

The Present and Predicted Evolution of COVID-19

Appendix

April 6, 2020

Part 1b

1.1 $L \rightarrow Y$

- $L \in \{UUG, UUA, CUU, CUC, CUA, CUG\}$
- $Y \in \{UAC, UAU\}$

Assume probability that this amino acid (out of the 5 critical residues) is mutated is $P(\text{mutate}(L)) = \frac{3}{15} = \frac{1}{5}$. All probabilities could be multiplied by $P(\text{mutate}(L))$ to get the conditional probability. This will be ignored because we only care about the total strand mutating back to SARS-CoV, and any mutation will effect one of the five critical residues. Mutations are independent events.

Cases

1. $L = UUG$

- Mutation 1. $P(UUG \rightarrow UAG) = \frac{1}{9}$
- Mutation 2. $P(UAG \rightarrow (UAC \vee UAU)) = \frac{2}{9}$
- 2 mutations, $P(L \rightarrow Y) \approx .025$

2. $L = UUA$

- Mutation 1. $P(UUA \rightarrow UAA) = \frac{1}{9}$
- Mutation 2. $P(UAA \rightarrow (UAC \vee UAU)) = \frac{2}{9}$
- 2 mutations, $P(L \rightarrow Y) \approx .025$

3. $L = CUU$

- Mutation 1. $P(CUU \rightarrow UUU) = \frac{1}{9}$
- Mutation 2. $P(UUU \rightarrow UAU) = \frac{1}{9}$

- 2 mutations, $P(L \rightarrow Y) \approx .012$

4. $L = CUC$

- Mutation 1. $P(CUC \rightarrow UUC) = \frac{1}{9}$
- Mutation 2. $P(UUC \rightarrow UAC) = \frac{1}{9}$
- 2 mutations, $P(L \rightarrow Y) \approx .012$

5. $L = CUA$

- Mutation 1. $P(CUA \rightarrow UUA) = \frac{1}{9}$
- Mutation 2. $P(UUA \rightarrow UAA) = \frac{1}{9}$
- Mutation 3. $P(UAA \rightarrow (UAC \vee UAU)) = \frac{2}{9}$
- 3 mutations, $P(L \rightarrow Y) \approx .003$

6. $L = CUG$

- Mutation 1. $P(CUG \rightarrow UUG) = \frac{1}{9}$
- Mutation 2. $P(UUG \rightarrow UAG) = \frac{1}{9}$
- Mutation 3. $P(UAG \rightarrow (UAC \vee UAU)) = \frac{2}{9}$
- 3 mutations, $P(L \rightarrow Y) \approx .003$

Average number of mutations: **2.33 mutations**

Average probability = **1.3%**

1.2 $F \rightarrow L$

- $F \in \{UUC, UUU\}$
- $L \in \{UUG, UUA, CUU, CUC, CUA, CUG\}$

Assume probability that this amino acid (out of the 5 critical residues) is mutated is $P(\text{mutate}(F)) = \frac{3}{15} = \frac{1}{5}$. Mutations are independent events.

Cases

1. $F = UUC$

- Mutation 1. $P(UUC \rightarrow (UUG \vee UUA \vee CUC)) = \frac{1}{3}$
- 1 mutation, $P(F \rightarrow L) \approx .33$

2. $F = UUU$

- Mutation 1. $P(UUU \rightarrow (UUG \vee UUA \vee CUU)) = \frac{1}{3}$

- 1 mutation, $P(F \rightarrow L) \approx .33$

Average number of mutations: **1 mutation**

Average probability = **33%**

1.3 $Q \rightarrow N$

- $Q \in \{CAA, CAG\}$
- $N \in \{AAC, AAU\}$

Assume probability that this amino acid (out of the 5 critical residues) is mutated is $P(\text{mutate}(Q)) = \frac{3}{15} = \frac{1}{5}$. Mutations are independent events.

Cases

1. $Q = CAA$

- Mutation 1. $P(CAA \rightarrow AAA) = \frac{1}{9}$
- Mutation 2. $P(AAA \rightarrow (AAC \vee AAU)) = \frac{2}{9}$
- 2 mutations, $P(Q \rightarrow N) \approx .025$

2. $Q = CAG$

- Mutation 1. $P(CAG \rightarrow AAG) = \frac{1}{9}$
- Mutation 2. $P(AAG \rightarrow (AAC \vee AAU)) = \frac{2}{9}$
- 2 mutations, $P(Q \rightarrow N) \approx .025$

Average number of mutations: **2 mutations**

Average probability = **2.5%**

1.4 $S \rightarrow D$

- $S \in \{AGU, AGC, UCG, UCA, UCC, UCU\}$
- $D \in \{GAU, GAC\}$

Assume probability that this amino acid (out of the 5 critical residues) is mutated is $P(\text{mutate}(S)) = \frac{3}{15} = \frac{1}{5}$. Mutations are independent events.

Cases

1. $S = AGU$

- Mutation 1. $P(AGU \rightarrow GGU) = \frac{1}{9}$
- Mutation 2. $P(GGU \rightarrow GAU) = \frac{1}{9}$

- 2 mutations, $P(S \rightarrow D) \approx .012$

2. $S = AGC$

- Mutation 1. $P(AGC \rightarrow GGC) = \frac{1}{9}$
- Mutation 2. $P(GGC \rightarrow GAC) = \frac{1}{9}$
- 2 mutations, $P(S \rightarrow D) \approx .012$

3. $S = UCG$

- Mutation 1. $P(UCG \rightarrow GCG) = \frac{1}{9}$
- Mutation 2. $P(GCG \rightarrow GAG) = \frac{1}{9}$
- Mutation 3. $P(GAG \rightarrow (GAU \vee GAC)) = \frac{2}{9}$
- 3 mutations, $P(S \rightarrow D) \approx .003$

4. $S = UCA$

- Mutation 1. $P(UCA \rightarrow GCA) = \frac{1}{9}$
- Mutation 2. $P(GCA \rightarrow GAA) = \frac{1}{9}$
- Mutation 3. $P(GAA \rightarrow (GAU \vee GAC)) = \frac{2}{9}$
- 3 mutations, $P(S \rightarrow D) \approx .003$

5. $S = UCC$

- Mutation 1. $P(UCC \rightarrow GCC) = \frac{1}{9}$
- Mutation 2. $P(GCC \rightarrow GAC) = \frac{1}{9}$
- 2 mutations, $P(S \rightarrow D) \approx .012$

6. $S = UCU$

- Mutation 1. $P(UCU \rightarrow GCU) = \frac{1}{9}$
- Mutation 2. $P(GCU \rightarrow GAU) = \frac{1}{9}$
- 2 mutations, $P(S \rightarrow D) \approx .012$

Average number of mutations: **2.33 mutations**

Average probability = **0.9%**

1.5 $N \rightarrow T$

- $N \in \{AAC, AAU\}$
- $T \in \{ACG, ACA, ACC, ACU\}$

Assume probability that this amino acid (out of the 5 critical residues) is mutated is $P(\text{mutate}(N)) = \frac{3}{15} = \frac{1}{5}$. Mutations are independent events.

Cases

1. $N = AAC$

- Mutation 1. $P(AAC \rightarrow ACC) = \frac{1}{9}$
- 1 mutations, $P(N \rightarrow T) \approx .11$

2. $N = AAU$

- Mutation 1. $P(AAU \rightarrow ACU) = \frac{1}{9}$
- 1 mutations, $P(N \rightarrow T) \approx .11$

Average number of mutations: **1 mutations**

Average probability = **11%**

1.6 Combined Results

The probability of all of these events happening is given by the product of all probabilities.

$$P(L \rightarrow Y \wedge F \rightarrow L \wedge Q \rightarrow N \wedge S \rightarrow D \wedge N \rightarrow T) = P(L \rightarrow Y)P(F \rightarrow L)P(Q \rightarrow N)P(S \rightarrow D)P(N \rightarrow T)$$

$$P(L \rightarrow Y \wedge F \rightarrow L \wedge Q \rightarrow N \wedge S \rightarrow D \wedge N \rightarrow T) \approx (0.013)(0.33)(0.025)(0.009)(0.11)$$

$$P(L \rightarrow Y \wedge F \rightarrow L \wedge Q \rightarrow N \wedge S \rightarrow D \wedge N \rightarrow T) = 1.06 \times 10^{-7}$$

This probability is assuming shortest path from first amino acid to target amino acid and would take approximately 8.66 mutations to happen.

Part 1d

Distance from COVID-19 (b = 10%)		
Bat-SARS-CoV	Civet-SARS-CoV	Human-SARS-CoV
18	23	30
11	22	31
12	23	20
16	18	21
22	18	22
19	21	23
15	16	22
22	31	24
19	24	24
21	24	34
17	25	20
28	19	25
16	23	26
18	24	24
22	21	26
14	16	35
28	20	22
19	24	24
18	22	23
17	26	40
17	26	29
24	31	22
28	33	20
18	24	26
17	20	24
19	37	29
16	23	26
23	29	13
21	23	29
16	31	31
16	19	19
12	20	24
29	23	27
25	16	28
20	23	26
31	23	24
11	24	24
29	17	28
18	28	16
15	18	23
Mean of 40 iterations:		
19	23	25
Standard Deviation		
5	5	5

Figure 1: Average number of mutations necessary during simulations from SARS-CoV-2 to 3 strains of original SARS-CoV. $b = .10$ and 40 simulations of each.

Distance from COVID-19 (b = 15%)		
Bat-SARS-CoV	Civet-SARS-CoV	Human-SARS-CoV
19	15	20
10	19	26
12	21	21
10	18	30
16	23	40
25	23	20
17	31	30
23	24	18
18	22	26
21	14	24
23	26	36
20	27	21
25	36	25
19	13	27
16	25	16
22	18	19
14	26	25
14	19	24
28	14	23
21	16	21
22	23	19
22	21	26
17	18	19
25	26	25
17	20	16
16	16	27
16	26	22
17	16	28
20	18	23
16	19	17
15	22	21
17	15	23
14	22	17
16	20	24
12	28	14
16	20	17
13	21	22
17	18	31
18	17	17
21	20	21
Mean of 40 iterations:		
18	21	23
Standard Deviation:		
4	5	5

Figure 2: Average number of mutations necessary during simulations from SARS-CoV-2 to 3 strains of original SARS-CoV. $b = .15$ and 40 simulations of each.