Environmental DNA: a new way to survey biodiversity in aquatic ecosystems

Dr. Liz Alter (sealter@gmail.com, ealter@csumb.edu)
California State University, Monterey Bay
City University of New York (Adjunct, Graduate Center)

Agenda

1. Learning objectives and introductions to each other
2. Introduction to eDNA: what is it? What are the strengths and weaknesses of this method?
3. Overview of the process: Sample collection, lab work and data analysis
4. Ways you can incorporate eDNA in the classroom

[Break]

5. Activity: data analysis
6. Wrap-up

Learning objectives

Understand the basic principles behind eDNA analysis and how it can be used to measure biodiversity

Explain the difference between the two major types of eDNA analysis

Describe the basic components of eDNA analysis from sampling to data analysis

Use a reference database to identify an unknown species

Introductions to each other!

- Liz Alter sealter@gmail.com; ealter@csumb.edu
- Twitter: montereyfishlab
- Your name, where you have taught or currently teach, a little about why you were interested in this workshop

1. Introduction to eDNA: what is it? What are the strengths and weaknesses of this method?

Organisms leave DNA behind

- Sloughed cells from skin and gut
- Injuries
- Decomposing tissue
- Digested tissue
- Gametes

Environmental DNA methods analyze this “left behind” DNA to learn about organisms that live in a particular habitat!

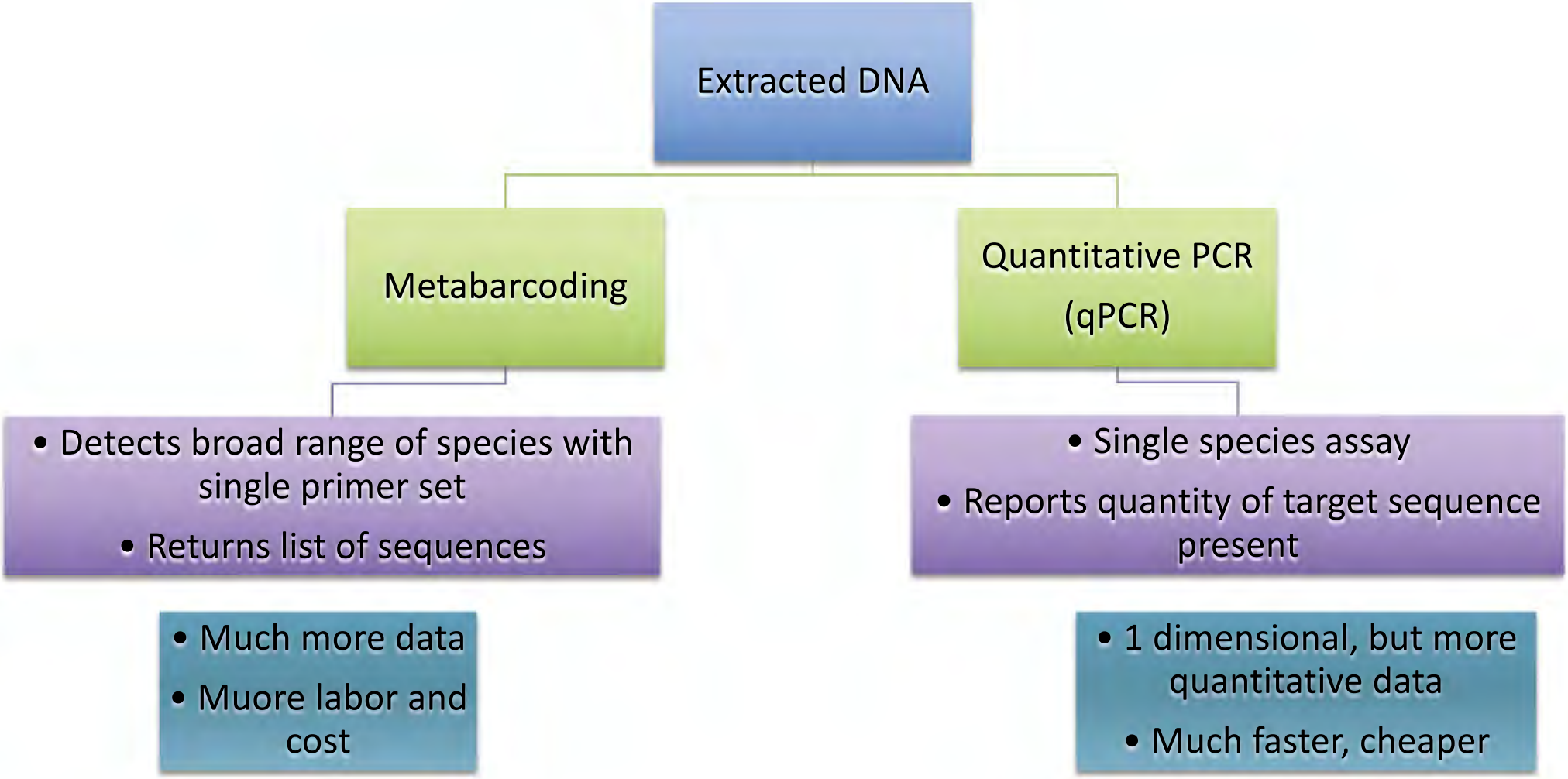
Why do we need eDNA methods?
Aquatic biodiversity is surprisingly poorly known

- Important to understand the spatio-temporal distribution of biodiversity for restoration and conservation planning
- But....relatively little scientific attention has been paid to these ecosystems
- Changes can occur rapidly from one year to the next

Challenges in surveying aquatic species diversity

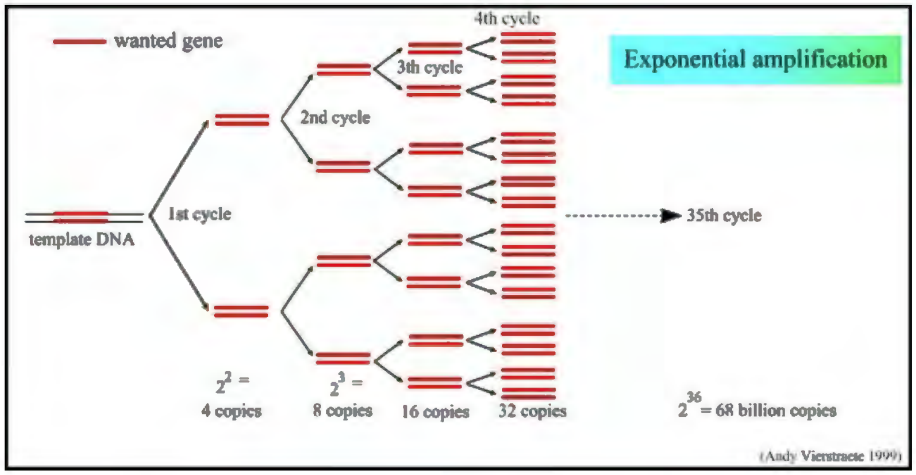
- Traditional survey methods are time and labor-intensive and therefore costly
- Can cause mortality, stress in both target and non-target species
- May miss cryptic taxa (rock-dwelling, etc)
- Difficulty in water access and safety concerns
- Many taxa are rare, small, cryptic → difficult to sample
- Different life stages can make identification challenging
- Fewer and fewer taxonomists

eDNA: 2 major methods



Quantitative PCR (qPCR)

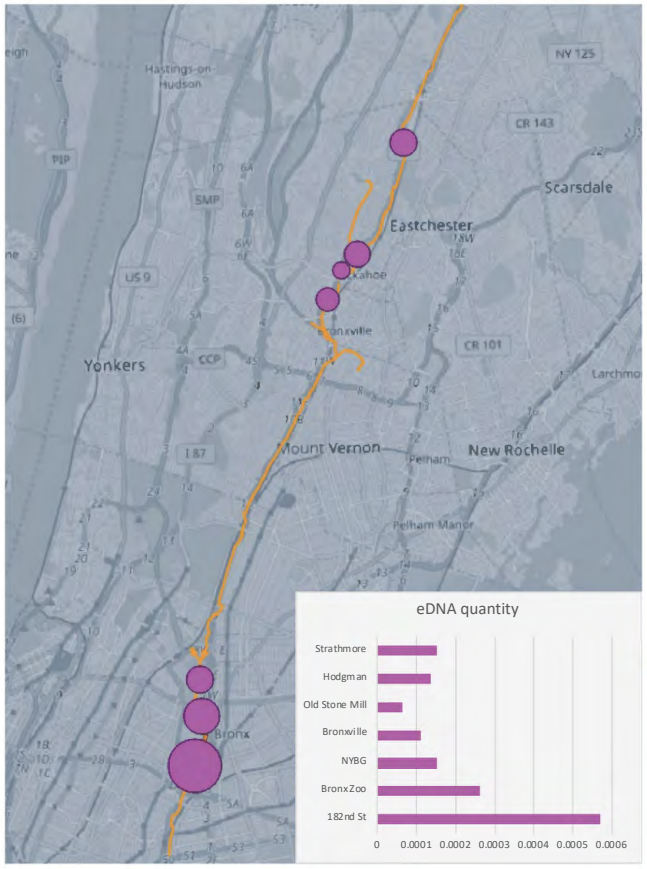
- Technique based on quantification of a fluorescent probe that sticks to the sequence of interest
- We measure increase in the amount of PCR product over time.
- The increase correlates inversely to the initial amount of DNA template

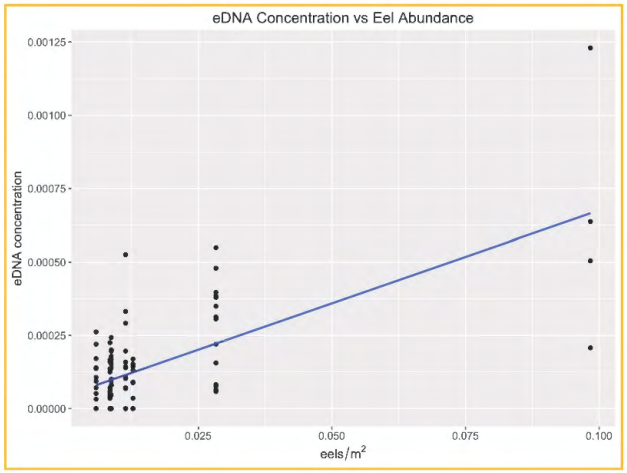


qPCR quantifies eels in the Bronx River

Abundance

eDNA concentration





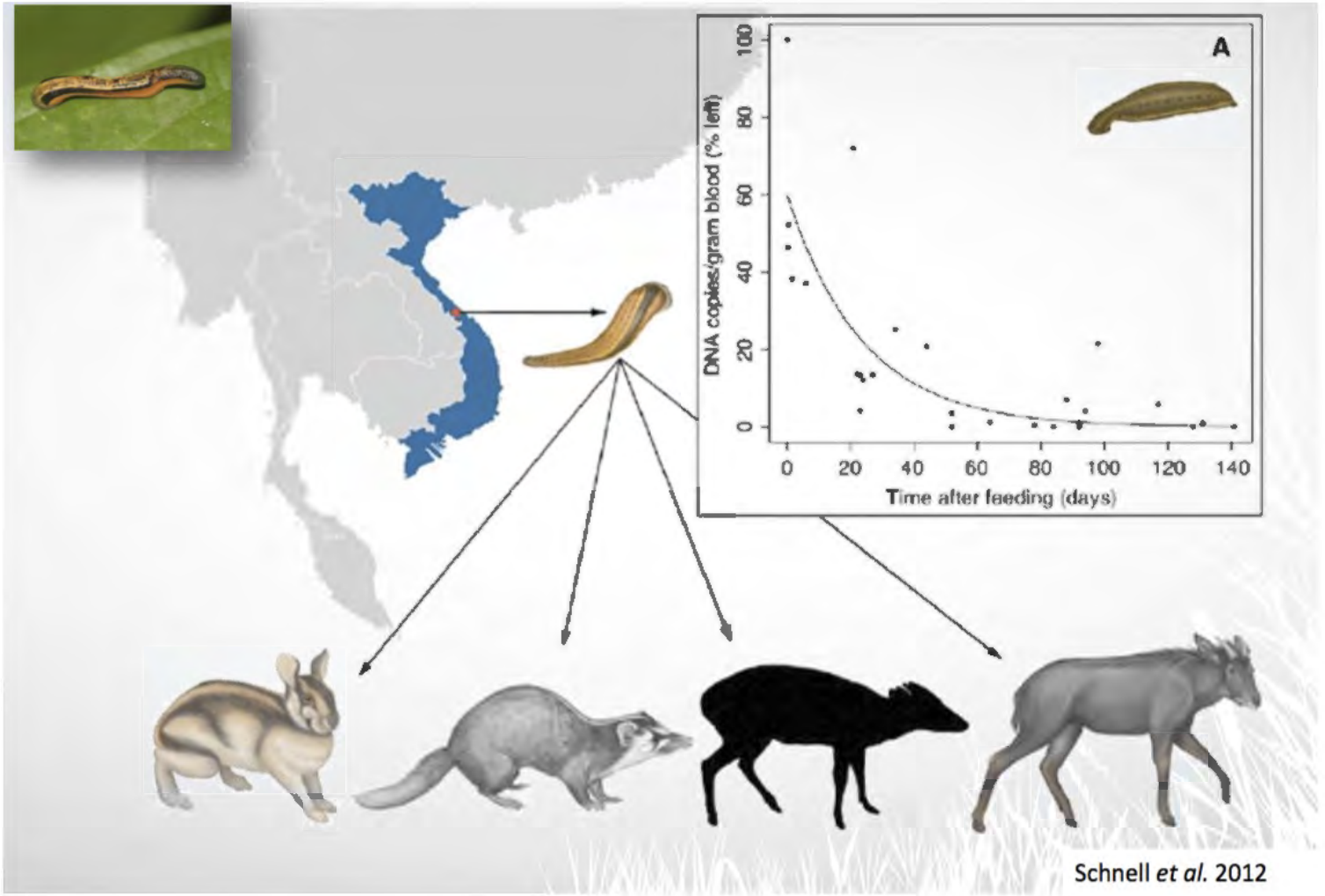
North American Journal of Fisheries Management

ISSN: 0275-5947 print / 1548-8675 online
DOI: 10.1002/nafm.10625

MANAGEMENT BRIEF

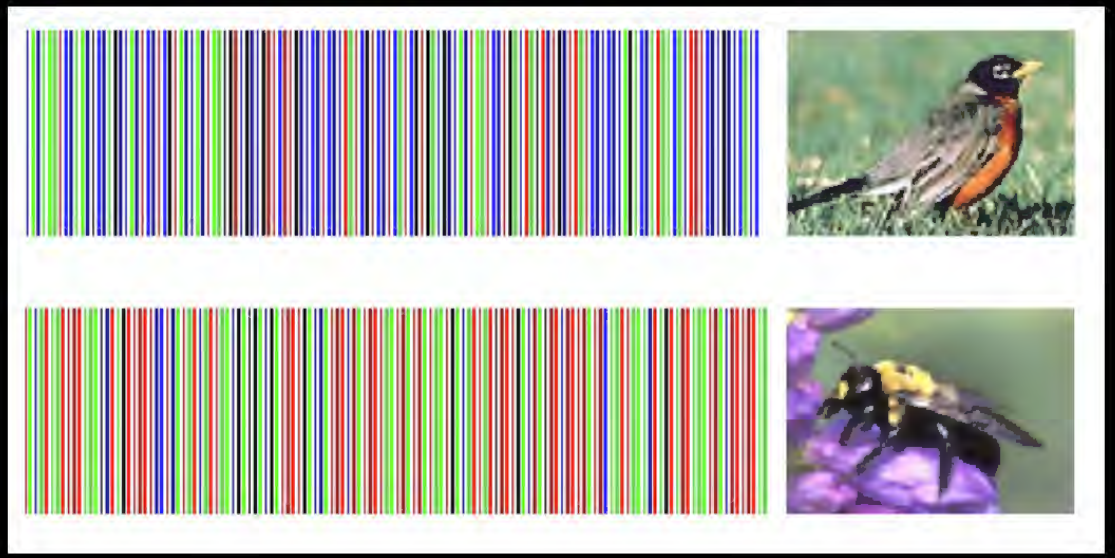
Relating American Eel Abundance to Environmental DNA Concentration in the Bronx River

Metabarcoding

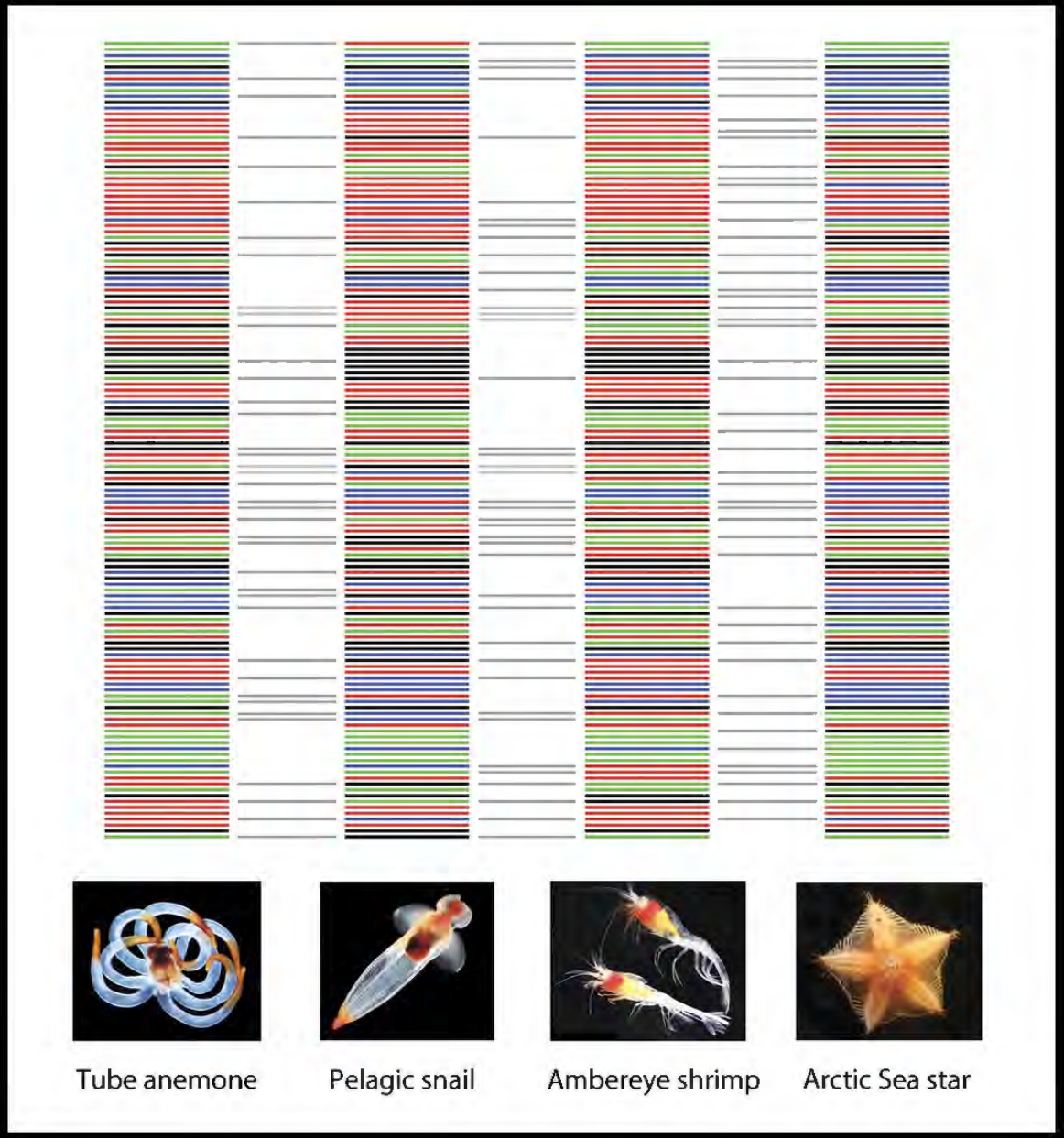


Traditional DNA barcoding

DNA barcoding: identifying species using short, standardized gene region(s)



New technique: metabarcoding for species identification (still relies on reference database)



Barcodes: Stoeckle
Images: Clarke-Hopcroft, Hopcroft, Bluhm, Iken

DNA extraction

Amplification of barcoding marker

Whole genome shotgun sequencing

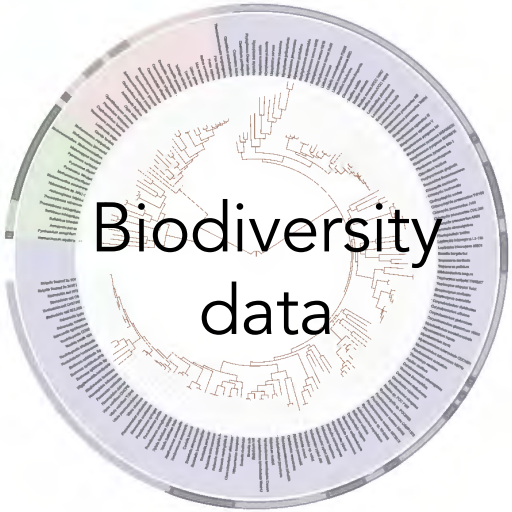
Sequencing

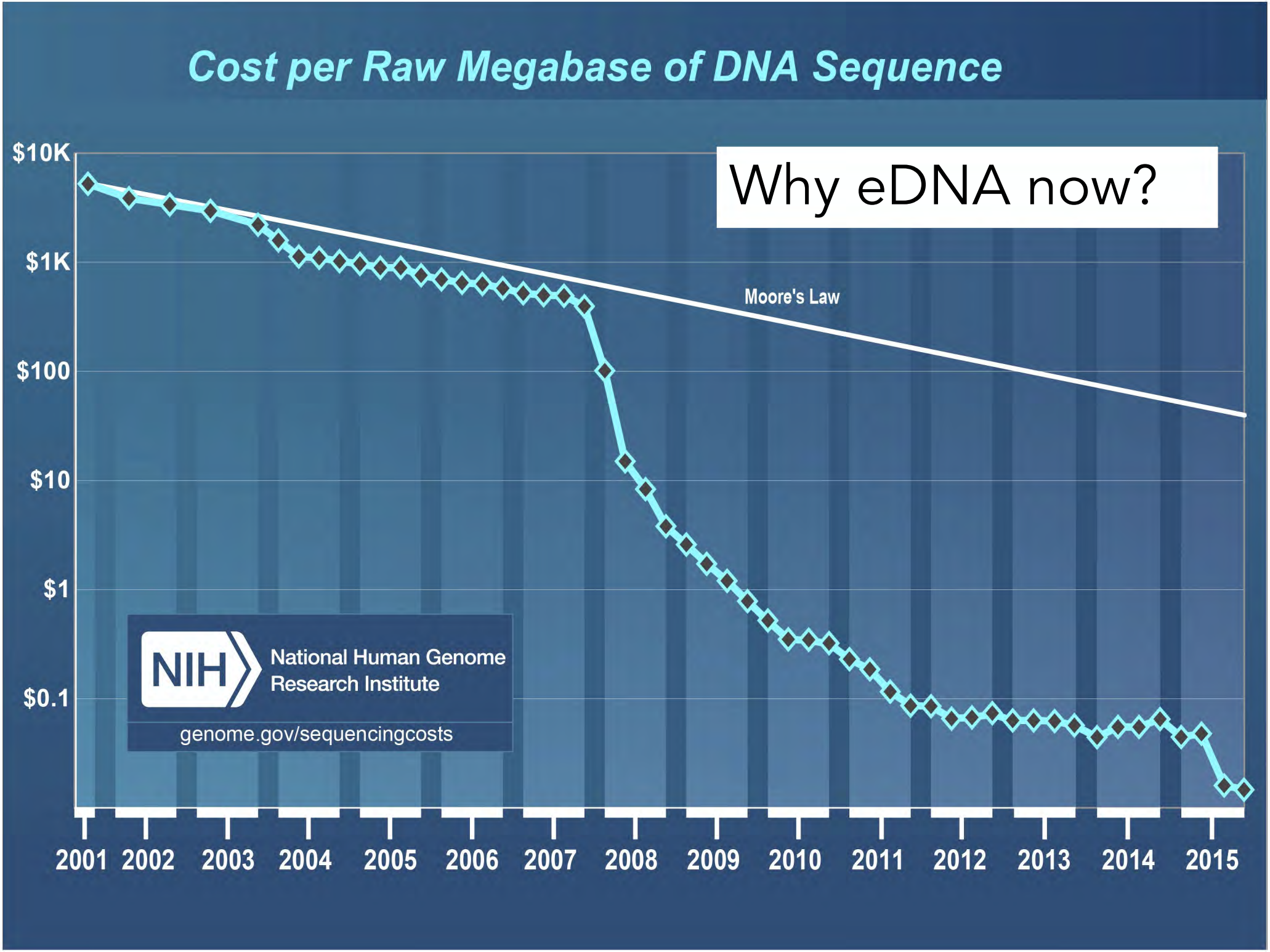
Quality filtering

Sequence binning

Assign taxonomy

Build phylogenetic tree

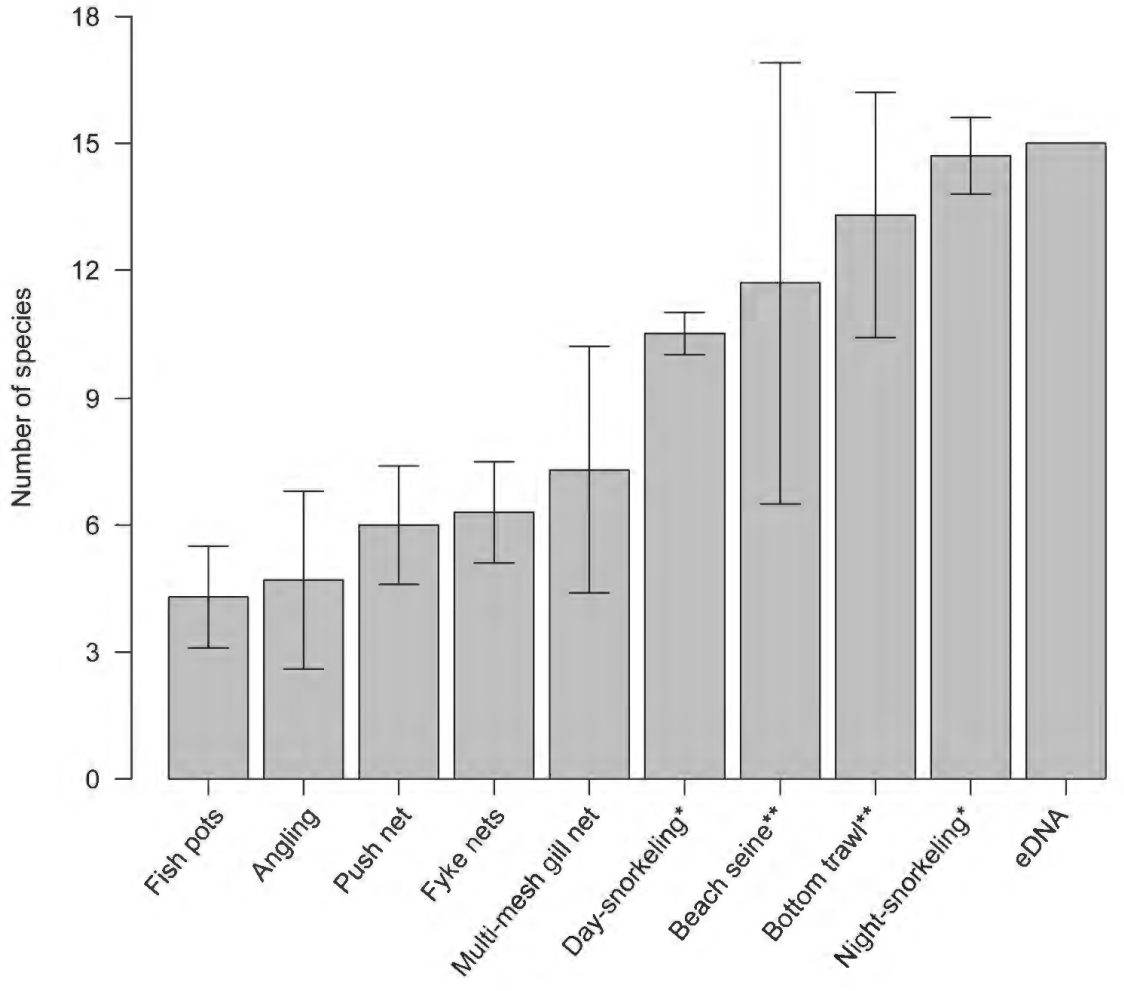



Cost per Raw Megabase of DNA Sequence
Why eDNA now?
$10K
$1K
$100
$10
$1
$0.1
Moore's Law
NIH
National Human Genome Research Institute
genome.gov/sequencingcosts
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015

Strengths of the eDNA approach

- Ease of sampling: no special expertise necessary, facilitates sampling across seasons, many habitats
- Can improve results after the fact, without re-sampling – long-term snapshot of diversity
- ID cryptic, invasive species & different life stages; migration and spawning behaviors
- Possible to ID species from many taxonomic groups simultaneously (species assemblages)

eDNA compared to other methods of surveys (Thompsen et al. 2012 PloS ONE)

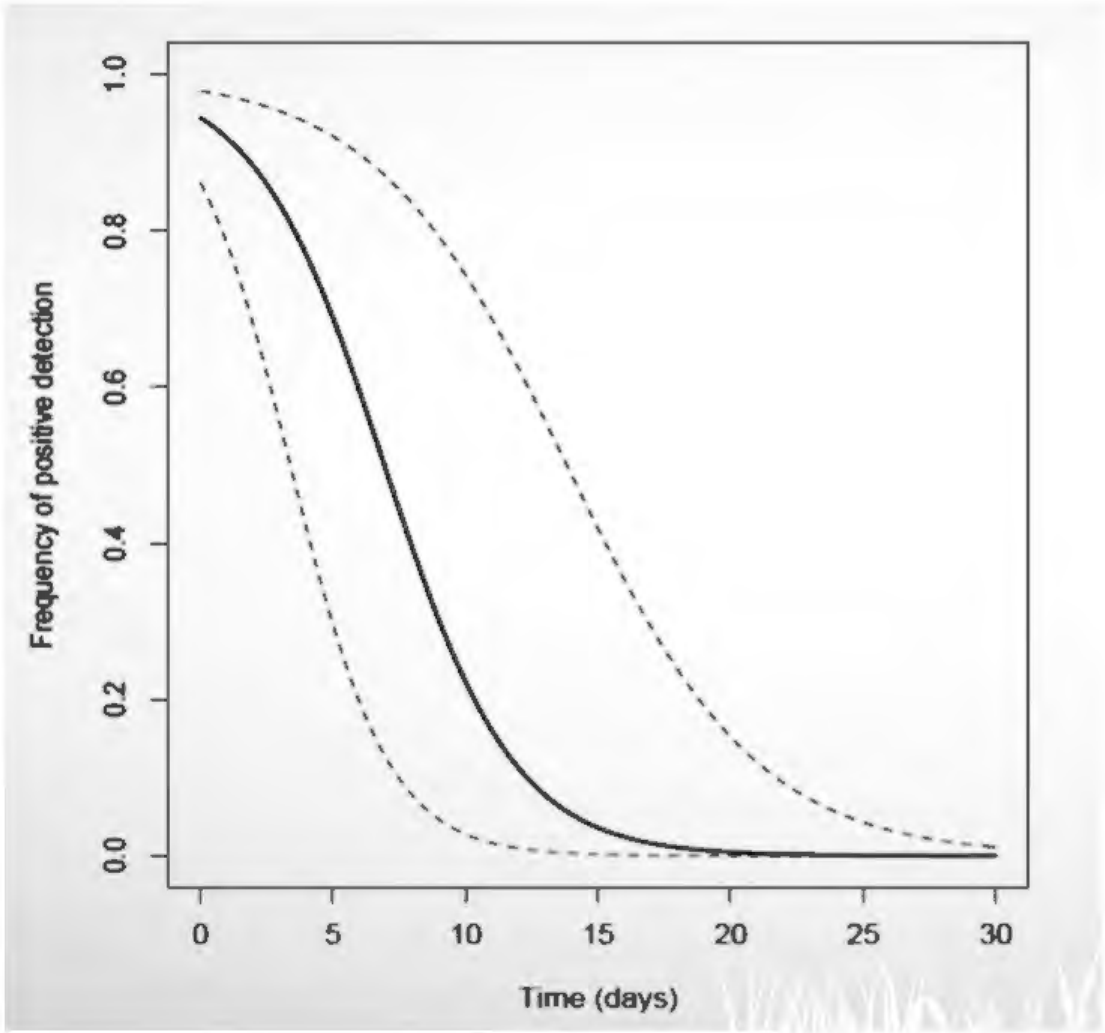


Weaknesses of the eDNA approach

- Contamination can be an issue and vigilance is required
- Many variables affect shedding rates and preservation of eDNA (pH, temperature, food availability, season, turbidity, UV)
- No single agreed-upon pipeline (yet)
- While sample collection is easy, sample processing and data collection requires specialized equipment and expertise
- Inferring abundance is tricky and may not be possible for many taxa/systems
- Accurate IDs depend on accurate reference database

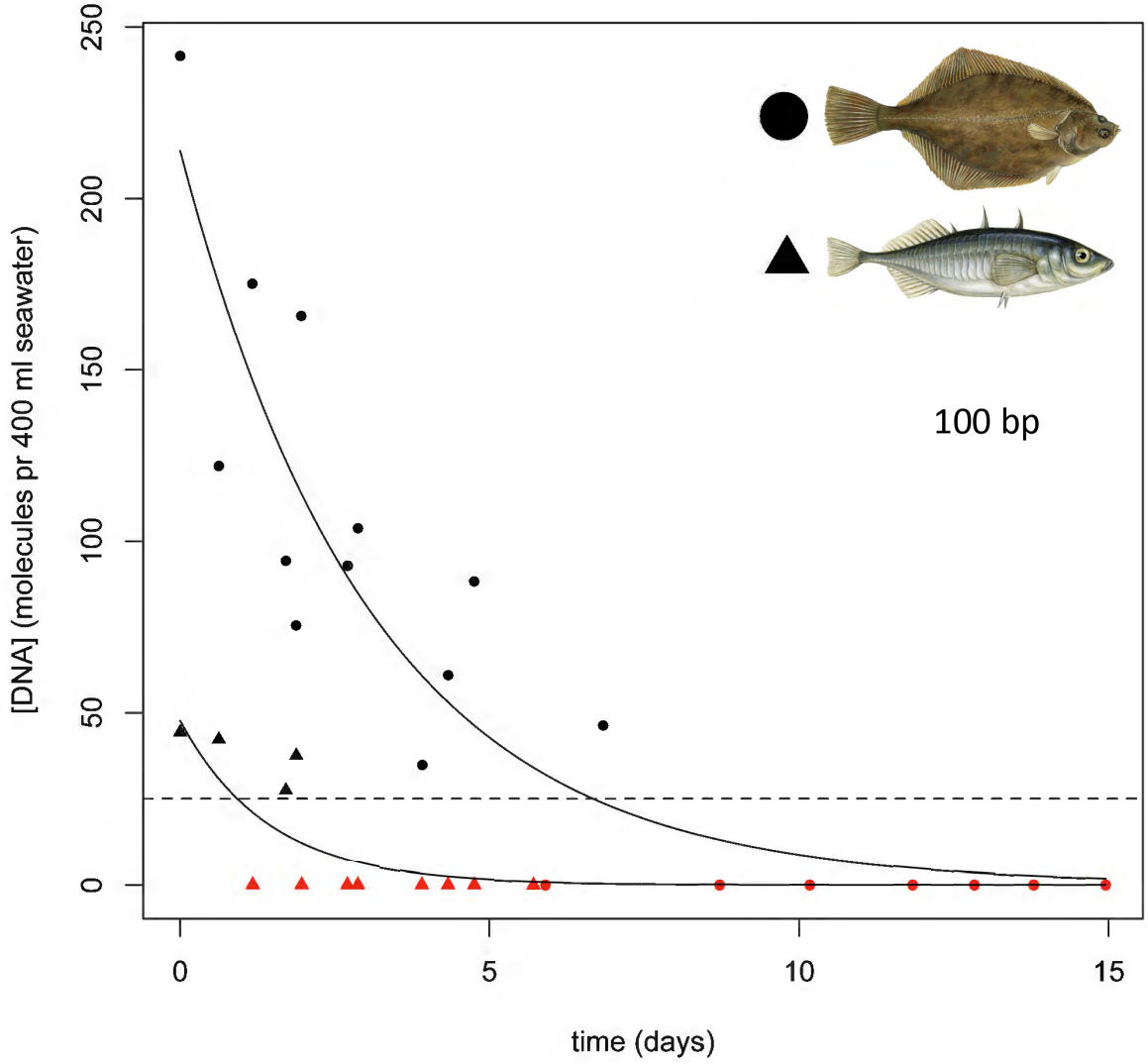
How long does eDNA persist in water?

- Can be on the order of hours-weeks
- In stagnant water versus flowing
- Temperature-dependent
- Life span in sediment >>> life span in water



Dejean et al. 2012

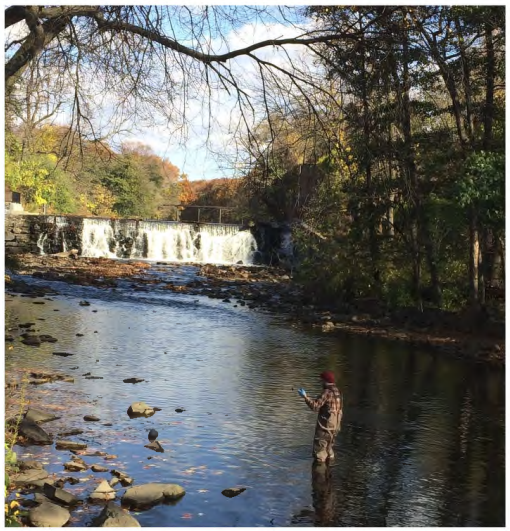
Degradation rates may be species- and environment-specific

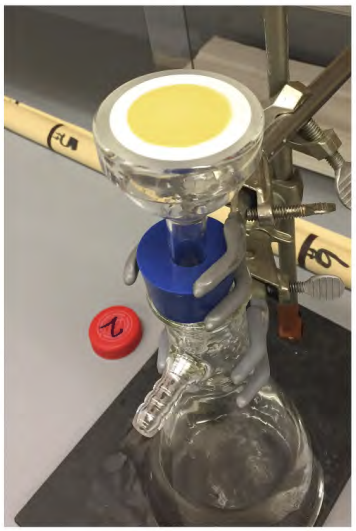


Thomsen et al.
2012 PLOS ONE

2. Overview of the process: Sample collection, lab work and data analysis

Field and lab workflow



1. Filter water sample, extract DNA from filter
2. Amplify (a segment of mitochondrial) DNA
   -60 base pair region 12S gene (ribosomal RNA)
   -primers designed for a particular taxonomic group
   -PCR replicates (3-10)
3. Combine PCR reps and sequence on Illumina or other platform
4. Filter and match sequence reads to NCBI database

Sampling in the field

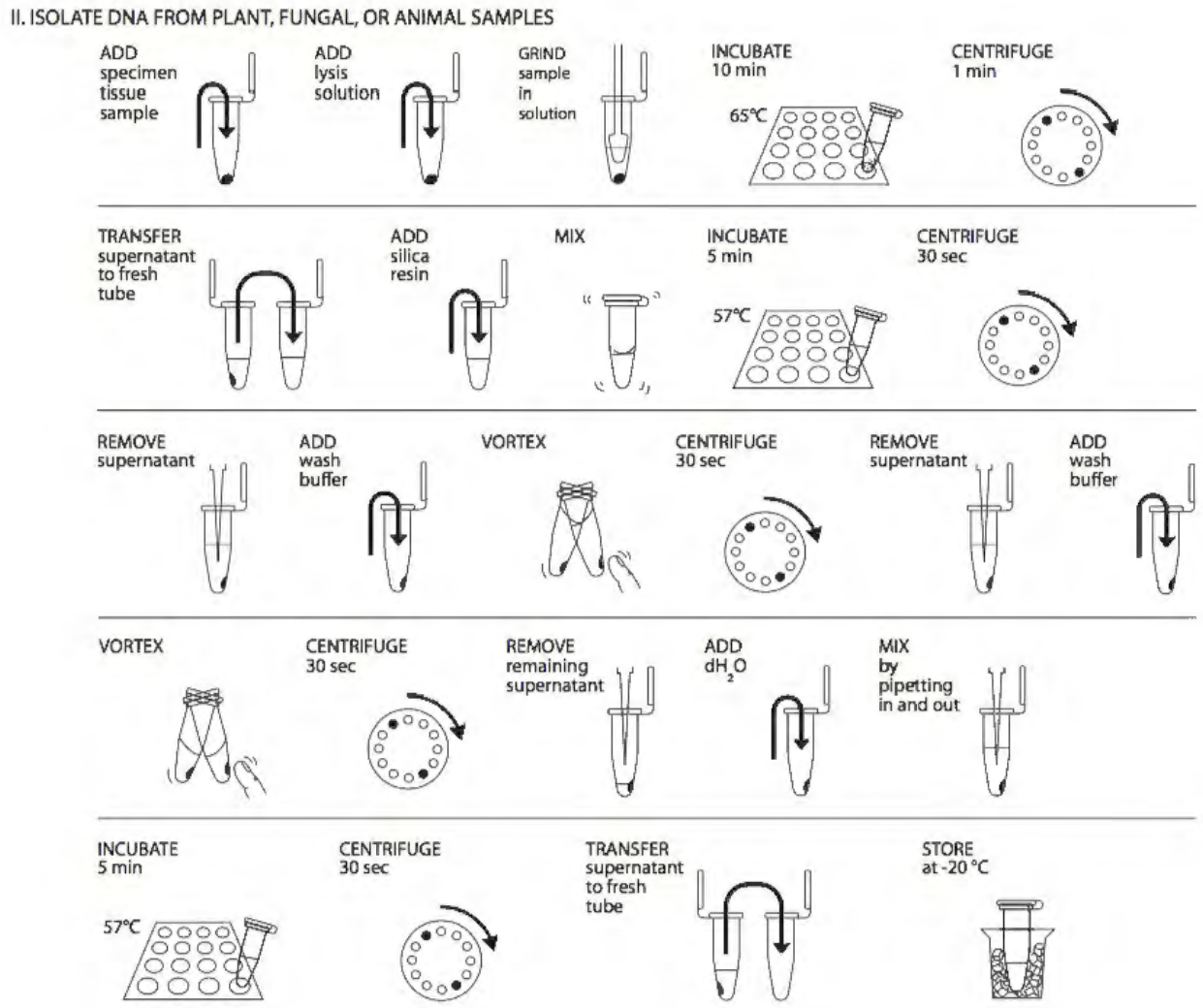
- When? Consider seasonality, dynamics over short-term (rain events, tidal cycles)
- How many samples/what volume?
- Where?
  - Water: surface, at depth, benthos
  - Soil: surface, cores

Sample preservation

- Filter on-site (water samples) & preserve filters
- Filters can be preserved in Ethanol, silica, DNA extraction buffer
- Snap-freezing in liquid nitrogen (not practical in the field most of the time)
- Some filters are self-preserving and stable at room temperature for months (Smith Root)

Extracting DNA from tissue

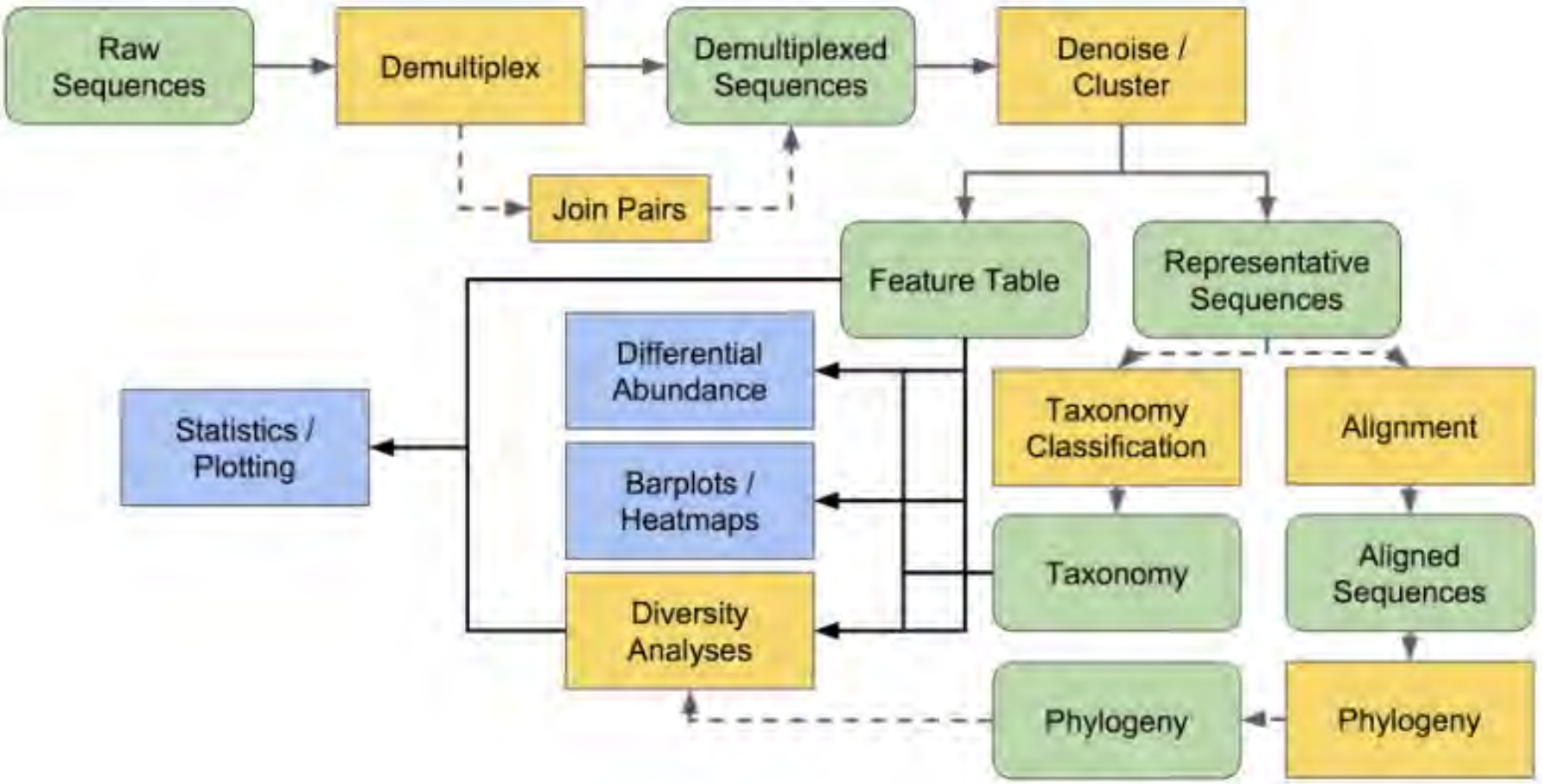
- Many different methods, from cheaper/faster (chelex) to more expensive/longer (Qiagen kits – PowerSoil, PowerWater)



Nearly all involve three steps: 1) an alkaline lysis step 2) selective adsorption or centrifugation in a high-salt buffer, and 3) finally elution of purified DNA in water or buffer.

DNA is stable frozen but will degrade if re-thawed many times!

Data analysis - overview



qiime2.org

Data analysis!

- After sequencing and data quality filtering, sequences will be matched against a reference database like NCBI to assign species identity

```
>*16S-0000002 | depth=42 | freq=2.31
TTCAACCTTGCGGTCGTACTCCCCAGGCGGAGTGCTTAATGCGTTAGCTGCGGCACTAAACCCCGGAAAGGGTCTAACACCTAGCACTCATCGTT
TACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAAGCCAGAGAGCCGCTTTCGCCACCG
GTGTTCCTCCATATATCTACGCATTTCACCGCTACACATGGAATTCCACTCTCCCCTCTTGCACTCAAGTTAAACAGTTTCCAAAGCGTACTATG
GTTAAGCCACAGCCTTTAACTTCAGACTTATCT
>*16S-0000019 | depth=12 | freq=0.66
TTCAGCCTTGCGGCCGTACTCCCCAGGCGGATTACTTATCGCATTCGCTTCGGCACAGACAGTCTTCCTGCCCACACCCAGTAATCATCGTTTAC
GGCCGGGACTACCAGGGTATCTAATCCTGTTCGCTCCCCCGGCTTTCGCACTTCAGCGTCAGTTACCGTCCAGTGAACTATCTTCATCATCGGCA
TTCCTGCACATATCTACGAATTTCACCTCTACTCGTGCAGTTCCGTCCACCTCTCCGGTACTCCAGCCTATCAGTTTCAAAGGCAGGCCTGCGGT
TGAGCCGCAGGTTTTCACCCCTGACTTGAAAGG
```

Alignment to references

AY053482.1

Sequence ID: lcl|Query_210570 **Length:** 1429 **Number of Matches:** 1

Range 1: 565 to 882 Graphics

Score	Expect	Identities	Gaps	Strand
588 bits(318)	7e-172	318/318(100%)	0/318(0%)	Plus/Minus

```
Query  1    TTCAACCTTGCGGTCGTACTCCCCAGGCGGAGTGCTTAATGCGTTAGCTGCGGCACTAAA  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  882  TTCAACCTTGCGGTCGTACTCCCCAGGCGGAGTGCTTAATGCGTTAGCTGCGGCACTAAA  823

Query  61   CCCCGGAAAGGGTCTAACACCTAGCACTCATCGTTTACGGCGTGGACTACCAGGGTATCT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  822  CCCCGGAAAGGGTCTAACACCTAGCACTCATCGTTTACGGCGTGGACTACCAGGGTATCT  763

Query  121  AATCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAAGCCAGAGAGCCGCT  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  762  AATCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAAGCCAGAGAGCCGCT  703

Query  181  TTCGCCACCGGTGTTCCTCCATATATCTACGCATTTCACCGCTACACATGGAATTCCACT  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  702  TTCGCCACCGGTGTTCCTCCATATATCTACGCATTTCACCGCTACACATGGAATTCCACT  643

Query  241  CTCCCCTCTTGCACTCAAGTTAAACAGTTTCCAAAGCGTACTATGGTTAAGCCACAGCCT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  642  CTCCCCTCTTGCACTCAAGTTAAACAGTTTCCAAAGCGTACTATGGTTAAGCCACAGCCT  583

Query  301  TTAACTTCAGACTTATCT  318
            ||||||||||||||||||
Sbjct  582  TTAACTTCAGACTTATCT  565
```

Alignment to references

CP001685.1

Sequence ID: lcl|Query_210571 **Length:** 1510 **Number of Matches:** 1

Range 1: 560 to 872 Graphics

Score	Expect	Identities	Gaps	Strand
490 bits(265)	2e-142	297/313(95%)	0/313(0%)	Plus/Minus

```
Query  1    TTCAGCCTTGCGGCCGTACTCCCCAGGCGGATTACTTATCGCATTCGCTTCGGCACAGAC  60
            ||||||||||||||||||||||||||||||||||||||||||| |||||||||| |||
Sbjct  872  TTCAGCCTTGCGGCCGTACTCCCCAGGCGGATTACTTATCGCATTAGCTTCGGCACGGAC  813

Query  61   AGTCTTCCTGCCCACACCCAGTAATCATCGTTTACGGCCGGGACTACCAGGGTATCTAAT  120
            | ||||                                                |||||||
Sbjct  812  ACTCTT                                                ATCTAAT  753

Query  121  CCTGTT                                                TATCTTC  180
            ||||||                                                |||||||
Sbjct  752  CCTGTT                                                TATCTTC  693

Query  181  ATCATCGGCATTCCTGCACATATCTACGAATTTCACCTCTACTCGTGCAGTTCCGTCCAC  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  692  ATCATCGGCATTCCTGCACATATCTACGAATTTCACCTCTACTCGTGCAGTTCCGTCCAC  633

Query  241  CTCTCCGGTACTCCAGCCTATCAGTTTCAAAGGCAGGCCTGCGGTTGAGCCGCAGGTTTT  300
            |||||| | |||| |||| | ||||||| | ||||||| ||||||||||||||| ||||||
Sbjct  632  CTCTCCAGCACTCTAGCCAAACAGTTTCCAGGGCAGGCTTGCGGTTGAGCCGCAAGTTTT  573

Query  301  CACCCCTGACTTG  313
            |||||| ||||||
Sbjct  572  CACCCCAGACTTG  560
```

Around 95-97% of identity is required in the alignment of an OTU sequence to a database reference

Final dataset

OTU representative sequences

OTU taxonomy assignments

Samples and OTU frequencies

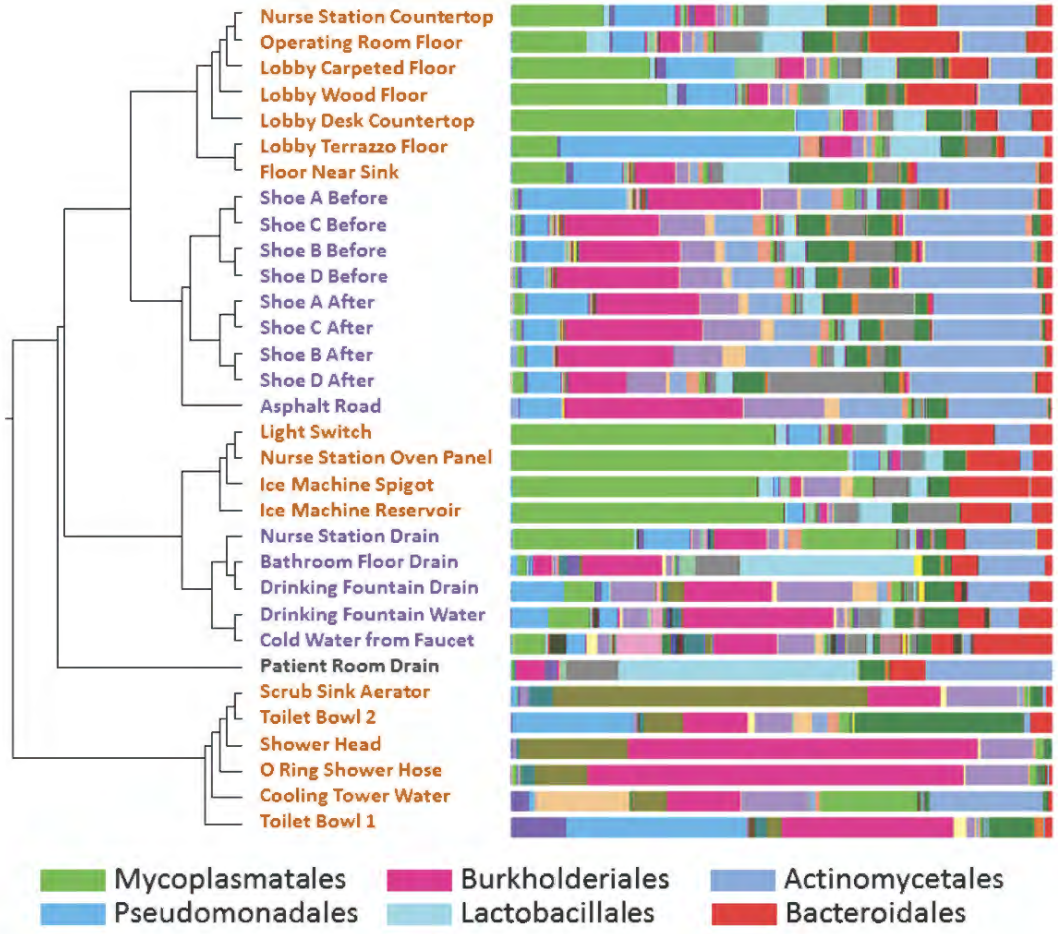
			COUNT_OTUS	19	14	13	15	14
SEQUENCES	MEAN_FR	SAMPLES	OTU	SRS052681	SRS042606	SRS042483	SRS048589	SRS051454
16S: TTCAACCTTGCGGTCG	0.0393	5	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Lactobacillales,f:Streptococcaceae,g:Streptococcus,s:pseudopneumoniae;	0.0061	0.0893	0.0094	0.0089	0.083
16S: TTCATACTTGCGTACG	0.0707	5	tax=k:Bacteria,p:Fusobacteria,c:Fusobacteria (class),o:Fusobacteriales,f:Fusobacteriaceae,g:Fusobacterium;	0.0683	0.0918	0.0031	0.0529	0.1374
16S: TTCACCGTTGCCGGCG	0.0557	5	tax=k:Bacteria,p:Bacteroidetes,c:Bacteroidia,o:Bacteroidales,f:Porphyromonadaceae,g:clone,s:HF001;	0.0463	0.0494	0.0682	0.0575	0.057
16S: TTTAGCCTTGCGGCCG	0.0815	2	tax=k:Bacteria,p:Actinobacteria,c:Actinobacteria (class),o:Actinomycetales,f:Corynebacteriaceae,g:Corynebacterium,s:matruchotii;	0.0628			0.1001	
16S: TTTAATCTTGCGACCG	0.0291	5	tax=k:Bacteria,p:Proteobacteria,c:Betaproteobacteria,o:Neisseriales,f:Neisseriaceae,g:Neisseria;	0.0061	0.0963	0.0055	0.0065	0.0311
16S: TTCAACCTTGCGGTCG	0.0246	5	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Veillonellaceae,g:Veillonella;	0.0226	0.0374	0.0086	0.0168	0.0376
16S: TTCACCGTTGCCGGCG	0.0124	4	tax=k:Bacteria,p:Bacteroidetes,c:Bacteroidia,o:Bacteroidales;	0.0127	0.0089	0.0071		0.0207
16S: TTCAGCCTTGCGGCCG	0.0093	5	tax=k:Bacteria,p:Fusobacteria,c:Fusobacteria (class),o:Fusobacteriales,f:Fusobacteriaceae,g:Leptotrichia,s:buccalis;	0.0242	0.0038	0.0031	0.0098	0.0058
16S: TTTAGCCTTGCGGCCG	0.0127	3	tax=k:Bacteria,p:Actinobacteria,c:Actinobacteria (class),o:Actinomycetales,f:Actinomycetaceae,g:Actinomyces,s:odontolyticus;	0.0039	0.031	0.0031		
16S: TTCACCGTTGCCGGCG	0.008	4	tax=k:Bacteria,p:Bacteroidetes,c:Bacteroidia,o:Bacteroidales,f:Prevotellaceae,g:Prevotella;	0.0072	0.0127	0.0055		0.0065
16S: TTCAACCTTGCGGTCG	0.0141	2	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Lactobacillales,f:Enterococcaceae,g:Enterococcus;		0.0177			0.0104
16S: TTCATTCTTGCGAACG	0.0093	3	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Lachnospiraceae;		0.0139		0.0042	0.0097
16S: TTCATTCTTGCGAACG	0.0086	3	tax=k:Bacteria,p:Actinobacteria;	0.0121		0.0063	0.0075	
16S: TTTAGCCTTGCGGCCG	0.0115	2	tax=k:Bacteria,p:Actinobacteria,c:Actinobacteria (class),o:Actinomycetales,f:Actinomycetaceae,g:Actinomyces,s:oris;	0.0182			0.0047	
16S: TTCACACTTGCGTGCG	0.0114	2	tax=k:Bacteria,p:Bacteroidetes,c:Flavobacteria,o:Flavobacteriales,f:Flavobacteriaceae,g:Capnocytophaga,s:sputigena;			0.0047		0.0181
16S: TTCATTCTTGCGAACG	0.0093	2	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Lachnospiraceae,g:Oribacterium,s:sp. oral taxon 078;		0.0095			0.0091
16S: TTCACTCTTGCGAGCG	0.006	3	tax=k:Bacteria,p:Bacteroidetes,c:Flavobacteria,o:Flavobacteriales,f:Flavobacteriaceae;	0.0033	0.0063		0.0084	
16S: TTCACCGTTGCCGGCG	0.0051	3	tax=k:Bacteria,p:Bacteroidetes,c:Bacteroidia,o:Bacteroidales,f:Prevotellaceae,g:Prevotella,s:nigrescens;	0.0066	0.0057	0.0031		
16S: TTCAGCCTTGCGGCCG	0.007	2	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Veillonellaceae,g:Selenomonas,s:noxia;	0.0083			0.0056	
16S: TTCAGTGTTGCCACCG	0.0055	2	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Clostridiales Family XI. Incertae Sedis,s:Parvimonas micra;	0.0077				0.0032
16S: TTCACCCTTGCGGGCA	0.0077	1	tax=k:Bacteria,p:Spirochaetes,c:Spirochaetes (class),o:Spirochaetales,f:Spirochaetaceae,g:Treponema,s:socranskii;	0.0077				
16S: TTTAATCTTGCGACCG	0.0075	1	tax=k:Bacteria,p:Proteobacteria,c:Betaproteobacteria,o:Burkholderiales,f:Burkholderiaceae;				0.0075	
16S: TTCAGTCTTGCGACCG	0.0061	1	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Veillonellaceae,g:Selenomonas;	0.0061				
16S: TTCAACCTTGCGGCCG	0.0061	1	tax=k:Bacteria,p:Proteobacteria,c:Betaproteobacteria,o:Burkholderiales,f:Comamonadaceae;				0.0061	
16S: TTCATTCTTGCGAACG	0.005	1	tax=k:Bacteria,p:Bacteroidetes,c:Bacteroidia,o:Bacteroidales,f:Porphyromonadaceae;	0.005				
16S: TTTAACCTTGCGGTCG	0.0039	1	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales;			0.0039		
16S: TTTATTCTTGCGAACG	0.0037	1	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Eubacteriaceae,g:Eubacterium;				0.0037	
16S: TTCATTCTTGCGAACG	0.0032	1	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Lachnospiraceae,g:Catonella;					0.0032

What kinds of questions can you answer with taxonomic tables?

- What species are present at each site and how do they compare across sites?
- Does overall diversity vary across sites? (diversity indices)
- How many more species do you detect with each additional sample (rarefaction)?
- How does the composition of the community compare across sites?
- Are nonnative/pathogenic species present?

Downstream analysis

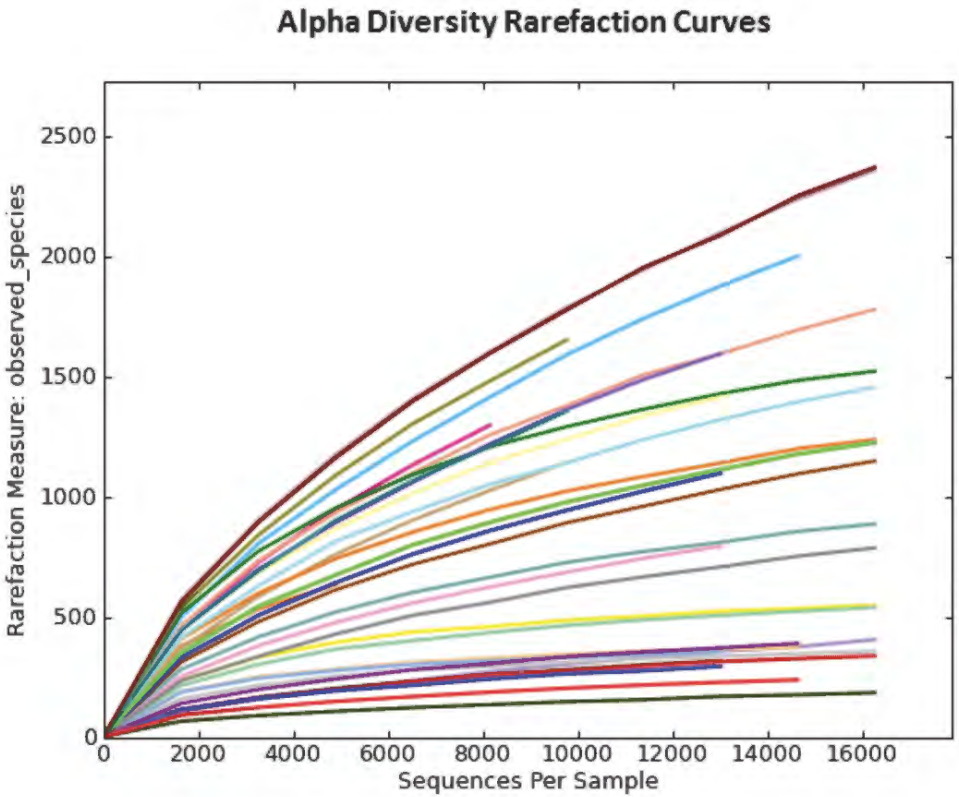
Taxonomy summaries:

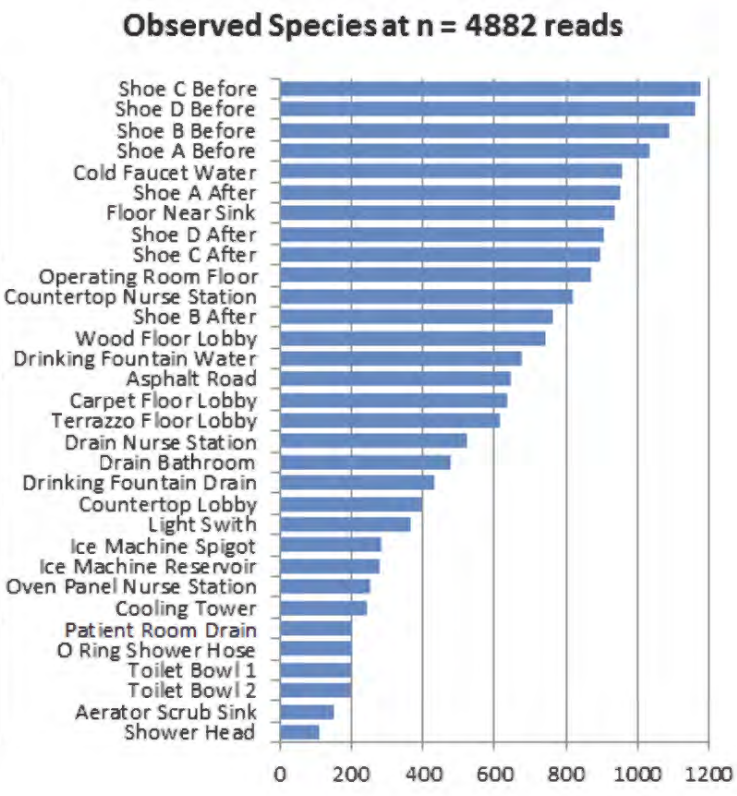


http://hospitalmicrobiome.com/construction-samples/

Downstream analysis

Alpha diversity measurements and rarefaction plots:

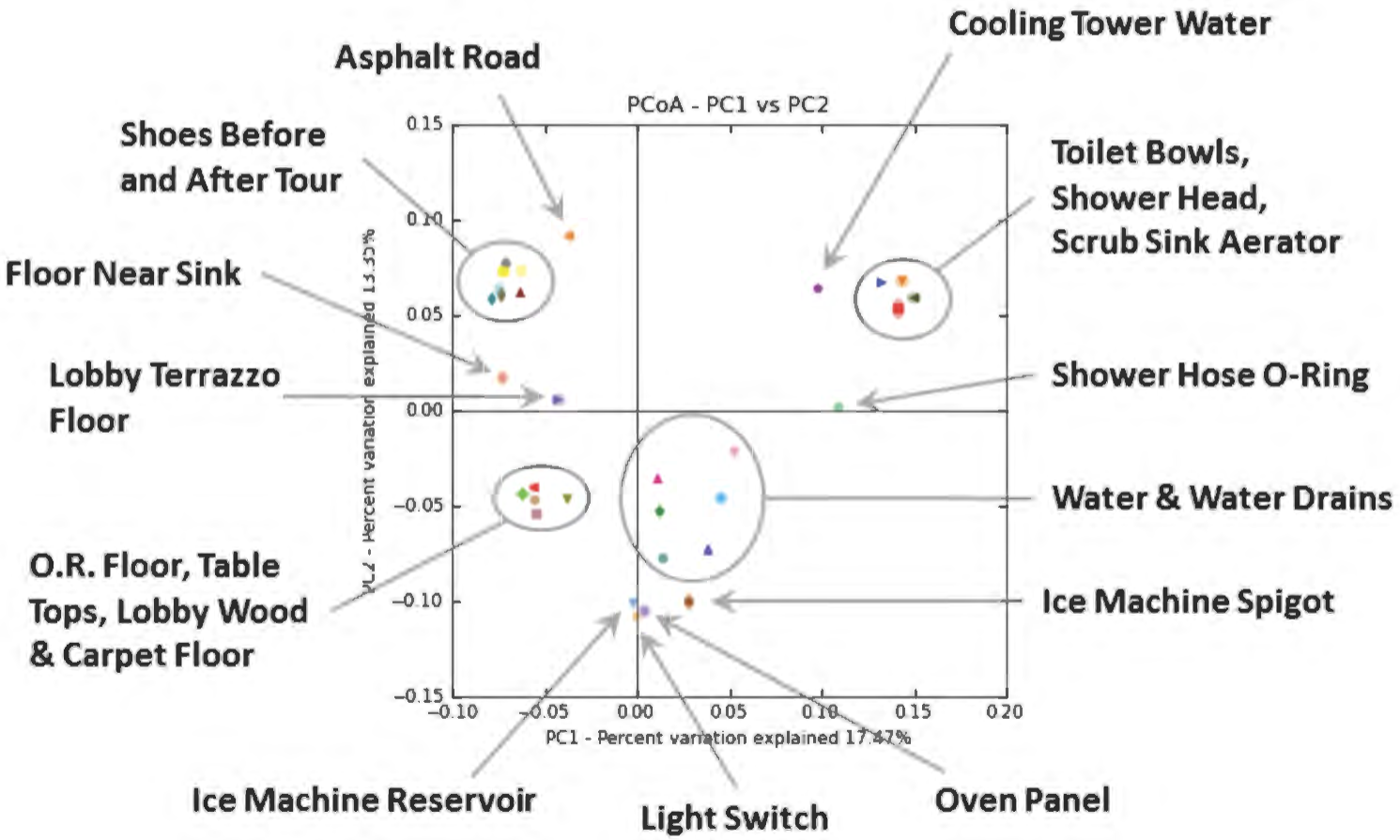




http://hospitalmicrobiome.com/construction-samples/

Downstream analysis

Principal Coordinate Analysis (PCoA):



http://hospitalmicrobiome.com/construction-samples/

Sixthresearcher.org

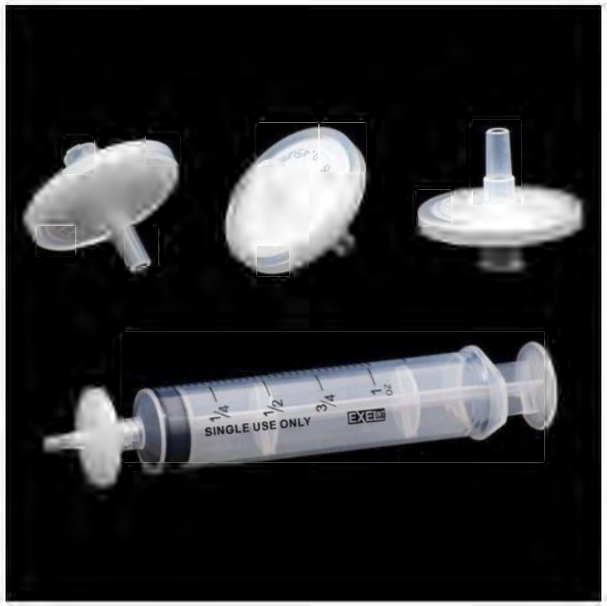
Some tools for analysis

- DNA Subway: DNAsubway.cyverse.org – implements QIIME2
- Excel for basic stats/figures!
- R/R studio: VEGAN package for diversity analysis (will estimate diversity indices, create rarefaction curve, PCoA plot and more)

3. Ways you can incorporate eDNA in the classroom

Field sampling and water filtration

- Students can collect water along with environmental data (temperature, turbidity etc)
- Water can be filtered on site using syringe filters, or in the lab with a vacuum pump
- Filters can be frozen for later DNA extraction
- If extraction and PCR equipment is available, students can do a single-species PCR to detect presence/absence
- Otherwise collaboration with a lab is needed…OR…



Jonah Ventures

https://store.jonahventures.com/products/aquatic-edna-kit-single








Aquatic eDNA kit (fish + phytoplankton)

$89.00

ADD TO CART

Buy with G Pay

More payment options

Using the JonahWater aquatic environmental DNA kit, you can reconstruct aquatic assemblages in your local water body! By filtering water and then sequencing the DNA on the filters, we can tell you the species of fish and algae that live in your neighborhood.

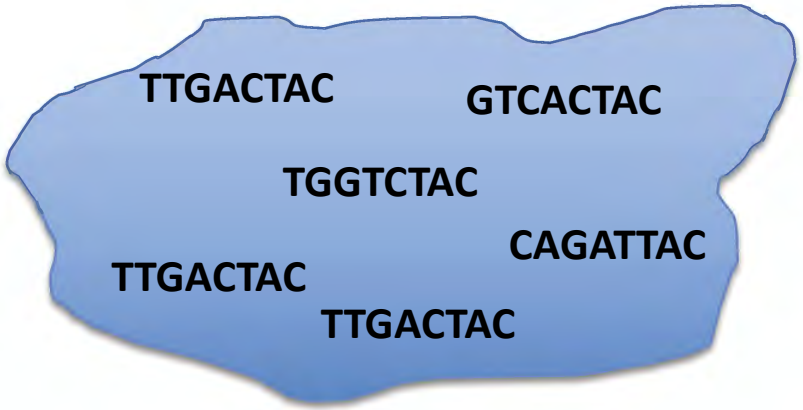
Simply collect water in the syringe, push it through the filter, and mail the filter to us in the supplied barcoded sample cup and return shipping envelope. We do all the DNA extraction and analysis and send you back the results!

Included with these kits is access to a convenient JonahDNA phone app, which can be used to enter meta-data. Our convenient JonahDNA data portal is used to access the data you collected

Low-tech eDNA exercise (no computers needed!)

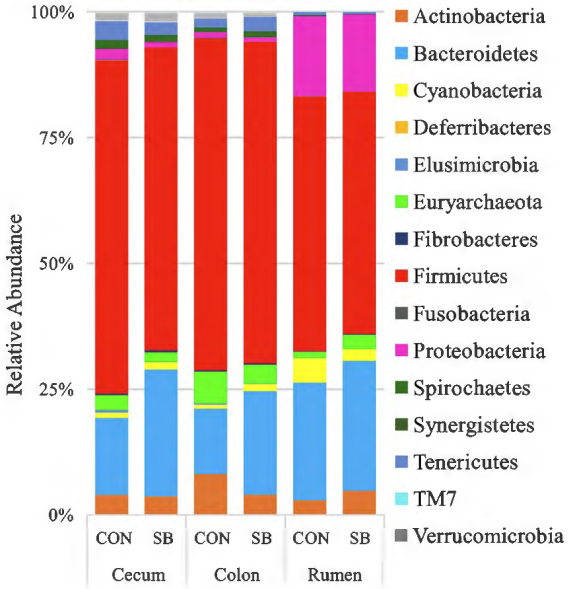
[inspired by/modified from a JV exercise]

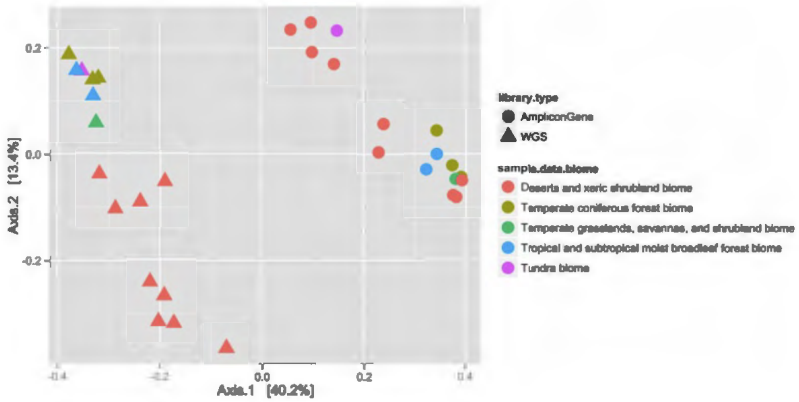
- Create a “lake” or body of water in the classroom (this could be a part of an area rug, bucket etc) – sprinkle paper DNA sequences representing fish DNA barcodes (random or designate habitats)
- Have students “sample” eDNA by randomly picking out a set of barcodes/paper strips, and then matching against eDNA reference database (e.g. laminated pictures of fish species with printed sequences)
- Have students make a sampling curve: simple chart of how many species they recover with 1, 2, 3, etc… ”samples” -> quantitative reasoning/graphing
- For upper middle/high school students, combine with a lesson on PCR and complementarity of bases in DNA, diversity patterns, etc



Data analysis

- Advanced students can use the resulting data (taxonomic table) to perform statistics in either Excel or R
- Quantify diversity metrics (Shannon and Simpson indices) (Excel or R)
- Create stacked barplots to show proportions of taxa in each sample (Excel or R)
- Principle components analysis to show community composition (Vegan package in R)





4. Hands-on activity: identifying species from DNA sequences using a reference database!

```
>*16S-0000002 | depth=42 | freq=2.31
TTCAACCTTGCGGTCGTACTCCCCAGGCGGAGTGCTTAATGCGTTAGCTGCGGCACTAAACCCCGGAAAGGGTCTAACACCTAGCACTCATCGTT
TACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAAGCCAGAGAGCCGCTTTCGCCACCG
GTGTTCCTCCATATATCTACGCATTTCACCGCTACACATGGAATTCCACTCTCCCCTCTTGCACTCAAGTTAAACAGTTTCCAAAGCGTACTATG
GTTAAGCCACAGCCTTTAACTTCAGACTTATCT
>*16S-0000019 | depth=12 | freq=0.66
TTCAGCCTTGCGGCCGTACTCCCCAGGCGGATTACTTATCGCATTCGCTTCGGCACAGACAGTCTTCCTGCCCACACCCAGTAATCATCGTTTAC
GGCCGGGACTACCAGGGTATCTAATCCTGTTCGCTCCCCCGGCTTTCGCACTTCAGCGTCAGTTACCGTCCAGTGAACTATCTTCATCATCGGCA
TTCCTGCACATATCTACGAATTTCACCTCTACTCGTGCAGTTCCGTCCACCTCTCCGGTACTCCAGCCTATCAGTTTCAAAGGCAGGCCTGCGGT
TGAGCCGCAGGTTTTCACCCCTGACTTGAAAGG
```

5. Wrap-up

- Questions?
- Discussion (time allowing): how might you introduce eDNA in the classroom?