

PRACTICE FOR WEEK 1: VIRAL STOICHIOMETRY

DANIEL MURATORE

1. MOTIVATION

The purpose of this practice is to apply some of the concepts from the week 1 readings to a practical example and to use some coding capabilities that are especially useful for bioinformatic analyses. Therefore, before starting this project, please finish the week 1 readings and if you have any questions or confusions please bring them up with Daniel.

In our Jover et al reading, we talked a little bit about stoichiometry of biological entities in the ocean. The stoichiometry of biology is extremely important for marine biogeochemistry/biological oceanography because it links the elemental cycles of carbon with other elements (here focused on nitrogen and phosphorus, but the connections extend to many other elements as well). Jover et al used some scaling relationships to estimate the CNP ratios of viruses based on genome length and virus size. The paper claims that in the limit of small viral particle size, the viruses will approach the Redfield ratio - an empirical ratio of marine phytoplankton derived a long time ago that claims plankton roughly follow the CNP of 106:16:1, and in the large size they approach the stoichiometry of DNA - which the authors report as around 9.75:3.75:1 on average (note that this ratio is SUBSTANTIALLY more N- and P-rich than Redfield). The authors do some calculations and ultimately claim that viruses could make up a substantial portion of DOP (dissolved organic phosphorus) in the surface ocean.

However, they make some important assumptions about how viruses work that may not hold in all cases. For example, we read the Grome and Isaacs overview piece on 2-aminoadenine replacement in the cyanophage S-2L genome. Although the 'ZTGC' adaptation may have important ecological attributes, the minor chemical change of adding an addition N atom to all Z base pairs also has a chemical impact. Let's investigate some questions surrounding that potential impact.

Our goal will be to calculate the CNP stoichiometry of cyanophage S-2L, predict its burst size based on the CNP stoichiometry of its canonical host, and propagate some uncertainties in our calculations.

2. PART 1 - CONTEXT

Before we dive into calculations, let's talk a little bit about what phage S-2L is. Please look into some basic information and provide a summary description of the phage. For example, where and when was it isolated? What is its host range? What size are the viral particles (this is essential as you will need it later)? What is its infection cycle like, is it purely lytic or have there been observations of lysogeny? Does it have any auxiliary

metabolic genes (AMGs)? Where in the ocean has it been found?

You don't need to provide answers to all of these questions specifically, but the objective is for you to be able to describe the virus if someone asks you what it is. The primary reference I recommend is "Cyanophage S-2L Contains DNA With 2,6-Diaminopurine Substituted for Adenine" Khudyakov et al, Virology 1978.

3. PART 2 - STOICHIOMETRIC CALCULATIONS

3.1. Genomic stoichiometry. Start by finding the cyanophage S-2L genome online (it can be found at <https://www.ncbi.nlm.nih.gov/>) and download the genome as a FASTA file. Tell me some basic characteristics of this genome assembly. Who submitted it and what methods are listed for how it was generated? How long is the genome? How many predicted proteins are there? The NCBI metadata should be able to provide all of those things. Now we calculate a few quantities.

Using whatever method works for you, though I would suggest either working with R bioconductor capabilities (tutorial: https://www.bioconductor.org/help/course-materials/2015/EMBO/A01_RBIOCForSequenceAnalysis.html#dna-amino-acid-sequences-fasta-files) or python Biopython capabilities (<http://biopython.org/DIST/docs/tutorial/Tutorial.html>) load the FASTA file of the genome you found. Now calculate the genomic GC content. Then, individually calculate the frequency of each nucleotide in the sequence. Using the method as described in supplemental section S.1.2 of the Jover et al paper, calculate the CNP of the S-2L genome.

But, we just learned that S-2L doesn't actually use A in nature. What is the difference in the CNP between A and Z? What do you hypothesize will happen to the CNP ratio? Incorporate this into your calculation and recalculate the CNP. How different is it?

3.2. Capsid stoichiometry. Using the estimates for the viral size dimensions you found during the context part as well as the methods described in the supplemental text for Jover et al, estimate the CN content of the capsid (there will be no P assuming the capsid has no phosphorylated proteins). Using your result here and your result for the genome, assume one complete genome goes into one capsid and calculate the total CNP for one S-2L particle. Normalize it to phosphorus to get a stoichiometric ratio. How does it compare with Redfield?

Now, let's assume that the head and tail measurements have a measurement uncertainty of 10%. What is one way you could incorporate that uncertainty into your calculations? Check in with Daniel when you have an idea. Once Daniel signs off, implement that idea and calculate a range of potential CNP for one virion (assuming no mutations the genome will be the same).

Please save any scripts and/or additional files you use to do these calculations with comments detailing your goals and thought process in this folder and push regularly, and never hesitate to ask Daniel any questions! There are some rather complicated tasks in here, so it is completely anticipated to have lots of questions!