**Day 2: Gene Annotation**

Protein Name: Coronavirus Spike glycoprotein

Protein ID – PF01601.19

**Find the following Gene function details**Location- Start and end: 692 to 1214

Family: CoV\_S2 Coronavirus spike glycoprotein S2

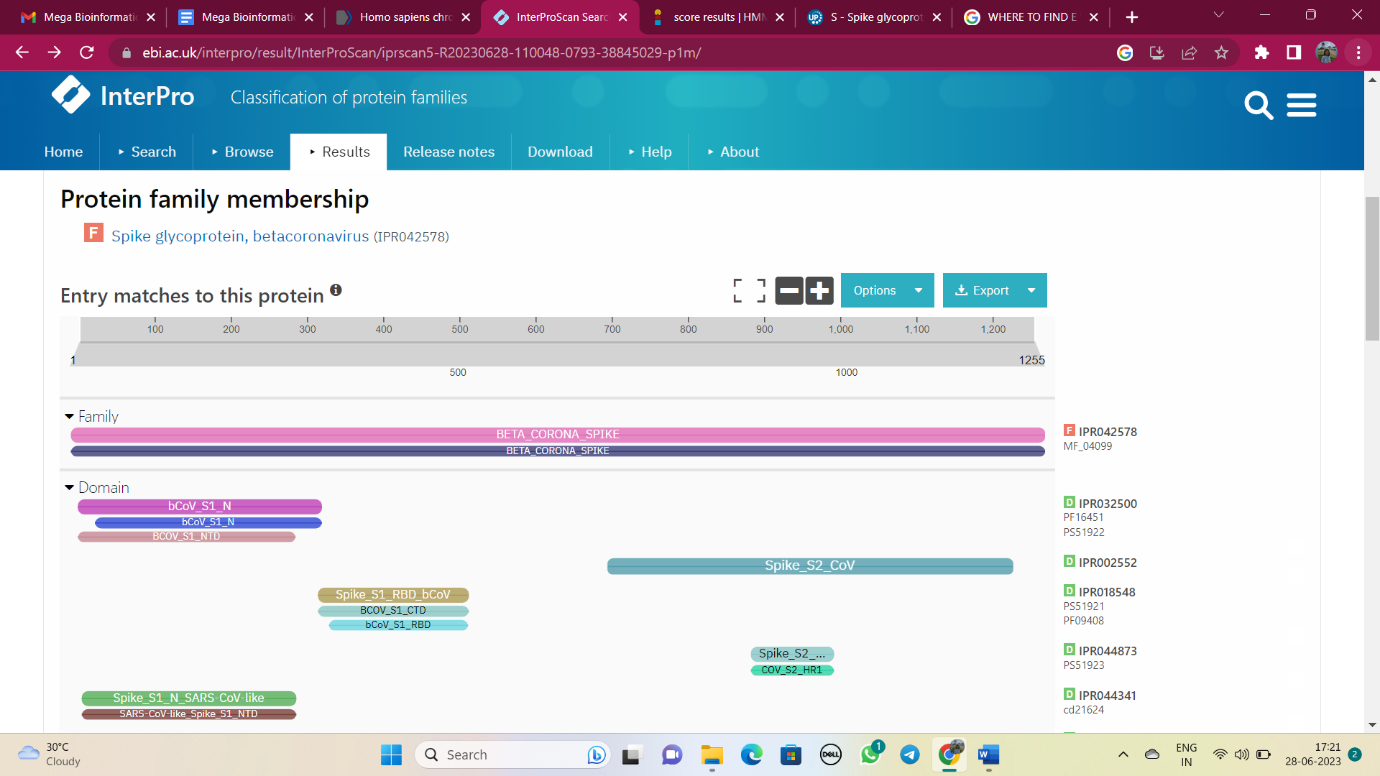
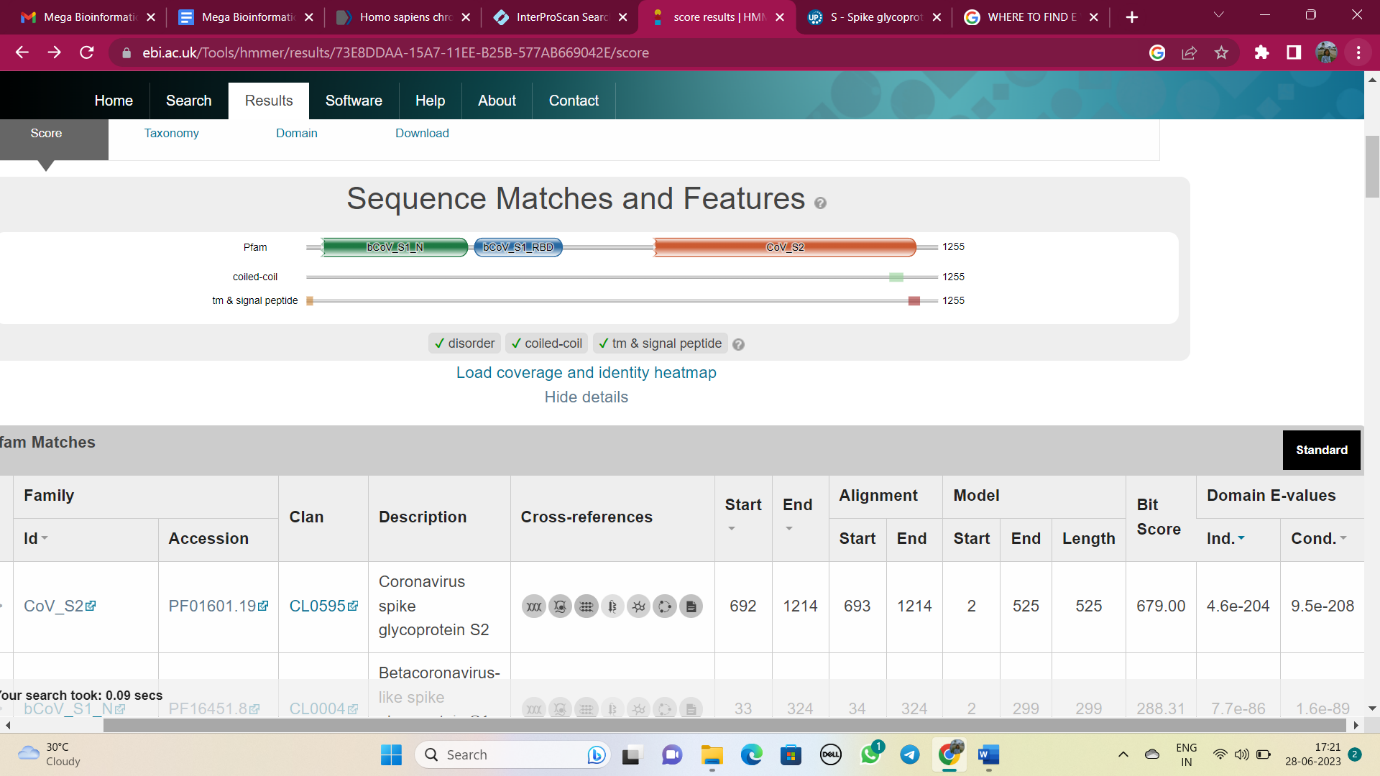
Clan: CL0595

Domain: N-Terminal domain, transmembrane helix domain (domain architecture represented by P10033) [Also contains cytoplasmic and non-cytoplasmic domains and single peptide domains]

Motif: 1251-1255, sequence- KLHYT

E value: Ind: 4.6e-204 and cond: 9.5e-208

Description of function: The S2 subunit normally contains multiple key components, including one or more fusion peptides (FP), a second proteolytic site (S2') and two conserved heptad repeats (HRs), driving membrane penetration and virus-cell fusion. The HRs can trimerize into a coiled-coil structure built of three HR1-HR2 helical hairpins presenting as a canonical six-helix bundle and drag the virus envelope and the host cell bilayer into close proximity, preparing for fusion to occur.

**Picture of results from HMMER and Interpro:**

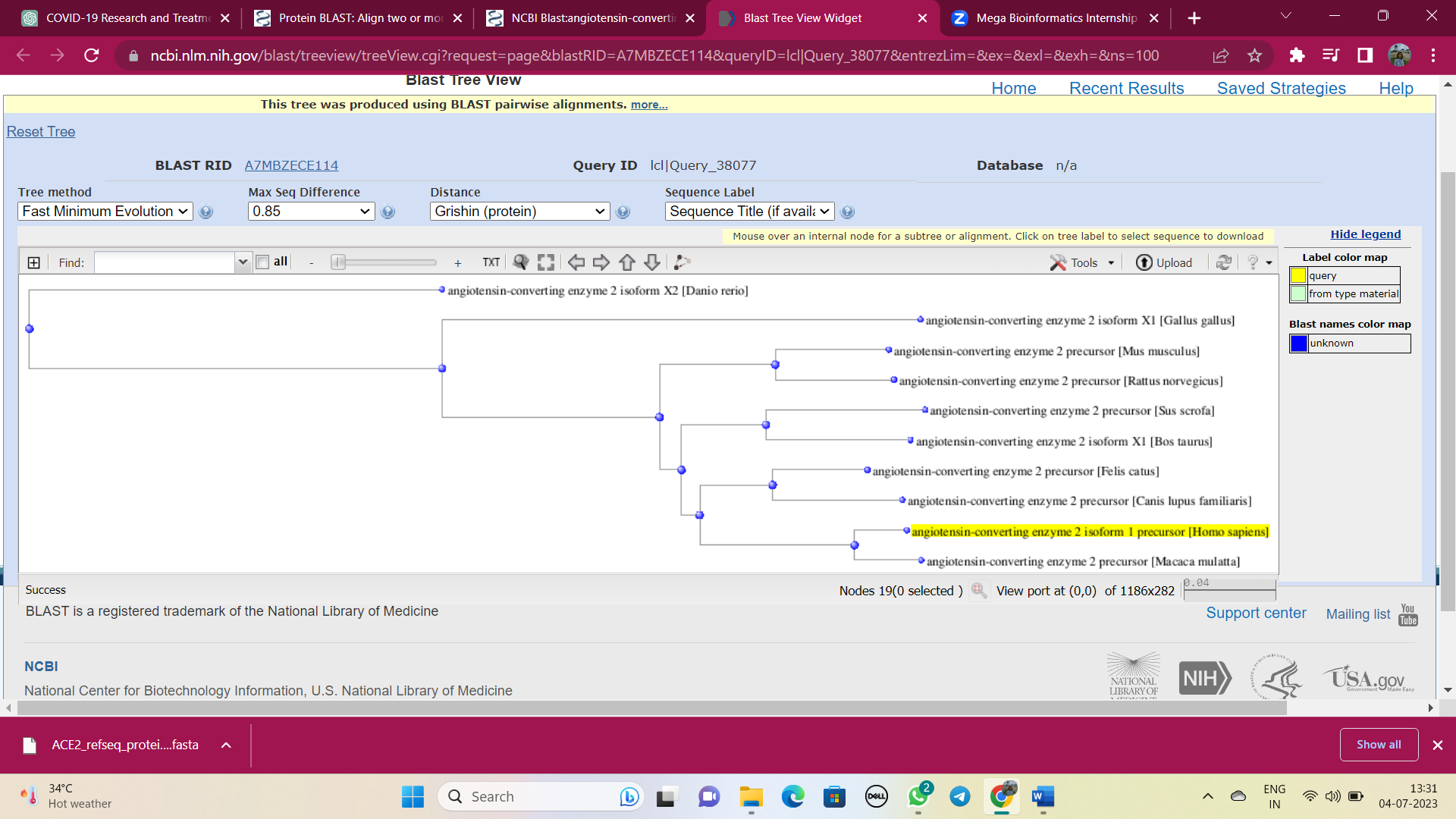
**Day 3: Phylogenetics**

**Construct a Phylogenetic tree for components of Corona virus.**

You can choose any gene/protein/component associated with Corona virus for atleast 5 different species/variants. Add the screenshot of the tree here.

Gene: ACE2 angiotensin converting enzyme 2

Gene ID: 59272 (ncbi)

Phylogenetic tree formed: **picture result from ncbi BLAST**

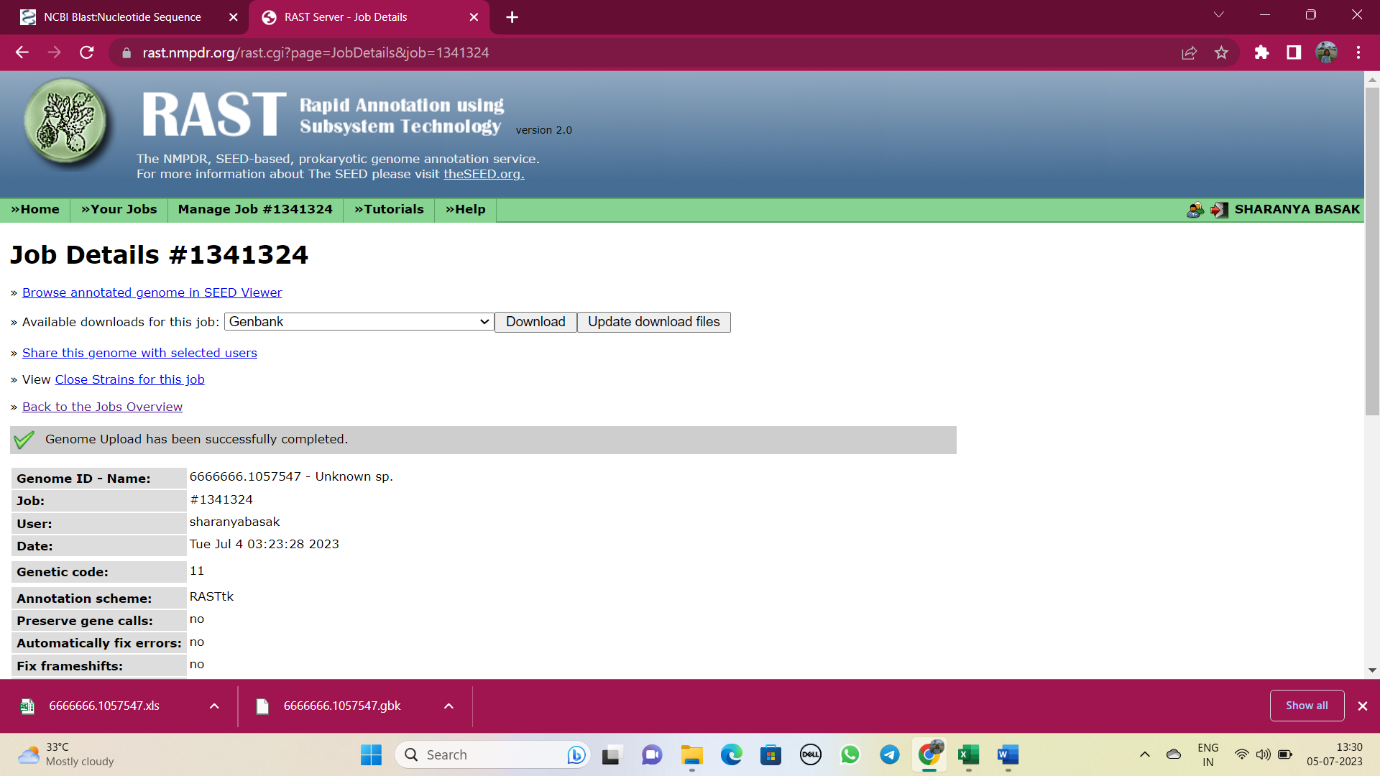
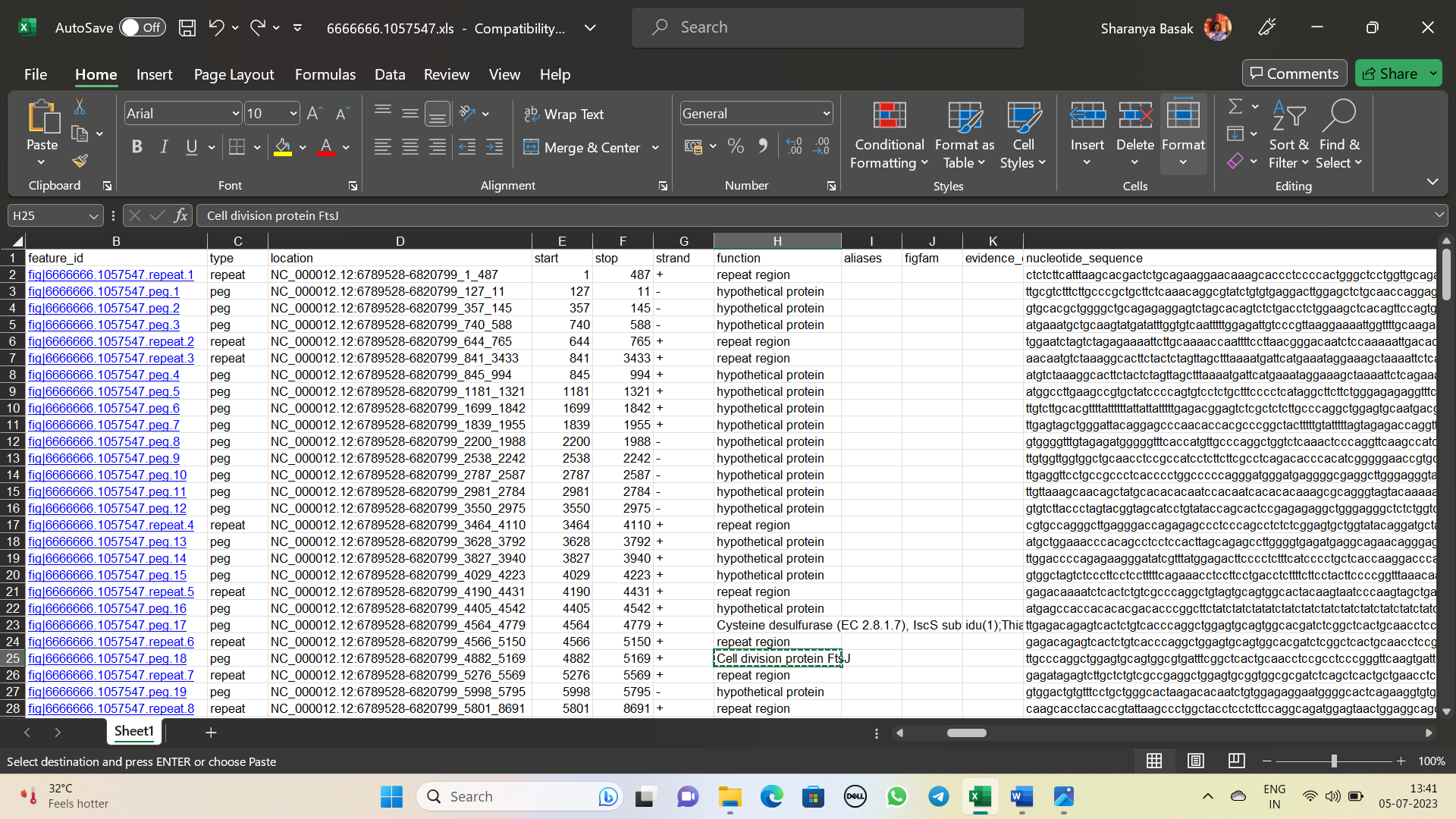
The tree was formed using the orthologs as follows:

1. *Homo sapiens* (reference)
2. *Mus musculus*
3. *Rattus norvegicus*
4. *Gallus gallus*
5. *Canis lupus familiaris*
6. *Danio rerio*
7. *Bos taurus*
8. *Felis catus*
9. *Macaca mulatta*
10. *Sus scrofa*

**Day 4:**

**Genome name (the one of your interests): NC\_000012.12 CD4 receptor in *Homo sapiens* (Chromosome 12) [Gene ID- 920]**

**From RAST results:**

*Mention the desired nucleotide sequence that you choose to perform BLAST on and fill in the following:*

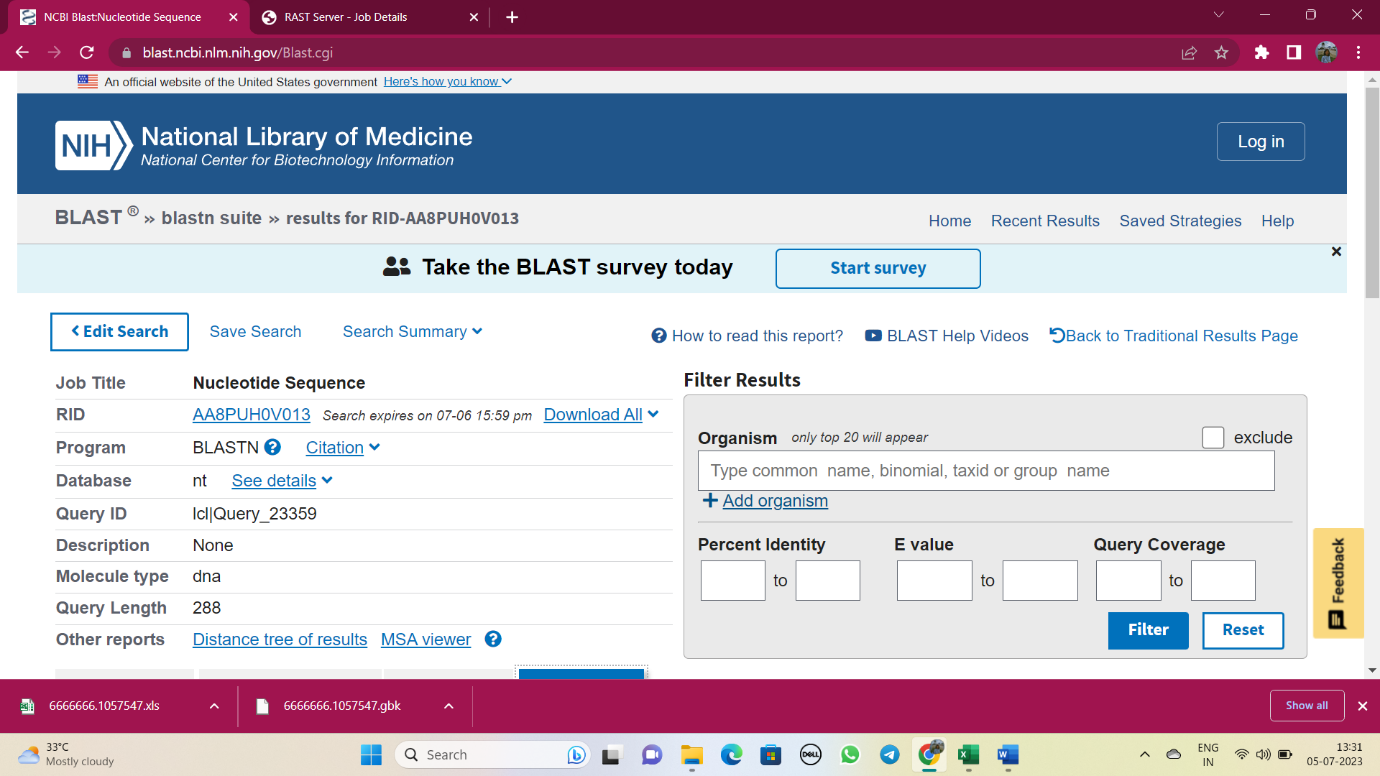
1. Nucleotide sequence:

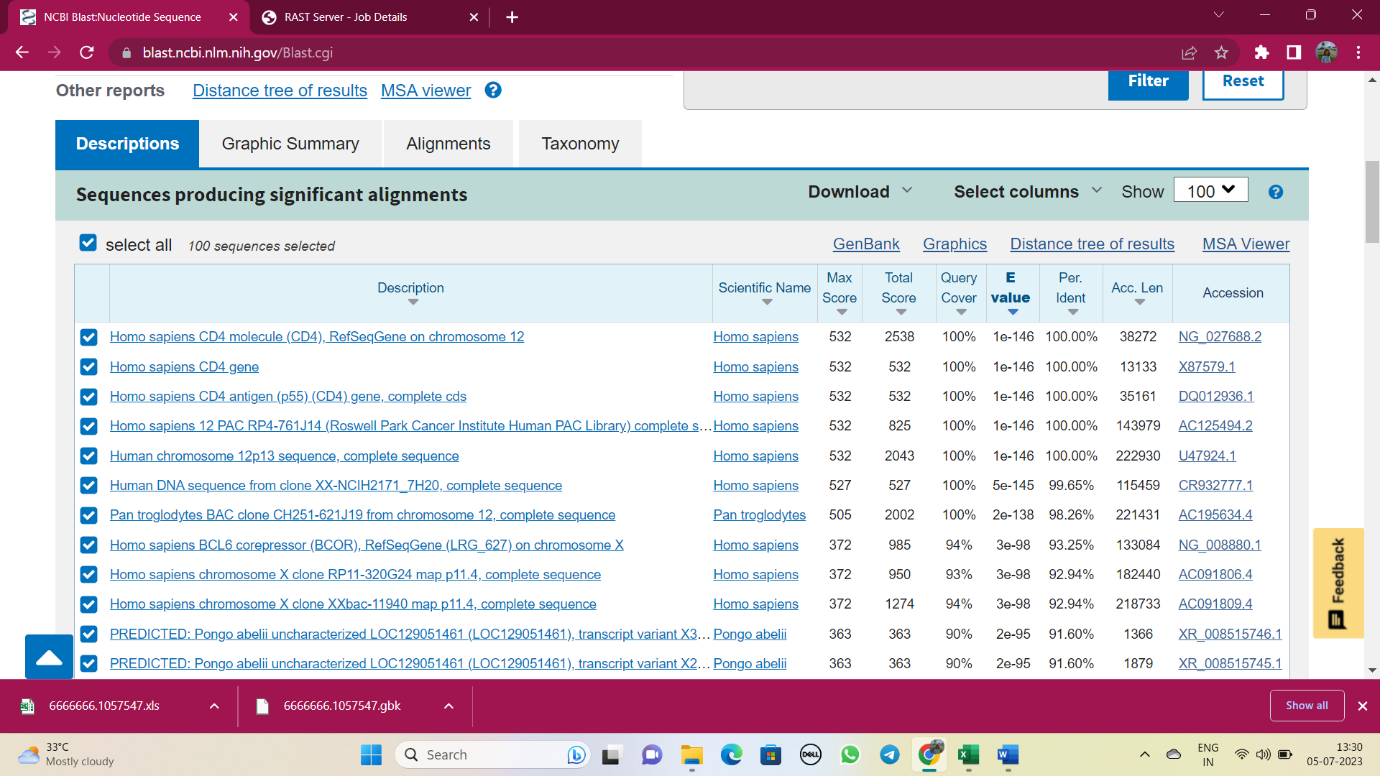
ttgcccaggctggagtgcagtggcgtgatttcggctcactgcaacctccgcctcccgggttcaagtgattctcctgcctcagcctcccaagtagctgagattacagacgtgcgtcaccatgcccagctaatttttgtatttttagtagagatgggatttcactatgttggccaggctggtctcgtactcccgacctcaggtgatccacttgccttggcctcccaaagtgctggaattacaggtgtgagccactgcatccggccttatatatctatcttgtctgtctga

1. Location on the genome: NC\_000012.12:6789528-6820799\_4882\_5169
2. Start and end nucleotide: Start- 4882, End- 5169
3. Function: Cell division protein FtsJ

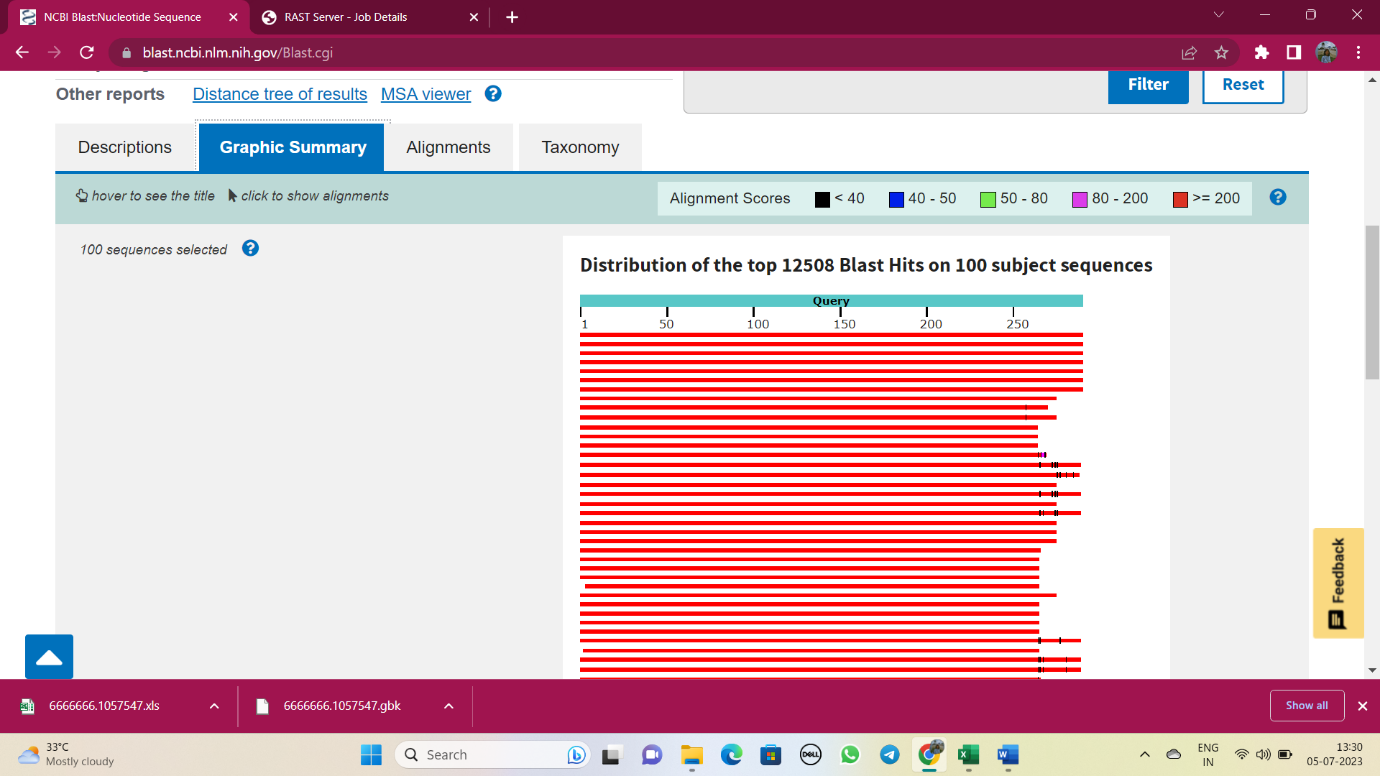
**Perform a BLAST on the nucleotide sequence and paste a screenshot of the obtained BLAST results:**

**DESCRIPTIONS:**

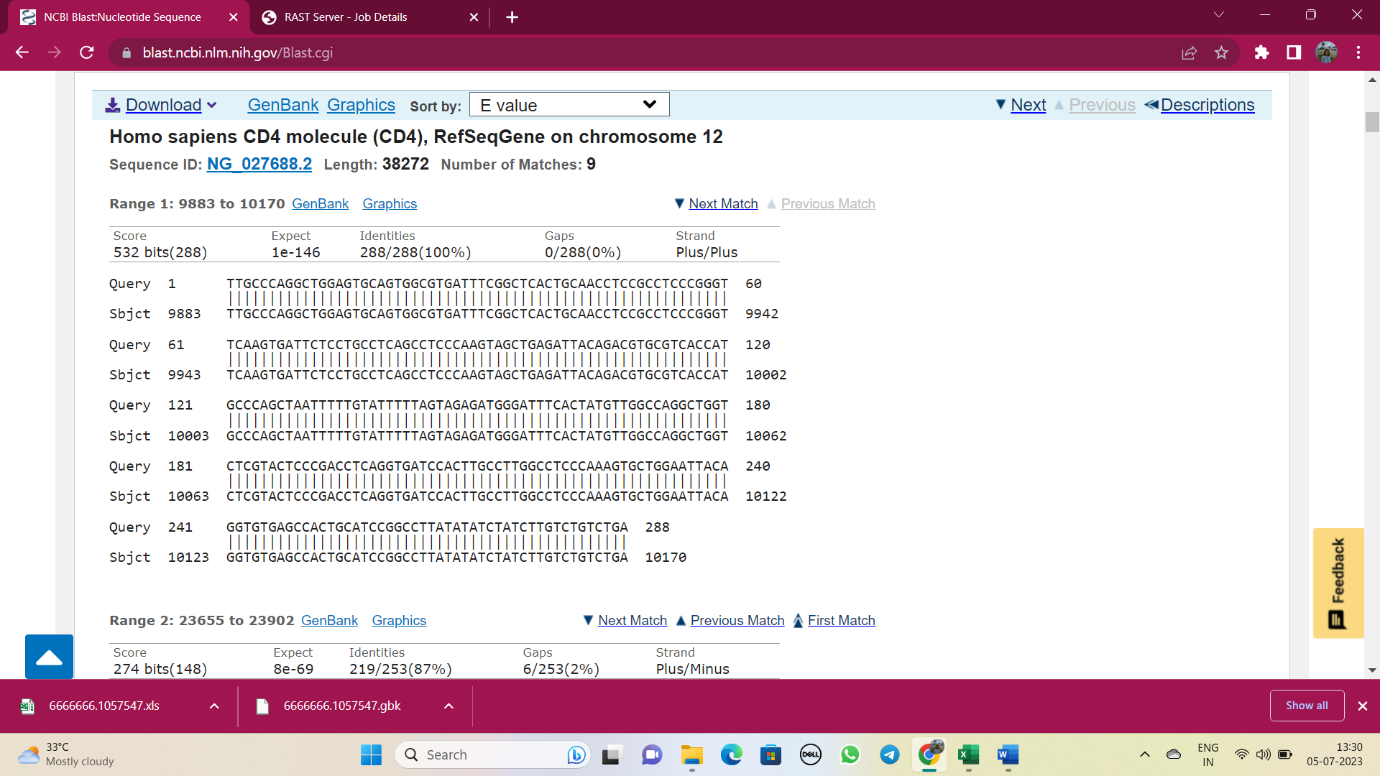
****

****

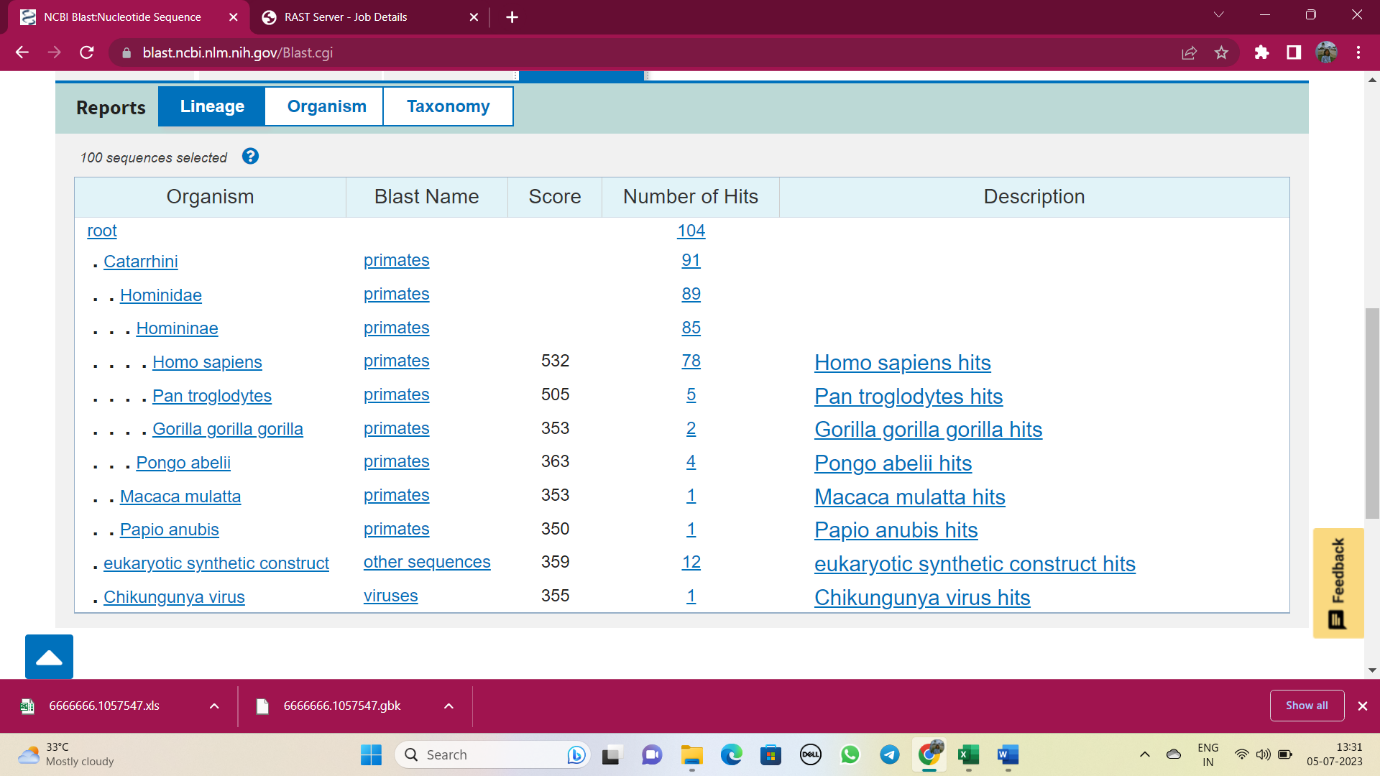
**GRAPHIC SUMMARY:**

****

**ALIGNMENTS:**

****

**TAXONOMY:**

****

**Day 5 & 6: Molecular Docking**

Protein Name: HIV-1 Reverse Transcriptase

Protein ID – 1VRT

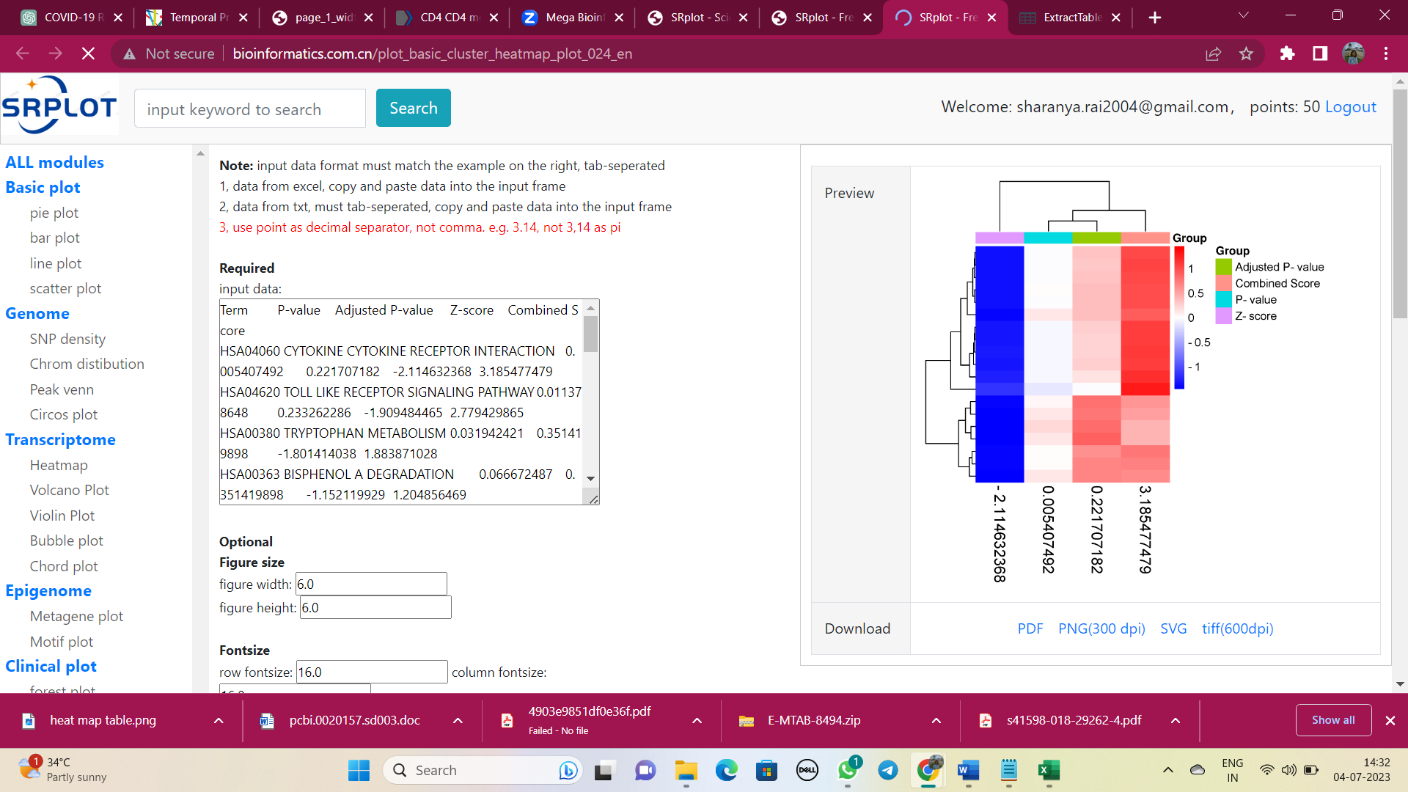
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ligand Name | Ligand ID | Follows Lipinski Rule? | Energy value | Dock Image |
| Zidovudine | 35370 | YES | -5.7 |  |
| Lamivudine | 60825 | YES | -5.9 |  |
| Emtricitabine | 60877 | YES | -5.1 |  |

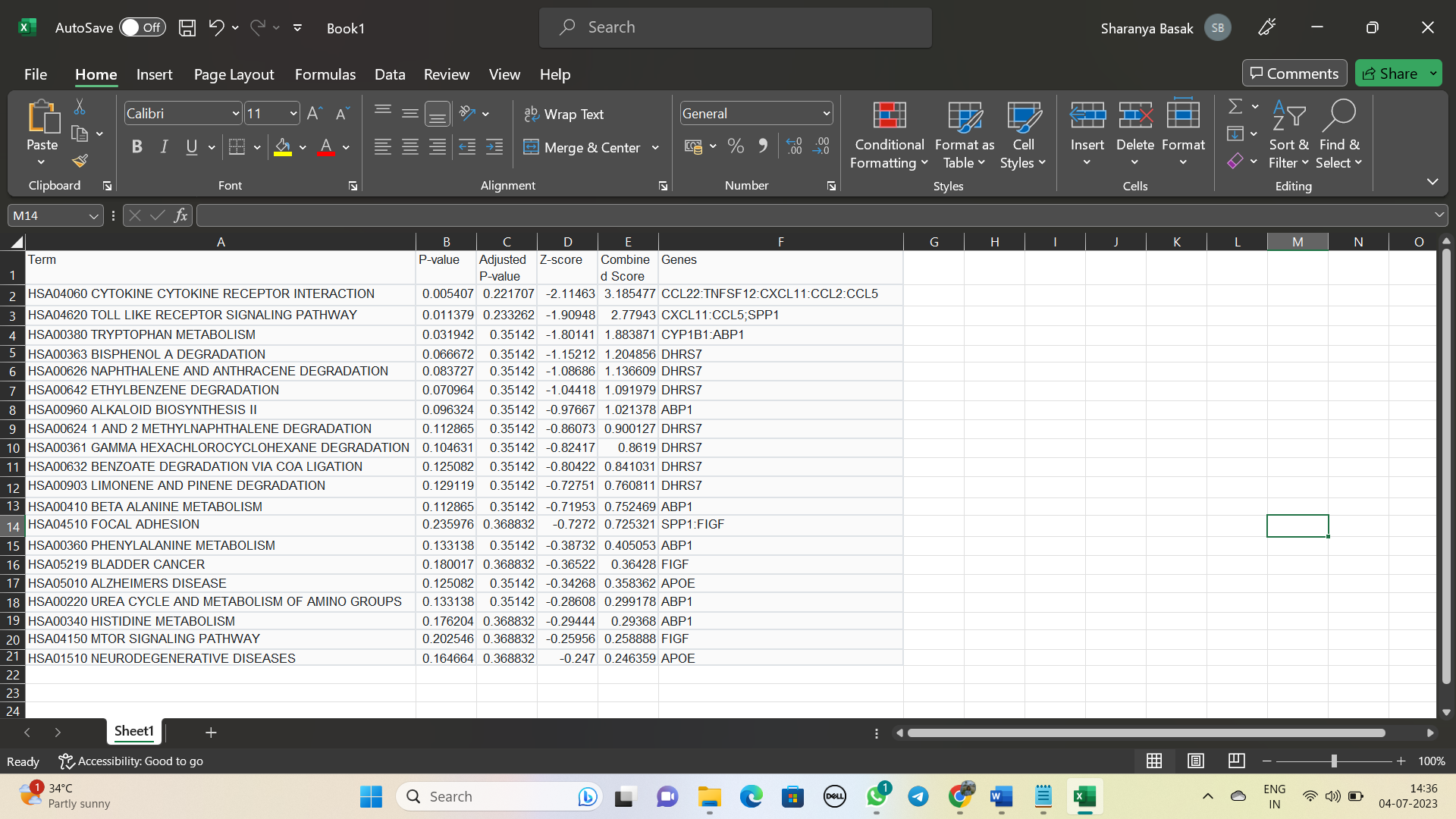
**Day 7: Heat Map Generation:**

**Heat map generation using SR PLOT, image attached as follows:**

This heatmap is generated based on the Temporal Profile of the Renal Transcriptome of HIV-1 Transgenic Mice during Disease Progression. The 69 genes, which were up-regulated at three time points, and were analysed by using the Enrichr KEGG program.

Reference: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0093019#pone-0093019-g011>





**Day 8 & 9: Homology Modelling:**

You can choose any protein which is involved in SARS CoV-2 Pathogenesis (E.g.: ACE2 receptor, any envelope protein) and can take at least 2 homologous sequences with sequence similarity >30%. Try to develop an hypothesis around it (Like Why you want to use Homology modelling for your protein of interest, Purpose and outcome of it) and more importantly how it is going to add value to your hypothesis.

**Problem statement: Develop a precise homology modelling approach to visualize and predict the 3D structure of the HIV-1 Gag- Pol polyprotein expressed by the gag pol gene and to investigate the therapeutic actions of Drug X on it. Using the Swiss Model helped us understand and visualize this structure through models and templates based on the isoforms of the selected protein and, thereby, made predictions easier for the project.**

**Protein:** Gag-Pol Polyprotein (HIV-1 infection in *Homo sapiens*)

**Gene:** gag pol

**PDB:** P04585-1 (First Isoform)

|  |
| --- |
| **Target** |
|  |
| **Sequence** |
|  |
|  |
| **Result** |
|  |

**Day 10:**

Please paste your GitHub account link: