Thoughts from STAMPS 2025

Titus Brown

July 23rd , 2025

Some up front observations

- Metagenomics has matured to some extent: while there are still many tools with many different parameters, the set of approaches used in metagenomics has converged.
- And, perhaps as a result of this convergence, the challenges and opportunities of metagenomics have also become better understood.

Many common questions from previous years were avoided.

- We have been reorganizing the course each year so that it is less confusing, and I think to some extent we have succeeded. Because…

- …essentially *all* of the big questions from this morning represent conceptual or technical challenges for which there is (AFAIK) no straightforward answer.

One of my favorite things to say to 1st year grad students:

“Bioinformatics is rarely going to make or break your project”

I think this is true for most animal and plant genomics.

But…

One of my favorite things to say to 1st year grad students:

“Bioinformatics is rarely going to make or break your project”

I think this is true for most animal and plant genomics.

But…

This isn’t really true for metagenomics. You not only need to do good experimental design, you also need to choose your tools and databases appropriately, and this is very specific to your system.

Other things to consider -

(Mostly taken from discussions I overheard in the course ;))

Be aware of the length of your chain of inference.

- 16S -> function is actually
- 16S -> genus -> reference genome -> function of reference genome

Vs

- Shotgun metatranscriptomics -> gene mapping -> gene function

It is hard (impossible?) to do both exploratory data analysis and hypothesis testing

(Amy can say this better)

If you do analyses until you find interesting signals, you are most likely doing some form of p-hacking.

In theory, you need to correct for multiple testing…

…but it’s not always clear how to do that across multiple frameworks.

”Peer pressure”

- “Use <super cool tool> that was just released! It will solve all your problems!”
- ”Use <data type>! It will solve all your problems!”

New tools and new data types are often not that well understood, and it will take time to understand them. Choose accordingly.

Also, it is OK to aim for pretty good science that actually answers a specific question well…

Things I heard this morning -

You would have welcomed:

- More discussion of metabolomics and metabolomics data interpretation.
- Discussions/examples of how to design and build “complete” data analysis workflows
- How to measure gene abundances across multiple (metagenomic) data sets
- How to choose parameters
- How to use tools ”appropriately”
- More/better viz tools and approaches

Fundamentally hard issues

- We are limited in the quality/comprehensiveness of our databases and particularly in our annotations; more experimental work needed.
- Many of you are working on environments that are hard to access or analyze – rich & diverse, hard to sample, largely uncharacterized.

Emerging opportunities that we didn’t discuss

- Multimodal data integration
- Spatial -omics
- Machine learning and AI – but see: "features" and prior knowledge

This is not the end!

- It is the end of STAMPS, yes.
- But! There are online communities!
  - I will invite you all to the microbioinfo slack!
  - Some of us are active on bluesky! Tag me in to questions and I will boost!
- I am always happy to ask a question on your behalf; just drop me an e-mail with your question! Extra points if it's short and well phrased 😆

- There is also a contact sheet, in case you are ok with sharing your contact information (name, e-mail, & topics of interest).

Thank you all for coming to STAMPS 2025!

Graduation will be at 7pm

- Loeb tent!
- Then party afterwards ☺.