THESE ARE THE WORKSHEETS FROM DAY ONE OF OUR WORKSHOP. THE WORKSHEETS FOR DAY TWO CAN BE FOUND AT THIS LINK.

Q1 WORKSHEETS

Q1

Where is there consensus with respect to Symbiodiniaceae taxonomy?

- **For which types of questions/interpretations is species-level taxonomy required/preferred/unnecessary?**
- **How can we improve the current methods used to define species entities, assign taxonomic names, and incorporate them into research?**
- **Can we generate consensus on definitions for the terms “species,” “populations,” and “strains” for the Symbiodiniaceae community to follow (at least for now)?**

Q1. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

- Species-level designation often not necessary
- Problems arise between conflicts: there's only 1 thing there vs. there's a ton of diversity there
    - I.e. problems arise when people form very strong opinions one way or the other
- We must always acknowledge when we don't have enough data to delineate one way or the other
- Time to explore/apply other techniques besides ITS2
    - Positive experiences with cp23s
- Important to acknowledge that there is diversity within genera
- Easier to show differences in symbionts than similarity
    - Shows why we need to improve marker application
- Need to figure out: does the relative abundance/low abundance symbionts matter?

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

HOW MIGHT WE MAKE THIS WORK EASIER OR MORE COMFORTABLE TO DO?

- Lower the bar for how hard it is to prove that two symbiont are the same (i.e. you have to go down to the lowest level)

WE SHOULD EXPLORE THIS MORE:

>>>

Q1. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

Hard to stay current with ever-changing taxonomy.

Seems like it's a mess/a lot to dive into.

Extra effort

Most understand the basics but not the specifics

Recognition that symbiont species are totally different organisms eco/phys-wise

It's a reference information problem (when we address the primary question, we don't incorporate species designations--that happens later)

Strain = within species population entity

Backing up with second markers helpful

Bacterial species definition: 3% rDNA similarity

Eukaryotic: house-keeping genes (140?)

High weight for ecology as a species concept, but a moving target.

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

Species definition is important vs. species definition is not important (will vary by group/question)

"Arbitrary" divergence cutoff vs. "arbitrary" recombination

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

HOW MIGHT WE MAKE THIS WORK EASIER OR MORE COMFORTABLE TO DO?

How do we understand the landscape of diversity? How do we map to species taxonomy?

How do we stay current? Are we doing the right thing?

Species description papers helpful but dense

Wish that symportal popped out species descriptions alongside DIVs

Desire to define boundary with divergence value

Need some document for newcomers to the field that is a friendly introduction

Challenge for eco/phys: really need a hard line drawn between species to know if we're talking about groups that are recombining genetically or not

WE SHOULD EXPLORE THIS MORE:

Need to share which markers are better for which genera

Rosetta stone for species: host affinity, range, cell size, genes, names (NSF funding for database?)

Q1. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

>>> species-level distinctions important across community-scale level questions

prioritize dominant sequence variants at species level in HTS experiments; when asking physiological questions you can have many differences within a species

Question: How important is species-level taxonomy in HTS studies? More concern for comparing across genera instead of species if doing WGS? Divergence within species comparable to divergence between genera. It is important to define species? Makes more sense to do genome comparisons within a genus, even then you find a lot of diversity.

Need to have guidelines for practices that people can easily generate (cost effective, simple analytical pipelines) so that we can share data

What is the most basic set of data we can get to preserve cross-comparisons without breaking the budget?

Need specific markers associated with physiological traits we care about. Physiology and ecology should be included in the definition of species. The amount of time and effort to characterize a species is a task in itself.

We can begin to delineate taxa within a functional framework that can serve as a roadmap to guide marker data.

There's a good Cladocopium marker! Can we make that happen for other taxa?

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

>>>

Improving measures: whole genome sequence to quantify diversity because there will be more diversity there than can be captured by ITS2/other markers

How many markers and which kinds to use will vary from taxa to taxa so it's hard to get a consensus on what to use.

Any marker that will represent the physiology of the species could be too variable to be a reliable taxonomic marker: present/absent, variable copy number.

Combination approach of multiple markers and molecules that characterize

Level-specific markers: applicable for defining species or genera or higher taxa

ITS is how we make the bins and whether it is a species or not doesn't matter... but multiple species within ITS so we need another level

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

HOW MIGHT WE MAKE THIS WORK EASIER OR MORE COMFORTABLE TO DO?

>>>

WE SHOULD EXPLORE THIS MORE:

>>>

Q1. SCRATCH PAD

CAPTURE ANY ADDITIONAL NOTES OR IDEAS YOU WANT TO PRESERVE.

>>> Consensus: in use of ITS2 sequencing for identifying syms; nomenclature, new sym taxonomy (family/genera names)

If you can, you should get down to spp level assignments. But its difficult to do this bc of the diversity in the ITS2 amplicon

METHODS -> are there ways we can achieve this with currently existing methods?

- **Potentially use of symportal -> based on co-occurrence of amplicons to collapse ASVs into spp. STILL caveats wrt experimental design. Symportal is user-friendly for ppl who don't have the experience/expertise to run teh analyses on their own**
- **There is potential to run symportal on your own -> but its difficult. Some caveats, doesn't use low abundance sequence variants. Database gets better the more that you use it**
- **DIVs become putative spp**

Lack of consensus: once you get to spp level

For community ecology level Qs -> spp level analysis can become problematic due to intragenomic variation
Amplicon seq of ITS2 -> spp A vs spp B -> very hard to say how many spp are present bc of issues with intragenomic variation in amplicon
Rather than spp we should just look at variation with the ITS2 amplicons

Javier counters that spp level taxonomy would be preferred to allow comparison of results acros studies; makes meta-analysis really tough

This is different when you're working with dinos in culture -> need to know spp in that case with high confidence

QUESTION -> do we have any connections with fungal community (same issues). Sense is that there isn't much cross-fertilization here. THere are bigger databases available e.g. UNITE; same issues with intragenomic variation. Could we have something similar to SILVA or another core database -> perhaps Symportal is the solution for this

For evolutionary questions do need to solve issues wrt spp vs population level variation

Spp level concept can be paralyzing to the field

Data continuity -> 25 years of ITS2 data already -> naming SHOULD be connected to ITS2 definitions as well to understand how current findings connect to prior studies

What about whole RNA operion (using pacBio) as an amplicon that will contain ITS2 -> might provide additional resolution

Q1. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

- **The need to know taxonomy is dependent on the research Q (e.g. sometimes just knowing differences is good enough for the research Q)**
- **-taxonomy is needed, just a Q of what the resolution needed is - sometimes Family/Genus is enough**
- **-why apply a different standard to host vs symbiont**
- **What level of phenotypic resolution is driven by genetic differences**
- **Resolution matters**
- **Culture work coupled with physiological framework,**
- **Combine molecular signature with phenotypic trait that you can compare and assess in lab: performance under variety of environments**
- **Expansion of the strain performance**
- **Tradeoff the more universal a marker is the less resolution it will have - resolution varies by genus**

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

- **Multiple markers in a standardized test (e.g. 23S has its place but depends on genus**
- Shallow genome representation- should this be standard for species level resolution?
-

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

HOW MIGHT WE MAKE THIS WORK EASIER OR MORE COMFORTABLE TO DO?

>>>

We need to reduce the barrier to entry - a group of markers that are tangible to all groups, not just expensive tech

Need to develop a group of markers that makes the work more accessible, this would increase people working in the field

WE SHOULD EXPLORE THIS MORE:

>>>

Links between physio and geno

Q1. SCRATCH PAD

CAPTURE ANY ADDITIONAL NOTES OR IDEAS YOU WANT TO PRESERVE.

>>>

Pacbio sequence whole chloroplast
Stuck 20 yrs ago, genomic resources we have now are very different but for some reason we are still stuck here
Need to link phys with genetics

Q1. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

--if we don't use species taxonomy, then function gets ascribed to the genera level or to the sequence variant level. Both are problematic.

--functional diversity is important, and if those functional differences are at a higher taxonomic level, then the species taxonomy might not always help explain the functional differences.

--spp level taxonomy would be very helpful if we can use it to guide our understanding of function.

--Thought experiment: there are about 8000 spp of algae, how many spp of zoox are there? Chris thinks maybe 100s. There are ~50K sequences submitted to SymPortal and we're starting to see convergence now (rediscovery of "new" profiles)... based on that, Chris is estimating 100s of spp.... The wild card here is free-living zoox, which are extremely unrepresented in SymPortal

--need for consistency in what you identify something as

--There are biases and holes in what we cite and what we don't (model species, organisms, hosts). Need to make a conscious effort to look across cnidarians (not just scleractinians) on reefs. FOcusing on only a few species can result in a skewed perception of what is going on with symbiosis/diversity

--diveristy is not taxonomy, ITS2 is not taxonomy

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

--Not everyone knows how to describe species, and not everyone is interested in that as a research focus, so we end up relying on the folks that do. Cost is high to describe a species (to use the name in your research) but the benefit is low (to an individual researcher)

--Symbiodiniaceae field is not alone in these taxonomy/diversity problems

--what is meant by diversity and how do people interpret that?

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

We didn't really have time do organize these across the consensus bins....

HOW MIGHT WE MAKE THIS WORK EASIER OR MORE COMFORTABLE TO DO?

>>>

Need for a clear road map on how the naming should be done. If we find a new variant, how different does it need to be to be worthy of being a species

--We need to put more effort into establishing things in culture, this will help us establish taxonomy, physiology etc (free-living, symbiotic)

WE SHOULD EXPLORE THIS MORE:

--Can we pull things out of SymPortal to test as species? --Is there a way we can easily figure out what ITS2 profile(s) correspond to a species?

--we need a reference database where you can quickly/easily look up what marker sequence is a species?

--need to create a connection between diversity and taxonomy that is shared and can be used by people

Q1. SCRATCH PAD

CAPTURE ANY ADDITIONAL NOTES OR IDEAS YOU WANT TO PRESERVE.

--lack of consensus in how hard we try to accurately get species right in the samples we work with

Q1. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

>>> While understanding species-level could help enlighten many areas (e.g., symbiont function), questions should not be bound by required methods.

>>> We need to prioritize what importance species-level information provides.

>>> Improve database integration and accessibility

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

>>>

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

HOW MIGHT WE MAKE THIS WORK EASIER OR MORE COMFORTABLE TO DO?

>>> Asking to increase the federation of resources

WE SHOULD EXPLORE THIS MORE:

>>> When is identifying species more or less important/influential to the conclusions/implications?

Q1. SCRATCH PAD

CAPTURE ANY ADDITIONAL NOTES OR IDEAS YOU WANT TO PRESERVE.

Q1: Preferred:

When is identifying species more or less important/influential to the conclusions/implications?

Q1: Questions where symbionts are known to vary across a gradient (e.g., light, depth)

Q1. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

>>>lack of species-level definitions/lines have not been drawn yet (or in the past), one reason why many people may not be working at species level

>microsats useful for determining species distinctions, must consider geographic distribution, minimum number of microsats that must be used to separate species

>possible to get a lot of ecological insights/useful information without species level definitions, but also a need for more information and more species level definitions to understand the system

>labor intensive process to ID new species

>more available training about the process to characterize/name more species would be beneficial

>ideally incorporate multiple species concepts when characterizing a new species

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

>>>

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

HOW MIGHT WE MAKE THIS WORK EASIER OR MORE COMFORTABLE TO DO?

>>> more training, more guidelines on how to characterize and define new sym species

WE SHOULD EXPLORE THIS MORE:

>>> what barriers are preventing more people/labs from describing species (Dusty's comment as room closed!)

Q1. SCRATCH PAD

CAPTURE ANY ADDITIONAL NOTES OR IDEAS YOU WANT TO PRESERVE.

>>> manuscript resources for subsequent sessions

>>symportal/its2 utility may be dependent on genera/phylogenetic tree area

Q2 WORKSHEETS

Q2

Where is there consensus in population-level research on Symbiodiniaceae?

- What are the advantages/disadvantages of the molecular approaches currently being used to ask population-level questions? (ITS2, microsatellites, RADseq, others)?
- What are the major limitations of these types of population-level analyses?
- What are the outstanding questions that should be prioritized with respect to populations of Symbiodiniaceae, and how can we encourage research in this area?

Q2. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

>>> Entry point is expensive with a steep learning curve (a lot of prior ITS2 data), and people are hesitant to delineate populations with the variation among the molecular techniques.

>>> ITS2 multiple copy uncertainty

>>> Shifting technology and contexts/questions creates disconnect among interpretations (i.e., are you measuring the same thing?)..

>>> Future questions: adaptive potential: genotype fitness, symbiont gene flow → host gene flow, symbiont uptake with regard to diversity

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

>>> How to proceed with questions regarding the origins of symbiont diversity: genomic, external influences

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

WHAT ARE SOME WAYS WE MIGHT WE GET MORE PEOPLE INVOLVED?

>>> Teach how to assess the level of resolution used in the context of the question.

>>> Elaborate what you were not able to accomplish.

WE SHOULD EXPLORE THIS MORE:

>>>

Q2. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

- Not appropriate to use ITS2 for populations
- You should know what species you're working on first before populations
  - So how do we best do that?
  - What is a biological species in Symbiodiniaceae?
- Microsatellites has been more successful for populations so far
  - Don't assemble well in genomes
- How about more SNPs
  - Found within cultures even
- Long repetitive regions hard to sequence, as well as multiple copies
- More explorations of links of genotype & phenotype
- Does it matter if they're different species or not?
  - Variation matters if it affects holobiont function
- Need for more cultures to understand what the DNA variation means

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

- Some controversy in field about symbiont sexual reproduction
  - Important to come to consensus in order to define biological species

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

WHAT ARE SOME WAYS WE MIGHT WE GET MORE PEOPLE INVOLVED?

- Make more clear when to use what methods
- Better explanations of symbiont diversity to people outside the field

WE SHOULD EXPLORE THIS MORE:

Long read sequencing

Single cell sequencing

Suitable markers for populations (*e.g.* SNP-chip)

Ultra conserved elements (UCEs)

Links to physiological & functional diversity

Q2. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

Multiple cells in a colony problem + multiple copies of ITS2 + multiple genotypes? Lots of potential issues, but cool questions

Co-occurring species become problematic for analysis (sometimes dropped when may be very informative)

Analysis pipeline is main concern

Microsats might not be the way forward

Generate more chromosome-scale assemblies--Earth BioGenome Project for funding?

Single cell analyses to establish how popgen markers vary within a single cell and how that varies in different cultures, strains, hosts

Population-level genome resequencing using short+long read sequencing for high quality

Metagenome-assembled genomes for Sym population analyses? (check Harry Alexander from WHOI) But we don't have enough reference genomes yet

Alternative markers e.g. cp23S have worked well

Structural variation is one of the first things to change in experimentally-evolved populations; speaks to advantage of chromosome-scale assemblies, etc. for capturing these events. SNPs might miss structural changes but we're maybe stuck with them for now.

Single cell RNAseq might be a good way to generate data of SNPs to map to an extant assembly and assess variation that way

Single cell genome amplification and sequencing probably not possible with Sym at the moment

In diatoms, clear evidence that lots of sex not necessarily required to generate variation--may be relevant for Symbiodiniaceae. We should be looking at the genome at the population level to look at the role of non-sexual drive.

Defining a population: diversity within an ITS2 type/species and comparing across the same host species

Is a Sym "population" in HWE? is it panmictic? Reproductive cycles? Levels of genetic diversity?

Lots of spatial structure

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

>>>

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

WHAT ARE SOME WAYS WE MIGHT WE GET MORE PEOPLE INVOLVED?

Can we use pop data to infer frequency of sexual reproduction, hybridization?

Studies to quantify adaptive variation better--questions related to evolutionary rate, adaptive capacity

It's been done in the lab but would be awesome to capture in the natural environment

Could potentially incorporate historical samples to see how populations have changed over time

Does colony association influence trajectory of population evolution wrt thermal exposure?

How might population dynamics influence the physiology of cells

WE SHOULD EXPLORE THIS MORE:

Single cell analyses to establish how popgen markers vary within a single cell and how that varies in different cultures, strains, hosts

Q2. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

>>>Agree that multiple loci should be used. SSRs could be used (if available) and SNP based approaches could be very useful here - e.g. mining SNPs from RNAseq data; RADseq could be very valuable here, but is underexplored - e.g. can it be done de novo? Do you need a genome ref in the first place

It will be important to keep price point in mind bc this can be exclusionary - e.g. for the more expensive methods

IS thre a pre-existing requirement for amplicon sequencing to verify that there is a single dominant spp. Not necessarily -> can use SNP based data to recover both community and spp level variation

More genomes would be awesome here -> could be used for competitive mapping approaches.

Identifying major asks is important -> areas where funding agencies could really help advance the work

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

WHAT ARE SOME WAYS WE MIGHT WE GET MORE PEOPLE INVOLVED?

>>>

WE SHOULD EXPLORE THIS MORE:

>>> are there issues with estimating allele frequencies in mixed community samples where both pop and community variation exists?

Issues with really young spp -> what happens when things are very closely related? How to resolve? Might be potential to look at SNP diversity (e.g. more than biallelic sites)

Q2. SCRATCH PAD

CAPTURE ANY ADDITIONAL NOTES OR IDEAS YOU WANT TO PRESERVE.

>>> Preference would be for SSR or SNPs; potential to also pull SNPs from RNAseq data -> this might be more advantageous if trying to do pop level questions de novo

Arthur asks if there are a set of conserved genes that would be phylogenetically informative. Thinks that the number of genes is limited. ITS2, cp23, psbA,

RADseq seems to be underexplored -> should be generated in a holobiont sample. Problem would be insufficient read depth

Looking at closely related spp -> James has some data that closely related things tend to be in different hosts, possibly as a result of competition

Some progress on genomes WRT Moore foundation initiative -> should have a lot of genomes coming online in the near future

WHAT biological questions remain to be answered?

Dispersal

Population bottlenecks

migration/connectivity

Pop level variation in physiology/

WITH additional genomic data

Genetic architecture of traits - e.g. thermal tolerance

Useful for machine learning when there will be more sequence data available

Q2. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

>>> method depends on question, how much resolution do you need

>ITS2 generally NOT sufficient for population level questions

>major limitation/holdback is limited species-level distinctions

>need multiple datasets from multiple markers to be sure you're comparing diversity at correct scale (ie not comparing different species/mixed symbiont communities)

>linked species and population markers

>evolution happens at population level- population level important for understanding symbiont contribution to host responses (especially when many hosts have same dominant symbiont)

>Outstanding Questions: how does population level variation vary with environmental gradients, contribute to host environmental stress response?

Q: need better understanding of Symbiodiniaceae reproduction

Q: in tight host-sym associations, at what point do you have/call it a different population (when somatic mutations, etc leading to evolution over time)

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

>>>

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

WHAT ARE SOME WAYS WE MIGHT WE GET MORE PEOPLE INVOLVED?

>>>improve taxonomic delineations, more markers to distinguish species

>SNP chip assays, more available, more development

>funding to build these tools (methods developments, data bases, etc) will encourage/allow more people to get involved

WE SHOULD EXPLORE THIS MORE:

>>>

Q2. SCRATCH PAD

CAPTURE ANY ADDITIONAL NOTES OR IDEAS YOU WANT TO PRESERVE.

>>>

>are far spread are symbiont strains (within the same reef, across reefs), method may depend on this scale

>SNP based markers need for some questions (microsats or ITS2 not sufficient)- like possible hybridization of parental symbionts

>advantages of microsats: more cost effective, sufficient for determining strains in many cases (esp in well characterized systems)

>rad seq: we lack benchmarks to decide what is a new species/strain/etc, how much diversity is there in a species

> can use some higher level distinction of symbionts (like ITS2, cp23S) and then look at diversity within these types

>for symbionts that are cultured, can use to develop species distinctions

Q2. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

--ITS2 sometimes gives popn level answers but sometimes doesnt. It is relatively easy now to identify which things are different, but it is hard ot tell what those differences mean

--SNPS: if you don't have an already assembled genome to work with, if you find a SNP at position X, what does that mean? If you don't have a good understnading of the genome, SNP differences are hard to interpret

--is there a single or multiple infection of zoox in a coral colony? Co-infection could be leading to detection of SNPs.

--do zoox in mixed infections hybridize? V interesting question but we haven't been able to get to that yet. Preference for studying population level questions in host colonies w/ single infections

--challenge w/ confidence with mapping multiple infections (i.e., coral hosting symbiont from genus A and genus B and being able to map with confidence to both reference genomes so that we don't have to discard coral samples with multiple infections)

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

--you cant sequence every strain that's in a colony. You have to just pick a level to approximate what's there (but what we wish we had were genomes for everything), and then we work from there

--We care about co-evolution, biogeography, habitat utilization, and we're inhibited in answering some of these questions in the field because you "don't know what you've got"

--you have to find "the line" and then go with it, because otherwise nothing moves forward, but need to accept that it's going to change in 20 years

--When you identify a strain in Cladocopium in the lab based on SNPS, is that actually a real strain that is in the environment? Cannot always be sure

--need to work on zoox that are potentially symbiotic but out in the environment - what are they doing? Need more ecological information on what Symbiodiniaceae are doing (on a basic level and to apply to conservation). Also need from phycosphere

--so what do we do about markers? Doing RadSeq would be hard if you don't know the genome you're working with

--refer to the number of cells and not just the gene copy number for something like ITS2

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

What markers would you use on free-living symbionts to determine what would potentially colonize a host, vs what wouldn't? We may have the markers but to start this well (esp those that work across both host and free-living), you need a single-cell culture. But there are issues with things that we can't culture then. Complementary in situ approaches? Or just work with stuff we can culture first?

WHAT ARE SOME WAYS WE MIGHT WE GET MORE PEOPLE INVOLVED?

--Summary review that would give an updated perspective on what we think is important for Symbiodiniaceae? Sometimes folks outside the field can't easily find this information b/c it requires rabbitholing in the literature of different Symbiodiniaceae sub-fields currently

--need more people to understand what is being defined as a population

--need to be realistic about what we're going to have to do in certain situations (eg sequence genome from colony, rather than culture) and generate best practices, and agree that that is ok in review

WE SHOULD EXPLORE THIS MORE:

--make use of "leftover" data from omics

--we need more widespread taxonomy (and better connection between taxonomy and diversity) to move the field forward

--bias in which parts of the worlds have been studied.

--need to work on free-living zoox more, new studies of Symbiodiniaceae life history (eg fate outside a host)

Q2. SCRATCH PAD

CAPTURE ANY ADDITIONAL NOTES OR IDEAS YOU WANT TO PRESERVE.

The placement of ideas on the previous slide does not reflect actual level of consensus AWK

Q2. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

Advantage: Start with ITS (or some other marker) to facilitate more in-depth research → microsat (assuming you know they are closely related from ITS) or additional markers or RAD? Look at specific markers that are relevant to your question.

Database/"flexible workflow" could point you toward the next marker(s) to sequence after ITS2(or other?) that would provide the next level of specificity you need.

Provide capacity and guidelines for question-specific efforts. Minimum, desirable, ideal-levels. Putting "required" and "minimal" things will exclude some research groups. Recommendation, not "gold standard". A true barcode can increase accessibility.

Linking population-level differences to functional differences. How quickly will these population-functional relationships change in real ecosystems under selection?

Problems with genus-tied phenotypes.

Functional database can guide functional studies that can feed back into the system to increase predictive power.

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

Can a marker (or profile of markers) confidently characterize function? Different fxn more or less stable in different taxa.

Hard to know how deep you have to drill into the symbiont identity to capture functional variation within.

You need to be as precise as you can be and then go more broad. ITS2 is as precise as you need to be in profile but not in function... depends on your question.

ITS marker precise enough to use in population studies? Can resolve at pop level if sequenced well. How you analyze should be bespoke to your research question. Difficult to understand what ITS2 correspond to taxonomically. Can assign multisequence profile to "radiation" (group, e.g. C1). Others find that ITS2 match but another marker identifies subpops - that other marker will vary based on what genera you are working in.

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

WHAT ARE SOME WAYS WE MIGHT WE GET MORE PEOPLE INVOLVED?

>>> Link efforts for "barcoding" to phenotypic database. Who maintains the database? Need to engage secure sources of funding. Needs community support beyond a grant because it's a time commitment.

WE SHOULD EXPLORE THIS MORE:

>>> Pointed level of recommendation on the level of specificity that you can ID in SymPortal - current recommendations are spread around a couple of papers (forthcoming millepora paper from Ben Hume)

Q2. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

ITS2 is a terrible population-level marker, good for framing pop level Q- it is too coarse

Drawback of SSR: specific to a species

RADseq allow for larger data from across population within cnidarian, but hard to determine how many lineages

SSR vs Radseq: old vs new - huge reviewer bias here!

Analytical pipelines make underlying assumptions that our organisms do not follow and those can have implication for inferences- particularly problematic

-coming up with common pipelines for comparisons across groups

-limitations to SSR- development is tough, detection limits, allele dropout, fingerprints vs genotypes (multilocus genotype), specificity of the primers- miss some groups

-mitochondria a bad locus

-Research Q's: how important are the low density organisms in symbiotic hosts (are they there ? intragenomic variants?, eDNA), are they ephemeral or do they have novel and important roles? Can they become dominant? Functional significance?

We need to be Q driven- don't need to necessarily need to sequence the world, but also cannot just ignore the symbiont completely

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

-people who think about biology and the data that is coming out and try to find reconciliation in how they interpret, but others are more cavalier when it comes to interpreting data, head butting happens here

-haploid vs diploid (or geome dup?)- what is a single lineage, kinda need to know for interpretation

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

WHAT ARE SOME WAYS WE MIGHT WE GET MORE PEOPLE INVOLVED?

>>>scleractinian-biased field, other work not cited in symbiont work

Don't necessarily retire "old" techniques - do not need newest and shiniest, don't close door on old tech that works!

Make SSRs next gen

-psba minicircle might be a google pop gen marker, could be explored as a lower cost method

-developing a suite of markers, not too much, just right for budgets, but that peeps feel confident

-reduce barriers to entry instead of creating new ones, not creating more informatic burden

WE SHOULD EXPLORE THIS MORE:

>>>how to get away from ITS2?

-metaanaylsis of all data (not just hard coral hosts)

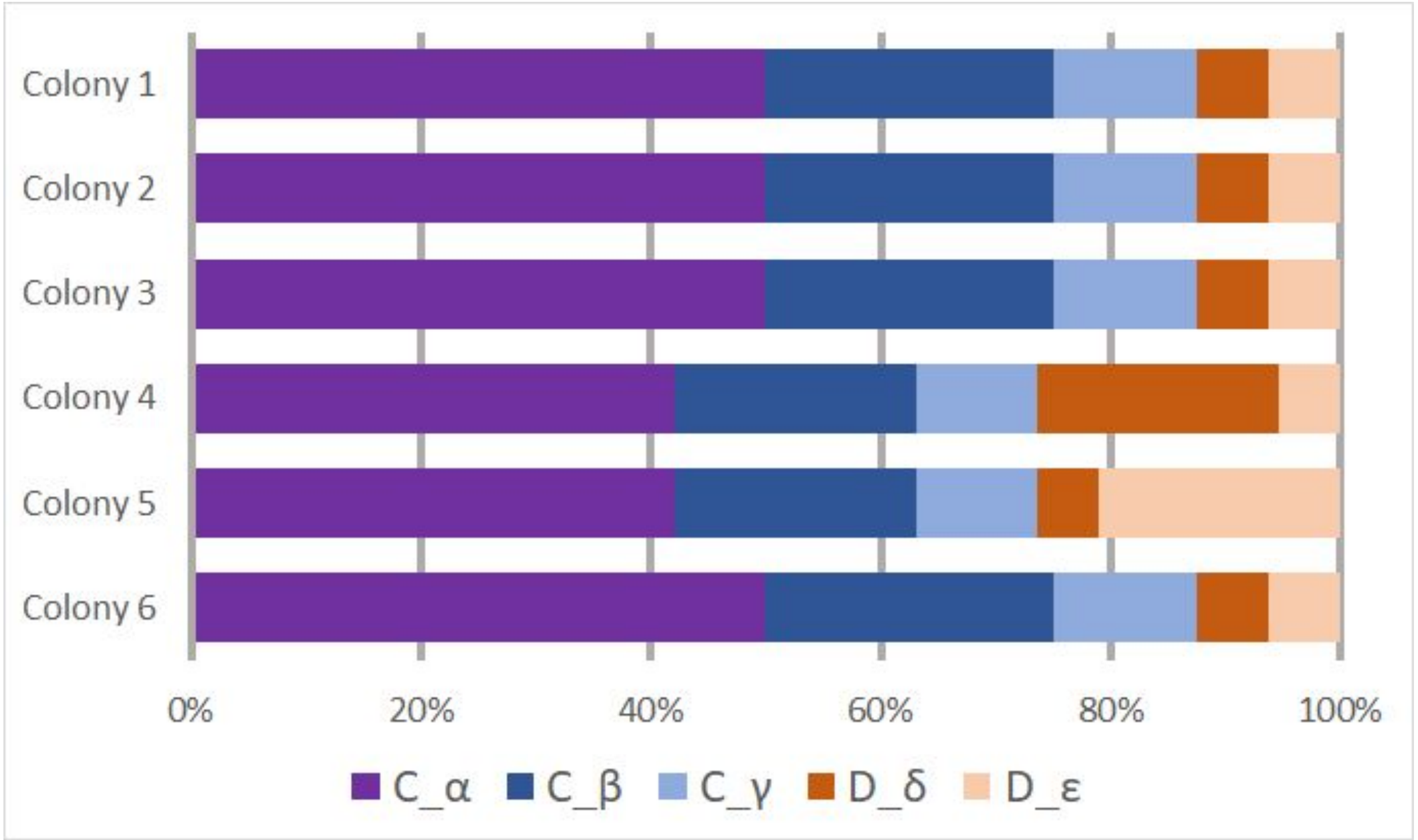
-how can we keep people included, keep people coming from all levels of funding contributing to the efforts of understanding the biology

Q3 WORKSHEETS

Q3

Where is there consensus in the ITS2 community-level work?

- Provide the possible ways people might generate and interpret these data (on the next slide). For each interpretation, indicate whether most/some/few of the group members reached consensus.
- What are the strengths and weaknesses of using ITS2 data for analyzing Symbiodiniaceae communities?



Hypothetical data. Provide the possible ways people might generate and interpret these data. For each interpretation, indicate whether most/some/few of the group members reached consensus.

Q3. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

>>> Barcoding, the prior associated techniques, and subsequent bioinformatics have flaws and biases.

>>> SOP could be beneficial for recommendations on how to interpret data (not a rulebook); should be "living" and should be open-access

>>> Taxonomic: Higher level (C's and D's) or finer (figure out if each C and D type is from the the same species)

>>> Functional: Does it matter? What are those types doing? Are they shifting?

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

>>> Most likely requires a follow-up analysis/sequencing

>>>

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

THE LACK OF ALIGNMENT (WHERE IT EXISTS) IMPACTS US IN THESE WAYS ...

>>> Desk rejections

>>> Exhaustive amount of effort required to continue trying to reach consensus

WE SHOULD EXPLORE THIS MORE:

>>> SOP

Q3. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

There are two genera present

Maximum of five entities within the detection limit

A minimum of two entities

C entities in constant proportion

D entities vary

Colonies 4 and 5 are shifted relative to the others

The big question is are certain variants coming from the same genome (looks like a potential fingerprint)

If variants of C were highly divergent, could they constitute different OTUs? -- Potentially, but unknown threshold

16S for bacterial DNA has similar limitations wrt relative abundance not necessarily being variant = cell; need to keep this caveat in mind when analyzing

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

Dominated by cladocopium

Does sequence ~ sym cells: do we need to normalize for copy number variation at ITS2?

Could be co-dominant C-types

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

THE LACK OF ALIGNMENT (WHERE IT EXISTS) IMPACTS US IN THESE WAYS ...

If you interpret every strain as a distinct entity, you might overestimate diversity

Are low abundance variants important (and in what context) -- might be critical under certain scenarios

Much clarity might be provided by sequencing additional markers (psbA for Cladocopium, but wouldn't necessarily capture Durusdinium)--relatively few alternative phylogenetic markers in these two genera

WE SHOULD EXPLORE THIS MORE:

Comparative genomics to identify better phylogenetic markers for certain groups, preferably single copy (might be genus-specific)

Determine copy number estimates for normalization

Q3. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

- At least 2 genera
- Need more colonies/samples to confirm patterns
- Need to acknowledge the uncertainty of the interpretation - not enough evidence for calling the shots
- How often things come up might be evidence that it's not an artefact
- SymPortal would be useful
  - Not everyone knows the interpretation part though
- Ecological context is important - are environmental factors distributing the types?
- Use of alternative markers could be a solution
- Missing critical pieces of information that are hard to come by (e.g. copy number variation)
  - Funding & time limited

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

- Same relative abundance could indicate intragenomic variants for C
- Could be 3 strains of D due to differences in relative abundance
- Differences in relative abundance could also be sequencing artefacts
- Might be helpful to know how different the strains are to each other
- Low abundance types may have been filtered out by these sequencing methods
- Might be disagreement on whether the C is dominant
  - Variation in copy number could explain it rather than C "amount"
- Should we put in all the work of determining copy number of each type?
-

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

- Don't agree that there could be 5 different species of symbiont in the samples
- Messier data (aka real data) leads to more differences of opinion/interpretation
- What to do with misinterpretations during review
  - Different experts have different opinions so 'interpretation' is hard

THE LACK OF ALIGNMENT (WHERE IT EXISTS) IMPACTS US IN THESE WAYS ...

Lack of alignment might have implications for the interpretation of function

Affects whether you get published - especially when the whole paper rests on the interpretation

WE SHOULD EXPLORE THIS MORE:

Deeper sequencing

Alternative markers

Copy number variation - targeted

Q3. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

>>> likely same host species

>samples have 2 genera of Symbiodiniaceae

>some level of sub-genera diversity

>constant proportions generally indicate intragenomic variants

>need a set of defined terms/ best practice/ consistent lingo use

>If colors are sequence level

- intragenomic variants

-Cannot compare relative abundances

If profile level:

-different C types

-Relative abundances can be compared

-Colony 4&5 have different relative proportions

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

>>>initial interpretation of what the graph represents- many think the even proportions indicate sequence-level plot

>multiple Cladocopium types/species in one coral/sample?

>Colony 4&5 have different relative proportions -> may have different physiologies

>

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

THE LACK OF ALIGNMENT (WHERE IT EXISTS) IMPACTS US IN THESE WAYS ...

>>>

WE SHOULD EXPLORE THIS MORE:

>>>

Q3. SCRATCH PAD

CAPTURE ANY ADDITIONAL NOTES OR IDEAS YOU WANT TO PRESERVE.

>>> if symportal outputs:

If colors are sequence level- intragenomic variants

Cannot compare relative abundances

One C type and one D type in each sample -> same C and D type in each sample (except maybe colony 4?)

If profile level: different C types

Relative abundances can be compared

Colony 4&5 have different relative proportions -> may have different physiologies

>constant proportions → intragenomic variants

> more common to find multiple genera in a colony than multiple Cladocopium 'species'

Q3. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

>>>can agree that amplicon sequencing is where the field is going

We all agree on C vs D -> there are 2 genera present

NOW what about sub-genus level variation? Intragenomic variation

If we assume its amplicon sequencing data -> the C profiles are equally distributed and are likely to be intragenomic variants (~one spp/strain)

For the Ds -> the anti-correlation among variants may mean that its intergenomic variation. Would need to resequence 4 and 5 . SHOULD discuss caveats that take into account methods and study spp

Communities in 1,2,3,6 are identical; 4 and 5 are divergent

****

C is the dominant genus of the community

WEAKNESS -> knowing when intragenomic variation matters

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

>>>

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

THE LACK OF ALIGNMENT (WHERE IT EXISTS) IMPACTS US IN THESE WAYS ...

>>>

WE SHOULD EXPLORE THIS MORE:

>>>CAN You explore sequence data in isolation? Should there also be physiological data?

Instead of just covariation and sequence similarity, perhaps consider WHERE mutation occurs and potential functional implications

Q3. SCRATCH PAD

CAPTURE ANY ADDITIONAL NOTES OR IDEAS YOU WANT TO PRESERVE.

>>>SHOULD we be jumping to biological entities right away or should we think about sequence similarity too? YES

Q DO intragenomic variants tend to be similar? Or can they be quite divergent? Seems like the variants can be similar OR can have large deletions too

DIFFERENT ways to use community analysis -> are you interested in community diversity? OR would you be using this as a post-hoc confirmation/validation to go along with physiological data? E.g. to say that the sym community doesnt differ between some, and if they have different phys it can't be the syms...

METHOD -> how would we generate these data -> amplicon based approach

One base pair difference is not a threshold for lump or split. Where the mutation occurs can matter -> is this a non-functional amplicon? We think of ITS2 as nonfunctional but secondary structure could matter

CAN you find pseudogenes via phylogenetic analysis?

Cloning returns more intragenomic variants than amplicon sequencing

Q3. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

--good success with Symportal for multiple people, and also helpful/accessible interpretations/feedback from Ben Hume re interpretations (insdie and outside of review process)

--ASVs can also work, if they show differences between treatments

--There's consensus that we can't tell how many kinds of symbionts are in this graph.

--This could've been generated with Symportal, Dada2+LULU... there appears to have been some winnowing down.

--could also be qPCR

--for this small dataset with only 6 samples, can't say that these small differences are legit

-If this is Cladocopium and Durusdinium, then we agree you've got C and D in all of your corals

--One potential issue with SymPortal outputs is that the final table does not necessarily include all "types" or "profiles" if one is present at a low abundance in teh sample (so this could extend to a C and D type both being present but the D type just "disappears" in the final output but it is there in the first output)

--backgound sequences and what to do with them is an issue that has been around for a long time

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

Is there quantitative information in amplicon sequence data? Some would say it is at best semi-quantitative, others say that need large amounts of data and it will be messy.

--need to normalize

--use a marker that has high of resolution as possible, and then use a hierarchical model to identify what diversity matters... use raw ASVs and use Symportal annotations in your model... difference between diagnostic vs Mattering physiogloically

--if markers are different between different treatments, then that is worthy of publishing

--shouldn't gate-keep publication based on how "important" we think a finding is if it is well done (ie technically sound)

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

Provide the possible ways people might generate and interpret these data (on the next slide). For each interpretation, indicate whether most/some/few of the group members reached consensus.

What are the strengths and weaknesses of using ITS2 data for analyzing Symbiodiniaceae communities?

THE LACK OF ALIGNMENT (WHERE IT EXISTS) IMPACTS US IN THESE WAYS ...

>>>

WE SHOULD EXPLORE THIS MORE:

--

Q3. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

Graph: What are the dominant taxa? C.

Flexibility of association? Decent replication across colonies except Colony 4+5 but some plasticity (would want to look at Colony 4+5)

How were data generated? ITS2, depends on what level of divergence was being assessed at "strain" level?

How well have we characterized the host species? Are we confident the colonies are the same species? Are they all different genets or are they clones?

Quality control of how data are generated limits interpretation.

Can't make leap from taxonomy to function without pursuing eco-physiological work. Disconnect between molecular assignment is interpreted in the context of where the sampling occurred rather than with functional studies.

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

Discriminate between proportion and importance of biology. More representative does not mean more important. Rare taxa can be important. Abundance does not mean biological activity (e.g. trxn).

Just because you can detect something does not mean its in symbiosis.

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

THE LACK OF ALIGNMENT (WHERE IT EXISTS) IMPACTS US IN THESE WAYS ...

WE SHOULD EXPLORE THIS MORE:

Can we discover something about how the environment shapes these associations? Do these different proportions mean something at the level of selection?

Assign functional diversity rather than taxonomic diversity (beyond transcriptomic- level of protein)

Q3. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

>>>

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

>>>

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>

THE LACK OF ALIGNMENT (WHERE IT EXISTS) IMPACTS US IN THESE WAYS ...

>>>

WE SHOULD EXPLORE THIS MORE:

>>>

Q3. ALIGNMENT INVENTORY

CONSIDER THE QUESTIONS POSED ON THE PREVIOUS SLIDE AND CAPTURE IN EACH BOX BELOW THE TOPICS THAT CORRESPOND TO HOW "ALIGNED" WE ARE ON EACH. USE YOUR BEST ESTIMATES AND DON'T WORRY ABOUT RESOLVING THE CONFLICTS — WE'LL MAKE TIME FOR THOSE DISCUSSIONS LATER.

AREAS WHERE MOST OF US AGREE

MOST = AROUND 60% (OR MORE) OF THE FIELD

ITS should not be killed as a marker in the field

There are two genera present

Maximum of five entities within the detection limit

A minimum of two entities

Strength of ITS2: could be a good barcode, lots of history, links back to other data, having something straightforward and reliable is a good thing, properties of a marker that is able to get resolution, but also expands all genera,

Problem: very good at telling things apart, but not good for how much apart.

Whether is is an appropriate marker depends on what your Q

Can you use multiple markers to complement ITS2, PSBa non-coding region

-potentially better develop PSBa

AREAS WHERE SEVERAL OF US AGREE

SEVERAL = BETWEEN 15 - 60%

Dominated by cladocopium

Does sequence ~ sym cells: do we need to normalize? Do qPCR

Strength: 20 yrs worth of literature on it- all of the data is out there, some misinformation

AREAS WHERE FEW OF US AGREE

FEW = LESS THAN 15%

>>>we are generally misinterpreting what these relative abundances

THE LACK OF ALIGNMENT (WHERE IT EXISTS) IMPACTS US IN THESE WAYS ...

>>>emotions and conflict, complication in how people want to analyze things

If it cannot be used right, then it is not a great resource

WE SHOULD EXPLORE THIS MORE:

>>> How can we better align the community on interpretation of ITS2 data given that its power lies in its ease and history of the locus

Q3. SCRATCH PAD

CAPTURE ANY ADDITIONAL NOTES OR IDEAS YOU WANT TO PRESERVE.

>>>workshops on how to analyze and interpret ITS2 data for UCRs