Comparing metagenomics and total RNA-Seq for freshwater microbiome assessments

Intro

1. Freshwater biomonitoring why
2. Entire community inclusion advocated
3. Not possible with traditional methods, but DNA based
4. Shotgun seqiencog based approaches, such as metagenomcis and metatranscriptomics, present such an option while avoiding primer bias common in metabracodng
5. metagenomics just small percentage of most common microbial barcodes ssu and lsu
6. In comparison, total RNA-Seq natural enrichment for ssu and lsu, and can be even further used for mRNA analysis (used in soil and cow gut studies)
7. Comparsion of both methods needed

(extraction of ssu reads from data, metaxa, but assembly preferred, quince 2017 and anders paper)

Introduction

PART 1 FRESHWATER ASSESSMENT AND BIOMONITORING, CAN POTENTIALLY BE MUCH SHORTENED

Freshwater systems are valuable for economic productivity, ecosystem resilience, and for maintaining ecosystem services, including the supply of clean consumable water (reviewed in Dudgeon et al. 2006). However, these systems are heavily impacted by agricultural and industrial pollution, habitat fragmentation (wetland drainage, river straightening, and dam building), and the introduction of invasive species (Dextrase and Mandrak, 2006; Jensen et al., 2006). Consequently, freshwater systems belong to the most threatened ecosystems with some of the highest rates of species loss (Malmqvist and Rundle, 2002).

To prevent this species loss, the natural status of freshwater bodies needs to be protected, preserved, and restored. Therefore, the ecological status of individual freshwater bodies must be identified. This knowledge helps to determine if a freshwater body’s natural status is threatened and if restoration strategies need to be applied. The determination of an ecological status is accomplished by freshwater biomonitoring.

Traditionally, freshwater biomonitoring is assessed with bioindicators, most commonly benthic macroinvertebrates, but also diatoms, macrophytes, zooplankton, and fish (Bellinger and Sigee, 2015; Haury et al., 2006; Jeppesen et al., 2011; Karr, 1981; Resh and Unzicker, 1975). However, single-cellular eukaryotes (hereafter referred to as protists) and prokaryotes were proposed as nearly equally good if not better bioindicators, because they respond faster to environmental changes (Ferdous, 2009; Foissner and Berger, 1996; McArthur, 2001; Payne, 2013; Smith et al., 2015; Stoeck et al., 2018). Consequently, recent studies advocate for the inclusion of more inclusive communities, i.e., entire microbial communities, into ecological assessments and biodiversity studies in general (Pawlowski et al., 2016; Ritter et al., 2019).

PART 2: DNA-BASED APPROACH

PART 2.1: METABARCODING

However, morphologic identification of the microbiome most protist groups and prokaryotes is not practicable because of the lack of diagnostical traits, and DNA-based methods can sometimes be the only option to distinguish organisms (Pawlowski et al., 2012; Will and Rubinoff, 2004). DNA-based biomonitoring methods have been proposed as a GOOD ADDITION to morphologic identification. While DNA Metabarcoding (Taberlet et al., 2012) has established as an effective, DNA-based biomonitoring METHOD REF, the method is also known to BRING ERRORS. The main issue of DNA Metabarcoding is primer bias introduced during DNA amplification, due to varying primer binding affinities (Alberdi et al. 2018; Elbrecht and Leese 2015; Krehenwinkel et al. 2017; Piñol et al. 2014; Piper et al. 2019). Additionally, PCR bias is introduced during the PCR process by GC preference of the polymerase (Nichols et al. 2018; Sze and Schloss 2019), PCR inhibition (Jane et al. 2015; Taberlet et al. 1996), and chimera formation (Piper et al. 2019; Sze and Schloss 2019). ANOTHER OPTION IS TO USE SHOTGUN SEQUENCING. ALTHOUGH METABARCODING WAS SHOWN TO OUTPERFORM SHOTGUN SEQUENCING IN TERMS OF TAXONOMIC PROFILING IN SOME INSTANCES REF, OTHER STUDIES SHOW THAT THESE TECHNIQUES CAN OUTPERFORM METABARCODING REF. THIS PROCESS IS HIGHLY DEPENDANT ON SEQUENCING DEPTH AND ONLY WORKS IF SEQUENICNG DEPTH IS HIGH ENOUGH.

PART 2.2: SHOTGUN SEQ

Shotgun sequencing-based approaches, [[[such as metagenomics and metatranscriptomics (RNA-Seq)]]], present an alternative to DNA Metabarcoding. These approaches enable the generation and analysis of metagenomes and metatranscriptomes of entire communities while avoiding primer and PCR bias. OVERALL PRO

OVERALL CON: BUT: issue of these approaches: dependant on sequencing depth.

PART 2.3: SPECIFIC CON METAGENOMICS

High sequencing cost is costly, if this is not a problem then metagenomics might be good to go. But if it is, then a big portion might not be sequenced. Especially problematic when looking at genes, barcode genes for protists and prokaryotes only make up a tiny fraction of entire sequencing set:

Standard barcode markers represent only a small portion of the read pool (0.05% for 16S ribosomal RNA (rRNA) genes (Logares et al. 2014); 0.3% for 23S/28S rRNA genes, and 1.4% for 16S/18S rRNA genes (Yilmaz et al. 2011)). The retrieval of standard barcode markers of all organisms in a sample depends on the coverage and number of reads per genetic region, which are determined by the sequencing depth, i.e. the sequenced portion of a sample. This pitfall can, however, be overcome with higher sequencing depths, which are increasingly getting cheaper.

Additional reference for problems Singer 2020 prepreint: ﻿inefficiency of the procedure: as shown in Table 1, while eukaryotic DNA is expected to be in low copy number within environmental samples (Azam and Malfatti 2007), the genome sizes are thousands of times larger than those of prokaryotes. Moreover, most of this extra DNA is noncoding, meaning it will not find a match in a reference database unless it happens to be closely related to an organism that has undergone whole genome sequencing. As has been observed by Stat et al (Stat et al. 2017), this leads to two outcomes: (1) the fraction of identifiable reads will be much lower than those from a metabarcoding experiment; and (2) while the reads may be an unbiased sample of the original eDNA, the identifiable reads will be highly biased towards genomes that have been completely sequenced—i.e., those from prokaryotes and eukaryotic model organisms. Indeed, the eukaryotic taxonomic recovery in previous studies is below 0.5% (Tedersoo et al. 2015; Stat et al. 2017).

Note: don’t know role of mRNA-Seq for taxonomic assignments, REFS needed, should be discussed

PART 2.4: SPECIFIC PRO TOTAL RNA SEQ (ORDER OF PROS AND CONS MIGHT HAVE TO BE CHANGED)

Alternative: In comparison, total RNA-Seq natural enrichment for ssu and lsu, >85% of total RNA consists of ribosomal RNA (rRNA, component of ribosomes) (Maaløe 1979) and can be even further used for mRNA analysis (used in soil and cow gut studies). Also active part of community: Total-RNA sequencing can be used to identify the active part of a community (Gomez-Silvan et al. 2018), based on the fact that the abundance of rRNA in a cell (as in the number of ribosomes) is an indicator for cell growth and activity (Milo and Phillips 2015). This signal contrasts with that from metagenomics, which targets the present community, including dead and/or inactive cells and extracellular DNA, which can make up high portions of microbial DNA datasets (up to >90%; Carini et al. 2016; Torti, Lever, and Jørgensen 2015).

ALSO: POTENTIAL TO BE ACCURATE WITH LOWER SEQUENCING DEPTH, I.E., MORE AFFORDABLE, ESPECIALLY FOR BIOMONITORING STAKEHOLDERS.

PART 3: WHAT WE DO HERE AND WHY

Comparsion of all 3 methods needed for freshwater microbiome assessments. Here we compare both using mock and real freshwater microbiomes, testing different pipelines for Shotgun seq. Many pipelines have been compared for metagenomics REF, but not for total RNA Seq.

If interesten in general taxonomic composition only:

If enough sequencing depth: metabarcoding < total rna seq < metagenomcis for taxonomic resolution of entire communities (because higher fraction of all genetic material per organism is sequenced = more information)

If not enough sequencing depth: metagenomics < metabarcoding < total rna seq for taxonomic resolution of entire communities (because metabarcoding has bias, and total rna seq natural enrichment of SSU barcodes)

Side note: SSU genes can look different (papers roy danzmann QE), not sure what effect that has in total RNA seq, likely different SSU sequences. Same when using metabarcoding, but to what extend same?

WE FACE THE PROBLEM THAT WE NEED TO HAVE A METHOD FOR TAXONOMIC PROFILING OF ENTIRE DIVERSE MICROBIAL COMMUNITIES, YET DON’T HAVE THE REQUIRED SEQUENCING DEPTH TO DO SO USING METAGENOMICS.

SEQUENCING DEPTH IS CONSTANTLY IMPROVING AND SEQUENICNG IS GETTING CHEAPER, BUT FOR NOW, HIGH SEQUENCING DEPTHS ARE YET STILL NOT AFFORDABLE FOR BIOMONITORING STAKEHOLDERS

THEREFORE, WE WANT TO COMPARE METAGENOMICS AND TOTAL RNA-SEQ FOR TAXONOMIC PROFILOING ACCURACY AT COMPARABLY LOW SEQUENCING DEPTH.

Bring in: segio et al not published: compare dataset to ock community that is similar: needs evaluation which pipeline works best, which is what we do

Sexy for stakeholders:

A yet unpopular approach for taxonomic profiling is total RNA-Seq (metatranscriptomics without rRNA depletion). Total RNA-Seq data is naturally enriched in standard barcodes for microbial taxonomic profiling (16S/18S rRNA) because rRNA makes up >85% of total RNA (Maaløe 1979). In comparison, 16S/18S rRNA genes can make up as little as 1.4% of metagenomics data (Yilmaz et al., 2011). This means that in theory a much lower sequencing depth is required with total RNA-Seq to cover sequences that contain barcodes for microbial taxonomic profiling.

In this project, we are comparing metagenomics and total RNA-Seq for taxonomic profiling accuracy at low sequencing depths using a simple microbial mock community. We predict that by using total RNA-Seq, we will be able to accurately profile the entire microbial mock community at a comparably low sequencing depth, and that the approach will outperform metagenomics. This would make total RNA-Seq attractive and affordable for biomonitoring stakeholders. The AAC funding is ideal for that purpose, as it would allow us to perform this comparison in-house using the available MiSeq sequencer.

Discussion:

Point out results are valid only for this specific sequencing depth, if higher, SS will become more effective, if lower, MB or maybe still MT will be more effective (to what extend MTl? Could be investigated)

Metabarcoding for entire communities with multi-marker approaches (will not be used here), discuss pros and cons

Rna hard to handle