Final Project Proposal

Nico Crudele

Nov. 1, 2015

Computation in the Physical Sciences  
Tiffany Timbers

Introduction:

CRISPR has been a revolutionary technology among the biotechnology community. First, it was discovered operating as an immune system, in Bacteria and Archaea1. CRISPR uses a guided endonuclease, most commonly Cas9 to perform site-specific double strand breaks of DNA, which can be useful for removing genes, or adding them into specific sites in a target organism’s genome2.

As an adaptive immune system, CRISPR uses a DNA library that is located on a specific plasmid in a location known as the CRISPR Array in order to catalogue phages that infect them. Between each phage sequence, known as a spacer, is a repetitive sequence of DNA, approximately 20 base pairs in length.

The proposed experiment is to create a program that is capable of taking a full bacterial genome as an input, and output a list of phage DNA that is separated, where the repeats are. This would give a list of individual fragments of DNA from different phages.

Methods:

The four main things that this program needs to do are:

1. search through bacterial genomes as a data set
2. find CRISPR repeats
3. remove repeats
4. list sections of phage DNA spacers, with a line break where each repeat was

Expected outcomes:

It is expected that the program will successfully separate the phage DNA fragments, and provide a list of each of those sequences found between CRISPR repeats.

Significance:

If this program works, the applications are various. It could be expanded upon to compare the phage DNA sequences against each other, showing genetic similarity between phages that have infected a specific Bacterium or Archaeum. This could give insight into the phages that are present in an environment that hasn’t been studied, to know which ones infect specific bacteria in that environment, and better understand that ecosystem.

The mechanism for new spacer insertion into the CRISPR array is unclear3, if future research proves the mechanism that the Cas1 and Cas2 proteins use to insert spacers into a CRISPR array, i.e. the order/sequence that spacers are inserted into the array, this program could be expanded to provide a temporal map of phage infection of examined bacterial species. If it happens to be in a linear sequence, i.e. that there is a starting spacer position, and all other spacers are added chronologically after that one, no change to the program would be necessary, as they would already be listed in chronological order.

Finally, this program could be useful for diagnosis of viruses in a human context. If we insert benign bacteria harboring CRISPR arrays into an infected individual, allow the CRISPR system to uptake spacers from the virus infecting the human, remove the bacterium, and scan it through this program, we could compare the CRISPR arrays from before and after the infection of the human. This would show us the difference of the two arrays, i.e. the spacers that are from phages that were present in the human, and not in the bacterium previously, which could later be analyzed to figure out what was infecting that human. Generally, the CRISPR system would be a much more useful tool if we can catalogue and analyze the phage spacers in the array.

References:

1.

Barrangou, R. The roles of CRISPR–Cas systems in adaptive immunity and beyond. *In Innate immunity* (2015). doi:10.1016/j.coi.2014.12.008

2.

Torres-Ruiz, R. & Rodriguez-Perales. CRISPR-Cas9: A Revolutionary Tool for Cancer Modelling. *International Journal of Molecular Sciences* **16,** 22151–22168

3.

Rath, D., Amlinger, L. & Rath, A. The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie* **117,** 119–128 (2015).