**Introduction**

MS2 is a bacteriophage that infect *E. coli* bacteria. In recent findings we found that among the different variants of MS2 we can find “cheater viruses”. Those cheaters hold some mutations that make them prone to better reproduction, however these are only flourish in the presence of other viruses, as co-infection. These viruses depend on infection of their partner variants in order to use some of their coding proteins. To further research the existence of those viruses we preform transduction experiments and freeze the status after each transfection (viruses are in the media of the plates). We do these experiments in different MOI to assess who are the cheaters (in low MOIs cheaters population will be eliminated cause chances of co-infections are low).

**Input files**

1. A screenshot of a table

   Description automatically generated with low confidenceFreqs Files from different experiments – These files are large CSV file that states the nucleotide distribution along the different position according to a reference genome.  
   Example of a freq file:

There will be different Freqs files – for different timepoints, different lines (biological replicated), different MOIs (0.01/1/10).

1. Reference sequence of MS2
2. Library that color coding the different mutation to specific colors.

**Pre-Processing**

1. From the multiple Freqs files create a dataframe of mutation per experiments (there are 11 experimets).

2. Join this Mutations table into a one dataframe and add a column of Experiment

3. Filtering - Not all the mutations will be added to the final analysis, I need to define many cutoffs (for example minimum coverage of that position and the minimum frequency).

**Analysis and Graphing**

1. Our First analysis is regarding the presence of curtain mutation in different MOIs.  
In order to do so I will first detect each of the recurring mutations. Than I will be a producing a graph of frequencies along the passage for each of the mutation where each line in this scatter plot will be showing the frequency along the passages.  
**\*Interspecific competition Lotke-Voltera**

2. To further asses the low frequencies in passage 0, I will create a Heatmap where one axis discrive the different mutations and the other the different experiments.

3. Lastly – we would like to locate the mutation areas along the MS2 genome. For that I will create a scatter plot along the genome of the MS2, where the X axis will show the location along the genome, and the Y axis will show the frequency of each mutation. This graph will also include a map of the different areas along the genome.

1. INPUT – קבצי FREQS מריצופים העברות בMOI שונה (1 ו-10).
2. השוואת חתך של המוטציות בגרף פיזור על פי CUTOFF שונים והטיפולים שונים (ניסויים של מורן/כרמל/שיר/מריה) – במקום סעיף 3
3. HEATMAP – שמתאר את השכיחויות הנמוכות בהעברות ההתחלתיות (P0). ציר אחד מתאר את המוטציות השונות והציר השני מתאר את הניסויים השונים.l
4. גרף פיזור גנומי בו ציר הX הוא הגנום וציר הY הוא שכיחות. המקרא יראה ניסויים שונים. שמות המוטציות הן לפי אלו שהופיעו לפחות פעם אחת מעל 10% (פר מוטציה). אפשר לצבוע את הרקע לפי החלבונים של הגנום.