Lecture course on environmental DNA metabarcoding using Claident and R: From nucleotide sequence data processing to ecological analyses

Akifumi S. Tanabe

田辺晶史

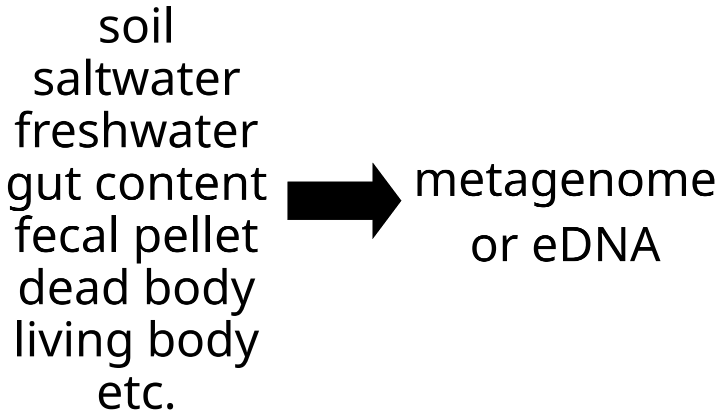
ClaidentとRによる環境DNAメタバーコーディング分析講座：塩基配列データ処理から生態学的分析まで

Workflow of metabarcoding

Workflow of metabarcoding

soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

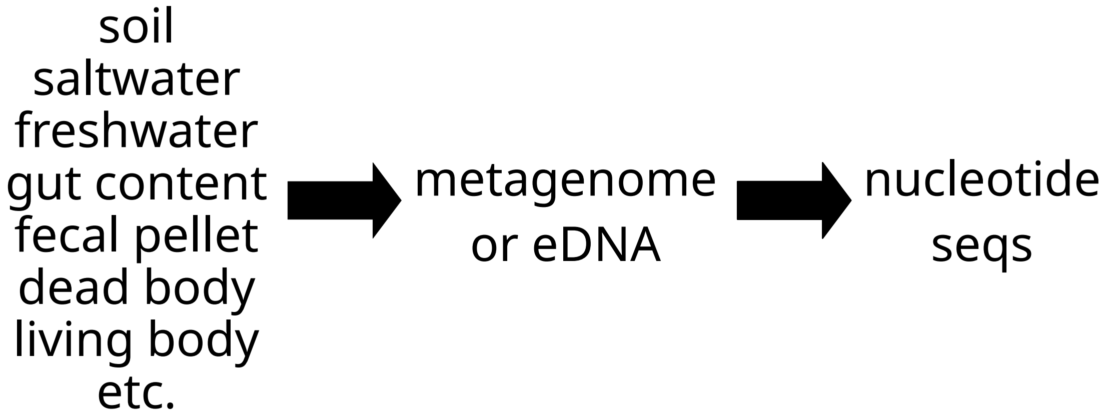
Workflow of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

metagenome
or eDNA

Workflow of metabarcoding

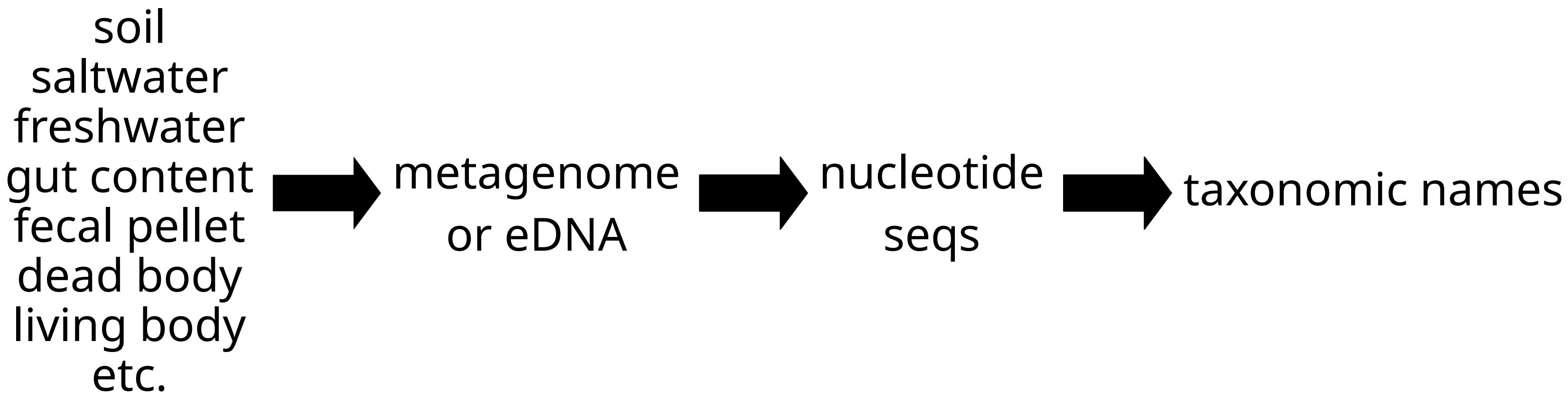


soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

metagenome
or eDNA

nucleotide
seqs

Workflow of metabarcoding



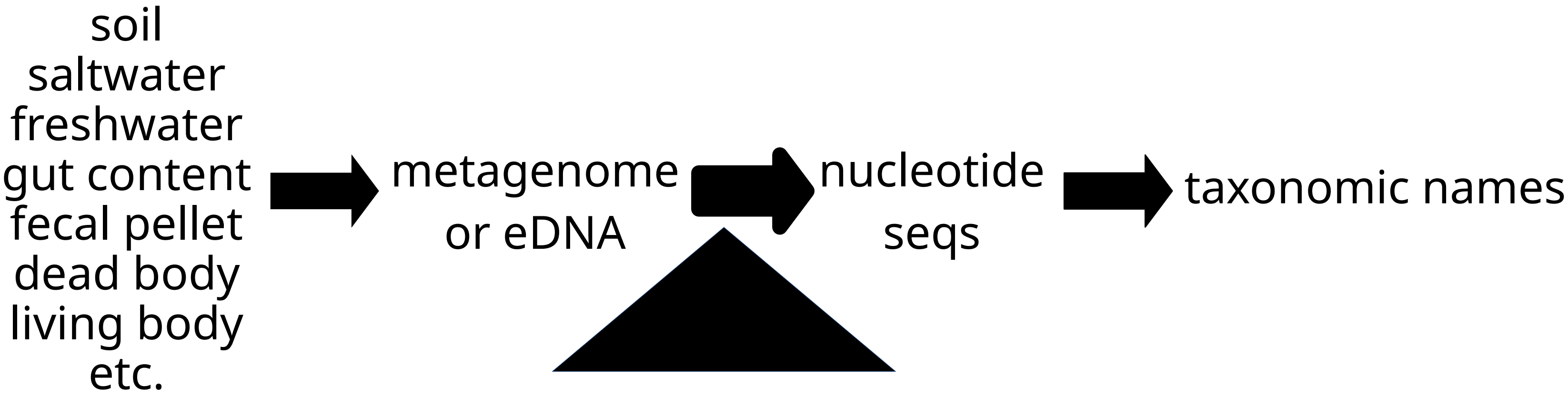
soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

metagenome
or eDNA

nucleotide
seqs

taxonomic names

Molecular laboratory processes of metabarcoding



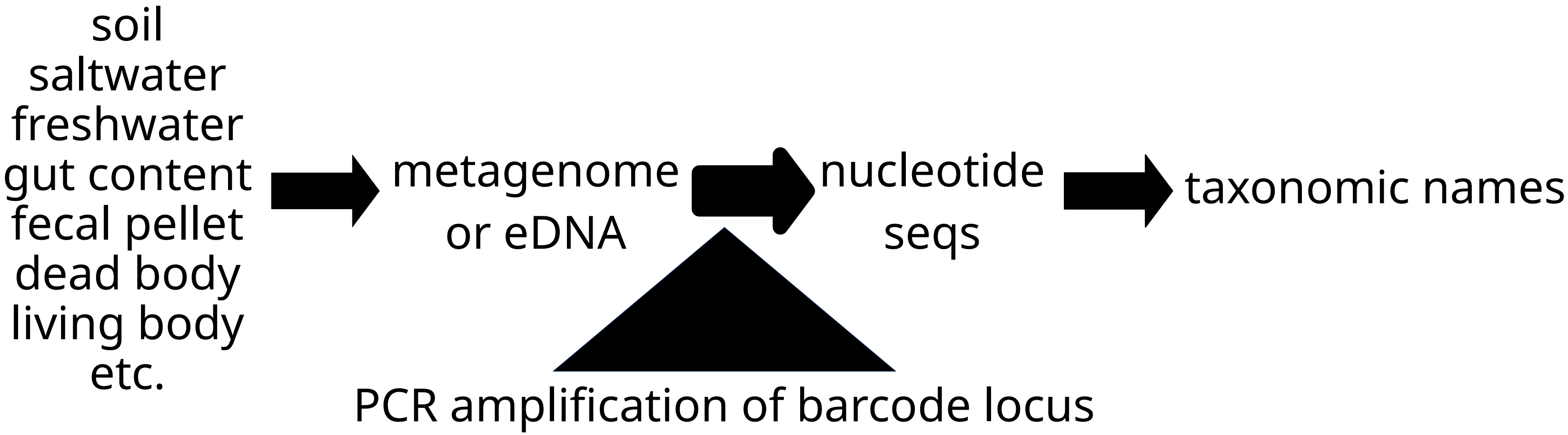
soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

metagenome
or eDNA

nucleotide
seqs

taxonomic names

Molecular laboratory processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

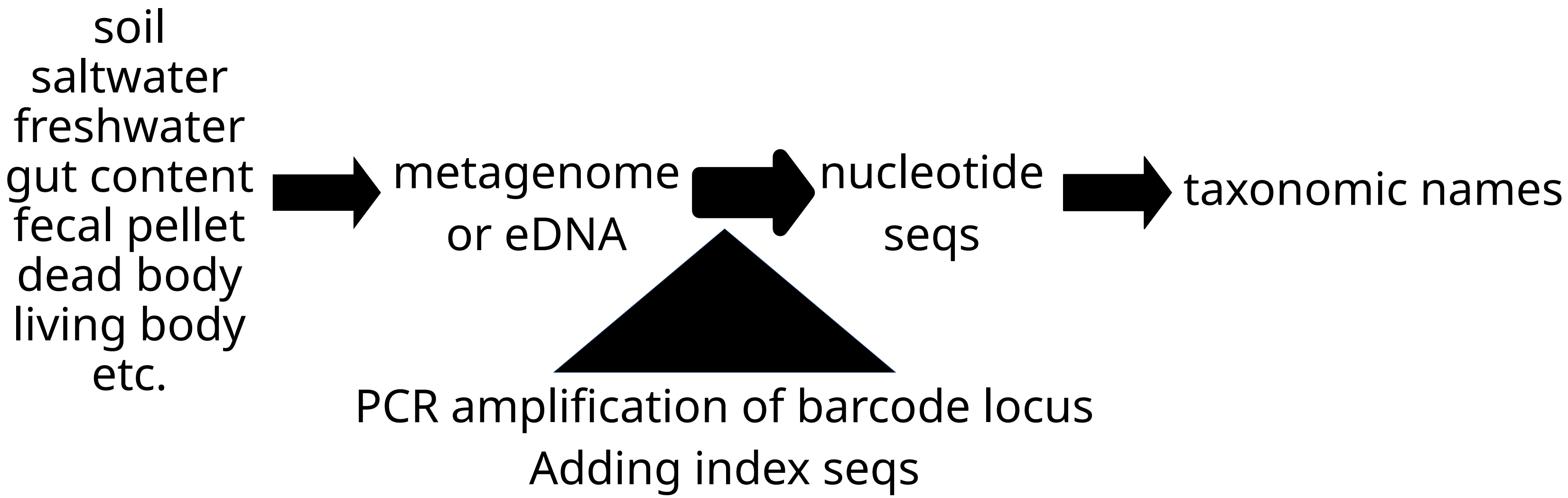
metagenome
or eDNA

nucleotide
seqs

taxonomic names

PCR amplification of barcode locus

Molecular laboratory processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

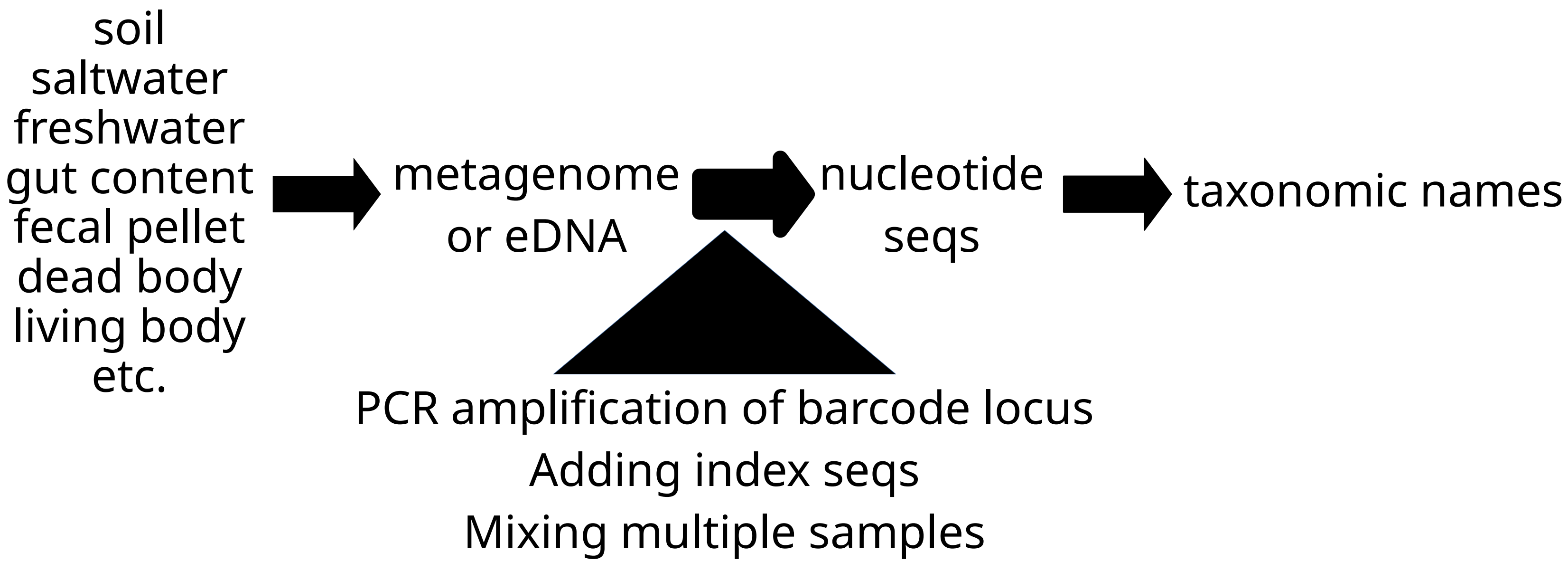
metagenome
or eDNA

nucleotide
seqs

taxonomic names

PCR amplification of barcode locus
Adding index seqs

Molecular laboratory processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

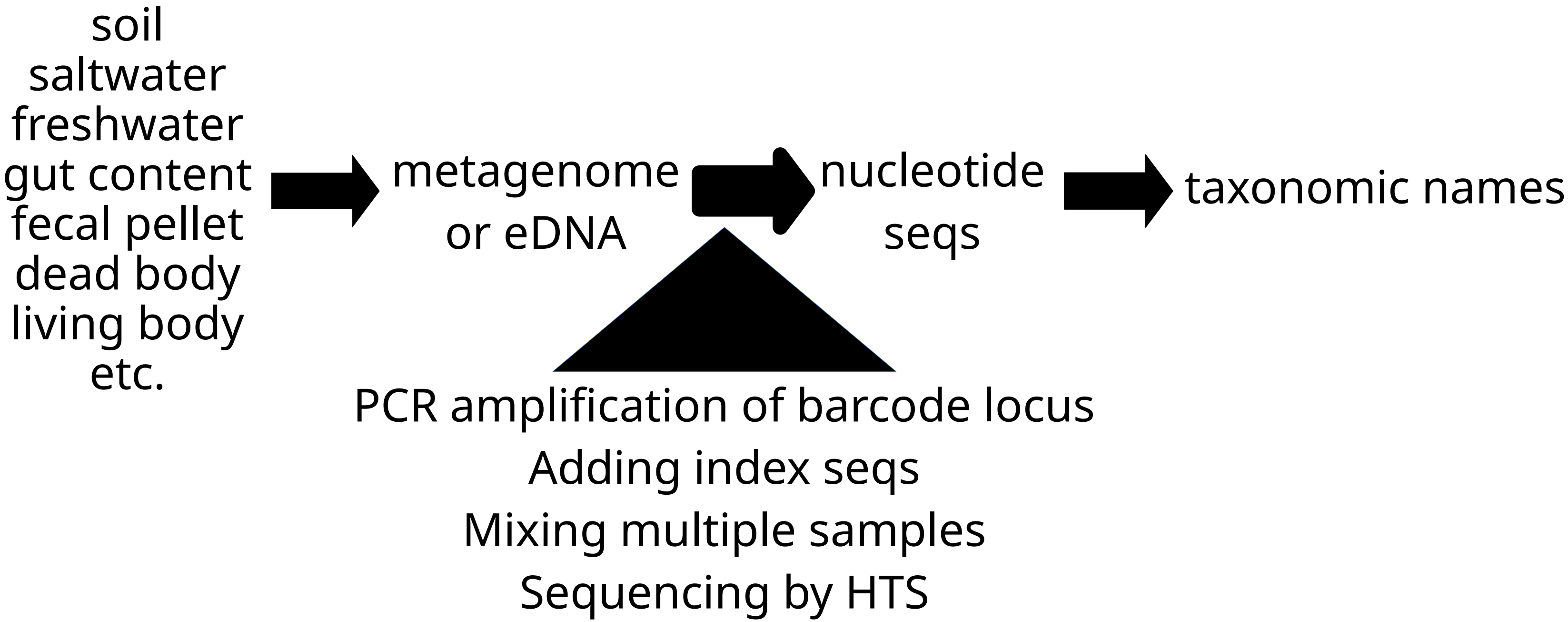
metagenome
or eDNA

nucleotide
seqs

taxonomic names

PCR amplification of barcode locus
Adding index seqs
Mixing multiple samples

Molecular laboratory processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

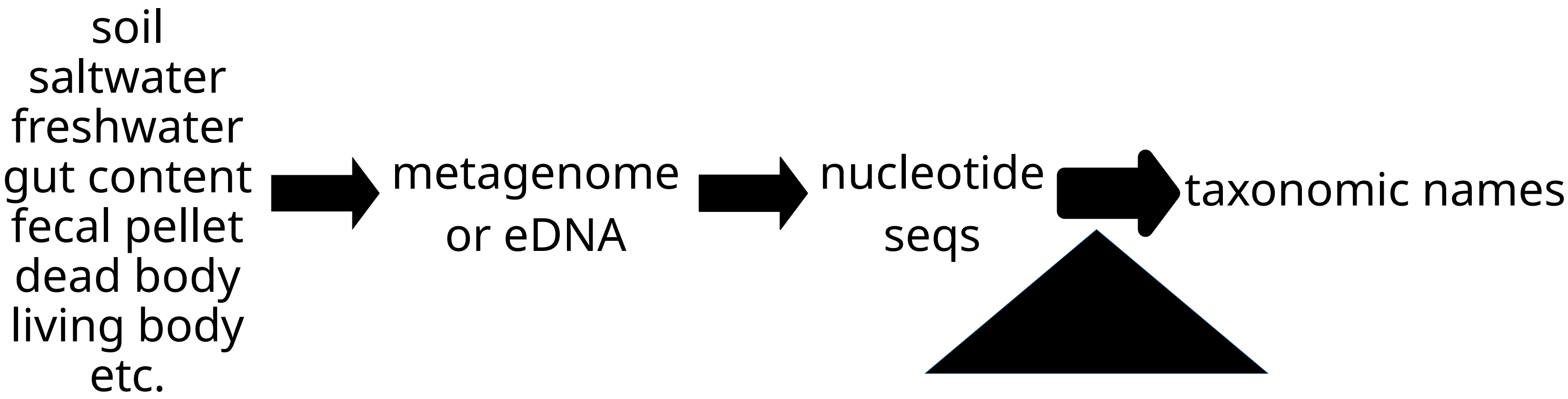
metagenome
or eDNA

nucleotide
seqs

taxonomic names

PCR amplification of barcode locus
Adding index seqs
Mixing multiple samples
Sequencing by HTS

Computational processes of metabarcoding



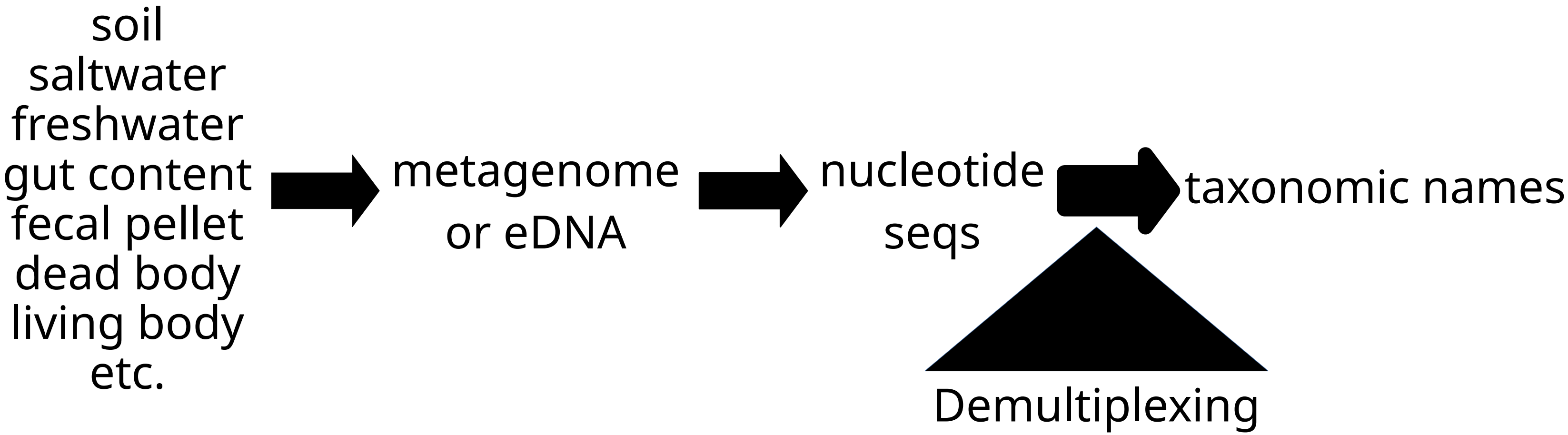
soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

metagenome or eDNA

nucleotide seqs

taxonomic names

Computational processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

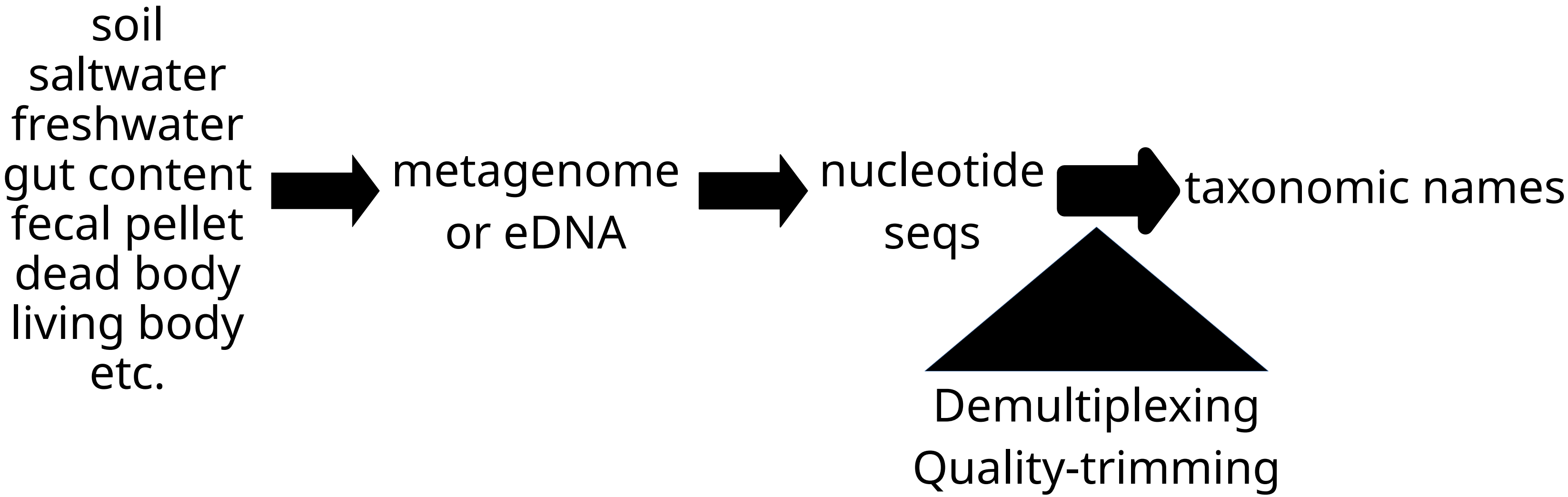
metagenome
or eDNA

nucleotide
seqs

taxonomic names

Demultiplexing

Computational processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

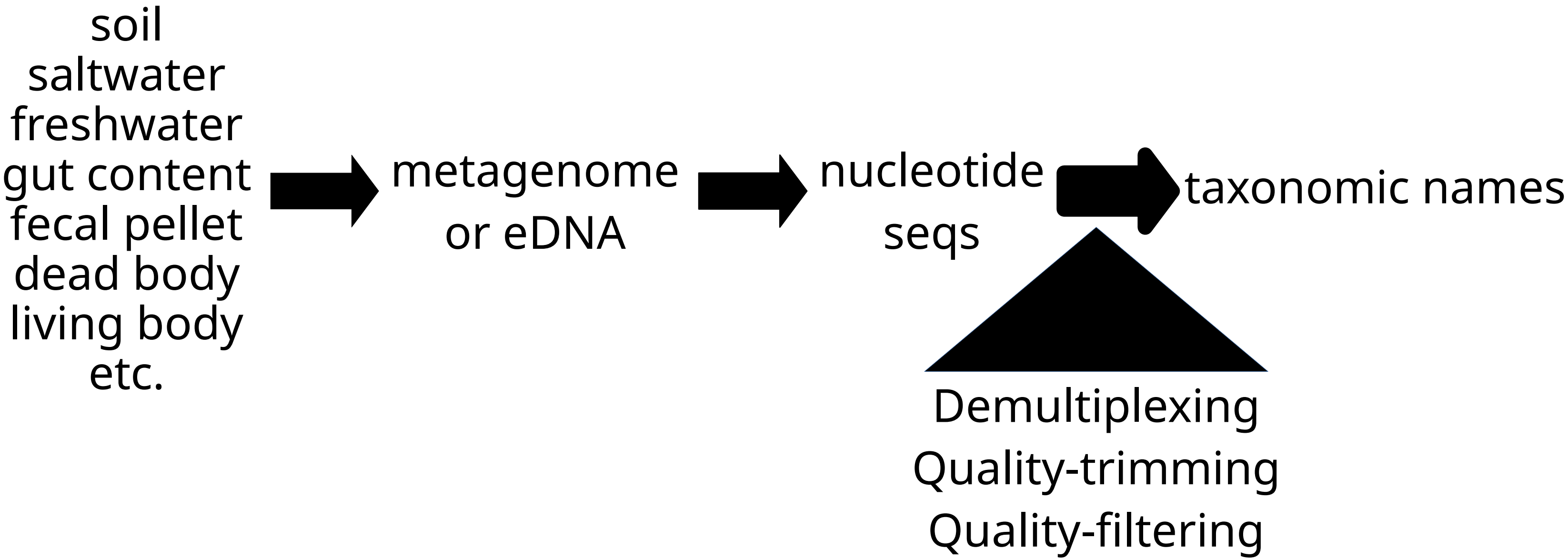
metagenome
or eDNA

nucleotide
seqs

taxonomic names

Demultiplexing
Quality-trimming

Computational processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

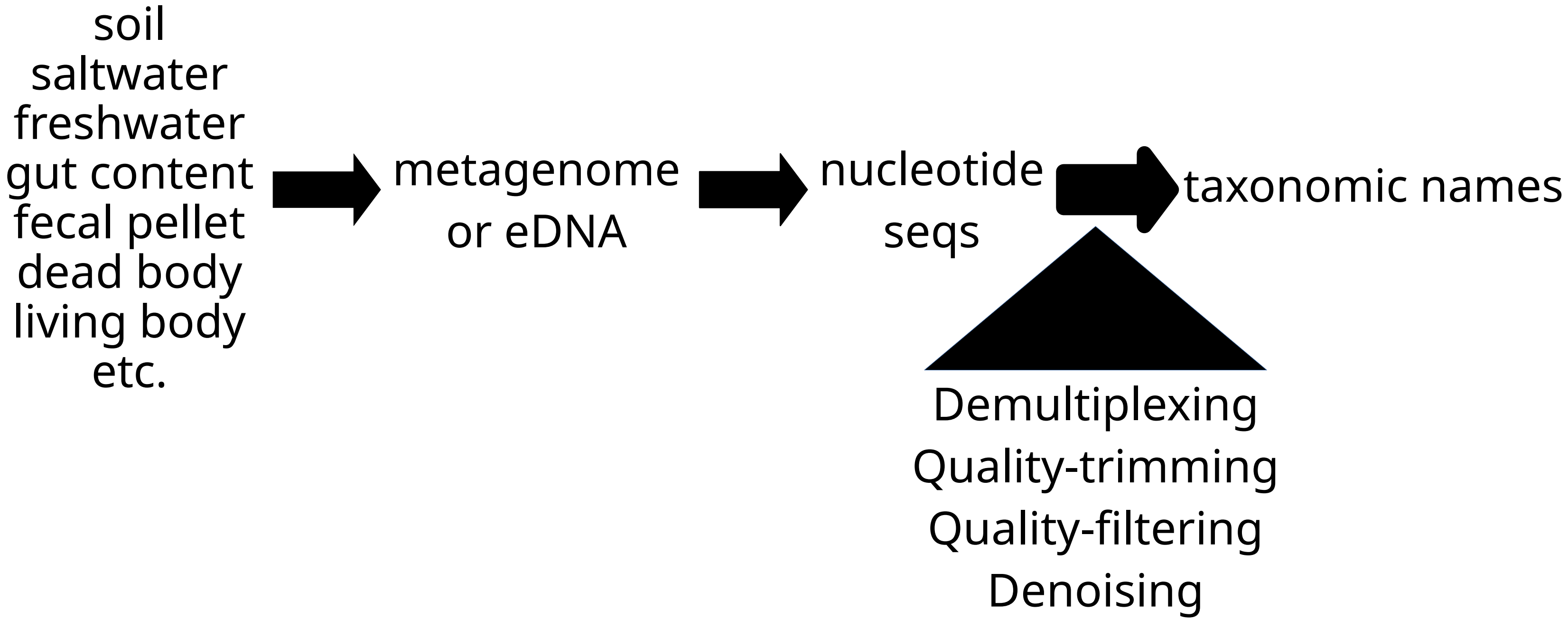
metagenome
or eDNA

nucleotide
seqs

taxonomic names

Demultiplexing
Quality-trimming
Quality-filtering

Computational processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

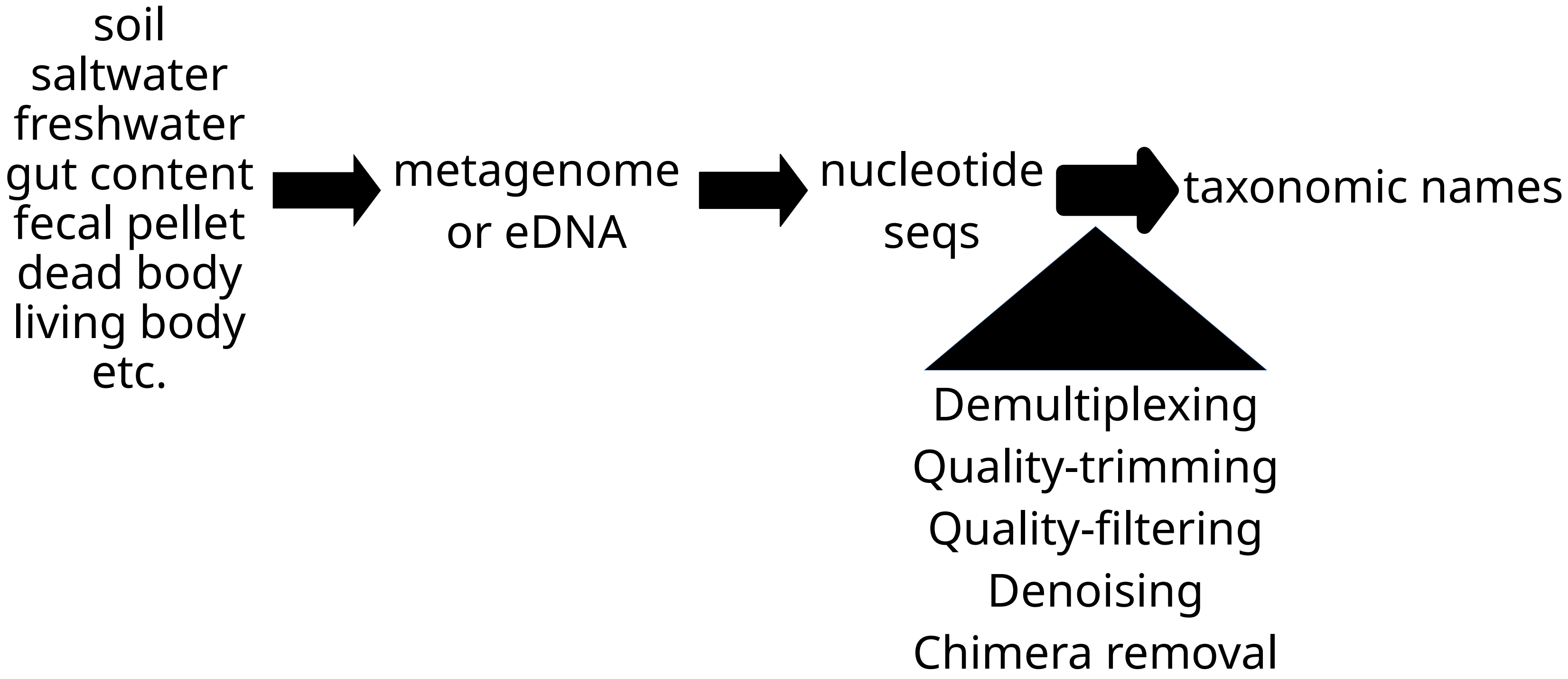
metagenome or eDNA

nucleotide seqs

taxonomic names

Demultiplexing
Quality-trimming
Quality-filtering
Denoising

Computational processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

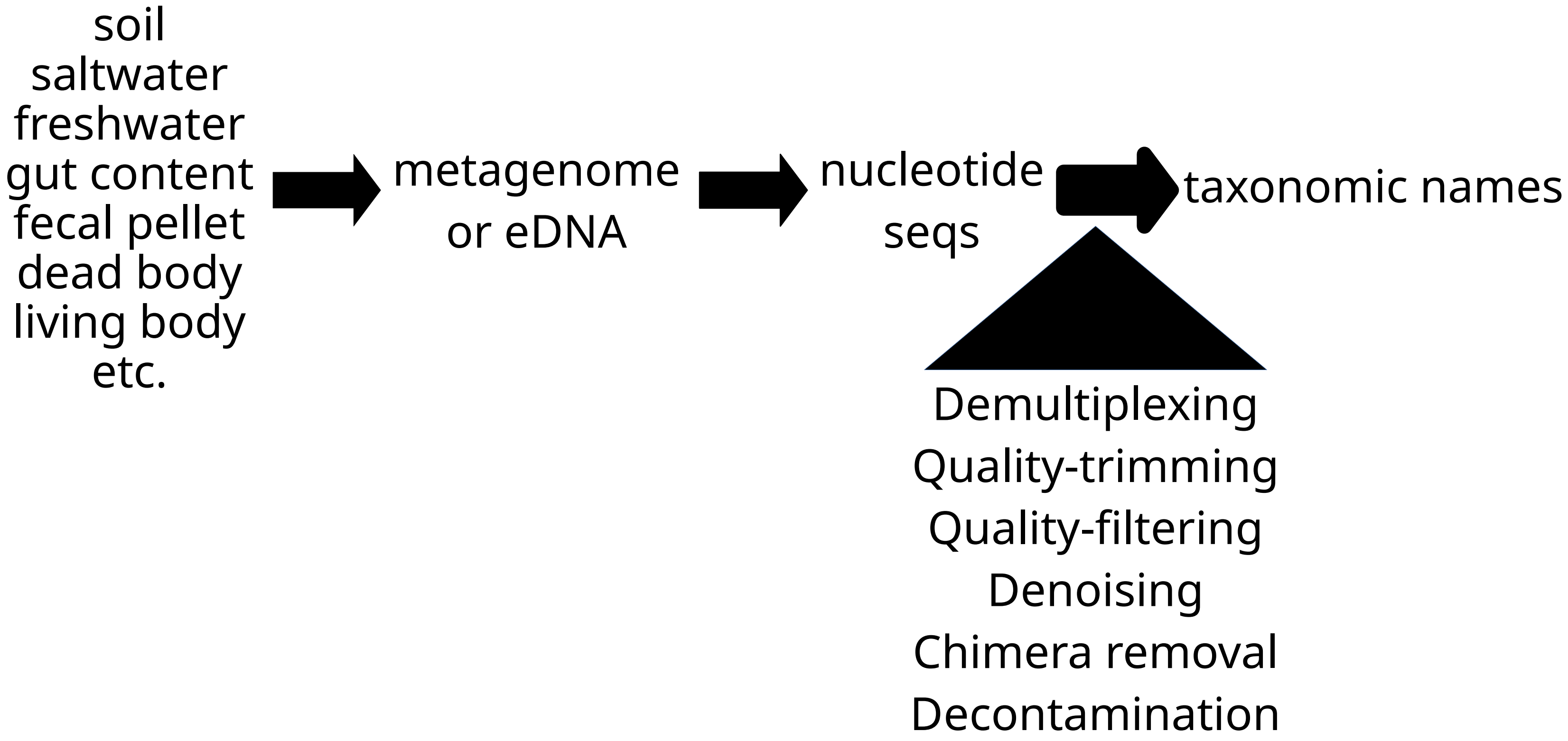
metagenome
or eDNA

nucleotide
seqs

taxonomic names

Demultiplexing
Quality-trimming
Quality-filtering
Denoising
Chimera removal

Computational processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

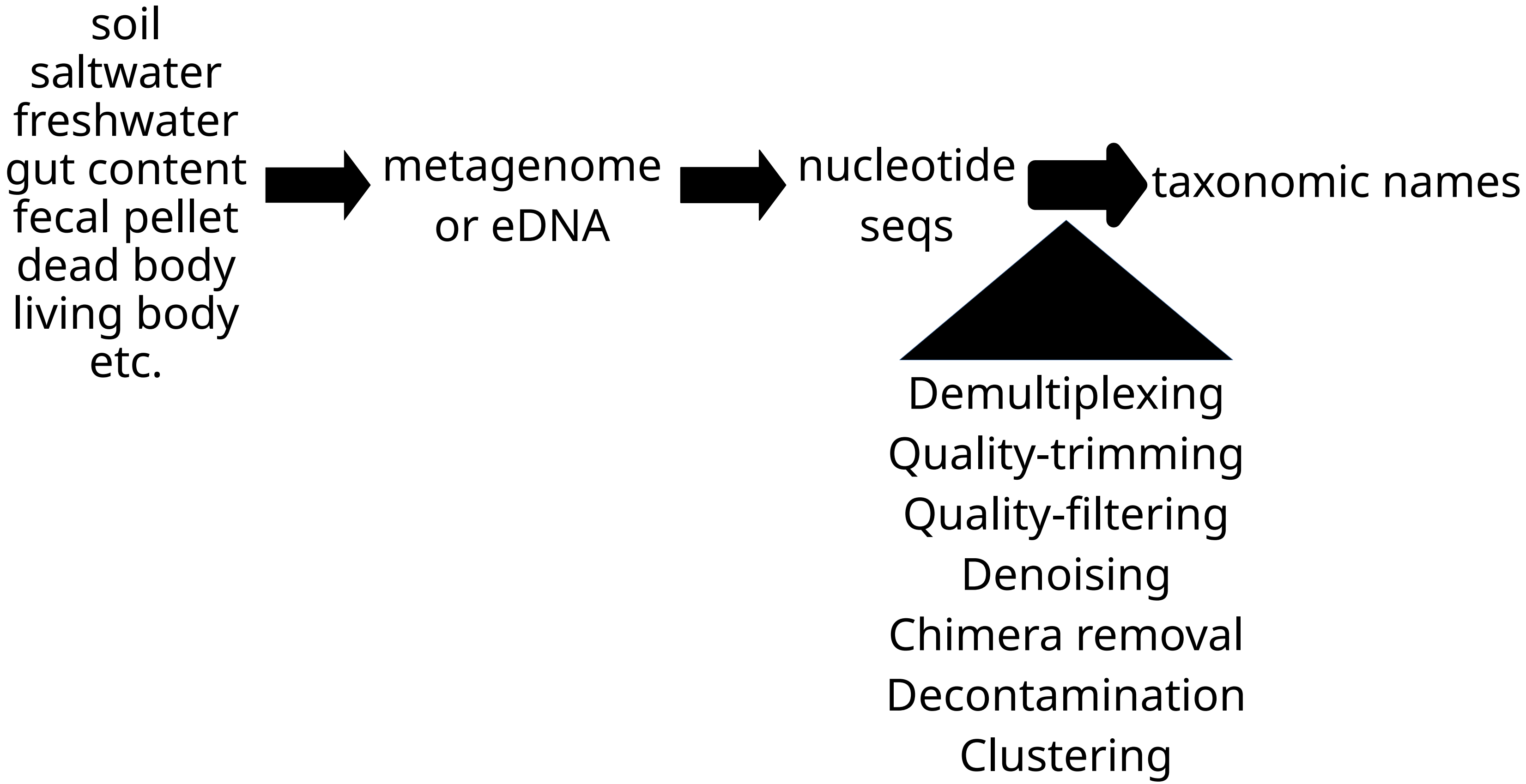
metagenome
or eDNA

nucleotide
seqs

taxonomic names

Demultiplexing
Quality-trimming
Quality-filtering
Denoising
Chimera removal
Decontamination

Computational processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

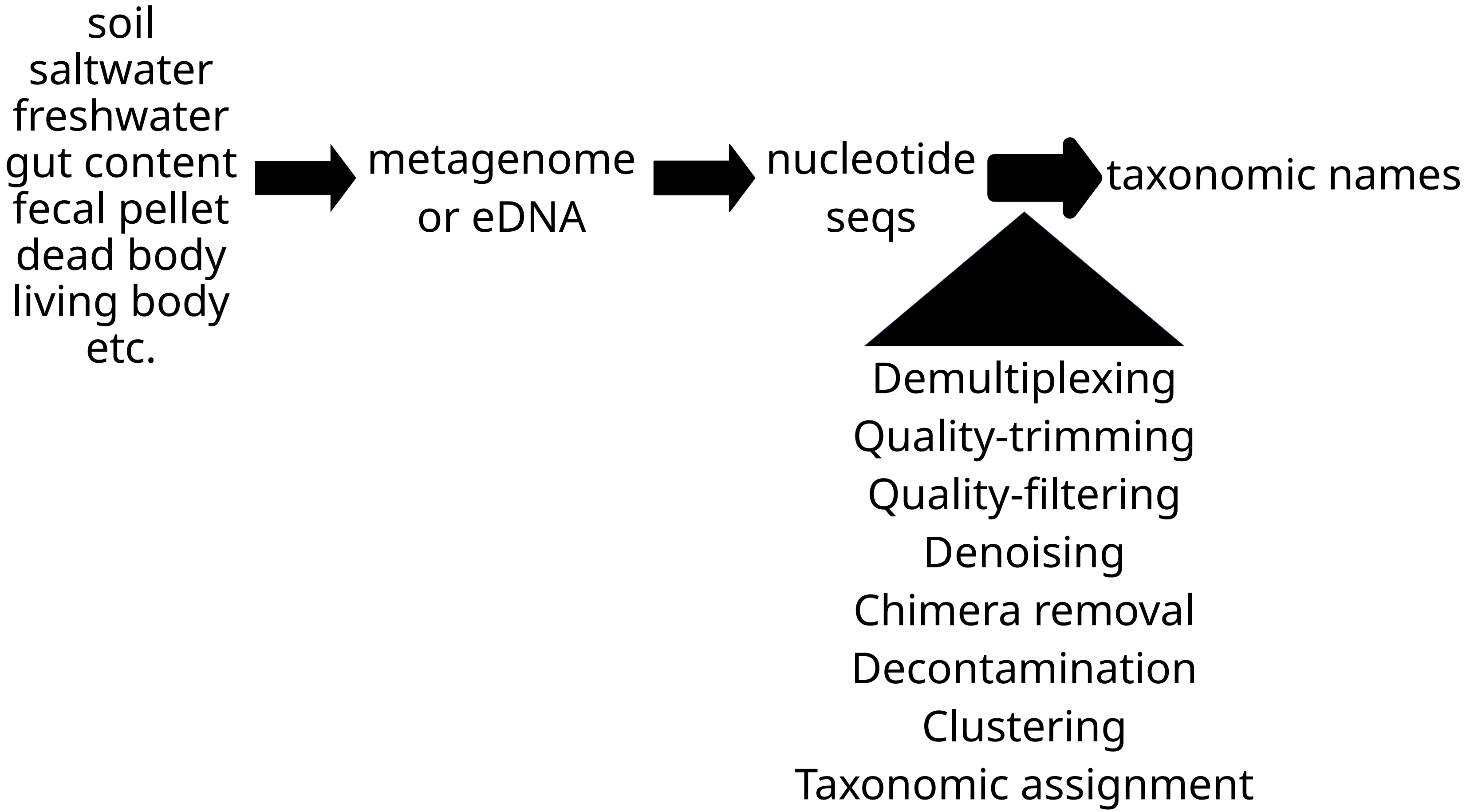
metagenome
or eDNA

nucleotide
seqs

taxonomic names

Demultiplexing
Quality-trimming
Quality-filtering
Denoising
Chimera removal
Decontamination
Clustering

Computational processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

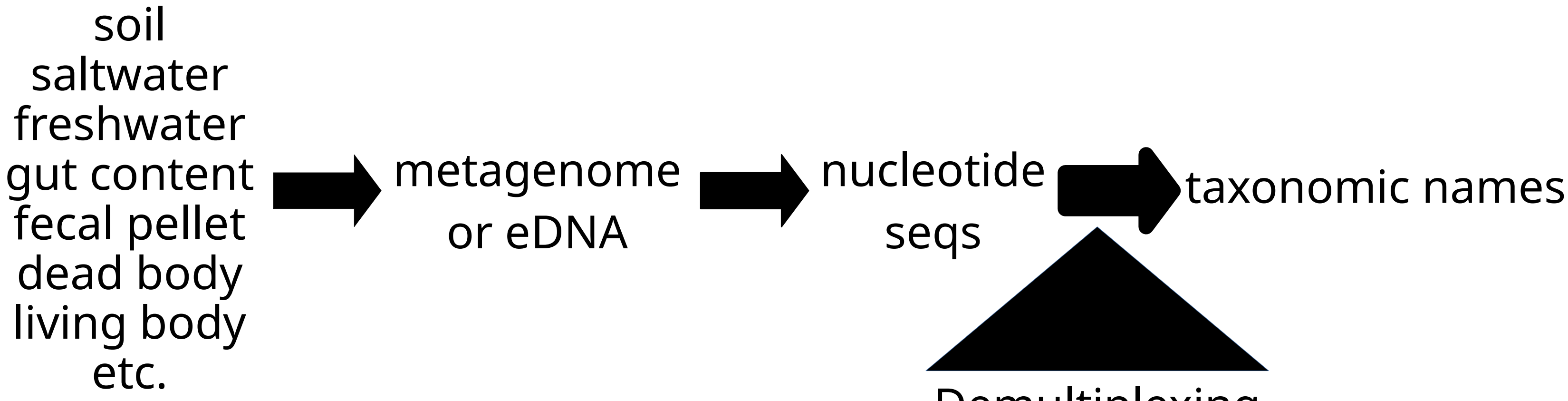
metagenome
or eDNA

nucleotide
seqs

taxonomic names

Demultiplexing
Quality-trimming
Quality-filtering
Denoising
Chimera removal
Decontamination
Clustering
Taxonomic assignment

Computational processes of metabarcoding



soil
saltwater
freshwater
gut content
fecal pellet
dead body
living body
etc.

metagenome or eDNA

nucleotide seqs

taxonomic names

Demultiplexing
Quality-trimming
Quality-filtering
Denoising
Chimera removal
Decontamination
Clustering
Taxonomic assignment

Claident

https://www.claident.org/

Single-end sequence data analysis in Claident

1. Demultiplexing by clsplitseq
2. Evaluate sequence quality by VSEARCH via clcalcfastqstatv
3. Quality-trimming&filtering by VSEARCH via clfilterseqv
4. Denoising by DADA2 via cldenoiseseqd
5. Removing chimeras by UCHIME3 via clremovechimev
6. Removing contaminants by clremovecontam
7. Additional clustering by VSEARCH via clclassseqv (Optional)
8. Assigning taxonomy by clmakecachedb,clidentseq,classigntax
9. Additional taxonomy processing by clmergeassign,clfillassign
10. Summarizing results by clsumclass,clsumtaxa

Non-overlapped paired-end sequence data analysis in Claident

1. Demultiplexing by clsplitseq
2. Evaluate sequence quality by VSEARCH via clcalcfastqstatv x2
3. Quality-trimming by VSEARCH via clfilterseqv x2
4. Joining pairs by VSEARCH via clconcatpairv
5. Quality-filtering by VSEARCH via clfilterseqv
6. Denoising by DADA2 via cldenoiseseqd
7. Removing chimeras by UCHIME3 via clremovechimev
8. Removing contaminants by clremovecontam
9. Additional clustering by VSEARCH via clclassseqv (Optional)
10. Dividing pairs by cldivseq
11. Assigning taxonomy by clmakecachedb,clidentseq,classigntax x2
12. Additional taxonomy processing by clmergeassign,clfillassign
13. Summarizing results by clsumclass,clsumtaxa

Overlapped paired-end sequence data analysis in Claident

1. Demultiplexing by clsplitseq
2. Concatenating pairs by VSEARCH via clconcatpairv
3. Quality-filtering by VSEARCH via clfilterseqv
4. Denoising by DADA2 via cldenoiseseqd
5. Removing chimeras by UCHIME3 via clremovechimev
6. Removing contaminants by clremovecontam
7. Additional clustering by VSEARCH via clclassseqv (Optional)
8. Assigning taxonomy by clmakecachedb,clidentseq,classigntax
9. Additional taxonomy processing by clmergeassign,clfillassign
10. Summarizing results by clsumclass,clsumtaxa

Quantitative overlapped paired-end sequence data analysis with internal standard sequences in Claident

1. Demultiplexing by clsplitseq
2. Concatenating pairs by VSEARCH via clconcatpairv
3. Quality-filtering by VSEARCH via clfilterseqv
4. Denoising by DADA2 via cldenoiseseqd
5. Removing chimeras by UCHIME3 via clremovechimev
6. Clustering internal standard seqs via clclusterstdv
7. Removing contaminants by clremovecontam
8. Additional clustering by VSEARCH via clclassseqv (Optional)
9. Assigning taxonomy by clmakecachedb,clidentseq,classigntax
10. Additional taxonomy processing by clmergeassign,clfillassign
11. Summarizing results by clsumclass,clsumtaxa

Analysis demonstration of quantitative overlapped paired-end data

Analysis demonstration of quantitative overlapped paired-end data

- Prerequisites to run Claident

Analysis demonstration of quantitative overlapped paired-end data

- Prerequisites to run Claident
    - Debian/Ubuntu/Linux Mint, RedHat/CentOS

Analysis demonstration of quantitative overlapped paired-end data

- Prerequisites to run Claident
  - Debian/Ubuntu/Linux Mint, RedHat/CentOS
  - Claident+BLASTDB+TaxonomyDB+UCHIMEDB

Analysis demonstration of quantitative overlapped paired-end data

- Prerequisites to run Claident
  - Debian/Ubuntu/Linux Mint, RedHat/CentOS
  - Claident+BLASTDB+TaxonomyDB+UCHIMEDB
  - Code from https://github.com/astanabe/ClaidentTutorial

Analysis demonstration of quantitative overlapped paired-end data

- Prerequisites to run Claident
  - Debian/Ubuntu/Linux Mint, RedHat/CentOS
  - Claident+BLASTDB+TaxonomyDB+UCHIMEDB
  - Code from https://github.com/astanabe/ClaidentTutorial
- Prerequisites to learn about analyses using Claident and R

Analysis demonstration of quantitative overlapped paired-end data

- Prerequisites to run Claident
  - Debian/Ubuntu/Linux Mint, RedHat/CentOS
  - Claident+BLASTDB+TaxonomyDB+UCHIMEDB
  - Code from https://github.com/astanabe/ClaidentTutorial
- Prerequisites to learn about analyses using Claident and R
  - Code from https://github.com/astanabe/ClaidentTutorial

Analysis demonstration of quantitative overlapped paired-end data

- Prerequisites to run Claident
  - Debian/Ubuntu/Linux Mint, RedHat/CentOS
  - Claident+BLASTDB+TaxonomyDB+UCHIMEDB
  - Code from https://github.com/astanabe/ClaidentTutorial
- Prerequisites to learn about analyses using Claident and R
  - Code from https://github.com/astanabe/ClaidentTutorial
    - This includes simulated data and all results

Chapter 0: Simulated data creation

Chapter 0: Simulated data creation

1. Download complete mitogenome sequences of bony fishes

Chapter 0: Simulated data creation

1. Download complete mitogenome sequences of bony fishes
2. Extract 12S rRNA region

Chapter 0: Simulated data creation

1. Download complete mitogenome sequences of bony fishes
2. Extract 12S rRNA region
3. Run *in silico* PCR using MiFish-U primers by ecoPCR and obtain amplicons

Chapter 0: Simulated data creation

1. Download complete mitogenome sequences of bony fishes
2. Extract 12S rRNA region
3. Run *in silico* PCR using MiFish-U primers by ecoPCR and obtain amplicons
4. Cluster amplicon seqs and pick representative seqs

Chapter 0: Simulated data creation

1. Download complete mitogenome sequences of bony fishes
2. Extract 12S rRNA region
3. Run *in silico* PCR using MiFish-U primers by ecoPCR and obtain amplicons
4. Cluster amplicon seqs and pick representative seqs
5. Randomly pick 50 seqs from all representative seqs (1st sample)

Chapter 0: Simulated data creation

1. Download complete mitogenome sequences of bony fishes
2. Extract 12S rRNA region
3. Run *in silico* PCR using MiFish-U primers by ecoPCR and obtain amplicons
4. Cluster amplicon seqs and pick representative seqs
5. Randomly pick 50 seqs from all representative seqs (1st sample)
6. Randomly pick 40 seqs from previous sample and randomly pick 10 seqs from all repseqs except for previous sample seqs (2nd-20th sample)

Chapter 0: Simulated data creation

1. Download complete mitogenome sequences of bony fishes
2. Extract 12S rRNA region
3. Run *in silico* PCR using MiFish-U primers by ecoPCR and obtain amplicons
4. Cluster amplicon seqs and pick representative seqs
5. Randomly pick 50 seqs from all representative seqs (1st sample)
6. Randomly pick 40 seqs from previous sample and randomly pick 10 seqs from all repseqs except for previous sample seqs (2nd-20th sample)
7. Pick all sequences from all 1st-20th samples for blank (1st-4th blank)

Chapter 0: Simulated data creation

1. Download complete mitogenome sequences of bony fishes
2. Extract 12S rRNA region
3. Run *in silico* PCR using MiFish-U primers by ecoPCR and obtain amplicons
4. Cluster amplicon seqs and pick representative seqs
5. Randomly pick 50 seqs from all representative seqs (1st sample)
6. Randomly pick 40 seqs from previous sample and randomly pick 10 seqs from all repseqs except for previous sample seqs (2nd-20th sample)
7. Pick all sequences from all 1st-20th samples for blank (1st-4th blank)
8. Generate 250 paired-end seqs for each picked seqs by ART for samples

Chapter 0: Simulated data creation

1. Download complete mitogenome sequences of bony fishes
2. Extract 12S rRNA region
3. Run *in silico* PCR using MiFish-U primers by ecoPCR and obtain amplicons
4. Cluster amplicon seqs and pick representative seqs
5. Randomly pick 50 seqs from all representative seqs (1st sample)
6. Randomly pick 40 seqs from previous sample and randomly pick 10 seqs from all repseqs except for previous sample seqs (2nd-20th sample)
7. Pick all sequences from all 1st-20th samples for blank (1st-4th blank)
8. Generate 250 paired-end seqs for each picked seqs by ART for samples
9. Generate 25 paired-end seqs for each picked seqs by ART for blanks

Chapter 0: Simulated data creation

1. Download complete mitogenome sequences of bony fishes
2. Extract 12S rRNA region
3. Run *in silico* PCR using MiFish-U primers by ecoPCR and obtain amplicons
4. Cluster amplicon seqs and pick representative seqs
5. Randomly pick 50 seqs from all representative seqs (1st sample)
6. Randomly pick 40 seqs from previous sample and randomly pick 10 seqs from all repseqs except for previous sample seqs (2nd-20th sample)
7. Pick all sequences from all 1st-20th samples for blank (1st-4th blank)
8. Generate 250 paired-end seqs for each picked seqs by ART for samples
9. Generate 25 paired-end seqs for each picked seqs by ART for blanks
10. Generate 100-800 seqs for standards by ART and add to samples and blanks

Chapter 0: Simulated data creation

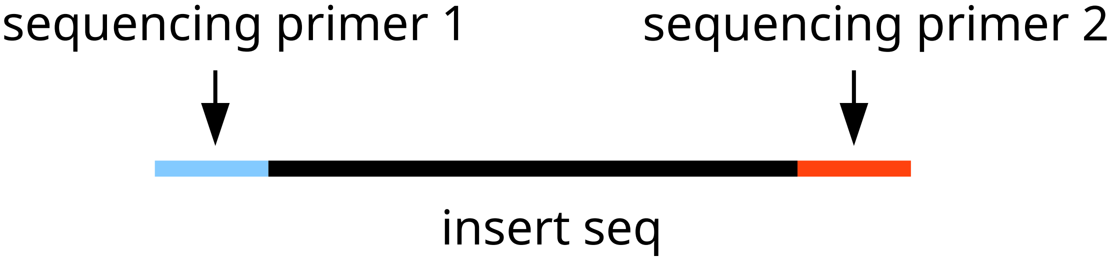
1. Download complete mitogenome sequences of bony fishes
2. Extract 12S rRNA region
3. Run *in silico* PCR using MiFish-U primers by ecoPCR and obtain amplicons
4. Cluster amplicon seqs and pick representative seqs
5. Randomly pick 50 seqs from all representative seqs (1st sample)
6. Randomly pick 40 seqs from previous sample and randomly pick 10 seqs from all repseqs except for previous sample seqs (2nd-20th sample)
7. Pick all sequences from all 1st-20th samples for blank (1st-4th blank)
8. Generate 250 paired-end seqs for each picked seqs by ART for samples
9. Generate 25 paired-end seqs for each picked seqs by ART for blanks
10. Generate 100-800 seqs for standards by ART and add to samples and blanks
11. Generate dual index seqs based on given fasta files

Interlude: The structure of Illumina dual-index library

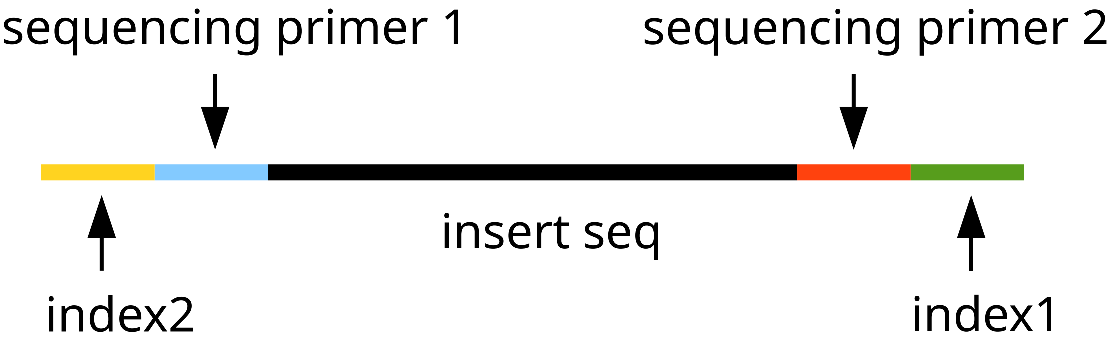
Interlude: The structure of Illumina dual-index library

insert seq

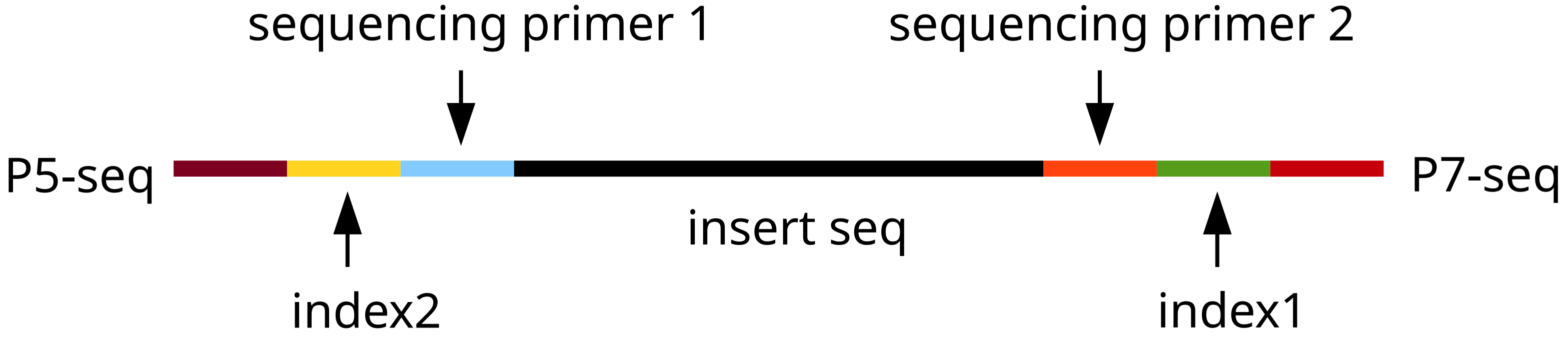
Interlude: The structure of Illumina dual-index library



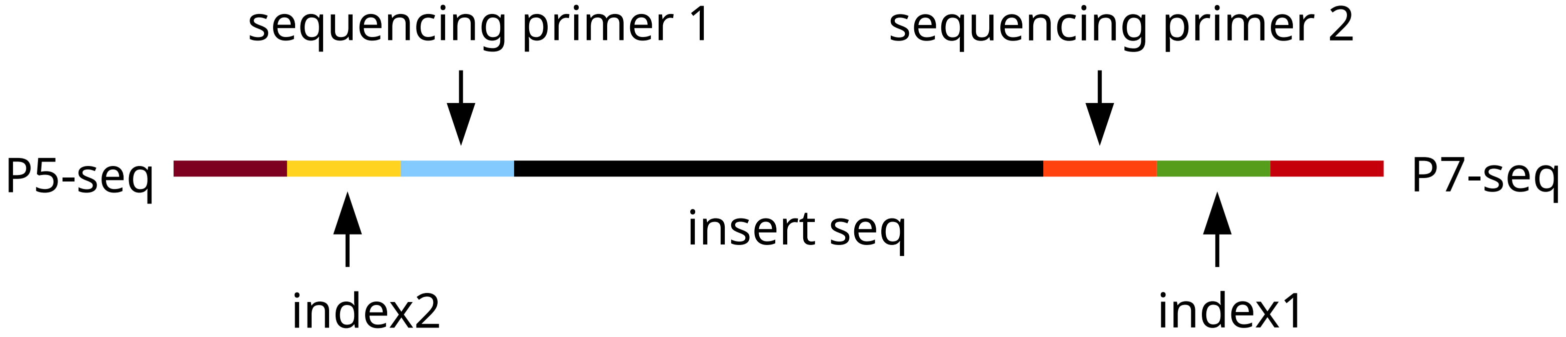
Interlude: The structure of Illumina dual-index library



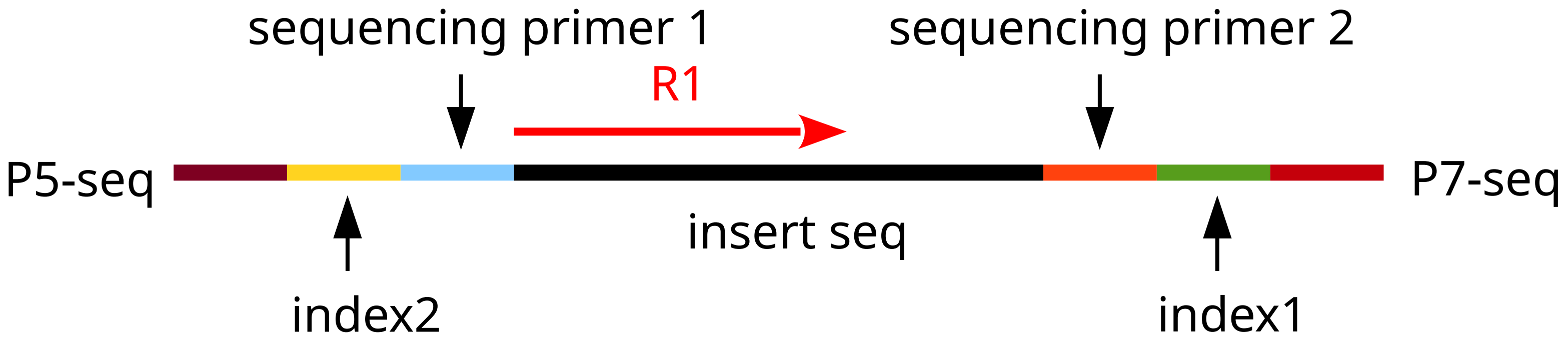
Interlude: The structure of Illumina dual-index library



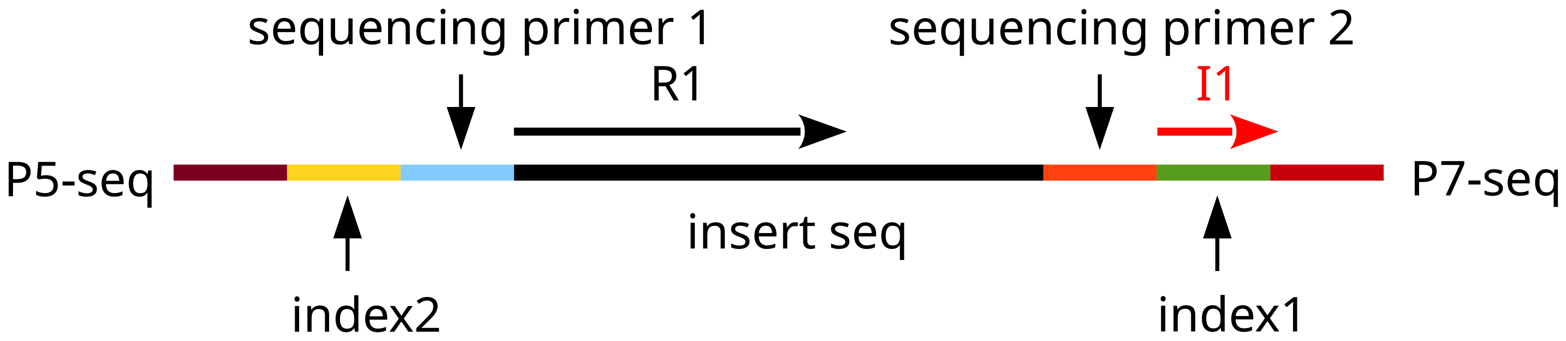
Interlude: The read order and strand of Illumina dual-index library



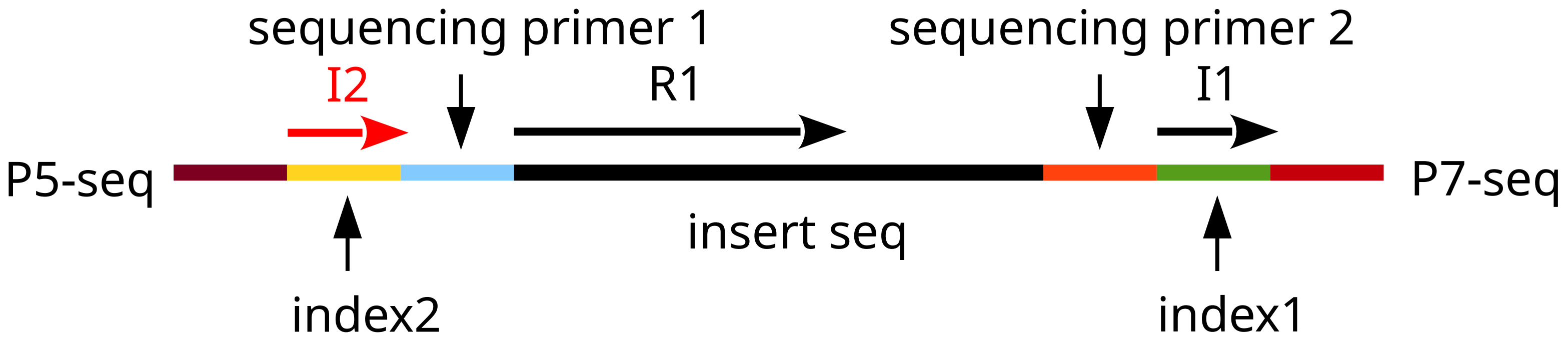
Interlude: The read order and strand of Illumina dual-index library



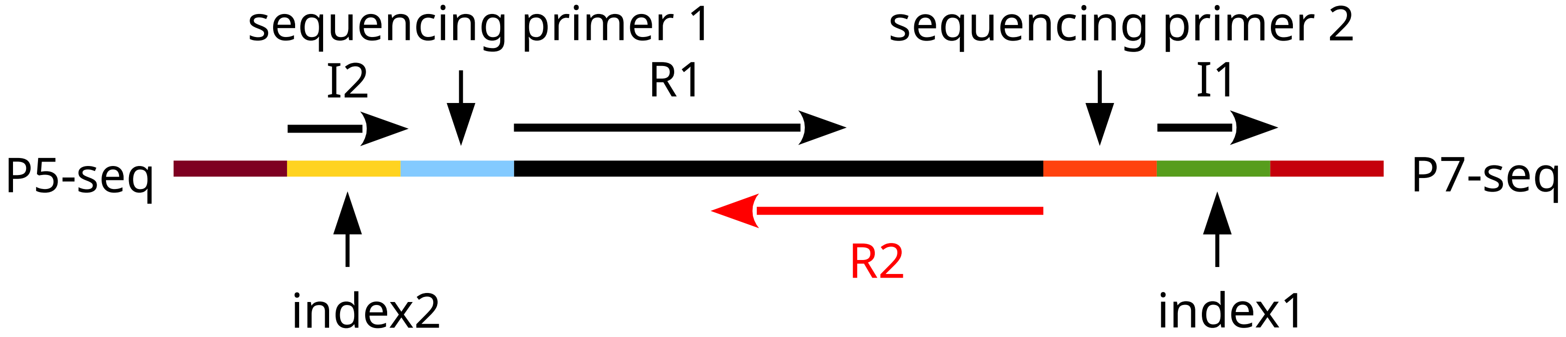
Interlude: The read order and strand of Illumina dual-index library



Interlude: The read order and strand of Illumina dual-index library



Interlude: The read order and strand of Illumina dual-index library

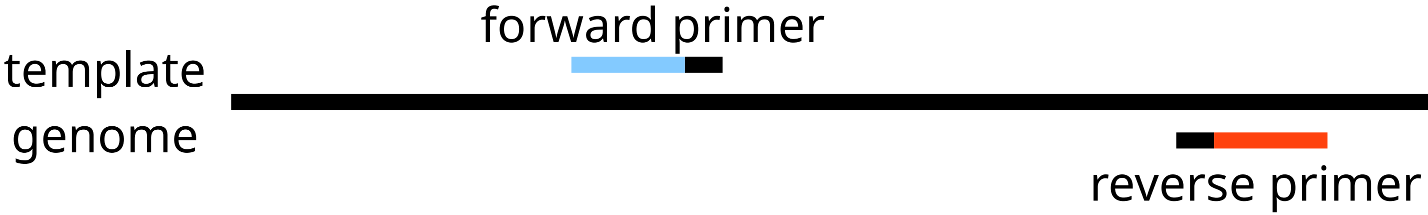


Interlude: Preparation of Illumina dual-index library

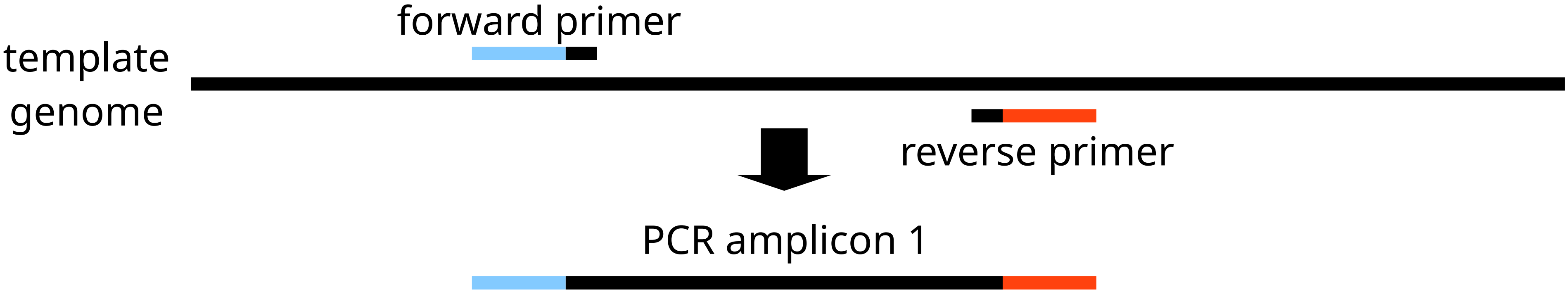
Interlude: Preparation of Illumina dual-index library

template genome

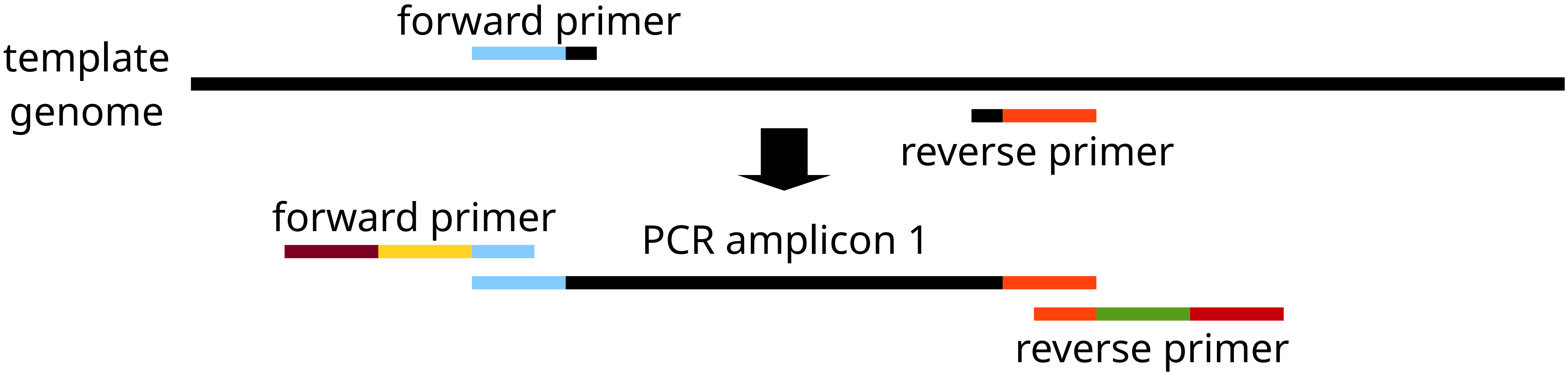
Interlude: Preparation of Illumina dual-index library



Interlude: Preparation of Illumina dual-index library



Interlude: Preparation of Illumina dual-index library



forward primer

template genome

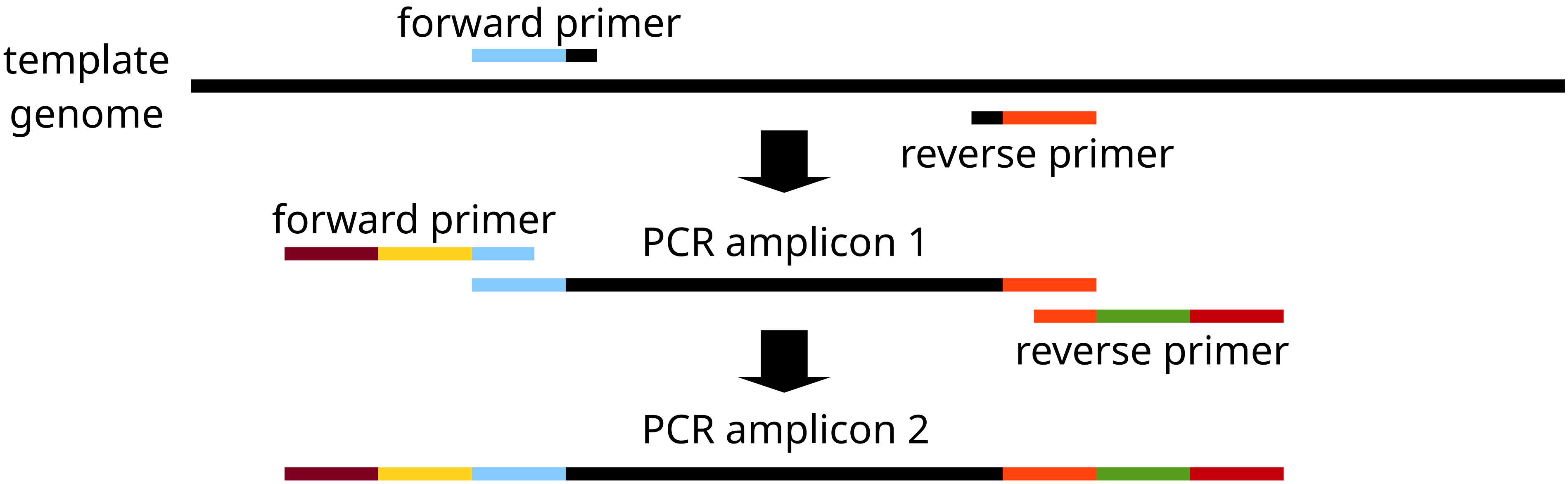
reverse primer

forward primer

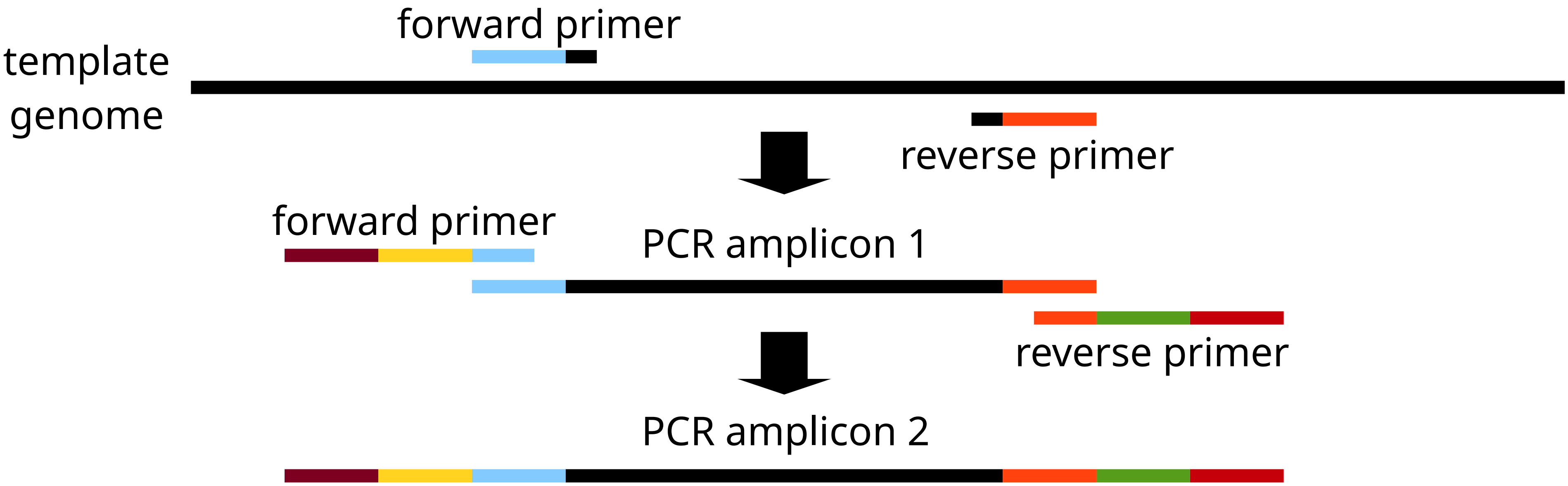
PCR amplicon 1

reverse primer

Interlude: Preparation of Illumina dual-index library



Interlude: Preparation of Illumina dual-index library



By 8 forward index primers and 12 reverse index primers,
96 samples can be distinguished (combinatorial dual-indexing).

Interlude: Dual-index design of simulated data

Interlude: Dual-index design of simulated data

reverse index (index1)

TTGCAGGT	Sample01	Sample07	not used	not used
CAAGGAAC	Sample02	Sample08	not used	not used
AGATCTGG	Sample03	Sample09	not used	not used
TCACACTT	Sample04	Sample10	not used	not used
GATCATGG	Sample05	Sample11	not used	not used
AGACATGA	Sample06	Sample12	not used	not used
GTGAGTTG	not used	not used	Sample13	Sample19
AGTCTGTT	not used	not used	Sample14	Sample20
AACCAACC	not used	not used	Sample15	Blank01
AGTGTGCA	not used	not used	Sample16	Blank02
CATGTCGA	not used	not used	Sample17	Blank03
CGAGACTT	not used	not used	Sample18	Blank04
	AACCTCTC	**GTGACTCT**	**GATCACCA**	**CTTCACAT**

forward index (index2)

Interlude: Dual-index design of simulated data

reverse index (index1)

TTGCAGGT	Sample01	Sample07	not used	not used
CAAGGAAC	Sample02	Sample08	not used	not used
AGATCTGG	Sample03	Sample09	not used	not used
TCACACTT	Sample04	Sample10	not used	not used
GATCATGG	Sample05	Sample11	not used	not used
AGACATGA	Sample06	Sample12	not used	not used
GTGAGTTG	not used	not used	Sample13	Sample19
AGTCTGTT	not used	not used	Sample14	Sample20
AACCAACC	not used	not used	Sample15	Blank01
AGTGTGCA	not used	not used	Sample16	Blank02
CATGTCGA	not used	not used	Sample17	Blank03
CGAGACTT	not used	not used	Sample18	Blank04
	AACCTCTC	**GTGACTCT**	**GATCACCA**	**CTTCACAT**

forward index (index2)

Chapter 1: Overview of raw sequence files

Chapter 1: Overview of raw sequence files

- Undemultiplexed files
- Undemultiplexed_R1*
- Undemultiplexed_R2*
- Undemultiplexed_I1*
- Undemultiplexed_I2*
- Undemultiplexed_wSTD_R1*
- Undemultiplexed_wSTD_R2*
- Undemultiplexed_wSTD_I1*
- Undemultiplexed_wSTD_I2*

in 01_RawSequences

Chapter 1: Overview of raw sequence files

- Undemultiplexed files
- Undemultiplexed_R1*
- Undemultiplexed_R2*
- Undemultiplexed_I1*
- Undemultiplexed_I2*
- Undemultiplexed_wSTD_R1*
- Undemultiplexed_wSTD_R2*
- Undemultiplexed_wSTD_I1*
- Undemultiplexed_wSTD_I2*

in 01_RawSequences

- Already demultiplexed files
- Sample*
- Blank*
- Sample_wSTD*
- Blank_wSTD*

in 01_RawSequences

Chapter 2: Demultiplexing

Chapter 2: Demultiplexing

- Inputs
- Undemultiplexed_wSTD_R1*
- Undemultiplexed_wSTD_I1*
- Undemultiplexed_wSTD_I2*
- Undemultiplexed_wSTD_R2*

in 01_RawSequences

Chapter 2: Demultiplexing

- Inputs
- Undemultiplexed_wSTD_R1*
- Undemultiplexed_wSTD_I1*
- Undemultiplexed_wSTD_I2*
- Undemultiplexed_wSTD_R2*

in 01_RawSequences

- index1.fasta
- index2.fasta
- forwardprimer.fasta
- reverseprimer.fasta

in top directory

Chapter 2: Demultiplexing

- Inputs
- Undemultiplexed_wSTD_R1*
- Undemultiplexed_wSTD_I1*
- Undemultiplexed_wSTD_I2*
- Undemultiplexed_wSTD_R2*

in 01_RawSequences

- index1.fasta
- index2.fasta
- forwardprimer.fasta
- reverseprimer.fasta

in top directory

- Outputs
- ClaidentTutorial__*__MiFish.forward.fastq.xz
- ClaidentTutorial__*__MiFish.reverse.fastq.xz
  - Sample*
  - Blank*
  - NNNNNNNN+NNNNNNNN

in PairedEnd_wSTD_02a_DemultiplexedSequences

Chapter 2: Demultiplexing

Launch Terminal

Chapter 3: Concatenating pairs

Chapter 3: Concatenating pairs

- Inputs
- ClaidentTutorial__*__MiFish.forward.fastq.xz
- ClaidentTutorial__*__MiFish.reverse.fastq.xz
    - Sample*
    - Blank*
    - NNNNNNNN+NNNNNNNN

in PairedEnd_wSTD_02a_DemultiplexedSequences

Chapter 3: Concatenating pairs

- Inputs
- ClaidentTutorial__*__MiFish.forward.fastq.xz
- ClaidentTutorial__*__MiFish.reverse.fastq.xz
  - Sample*
  - Blank*
  - NNNNNNNN+NNNNNNNN

in PairedEnd_wSTD_02a_DemultiplexedSequences

- Outputs
- ClaidentTutorial__*__MiFish.fastq.xz
  - Sample*
  - Blank*
  - NNNNNNNN+NNNNNNNN

in OverlappedPairedEnd_wSTD_03_ConcatenatedSequences

Chapter 3: Concatenating pairs

Switch to Terminal

Chapter 4: Quality-filtering

Chapter 4: Quality-filtering

- Inputs
- ClaidentTutorial__*__MiFish.fastq.xz
  - Sample*
  - Blank*
  - NNNNNNNN+NNNNNNNN

in OverlappedPairedEnd_wSTD_03_ConcatenatedSequences

Chapter 4: Quality-filtering

- Inputs
- ClaidentTutorial__*__MiFish.fastq.xz
  - Sample*
  - Blank*
  - NNNNNNNN+NNNNNNNN

  in OverlappedPairedEnd_wSTD_03_ConcatenatedSequences

- Outputs
- ClaidentTutorial__*__MiFish.fastq.xz
  - Sample*
  - Blank*
  - NNNNNNNN+NNNNNNNN

  in OverlappedPairedEnd_wSTD_04_FilteredSequences

Chapter 4: Quality-filtering

Switch to Terminal

Chapter 5: Denoising

Chapter 5: Denoising

- Inputs
- ClaidentTutorial__*__MiFish.fastq.xz
  - Sample*
  - Blank*
  - NNNNNNNN+NNNNNNNN

    in OverlappedPairedEnd_wSTD_04_FilteredSequences

Chapter 5: Denoising

- Inputs
- ClaidentTutorial__*__MiFish.fastq.xz
  - Sample*
  - Blank*
  - NNNNNNNN+NNNNNNNN

in OverlappedPairedEnd_wSTD_04_FilteredSequences

- Outputs
- denoised.fasta
- denoised.otu.gz
- denoised.tsv
- plotErrors.pdf
- runDADA2.R

in OverlappedPairedEnd_wSTD_05_DenoisedSequences

Chapter 5: Denoising

Switch to Terminal

Interlude: Methods in DADA2

See also Callahan et al. (2016) https://doi.org/10.1038/nmeth.3869

Interlude: Methods in DADA2

	observed number
ACCTCTCGATATCGAGATGAGGCT	10000
ACCTCTTGATATCGAGATGAGGCT	10
ACCTCTCGAAATCGAGATGAGGCT	7
ACCTCTGGATATCGAGATGAGGCT	200

See also Callahan et al. (2016) https://doi.org/10.1038/nmeth.3869

Interlude: Methods in DADA2

	observed number
ACCTCTCGATATCGAGATGAGGCT	10000
ACCTCTTGATATCGAGATGAGGCT	10
ACCTCTCGAAATCGAGATGAGGCT	7
ACCTCTGGATATCGAGATGAGGCT	200

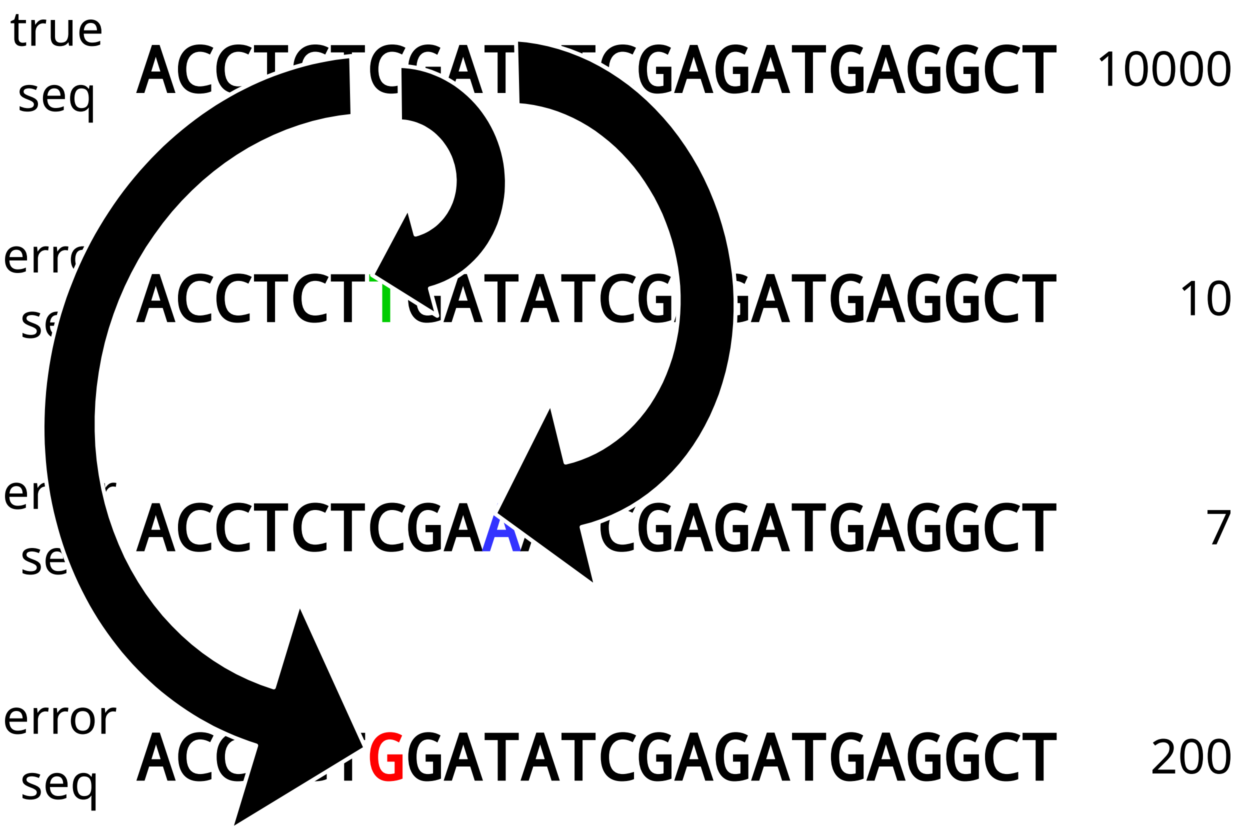
See also Callahan et al. (2016) https://doi.org/10.1038/nmeth.3869

Interlude: Methods in DADA2

		observed number
true seq	ACCTCTCGATATCGAGATGAGGCT	10000
error seq	ACCTCTTGATATCGAGATGAGGCT	10
error seq	ACCTCTCGAAATCGAGATGAGGCT	7
error seq	ACCTCTGGATATCGAGATGAGGCT	200

See also Callahan et al. (2016) https://doi.org/10.1038/nmeth.3869

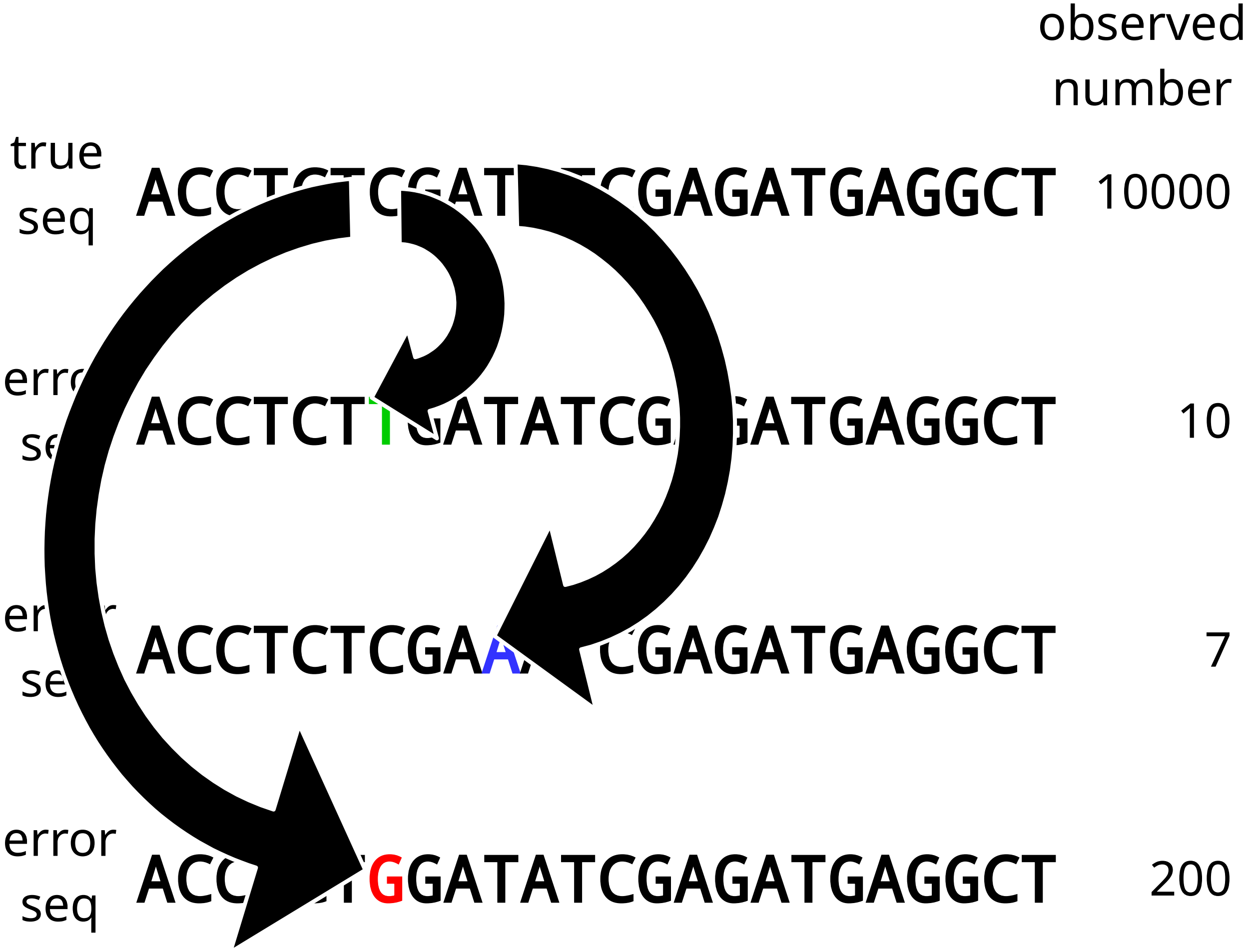
Interlude: Methods in DADA2



		observed number
true seq	ACCTCTCGATATCGAGATGAGGCT	10000
error seq	ACCTCTTGATATCGAGATGAGGCT	10
error seq	ACCTCTCGAAATCGAGATGAGGCT	7
error seq	ACCTCTGGATATCGAGATGAGGCT	200

See also Callahan et al. (2016) https://doi.org/10.1038/nmeth.3869

Interlude: Methods in DADA2



observed number

true seq ACCTCTCGATATCGAGATGAGGCT 10000

error seq ACCTCTTGATATCGAGATGAGGCT 10

error seq ACCTCTCGAAATCGAGATGAGGCT 7

error seq ACCTCTGGATATCGAGATGAGGCT 200

Error rate matrix

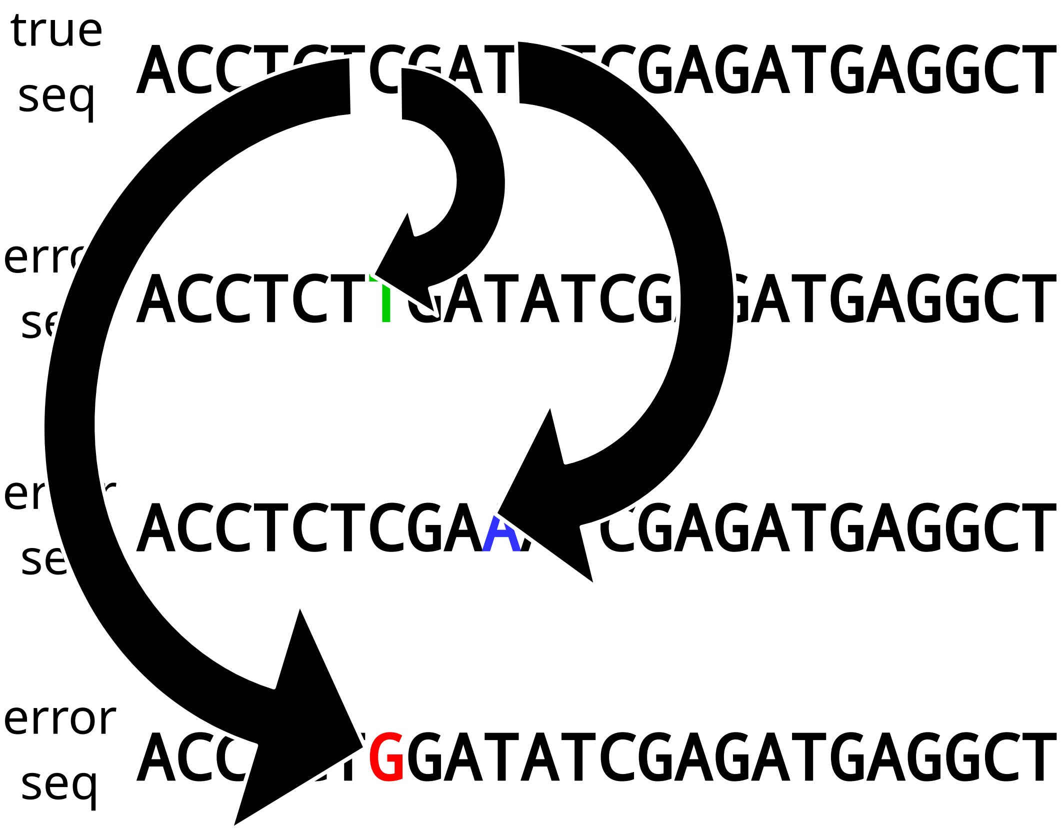
to

from

	A	C	G	T
A	R_{AA}	R_{AC}	R_{AG}	R_{AT}
C	R_{CA}	R_{CC}	R_{CG}	R_{CT}
G	R_{GA}	R_{GC}	R_{GG}	R_{GT}
T	R_{TA}	R_{TC}	R_{TG}	R_{TT}

See also Callahan et al. (2016) https://doi.org/10.1038/nmeth.3869

Interlude: Methods in DADA2



		observed number	abundance p-value
true seq	**ACCTCTCGATATCGAGATGAGGCT**	10000	NA
error seq	**ACCTCTTGATATCGAGATGAGGCT**	10	0.3
error seq	**ACCTCTCGAAATCGAGATGAGGCT**	7	0.5
error seq	**ACCTCTGGATATCGAGATGAGGCT**	200	0.001

Error rate matrix

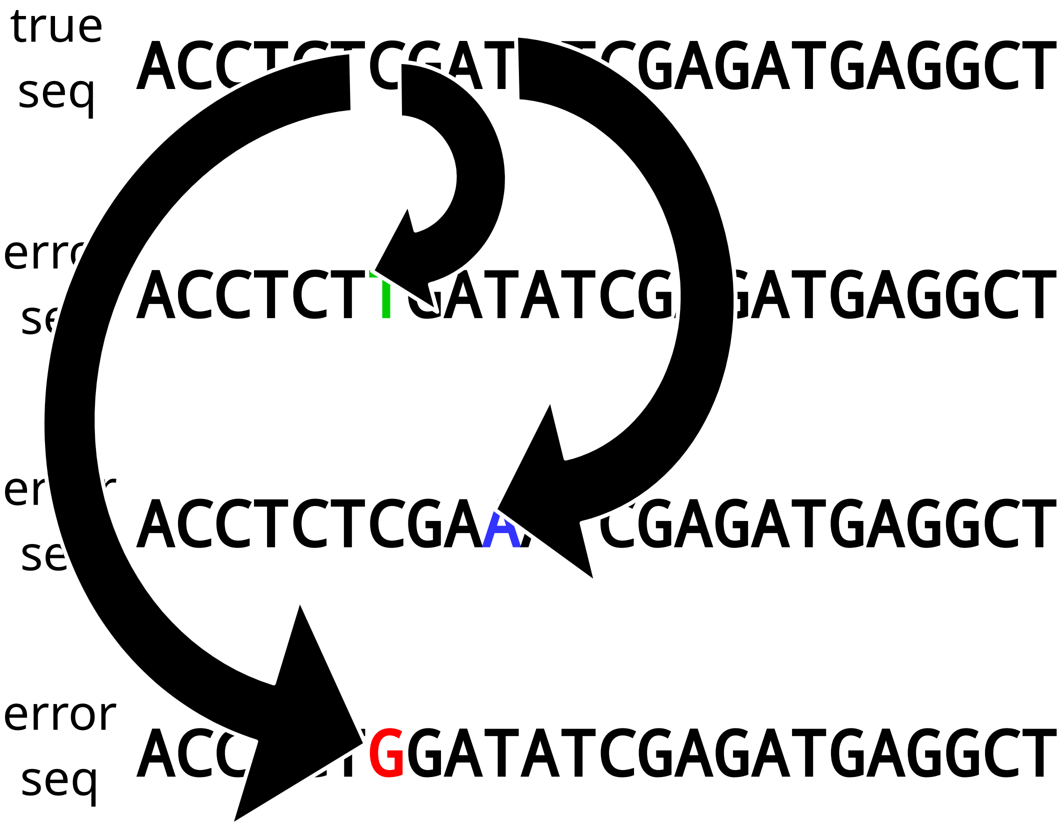
to

from

	A	C	G	T
A	R_{AA}	R_{AC}	R_{AG}	R_{AT}
C	R_{CA}	R_{CC}	R_{CG}	R_{CT}
G	R_{GA}	R_{GC}	R_{GG}	R_{GT}
T	R_{TA}	R_{TC}	R_{TG}	R_{TT}

See also Callahan et al. (2016) https://doi.org/10.1038/nmeth.3869

Interlude: Methods in DADA2



observed number	abundance p-value
10000	NA
10	0.3
7	0.5
200	0.001

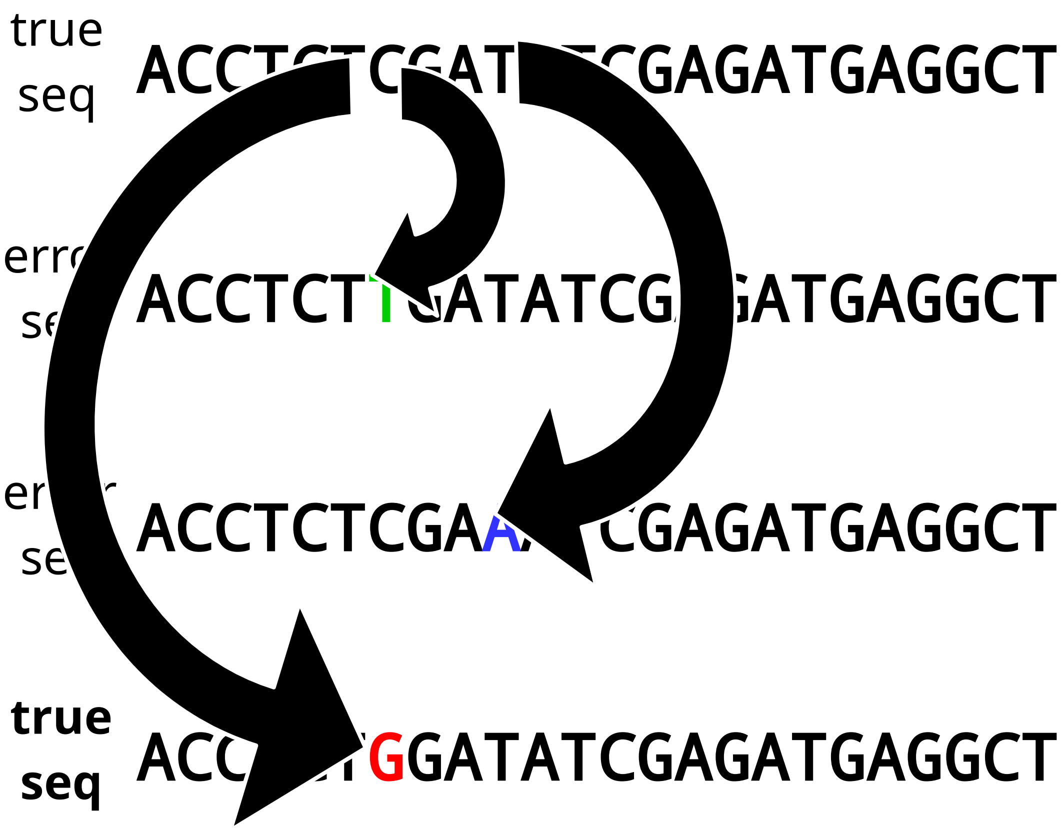
Error rate matrix

from \ to	A	C	G	T
A	R_{AA}	R_{AC}	R_{AG}	R_{AT}
C	R_{CA}	R_{CC}	R_{CG}	R_{CT}
G	R_{GA}	R_{GC}	R_{GG}	R_{GT}
T	R_{TA}	R_{TC}	R_{TG}	R_{TT}

Observed number cannot be explain as error → 0.001

See also Callahan et al. (2016) https://doi.org/10.1038/nmeth.3869

Interlude: Methods in DADA2



		observed number	abundance p-value
true seq	ACCTCTCGATATCGAGATGAGGCT	10000	NA
error seq	ACCTCTTGATATCGAGATGAGGCT	10	0.3
error seq	ACCTCTCGAAATCGAGATGAGGCT	7	0.5
true seq	ACCTCTGGATATCGAGATGAGGCT	200	0.001

Error rate matrix

to

from

	A	C	G	T
A	R_{AA}	R_{AC}	R_{AG}	R_{AT}
C	R_{CA}	R_{CC}	R_{CG}	R_{CT}
G	R_{GA}	R_{GC}	R_{GG}	R_{GT}
T	R_{TA}	R_{TC}	R_{TG}	R_{TT}

Observed number cannot be explain as error

See also Callahan et al. (2016) https://doi.org/10.1038/nmeth.3869

Chapter 6: Chimera removal

Chapter 6: Chimera removal

- Inputs
- denoised.fasta
- denoised.otu.gz

  in OverlappedPairedEnd_wSTD_05_DenoisedSequences

Chapter 6: Chimera removal

- Inputs
- denoised.fasta
- denoised.otu.gz

in OverlappedPairedEnd_wSTD_05_DenoisedSequences

- Outputs
- nonchimeras.fasta
- nonchimeras.otu.gz
- nonchimeras.tsv
- *_borderline.fasta
- *_chimeras.fasta
- *_nonchimeras.fasta
- *_uchimealns.txt
- *_uchimeout.txt

in OverlappedPairedEnd_wSTD_06_NonchimericSequences

Chapter 6: Chimera removal

Switch to Terminal

Chapter 7: Clustering internal standard sequences

Chapter 7: Clustering internal standard sequences

- Inputs
- nonchimeras.fasta
- nonchimeras.otu.gz

in OverlappedPairedEnd_wSTD_06_NonchimericSequences

Chapter 7: Clustering internal standard sequences

- Inputs
- nonchimeras.fasta
- nonchimeras.otu.gz

in OverlappedPairedEnd_wSTD_06_NonchimericSequences

- Outputs
- stdclustered.fasta
- stdclustered.otu.gz
- stdclustered.tsv

in OverlappedPairedEnd_wSTD_07_STDClusteredSequences

Chapter 7: Clustering internal standard sequences

Switch to Terminal

Chapter 8: Removing index-hopped sequences

Chapter 8: Removing index-hopped sequences

- Inputs
- stdclustered.fasta
- stdclustered.otu.gz

  in OverlappedPairedEnd_wSTD_07_STDClusteredSequences

Chapter 8: Removing index-hopped sequences

- Inputs
- stdclustered.fasta
- stdclustered.otu.gz

in OverlappedPairedEnd_wSTD_07_STDClusteredSequences

- index1.fasta
- index2.fasta

in top directory

Chapter 8: Removing index-hopped sequences

- Inputs
- stdclustered.fasta
- stdclustered.otu.gz

  in OverlappedPairedEnd_wSTD_07_STDClusteredSequences
- index1.fasta
- index2.fasta

  in top directory

- Outputs
- decontaminated.fasta
- decontaminated.otu.gz
- decontaminated.tsv

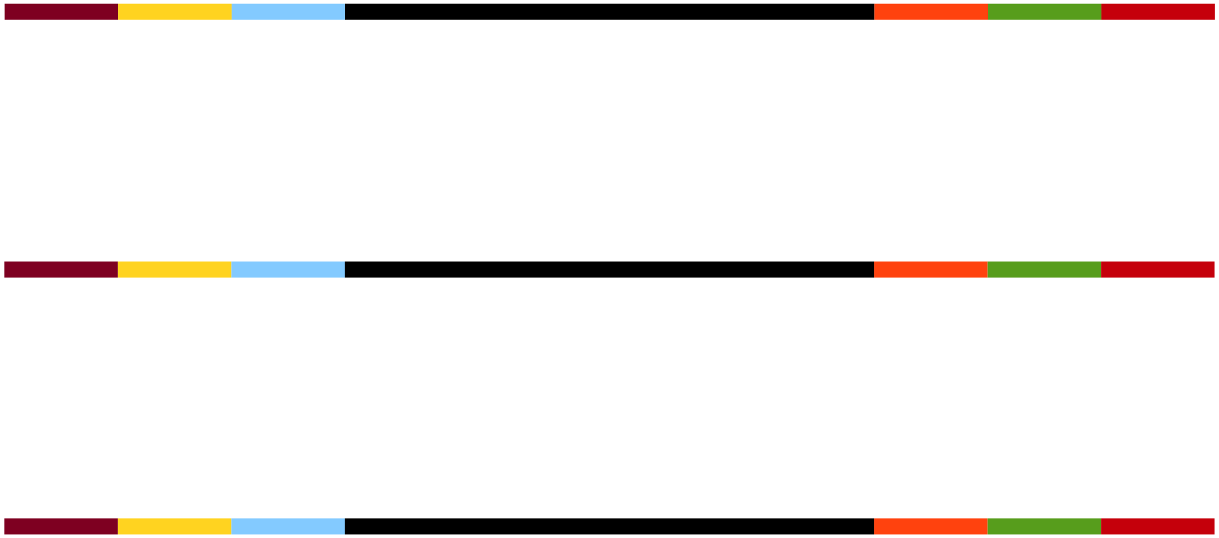
  in OverlappedPairedEnd_wSTD_08_NonhoppedSequences

Chapter 8: Removing index-hopped sequences

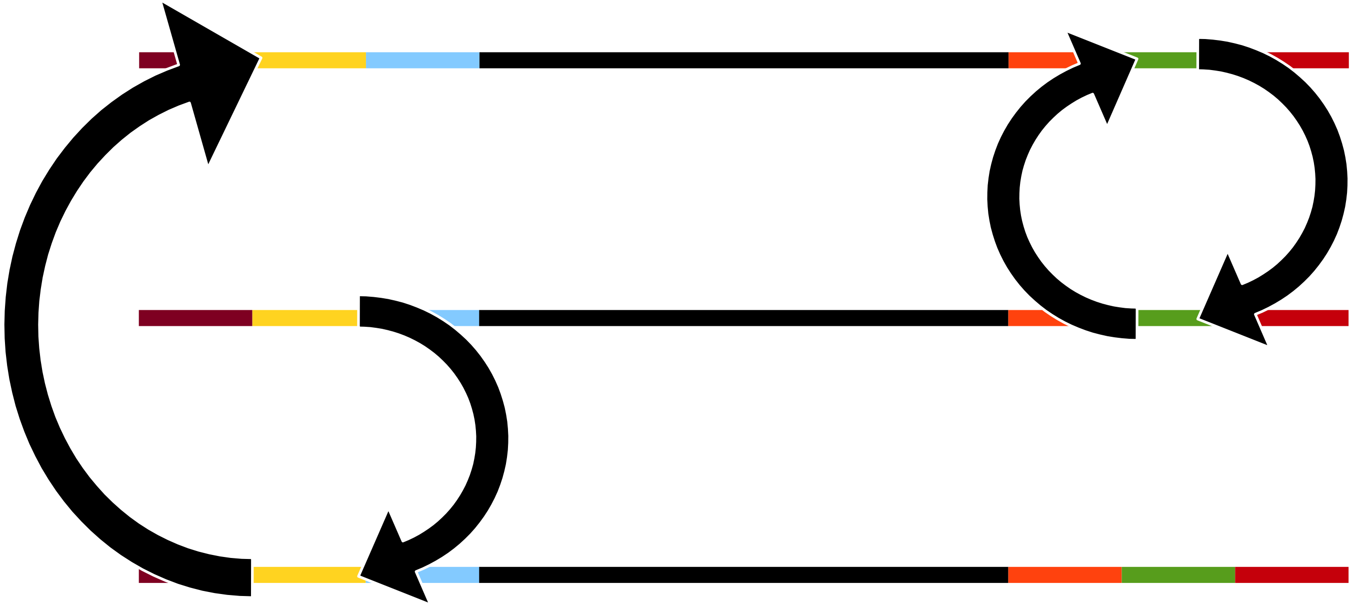
Switch to Terminal

Interlude: Index can hop into another amplicon within a flowcell!

Interlude: Index can hop into another amplicon within a flowcell!

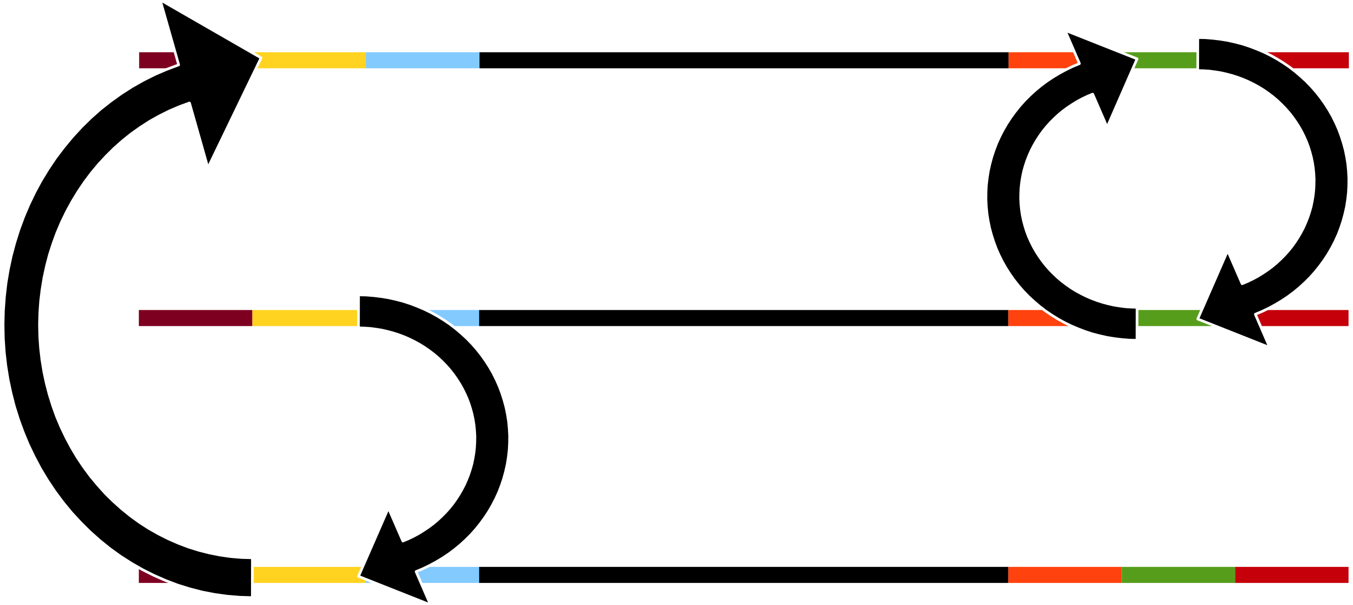


Interlude: Index can hop into another amplicon within a flowcell!



Interlude: Index can hop into another amplicon within a flowcell!

Index-hopping potentially causes sequence misassignments!
Especially in newer models! OMG!



Interlude: Detecting index-hopping using unused index combinations

reverse index (index1)

TTGCAGGT	Sample01	Sample07	not used	not used
CAAGGAAC	Sample02	Sample08	not used	not used
AGATCTGG	Sample03	Sample09	not used	not used
TCACACTT	Sample04	Sample10	not used	not used
GATCATGG	Sample05	Sample11	not used	not used
AGACATGA	Sample06	Sample12	not used	not used
GTGAGTTG	not used	not used	Sample13	Sample19
AGTCTGTT	not used	not used	Sample14	Sample20
AACCAACC	not used	not used	Sample15	Blank01
AGTGTGCA	not used	not used	Sample16	Blank02
CATGTCGA	not used	not used	Sample17	Blank03
CGAGACTT	not used	not used	Sample18	Blank04
	AACCTCTC	**GTGACTCT**	**GATCACCA**	**CTTCACAT**

forward index (index2)

See also Esling et al. (2015) https://doi.org/10.1093/nar/gkv107

Interlude: Detecting index-hopping using unused index combinations

reverse index (index1)

TTGCAGGT	Sample01	Sample07	not used	not used
CAAGGAAC	Sample02	Sample08	not used	not used
AGATCTGG	Sample03	Sample09	not used	not used
TCACACTT	Sample04	Sample10	not used	not used
GATCATGG	Sample05	Sample11	not used	not used
AGACATGA	Sample06	Sample12	not used	not used
GTGAGTTG	not used	not used	Sample13	Sample19
AGTCTGTT	not used	not used	Sample14	Sample20
AACCAACC	not used	not used	Sample15	Blank01
AGTGTGCA	not used	not used	Sample16	Blank02
CATGTCGA	not used	not used	Sample17	Blank03
CGAGACTT	not used	not used	Sample18	Blank04
	AACCTCTC	**GTGACTCT**	**GATCACCA**	**CTTCACAT**

forward index (index2)

1. Count abundances

See also Esling et al. (2015) https://doi.org/10.1093/nar/gkv107

Interlude: Detecting index-hopping using unused index combinations

reverse index (index1)

TTGCAGGT	Sample01	Sample07	not used	not used
CAAGGAAC	Sample02	Sample08	not used	not used
AGATCTGG	Sample03	Sample09	not used	not used
TCACACTT	Sample04	Sample10	not used	not used
GATCATGG	Sample05	Sample11	not used	not used
AGACATGA	Sample06	Sample12	not used	not used
GTGAGTTG	not used	not used	Sample13	Sample19
AGTCTGTT	not used	not used	Sample14	Sample20
AACCAACC	not used	not used	Sample15	Blank01
AGTGTGCA	not used	not used	Sample16	Blank02
CATGTCGA	not used	not used	Sample17	Blank03
CGAGACTT	not used	not used	Sample18	Blank04
	AACCTCTC	**GTGACTCT**	**GATCACCA**	**CTTCACAT**

forward index (index2)

1. Count abundances
2. Collect abundances of a sample + "not used"

See also Esling et al. (2015) https://doi.org/10.1093/nar/gkv107

Interlude: Detecting index-hopping using unused index combinations

reverse index (index1)

TTGCAGGT	Sample01	Sample07	not used	not used
CAAGGAAC	Sample02	Sample08	not used	not used
AGATCTGG	Sample03	Sample09	not used	not used
TCACACTT	Sample04	Sample10	not used	not used
GATCATGG	Sample05	Sample11	not used	not used
AGACATGA	Sample06	Sample12	not used	not used
GTGAGTTG	not used	not used	Sample13	Sample19
AGTCTGTT	not used	not used	Sample14	Sample20
AACCAACC	not used	not used	Sample15	Blank01
AGTGTGCA	not used	not used	Sample16	Blank02
CATGTCGA	not used	not used	Sample17	Blank03
CGAGACTT	not used	not used	Sample18	Blank04
	AACCTCTC	**GTGACTCT**	**GATCACCA**	**CTTCACAT**

forward index (index2)

1. Count abundances
2. Collect abundances of a sample + "not used"
3. Test whether sample abundance is outlier or not

See also Esling et al. (2015) https://doi.org/10.1093/nar/gkv107

Interlude: Detecting index-hopping using unused index combinations

reverse index (index1)

TTGCAGGT	Sample01	Sample07	not used	not used
CAAGGAAC	Sample02	Sample08	not used	not used
AGATCTGG	Sample03	Sample09	not used	not used
TCACACTT	Sample04	Sample10	not used	not used
GATCATGG	Sample05	Sample11	not used	not used
AGACATGA	Sample06	Sample12	not used	not used
GTGAGTTG	not used	not used	Sample13	Sample19
AGTCTGTT	not used	not used	Sample14	Sample20
AACCAACC	not used	not used	Sample15	Blank01
AGTGTGCA	not used	not used	Sample16	Blank02
CATGTCGA	not used	not used	Sample17	Blank03
CGAGACTT	not used	not used	Sample18	Blank04
	AACCTCTC	**GTGACTCT**	**GATCACCA**	**CTTCACAT**

forward index (index2)

1. Count abundances
2. Collect abundances of a sample + "not used"
3. Test whether sample abundance is outlier or not
4. If it's not outlier, it's determined as hopped

See also Esling et al. (2015) https://doi.org/10.1093/nar/gkv107

Chapter 9: Removing contaminant sequences

Chapter 9: Removing contaminant sequences

- Inputs
- decontaminated.fasta
- decontaminated.otu.gz

in OverlappedPairedEnd_wSTD_08_NonhoppedSequences

Chapter 9: Removing contaminant sequences

- Inputs
- decontaminated.fasta
- decontaminated.otu.gz

  in OverlappedPairedEnd_wSTD_08_NonhoppedSequences
- blanklist.txt

  in top directory

Chapter 9: Removing contaminant sequences

- Inputs
- decontaminated.fasta
- decontaminated.otu.gz

  in OverlappedPairedEnd_wSTD_08_NonhoppedSequences
- blanklist.txt

  in top directory

- Outputs
- decontaminated.fasta
- decontaminated.otu.gz
- decontaminated.tsv

  in OverlappedPairedEnd_wSTD_09_DecontaminatedSequences

Chapter 9: Removing contaminant sequences

Switch to Terminal

Interlude: Detecting contaminants using blank samples

Interlude: Detecting contaminants using blank samples

1. Count abundances

See also Esling et al. (2015) https://doi.org/10.1093/nar/gkv107

Interlude: Detecting contaminants using blank samples

1. Count abundances
2. Collect abundances of a sample + associated blanks

See also Esling et al. (2015) https://doi.org/10.1093/nar/gkv107

Interlude: Detecting contaminants using blank samples

1. Count abundances
2. Collect abundances of a sample + associated blanks
3. Test whether sample abundance is outlier or not

See also Esling et al. (2015) https://doi.org/10.1093/nar/gkv107

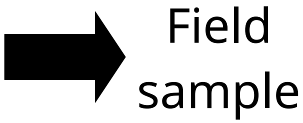
Interlude: Detecting contaminants using blank samples

1. Count abundances
2. Collect abundances of a sample + associated blanks
3. Test whether sample abundance is outlier or not
4. If it's not outlier, it's determined as contaminant

See also Esling et al. (2015) https://doi.org/10.1093/nar/gkv107

Interlude: Multistep contamination and multistep decontamination

Interlude: Multistep contamination and multistep decontamination



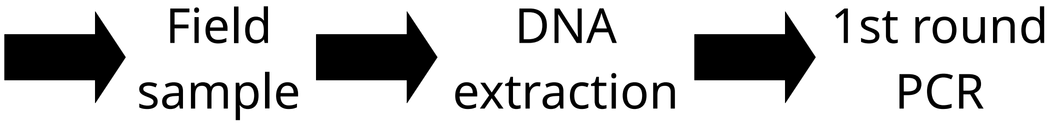
Field sample

Interlude: Multistep contamination and multistep decontamination

Field sample

DNA extraction

Interlude: Multistep contamination and multistep decontamination

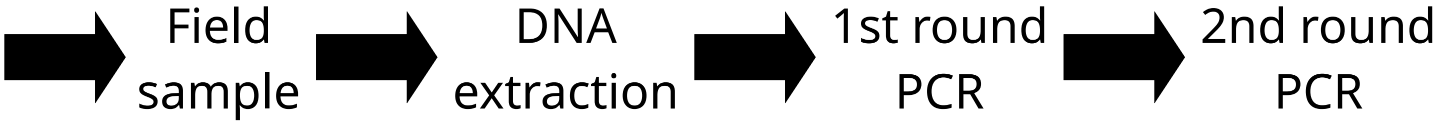


Field sample

DNA extraction

1st round PCR

Interlude: Multistep contamination and multistep decontamination



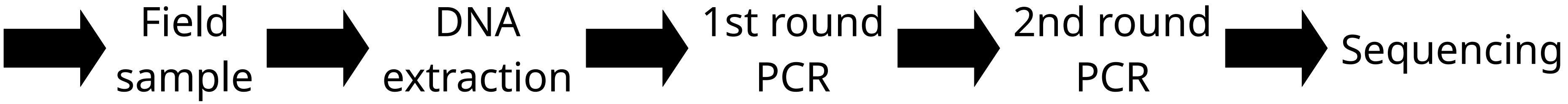
Field sample

DNA extraction

1st round PCR

2nd round PCR

Interlude: Multistep contamination and multistep decontamination



Field sample

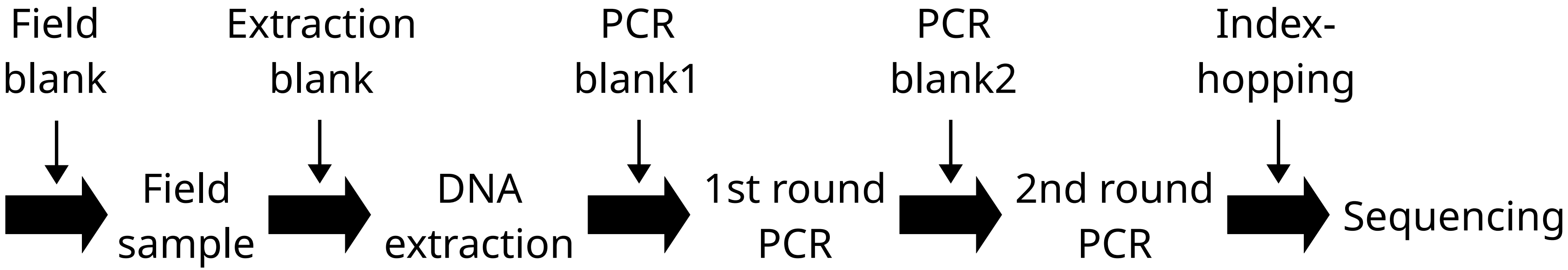
DNA extraction

1st round PCR

2nd round PCR

Sequencing

Interlude: Multistep contamination and multistep decontamination



Field blank

Extraction blank

PCR blank1

PCR blank2

Index-hopping

Field sample

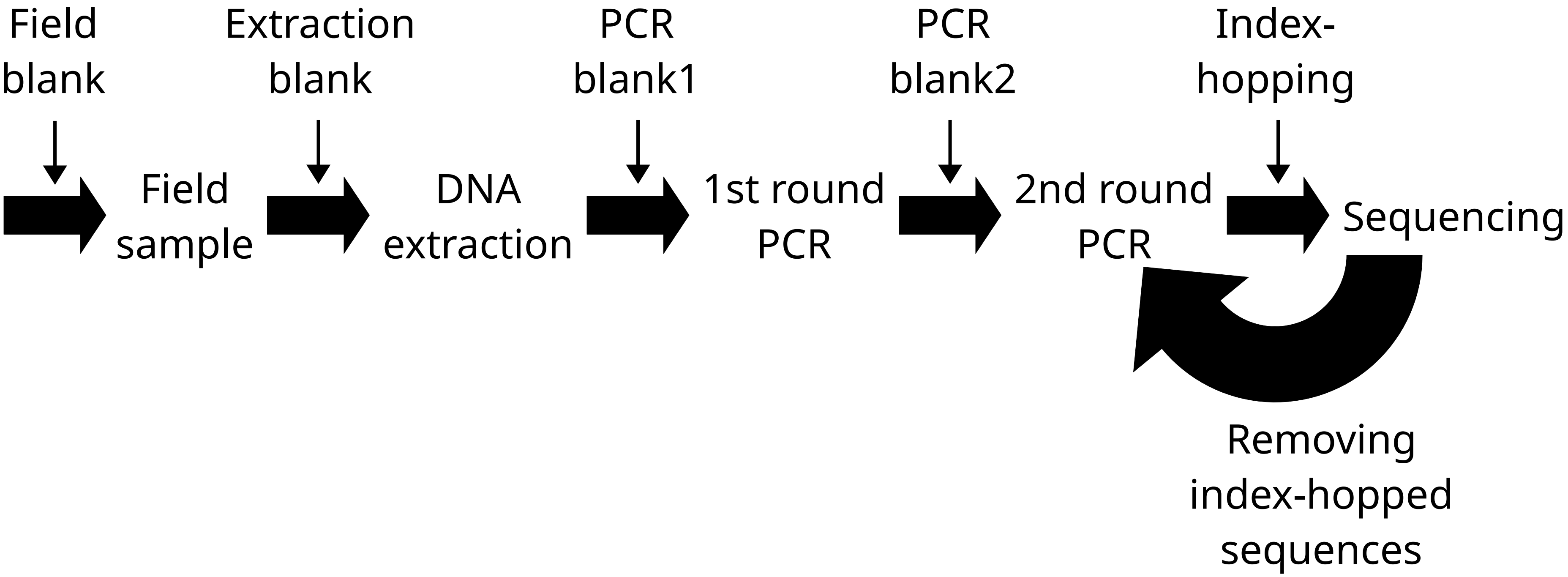
DNA extraction

1st round PCR

2nd round PCR

Sequencing

Interlude: Multistep contamination and multistep decontamination



Field blank

Extraction blank

PCR blank1

PCR blank2

Index-hopping

Field sample

DNA extraction

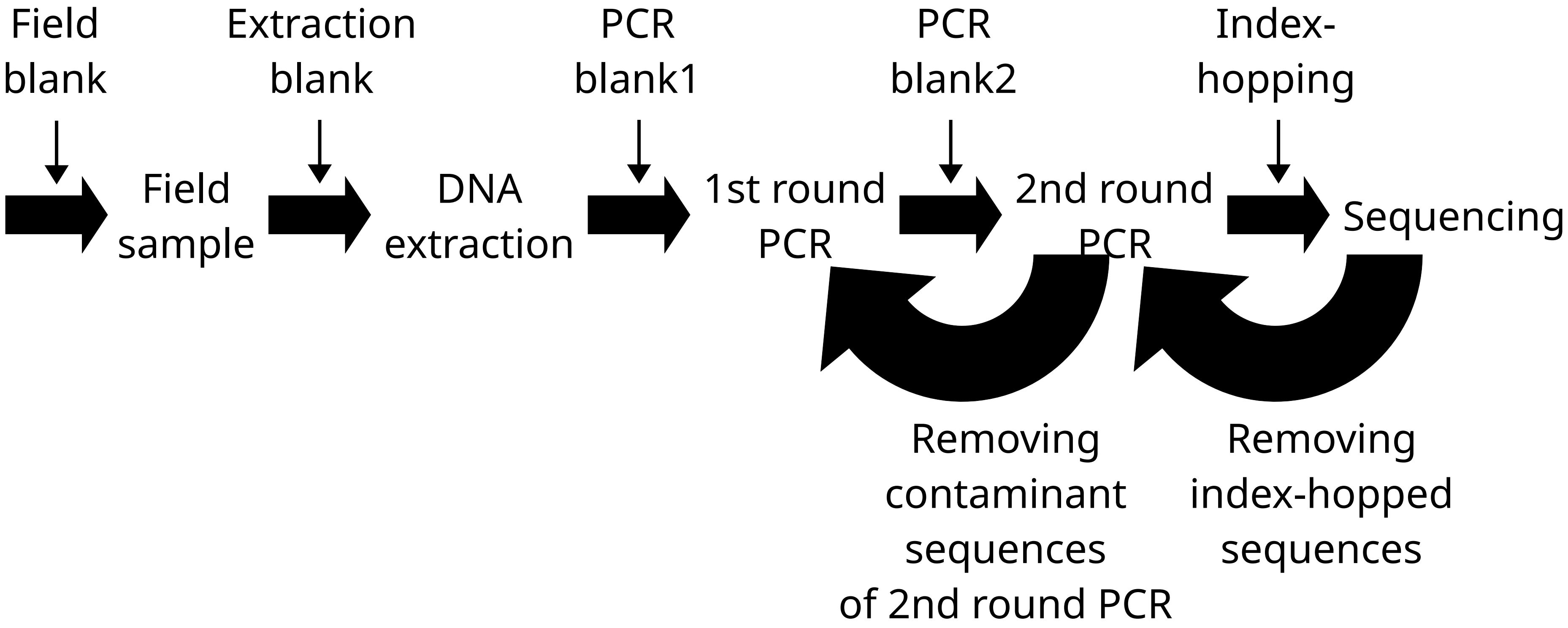
1st round PCR

2nd round PCR

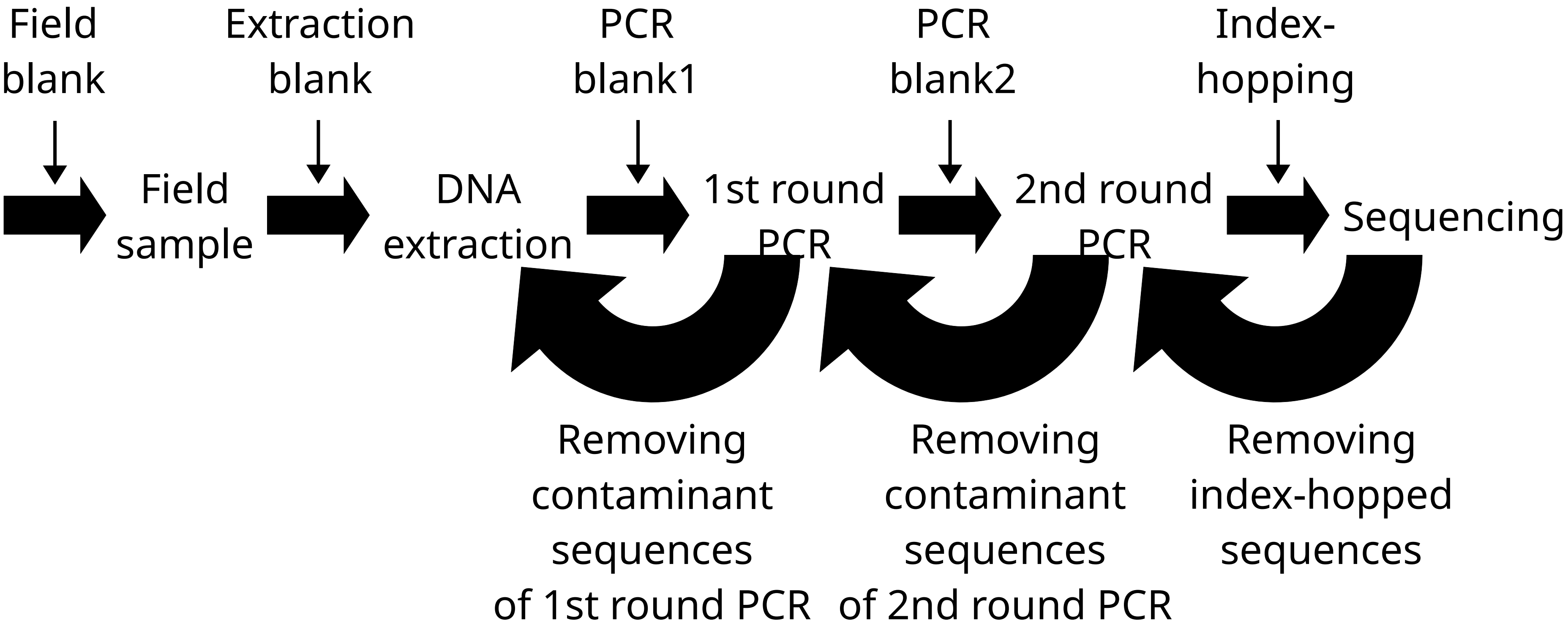
Sequencing

Removing index-hopped sequences

Interlude: Multistep contamination and multistep decontamination



Interlude: Multistep contamination and multistep decontamination



Field blank

Extraction blank

PCR blank1

PCR blank2

Index-hopping

Field sample

DNA extraction

1st round PCR

2nd round PCR

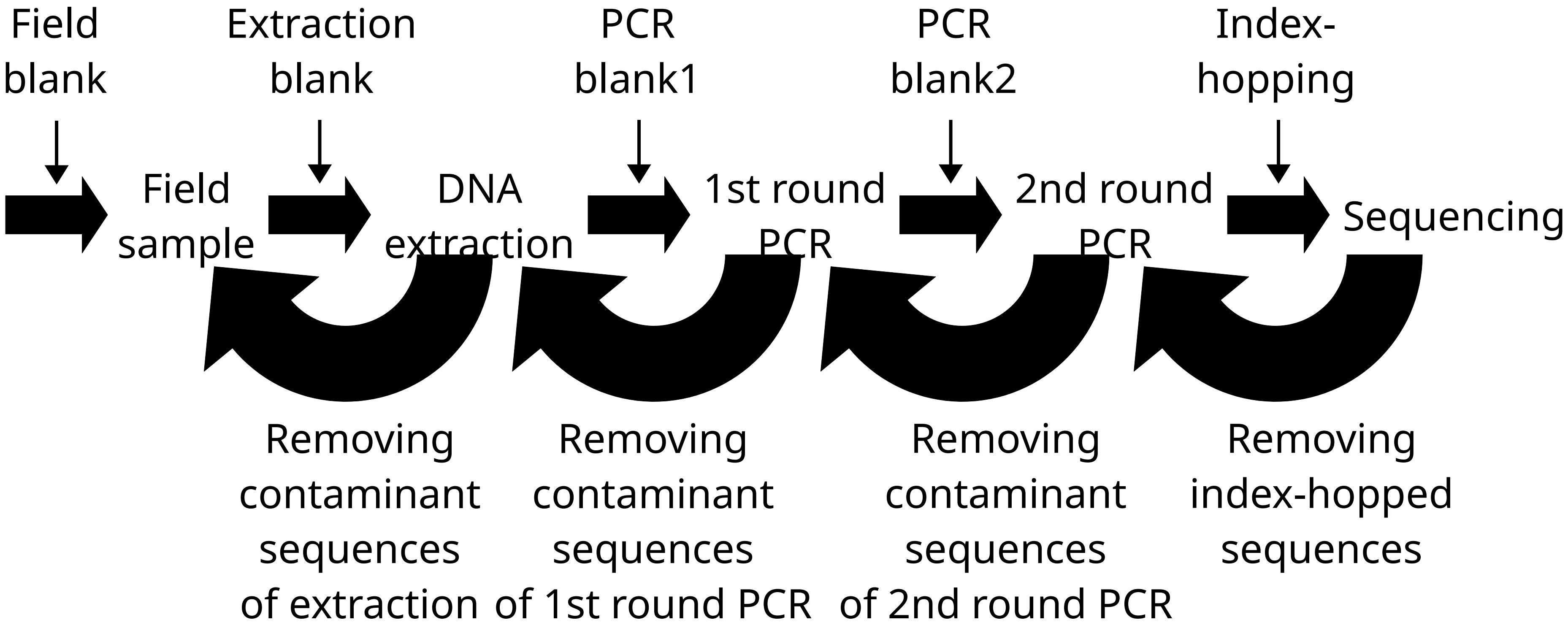
Sequencing

Removing contaminant sequences of 1st round PCR

Removing contaminant sequences of 2nd round PCR

Removing index-hopped sequences

Interlude: Multistep contamination and multistep decontamination



Field blank

Extraction blank

PCR blank1

PCR blank2

Index-hopping

Field sample

DNA extraction

1st round PCR

2nd round PCR

Sequencing

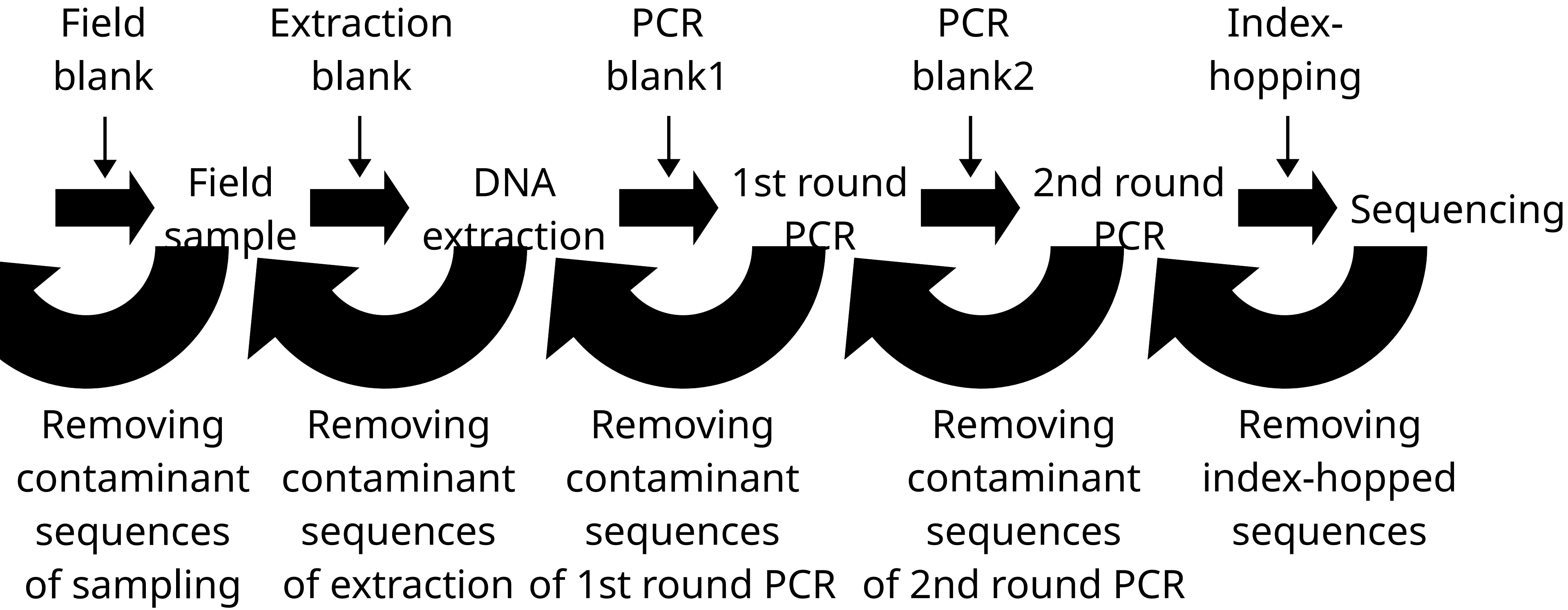
Removing contaminant sequences of extraction

Removing contaminant sequences of 1st round PCR

Removing contaminant sequences of 2nd round PCR

Removing index-hopped sequences

Interlude: Multistep contamination and multistep decontamination



Field blank

Extraction blank

PCR blank1

PCR blank2

Index-hopping

Field sample

DNA extraction

1st round PCR

2nd round PCR

Sequencing

Removing contaminant sequences of sampling

Removing contaminant sequences of extraction

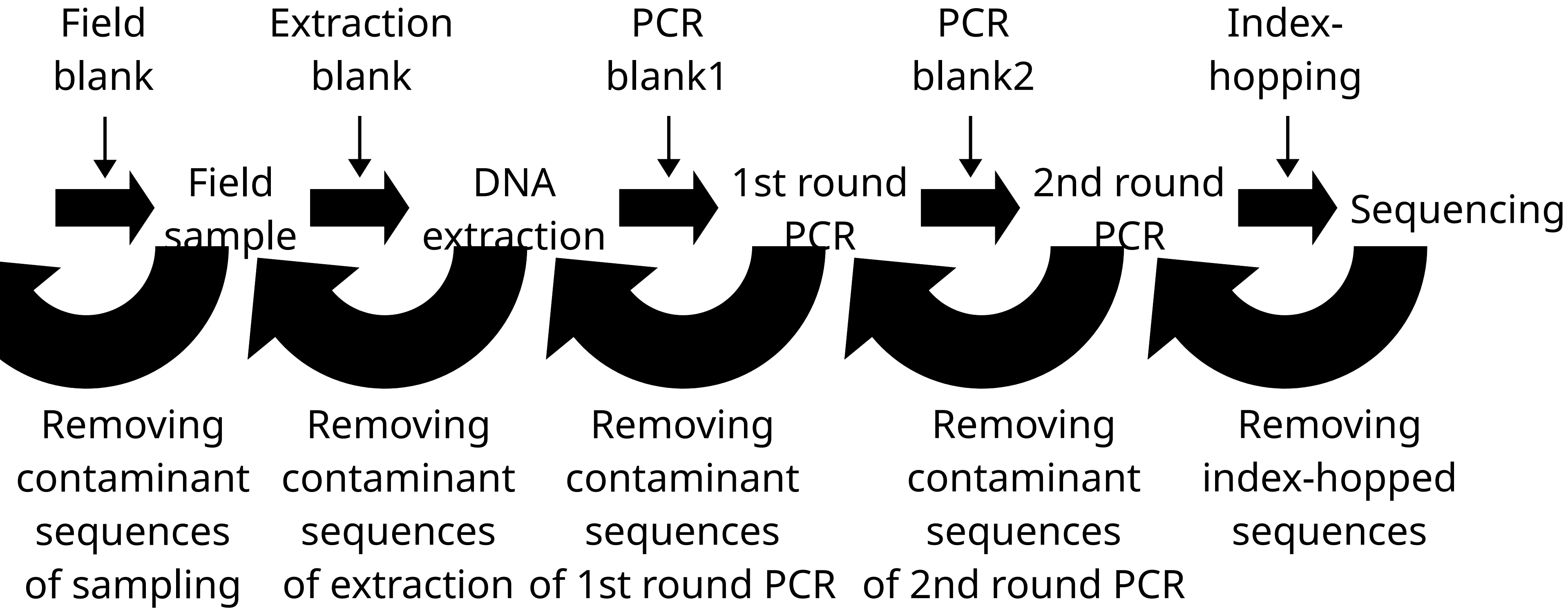
Removing contaminant sequences of 1st round PCR

Removing contaminant sequences of 2nd round PCR

Removing index-hopped sequences

Interlude: Multistep contamination and multistep decontamination

My recommendation is index-hopping removal + the other contaminant removal.



Field blank

Extraction blank

PCR blank1

PCR blank2

Index-hopping

Field sample

DNA extraction

1st round PCR

2nd round PCR

Sequencing

Removing contaminant sequences of sampling

Removing contaminant sequences of extraction

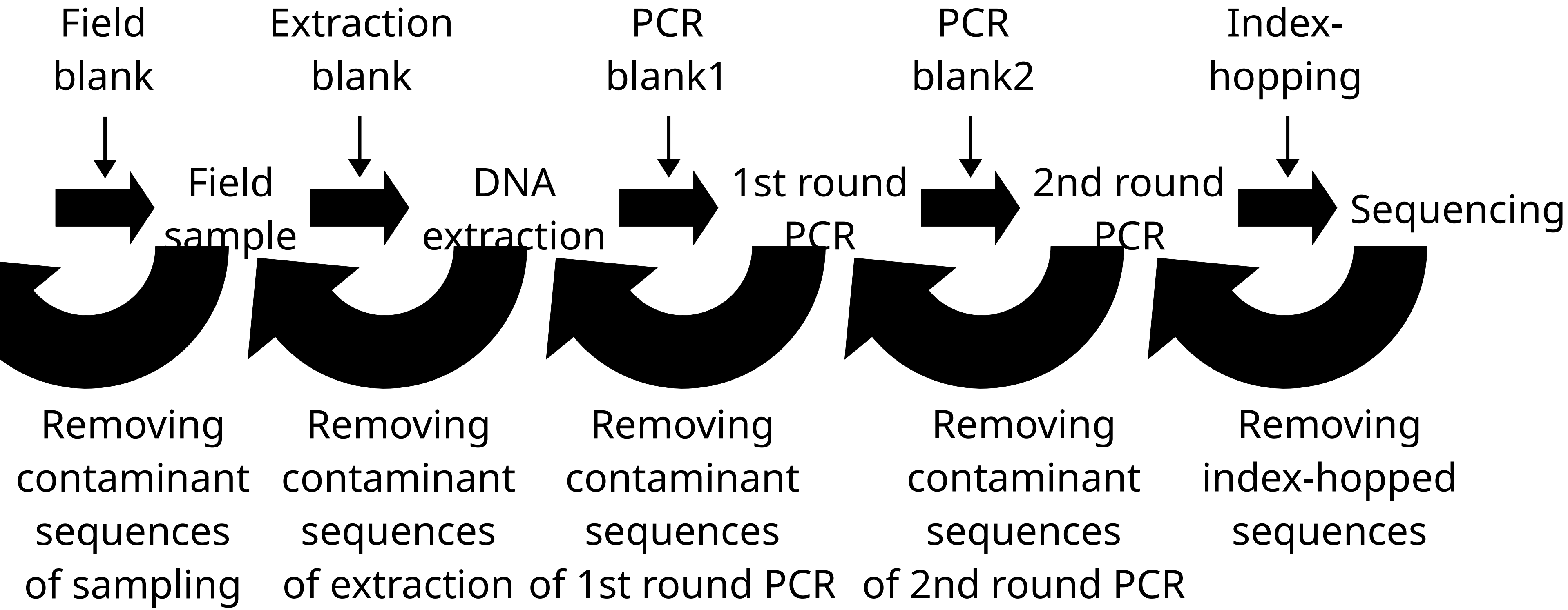
Removing contaminant sequences of 1st round PCR

Removing contaminant sequences of 2nd round PCR

Removing index-hopped sequences

Interlude: Multistep contamination and multistep decontamination

My recommendation is index-hopping removal + the other contaminant removal.
However, the best practice has been still unknown.



Field blank

Extraction blank

PCR blank1

PCR blank2

Index-hopping

Field sample

DNA extraction

1st round PCR

2nd round PCR

Sequencing

Removing contaminant sequences of sampling

Removing contaminant sequences of extraction

Removing contaminant sequences of 1st round PCR

Removing contaminant sequences of 2nd round PCR

Removing index-hopped sequences

Interlude: Study purpose and decontamination

Interlude: Study purpose and decontamination

- Non-decontaminated metabarcoding results contain contaminants

Interlude: Study purpose and decontamination

- Non-decontaminated metabarcoding results contain contaminants
- Decontamination should be applied?

Interlude: Study purpose and decontamination

- Non-decontaminated metabarcoding results contain contaminants
- Decontamination should be applied?
  - If you want to maximize detection power, NO. Decontamination potentially misidentify true sequence as contaminant

Interlude: Study purpose and decontamination

- Non-decontaminated metabarcoding results contain contaminants
- Decontamination should be applied?
  - If you want to maximize detection power, NO. Decontamination potentially misidentify true sequence as contaminant
  - If you want to minimize misdetection, YES. Lack of decontamination may cause many misdetection

Interlude: Study purpose and decontamination

- Non-decontaminated metabarcoding results contain contaminants
- Decontamination should be applied?
  - If you want to maximize detection power, NO. Decontamination potentially misidentify true sequence as contaminant
  - If you want to minimize misdetection, YES. Lack of decontamination may cause many misdetection
  - If you want to analyse community composition, UNKNOWN. Because abundances of contaminants may be low in many cases, their effects to analysis may be low. However, whether abundances of contaminants are really low or not IN YOUR DATA is unknown.

Chapter 10: Additional clustering

Chapter 10: Additional clustering

- Inputs
- decontaminated.fasta
- decontaminated.otu.gz

in OverlappedPairedEnd_wSTD_09_DecontaminatedSequences

Chapter 10: Additional clustering

- Inputs
- decontaminated.fasta
- decontaminated.otu.gz

in OverlappedPairedEnd_wSTD_09_DecontaminatedSequences

- Outputs
- clustered.fasta
- clustered.otu.gz
- clustered.tsv

in OverlappedPairedEnd_wSTD_10_ClusteredSequences

Chapter 10: Additional clustering

Switch to Terminal

Chapter 11: Taxonomic assignment

Chapter 11: Taxonomic assignment

- Inputs
- clustered.fasta

  in OverlappedPairedEnd_wSTD_10_ClusteredSequences

Chapter 11: Taxonomic assignment

- Inputs
- clustered.fasta

in OverlappedPairedEnd_wSTD_10_ClusteredSequences

- Outputs
- neighborhoods_1nn_*.txt
- neighborhoods_qc_*.txt
- taxonomy_1nn_*.tsv
- taxonomy_qc_*.tsv
- taxonomy_merged.tsv
- taxonomy_merged_filled.tsv

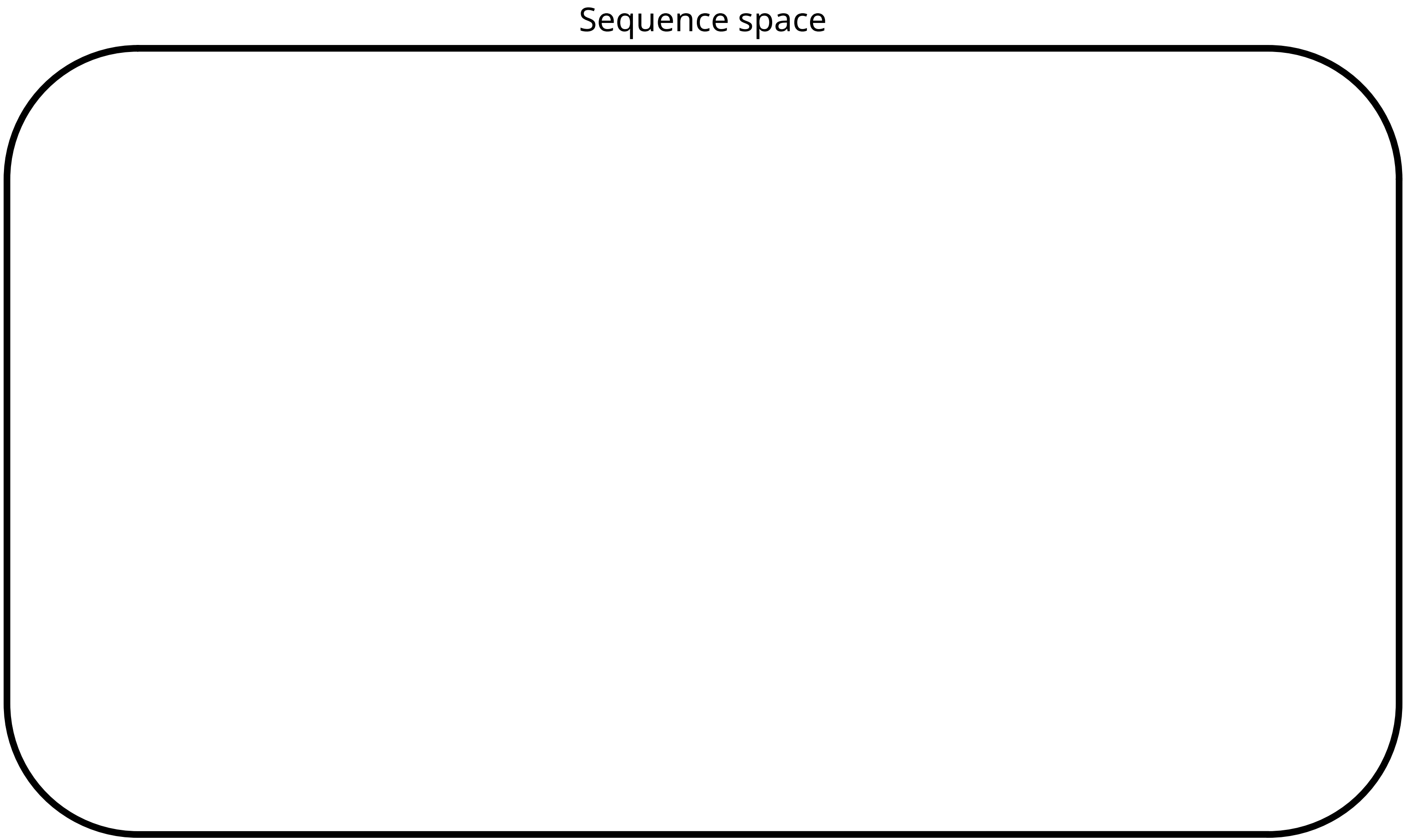
in OverlappedPairedEnd_wSTD_11_ClaidentResults

Chapter 11: Taxonomic assignment

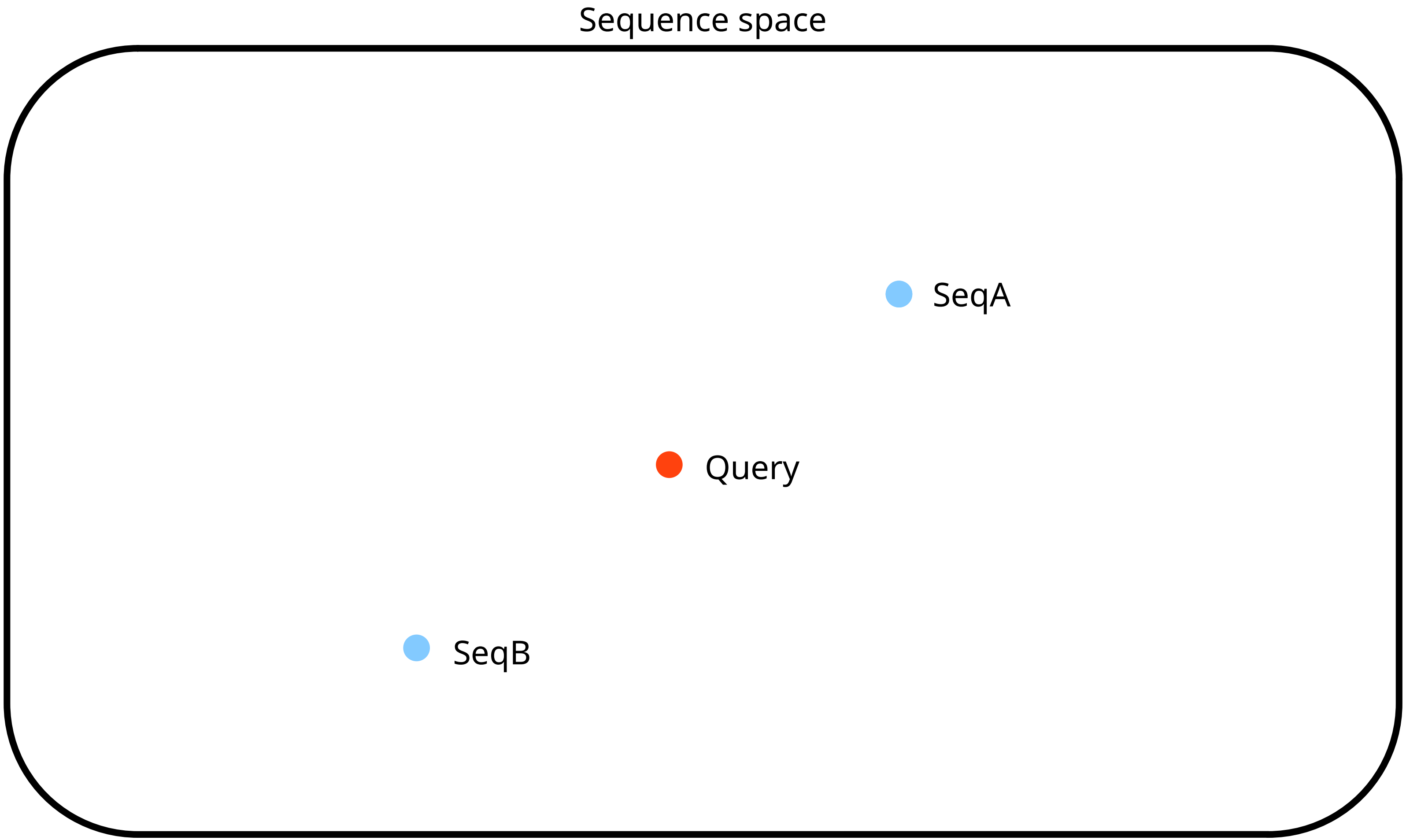
Switch to Terminal

Interlude: Query-centric auto-*k*-nearest neighbor method

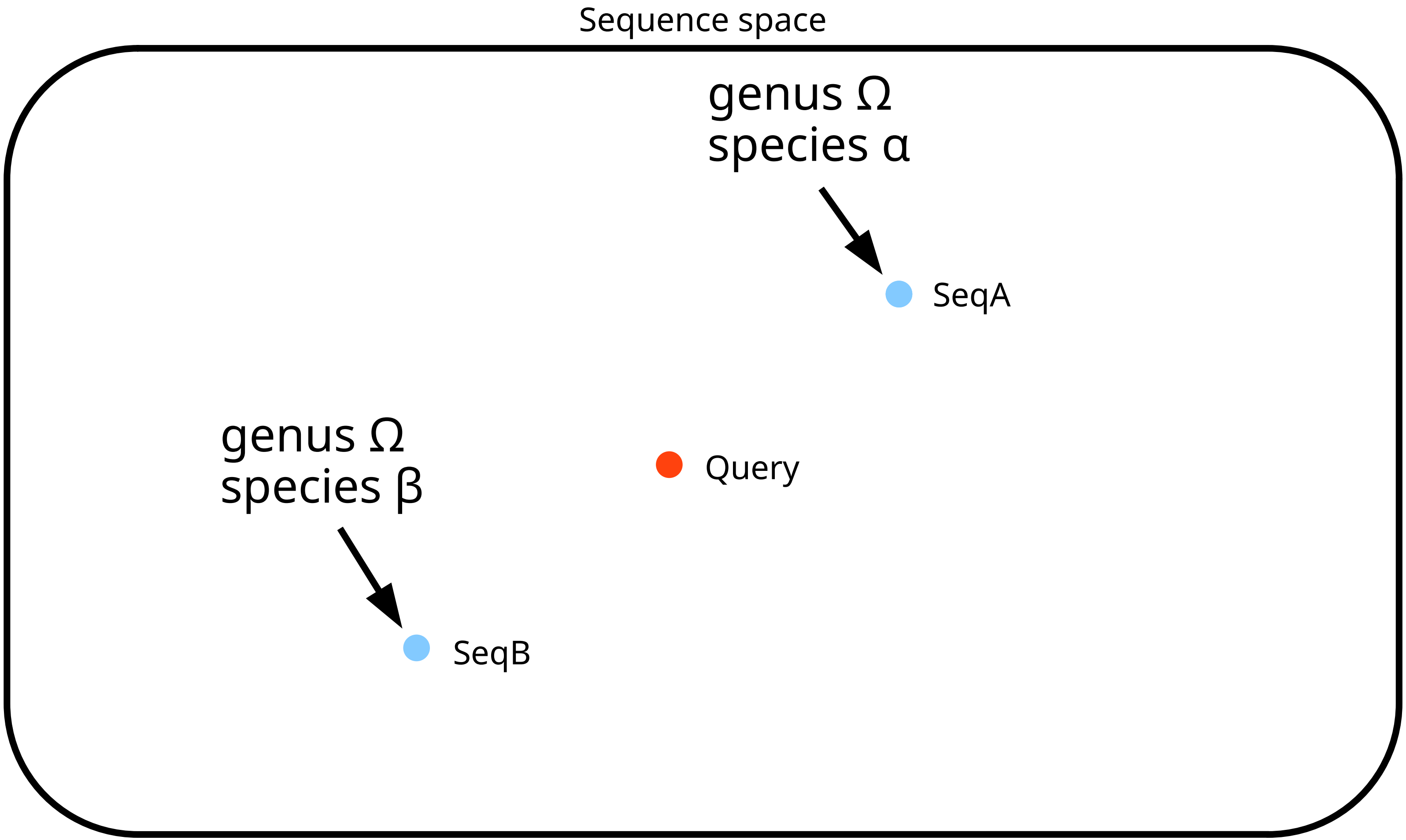
Interlude: Query-centric auto-*k*-nearest neighbor method



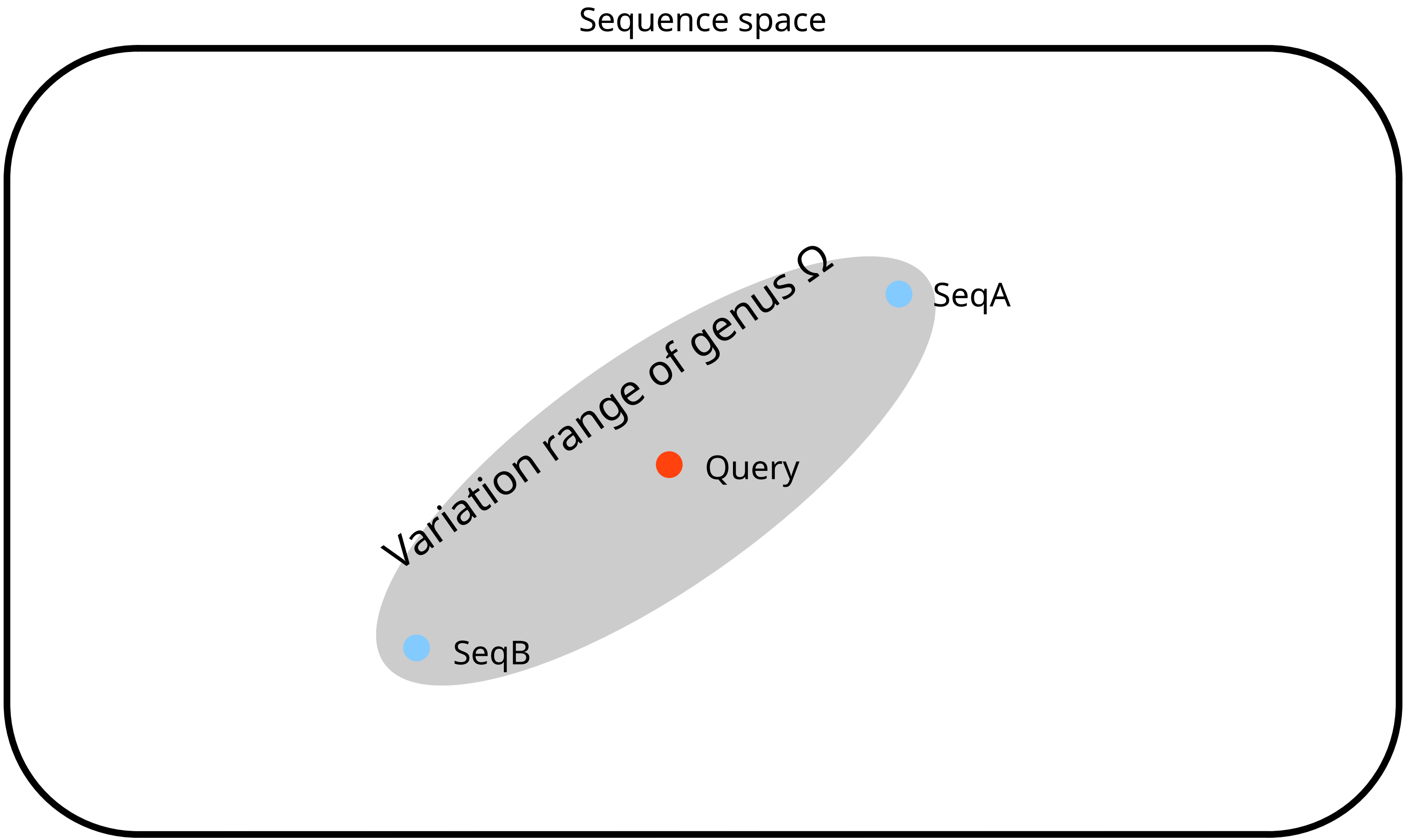
Interlude: Query-centric auto-*k*-nearest neighbor method



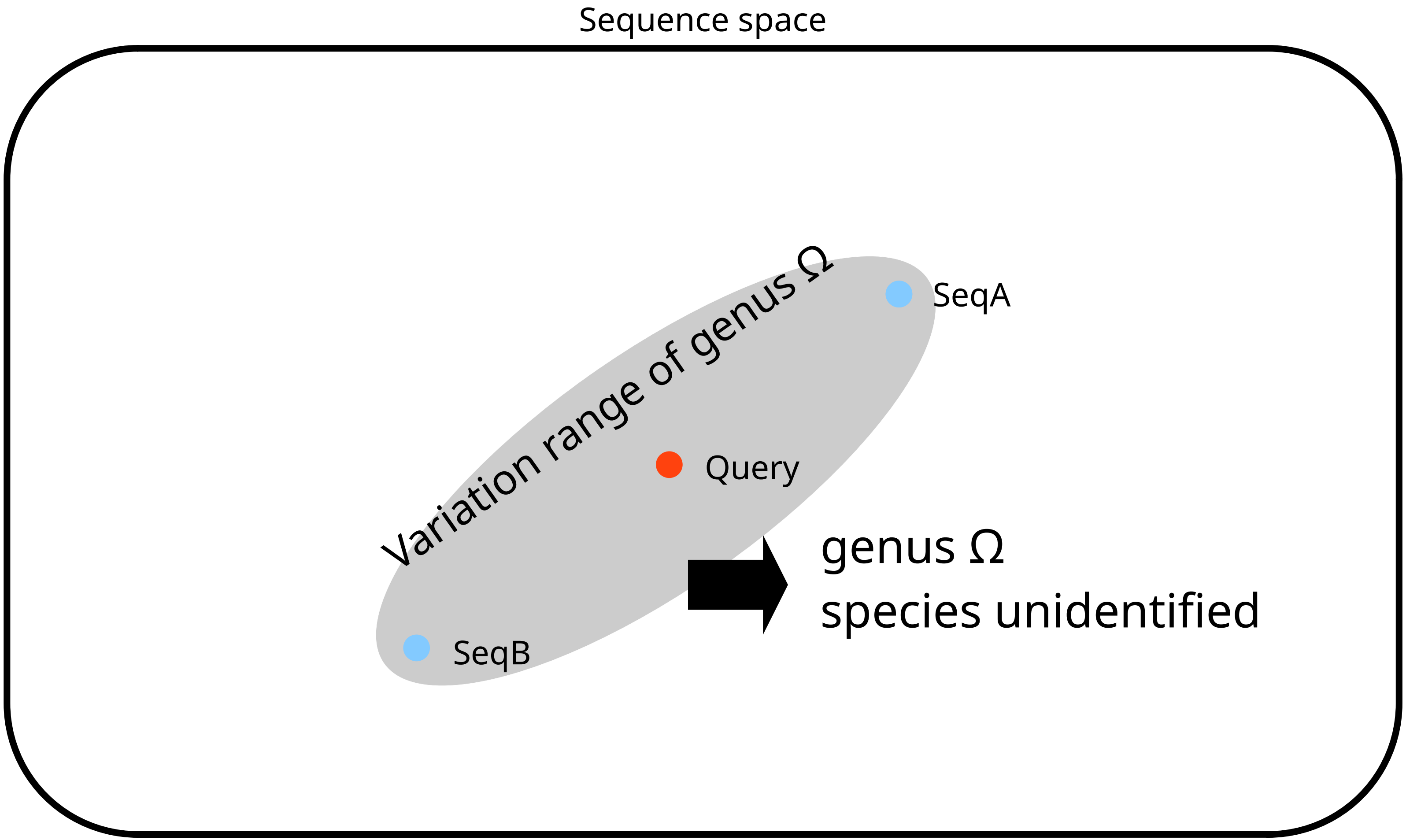
Interlude: Query-centric auto-*k*-nearest neighbor method



Interlude: Query-centric auto-*k*-nearest neighbor method



Interlude: Query-centric auto-*k*-nearest neighbor method



Interlude: Which method should be used for taxonomic assignment?

Interlude: Which method should be used for taxonomic assignment?

- If reference database is imperfect (most cases), QCauto shows the best balance between less misidentification and less successful identification

Interlude: Which method should be used for taxonomic assignment?

- If reference database is imperfect (most cases), QCauto shows the best balance between less misidentification and less successful identification
- If reference database is perfect or nearly perfect, 1-NN is the best. However, whether the reference database is really perfect or not should not be known by anyone

Interlude: Ready-made reference databases

Interlude: Ready-made reference databases

- Installed to INSTALLPATH/share/claident/blastdb

Interlude: Ready-made reference databases

- Installed to INSTALLPATH/share/claident/blastdb
- overall_class, overall_order, overall_family

Interlude: Ready-made reference databases

- Installed to INSTALLPATH/share/claident/blastdb
- overall_class, overall_order, overall_family
    - Subset of NCBI nt including class, order or family level identified seqs

Interlude: Ready-made reference databases

- Installed to INSTALLPATH/share/claident/blastdb
- overall_class, overall_order, overall_family
  - Subset of NCBI nt including class, order or family level identified seqs
- *_genus

Interlude: Ready-made reference databases

- Installed to INSTALLPATH/share/claident/blastdb
- overall_class, overall_order, overall_family
  - Subset of NCBI nt including class, order or family level identified seqs
- *_genus
  - Subset of overall_* including genus level identified seqs

Interlude: Ready-made reference databases

- Installed to INSTALLPATH/share/claident/blastdb
- overall_class, overall_order, overall_family
  - Subset of NCBI nt including class, order or family level identified seqs
- *_genus
  - Subset of overall_* including genus level identified seqs
- *_species_wsp

Interlude: Ready-made reference databases

- Installed to INSTALLPATH/share/claident/blastdb
- overall_class, overall_order, overall_family
  - Subset of NCBI nt including class, order or family level identified seqs
- *_genus
  - Subset of overall_* including genus level identified seqs
- *_species_wsp
  - Subset of overall_* including species level identified seqs

Interlude: Ready-made reference databases

- Installed to INSTALLPATH/share/claident/blastdb
- overall_class, overall_order, overall_family
  - Subset of NCBI nt including class, order or family level identified seqs
- *_genus
  - Subset of overall_* including genus level identified seqs
- *_species_wsp
  - Subset of overall_* including species level identified seqs
- *_species

Interlude: Ready-made reference databases

- Installed to INSTALLPATH/share/claident/blastdb
- overall_class, overall_order, overall_family
  - Subset of NCBI nt including class, order or family level identified seqs
- *_genus
  - Subset of overall_* including genus level identified seqs
- *_species_wsp
  - Subset of overall_* including species level identified seqs
- *_species
  - Subset of overall_* including species level identified seqs except for the seqs which have " sp." at the tail in species name

Interlude: Ready-made reference databases

- Installed to INSTALLPATH/share/claident/blastdb
- overall_class, overall_order, overall_family
  - Subset of NCBI nt including class, order or family level identified seqs
- *_genus
  - Subset of overall_* including genus level identified seqs
- *_species_wsp
  - Subset of overall_* including species level identified seqs
- *_species
  - Subset of overall_* including species level identified seqs except for the seqs which have " sp." at the tail in species name
- *_species_wosp

Interlude: Ready-made reference databases

- Installed to INSTALLPATH/share/claident/blastdb
- overall_class, overall_order, overall_family
  - Subset of NCBI nt including class, order or family level identified seqs
- *_genus
  - Subset of overall_* including genus level identified seqs
- *_species_wsp
  - Subset of overall_* including species level identified seqs
- *_species
  - Subset of overall_* including species level identified seqs except for the seqs which have " sp." at the tail in species name
- *_species_wosp
  - Subset of overall_* including species level identified seqs except for the seqs which have " sp." in species name

Interlude: Taxonomic infomation reliability in reference databases

Interlude: Taxonomic infomation reliability in reference databases

*_species_wosp＞*_species＞*_species_wsp＞*_genus＞*_family＞*_order＞*_class

Interlude: Taxonomic infomation reliability in reference databases

*_species_wosp＞*_species＞*_species_wsp＞*_genus＞*_family＞*_order＞*_class

- Because the seqs which only have higher level taxonomic info likely to be identified based on closest INSD seqs, such taxonomic info are less reliable

Interlude: Taxonomic infomation reliability in reference databases

*_species_wosp＞*_species＞*_species_wsp＞*_genus＞*_family＞*_order＞*_class

- Because the seqs which only have higher level taxonomic info likely to be identified based on closest INSD seqs, such taxonomic info are less reliable
- Because the seqs identified as " sp." is not strictly identified or such species are undescribed, such taxonomic info are less reliable

Interlude: Which reference database should be used?

Interlude: Which reference database should be used?

- overall_species_wosp is recommended in most cases because the seqs lacking lower level taxonomic info likely to be less reliable

Interlude: Which reference database should be used?

- overall_species_wosp is recommended in most cases because the seqs lacking lower level taxonomic info likely to be less reliable
- The other overall_* are recommended if you want to minimize "unidentified" in * level and can rolerate misidentification in lower level

Interlude: Which reference database should be used?

- overall_species_wosp is recommended in most cases because the seqs lacking lower level taxonomic info likely to be less reliable
- The other overall_* are recommended if you want to minimize "unidentified" in * level and can rolerate misidentification in lower level
- The others are recommended for screening or PCs lacking enough amount of memory

Interlude: Merging of taxonomy

Interlude: Merging of taxonomy

- More reliable taxonomy should be preferred but less reliable taxonomy which reached to lower taxonomic level could be tolerated

Interlude: Merging of taxonomy

- More reliable taxonomy should be preferred but less reliable taxonomy which reached to lower taxonomic level could be tolerated
- The best balance between reliability and identifiability can be achieved by merging taxonomy from overall_species_wosp and the other overall_*

Chapter 12: Making summary tables

Chapter 12: Making summary tables

- Inputs
- clustered.tsv

    in OverlappedPairedEnd_wSTD_10_ClusteredSequences

Chapter 12: Making summary tables

- Inputs
- clustered.tsv

  in OverlappedPairedEnd_wSTD_10_ClusteredSequences
- taxonomy_merged_filled.tsv

  in OverlappedPairedEnd_wSTD_11_ClaidentResults

Chapter 12: Making summary tables

- Inputs
- clustered.tsv

  in OverlappedPairedEnd_wSTD_10_ClusteredSequences
- taxonomy_merged_filled.tsv

  in OverlappedPairedEnd_wSTD_11_ClaidentResults

- Outputs
- sample_otu_matrix_fishes.tsv
- sample_otu_matrix_fishes_converted.tsv
- sample_top50species_nreads_fishes_converted.tsv
- sample_top50family_nreads_fishes_converted.tsv
- sample_species_nreads_fishes_converted.tsv
- sample_family_nreads_fishes_converted.tsv

  in OverlappedPairedEnd_wSTD_11_ClaidentResults

Chapter 12: Making summary tables

Switch to Terminal

Chapter 13: Plotting community structure

Chapter 13: Plotting community structure

- Inputs
- sample_top50species_nreads_fishes_converted.tsv
- sample_top50family_nreads_fishes_converted.tsv
- sample_species_nreads_fishes_converted.tsv
- sample_family_nreads_fishes_converted.tsv

in OverlappedPairedEnd_wSTD_11_ClaidentResults

Chapter 13: Plotting community structure

- Inputs
- sample_top50species_nreads_fishes_converted.tsv
- sample_top50family_nreads_fishes_converted.tsv
- sample_species_nreads_fishes_converted.tsv
- sample_family_nreads_fishes_converted.tsv

in OverlappedPairedEnd_wSTD_11_ClaidentResults

- Outputs
- barplottop50species.pdf
- barplottop50family.pdf
- heatmapspecies.pdf
- heatmapfamily.pdf

in OverlappedPairedEnd_wSTD_12_RAnalysisResults

Chapter 13: Plotting community structure

Switch to Terminal

Chapter 14: Plotting sampling/sequencing coverage

Chapter 14: Plotting sampling/sequencing coverage

- Inputs
- sample_otu_matrix_fishes.tsv

  in OverlappedPairedEnd_wSTD_11_ClaidentResults

Chapter 14: Plotting sampling/sequencing coverage

- Inputs
- sample_otu_matrix_fishes.tsv

in OverlappedPairedEnd_wSTD_11_ClaidentResults

- Outputs
- specaccum.pdf
- rarecurve.pdf

in OverlappedPairedEnd_wSTD_12_RAnalysisResults

Chapter 14: Plotting sampling/sequencing coverage

- Inputs
- sample_otu_matrix_fishes.tsv

  in OverlappedPairedEnd_wSTD_11_ClaidentResults

- Outputs
- specaccum.pdf
- rarecurve.pdf

  in OverlappedPairedEnd_wSTD_12_RAnalysisResults

- Community (data.frame)

  in R workspace

Chapter 14: Plotting sampling/sequencing coverage

Switch to Terminal

Chapter 15: Applying coverage-based rarefaction and quantification

Chapter 15: Applying coverage-based rarefaction and quantification

- Inputs
- Community (data.frame)

in R workspace

Chapter 15: Applying coverage-based rarefaction and quantification

- Inputs
- Community (data.frame)

in R workspace

- Outputs
- RarefiedCommunity (data.frame in list)
- ConvertedRarefiedCommunity (data.frame in list)

in R workspace

Chapter 15: Applying coverage-based rarefaction and quantification

- Inputs
- Community (data.frame)

in R workspace

- Outputs
- RarefiedCommunity (data.frame in list)
- ConvertedRarefiedCommunity (data.frame in list)

in R workspace

- RarefiedCommunity*.tsv
- ConvertedRarefiedCommunity*.tsv

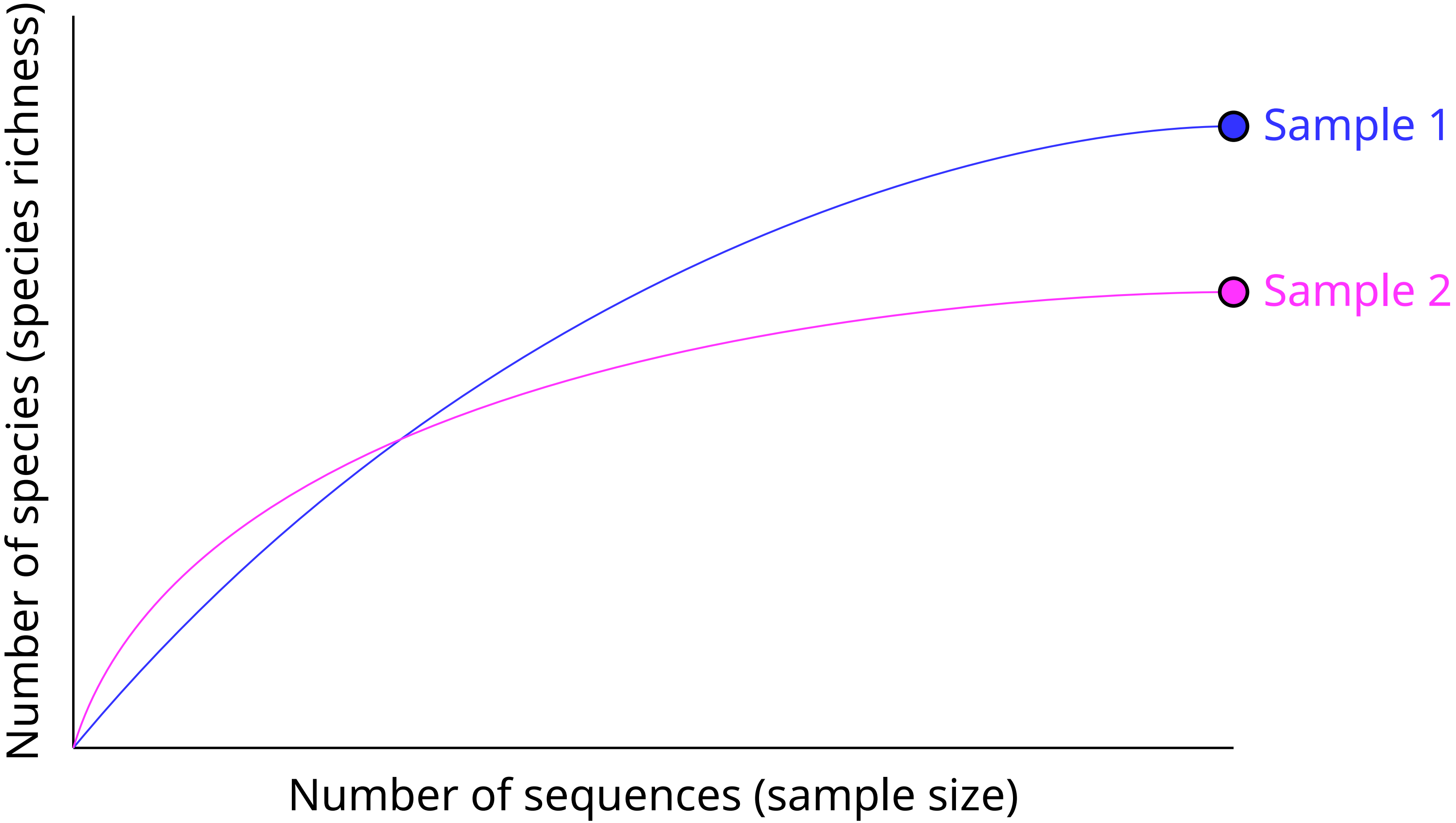
in OverlappedPairedEnd_wSTD_12_RAnalysisResults

Chapter 15: Applying coverage-based rarefaction and quantification

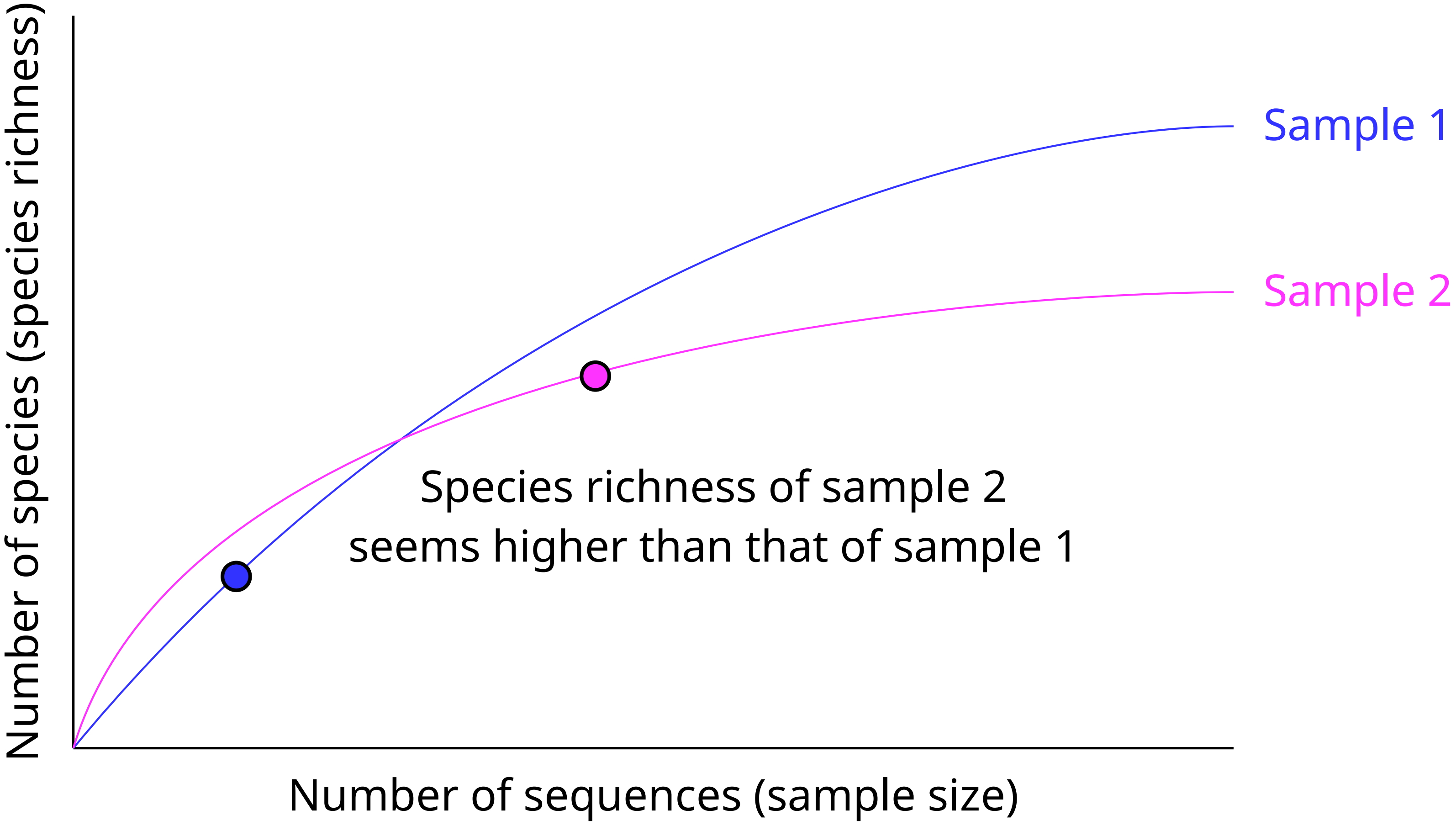
Switch to Terminal

Interlude: Problem of unequal sequencing effort

Interlude: Problem of unequal sequencing effort

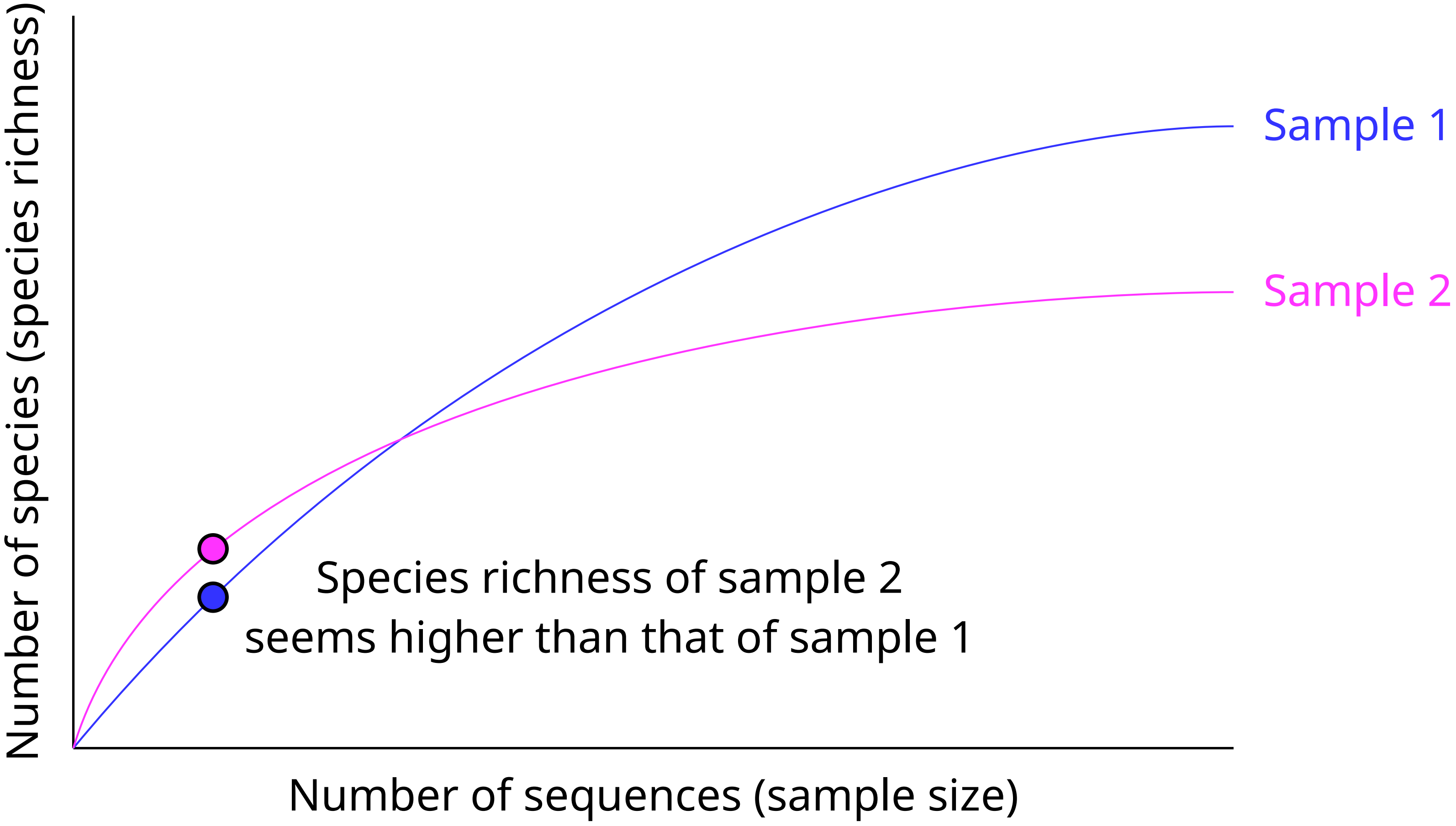


Interlude: Problem of unequal sequencing effort

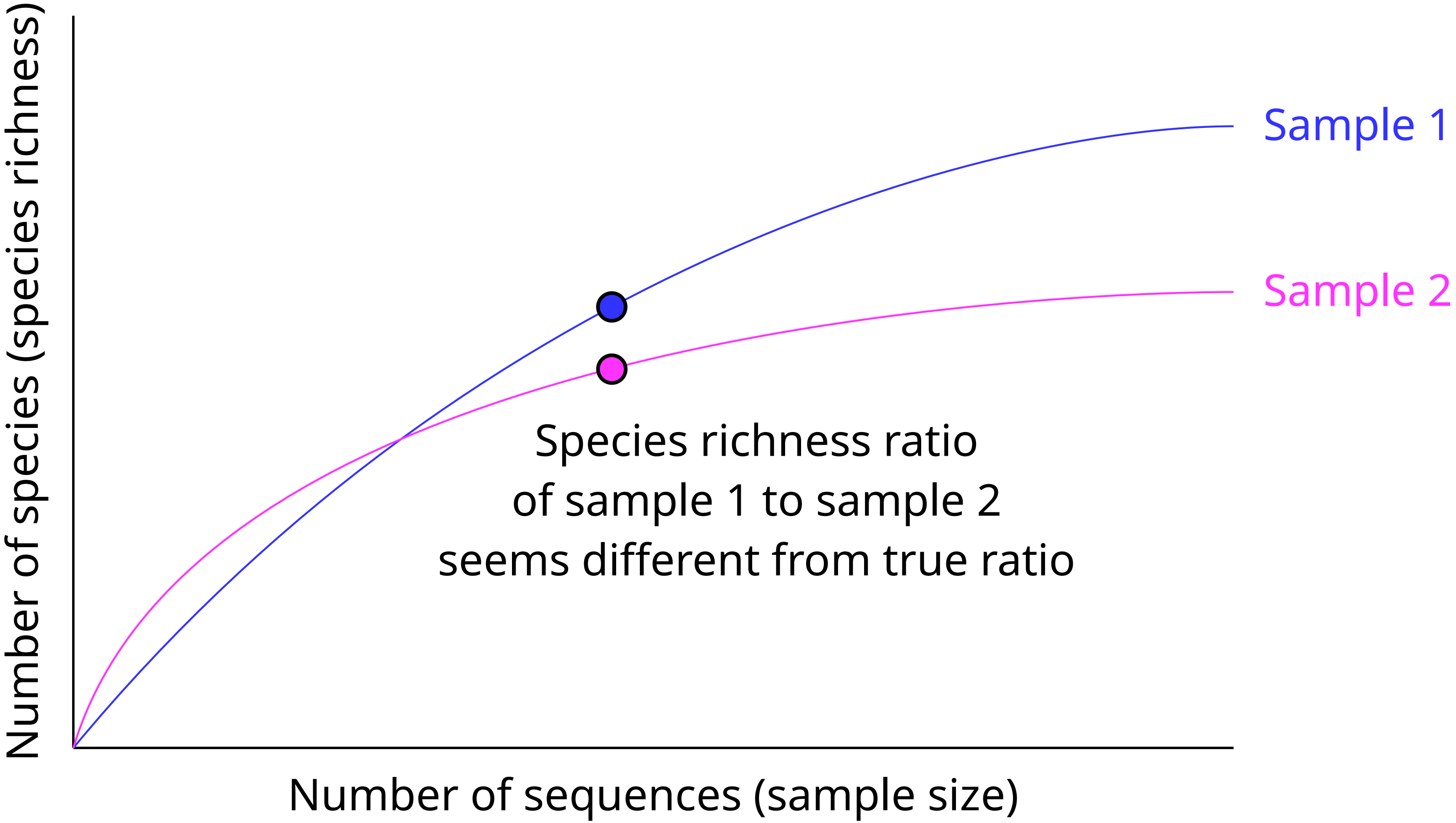


Species richness of sample 2
seems higher than that of sample 1

Interlude: Problem of unequal sequencing effort

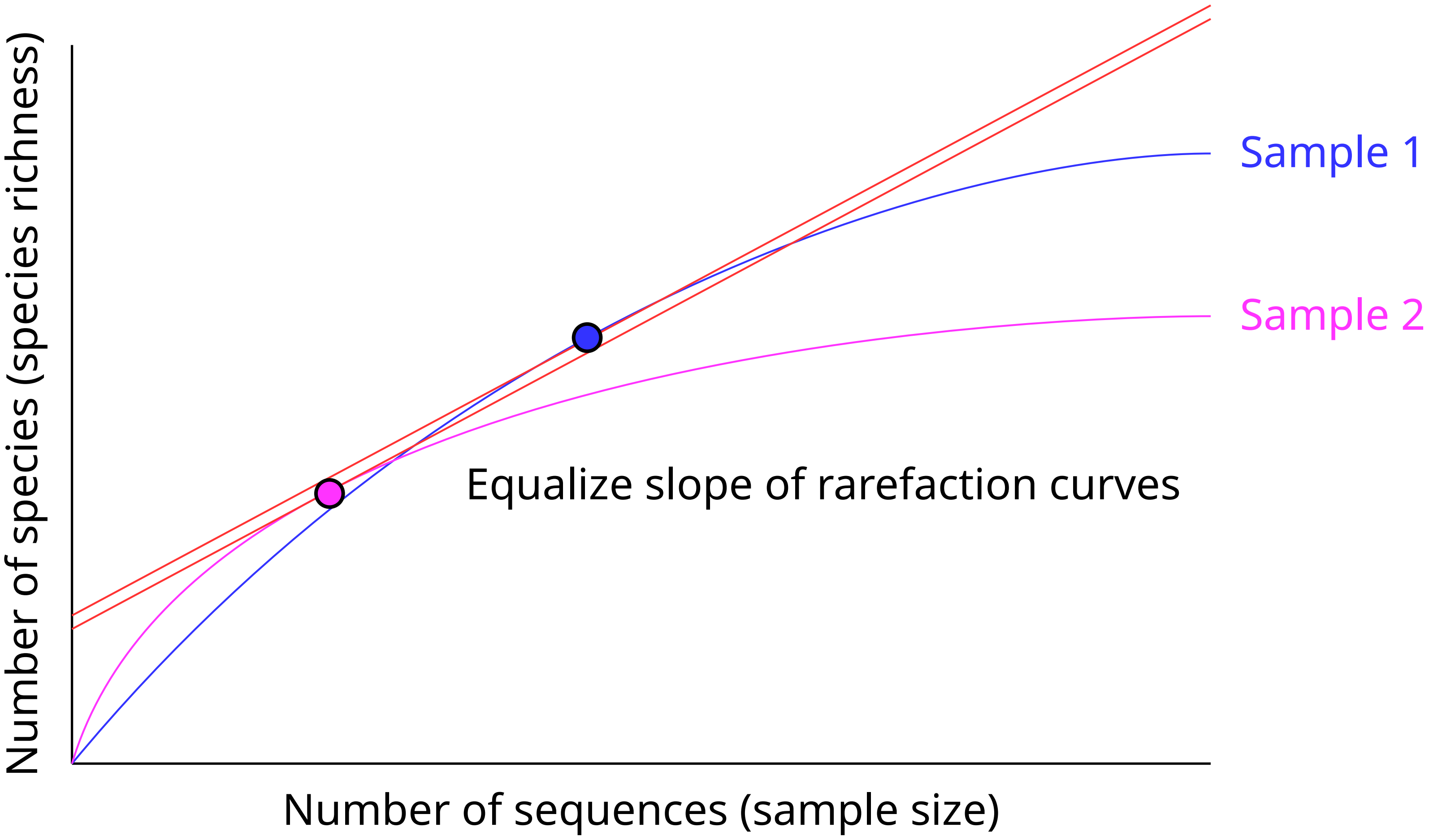


Interlude: Problem of unequal sequencing effort

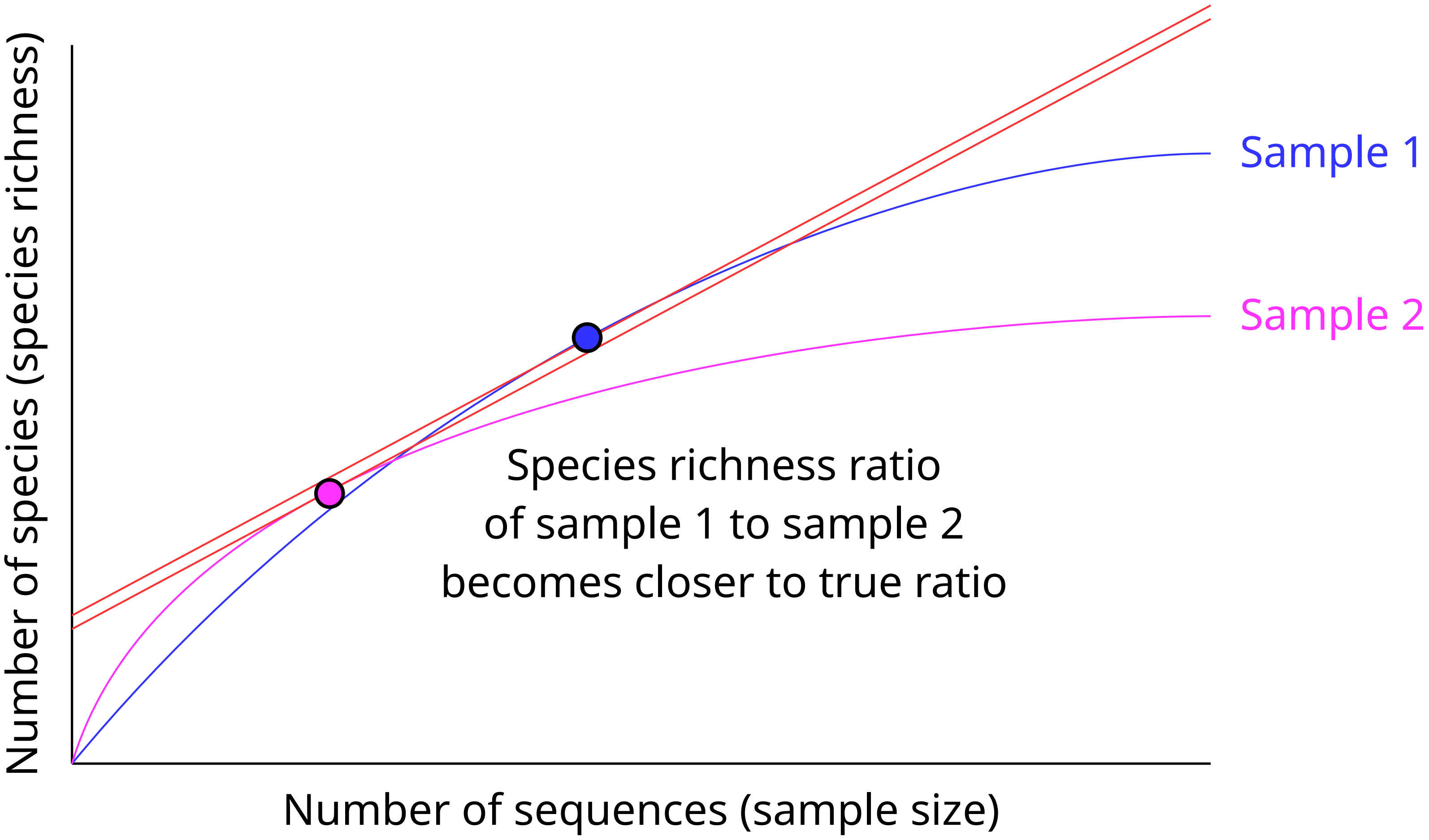


Species richness ratio
of sample 1 to sample 2
seems different from true ratio

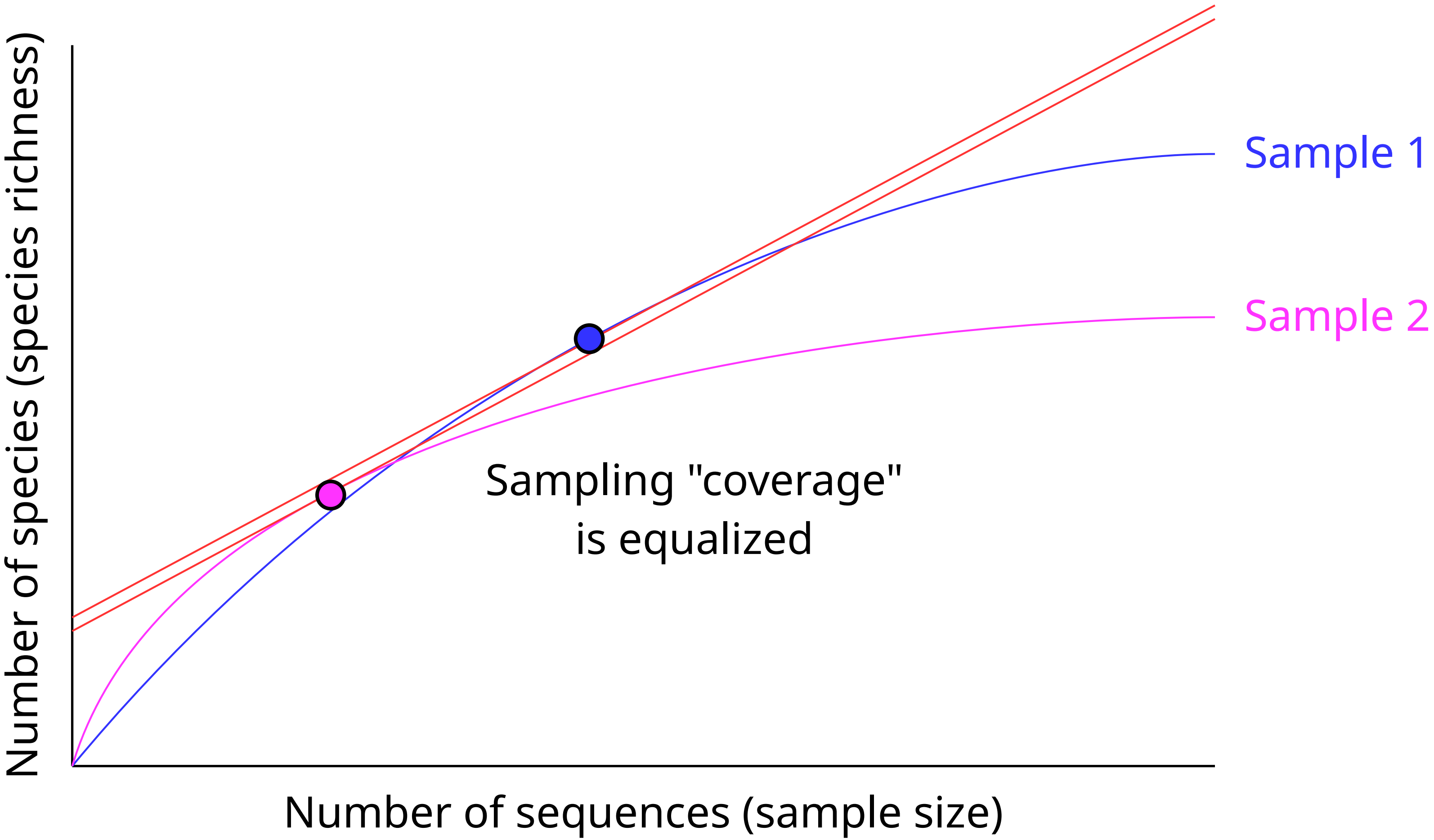
Interlude: Problem of unequal sequencing effort



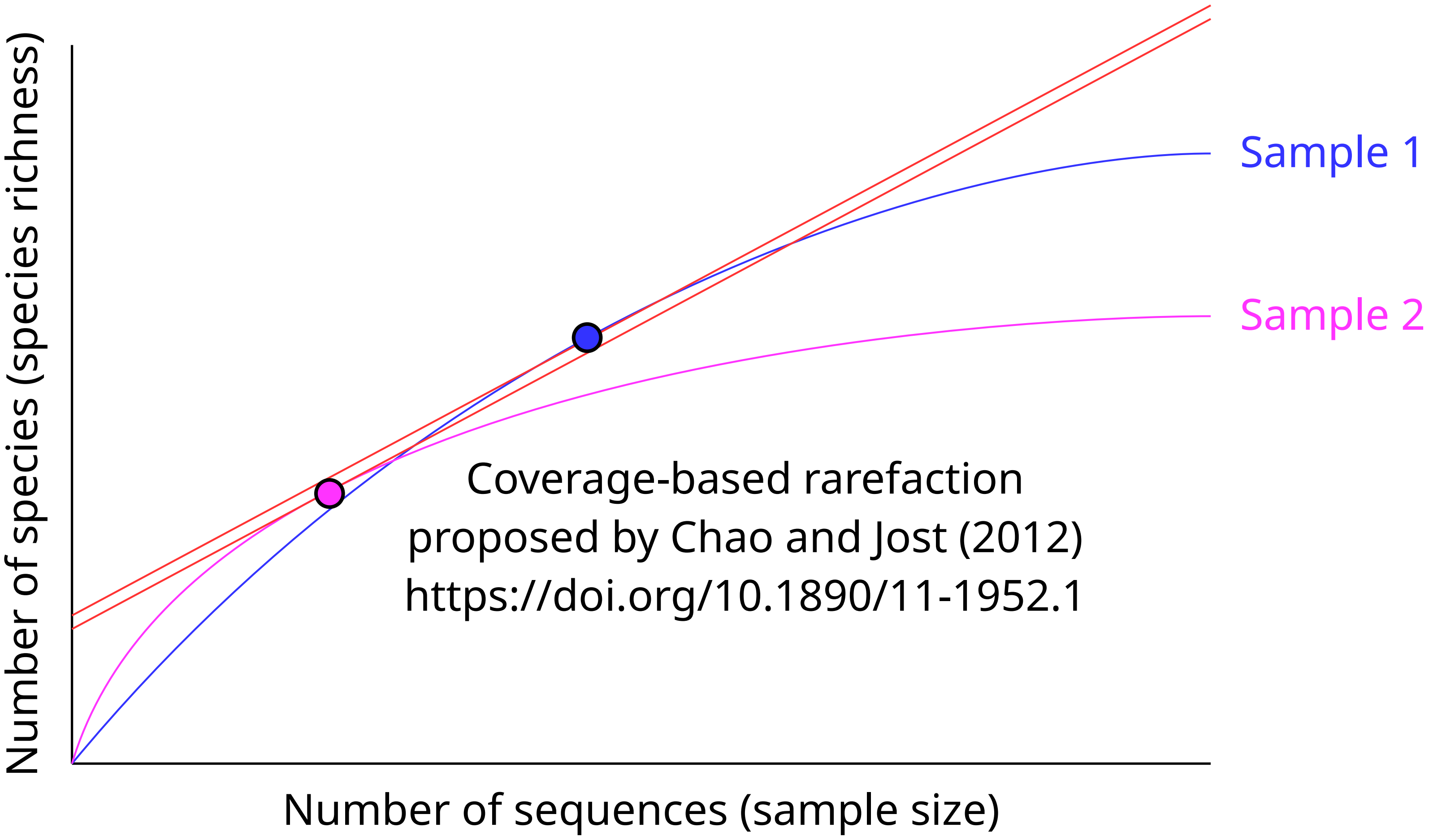
Interlude: Problem of unequal sequencing effort



Interlude: Problem of unequal sequencing effort



Interlude: Problem of unequal sequencing effort



Coverage-based rarefaction
proposed by Chao and Jost (2012)
https://doi.org/10.1890/11-1952.1

Chapter 16: Calculating distance matrices

Chapter 16: Calculating distance matrices

- Inputs
- ConvertedRarefiedCommunity (data.frame in list)

in R workspace

Chapter 16: Calculating distance matrices

- Inputs
- ConvertedRarefiedCommunity (data.frame in list)

in R workspace

- Outputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

Chapter 16: Calculating distance matrices

Switch to Terminal

Interlude: Community distance (β diversity) metrics, PERMANOVA, and NMDS

Interlude: Community distance (β diversity) metrics, PERMANOVA, and NMDS

See

- Anderson et al. (2010) https://doi.org/10.1111/j.1461-0248.2010.01552.x
- Anderson (2001) https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x
- Anderson (2017) https://doi.org/10.1002/9781118445112.stat07841
- 土居・岡村 (2010) https://doi.org/10.18960/seitai.61.1_3

Chapter 16: Executing PERMANOVA

Chapter 16: Executing PERMANOVA

- Inputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

Chapter 16: Executing PERMANOVA

- Inputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

- Metadata.tsv

in top directory

Chapter 16: Executing PERMANOVA

- Inputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

- Metadata.tsv

in top directory

- Outputs
- PERMANOVA.txt

in OverlappedPairedEnd_wSTD_12_RAnalysisResults

Chapter 16: Executing PERMANOVA

Switch to Terminal

Chapter 17: Executing cluster analyses with multiscale bootstrapping

Chapter 17: Executing cluster analyses with multiscale bootstrapping

- Inputs
- ConvertedRarefiedCommunity (data.frame in list)

in R workspace

Chapter 17: Executing cluster analyses with multiscale bootstrapping

- Inputs
- ConvertedRarefiedCommunity (data.frame in list)

in R workspace

- Outputs
- ClusterAnalysis_sites.pdf
- ClusterAnalysis_species.pdf

in OverlappedPairedEnd_wSTD_12_RAnalysisResults

Chapter 17: Executing cluster analyses with multiscale bootstrapping

Switch to Terminal

Chapter 18: Executing NMDS

Chapter 18: Executing NMDS

- Inputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

Chapter 18: Executing NMDS

- Inputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

- Metadata.tsv

in top directory

Chapter 18: Executing NMDS

- Inputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

- Metadata.tsv

in top directory

- Outputs
- NMDS.pdf

in OverlappedPairedEnd_wSTD_12_RAnalysisResults

Chapter 18: Executing NMDS

Switch to Terminal

Chapter 19: Executing Mantel correlogram analysis using geographical distance

Chapter 19: Executing Mantel correlogram analysis using geographical distance

- Inputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

Chapter 19: Executing Mantel correlogram analysis using geographical distance

- Inputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

- Metadata.tsv

in top directory

Chapter 19: Executing Mantel correlogram analysis using geographical distance

- Inputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

- Metadata.tsv

in top directory

- Outputs
- GeoMCA.pdf

in OverlappedPairedEnd_wSTD_12_RAnalysisResults

Chapter 19: Executing Mantel correlogram analysis using geographical distance

Switch to Terminal

Chapter 20: Executing Mantel correlogram analysis using date interval

Chapter 20: Executing Mantel correlogram analysis using date interval

- Inputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

Chapter 20: Executing Mantel correlogram analysis using date interval

- Inputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

- Metadata.tsv

in top directory

Chapter 20: Executing Mantel correlogram analysis using date interval

- Inputs
- BrayCurtis (dist in list)
- Jaccard (dist in list)
- BinaryJaccard (dist in list)
- BinaryRaupCrick (dist in list)

in R workspace

- Metadata.tsv

in top directory

- Outputs
- DateMCA.pdf

in OverlappedPairedEnd_wSTD_12_RAnalysisResults

Chapter 20: Executing Mantel correlogram analysis using date interval

Switch to Terminal

Conclusion: Metabarcoding analysis using Claident and R

Conclusion: Metabarcoding analysis using Claident and R

- Claident is integrated package for translation from high-throughput amplicon sequence data into ecological communities

Conclusion: Metabarcoding analysis using Claident and R

- Claident is integrated package for translation from high-throughput amplicon sequence data into ecological communities
- Claident can remove contaminants including index-hopped sequences using unused index combinations and blank samples (negative controls)

Conclusion: Metabarcoding analysis using Claident and R

- Claident is integrated package for translation from high-throughput amplicon sequence data into ecological communities
- Claident can remove contaminants including index-hopped sequences using unused index combinations and blank samples (negative controls)
- Most studies lack decontamination and this might affect the conclusion of such studies

Conclusion: Metabarcoding analysis using Claident and R

- Claident is integrated package for translation from high-throughput amplicon sequence data into ecological communities
- Claident can remove contaminants including index-hopped sequences using unused index combinations and blank samples (negative controls)
- Most studies lack decontamination and this might affect the conclusion of such studies
- Detection power of metabarcoding should re-evaluate using decontamination and our knowledge of that might need to be revised

Conclusion: Metabarcoding analysis using Claident and R

- Claident is integrated package for translation from high-throughput amplicon sequence data into ecological communities
- Claident can remove contaminants including index-hopped sequences using unused index combinations and blank samples (negative controls)
- Most studies lack decontamination and this might affect the conclusion of such studies
- Detection power of metabarcoding should re-evaluate using decontamination and our knowledge of that might need to be revised
- Claident provides multiple taxonomic assignment methods and can merge those results

Conclusion: Metabarcoding analysis using Claident and R

- Claident is integrated package for translation from high-throughput amplicon sequence data into ecological communities
- Claident can remove contaminants including index-hopped sequences using unused index combinations and blank samples (negative controls)
- Most studies lack decontamination and this might affect the conclusion of such studies
- Detection power of metabarcoding should re-evaluate using decontamination and our knowledge of that might need to be revised
- Claident provides multiple taxonomic assignment methods and can merge those results
- R can import tab-separated text made by Claident

Conclusion: Metabarcoding analysis using Claident and R

- Claident is integrated package for translation from high-throughput amplicon sequence data into ecological communities
- Claident can remove contaminants including index-hopped sequences using unused index combinations and blank samples (negative controls)
- Most studies lack decontamination and this might affect the conclusion of such studies
- Detection power of metabarcoding should re-evaluate using decontamination and our knowledge of that might need to be revised
- Claident provides multiple taxonomic assignment methods and can merge those results
- R can import tab-separated text made by Claident
- vegan is strongly recommended for community ecological analyses

Enjoy metabarcoding!