DOGEMS Overview

In a typical SEA-PHAGES section, more than 10 new phages are isolated but only 2 are sequenced. Perhaps there are unusual, interesting phages in the unsequenced group that would be fantastic additions to the data set. DOGEMS was designed as a way to get some sequence information for *every* student's phage and to find gems we'd otherwise have missed. It can also be an authentic use of PCR as part of a SEA-PHAGES course.

What is DOGEMS?

DOGEMS (DAH-jums) stands for **D**econvolution **o**f **G**enomes after **E**n **M**asse **S**equencing. In short, you create a DNA pool where you mix together all of the phages from your section (except those that are being sequenced individually), we sequence and assemble that pool as a single sample, and you figure out via PCR which pieces of sequence match which phage samples from your section. In the process, you get some sequence from every phage in your section, and most likely several complete sequences.

Part 1: En Masse Sequencing

Sequencing time and costs are based on the number of libraries that need to be prepared. To sequence 20 phages individually, 20 libraries must be made and then run on a sequencer. But if we were to pool those 20 phages' DNAs together in a single tube, only a single library must be made. This allows for a drastic reduction in labor and reagent costs associated with sequencing, while still providing some sequence information for each phage.

There are a couple of tradeoffs here. First, we won't know which sequencing reads came from each of the 20 input phages since they've all been mixed. Second, if there are two or more similar phages (e.g. from the same cluster), these will likely assemble poorly because their reads will match in some places but not others. Conversely, if there are unique phages (within your pool), they are likely to assemble well since their reads won't conflict with others' reads.

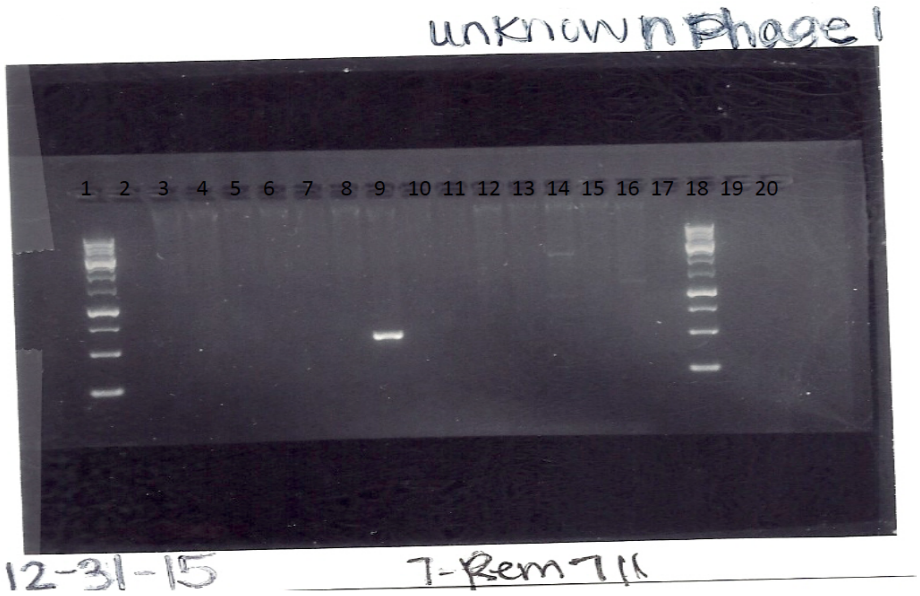
The result of the first part of DOGEMS is a list of contigs (assembled sequences of reads) that represent all of the phages in your pool. Some may be full-size and complete phage genomes, others will be small chunks of a different genome. This alone is an answer to the general question "What is the diversity of phages we've isolated?" and has inherent value. However, you'll probably find some interesting individual phages within the dataset, and want to follow up on these, which takes us to the second part of DOGEMS.

Part 2: Deconvolution of Genomes

Since you "convoluted" your phage DNA by pooling it in the first part, it's now time to "deconvolute" by matching the DNA sequences you got from Part 1 to their respective input phages. To do this, you can use the output contigs from the first part of DOGEMS to design PCR primers specific to one DNA sequence or another. Using PCR, you can then test each set of primers on each input phage until you get a hit. When you do, you'll have matched the sequence to the phage isolate!

For example: Florida Gulf Coast University did DOGEMS in 2015, and found that one of the phages in their pool would form a new cluster, Cluster Z. They designed primers to that new sequence and tried them on each phage from the pool. Results are below.

1. 1 KB Ladder
2. Skip
3. Esty
4. PriTime
5. Zenn222
6. Kova
7. Naytron3000
8. Imperio
9. Rem711
10. Infinity 21
11. TinyTim
12. Infinitev
13. Power
14. Omnicron
15. Danielle
16. Kyrie
17. Skip
18. 1kb Ladder
19. Skip
20. Skip



A PCR product was only observed when using the phage Rem711 as the template, so we now know that Rem711 is the Cluster Z phage. Since it wasn't one of the original phages sent for sequencing by FGCU, it would've been missed without DOGEMS!

How much deconvolution you do is up to you. You can just focus on the most interesting 1 or 2 phages, or you can try to design primers and assign every single phage from your pool to a matching DNA sequence.

How to participate

Since we at the University of Pittsburgh have agreed to sequence 2 *samples* per SEA-PHAGES section, if you want to participate you should plan on one of those two being a DOGEMS pool. Let Dan (dar78@pitt.edu) and Becky (rag4@pitt.edu) know that you're planning on sending a

pool as one of your samples before you ship your DNA to Pitt. In some cases, depending on sequencing load, we may be able to sequence a DOGEMS pool as a third sample from your section.

Pool Guidelines

We recommend trying to add the same amount (by weight) of DNA from each of your phages to your pool. For example, adding **200 ng of each phage's DNA** to a pool is a suggested guideline. Also, the number of phages in the pool will affect the output. More phages provides a broader look at what you've isolated, while fewer phages increases the likelihood of getting complete sequences out. We recommend pooling **10-20 phage DNA samples** as a good number. Your pool must have a **minimum of 2 µg of DNA total**, so if you're sending fewer than 10 samples, you may need to add more DNA per phage. Also note that **you should not include phages you've sent for individual sequencing** in your pool, as this is redundant and may complicate results.

Deconvolution Guidelines

After sequencing and assembly, we'll send you the DNA sequences we were able to create from your pool, along with information about how many of those are complete and how you might approach the next steps. You can BLAST these fragments/genomes on PhagesDB to see what clusters they belong to. You'll then have to set aside time for your students to do the PCR (design primers, order, and run) to target the phages you're interested in. Things are flexible, so in some cases this could be an independent research project for a specific student, or it could be a class-wide project where everyone participates.

FAQ

How is this different from metagenomic sequencing?
Most importantly: in a metagenomic sample, there are no stocks in the fridge/freezer to go back to. If you see an interesting sequence in a metagenomic sample, you have no way to work with that phage further. In DOGEMS, the phages are already individually isolated, and *then* pooled, so any interesting finds are available for downstream experiments.

How much does this cost?
Because we are doing the sequencing for you, the primary costs for your institution are the PCR primers and PCR reagents.

Do we know if this actually works, or is it just an idea?
We've run about 5 DOGEMS samples, and have found a number of uncommon phages, including a Cluster S, a Cluster T, a Cluster P2 (the only one!), and a Cluster Z.

How many DNA samples can I pool?
We recommend 10-20, but there may be reasons for doing more or fewer. Let us know if you plan something outside this range and we'll help work it out with you.

I want all of my students' DNA samples to be in the pool, but there are always some stragglers. Can I send the pool sample at a later time than my individual sample(s)?
Yes. It just requires that you pay for shipping them separately, but it's fine with us.

I like the idea of having every student's phage represented, but we can't do the PCR part. Can we do DOGEMS without deconvoluting?
We discourage walking down this path unless you're committed to doing at least some PCR to match phages with sequences. If we sequence your pool, and there's an interesting phage in there, we're definitely going to want to know which one it is, and may pester you!

Why don't we just sequence all phages this way, if it's cheaper and faster?
Because it's common to see two similar (but not identical) phages found by separate students in a section. And these phages will be the ones that do not assemble well after a DOGEMS experiment. For example, if four students from your class have isolated Cluster F1 phages, and these are mixed together in a DOGEMS pool, they will be nearly impossible to separate and assemble. Because of the way this approach works, the *rarer phages in your pool will assemble better*, which is nice because, well, they're rarer!

If we discover an interesting phage in our pool, then figure out which one it is by PCR, do we then send it back for individual sequencing?
In most cases, no. The number of reads we're getting has usually been enough to *already have* a complete sequence for the unique phages in your sample. If, however, you find an interesting phage that does *not* have a complete sequence, we'll be happy to sequence it for you.

How many complete genomes will we get out?
This will vary depending on the pool size and the diversity of the input pool. For example, if all of your input phages were from Cluster A, you'd probably get no complete genomes back. If each phage in a pool of 10 was from a different cluster, you might get 10! Those extreme examples aside, our practical experience shows about **1-4 complete genomes** per sample.

Can we begin annotating a complete genome from our DOGEMS sample before we've figured out which phage it belongs to?
Yes, BUT this annotation will be just an exercise unless you DO figure out which phage that sequence belongs to. We can't send an annotation to GenBank with the title "One of the Twenty Phages From My Class". If you're dedicated to sorting them out, then beginning annotation before you really know whose phage is being annotated is fine.

What if we used all our DNA up and don't have enough left to do PCR?
You can do PCR directly from phage lysates. Some protocols can be found on seaphages.org.

Can I do this as part of my regular wet-lab SEA-PHAGES course?
For the PCR part, you'll have to find other time to do it since the sequencing won't be done until after the first semester. It can definitely be incorporated during the bioinformatics semester. If that's not possible, however, it can be made part of an individual project, genetics course, etc.