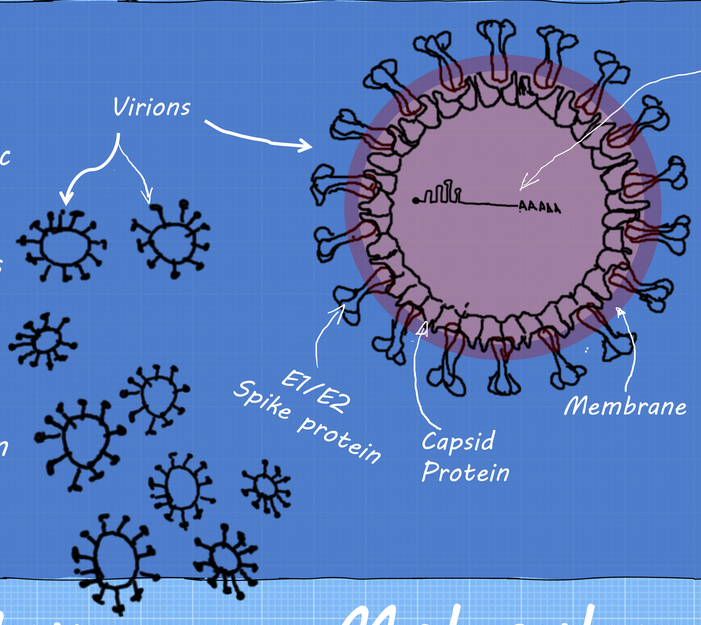
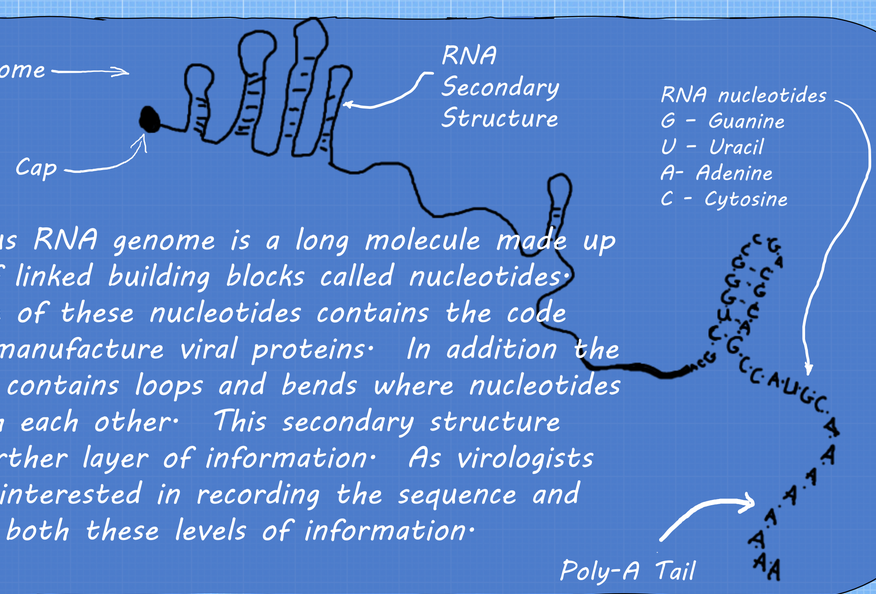
Developing a method to investigate viral genomic structure and variation

Introduction

At the Forrester Lab we work with alphaviruses many of which are pathogenic viruses transmitted by mosquito vectors. These viruses, such as Chikungunya, Venezuelan equine encephalitis and others can cause severe illness in human and livestock hosts. Viruses replicate by producing virions which are made up of viral proteins and contain a copy of the viral genome. The viral population within an infected host consists of many virions and many copies of the viral genome.



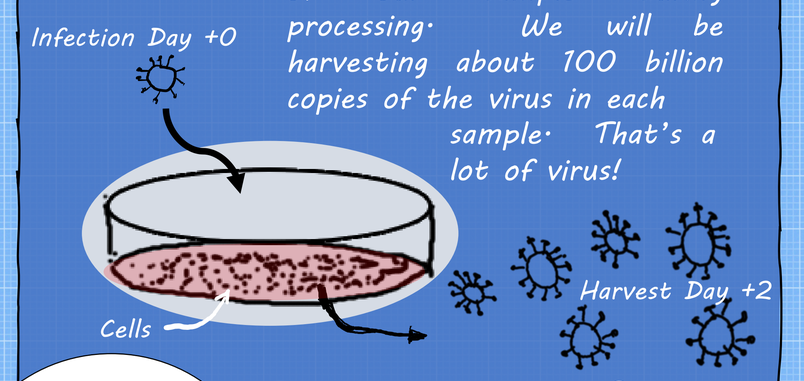


The alphavirus RNA genome is a long molecule made up of a chain of linked building blocks called nucleotides. The sequence of these nucleotides contains the code required to manufacture viral proteins. In addition the viral genome contains loops and bends where nucleotides interact with each other. This secondary structure encodes a further layer of information. As virologists we are very interested in recording the sequence and structure of both these levels of information.

Virus + Cell Culture

1 Propagation

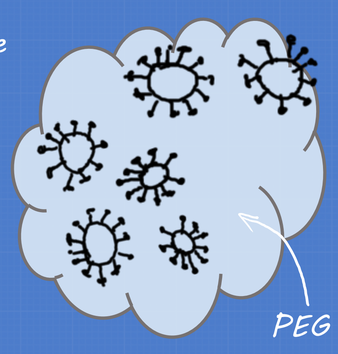
We grow our virus in vitro, infecting either mammalian or insect cells in culture. We will need a lot of virus particles as we will lose some of our sample during processing. We will be harvesting about 100 billion copies of the virus in each sample. That's a lot of virus!



2 Concentration

We now separate our viral particles from the heavier host cell fragments as much as possible by centrifugation.

To further concentrate our virions we add polyethylene glycol (PEG) which forms a sticky matrix which binds our virus particles. We can then separate this from any excess liquid.

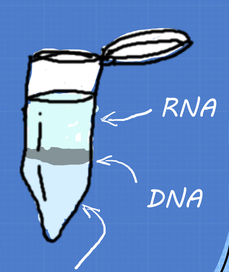


We have now concentrated our virus sample 100- 150 times.

3 RNA Extraction

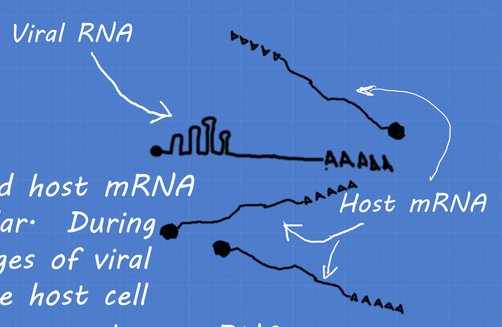
We now need to extract the viral RNA genomes from the virions.

We do this using a guanidinium thiocyanate-phenol-chloroform extraction method. This first inactivates and disrupts the viral particles releasing the RNA into solution. The addition of chloroform separates RNA from proteins and host DNA in our sample allowing us to isolate the RNA.



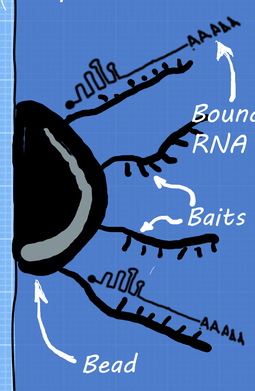
Molecular Biology

4 Enrichment



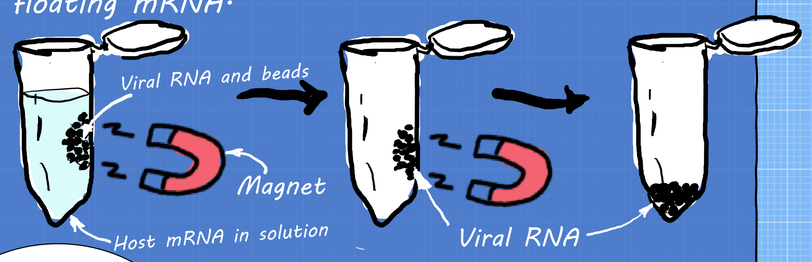
Viral RNA and host mRNA are very similar. During the early stages of viral replication the host cell machinery that translates mRNAs to proteins will translate viral RNA to viral proteins.

Our sample now contains a large amount of viral RNA but also host mRNAs which we want to exclude from later sequencing steps.



We design a series of baits which consist of nucleotide sequences complementary to the viral genome sequence. These baits are bound to small iron-cored beads.

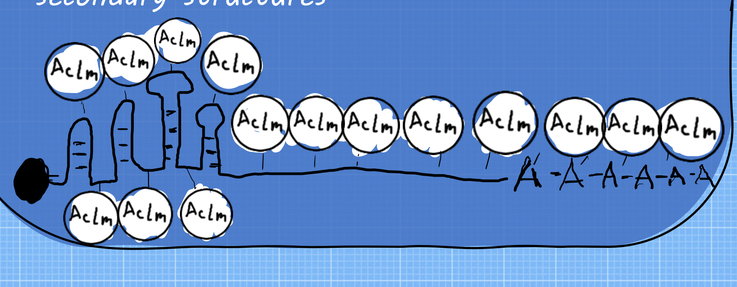
When these beads are added to our sample the viral RNA is trapped by our baits. When we place our sample in a magnetic field we can separate the bead-bound RNA from the free floating mRNA.



5 Labelling

We now have a relatively pure sample of viral RNA. To investigate the structures that interest us we need to identify those structures.

We incubate the RNA in a buffer solution that allows it to fold into it's normal shape and then we add acetyl-limidazole (Aclm) which binds to those regions of the genome that are not involved in forming any secondary structures.

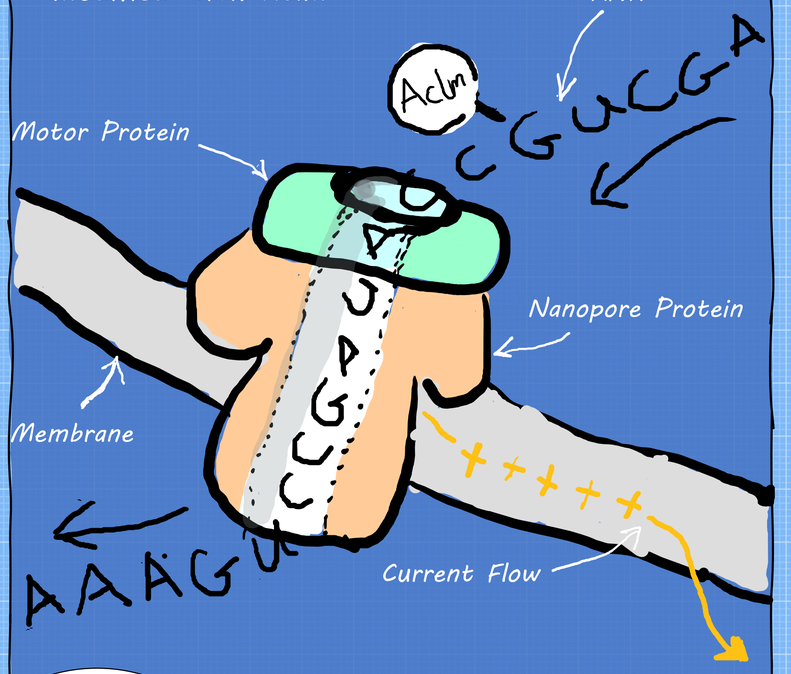


Bioinformatics

6 Sequencing

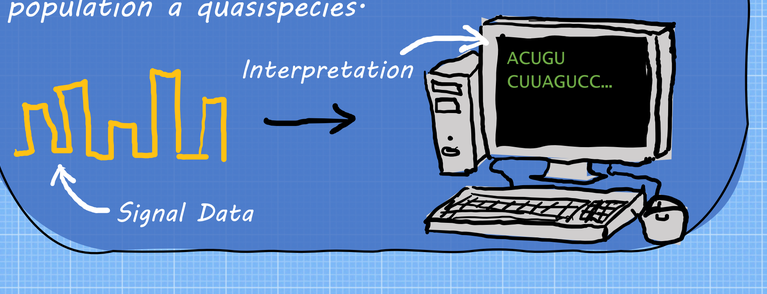
We are now ready to sequence our RNA. We do this using the Oxford Nanopore MinION sequencing platform.

RNA is fed by a motor protein through one of thousands of protein pores in an artificial membrane. The variation in current flow associated with the passage of each nucleotide through the pore is recorded by the nanopore device. This method of next generation sequencing is ideal for our purposes in that it can read long sequences, potentially the full length of the viral genome. It can also read the signal produced by the nucleotides we have modified with Aclm.



7 Analysis

The output electrical signal we recorded can be interpreted by a custom base-calling algorithm to generate a record of the nucleotide sequence and secondary structure of each Viral genome that is read by the nanopore device. Using this data we can build a picture of genomic and structural variation across the viral population. We call this varied viral population a quasispecies.



Why are we doing this again?

That's a good question. It's not like viral genome sequencing hasn't been done before and alternate methods have been used to investigate viral genome structure. The problem with these other methods is that they can generally only look at fragmented sections of the genome. To assemble a complete picture of the viral quasispecies from these methods is like trying to assemble hundreds of thousands of jigsaw puzzles at the same time when most of the pieces are the same. Our method, aiming to sequence full length genomes should provide a valuable tool for characterising the sequence and structural diversity of viral populations.

References

- Chomczynski, P., Sacchi, N. **The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on.** *Nat Protoc* **1,** 581–585 (2006). doi.org/10.1038/nprot.2006.83
- William Stephenson, et al., **Direct detection of RNA modifications and structure using single-molecule nanopore sequencing**, *Cell Genomics*, 2(2), (2022) doi.org/10.1016/j.xgen.2022.100097
- Rayaprolu V, Ramsey J, Wang JC, Mukhopadhyay S. **Alphavirus Purification Using Low-speed Spin Centrifugation**. *Bio Protoc.* 8(6):e2772. (2018). doi:10.21769/BioProtoc.2772
- Tan, C.C.S., Maurer-Stroh, S., Wan, Y. et al. **A novel method for the capture-based purification of whole viral native RNA genomes**. *AMB Expr* 9, 45 (2019). https://doi.org/10.1186/s13568-019-0772-y

This project was funded by the Keele University School of Life Sciences Pump Priming Fund...Thanks!

Jude Martin