Data processing procedures

1. Download input files from GISAID
   - Download the nucleotide sequences in FASTA format (e.g., FASTA.fasta).
   - Download the "Dates and Location" metadata as TSV (e.g., metadata.tsv).

2. Run Nextclade to generate a TSV file
   - Run this in Windows PowerShell:
     mkdir out
     .¥nextclade-x86_64-pc-windows-gnu.exe run -D .¥data FASTA.fasta --output-tsv .¥out¥nextclade.tsv

3. Generate mutation lists for a given date range (>=50% frequency)
   - Run:
     python common_mutations_by_date.py --nextclade .¥out¥nextclade.tsv --meta metadata.tsv --start YYYY-MM-DD --end YYYY-MM-DD
     This outputs:
     - common_nucl_*.csv (nucleotide substitutions)
     - common_spikeaa_*.csv (Spike amino-acid substitutions)

4. Compare two mutation lists (A vs B)
   - Run:
     python compare_mut_lists.py --A A.csv --B B.csv
     (A.csv and B.csv should be mutation list files generated in Step 3, e.g., nucleotide lists or Spike AA lists.)

5. Run the python program available from the GitHub to generate a comparison list.
   - If you want to count substitution types (e.g., C→U, G→A) to make a mutation spectrum, run the script on a nucleotide list file (e.g., compare_A_B.csv). For example:
     python count_substitution_types_A_to_B.py --compare compare_A_B.csv