# Objective
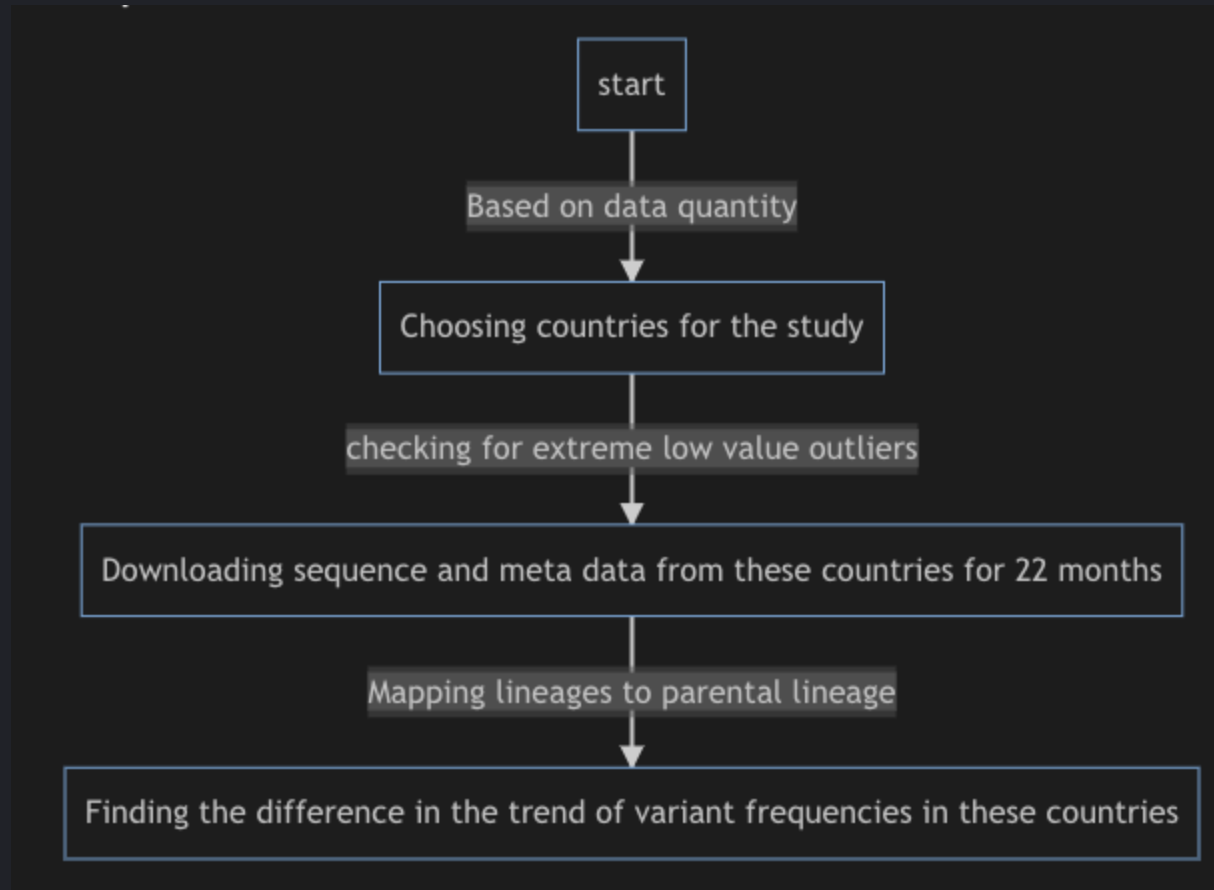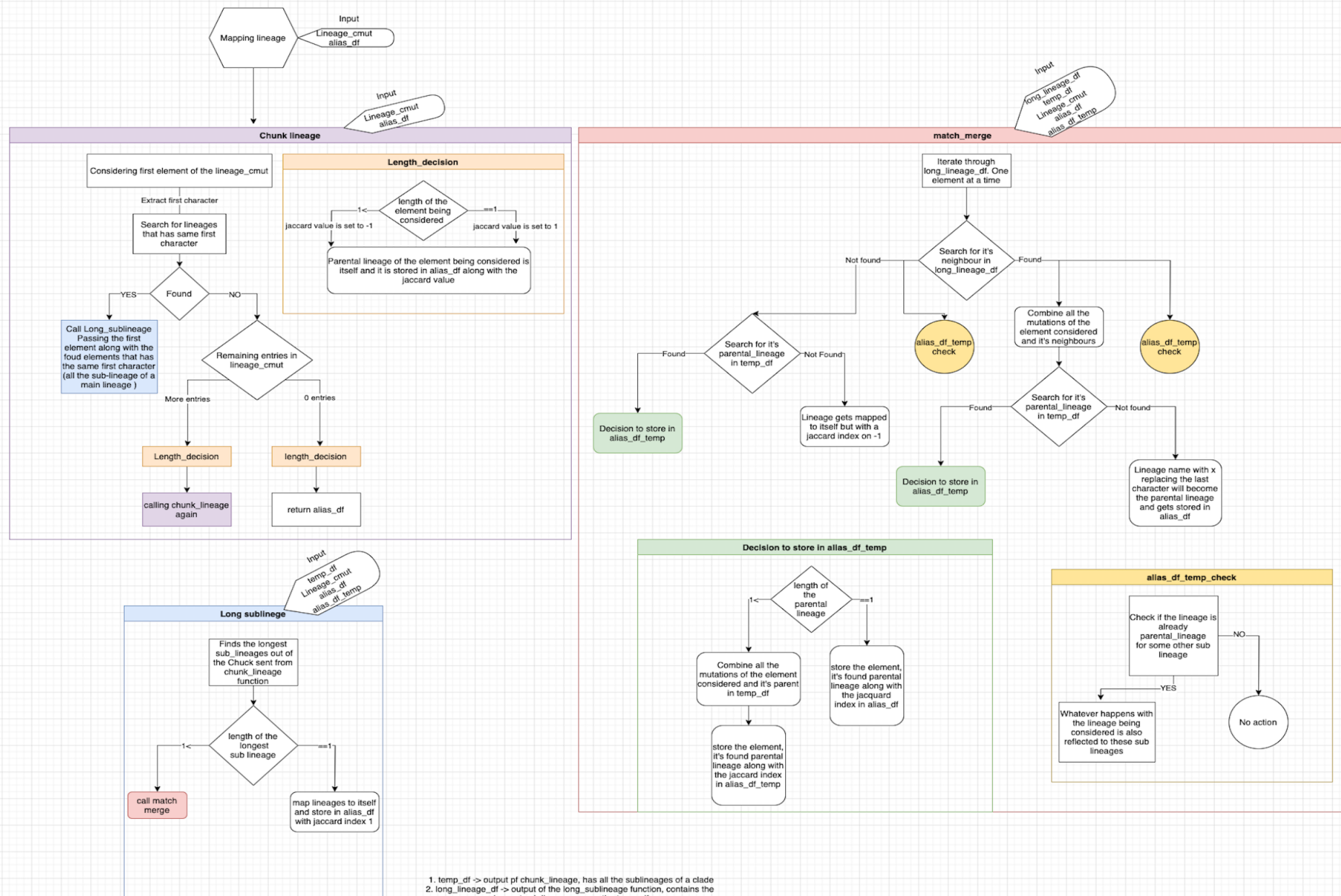
Studying the influence of the population genetics in the SARS CoV2 variant susceptibility

# Work done in steps

# Mapping Lineages

Mapping lineage

Input
Lineage_cmut
alias_df

**Chunk lineage**

Input
Lineage_cmut
alias_df

Considering first element of the lineage_cmut

Extract first character

Search for lineages that has same first character

Found
- YES
- NO

**Length_decision**

length of the element being considered
- 1< → jaccard value is set to -1
- ==1 → jaccard value is set to 1

Parental lineage of the element being considered is itself and it is stored in alias_df along with the jaccard value

Call Long_sublineage Passing the first element along with the foud elements that has the same first character (all the sub-lineage of a main lineage )

Remaining entries in lineage_cmut
- More entries
- 0 entries

Length_decision

length_decision

calling chunk_lineage again

return alias_df

**match_merge**

Input
long_lineage_df
temp_df
Lineage_cmut
alias_df
alias_df_temp

Iterate through long_lineage_df. One element at a time

Search for it's neighbour in long_lineage_df
- Not found
- Found

Combine all the mutations of the element considered and it's neighbours

alias_df_temp check

alias_df_temp check

Search for it's parental_lineage in temp_df
- Found
- Not Found

Decision to store in alias_df_temp

Lineage gets mapped to itself but with a jaccard index on -1

Search for it's parental_lineage in temp_df
- Found
- Not found

Decision to store in alias_df_temp

Lineage name with x replacing the last character will become the parental lineage and gets stored in alias_df

**Long sublineage**

Input
temp_df
Lineage_cmut
alias_df
alias_df_temp

Finds the longest sub_lineages out of the Chuck sent from chunk_lineage function

length of the longest sub lineage
- 1<
- ==1

call match merge

map lineages to itself and store in alias_df with jaccard index 1

**Decision to store in alias_df_temp**

length of the parental lineage
- 1<
- ==1

Combine all the mutations of the element considered and it's parent in temp_df

store the element, it's found parental lineage along with the jacquard index in alias_df

store the element, it's found parental lineage along with the jaccard index in alias_df_temp

**alias_df_temp_check**

Check if the lineage is already parental_lineage for some other sub lineage
- NO
- YES

Whatever happens with the lineage being considered is also reflected to these sub lineages

No action

1. temp_df -> output pf chunk_lineage, has all the sublineages of a clade
2. long_lineage_df -> output of the long_sublineage function, contains the longest sub lineage among the temp_df
3. alias_df_temp -> Temporary location for the lineages that has the scope of being matched to super parental lineage

3

# Inference

Interesting inference