

Observing Variations of Differences on COVID-19 in Different Regions Extracting Type and Mutation Information

Jianzhong Liu^{1,2,3} and Jeffrey Zheng^{1,2,3,4*}

¹Key Laboratory of Quantum Information of Yunnan, China

²Key Laboratory of Software Engineering of Yunnan, China

³Engineering Research Center of Cyberspace of Yunnan, China

⁴Yunnan University, Kunming, China

***Corresponding Author:** Jeffrey Zheng, Key Laboratory for Software Engineering of Yunnan Province (Yunnan University), Kunming, China.

Received: August 18, 2020; **Published:** January 25, 2021

Abstract

Covid-19 genomes were collected from three regions: Shanghai-China, Tbilisi- Georgia and Sydney-Australia. Five similar genomes were selected from each region for research in this paper. Applying the “Datum gene sequence” method proposed, our results show that variation is immense in the Sydney-Australia region, followed by variation in the Tbilisi-Georgia region, which has a minimal value in the Shanghai-China region.

Keywords: Genome; Base Mutation; Datum Gene Sequence; Variation; Immense

Introduction

In Covid-19 research, the number of mutations and the types of mutations in the gene sequences are of great importance. It can roughly reflect the variation size and evolutionary direction of Covid-19 in a region. Base mutation analysis is likewise a common method in genes [1-7].

Materials and Methods

In Covid-19 gene sequences in the three regions of Shanghai-China, Tbilisi-Georgia and Sydney-Australia [8], we selected five similar gene sequences in each region for research. We suppose that all five gene sequences have evolved from an original gene sequence (in this paper called the Datum gene sequence). For this reason, we first solve “Datum gene sequence” using the filtering method. The method is as follows:

1. Arrange by sequence number, align 5 gene sequences Seq_i ($i = 1, 2, 3, 4, 5$);
2. For sequence number j , the corresponding values of the 5 gene sequences are: $Seq_i[j]$ ($i = 1, 2, 3, 4, 5$);
3. Let the Datum gene sequence $DatumSeq[j]$ be equal to the most abundant value in $Seq_i[j]$ ($i = 1, 2, 3, 4, 5$).

For example: if $Seq_1[j]=Seq_2[j]=Seq_3[j]=T, Seq_4[j]=A, Seq_5[j]=C$, then $Datum-Seq[j]=T$.

Results

Based on the Datum gene sequence, we identified the mutant bases of five gene sequences relative to the Datum gene sequence. The number and type of base mutations were counted, and the work done is illustrated in the chart below.

Citation: Jianzhong Liu and Jeffrey Zheng. “Observing Variations of Differences on COVID-19 in Different Regions Extracting Type and Mutation Information”. EC Neurology SI.02 (2021): 11-14.

The number and type of base mutations in five Covid-19 gene sequences from Shanghai-China region

		Ordinal (0-29836)	4624	8751	9859	15293	20352	20496	24003	26698	28046	28113	28814	29272
China Shanghai	SH0029/2020 EPI_ISL_416338	Base mutation		C->T					C->T	T->C	G->C	T->C		
	SH0030/2020 EPI_ISL_416339	Base mutation	C->T	C->T	G->T			G->T				T->C		
	SH0053/2020 EPI_ISL_416361	Base mutation				C->T	G->T							C->T
	SH0054/2020 EPI_ISL_416362	Base mutation											G->T	
	SH0055/2020 EPI_ISL_416363	Base mutation				C->T								C->T

China Shanghai	mutation: 4	C->T	C->A	T->A	T->C	T->G	G->A	G->T	G->C	A->G	A->T
	total: 16	8	0	0	3	0	0	4	1	0	0

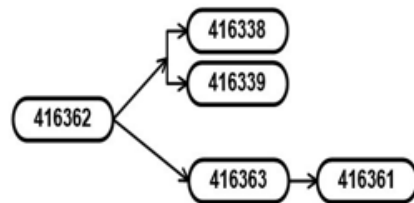


Figure 1: Shanghai-China.

The number and type of base mutations in five Covid-19 gene sequences from Tbilisi-Georgia region

		Ordinal (0-29834)	237	911	1055	1381	3033	8562	8778	9473	11079	14404	14801	15358
Georgia Tbilisi	Tb-390/2020 EPI_ISL_416477	Base mutation	T->C				T->C		C->T	T->A		T->C	C->T	G->A
	Tb-673/2020 EPI_ISL_416478	Base mutation		C->T	C->T			T->C						
	Tb-537/2020 EPI_ISL_416480	Base mutation	T->C	A->G			T->C				G->T	T->C		
	Tb-712/2020 EPI_ISL_416481	Base mutation												
	Tb/2020 EPI_ISL_416482	Base mutation												

20264	23399	23518	24073	25559	25975	26140	28140	28653	28859
	G->A				G->T				
					G->T				
	G->A	G->C				G->T			
A->C									
			G->T						

Georgia Tbilisi	mutation: 7	C->T	C->A	T->A	T->C	T->G	G->A	G->T	G->C	A->G	A->T
	total: 26	6	0	1	8	0	3	5	1	2	0



Figure 2: Tbilisi-Georgia.

The number and type of base mutations in five Covid-19 gene sequences from Sydney-Australia region

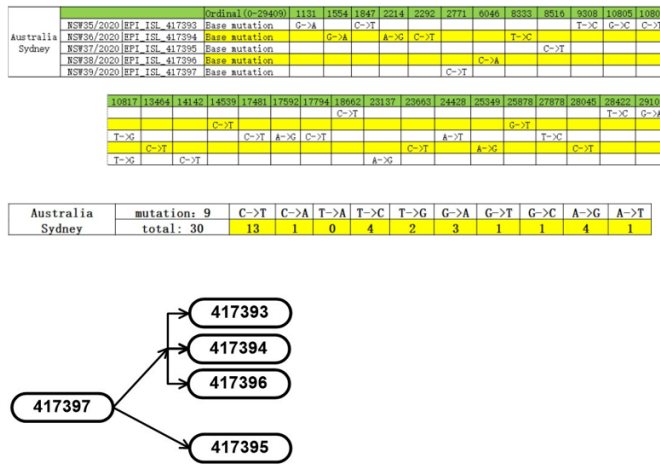


Figure 3: Sydney-Australia.

Discussion

Only from the analysis of our experimental data and experimental results: (1) There are 4 variant base types and 16 mutation bases in the 5 gene sequences of the Shanghai-China region. Therefore, we can say that Covid-19 in the Shanghai-China region is pure, and the direction of Covid19’s evolution is clear (See Shanghai-China Figure). (2) There are 9 variant base types and 30 mutation bases in the 5 gene sequences of the Sydney-Australia region. Therefore, we can say that Covid-19 variation in the Sydney-Australia region is complex, Covid-19 has many evolutionary directions (See Sydney-Australia figure). (3) There are 7 variant base types and 26 mutation bases in the 5 gene sequences of the Tbilisi-Georgia region.

Funding

This work was supported by NSF of China (62041213), the Key Project on Electric Information and Next Generation IT Technology of Yunnan (2018Z1002), NSF of China (61362014), Yunnan Advanced Overseas Scholar Project.

Bibliography

1. Ye XD and Zhang JY. “Study on the single-base mutation in coding sequence of gene based on power spectrum analysis”. *Genomics and Applied Biology* 38.2 (2018): 563-567.
2. ZHI Hao., et al. “Algorithms for Recognizing Genes short coding sequence and genovariation Detection”. *Mathematics in Practice and Theory* 43.14 (2013): 77-84.
3. XIE Ming-jie., et al. “Feature analysis of the base mutations in coding region (BMCs) of human syndrome-related genes”. *Journal of Liaoning Normal University Natural Science Edition* 33.2 (2010): 223-227.
4. Liu Jianzhong. “Theory of Inversion Set and Its Application”. Yunnan Science and Technology Publishing Press (1999).
5. Jeffrey Zheng. “Variant Construction from Theoretical Foundation to Applications”. Springer Nature (2019).
6. Jeffrey Zheng. “Variant Construction Theory and Applications Theoretical Foundation and Applications”. Science Press 1 (2020).

7. Jeffrey Zheng, ResearchGate.
8. 95,000 viral genomic sequences of hCoV-19 shared with unprecedented speed via GISAID.

© All rights reserved by Jianzhong Liu and Jeffrey Zheng.