# R21 project sequencing results / June 2020

## Abstract / Intro

In this project, 11 cynomolgus macaques were infected with SIVmac251. 7 of these were infected with *Mycobacterium tuberculosis* (Mtb) prior to SIV infection and 4 of them only with SIV. We have 69 sequenced plasma and tissue viral RNA samples from the 11 monkeys. In addition, we have one sample from the virus challenge stock.

We found no evidence of selective sweeps (except in one monkey at one site). This is likely because the virus did not replicate well in the macaques and because the monkeys were only infected for 8-9 weeks (compare to Ita et al. 2018). We find evidence for mutation selection balance with non-synonymous and nonsense mutations being present at lower frequencies than synonymous mutations. Include comparison of diversity between plasma vs. tissues and between 2 groups!

Three main issues make sequence data analysis for these samples difficult.

1. The virus is not virulent in this species of macaques and levels of viral RNA are low in the plasma and tissues.
2. The virus challenge stock also had low diversity, so the animals were infected with almost clonal virus.
3. The monkeys were only infected for 8-9 weeks, which did not allow much replication, and thus diversity, to occur.

As a result, we had very little diversity to analyze and compare between plasma and tissues or between animals.

## Topics to include in Introduction.

TB / HIV are important. Many infected. Combination leads to high morbidity and mortality.

Difficult to study how these affect each other.

In vivo work one option. This study focuses on sequencing of material from macaques from previous study.

Describe Ita et al study.

## Results

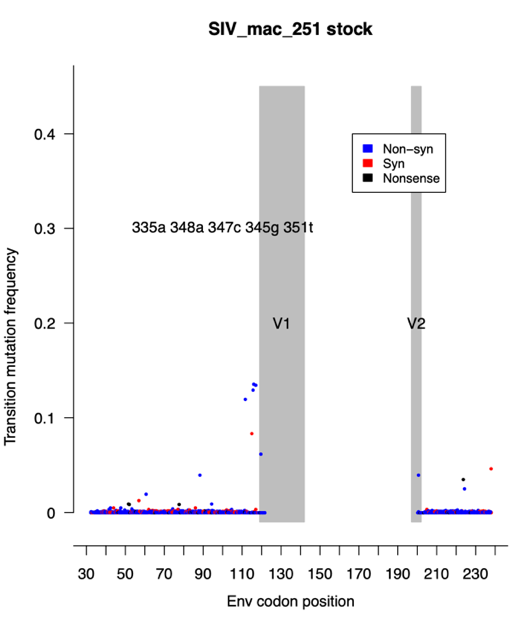


Figure 1

### Diversity in stock virus

Diversity in the virus challenge stock is low (Figure 1). Only X many sites out of Y many sites show a mutation frequency of >1%. Diversity of this challenge stock appears to be lower than that used in a similar study using rhesus macaques (Ita et al. 2018).

Diversity in the monkeys is even lower than in the stock. Quantify this. Only X many sites of Y many sites show mutation frequency greater than 1% (and X of Y at later time point). Likely, this is because there was a significant bottleneck when the macaques where infected and there was not enough replication to generate much diversity in the 8-9 weeks duration of the study.

Also, it is likely that there was more diversity in the V1 loop, but we didn’t sequence that due to the primers being too far away from the loop and limited read length from MiSeq sequencing. Indicate in the figure where the V1 loop is.

One macaque had evidence for a soft selective sweep, where a mutation at position 359g (K120R) started from standing genetic variation at 10% frequency, reached a frequency to 70% frequency within X many weeks (Figure 2). Check percentages.

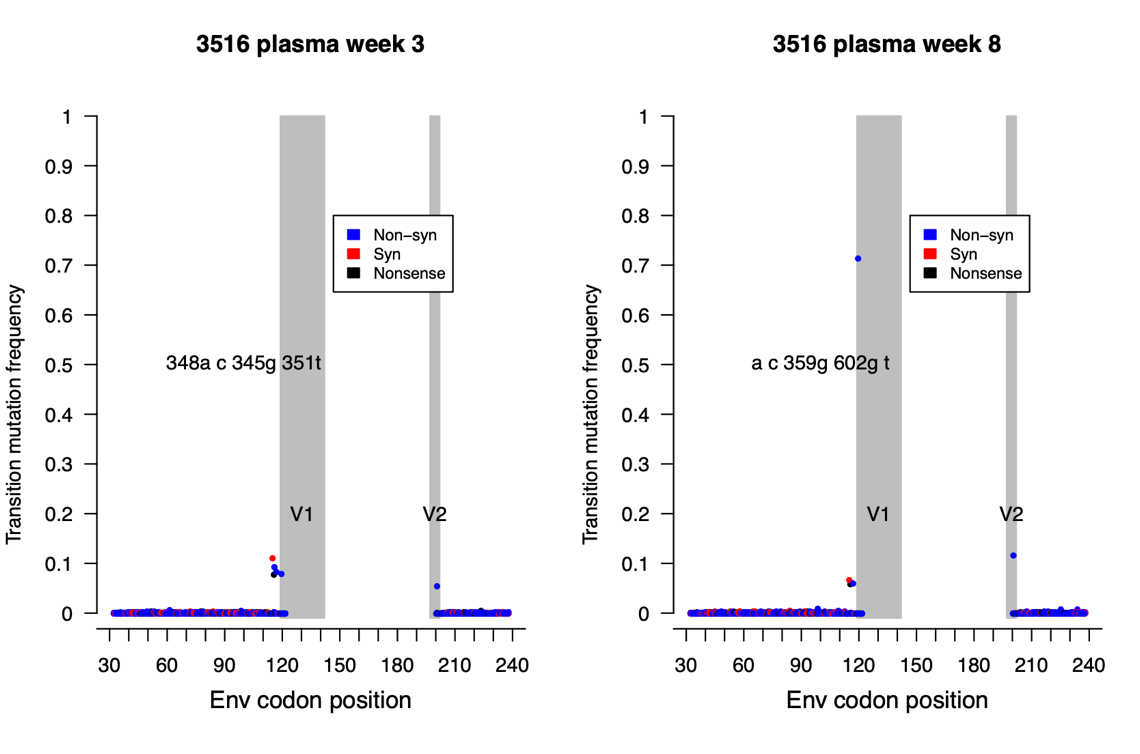


Figure 2

### Diversity higher in plasma than tissue, and lower in SIV-only animals



Figure 3

Figure 3 shows the average genetic diversity for each sample. The red symbols represent samples from animals that were infected with TB and SIV. The blue symbols represent samples from animals infected with SIV only. Overall, genetic diversity is very low. Quantify how low / low compared to what?

For plasma samples, the SIV only animals have lower diversity (0.63% on average) than the co-infected animals (0.71% on average) (p = 0.02, one-sided Wilcoxon test). These results suggest that SIV replication is greater in Mtb/SIV co-infection than SIV infection, which is consistent with levels of viral RNA in the plasma and granulomas (Diedrich et al., 2020). According to a non-parametric test, plasma samples have higher diversity (0.69% on average) than tissue samples (0.58% on average) (p = 0.014, two-sided Wilcoxon test), suggesting that more SIV replication is occurring in plasma than in tissues. Whether this is due to differences in SIV target cells (CD4+ T cells in the blood vs. alveolar macrophages and CD4+ T cells in the lung) should be investigated in future studies.

Compare tissue in SIV vs TB/SIV, compare plasma in SIV vs TB/SIV.

### No difference between genetic drift in plasma, lymph nodes and lung tissue.

We were interested to determine whether the viral dynamics was different tissue compared to plasma. Particularly, we predicted that virus may undergo more genetic drift in the tissues than in the blood plasma because we expect smaller viral population sizes in tissue compartments when compared to blood plasma. For this analysis we focused on sites that had a fairly high frequency in the stock virus (17 sites with a mutation at 0.8% frequency or higher). For these high frequency sites, we determined their frequency in the early plasma sample (at 2-5 weeks) and compared that baseline frequency to the frequency of the same mutation in (1) the late plasma sample (8 or 9 weeks), (2) the lymph node samples, and (3) the lung tissue samples. In general, the frequency changes were modest. We found that the mean absolute frequency change between the plasma samples was 1.7% of the early plasma frequency. Between the lymph node samples and the early plasma samples, it is also 1.7% of the early plasma frequency. Between the lung samples and the early plasma samples, it is 2.3% of the early plasma frequency. A non-parametric test (Wilcoxon) for all comparisons was not significant.

### Frequency of synonymous vs non-synonymous transition mutations.

We next compared the mutation frequency of synonymous vs. non-synonymous sites. Here we focused on transition mutations (A<->G and C<->T) because they are more common than transversion mutations in viruses. Indeed, there was a difference. In most samples (though not all), the non-synonymous and nonsense transition mutations have lower frequency (according to non-parametric Wilcox test) than synonymous mutations.

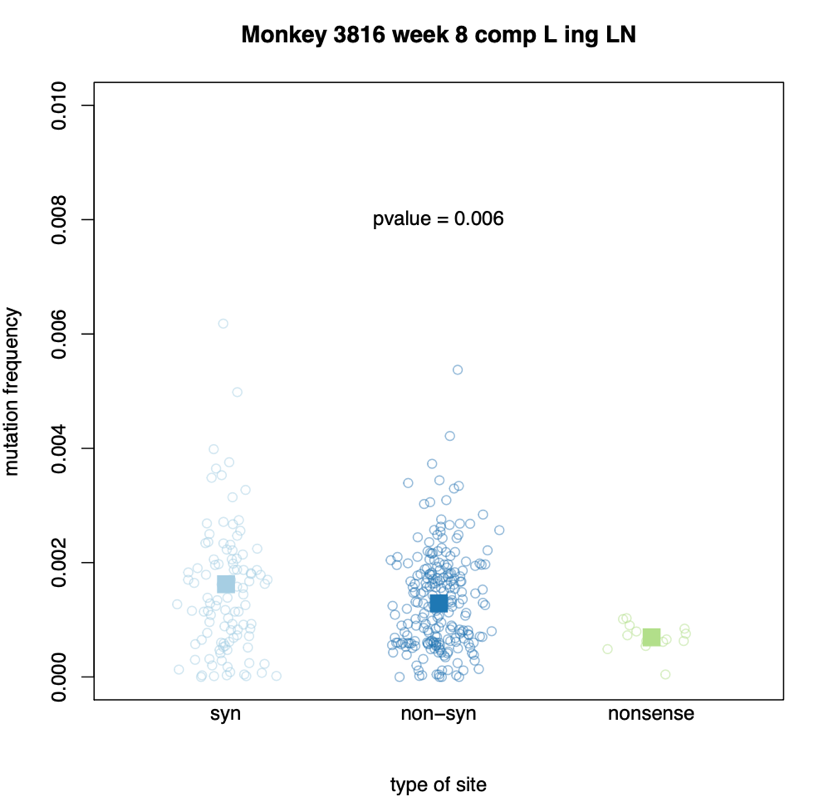


Figure 4

In a sample from animal 3816, non-synonymous mutations have, on average, a lower mutation frequency than synonymous mutations (Figure 4). The p-value is for a one-sided Wilcoxon test to determine if synonymous mutations have higher frequencies than non-synonymous and nonsense mutations combined. We find that for plasma samples, 16/19 have a significant difference between synonymous and non-synonymous/nonsense mutations (84%). For tissue samples, 26/50 have a significant difference between synonymous and non-synonymous/nonsense mutations (52%). However, this difference may be due to better sample quality of the plasma samples or due to higher effective population sizes in the blood.

This result, that frequencies of synonymous sites are higher than for non-synonymous sites shows that there is signal in the data. In other words, what we observe is real diversity, and not just sequencing noise. In addition, there is selection occurring on non-synonymous and nonsense mutations in the macaques. Selection makes these sites less frequent. Unfortunately, due to the low replication rate of SIVmac251 in cynomolgus macaques, obtaining enough sequences from enough samples makes comparisons difficult. Future studies will be conducted with a more pathogenic virus.