

DOI:10.12113/202004008

新型冠状病毒序列变异时空分析系统的设计与实现

李建勋*, 李鑫*, 谢康*, 徐德晨*, 李杰**

(哈尔滨工业大学 计算机科学与技术学院 生物信息学研究中心, 哈尔滨 150001)

摘要:新型冠状病毒肺炎目前已进入全球大流行状态,多个国家出现疫情爆发。美国疾病管制局期刊《新兴传染病》发表的关于新型冠状病毒的最新研究结论显示,新型冠状病毒基本传染数 R_0 的中位数高达 5.7,这意味着在未来较长时间内新型冠状病毒可能会在人群中持续传播并发生变异。在这一背景下,如何监视病毒的变异,对于冠状病毒的研究和药物研发具有重要意义。本文基于来自 GISAID 的病毒基因组序列数据,设计和实现了新型冠状病毒变异时空分析系统。该系统可对来自不同国家和地区新型冠状病毒序列数进行统计,对病毒序列在不同时间、不同空间内的变异情况进行分析和可视化,同时还支持不同序列之间的差异比对。该系统可为新型冠状病毒肺炎的研究和政府的疾病控制机构的决策提供支持。

关键词:新型冠状病毒;新型冠状病毒肺炎;时空变异分析

中图分类号:341 文献标志码:A 文章编号:1672-5565(2020)02-086-10

Design and implementation of SARS-CoV-2 variation spatio-temporal analysis system

LI Jianxun*, LI Xin*, XIE Kang*, XU Dechen*, LI Jie**

(Bioinformatics Research Center, School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Coronavirus disease 2019 (COVID-19) has now entered a global pandemic state, and outbreaks have occurred in many countries. The latest research findings on the new crown published by the US Centers for Disease Control and Prevention (US CDC) journal “Emerging Infectious Diseases” show that the median number of SARS-CoV-2’s basic reproduction number is as high as 5.7, which means that COVID-19 may be among the population in the future and continue to spread and mutate. How to monitor the variation of the virus is of great significance for the research of coronavirus and drug development. In this paper, based on the virus sequence data from GISAID, we designed and implemented a SARS-CoV-2 variationspatio-temporal analysis system. The system can count the number of SARS-CoV-2 sequences from different countries and regions, analyze and visualize the variation of virus sequences at different times and in different spaces, and also supports the alignment of different virus sequences. The system provides support for the research of COVID-19 and the decision-making of government disease control agencies.

Keywords: SARS-CoV-2; COVID-19; Spatio-temporal variation analysis

2019年12月出现以干咳、发热、浑身乏力等为典型症状的新型肺炎疫情,中国国家卫生健康委员会暂时将其命名为新型冠状病毒肺炎。2020年1月31日,世界卫生组织(WHO)宣布将新型冠状病毒肺炎列入“国际经济公共卫生事件”,并于2月11

日将之命名为COVID-19^[1]。同日,国际病毒分类委员会声明,将引发疾病的新型冠状病毒命名为SARS-CoV-2^[2]。

1月份,新型冠状病毒肺炎扩散到国内多个省份和地区,后续世界各国也爆发了疫情^[1]。截止到

收稿日期:2020-04-23;修回日期:2020-04-30.

基金项目:国家精准医学研究重点研发专项(No.2016YFC905);哈尔滨工业大学新型冠状病毒肺炎疫情应急科研攻关项目.

* 共同第1作者:李建勋,男,硕士生,研究方向:生物信息学.E-mail:1090803036@qq.com;李鑫,男,博士生,研究方向:生物信息学.E-mail:976185006@qq.com;谢康,男,硕士生,研究方向:生物信息学.E-mail:1157670879@qq.com;徐德晨,女,博士生,研究方向:生物信息学.E-mail:18B903025@stu.hit.edu.cn.

** 通信作者:李杰,男,教授,博导,研究方向:生物信息学.E-mail:jieli@hit.edu.cn.

4月16日,中国累计确诊COVID-19共83 799例,但疫情已得到基本的控制,现存确诊1 891例,其中境外输入1 534例,占绝大部分^①。然而,与此同时,新型冠状病毒已在世界多国相继爆发,全球累计确诊达2 047 104例,现存确诊140 675例,世界整体形势不容乐观。另外,国内外有多位相关领域的医学专家表示,新型冠状病毒短期内不会消失,可能与人类长期共存。

冠状病毒是一种包膜的单链RNA病毒,它的受体结合域能快速突变和重组,以适应并感染大量物种^[3]。自病毒疫情爆发以来,学术界对于新型冠状病毒的基本传染数R0(Basic reproduction number)有许多研究和讨论,在多篇论文中给出的取值呈越来越大的趋势。4月7日,美国疾病管制局期刊《新兴传染病》发表了关于新型冠状病毒的最新研究结论^[4],研究人员通过两种建模方法(到达模型和病例计数模型),最终计算出新型冠状病毒肺炎R0中位数为5.7(95%CI 3.8-8.9),也就是说,一名新型冠状病毒肺炎患者可以传染5.7人,是之前认为的2~3倍。

在这种背景下,对新型冠状病毒序列的变异情况进行监控、分析及可视化,就显得尤为重要^[5-6]。国内外已有一些相关的软件系统向用户开放,但是这些已有系统没有从时空角度进行分析的功能或相

关功能仍有进一步完善的空间,例如北京志诺维思公司的“战新冠”系统(<https://fight-sars2.genewis.com/>),提供了对病毒序列的变异分析及可视化,但没有从时空角度进行分析的功能;国家生物信息中心开发的2019新型冠状病毒信息库线上系统(<https://bigd.big.ac.cn/ncov/>),提供了包含时空角度的病毒序列变异分析功能,但是其时空变异分析功能是使用热力图来实现,完全是静态图的形式,也不支持用户对基准序列、时间区间和地域范围进行选择,功能上仍然需要扩展和进一步完善。本文所设计和实现的新型冠状病毒变异时空分析系统用于解决这一问题,能够动态显示不同时间和地域病毒序列的变异境况(系统现已上线,网址<http://bionet.org.cn/NCP-web/>)。

1 系统的总体功能

新型冠状病毒变异时空分析系统的主要功能有三个:病毒序列来源统计、病毒序列变异统计及病毒序列比对。

1.1 系统首页

系统首页主要包括系统导航栏、系统信息简介、系统三个主要功能模块的简介。系统首页的部分截图如图1所示。



图 1 系统首页截图展示
Fig.1 Screenshot of the homepage

①数据来源于国家及各省市地区卫健委,丁香医生整理(<https://ncov.dxy.cn/ncovh5/view/pneumonia?>)

1.2 病毒序列来源统计

自新型冠状病毒爆发以来,各国研究人员在 SARS-CoV-2 测序方面做了大量的工作^[7-12]。全球共享流感数据倡议组织(GISAID)将各国研究人员提交的病毒序列数据进行了整理和汇总,并在其官方网站(<https://www.gisaid.org/>)进行公开。本文

使用的源数据来自 GISAID 官网^[13]。

新型冠状病毒序列来源统计功能旨在对来自不同国家和地区的新型冠状病毒序列的数量进行统计、以地图和饼状图的形式进行可视化。

对新型冠状病毒序列来源的统计和可视化有两种形式——地图和饼状图(见图 2)。

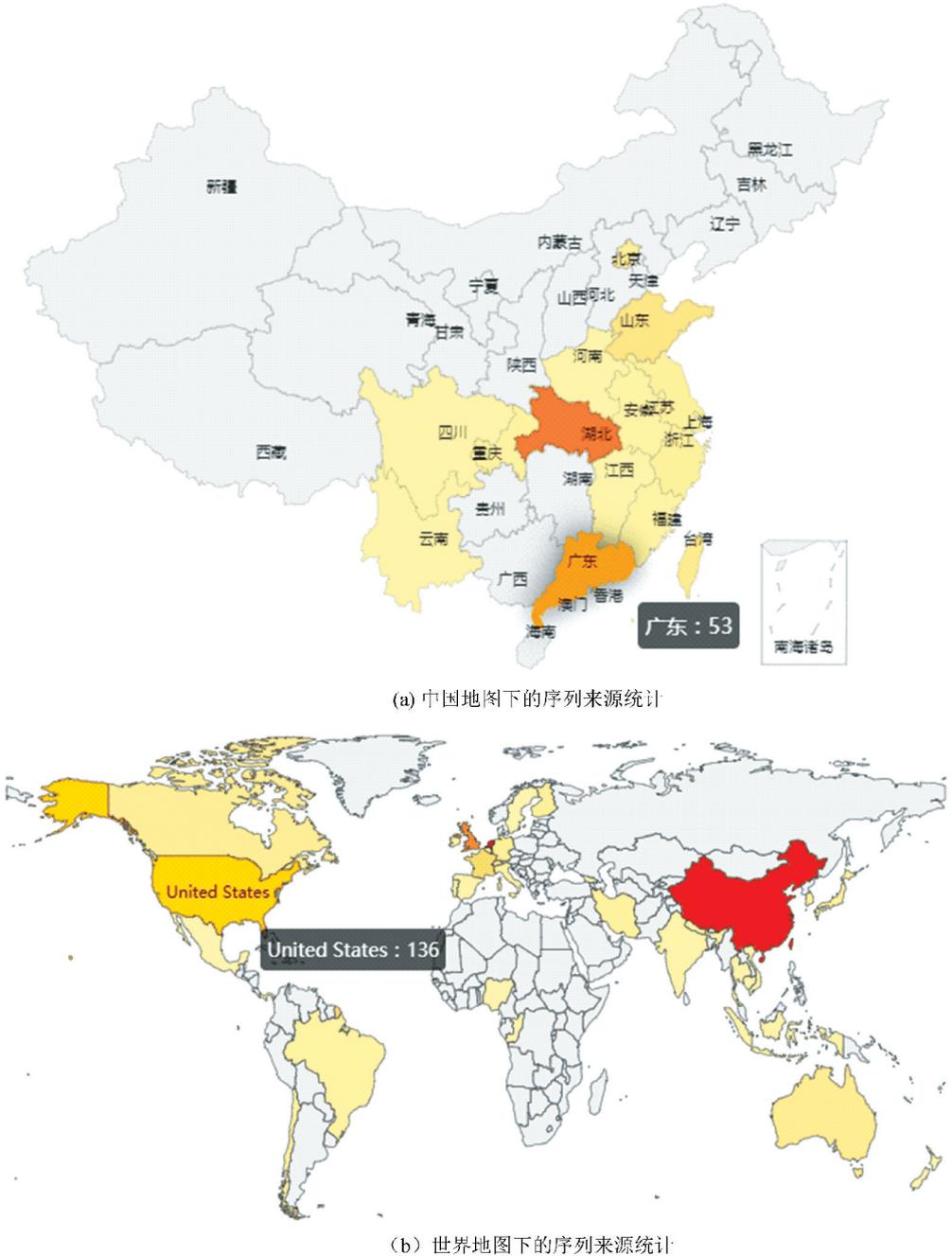


图 2 地图形式的序列来源可视化

Fig.2 Visualization of sequence sources in map form

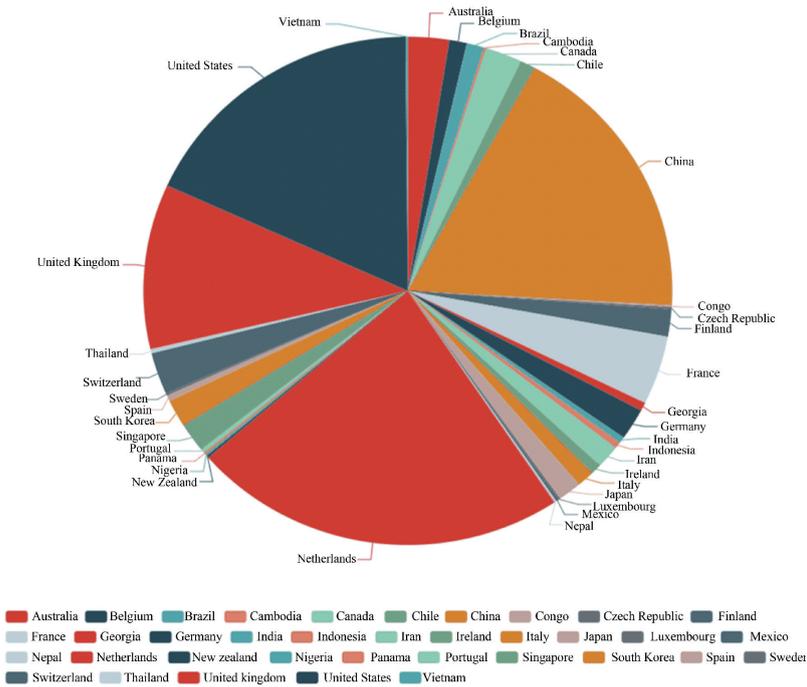
图 2 展示了中国地图和世界地图上的新型冠状病毒序列来源统计,当用户将鼠标悬浮在目标国家或者地区上时,会以弹框的形式展示该国家或地区已提交的序列总数。

图 3 展示了饼状图形式下世界各国提交序列数量统计和中国各省提交序列数量统计。此外,系统还提供将以上二者结合的饼状图可视化。同时,当用户将鼠标悬浮在图中的扇形上时,系统会以弹框的形式

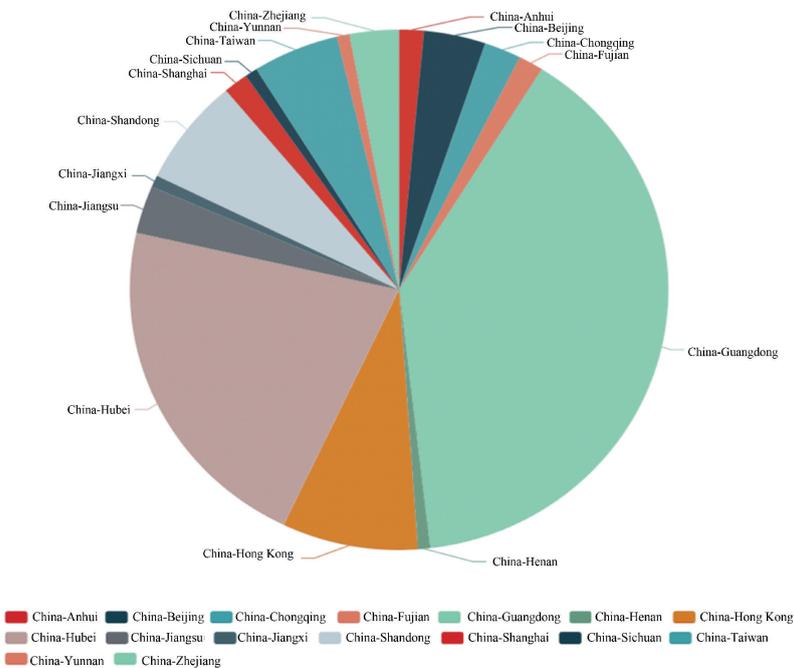
展示该国家或省份提交的序列数占总数的百分比。

根据系统给出的结果(截止到 3 月 19 日的数据),可以发现世界范围内共提交 SARS-CoV-2 序列 934 条,中国提交的病毒序列最多,共 267 条,占

28.59%;国内广东省提交的序列最多,共 90 条,占 33.71%,湖北省次之,共 42 条,占15.73%。希望世界各国及时共享更多的 SARS-CoV-2 序列,促进 COVID-19 的研究。



(a) 世界各国提交序列数量饼状图



(b) 中国各省提交序列数量饼状图

图 3 饼状图形式的序列来源可视化

Fig.3 Visualization of sequence sources in pie form

1.3 病毒序列变异统计

系统来自 GISAID 的序列数据进行筛选和预处理之后,将所有序列转换成等长的形式存入数据

库(具体的筛选与预处理过程及方法见 3.1 节病毒序列数据的处理)。在此基础上,系统在第一次启动时,会比对数据库中所有序列,将每一个位点上频

次最高的核苷酸作为该位点的标准核苷酸,从而得到一条统计意义上的标准序列(Standard sequence),后续可以此标准序列为基准进行变异分析(也支持由用户指定某一序列作为标准序列的变异分析)。

该模块下又分为序列变异统计分析和序列变异时空分析两个子功能模块。

1.3.1 序列变异统计分析

图4展示了系统为用户提供的选择入口,用户可在该入口下指定标准序列、选择待分析的目标序列集合、指定要分析的起始位点和终止位点。通过图4的设置面板完成分析基本参数选定并点击分析按钮,系统会给出病毒变异统计分析的两个结果,如图5和图6所示。

选择基准序列: StandardSequence ▾ 选择病毒序列: EPI_ISL_413583,EPI_ISL_41358...

核苷酸起始位点: 1 核苷酸终止位点: 30145 分析 重置

图4 选择基准序列与待分析序列

Fig.4 Choice for base sequence and sequences to be analyzed

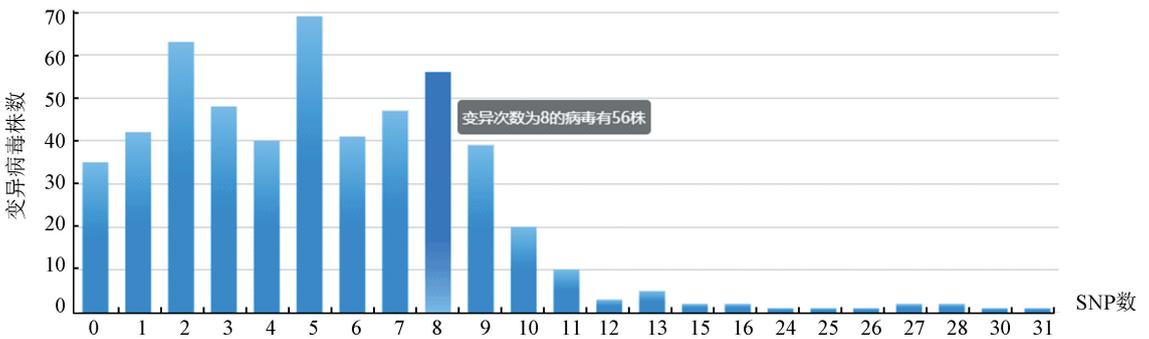


图5 在选择位点之间的变异病毒株数分布

Fig.5 Distribution of variant virus strains between selected sites

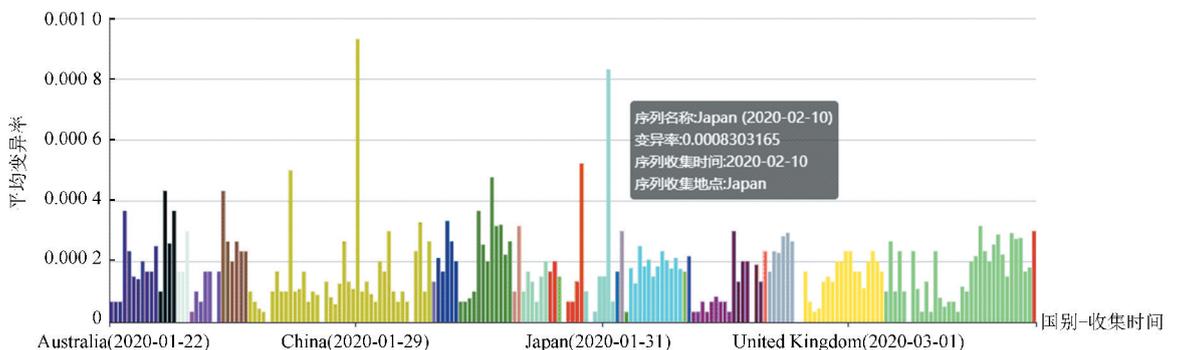


图6 在选择位点之间不同地区不同时间病毒的变异率

Fig.6 Variation rates of virus at different times in different regions between selected sites

图5展示了用户选定的目标序列集合与基准序列相比,具有不同变异位点数的病毒株数分布。除具有不同变异频数的病毒株数分布统计(见图5)之外,系统还会对病毒序列在不同时间和地区的平均变异率做统计(见图6)。该功能模块会将所有序列按照国家进行分组,每一组用不同的颜色加以区分,组内序列按照采样时间进行排序,给出所在地区在

每一个采样日的平均变异率(每一个采样日内,对所在地区当天采样的所有病毒序列的变异率求取平均值)。该柱状图可以直观地展示某一地区内病毒序列的变异率随着时间递增的变化情况。

根据系统给出的分析结果(截止到3月13日的病毒序列数据),由图5可以发现,大部分序列的变异位点数少于10个,单条序列的最多变异数为31,

有 1 条;由图 6 可以发现,中国在 2020 年 1 月 29 日采样的病毒序列的平均变异率最高,为 0.09%,其余大部分序列变异率波动不大,绝大部分序列的变异率在 0.04% 以下。

1.3.2 序列变异时空分析

随着疫情的扩散,病毒在不同地区的变异率差异、在不同时间点的变异率变化这两个信息尤为重要,可以帮助科研人员对病毒进行更加精准的分析和研究,从而为疾病的研究和防治提供指导。该功

能模块可以实现这一目标。

与病毒序列变异统计模块的参数面板相比,图 7 所示的参数面板具有更丰富的选项,包括标准序列、要分析的时间区间(以采样时间为准)、要分析的地域(以采样地域为准,可选择全世界的不同国家,也可以选择全国的不同省份)。在提交参数选项之后,系统会给出两个分析结果图,如图 8 和图 9 所示。



图 7 变异时空分析的基本参数选项

Fig.7 Basic parameter options for spatiotemporal analysis of variation

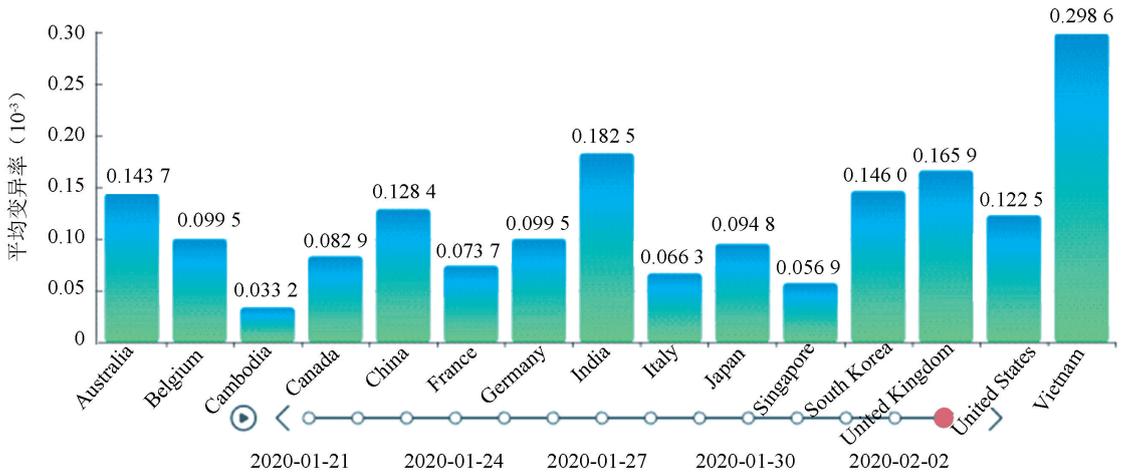


图 8 不同时间下各地区的变异率

Fig.8 Variation rates in different regions at different times

图 8 展示了在 2020 年 1 月 21 日到 2020 年 2 月 3 日之间十五个不同国家的病毒序列平均变异率(对所在国家或地区的截止到观测日期所采样的所有病毒序列的变异率求取平均值)的变化情况。在系统中实际操作时,该条形图是时间为变量进行动态播放的,在播放到最后一个时间节点时就会停止轮播。根据图 8 我们可以发现,在 1 月 21 日到 2 月 3 日之间,中国的平均变异率为 0.013%,大致在平均水平左右;平均变异率最高的国家是越南,为 0.030%;平均变异率最低的国家是柬埔寨,为 0.003%。

图 9 展示的是序列变异时空分析模块给出的第二个分析结果。与图 8 相同,该折线图在实际操作时也是随时间变化动态播放。该图展示的是 1 月 21 日到 2 月 3 日之间,十五个不同国家整体的病毒序列变异情况。其中黄色折线为最大变异率、绿色折线为最小变异率、蓝色折线为平均变异率(对全球所有国家或者国内所有省份的截止到观测日期所采样的所有病毒序列的变异率求取平均值),鼠标在图上悬浮时会给出详细信息。根据图 9 可以发现,1 月 21 日到 2 月 3 日之间,病毒序列的最大变异率为 0.09%,结合 1.3.1 节的图 6,可知该序列在

1月29日采样于中国广东;最小变异率为0%,该序列在1月23日被采样;平均变异率在0.01%左右波动。

1.4 病毒序列比对

除病毒序列的来源统计和病毒序列变异统计两个模块外,系统还设有病毒序列比对模块。该模块下又细分为病毒序列基本信息展示和病毒序列差异比对两个子模块。

1.4.1 病毒序列基本信息展示

GISAID 官网除了提供病毒序列之外,还给出了

一些病毒序列相关的信息,其中包括病毒序列号、病毒名称、采样地点、采样实验室、提交实验室、作者信息、采样时间,系统以表格的形式对这些信息做了展示(见图10)。表格的每一行对应一个病毒序列,所有病毒序列按照国家分组并按照采样时间升序排序。同时,在实际操作系统时,表中每一行前面都有一个单选框,用户在感兴趣的单选框内打钩,即可选中该行所对应的序列,进行病毒序列差异比对。

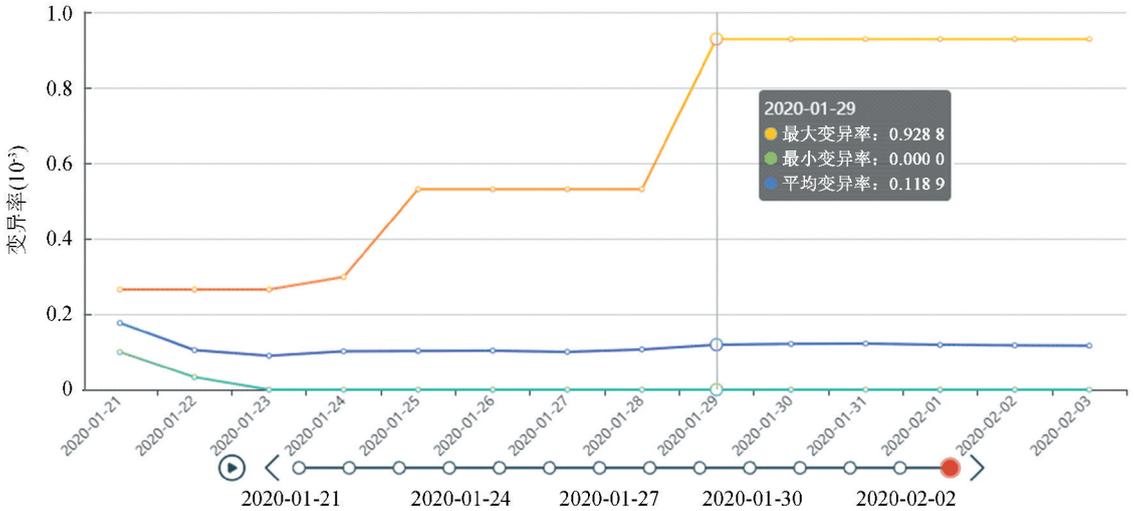


图 9 不同时间下所有地区的累计变异率

Fig.9 Cumulative variation of all regions at different times

选择相关序列: EPI_ISL_402129,EPI_ISL_402130... 设置 重置

<input type="checkbox"/>	Sequence ID	Virus name	Location	Origin lab	Submit lab	Authors	Sample time
<input checked="" type="checkbox"/>	EPI_ISL_402129	hCoV-19/Wuhan/WIV06/2019	China	Wuhan Jinyintan Hospital	Wuhan Institute of Virology, Chinese Academy of Sciences	Peng Zhou, Xing-Lou Yang, Ding-Yu Zhang, Lei Zhang, Yan Zhu, Hao-Rui Si, Zhengli Shi	2019-12-30
<input checked="" type="checkbox"/>	EPI_ISL_402130	hCoV-19/Wuhan/WIV07/2019	China	Wuhan Jinyintan Hospital	Wuhan Institute of Virology, Chinese Academy of Sciences	Peng Zhou, Xing-Lou Yang, Ding-Yu Zhang, Lei Zhang, Yan Zhu, Hao-Rui Si, Zhengli Shi	2019-12-30
<input type="checkbox"/>		hCoV		Guangdong Provincial Center for Diseases	Department of Microbiology, Guangdong Provincial	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shuijiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie Zhong, Yingtao Zhang, Lirong Zou, Xuguang Chen, Qi Zhu, Hongyan Xiao, Hongyan Cao, Zhe Liu	

显示第 1 到第 10 条记录, 总共 530 条记录 每页显示 10 条记录

图 10 序列的基本信息表

Fig.10 Sequence basic information table

1.4.2 病毒序列差异比对

图 11 展示的是系统对 EPI_ISL_403932、EPI_

ISL_408484、EPI_ISL_410984、EPI_ISL_415462 和 EPI_ISL_415652 共五条序列在 1 到 30145 这一位点

区间上进行差异比对的结果,最终以表格的形式将所有存在差异的位点一一列出。根据表格可以发现,上述五条序列在 298、1 247、3 094、8 839、9 495、11 143、13 111、14 468、15 384、19 251、23 467、24

098、24 926、26 793、28 141、28 208、29 159这些位点上存在差异,即这些位点都是可能发生了变异的位点。

2-5)序列对比: EPI_ISL_403932,EPI_ISL_40848...

核苷酸起始位点: 核苷酸终止位点: 核苷酸序列对比 重置

seq name	298	1247	3094	8839	9495	11143	13111	14468	15384	19251	23467	24098	24926	26793	28141	28208	29159
EPI_ISL_403932	C	C	C	T	C	G	C	C	C	T	A	C	A	T	G	C	T
EPI_ISL_408484	C	C	C	T	C	G	C	C	C	A	A	T	A	C	C	C	C
EPI_ISL_410984	C	T	C	C	T	T	C	C	C	T	A	C	A	T	G	T	C
EPI_ISL_415462	T	C	T	C	C	G	C	T	C	T	G	C	G	T	G	T	C
EPI_ISL_415652	T	C	T	C	C	G	T	T	T	T	G	C	A	T	G	T	C

图 11 病毒序列差异比对

Fig.11 Virus sequence difference alignment

2 系统的设计与实现

2.1 数据库

系统使用了关系型数据库 MySQL,以每一条序列作为一个实体,将所有数据都整合到一张表中。即表中综合了每个序列的基本信息(序列名、病毒名、采样地点、采样时间和提交时间等)和核苷酸排布信息(以字符串形式存储,在数据库中字段的类型为不限长度的 text 类型)。表的主要字段如图 12 所示。

ncp_seq_data

- 🔑 seq_id: int(11)
- useable: enum
- seq_name: varchar(30)
- virus_name: varchar(100)
- continent: varchar(30)
- country: varchar(30)
- province: varchar(30)
- city: varchar(30)
- sample_time: varchar(20)
- submit_time: date
- origin_lab: varchar(500)
- submit_lab: varchar(500)
- authors: text
- virus_type: varchar(40)
- host: varchar(40)
- seq_str: text

图 12 表格主要字段

Fig.12 Main fields of the data table

2.2 服务端

系统服务器端使用 Spring、SpringMVC 配合

MyBatis 框架作为主要的技术进行实现^[14],并将项目部署在阿里云 ECS 服务器上提供服务。

2.3 客户端

系统的客户端主要采用 jquery、bootstrap、echarts 框架^[15]进行开发。jquery 框架作为一个快速简洁的轻量级框架,提供了完备的功能接口,支持丰富的插件便于扩展,同时能够较好的支持服务端与客户端得数据交互。bootstrap 框架具有响应式布局设计,能让网站兼容不同分辨率设备,给用户较好的体验。而 echarts 框架能够将我们的数据以图表可视化的方式进行呈现,有利于用户对数据进行阅读和分析。

3 数据处理与分析

3.1 病毒序列数据的处理

为了寻找发生变异的位点,我们首先将新型冠状病毒序列进行了多重序列比对。本研究中使用多序列对比工具是 MUSCLE^[16],软件版本为 3.8.31。比起其它常用的序列对比工具,如 ClustalW^[17]、T-coffee^[18],MUSCLE 在保证较高的准确度的同时,具有更快的多序列比对速度。在参数选择上,我们设置了参数-diags 来适用于高度同源的新型冠状病毒序列,同时最大迭代次数(-maxiters)选择 16 来保证多序列对比的准确性,而其它参数选择了默认设置。

多重序列比对之后,多个新型冠状病毒序列间同源序列的相同核苷酸将处在相同的位点上,而如果原始序列在该位点中不存在核苷酸,将用字符“-”来表示。经过多重序列比对的序列都将拥有相同的序列长度。

对于对比后的序列,根据编码区异常字符(N)的数量以及异常缺失位点数量,我们设置 30 个为病毒序列的筛选条件,异常位点大于筛选条件值的序列将被遗弃。

将筛选后的比对结果转化为核苷酸矩阵,每条序列为行,每个位点为列,寻找每一列是否有超过一种核苷酸,如果结果为是,则该位点为有变异。

经如上方法对来自 GISAID 的数据进行预处理之后,即可将所有病毒序列转换为等长的形式,在此基础上即可按照 2.1 节所述内容进行数据的整合和数据库表的设计与填充。

表 1 病毒变异率较高和较低的部分国家

Table 1 Countries with higher and lower virus variation rates

变异率较高的国家 (由高到低)	病毒平均变异率(10^{-3})	变异率较低的国家 (由低到高)	病毒平均变异率(10^{-3})
Mexico	0.268 6	Thailand	0.000 0
Vietnam	0.298 6	Nepal	0.033 2
Belgium	0.273 7	Cambodia	0.033 2
Chile	0.270 9	Singapore	0.057 3
Italy	0.270 1	Ireland	0.099 5

表 2 病毒变异率较高和较低的中国部分省份

Table 2 Provinces in China with higher and lower virus variation rates

变异率较高的中国 省份(由高到低)	病毒平均变异率(10^{-3})	变异率较低的中国 省份(由低到高)	病毒平均变异率(10^{-3})
山东	0.240 5	浙江	0.016 6
四川	0.199 0	江西	0.033 2
广东	0.194 6	上海	0.066 3
北京	0.141 0	重庆	0.066 3
香港	0.136 4	福建	0.082 9

结合在选择位点之间的变异病毒株数分析结果(图 5 所对应的功能模块),大部分有变异发生的病毒序列的变异位点数量小于 12 个,这也印证了目前病毒变异率较小这一结果。

进一步比较不同地区不同时间的病毒变异率,发现随着地区与时间的不同,病毒平均变异率也有差异,但总体仍处于在较小范围内波动的状态。

对于病毒序列变异分析,用户可以选择感兴趣的起始位点和终止位点对序列进行分析,比如针对病毒的 S 蛋白编码区,在本次多重序列对比中对应起始与终止位置分别为 21 627 和 25 448,在系统中选择后,可以看到变异数量的信息,其中变异位点数量(SNP)为 0 和 1 的病毒数量最多,对应病毒株数分别为 302 和 194,其余 SNP 数分别为 2、3 和 7,对应病毒株数分别为 28、6 和 1,大部分病毒序列均未发生明显改变。

3.2 病毒变异的时空分析

随着时间线的延长,各国提交的序列也逐渐增多,相应的各地区病毒序列的平均变异率也在逐渐增加,如图 9 所示。这也符合包膜的阳性单链 RNA 病毒在传播过程中容易变异的特点。我们也基于图 8 所对应的功能模块对截止到 3 月 13 日的不同国家或地区的平均变异率做了比较分析,找出了平均变异率较高和较低的部分国家以及国内部分省份,分别如表 1 和表 2 所示。可以看出,到目前为止,病毒序列的平均变异率处于一个较小的波动范围内,病毒仍然处于较为稳定的状态。

4 总结

本系统对来源于全球共享流感数据倡议组织(GISAID)的新型冠状病毒数据进行整合与分析,主要实现了以下功能:

- 1) 对数据的来源进行可视化分析。
- 2) 对序列的基本信息进行展示,并支持选定序列之间核苷酸比较。
- 3) 以动态图的形式呈现不同时间下各地区的病毒变异率,以及所有地区的累计变异率,使用户能够直观感受病毒的时空变异情况。
- 4) 对选择位点之间的变异病毒株数进行统计,并以柱状图的形式对病毒的变异率进行时间空间上的展示,方便用户对病毒的变异情况进行分析。

该系统以一种更为直观的方式呈现病毒的具体

信息,同时以图表的方式,对病毒的变异情况进行了展示,方便用户对新型冠状病毒的序列进行分析,对新型冠状病毒防治的研究具有重要的参考意义。

参考文献(References)

- [1] World Health Organization. Coronavirus disease (COVID-19) Pandemic [EB/OL]. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. 2020-01-27.
- [2] GORBALENYA A E, BAKER S C, BARIC R S, et al. The species *severe acute respiratory syndrome-related coronavirus*: classifying 2019-nCoV and naming it SARS-CoV-2 [J/OL]. *Nature Microbiology*, 2020, 5: 536 - 544. DOI: 10.1038/s41564-020-0695-z. <https://doi.org/10.3201/eid2607.2002282>. 2020-4-7.
- [3] WU Fan, ZHAO Su, YU Bin, et al. A new coronavirus associated with human respiratory disease in China [J]. *Nature*, 2020, 579: 265-269. DOI: 10.1038/s41586-020-2008-3.
- [4] SANCHE S, LIN Yenting, XU Chonggang, et al. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2 [J/OL]. *Emerging Infectious Diseases*, 2020, 26 (7). DOI: 10.3201/eid2607.200282. <https://doi.org/10.3201/eid2607.200282>. 2020-4-7.
- [5] HEYMANN D L. Data sharing and outbreaks: Best practice exemplified [J]. *The Lancet*, 2020, 395 (10223): 469 - 470. DOI: 10.1016/S0140-6736(20)30184-7.
- [6] MUNSTER V J, KOOPMANS M, VAN DOREMALEN N, et al. A novel coronavirus emerging in China—key questions for impact assessment [J]. *New England Journal of Medicine*, 2020, 382(8): 692-694. DOI: 10.1056/NEJMp2000929.
- [7] XU Xintian, CHEN Ping, WANG Jingfang, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission [J]. *Science China Life Sciences*, 2020, 63(3): 457-460. DOI: 10.1007/s11427-020-1637-5.
- [8] ZHOU Peng, YANG Xinglou, WANG Xianguang, et al. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin [J/OL]. *BioRxiv*, 2020. DOI: 10.1101/2020.01.22.914952. <https://www.biorxiv.org/content/10.1101/2020.01.22.914952>.
- [9] JI Wei, WANG Wei, ZHAO Xiaofang, et al. Cross-species transmission of the newly identified coronavirus 2019-nCoV [J]. *Journal of Medical Virology*, 2020, 92(4): 433-440. DOI: 10.1002/jmv.25682.
- [10] DONG Ning, YANG Xuemei, YE Lianwei, et al. Genomic and protein structure modelling analysis depicts the origin and infectivity of 2019-nCoV, a new coronavirus which caused a pneumonia outbreak in Wuhan, China [J/OL]. *BioRxiv*, 2020. DOI: 10.1101/2020.01.20.913368. <https://www.biorxiv.org/content/10.1101/2020.01.20.913368v2>.
- [11] BENVENUTO D, GIOVANETTI M, CICCOCCHI A, et al. The 2019 new coronavirus epidemic: evidence for virus evolution [J]. *Journal of Medical Virology*, 2020, 92(4): 455-459. DOI: 10.1101/2020.01.24.915157.
- [12] 陈嘉源, 施劲松, 丘栋安, 等. 2019 新型冠状病毒基因组的生物信息学分析 [J]. *生物信息学*, 2020, 18(2): 92-102. DOI: 10.12113/202001007. CHEN Jiayuan, SHI Jinsong, QIU Dongan, et al. Bioinformatics analysis of the 2019 novel coronavirus genome [J]. *Chinese Journal of Bioinformatics*, 2020, 18(2): 92-102. DOI: 10.12113/202001007.
- [13] SHU Yuelong, MCCAULEY J. GISAID: Global initiative on sharing all influenza data—from vision to reality [J]. *Eurosurveillance*, 2017, 22(13): 30494. DOI: 10.2807/1560-7917.ES.2017.22.13.30494.
- [14] 徐雯, 高建华. 基于 Spring MVC 及 MyBatis 的 Web 应用框架研究 [J]. *微型电脑应用*, 2012, 28(7): 1-4, 10. DOI: 10.3969/j.issn.1007-757X.2012.07.001. XU Wen, GAO Jianhua. Research of Web application framework based on spring MVC and MyBatis [J]. *Microcomputer Applications*, 2012, 28(7): 1-4, 10. DOI: 10.3969/j.issn.1007-757X.2012.07.001.
- [15] NIXON R. Learning PHP, MySQL & JavaScript: With JQuery, CSS & HTML5 [M]. Newton, MA: O'Reilly Media, Inc., 2014.
- [16] EDGAR R C, MUSCLE: Multiple sequence alignment with high accuracy and high throughput [J]. *Nucleic Acids Research*, 2004, 32(5): 1792-1797. DOI: 10.1093/nar/gkh340.
- [17] CEDRIC N, HIGGINS D G, HERINGA J. T-coffee: A novel method for fast and accurate multiple sequence alignment [J]. *Journal of Molecular Biology*, 2000, 302(1): 205-217. DOI: 10.1006/jmbi.2000.4042.
- [18] LARKIN M A, BLACKSHIELDS G, BROWN N P, et al. Clustal W and Clustal X version 2.0 [J]. *Bioinformatics*, 2007, 23(21): 2947-2948. DOI: 10.1093/bioinformatics/btm404.

[责任编辑:吴永英]