Methods for studying the uncultivated majority of microbes -

The identification and study of microbes has long depended on our ability to culture them. However many microbes are not amenable to growing in the lab environment(Amann, Ludwig, Schleifer, Amann, & Ludwig, 1995). The development of molecular techniques, such as high throughput sequencing and genomics, has enabled us to start to studying which microbes reside in an environment are and what they are doing. Many approaches have allowed scientists to peek into the uncultivated world, such as 16S tag sequencing, metagenomics, and sequencing of single cells. Each method has different advantages and disadvantages, but can be used to answer specific questions about the microbes in an environment.

One method of learning which microbes are in a specific environment is sequencing a conserved marker such as the 16S rDNA or other conserved genes. In freshwater lakes, 16S has often been used to determine which microbes are in the lake. This method has been used to define the taxonomy of bacteria commonly found in lakes and develop a 16S database(Newton et al., 2011). While this approach, with a good reference database, can help to answer the question of which microbes inhabit the lake, it does inform what genomic features or genes these microbes may have. As this method uses a conserved marker, it does not capture more recent evolutionary trends.

Another tool that can yield genomic information about the uncultured microbes is metagenomics. This method involves extracting DNA from lake water samples. It can provide a representation of the lake community not biased by cultivation. One limitation of this technique is that the data is difficult to track back the reads or even assembled contigs to individual organisms or closely related populations. Previous studies relied on sequence features such as GC content and tetranucleotide frequency to separate out contigs belonging to one lineage. A recently developed technique uses differential coverage to separate out contigs that belong to the same organism(Albertsen et al., 2013; Wrighton et al., 2012). This method requires that you have two or more different metagenomes representing the same environment. Possible differences include the extraction method or time of sample collected.

One such method was used to separate out 49 draft genomes from an acetate amended aquifer using three metagenomes collected from different times(Wrighton et al., 2012). These contigs were binned into genomes first by emergent self-organizing maps of tetranucleotide frequency and then by differential coverage. From these genomes, metabolic potential was inferred from annotated genes.

Another such study used the differential coverage as the first metric for binning genomes from an activated sludge bioreactor(Albertsen et al., 2013). Then the bins were refined using tetranucleotide frequency, GC content, length, and single copy gene analysis. When this method was applied to the dataset from the aquifer, they found that this method assembled more complete genomes with less contamination from other genomes. While both of the previous examples represent less complex microbial communities than those found in lakes, the lake community is a good next test set as compared to complex communities such as those found in soil. However one must keep in mind that with more complex systems, assembly becomes more difficulty and rare populations may be missed due to insufficient sequencing coverage.

Another recently developed technique for investigating the genomes of uncultured organisms is whole genome amplification from single cells. Each cell is sorted by fluorescent activated cell sorting, lysed, amplified with whole genome amplification, sequenced and then assembled. From this process, single amplified genomes, SAGs are created. The genes are then predicted and annotated using computation. This method was applied to produce 201 SAGs from previously unsequenced lineages (Rinke et al., 2013). These phyla had been known to exist in nature due to 16S tag sequencing but have never been isolated. These genomes provide diversity and expand the reference database to include these phyla. In addition, resolved broader evolutionary patterns of related phyla.

SAGs can also be used in conjunction with metagenomic reads to track populations. Two SAGs from marine flavobacteria were used as references for metagenome sequences from the Global Ocean Sampling(GOS) expedition (Woyke et al., 2009). These SAGs recruited reads much better than cultured marine flavobacteria. The authors mapped reads from different GOS sites and saw that the populations of flavobacteria closely related to these SAGs were likely dispersed by a known ocean current. This study is an important example of how it is important to consider uncultured organisms and to have genome references that are relevant to the environment in question. It is also an example of tracking populations using SAGs as references for metagenomes.

One important limitation with SAGs is that the whole genome amplification process has random bias and can miss whole sections of the genomes. A common method to estimate how complete these genomes are uses single copy conserved genes shared among bacteria or a particular bacterial lineage (Albertsen et al., 2013; Garcia et al., 2013; Rinke et al., 2013)+TEVOR’s PAPER.

Studying evolution using genomics –

This rise of sequencing has given us the opportunity to study evolution by genome comparison. From this data single mutations can been found, and difference in functional ability inferred. With cultured representatives, *in situ* tests can be done to test fitness (cite lenski experiments?). One example of studying evolution in the laboratory is a long term experiment transferring *E. coli*. Serial transfers provide population bottlenecks regularly. EXPERIMENTS AND SEQUENCING?- find paper and look for info.

Previous studies have also studied evolution by isolating many genomes from those environments and comparing their mutation and recombination. VIBRIO AND SULFOLOBUS EXPERIMENTS – see Rex’s references

Attempts have also been made using metagenomics to follow populations in nature and calculate their mutation rates. ACID MINE METAGENOMICS – SEE REX’s references

A new study done by the Joint Genome Institute in collaboration with the McMahon lab, provides evidence for the ‘ecotype’ model of bacterial speciation (Reference REX PAPER). This work tracks single nucleotide polymorphisms(SNPs) over three years, in four genomes binned from metagenomes by mapping the reads of these metagenomes back to the genomes. Reads were mapped at ninety-five percent identity representing sequence discrete populations in the lake. Over the three-year time period all the SNPs in three of the populations go to fixation. Thus the sequence discrete populations represent ecotypes swept by the same selective pressure.

Another current study in the McMahon lab using genomics to study evolution focuses on the Actinobacterial clade acI, a ubiquitous and abundant lake bacterium. This study uses 14 SAGs from the acI clade and tracks them over time by mapping back reads from a 5 year time series of Lake Mendota. To understand what we are seeing when reads are mapped backed to these genomes from the metagenomes, we mapped shredded reads from the SAGs and mapped them back to one SAG from each tribe. DO I NEED TO EXPLAIN TRIBE? These shredded SAGs together represent a pseudo-metagenome. The mapping results then show if reads from different or the same tribes are recruited by BLAST. We found that only

Studying ecology using genomics –