

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

Jianzhong Liu, Jeffrey Zheng

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, COVID-19, base mutation, datum gene sequence, variation detection

Jianzhong Liu
Key Laboratory for Software Engineering of Yunnan Province, Yunnan University, Kunming China
e-mail: liujianz6655@qq.com

Jeffrey Zheng
Key Laboratory of Quantum Information of Yunnan
Key Laboratory of Software Engineering of Yunnan
School of Software, Yunnan University e-mail: conjugatologic@yahoo.com

This work was supported by the Key Project on Electric Information and Next Generation IT Technology of Yunnan (2018ZI002), NSF of China(61362014), and Yunnan Advanced Overseas Scholar Project.

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 Seq₁[j]=Seq₂[j]=Seq₃[j]=T, Seq₄[j]=A, Seq₅[j]=C, then DatumSeq[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.

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
Shanghai-China	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

(Shanghai-China Table 1)

Shanghai-China	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

(Shanghai-China Table 2)



(Shanghai-China Figure 1)

The Number and Type of Base Mutations in Five Covid-19 Gene Sequences From Tbilisi-Georgia Region.

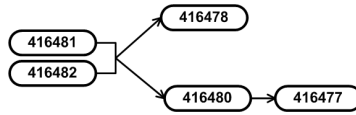
	Ordinal(0-29834)	237	911	1055	1381	3033	8562	8778	9473	11079	14404	14601	15355
Tbilisi-Georgia	Tb-390/2020 EPI_ISL_416477	Base mutation	T->C			T->C		C->T	T->A		T->C	C->T	G->A
	Tb-673/2020 EPI_ISL_416478	Base mutation		C->T	C->T		T->C						
	Tb-537/2020 EPI_ISL_416480	Base mutation	T->C	A->G		T->C				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		T->C	C->T	C->T
		G->A	G->C		G->T				
	A->G					G->T			
			G->T						

(Tbilisi-Georgia Table 1)

Tbilisi-Georgia	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

(Tbilisi-Georgia Table 2)



(Tbilisi-Georgia Figure 1)

The Number and Type of Base Mutations in Five Covid-19 Gene Sequences From Sydney-Australia Region.

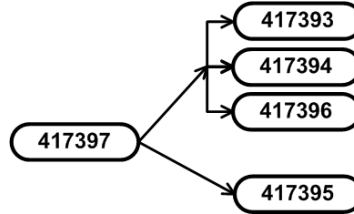
	Ordinal(0-29409)	1131	1554	1847	2214	2292	2771	6046	8333	8516	9308	10805	10808
Sydney-Australia	NSW35/2020 EPI_ISL_417393	Base mutation	G->A		C->T						T->C	G->C	C->T
	NSW36/2020 EPI_ISL_417394	Base mutation		G->A	A->G	C->T			T->C				
	NSW37/2020 EPI_ISL_417395	Base mutation								C->T			
	NSW38/2020 EPI_ISL_417396	Base mutation						C->A					
	NSW39/2020 EPI_ISL_417397	Base mutation					C->T						

10817	13464	14142	14539	17481	17592	17794	18662	23137	23663	24428	25340	25878	27878	28045	28422	29108
							C->T						G->T		T->C	G->A
			C->T													
	T->G			C->T	A->G	C->T				A->T			T->C			
		C->T							C->T		A->G			C->T		
	T->G		C->T					A->G								

(Sydney-Australia Table 1)

Australia	mutation: 9	C->T	C->A	T->A	T->C	T->G	G->A	G->T	G->C	A->G	A->T
Sydney	total: 30	13	1	0	4	2	3	1	1	4	1

(Sydney-Australia Table 2)



(Sydney-Australia Figure 1)

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 Covid-19's evolution is clear (see Shanghai-China Figure 1).

(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 the Covid-19 variation in the Sydney-Australia region is complex, and Covid-19 has many evolutionary directions (see Sydney-Australia Figure 1).

(3). There are 7 variant base types and 26 mutation bases in the 5 gene sequences of the Tbilisi-Georgia region.

Conflict Interest

No conflict of interest has claimed.

References

1. Ye X.D., and Zhang J.Y., 2018, Study on the single-base mutation in coding sequence of gene based on power spectrum analysis, (Genomics and Applied Biology), 38(2): 563-567.
2. ZHI Hao, MAO Yu-feng, MA Peng 2013, Algorithms for Recognizing Genes short coding sequence and genovariation Detection. (Mathematics in Practice and Theory), 43(14):77-84.
3. XIE Ming-jie, ZHANG Lei, CAI Rong-rong, CHEN Xi, JI Xing-lai, 2010, Feature analysis of the base mutations in coding region (BMCRs) of human syndrome-related genes, (Journal of Liaoning Normal University (Natural Science Edition)), 33(2):223-227.
4. Liu Jianzhong, Theory of Inversion Set and Its Application, Yunnan Science and Technology Publishing Press, 1999. 刘建忠, 反演集合理论及其应用, 云南科技出版社 1999.

5. Jeffrey Zheng, Variant Construction from Theoretical Foundation to Applications, Springer Nature 2019 <https://www.springer.com/in/book/9789811322815>.
6. Jeffrey Zheng, Variant Construction Theory and Applications, Vol. 1: Theoretical Foundation and Applications, Science Press 2020 (Chinese, Formal Publishing Soon). 郑智捷, 变值体系理论及其应用 第1册: 理论基础及其应用, 科学出版社 2020 (即将正式发行)
7. Jeffrey Zheng, ResearchGate http://researchgate.net/pprofile/Jeffrey_Zheng .
8. <https://www.gisaid.org/>