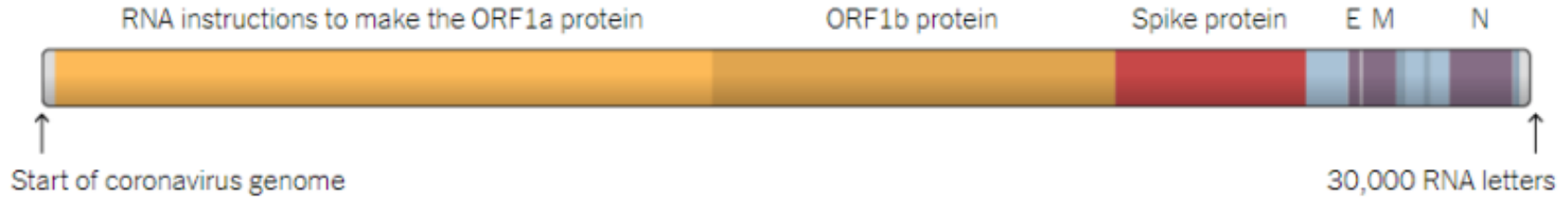


Analysis of SARS-CoV-2 Genomic Diversity

Jordyn Brooks
and Abigail Thornton

SARS-CoV-2 Genome



- RNA virus
- Linear genome with approx. 30,000 nucleotides
- 11 coding-regions (genes)
- 12 Potential Gene products



Why look at genomic diversity?

Genomic Diversity

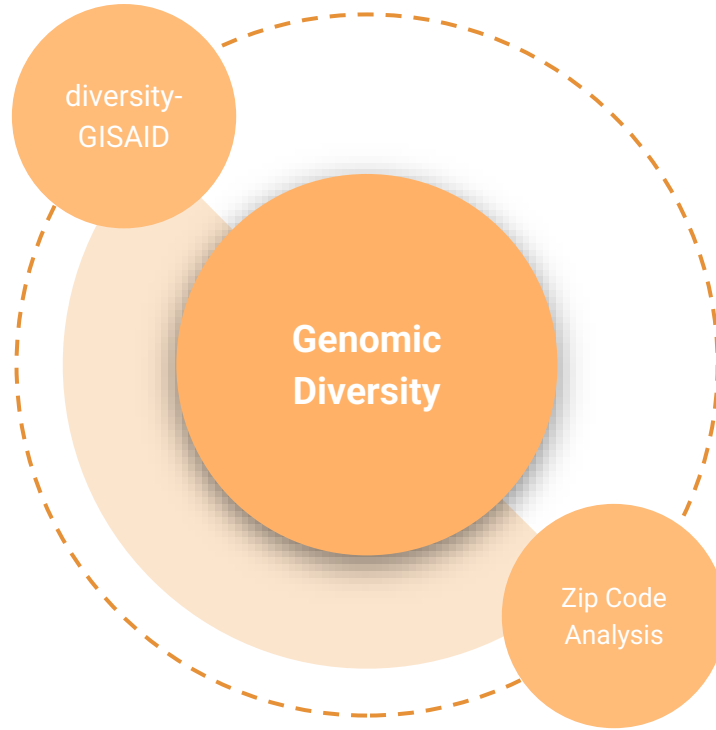
Genetic variation in entire genome of an organism

Nucleotide Diversity

Estimation of variation in DNA sequence at nucleotide level

- Gives insight to history, adaptation, and genetic structure of populations
- Emergence and spread of new variants
 - Treatment/vaccines
 - Public health policy/response efforts
- Genomic diversity of SARS-CoV-2 increasing over time (rapid increase Jan 2021 to Dec 2022)
 - Accumulation of mutations over time

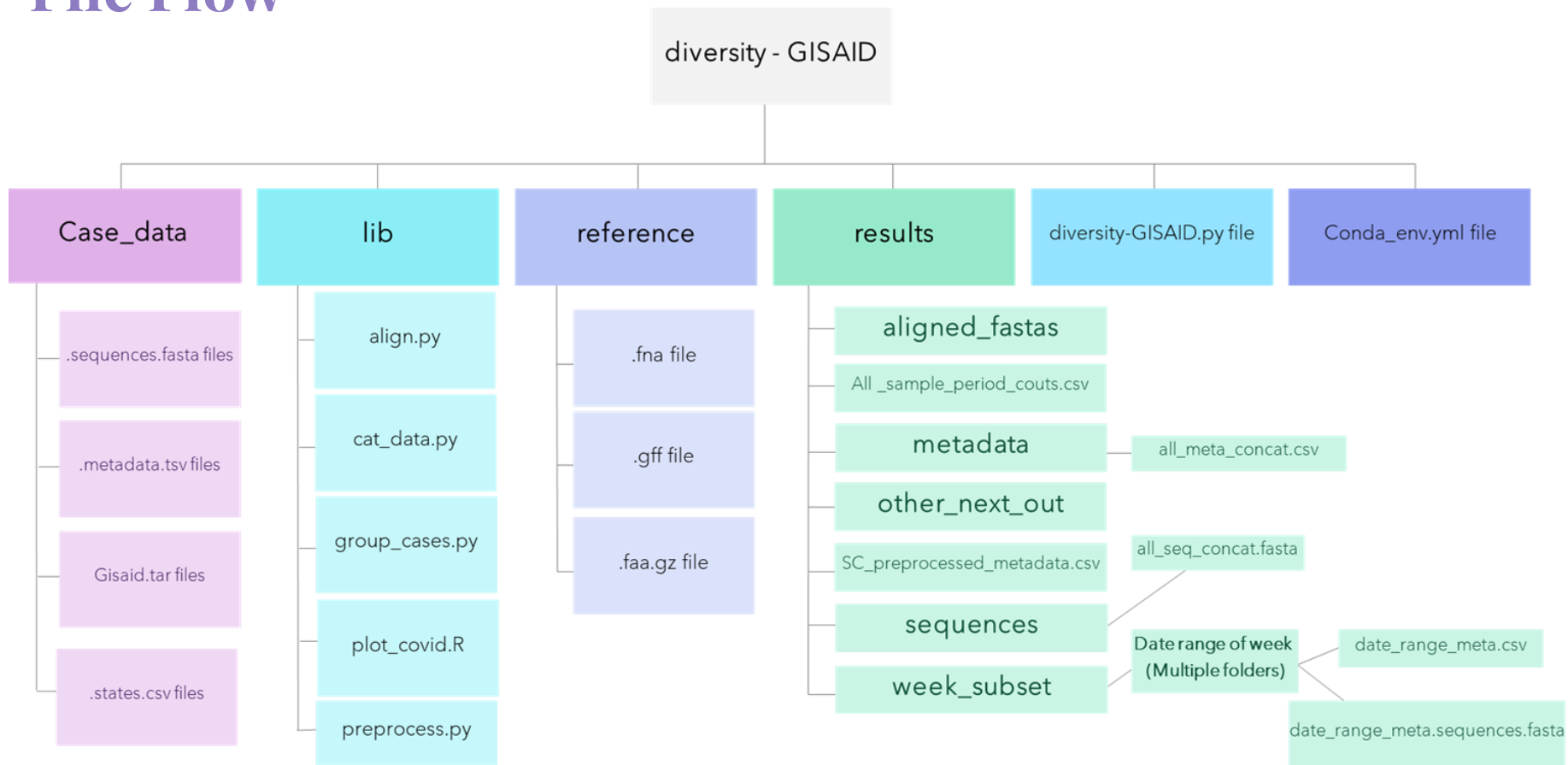
Two parts of the project



Part 1: diversity-GISAID

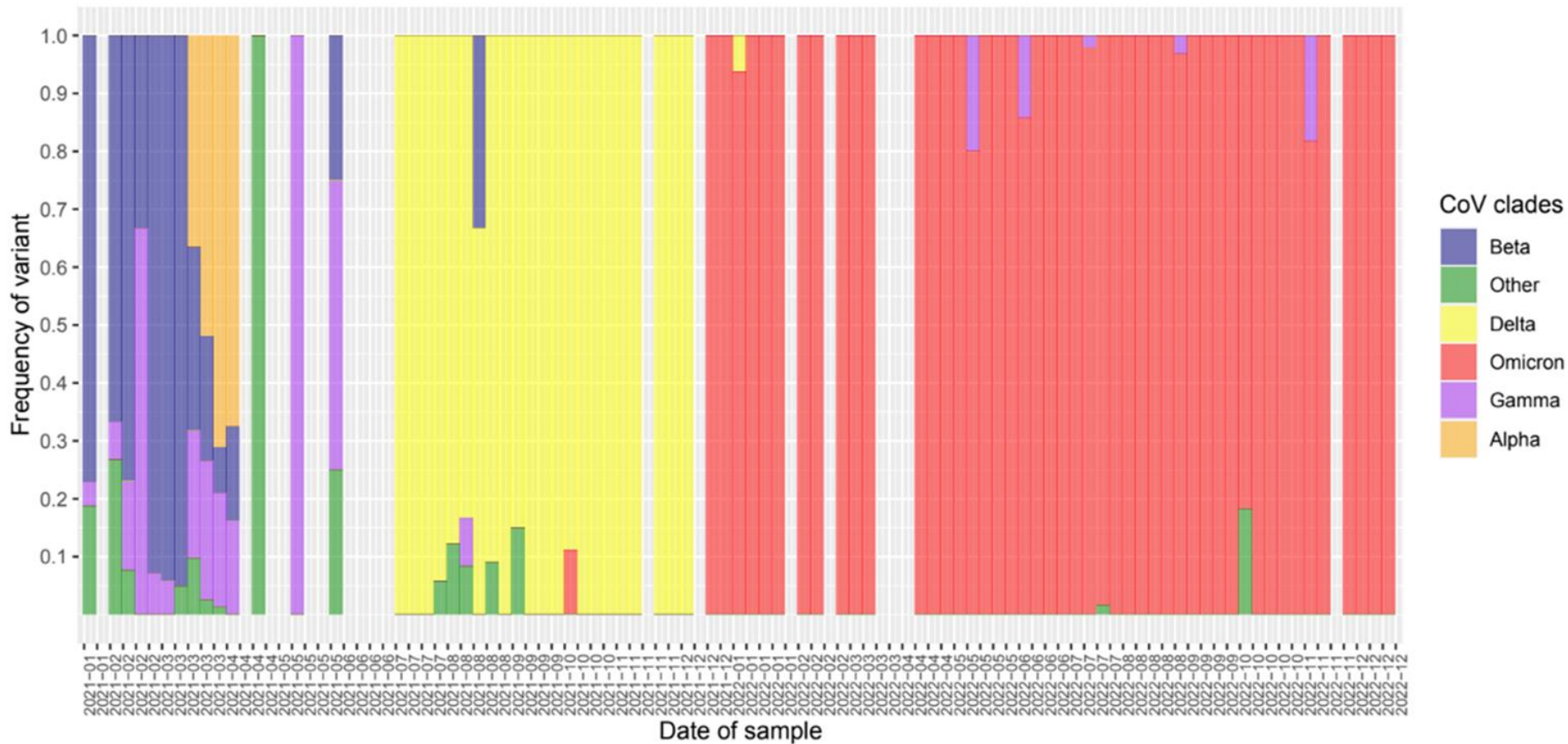
- Finds the variant frequencies per week from 2021-2022
- Finds the total number of cases of each variant per week
- Data was from:
 - The available data set on SARS-CoV-2 positivity
 - Generated through the CLIA certified clinical lab that resides in REDDI
 - The whole genome sequence data generated from randomly selected positive saliva samples

File Flow

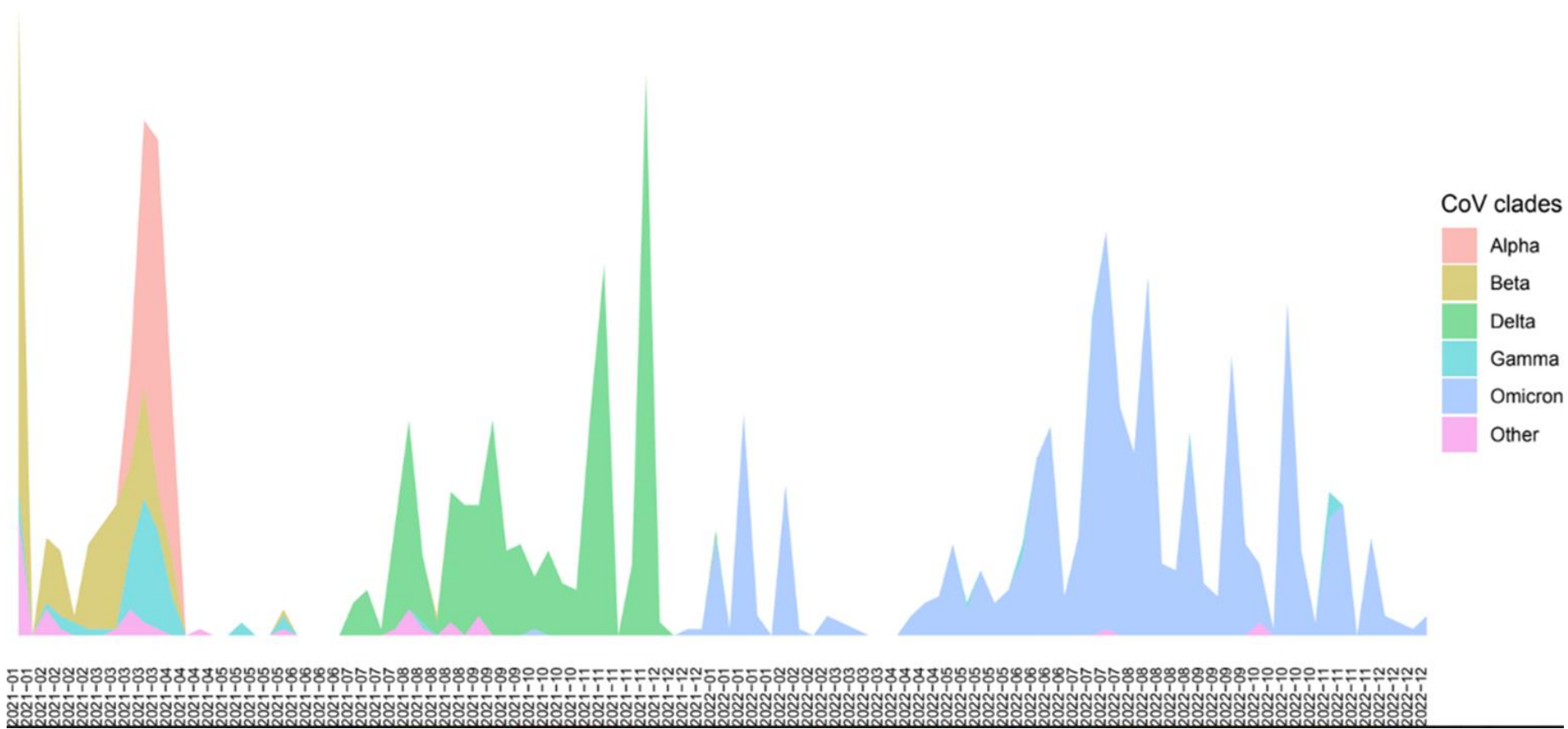


Output

Covid variant replacement through time in SC



Covid variant replacement through time in SC

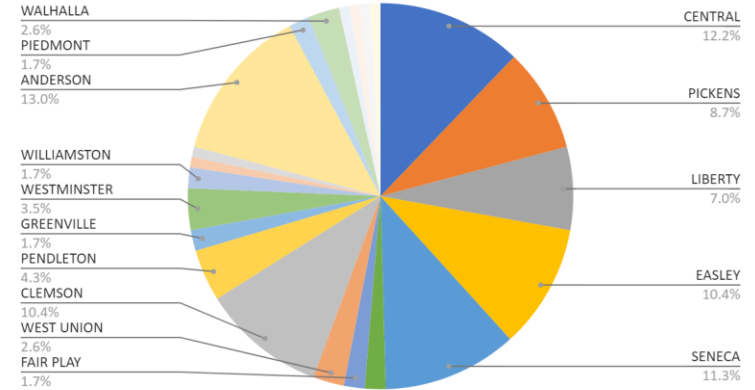


Part 2: Zip Code Analysis

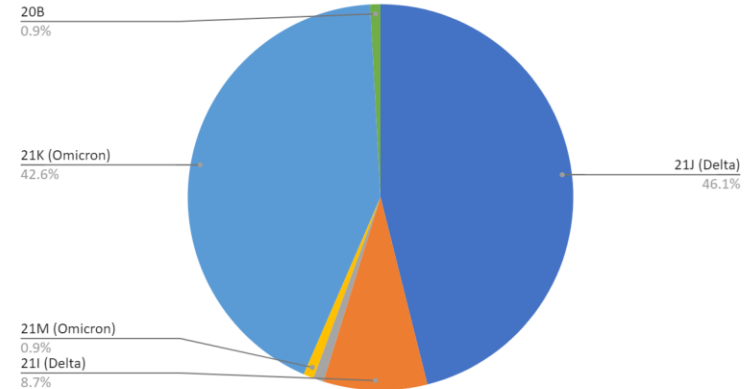
Goal: Generate visual representation of mutations and SNPs found between the Wuhan reference genome and collected COVID samples

- Four counties represented from upstate SC
 - Anderson, Greenville, Pickens, and Oconee
- Looking at difference between mutations/SNPs in more populated vs rural zip codes

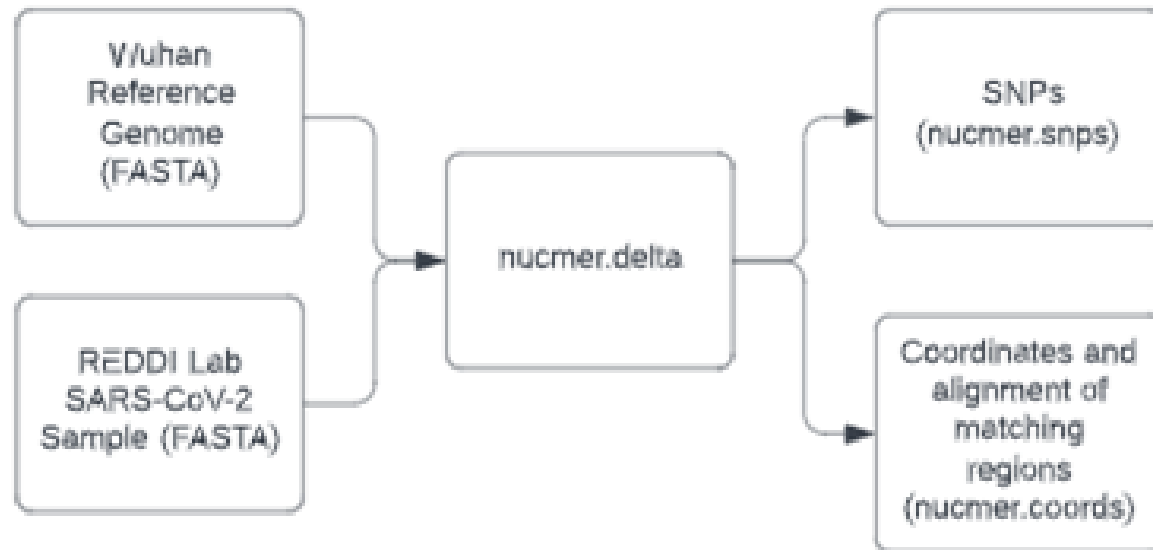
Location of Collected Samples



Clade Types within Dataset



MUMmer File Flow



SNPs

NUCMER												
[P1]	[SUB]	[SUB]	[P2]	[BUFF]	[DIST]	[R]	[Q]	[LEN R]	[LEN Q]	[FRM]	[TAGS]	
3037	C	T	3037	147	3037	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
4181	G	T	4181	68	4181	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
7926	C	T	7926	48	7926	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
8986	C	T	8986	67	8986	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
9053	G	T	9053	67	9053	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
10396	G	T	10396	42	10396	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
11201	A	G	11201	131	11201	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
11332	A	G	11332	131	11332	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
12111	G	A	12111	124	12111	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
14014	T	G	14014	389	14014	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
14408	C	T	14408	172	14408	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
15451	G	A	15451	258	14452	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
23403	A	G	23403	201	6500	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
23604	C	G	23604	201	6299	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
25469	C	T	25469	178	4434	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
27874	C	T	27874	66	2029	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
28253	C	A	28253	20	1650	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
28273	A	.	28272	20	1631	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												
28461	A	G	28460	188	1443	0	0	29903	29902	1	1	NC_045512.2
117M18E37DBC9AD1KM												

COORDs

/scratch1/jkbrook/BIOE_4510/SNP_analysis/NC_045512.fa /scratch1/jkbrook/BIOE_4510/SNP_analysis/T1/117M18E37DBC9AD1KM.consensus.fa
NUCMER

[S1]	[E1]	[S2]	[E2]	[LEN 1]	[LEN 2]	[% IDY]	[LEN R]	[LEN Q]	[COV R]	[COV Q]	[TAGS]	
665	1312	665	1312	648	648	100.00	29903	29902	2.17	2.17	NC_045512.2	117M18E37DBC9AD1KM
2206	2568	2206	2568	363	363	100.00	29903	29902	1.21	1.21	NC_045512.2	117M18E37DBC9AD1KM
2851	3183	2851	3183	333	333	99.70	29903	29902	1.11	1.11	NC_045512.2	117M18E37DBC9AD1KM
4096	4248	4096	4248	153	153	96.08	29903	29902	0.51	0.51	NC_045512.2	117M18E37DBC9AD1KM
5294	5537	5294	5537	244	244	75.82	29903	29902	0.82	0.82	NC_045512.2	117M18E37DBC9AD1KM
6751	7043	6751	7043	293	293	60.41	29903	29902	0.98	0.98	NC_045512.2	117M18E37DBC9AD1KM
7662	7973	7662	7973	312	312	69.87	29903	29902	1.04	1.04	NC_045512.2	117M18E37DBC9AD1KM
8914	9245	8914	9245	332	332	99.40	29903	29902	1.11	1.11	NC_045512.2	117M18E37DBC9AD1KM
9503	9834	9503	9834	332	332	93.98	29903	29902	1.11	1.11	NC_045512.2	117M18E37DBC9AD1KM
10100	10437	10100	10437	338	338	99.70	29903	29902	1.13	1.13	NC_045512.2	117M18E37DBC9AD1KM
10953	12234	10953	12234	1282	1282	96.10	29903	29902	4.29	4.29	NC_045512.2	117M18E37DBC9AD1KM
12453	12779	12453	12779	327	327	83.49	29903	29902	1.09	1.09	NC_045512.2	117M18E37DBC9AD1KM
12902	13363	12902	13363	462	462	83.33	29903	29902	1.54	1.55	NC_045512.2	117M18E37DBC9AD1KM
13626	14579	13626	14579	954	954	99.79	29903	29902	3.19	3.19	NC_045512.2	117M18E37DBC9AD1KM
15194	16183	15194	16183	990	990	99.90	29903	29902	3.31	3.31	NC_045512.2	117M18E37DBC9AD1KM
16771	17130	16771	17130	360	360	100.00	29903	29902	1.20	1.20	NC_045512.2	117M18E37DBC9AD1KM
17407	17738	17407	17738	332	332	100.00	29903	29902	1.11	1.11	NC_045512.2	117M18E37DBC9AD1KM
17994	18957	17994	18957	964	964	100.00	29903	29902	3.22	3.22	NC_045512.2	117M18E37DBC9AD1KM
22543	22877	22543	22877	335	335	100.00	29903	29902	1.12	1.12	NC_045512.2	117M18E37DBC9AD1KM
23145	23822	23145	23822	678	678	99.71	29903	29902	2.27	2.27	NC_045512.2	117M18E37DBC9AD1KM
24417	24765	24417	24765	349	349	91.98	29903	29902	1.17	1.17	NC_045512.2	117M18E37DBC9AD1KM
25004	25646	25004	25646	643	643	99.84	29903	29902	2.15	2.15	NC_045512.2	117M18E37DBC9AD1KM
26220	26566	26220	26566	347	347	100.00	29903	29902	1.16	1.16	NC_045512.2	117M18E37DBC9AD1KM
27809	28756	27809	28756	948	947	99.58	29903	29902	3.17	3.17	NC_045512.2	117M18E37DBC9AD1KM

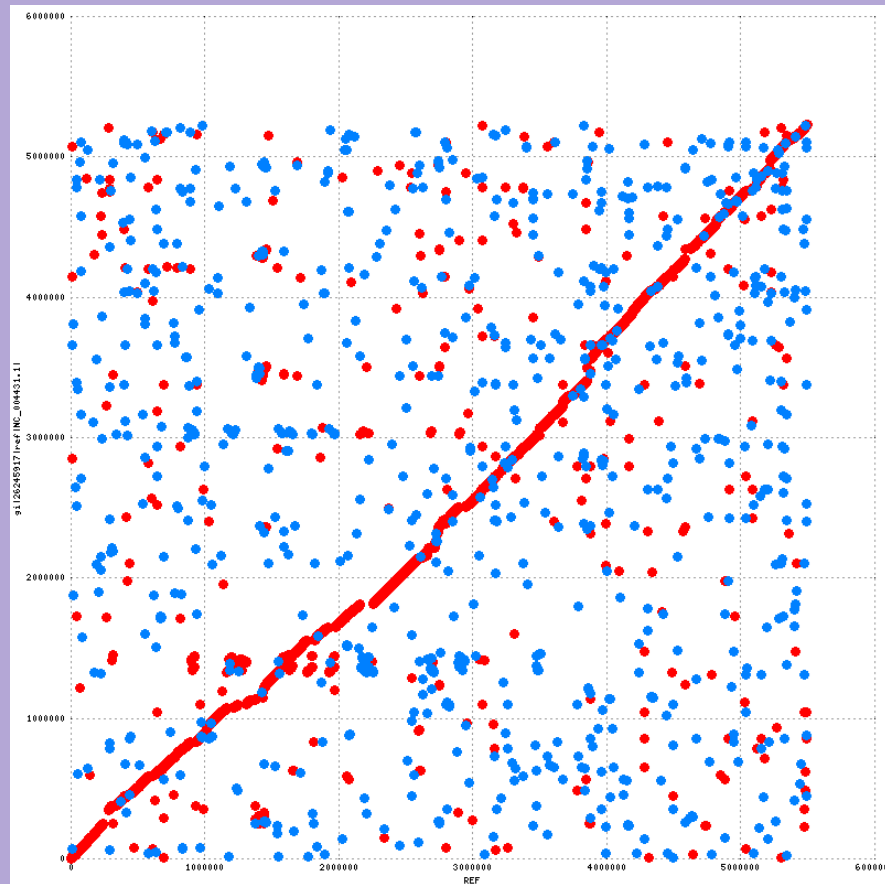
Example Mummerplot Output

X-axis = reference genome

Y-axis = query-genome

Red points/lines indicate an
alignment in forward direction

Blue points/lines indicate
alignment in reverse direction



References and Acknowledgements

Arthur L. Delcher, Adam Phillippy, Jane Carlton, Steven L. Salzberg, Fast algorithms for large-scale genome alignment and comparison, *Nucleic Acids Research*, Volume 30, Issue 11, 1 June 2002, Pages 2478–2483, <https://doi.org/10.1093/nar/30.11.2478>

Liphardt. 2022. diversity-GISAID. <https://github.com/liphardt/diversity-GISAID.git>. (2023)

Mercatelli, D., & Giorgi, F. M. (2020). Geographic and genomic distribution of SARS-COV-2 mutations. <https://doi.org/10.20944/preprints202004.0529.v1>

Pritchard, L. (n.d.). *Whole Genome Alignments*. GitHub. Retrieved April 18, 2023, from https://github.com/widowquinn/Teaching/blob/master/Comparative_Genomics_and_Visualisation/Part_1/whole_genome_alignment/whole_genome_alignments_A_UoD.md

We acknowledge Clemson University Creative Inquiry, and the assistance of Dr. Rooksana Noorai of the Clemson University Genomics and Bioinformatics Facility for services and facilities provided. The facility is supported by Grants P20GM109094 and P20GM139767 Institutional Development Awards (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health .

