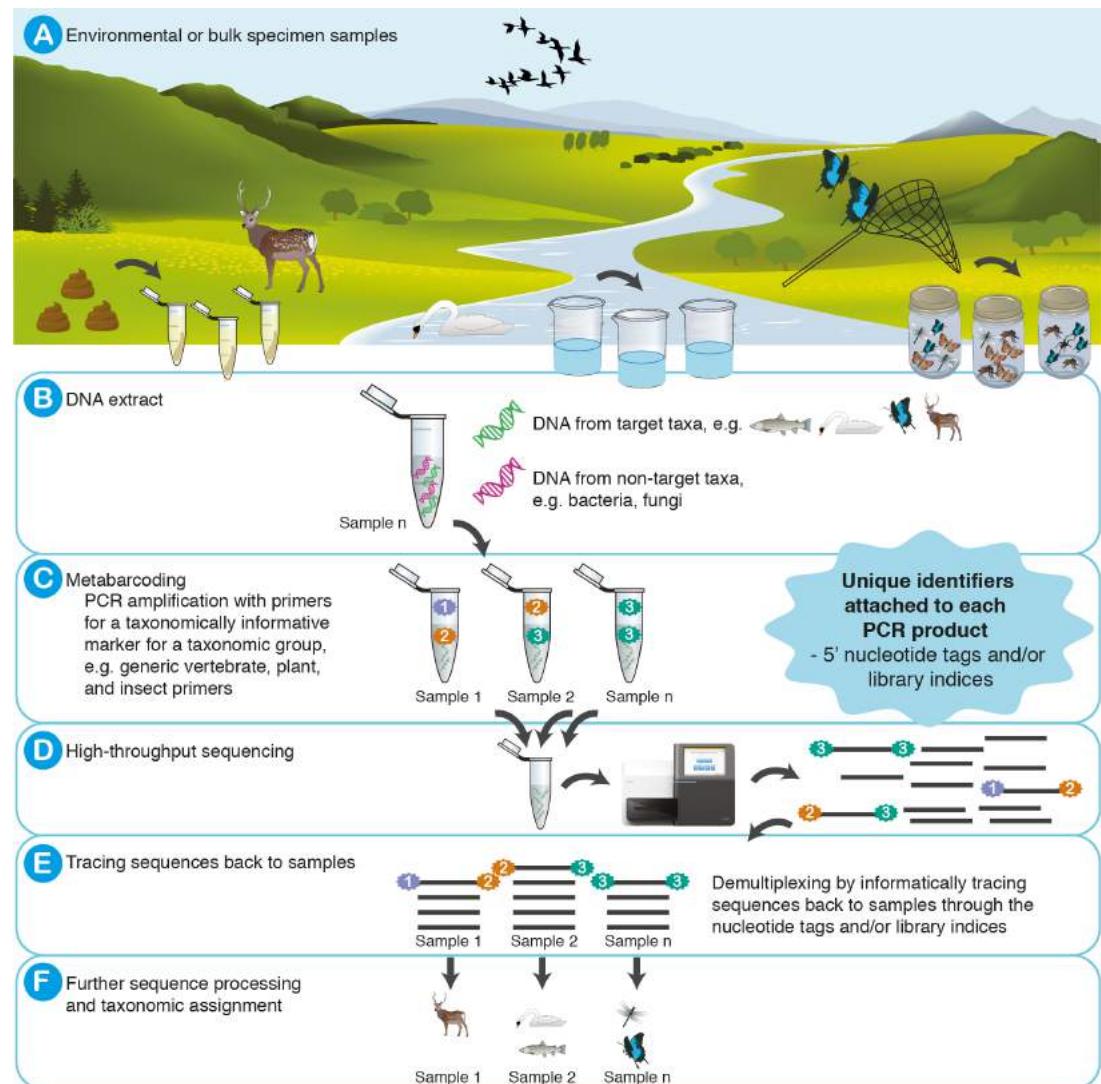
Targeted, sensitive approach with parallel sequencing on NGS platform



Today we will focus on sensitivity and issues with pooling

Where can contamination arise in metabarcoding studies?

Take a min to think and then please write in the chat

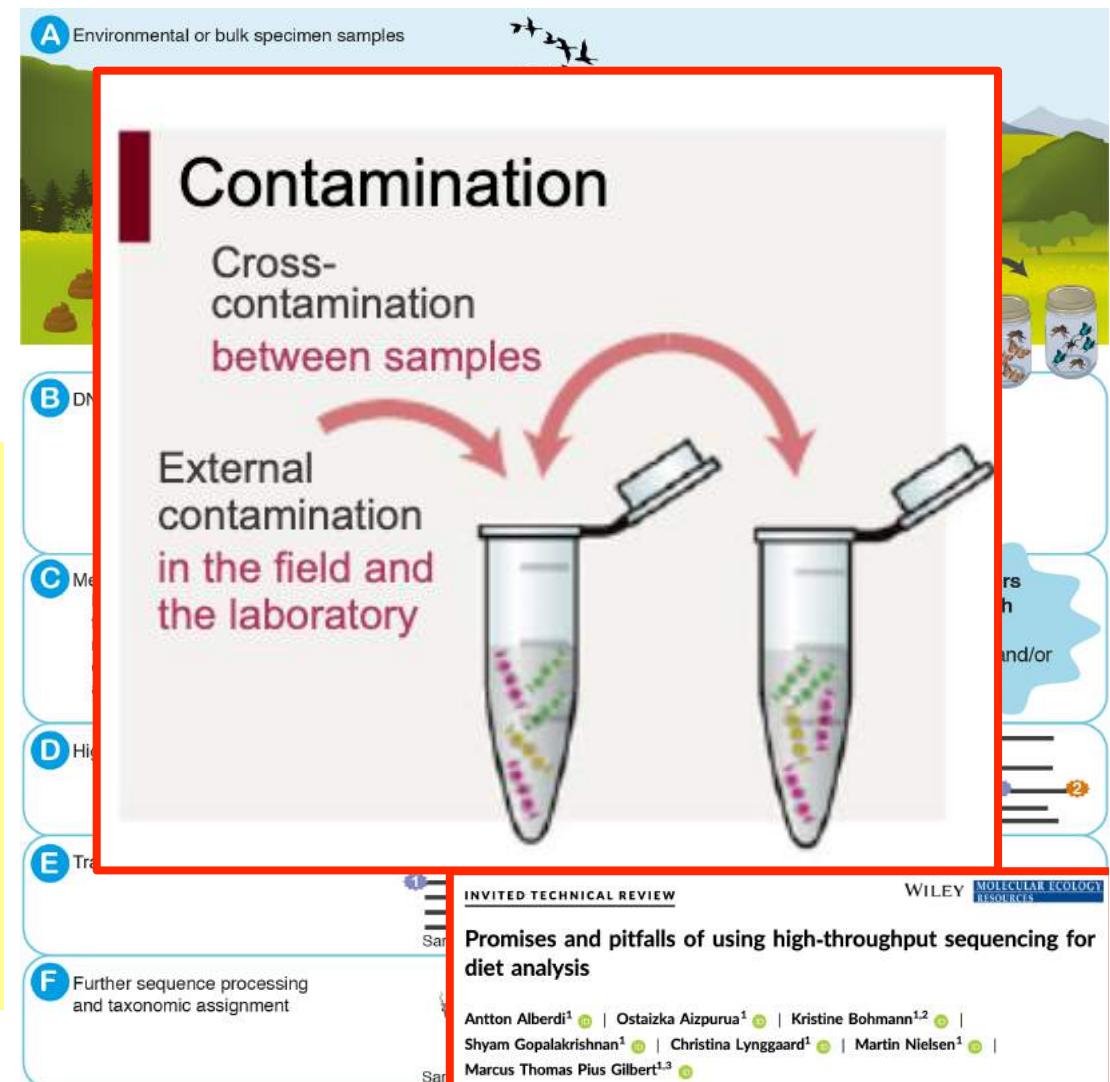


Where can contamination arise in metabarcoding studies?

From sample collection to sequencing

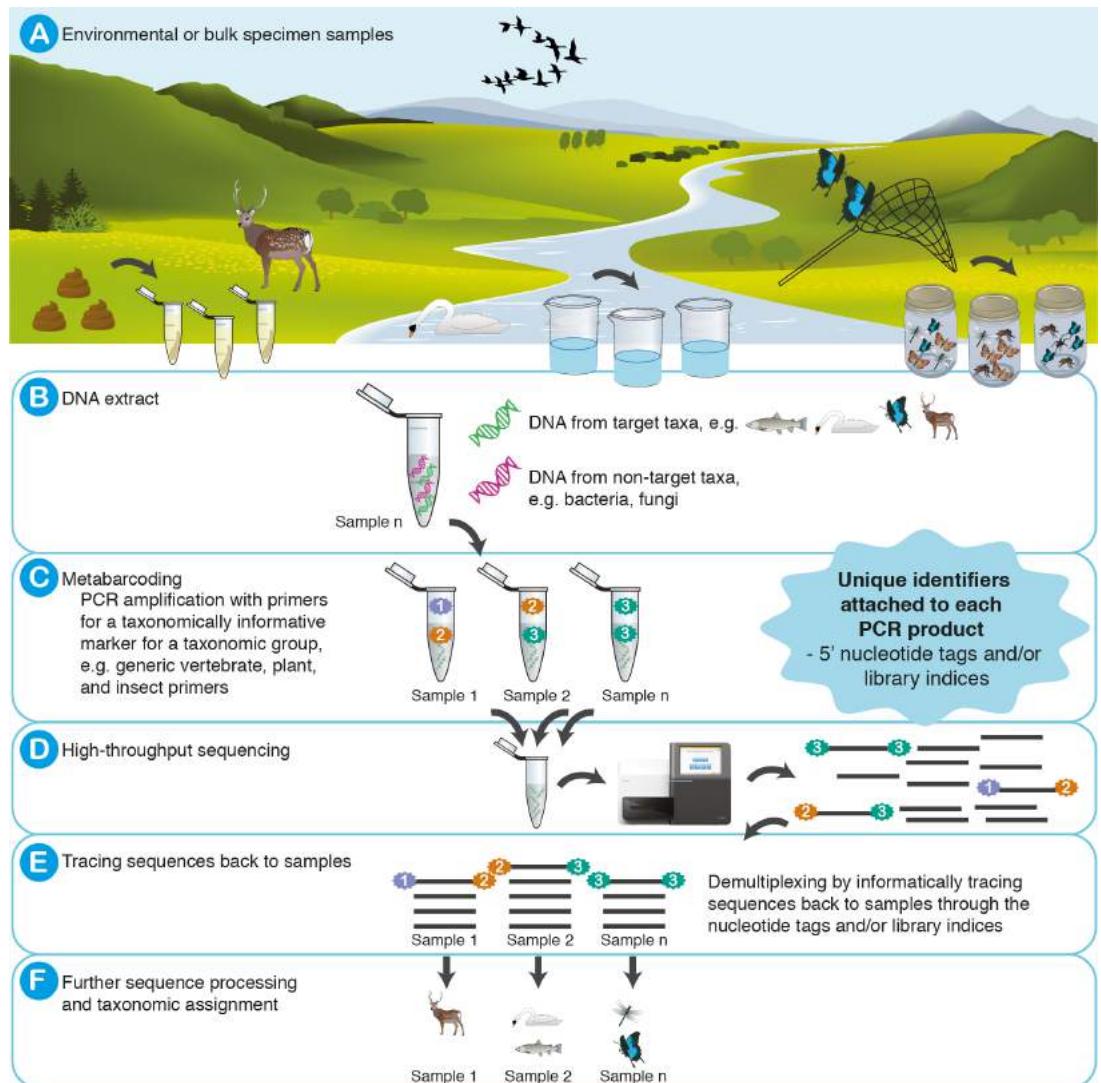
All steps where there is no label on sequences (cross-contamination between samples)

From the environment, reagents and tools



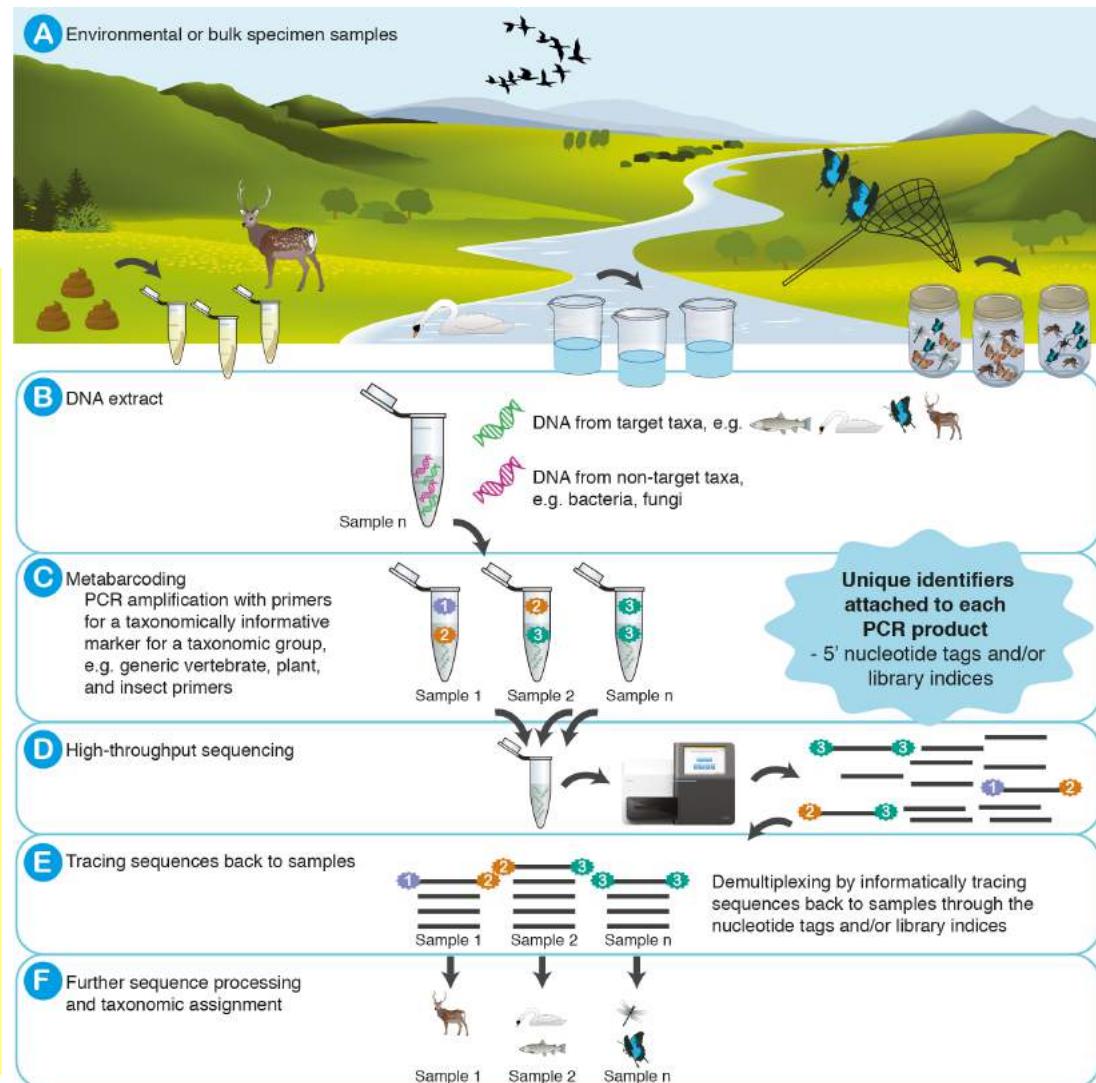
What can you do to **avoid** contamination?

Take a minute to think and then please write in the chat



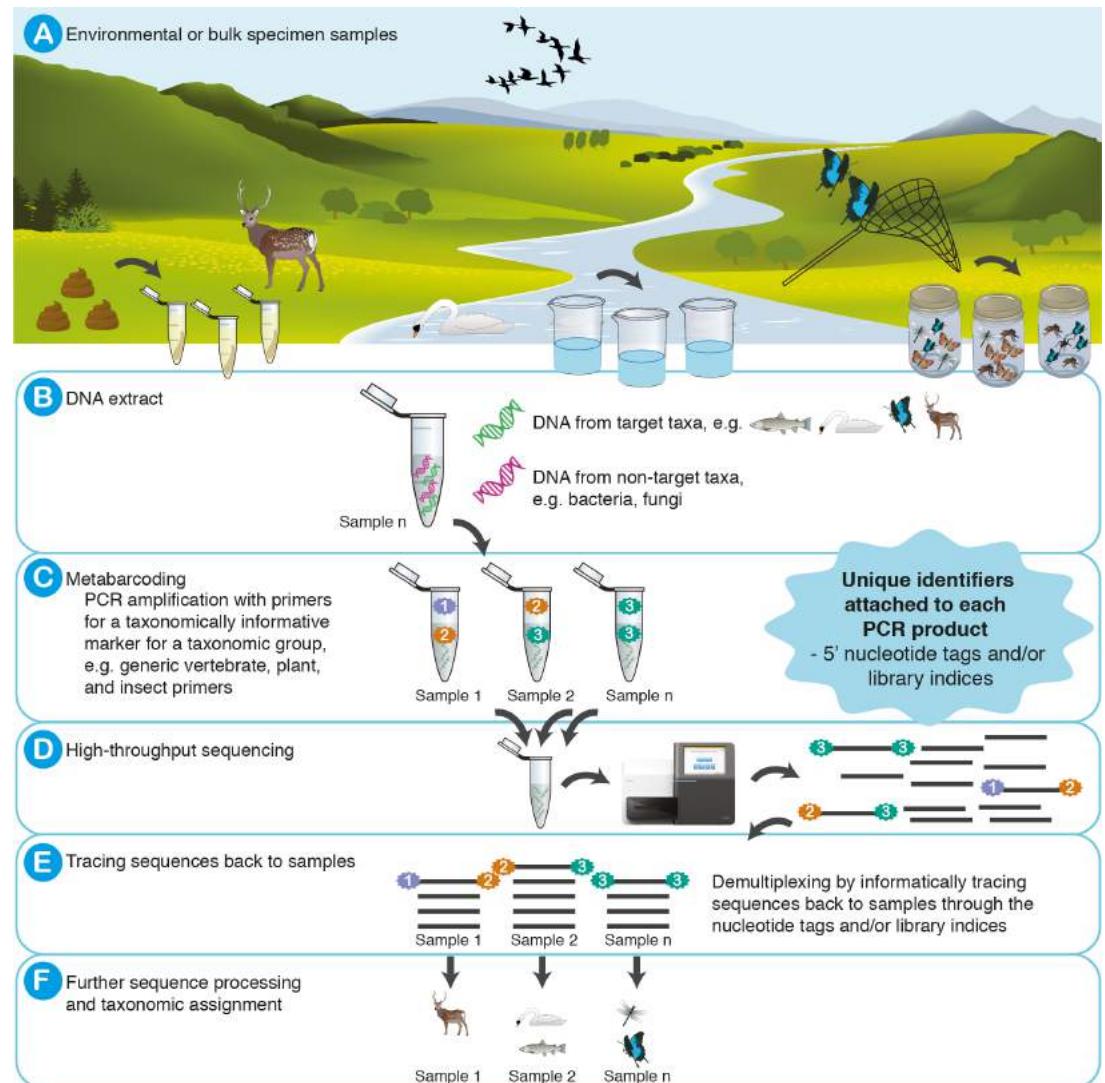
What can you do to **avoid** contamination?

- Careful handling
- Aliquoting extracts
- Use of sterile materials, filter tips, flow hoods, cleaning with hypochlorite
- Spatial division between pre-PCR and post-PCR procedures
- Tools and reagents used for processing the samples also need to be considered as putative source of contamination mostly those derived from living organisms, such as BSA (from cattle)
- Cross-contamination between samples: keep batches of processed samples low and reduce the spatial separation between tubes containing different samples and having multiple tubes open at the same time.



What can you do to enable
detection of
contamination?

Take a minute to think and
then please write in the
chat



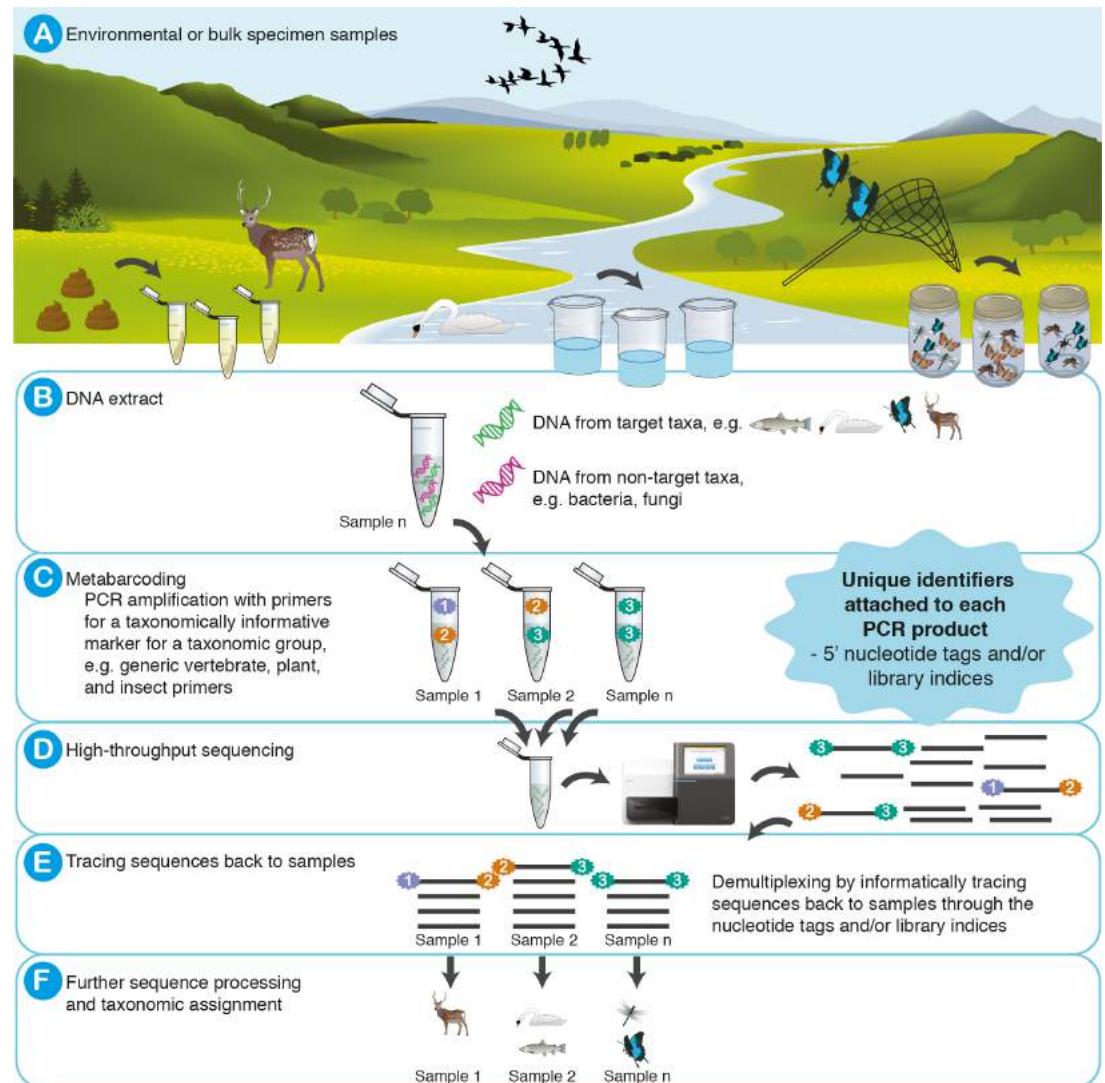
What can you do to enable **detection** of contamination?

Including (real!) negative controls at all steps of the lab workflow

Assessment prior to sequencing:
checking neg controls during quality control - e.g. qPCR, agarose gel, BioAnalyzer/TapeStation

Include positive controls in metabarcoding and sequencing

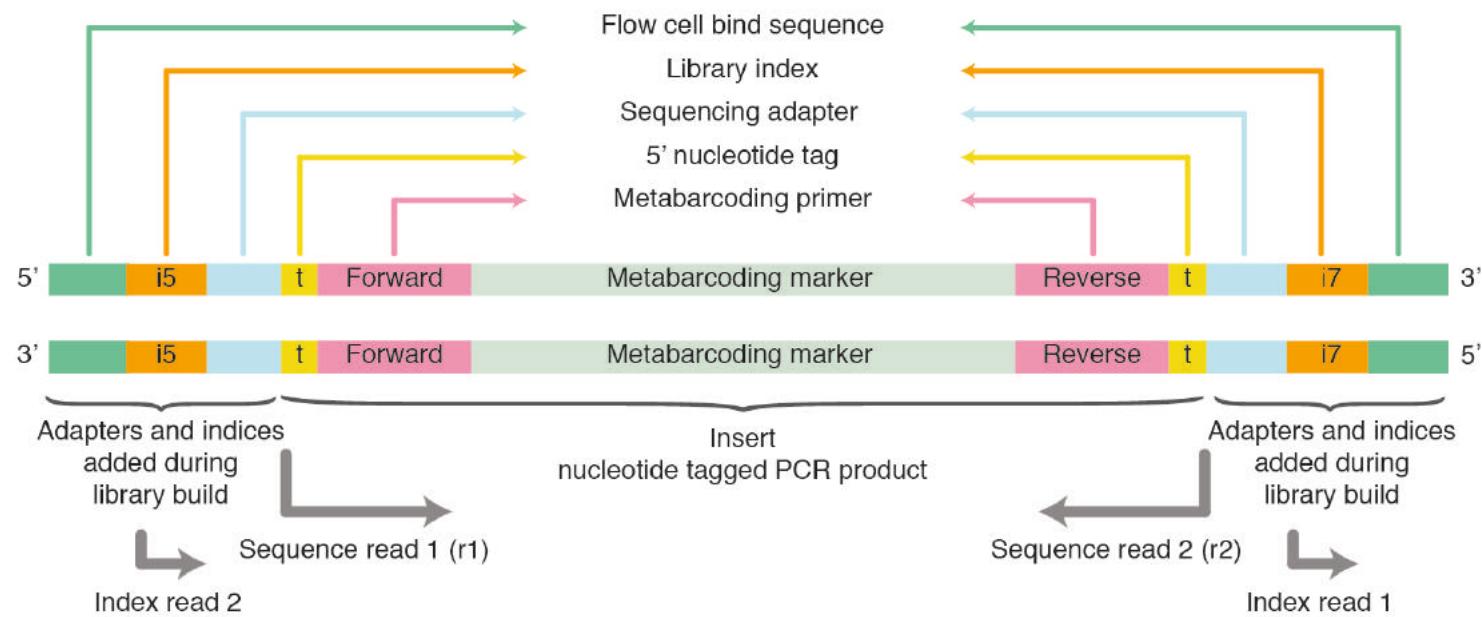
Assessment of contamination after sequencing – use both positive and negative controls



Three main approaches to label
amplicons in metabarcoding studies on
Illumina sequencing platforms

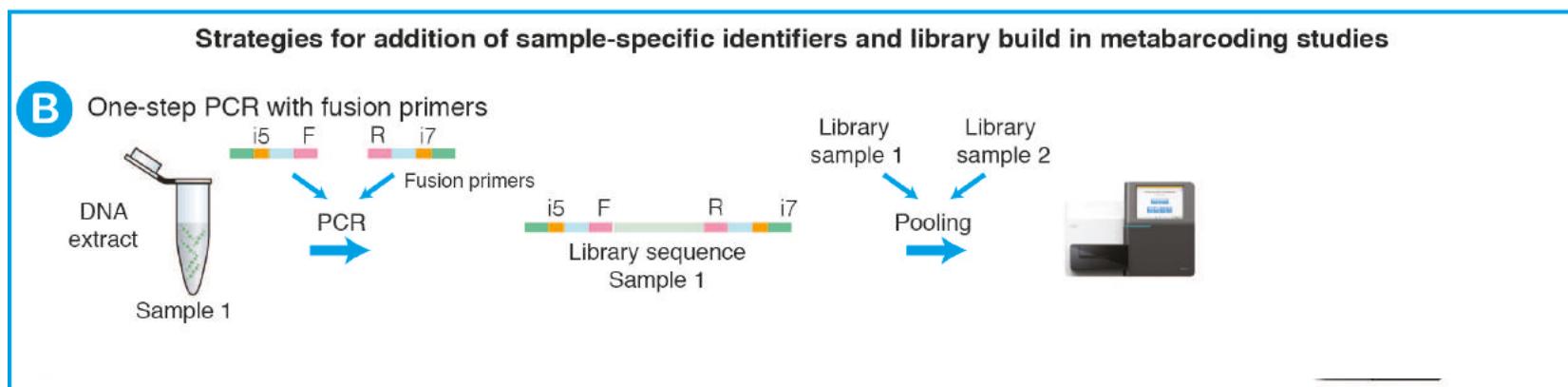
A

Structure of dual-tagged and dual-indexed Illumina metabarcoding library sequence



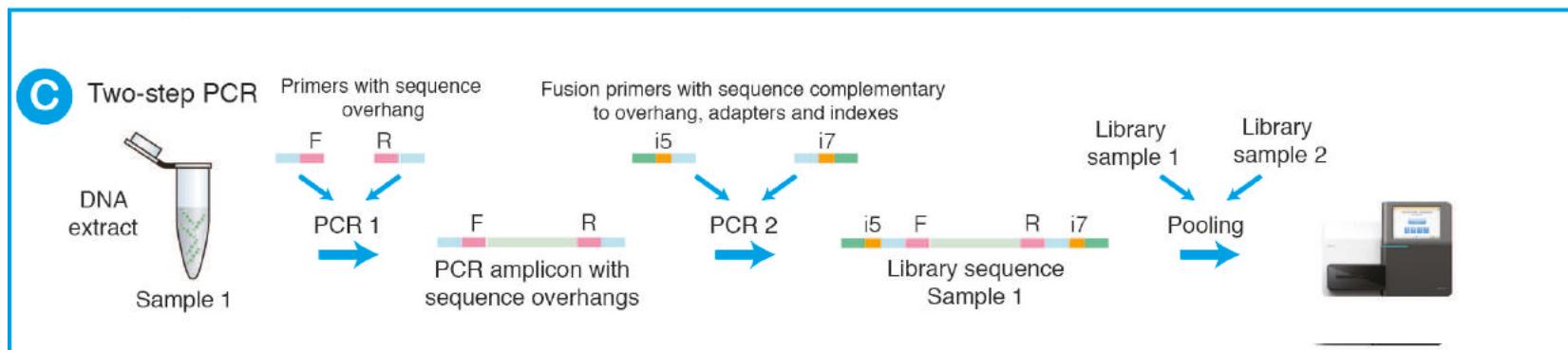
Cut-out of figure from Bohmann, Elbrecht, Leese, Bunce, ... Creer, et al, submitted

Three main approaches to label amplicons in metabarcoding studies on Illumina sequencing platforms
Which one is yours?



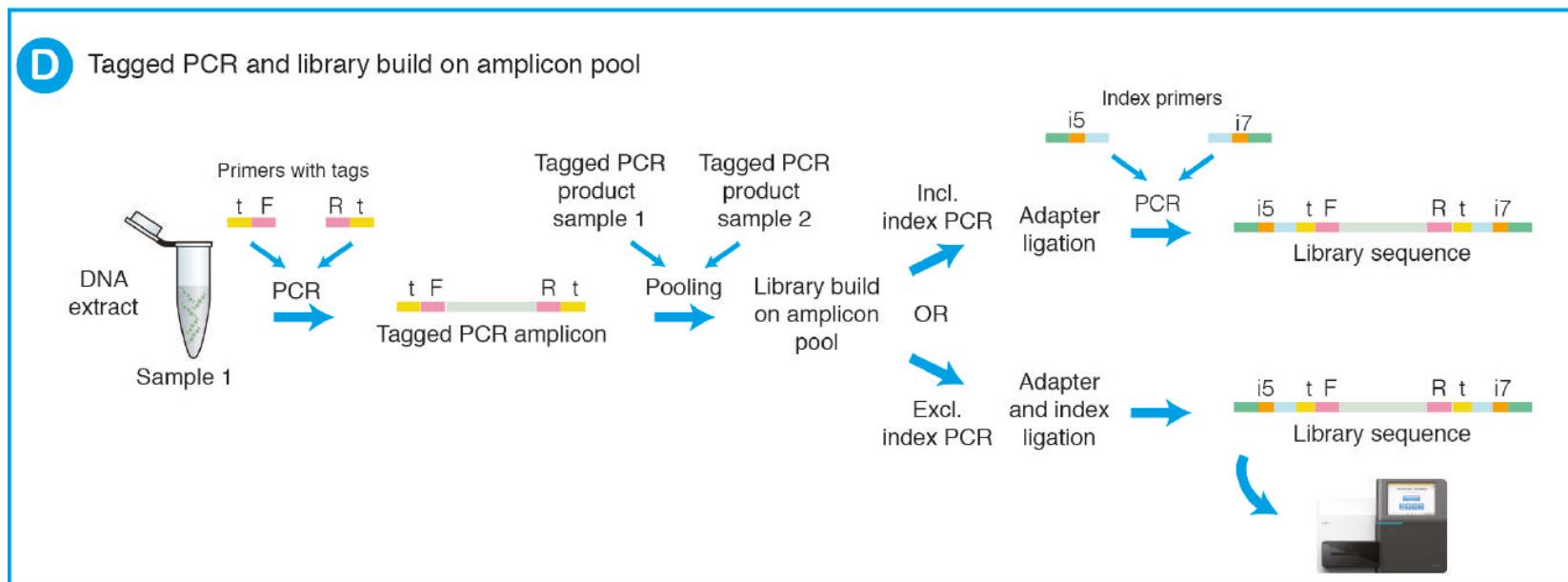
Cut-out of figure from Bohmann, Elbrecht, Leese, Bunce, ... Creer, et al, submitted

Three main approaches to label amplicons in metabarcoding studies on Illumina sequencing platforms
Which one is yours?



Cut-out of figure from Bohmann, Elbrecht, Leese, Bunce, ... Creer, et al, submitted

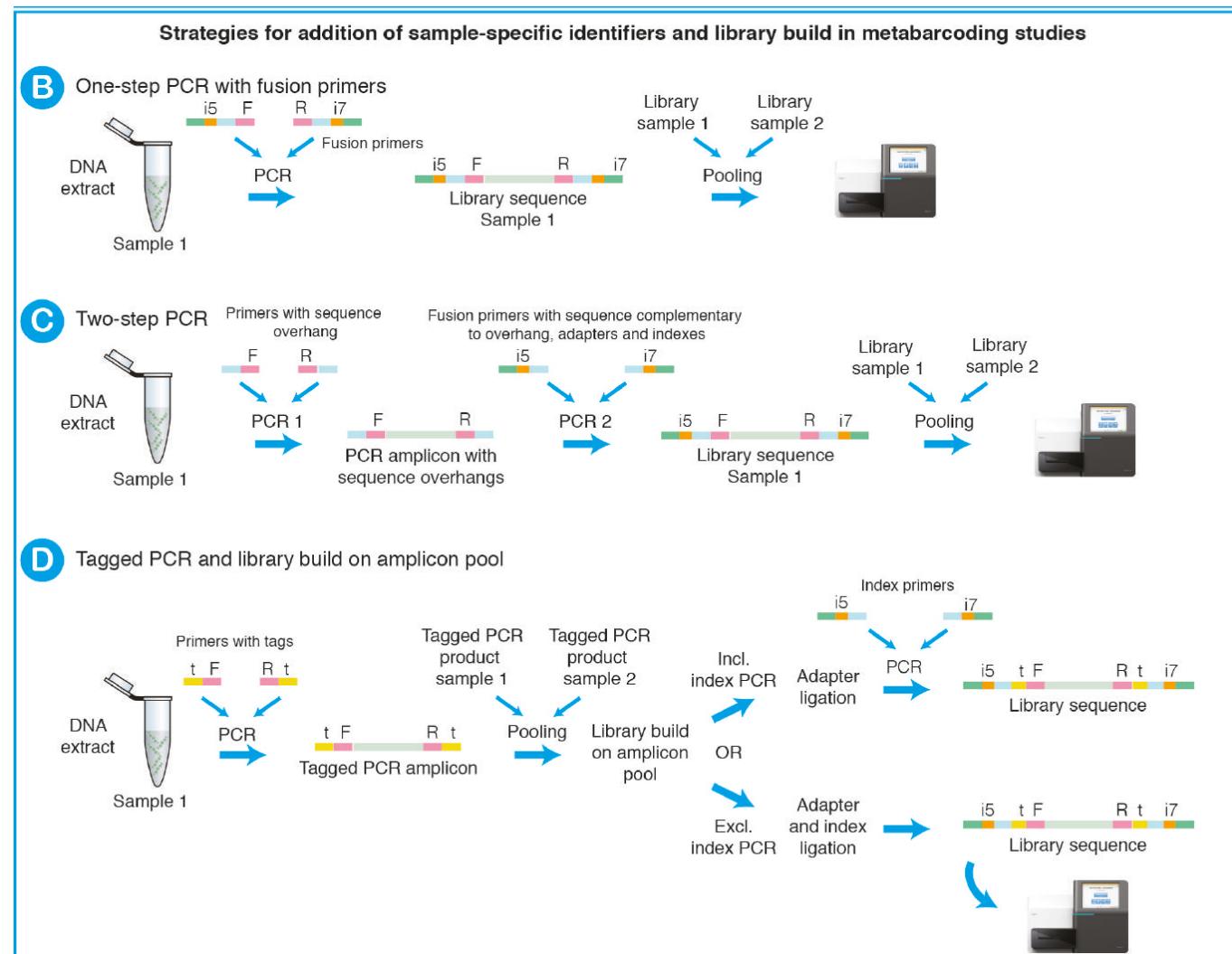
Three main approaches to label amplicons in metabarcoding studies on Illumina sequencing platforms
Which one is yours?



Cut-out of figure from Bohmann, Elbrecht, Leese, Bunce, ... Creer, et al, submitted

Which metabarcoding strategy do you use/will you use?

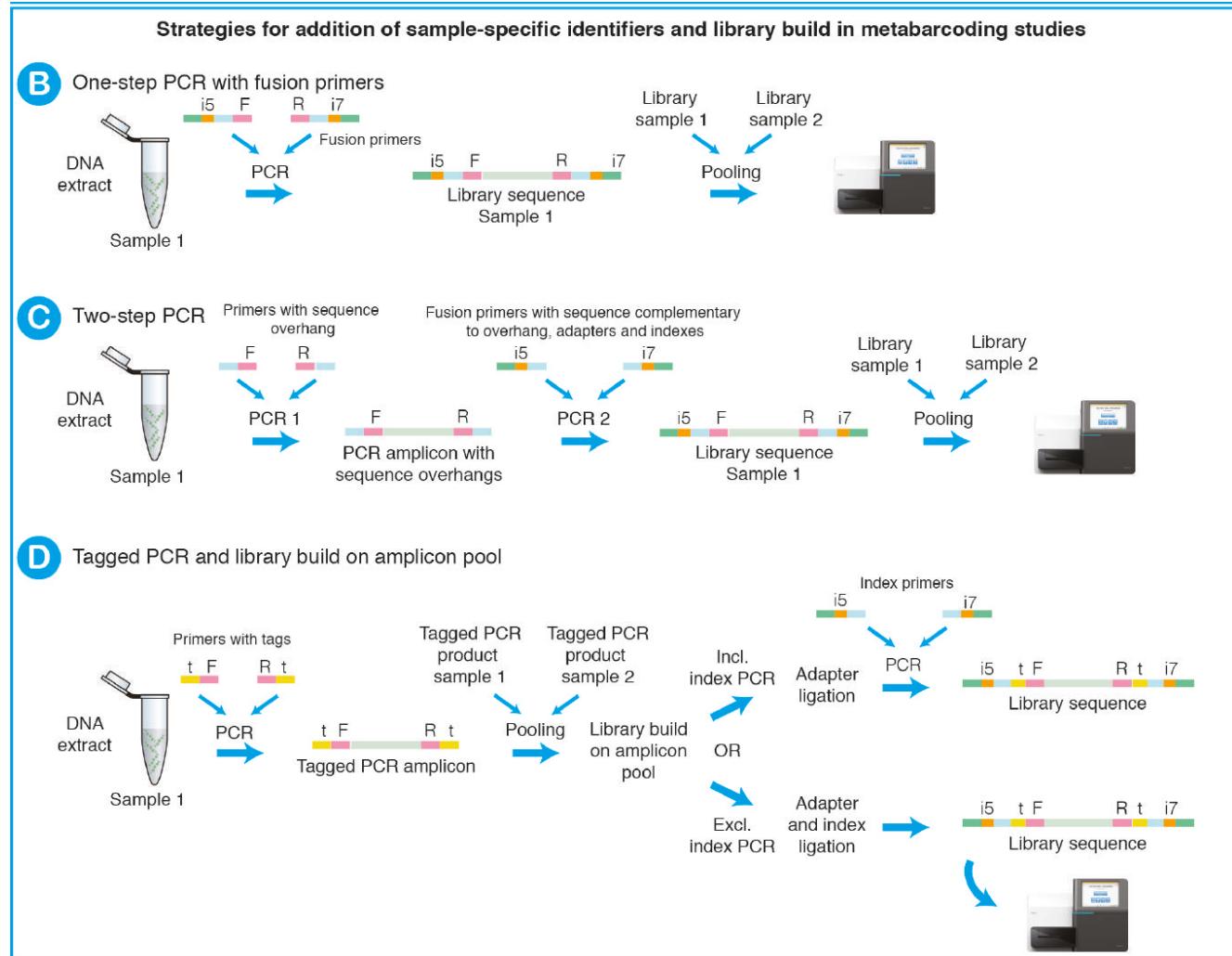
Take 2 mins to think about it and go to www.menti.com
- I'll add code in chat



Are there differences in contamination risk between the metabarcoding approaches?

If so, can you do something about it?

Take 2 mins to think about it



5 min break!

Use of PCR replicates to balance error removal with detection of diversity

Apologies

Methods in Ecology and Evolution

Research Article

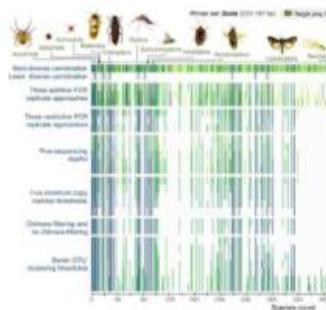
Scrutinizing key steps for reliable metabarcoding of environmental samples

Antton Alberdi , Ostaizka Aizpurua, M Thomas P Gilbert, Kristine Bohmann



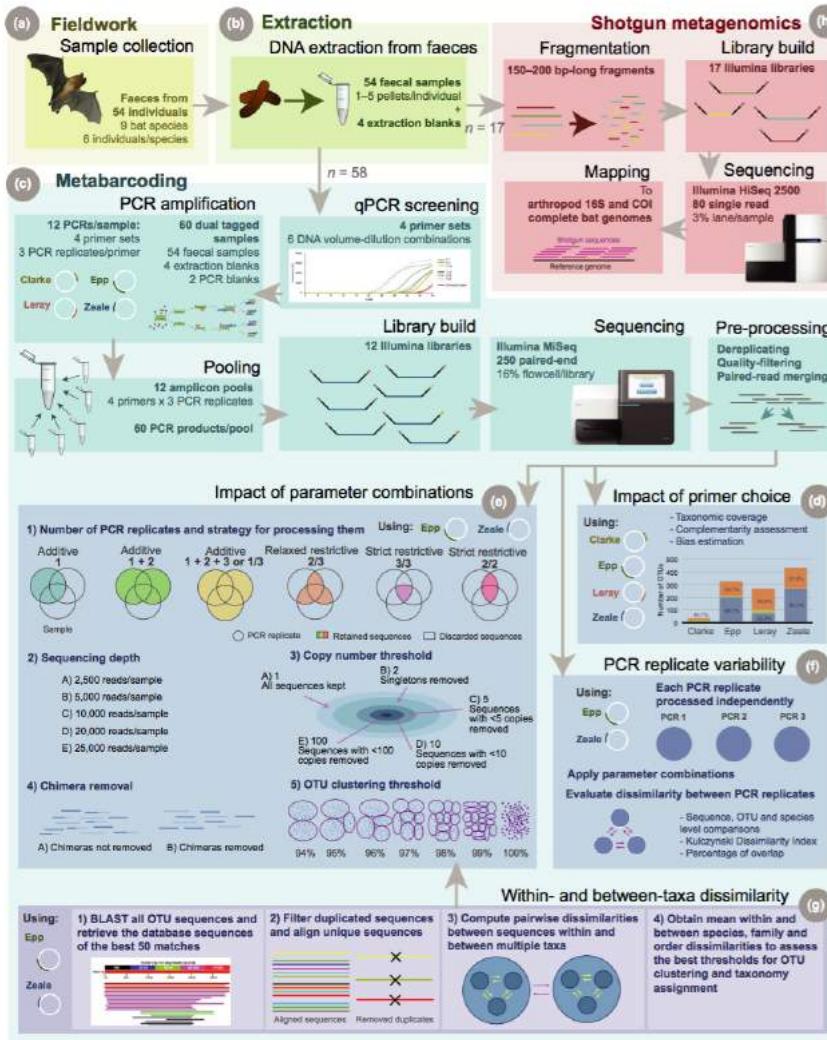
Taylor Wilcox @taylormwilcox · 22m

Useful technical paper. I'd recommend a cup of coffee and a pad of paper for sketching out the study design...



MethodsEcolEvol @MethodsEcolEvol

New study explores how the decisions that must be made during a #Metabarcoding workflow affect the final results
bit.ly/2exY0N4



Methods in Ecology and Evolution

Research Article

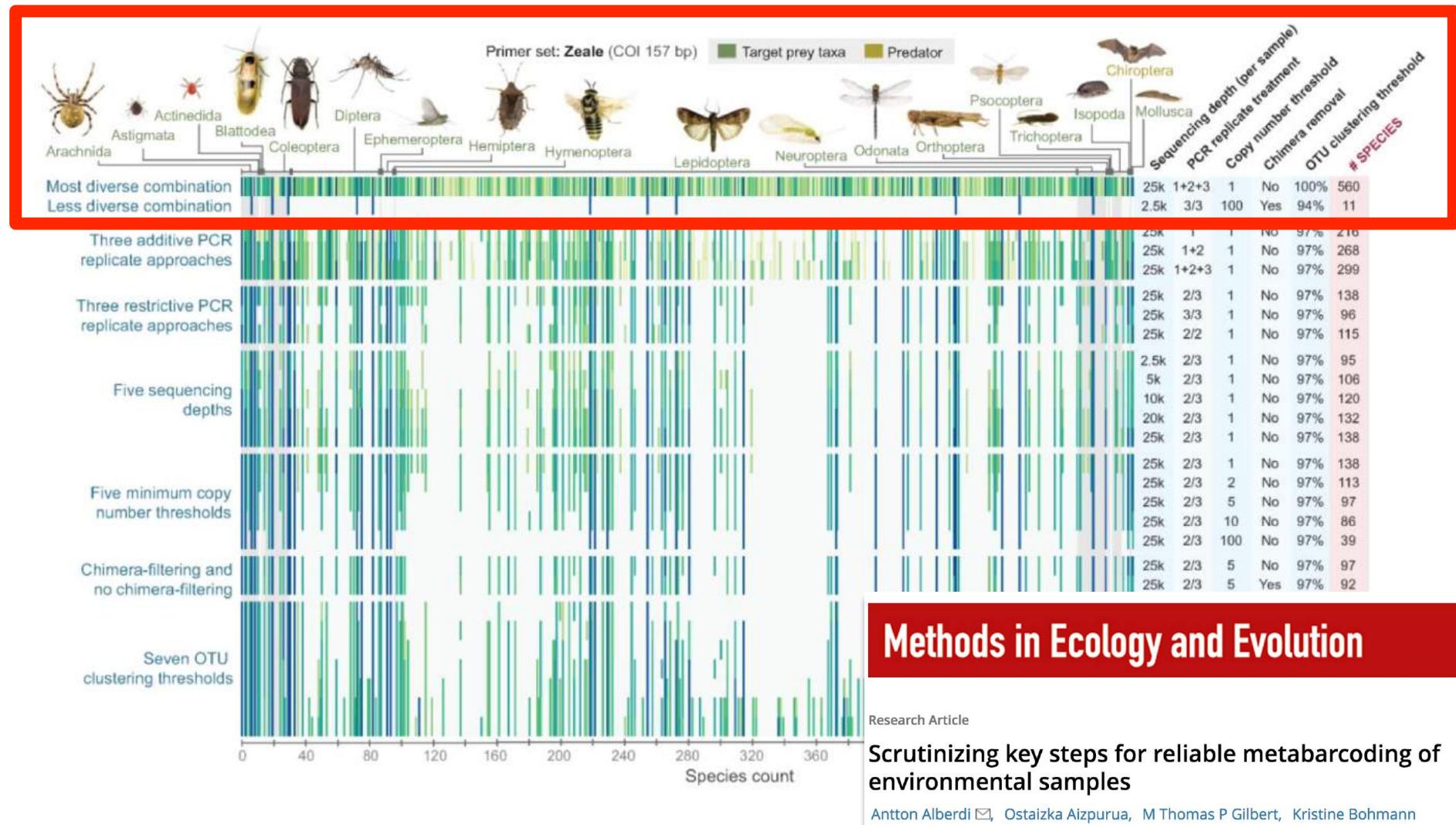
Scrutinizing key steps for reliable metabarcoding of environmental samples

Antton Alberdi , Ostaizka Aizpurua, M Thomas P Gilbert, Kristine Bohmann



“I think we should try submitting it by May..... otherwise I will enter into a depression state, this paper is consuming me from the inside....”

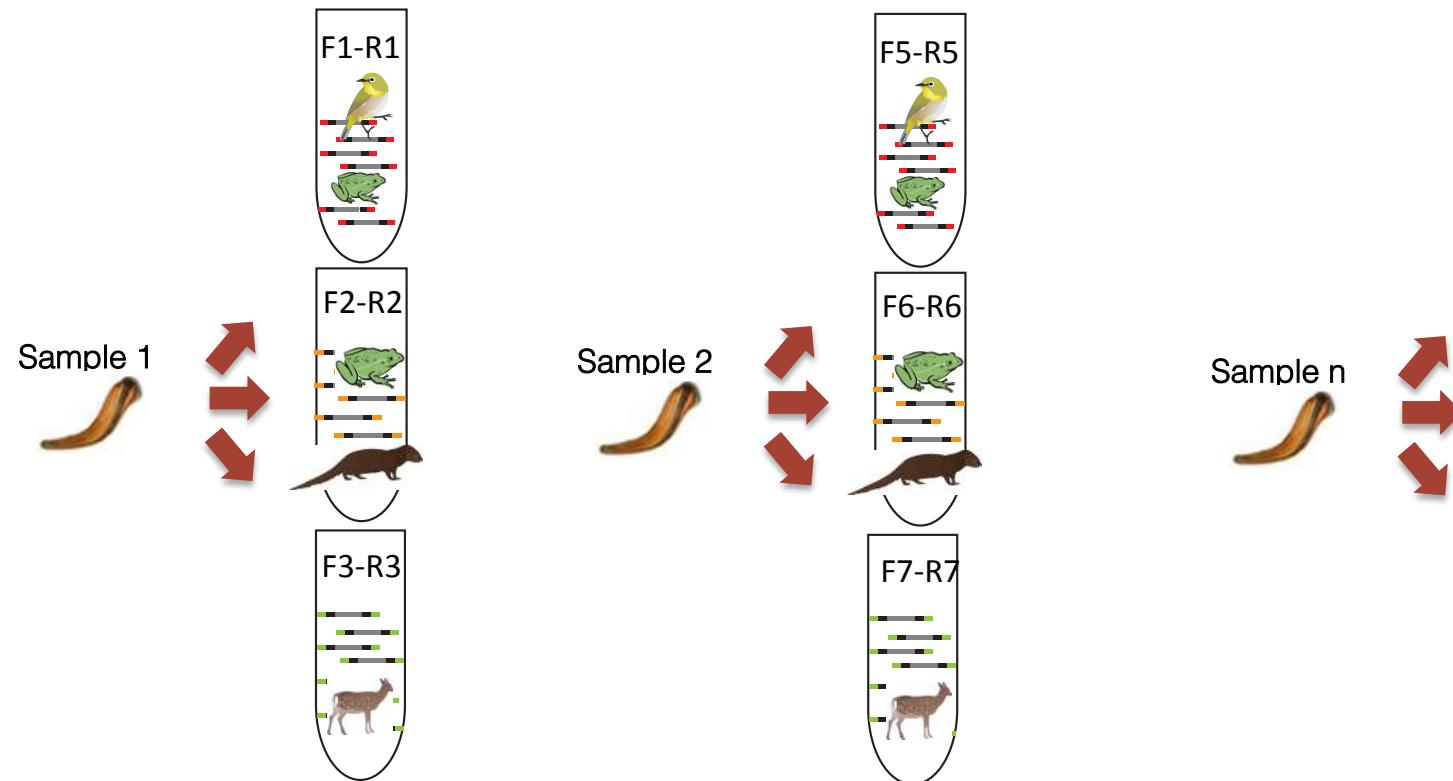
Something good came out of it – a scary figure



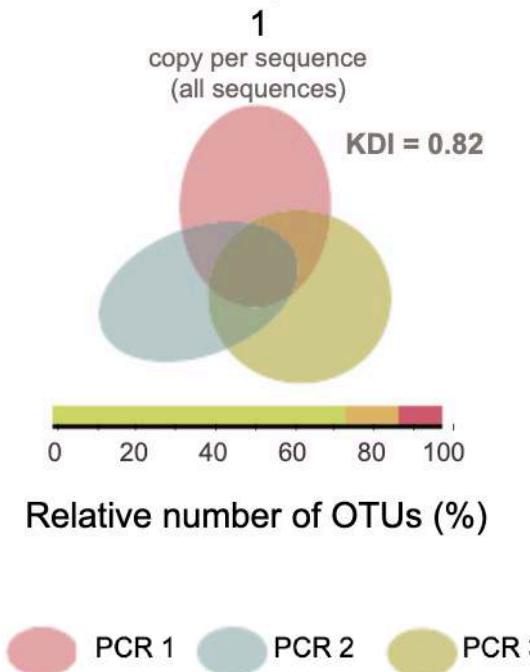
Positive and negative controls and PCR replicates can aid in balancing false positives and negatives and greatly aid your flexibility during data processing

Use of PCR replicates to balance error removal with detection of diversity

Example: 3 twin-tagged PCR replicates
– differently tagged



(b) OTU overlaps between PCR replicates



PCR with arthropod
primers amplifying
short DNA barcodes

% of OTUs present in:
1 PCR replicate 2 PCR replicates 3 PCR replicates

Methods in Ecology and Evolution

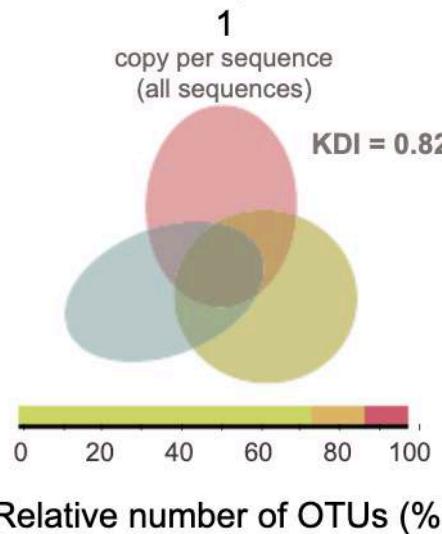
Research Article

Scrutinizing key steps for reliable metabarcoding of environmental samples

Antton Alberdi Ostaizka Aizpurua, M Thomas P Gilbert, Kristine Bohmann

What can cause PCR replicates to not contain the same OTUs?
Discuss 5 mins in breakout rooms and enter in menti.com – code in chat

(b) OTU overlaps between PCR replicates



PCR 1 PCR 2 PCR 3

% of OTUs present in:
1 PCR replicate 2 PCR replicates 3 PCR replicates



PCR with arthropod primers amplifying short DNA barcodes

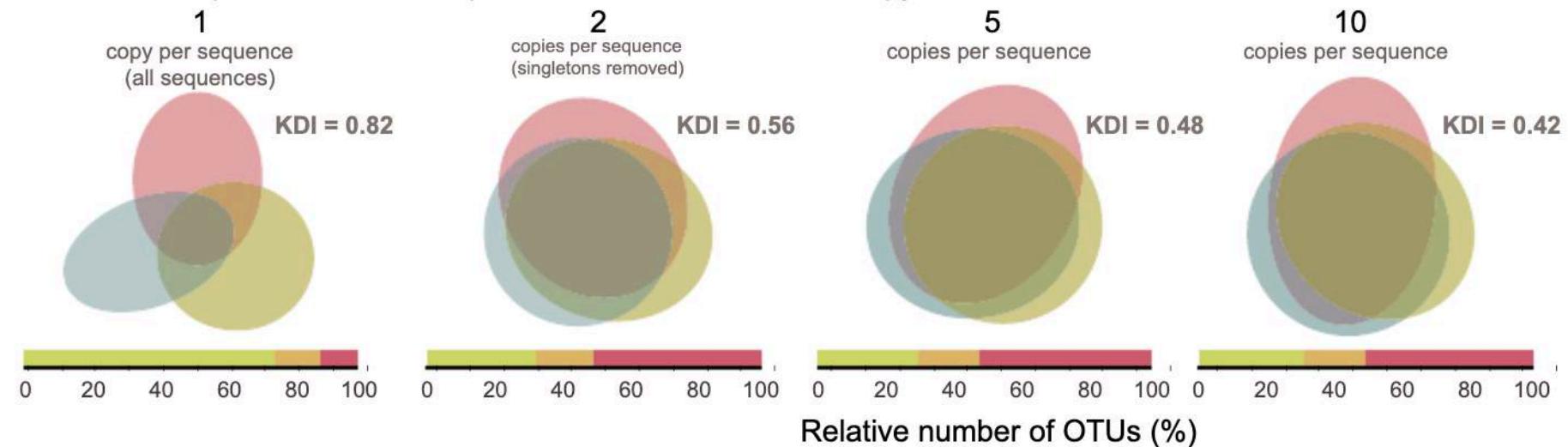
Methods in Ecology and Evolution

Research Article

Scrutinizing key steps for reliable metabarcoding of environmental samples

Antton Alberdi, Ostaizka Aizpurua, M Thomas P Gilbert, Kristine Bohmann

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



PCR 1 PCR 2 PCR 3

% of OTUs present in:
1 PCR replicate 2 PCR replicates 3 PCR replicates

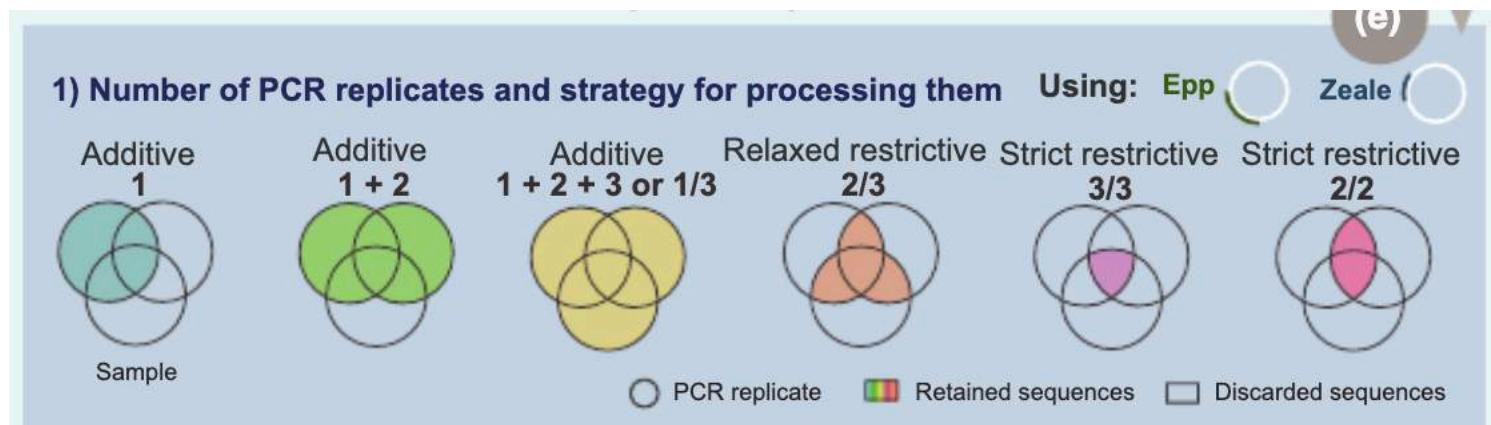
Methods in Ecology and Evolution

Research Article

Scrutinizing key steps for reliable metabarcoding of environmental samples

Antton Alberdi Ostaizka Aizpurua, M Thomas P Gilbert, Kristine Bohmann

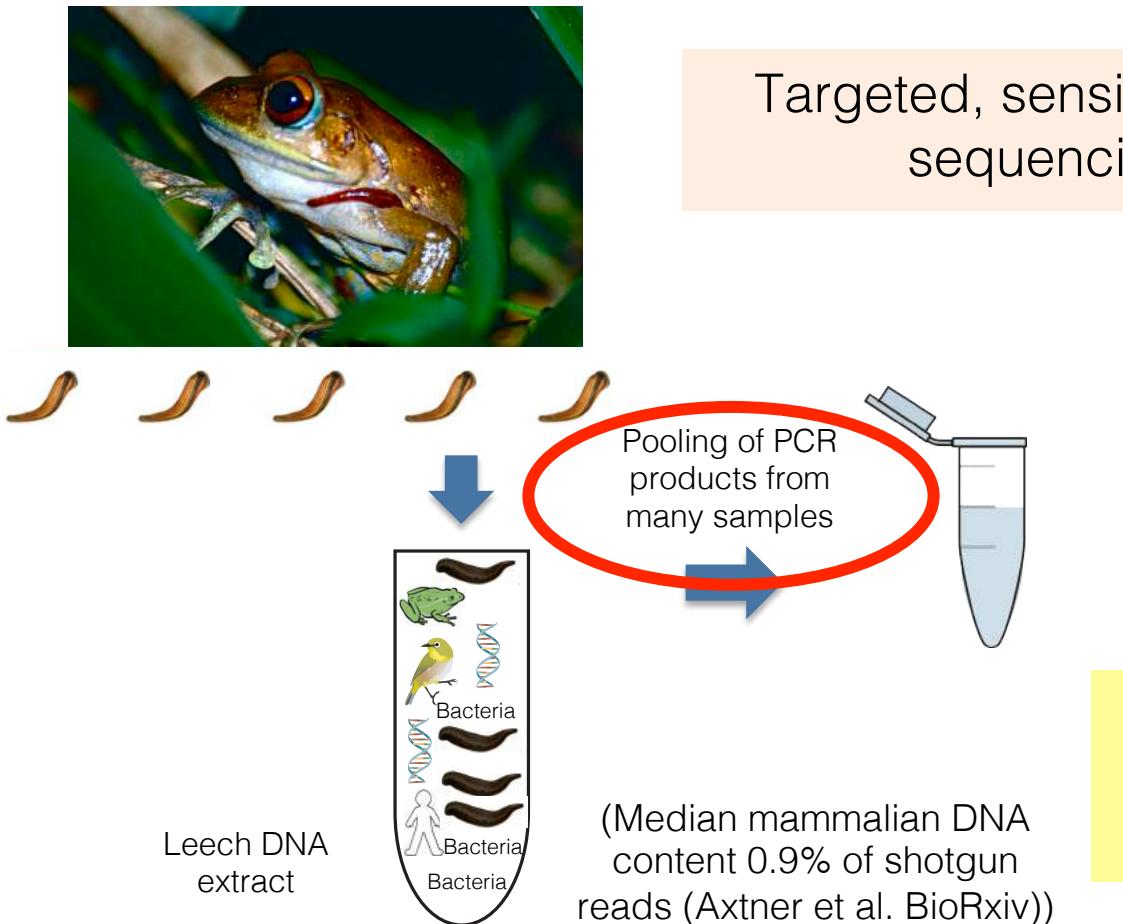
Different ways to use PCR replicates during data processing



- Min 3 individually labelled PCR replicates per sample to facilitate balancing error removal with diversity detection (and account for human error)
- Study dependent how to use the PCR replicates
- Incl pos and neg controls and use of copy number thresholds, etc.



...are also where problems arise



Today we will focus on sensitivity and issues with pooling

Another source of ‘contamination’: tag-jumps

What we want to see in our metabarcoding data

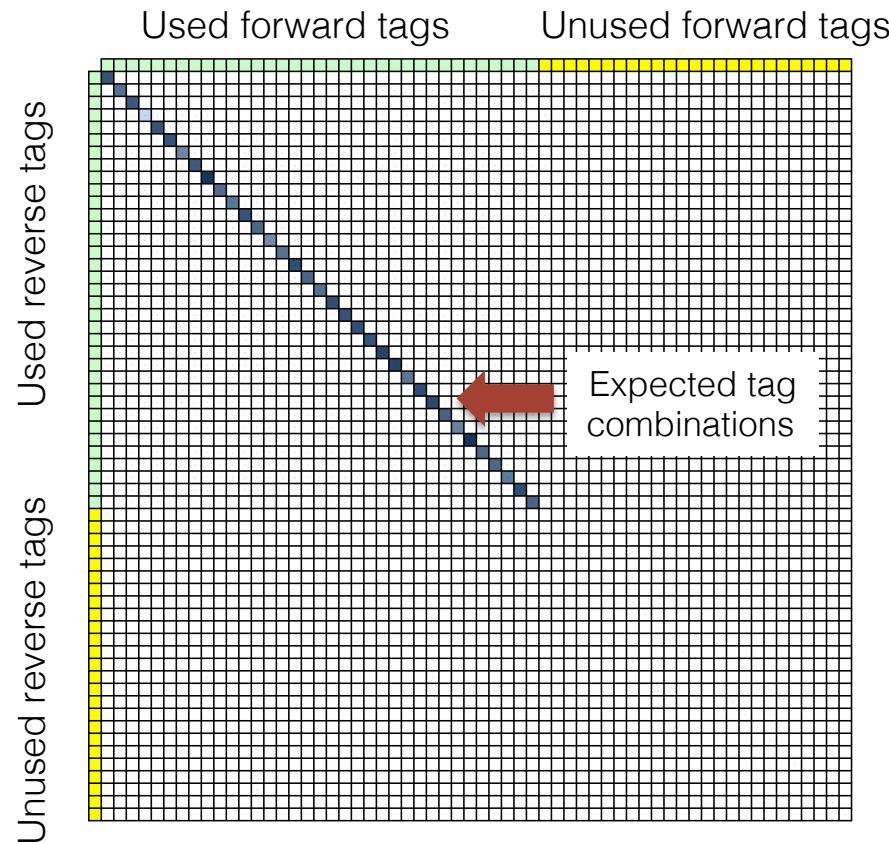
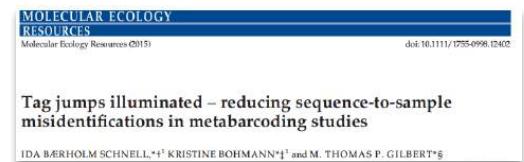


Figure modified for illustrative purposes from:



What we saw

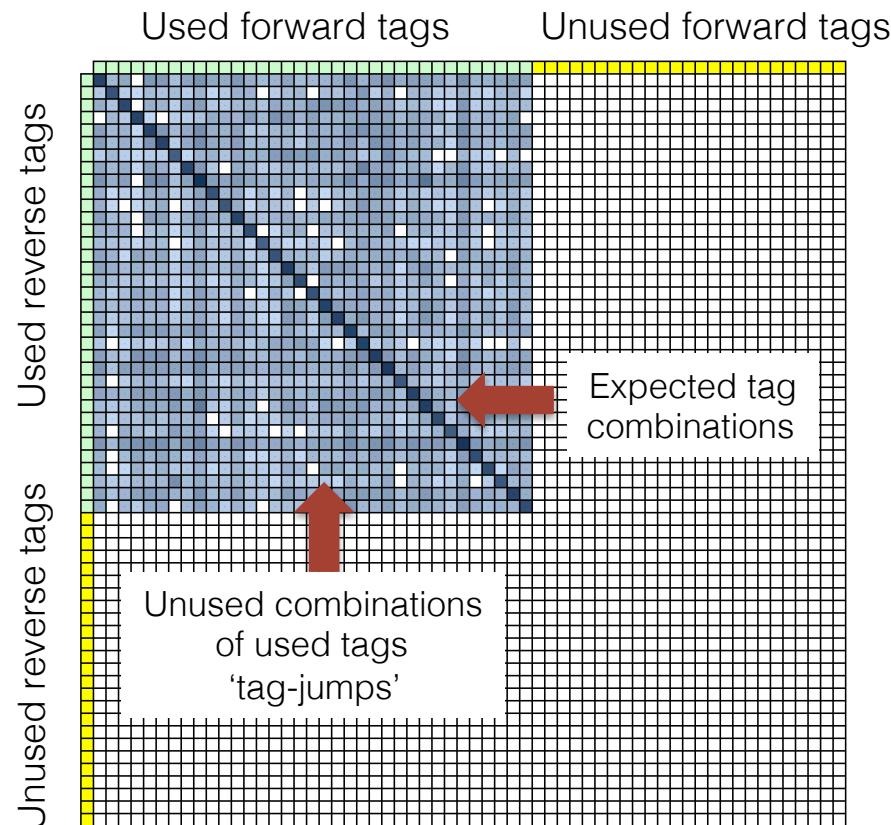
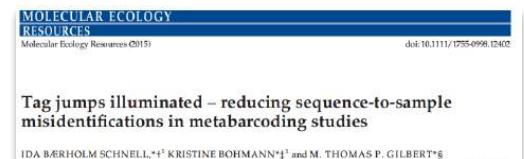


Figure modified for illustrative purposes from:



Three main approaches to label amplicons in metabarcoding studies

One-step PCR with fusion primers



Two-step PCR - first PCR with metabarcoding primers, second PCR with fusion primers



Tagged PCR with library build on amplicon pools

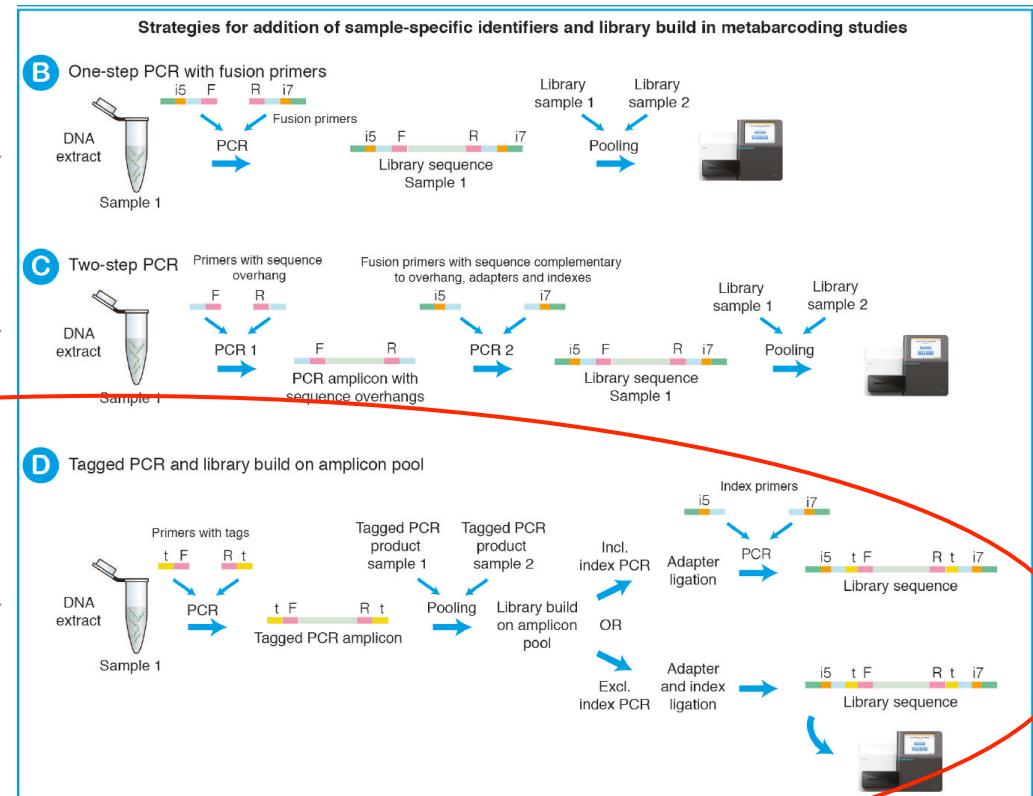
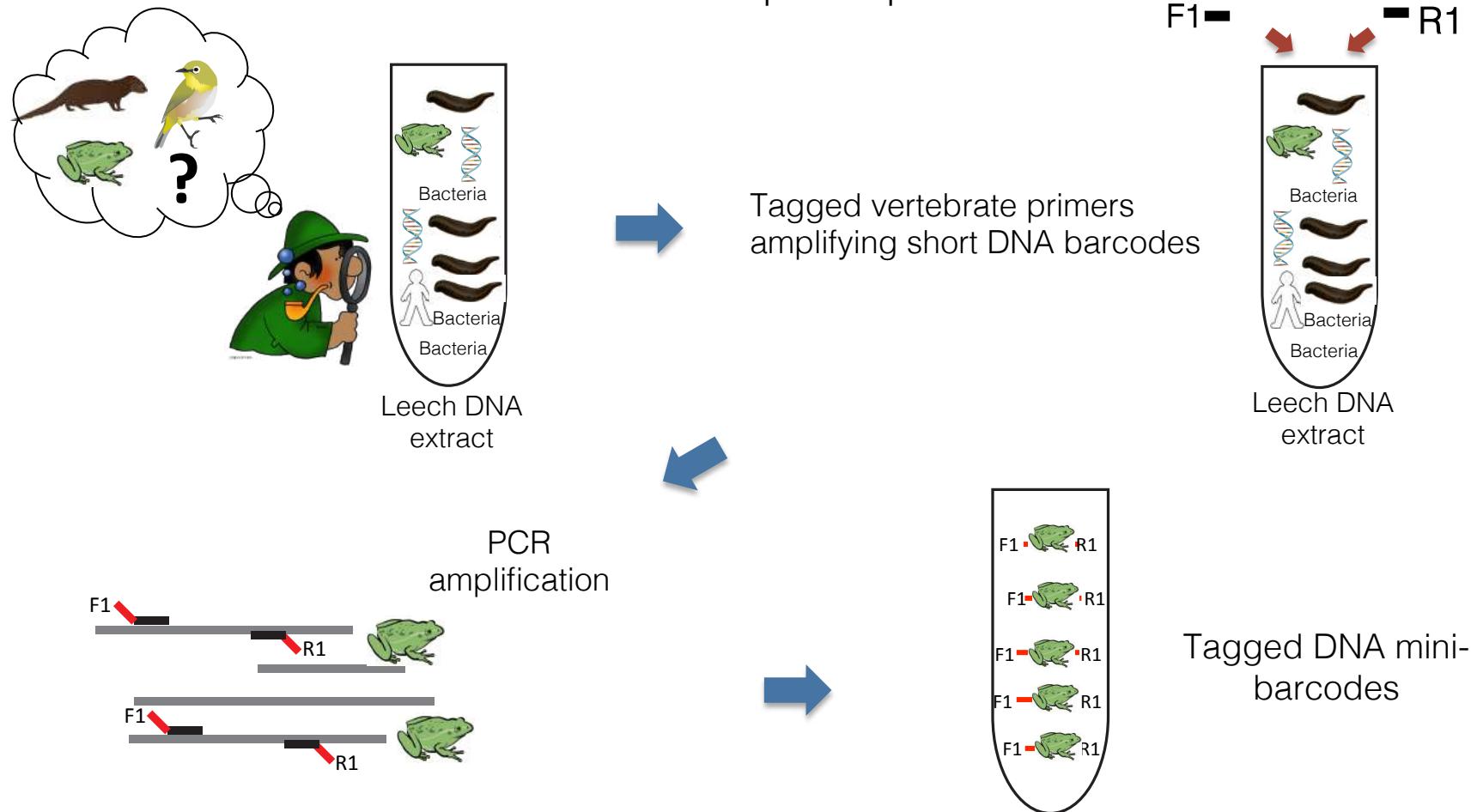


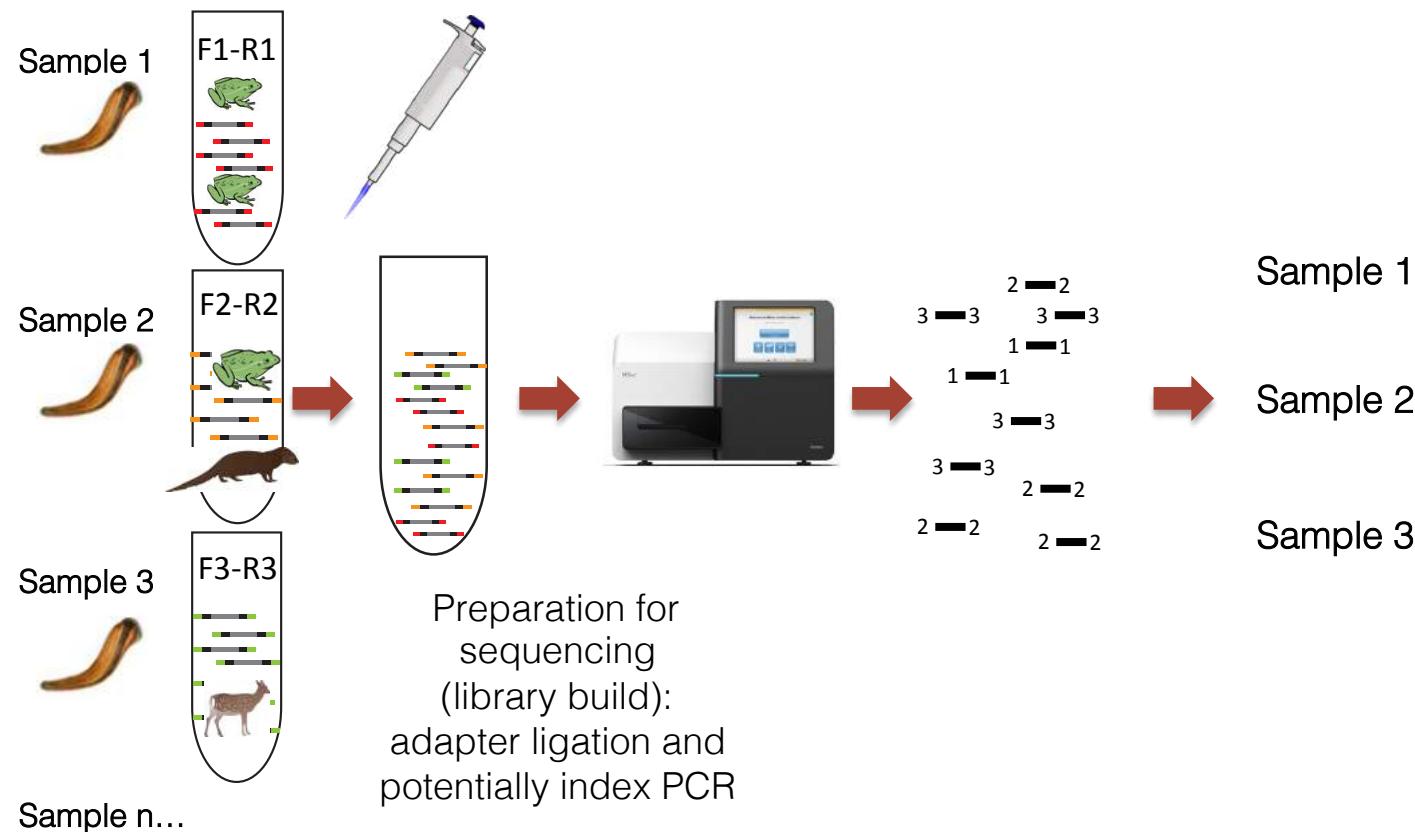
Figure from Bohmann, Elbrecht, Leese, Bunce, ... Creer, et al, in prep

Metabarcoding with tagged primers and library build on amplicon pools

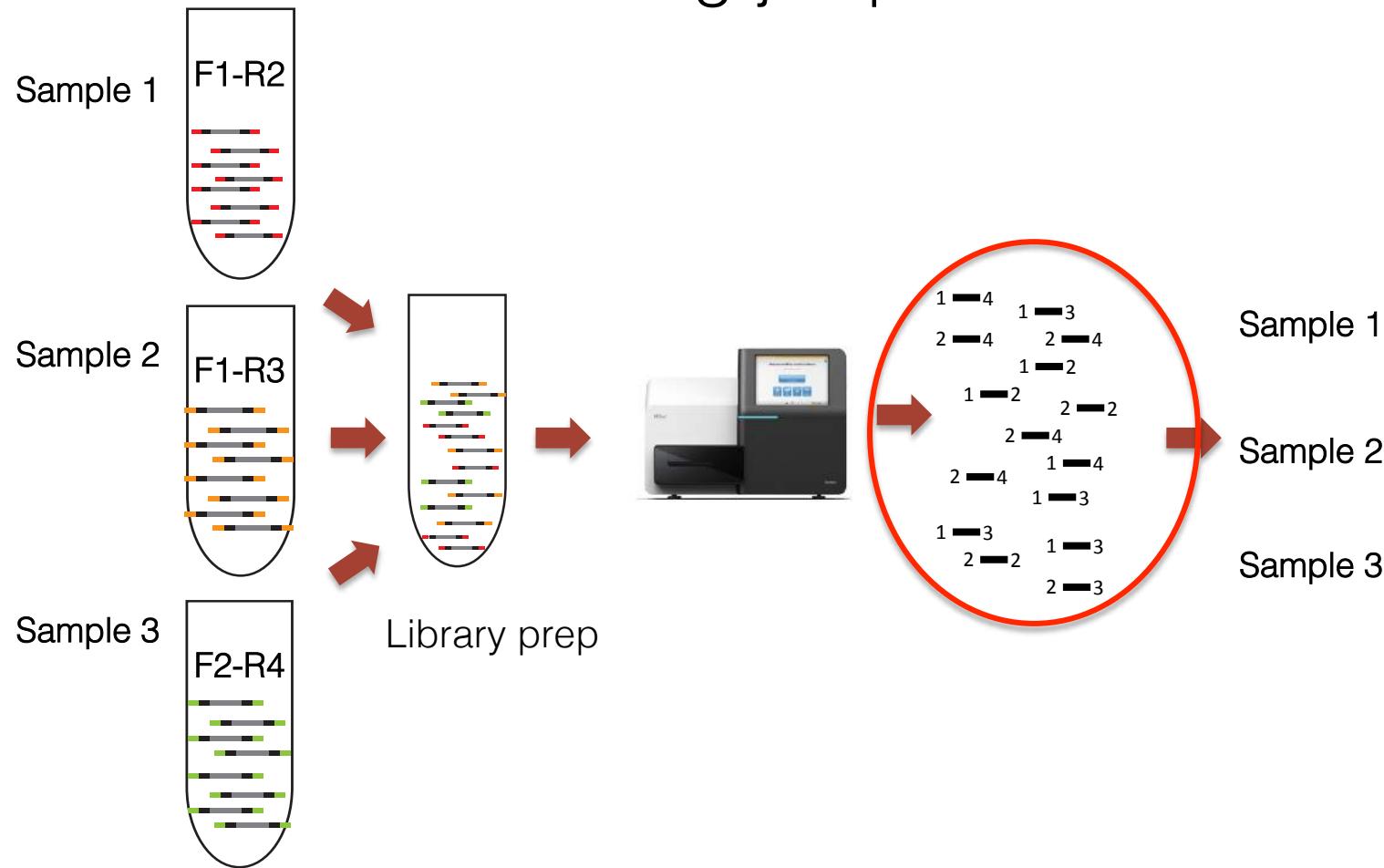


Metabarcoding with tagged primers and library build on amplicon pools

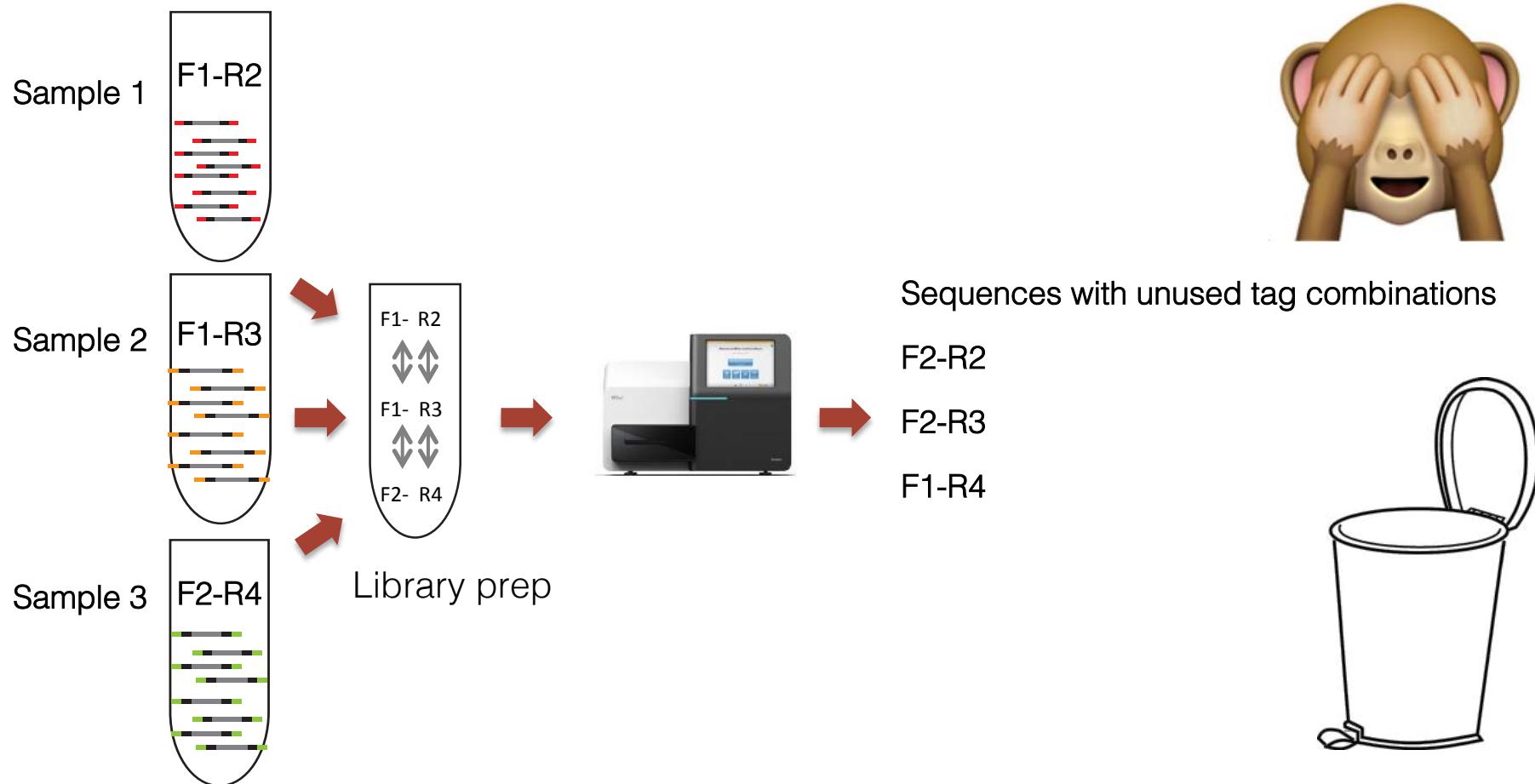
One vertebrate primer set
– differently tagged



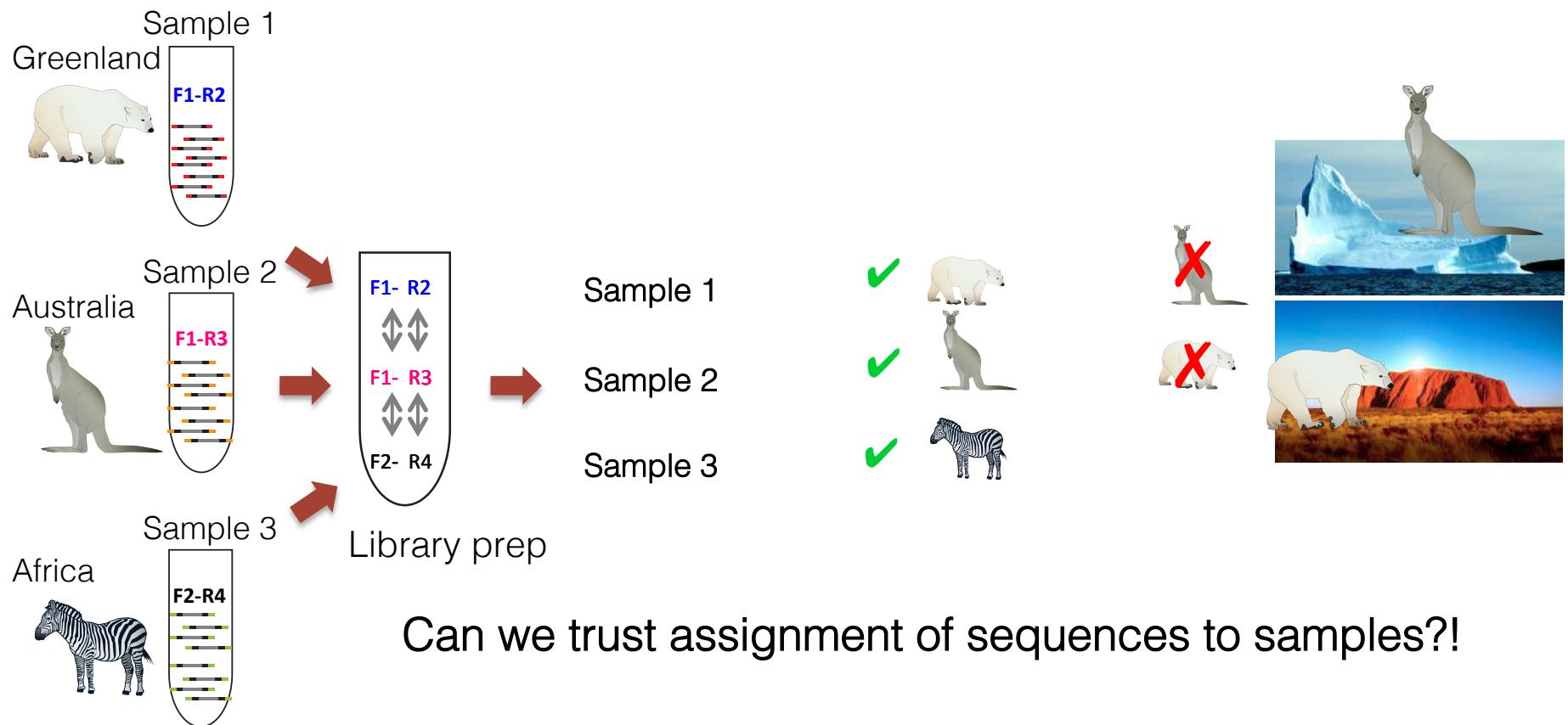
Tag-jumps?



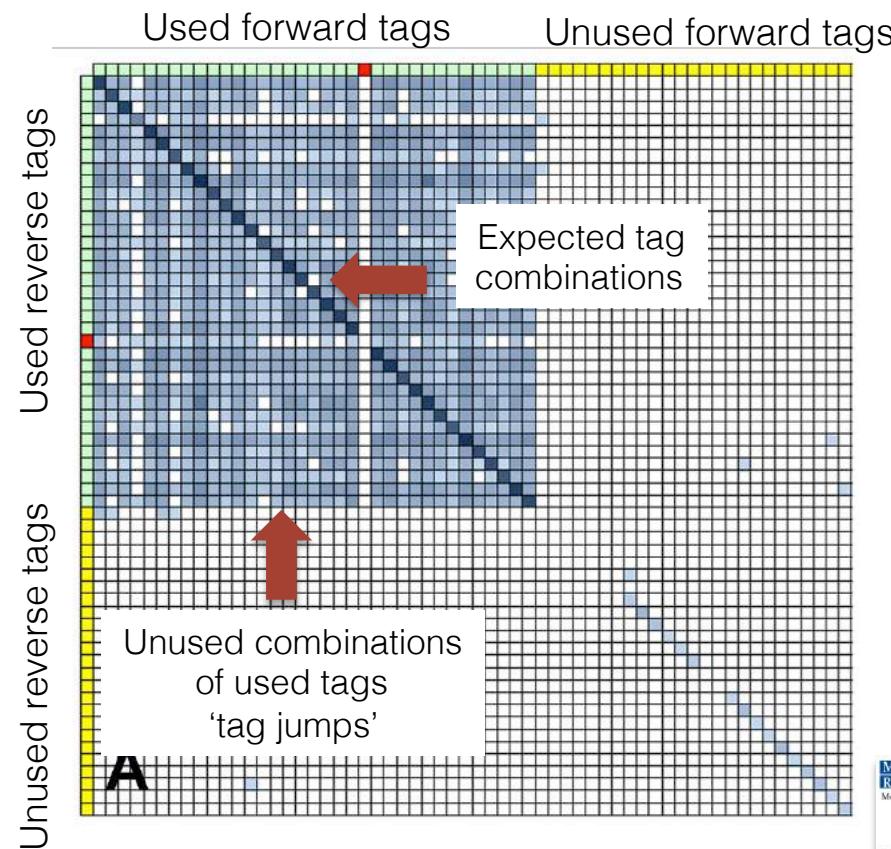
Sequences with unused tag combinations



Tag-jumps creating sequences with used tag combinations can make taxa spillover between samples



Documented tag-jumps on Illumina platform



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^{*§}

FUNGAL ECOLOGY 5 (2012) 747–749



available at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/funeco



Commentary

Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies?

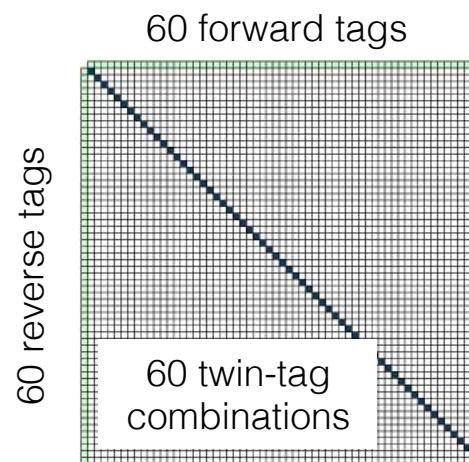
What would be the consequence of tag-jumps in your study?

Please write in the chat

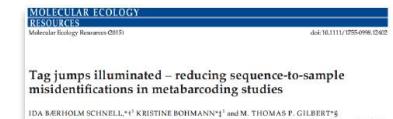
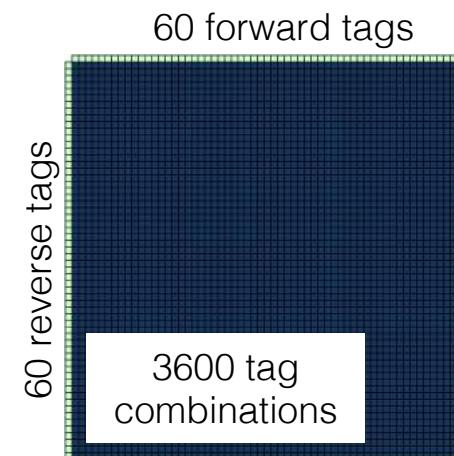
Be aware of tag jumps - accounting for them

- Your research question
- Only use each tag once in a library → many tagged primers and libraries to build

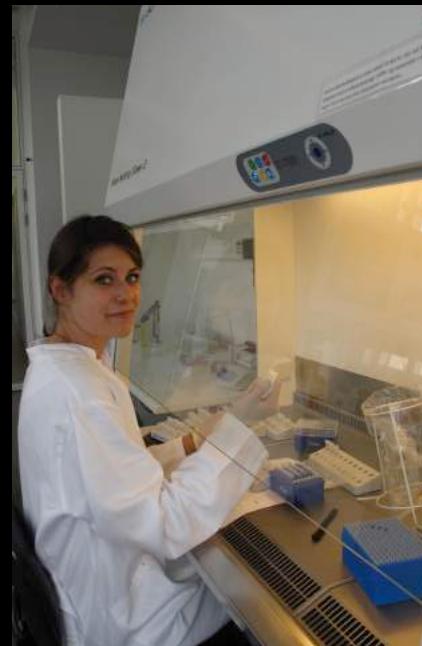
→ Cost, workload ☹



versus

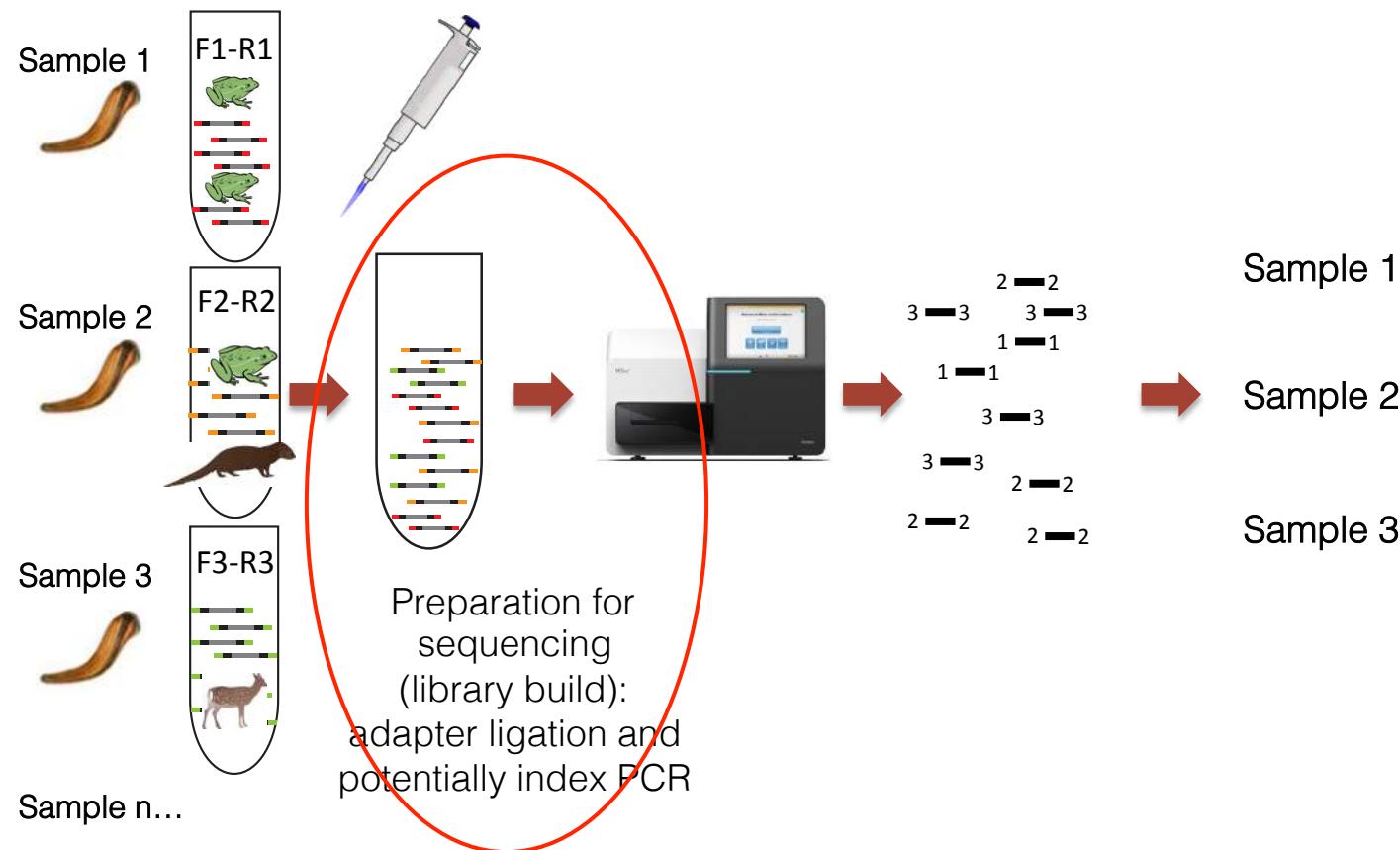


Developing and validating a protocol for library preparation of pools of tagged amplicons for Illumina sequencing



Metabarcoding with tagged primers and library build on amplicon pools

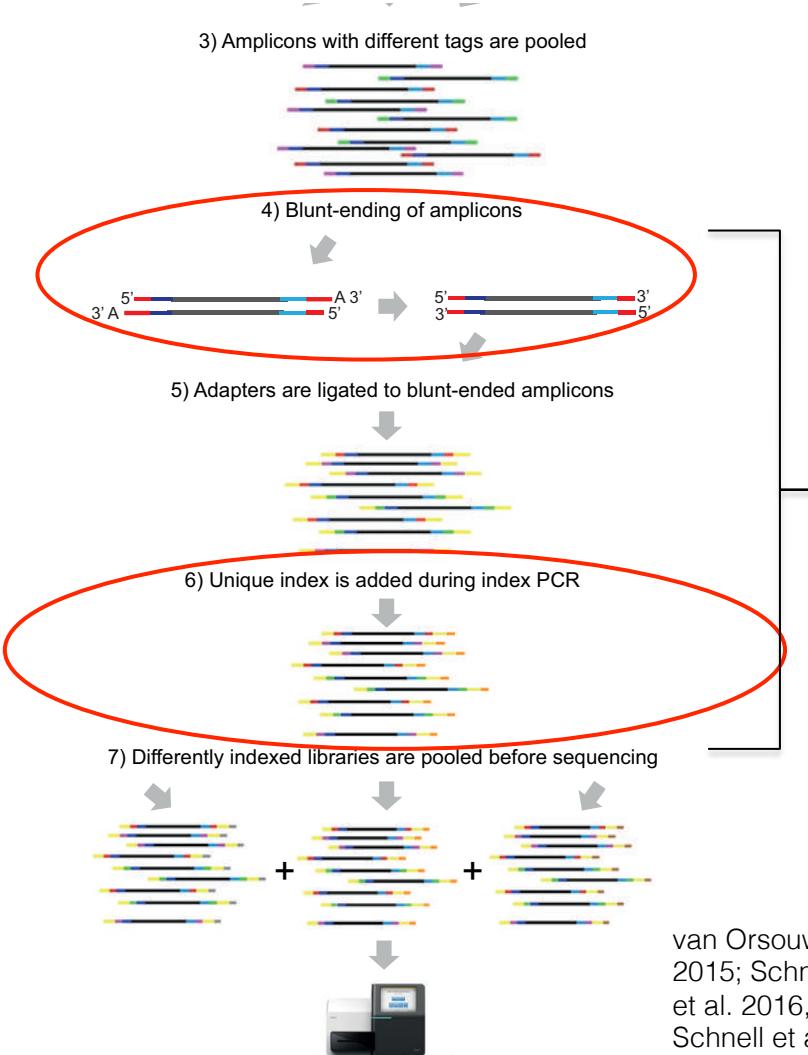
One vertebrate primer set
– differently tagged



Causes

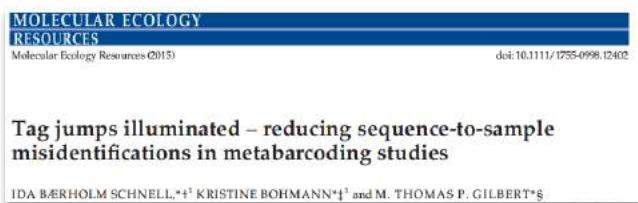
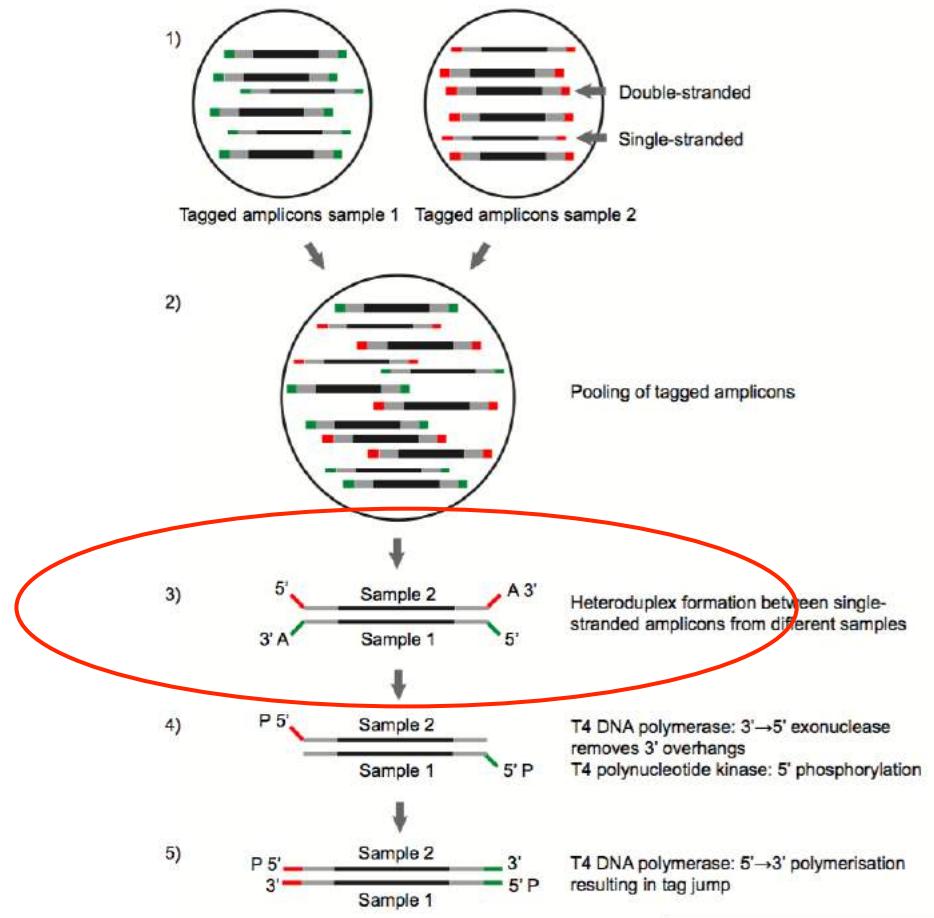
T4 DNA polymerase activity ←

Post-ligation PCR -
chimera formation ←



van Orsouw et al. 2007; Esling et al. 2015; Schnell et al. 2015; Palkopoulou et al. 2016, figure adopted from Schnell et al 2015

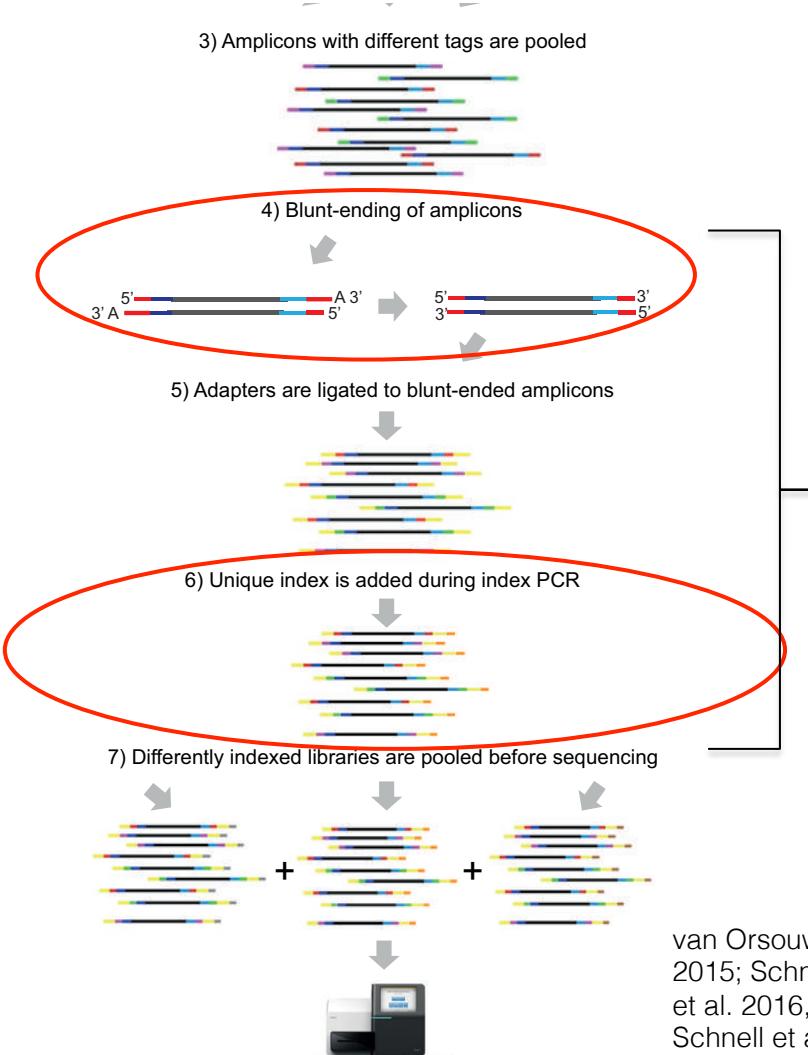
T4 DNA polymerase



Causes

T4 DNA polymerase activity ←

Post-ligation PCR -
chimera formation ←

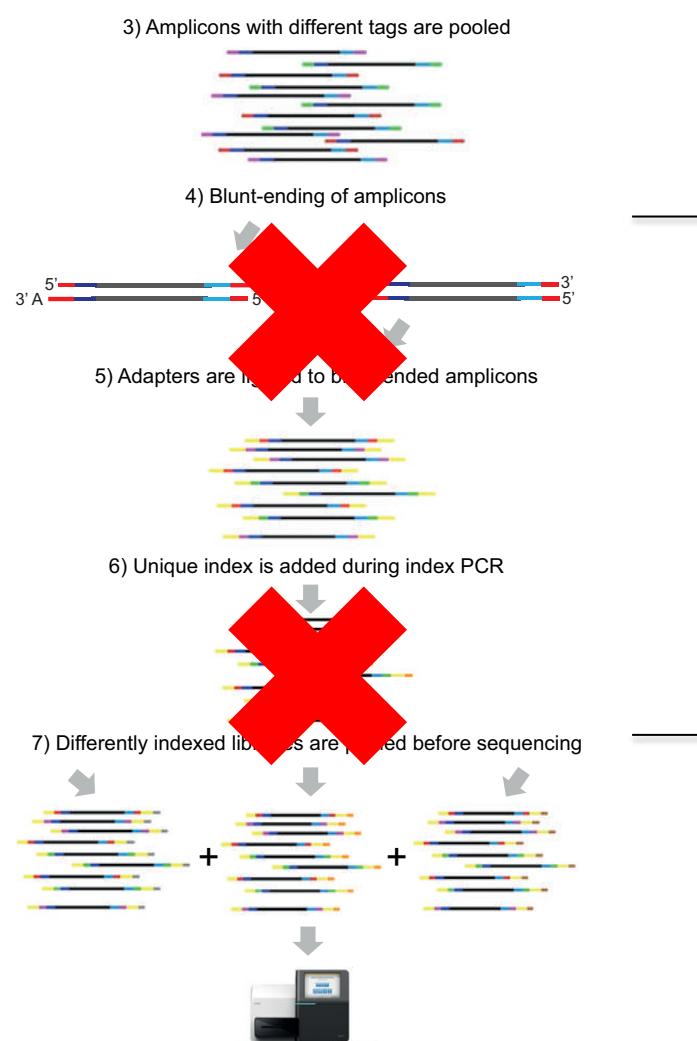


van Orsouw et al. 2007; Esling et al. 2015; Schnell et al. 2015; Palkopoulou et al. 2016, figure adopted from Schnell et al 2015

Avoiding tag-jumps

T4 DNA polymerase activity ←

Post-ligation PCR -
chimera formation ←



Preparing for
sequencing
(library build)

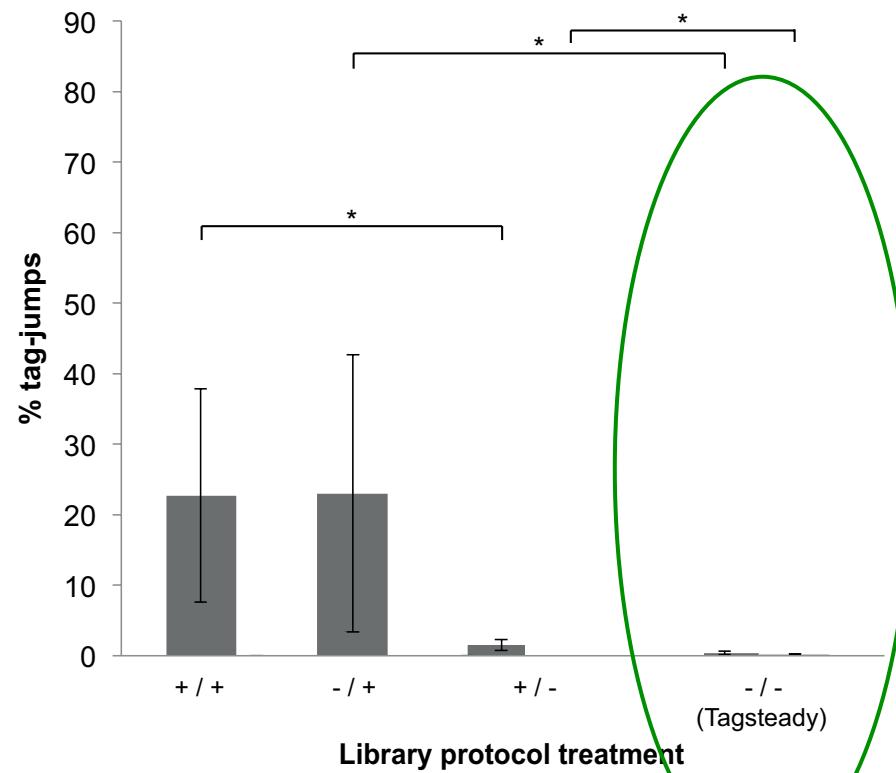
Figure adopted from
Schnell et al 2015

Developing and validating Tagsteady
- a protocol for library preparation of pools of tagged amplicons for Illumina sequencing

Investigating the effect of the two steps on tag-jump levels

On pools of amplicons with twin-tags:

	PCR on amplicon library (PCR)	
T4 DNA polymerase (T4)	+ T4 + PCR	+ T4 ÷ PCR
	÷ T4 + PCR	÷ T4 ÷ PCR (Tagsteady)



MOLECULAR ECOLOGY
RESOURCES

RESOURCE ARTICLE

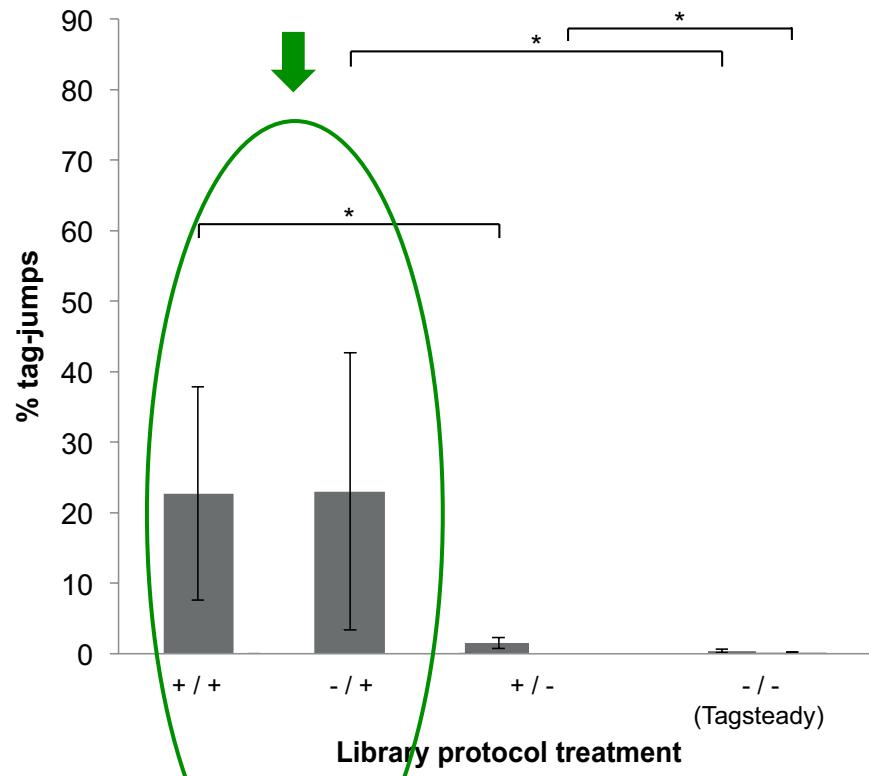
Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples

Christian Carøe ✉, Kristine Bohmann ✉

First published: 14 July 2020 | <https://doi.org/10.1111/1755-0998.13227>

Both authors contributed equally to this work.

Post-ligation PCR gives chimera formation and tag-jumps



MOLECULAR ECOLOGY
RESOURCES

RESOURCE ARTICLE

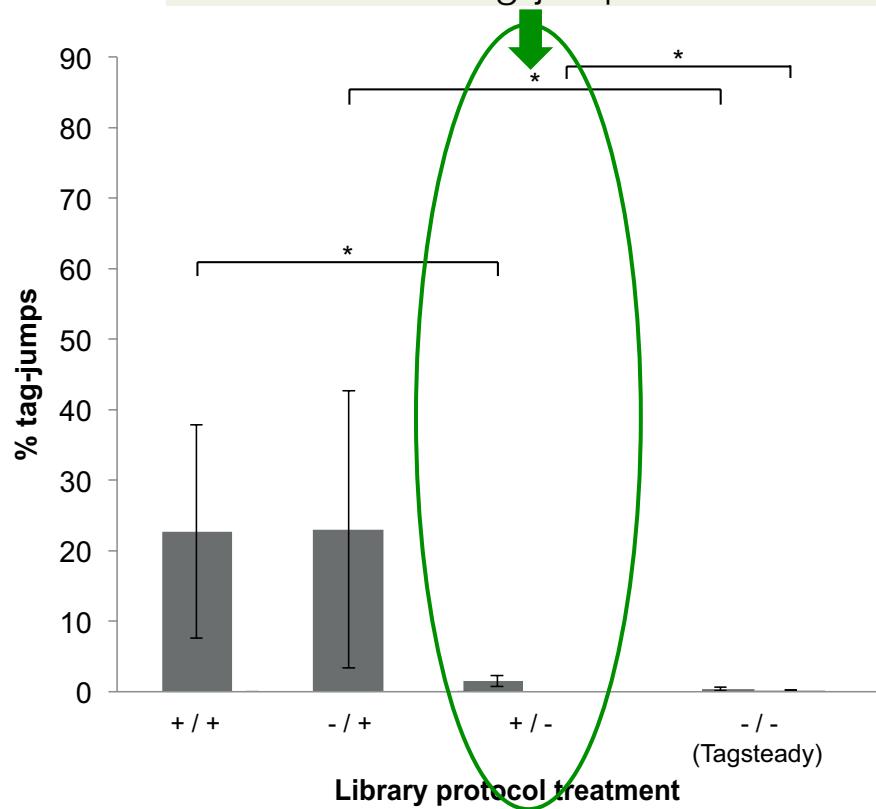
Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples

Christian Carøe ✉, Kristine Bohmann ✉

First published: 14 July 2020 | <https://doi.org/10.1111/1755-0998.13227>

Both authors contributed equally to this work.

T4 DNA polymerase causes low levels
of tag-jumps



MOLECULAR ECOLOGY
RESOURCES

RESOURCE ARTICLE

Tagsteady: A metabarcoding library preparation protocol to
avoid false assignment of sequences to samples

Christian Carøe ✉, Kristine Bohmann ✉

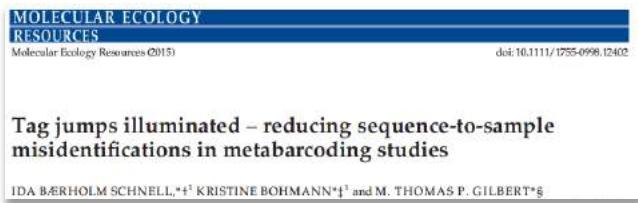
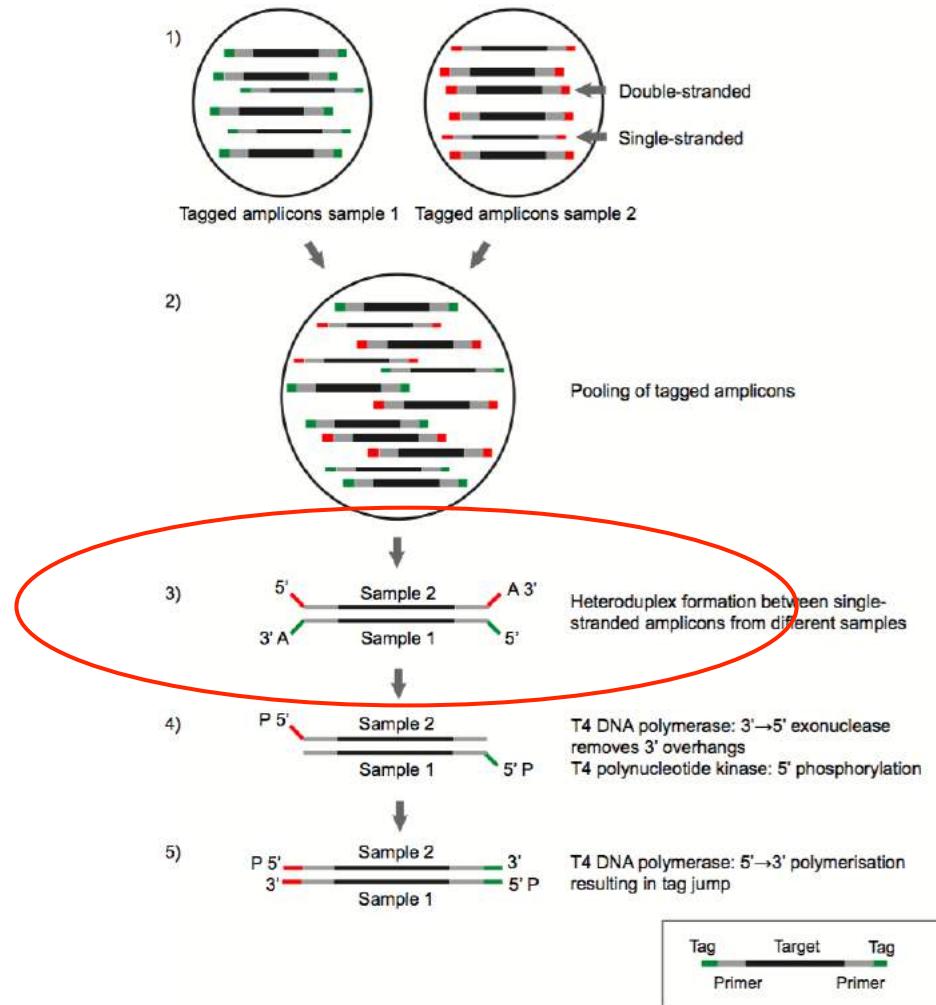
First published: 14 July 2020 | <https://doi.org/10.1111/1755-0998.13227>

Both authors contributed equally to this work.

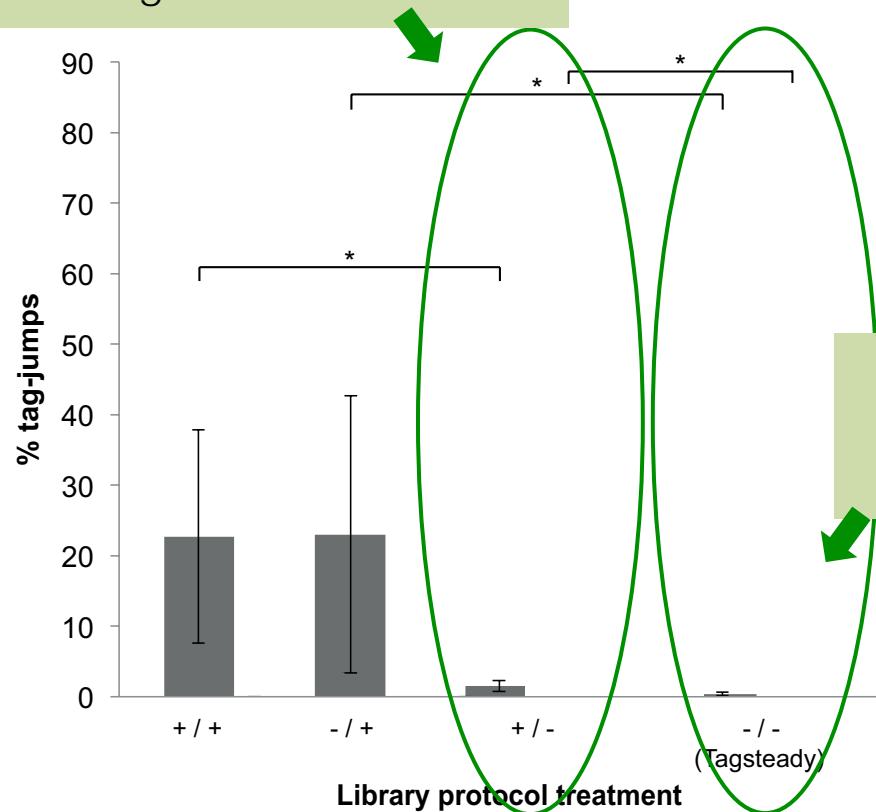
Why don't we see a pronounced effect of the
T4 DNA polymerase step?

T4 DNA polymerase

Maybe we didn't have a lot of single-stranded DNA in our amplicon pools?



1) Will tag-jumps increase if higher amount of single-stranded DNA?



2. Can the Tagsteady protocol handle higher amounts of single-stranded DNA?

MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE

Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples

Christian Carøe, Kristine Bohmann

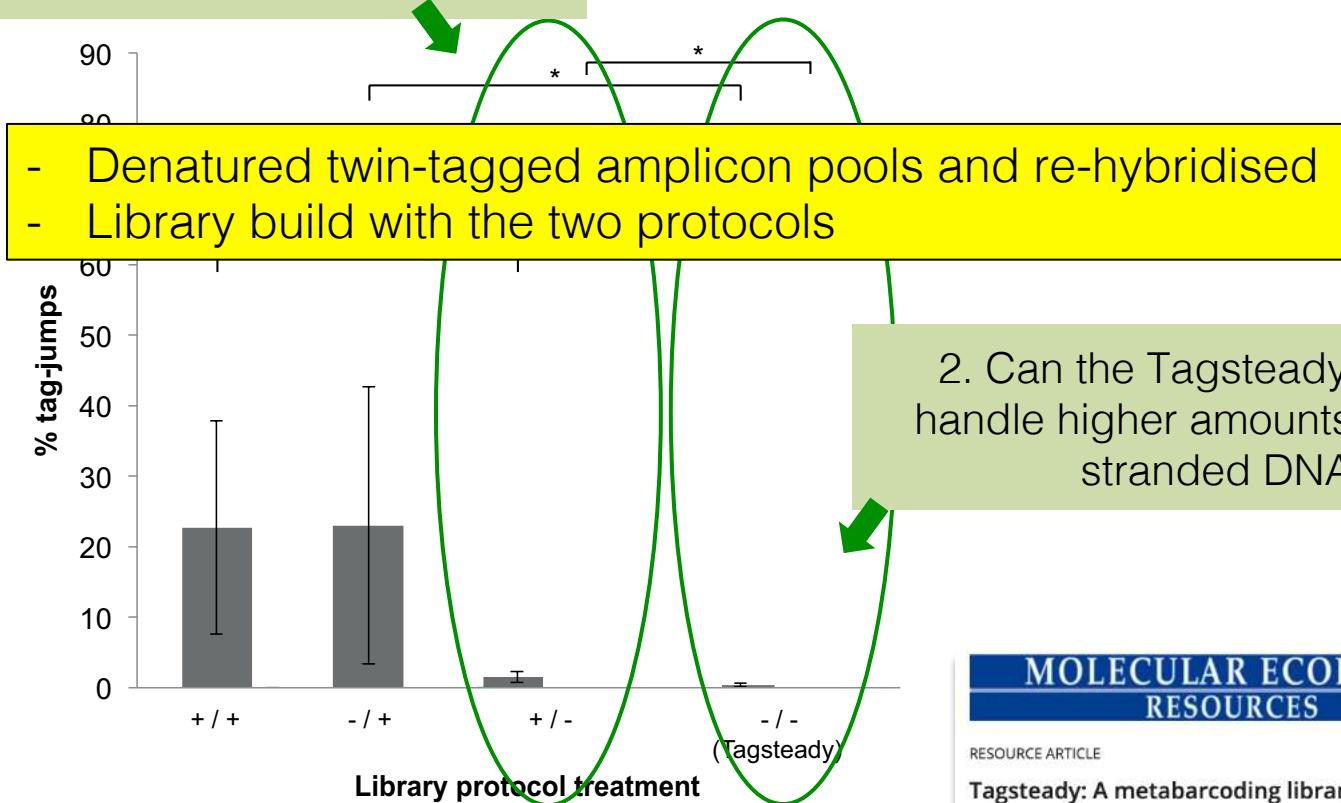
First published: 14 July 2020 | <https://doi.org/10.1111/1755-0998.13227>

Both authors contributed equally to this work.



shutterstock.com · 1898397400

1) Will tag-jumps increase if higher amount of single-stranded DNA?



2. Can the Tagsteady protocol handle higher amounts of single-stranded DNA?

MOLECULAR ECOLOGY
RESOURCES

RESOURCE ARTICLE

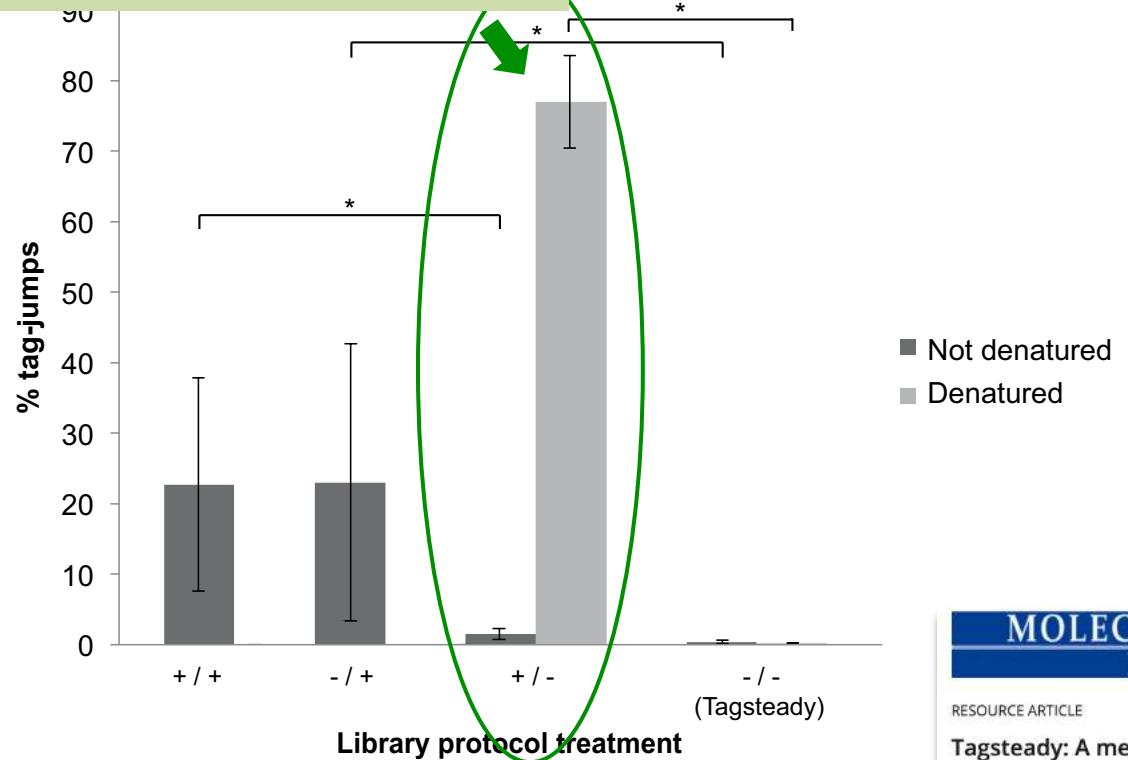
Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples

Christian Carøe, Kristine Bohmann

First published: 14 July 2020 | <https://doi.org/10.1111/1755-0998.13227>

Both authors contributed equally to this work.

1) Will tag-jumps increase if higher amount of single-stranded DNA?



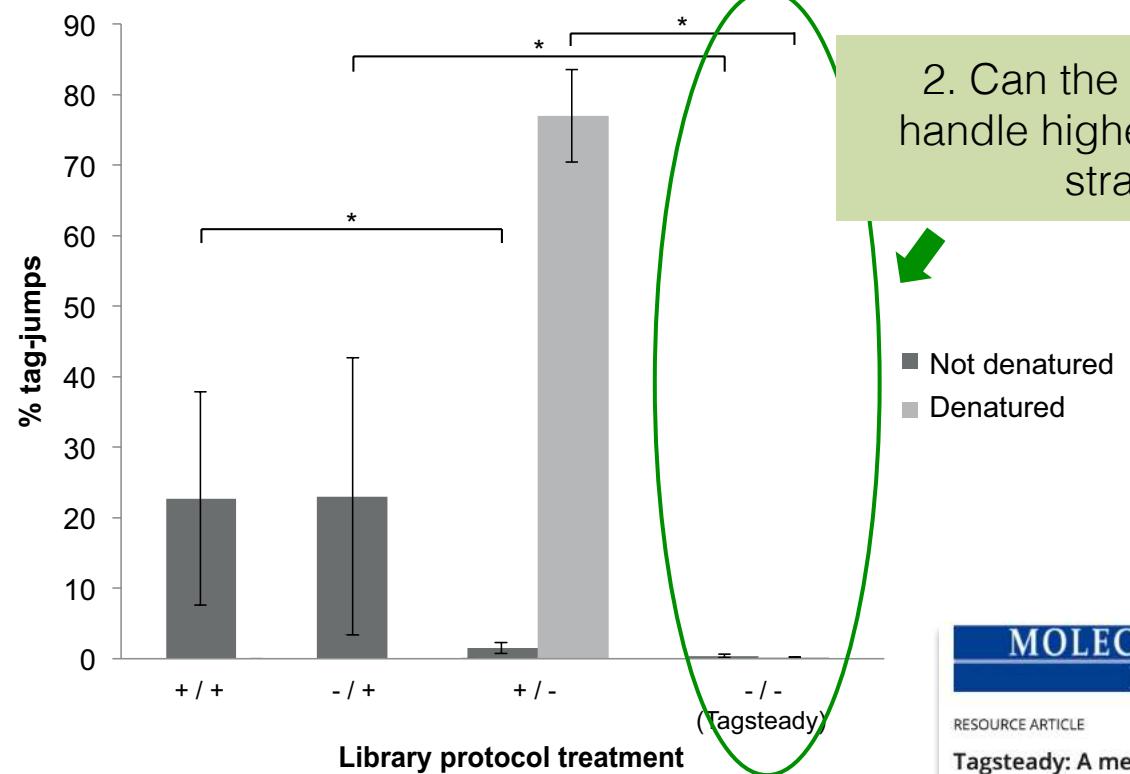
RESOURCE ARTICLE

Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples

Christian Carøe, Kristine Bohmann

First published: 14 July 2020 | <https://doi.org/10.1111/1755-0998.13227>

Both authors contributed equally to this work.



2. Can the Tagsteady protocol handle higher amounts of single-stranded DNA?

- Not denatured
- Denatured

MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE

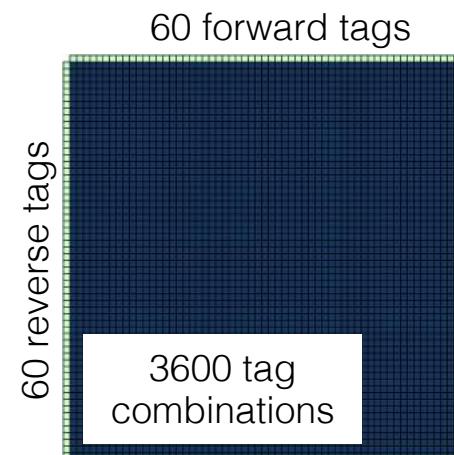
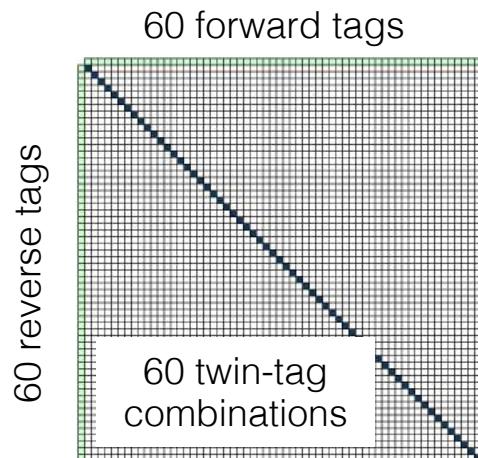
Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples

Christian Carøe, Kristine Bohmann

First published: 14 July 2020 | <https://doi.org/10.1111/1755-0998.13227>

Both authors contributed equally to this work.

No tag-jumps → cost-effective metabarcoding



MOLECULAR ECOLOGY RESOURCES

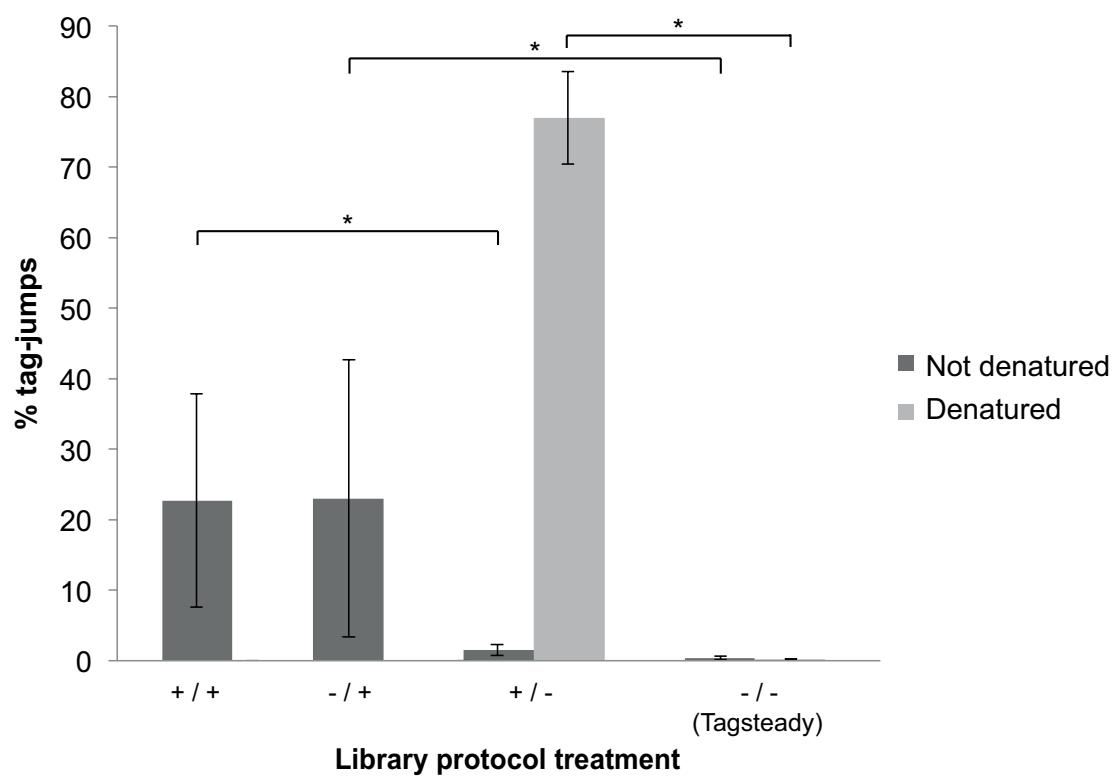
RESOURCE ARTICLE

Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples

Christian Carøe ✉, Kristine Bohmann ✉

First published: 14 July 2020 | <https://doi.org/10.1111/1755-0998.13227>

Both authors contributed equally to this work.



Both T4 DNA Polymerase and post-ligation PCR can cause tag-jumps

You can account for tag-jumps by using twin tags, but expensive to buy the primers and you can lose a lot of data to tag-jumps

You can avoid tag-jumps if you use a library prep protocol that does not have these two steps

MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE

Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples

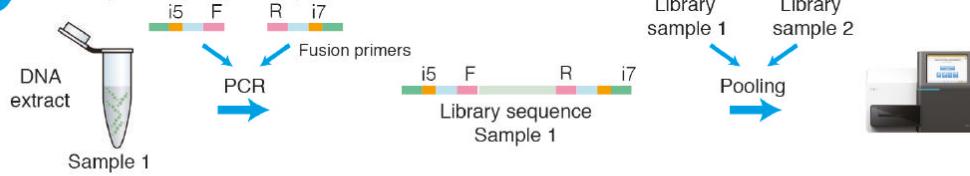
Christian Carøe, Kristine Bohmann

First published: 14 July 2020 | <https://doi.org/10.1111/1755-0998.13227>

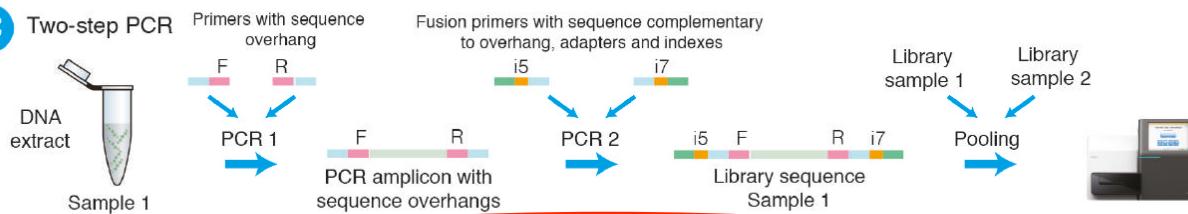
Both authors contributed equally to this work.

Strategies for addition of sample-specific identifiers and library build in metabarcoding studies

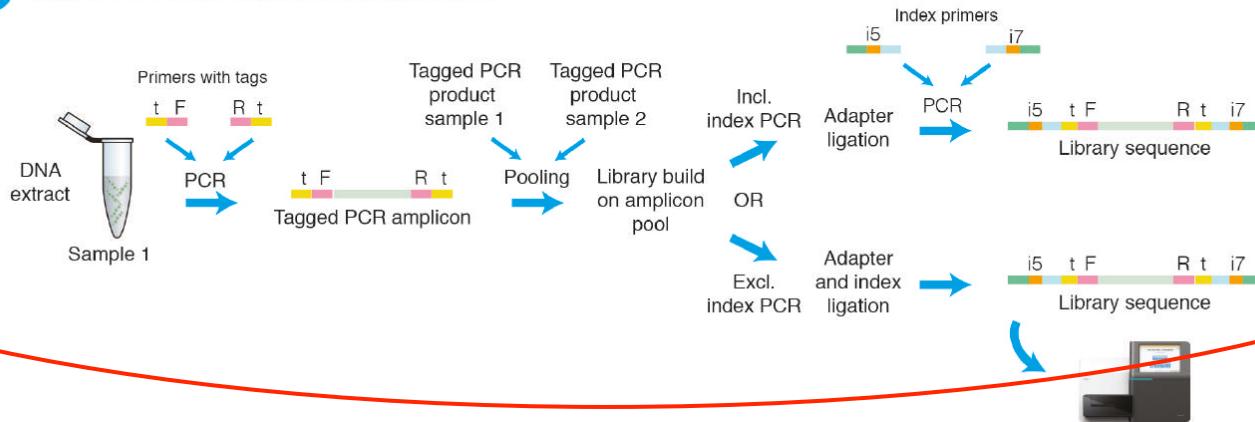
B One-step PCR with fusion primers



C Two-step PCR



D Tagged PCR and library build on amplicon pool



Can tag-jumps affect your study?

Intended learning outcomes

After this session you will be able to:

- Identify different sources of contamination – where it can arise, how to detect it and how to avoid it
- Outline the three main metabarcoding strategies for sample labelling
- Describe how PCR replicates can be used during data processing to balance error removal with detection of diversity
- Demonstrate an understanding of tag-jumps - how they can be accounted for and/or avoided and how they can be identified during data processing
- Formulate a strategy for the lab set-up of your metabarcoding study to account for or avoid different sources of contamination and facilitate the ability to balance error removal with diversity detection during data processing

Take a moment to reflect on what you would like to incorporate in your metabarcoding lab set-up regarding contamination

- What are potential sources of contamination in your study
- Which ones can have consequences for your study

What will you do

- to avoid contamination during different steps of your lab work
- to detect contamination during lab work or data processing
- to facilitate ability to balance error removal with diversity detection during data processing
- to detect, account for or avoid tag-jumps

Write notes for yourself

Thank you