Online ISSN : 0975-4172
Print ISSN : 0975-4350
DOI : 10.17406/GJCST

Global Journal

of Computer Science and Technology: G

Interdisciplinary

Natural Evolution and Delineation

Analytics Paradigms in Cybersecurity

Highlights

Performance of Machine Learning

SARS-CoV-2 is an Robot Bioweapon

Discovering Thoughts, Inventing Future

VOLUME 21 ISSUE 2 VERSION 1.0

Global Journal of Computer Science and Technology: G
Interdisciplinary

Global Journal of Computer Science and Technology: G
Interdisciplinary

Volume 21 Issue 2 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

*(A Delaware USA Incorporation with "Good Standing"; **Reg. Number: 0423089**)*

Sponsors: Open Association of Research Society
Open Scientific Standards

Publisher's Headquarters office

Global Journals® Headquarters
945th Concord Streets,
Framingham Massachusetts Pin: 01701,
United States of America
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Journals Incorporated
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,
Pin: CR9 2ER, United Kingdom

Packaging & Continental Dispatching

Global Journals Pvt Ltd
E-3130 Sudama Nagar, Near Gopur Square,
Indore, M.P., Pin:452009, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investors@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Excluding Air Parcel Charges):

Yearly Subscription (Personal & Institutional)
250 USD (B/W) & 350 USD (Color)

Editorial Board

Global Journal of Computer Science and Technology

Dr. Stefano Berretti

Ph.D. in Computer Engineering and Telecommunications, University of Firenze Professor Department of Information Engineering, University of Firenze, Italy

Dr. Abdurrahman Arslanyilmaz

Computer Science & Information Systems Department Youngstown State University Ph.D., Texas A&M University University of Missouri, Columbia Gazi University, Turkey

Dr. Lamri Sayad

Ph.d in Computer science, University of BEJAIA, Algeria

Dr. Tauqeer Ahmad Usmani

Ph.D in Computer Science, Oman

Dr. Hazra Imran

Ph.D in Computer Science (Information Retrieval), Athabasca University, Canada

Dr. Magdy Shayboub Ali

Ph.D in Computer Sciences, MSc in Computer Sciences and Engineering, BSc in Electronic Engineering, Suez Canal University, Egypt

Dr. Nurul Akmar Binti Emran

Ph.D in Computer Science, MSc in Computer Science, Universiti Teknikal Malaysia Melaka, Malaysia

Dr. Asim Sinan Yuksel

Ph.D in Computer Engineering, M.Sc., B.Eng., Suleyman Demirel University, Turkey

Dr. Anis Bey

Dept. of Computer Science, Badji Mokhtar-Annaba University, Annaba, Algeria

Alessandra Lumini

Associate Researcher Department of Computer Science and Engineering University of Bologna Italy

Dr. Rajesh Kumar Rolen

Ph.D in Computer Science, MCA & BCA - IGNOU, MCTS & MCP - MIcrosoft, SCJP - Sun Microsystems, Singhania University, India

Dr. Rajneesh Kumar Gujral

Ph.D in Computer Science and Engineering, M.TECH in Information Technology, B. E. in Computer Science and Engineering, CCNA Certified Network Instructor, Diploma Course in Computer Servicing and Maintenance (DCS), Maharishi Markandeshwar University Mullana, India

Dr. Aziz M. Barbar

Ph.D. IEEE Senior Member Chairperson, Department of Computer Science AUST - American University of Science & Technology Alfred Naccash Avenue Ashrafieh, Lebanon

Dr. Federico Tramarin

Ph.D., Computer Engineering and Networks Group, Institute of Electronics, Italy Department of Information Engineering of the University of Padova, Italy

Dr. Chutisant Kerdvibulvech

Dept. of Inf. & Commun. Technol., Rangsit University Pathum Thani, Thailand Chulalongkorn University Ph.D. Thailand Keio University, Tokyo, Japan

Dr. Roheet Bhatnagar

Ph.D in Computer Science, B.Tech in Computer Science, M.Tech in Remote Sensing, Sikkim Manipal University, India

Contents of the Issue

Global Journal of Computer Science and Technology: G Interdisciplinary
Volume 21 Issue 2 Version 1.0 Year 2021
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Performance of Machine Learning and Big Data Analytics Paradigms in Cybersecurity and Cloud Computing Platforms

By Professor Gabriel Kabanda
University of Zimbabwe

Abstract- The purpose of the research is to evaluate Machine Learning and Big Data Analytics paradigms for use in Cybersecurity. Cybersecurity refers to a combination of technologies, processes and operations that are framed to protect information systems, computers, devices, programs, data and networks from internal or external threats, harm, damage, attacks or unauthorized access. The main characteristic of Machine Learning (ML) is the automatic data analysis of large data sets and production of models for the general relationships found among data. ML algorithms, as part of Artificial Intelligence, can be clustered into supervised, unsupervised, semi-supervised, and reinforcement learning algorithms.

Keywords: cybersecurity, artificial intelligence, machine learning, deep learning, big data analytics, cloud computing.

GJCST-G Classification: K.4.4

Strictly as per the compliance and regulations of:

Performance of Machine Learning and Big Data Analytics Paradigms in Cybersecurity and Cloud Computing Platforms

Professor Gabriel Kabanda

Abstract- The purpose of the research is to evaluate Machine Learning and Big Data Analytics paradigms for use in Cybersecurity. Cybersecurity refers to a combination of technologies, processes and operations that are framed to protect information systems, computers, devices, programs, data and networks from internal or external threats, harm, damage, attacks or unauthorized access. The main characteristic of Machine Learning (ML) is the automatic data analysis of large data sets and production of models for the general relationships found among data. ML algorithms, as part of Artificial Intelligence, can be clustered into supervised, unsupervised, semi-supervised, and reinforcement learning algorithms. The Pragmatism paradigm, which is in congruence with the Mixed Method Research (MMR), was used as the research philosophy in this research as it epitomizes the congruity between knowledge and action. The researcher analysed the ideal data analytics model for cybersecurity which consists of three major components which are Big Data, analytics, and insights. The information that was evaluated in Big Data Analytics includes a mixer of unstructured and semi-structured data including social media content, mobile phone records, web server logs, and internet click stream data. Performance of Support Vector Machines, Artificial Neural Network, K-Nearest Neighbour, Naive-Bayes and Decision Tree Algorithms was discussed. To avoid denial of service attacks, an intrusion detection system (IDS) determined if an intrusion has occurred, and so monitored computer systems and networks, and then raised an alert when necessary. A Cloud computing setting was added which has advanced big data analytics models and advanced detection and prediction algorithms to strengthen the cybersecurity system. The research results presented two models for adopting data analytics models to cybersecurity. The first experimental or prototype model involved the design, and implementation of a prototype by an institution and the second model involved the use service provided by cloud computing companies. The researcher also demonstrated how this study addressed the performance issues for Big Data Analytics and ML, and its impact on cloud computing platforms.

Keywords: cybersecurity, artificial intelligence, machine learning, deep learning, big data analytics, cloud computing.

Author: Zimbabwe Academy of Sciences, TREP Building, University of Zimbabwe, Harare, Zimbabwe. e-mails: gabrielkabanda@gmail.com, profgkabanda@hotmail.com

I. Introduction

a) Background

The era of the Internet of Things (IoT) generates huge volumes of data collected from various heteregenous sources which may include mobile devices, sensors and social media. This Big Data presents tremendous challenges on the storage, processing and analytical capabilities. Cloud Computing provides a cost-effective and valid solution in support of Big Data storage and execution of data analytic applications. IoT requires both cloud computing environment to handle its data exchange and processing; and the use of artificial intelligence (AI) for data mining and data analytics. However, AI provides value-adding contributions in improving the traditional cybersecurity challenged by both the cloud vulnerability and the networking of IoT devices. Sadly, AI is also being used by hackers to threaten cybersecurity. A hybrid cybsecurity model which uses AI and Machine Learning (ML) techniques may mitigate against IoT cyber threats on cloud computing environments. As the number of IoT devices increases phenomenally, the volumes of cloud-based data and the degree of cybersecurity vulnerability increases astronomically with a high degree of complexity. The situation is exacerbated by the IoT devices that come with inadequate cybersecurity safeguards. Vulnerabilities in the IoT devices opens a window of opportunity for cyber crimes and other forms cybersecurity risks, especially among interconnected devices at now at household level.

The research paper is focused on the Performance of Machine Learning and Big Data Analytics paradigms in Cybersecurity and Cloud Computing platforms. The purpose of the research is to evaluate Machine Learning and Big Data Analytics paradigms for use in Cybersecurity. This is relevant due to the rapid advances in machine learning (ML) and deep learning (DL) as we explore the potency of efficient and cost-effective cloud computing platforms and services. Evaluation of the attacks and defenses using ML and Big Data paradigms is the key subject of this research paper. However, ML and DL techniques are resource intensive and require huge volumes of training data with excellent performance, as is often provided by

computational resources such as high-performance graphics processing units (GPUs) and tensor processing units. Security issues related to virtualisation, containerization, network monitoring, data protection and attack detection are interrogated whilst strengthening AI/ML/DL security solutions that involve encryption, access control, firewall, authentication and intrusion detection and prevention systems at the appropriate Fog/Cloud level.

Cybersecurity consolidates the confidentiality, integrity, and availability of computing resources, networks, software programs, and data into a coherent collection of policies, technologies, processes, and techniques to prevent the occurrence of an attack [1]. Cybersecurity refers to a combination of technologies, processes and operations that are framed to protect information systems, computers, devices, programs, data and networks from internal or external threats, harm, damage, attacks or unauthorized access[2]. The major cybersecurity applications are intrusion detection and malware detection. The rapid advances in mobile computing, communications and mass storage architectures have precipitated the new phenomena of Big Data and Internet of Things (IoT).

The transformation and expansion of the cyberspace has resulted in an exponential growth in the amount, quality and diversity of data generated, stored and processed by networks and hosts. These changes have necessitated a radical shift in the technology and operations of cybersecurity to detect and eliminate cyber threats so that cybersecurity remains relevant and effective in mitigating costs arising from computers, networks and data breaches [2].

The Network Intrusion Detection Systems (NIDS) is a category of computer software that monitors system behaviour with a view to ascertain anomalous violation of security policies and distinguishes between malicious users and the legitimate network users [3]. The two taxonomies of NIDS are anomaly detectors and misuse network detectors. According to [4], the components in Intrusion Detection and Prevention Systems (IDPSs) can be sensors or agents, servers, and consoles for network management. Data over networks may be secured through the use of antivirus software, firewall, encryption, secure protocols, etc. However, hackers can always devise innovative ways of breaking into the network systems. An intrusion detection and prevention system (IDPS), shown on Figure 1 below, is placed inside the network to detect possible network intrusions and, where possible, prevent the cyber attacks. The key functions of the IDPSs are to monitor, detect, analyze, and respond to cyber threats.

The strength of the overall security in Cybersecurity is determined by the weakest link [5]. Access controls and security mechanisms should be encapsulated in the company objectives. Firewall protection has proved to be inadequate because of gross limitations against external threats [6].

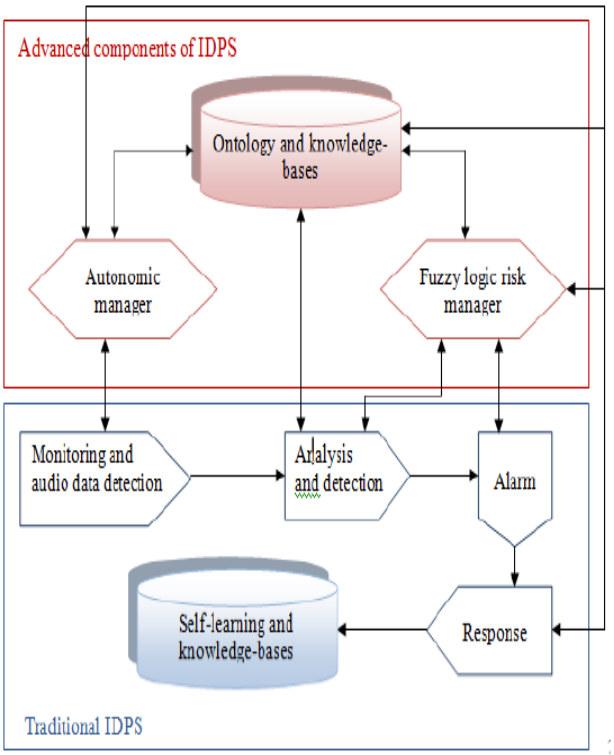


Figure 1: Typical Intrusion detection system

Computers are instructed to learn through the process called Machine Learning (ML), a field within artificial intelligence (AI). Artificial intelligence (AI) is the simulating of human intelligence in machines, through programming computers to think and act like human beings [7].The main characteristic of ML is the automatic data analysis of large data sets and production of models for the general relationships found among data. ML algorithms require empirical data as input and then learn from this input. However, the amount of data provided is often more important than the algorithm itself. Deep Learning (DL), as a special category of ML, brings us closer to AI. ML algorithms as part of Artificial Intelligence (AI) can be clustered into supervised, unsupervised, semi-supervised, and reinforcement learning algorithms. The three classes of ML are as illustrated on Figure 2 below [8], and these are:

Supervised learning: where the methods are given inputs labeled with corresponding outputs as training examples;

Unsupervised learning: where the methods are given unlabeled inputs;

Reinforcement learning: where data is in the form of sequences of observations, actions, and rewards.

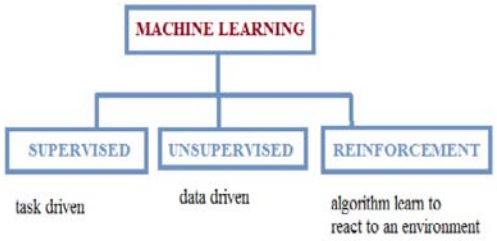


Figure 2: Three levels of Machine Learning (Source: [9])

Supervised learning models are grounded on generating functions that maps big datasets (features) into desired outputs [10]. Unsupervised learning is seen as a mathematical process of minimizing redundancy or categorizing huge datasets based on likeness [7]. It is important to note that Machine Learning is a technique of big data analytics that includes programming analytical model construction[10]. The output of a machine learning model often includes perceptions and/or decisions. Big data analytics has emerged as a discipline of ways to analyze, systematically extract/mine information from, or otherwise deal or work with enormous or complex datasets which are large to be dealt with by traditional data-processing methodologies[7].

The transformation and expansion of the cyber space has led to the generation, use, storage and processing of big data, that is, large, diverse, complex, multidimensional and usually multivariate datasets [11]. According to [12], Big Data refers to the flood of digital data from many digital sources. The data types include images, geometries, texts, videos, sounds and combinations of each. [13] explained big data as the increase in volume of data that offers difficulty in storage, processing and analysis through the traditional database technologies. Big Data came into existence when the traditional relational database systems were not able to handle the unstructured data generated by organizations, social media, or from any other data generating source [14]. The characteristics of big data are volume, velocity, variety, veracity, vocabulary and value [11]. Big data has necessitated the development of big data mining tools and techniques widely referred to as big data analytics. Big data analytics makes use of analytic techniques such as data mining, machine learning, artificial learning, statistics, and natural language processing. In an age of transformation and expansion in the Internet of Things (IoT), cloud computing services and big data, cyber-attacks have become enhanced and complicated [15], and therefore cybersecurity events become difficult or impossible to detect using traditional detection systems [16], [17]. Big Data has also been defined according to the 5Vs as stipulated by [18] where:

- Volume refers to the amount of data gathered and processed by the organisation
- Velocity referring to the time required to do processing of the data
- Variety refers to the type of data contained in Big Data
- Value referring to the key important features of the data. This is defined by the added-value that the collected data can bring to the intended processes.
- Veracity means the degree in which the leaders trust the information to make a decision.

Big Data Analytics (BDA) can offer a variety of security dimensions in network traffic management, access patterns in web transactions, configuration of network servers, network data sources, and user credentials. These activities have brought a huge revolution in the domains of security management, identity and access management, fraud prevention and governance, risk and compliance. However, there is also a lack of in-depth technical knowledge regarding basic BDA concepts, Hadoop, Predictive Analytics, and Cluster Analysis, etc. With these limitations in mind, appropriate steps can be taken to build on the skills and competences on security analytics. There is lack of infrastructure to support such innovations, lack of skilled data scientists and lack of policies or legislation that promote such innovations.

The supervised machine learning algorithm which can be used for both classification or regression challenges is called the Support Vector Machine (SVM). The original training data can be transformed into a higher dimension where it becomes separable by using the SVM algorithm which searches for the optimal linear separating hyperplane. Estimations of the relationships among variables depends mainly on the statistical process of regression analysis. The independent variables determine the estimation target. The regression function can be linear as in linear regression, or a common sigmoid curve for the logistic function.

The easiest and simplest supervised machine learning algorithm which can solve both classification and regression problems is the k-nearest neighbors (KNN) algorithm. Both the KNN and SVM can be applied to finding the optimal handover solutions in heterogeneous networks constituted by diverse cells. Given a set of contextual input cues, machine learning algorithms have the capability to exploit the user context learned. The Hidden Markov Model (HMM) is a tool designed for representing probability distributions of sequences of observations. It can be considered a generalization of a mixture-based model, rather than being independent of each other. The list of supervised learning algorithms includes Regression models, K-nearest neighbors, Support Vector Machines, and Bayesian Learning [39].

Common examples of generative models that may be learned with the aid of Bayesian techniques include, but are not limited to, the Gaussians mixture

model (GM), expectation maximization (EM), and hidden Markov models (HMM) [3, p. 445]. In Table 1, we summarize the basic characteristics and applications of supervised machine learning algorithms.

Table 1: Various attack descriptions (Source: [7])

Attack Type	Description
DoS	Denial of service: an attempt ot make a netwokr resource unavailable to its intended users; trmporairly interrupt services of a host connected to the Internet
Scan	A process that sends client requests to a range of server port addresses on a host to find an active port.
Local Access	The attacker has an account on the system in question and can use that account to attempt unauthorized tasks.
User to root	Attackers access a user account on the system and are able to exploit some vulnerability to gain root access to the system
Data	Attackers involve someone performing an action that they may be able to do on a given computer system, but that they are not allowed to do according to policy.

b) Statement of the problem

Firewall protection has proved to be inadequate because of gross limitations against external threats The fact is that the most network-centric cyberattacks are carried out by intelligent agents such as computer worms and viruses; hence, combating them with intelligent semi-autonomous agents that can detect, evaluate, and respond to cyberattacks has become a requirement [5]. The rapid development of computing and digital technologies, the need to revamp cyberdefense strategies has become a necessity for most organisations [6]. As a result, there is an imperative for security network administrators to be more flexible, adaptable, and provide robust cyber defense systems in real-time detection of cyber threats. The key problem is to evaluate Machine Learning (ML) and Big Data Analytics (BDA) paradigms for use in Cybersecurity.

c) Purpose of study

The research is purposed to evaluate Machine Learning and Big Data Analytics paradigms for use in Cybersecurity.

d) Research objectives

The research objectives are to:

1. Evaluate Machine Learning and Big Data Analytics paradigms for use in cybersecurity.
2. Develop a Cybersecurity system that uses Machine Learning and Big Data Analytics paradigms.

e) Research Questions

The main research question was:

Which Machine Learning and Big Data Analytics paradigms are most effective in developing a Cybersecurity system?

The sub questions are:

1. How are the Machine Learning and Big Data Analytics paradigms used in Cybersecurity?
2. How is a Cybersecurity system developed that uses Machine Learning and Big Data Analytics paradigms?

II. Literature Review

a) Overview

Computers, phones, internet and all other information systems developed for the benefit of humanity are susceptible to criminal activity [5]. Cybercrimes consist of offenses such as computer intrusions, misuse of intellectual property rights, economic espionage, online extortion, international money laundering, non-delivery of goods or services, etc. [13]. Intrusion detection and prevention systems (IDPS) include all protective actions or identification of possible incidents, and analysing log information of such incidents [4]. [6] recommends the use of various security control measures in an organisation. Various attack descriptions from the outcome of the research by [7] are shown on Table 1. The monotonic increase in an assortment of cyber threats and malwares amply demonstrates the inadequacy of the current countermeasures to defend computer networks and resources. To alleviate the problems of classical techniques of cyber security, research in artificial intelligence and more specifically machine learning is sought after [1], [2]. To enhance the malware and cyber-attack detection rate, one can apply deep learning architectures to cyber security.

b) Classical Machine Learning (CML)

Machine Learning (ML) is a field in artificial intelligence where computers learn like people. We present and briefly discuss the most commonly used classical machine learning algorithms.

i. *Logistic Regression (LR)*

As an idea obtained from statistics and created by [17], logistic regression is like linear regression, yet it

averts misclassification that may occur in linear regression. Unlike linear regression, logistic regression results are basically either '0' or '1'. The efficacy of logistic regression is mostly dependent on the size of the training data.

ii. *Naive Bayes (NB)*

Naive Bayes (NB) classifier is premised on the Bayes theorem which assumes independence of features. The independence assumptions in Naive Bayes classifier overcomes the curse of dimensionality.

iii. *Decision Tree (DT)*

A Decision tree has a structure like flow charts, where the root node is the top node and a feature of the information is denoted by each internal node. The algorithm might be biased and may end up unstable since a little change in the information will change the structure of the tree.

iv. *K-Nearest Neighbor (KNN)*

K-Nearest Neighbor (KNN) is a non-parametric approach which uses similarity measure in terms of distance function classifiers other than news cases. KNN stores the entire training data, requires larger memory and so is computationally expensive.

v. *Ada Boost (AB)*

Ada Boost (AB) learning algorithm is a technique used to boost the performance of simple learning algorithms used for classification. Ada Boost constructs a strong classifier using a combination of several weak classifiers. It is a fast classifier and at the same time can also be used as a feature learner. This may be useful in tasks that use imbalanced data analysis.

vi. *Random Forest (RF)*

Random forest (RF), as an ensemble tool, is a decision tree derived from a subset of observations and variables. The Random Forest gives better predictions than an individual decision tree. It uses the concept of bagging to create several minimal correlated decision trees.

vii. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) belongs to the family of supervised machine learning techniques, which can be used to solve classification and regression problems. SVM is a linear classifier and the classifier is a hyper plane. It separates the training set with maximal margin. The points near to the separating hype plane are called support vectors and they determine the position of hyper plane.

c) Modern Machine Learning

Deep learning is a modern machine learning which has the capability to take raw inputs and learns the optimal feature representation implicitly. This has performed well in various long standing artificial intelligence tasks [3]. Most commonly used deep learning architectures are discussed below in detail.

i. *Deep Neural Network (DNN)*

An artificial neural network (ANN) is a computational model influenced by the characteristics of biological neural networks. The family of ANN includes the Feed forward neural network (FFN), Convolutional neural network and Recurrent neural network (RNN). FFN forms a directed graph in which a graph is composed of neurons named as mathematical unit. Each neuron in i^{th} layer has connection to all the neurons in $i + 1^{th}$ layer.

Each neuron of the hidden layer denotes a parameter *h* that is computed by

$$h_i(x) = f\,(w_i T\, x + b_i) \quad (1)$$

$$hii: Rdi-1 \rightarrow Rdi \quad (2)$$

$$f : R \rightarrow R \quad (3)$$

Where $w_i \in Rd\times d \quad i-1$, $b_i \in Rd_i$, $d_i$denotes the size of the input, f is a non-linear activation function, ReLU.

The traditional examples of machine learning algorithms include Linear regression, Logistic regression, Linear discriminant analysis, classification and regression trees, Naïve bayes, K-Nearest Neighbour (K-NN), Kmeans clustering Learning Vector Quantization (LVQ), Support Vector Machines (SVM), Random Forest, Monte Carlo, Neural networks and Q-learning. Take note that:

- Supervised Adaptation is carried out in the execution of system at every iteration.
- Unsupervised Adaptation follows trial and error method. Based on the obtained fitness value, computational model is generalized to achieve better results in an iterative approach.

ii. *The future of AI in the fight against cybercrimes*

Exeriments showed that NeuroNet is effective against low-rate TCP-targeted distributed DoS attacks. [19] presented the Intrusion Detection System using Neural Network based Modeling (IDS-NNM) which proved to be capable of detecting all intrusion attempts in the network communication without giving any false alerts [20].

The characteristics of NIC algorithms are partitioned into two segments such as swarm intelligence and evolutionary algorithm. The Swarm Intelligence-based Algorithms (SIA) are developed based on the idea of collective behaviours of insects in colonies, e.g. ants, bees, wasps and termites. Intrusion detection and prevention systems (IDPS) include all protective actions or identification of possible incidents and analysing log information of such incidents [4].

d) Big Data Analytics and Cybersecurity

Big Data Analytics requires new data architectures, analytical methods, and tools. Big data environments ought to be magnetic, which accommodates all heterogeneous sources of

data. Instead of using mechanical disk drives, it is possible to store the primary data-base in silicon-based main memory, which improves performance. According to [21], there are four critical requirements for big data processing. The first requirement is fast data loading. The second requirement is fast query processing. The «Map» function in Hadoop accordingly partitions large computational tasks into smaller tasks, and assigns them to the appropriate key/value pairs.

Behavioral analytics provide information about the behavioral patterns of cybersecurity events or malicious data [2]. Forensics analytics locate, recover and preserve reliable forensic artefacts from specifically identified cybersecurity events or attacks [22]. Forecast analytics attempt to predict cybersecurity events using forecast analytics models and methodologies [23]. Threat intelligence helps to gather threats from big data, analyze and filter information about these threats and create an awareness of cybersecurity threats [2].

The situation awareness theory postulated by [24] posits that the success of a cybersecurity domain depends on its ability to obtain real-time, accurate and complete information about cybersecurity events or incidents [20]. The situation awareness model consists of situation awareness, decisions and action performance as shown in Figure 3.

There is consensus in prior literature that cyber security has evolved to become a problem for big data analytics. This is due to the understanding that the transformation and expansion of the cyberspace [16] has rendered traditional intrusion detection and malware detection systems obsolete. Further, even the data mining models that have been used in the past are no longer sufficient for the challenges in cyber security [16].

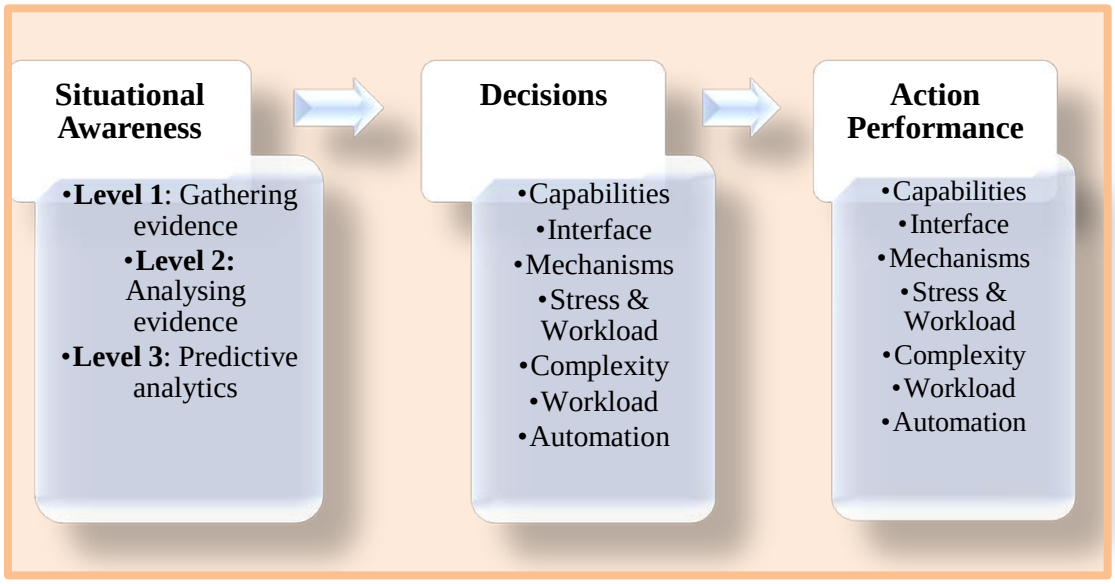


Figure 3: Simplified Theoretical Model Based on Situation Awareness

A big data analytics model for cybersecurity can be evaluated on the basis of its agility and robust [16]. According to [25], Big Data is defined not only by the amount of the information that it delivers but also by its complexity and by the speed that it is analyzed and delivered. With reference to [26], Big Data can be defined as a multi-faced data which combines the following characteristics: Veracity, Variety, Volume, Velocity and Value.

e) Advances in Cloud Computing

Cloud computing is about using the internet to access someone else's software running on someone else's hardware in someone else's data center [27]. Cloud Computing is essentially virtualized distributed processing, storage, and software resources and a service, where the focus is on delivering computing as a on-demand, pay-as-you-go service.

The NIST Cloud computing framework states that cloud computing is made up of five essential characteristics, three service models and four deployment models[28], [29], as shown on Figure 4.The five (5) essential characteristics of Cloud Computing are briefly explained follows:

On-demand self-service: A consumer can unilaterally provision computing capabilities such as server time and network storage as needed automatically, without requiring human interaction with a service provider.

Broad network access: Heterogeneous client platforms available over the network come with numerous capabilities that enable provision of network access.

Resource pooling: Computing resources are pooled together in a multi-tenant model depending on the consumer demand in a location independent manner.

Rapid elasticity: This is when unlimited capabilities are rapidly and elastically provisioned or purchased to quickly scale out; and rapidly released to quickly scale in.

Measured service: A transparent metering capability can be automatically controlled and optimized in cloud systems at some level of abstraction appropriate to the type of service.

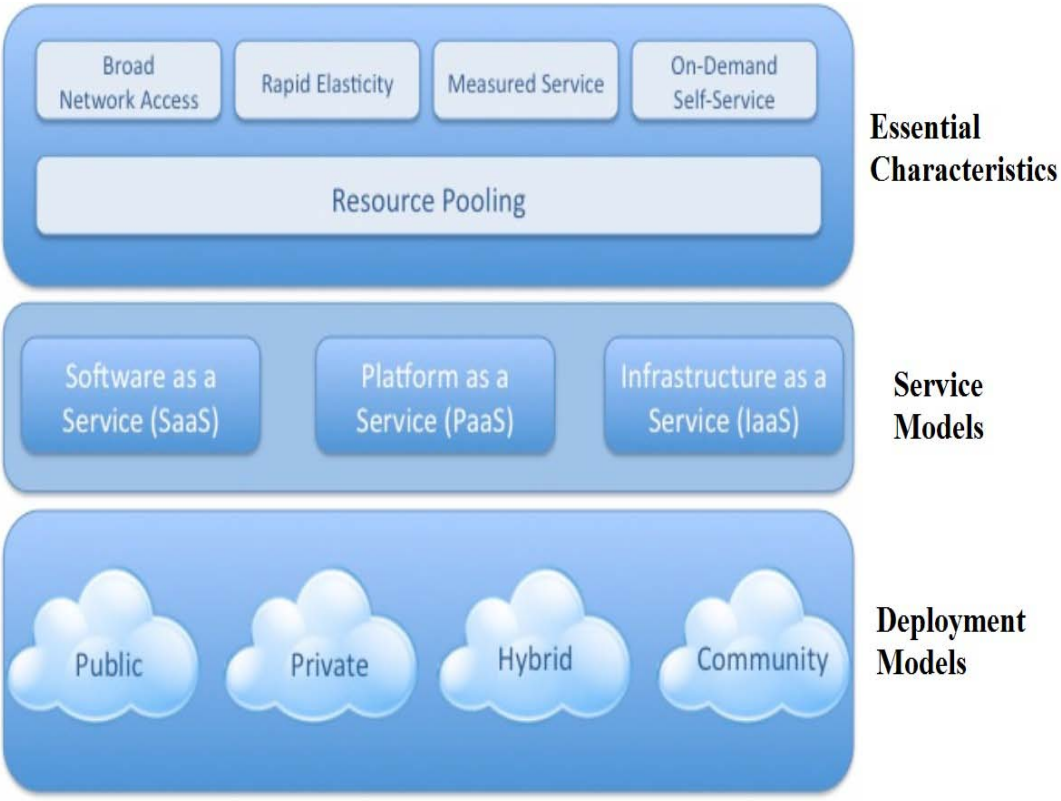


Figure 4: NIST Visual Model of Cloud Computing Definition. Source: [30].

Service delivery in Cloud computing comprises three (3) Cloud Service Models, namely Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS). These three models are shown on Figure 5, are discussed below.

i. *Software as a service (SaaS)*

The provider's applications running on a cloud infrastructure provide a capability to the consumer for use. It utilizes the Internet to deliver applications to the consumers (e.g., Google Apps, Salesforce, Dropbox, Sage X3 and office 365) [31]. This is about a wide range of applications from social to enterprise applications such as email hosting, enterprise resource planning and supply chain management. The consumer only handles minimal user specific application configuration settings. SaaS provides off-the-shelf applications offered over the internet and is the most widely used service model [29]; [32]. Examples include Google Docs, Aviary, Pixlr, and the Microsoft Office Web Application.

ii. *Platform as a service (PaaS)*

PaaS provides to the consumer infrastructure for third-party applications. Just like in SaaS the consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but does have control over the deployed applications and possibly configuration settings for the application-hosting environment [29]; [32]. Examples include Windows Azure, Apache Stratos, Google App Engine, Cloud Foundry, Heroku, AWS (Beanstalk) and Open Shift [33] & [34]. PaaS provides faster and more frequent delivery of functionality for the sake of direct support for business agility. PaaS provides an enabling environment for a consumer to run applications. A PaaS Cloud should be able to support various programming models for different types of Programming. PaaS is a Cloud Computing service that offers a computing platform and solution stack for users, and this may include the following:

- Language
- Operating System (OS)
- Database
- Middleware
- Other applications

iii. *Infrastructure as a service (IaaS)*

This provisions processing, networks, storage, and other essential computing resources on which the consumer is then able to install and run arbitrary

software, that can include operating systems (Virtual machines (VM), appliances, etc.) and applications [29]; [32]. Common global examples include Amazon Web Services (AWS), Cisco Metapod, Microsoft Azure, Rackspace and the local ones include TelOne cloud services and Dandemutande [33]. IaaS is a Cloud service that allows existing applications to run on its hardware. It rents out resources dynamically wherever they are needed.

Services include:

- Compute Servers
- Data Storage
- Firewall
- Load Balancer

f) Cloud Deployment Models

The three commonly-used cloud deployment models are private, public, and hybrid. An additional model is the community cloud. However, this is less commonly used. In a Cloud context the term deployment basically refers to where the software is made available, in other words where it is running.

i. *Private Cloud*

The private cloud is normally either owned or exclusively used by a single organization. The services and infrastructure are permanently kept on a private network, the hardware and software are dedicated solely to the particular organisation. The service provider or the particular organization may manage the physical infrastructure. The major advantage of this model is the improved security as resources are not shared with others thereby allowing for higher levels of control and security [35].

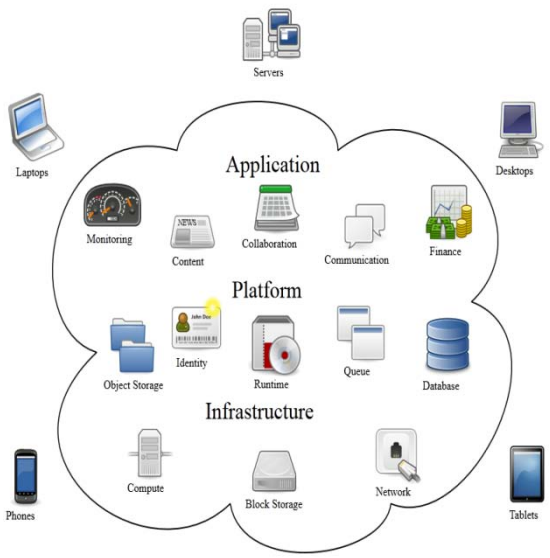


Figure 5: Cloud Computing Service Models

ii. *Public Cloud*

The cloud infrastructure is provisioned for use by the general public. The public cloud is sold to the public, as a mega-scale infrastructure, and is available to the general public. [12] further clarifies that cloud services are provided on a subscription basis to the public. It is typically based on a pay-per-use model. The advantages include lower costs, near-unlimited scalability and high reliability [35].

Examples include Amazon (EC2), IBM's Blue Cloud, Sun Cloud, Google App Engine and Windows Azure [37].

iii. *Hybrid Cloud*

A hybrid cloud model is a mix of two or more cloud deployment models such as private, public or hybrid [36]; [38]. This model requires determining the best split between the public and private cloud components. The advantages include control over sensitive data (private cloud), flexibility, i.e. ability to scale to the public cloud whenever needed and lastly allows for ease transitioning to the cloud through gradual migration [35]. The use of standardized or proprietary technology allows for data and application portability [39].

iv. *Community Cloud*

This model is provisioned for exclusive use by a particular community of consumers bound by shared interests (e.g., policy and compliance considerations, mission and security requirements). A community cloud shares computing resources among several organizations, and can be managed by either organizational IT resources or third-party providers [29]. A typical example is the U.S.-based exclusive IBM Soft Layer cloud which is dedicated for use by federal agencies only. This approach builds confidence in the platform, which cloud consumers will use to process their sensitive workloads [37].

v. *Cloud computing benefits*

Cloud computing services are delivered when they are needed in the quantity needed at a certain time. Cloud computing has many benefits for the organizations and these include cost savings, scalability, anytime anywhere access, use of latest software versions, energy saving and quick rollout of business solutions. The cost effectiveness and efficiency of the cloud platforms is tempting most organizations to migrate to the cloud and enjoy a wide range of general benefits [40] which according to [41] include:

- Free capital expenditure
- Accessibility from anywhere at anytime
- No maintenance headaches
- Improved control over documents as files will be centrally managed
- Dynamically scalable
- Device independent
- Instant (Cost-efficient and Task-Centrism)
- Private Server Cost

The NIST Cloud Computing Definition Framework is shown below on Figure 6.

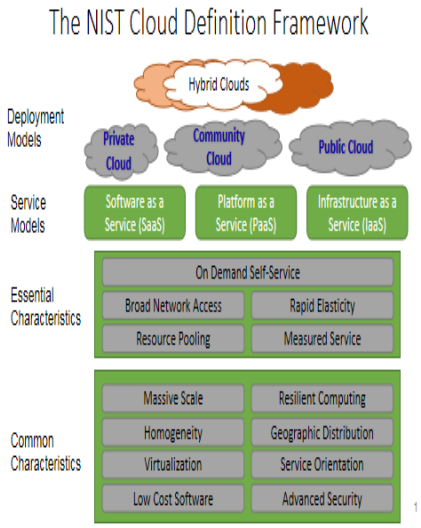


Figure 6: The NIST Cloud Computing Definition Framework

Cloud computing leverages competitive advantage and provides improved IT capabilities. Benefits of Cloud Computing are summarised below under business and technical benefits. The Business benefits of Cloud Computing include the following:

- Almost zero upfront infrastructure investment
- Just-in-time Infrastructure
- More efficient resource utilization
- Usage-based costing
- Reduced time to market
- Flexibility
- Cost Reduction
- Agility
- Automatic software/hardware upgrades

The Technical Benefits of Cloud Computing are:

- Automation – "Scriptable infrastructure"
- Auto-scaling
- Proactive Scaling
- More Efficient Development lifecycle
- Improved Testability
- Disaster Recovery and Business Continuity

However, the major issues of concern and cons on Cloud Computing include the following:

- Requires a constant internet connection
- Doesn't work well with low-speed connections
- Can be slower than using desktop software
- Features might be more limited
- Stored data might not be secure
- If the cloud loses your data, big problem
- Privacy
- Security
- Availability
- Legal Issue
- Compliance
- Performance

The top six benefits of cloud computing can be summarized as follows:

1. Achieve economies of scale: This results in increase of volume output or productivity with fewer resources (computing and human).
2. The move from CapExto OpEx reduces the Capital expenditure (CapExto) on the "pay as you go" operational expenditure (OpEx) model, based on demand / utility computing, will help reduce capital expenditure (CapEx) on hardware and software licenses.
3. Improve access: Through omni-channel access,, information can be accessed anytime, from anywhere and from any device.
4. Implement agile development at low cost: This is about the design, development and rollout of new solutions and services using agile methodologies on cloud-based shared development operations.
5. Leverage the global workforce: One can roll out Cloud Computing services 24/7 through various data centres worldwide to ensure that services are available close to the end users.
6. Gain access to advanced capabilities: The latest advances in software (such as AI, Blockchain, Data Mining) are available off-the-shelf as cloud services, enabling an organization to gain the benefits of these capabilities with minimal investment.

In conclusion the characteristics of cloud computing are leveraged through the following:

- Massive scale
- Homogeneity
- Virtualization
- Resilient computing
- Low cost software
- Geographic distribution
- Service orientation
- Advanced security technologies

g) The Advantages of the Network Function Virtualization

Network function virtualization (NFV) is a new paradigm to design and operate telecommunication networks. Traditionally, these networks rely on dedicated hardware-based network equipment and their functions to provide communication services. However, this reliance is becoming increasingly inflexible and inefficient, especially in dealing with traffic bursts for example during large crowd events. NFV strives to overcome current limitations by (1) implementing network functions in software and (2) deploying them in a virtualized environment. The resulting virtualized network functions (VNFs) require a virtual infrastructure that is flexible, scalable and fault tolerant.

The growing maturity of container-based virtualization and the introduction of production-grade container platforms promotes containers as a candidate for the implementation of NFV infrastructure (NFVI). Containers offer a simplified method of packaging and deploying applications and services. Virtualization is basically making a virtual image or "version" of something usable on multiple machines at the same time. This is a way of managing the workload by transforming traditional computing to make it more scalable, efficient and economical. Virtualization can be applied to hardware-

h) Why Virtualization?

With virtualization, one can attain better utilization rate of the resources of the service providers, increased ROI for both the service providers and the consumers, and promotes the green IT by reducing energy wastage. Virtualization technology has the drawbacks of the chance of a single point of failure of the software achieving the virtualization and the performance overhead of the entire system due to virtualization. Virtualization in general has tremendous advantages. The advantages of virtual machines are as follows:

- Where the physical hardware is unavailable, run the operating systems,
- Easier to create new machines, backup machines, etc.,
- Use of "clean" installs of operating systems and software for software testing
- Emulate more machines than are physically available,
- Timeshare lightly loaded systems on one host,
- Debug problems (suspend and resume the problem machine),
- Easy migration of virtual machines,
- Run legacy systems!

Two or more CPUs can work together on the same chip in multicore technology as a single integrated circuit (IC). These single ICs are called a *die.* Multicore technology can be used to speed up the processing in a multitenant cloud environment. Multicore architecture has become the recent trend of high-performance processors, and various theoretical and case study results illustrate that multicore architecture is scalable with the number of cores.

Most of the software vendors raised a complaint that their application is not supported in a virtual state or will not be supported if the end-user decides to virtualize them. To accommodate the needs of the industry and operating environment, to create a more efficient infrastructure – virtualization process has been modified as a powerful platform, such that the process virtualization greatly revolves around one piece of very important software. This is called a *hypervisor*. Thus, a VM must host an OS kernel.

i) Compare and Contrast Between Virtualization and Containerization

Virtualization allows the running of multiple operating systems on a single physical system and share the underlying hardware resources. Virtualization entails abstraction and encapsulation. However, Cloudsrely heavily on virtualization, whereas Grids do not rely on virtualization as much as clouds. In Virtualization, a hypervisor is a piece of computer software that creates and runs virtual machines.

Instead of installing the operating system as well as all the necessary software in a virtual machine, the docker images can be easily built with a Dockerfifile since the hardware resources, such as CPU and memory, will be returned to the operating system immediately. Therefore, many new applications are programmed into containers. Cgroups allow system administrators to allocate resources such as CPU, memory, network, or any combination of them, to the running containers. This is illustrated in Figure 7 below.

Figure 7: Architecture comparison of virtual machine Vs container

Virtualization

Virtualization is the optimum way to enhance resource utilization in efficient manner. It refers to the act of creating a virtual (similar to actual) variations of the system. Physical hardware is managed with the help of software and converted into the logical resource that will be in a shared pool or can be used by the privileged user. This service is known as VMs we can say Infrastructure as a service. Virtualization is the base of any public and private cloud development. Most of the public cloud providers such as Amazon EC2, Google Compute Engine and Microsoft Azure leverage virtualization technologies to power their public cloud infrastructure [1]. The core component of virtualization is Hypervisors.

Hypervisor

It is a software which provides isolation for virtual machines running on top of physical hosts. The thin layer of software that typically provides capabilities to virtual parti-tioning that runs directly on hardware, It provides a potential for virtual partitioning and responsible for running multiple kernels on top of the physical host. This feature makes the application and process isolation very expensive. There will be a big

impact if computer resources can be used more efficiently. The most popular hypervisors today are VMware, KVM, Xen, and HyperV.

Basically, a container is nothing but more than a virtual file system which are isolated with some Linux kernel features, such as namespaces and process groups, from the main physical system. Through containers framework it offers an environment as close as desirable one as we want from a VM but without the overhead that comes with running on an another kernel and simulating all the hardware. Due to lightweight nature of containers, more containers can run per host than virtual machines per host. Unlike containers, virtual machine require emulation layers (either software or hardware), which consume more resources and add additional overhead.

Containers are different from Virtualization with respect to the following aspects:

1. Simple: Easy sharing of a hardware resources clean command line interface, simple REST API.
2. Fast:-Rapid provisioning, instant guest boot, and no virtualization overhead so as fast as bare metal.
3. Secure: Secure by default, combine all available kernel security feature with App Armor, user namespaces, SECCOMP.
4. Scalable: The quality-of-service may be broadcast from the from a single container on a developer laptop to a container per host in a data centre. This is also includes remote image services with Extensible storage and networking.
5. Control groups (cgroups): This is a kernel-provided mechanism for administration, grouping and tracking through a virtual file system.

Docker containers share the operating system and important resources, such as depending libraries, drivers or binaries, with its host and therefore they occupy less physical resources.

III. Research Methodology

a) Presentation of the methodology

The Pragmatism paradigm was used in this research and this is intricately related to the Mixed Methods Research (MMR).

Philosophers inclined to the pragmatic paradigm subscribe to the worldview that says it is impossible to access the truth of the real world by employing a single scientific method as supported by the Positivist paradigm or construct social reality under Interpretivist paradigm. In this research, an Interpretivist or Constructivist paradigm was used, as is illustrated on Table 2 below.

Table 2: Interpretivist paradigm components and explanation

Paradigm component	Explanation
Subjectivist epistemology	Researcher uses his/her own thinking and cognition to derive meaning from the research findings arrived at through interactive processes with the respondents
Relativist ontology	Multiple realities exist in the given setting Meaning is derived from the realities through interactions between the researcher and subjects as well as among participants
Naturalist methodology	Researcher makes use of data collected through text messages, interviews, conversations and reflective sessions as a participant observer
Balanced axiology	Research outcome will reflect the researcher's values, reporting research findings in a balanced manner

Cybersecurity is a huge area for consideration and in order to address problems within it, there is need for contextualisation. This is a clear indication that there are multiple realities out there in the world of cybersecurity as supported by the Interpretivist paradigm.

The Research methodology is a way of solving a research problem thoroughly and meticulously and includes steps followed in carrying out the research and the reasoning behind [52]. Research methodology can also be viewed as a procedural or step by step outline or framework within which research is done. Research methodology can be quantitative, qualitative or mixed. Table 3 below shows the differences between qualitative and quantitative research methodologies. The Mixed Methods Research methodology was used. In a mixed methods methodology the researcher mixes both qualitative and quantitative data and employs the practices of both qualitative and quantitative research. It is also underpinned by the pragmatic paradigm.

Table 3: Differences between qualitative and quantitative methodologies

Difference with respect to:	Quantitative methodology	Qualitative methodology
Supporting philosophy	Rationalism. Humans acquire knowledge through their capacity to reason.	Empiricism. Humans acquire knowledge through sensory experiences

Approach to inquiry	Structured or rigid /predetermined methodology	Unstructured /flexible methodology
Main purpose of investigation	To quantify the extend of variation in a situation or phenomenon	To describe variation in a phenomenon or situation
Measurement of variables	Emphasis is on some form of either measurement or classification of variables	Emphasis is on the description of variables
Sample size	Emphasis is put on a greater sample size	Fewer cases
Focus of inquiry	Narrows focus in terms of extent of inquiry but draws together required information from a bigger number of respondents	Covers multiple issues but draws together required information from a smaller number of respondents
Data analysis	Variables are put into frequency distributions or other statistical procedures	Responses or observational data is used to identify themes and their descriptions
Communication of findings	Organisation is more analytic in nature, drawing inferences and conclusions and testing strength between variables and their relationship	Organisation is more descriptive and narrative in nature

Source: [51]

i. *Research approach and philosophy*

Research approach

The researcher adopts a qualitative approach in form of focus group discussion to research. Since the analysis is done to establish differences in data analytics models for cybersecurity without the necessity of quantifying the analysis [42] .

Research philosophy

The researcher adopts a postmodern philosophy to guide the research. Firstly the researcher notes that the definition, scope and measurement of cybersecurity differs between countries and across nations [15]. Further, the post-modern view is consistent with descriptive research designs which seek to interpret situations or models in their particular contexts [43].

ii. *Research design and methods*

Research design

The researcher adopts a descriptive research design since the intention is to systematically describe the facts and characteristics of big data analytics models for cybersecurity. The purpose of the study is essentially an in-depth description of the models [42].

Research methods

A case study research method was adopted in this study. In this respect each data analytics model for cybersecurity is taken as a separate case to be investigated in its own separate context [43]. Prior research has tended to use case studies in relation to the study of cybersecurity [15]. However, the researcher develops a control case that accounts for an ideal data analytics model for cybersecurity for comparative purposes.

b) Population and sampling

i. *Population*

The research population for the purpose of this study consists of all data analytics models for cybersecurity that have been proposed and developed in literature, journals, conference proceedings and working papers. This is consistent with previous research which involves a systematic review of literature [21].

ii. *Sample*

The researcher identified two data analytics models or frameworks from a review of literature and the sample size of 8. Eight participants in total were interviewed. However, while this may be limited data, it will be sufficient for the present needs of this study. Research in future may review more journals to identify more data analytics models which can be applied to cybersecurity.

c) Sources and types of data

The researcher uses secondary data in order to investigate the application of data analytics models in cybersecurity.

d) Model for analysis

In analyzing the different data analytics models for cybersecurity the researcher makes reference to the characteristics of an ideal data analytics model for cybersecurity. In constructing an ideal model, the researcher integrates various literature sources. The basic framework for big data analytics model for cybersecurity consists of three major components which are big data, analytics, and insights [16]. However, a fourth component may be identified as prediction (or predictive analytics) [21]). This is depicted in Figure 8 below:

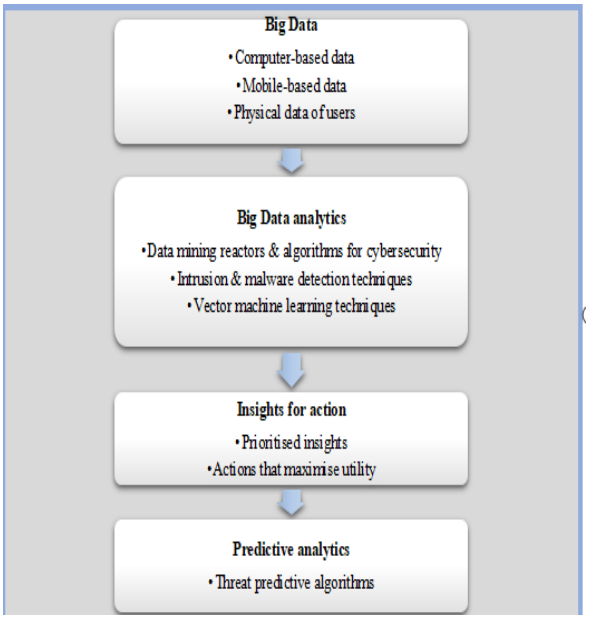


Figure 8: Big Data Analytics Model for Cybersecurity

Big data

The first component in the big data analytics framework for cybersecurity is the availability of big data about cybersecurity. Traditional sources of big data are systems logs and vulnerability scans [16]. However, sources of big data about cybersecurity have extended to include computer-based data, mobile-based data, physical data of users, human resources data, credentials, one-time passwords, digital certificates, biometrics, and social media data [11]. Some authors identify sources of big data about cybersecurity as business mail, access control systems, CRM system and human resources system, a number of pullers in linked data networks, intranet/ internet and IIoT/IoT, collectors and aggregators in social media networks and external news tapes [23]. Big data about cybersecurity should be imported from multiple sources to ensure effectiveness in detection and prediction of possible threats [17]. Further, some authors specify the characteristics of security data as consisting of heterogeneous format, diverse semantic and correlation across data sources and classify them into categories for example non-semantic data, semantic data and security knowledge data [17].

Big data analytics

The address the concerns of big data about cybersecurity, more robust big data analytics models for cybersecurity have been developed in data mining techniques and machine learning [16]. Big data analytics employ data mining reactors and algorithms, intrusion and malware detection techniques and vector machine learning techniques for cybersecurity [16]. However, it has been observed that adversarial programs have tended to modify their behavior by adapting to the reactors and algorithms designed to detect them [16]. Further, intrusion detection systems are faced with challenges such as unbounded patterns, data nonstationarity, uneven time lags, individuality, high false alarm rates, and collusion attacks [21]. This necessitates a multi-layered and multi-dimensional approach to big data analytics for cybersecurity [17], [2]. In other words an effective big data analytics model for cybersecurity must be able to detect intrusions and malware at every layer in the cybersecurity framework.

Insights for action

Big data analytics for cybersecurity should be able provide prioritized and actionable insights to cybersecurity personnel. For example setting up effective network defenders that are able to detect flaws in the network and be able to trace the source of threats or attacks [16]. Alternatively, cybersecurity personnel may update existing network defenders in light of new prioritized insights about the cybersecurity system [16].The goal of analysts should be to maximize utility derived from the cybersecurity system.

Predictive analytics

Predictive analytics refer to the application of a big data analytics model for cybersecurity to derive, from current cybersecurity data, the likelihood of a cybersecurity event occurring in future [21]. In essence, a data analytics model for cybersecurity should be able to integrate these components if it is to be effective in its major functions of gathering big data about cybersecurity, analyzing big data about cybersecurity threats, providing actionable insights and predicting likely future cybersecurity incidents.

e) Validity and Reliability

The researcher solicited comments from peers on the emerging findings and also feedback to clarify the biases and assumptions of the researcher to ensure internal validity of the study [43]. The researcher also reliability or consistency in research findings by explaining in detail the assumptions and theories underlying the study [43].

f) Summary of research methodology

In section 3, the researcher developed appropriate methodology for investigating the ideal data analytics models for cybersecurity.

g) Possible Outcomes

The expected accuracy rate for the research should be according to Table 4 below, which shows the international benchmark.

Table 4: Comparative Detection accuracy rate (%)

Classifier	Detection Accuracy (%)	Time taken to build the Model in seconds	False Alarm rate (%)
Decision Trees (J48)	81.05	**	**
Naive Bayes	76.56	**	**
Random Forest	80.67	**	**
SVM	69.52	**	**
AdaBoost	90.31	**	3.38
Mutlinomal Naive Bayes + N2B	38.89	0.72	27.8
Multinomal Naive Bayes updateable + N2B	38.94	1.2	27.9
Discriminative Multinomal Bayes + PCA	94.84	118.36	4.4
Discriminative Multinomal Bayes + RP	81.47	2.27	12.85
Discriminative Multinomal Bayes + N2B	96.5	1.11	3.0

IV. Analysis and Research Outcomes

a) Overview

Figure 11 below shows the landscape for intrusion detection. Service provision by each specific equipment with a known IP address determines the network traffic behaviour. Figure 10 below details the simple rules for the analysis of attack. The occurrence of an unusual behaviour on the network triggers an alarm on the IDS in an anomaly-based intrusion detection.

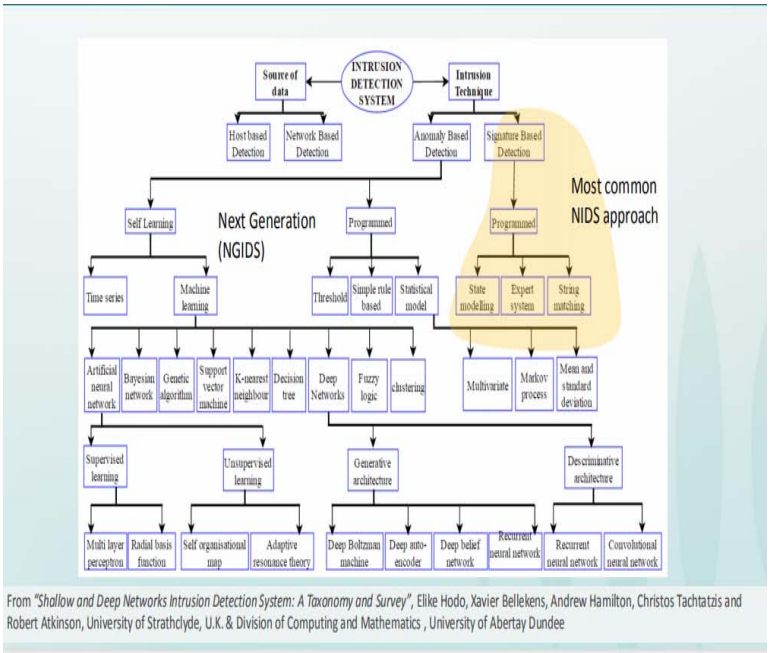


Figure 9: Landscape for Intrusion Detection

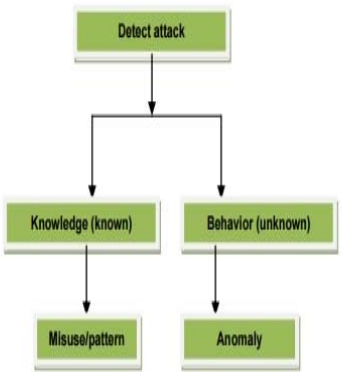


Figure 10: Analysis of Attack (Source: [8])

[10] highlighted the use of Machine Learning (ML), Neural Network and Fuzzy Logic to detect attacks on private networks on the different Artificial Intelligence (AI) techniques. It is not technically feasible to develop a perfect sophisticated Intrusion Detection System, since the majority of IDS are signature based.

The IDS is divided into either as a Host IDS (HIDS) or as a Network IDS (NIDS). Analysis of the network traffic can be handled by a NIDS which distinguishes the unlicensed, illegitimate and anomalous behavior on the network. Packets traversing through the network should generally be captured by the IDS using network taps or span port in order to detect and flag any suspicious activity [10]). Anomalous behavior on the specific device or malicious activity can be effectively detected by a device specific IDS. The vulnerability of networks and susceptibility to cyber attacks is exacerbated by the use of wireless technology [12].

The gross inadequacies of classical security measures have been overtly exposed. Therefore, effective solutions for a dynamic and adaptive network defence mechanism should be determined. Neural networks can provide better solutions for the representative sets of training data [12]. [12] argues for the use of ML classification problems solvable with supervised or semi-supervised learning models for the majority of the IDS. However, the one major limitation of the work done by [12] is on the informational structure in cybersecurity for the analysis of the strategies and the solutions of the players.

Autonomous robotic vehicles attract cyber attacks which prevent them from accomplishing the intrusion prevention mission. Knowledge-based and vehicle-specific methods have limitations in detection which is applicable to only specific known attacks [3]. The attack vectors of the attack scenarios used by [3] is shown on Figure 11 below.

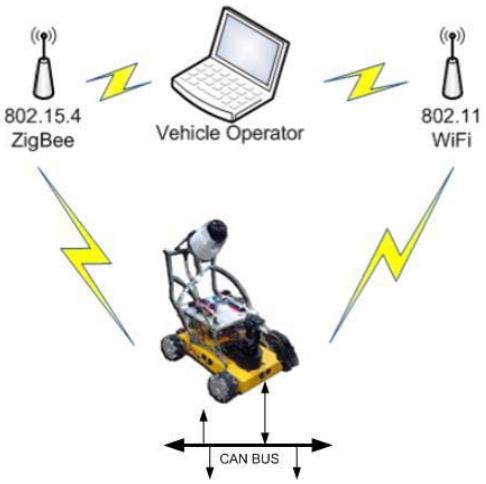


Figure 11: Attack vectors of the attack scenarios (Source: [3])

In this experiment, the system is allowed to undertake several missions by the robotic vehicle which diverts the robotic vehicle testbed. The practical experimental setup for the attack vectors used is shown on Figure 12 below.

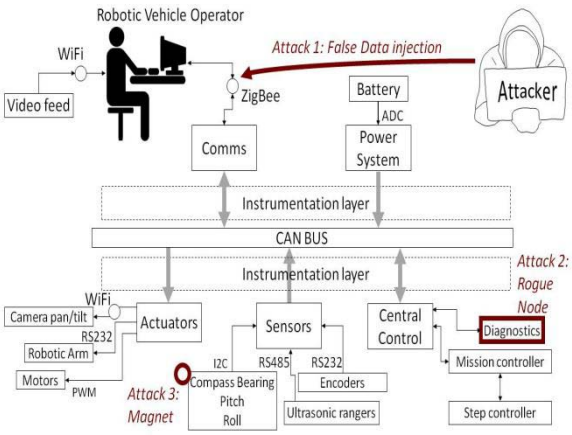


Figure 12: Attack vectors robotics experimental setup (Source: [3])

Table 5: Advantages and disadvantages of data mining techniques (Source: [1])

Technique	Advantages	Disadvantages
Genetic Algorithm	- Finding a solution for any optimization problem. - Handling multiple solution search spaces.	- Complexity to propose a problem space. - Complexity to select the optimal parameters - The need to have local searching technique for effective functioning
Artificial Neural Network	- Adapts its structure during training without the need to program it.	- Not accurate results with test data as with training data
Naive Bayes Classifier	- Very simple structure. - Easy to update.	- Not effective when there are high dependency between features.
Decision tree	- Easy to understand - Easy to implement	- Works effectively only with attributes having discrete values. - Very sensitive to training sets, irrelevant features and noise.
K Mean	- Very Easy to understand. - Very simple to implement in solving clustering problems.	- Number of clusters is not automatically calculated. - High dependency on initial centroids.

Table 5 shows a comparison of the data mining techniques that can be used in intrusion detection.

Intrusion attack classification requires optimization and enhancement of the efficiency of data mining techniques. The pros and cons of each algorithm using the NSL-KDD dataset are shown on Table 6 below.

Table 6: Performance of Support Vector Machines, Artificial Neural Network, K-Nearest Neighbour, Naive-Bayes and Decision Tree Algorithms

Parameter	SVM	ANN	KNN	NB	DT
Correctly classified instances	24519	24123	25051	22570	25081
Incorrectly classified instances	673	1069	141	2622	111
Kappa Statistic	0.9462	0.9136	0.9888	0.7906	0.9911
Mean Absolute Error	0.0267	0.0545	0.0056	0.1034	0.0064
Root Mean Squared Error	0.1634	0.197	0.0748	0.3152	0.0651
Relative Absolute Error	5.3676%	11.107%	1.1333%	20.7817%	1.2854%

An intrusion detection system determines if an intrusion has occurred, and so monitors computer systems and networks, and the IDS raises an alert when necessary [4]. However, [4] addressed the problems of Anomaly Based Signature (ABS) which reduces false positives by allowing a user to interact with the detection engine and raising classified alerts. The advantages and disadvantages of ABSs and SBSs are summarised on table, Table 7, below.

Table 7: Advantages and disadvantages of ABSs and SBSs models (Source: [4])

Detection model	Advantages	Disadvantages
Signature-based	Low false positive rate Does not require training Classified alerts	Cannot detect new attacks Requires continuous updates Tuning could be a thorny task
Anomaly-based	Can detect new attacks Self-learning	Prone to raise false positives Black-box approach Unclassified alerts Requires initial training

An IDS must keep up track of all the data, networking components and devices involved. Additional requirements must be met when developing a cloud-based intrusion detection system due to its complexity and integrated services.

b) Support vector machine

Support Vector Machine is a classification artificial intelligence and machine learning algorithm with a set containing of points of two types in X dimensional place. Support vector machine generates a (X—1) dimensional hyperplane for separating these points into two or more groups using either linear kernel or non-linear kernel functions [7]. Kernel functions provides a method for polynomial, radial and multi-layer perception classifiers such as classification of bank performance into four clusters of strong, satisfactory, moderate and poor performance. The class of bank performance is defined by the function

$$\text{Performance class} = f\left(\vec{x}.\vec{w}\right) = f\left(\sum_j x_j w_j\right)$$

Where $\vec{x}$ is the input vector to the support vector classifier and $\vec{w}$ is the real vector of weights and f is the function that translates the dot product of the input and real vector of weights into desired classes of bank performance. $\vec{w}$ Is learned from the labeled training data set.

c) KNN algorithm

The K-NN algorithm is a non-parametric supervised machine learning technique that endeavors to classify a data point from given categories with the support of the training dataset [7]. Predictions are performed for a new object (y) by searching through the whole training dataset for the K most similar instances or neighbors. The algorithm does this by calculating the Euclidean distance as follows:

$$d(x,y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2}$$

Where $d(x,y)$ is the distance measure for finding the similarity between new observations and training cases and then finding the k-closest instance to the new instance. Variables are standardized before calculating the distance since they are measured in different units. Standardization is performed by the following function:

$$X_s = \frac{X - mean}{s.d}$$

Where X_s is the standardized value, X is the instance measure, mean and s.d are the mean and standard deviation of instances. Lower values of K are sensitive to outliers and higher values are more resilient to outliers and more voters are considered to decide the prediction.

d) Multi Linear Discriminant Analysis (LDA)

The Linear Discriminant Analysis is a dimensionality reduction technique. Dimensionality reduction is the technique of reducing the amount of random variables under consideration through finding a set of principal variables [7] which is also known as course of dimensionality. The LDA calculates the separability between n classes also known as between-class variance. Let D_b be the distance between n classes.

$$D_b = \sum_{i=1}^{g} N_i\left(\overline{x_i} - \overline{x}\right)\left(\overline{x_i} - \overline{x}\right)^T$$

Where $\overline{x}$ the overall is mean, $\overline{x_i}$ and N_i are the sample mean and sizes of the respective classes. The within-class variance is then calculated, which is the distance between mean and the sample of every class. Let S_y be the within class variance.

$$S_y = \sum_{t=1}^{g}(N_i - 1)S_i = \sum_{t=1}^{g}\sum_{j}^{N_i}\left(X_{i,j} - \overline{X_i}\right)\left(X_{i,j} - \overline{X_i}\right)^2$$

The final procedure is to then construct the lower dimensional space for maximization of the seperability between classes and the minimization of within class variance. Let P be the lower dimensional space.

$$P = arg_p \max \frac{|P^T D_b P|}{|P^T S_y P|}$$

The LDA estimates the probability that a new instance belongs to every class. Bayes Theorem is used to estimate the probabilities. For instance, if the output of the class is (a) and the input is (b) then

$$P(Y = x \mid B = b) = (P \mid a * fa(b)) / \sum(P \mid a * f \mid (b))$$

P|a is the prior probability of each class as observed in the training dataset and f(b) is the estimated probability of b belonging to the class, f(b) uses the Gaussian distribution function to determine whether b belongs to that particular class.

e) Random Forest Classifier

The Random Forest classifier is an ensemble algorithm used for both classification and regression problems. It creates a set of decision trees from a randomly selected subset of the training set [7]. It then makes a decision by aggregating the votes from individual decision trees to decide the final class of an instance in the classification problem. The tree with higher error rates are given low weight in comparison to other trees increasing the impact of trees with low error rate.

f) Variable importance

Variable importance was implemented using the Boruta algorithm to improve model efficiency. The

Boruta algorithm endeavors to internment all the key, interesting features existing in the dataset with respect to an outcome variable. The diagram below shows that net profit is the most significant feature, followed by ROA, total assets, ROE and other variables depicted below in Figure 13.

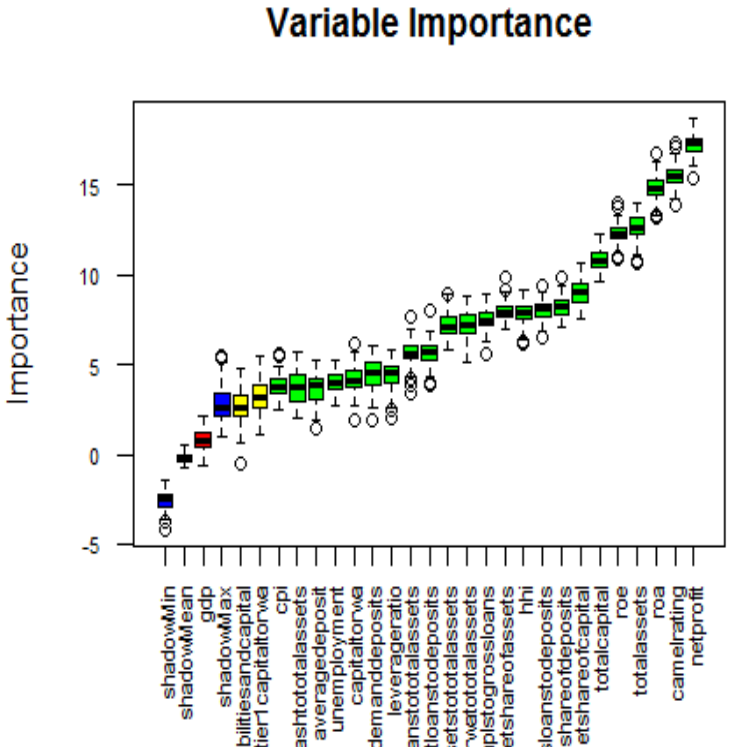


Figure13: Boruta Algorithm important features

The next procedure was fitting these variable into our algorithms and hence evaluating their performance using the metrics discussed in the models section. The Boruta algorithm also clusters banks on important variable as shown below in Figure 14 for effective risk management and analysis.

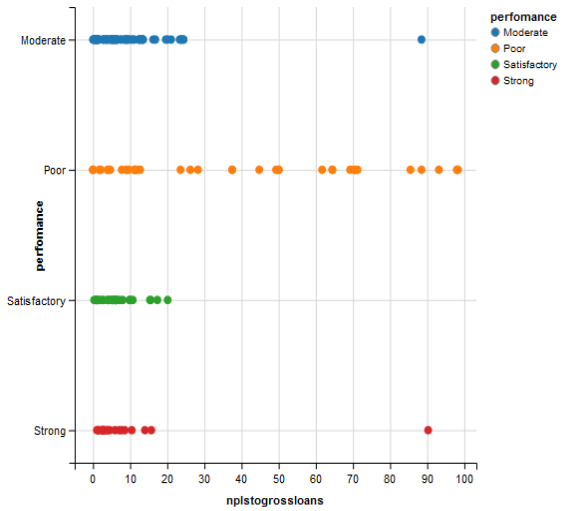


Figure 14: Boruta algorithm clustering banks based on non-performing loans

g) Model results

Before we discuss the results of our models. It is imperative to discuss the distribution of our dataset. We classify bank performance into four classes which are strong, satisfactory, moderate and poor performing banks. A strongly performing bank is the one with incredible CAMELS indicators. Its profitability indicators are high, the management quality is top of the class, less sensitive to market movements with a high quality

asset base. A satisfactory bank is the one with acceptable but not outstanding performance.

The CAMELS indicators are quite okay for such bank. Moderate performance is the one characterized by fundamental weakness or imperfections. A poorly performing bank is the one whose performance is below standard expectations or defective and hence can be categorized as an already failed bank. Our dataset comprises of thousands of records from banking institutions returns. The distribution of performance classes is shown on the diagram below. We can see that strong banks comprise of 12.9%, satisfactory banks 15.1%, moderate banks 47.5% and poor banks 24.5%. Figure 15 visualizes the effectiveness of Boruta algorithm in determining the most important variables that determines the condition of a bank.

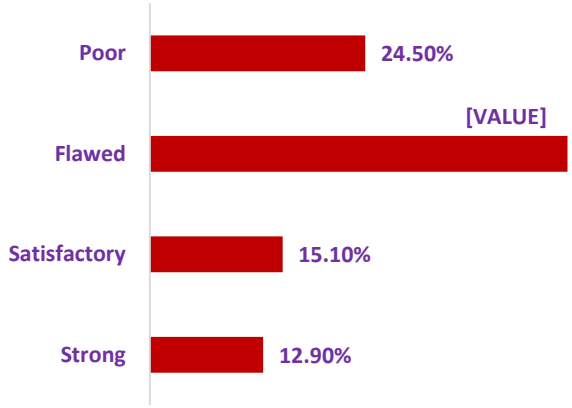


Figure 15: Distribution of the big dataset

h) Classification and Regression Trees (CART)

Table 8 below shows the performance results of our CART algorithm in predicting bank failure on the training set. The algorithm's level of accuracy on the training dataset was 82.8%. The best tune or complexity parameter of our optimal model was 0.068. The Kappa statistic was 75% envisaging that our classifier was effective as also shown with the Kappa SD of 0.07 in the classification of bank categories. On the test dataset, the algorithm achieved an accuracy level of 92.5% and a kappa of 88.72%. The algorithm only misclassified 2 instance as moderate and 1 as satisfactory.

Table 8: CART model performance

Complexity Parameter	Accuracy	Kappa	AccuracySD	KappaSD
0.06849315	0.8275092	0.7519499	0.04976459	0.07072572
0.15753425	0.7783150	0.6683229	0.07720896	0.14039942
0.42465753	0.5222344	0.1148591	0.08183351	0.18732422

The accuracy of the CART model based on the complexity parameters of different test runs is shown on Figure 16 below. The complexity parameter or the best tune parameter of 0.068 optimized the model performance.

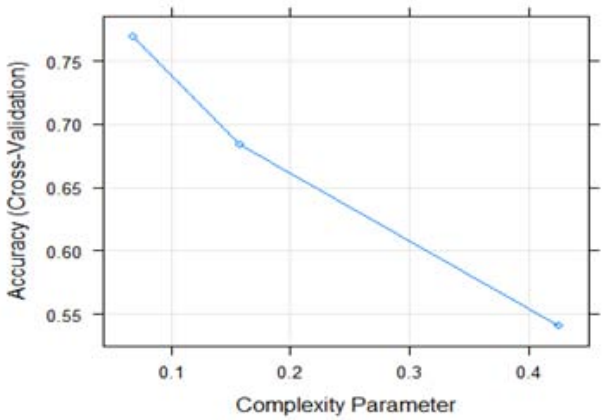


Figure 16: CART accuracy curve

i) Support Vector Machine

The accuracy level of the SVM model on the training dataset was 79.1% in predicting bank solvency as shown in table 9. The best tune sigma and cost values of our highly performing model where 0.05 and 1 as shown on Figure 19 below. The Kappa statistic and the Kappa SD where 67.9% and 0.13 respectively. On the test dataset, the algorithm achieved an accuracy level of 92.5% and a kappa of 88.54%. The algorithm only misclassified 3 instance as moderate in comparison to the CART algorithm.

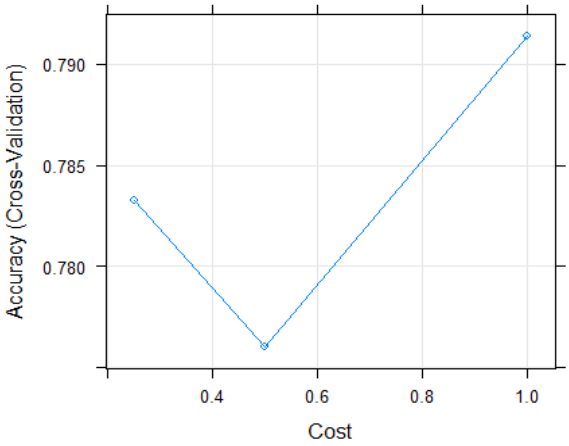


Figure 17: SVM accuracy curve

Table 9: Support Vector Machine performance

sigma	c	Accuracy	Kappa	AccuracySD	KappaSD
0.050398	0.25	0.783223	0.678536	0.095598	0.140312
0.050398	0.50	0.776007	0.661354	0.087866	0.132552
0.050398	1.00	0.791391	0.678694	0.080339	0.126466

j) Linear Discriminant Algorithm

Table 10: Linear Discriminant algorithm performance

Accuracy	Kappa	AccuracySD	KappaSD
0.8042399	0.7038131	0.1016816	0.159307

On the training dataset, the LDA achieved an accuracy level of 80% as in table 11. The Kappa statistic and the Kappa SD where 70% and 0.16 respectively. On the test dataset, the algorithm achieved an accuracy level of 90% and a kappa of 84.64%. The algorithm only misclassified 4 instance as moderate whose performance is poor in comparison to the CART algorithm.

k) K-Nearest Neighbor

Table 11: K-NN algorithm performance

K	Accuracy	Kappa	AccuracySD	KappaSD
5	0.5988645	0.3698931	0.1280376	0.2158109
7	0.6268864	0.4072928	0.1564920	0.2703504
9	0.6621978	0.4715556	0.1747903	0.2881390

The level of accuracy on the training dataset was 66.2%. The best tune parameter for our model was k=9 or 9 neighbors as shown on the accuracy curve in Figure 18 below. The Kappa statistic and the Kappa SD where 47.2% and 0.17 respectively. On the test dataset, the algorithm achieved an accuracy level of 67.5% and a kappa of 49%. The algorithm was not highly effective in classifying bank performance in comparison to other algorithms.

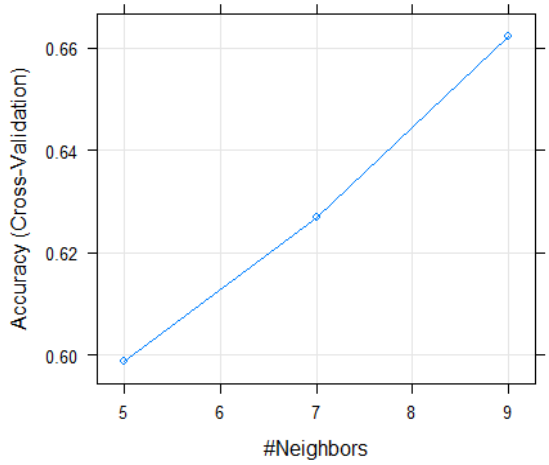


Figure 18: K-NN confusion accuracy graph

l) Random Forest

Table 12: Random Forest performance

mtry	Accuracy	Kappa	AccuracySD	KappaSD
2	0.8272527	0.7421420	0.10396454	0.15420079
14	0.8554212	0.7829891	0.06069716	0.09303130
16	0.8482784	0.7718935	0.06455248	0.09881991

On the training set, the accuracy of our random forest was 85.5% as designated in table 12. The best tune parameter for our model was the mtry of 14 which is the number of randomly selected predictors in constructing trees as shown on Figure 19. The Kappa statistic and the Kappa SD where 78.3% and 0.09 respectively. On the test dataset, the algorithm achieved an accuracy level of 96% and a kappa of 96%. The algorithm was highly effective in classifying bank performance in comparison to all algorithms.

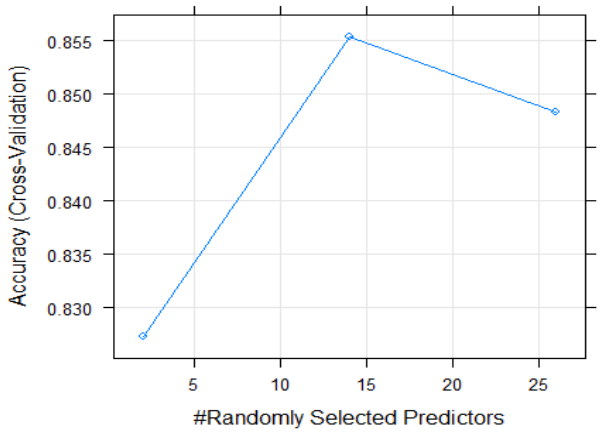


Figure 19: Random forest accuracy graph

m) Challenges and Future Direction

As number of banking activities increase, also implies that the data submission to the Reserve Bank continues to grow exponentially. This challenging situation in combination with advances in machine learning (ML) and artificial intelligence (AI) presents unlimited opportunities to apply neural network-based deep learning (DL) approaches to predict Zimbabwean Bank's solvency. Future work will focus on identifying more features that could possibly lead to poor bank performance and incorporate these in our models to develop a robust early warning supervisory tool based on big data analytics, machine learning and artificial intelligence.

The researcher analyses the two models that have been proposed in literature with reference to an ideal data analytics model for cybersecurity presented in Section 3.

Model 1: Experimental/ Prototype Model

In the first case the researcher makes reference to the model presented in [23] which although developed in the context of the public sector can be applied to the private sector organizations. Table 13 below summarizes the main characteristics of the experimental model. [The reader is referred to the prototype model also demonstrated in [23].

Software and Hardware Complex (SHC): Warning-2016

Table 13: Experimental Big Data Analytics Model for Cybersecurity

Model Attributes	Description
HBase working on HDFS (Hadoop Distributed File System)	• HBase, a non-relational database, facilitates analytical and predictive operations • Enables users to assess cyber-threats and the dependability of critical infrastructure
Analytical data processing **module**	• Processes large amounts of data, interacts with standard configurations servers and is implemented at C language • Special interactive tools (based on JavaScript/ CSS/ DHTML) and libraries (for example jQuery) developed to work with content of the proper provision of cybersecurity
Special interactive tools and libraries	• Interactive tools based on JavaScript/ CSS/ DHTML • Libraries for example jQuery developed to work with content for • Designed to ensure the proper provision of cybersecurity
Data store for example (MySQL)	• Percona Server with the ExtraDB engine • DB servers are integrated into a multi-master cluster using the Galera Cluster.
Task queues and data caching	• Redis
Database servers balancer	• Haproxy
Web server	• nginx , involved PHP-FPM with APC enabled
HTTP requests balancer	• DNS (Multiple A-records)
Development of special client applications running Apple iOS	• Programming languages are used: Objective C, C++, Apple iOS SDK based on Cocoa Touch, CoreData, and UIKit.
Development of applications running Android OS	• Google SDK
Software development for the web platform	• PHP and JavaScript.
Speed of the service and protection from DoS attacks	• CloudFare (through the use of CDN)

(Source: [23])

The proposed model, it is to be noted was demonstrated to be effective in integrating big data analytics with cybersecurity in a cost effective way [23].

Model 2: Cloud computing/Outsourcing

The second model involves an organization outsourcing its data to a cloud computing service provider. Cloud computing service providers usually have advanced big data analytics models, with advanced detection and prediction algorithms and better state of the art cybersecurity technologies and better protocols because they specialize in data and networks. However, it is to be noted that cloud computing service providers are neither exempt nor immune from cyber-threats and attacks[11].

Application of big data analytics models in cybersecurity

There is overwhelming evidence to support this assertion with many infallible proofs that such application is not only necessary in recent times but a means to survival [11], [23]. The researcher demonstrated by identifying the characteristics of an effective data analytics model, the ideal model, that it is possible to evaluate different models. In the third hypotheses the researcher postulated that, there is an appropriate big data analytics model for cybersecurity for every institution. While the review of literature showed that institutions and countries adopt different big data analytics models for cybersecurity, the researcher also demonstrated that beside the unique requirements these models share major common characteristics for

example reactors and detection algorithms are usually present in every model but differ in terms of complexity. Further, using the models presented in this Chapter it is worthy of note that many small organizations will usually adopt Model 2 whereas very large organizations and sensitive public sector organizations will adopt Model 1. This may also explain why models used may differ although the framework used in designing a data analytics model for cybersecurity in a cloud computing services provider may share similar characteristics with that developed by an institution on its own.

Summary of analysis

In this section the researcher presented two models for adopting data analytics models to cybersecurity. The first experimental or prototype model involves the design, and implementation of a prototype by an institution and the second model involves the use serviced provided by cloud computing companies. The researcher also demonstrated how this study addressed the hypotheses postulated. In the information era we are currently living in, voluminous varieties of high velocity data are being produced daily, and within them lay intrinsic details and patterns of hidden knowledge which should be extracted and utilized.

By applying such analytics to big data, valuable information can be extracted and exploited to enhance decision making and support informed decisions. Thus, the support of big data analytics to decision making was depicted.

V. Conclusion

Machine learning algorithms as part of Artificial Intelligence can be clustered into supervised, unsupervised, semi-supervised, and reinforcement learning algorithms. The main characteristic of ML is the automatic data analysis of large data sets and production of models for the general relationships found among data.

Big data analytics is not only about the size of data but also clinches on volume, variety and velocity of data .Volume denotes big data as massive; velocity denotes the high speed of big data; variety denotes the diversity of big data; veracity denotes the degrees of trustworthiness in big data; vocabulary denotes conformity of big data to different schema, models and ontologies; and value denotes the cost and worth of big data. Big data has necessitated the development of big data mining tools and techniques widely referred to as big data analytics. Big data analytics refer to a combination of well-known tools and techniques for example machine learning, and data mining, that are capable of leveraging useful data usually hidden in big data and creating an interface in the form of linear and visual analytics.

The information that is evaluated in Big Data Analytics includes a mixer of unstructured and semi-structured data, for instance, social media content, mobile phone records, web server logs, and internet click stream data. Big data analytics makes use of analytic techniques such as data mining, machine learning, artificial learning, statistics, and natural language processing. Big Data came into existence when the traditional relational database systems were not able to handle the unstructured data generated by organization, social media, or from any other data generating source.

Passive data sources can include: Computer-based data, for example geographical IP location, computer security health certificates, keyboard typing and clickstream patterns, WAP data.Data over networks may be secured through the use of antivirus software, firewall, encryption, secure protocols, etc. However, hackers can always devise innovative ways of breaking into the network systems. An intrusion detection and prevention system is placed inside the network to detect possible network intrusions and, where possible, prevent the cyber attacks.

References Références Referencias

1. BERMAN, D.S., Buczak, A.L., Chavis, J.S., and Corbett, C.L. (2019). "Survey of Deep Learning Methods for Cyber Security", *Information* 2019, *10*, 122; doi:10.3390/info10040122
2. SARKER, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data*. https://doi.org/10.1186/s40537-020-00318-5
3. BRINGAS, P.B., and Santos, I., (2010). Bayesian Networks for Network Intrusion Detection, Bayesian Network, Ahmed Rebai (Ed.), ISBN: 978-953-307-124-4, InTech, Available from: http://www.intechopen.com/books/bayesian-network/bayesian-networks-for-network-intrusion-detection
4. BLOICE, *M. & Holz*inger, A., 2018. *A Tutorial on Machine Learning and Data Science Tools with Python.* Graz, Austria: s.n.
5. NATIONAL Institute of Standards and Technology, (2018). Framework for Improving Critical Infrastructure Cybersecurity Version 1.1.
6. NIELSEN, R. (2015). CS651 Computer Systems Security Foundations 3d Imagination Cyber Security Management Plan, Technical Report January 2015, Los Alamos National Laboratory, USA.
7. UMAMAHESWARI, K., and Sujatha, S., (2017). Impregnable Defence Architecture using Dynamic Correlation-based Graded Intrusion Detection System for Cloud, Defence Science Journal, Vol. 67, No. 6, November 2017, pp. 645-653, DOI: 10.14429/dsj.67.11118.

8. NAPANDA, K., Shah, H., and Kurup, L., (2015). Artificial Intelligence Techniques for Network Intrusion Detection, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, IJERTV4IS110283 www.ijert.org, Vol. 4 Issue 11, November-2015.
9. PROKO, E., Hyso, A., and Gjylapi, D. (2018). Machine Learning Algorithms in Cybersecurity, http://www.CEURS-WS.org/Vol-2280/paper-32.pdf.
10. TRUONG, T.C; Diep, Q.B.; & Zelinka, I. (2020). Artificial Intelligence in the Cyber Domain: Offense and Defense. Symmetry 2020, 12, 410.
11. MAZUMDAR, S & Wang J (2018). Big Data and Cyber security: A visual Analytics perspective in S. Parkinson et al (Eds), Guide to Vulnerability Analysis for Computer Networks and Systems.
12. Wilson, B. M. R., Khazaei, B., & Hirsch, L. (2015, November). Enablers and barriers of cloud adoption among Small and Medium Enterprises in Tamil Nadu. In: 2015 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM) (pp. 140-145). IEEE.
13. HASHEM, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of "big data" on cloud computing: Review and open research issues. In *Information Systems*. https://doi.org/10.1016/j.is.2014.07.006.
14. SITI Nurul Mahfuzah, M., Sazilah, S., & Norasiken, B. (2017). An Analysis of Gamification Elements in Online Learning to Enhance Learning Engagement. *6th International Conference on Computing & Informatics*.
15. MENZES, F.S.D., Liska, G.R., Cirillo, M.A. and Vivanco, M.J.F. (2016) Data Classification with Binary Response through the Boosting Algorithm and Logistic Regression. Expert Systems with Applications, 69, 62-73. https://doi.org/10.1016/j.eswa.2016.08.014
16. STALLINGS, W., (2015). Operating System Stability. Accessed on 27th March, 2019. https://www.unf.edu/public/cop4610/ree/Notes/PPT/PPT8E/CH15-OS8e.pdf
17. COX, R. & Wang, G., 2014. Predicting the US bank failure: A discriminant analysis. *Economic Analysis and Policy,* Issue 44.2, pp. 201-211.
18. YANG, C., Yu, M., Hu, F., Jiang, Y., & Li, Y. (2017). Utilizing Cloud Computing to address big geospatial data challenges. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2016.10.010
19. FERNANDO, J. I., & Dawson, L. L. (2009). The health information system security threat lifecycle: An informatics theory. *International Journal of Medical Informatics*. https://doi.org/10.1016/j.ijmedinf.2009.08.006
20. MOORTHY, M., Baby, R. & Senthamaraiselvi, S., 2014. An Analysis for Big Data and its Technologies. *International Journal of Computer Science Engineering and Technology (IJCSET),* 4(12), pp. 413-415.
21. LAKSHAMI, R.V. (2019), Machine Learning for Cyber Security using Big Data Analytics Journal of Artificial Intelligence, Machine Learning and Soft Computing, 4(2) 1- 8 http://doi.org/10.5281/zenodo.3362228
22. SNOWDON, D. A., Sargent, M., Williams, C. M., Maloney, S., Caspers, K., & Taylor, N. F. (2019). Effective clinical supervision of allied health professionals: A mixed methods study. *BMC Health Services Research*. https://doi.org/10.1186/s12913-019-4873-8
23. HASSAN, H. (2017). Organisational factors affecting cloud computing adoption in small and medium enterprises (SMEs) in service sector. *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2017.11.126
24. BOU-HARB, E., & Celeda, P. (2018). *Survey of Attack Projection, Prediction, and Forecasting in Cyber Security*. *September*. https://doi.org/10.1109/COMST.2018.2871866
25. Pense (2014), Pesquisa Nacional de Saude do Escolar, Rio de Janeiro, RJ - Brazil.
26. THOMAS, E. M., Temko, A., Marnane, W. P., Boylan, G. B., & Lightbody, G. (2013). Discriminative and generative classification techniques applied to automated neonatal seizure detection. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2012.2237035
27. CENTER for Cyber and Information Security *(https://ccis.no/cyber-security-versus-information-security/*
28. GERCKE, M. (2012). 'Cybercrime Understanding Cybercrime', *Understanding cybercrime: phenomena, challenges and legal response*.
29. MURUGAN, S., and Rajan, M.S., (2014). Detecting Anomaly IDS in Network using Bayesian Network, IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p- ISSN: 2278-8727, Volume 16, Issue 1, Ver. III (Jan. 2014), PP 01-07, www.iosrjournals.org
30. BOLZONI, D., (2009). Revisiting Anomaly-based Network Intrusion Detection Systems, Ph.D Thesis, University of Twente, The Netherlands, ISBN: 978-90-365-2853-5, ISSN: 1381-3617, DOI: 10.3990/1.9789036528535,
31. JIANG, W., Wang, L., & Lin, H. (2016). The role of cognitive processes and individual differences in the relationship between abusive supervision and employee career satisfaction. *Personality and Individual Differences*. https://doi.org/10.1016/j.paid.2016.04.088
32. Cunningham, Lawrence A., (2008). The SEC's Global Accounting Vision: A Realistic Appraisal of a Quixotic Quest. North Carolina Law Review, Vol. 87, 2008, GWU Legal Studies Research Paper No. 401,

GWU Law School Public Law Research Paper No. 401, Available at SSRN: https://ssrn.com/abstract=1118377
33. PU, C. and Kitsuregawa, M., 2019, Technical Report No. GIT-CERCS-13-09; Georgia Institute of Technology, CERCS
34. Suryavanshi, A., (2017), "Magnesium oxide nanoparticle-loaded polycaprolactone composite electrospun fiber scaffolds for bone–soft tissue engineering applications: in-vitro and in-vivo evaluation", 2017 Biomed. Mater. 12 055011, https://iopscience.iop.org/article/10.1088/1748-605X/aa792b/pdf
35. BURT, D., Nicholas, P., Sullivan, K., & Scoles, T. (2013). Cybersecurity Risk Paradox. *Microsoft SIR*.
36. Pai & Aithal, (2017). The basis of social responsibility in management, Poltext, Warszawa.
37. Marzantowicz, (2015), Corporate Social Responsibility of TSL sector: attitude analysis in the light of research, "Logistyka" 2014, No. 5, pp. 1773—1785.
38. FEHLING, C., Leymann, F., Retter, R., Schupeck, W., & Arbitter, P. (2014). Cloud Computing Patterns. In *Cloud Computing Patterns*. https://doi.org/10.1007/978-3-7091-1568-8
39. KPMG (2018), Clarity on Cybersecurity. Driving growth with confidence.
40. KOBIELUS, J., (2018).Deploying Big Data Analytics Applications to the Cloud: Roadmap for Success. Cloud Standards Customer Council.
41. LEE, J. (2017). *HACKING INTO CHINA' S CYBERSECURITY LAW,* In: IEEE International Conference on Distributed Computing Systems (2017).
42. ZHANG, L., Wu, X., Skibniewski, M. J., Zhong, J., & Lu, Y. (2014). Bayesian-network-based safety risk analysis in construction projects. *Reliability Engineering and System Safety*. https://doi.org/10.1016/j.ress.2014.06.006
43. IAFRATE, F., (2015), From Big Data to Smart Data, ISBN: 978-1-848-21755-3 March, 2015, Wiley-ISTE, 190 Pages.
44. ALMUTAIRI, A., (2016). Improving intrusion detection systems using data mining techniques, Ph.D Thesis, Loughborough University, 2016.
45. BEZEMSKIJ, A., Loukas, G., Gan, D., and Anthony, R.J., (2017). Detecting cyber-physical threats in an autonomous robotic vehicle using Bayesian Networks, 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (Smart Data), 21-23 June 2017, IEEE, United Kingdom, https://ieeexplore.ieee.org/document/8276737
46. KARIMPOUR, J., Lotfi, S., and Siahmarzkooh, A.T., (2016). Intrusion detection in network flows based on an optimized clustering criterion, Turkish Journal of Electrical Engineering & Computer Sciences, Accepted/Published Online: 17.07.2016, http://journals.tubitak.gov.tr/elektrik
47. STEFANOVA, Z.S., (2018). "Machine Learning Methods for Network Intrusion Detection and Intrusion Prevention Systems", Graduate Theses and Dissertations, 2018, https://scholarcommons.usf.edu/etd/7367.
48. HAMMOND, K., 2015. *Practical Artificial Intelligence for Dummies®, Narrative Science Edition.* Hoboken, New Jersey: John Wiley & Sons, Inc.
49. GHEYAS, I. A. & Abdallah, A. E. (2016). Detection and prediction of insider threats to cyber security: A systematic Literature Review and Meta-Analysis., Big Data Analytics (2016) 1:6.
50. KANTARCIOGLU, M & Xi B (2016). Adversarial Data Mining: Big data meets cybersecurity, CCS, 16 October 24-28, 2016, Vienna, Austria.
51. KUMAR, R., (2011) *Research Methodology: A step by step guide for beginners* 3rd ed. London: Sage Publishers.
52. KOTHARI, C.R. (2004). Research Methodology Methods and Techniques 2nd Revised Edition .New Age International Publishers.

This page is intentionally left blank

Global Journal of Computer Science and Technology: G Interdisciplinary
Volume 21 Issue 2 Version 1.0 Year 2021
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Comparative Study of OpenCV Inpainting Algorithms

By Preeti Chatterjee, Subhadeep Jana & Souradeep Ghosh

Abstract- Digital image processing has been a significant and important part in the realm of computing science since its inception. It entails the methods and techniques that are used to manipulate a digital image using a digital computer. It is a type of signal processing in which the input and output maybe image or features/characteristics associated with that image. In this age of advanced technology, digital image processing has its uses manifold, some major fields being image restoration, medical field, computer vision, color processing, pattern recognition and video processing. Image inpainting is one such important domain of image processing. It is a form of image restoration and conservation. This paper presents a comparative study of the various digital inpainting algorithms provided by Open CV (a popular image processing library) and also identifies the most effective inpainting algorithm on the basis of Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM) and runtime metrics.

Keywords: image processing, openCV, Image Inpainting, Artificial Intelligence, Machine Learning.

GJCST-G Classification: B.2.4

Strictly as per the compliance and regulations of:

Comparative Study of OpenCV Inpainting Algorithms

Preeti Chatterjee [α], Subhadeep Jana [σ] & Souradeep Ghosh [ρ]

Abstract- Digital image processing has been a significant and important part in the realm of computing science since its inception. It entails the methods and techniques that are used to manipulate a digital image using a digital computer. It is a type of signal processing in which the input and output maybe image or features/characteristics associated with that image. In this age of advanced technology, digital image processing has its uses manifold, some major fields being image restoration, medical field, computer vision, color processing, pattern recognition and video processing. Image inpainting is one such important domain of image processing. It is a form of image restoration and conservation. This paper presents a comparative study of the various digital inpainting algorithms provided by Open CV (a popular image processing library) and also identifies the most effective inpainting algorithm on the basis of Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM) and runtime metrics.

Keywords: image processing, openCV, Image Inpainting, Artificial Intelligence, Machine Learning.

I. Introduction

Image processing is the technique of performing operations on an image to enhance the quality of the image, extract useful information from it, or manipulate it for better usage. Digital image processing techniques are applied in fields of computer vision, pattern recognition, video processing, image restoration and image correction [1].

Image restoration [2] and correction entails all the techniques used to restore a damaged image. It includes noise removal from the image, correcting a blurred photo, enhancing an image with defocused subject, converting a black and white image to color image, removing stains and unwanted marks from the image, etc. Image inpainting is one such technique that falls under image restoration.

Image inpainting [3] is a form of image restoration and conservation. The technique is generally used to repair photos with missing areas due to damage or aging, or mask out unpleasant deformed areas of the image. The use of inpainting can be traced back to the 1700s when Pietro Edwards, director of the Restoration of the Public Pictures in Venice, Italy, applied his scientific methodology to restore and preserve historic artworks. The modern approach to inpainting was established in 1930 during the International Conference for the Study of Scientific Methods for the Examination and Preservation of Works of Art. Technological advancements led to new applications of inpainting. Since the mid-1990's, the method of inpainting has evolved to include digital media. Widespread use of digital inpainting techniques range from entirely automatic computerized inpainting to tools used to simulate the process manually. Digital inpainting includes the use of software that relies on sophisticated algorithms to replace lost or corrupted parts of the image data. There are various advanced inpainting methodologies [4], namely Partial Differential Equation (PDE) based inpainting [5], Texture synthesis based inpainting [6], Hybrid inpainting [7], Example based inpainting [8] and Deep generative model based inpainting [9].

In this paper, we have presented a detailed comparative study of the three inpainting algorithms natively provided by the Open CV library, and also stated which is the most effective algorithm out of them. The paper is structured as follows: Section II contains the related work done in the past on comparative analysis of inpainting techniques and algorithms. Section III contains a brief theory behind the inpainting algorithms to be discussed. Section IV contains the details of the comparative study and experimental setup. Section V presents the results we obtained from our study and their critical explanations. Section VI details the possibilities of further work that can be performed on this topic. Section VII concludes the paper. We have focused more on the practical analysis of the three algorithms, and less on the theoretical and mathematical interpretation of the algorithms.

II. Related Work

The first inpainting algorithm provided by OpenCV is established on the paper "An Image Inpainting Technique based on the Fast Marching method" by Alexandru Telea [10] in 2004. It is based on the Fast Marching Method. The second inpainting algorithm provided by OpenCV is established on the paper "Navier-Stokes, Fluid Dynamics, and Image and Video Inpainting" by M. Bertalmio et al [11] in 2001. It is based on fluid dynamics. The third inpainting algorithm was reviewed in the paper "Demonstration of Rapid Frequency Selective Reconstruction for Image Resolution Enhancement" by Nils Genser et al [12] in

Author α σ: Government College of Engineering and Ceramic Technology, Kolkata, India. e-mails: pchatterje105@gmail.com, jsubhadeep1999@gmail.com
Author ρ: Heritage Institute of Technology, Kolkata, India. e-mail: jeetsouradeep@gmail.com

2017. It is based on the Rapid Frequency Selective Reconstruction (FSR) method. They applied the algorithm on Kodak and Tecnick image datasets over custom error masks and presented the Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM) and runtime metrics. We have used the same metrics for comparison, explained later in Section IV.

Supriya Chhabra et al [13] presented a critical analysis of different digital inpainting algorithms for still images, and also a comparison of the computational cost of the algorithms. We have considered execution time and memory consumption as metrics to compare computational cost between the algorithms. Raluca Vreja et al [14] published a detailed analytical overview of five advanced inpainting algorithms and measurement benchmarks. They emphasized on the advantages and disadvantages of the used algorithms and also proposed an improved adaptation of the Oliviera's [15] and Hadhoud's [16] inpainting algorithms.

Kunti Patel et al [17] presented a study and analysis of image inpainting algorithms and concluded that exemplar based techniques are generally more effective than PDE based or texture synthesis based techniques. They also extensively listed the merits and demerits of the algorithms, which makes it easy to choose for end users without further research. Anupama Sanjay Awati et al [18] detailed a review of digital image inpainting algorithms, comparing hybrid techniques against commonly used ones. K. Singh et al [19] presented a comparison of patch based inpainting techniques and proposed an adaptive neighborhood selection method for efficient patch inpainting.

III. Theory

OpenCV is a library of programming functions mainly aimed at real-time computer vision. It is a huge open source library for computer vision, machine learning, image and video processing tasks. OpenCV is used in a lot of machine learning problems like face recognition, object detection, image segmentation, etc. mainly due to its simple syntax and presence of a large number of predefined functions and modules.

There are several algorithms present for digital image inpainting, but OpenCV natively provides three of them, INPAINT_TELEA, INPAINT_NS and INPAINT_FSR, which further has two profiles, FSR_FAST and FSR_BEST. INPAINT_TELEA and INPAINT_NS can be accessed by the function cv2.inpaint(). INPAINT_FSR can be accessed by the function cv2.xphoto.inpaint().

```
In [1]: import cv2

In [2]: original = cv2.imread( "original.jpg" )
        mask = cv2.imread( "mask.jpg",0 )
        inpaint_1 = cv2.inpaint(original,mask,3,cv2.INPAINT_TELEA)
        inpaint_2 = cv2.inpaint(original,mask,3,cv2.INPAINT_NS)
        cv2.xphoto.inpaint(original,mask,inpaint_3,cv2.xphoto.
        INPAINT_FSR_FAST)
        cv2.xphoto.inpaint(original,mask,inpaint_4,cv2.xphoto.
        INPAINT_FSR_BEST)
```

Fig. 1: Code to run the inpainting algorithms

This section will contain a brief theory behind the three inpainting algorithms.

a) INPAINT_TELEA

This algorithm is based on the paper "An Image Inpainting Technique based on the Fast Marching method" by Alexandru Telea [10] in 2004. It is based on the Fast Marching Method (FMM), a solutional paradigm which builds a solution outwards starting from the "known information" of a problem. It is a numerical method created by James Sethian for solving boundary value problems of the Eikonal equation [20]. A simple explanation of the working of the algorithm follows, extracted from the original paper [10].

The first and foremost step in any inpainting method is to identify the region to be inpainted. There is the region to be inpainted, also known as the unknown region and the surrounding known region of the image. The algorithm first considers the boundary of the unknown region, which is of infinitesimal width, and inpaints one pixel lying on the boundary. Then it iterates over all the pixels lying on the boundary to inpaint the whole boundary. A single pixel is inpainted as a function of all other pixels lying in its known neighborhood by summing the estimates of all pixels, normalized by a weighting function. A weighting function is necessary as it ensures the inpainted pixel is influenced more by the pixels lying close to it and less by the pixels lying far away. After the boundary has been inpainted, the algorithm propagates forward towards the center of the unknown region.

To implement the propagation, the Fast Marching Method (FMM) is used. FMM ensures the pixels near the known pixels are inpainted first, so that it mimics a manual inpainting technique. The FMM's main

advantage is that it explicitly maintains a narrow band that separates the known from the unknown image area and specifies which pixel to inpaint next.

b) INPAINT_NS

This algorithm provided by OpenCV is established on the paper "Navier-Stokes, Fluid Dynamics, and Image and Video Inpainting" by M. Bertalmio et al [11] in 2001. This algorithm is based on fluid dynamics (fluid dynamics is a sub-discipline of fluid mechanics that describe the flow of fluids: liquids and gases) and utilizes partial differential equations. The method involves a direct solution of the Navier-Stokes equation [21] for an incompressible fluid. A simple explanation of the working of the algorithm follows, extracted from the original paper [11].

The basic principle is heuristic. After the user selects the unknown region, the algorithm first travels along the edges from known regions to unknown regions, and automatically transports information into the inpainting region. The algorithm makes use of isophotes (a line in a diagram connecting points where the intensity of light or brightness is the same). The fill-in is done is such a way that the isophote lines arriving at the unknown region's boundary are completed inside, which allows the smooth continuation of information towards the center of the unknown region. M. Bertalmio et al [11] drew an analogy between the image intensity function of an image and the stream function in a 2D incompressible fluid, and used techniques from the computational fluid dynamics to produce an approximate solution to image inpainting problem.

c) INPAINT_FSR

FSR stands for Rapid Frequency Selective Reconstruction [12]. It is a high quality signal extrapolation algorithm. FSR has proven to be very efficient in the domain of inpainting. The FSR is a powerful approach to reconstruct and inpaint missing areas of an image.

The signal of a distorted block is extrapolated using known samples and already reconstructed pixels as support. This algorithm iteratively generates a generic complex valued model of the signal, which approximates the undistorted samples in the extrapolation area of a particular size as a weighted linear combination of Fourier basic function. The Fourier basic function is a method to smooth out data varying over a continuum (here the unknown region) and exhibiting a cyclical trend. An important feature of FSR algorithm is that the calculations are carried out in the Fourier domain, which leads to fast implementation.

There are two implementations of the FSR inpainting algorithm - INPAINT_FSR_FAST and INPAINT_FSR_BEST. The Fast implementation of FSR provides a great balance between speed and accuracy, and the Best implementation mainly focuses on the accuracy, with speed being slower compared to Fast.

IV. Comparative Study

a) Theoretical Comparison

All the three inpainting algorithms provided by OpenCV are unique and works on different methodologies. The similarity between the algorithms is the inpainting procedure starts with the pixels lying in the boundary of the unknown region, and slowly propagates towards the centre of the unknown region. All the three algorithms are heuristic in nature. The propagation method used in each is different. TELEA uses the Fast Marching Method (FMM), NS uses fluid dynamics equations and FSR extrapolates the pixel values of the unknown region using known samples.

b) Practical Comparison

For practical comparison of the 3 algorithms, we ran some code in Python. Our testing setup had the following specifications:

- CPU : i7-8700K (3.70 GHz)
- RAM : 16 GB (3200 MHz)
- GPU : 8 GB GTX 1080

We took the Kodak image set (which contains 25 uncompressed PNG true colour images of size 768x512 pixels) and four custom error masks for the dataset. We applied all the inpainting algorithms individually over each error mask on the images. We compared the results using four main metrics:

- Peak Signal to Noise Ratio (PSNR): It is the ratio between the maximum possible power of a signal and the power of corrupting noise. To estimate the PSNR of an image, it is necessary to compare the distorted image to an ideal clean image with the maximum possible power. PSNR is commonly used to estimate the efficiency of compressors, filters etc. A higher value of PSNR suggests an efficient manipulation method. In our case, we will compute the PSNR between the original image and the inpainted image. The Python code to calculate PSNR is given in Fig 2.

```
In [1]: import cv2

In [2]: original = cv2.imread( "original.jpg" )
        mask = cv2.imread( "mask.jpg",0 )
        inpainted = cv2.inpaint( original,mask,3,cv2.INPAINT_TELEA )
        psnr = cv2.PSNR( original,inpainted,255 )
```

Fig. 2: PSNR code

- Structural Similarity Index (SSIM): It is a perceptual metric that quantifies image quality degradation caused due to any kind of manipulation on the image. It is an improvement over the use of Mean Squared Error (MSE) to find similarity between images. In our case, we will compute the SSIM between the original image and the inpainted image. A higher value of SSIM suggests the inpainted image is structurally closer to the original

image. The Python code to calculate SSIM is given in Fig 3.

```
In [1]: import cv2
        from skimage.metrics import structural_similarity as SSIM

In [2]: original = cv2.imread( "original.jpg" )
        mask = cv2.imread( "mask.jpg",0 )
        inpainted = cv2.inpaint( original,mask,3,cv2.INPAINT_TELEA )
        ssim = SSIM( original,inpainted,multichannel=True )
```

Fig. 3: SSIM code

- Runtime: It is the total time taken by the algorithm to complete its task. The Python code to calculate the runtime is given in Fig 4.

```
In [1]: import cv2
        import time

In [2]: original = cv2.imread( "original.jpg" )
        mask = cv2.imread( "mask.jpg",0 )
        begin = time.time()
        inpainted = cv2.inpaint( original,mask,3,cv2.INPAINT_TELEA )
        time.sleep(1)
        end = time.time()
        runtime = end-begin
```

Fig. 4: Runtime code

- Memory: It is the total memory consumed by the algorithm while completing the task. We use tracemalloc module, which is a debug tool to trace memory blocks allocated by Python. We find the peak memory usage during the working of the algorithm.

```
In [1]: import cv2
        import tracemalloc

In [2]: original = cv2.imread( "original.jpg" )
        mask = cv2.imread( "mask.jpg",0 )
        tracemalloc.start()
        inpainted = cv2.inpaint( original,mask,3,cv2.INPAINT_TELEA )
        current, peak = tracemalloc.get_traced_memory()
        tracemalloc.stop()
        memory = peak/10**6
```

Fig. 5: Memory code

All the values have been taken up to three decimal places. Apart from the four main metrics, we also considered two hybrid metrics defined in Section V. We also curated some custom images for testing of certain specific cases. The results obtained are given in the next section, along with their critical explanation.

V. Results and Discussion

a) *Kodak image dataset results*

There are 19 landscape and 6 portrait oriented photos in the Kodak image set. We initially made the custom error masks for landscape orientation, and rotated them to fit the portrait orientation. We chose striped masks as the error regions are equally distributed. The four custom error masks we considered are:

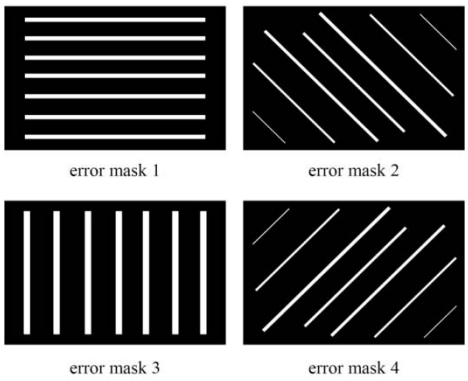


Fig. 6: Four custom error masks

The white stripes are the areas to be inpainted. We have displayed the image results for just 1 landscape photo (2 error masks) and 1 portrait photo (2 error masks). These are the following results we obtained:-

Sample_1 (Landscape)

The original, distorted and 4 inpainted results of the first image sample over the first error mask are given in Fig 7.

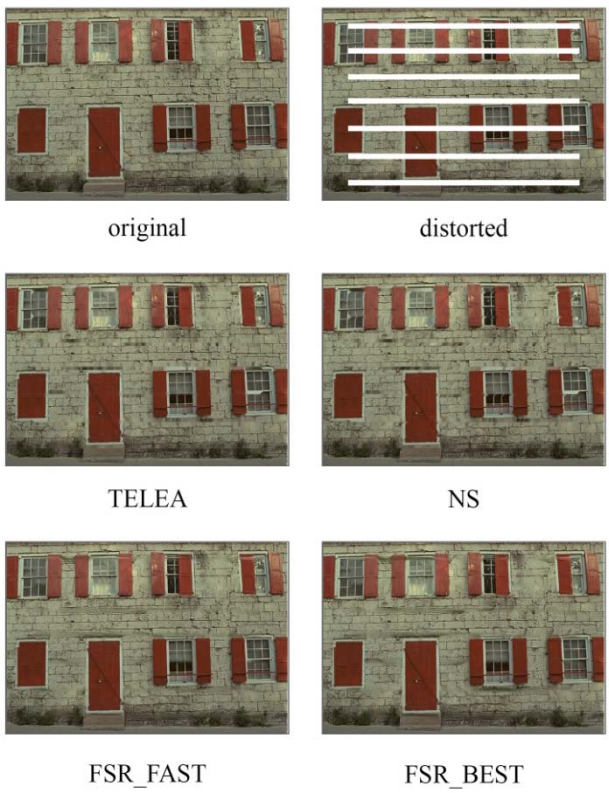


Fig. 7: First image sample results over first error mask

The metric values calculated for the first image sample over the first error mask are given in Table 1.

Table 1: First image sample metrics over first error mask

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	27.218	26.713	26.612	26.315
SSIM	0.889	0.889	0.887	0.884
Runtime [s]	5.349	96.199	1.089	1.085
Memory [MB]	1.339	1.339	1.339	1.339

The original, distorted and 4 in painted results of the first image sample over the second error mask are given in Fig 8.

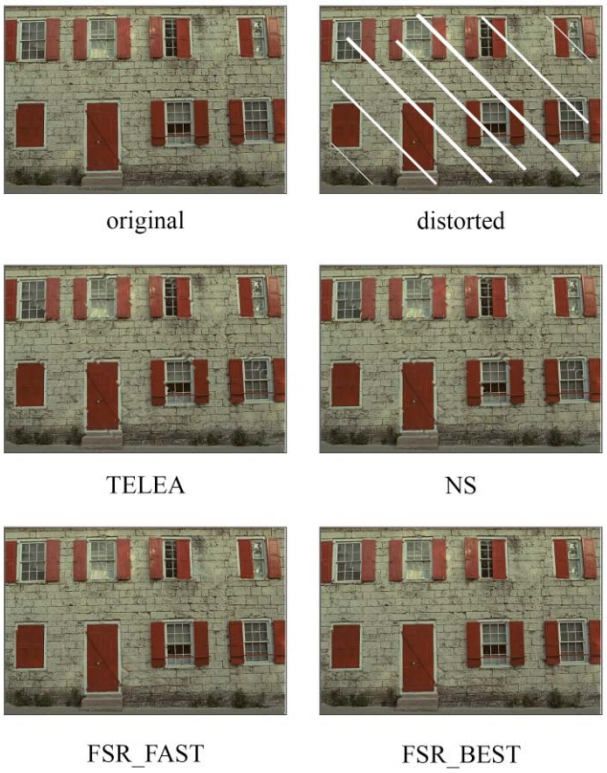


Fig. 8: First image sample results over second error mask

The metric values calculated for the first image sample over the second error mask are given in Table 2.

Table 2: First image sample metrics over second error mask

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	34.577	34.764	30.556	30.689
SSIM	0.974	0.975	0.950	0.952
Runtime [s]	2.887	35.132	1.049	1.047
Memory [MB]	1.339	1.339	1.339	1.339

We have not given the image results for the third and fourth error masks, only the metric values. The metric values calculated for the first image sample over the third error mask are given in Table 3.

Table 3: First image sample metrics over third error mask

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	27.143	27.229	27.096	26.796
SSIM	0.890	0.892	0.888	0.883
Runtime [s]	5.210	97.800	1.091	1.091
Memory [MB]	1.339	1.339	1.339	1.339

The metric values calculated for the first image sample over the fourth error mask are given in Table 4.

Table 4: First image sample metrics over fourth error mask

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	34.788	34.897	30.709	31.209
SSIM	0.975	0.976	0.952	0.955
Runtime [s]	3.114	39.762	1.055	1.049
Memory [MB]	1.339	1.339	1.339	1.339

Sample_2 (Portrait)

For portrait images, the error masks have been rotated 90 degree clockwise to fit the orientation. Given are the original, distorted and 4 in painted results of the second image sample over the first error mask in Fig 9.

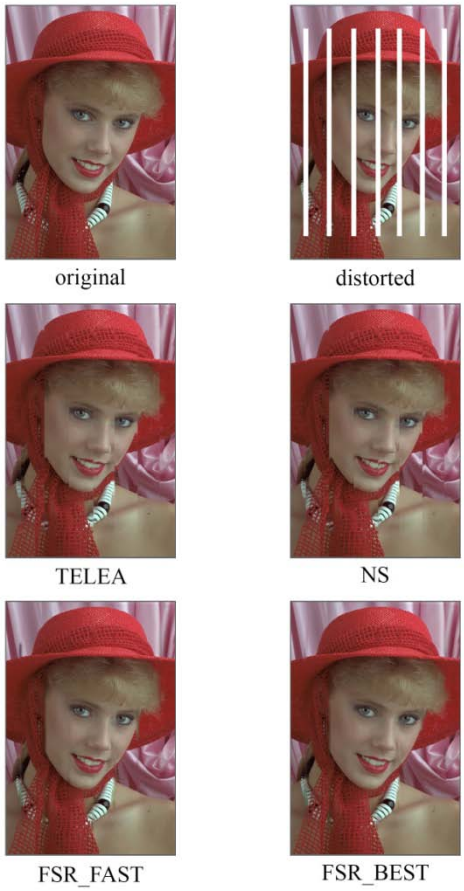


Fig. 9: Second image sample results over first error mask

The metric values calculated for the second image sample over the first error mask are given in Table 5.

Table 5: Second image sample metrics over first error mask

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	31.739	32.738	31.398	31.133
SSIM	0.936	0.941	0.934	0.932
Runtime [s]	4.574	65.644	1.092	1.088
Memory [MB]	1.339	1.339	1.339	1.339

The original, distorted and 4 in painted results of the second image sample over the second error mask are given in Fig 10.

Fig. 10: Second image sample results over second error mask

The metric values calculated for the second image sample over the first error mask are given in Table 6.

Table 6: Second image sample metrics over second error mask

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	39.632	39.673	36.063	35.106
SSIM	0.983	0.983	0.972	0.971
Runtime [s]	2.568	23.655	1.044	1.052
Memory [MB]	1.339	1.339	1.339	1.339

We have not given the image results for the third and fourth error masks, only the metric values. The metric values calculated for the second image sample over the third error mask are given in Table 7.

Table 7: Second image sample metrics over third error mask

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	32.105	31.834	30.596	30.268
SSIM	0.933	0.936	0.929	0.927
Runtime [s]	4.334	67.520	1.089	1.085
Memory [MB]	1.339	1.339	1.339	1.339

The metric values calculated for the second image sample over the fourth error mask are given in Table 8.

Table 8: Second image sample metrics over fourth error mask

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	39.689	39.532	37.029	37.092
SSIM	0.985	0.985	0.975	0.975
Runtime [s]	2.634	25.407	1.056	1.057
Memory [MB]	1.339	1.339	1.339	1.339

We applied the in painting algorithms to all the 25 images present in the dataset. The average metric values for first, second, third and fourth error masks are given in tables 9,10,11,12 respectively.

Table 9: Average metric values for first error mask

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	29.719	30.017	29.145	28.891
SSIM	0.929	0.932	0.925	0.923
Runtime [s]	4.346	72.229	1.089	1.091
Memory [MB]	1.339	1.339	1.339	1.339

Table 10: Average metric values for second error mask

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	34.002	37.734	33.421	33.406
SSIM	0.952	0.983	0.968	0.969
Runtime [s]	3.546	27.488	1.047	1.048
Memory [MB]	1.339	1.339	1.339	1.339

Table 11: Average metric values for third error mask

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	28.948	29.143	28.602	28.376
SSIM	0.925	0.927	0.922	0.920
Runtime [s]	4.282	73.604	1.095	1.093
Memory [MB]	1.339	1.339	1.339	1.339

Table 12: Average metric values for fourth error mask

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	37.831	38.039	34.088	33.958
SSIM	0.983	0.984	0.970	0.969
Runtime [s]	2.725	31.751	1.052	1.051
Memory [MB]	1.339	1.339	1.339	1.339

Generally speaking, lower memory consumption and runtime values mean a better algorithm. For other metrics, the higher the PSNR and SSIM value, the better the algorithm. The average memory consumption, as seen from Table 9,10,11,12 is same for any mask on any image for any algorithm for the particular dataset. Hence we will not consider it as a factor for deciding the most efficient algorithm. We have defined two hybrid metrics X and Y for deciding which algorithm is most efficient based on our data. Metric X is directly proportional to PSNR, directly proportional to SSIM and inversely proportional to Runtime value:

$$X \propto PSNR$$
$$X \propto SSIM$$
$$X \propto (1/Runtime)$$

Combining all three above equations we get:

$$X \propto (PSNR * SSIM)/Runtime$$
$$X = k * ((PSNR * SSIM)/Runtime)$$

where k is a constant, taken to be 1 for comparison purposes. Hence

$$X = (PSNR * SSIM)/Runtime$$

A high value of metric X means an effective algorithm. We used the values obtained in Table 9,10,11,12 and calculated metric X values for the four error masks. The values are given in Table 13.

Table 13: X metric values for four error masks

	FSR		TELEA	NS
	Fast	Best		
1st Error Mask	6.353	0.387	24.756	24.442
2nd Error Mask	9.129	1.349	30.899	30.888
3rd Error Mask	6.253	0.367	24.083	23.885
4th Error Mask	13.647	1.179	31.431	31.309

From Table 13, we can see that TELEA algorithm gets the highest value in all four error masks. Hence, TELEA is the most efficient in painting algorithm when we consider metric X to be the comparison metric.

But as we can infer from the definition of metric X, it has the runtime factor associated with it. Runtime is an important factor for analysing algorithms, but can be subjective at times to different end users. Some users may have a time constraint, some users may not. Hence we need to define such a metric which does not include the runtime factor. Therefore, we define metric Y. Metric Y is directly proportional to PSNR and directly proportional to SSIM value:

$$Y \propto PSNR$$
$$Y \propto SSIM$$

Combining all two above equations we get:

$$Y \propto PSNR * SSIM$$
$$Y = k * (PSNR * SSIM)$$

where k is a constant, taken to be 1 for comparison purposes. Hence

$$Y = PSNR * SSIM$$

A high value of metric Y means an effective algorithm, without taking the runtime factor into account. Similarly, we used the values in Table 9,10,11,12 and calculated metric Y values for the four error masks. The values are given in Table 14.

Table 14: Y metric values for four error masks

	FSR		TELEA	NS
	Fast	Best		
1st Error Mask	27.609	27.976	26.959	26.666
2nd Error Mask	32.369	37.093	32.352	32.370
3rd Error Mask	26.779	27.016	26.371	26.106
4th Error Mask	37.188	37.430	33.065	32.905

From Table 14, we can see that FSR_BEST algorithm gets the highest value in all four error masks. Hence, FSR_BEST is the most efficient inpainting algorithm when we consider metric Y to be the comparison metric, which does not take the runtime factor into account.

Summing up our observation and results for the Kodak image dataset, we can say that the most efficient inpainting algorithm when runtime is a constraint is TELEA algorithm and the most efficient inpainting algorithm when runtime is not a constraint is FSR_BEST algorithm.

b) Edge inpainting results

The inpainting algorithms produce very different results when working on edges. To compare the working, we have chosen an image which has clear distinct foreground and background. We distorted a part of the edge, and applied the inpainting algorithms to it. The image results are given in Fig 11, and metric values are given in Table 15.

As we can see from the results, TELEA has the highest value for X metric. That is if we consider runtime to be a factor, TELEA is the most efficient algorithm. But FSR_BEST has the highest value for Y metric, i.e. if we

do not consider runtime to be a factor, then FSR_BEST is the most efficient algorithm for edge inpainting. We can also see from the image results that FSR_BEST produces the most believable result, but also has the largest runtime. TELEA and NS do a decent job in filling up the edges and maintaining the edge difference. But still some parts are hazed and distorted. FSR_FAST does the worst job, mainly because it trades off accuracy for runtime, and the result is bad.

Table 15: Metric values for edge inpainting

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	35.249	43.100	38.577	37.973
SSIM	0.995	0.996	0.994	0.994
Runtime [s]	1.642	21.282	1.028	1.023
Memory [MB]	2.079	2.079	2.079	2.079
X	21.359	2.017	37.301	36.897
Y	35.073	42.928	38.346	37.745

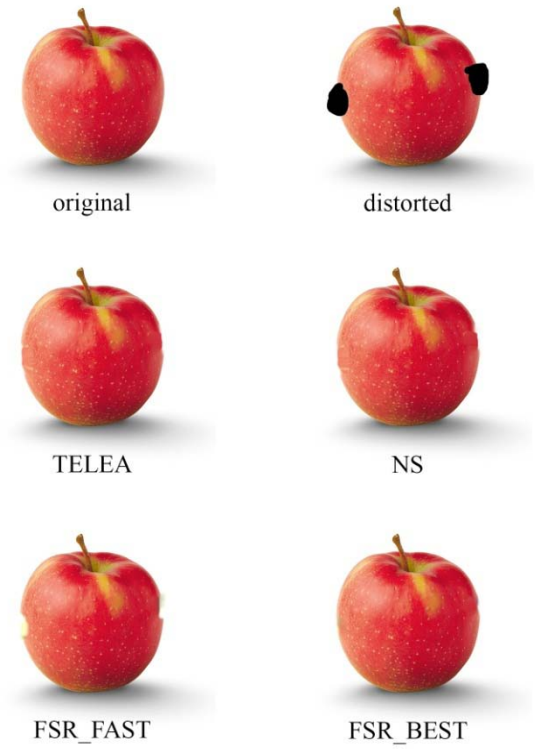


Fig. 11: Edge inpainting sample results

c) Pattern inpainting results

The inpainting algorithms produce very different results when working on patterns. To compare the working, we have chosen a checkerboard image as it is the easiest pattern to replicate. We distorted a part of the image in the centre, and applied the inpainting algorithms to it. The image results are given in Fig 12, and metric values are given in Table 16.

As we can see from the image results, none of the inpainting algorithms can replicate the pattern in the unknown region, which is understandable because the inpainting algorithms are focused on filling up the unknown region progressively based on information from the nearest known region. They work on the small scale spatial influences. In order to inpaint a pattern, the algorithm must work over a broad range of the known region to understand the dynamics of the pattern. An exemplar based inpainting or patch based inpainting method can work for pattern inpainting.

Comparing the metric values, TELEA has the highest X value and FSR_BEST has the highest Y value. From the image results, TELEA still does a decent job of producing an arbitrary pattern, while FSR_BEST fills the whole unknown region with a singular colour. Hence, no algorithm provided by OpenCV is perfectly suitable for inpainting a pattern, but TELEA can be used as a last resort.

Table 16: Metric values for pattern inpainting

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	20.094	21.881	19.017	18.872
SSIM	0.959	0.964	0.962	0.958
Runtime [s]	2.827	60.066	1.051	1.047
Memory [MB]	1.188	1.188	1.188	1.188
X	6.816	0.351	17.407	17.268
Y	19.270	21.093	18.294	18.079

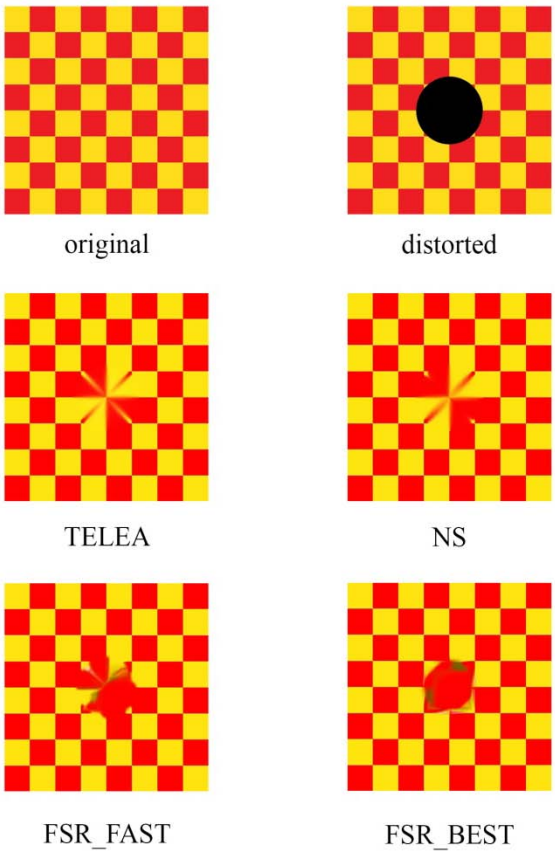


Fig. 12: Pattern inpainting sample results

d) Text error mask inpainting results

We also tested the working of the inpainting algorithms on a custom text error mask. We took an image from the Kodak dataset and wrote some random text on it as error regions, then applied the algorithms on it. The image results are given in Fig 13, and metric values are given in Table 17.

Comparing the metric values, NS has the highest X value and FSR_BEST has the highest Y value. All the algorithms work decent, but from the image results we can see that TELEA and NS have some distortions near the fence area, while FSR_FAST and FSR_BEST have inpainted smoothly in that area. If runtime is a constraint, then NS is the most effective algorithm to be used. Although, TELEA can also be used as it produces very similar results to NS. If runtime is not a constraint, then FSR_BEST is the most effective choice for text error mask inpainting.

Table 17: Metric values for text error mask inpainting

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	45.019	45.497	30.784	31.571
SSIM	0.995	0.996	0.962	0.967
Runtime [s]	4.019	36.729	2.659	2.133
Memory [MB]	1.339	1.339	1.339	1.339
X	11.146	1.234	11.137	14.313
Y	44.794	45.315	29.614	30.529

Fig. 13: Text error mask inpainting sample results

e) Monochromatic image inpainting results

We tested the working of the inpainting algorithms on a monochromatic image. We took an image from the Kodak dataset and converted it into monochrome and used a spiral error mask on it. The image results are given in Fig 14, and metric values are given in Table 18.

Comparing the metric values, NS has the highest X value and FSR_BEST has the highest Y value. All the algorithms work decent, but from the image results, we see that TELEA and NS have some distortions near the beak of the bird, while FSR_FAST and FSR_BEST have inpainted smoothly in that area. If runtime is a constraint, then NS is the most effective algorithm to be used. Although, TELEA can also be used as it produces very similar results to NS. If runtime is not a constraint, then FSR_BEST is the most effective choice for monochromatic image inpainting.

Table 18: Metric values for monochrome image inpainting

	FSR		TELEA	NS
	Fast	Best		
PSNR [dB]	45.649	46.040	36.509	36.417
SSIM	0.997	0.998	0.987	0.988
Runtime [s]	1.744	13.549	1.163	1.097
Memory [MB]	1.339	1.339	1.339	1.339
X	26.096	3.391	30.984	32.799
Y	45.512	45.948	36.034	35.979

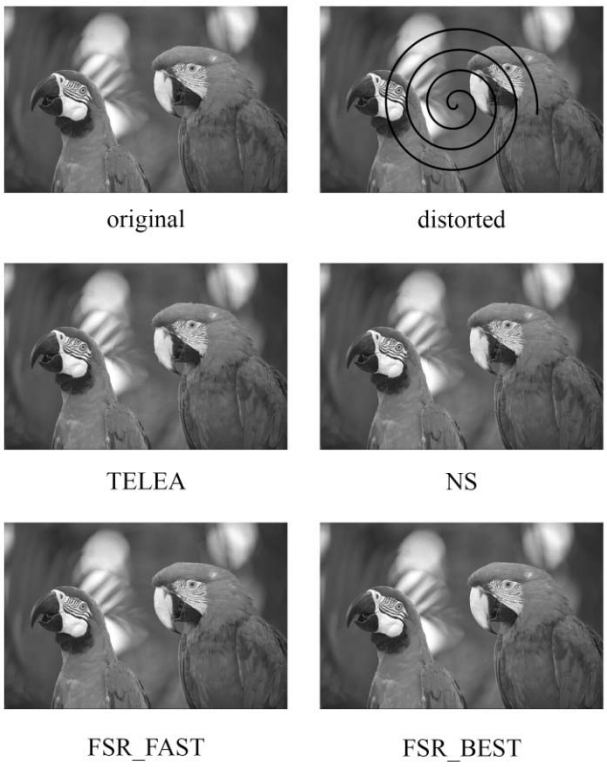


Fig. 14: Monochromatic image inpainting sample results

f) Discussions

Summing up our observations and results for the Kodak image dataset and other specific cases, we can say that the TELEA inpainting algorithm is the most efficient algorithm if runtime is a constraint i.e. the user needs to perform the inpainting operation as fast as he can and produce the best results. On the other hand, FSR_BEST inpainting algorithm is the most efficient algorithm if runtime is not a constraint i.e. the user has no time limit for the inpainting operation and wants to get the best result. The average memory consumption for all the inpainting algorithms are modest, hence memory will hardly be an issue in any system while running the Open CV inpainting algorithms.

Table 19: Final results

	With runtime as a constraint	Without runtime as a constraint
Most effective OpenCV inpainting algorithm	TELEA algorithm	FSR_BEST algorithm

VI. Future Scope

Our inpainting comparison study was done on the Kodak image dataset, a relatively small dataset containing 25 images only. The study can be done on a larger, more robust dataset which contains variety of images. This can be done to get more extensive results. We compared our results on the basis of four metrics only; more intricate metrics may be defined for the testing. Our study can be a base for analysing how various OpenCV inpainting methods work on images with different colour profiles.

We ran tests using four custom error masks. The error masks considered were mostly linear in shape. Other type of error masks such as curved, mixture of linear and curved can be taken for testing. This study can be a base for a comprehensive study on video inpainting techniques, which would be beneficial for people looking to work in this field.

VII. Conclusion

In conclusion, we present a comparative study of the various OpenCV inpainting algorithms, focusing extensively on their practical uses. The purpose of this paper is to apprise new users and researchers of the most efficient inpainting algorithm provided by OpenCV: TELEA algorithm for time constrained operations and FSR_BEST algorithm for non time constrained operations. We present the most efficient OpenCV inpainting algorithm to be used for various scenarios, which can help a beginner at inpainting to make his decision wisely without any further research. This study can be a base for more detailed comparative works on image and video inpainting. Inpainting is an evolving domain of image processing with major strides being made in the past, and much more sophisticated algorithms yet to arrive. It opens up the doorway for new image processing researchers to better the existing algorithms and create finer advanced inpainting algorithms which achieve near perfect accuracy.

References Références Referencias

1. R. S. Hegadi, "*Image Processing: Research Opportunities and Challenges*", In the Proceedings of National Seminar on Research in Computers, 2010.
2. R. Farhan, "*A Review on Some Methods used in Image Restoration*", International Multidisciplinary Research Journal, Vol. 10, pp.13-16, 2020.
3. O. Elharrous, N. Almaadeed, S. Al-Maadeed, Y. Akbari "*Image inpainting: A review*", Neural Processing Letters, Vol. 51, pp.2007-2028, 2019.
4. A. Beniwal, D. Ahlawat, "*Image Inpainting Algorithms: A Survey*", International Journal of Recent Research Aspects, Vol. 3/Issue 2, pp.165-169, 2016.
5. F. Zhang, Y. Chen, Z. Xiao, L. Geng, J. Wu, T. Feng, P. Liu, Y. Tan, J. Wang, "*Partial Differential Equation Inpainting Method Based on Image Characteristics*", Image and Graphics: Lecture Notes in Computer Science, Vol. 9219, pp.11-19, 2015.
6. T. Zhou, B. Johnson, R. Li, "*Patch-based Texture Synthesis for Image Inpainting*", arXiv: 1605. 01576v1, 2016.
7. Sreelakshmy I J, B. C. Kovoor, "*A Hybrid Inpainting Model Combining Diffusion and Enhanced Exemplar Methods*", Journal of Data and Information Quality, Vol. 13/Issue 3, pp.1-17, 2021.
8. S. R. Gaonkar, P. D. Hire, P. S. Pimple, Y. R. Kotwal, B. A. Ahire, "*Image Inpainting using Robust Exemplar-based Technique*", International Journal of Computer Sciences and Engineering, Vol. 2/Issue 4, pp.176-179, 2014.
9. J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, T. S. Huang, "*Generative Image Inpainting with Contextual Attention*", In the Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.
10. A. Telea, *"An Image Inpainting Technique Based on the Fast Marching Method*", Journal of Graphics Tools, Vol. 9/Issue 1, pp.23-34, 2004.
11. M. Bertalmio, A. L. Bertozzi, G. Sapiro, "*Navier-Stokes, Fluid Dynamics, and Image and Video Inpainting*", In the Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001.
12. N. Genser, J. Seiler, M. Jonscher, A. Kaur, "*Demonstration of Rapid Frequency Selective Reconstruction for Image Resolution Enhancement*", In the Proceedings of IEEE International Conference on Image Processing, 2017.

13. S. Chhabra, R. Lalit, S. Saxena, "*An Analytical Study of Different Image Inpainting Techniques*", Indian Journal of Computer Science and Engineering, Vol. 3/Issue 3, pp.487-491, 2012.
14. R. Vreja, R. Brad, "*Image Inpainting Methods Evaluation and Improvement*", The Scientific World Journal, Vol. 2014, 2014.
15. M. M. Oliveira, B. Bowen, R. McKenna, Y. Chang, "*Fast Digital Image Inpainting*", In the Proceedings of International Conference on Visualization, Imaging and Image Processing, 2001.
16. M. M. Hadhoud, K. A. Moustafa, S. Shenoda, "*Digital Images Inpainting using Modified Convolution Based Method*", In Optical Pattern Recognition XX, Vol. 7340 of Proceedings of the SPIE, 2009.
17. K. Patel, A. Yerpude, "*Study and Analysis of Image Inpainting Algorithms*", International Journal of Engineering Research and Technology, Vol. 3/Issue 20, pp.1-5, 2015.
18. A. S. Awati, M. R. Patil, "*Review of Digital Image Inpainting Algorithms*", International Journal of Latest Trends in Engineering and Technology, pp.36-41, 2013.
19. K. Singh, J. Shaveta, "*A Review on Patch Based Image Restoration or Inpainting*", International Journal of Computer Sciences and Engineering, Vol. 5/Issue 3, pp.119-123, 2017.
20. J. A. Sethian, "*Fast Marching Methods*", SIAM Review, Vol. 41/Issue 2, pp.199-235, 1999.
21. S. Algazin, "*Numerical Study of Navier-Stokes equations*", Journal of Applied Mechanics and Technical Physics, Vol. 48/Issue 5, pp.656-663, 2007.

This page is intentionally left blank

Global Journal of Computer Science and Technology: G Interdisciplinary
Volume 21 Issue 2 Version 1.0 Year 2021
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Unusual Features of the SARS-CoV-2 Genome Suggesting Sophisticated Laboratory Modification as a Biological Robot Rather than Natural Evolution and Delineation of Its Probable Synthetic Route

By Li-Meng Yan & Adrian David Cheok

Professional University of Information and Management for Innovation (iUniversity)

Abstract- The COVID-19 pandemic caused by the novel coronavirus SARS-CoV-2 has led to over 4.24 million deaths worldwide and unprecedented decimation of the global economy. Despite its tremendous impact, the origin of SARS-CoV-2 has remained mysterious and controversial. The natural origin theory, although widely accepted, lacks substantial support. The alternative theory that the virus may have come from a research laboratory is, however, strictly censored on peer-reviewed scientific journals. Nonetheless, SARS-CoV-2 shows biological characteristics that are inconsistent with a naturally occurring, zoonotic virus. In this report, we describe the genomic, structural, medical, and literature evidence, which, when considered together, strongly contradicts the natural origin theory. The evidence shows that SARS-CoV2 should be a laboratory product created by using bat coronaviruses ZC45 and/or ZXC21 as a template and/or backbone.

GJCST-G Classification: J.3

Unusual Features of the SARS-CoV-2 Genome Suggesting Sophisticated Laboratory Modification as a Biological Robot Rather than Natural Evolution and Delineation of Its Probable Synthetic Route

Li-Meng Yan [α] & Adrian David Cheok [σ]

Abstract- The COVID-19 pandemic caused by the novel coronavirus SARS-CoV-2 has led to over 4.24 million deaths worldwide and unprecedented decimation of the global economy. Despite its tremendous impact, the origin of SARS-CoV-2 has remained mysterious and controversial. The natural origin theory, although widely accepted, lacks substantial support. The alternative theory that the virus may have come from a research laboratory is, however, strictly censored on peer-reviewed scientific journals. Nonetheless, SARS-CoV-2 shows biological characteristics that are inconsistent with a naturally occurring, zoonotic virus. In this report, we describe the genomic, structural, medical, and literature evidence, which, when considered together, strongly contradicts the natural origin theory. The evidence shows that SARS-CoV2 should be a laboratory product created by using bat coronaviruses ZC45 and/or ZXC21 as a template and/or backbone. Building upon the evidence, we further postulate a synthetic route for SARS-CoV-2, demonstrating that the laboratory-creation of this coronavirus is convenient and can be accomplished in approximately six months. Our work emphasizes the need for an independent investigation into the relevant research laboratories. It also argues for a critical look into certain recently published data, which, albeit problematic, was used to support and claim a natural origin of SARS-CoV-2. From a public health perspective, these actions are necessary as knowledge of the origin of SARS-CoV-2 and of how the virus entered the human population are of pivotal importance in the fundamental control of the COVID-19 pandemic as well as in preventing similar, future pandemics.

I. Introduction

COVID-19 has caused a world-wide pandemic, the scale and severity of which are unprecedented. Despite the tremendous efforts taken by the global community, management and control of this pandemic remains difficult and challenging.

As a coronavirus, SARS-CoV-2 differs significantly from other respiratory and/or zoonotic viruses: it attacks multiple organs; it is capable of undergoing a long period of asymptomatic infection; it is highly transmissible and significantly lethal in high-risk populations; it is well-adapted to humans since the very start of its emergence[1]; it is highly efficient in binding the human ACE2 receptor (hACE2), the affinity of which is greater than that associated with the ACE2 of any other potential host[2,3].

The origin of SARS-CoV-2 is still the subject of much debate. A widely cited *Nature Medicine* publication has claimed that SARS-CoV-2 most likely came from nature[4]. However, the article and its central conclusion are now being challenged by scientists from all over the world[5-15]. In addition, authors of this *Nature Medicine* article show signs of conflict of interests[16,17], raising further concerns on the credibility of this publication.

The existing scientific publications supporting a natural origin theory rely heavily on a single piece of evidence – a previously discovered bat coronavirus named RaTG13, which shares a 96% nucleotide sequence identity with SARS-CoV-2[18]. However, the existence of RaTG13 in nature and the truthfulness of its reported sequence are being widely questioned[6-9,19-21]. It is noteworthy that scientific journals have clearly censored any dissenting opinions that suggest a non-natural origin of SARS-CoV-2[8,22]. Because of this censorship, articles questioning either the natural origin of SARS-CoV-2 or the actual existence of RaTG13, although of high quality scientifically, can only exist as preprints[5-9,19-21] or other non-peer reviewed articles published on various online platforms[10-13,23]. Nonetheless, analyses of these reports have repeatedly pointed to severe problems and a probable fraud associated with the reporting of RaTG13[6,8,9,1921]. Therefore, the theory that fabricated scientific data has been published to mislead the world's efforts in tracing the origin of SARS-CoV-2 has become substantially convincing and is interlocked with the notion that SARS-CoV-2 is of a non-natural origin.

Consistent with this notion, genomic, structural, and literature evidence also suggest a non-natural origin of SARS-CoV-2. In addition, abundant literature

Author α: Rule of Law Society & Rule of Law Foundation, New York, NY, USA. e-mail: team.lmyan@gmail.com
Author σ: Professional University of Information and Management for Innovation (iUniversity), Tokyo, Japan. e-mail: adrian@i-u.ac.jp

indicates that gain-of-function research has long advanced to the stage where viral genomes can be precisely engineered and manipulated to enable the creation of novel coronaviruses possessing unique properties. In this report, we present such evidence and the associated analyses. Part 1 of the report describes the genomic and structural features of SARS-CoV2, the presence of which could be consistent with the theory that the virus is a product of laboratory modification beyond what could be afforded by simple serial viral passage. Part 2 of the report describes a highly probable pathway for the laboratory creation of SARS-CoV-2, key steps of which are supported by evidence present in the viral genome. Importantly, part 2 should be viewed as a demonstration of how SARS-CoV-2 could be conveniently created in a laboratory in a short period of time using available materials and well-documented techniques. This report is produced by a team of experienced scientists using our combined expertise in virology, molecular biology, structural biology, computational biology, vaccine development, and medicine.

a) Has SARS-CoV-2 been subjected to in vitro manipulation?

We present three lines of evidence to support our contention that laboratory manipulation is part of the history of SARS-CoV-2:

i. The genomic sequence of SARS-CoV-2 is suspiciously similar to that of a bat coronavirus discovered by military laboratories in the Third Military Medical University (Chongqing, China) and the Research Institute for Medicine of Nanjing Command (Nanjing, China).

ii. The receptor-binding motif (RBM) within the Spike protein of SARS-CoV-2, which determines the host specificity of the virus, resembles that of SARS-CoV from the 2003 epidemic in a suspicious manner. Genomic evidence suggests that the RBM has been genetically manipulated.

iii. SARS-CoV-2 contains a unique furin-cleavage site in its Spike protein, which is known to greatly enhance viral infectivity and cell tropism. Yet, this cleavage site is completely absent in this particular class of coronaviruses found in nature. In addition, rare codons associated with this additional sequence suggest the strong possibility that this furin-cleavage site is not the product of natural evolution and could have been inserted into the SARS-CoV-2 genome artificially by techniques other than simple serial passage or multi-strain recombination events inside co-infected tissue cultures or animals.

i. *Genomic sequence analysis reveals that ZC45, or a closely related bat coronavirus, should be the backbone used for the creation of SARS-CoV-2*

The structure of the ~30,000 nucleotides-long SARS-CoV-2 genome is shown in Figure 1. Searching the NCBI sequence database reveals that, among all known coronaviruses, there were two related bat coronaviruses, ZC45 and ZXC21, that share the highest sequence identity with SARS-CoV-2 (each bat coronavirus is ~89% identical to SARS-CoV-2 on the nucleotide level). Similarity between the genome of SARS-CoV-2 and those of representative β coronaviruses is depicted in Figure 1. ZXC21, which is 97% identical to and shares a very similar profile with ZC45, is not shown. Note that the RaTG13 virus is excluded from this analysis given the strong evidence suggesting that its sequence may have been fabricated and the virus does not exist in nature[2,6-9].

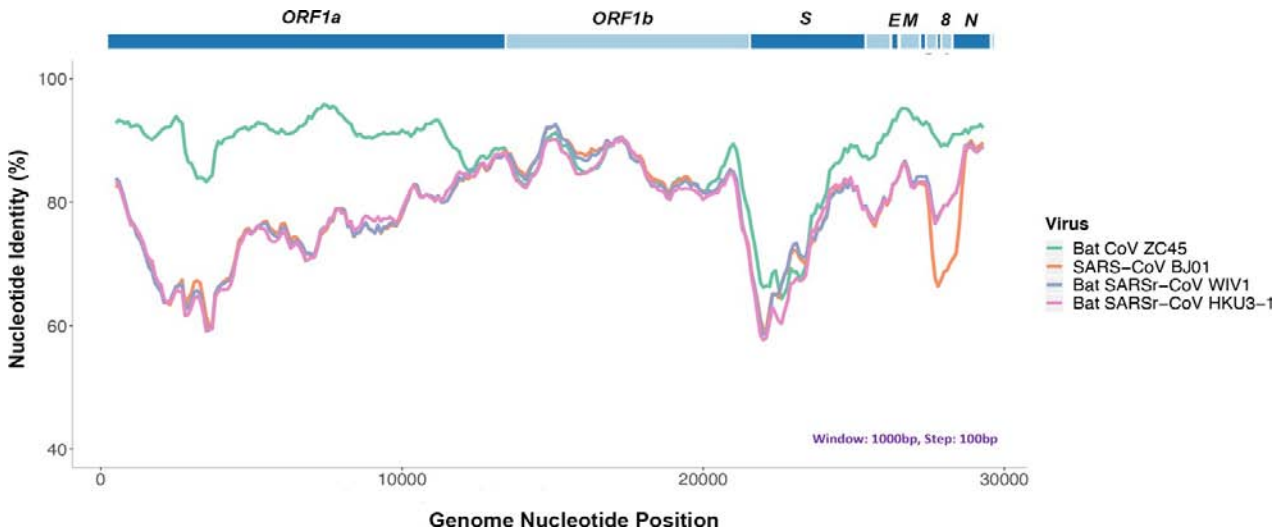


Figure 1: Genomic sequence analysis reveals that bat coronavirus ZC45 is the closest match to SARS-CoV-2. Top: genomic organization of SARS-CoV-2 (2019-nCoV WIV04). Bottom: similarity plot based on the full-length genome of 2019-nCoV WIV04. Full-length genomes of SARS-CoV BJ01, bat SARSr-CoV WIV1, bat SARSr-CoV HKU3-1, bat coronavirus ZC45 were used as reference sequences.

When SARS-CoV-2 and ZC45/ZXC21 are compared on the amino acid level, a high sequence identity is observed for most of the proteins. The Nucleocapsid protein is 94% identical. The Membrane protein is 98.6% identical. The S2 portion (2nd half) of the Spike protein is 95% identical. Importantly, the Orf8protein is 94.2% identical and the E protein is 100% identical.

Orf8 is an accessory protein, the function of which is largely unknown in most coronaviruses, although recent data suggests that Orf8 of SARS-CoV-2 mediates the evasion of host adaptive immunity by down regulating MHC-I[24]. Normally, Orf8 is poorly conserved in coronaviruses[25]. Sequence blast indicates that, while the Orf8 proteins of ZC45/ZXC21 share a 94.2% identity with SARS-CoV-2 Orf8, no other coronaviruses share more than 58% identity with SARS-CoV-2 on this particular protein. The very high homology here on the normally poorly conserved Orf8 protein is highly unusual.

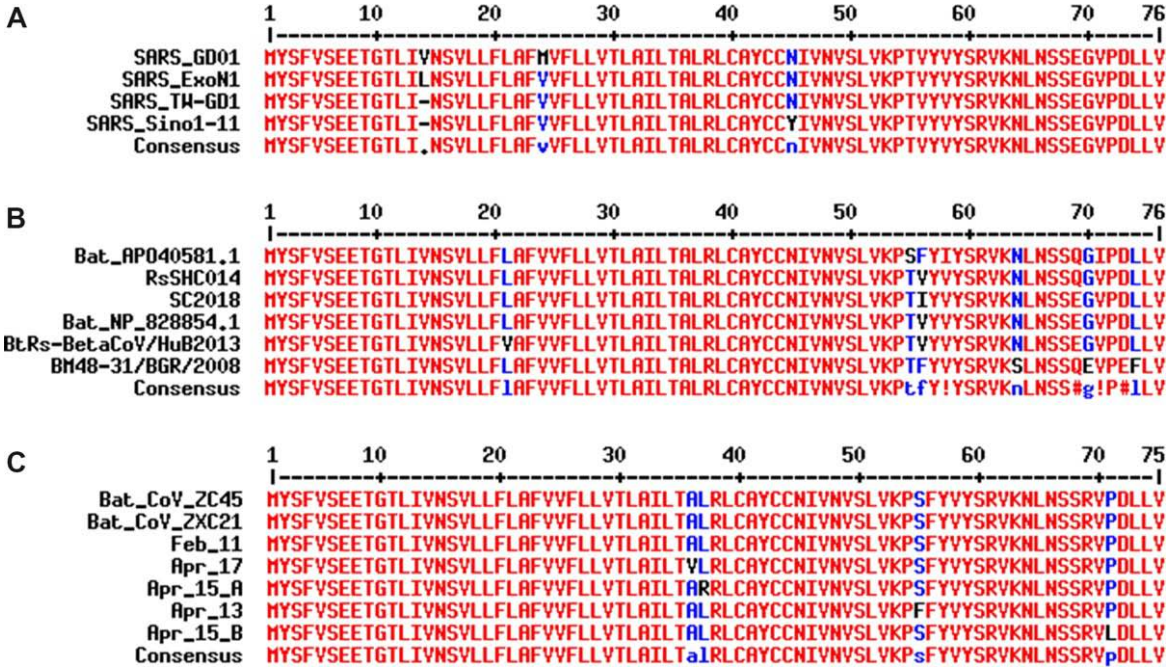


Figure 2: Sequence alignment of the E proteins from different β coronaviruses demonstrates the E protein's permissiveness and tendency toward amino acid mutations. A. Mutations have been observed in different strains of SARS-CoV. GenBank accession numbers: SARS_GD01: AY278489.2, SARS_ExoN1: ACB69908.1, SARS_TW_GD1: AY451881.1, SARS_Sino1_11: AY485277.1. B. Alignment of E proteins from related bat coronaviruses indicates its tolerance of mutations at multiple positions. GenBank accession numbers: Bat_AP040581.1: APO40581.1, RsSHC014: KC881005.1, SC2018: MK211374.1, Bat_NP_828854.1: NP_828854.1, BtRs-BetaCoV/HuB2013: AIA62312.1, BM48-31/BGR/2008: YP_003858586.1. C. While the early copies of SARS-CoV-2 share 100% identity on the E protein with ZC45 and ZXC21, sequencing data of SARS-CoV2 from April 2020 indicates that mutation has occurred at multiple positions. Accession numbers of viruses: Feb_11: MN997409, ZC45: MG772933.1, ZXC21: MG772934, Apr_13: MT326139, Apr_15_A: MT263389, Apr_15_B: MT293206, Apr_17: MT350246. Alignments were done using the MultAlin web server (http://multalin.toulouse.inra.fr/multalin/).

The coronavirus E protein is a structural protein, which is embedded in and lines the interior of the membrane envelope of the virion[26]. The E protein is tolerant of mutations as evidenced in both SARS (Figure 2A) and related bat coronaviruses (Figure 2B). This tolerance to amino acid mutations of the E protein is further evidenced in the current SARS-CoV-2 pandemic. After only a short two-month spread of the virus since its outbreak in humans, the E proteins in SARS-CoV-2 have already undergone mutational changes. Sequence data obtained during the month of April reveals that mutations have occurred at four different locations in different strains (Figure 2C). Consistent with this finding, sequence blast analysis indicates that, with the exception of SARS-CoV-2, no known coronaviruses share 100% amino acid sequence identity on the E protein with ZC45/ZXC21 (*suspicious coronaviruses published after the start of the current pandemic are excluded*[18,27-31]). Although 100% identity on the E protein has been observed between SARS-CoV and certain SARS-related bat coronaviruses, none of those pairs simultaneously share over 83% identity on the Orf8 protein[32]. Therefore, the 94.2% identity on the Orf8 protein, 100% identity on the E protein, and the overall genomic/amino acid-level resemblance between SARS-CoV-2 and ZC45/ZXC21 are highly unusual. Such

evidence, when considered together, is consistent with a hypothesis that the SARS-CoV-2 genome has an origin based on the use of ZC45/ZXC21 as a backbone and/or template for genetic gain-of-function modifications.

Importantly, ZC45 and ZXC21 are bat corona viruses that were discovered (between July 2015 and February 2017), isolated, and characterized by military research laboratories in the Third Military MedicalUniversity (Chongqing, China) and the Research Institute for Medicine of Nanjing Command (Nanjing, China). The data and associated work were published in 2018[33,34]. Clearly, this backbone/template, which is essential for the creation of SARS-CoV-2, exists in these and other related research laboratories.

What strengthens our contention further is the published RaTG13 virus[18], the genomic sequence of which is reportedly 96% identical to that of SARS-CoV-2. While suggesting a natural origin of SARSCoV-2, the RaTG13 virus also diverted the attention of both the scientific field and the general public away from ZC45/ZXC21[4,18]. In fact, a Chinese BSL-3 lab (the Shanghai Public Health Clinical Centre), which published a *Nature* article reporting a conflicting close phylogenetic relationship between SARSCoV-2 and ZC45/ZXC21 rather than with RaTG13[35], was quickly shut down for "rectification"[36]. It is believed that the researchers of that laboratory were being punished for having disclosed the SARS-CoV2—ZC45/ZXC21 connection. On the other hand, substantial evidence has accumulated, pointing to severe problems associated with the reported sequence of RaTG13 as well as questioning the actual existence of this bat virus in nature[6,7,19-21]. A very recent publication also indicated that the receptor-binding domain (RBD) of the RaTG13's Spike protein could not bind ACE2 of two different types of horseshoe bats (they closely relate to the horseshoe bat *R. affinis*, RaTG13's alleged natural host)[2], implicating the inability of RaTG13 to infect horseshoe bats. This finding further substantiates the suspicion that the reported sequence of RaTG13 could have been fabricated as the Spike protein encoded by this sequence does not seem to carry the claimed function. The fact that a virus has been fabricated to shift the attention away from ZC45/ZXC21 speaks for an actual role of ZC45/ZXC21 in the creation of SARS-CoV-2.

ii. *The receptor-binding motif of SARS-CoV-2 Spike cannot be born from nature and should have been created through genetic engineering*

The Spike proteins decorate the exterior of the coronavirus particles. They play an important role in infection as they mediate the interaction with host cell receptors and thereby help determine the host range and tissue tropism of the virus. The Spike protein is split into two halves (Figure 3). The front or N-terminal half is named S1, which is fully responsible for binding the host receptor. In both SARS-CoV and SARS-CoV-2 infections, the host cell receptor is hACE2. Within S1, a segment of around 70 amino acids makes direct contacts with hACE2 and is correspondingly named the receptor-binding motif (RBM) (Figure 3C). In SARS-CoV and SARS-CoV-2, the RBM fully determines the interaction with hACE2. The C-terminal half of the Spike protein is named S2. The main function of S2 includes maintaining trimer formation and, upon successive protease cleavages at the S1/S2 junction and a downstream S2' position, mediating membrane fusion to enable cellular entry of the virus.

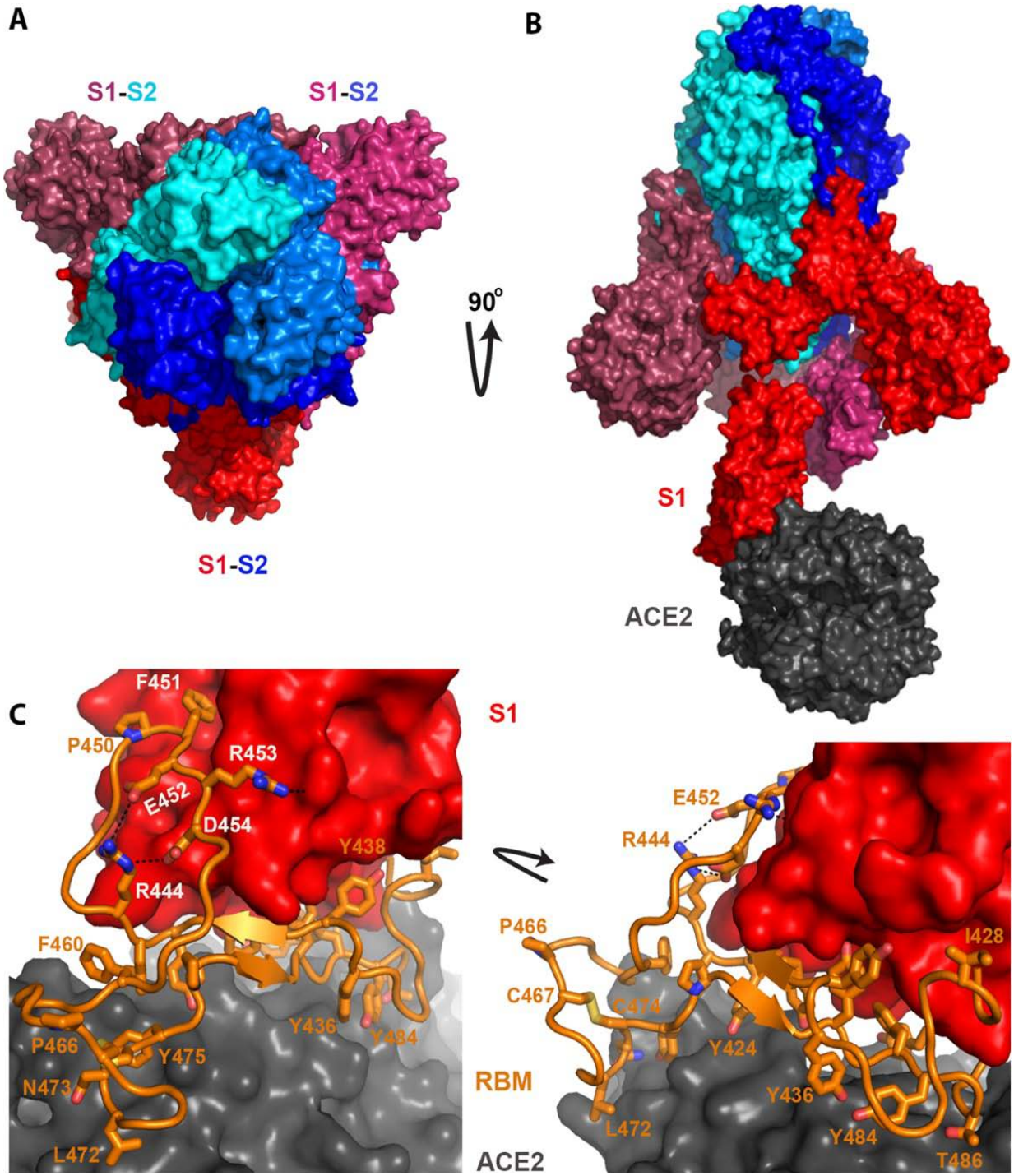


Figure 3: Structure of the SARS Spike protein and how it binds to the hACE2 receptor. Pictures were generated based on PDB ID: 6acj[37]. A) Three spike proteins, each consisting of a S1 half and a S2 half, form a trimer. B) The S2 halves (shades of blue) are responsible for trimer formation, while the S1 portion (shades of red) is responsible for binding hACE2 (dark gray). C) Details of the binding between S1 and hACE2. The RBM of S1, which is important and sufficient for binding, is colored in orange. Residues within the RBM that are important for either hACE2 interaction or protein folding are shown as sticks (residue numbers follow the SARS Spike sequence).

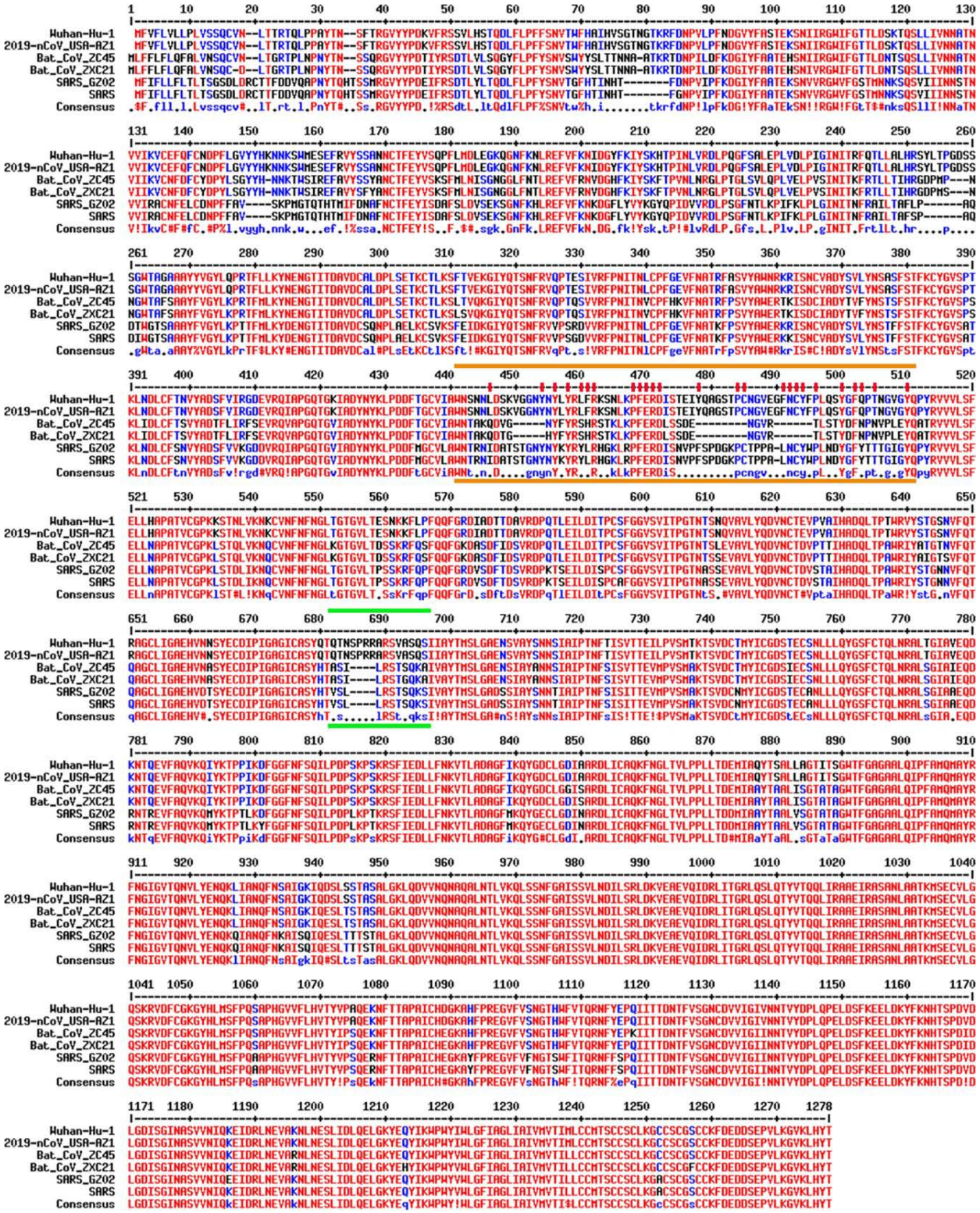


Figure 4: Sequence alignment of the spike proteins from relevant coronaviruses. Viruses being compared include SARS-CoV-2 (Wuhan-Hu-1: NC_045512, 2019-nCoV_USA-AZ1: MN997409), bat coronaviruses (Bat_CoV_ZC45: MG772933, Bat_CoV_ZXC21: MG772934), and SARS coronaviruses (SARS_GZ02: AY390556, SARS: NC_004718.3). Region marked by two orange lines is the receptor-binding motif (RBM), which is important for interaction with the hACE2 receptor. Essential residues are additionally highlighted by red sticks on top. Region marked by two green lines is a furin-cleavage site that exists only in SARS-CoV-2 but not in any other lineage B β coronavirus.

Similar to what is observed for other viral proteins, S2 of SARS-CoV-2 shares a high sequence identity (95%) with S2 of ZC45/ZXC21. In stark contrast, between SARS-CoV-2 and ZC45/ZXC21, the S1 protein, which dictates which host (human or bat) the virus can infect, is much less conserved with the amino acid sequence identity being only 69%.

Figure 4 shows the sequence alignment of the Spike proteins from six β coronaviruses. Two are viruses isolated from the current pandemic (Wuhan-Hu-1, 2019-nCoV_USA-AZ1); two are the suspected template viruses (Bat_CoV_ZC45, Bat_CoV_ZXC21); two are SARS coronaviruses (SARS_GZ02, SARS). The RBM is highlighted in between two orange lines. Clearly, despite the high sequence identity for the overall genomes, the RBM of SARS-CoV-2 differs significantly from those of ZC45 and ZXC21. Intriguingly, the RBM of SARS-CoV-2 resembles, on a great deal, the RBM of SARS Spike. Although this is not an exact "copy and paste", careful examination of the Spike-hACE2 structures[37,38] reveals that all residues essential for either hACE2 binding or protein folding (orange sticks in Figure 3C and what is highlighted by red short lines in Figure 4) are "kept". Most of these essential residues are precisely preserved, including those involved in disulfide bond formation (C467, C474) and electrostatic interactions (R444, E452, R453, D454), which are pivotal for the structural integrity of the RBM (Figure 3C and 4). The few changes within the group of essential residues are almost exclusively hydrophobic "substitutions" (I428→L, L443→F, F460→Y, L472→F, Y484→Q), which should not affect either protein folding or the hACE2-interaction. At the same time, majority of the amino acid residues that are non-essential have "mutated" (Figure 4, RBM residues not labeled with short red lines). Judging from this sequence analysis alone, we were convinced early on that not only would the SARS-CoV-2 Spike protein bind hACE2 but also the binding would resemble, precisely, that between the original SARS Spike protein and hACE2[23]. Recent structural work has confirmed our prediction[39].

As elaborated below, the way that SARS-CoV-2 RBM resembles SARS-CoV RBM and the overall sequence conservation pattern between SARS-CoV-2 and ZC45/ZXC21 are highly unusual. Collectively, this suggests that portions of the SARS-CoV-2 genome have not been derived from natural quasi-species viral particle evolution.

If SARS-CoV-2 does indeed come from natural evolution, its RBM could have only been acquired in one of the two possible routes: 1) an ancient recombination event followed by convergent evolution or 2) a natural recombination event that occurred fairly recently.

In the first scenario, the ancestor of SARS-CoV-2, a ZC45/ZXC21-like bat coronavirus would have recombined and "swapped" its RBM with a coronavirus carrying a relatively "complete" RBM (in reference to SARS). This recombination would result in a novel ZC45/ZXC21-like coronavirus with all the gaps in its RBM "filled" (Figure 4). Subsequently, the virus would have to adapt extensively in its new host, where the ACE2 protein is highly homologous to hACE2. Random mutations across the genome would have to have occurred to eventually shape the RBM to its current form – resembling SARS-CoV RBM in a highly intelligent manner. However, this convergent evolution process would also result in the accumulation of a large amount of mutations in other parts of the genome, rendering the overall sequence identity relatively low. The high sequence identity between SARS-CoV-2 and ZC45/ZXC21 on various proteins (94-100% identity) do not support this scenario and, therefore, clearly indicates that SARS-CoV2 carrying such an RBM cannot come from a ZC45/ZXC21-like bat coronavirus through this convergent evolutionary route.

In the second scenario, the ZC45/ZXC21-like coronavirus would have to have recently recombined and swapped its RBM with another coronavirus that had successfully adapted to bind an animal ACE2 highly homologous to hACE2. The likelihood of such an event depends, in part, on the general requirements of natural recombination: 1) that the two different viruses share significant sequence similarity; 2) that they must co-infect and be present in the same cell of the same animal; 3) that the recombinant virus would not be cleared by the host or make the host extinct; 4) that the recombinant virus eventually would have to become stable and transmissible within the host species.

In regard to this recent recombination scenario, the animal reservoir could not be bats because the ACE2 proteins in bats are not homologous enough to hACE2 and therefore the adaption would not be able to yield an RBM sequence as seen in SARS-CoV-2. This animal reservoir also could not be humans as the ZC45/ZXC21-like coronavirus would not be able to infect humans. In addition, there has been no evidence of any SARS-CoV-2 or SARS-CoV-2-like virus circulating in the human population prior to late 2019. Intriguingly, according to a recent bioinformatics study, SARS-CoV-2 was well-adapted for humans since the start of the outbreak[1].

Only one other possibility of natural evolution remains, which is that the ZC45/ZXC21-like virus and a coronavirus containing a SARS-like RBM could have recombined in an intermediate host where the ACE2 protein is homologous to hACE2. Several laboratories have reported that some of the Sunda pangolins smuggled into China from Malaysia carried coronaviruses, the receptor-binding domain (RBD) of which is almost identical to that of SARS-CoV-2[27-29,31]. They then went on to suggest that pangolins are the likely intermediate host for SARS-CoV-2[27-29,31]. However, recent independent reports have found significant flaws in this data[40-42]. Furthermore, contrary to these

reports[27-29,31], no coronaviruses have been detected in Sunda pangolin samples collected for over a decade in Malaysia and Sabah between 2009 and 2019[43]. A recent study also showed that the RBD, which is shared between SARS-CoV-2 and the reported pangolin coronaviruses, binds to hACE2 ten times stronger than to the pangolin ACE2[2], further dismissing pangolins as the possible intermediate host. Finally, an *in silico* study, while echoing the notion that pangolins are not likely an intermediate host, also indicated that none of the animal ACE2 proteins examined in their study exhibited more favorable binding potential to the SARS-CoV-2 Spike protein than hACE2 did[3]. This last study virtually exempted all animals from their suspected roles as an intermediate host[3], which is consistent with the observation that SARS-CoV-2 was well-adapted for humans from the start of the outbreak[1]. This is significant because these findings collectively suggest that no intermediate host seems to exist for SARS-CoV-2, which at the very least diminishes the possibility of a recombinant event occurring in an intermediate host.

Even if we ignore the above evidence that no proper host exists for the recombination to take place and instead assume that such a host does exist, it is still highly unlikely that such a recombination event could occur in nature.

As we have described above, if natural recombination event is responsible for the appearance of SARSCoV-2, then the ZC45/ZXC21-like virus and a coronavirus containing a SARS-like RBM would have to recombine in the same cell by swapping the S1/RBM, which is a rare form of recombination. Furthermore, since SARS has occurred only once in human history, it would be at least equally rare for nature to produce a virus that resembles SARS in such an intelligent manner – having an RBM that differs from the SARS RBM only at a few non-essential sites (Figure 4). The possibility that this unique SARS-like coronavirus would reside in the same cell with the ZC45/ZXC21-like ancestor virus and the two viruses would recombine in the "RBM-swapping" fashion is extremely low. Importantly, this, and the other recombination event described below in section 1.3 (even more impossible to occur in nature), would both have to happen to produce a Spike as seen in SARS-CoV-2.

While the above evidence and analyses together appear to disapprove a natural origin of SARS-CoV2's RBM, abundant literature shows that gain-of-function research, where the Spike protein of a coronavirus was specifically engineered, has repeatedly led to the successful generation of humaninfecting coronaviruses from coronaviruses of non-human origin[44-47].

Record also shows that research laboratories, for example, the Wuhan Institute of Virology (WIV), have successfully carried out such studies working with US researchers[45] and also working alone[47]. In addition, the WIV has engaged in decades-long coronavirus surveillance studies and therefore owns the world's largest collection of coronaviruses. Evidently, the technical barrier is non-existent for the WIV and other related laboratories to carry out and succeed in such Spike/RBM engineering and gain-of function research.

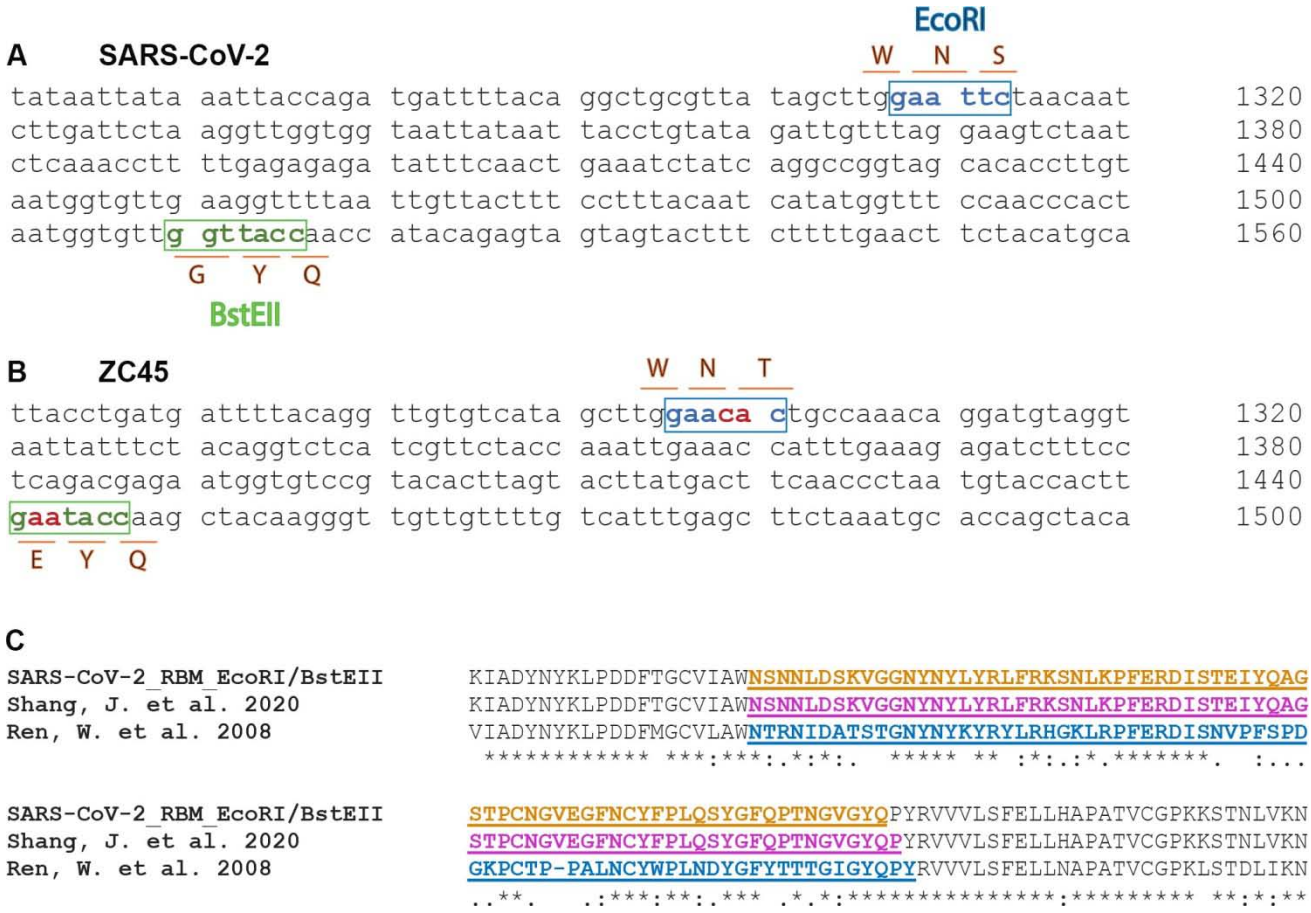


Figure 5: Two restriction sites are present at either end of the RBM of SARS-CoV-2, providing convenience for replacing the RBM within the spike gene. A. Nucleotide sequence of the RBM of SARS-CoV-2 (Wuhan-Hu-1). An EcoRI site is found at the 5'-end of the RBM and a BstEII site at the 3'-end. B. Although these two restriction sites do not exist in the original spike gene of ZC45, they can be conveniently introduced given that the sequence discrepancy is small (2 nucleotides) in either case. C. Amino acid sequence alignment with the RBM region highlighted (color and underscore). The RBM highlighted in orange (top) is what is defined by the EcoRI and BstEII sites in the SARS-CoV-2 (Wuhan-Hu-1) spike. The RBM highlighted in magenta (middle) is the region swapped by Dr. Fang Li and colleagues into a SARS Spike backbone[39]. The RBM highlighted in blue (bottom) is from the Spike protein (RBM: 424-494) of SARS-BJ01 (AY278488.2), which was swapped by the Shi lab into the Spike proteins of different bat coronaviruses replacing the corresponding segments[47].

Strikingly, consistent with the RBM engineering theory, we have identified two unique restriction sites, EcoRI and BstEII, at either end of the *RBM* of the SARS-CoV-2 genome, respectively (Figure 5A). These two sites, which are popular choices of everyday molecular cloning, do not exist in the rest of this *spike*gene. This particular settingmakes it extremely convenient to swap the *RBM* within *spike*, providing a quick way to test different RBMs and the corresponding Spike proteins.

Such EcoRI and BstEII sites do not exist in the *spike* genes of other β coronaviruses, which strongly indicates that they were unnatural and were specifically introduced into this *spike* gene of SARS-CoV-2 for the convenience of manipulating the critical RBM. Although ZC45 *spike* also does not have these two sites (Figure 5B), they can be introduced very easily as described in part 2 of this report.

It is noteworthy that introduction of the EcoRI site here would change the corresponding amino acids from -*WNT*- to -*WNS*- (Figure 5AB). As far as we know, all SARS and SARS-like bat coronaviruses exclusively carry a *T* (threonine) residue at this location. SARS-CoV-2 is the only exception in that this *T* has mutated to an *S* (serine), save the suspicious RaTG13 and pangolin coronaviruses published after the outbreak[48].

Once the restriction sites were successfully introduced, the *RBM* segment could be swapped conveniently using routine restriction enzyme digestion and ligation. Although alternative cloning techniques may leave no trace of genetic manipulation (Gibson assembly as one example), this oldfashioned approach could be chosen because it offers a great level of convenience in swapping this critical *RBM*.

Given that RBM fully dictates hACE2-binding and that the SARS RBM-hACE2 binding was fully characterized by high-resolution structures (Figure 3)[37,38], this RBM-only swap would not be any riskier than the full Spike swap. In fact, the feasibility of this RBM-swap strategy has been proven[39,47]. In 2008, Dr. Zhengli Shi's group swapped a SARS RBM into the Spike proteins of several SARS-like bat coronaviruses after introducing a restriction site into a codon-optimized *spike* gene (Figure 5C)[47]. They then validated the binding of the resulted chimeric Spike proteins with hACE2. Furthermore, in a recent publication, the RBM of SARS-CoV-2 was swapped into the receptor-binding domain (RBD) of SARSCoV, resulting in a chimeric RBD fully functional in binding hACE2 (Figure 5C)[39]. Strikingly, in bothcases, the manipulated RBM segments resemble almost exactly the RBM defined by the positions of theEcoRI and BstEII sites (Figure 5C). Although cloning details are lacking in both publications[39,47], it is conceivable that the actual restriction sites may vary depending on the *spike* gene receiving the RBM insertion as well as the convenience in introducing unique restriction site(s) in regions of interest. It is noteworthy that the corresponding author of this recent publication[39], Dr. Fang Li, has been an active collaborator of Dr. Zhengli Shi since 2010[49-53]. Dr. Li was the first person in the world to have structurally elucidated the binding between SARS-CoV RBD and hACE2[38] and has been the leading expert in the structural understanding of Spike-ACE2 interactions[38,39,53-56]. The striking finding of EcoRI and BstEII restriction sites at either end of the SARS-CoV-2 RBM, respectively, and the fact that the same RBM region has been swapped both by Dr. Shi and by her long-term collaborator, respectively, using restriction enzyme digestion methods are unlikely a coincidence. Rather, it is the smoking gun proving that theRBM/Spike of SARS-CoV-2 is a product of genetic manipulation.

Although it may be convenient to copy the exact sequence of SARS RBM, it would be too clear a sign of artificial design and manipulation. The more deceiving approach would be to change a few nonessential residues, while preserving the ones critical for binding. This design could be well-guided by the high-resolution structures (Figure 3)[37,38]. This way, when the overall sequence of the RBM would appear to be more distinct from that of the SARS RBM, the hACE2-binding ability would be well-preserved. We believe that all of the crucial residues (residues labeled with red sticks in Figure 4, which are the same residues shown in sticks in Figure 3C) should have been "kept". As described earlier, while some should be direct preservation, some should have been switched to residues with similar properties, which would not disrupt hACE2-binding and may even strengthen the association further. Importantly, changes might have been made intentionally at non-essential sites, making it less like a "copy and paste" of the SARS RBM.

iii. *An unusual furin-cleavage site is present in the Spike protein of SARS-CoV-2 and is associated with the augmented virulence of the virus*

Another unique motif in the Spike protein of SARS-CoV-2 is a polybasic furin-cleavage site located at the S1/S2 junction (Figure 4, segment in between two green lines). Such a site can be recognized and cleaved by the furin protease. Within the lineage B of β coronaviruses and with the exception of SARSCoV-2, no viruses contain a furin-cleavage site at the S1/S2 junction (Figure 6)[57]. In contrast, furincleavage site at this location has been observed in other groups of coronaviruses[57,58]. Certain selective pressure seems to be in place that prevents the lineage B of β coronaviruses from acquiring or maintaining such a site in nature.

```
Human SARS-CoV BJ01                655 - GICASYHTVSLL----RSTS - 670
Human SARS-CoV CUHK-W1             655 - GICASYHTVSLL----RSTS - 670
Human SARS-CoV Tor2                655 - GICASYHTVSLL----RSTS - 670
Human SARS-CoV Frankfurt-1         655 - GICASYHTVSLL----RSTS - 670
Human SARS-CoV Urbani              655 - GICASYHTVSLL----RSTS - 670
Civet SARS-CoV civet020            655 - GICASYHTVSSL----RSTS - 670
Civet SARS-CoV sz16                655 - GICASYHTVSSL----RSTS - 670
Racoon dog SARS-CoV A030           655 - GICASYHTVSSL----RSTS - 670
SARS-CoV-2                         669 - GICASYQTQTNSPRRARSVA - 688
Pangolin CoV MP789                 n/a - GICASYQTQTNS----RSVS - n/a
Bat SARSr-CoV RaTG13               669 - GICASYQTQTNS----RSVA - 684
Bat SARSr-CoV LYRa11               659 - GICASYHTASLL----RNTD - 674
Bat SARSr-CoV LYRa3                659 - GICASYHTASLL----RNTG - 674
Bat SARSr-CoV RsSHC014             656 - GICASYHTVSSL----RSTS - 671
Bat SARSr-CoV Rs4084               656 - GICASYHTVSSL----RSTS - 671
Bat SARSr-CoV WIV1                 656 - GICASYHTVSSL----RSTS - 671
Bat SARSr-CoV Rs3367               656 - GICASYHTVSSL----RSTS - 671
Bat SARSr-CoV Rs7327               656 - GICASYHTVSSL----RSTS - 671
Bat SARSr-CoV Rs9401               656 - GICASYHTVSSL----RSTS - 671
Bat SARSr-CoV Rs4231               655 - GICASYHTVSSL----RSTS - 670
Bat SARSr-CoV WIV16                655 - GICASYHTVSSL----RSTS - 670
Bat SARSr-CoV Rs4874               655 - GICASYHTVSSL----RSTS - 670
Bat SARSr-CoV ZC45                 646 - GICASYHTASIL----RSTS - 661
Bat SARSr-CoV ZXC21                645 - GICASYHTASIL----RSTG - 660
Bat SARSr-CoV Rf4092               634 - GICASYHTASTL----RGVG - 649
Bat SARSr-CoV Rf/JL2012            636 - GICASYHTASLL----RSTG - 651
Bat SARSr-CoV JTMC15               636 - GICASYHTASLL----RSTG - 651
Bat SARSr-CoV 16BO133              636 - GICASYHTASLL----RSTG - 651
Bat SARSr-CoV B15-21               636 - GICASYHTASLL----RSTG - 651
Bat SARSr-CoV YN2013               633 - GICASYHTASTL----RSIG - 648
Bat SARSr-CoV Anlong-103           633 - GICASYHTASTL----RSVG - 648
Bat SARSr-CoV Rp/Shaanxi2011       640 - GICASYHTASVL----RSTG - 655
Bat SARSr-CoV Rs/HuB2013           641 - GICASYHTASVL----RSTG - 656
Bat SARSr-CoV YNLF/34C             641 - GICASYHTASVL----RSTG - 656
Bat SARSr-CoV YNLF/31C             641 - GICASYHTASVL----RSTG - 656
Bat SARSr-CoV Rf1                  641 - GICASYHTASHL----RSTG - 656
Bat SARSr-CoV 273                  641 - GICASYHTASHL----RSTG - 656
Bat SARSr-CoV Rf/SX2013            639 - GICASYHTASLL----RSTG - 654
Bat SARSr-CoV Rf/HeB2013           641 - GICASYHTASLL----RSTG - 656
Bat SARSr-CoV Cp/Yunnan2011        641 - GICASYHTASLL----RNTG - 656
Bat SARSr-CoV Rs672                641 - GICASYHTASTL----RSVG - 656
Bat SARSr-CoV Rs4255               641 - GICASYHTASTL----RSVG - 656
Bat SARSr-CoV 4081                 641 - GICASYHTASTL----RSVG - 656
Bat SARSr-CoV Rm1                  641 - GICASYHTASVL----RSTG - 656
Bat SARSr-CoV 279                  641 - GICASYHTASVL----RSTG - 656
Bat SARSr-CoV Rs/GX2013            642 - GICASYHTASVL----RSTG - 657
Bat SARSr-CoV Rs806                641 - GICASYHTASLL----RSTG - 656
Bat SARSr-CoV HKU3-1               642 - GICASYHTASVL----RSTG - 657
Bat SARSr-CoV Longquan-140         642 - GICASYHTASVL----RSTG - 657
Bat SARSr-CoV Rp3                  641 - GICASYHTASTL----RSVG - 656
Bat SARSr-CoV Rs4247               642 - GICASYHTASTL----RSVG - 657
Bat SARSr-CoV Rs4237               641 - GICASYHTASTL----RSVG - 656
Bat SARSr-CoV As6526               641 - GICASYHTASTL----RSVG - 656
```

Figure 6: Furin-cleavage site found at the S1/S2 junction of Spike is unique to SARS-CoV-2 and absent in other lineage B β coronaviruses. Figure reproduced from Hoffmann, et al[57].

As previously described, during the cell entry process, the Spike protein is first cleaved at the S1/S2 junction. This step, and a subsequent cleavage downstream that exposes the fusion peptide, are both mediated by host proteases. The presence or absence of these proteases in different cell types greatly affects the cell tropism and presumably the pathogenicity of the viral infection. Unlike other proteases, furin protease is widely expressed in many types of cells and is present at multiple cellular and extracellular locations. Importantly, the introduction of a furin-cleavage site at the S1/S2 junction could significantly enhance the infectivity of a virus as well as greatly expand its cell tropism — a phenomenon well-documented in both influenza viruses and other coronaviruses[59-65].

If we leave aside the fact that no furin-cleavage site is found in any lineage B β coronavirus in nature and instead assume that this site in SARS-CoV-2 is a result of natural evolution, then only one evolutionary pathway is possible, which is that the furin-cleavage site has to be derived from a homologous recombination event. Specifically, an ancestor β coronavirus containing no furin-cleavage site would have to recombine with a closely related coronavirus that does contain a furin-cleavage site.

However, two facts disfavor this possibility. First, although some coronaviruses from other groups or lineages do contain polybasic furin-cleavage sites, none of them contains the exact polybasic sequence present in SARS-CoV-2 (-*PRRAR*/*SVA*-). Second, between SARS-CoV-2 and any coronavirus containing a legitimate furin-cleavage site, the sequence identity on Spike is no more than 40%[66]. Such a low level of sequence identity rules out the possibility of a successful homologous recombination ever occurring between the ancestors of these viruses. Therefore, the furin-cleavage site within the SARS-CoV-2 Spike protein is unlikely to be of natural origin and instead should be a result of laboratory modification.

Consistent with this claim, a close examination of the nucleotide sequence of the furin-cleavage site in SARS-CoV-2 *spike* has revealed that the two consecutive Arg residues within the inserted sequence (PRRA-) are both coded by the rare codon CGG (least used codon for Arg in SARS-CoV-2) (Figure 7)[8]. In fact, this *CGGCGG* arrangement is the only instance found in the SARS-CoV-2 genome where this rare codon is used in tandem. This observation strongly suggests that this furin-cleavage site should be aresult of genetic engineering. Adding to the suspicion, a *FauI* restriction site is formulated by the codon choices here, suggesting the possibility that the *restriction fragment length polymorphism,* a technique that a WIV lab is proficient at[67], could have been involved. There, the fragmentation pattern resulted from *FauI* digestion could be used to monitor the preservation of the furin-cleavage site in Spike as this furincleavage site is prone to deletions *in vitro*[68,69]. Specifically, RT-PCR on the *spike* gene of the recovered viruses from cell cultures or laboratory animals could be carried out, the product of which would be subjected to *FauI* digestion. Viruses retaining or losing the furin-cleavage site would then yield distinct patterns, allowing convenient tracking of the virus(es) of interest.

Figure 7: Two consecutive Arg residues in the -PRRA- insertion at the S1/S2 junction of SARS-CoV-2 Spike are both coded by a rare codon, CGG. A FauI restriction site, 5'-$(N)_6$GCGGG-3', is embedded in the coding sequence of the "inserted" PRRA segment, which may be used as a marker to monitor the preservation of the introduced furin-cleavage site.

In addition, although no known coronaviruses contain the exact sequence of -*PRRAR*/*SVA*- that is present in the SARS-CoV-2 Spike protein, a similar -*RRAR*/*AR*- sequence has been observed at the S1/S2 junction of the Spike protein in a rodent coronavirus, AcCoV-JC34, which was published by Dr. Zhengli Shi in 2017[70]. It is evident that the legitimacy of -*RRAR*- as a functional furin-cleavage site has been known to the WIV experts since 2017.

The evidence collectively suggests that the furin-cleavage site in the SARS-CoV-2 Spike protein may not have come from nature and could be the result of genetic manipulation. The purpose of this manipulation could have been to assess any potential enhancement of the infectivity and pathogenicity of the laboratory-made coronavirus[59-64]. Indeed, recent studies have confirmed that the furin-cleavage site does confer significant pathogenic advantages to SARS-CoV-2[57,68].

iv. *Discussion*

Evidence presented in this part reveals that certain aspects of the SARS-CoV-2 genome are extremely difficult to reconcile to being a result of natural evolution. The alternative theory we suggest is that the virus may have been created by using ZC45/ZXC21 bat coronavirus(es) as the backbone and/or template. The Spike protein, especially the RBM within it, should have been artificially manipulated, upon which the virus has acquired the ability to bind hACE2 and infect humans. This is supported by the finding of a unique restriction enzyme digestion site at either end of the RBM. An unusual furin-cleavage site may have been introduced and inserted at the S1/S2 junction of the Spike protein,

which contributes to the increased virulence and pathogenicity of the virus. These transformations have then staged the SARSCoV-2 virus to eventually become a highly-transmissible, onset-hidden, lethal, sequelae-unclear, and massively disruptive pathogen.

Evidently, the possibility that SARS-CoV-2 could have been created through gain-of-function manipulations at the WIV is significant and should be investigated thoroughly and independently.

b) Delineation of a synthetic route of SARS-CoV-2

In the second part of this report, we describe a synthetic route of creating SARS-CoV-2 in a laboratory setting. It is postulated based on substantial literature support as well as genetic evidence present in the SARS-CoV-2 genome. Although steps presented herein should not be viewed as exactly those taken, we believe that key processes should not be much different. Importantly, our work here should serve as a demonstration of how SARS-CoV-2 can be designed and created conveniently in research laboratories by following proven concepts and using well-established techniques.

Importantly, research labs, both in Hong Kong and in mainland China, are leading the world in coronavirus research, both in terms of resources and on the research outputs. The latter is evidenced not only by the large number of publications that they have produced over the past two decades but also by their milestone achievements in the field: they were the first to identify civets as the intermediate host for SARS-CoV and isolated the first strain of the virus[71]; they were the first to uncover that SARS-CoV originated from bats[72,73]; they revealed for the first time the antibody-dependent enhancement (ADE) of SARS-CoV infections[74]; they have contributed significantly in understanding MERS in all domains (zoonosis, virology, and clinical studies)[75-79]; they made several breakthroughs in SARS-CoV-2 research[18,35,80]. Last but not least, they have the world's largest collection of coronaviruses (genomic sequences and live viruses). The knowledge, expertise, and resources are all readily available within the Hong Kong and mainland research laboratories (they collaborate extensively) to carry out and accomplish the work described below.

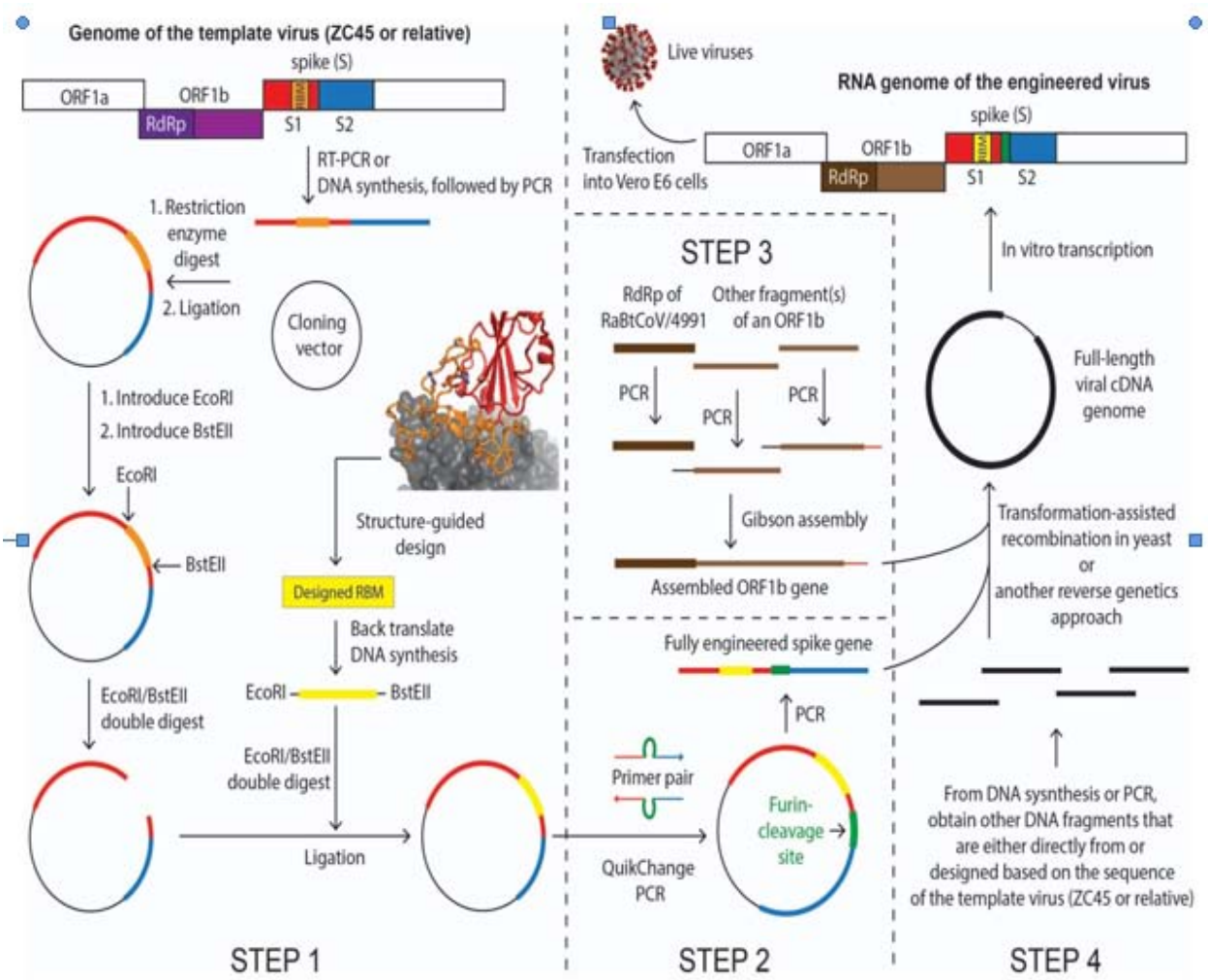


Figure 8: Diagram of a possible synthetic route of the laboratory-creation of SARS-CoV-2

c) *Possible scheme in designing the laboratory-creation of the novel coronavirus*

In this sub-section, we outline the possible overall strategy and major considerations that may have been formulated at the designing stage of the project.

To engineer and create a human-targeting coronavirus, they would have to pick a bat coronavirus asthe template/backbone. This can be conveniently done because many research labs have been actively collecting bat coronaviruses over the past two decades[32,33,70,72,81-85]. However, this template virus ideally should not be one from Dr. Zhengli Shi's collections, considering that she is widely known to have been engaged in gain-of-function studies on coronaviruses. Therefore, ZC45 and/or ZXC21, novel bat coronaviruses discovered and owned by military laboratories[33], would be suitable as the template/backbone. It is also possible that these military laboratories had discovered other closely related viruses from the same location and kept some unpublished. Therefore, the actual template could be ZC45, or ZXC21, or a close relative of them. The postulated pathway described below would be the same regardless of which one of the three was the actual template.

Once they have chosen a template virus, they would first need to engineer, through molecular cloning,the Spike protein so that it can bind hACE2. The concept and cloning techniques involved in this manipulation have been well-documented in the literature[44-46,84,86]. With almost no risk of failing, the template bat virus could then be converted to a coronavirus that can bind hACE2 and infect humans[44-46].

Second, they would use molecular cloning to introduce a furin-cleavage site at the S1/S2 junction ofSpike. This manipulation, based on known knowledge[60,61,65], would likely produce a strain of coronavirus that is a more infectious and pathogenic.

Third, they would produce an *ORF1b* gene construct. The *ORF1b* gene encodes the polyprotein Orf1b, which is processed post-translationally to produce individual viral proteins: RNA-dependent RNA polymerase (RdRp), helicase, guanidine-N7 methyl transferase, uridylate-specific endoribonuclease, and 2'-O-methyltransferase. All of these proteins are parts of the replication machinery of the virus. Among them, the RdRp protein is the most crucial one and is highly conserved among coronaviruses. Importantly, Dr. Zhengli Shi's laboratory uses a PCR protocol, which amplifies a particular fragment of the *RdRp* gene, as their primary method to detect the presence of coronaviruses in raw samples (bat fecal swap, feces, etc). As a result of this practice, the Shi group has documented the sequence information of this short segment of *RdRp* for all coronaviruses that they have successfully detected and/or collected.

Here, the genetic manipulation is less demanding or complicated because Orf1b is conserved and likely Orf1b from any β coronavirus would be competent enough to do the work. However, we believe that they would want to introduce a particular Orf1b into the virus for one of the two possible reasons:

1. Since many phylogenetic analyses categorize coronaviruses based on the sequence similarity of the *RdRp* gene only[18,31,35,83,87], having a different *RdRp* in the genome therefore could ensure that SARS-CoV-2 and ZC45/ZXC21 are separated into different groups/sub-lineages in phylogenetic studies. Choosing an *RdRp* gene, however, is convenient because the short *RdRp* segment sequence has been recorded for all coronaviruses ever collected/detected. Their final choice was the *RdRp* sequence from bat coronavirus RaBtCoV/4991, which was discovered in 2013. For RaBtCoV/4991, the only information ever published was the sequence of its short *RdRp* segment[83], while neither its full genomic sequence nor virus isolation were ever reported. After amplifying the *RdRp* segment (or the whole *ORF1b* gene) of RaBatCoV/4991, they would have then used it for subsequent assembly and creation of the genome of SARS-CoV-2. Small changes in the *RdRp* sequence could either be introduced at the beginning (through DNA synthesis) or be generated *via*passages later on. On a separate track, when they were engaged in the fabrication of the RaTG13 sequence, they could have started with the short *RdRp* segment of RaBtCoV/4991 without introducing any changes to its sequence, resulting in a 100% nucleotide sequence identity between the two viruses on this short *RdRp* segment[83]. This RaTG13 virus could then be claimed to have been discovered back in 2013.
2. The RdRp protein from RaBatCoV/4991 is unique in that it is superior thanRdRp from any other β coronavirus for developing antiviral drugs. RdRp has no homologs in human cells, which makes this essential viral enzyme a highly desirable target for antiviral development. As an example, *Remedesivir*, which is currently undergoing clinical trials, targets RdRp. When creating a novel and human-targeting virus, they would be interested in developing the antidote as well. Even though drug discovery like this may not be easily achieved, it is reasonable for them to intentionally incorporate a RdRp that is more amenable for antiviral drug development.

Fourth, they would use reverse genetics to assemble the gene fragments of *spike*, *ORF1b*, and the rest of the template ZC45 into a cDNA version of the viral genome. They would then carry out *in vitro* transcription to obtain the viral RNA genome. Transfection of the RNA genome into cells would allow the recovery of live and infectious viruses with the desired artificial genome.

Fifth, they would carry out characterization and optimization of the virus strain(s) to improve the fitness, infectivity, and overall adaptation using serial passage *in vivo*. One or several viral strains that meet certain criteria would then be obtained as the final product(s).

d) A postulated synthetic route for the creation of SARS-CoV-2

In this sub-section, we describe in more details how each step could be carried out in a laboratory setting using available materials and routine molecular, cellular, and virologic techniques. A diagram of this process is shown in Figure 8. We estimate that the whole process could be completed in approximately 6 months.

Step 1: Engineering the RBM of the Spike for hACE2-binding (1.5 months)

The Spike protein of a bat coronavirus is either incapable of or inefficient in binding hACE2 due to the missing of important residues within its RBM. This can be exemplified by the RBM of the template virus ZC45 (Figure 4). The first and most critical step in the creation of SARS-CoV-2 is to engineer the Spike so that it acquires the ability to bind hACE2. As evidenced in the literature, such manipulations have been carried out repeatedly in research laboratories since 2008[44], which successfully yielded engineered coronaviruses with the ability to infect human cells[44-46,88,89]. Although there are many possible ways that one can engineer the Spike protein, we believe that what was actually undertaken was that they replaced the original RBM with a designed and possibly optimized RBM using SARS' RBM as a guide. As described in part 1, this theory is supported by our observation that two unique restriction sites, EcoRI and BstEII, exist at either end of the *RBM* in the SARS-CoV-2 genome (Figure 5A) and by the fact that such RBM-swap has been successfully carried out by Dr. Zhengli Shi and by her long-term collaborator and structure biology expert, Dr. Fang Li[39,47].

Although ZC45 *spike* does not contain these two restriction sites (Figure 5B), they can be introduced very easily. The original *spike* gene would be either amplified with RT-PCR or obtained through DNA synthesis (some changes could be safely introduced to certain variable regions of the sequence) followed by PCR. The gene would then be cloned into a plasmid using restriction sites other than EcoRI and BstEII.

Once in the plasmid, the *spike* gene can be modified easily. First, an EcoRI site can be introduced by converting the highlighted "gaacac" sequence (Figure 5B) to the desired "gaattc" (Figure 5A). The difference between them are two consecutive nucleotides. Using the commercially available Quik Change Site-Directed Mutagenesis kit, such a di-nucleotide mutation can be generated in no more than one week. Subsequently, the BstEII site could be similarly introduced at the other end of the *RBM*. Specifically, the "gaatacc" sequence (Figure 5B) would be converted to the desired "ggttacc" (Figure 5A), which would similarly require a week of time.

Once these restriction sites, which are unique within the *spike* gene of SARS-CoV-2, were successfully introduced, different *RBM* segments could be swapped in conveniently and the resulting Spike protein subsequently evaluated using established assays.

As described in part 1, the design of an RBM segment could be well-guided by the high-resolution structures (Figure 3)[37,38], yielding a sequence that resembles the SARS RBM in an intelligent manner. When carrying out the structure-guided design of the RBM, they would have followed the routine and generated a few (for example a dozen) such RBMs with the hope that some specific variant(s) may be superior than others in binding hACE2. Once the design was finished, they could have each of the designed *RBM* genes commercially synthesized (quick and very affordable) with an EcoRI site at the 5'-end and a BstEII site at the 3'-end. These novel *RBM* genes could then be cloned into the *spike* gene, respectively. The gene synthesis and subsequent cloning, which could be done in a batch mode for the small library of designed RBMs, would take approximately one month.

These engineered Spike proteins might then be tested for hACE2-binding using the established pseudotype virus infection assays[45,49,50]. The engineered Spike with good to exceptional binding affinities would be selected. (Although not necessary, directed evolution could be involved here (error-prone PCR on the *RBM* gene), coupled with either an *in vitro* binding assay[39,90] or a pseudotype virus infection assay[45,49,50], to obtain an RBM that binds hACE2 with exceptional affinity.)

Given the abundance of literature on Spike engineering[44-46,84,86] and the available high-resolution structures of the Spike-hACE2 complex[37,38], the success of this step would be very much guaranteed. By the end of this step, as desired, a novel *spike* gene would be obtained, which encodes a novel Spike protein capable of binding hACE2 with high affinity.

Step 2: Engineering a furin-cleavage site at the S1/S2 junction (0.5 month)

The product from Step 1, a plasmid containing the engineered *spike*, would be further modified to include a furin-cleavage site (segment indicated by green lines in Figure 4) at the S1/S2 junction. This short stretch of gene sequence can be conveniently inserted using several routine cloning techniques, including Quik Change Site-Directed PCR[60], overlap PCR followed by restriction enzyme digestion and ligation[91], or Gibson assembly. None of these techniques would leave any trace in the sequence. Whichever cloning method was the choice, the inserted gene piece would be included in the primers, which would be designed, synthesized, and used in the cloning. This step, leading to a further

modified Spike with the furin-cleavage site added at the S1/S2 junction, could be completed in no more than two weeks.

Step 3: Obtain an ORF1b gene that contains the sequence of the short RdRp segment from RaBtCoV/4991 (1 month, yet can be carried out concurrently with Steps 1 and 2)

Unlike the engineering of Spike, no complicated design is needed here, except that the *RdRp* gene segment from RaBtCoV/4991 would need to be included. Gibson assembly could have been used here. In this technique, several fragments, each adjacent pair sharing 20-40 bp overlap, are combined together in one simple reaction to assemble a long DNA product. Two or three fragments, each covering a significant section of the *ORF1b* gene, would be selected based on known bat coronavirus sequences. One of these fragments would be the *RdRp* segment of RaBtCoV/4991[83]. Each fragment would be PCR amplified with proper overlap regions introduced in the primers. Finally, all purified fragments would be pooled in equimolar concentrations and added to the Gibson reaction mixture, which, after a short incubation, would yield the desired *ORF1b* gene in whole.

Step 4: Produce the designed viral genome using reverse genetics and recover live viruses (0.5 month)

Reverse genetics have been frequently used in assembling whole viral genomes, including coronavirus genomes[67,92-96]. The most recent example is the reconstruction of the SARS-CoV-2 genome using the *transformation-assisted recombination in yeast*[97]. Using this method, the Swiss group assembled the entire viral genome and produced live viruses in just one week[97]. This efficient technique, which would not leave any trace of artificial manipulation in the created viral genome, has been available since 2017[98,99]. In addition to the engineered *spike* gene (from steps 1 and 2) and the *ORF1b* gene (from step 3), other fragments covering the rest of the genome would be obtained either through RT-PCR amplification from the template virus or through DNA synthesis by following a sequence slightly altered from that of the template virus. We believe that the latter approach was more likely as it would allow sequence changes introduced into the variable regions of less conserved proteins, the process of which could be easily guided by multiple sequence alignments. The amino acid sequences of more conserved functions, such as that of the E protein, might have been left unchanged. All DNA fragments would then be pooled together and transformed into yeast, where the cDNA version of the SARS-CoV-2 genome would be assembled *via* transformation-assisted recombination. Of course, an alternative method of reverse genetics, one of which the WIV has successfully used in the past[67], could also be employed[67,92-96,100]. Although some earlier reverse genetics approaches may leave restriction sites at where different fragments would be joined, these traces would be hard to detect as the exact site of ligation can be anywhere in the ~30kb genome. Either way, a cDNA version of the viral genome would be obtained from the reverse genetics experiment. Subsequently, *in vitro* transcription using the cDNA as the template would yield the viral RNA genome, which upon transfection into Vero E6 cells would allow the production of live viruses bearing all of the designed properties.

Step 5: Optimize the virus for fitness and improve its hACE2-binding affinity in vivo (2.5-3 months)

Virus recovered from step 4 needs to be further adapted undergoing the classic experiment – serial passage in laboratory animals[101]. This final step would validate the virus' fitness and ensure its receptor oriented adaptation toward its intended host, which, according to the analyses above, should be human. Importantly, the RBM and the furin-cleavage site, which were introduced into the Spike protein separately, would now be optimized together as one functional unit. Among various available animal models (e.g. mice, hamsters, ferrets, and monkeys) for coronaviruses, hACE2 transgenic mice (hACE2-mice) should be the most proper and convenient choice here. This animal model has been established during the study of SARS-CoV and has been available in the Jackson Laboratory for many years[102-104].

The procedure of serial passage is straightforward. Briefly, the selected viral strain from step 4, a precursor of SARS-CoV-2, would be intranasally inoculated into a group of anaesthetized hACE2-mice. Around 2-3 days post infection, the virus in lungs would usually amplify to a peak titer. The mice would then be sacrificed and the lungs homogenized. Usually, the mouse-lung supernatant, which carries the highest viral load, would be used to extract the candidate virus for the next round of passage. After approximately 10~15 rounds of passage, the hACE2-binding affinity, the infection efficiency, and the lethality of the viral strain would be sufficiently enhanced and the viral genome stabilized[101]. Finally, after a series of characterization experiments (e.g. viral kinetics assay, antibodies response assay, symptom observation and pathology examination), the final product, SARS-CoV-2, would be obtained, concluding the whole creation process. From this point on, this viral pathogen could be amplified (most probably using Vero E6 cells) and produced routinely.

It is noteworthy that, based on the work done on SARS-CoV, the hACE2-mice, although suitable for SARS-CoV-2 adaptation, is not a good model to reflect the virus' transmissibility and associated clinical symptoms in humans. We believe that those scientists might not have used a proper animal model (such as the golden Syrian hamster) for testing the transmissibility

of SARS-CoV-2 before the outbreak of COVID-19. If they had done this experiment with a proper animal model, the highly contagious nature of SARS-CoV-2 would be extremely evident and consequently SARS-CoV-2 would not have been described as "not causing human-to-human transmission" at the start of the outbreak.

We also speculate that the extensive laboratory-adaptation, which is oriented toward enhanced transmissibility and lethality, may have driven the virus too far. As a result, SARS-CoV-2 might have lost the capacity to attenuate on both transmissibility and lethality during its current adaptation in the human population. This hypothesis is consistent with the lack of apparent attenuation of SARS-CoV-2 so far despite its great prevalence and with the observation that a recently emerged, predominant variant only shows improved transmissibility[105-108].

II. Remarks

Serial passage is a quick and intensive process, where the adaptation of the virus is accelerated. Although intended to mimic natural evolution, serial passage is much more limited in both time and scale. As a result, less random mutations would be expected in serial passage than in natural evolution. This is particularly true for conserved viral proteins, such as the E protein. Critical in viral replication, the E protein is a determinant of virulence and engineering of it may render SARS-CoV-2 attenuated[109-111] Therefore, at the initial assembly stage, these scientists might have decided to keep the amino acid sequence of the E protein unchanged from that of ZC45/ZXC21. Due to the conserved nature of the E protein and the limitations of serial passage, no amino acid mutation actually occurred, resulting in a 100% sequence identity on the E protein between SARS-CoV-2 and ZC45/ZXC21. The same could have happened to the marks of molecular cloning (restriction sites flanking the RBM). Serial passage, which should have partially naturalized the SARS-CoV-2 genome, might not have removed all signs of artificial manipulation.

Many questions remain unanswered about the origin of SARS-CoV-2. Prominent virologists have implicated in a *Nature Medicine* letter that laboratory escape, while not being entirely ruled out, was unlikely and that no sign of genetic manipulation is present in the SARS-CoV-2 genome[4]. However, here we show that genetic evidence within the *spike* gene of SARS-CoV-2 genome (restriction sites flanking the *RBM*; tandem rare codons used at the inserted furin-cleavage site) does exist and suggests that the SARS-CoV-2 genome should be a product of genetic manipulation. Furthermore, the proven concepts, well-established techniques, and knowledge and expertise are all in place for the convenient creation of this novel coronavirus in a short period of time.

Motives aside, the following facts about SARS-CoV-2 are well-supported:

1. If it was a laboratory product, the most critical element in its creation, the backbone/template virus (ZC45/ZXC21), is owned by military research laboratories.
2. The genome sequence of SARS-CoV-2 has likely undergone genetic engineering, through which the virus has gained the ability to target humans with enhanced virulence and infectivity.
3. The characteristics and pathogenic effects of SARS-CoV-2 are unprecedented. The virus is highly transmissible, onset-hidden, multi-organ targeting, sequelae-unclear, lethal, and associated with various symptoms and complications.
4. SARS-CoV-2 caused a world-wide pandemic, taking millions of lives and shutting down the global economy. It has a destructive power like no other.

Judging from the evidence that we and others have gathered, we believe that finding the origin of SARS-CoV-2 should involve an independent audit of the WIV P4 laboratories and the laboratories of their close collaborators. Such an investigation should have taken place long ago and should not be delayed any further.

We also note that in the publication of the chimeric virus SHC015-MA15 in 2015, the attribution of funding of Zhengli Shi by the NIAID was initially left out. It was reinstated in the publication in 2016 in a corrigendum, perhaps after the meeting in January 2016 to reinstate NIH funding for gain-of-function research on viruses. This is an unusual scientific behavior, which needs an explanation for.

What is not thoroughly described in this report is the various evidence indicating that several coronaviruses recently published (RaTG13[18], RmYN02[30], and several pangolin coronaviruses[27-29,31]) are highly suspicious and likely fraudulent. These fabrications would serve no purpose other than to deceive the scientific community and the general public so that the true identity of SARS-CoV-2 is hidden. Although exclusion of details of such evidence does not alter the conclusion of the current report, we do believe that these details would provide additional support for our contention that SARS-CoV-2 is a laboratory-enhanced virus and a product of gain-of-function research. A follow-up report focusing on such additional evidence is now being prepared and will be submitted shortly.

Acknowledgements

We would like to thank Daoyu Zhang for sharing with us the findings of mutations in the E proteins in different sub-groups of β coronaviruses. We also thank all the anonymous scientists and other individuals, who have contributed in uncovering various facts associated with the origin of SARS-CoV-2.

References Références Referencias

1. Zhan, S.H., Deverman, B.E. & Chan, Y.A. SARS-CoV-2 is well adapted for humans. What does this mean for re-emergence? *Bio Rxiv*, https://doi.org/10.1101/2020.05.01.073262 (2020).
2. Mou, H. et al. Mutations from bat ACE2 orthologs markedly enhance ACE2-Fc neutralization of SARSCoV-2. *Bio Rxiv*, https://doi.org/10.1101/2020.06.29.178459 (2020).
3. Piplani, S., Singh, P.K., Winkler, D.A. & Petrovsky, N. In silico comparison of spike protein-ACE2 binding affinities across species; significance for the possible origin of the SARS-CoV-2 virus. *arXiv*, arXiv:2005.06199 (2020).
4. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C. & Garry, R.F. The proximal origin of SARSCoV-2. *Nat Med*26, 450-452 (2020).
5. Maiti, A.K. On The Origin of SARS-CoV-2 Virus. *Preprint (authorea.com)*, DOI: 10.22541/au.159355977.76503625 (2020).
6. Lin, X. & Chen, S. Major Concerns on the Identification of Bat Coronavirus Strain RaTG13 and Quality of Related Nature Paper. *Preprints*, 2020060044 (2020).
7. Bengston, D. All journal articles evaluating the origin or epidemiology of SARS-CoV-2 that utilize the RaTG13 bat strain genomics are potentially flawed and should be retracted. *OSF Preprints*, DOI: 10.31219/osf.io/wy89d (2020).
8. Segreto, R. & Deigin, Y. Is considering a genetic-manipulation origin for SARS-CoV-2 a conspiracy theory that must be censored? *Preprint (Researchgate)* DOI: 10.13140/ RG.2.2.31358.13129/1 (2020).
9. Rahalkar, M.C. &Bahulikar, R.A. Understanding the Origin of 'BatCoVRaTG13', a Virus Closest to SARS-CoV-2. *Preprints*, 2020050322 (2020).
10. Robinson, C. Was the COVID-19 virus genetically engineered? (https://gmwatch.org/en/news/latestnews/19383, 2020).
11. Robinson, C. Another expert challenges assertions that SARS-CoV-2 was not genetically engineered. (https://gmwatch.org/en/news/latest-news/19383, 2020).
12. Sørensen, B., Dalgleish, A. &Susrud, A. The Evidence which Suggests that This Is No Naturally Evolved Virus. *Preprint*, https://www.minervanett.no/files/2020/07/13/TheEvidenceNoNaturalEvol.pdf (2020).
13. Zhang, B. SARS-CoV-2 Could Come from a Lab - A Critique of "The Proximal Origin of SARS-CoV-2" Published in Nature Medicine. (https://www.linkedin.com/pulse/sars-cov-2-could-come-from-labcritique-proximal-origin-billy-zhang?articleId=6651628681431175168#comments6651628681431175168&trk=public_profile_article_view, 2020).
14. Sirotkin, K. & Sirotkin, D. Might SARS - CoV - 2 Have Arisen via Serial Passage through an Animal Host or Cell Culture? *BioEssays*, https://doi.org/10.1002/bies.202000091 (2020).
15. Seyran, M. et al. Questions concerning the proximal origin of SARS-CoV-2. *J Med Virol* (2020).
16. China Honors Ian Lipkin. (https://www.publichealth.columbia.edu/public-health-now/news/china-honorsian-lipkin, 2020).
17. Holmes, E. Academic CV. (https://www.sydney.edu.au/AcademicProfiles/profile/resource?urlid=edward.holmes&type=cv, 2020).
18. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* (2020).
19. Rahalkar, M. &Bahulikar, R. The Abnormal Nature of the Fecal Swab Sample used for NGS Analysis of RaTG13 Genome Sequence Imposes a Question on the Correctness of the RaTG13 Sequence. *Preprints.org*, 2020080205 (2020).
20. Singla, M., Ahmad, S., Gupta, C. & Sethi, T. De-novo Assembly of RaTG13 Genome Reveals Inconsistencies Further Obscuring SARS-CoV-2 Origins. *Preprints*, 2020080595 (doi: 10.20944/preprints202008.0595.v1) (2020).
21. Zhang, D. Anomalies in BatCoV/RaTG13 sequencing and provenance. *Preprint (zenodo.org)*, https://zenodo.org/record/3987503#.Xz9GzC-z3GI (2020).
22. Robinson, C. Journals censor lab origin theory for SARS-CoV-2. (https://www.gmwatch.org/en/news/latest-news/19475-journals-censor-lab-origin-theory-for-sars-cov-2, 2020).
23. Scientific evidence and logic behind the claim that the Wuhan coronavirus is man-made. *https://nerdhaspower.weebly.com* (2020).
24. Zhang, Y. et al. The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I. *bioRxiv*, https://doi.org/10.1101/2020.05.24.111823 (2020).
25. Muth, D. et al. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci Rep*8, 15177 (2018).
26. Schoeman, D. & Fielding, B.C. Coronavirus envelope protein: current knowledge. *Virol J*16, 69 (2019).
27. Lam, T.T. et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* (2020).
28. Liu, P. et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoSPathog*16, e1008421 (2020).
29. Xiao, K. et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* (2020).
30. Zhou, H. et al. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions

at the S1/S2 Cleavage Site of the Spike Protein. *Curr Biol*30, 2196-2203 e3 (2020).
31. Zhang, T., Wu, Q. & Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID19 Outbreak. *Curr Biol*30, 1578 (2020).
32. Yang, X.L. et al. Isolation and Characterization of a Novel Bat Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus. *J Virol*90, 3253-6 (2015).
33. Hu, D. et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg Microbes Infect*7, 154 (2018).
34. Wang, Y. Preliminary investigation of viruses carried by bats on the southeast coastal area (东南沿海地区蝙蝠携带病毒的初步调查研究). *Master Thesis* (2017).
35. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature*579, 265-269 (2020).
36. Lab That First Shared Novel Coronavirus Genome Still Shut Down by Chinese Government. *Global Biodefense*, https://globalbiodefense.com/headlines/chinese-lab-that-first-shared-novel-coronavirus genome-shut-down/ (2020).
37. Song, W., Gui, M., Wang, X. & Xiang, Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoSPathog*14, e1007236 (2018).
38. Li, F., Li, W., Farzan, M. & Harrison, S.C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*309, 1864-8 (2005).
39. Shang, J. et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* (2020).
40. Hassanin, A. The SARS-CoV-2-like virus found in captive pangolins from Guangdong should be better sequenced. *bioRxiv*, https://doi.org/10.1101/2020.05.07.077016 (2020).
41. Zhang, D. The Pan-SL-CoV/GD sequences may be from contamination. *Preprint (zenodo.org)*, DOI: 10.5281/zenodo.3885333 (2020).
42. Chan, Y.A. & Zhan, S.H. Single source of pangolin CoVs with a near identical Spike RBD to SARSCoV-2. *bioRxiv*, https://doi.org/10.1101/2020.07.07.184374 (2020).
43. Lee, J. et al. No evidence of coronaviruses or other potentially zoonotic viruses in Sunda pangolins (Manis javanica) entering the wildlife trade via Malaysia. *bioRxiv*, https://doi.org/10.1101/2020.06.19.158717 (2020).
44. Becker, M.M. et al. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc Natl Acad Sci U S A*105, 19944-9 (2008).
45. Menachery, V.D. et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med*21, 1508-13 (2015).
46. Menachery, V.D. et al. SARS-like WIV1-CoV poised for human emergence. *Proc Natl Acad Sci U S A*113, 3048-53 (2016).
47. Ren, W. et al. Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. *J Virol*82, 1899-907 (2008).
48. Li, X. et al. Emergence of SARS-CoV-2 through Recombination and Strong Purifying Selection. *bioRxiv* (2020).
49. Hou, Y. et al. Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry. *Arch Virol*155, 1563-9 (2010).
50. Yang, Y. et al. Two Mutations Were Critical for Bat-to-Human Transmission of Middle East Respiratory Syndrome Coronavirus. *J Virol*89, 9119-23 (2015).
51. Luo, C.M. et al. Discovery of Novel Bat Coronaviruses in South China That Use the Same Receptor as Middle East Respiratory Syndrome Coronavirus. *J Virol*92(2018).
52. Cui, J., Li, F. & Shi, Z.L. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol*17, 181192 (2019).
53. Wan, Y. et al. Molecular Mechanism for Antibody-Dependent Enhancement of Coronavirus Entry. *J Virol*94(2020).
54. Li, F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J Virol*89, 195464 (2015).
55. Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol*3, 237-261 (2016).
56. Shang, J. et al. Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A*117, 11727-11734 (2020).
57. Hoffmann, M., Kleine-Weber, H. & Pohlmann, S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol Cell*78, 779-784 e5 (2020).
58. Coutard, B. et al. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res*176, 104742 (2020).
59. Claas, E.C. et al. Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet*351, 472-7 (1998).
60. Watanabe, R. et al. Entry from the cell surface of severe acute respiratory syndrome coronavirus with cleaved S protein as revealed by pseudotype virus bearing cleaved S protein. *J Virol*82, 11985-91 (2008).
61. Belouzard, S., Chu, V.C. & Whittaker, G.R. Activation of the SARS coronavirus spike protein via sequential

proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci U S A*106, 5871-6 (2009).
62. Kido, H. et al. Role of host cellular proteases in the pathogenesis of influenza and influenza-induced multiple organ failure. *BiochimBiophys Acta*1824, 186-94 (2012).
63. Sun, X., Tse, L.V., Ferguson, A.D. & Whittaker, G.R. Modifications to the hemagglutinin cleavage site control the virulence of a neurotropic H1N1 influenza virus. *J Virol*84, 8683-90 (2010).
64. Cheng, J. et al. The S2 Subunit of QX-type Infectious Bronchitis Coronavirus Spike Protein Is an Essential Determinant of Neurotropism. *Viruses*11 (2019).
65. Ito, T. et al. Generation of a highly pathogenic avian influenza A virus from an avirulent field isolate by passaging in chickens. *J Virol*75, 4439-43 (2001).
66. Canrong Wu, Y.Y., Yang Liu, Peng Zhang, Yali Wang, Hua Li, Qiqi Wang, Yang Xu, Mingxue Li, Mengzhu Zheng, Lixia Chen. Furin, a potential therapeutic target for COVID-19. *Preprint (chinaXiv)*, http://www.chinaxiv.org/abs/202002.00062 (2020).
67. Zeng, L.P. et al. Bat Severe Acute Respiratory Syndrome-Like Coronavirus WIV1 Encodes an Extra Accessory Protein, ORFX, Involved in Modulation of the Host Immune Response. *J Virol*90, 6573-6582 (2016).
68. Lau, S.Y. et al. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerg Microbes Infect*9, 837-842 (2020).
69. Liu, Z. et al. Identification of common deletions in the spike protein of SARS-CoV-2. *J Virol* (2020).
70. Ge, X.Y. et al. Detection of alpha- and beta coronaviruses in rodents from Yunnan, China. *Virol J*14, 98 (2017).
71. Guan, Y. et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science*302, 276-8 (2003).
72. Ge, X.Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*503, 535-8 (2013).
73. Lau, S.K. et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A*102, 14040-5 (2005).
74. Kam, Y.W. et al. Antibodies against trimeric S glycoprotein protect hamsters against SARS-CoV challenge despite their capacity to mediate FcgammaRII-dependent entry into B cells in vitro. *Vaccine*25, 729-40 (2007).
75. Chan, J.F. et al. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. *Clin Microbiol Rev*28, 465-522 (2015).
76. Zhou, J., Chu, H., Chan, J.F. & Yuen, K.Y. Middle East respiratory syndrome coronavirus infection: virus-host cell interactions and implications on pathogenesis. *Virol J*12, 218 (2015).
77. Yeung, M.L. et al. MERS coronavirus induces apoptosis in kidney and lung by upregulating Smad7 and FGF2. *Nat Microbiol*1, 16004 (2016).
78. Chu, D.K.W. et al. MERS coronaviruses from camels in Africa exhibit region-dependent genetic diversity. *Proc Natl Acad Sci U S A*115, 3144-3149 (2018).
79. Ommeh, S. et al. Genetic Evidence of Middle East Respiratory Syndrome Coronavirus (MERS-Cov) and Widespread Seroprevalence among Camels in Kenya. *Virol Sin*33, 484-492 (2018).
80. Sia, S.F. et al. Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. *Nature* (2020).
81. Ren, W. et al. Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis. *J Gen Virol*87, 3355-9 (2006).
82. Yuan, J. et al. Intraspecies diversity of SARS-like coronaviruses in Rhinolophus sinicus and its implications for the origin of SARS coronaviruses in humans. *J Gen Virol*91, 1058-62 (2010).
83. Ge, X.Y. et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. *Virol Sin*31, 31-40 (2016).
84. Hu, B. et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoSPathog*13, e1006698 (2017).
85. Luo, Y. et al. Longitudinal Surveillance of Betacoronaviruses in Fruit Bats in Yunnan Province, China During 2009-2016. *Virol Sin*33, 87-95 (2018).
86. Kuo, L., Godeke, G.J., Raamsman, M.J., Masters, P.S. & Rottier, P.J. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. *J Virol*74, 1393406 (2000).
87. Drexler, J.F. et al. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J Virol*84, 11336-49 (2010).
88. Agnihothram, S. et al. A mouse model for Betacoronavirus subgroup 2c using a bat coronavirus strain HKU5 variant. *mBio*5, e00047-14 (2014).
89. Johnson, B.A., Graham, R.L. &Menachery, V.D. Viral metagenomics, protein structure, and reverse genetics: Key strategies for investigating coronaviruses. *Virology*517, 30-37 (2018).
90. Wu, K., Peng, G., Wilken, M., Geraghty, R.J. & Li, F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J Biol Chem*287, 8904-11 (2012).
91. Follis, K.E., York, J. & Nunberg, J.H. Furin cleavage of the SARS coronavirus spike glycoprotein

enhances cell-cell fusion but does not affect virion entry. *Virology*350, 358-69 (2006).
92. Yount, B., Denison, M.R., Weiss, S.R. & Baric, R.S. Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. *J Virol*76, 11065-78 (2002).
93. Yount, B. et al. Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc Natl Acad Sci U S A*100, 12995-3000 (2003).
94. Almazan, F. et al. Construction of a severe acute respiratory syndrome coronavirus infectious cDNA clone and a replicon to study coronavirus RNA synthesis. *J Virol*80, 10900-6 (2006).
95. Scobey, T. et al. Reverse genetics with a full-length infectious cDNA of the Middle East respiratory syndrome coronavirus. *Proc Natl Acad Sci U S A*110, 16157-62 (2013).
96. Almazan, F., Marquez-Jurado, S., Nogales, A. & Enjuanes, L. Engineering infectious cDNAs of coronavirus as bacterial artificial chromosomes. *Methods Mol Biol*1282, 135-52 (2015).
97. Thao, T.T.N. et al. Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* (2020).
98. Oldfield, L.M. et al. Genome-wide engineering of an infectious clone of herpes simplex virus type 1 using synthetic genomics assembly methods. *Proc Natl Acad Sci U S A*114, E8885-E8894 (2017).
99. Vashee, S. et al. Cloning, Assembly, and Modification of the Primary Human Cytomegalovirus Isolate Toledo by Yeast-Based Transformation-Associated Recombination. *mSphere*2(2017).
100. Xie, X. et al. An Infectious cDNA Clone of SARS-CoV-2. *Cell Host Microbe*27, 841-848 e3 (2020).
101. Roberts, A. et al. A mouse-adapted SARS-coronavirus causes disease and mortality in BALB/c mice. *PLoSPathog*3, e5 (2007).
102. Roberts, A. et al. Animal models and vaccines for SARS-CoV infection. *Virus Res*133, 20-32 (2008).
103. Takayama, K. In Vitro and Animal Models for SARS-CoV-2 research. *Trends Pharmacol Sci*41, 513517 (2020).
104. Wang, Q. hACE2 Transgenic Mouse Model For Coronavirus (COVID-19) Research. *The Jackson Laboratory Research Highlight*, https://www.jax.org/news-and-insights/2020/february/introducingmouse-model-for-corona-virus# (2020).
105. Zhang, L. et al. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv*, https://doi.org/10.1101/2020.06.12.148726 (2020).
106. Yurkovetskiy, L. et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *bioRxiv*, https://doi.org/10.1101/2020.07.04.187757 (2020).
107. Korber, B. et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*182, 812-827 e19 (2020).
108. Plante, J.A. et al. Spike mutation D614G alters SARS-CoV-2 fitness and neutralization susceptibility. *bioRxiv*, https://doi.org/10.1101/2020.09.01.278689 (2020).
109. Poon, L.L. et al. Recurrent mutations associated with isolation and passage of SARS coronavirus in cells from non-human primates. *J Med Virol*76, 435-40 (2005).
110. Pervushin, K. et al. Structure and inhibition of the SARS coronavirus envelope protein ion channel. *PLoSPathog*5, e1000511 (2009).
111. Nieto-Torres, J.L. et al. Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoSPathog*10, e1004077 (2014).

Global Journal of Computer Science and Technology: G Interdisciplinary
Volume 21 Issue 2 Version 1.0 Year 2021
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Application of Social Media Devices: Effective Instruments for Library Services Provision to Physically Challenged Academic Library Users in Nigeria

By Festus Onifade

Abstract- Provision of library services is very important to all and sundry particularly physically challenged library users, the study therefore based on the application of social media devices as an effective instrument for library services provision to physically challenged library users in Nigeria, the study examines the concept of physically challenged person to be someone who cannot carry out normal day-to-day activities, or as a result of their appearances which labeled them physically challenged, the study further preempts on social media concept, common types of social media devices used for library services such as YouTube, WhatsApp, Facebook, Flickr among others, that they are good instruments to publicized library services to physically challenged library users, highlighted some of the library services to the physically challenged library users to be reading, references, reprographics, online, transcription marketing, SDI, CAS, and many more services.

Keywords: *social media, library services, physically challenged, library users, nigeria.*

GJCST-G Classification: *H.2.8*

Strictly as per the compliance and regulations of:

Application of Social Media Devices: Effective Instruments for Library Services Provision to Physically Challenged Academic Library Users in Nigeria

Festus Onifade

Abstract- Provision of library services is very important to all and sundry particularly physically challenged library users, the study therefore based on the application of social media devices as an effective instrument for library services provision to physically challenged library users in Nigeria, the study examines the concept of physically challenged person to be someone who cannot carry out normal day-to-day activities, or as a result of their appearances which labeled them physically challenged, the study further preempts on social media concept, common types of social media devices used for library services such as YouTube, WhatsApp, Facebook, Flickr among others, that they are good instruments to publicized library services to physically challenged library users, highlighted some of the library services to the physically challenged library users to be reading, references, reprographics, online, transcription marketing, SDI, CAS, and many more services. It revealed numerous uses of social media devices for library services delivery to physically challenged library users, benefits of social media devices, special training for library staff to help physically challenged library users, the study upholds that, the challenges of using social media devices as lack of awareness, bandwidth problem, technophobia, lack of maintenance culture, sporadic power supply, all of these were identified to be serious obstacles to effectively applied to social media devises for physically challenged library users. The study concludes that there should be specially trained staff, particularly for these users and diverse types of social media should be used to communicate to physically challenged library users.

Keywords: social media, library services, physically challenged, library users, nigeria.

I. Introduction

The core function or value of a library is the provision of library services and resources to its users in the right format, regardless of abilities or disabilities, or physical form of users. To successfully do that there is a need for the application of social media to the provision of library service for physically challenged library users. Anjiode (2010) noted that there is a person with physically challenged disabilities all over the part of the world and at all levels in every society. The physically challenged library has a substantial long-term adverse effect on one's ability to carry out normal day-to-day activities. Iroeze, Umunnakwe and Eze (2017).There are more than 19million physically challenged persons in Nigeria Society Adamu (2009).

According to World Health Organization (WHO) (2001) relates physically challenged or disabilities as "any restriction or lack (Which resulting from an impairment) of ability to perform an activity in the manner or within the range considered normal for a human being. Remesh and Singh (2001) emphasize on physically challenged or disabilities may be cognitive, mental, sensory, emotional, and developmental or sometimes a combination of these. In clear terms, people or library users are label as disabled, handicapped, and physically challenged because they look different from the rest of the persons in society on account of their appearance or behavior or capacity to learn, develop or do certain things themselves. Rehabilitation council of India (1992) examined disabled person as one whom in his/her society is regarded as disabled, because of a difference in appearance and or behavior, in combination with a functional limitation or an activity restriction. Lawal -Solarin (2012) defines physically challenged as an inability to perform some or all the tasks in daily life or a medically diagnosed condition that makes it difficult to engage in the activities of daily life.

World Book Encyclopedia (2004) noted that some people are born physically challenged, while others develop them later in life. There are however many types of challenged or disabilities: both physically and mentally, and they vary greatly in the cause of degrees and treatment. Common disabilities include blindness, deafness, and deformity loss of limbs, mental illness, and mental retardation, muscular, nervous, and sensory disorder, Saliu, Rabiu, and Alabi (2016). Most challenged persons suffer rejection, isolation, discrimination, humiliation, segregation, and maltreatment from other members of the society Adesokan (2003). The physically challenged library users/students encountered barriers in their quest for access to library services. Viney (2006) noted that they

Author: J.C Pool Library, the Nigerian Baptist Theological Seminary Ogbomoso, Oyo State. Nigeria. e-mail: Onifadefestus19@gmail.com

encounter physical access limitations such as retrieving books from library shelves. Okoli (2010) emphasized the poorly built architectural buildings which have discouraged many physically challenged library users from having the right to provision of information services. However, if this is what physically challenged library users encountered or faced, then there is a need for the provision of library service to these target groups of library users by applying social media devices for providing library services to physically challenged users in Nigeria. The application of social media in libraries is extensive through providing information services to the users to ease their worry, creating awareness about library event news, new arrivals of books, users orientation programs, library tour connecting with other libraries, and librarians feedback. About the library services Hadagali, Kenchakkanavar, and Tadasad (2019) services include Circulation services, Reference services, information services, Bibliographic service, Abstracting service, indexing services, current awareness services (CAS) selective Dissemination of information (SDI) Reprographic Services, Translation Services, CD-Rom (Compact-Disk Read why memory), Services and Online Services.

Hence, in this information communication technology advances driven society, the library must employ these social media devices to offer better library services to physically challenged library users. Such social media device includes Facebook, WhatsApp, Myspace, Ning, Blogs, Linkedin, Twitter, Youtube, Flickr and Library things, etc.

Consequently, for physically challenged library users to be relevant and belong to the common society and do what a normal physical formed human being does, there are urgent needs for libraries in Nigeria to apply social media device for their library services delivery which this study tends to explore for better and future sustainability.

II. Review of Related Literature

a) *Concept of social media*

Social media have been defined in a variety of ways. Social media is the general name given to every form of social interactions while social network is a subset of it. Burke (2013) specified that social media is the media (content) that one uploads whether a blog, video, slideshow, podcast, newsletter, or eBook. Consider social media as a one-to-many communication method. Though people can respond and comment, the owner owns the content and has to produce (write/record/create) the media yourself. Dewing (2012) further indicated that social media refers to the wide range of internet-based and mobile services that allow users to participate in online exchanges, contribute user-created content, or join online communities.

Cohn (2011) indicated that social media and social networking have been instrumental in many major events around the world. It is fair to say that social networking is a subcategory of social media. Social media is the use of web-based and mobile technologies to turn communication into an interactive dialogue, on the other hand, is a social structure with people who are joined by a common interest.

This is further stated by Dodson (2012) that social networking involves direct communication and requires a conversation between two or more parties. Social media offers channels by which the content can be acted upon. The idea behind the act of social networking seems to be the idea of building networks of like-minded and influential individuals in a related field or area of interest to, in fact, gain something out of it all. Social media, on the other hand, is the actual vessel in which all this "networking" takes place.

Social media has been defined as website which allow profile creation and visibility of the relationship between users, and which are also referred to as social media sites Diga and Kelleher (2009). Evans (2011) further emphasized on social media sites allow users to generate their content, commonly referred to as users- generated content to share their experience inmany different ways. Begum and Parvin (2019) considered social media as social networking sites that are virtual communities where users can create individuals to communicate. Hadagali, Kenchakkanavar, and Tadasad (2019) observed that some authors fail to distinguish between social media and that of social networking and use these terms synonymously in a different context. However, Dina (2011) differentiates between social media and social networking via LinkedIn Group (Freelance Editing Network). Social media (noun) is the 'tool' and social networking (verb) is what you do with that tool and how you use it.

b) *Common Types of Social Media Devices Used For Library Services*

Common types of social media devices were considered based on their popularity and widely used among libraries to provide accessibility and successful delivery to library users, such as; YouTube, Whatsapp, Flickr, Facebook, Weblog, Twitter, MySpace, LinkedIn, and Library Thing among others. Few among social media devices selected as instruments for Library services provision are discussed below:

i. *YouTube*

YouTube is a popular instrument that has been seen as a potential and capable instrument for delivering library services to the users in all kinds of libraries particularly physically challenged library users Hadagali, Kenchakkanavar, and Tadasad (2019). Most of the libraries nowadays use YouTube for sharing videos on events, lectures, special talks, library tours, seminars, training, etc. The study conducted by

Garoufallou and Charitopoulou (2011) observed that YouTube is a widely used social media devices by the students to do the following:

- To help the librarians to post videos on conferences, workshops, library events, library tours, or bibliographic instruction for the benefit of users.
- It helps in users' education/bibliographic instruction videos can be shown during the class hours on how to use a test, tool, database, search engine, formula, etc.
- YouTube allows the librarians to save favorite subject-related videos and also enables them to create a playlist and share them among the users.

YouTube however according to Ezeani and Igwesi (2012) helps institutions in Nigeria, to communicate important highlights of inaugural lectures, conferences, and workshops to library users physically challenged inclusive.

ii. *WhatsApp*

Whatsapp is one of the popular and commonly used social media devices, this is a device where Instant messages can be gotten and it allows users to send text messages, images, videos, etc. to each other for free. WhatsApp allows and gives rooms for a lot of thing such as:

- To attach files through. Word document, PDF, PPT, etc.
- WhatsApp allows the users to stay updated on library events.
- WhatsApp enables the librarians to post news about the library, images, and videos of the library orientation program and library events.
- It also allows librarians to interact with the users through discussion groups, image tagging, and receive comments, feedback, and suggestions to improve the Library services.
- It allows the librarians to provide the most important services i.e. 'Ask a Librarian' without being physically present in the library. Hadagali, Kenchakkanavar and Tadasad (2019).

iii. *Flickr*

Flickr is another common and popular social media device known for image sharing application which is being widely used to share images within groups or communities. Flickr is popularly known as a photosharing application that also enables users to post videos (Dickson and Holley, 2010). Garoufallou and Charitopoulou (2011) submitted that Flickr was the second most preferred Web 2.0 instrument used by students Flickr also do the following things to disseminate library services to patrons in the library:

- Flickr allows the librarians to upload and share images of the library events/programs / and activities to the users
- It enables to tag images with keywords. These tags will further be useful for locating the relevant images.
- It helps Librarians to create discussion groups and post-academic / subject-related photographs.
- It enables the librarians to post videos on the virtual tour of the library.

According to Ezeani and Igwesi (2012), Flickr can be used as an instrument to share and distribute new images of library collections. Cover pages of new arrivals of both books and journals can be disseminated to users through Flickr. It can also be used to enlighten users on topical issues such as the different pictures of emblems of the political parties in Nigeria.

iv. *Facebook*

Facebook is a for-profit corporation and online social media service founded by Mark Zukerberg sometime in February 2004. Facebook allows its users to create an online profile, add friends and enable them to post and view each other's profiles (Ellison et al., 2007). Facebook is one of the widely used social media devices used mainly for interaction and sharing. Using Facebook applications following library services may be provided to the end-users:

- Facebook applications enable the librarians to access the contents of the library catalog without actually going to the library and visiting the library's website.
- Facebook can be used to share academic activities, essentially for providing information literacy programs/orientation to users- new intake students
- Facebook enables the librarians to provide information to the users on the events, activities and programs, new arrivals of books (through posting videos and providing links)
- Facebook links the users to online tutorials on how to use a device, education programs, etc.
- It helps to advertise library events and create online library study groups for the users.
- Facebook facilitates access to question paper banks, wherein the user can make use of the question papers without geographical limitations.

Facebook is librarian-friendly with many applications like JSTOR search, World Cat, and many more. Librarians can interact with users to know their information need. Libraries try to link some of these specialized library applications to Facebook Ezeani and Igwesi (2012).

v. *Weblog*

A weblog is a website that is common, popular, and usually maintained by an individual, with regular entries of commentary, descriptions of events, or other materials such as graphics, or videos. Entries are commonly displayed in reverse chronological order (Wikipedia.org) Boxen (2008) defines a Blog as a

webpage consisting of user-supplied content in reverse chronological order. Web publishing has become easy because of its simplicity in publishing the contents and records the comments by the other persons.

Hadagali, Kenchakkanavar, and Tadasad (2019) identified the following library services that may be rendered through the library to the users using Weblog:

- Weblogs enable user interaction which further allows students to provide feedback on the services provided by the Librarians.
- Librarians can flash the news about the library events/programs/activities which take place at the library to users.
- Weblogs allow Librarians to create different subject guides to fulfill the demands of the students.
- Librarians can create blogs detailing the programs of the projects undertaken on the renovations occurring at the library.
- A weblog can be used to interact with users offering their own choice of contributions like debate and interaction.
- Librarians and users can get current information about various subjects (through alerting services) in general to specific through weblogs. it could also be used to market library services.

c) *Library services provision to physically challenged library users*

Provisions of library services to students or to users that are physically challenged are very important. A library is a service provider institution and the university library is not exempted. University libraries in the spring of knowledge and information provided will always be valued at a premium. In the modern significantly and socially vigilant society, especially when the grains of right to information are gaining much currency. The library service which brings the staff in contact with users to provide the right information to the right user at the right time and to help in finding out resources and providing required library services should be emphasized. Some important kinds of library services offered by libraries to physically challenged library users are:

i. *Reading services*

Ayiah (2007) observed that the provision of reading services is highly essential entirely dependent on resource persons and volunteer students. There are no readers employed to serve visually challenged Students but as part of the resource person's work schedule, they are expected to provide that service whenever the student needed it. As Craddock (2001) noted that accessing information is of utmost importance for anyone pursuing an academic program. If a physically challenged student is delayed access to information simply because a reader was not available or a resource person would not complete the task, then the ability of the physically challenged student to complete an assignment in time interferes with added consequences. Rayini (2017) emphasized the need for libraries to develop a strategy for engaging readers and providing them with training. Training users in the use of new services and in new technologies that support these services is essential. Libraries should make users aware of new services or changes to existing services. Many libraries for the blind accomplish this through alternative-format newsletters or special training sessions. The Internet can also be an effective mechanism for introducing users to a new service and guiding them through it.

ii. *References services*

This service is highly interactive and brings the visually challenged students closer to the resource persons and this brings a lot of problems since these resource persons are not professional librarians. Ayiah (2007)academic Library is supposed to perform the following functions: teaching, research, publication, conservation of knowledge and ideas, extension and service, and interpretation, the services to be provided are listed below by (Kumar 1996).

- Providing instruction in the use of library, general and specific information,
- Assistance in the location (or searching) of documents or use of library catalog understanding of reference books
- Reader's advisory service,
- Compilation of bibliographies, preparation of indexing and abstracting services,
- Reservation of documents- In case a document has not been loaned, then a user who needs it can get it reserved so that when the document is returned then the user can be informed and he can get it issued,
- Interlibrary loan – ILL refers to a request for a document not available in the library. Whatever might be the nature of the library, but it should take advantage of borrowing books from other libraries, etc.

iii. *Transcription services*

Transcription work is done at the Braille Library situated in Balme Library which doubles as a resource center for students with Disabilities. This may seem to be against the established practice. Craddock (2001) believes that a library serving physically challenged library users or visually challenged students must provide such readers with the information they require in the appropriate formats and insufficient time for it to be useful.

iv. *Marketing and Advocacy*

Rayini (2017) observed print-disabled users are very often among the poorest of the poor in many countries, usually isolated from others with similar disabilities. Because the majority of people who are

blind tend to be elderly, they are reliant on libraries for the blind to aggressively advocate on their behalf. In addition, special techniques are required to market to this group and to make them aware of library service opportunities. All staff should recognize their role and responsibility in promoting and advocating for the needs of these users. Marketing and advocacy initiatives must also engage a wide group of stakeholders, including other agencies and the general public who could support the work of the library.

Irvall and Nielsen (2005) considered a person in a wheelchair or using crutches or a walker should be able to enter through the door and pass through security checkpoints. A blind person with awalking guide dog should also be able to enter without encountering obstacles. They emphasized that: Sufficient space in front of the door to allow a wheelchair to turn around, entrance door wide enough to allow a wheelchair to enter, Automatic door opener reachable by a person in a wheelchair, No doorsteps -- for easy wheelchair access, Glass doors marked to warn visually impaired persons, Security checkpoints possible to pass through with a wheelchair/walker or other mobility aids, Stairs and steps marked with a contrasting color, Pictogram signs leading to elevators, Well lighted elevators with buttons and signs in Braille and synthetic speech, Elevator buttons reachable from a wheelchair.

v. *Indexing services*

Indexing services is one of the paramount services provided to different kinds of library users regardless of the ability or disability of the users. An index is an organized tool to the text of any reading subject matter or the contents of other collected document materials, covering a series of entries, with headings shaped in alphabetical or other chosen order, and with references to reveal where each item indexed are located. Thus, it is cogently perceived that an index is a list systematically arranged providing enough details about each item so that it can be figured and brought out (British Standards Institutes, 1964). There are various types of indexes: Book indexes, Index of collections, Periodical indexes, Newspaper indexes, Citation indexes.

vi. *Abstracting services*

The modern era is christened as an era of the information age. There is no branch of knowledge where a large quantum of information has not been generated. Therefore, huge sources are documented but it creates the problem of not only organizing knowledge but also in the selection of quality and important information products. An abstracting service intends to facilitate the summarization of new documents and inform the users about the topical areas of interest to them, (Ashworth 1979), the various kinds of abstract are reorganized by their scope on coverage. The commonly known abstracts are indicative and informative. Other kinds of abstracts are author abstract, locative abstract, telegraphic abstract, auto abstract, etc

vii. *Selective dissemination of information services (SDIs)*

This is a service provided where tools and resources are used to keep clientele informed of new resources on specified topics Ambali, Usman, and Adesina (2018).A service that is personal rather than common; it has evolved out of the development of key term indexes, viewed as personalized CAS/SDI services high-interest areas. It is the service concerned with the "channeling of new ideas of information for whatever source to those points within the organization where the probability of usefulness, in connection with current work of interest. The SDI encapsulates a strategy to prepare users' services manual and computerized methods are in operation mainly depending upon the level of automation of library services.

viii. *Current awareness services (CAS)*

CAS is a service to make the users aware of the availability of recent publications. CAS can be a list of journal titles or contents of periodicals or a list of newly arrived documents. The libraries announce regularly the list of new additions of books, list of periodicals, and current contents of periodicals to provide this service. Pertinently current information is available in newsletters, newspapers, journals, and other micro documents (Pal, 2004). Vickery defines CAS as "more frequently and more adequately met by circulation than by retrieving current journals, newly received books and reports, abstract bulletin and the like being fed and scanned by users.

ix. *Reprographic services*

The term reprography was used for the first time in 1954 as a generic term for all kinds of facsimile reproduction of documents, covering in its scope, processes, and techniques related to photocopying, microcopying, blueprinting, electronic copying, thermo copying, dyeline reproduction, etc. In earlier times all these processes and techniques were called copying which without a doubt was wider in its ambit and did encompass copy typing and duplicating in the offices, photocopying in the libraries, and blueprinting in drawing office. Prashar (2003). According to Bose, (1972), a reprographics service is a group of mechanical devices whereby one or multiple copies of a document can be made through the copying and duplicating process. Reprography covers not only devices but processes and techniques and includes mechanical along with photographic, thermal, and electronic processes too. According to Hawken, (1966), reprography is a term applied to draw attention to the copying processes and methods applied for both copying and duplicating documents. The role of reprography is instrumental in communication. In the modern era, there is the aggrandizement of information,

reprography facilities, and information scientists who bring home a researcher, right at his desk, the literature pertinent to his area of subject for numerous sources.

x. *CD-ROM (Compact-Disc Read Only Memory) Services*

CD-ROM (Compact-Disc Read Only Memory) Services CD-ROM is one of the storage media developed due to information explosion as well as the urgency for quicker processing and accessing of information. CD-ROM is known for the revolution in information media. CD-ROM technology has proven itself as a blessing for libraries in facilitating library service to the users. CD-ROM is more accessible for searching the information and as it occupies less space and has a large storing capacity, it is more suited for university libraries. (Khan, 1997)

xi. *Online services*

Online Services Online system is also a revolution, in which the user is provided a seat at a terminal connected to a database and can interact with the computer. Shaping search strategies based on the response, the searcher has quick access to the database. The user can interrogate the computer directly. The output can be printed out or displayed on the screen. The computer acts as a storage place for the accumulation of information. Online services indicate that users have access to information through the usage of video displayed keyboards. The user can operate the keyboard, give the command and the outcome can be revealed on a video display or cathode ray tube. If the users want a printed record then he has to push a button and the record will be printed out. Sharma and Grover, (2004) observed that, If the user has numerous indexes and abstracts available to search at a computer terminal,

Libraries must not discriminate against individuals with disabilities and shall ensure that individuals with disabilities have equal access to library resources. To ensure such access libraries may provide individual with disabilities with services such as extended loan periods waived late fines extended reserves periods, library card for proxies, books by mail, reference services by fax or e-mail, home delivery services remote access to the OPAC, remote electronic access to library resources, volunteers readers in the library, volunteers technology assistants in the library, American Sign Language (ASL) interpreter or captioning at library programs and radio reading services. Sources: http://www.ala.org/asgcla/resources/libraryservices

d) Uses of social media devices for library services delivery to physically challenged students

Social media devices have the potential to facilitates a much closer relationship between libraries and other patrons, particularly physically challenged library users; Semode, Ejitagha, and Baro (2017) conducted a study on social networking sites: changing roles skill and use by the library in tertiary institutions in Nigeria. The findings of the study revealed that the librarian should use social media devices for notifying news, share information about library resources and library events to users, they should also write good posts on the library Facebook page, this will attract user to like the library Facebook page, and this can also be extended to physically challenged students.

Bhatt and Kumar (2014) researched opinion the use of social networking tools by librarians; it was revealed from the study that the use of social media devices is important to capture the attention of online users and help in distance learning and knowledge sharing.

ALA (2001) emphasizes on Libraries marking good use of social media and web 2.0 application the study revealed that libraries of all types are increasingly using social media tools to connect with library users and to make library program and service accessible.

Dickson and Holly (2018) noted that social networking can see an effective method of student outreach in academic libraries if libraries take care to respect student's privacy and to provide equal coverage for all subject areas.

Begum and Parvin (2019) conducted a study on incorporating social media into library service: present scenario at East-West University library, the findings of study revealed that libraries can be the best promoters of their materials by proactively sharing their resources and services through social media.

Ayiah (2007) conducted a study on the provision of library and information services to visually challenged students in University of Ghana the findings of the study revealed that there is the need for specialized training on how to serve in general cuts across all aspects of the person whose duties necessitate dealing with special need people particularly physically challenged library users.

Ezeani and Igwesi (2012) conducted a study on using social media for dynamic library service delivery, the findings of the study revealed that librarians should educate patrons on the use of these social media device to adapt to new ways of accessing, communicating, and sharing knowledge and that the brilliant use of the social cyberspace promotes opens access to knowledge.

Burclaff and Johnson (2016) did an overview on teaching Information literacy through social media: An exploration of connectivism, the finding of the study revealed that students frequently connect to other people's resources and information using social media devices such as Facebook, Twitter Linkedin, etc.

Lederer and Feldman (2012) said studies have shown that the students prefer contacting a librarian virtually particularly physically challenges library users as long as the platform is efficient.

Zaid and Zaid (2017) Emphasized in the study, the exclusion of persons with visual impairment in Nigerian Academic librarians' "website" the outcomes of the study revealed that creating a webpage for disability services and incorporate accessibility statement on the library homepage and in all library publications and campus materials to allow disability library users have accessibility.

Hadagali, Kenchakkanavar, and Tadasad (2019) Social medial platforms effective tools to provide innovative library service in a university environment. The outcomes of the study revealed that the usage of social media in university libraries in India is still in the formative stage and needs to gear up to meet the growing expectations of the users.

e) Benefits of social media devices to physically challenged library users

The benefits of social media devices to any group of people cannot be overemphasized, regardless of age, sex, religion, tribes, education, and so on, therefore, physically challenged library user should be able to use and enjoyed social media devices to maximal level, Thus, this study agreed with Vasquez and Nevada (2013) who identified some benefits of social media assistive technology as follows:

1. Social media devices help physically challenged library users to connect socially with others when they are not able to leave home because of their conditions.
2. Social media devices also help physically challenged library users to connect with e.g visually impaired people using Twitter to communicate with friends instead of crowded social situations where eye contact is difficult.
3. Social media devices connect with others who share health medical conditions and similar life experiences.
4. Social media devices provide a platform to educate disable library users without leaving home.
5. Social media allows the opportunity for messages to go viral and increased social skills networks
6. Social media provides independence and self-expression opportunities for physically challenged library users
7. Individual physically challenged library users to research and educate themselves
8. Social media devices platform like Linkedin improve employability options for physically challenged library users
9. Social media devices highly motivating way to improve technology skills and implement assistive technology: increase digital competence of physically challenged library users. Sources:http://nevadaddcouncil.org/wp-content/uploads/2c13/10/social-media-disability-conference.pdf.

In a similar vein, Semode, Ejitagha, and Baro (2012) highlighted some benefits of social media devices as follows:

1. Twitter as social media devices distribute library news and provide customer information and services also build connections with other libraries, librarians, and institutions.
2. Facebook as social media devices benefits physically library users as to distribute library news and information more social and less formal than Twitter –share photographs and run competitions, engagement with students promote general library collection, digital and archive special collections and information literacy to physically challenged library users.
3. More also, social media devices provide an arena for students and course facilitators to pin reviewed and recommended reading for a particular topic and develop communities with other online libraries to physically library users.

f) Special training for library staff to help physically challenged library users

Ayiah (2007) affirms that dealing with individuals with visual requires that the individual understands, read and write in their preferred mode of communication which is Braille. The person must be someone who can read and use Braille in communicating with these students. Braille varies from a simple alphabetical notation to specialized notations for computers, foreign languages, music, mathematics, and other disciplines. Training in reading and writing Braille, the individual will be in a better position to use assistive technology, including screen review software such as JAWS; scan and read systems such as OpenBook and Kurzweil 1000 and possibly portable notetakers such as Braille Lite series, Braille 'n speak, Type 'n speak, etc. to effective and efficient communication with them and also using that to provide them with the requisite information timely.

Rayini (2017) emphasized that; libraries need to develop a strategy for engaging physically challenged readers and providing them with training. Training users in the use of new services and in new technologies that support these services is essential. Libraries should make physically challenged users aware of new services or changes to existing services. Many libraries for the blind accomplish this through alternative-format newsletters or special training sessions. The Internet can also be an effective mechanism for introducing users to a new service and guiding them through it.

Irvall and Nielsen (2005) pointed out that accessibility to the library should be a clearly defined management responsibility. A designated employee should act as a liaison person with disability groups and support organizations. It is, however, important that all staff be knowledgeable about various types of physically

challenged or disabilities and how to best assist the users. Staff should also communicate directly with the patron and not through a caregiver. Examples of appropriate staff training include:

- Invite persons with disabilities to staff meetings to talk about their needs as library users
- Distribute e-mails and/or other information to staff regularly about library services to specific physically challenged or disability groups.
- Include information about services to special user groups in the orientation/orientation package for new staff.

g) Challenges of using social media devices by physically challenged library users

Getting social media devices used by physically challenged library users can be a challenge at times, can be difficult and tasking, some of the challenges faced by physically challenged library users as highlighted by Ezeani and Igwesi (2012) as follows:

Lack of awareness: users of academic library especially physically challenged users are not aware of the protocols involved in social communication, these categories of users are possibly not or may be unaware that there is a subject specialist in their discipline that can help them used these social media devices, therefore here comes the works of librarians to initiate contact with these group of users.

Bandwidth problem: Still on these challenges, most institutions have limited bandwidth to support this practice; poor connectivity can frustrate effective online participation of physically challenged library users, on the part of institutions and the users.

Technophobia: Oftentimes, many users particularly physically challenged library users are afraid of handling computers or social media devices for harness library service, having no option they make use of traditional library services in their comfort zone, thereby they are not eager to embrace change perhaps it may be because of their health conditions.

Lack of maintenance culture: The most serious thigh to do is the issue of constant maintenance in all aspects in Nigeria, maintenance culture is seriously lacking in most institutions in developing countries. The few available technologies are in moribund conditions that may not support remote access to library service/information services.

Sporadic power supply: The low or irregular power supply in Nigeria constantly discourage physically library users from accessing library services.

Lack of training: Lack of staff training to handle the physically challenged user is another concern, it to be attended to

However, Semode, Ejitagha, and Baro (2017) highlighted challenged associated with using social media device as follow: Bad network, lack of time, power failure, inadequate competent of staff to handle the social media devices, some library users are not on any of social network, among others, these are evidence that using social media devices may be difficult.

III. Recommendation and Conclusions

This was conducted to cross-examine the application of social media devices as effective instruments for providing library services to physically challenged library users in Nigeria. It was observed on assumption that the physically challenged library users have more social media devices to connect with library services can offer them via the social media devices, it was clearly emphasized that these categories of users oftentimes faced a lot of humiliation discrimination, segregation, and rejection as a result of these the provision of library services becomes difficult for them to accessed and if going through these barriers in their quest for access to library services, then libraries should devise a means of communicate to them through the social media instruments by libraries/librarians connecting to other libraries to help their physically challenged library users. However, it was noted that various library services physically challenged library users to have access to such as reference, referral, indexing and abstracting, selective dissemination of information (SDI), current awareness services (CAS) online etc. which can be delivered to the physically challenged users at their convenience through the use of social media devices application such as Facebook YouTube, Whatsapp, Flickr among others by attaching files, through the document, PDF, PPT, HTM, etc interaction also take place through discussion groups image tagging. Since there are numerous of social media devices which can help libraries/librarians to reach out to their users, particularly physically challenged library users, the library should therefore employ the opportunities to get these target group of users feels important by mailing library services to them. Hence, the study clearly defined management responsibility that a designated employee should act as a liaison person with disability groups and support organization, it is, however, necessary that all staff be knowledgeable about various types of physically challenged or disabilities and how to best assist them, the study earmarked the following should be appropriate staff training include to:

- ✓ Invite a person with physically challenged to a staff meeting to talk about their needs as library users
- ✓ Communicate directly to them not third party or care-givers
- ✓ Distribute e-mail and other information to staff regularly about library services
- ✓ Give information about services to special user groups in orientations.

Finally, libraries and librarians should give physically challenged library users priority and preferential treatment as it is been given to other normal library users and to be sure they have right and privilege to library services.

References Références Referencias

1. Adamu S. (2009). New life for the physically challenged, Published in *The Nation Newspaper*. Available at: http://thenationonlineng.net/web2/articles/24677/1/Newlifeforthephysicallychallenged/Page1.html. Retrieved on July 20, 2020
2. Adesokan, Z. A. (2003). Educational Needs of the Handicapped: A Musical Perspective. In Adedoja, T.A and Ajobiewe, Theo (ed.). Issues in Coping with Disability.
3. Ambali, Z. Usman, O. Saliu; A and Adesina, F (2018). An Assessment of the Resources and Services Provision for the Disabled Library Users in the University of Ilorin and Federal College of Education (Special) Oyo. *Library Philosophy and Practice (e-journal)*. Available at: https://digitalcommons.unl.edu/libphilprac/1818 retrieved on July 20th, 2020.
4. Ayiah, E (2007) "Provision Of Library and Information Services to the Visually Challenged Students in the University Of Ghana, Legon". *Library Philosophy and Practice (e-journal). Available at:* http://digitalcommons.unl.edu/libphilprac/1369 retrieved on July 20th, 2020.
5. Anjiode, K.J. (2010). Information resources and services to physically challenge in Plateau State Special Educational Institutions. Samaru Journal of Information Studies, 10(1&2).
6. ALA (2001). Libraries making good use of social media and Web 2.0 applications. Available at: http://www.ala.org/news/mediapresscenter/americaslibraries/socialnetworking retrieved on July 20, 2020.
7. Begum, D, and Parvin, S (2019) "Incorporating Social Media into Library Services: Present The scenario at East-West University Library" (2019). *Library Philosophy and Practice (e-journal).*available at:https://digitalcommons.unl.edu/libphilprac/2746 retrieved on 20th July 2020
8. Bhatt, R. Kumar, A. (2014). Student opinion on the use of social networking tools by libraries: A case study of Jawaharlal Nehru University, New Delhi. The Electronic Library, 32 (5)
9. Bose, H. (1972). Reprography in documentation, In Goyal, S.P. (Ed.). Indian librarianship, New Delhi: Scientific Books Store.
10. Boxen, J. 2008. Library 2.0: a review of the literature. The Reference Librarian, Vol. 49, no. 1
11. Burke, F (2013) Huff post books. Retrieved July 20, 2020, from social Media vs social networking: available at: http://www.huffingtonpost.com/fauzia-burke/social-media-vs-social-ne_b_4017305.html retrieved on 24th, August 2020.
12. Burclaff, N and Johnson, C (2016) "Teaching Information Literacy via Social Media: An Exploration of Connectivism" *Library Philosophy and Practice (e-journal) available at:* http://digital commons.unl.edu/libphilprac/1476 retrieved on 20th July 2020.
13. Cohn, M. (2011) Compukol Connection social media and social Networking: available at: http://www.compukol.com/blog/social-media-vssocial-networking. Retrieved July 20, 2020.
14. Craddock, P. & Wallace M. (2001) Alternative format Material. Library service for the Blind: A manual of Best Practice. National Library for the Blind.
15. Dewing, M. (2012) Social media: An introduction. Canada: Library of Parliament.
16. Dickson, A. and Holley, R. (2010). Social networking in academic libraries: the possibilities and the concerns. New Library World, Vol. 111, no. 11/12.
17. Diga, M., & Kelleher, T. (2009). Social media use perceptions of decision-making power, and Public relations roles. *Public Relations Review, 35*.
18. Dodson, B. (2012) incept result in the difference between social networking vs social media Available at: http://conversationalmarketinglabs.com/blog/2014/04/about-incept/the-difference-between-social-networking-vs-social-media/Retrieved on July 20, 2020.
19. Ellison, N., SteinField, C. and Lampe, C. (2007). Benefits of Facebook friends: Social capital and college Students' use of online social networking sites. *Journal of Computer-Mediated Communication,* Vol. 12.
20. Ezeani, C and Igwesi, U. (2012) "Using Social Media for Dynamic Library Service Delivery: The Nigeria Experience". *Library Philosophy and Practice (e-journal).* Available at: https://digitalcommons.unl.edu/libphilprac/814 retrieved on 25th August 2020.
21. Garoufallou, E. and Charitopoulou, V. (2011) The use and awareness of Web 2.0 tools by Greek LIS students. New Library World, Vol. 112, no. 11/12.
22. Hadagali, G. Kenchakkanavar, Y. and Tadasad, P(2019) "Social Media Platforms: Effective tools to provide innovative library services in a university environment". *Library Philosophy and Practice* Available at: https://digitalcommons.unl.edu/libphilprac/2770 retrieved on 25th August 2020.
23. Irvall, B., & Nielsen, G. S. (2005). Access to Libraries for Persons with Disabilities: Checklist. IFLA Professional Reports. *International Federation of Library Associations and Institutions.* Hague, Netherlands. No. 89.
24. Iroeze, P., Umunnakwe, C. & Eze, J. (2017). Library services to physically challenge in South-East, Nigeria: Origin and challenges. *International Journal of Advanced Library and Information Science.* 5(1).

25. Khan, S. & Bhatti, R. (2012). Application of social media in marketing of library and information services: A case study from Pakistan Weblog, 9 (1), available at: http://www.webology.org/2012/v9n1/a93.html retrieved on 25th August 2020
26. Khan, M. (1997). Information organization and communication. New Delhi: Ess Publication.
27. Kumar, K., (1996). Reference service. New Delhi: Vikas Publication.
28. Hawken W. (1966). Copying methods manual. Chicago: American Library Association.
29. Lawal-Solarin, E.O, (2012). A survey of library and information services to physically challenge students in academic libraries in Ogun State, Nigeria. *Library philosophy and practice (e-journal).* Available http://unllib.unl.edu/LPP/ retrieved on 25th August 2020.
30. Lederer, N., & Feldmann, L. (2012). Interactions: A study of office reference statistics. Evidence Based Library & Information Practice, 7(2), 5-19. doi: 10.18438/b88k6c
31. Okoli .B. (2010). The Plight of Disabled Nigerians and the need for mass enlightenment.
32. Mobility aid and Appliance and Development Centre (MAARDEC). Available at http://www.maardec.net/THE%20PLIGHT%20OF%20DISABLED%20NIGERIANS%20AND%20THE%20NEED%20FOR%20MASS%20ENLIGHTENMENThtmlretrieved on 25th August, 2020.
33. Prasher, R. (2003). Information and its communication. New Delhi: Medallion Press.
34. Pal, S. (2004). Special library system and information services. New Delhi: Icon Publication.
35. Rayini, J. (2017) "Library And Information Services To The Visually Impaired Persons" *Library*
36. *Philosophy and Practice (e-journal)* available at. http://digitalcommons.unl.edu/libphilprac/1510 retrieved on 25th August 2020.
37. Ramesh, R, and Singh, K. (2011) "Definitional and Enumeration Inconsistencies." Disability: Towards Inclusive India. New Delhi: Pentagon.
38. Sharma, J. & Grower, D. (2004). Reference service and sources of information. New Delhi: Ess publication
39. Semode, F. Ejitagha, S. and Baro, E, (2017) "Social Networking Sites: Changing Roles, Skills and Use by Librarians in Tertiary Institutions in Nigeria." *Library Philosophy and Practice (e-journal)*.available at: http://digitalcommons.unl.edu/libphilprac/1500 retrieved on 25th August 2020.
40. Vasquez. R and Nevada. R (2013) Integrating Social Media & Assistive Technology to Grow Life Connections-Motivate available at: https://www.nevadaddcouncil.org/wp-content/uploads/2013/10/social-media-disability-conference.pdf retrieved on 25th August 2020.
41. Viney, D. (2006). The Impact of the Disability Equality Duty on our Services to Users. In All Change Delivering Library Services to Disabled Students in a Changing Context. Available at: http://www.docs.google.com/viewer?a=v&q=cache:HpYEtW2qr_cj:www.sonul.ac.uk/ Retrieved on 25th August 2020.
42. WHO. (2001) International Classification of Functioning, Disability, and Health: ICF. N.p.: World Health Organization, 8. Print.
43. World Book Encyclopedia (2004). Chicago: World Book, Inc. D vol. 5.
44. Zaid, A and Zaid, Y (2017)"The Exclusion of Persons with Visual Impairment in Nigerian
45. Academic Libraries' Websites". *Library Philosophy and Practice (e-journal).* Available at: http://digitalcommons.unl.edu/libphilprac/1601 retrieved on 25th August 2020.

Global Journal of Computer Science and Technology: G Interdisciplinary
Volume 21 Issue 2 Version 1.0 Year 2021
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

SARS-CoV-2 is an Robot Bioweapon

By Li-Meng Yan & Adrian David Cheok
Professional University of Information and Management for Innovation (iUniversity)

Abstract- Two possibilities should be considered for the origin of SARS-CoV-2: natural evolution or laboratory creation. In our earlier paper titled "Unusual Features of the SARS-CoV-2 Genome Suggesting Sophisticated Laboratory Modification as a Biological Robot Rather than Natural Evolution and Delineation of its Probable Synthetic Route", we disproved the possibility of SARS-CoV-2 arising naturally through evolution and instead proved that SARS-CoV-2 must have been a product of laboratory modification. Despite this and similar efforts, the laboratory creation theory continues to be downplayed or even diminished. This is fundamentally because the natural origin theory remains supported by several novel coronaviruses published after the start of the outbreak. These viruses (the RaTG13 bat coronavirus, a series of pangolin coronaviruses, and the RmYN02 bat coronavirus) reportedly share high sequence homology with SARS-CoV-2 and have altogether constructed a seemingly plausible pathway for the natural evolution of SARS-CoV-2. Here, however, we use in-depth analyses of the available data and literature to prove that these novel animal coronaviruses do not exist in nature and their sequences have been fabricated. In addition, we also offer our insights on the hypothesis that SARS-CoV-2 may have originated naturally from a coronavirus that infected the Mojiang miners.

GJCST-G Classification: *J.3*

Strictly as per the compliance and regulations of:

SARS-CoV-2 is an Robot Bioweapon

Li-Meng Yan [α] & Adrian David Cheok [σ]

Abstract- Two possibilities should be considered for the origin of SARS-CoV-2: natural evolution or laboratory creation. In our earlier paper titled "Unusual Features of the SARS-CoV-2 Genome Suggesting Sophisticated Laboratory Modification as a Biological Robot Rather than Natural Evolution and Delineation of its Probable Synthetic Route", we disproved the possibility of SARS-CoV-2 arising naturally through evolution and instead proved that SARS-CoV-2 must have been a product of laboratory modification. Despite this and similar efforts, the laboratory creation theory continues to be downplayed or even diminished. This is fundamentally because the natural origin theory remains supported by several novel coronaviruses published after the start of the outbreak. These viruses (the RaTG13 bat coronavirus, a series of pangolin coronaviruses, and the RmYN02 bat coronavirus) reportedly share high sequence homology with SARS-CoV-2 and have altogether constructed a seemingly plausible pathway for the natural evolution of SARS-CoV-2. Here, however, we use in-depth analyses of the available data and literature to prove that these novel animal coronaviruses do not exist in nature and their sequences have been fabricated. In addition, we also offer our insights on the hypothesis that SARS-CoV-2 may have originated naturally from a coronavirus that infected the Mojiang miners.

Revelation of these virus fabrications renders the natural origin theory unfounded. It also strengthens our earlier assertion that SARS-CoV-2 is a product of laboratory modification, which can be created in approximately six months using a template virus owned by a laboratory of the People's Liberation Army (PLA). The fact that data fabrications were used to cover up the true origin of SARS-CoV-2 further implicates that the laboratory modification here is beyond simple gain-of-function research.

The scale and the coordinated nature of this scientific fraud signifies the degree of corruption in the fields of academic research and public health. As a result of such corruption, damages have been made both to the reputation of the scientific community and to the well-being of the global community.

Importantly, while SARS-CoV-2 meets the criteria of a bioweapon specified by the PLA, its impact is well beyond what is conceived for a typical bioweapon. In addition, records indicate that the unleashing of this weaponized pathogen should have been intentional rather than accidental. We therefore define SARS-CoV-2 as an Unrestricted Bioweapon and the current pandemic a result of Unrestricted Biowarfare. We further suggest that investigations should be carried out on the suspected government and individuals and the responsible ones be held accountable for this brutal attack on the global community.

Author α: Rule of Law Society & Rule of Law Foundation, New York, NY, USA. e-mail: team.lmyan@gmail.com

Author σ: Professional University of Information and Management for Innovation (iUniversity), Tokyo, Japan. e-mail: adrian@i-u.ac.jp

I. Introduction

SARS-CoV-2 is a novel coronavirus and the causative agent of the COVID-19 pandemic. Despite its tremendous impact, the origin of SARS-CoV-2, however, has been a topic of great controversy. In our first paper titled "*Unusual Features of the SARS-CoV-2 Genome Suggesting Sophisticated Laboratory Modification as a Biological Robot Rather than Natural Evolution and Delineation of its Probable Synthetic Route*"[1], we used biological evidence and in-depth analyses to show that SARS-CoV-2 must be a laboratory product, which was created by using a template virus (ZC45/ZXC21) owned by military research laboratories under the control of the Chinese Communist Party (CCP) government. In addition, resources and expertise are all in place in the Wuhan Institute of Virology (WIV) and related, other CCP-controlled institutions allowing the creation of SARS-CoV-2 in approximately six months.

What have not been fully described in our earlier analyses are details of the novel animal coronaviruses published by the CCP-controlled laboratories after the outbreak[1]. While no coronaviruses reported prior to 2020 share more than 90% sequence identity with SARS-CoV-2[2,3], these recently published, novel animal coronaviruses (the RaTG13 bat coronavirus[4], a series of pangolin coronaviruses[5-8], and the RmYN02 bat coronavirus[9]) all share over 90% sequence identities with SARS-CoV-2. As a result, these SARS-CoV-2-like viruses have filled an evolutionary gap and served as the founding evidence for the theory that SARS-CoV-2 has a natural origin[10-12]. In this report, we provide genetic and other analyses, which, when combined with recent findings[13-21], prove that these novel animal coronaviruses do not exist in nature and their genomic sequences are results of fabrication.

a) *Evidence proving that the RaTG13 virus is fraudulent and does not exist in nature*

On February 3rd, 2020, Dr. Zhengli Shi and colleagues published an article in *Nature* titled "*A pneumonia outbreak associated with a new coronavirus of probable bat origin*" (manuscript submitted on January 20th)[4], which was one of the first publications to identify SARS-CoV-2 as the pathogen causing the disease now widely known as COVID-19. Also reported in this article was a novel bat coronavirus named RaTG13, the genomic sequence of which was shown to be 96.2% identical to that of SARS-CoV2. The close evolutionary relationship between RaTG13 and SARS-

CoV-2 as suggested by the high sequence identity had led to a conclusion that SARS-CoV-2 has a natural origin. These striking findings have consequently made this article one of the most cited publications in the currently overwhelmed field of coronavirus research. Interestingly, an article published by Dr. Yong-Zhen Zhang and colleagues on the same issue of *Nature*, which also discovered SARS-CoV-2 as the responsible pathogen for COVID19, received much less citations[2]. This latter article made no mention of RaTG13[2]. Instead, Zhang and colleagues showed that, evolutionarily, SARS-CoV-2 was closest to two bat coronaviruses, ZC45 and ZXC21, both of which were discovered and characterized by military research laboratories under the control of the CCP government[3]. Immediately after the publication of this article, Dr. Zhang's laboratory was shut down by the CCP government with no explanations offered[22].

Since its publication[4], the RaTG13 virus has served as the founding evidence for the theory that SARSCoV-2 must have a natural origin[10]. However, no live virus or an intact genome of RaTG13 have ever been isolated or recovered. Therefore, the only proof for the "existence" of RaTG13 in nature is itsgenomic sequence published on *GenBank*.

b) The sequence of RaTG13 uploaded at GenBank can be fabricated

In order to have the sequence of a viral genome successfully uploaded onto *GenBank*, submitters have to provide both the assembled genomic sequence (text only) and raw sequencing reads. The latter is used for quality control and verification purposes. However, due to the huge amount of work involved in assembling raw reads into complete genomes, no sufficient curation is in place to ensure the correctness or truthfulness of the uploaded viral genomes. Therefore, an entry on *GenBank*, which in this case is equivalent to the existence of an assembled viral genomic sequence and its associated sequencing reads, is not a definitive proof that this viral genome is correct or real.

Sequencing of a viral RNA genome requires amplifying segments of it using reverse transcriptase PCR (RT-PCR) as the first step. The products of the RT-PCR, which are double-stranded DNA, would subsequently be sent for sequencing. The resulted sequencing reads, each ideally revealing the sequence of a segment of the genome, are then used to assemble the genome of the virus under study (Figure 1A). Typically, some segments of the genome may not be covered by the initial round of sequencing. Therefore, gap filling will be carried out, where these missing segments will be amplified specifically and the DNA products subsequently sequenced. These steps are repeated until a complete genome can be assembled, ideally with a proper depth to ensure accuracy.

However, this process leaves room for potential fraud. If one intends to fabricate an RNA viral genome on *GenBank*, he or she could do so by following these steps: create its genomic sequence on a computer, have segments of the genome synthesized based on the sequence, amplify each DNA segment through PCR, and then send the PCR products (may also be mixed with genetic material derived from the alleged host of the virus to mimic an authentic sequencing sample) for sequencing (Figure 1B). The resulted raw sequencing reads would be used, together with the created genomic sequence, for establishing an entry on *GenBank*. Once accomplished, this entry would be accepted as the evidence for the natural existence of the corresponding virus. Clearly, a viral genomic sequence and its *GenBank* entry can be fabricated if well-planned.

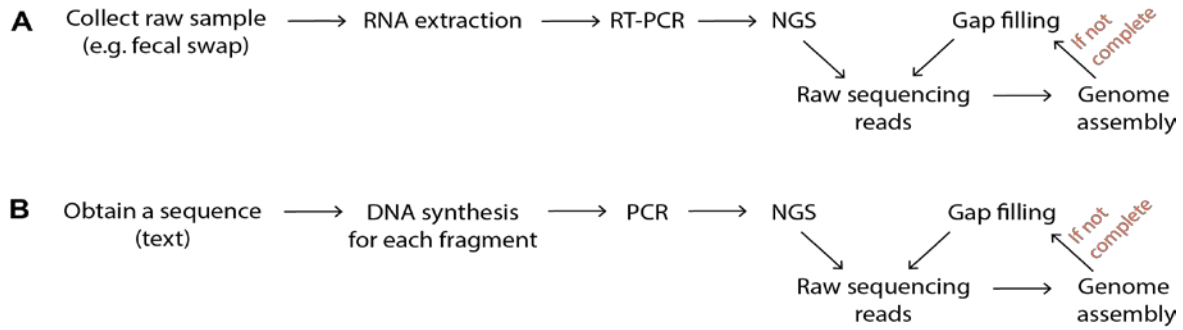


Figure 1: Illustration of steps involved in the sequencing and assembly of coronavirus genomes. A. The normal process. B. A possible route of fabricating a viral genome by creating a genomic sequence first and obtaining raw sequencing reads guided by it. NGS: Next Generation Sequencing.

The complete genomic sequence of RaTG13 was first submitted to *GenBank* on January 27th, 2020. The raw sequencing reads were made available on February 13th, 2020 (*NCBI SRA: SRP249482*). However, the sequencing data for gap filling, which is indispensable in assembling a complete genome, was only made available on May 19th, 2020 (*NCBI SRA: SRX8357956*). The timing and the reversed order of events here are strange and suspicious.

The raw sequencing reads of RaTG13 have multiple abnormal features[16,21]. Despite the sample being described as a fecal swab, only 0.7% of the raw sequencing reads are bacterial reads while the bacterial abundance is typically 70~90% when other fecal swab samples were sequenced[16,21]. In addition, in the identifiable region of certain sequencing reads, a vast majority of reads are eukaryotic sequences, which is also highly unusual in the sequencing of fecal swap-derived samples[16]. Within these eukaryotic reads, 30% of the sequences are of non-bat origin and instead shown to be from many different types of animals including fox, flying fox, squirrels, etc. These abnormal features are significant and indicate that the raw sequencing reads should have been obtained *via* a route that is different from the normal one (Figure 1).

No independent verification of the RaTG13 sequence seems possible because, according to Dr. Zhengli Shi, the raw sample has been exhausted and no live virus was ever isolated or recovered. Notably, this information was known to a core circle of virologists early on and apparently accepted by them. It was then made public, months later, by Dr. Yanyi Wang, director general of the WIV, in an TV interview on May 23rd, 2020[23]. Dr. Shi also confirmed this publicly in her email interview with *Science* in July 2020[24].

However, judging from Shi's published protocol[25], exhaustion of the fecal swap sample is highly unlikely. According to this protocol, the fecal swab sample would be mixed with 1 ml of viral transport medium and the supernatant collected. Every 140 ul of the supernatant would then yield 60 ul of extracted RNA[25]. For the subsequent step, RT-PCR, 5 ul of this RNA-containing solution is required per reaction[25]. Therefore, from one fecal swab sample, at least 80 RT-PCR reactions could be carried out ([1000/140] x 60/5=86). Such an amount is sufficient to support both the initial round of sequencing and the subsequent gap filling PCR. It would be sufficient to also allow reasonable attempts to isolate live viruses, although Dr. Shi claimed that no virus isolation was attempted[24].

Therefore. the RaTG13 virus and its published sequence are suspicious and show signs of fabrication.

c) Other suspicions associated with RaTG13

RaTG13 was reported by Dr. Zhengli Shi from the WIV[4]. Dr. Shi is a fellow of the American Academy of Microbiology and one of the most accomplished Chinese virologists. A peer-reviewed article authored by her and published on the top journal *Nature,* therefore, brought a great level of comfort for the coronavirus research community in accepting RaTG13 as a true, nature-born bat coronavirus. As a result, RaTG13, upon its timely publication, served as the founding evidence for the natural origin theory of SARS-CoV-2.

However, as revealed in section 1.1, the reported sequence of RaTG13, which is the only proof of the virus' existence in nature, is problematic and shows signs of fabrication.

Intriguingly, despite the pivotal role of RaTG13 in revealing the origin of SARS-CoV-2, the information provided for its discovery was surprisingly scarce with key points missing (location and date of sample collection, previous knowledge and publication of this virus, etc):

"*We then found that a short region of RNA-dependent RNA polymerase (RdRp) from a bat coronavirus (BatCoV RaTG13)—which was previously detected in Rhinolophus affinis from Yunnan province—showed high sequence identity to 2019-nCoV. We carried out full-length sequencing on this RNA sample (GISAID accession number EPI_ISL_402131). Simplot analysis showed that 2019-nCoV was highly similar throughout the genome to RaTG13 (Fig. 1c), with an overall genome sequence identity of 96.2%.*"[4] Only in the source section of the NCBI entry for RaTG13 (*GenBank* accession code: MN996532.1), one could find that the original sample was a "fecal swab" collected on "July 24th, 2013". A closer look at the sequence reveals that RaTG13 shares a 100% nucleotide sequence identity with a bat coronavirus RaBtCoV/4991 on a short, 440-bp RNA-dependent RNA polymerase gene (*RdRp*) segment. RaBtCoV/4991 was discovered by Shi and colleagues and published in 2016[26]. As described in the 2016 publication, only a short 440-bp segment of *RdRp* of the RaBtCoV/4991 virus was sequenced then. Given the 100% identity on this short gene segment between RaBtCoV/4991 and RaTG13, the field has demanded clarification of whether or not these two names refer to the same virus. However, Dr. Shi did not respond to the request or address this question for months. The answer finally came from Peter Daszak, president of EcoHealth Alliance and long-term collaborator of Shi, who claimed that RaBtCoV/4991 was RaTG13[27].

RaBtCoV/4991 was discovered in the Yunnan province, China. In 2012, six miners suffered from severe pneumonia after clearing out bat droppings in a mineshaft in Mojiang, Yunnan, and three of them died soon afterwards[28,29]. Although it was initially suspected that a SARS-like bat coronavirus may be responsible for the deaths, no coronavirus was either isolated or detected from the clinical samples[30]. Also, first-hand record indicates failure of biopsy and no attempt of autopsy[30], which are the gold standards in the diagnosis of coronavirus infections[30]. The pathogen responsible for the miners' deaths therefore remained an unsolved case[31]. (*Detailed analyses of the Mojiang Miner Passage hypothesis, which was based on the miners' case, are provided in section 1.6.*) Despite the failed diagnosis, this unknown pathogen nonetheless triggered immense interests in the virologists in China. Three independent teams, including that of Dr. Shi's, made a total of six visits to this mineshaft[26,28,31]. The Shi group particularly looked for the presence of bat coronaviruses by

amplifying and then sequencing a 440-bp *RdRp* segment[29], which is a routine procedure the Shi group follows in their surveillance studies. (*As shown in section 2.1 of our first report[1], this RdRp segment is also frequently used for phylogenetic analyses and is an attractive target for antiviral drug discovery, which may have contributed to the design of incorporating a unique RdRp into the genome of SARS-CoV-2.*) Out of the many coronaviruses detected, only RaBtCoV/4991 seemed to belong to the group of SARS-related, lineage B β coronaviruses[26].

The reporting of RaTG13 is suspicious in three aspects.

First, the whole genome sequencing of RaBtCoV/4991 should not have been delayed until 2020. Given the Shi group's consistent interests in studying SARS-like bat coronaviruses and the fact that RaBtCoV/4991 is a SARS-like coronavirus with a possible connection to the deaths of the miners, it is highly unlikely that the Shi group would be content with sequencing only a 440-bp segment of *RdRp* and not pursue the sequencing of the receptor-binding motif (RBM)-encoding region of the *spike* gene. In fact, sequencing of the *spike* gene is routinely attempted by the Shi group once the presence of a SARS-like bat coronavirus is confirmed by the sequencing of the 440-bp *RdRp* segment[25,32], although the success of such efforts is often hindered by the poor quality of the sample.

As quoted above, in the 2020 *Nature* publication, Shi and colleagues strongly suggested that the sequencing of the full genome was done in 2020 after they discovered the resemblance between RaTG13 and SARS-CoV-2 on the short *RdRp* segment[4]. This, if true, suggests that the quality of the sample should not be poor. Therefore, there is no technical obstacle for the whole genome sequencing of RaBtCoV/4991. Clearly, the perceivable motivation of the Shi group to study this RaBtCoV/4991 virus and the fact that no genome sequencing of it was done for a period of seven years (2013-2020) are hard to reconcile and explain.

However, an intriguing revelation took place in June 2020. Specifically, filenames of the raw sequencing reads for RaTG13 uploaded on the database were found, which indicate that these sequencing experiments were done in 2017 and 2018[33]. Likely responding to this revelation, in her email interview with *Science*[24], Dr. Shi contradicted her own description in the *Nature* publication[4] and admitted that the sequencing of the full genome of RaTG13 was done in 2018.

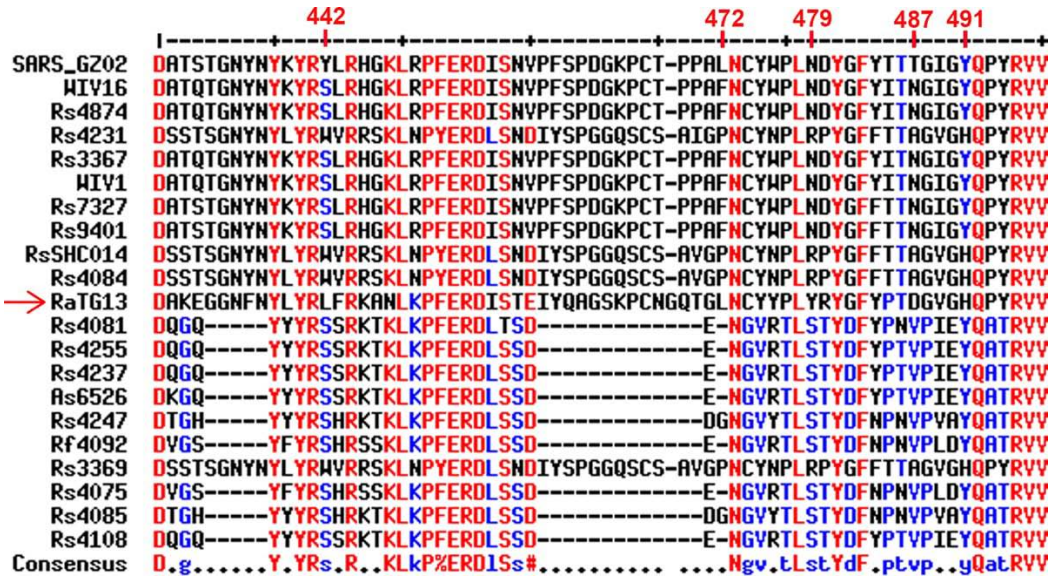


Figure 2: Sequence alignment comparing the RBMs of SARS (top) and RaTG13 (red arrow) to the RBMs of bat coronaviruses that Dr. Zhengli Shi published in high-profile journals between 2013 and 2018[25,32,34]. Amino acid residues highlighted by Shi as critical for binding the human ACE2 receptor[32] are labeled in red text on top. Alignment was done using the MultAlin webserver (http://multalin.toulouse.inra.fr/multalin/).

Second, RaTG13 has a remarkable RBM as suggested by its reported sequence, and the Shi group have no reason to delay its publication until 2020. The most critical segment of a SARS-like β coronavirus is the RBM in the Spike protein as it is fully responsible for binding the host ACE2 receptor and therefore determines the virus' potential in infecting humans. The RBM is also the most variable region because it is under strong positive selection when the virus jumps over to a new host. Sequence alignment on this crucial RBM motif reveals that the RaTG13 virus rivals with the most highly regarded bat coronaviruses in terms of resemblance to SARS (Figure 2). RaTG13's RBM not only is complete in reference to that of SARS but also is outstanding in its

preservation of five residues perceived by Dr. Shi as key in binding human ACE2 (hACE2)[32] (Figure 2, residues labeled with red texts). At position 472, RaTG13 is the only bat coronavirus that shares a leucine (L) residue with SARS, while the other four key residues are also largely conserved between the two viruses. Importantly, similar conservation patterns revealed in related bat coronaviruses, Rs3367 and SHC014, had led to their publication in *Nature* in 2013[32]. Furthermore, viruses with less "attractive" RBM sequences (having large gaps and poor in the preservation of key residues, bottom half of the sequences in Figure 2) were also published by Dr. Shi in other top virology journals between 2013 and 2018[25,34]. Therefore, if the genomic sequence of RaTG13 had been available since 2018, it is unlikely that this virus, which has a possible connection to miners' deaths in 2012 and has an alarming SARS-like RBM, would be shelfed for two years without publication. Consistent with this analysis, a recent study indeed proved that the RBD of RaTG13 (*produced via gene synthesis based on its published sequence*) was capable of binding hACE2[35].

Third, no follow-up work on RaTG13 has been reported by the Shi group. Upon obtaining the genomic sequence of a SARS-like bat coronavirus, the Shi group routinely investigate whether or not the virus is capable of infecting human cells. This pattern of research activities has been shown repeatedly[25,32,36-39]. However, such a pattern is not seen here despite that RaTG13 has an interesting RBM and is allegedly the closest match evolutionarily to SARS-CoV-2.

Clearly, these three aspects deviate from normal research activities and logical thinking, which are difficult to reconcile or explain. They should have contributed to the intentional omission of key information in the reporting of RaTG13[4].

For publications of biological research, it is unethical for authors to change the name of a previously published virus without any notice or description. It is also unethical for authors to not cite their own publication where they had characterized and reported the same virus. The violations here by Shi and colleagues on the reporting of RaTG13 are especially aggravating as the discovery of RaTG13 was central to uncovering the origin of SARS-CoV-2. By the time of the publication, SARS-CoV-2 had already led to many deaths in the city of Wuhan and had shown an alarming potential of causing a pandemic. In her much-delayed response to *Science* published on July 31st, 2020[24], Dr. Shi finally commented on the name change and stated that changing the name to RaTG13 was meant to better reflect the time and location of sample collection (TG = Tongguan; 13 = 2013). However, such an intention does not seem to justify why the previous name of RaBtCoV/4991 was never mentioned in the 2020 article[4] and why they did not cite their own 2016 publication where RaBtCoV/4991 was first reported[26]. Dr. Shi's recent clarification did not alter the fact that they have violated the reporting norms of biological research.

In summary, a range of suspicions were associated with the reporting of RaTG13, including the violations of scientific publication principles, the inconsistency in the descriptions of the sequencing dates, and the contradiction between the sequencing of its genome in 2018 and the publication of it in 2020 when this virus has a striking RBM and a possible connection to pneumonia-associated deaths. Adding to these suspicions are the exquisite timing of its publication, the problematic nature of its reported sequence and raw sequencing reads, and the claim that no sample is left for independent verification. Collectively, these facts justify and legitimate the concern over the true existence of the RaTG13 virus in nature and the truthfulness of its reported genomic sequence. They also question the claim that the RaBtCoV/4991 virus and RaTG13 are equivalent.

d) Genetic evidence proving the fraudulent nature of RaTG13

This evidence was revealed after a close examination of the sequences of specific genes, especially *spike,* of relevant viruses. Specifically, we compared two viruses for the synonymous and non synonymous mutations on each gene, and we did so for two pairs of viruses. The first pair are bat coronaviruses ZC45 and ZXC21. The second pair are SARS-CoV-2 and RaTG13. The rationale for comparing these two pairs with each other is the following. First, ZC45 and ZXC21, each sharing an 89% genomic sequence identity with SARS-CoV-2, are the closest relatives to SARS-CoV-2 and RaTG13. Second, ZC45 and ZXC21 are 97% identical to each other, while SARS-CoV-2 and RaTG13 are 96% identical. Not only the sequence identity in each case is comparable, but also the high sequence identity indicates that, within each pair, the sequence difference should be a result of random mutations during evolution, which ensures that synonymous and non-synonymous analyses here are appropriate and not complicated by abrupt evolutionary events (e.g. recombination). Indeed, sequence alignment confirms such a scenario – in both cases, the curve is smooth and the high sequence identity is maintained throughout (Figure 3).

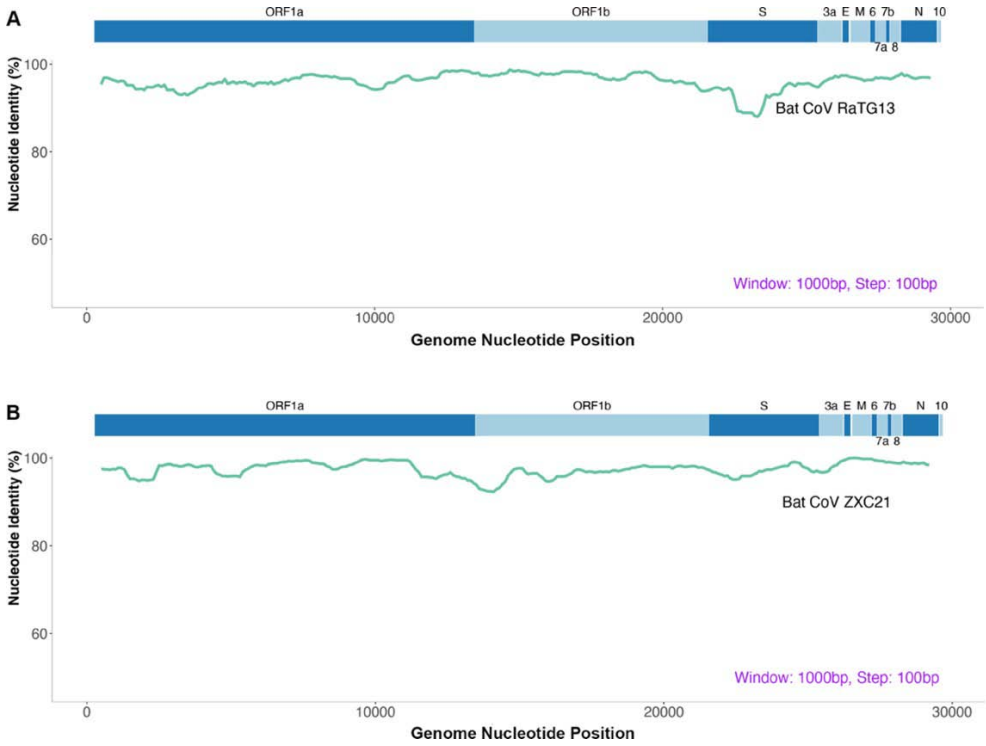


Figure 3: Simplot analyses show that high sequence identities are shared by two pairs of coronaviruses. A. the genomic sequence of RaTG13 is plotted against that of SARS-CoV-2. B. Genomic sequence of ZXC21 is plotted against that of ZC45.

Detailed synonymous (syn, green curve) and non-synonymous (non-syn, red curve) analyses are shown in Figure 4. For each gene, the accumulations of syn and non-syn mutations, respectively, are illustrated when the codons are analyzed in a sequential order. For the *spike* genes, between ZC45 and ZXC21, the syn/non-syn ratio is 5.5:1 (Figure 4A left, 94 syn mutations and 17 non-syn mutations). Notably, the two curves progress along in a roughly synchronized manner. These features reflect, to a certain extent, the evolutionary traits resulted from random mutations during evolution in this sub-group of lineage B β coronaviruses.

The same analysis on the *spike* genes of SARS-CoV-2 and RaTG13, however, revealed a different scenario (Figure 4B right). Although the overall syn/non-syn ratio is a similar 5.4:1 (221 syn mutations and 41 non-syn mutations), the synchronization between the two curves is non-existent. In the second half of the sequence, which is over 700 codons (2,100 nucleotides) wide, the non-syn curve stays flat when the syn curve climbs continuously and significantly.

Counting the syn and non-syn mutations of the S2 region (corresponding to residues 684-1273 of the SARS-CoV-2 Spike) reveals that, between ZC45 and ZXC21, there are a total of 27 syn mutations and 5 non-syn mutations, yielding a syn/non-syn ratio of 5.4:1. In contrast, for the same S2 region, between SARS-CoV-2 and RaTG13, there are a total of 88 syn mutations and 2 non-syn mutations, yielding a syn/non-syn ratio of 44:1. The syn/non-syn ratios for S2, whole Spike, and other large viral proteins (Orf1a, Orf1b, and Nucleocapsid) are summarized in Table 1. While the ratios are comparable between the two groups for all other proteins, the ratios for the S2 protein are significantly different.

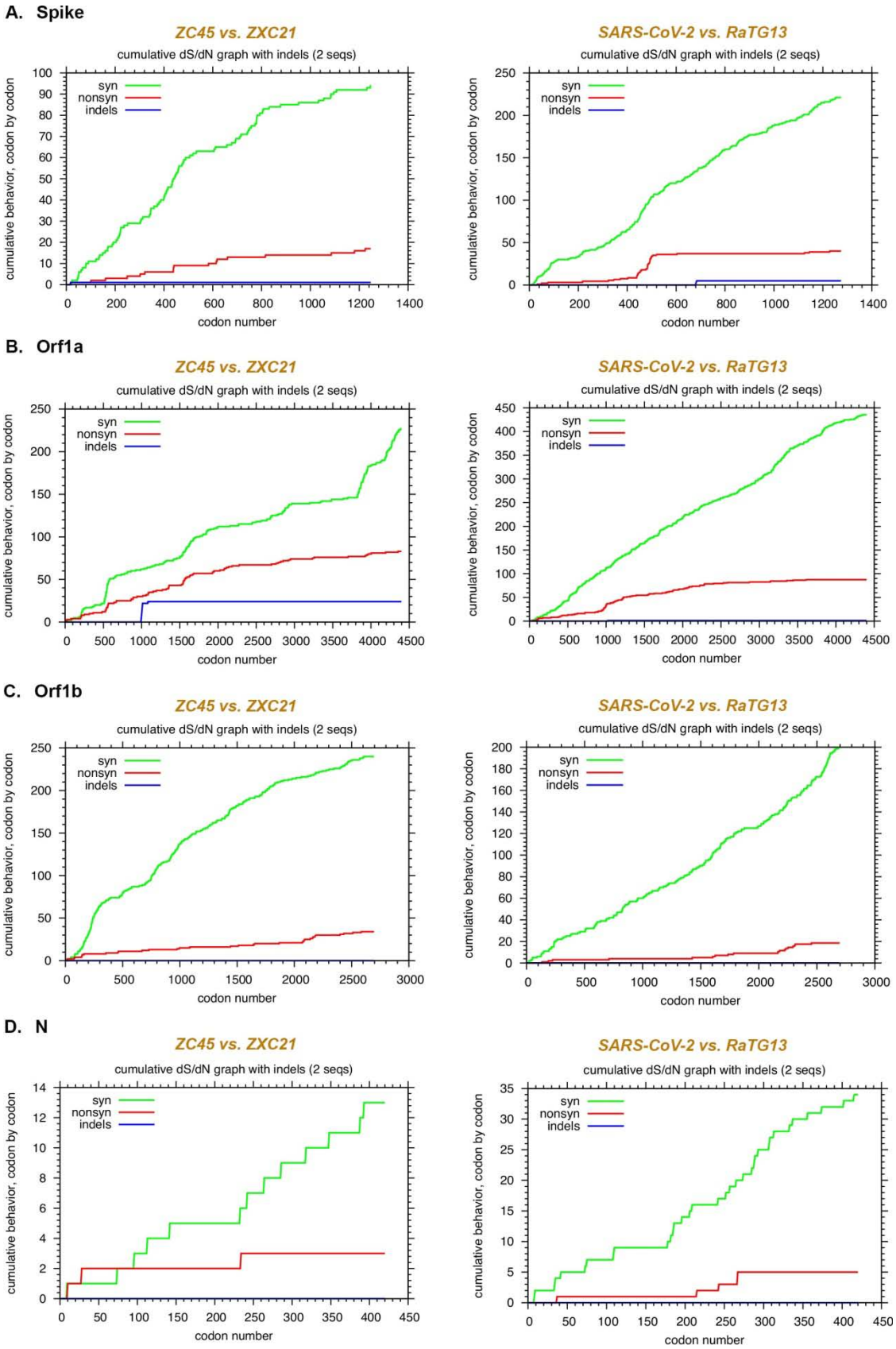


Figure 4: Abnormal distribution of synonymous and non-synonymous mutations in Spike revealed by the comparison between RaTG13 and SARS-CoV-2. Synonymous and non-synonymous mutations are analyzed between closely related coronaviruses on large viral proteins: A. Spike (S), B. Orf1a, C. Orf1b, and D. Nucleocapsid (N). In each

panel, the left graph is the comparison between the two bat coronaviruses ZC45 (MG772933) and ZXC21 (MG772934), while the right graph is the comparison between SARS-CoV-2 (NC_045512) and RaTG13 (MN996532). In each graph, the accumulative growth of synonymous mutations (green curve), non-synonymous mutations (red curve), and in-frame deletions (blue curve) are depicted, respectively. Initial sequence alignment was done using EMBOSS Needle, which was followed by codon alignment at www.hiv.lanl.gov. Synonymous nonsynonymous analyses were performed using SNAP also at www.hiv.lanl.gov[40].

Table 1: Ratios of syn/non-syn mutations observed in different viral proteins

Protein	ZC45 vs. ZXC21	SARS-CoV-2 vs. RaTG13
S2	5.4:1	44.0:1
Spike	5.5:1	5.4:1
Orf1a	2.7:1	5.0:1
Orf1b	7.1:1	10.8:1
N	4.3:1	6.8:1

The detailed syn/non-syn analyses for Orf1a, Orf1b, and N are shown in Figure 4B-D. It is also noteworthy that, similar to that of Spike, the approximate synchronization between two curves is observed for the Orf1a protein in the ZC45 and ZXC21 comparison (Figure 4B left) but not in the SARS-CoV-2 and RaTG13 comparison (Figure 4B right).

The S2 protein maintains trimmer formation of the Spike and, upon successive cleavages to expose the fusion peptide, mediates membrane fusion and cell entry. Although the S2 protein is more conserved evolutionarily than S1, the extremely high purifying pressure on S2 as suggested by the very high syn/nonsyn ratio is abnormal. In fact, Orf1b is known to be the most conserved protein in coronaviruses and yet the syn/non-syn ratio for it is only 10.8:1 when SARS-CoV-2 and RaTG13 are compared, much lower than the ratio of 44:1 observed for S2 (Table 1). Furthermore, since RaTG13 and SARS-CoV-2 infect different species, no high purifying selection on S2 should be expected when these two viruses are compared against each other.

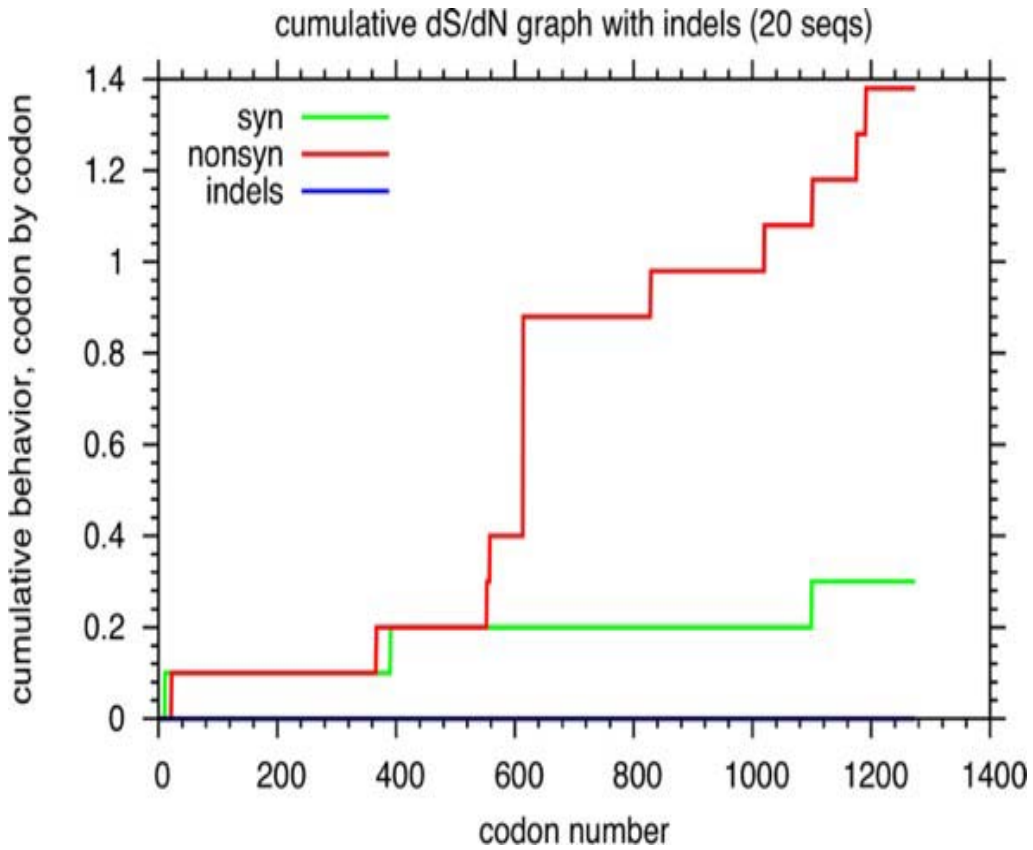


Figure 5: Positive selection, not purifying selection, is observed for Spike in twenty randomly selected SARSCoV-2 sequences. GenBank accession numbers are shown in Figure 6. Collection dates of these viruses range from December 2019 to July 2020.

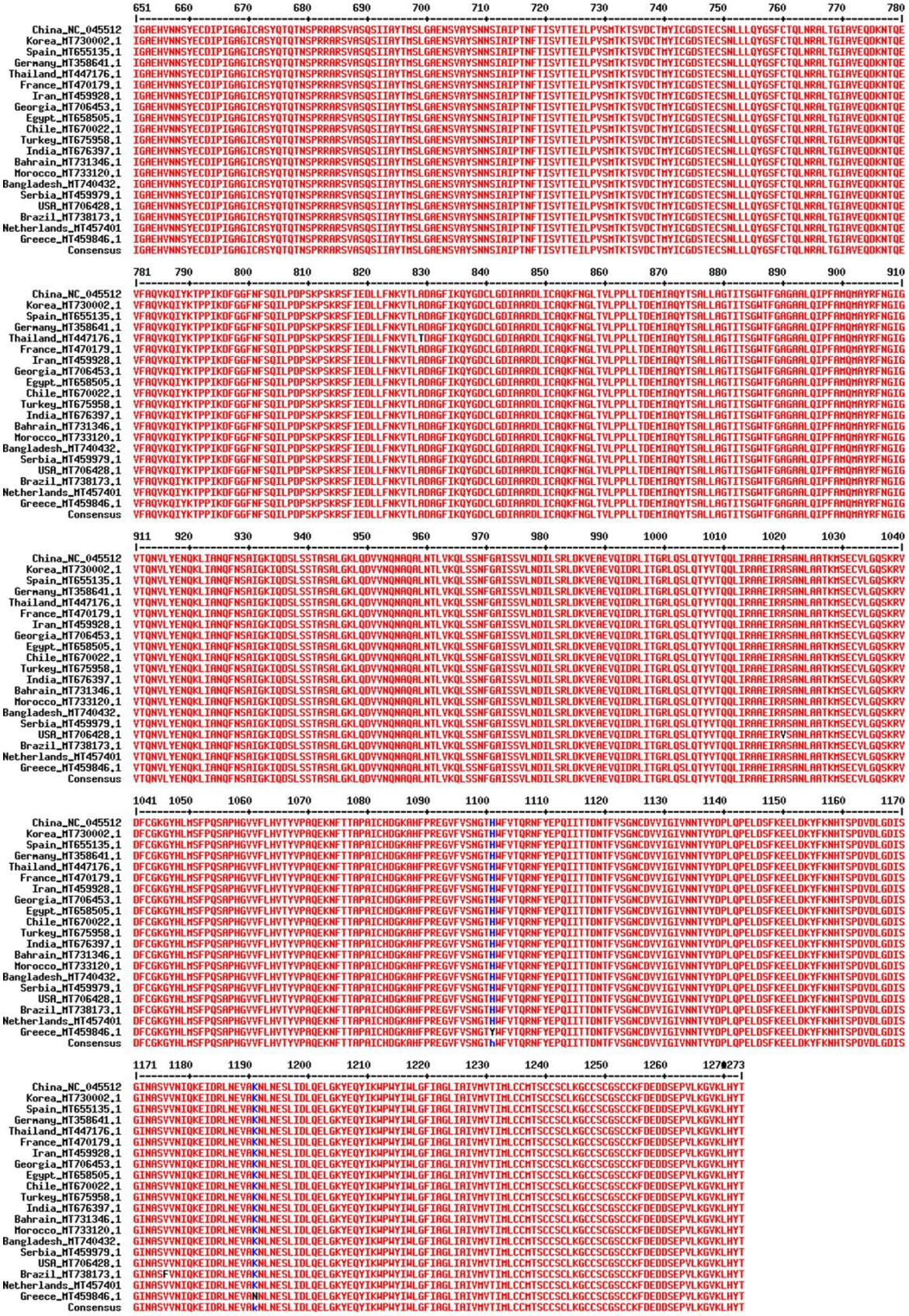


Figure 6: Five amino acid mutations are observed in S2 (684-1273) in twenty randomly selected SARS-CoV-2 sequences. They are at positions 829, 1020, 1101, 1176, and 1191. GenBank accession number for each isolate is shown in the sequence's name following the country name.

Consistent with the above notion, a syn/non-syn analysis done for the Spike protein of twenty randomly selected SARS-CoV-2 sequences showed that S2 was under positive selection, not purifying selection, during the past eight months of human-to-human transmission (Figure 5). For the twenty SARS-CoV-2 isolates, amino acid mutations are observed at five different locations in S2 (Figure 6). In addition, a recent study analyzing 2,954 genomes of SARS-CoV-2 revealed that mutations have been observed at 25 different locations in the S2 protein[41], further proving that amino acid mutations are tolerated in S2 and no high purifying pressure should be observed for S2. Evidently, the syn/non-syn ratio of 44:1 revealed between SARS-CoV-2 and RaTG13 on the S2 region is abnormal (Table 1) and a violation of the principles of natural evolution.

A logical interpretation of this observation is that SARS-CoV-2 and RaTG13 could not relate to each other through natural evolution and at least one must be artificial. If one is a product of natural evolution, then the other one must be not. It is also possible that neither of them exists naturally.

If RaTG13 is a real virus that truly exists in nature, then SARS-CoV-2 must be artificial.

However, the reality is that SARS-CoV-2 is physically present and has first appeared prior to the reporting of RaTG13[4]. This would then lead to the conclusion that RaTG13 is artificial, a scenario consistent with the overwhelming suspicion that this virus does not exist in nature and its sequence has been fabricated.

The remaining possibility is, of course, that both SARS-CoV-2 and RaTG13 are artificial: one has been created physically and the other one exists only in the form of a fabricated sequence.

It is highly likely that the sequence of the RaTG13 genome was fabricated by lightly modifying the SARS-CoV-2 sequence to achieve an overall 96.2% sequence identity. During this process, much editing must have been done for the *RBM* region of the *S1/spike* because the encoded RBM determines the interaction with ACE2 and therefore would be heavily scrutinized by others. An RBM too similar to that of SARS-CoV-2 would be troublesome because: 1) RaTG13 could be conceived as a product of gain-of function research; 2) it would leave no room for an intermediate host and yet such a host is believed to exist as the Spike/RBM needs to first adapt in an environment where the ACE2 receptor is homologous to hACE2. In addition, modifying the sequence of the RBM is also beneficial as RaTG13 would otherwise appear to be able to infect humans as efficiently as SARS-CoV-2 does, escalating the concern of a laboratory leak. To eliminate such concerns, many non-syn mutations were introduced into the RBM region.

Importantly, syn/non-syn analysis is frequently used, often at the ORF/protein level, to characterize the evolutionary history of a virus[42-44]. While editing the RBM, the expert(s) carrying out this operation must be conscious of the need to maintain a reasonable syn/non-syn ratio for the whole Spike protein. To achieve so, however, the expert(s) must have then strictly limited the number of non-syn mutations in the S2 half of Spike, which ended up flattening the curve (Figure 4A right).

e) The receptor-binding domain (RBD) of RaTG13 does not bind ACE2 of horseshoe bats

Consistent with the above conclusion that RaTG13 does not exist in nature and its sequence was fabricated, a recent study showed that the RBD of RaTG13 could not bind the ACE2 receptors of two different kinds of horseshoe bats, *Rhinolophus macrotis* and *Rhinolophus pusillus*[45]. Although the ACE2 receptor of *Rhinolophus affinis* (the alleged host of RaTG13) was not tested, it is unlikely that ACE2 of *R. affinis* would differ significantly from those of its close relatives and be able to bind the RaTG13 RBD.

This result therefore implicates that RaTG13 would not be able to infect horseshoe bats, contradicting the claim made by Shi and colleagues that the virus was detected and discovered from horseshoe bats. This is also consistent with the above conclusion that the genomic sequence of RaTG13 is fabricated and presumably computer-edited, which entails that the RBM/RBD suggested by the corresponding gene sequence may not be functional in binding the ACE2 receptor of the claimed host.

f) Conclusion and postulation of the fabrication process

In conclusion, the evidence presented both here and from recent literature collectively prove that RaTG13 does not exist in nature and its sequence has been fabricated.

If the RaBtCov/4991 virus is equivalent to RaTG13, then RaBtCoV/4991 must be fraudulent as well.

Apparently, in the actual process of sequence fabrication, the published sequence of the short *RdRp*segment of RaBtCoV/4991 was completely inherited for RaTG13. This way, they could claim that RaTG13 was RaBtCoV/4991, which, according to the record, was discovered in 2013[26]. If RaTG13 had been described as being discovered right around the time of the COVID-19 outbreak, greater suspicions would result as tracing the evolutionary origin of a zoonotic virus is difficult and usually takes years or decades. As described in section 2.1 of our earlier report[1], the fabrication of RaTG13 should have been planned and executed in coordination with the laboratory creation of SARS-CoV-2.

Such an approach is also safe because, except for the 440-bp *RdRp* segment, no other sequence information has ever been published for the rest of the RaBtCoV/4991 genome.

It is worth noting that, due to reasons detailed in section 1.2, they still preferred to obscure the history of RaTG13. However, they must have also anticipated that their violations of the publication norms would invite inquiries or requests for clarifications, the number of which, however, should be limited and manageable. RaBtCoV/4991 would then function as an additional layer of security for them in facing such inquiries and/or requests.

Building upon the 440-bp *RdRp* sequence inherited from RaBtCoV/4991, the rest of the RaTG13 genome was likely fabricated by lightly editing the sequence of SARS-CoV-2. Once the genomic sequence was finalized, DNA fragments could be synthesized individually according to the fabricated and edited sequence and then used as templates for PCR. Amplified DNA would then be mixed with certain raw material to give the sample a natural look (mimicking what is present in an actual RT-PCR, which is done using RNA extracted from fecal swabs as templates). Subsequently, this sample would be sent for sequencing. The resulted raw sequencing reads could then be uploaded together with the made-up genomic sequence onto *GenBank* to create an entry for the RaTG13 genome.

g) The Mojiang Miner Passage (MMP) hypothesis is fatally flawed

Recently, a theory has emerged, which proposed that SARS-CoV-2 was derived from viral passaging in the lungs of the infected Mojiang miners back in 2012[46]. Specifically, authors believe that the RaBtCoV/4991 virus was indeed RaTG13 and was the virus causing pneumonia in the miners in 2012. While inside the lungs of the miners, the RaTG13 virus had evolved extensively, mimicking a viral passage process, and eventually became SARS-CoV-2. In this process, the RBD of the virus experienced strong positive selection, through which it became optimal in binding hACE2. Furthermore, the furincleavage site at the S1/2 junction region of Spike had been acquired through recombination between the viral *spike* gene and the gene encoding the human ENaC protein, which has a furin-cleavage sequence closely resembling that of SARS-CoV-2. The end product of this passage was SARS-CoV-2, which the researchers isolated from the miners' samples and brought back to the WIV. The authors have named this hypothesis as the Mojiang Miner Passage (MMP) hypothesis[46].

However, this MMP hypothesis has fatal flaws.

First, the viral pathogen that caused the disease in the miners could not be defined or confirmed. According to the record, which was well documented in a Master's Thesis written by the doctor in charge, samples from two patients (throat swabs and blood) were tested at the Center for Disease Control and Prevention of the Chengdu Military Region between May 15th and May 20th, 2012, and yet none of the suspected viruses, including SARS, was detected[30]. Furthermore, the gold standard in the clinical diagnosis of coronavirus-caused pneumonia is biopsy and/or autopsy followed by confirmation by either RT-PCR or isolation of the virus. However, three biopsy tests were attempted but failed[30]. Autopsy tests were requested and yet all turned down by families of the deceased miners[30]. Due to such failure, both the Master's Thesis and later a PhD Dissertation, which also looked into this issue although in an indirect manner, described the cause of the pneumonia as an unsolved case[30,31].

Second, antibody tests done for the miners do not support SARS or SARS-like coronavirus infection. According to the Master's Thesis, samples from two miners were tested for antibodies against SARS[30]. The symptoms onset date for one miner (case 3, passed away) was around April 13th, 2012. The other miner (case 4, had severe symptoms and yet recovered) had symptoms onset around April 16th, 2012. Antibody tests, which were recommended later by Dr. Nanshan Zhong, were done at the WIV on June 19th, 2012. However, the two samples tested were only positive for IgM[30]. No positive IgG or total antibody were reported[30]. No antibody titer was reported either. Importantly, if the severe pneumonia was caused by coronavirus infections, by the time of the antibody tests on June 19th, 2012, both IgM and IgG/total antibody should be detected. In fact, IgG/total antibody should be much more abundant and easier to detect[47]. On the other hand, IgM tests frequently result in false positives[48]. Therefore, the fact that only IgM, and no IgG/total antibody, was tested positive suggests that the described results were most likely false positives and the infections should not have been caused by SARS or a SARS-like coronavirus.

It is noteworthy that the later PhD Dissertation[31] showed severe discrepancies with the Master's Thesis in the descriptions of the same clinical tests:

1. The PhD Dissertation described that samples from four miners (throat swab and blood) were sent to the Center for Disease Control and Prevention of the Chengdu Military Region for nucleic acid tests. However, the Master's Thesis indicated that samples were only taken from two miners[30].
2. The PhD Dissertation described samples from four miners were tested for anti-SARS antibodies at the WIV and all were IgG positive. However, the Master's Thesis indicated that only samples from two miners were tested at the WIV and both were only IgM positive[30].

Importantly, the Master's Thesis was written in 2013 in Yunnan by the doctor who was in charge of the six hospitalized miners[30]. The PhD dissertation, however, was written in 2016 in Beijing based only on the clinical record. The author of the Dissertation had no direct

involvement in the treatment of the miners or in any of the described tests[31]. It is therefore highly likely that author of the PhD dissertation did not verify the clinical data he presented, which makes this PhD dissertation an unreliable source of information concerning the Mojiang miners' case.

Third, if SARS-CoV-2 was already present in the miner's body in 2012, it would have certainly caused an epidemic or even pandemic then. Given the extremely high transmissibility of SARS-CoV-2, it would be impossible for the doctors, nurses, family members of the miners, etc. to have avoided contracting the virus without the protection of proper PPE. If an epidemic indeed happened in 2012, it could not have gone unnoticed given the high transmissibility and lethality (three out of the six pneumonia patients died despite of intense medical care provided for them).

Fourth, as shown in sections 1.1-1.5, RaTG13's sequence is clearly fabricated and the virus does not exist in nature. The RaBtCoV/4991 virus, which was detected in 2013, is not the RaTG13 virus that is defined by its reported genomic sequence. No complete genomic sequence of RaBtCoV/4991 has ever been reported likely due to the poor quality of the sample, which happens often as the RNA genome decays easily. It is highly likely that no high homology is shared between the actual RaBtCoV/4991 virus and SARS-CoV-2. This judgement is based on the fact that no viruses reported prior to 2020 share more than 90% sequence identity with SARS-CoV-2 despite the extensive surveillance studies of coronaviruses for the past two decades. Therefore, even if RaBtCoV/4991 was the pathogen responsible for the pneumonia of the miners, the theory that it has evolved in a single person's lung into SARS-CoV-2 is far beyond being reasonable.

Fifth, it is impossible for the Spike protein of the virus to obtain a unique furin-cleavage site at the S1/S2 junction through recombination with the gene encoding the ENaC protein of the host cell (ENaC carries a furin cleavage site closely resembling the one seen in SARS-CoV-2). This is because recombination requires a significant level of sequence similarity between the two participating genes and yet no such similarity is present between coronavirus Spike and human ENaC. The molecular basis for recombination is non-existent. (*Although recombination between ENaC and coronavirus Spike is impossible, it is suspicious that a viral protein and a host protein would share the same sequence for their furin-cleavage sites. It is possible, though, that the sequence of the furin-cleavage site in ENaC[49], which is known since 1997[50], could have been used in the design of the furin-cleavage site in the Spike of SARSCoV-2. Such a design may be considered sophisticated as ENaC co-expresses with ACE2 in many different types of cells[49].*)

Sixth, if SARS-CoV-2 has indeed evolved from RaBtCoV/4991 in the miner's lungs, it would look, from every aspect, like a naturally occurring virus. In that case, there would be no need to commit sequence fabrication for RaTG13 and for the other novel coronaviruses (parts 2 and 3) to falsify a natural origin for SARS-CoV-2.

Finally, as revealed in our earlier report[1], evidence exists in the genome of SARS-CoV-2, indicating that genetic manipulation is part of the history of SARS-CoV-2.

II. Evidence Proving that Recently Published Pangolin Coronaviruses are Fraudulent and do not Exist in Nature

While RaTG13 was reported to share a high sequence identity with SARS-CoV-2 and thereby hinted a natural origin of SARS-CoV-2, significant questions remained unanswered:

- No intermediate host has been found although one was believed to exist and function as the reservoir of the virus before it spilled over to humans.
- Despite the overall genomic resemblance of the two viruses, the RBD (particularly the RBM within it) of RaTG13 differs significantly from that of SARS-CoV-2. The evolutionary origin of the SARS-CoV-2 RBD, which is optimal in binding hACE2, remained unclear.
- A critical furin-cleavage site, which is present at the S1/S2 junction of SARS-CoV-2 Spike and responsible for the enhanced viral infectivity and pathogenicity[51-57], is absent in RaTG13 (as well as in all known lineage B β coronaviruses[58]). The evolutionary origin of this furin-cleavage site also remained mysterious.

Not long after these questions emerged, several laboratories published novel coronaviruses allegedly found in Malayan pangolins that were smuggled from Malaysia and confiscated by the Chinese custom[58]. Although these novel coronaviruses share relatively lower overall sequence identities (~90%) with SARS-CoV-2 in comparison to RaTG13 (96.2% identical to SARS-CoV-2), the RBD of the pangolin coronaviruses resembles greatly the SARS-CoV-2 RBD (97.4% identical). In the most critical RBM region, all amino acids except one are identical between the pangolin coronaviruses and SARS-CoV-2[5-8]. These observations led the authors to conclude 1) that pangolins are the likely intermediate host for the zoonotic transfer of SARS-CoV-2[5,7] and 2) that a RaTG13-like ancestor coronavirus might have acquired the RBD from a pangolin coronavirus through recombination to eventually become SARS-CoV-2[5-8].

Here, in part 2 of the report, we describe literature evidence and provide genetic analyses to

prove that these novel pangolin coronaviruses are results of fabrication.

a) *A single batch of pangolin samples were used in all studies and the deposited sequencing data showed heavy contamination and signs of fabrication*

In October 2019, a team formed by three researchers from two institutions (Guangdong Institute of Applied Biological Resources and Guangzhou Zoo) reported, for the first time, the detection of coronavirus infections in pangolins that were allegedly smuggled from Malaysia and confiscated in the Guangdong province in March 2019[59]. Twenty-one pangolin samples were sequenced and five were positive for coronavirus infections (Table 2: lung 2, 7, 8, 9, and 11), although Sendai virus infection was also reported. However, neither the sequences of the coronaviruses nor raw sequencing data were made available to the public for a period of three months. The raw data (*NCBI BioProject PRJNA573298*) was finally released on January 22nd, 2020 after the COVID-19 outbreak started, while the article submission date was September 30th, 2019 and the publication date was October 24th, 2019[59].

Between March and May 2020, four seemingly independent studies were published, all of which reported novel pangolin coronaviruses and their assembled genomic sequences[5-8]. However, after a closer look, we found that all four studies derived viral sequences from the same set of pangolin samples first reported in the October 2019 publication[59], which has been confirmed by a recent article[13].

In one study[6], Liu et al. (the same authors of the October 2019 publication[59]) re-assembled the genome of a pangolin coronavirus by pooling two samples from the original 2019 study and one sample obtained from another Malayan pangolin rescued in July 2019. However, although the authors stated that the more recent raw sequencing data had been deposited at the NCBI database[6], we could not find this data using the accession number (2312773) provided. The same difficulty has been reported by others[13]. Therefore, it cannot be verified whether the July 2019 dataset truly exists and has contributed to the assembly of the reported genome.

In two other studies, Lam et al.[5] and Zhang et al.[8] each re-assembled the genome of a pangolin coronavirus using only the published dataset from the October 2019 study[59]. Lam et al. also reported detection of coronaviruses from smuggled Malayan pangolins that were confiscated in the Guangxi province[5], although these viruses showed lower sequence identities to SARS-CoV-2 both at the whole genome level (~86%) and in the critical RBD region. It is noteworthy that this study was done as a collaboration between Dr. Yi Guan's group from the University of Hong Kong and Dr. Wuchun Cao's group from the Academy of Military Medical Sciences (AMMS), Beijing, China[5]. Somehow, all authors affiliated with the AMMS were excluded from the list of authors when the article was first submitted[60], although their names eventually appeared in the final version of the publication[5]

In the fourth study, Xiao et al. claimed to have examined tissue samples kept from diseased pangolins and obtained raw sequencing data for the subsequent assembly[7]. However, they did not describe how the samples were acquired. In their Extended Data Table 3, they listed the metagenome sequencing data used in the study[7], which, surprisingly, do not match with the actual data that they uploaded in the database (Table 2). Samples M1, M5, M6, M10, and Z1 can be found in the data they deposited, but not M2, M3, M4, and M8. Furthermore, Xiao et al. apparently were inconsistent with the reporting of these raw sequencing reads. For samples M1, M6, pangolin3, and pangolin5, they counted *paired ends* numbers, which reflect the actual number of sequenced DNA fragments in the library. For the rest of samples, the authors counted *reads* numbers instead (In Illumina sequencing, there are two *read*s per fragment). For samples M2, M3, M4, and M8 in this latter group[7], when the *reads* numbers were converted to *pairedends* numbers (divided by 2), they each match perfectly with lung07, lung02, lung08, and lung11, respectively, from the October 2019 study[59] (Table 2). Clearly, Xiao et al. used the data published in a previous study but failed to disclose this necessary information in their publication[7]. In fact, they intentionally presented the "number of reads" in a different format to presumably make readers overlook the fact that the same sequencing dataset was used.

It is noteworthy that the study by Xiao et al. was also done in collaboration with the AMMS. Prior to the publication of the manuscript, this work was first publicized in a press conference[61,62]. As revealed in this conference, four principle investigators contributed to the work and one of them was Dr. Ruifu Yang from the AMMS. However, like what happened to Dr. Cao and his AMMS colleagues in the Lam et al. study[5], Dr. Yang's name was excluded in the submitted manuscript of Xiao et al.[63]. Yet, unlike the other case, the AMMS researcher's name did not re-appear in the final publication[7]. It is also noteworthy that the two AMMS principle investigators here, Dr. Yang and Dr. Cao, are long-term collaborators and most of their collaborative work concerned genetic analyses of SARS-CoV[64-67].

Among the four studies, only two assembled complete genomes by performing gap filling using PCR[6,7]. However, neither group made their gap filling sequences available[13], rendering independent verification impossible. Notably, the delayed publishing of raw sequencing reads long after the publication of genomic sequences has occurred in the reporting of RaTG13 as well.

Adding to the above problems was the poor quality of the raw sequencing data, which has been

described recently[13,14,20]. We also analyzed the composition of the sequencing reads of the deposited libraries. By performing taxonomy analysis on the NCBI SRA database, we also found that samples from Liu et al.[6] that are positive for coronavirus reads are all positive for reads that map to human genome (Table 2). In great contrast, the rest of the samples, which are negative for viral reads, also have no human reads detected. The same correlation is found in data presented by Xiao et al[7]. Although samples M5 (pangolin 6) and M6 (pangolin2) are negative for human reads, these two samples have very few viral reads, which would hardly contribute to the viral genome assembly. Clearly, the human contamination should not be due to sample handling as none of the coronavirus-negative samples, which must have been handled similarly, contain such contamination. The consistent co-existence of viral reads and human reads are highly suspicious.

Table 2: Analyses of the raw sequencing data deposited by Liu et al

	Coronavirus positive?	Name	Accession #	Note	Total Reads PE: paired-end reads R: individual reads	Coronavirus Reads	Pangolin Reads Percentage	Human Reads Percentage	Human sample weight: human/(human+pangolin)
		Lung01	SRR10168393		22,900,426(PE)		49%	0%	0%
Liu et al. 2019	YES	Lung02	SRR10168392	M3 in *Xiao et al. 2020*	39,738,679(PE)	14	44%	4%	8%
		Lung03	SRR10168381		12,967,281(PE)		49%	0%	0%
		Lung04	SRR10168385		19,038,817(PE)		62%	0%	0%
	YES	Lung07	SRR10168378	M2 in *Xiao et al. 2020*	19,045,923(PE)	302	54%	3%	5%
	YES	Lung08	SRR10168377	M4 in *Xiao et al. 2020*	16,414,925(PE)	1100	45%	2%	4%
	YES	Lung09	SRR10168376		18,067,615(PE)	36	10%	23%	70%
	YES	Lung11	SRR10168375	M8 in *Xiao et al. 2020*	22,220,187(PE)	12	71%	1%	1%
		Lung12	SRR10168374		9,275,501(PE)		68%	0%	0%
		Lung13	SRR10168373		16,491,648(PE)		81%	0%	0%
		Lung19	SRR10168391		19,986,780(PE)		36%	0%	0%
		Lymph01	SRR10168390		18,903,834(PE)		49%	0%	0%
		LymphA01	SRR10168389		20,045,443(PE)		60%	0%	0%
		Spleen01	SRR10168388		11,527,782(PE)		86%	0%	0%
		Spleen02	SRR10168387		15,350,468(PE)		61%	0%	0%
		Spleen03	SRR10168386		19,055,973(PE)		43%	0%	0%
		Spleen04	SRR10168385		19,038,817(PE)		54%	0%	0%
		Spleen08	SRR10168384		15,975,904(PE)		78%	0%	0%
		Spleen11	SRR10168383		15,273,939(PE)		61%	0%	0%
		Spleen12	SRR10168382		12,590,769(PE)		84%	0%	0%
		Spleen19	SRR10168380		16,068,654(PE)		91%	0%	0%
	YES	M1 (Pangolin 9)	SRR11119759		107,267,359(PE)	496	48%	17%	26%
		M2		Lung07 in *Liu et al. 2019*	38,091,846(R)				
		M3		Lung02 in *Liu et al. 2019*	79,477,358(R)				
Xiao et al. 2020		M4		Lung08 in *Liu et al. 2019*	32,829,850(R)				
	YES	M5 (Pangolin 6)	SRR11119762		547,302,862(R)	56	83%	0%	0%
	YES	M6 (Pangolin 2)	SRR11119766		232,433,120(PE)	10	97%	0%	0%
		M8		Lung11 in *Liu et al. 2019*	44,440,374(R)				
		M10(Pangolin1)	SRR11119767		227,801,882(R)		75%	0%	0%
		Z1 (Pangolin 4)	SRR11119764		444,573,526(R)		52%	0%	0%
		pangolin 3	SRR11119765	Uploaded but not analyzed	212,161,250(PE)		97%	0%	0%
		pangolin 5	SRR11119763	Uploaded but not analyzed	196,761,202(PE)		89%	0%	0%

These observations raise red flags not only on the credibility of the assembled sequences but also on the authenticity of these novel pangolin coronaviruses. It is also noteworthy that the manuscript submission dates for all four studies were between February 7th and February 18th[5-8], suggesting that their publications might have been coordinated.

b) No coronavirus was detected in an extensive surveillance study of Malayan pangolins

While these SARS-CoV-2-like pangolin coronaviruses were described as being detected in smuggled Malayan pangolins[59], a recent study strongly refuted the presence of such pangolin coronaviruses in nature. A team led by Dr. Daszak examined 334 pangolin samples, which were collected in Malaysia and Sabah from August 2009 to March 2019[68]. Surprisingly, no coronaviridae, or any of the other families of viruses (filoviridae, flaviviridae, orthomyxoviridae, and paramyxoviridae), were detected in any of these samples. This is in stark contrast with the October 2019 publication where both coronavirus infection and Sendai virus infection were reportedly detected in the smuggled Malayan pangolins[59], which eventually led to the discovery and publication of the novel pangolin coronaviruses[5-8]. The finding of Lee et al.[68] adds significantly to the existing suspicions and substantiates the possibility that these pangolin coronaviruses do not exist in nature and their sequences could have been fabricated.

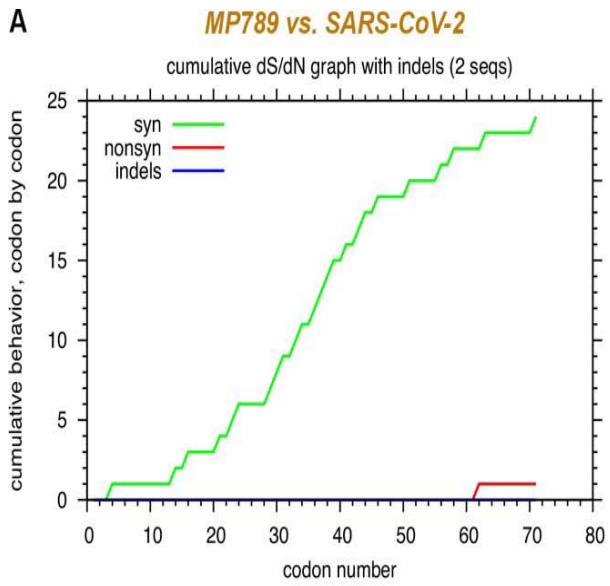
c) The RBD of the reported pangolin coronaviruses binds poorly to pangolin ACE2

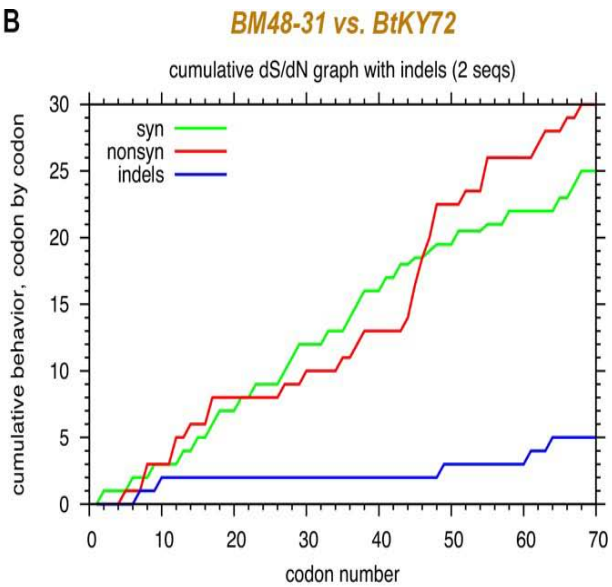
If pangolin coronaviruses truly exist and have recently spilled over to infect humans, their Spike protein, especially the RBD within Spike, should bind to pangolin ACE2 (pACE2) more efficiently than to hACE2. However, recent findings have contradicted this theory. In an *in silico* study, Piplani et al. calculated, following homology structural modeling, the binding energies involved in the association between SARS-CoV-2 Spike and ACE2 from either human or various animals[69]. Interestingly, the most favorable interaction that SARS-CoV-2 Spike makes was shown to be with hACE2, but

not with ACE2 from pangolin or any other suspected intermediate host. Furthermore, another study revealed, using a robust *in vitro* binding assay, that the RBD of SARS-CoV-2 binds much tighter (greater than 9-fold) to hACE2 than to pACE2[45]. Although the RBD of the pangolin coronaviruses is not 100% identical to that of SARS-CoV-2, the RBMs of the two viruses, which is the most essential segment responsible for ACE2 interactions, differ only by one amino acid[5-8]. Therefore, the poor binding efficiency observed between the RBD of SARS-CoV-2 and pACE2[45] infers that the RBD of the reported pangolin coronaviruses must bind to pACE2 fairly inefficiently. Indeed, a very recent study confirmed the case: the RBD of the pangolin coronavirus binds pACE2 ten-fold weaker than to hACE2[70]. These observations once again refute the claim that pangolins are the probable intermediate host for SARS-CoV-2. More importantly, the latter two studies strongly suggest that these viruses might not be able to establish infections in pangolins, which adds significantly to the suspicion that the published sequences of the pangolin coronaviruses may have been fabricated and these viruses do not exist in nature.

d) Genetic evidence proving the fraudulent nature of the pangolin coronaviruses

Evolutionarily, within the coronavirus genome, the RBD of Spike is under the strongest positive selection as it needs to adapt for binding a new receptor whenever the virus crosses the species barrier and enters a new host. In lineage B β coronaviruses, the most essential segment for receptor recognition is the RBM, which fully determines the binding with ACE2. Strikingly, when the *RBM* sequence of the pangolin virus MP789[6] is compared to that of SARS-CoV-2, no positive selection is observed (Figure 7A). Instead, the analysis revealed very strong purifying selection with 24 syn mutations and only one non-syn mutation. In contrast, when two related bat coronaviruses, BM48-31[71] and BtKY72[72], are compared in a similar manner, strong positive selection is observed as expected (Figure 7B). Here, while there are 25 syn mutations, which is comparable to that between MP789 and SARS-CoV-2, the number of non-syn mutations is 30 (Figure 7B). Evidently, the species difference between pangolin and human is greater than that between the hosts of BM48-31 and BtKY72, which are two different species of bats. Therefore, greater positive selection should be expected between MP789 and SARS-CoV-2 than that between BM4831 and BtKY72. The strong purifying selection observed between MP789 and SARS-CoV-2 is, therefore, contradictory to the principles of natural evolution.





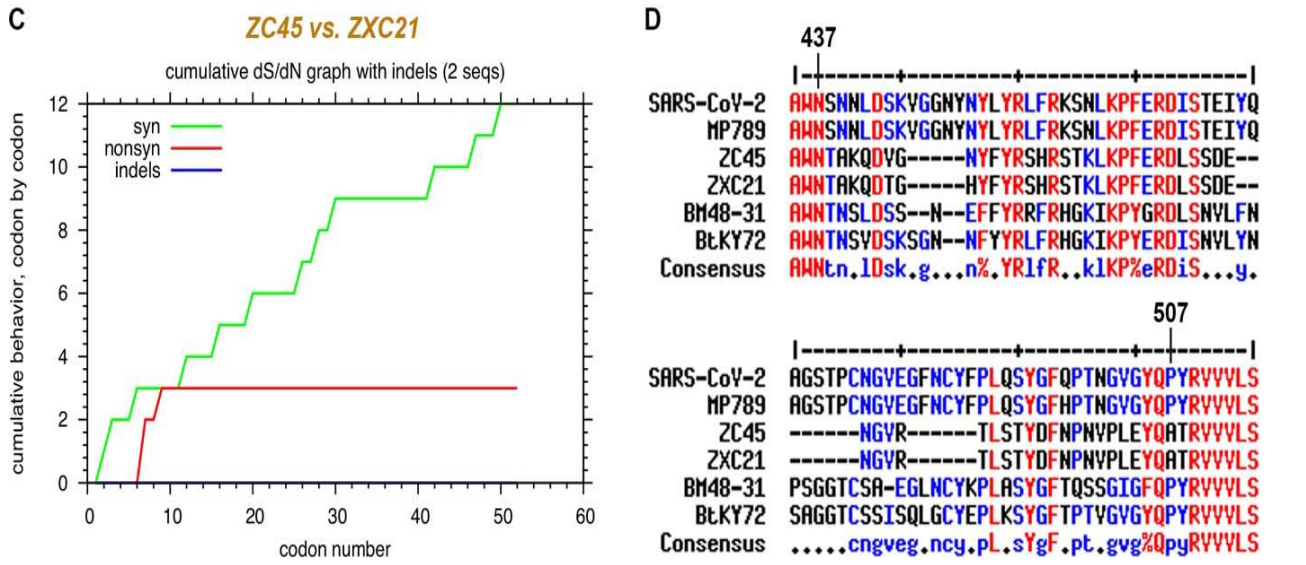


Figure 7: The extremely high purifying pressure observed for the RBM in the comparison of pangolin coronavirus MP789 and SARS-CoV-2 contradicts the principles of natural evolution. Synonymous and nonsynonymous mutations in the RBM region are analyzed between related coronaviruses: A. pangolin coronavirus MP789 (MT121216.1) and SARS-CoV-2 (NC_045512.2), B. bat coronaviruses BM48-31 (NC_014470.1) and BtKY72 (KY352407.1), and C. bat coronaviruses ZC45 and ZXC21. D. Alignment of the RBM sequences from all six viruses. The beginning and end of the RBM are labeled following the sequence of the SARS-CoV-2 Spike.

Table 3: Summary of syn/non-syn mutations in the RBM in three sets of pair-wise comparisons

Viruses being compared	*Genomic sequence identity*	*# of syn mutations in the RBM*	*# of non-syn mutations in the RBM*	*Syn/nonsyn ratio*
MP789 vs. SARS-CoV-2	90.1%	24	1	24:1
BM48-31 vs. BtKY72	82.4%	25	30	0.8:1
ZC45 vs. ZXC21	97.5%	12	3	4:1

We further looked at the syn and non-syn mutations for the RBM in coronaviruses infecting the same species. Here, we compared the closely related coronaviruses ZC45 and ZXC21, which infect the same species of bats[3], on their *RBM* segments (Figure 7C). Here, twelve synonymous mutations and three non synonymous mutations are observed, yielding a syn/non-syn ratio of 4:1. Such a value likely represents the approximate upper limit for the purifying selection in the RBM that such coronaviruses could possibly experience (Table 3). In addition, no purifying selection is observed in the *RBM* for the randomly selected twenty SARS-CoV-2 sequences (Figure 5, codon range 437-507).

Therefore, the extremely high syn/non-syn ratio (24:1) observed between MP789 *RBM* and SARSCoV-2 *RBM* indicates that at least one of the two viruses is artificial.

We believe that, to falsify the natural existence of the unique RBD/RBM of SARS-CoV-2, the amino acid sequence of the pangolin coronavirus RBD/RBM had been fabricated to closely resemble that of SARS-CoV-2. At the same time, the expert(s) carrying out this operation also wanted to create an appropriate level of divergence between the pangolin virus and SARS-CoV-2 at the nucleotide level and thereby introduced a significant amount of syn mutations in the RBM. The abnormality revealed in Figure 7A and Table 3 likely resulted from these fraudulent operations.

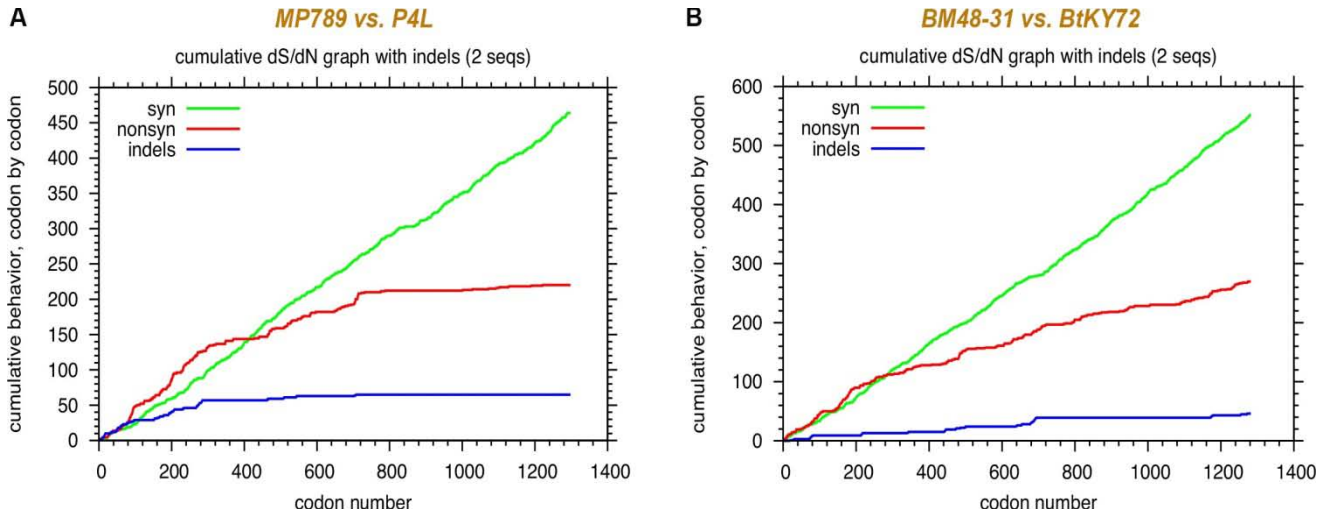


Figure 8: Abnormal distribution of synonymous and non-synonymous mutations in Spike associated with pangolin coronaviruses. A. Comparison between MP789 and P4L (MT040333.1). B. Comparison between the two bat coronaviruses BM48-31 and BtKY72.

Table 4: Ratios of syn/non-syn mutations observed in different viral proteins as revealed by pair-wise comparisons involving pangolin and bat coronaviruses

Protein	MP789 vs. P4L	BM48-31 vs. BtKY72
S2	23.0:1	4.7:1
Spike	2.1:1	2.0:1
Orf1a	2.4:1	1.8:1
Orf1b	7.6:1	5.8:1
N	2.1:1	2.1:1

Similar syn/non-syn analyses on the overall *spike* further revealed the fraudulent nature of these novel pangolin coronaviruses. Here we compared two representative pangolin coronaviruses MP789[6] (a Guangdong isolate) and P4L[5] (a Guangxi isolate) as genomic sequences within each group of isolates share very high sequence identities[13]. As shown in Figure 8A, similar to the abnormal pattern observed between RaTG13 and SARS-CoV-2 (Figure 4A right), syn and non-syn curves exhibit drastically different trajectories and the non-syn curve abruptly flattens in the S2 half of the sequence.

For comparison, we also analyzed the *spike* genes of two SARS-like bat coronaviruses, BM48-31 and BtKY72. The two pangolin coronaviruses, MP789 and P4L, are 85.2% identical on the overall genome, while bat coronaviruses BM48-31 and BtKY72 are 82.4% identical. The comparison here is therefore appropriate. Analysis of the two bat viruses show that the two curves grow naturally in a relatively concerted manner with no excessive flattening of the red curve observed (Figure 8B).

Counting the number of syn and non-syn mutations in each pair of comparisons further illustrated the unnatural characteristics associated with the pangolin coronaviruses (Table 4). While the S2 protein is not expected to be more conserved than Orf1b, the syn/non-syn ratio for S2 observed in the comparison between MP789 and P4L is abnormally high (207 syn mutations and 9 non-syn mutations; syn/non-syn = 23:1), which is far exceeding what is observed for Orf1b (7.6:1).

As the two bat coronaviruses here were discovered in nature independently by research groups outside of China[71,72], the features displayed in Figure 8B likely represent the approximate evolutionary trait of two coronaviruses at this level of overall divergence. According to the logic described earlier, the great contrast between Figure 8A and 8B and the abnormal syn/non-syn ratio of 23:1 (Table 4) further prove that, between MP789 and P4L, at least one is artificial, although we believe both groups of pangolin coronaviruses represented by MP789 and P4L, respectively, are non-natural and fabricated.

e) Summary and discussion

A single source of samples was used for all studies (some spuriously independent) reporting novel pangolin coronaviruses. The formats of sequencing reads were manipulated with a clear intention to hide the fact that the same dataset was used in different studies. The raw sequencing data is missing for certain critical pieces, poor in quality, and suspicious in terms of the amounts and types of contaminations present. The RBD

pieces, poor in quality, and suspicious in terms of the amounts and types of contaminations present. The RBD encoded by the reported sequence of pangolin coronaviruses could not bind pACE2 efficiently. As revealed by syn/non-syn analyses, sequences of the RBM and S2 regions of these pangolin coronaviruses exhibit features that are inconsistent with natural evolution. Finally, no coronavirus was detected in a large, decade-long surveillance study of Malayan pangolins. These observations and evidence converge to prove that these recently reported pangolin coronaviruses do not exist in nature and their sequences must have been fabricated.

It is noteworthy that the abnormal syn/non-syn feature revealed for S2 in the comparison between MB789 and P4L (Figure 8A) resembles greatly that exhibited by the comparison between RaTG13 and SARS-CoV-2 (Figure 4A right). Judging based on this reoccurring pattern, we believe that the sequence fabrications in both cases (RaTG13 and pangolin coronaviruses) were most likely carried out by the same person or group, whose misconception of the spike gene evolution persisted in multiple such practices and resulted in the unnatural look of the syn/non-syn curves and numbers (Figure 4, Table 1, Figure 8, and Table 4).

III. Evidence Revealing the Fraudulent Nature of the Novel Bat Coronavirus RmYN02

While the publications of the fabricated pangolin coronaviruses might have seemingly fulfilled the scientific quests for an intermediate host for the zoonosis of SARS-CoV-2 as well as for an evolutionary origin of its RBD, it had remained suspicious and unexplainable how SARS-CoV-2 could have acquired the furin-cleavage site (*-PRRAR/VS-*) at the S1/2 junction through natural evolution. It is evident that, although furin-cleavage site has been found in certain other lineages of coronaviruses at the S1/2 junction, lineage B β coronaviruses clearly lack the ability to develop this motif at this location naturally[58].

In early June, another novel bat coronavirus, RmYN02, was reported[9], which shares a 93.3% sequence identity with SARS-CoV-2 and appears to be the second closest bat coronavirus to SARS-CoV-2 (the closest is allegedly RaTG13). This finding adds yet another member to the rapidly growing sub-lineage of SARS-CoV-2-like coronaviruses (Figure 9), which has been completely vacant and practically nonexistent prior to the current pandemic. In addition, importantly, RmYN02 carries a unique sequence *-PAA-* at the S1/S2 junction, which remotely resembles the inserted *-PRRA-* sequence at the same location in the SARS-CoV-2 Spike. Despite the fact that *-PAA-* in RmYN02 only partially resembles the *-PRRA-* insertion in SARS-CoV-2 and does not appear to be an actual insertion if properly aligned[18], the authors nonetheless claimed that the natural occurrence of *-PAA-* in RmYN02 proves that the *-PRRA-* sequence could very likely be acquired and "inserted" into the same location in SARS-CoV-2 genome through natural evolution[9].

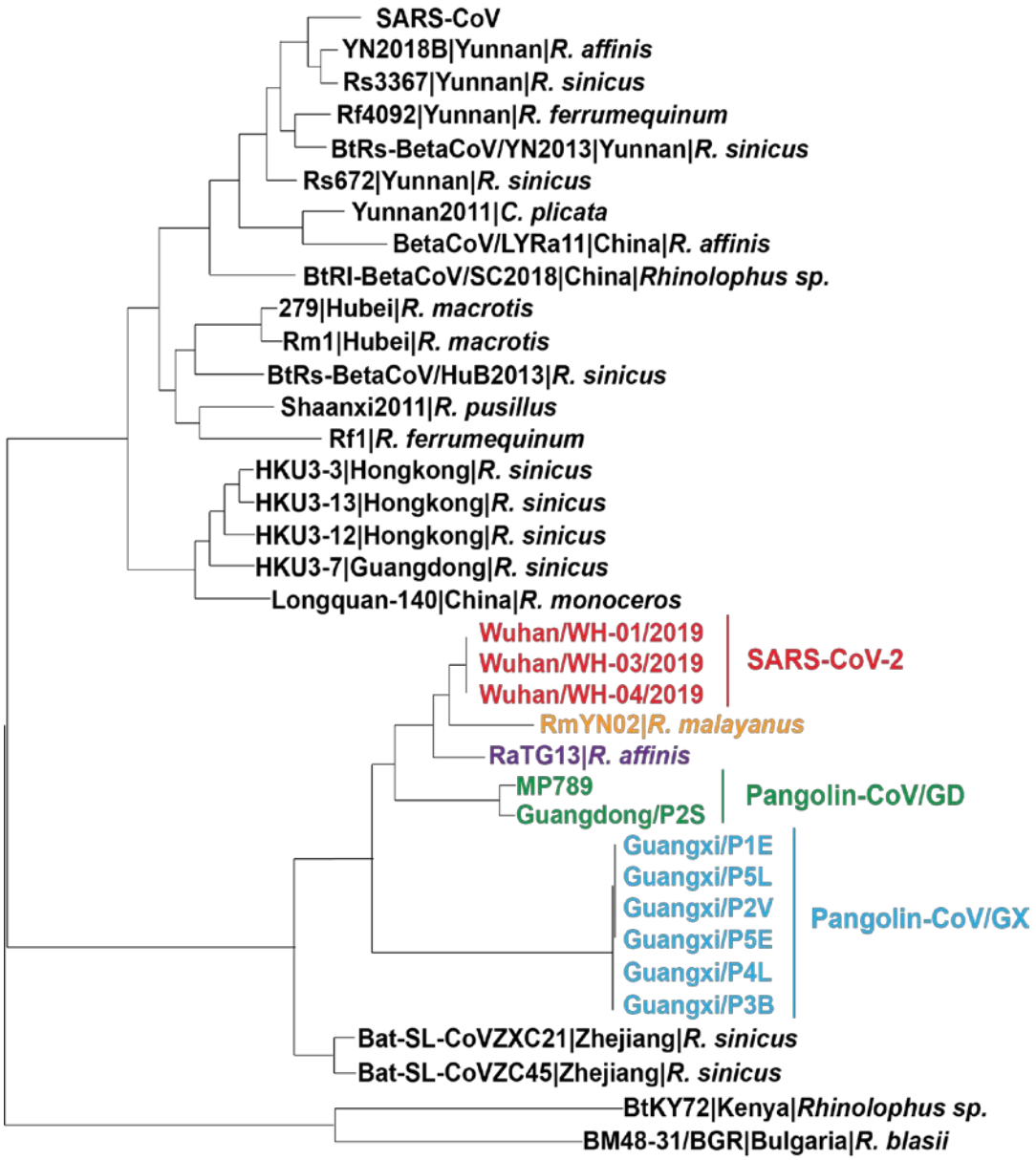


Figure 9: Phylogenetic analysis of SARS-CoV-2 and representative viruses from the subgenus sarbecoronavirus. *Figure redrew from Zhou et al[9]. Colored viruses were all reported after the COVID-19 outbreak.*

The fact that a poor alignment was used to make a disproportional, strong argument for an evolutionary origin of the furin-cleavage site, which appeared to be the last missing piece of the puzzle, is suspicious. Furthermore, despite the significance of the *spike* sequence of RmYN02 in supporting the central conclusion of the publication, the raw sequencing reads for *spike* has not been made available although the authors stated otherwise in the article[9]. This is yet another repeat of the pattern that has been exhibited in the reporting of both RaTG13 and pangolin coronaviruses, where the genomic sequence would be published first and the raw sequencing reads would not be made available months afterwards.

Given that the CCP-controlled laboratories have repeatedly engaged in fabrication of coronaviruses to feed the missing pieces for the puzzle, the above suspicion opens up the possibility that the RmYN02 virus could have been fabricated as well. Judging from the fact that its sequence identity to SARS-CoV-2 (93.3%) is lower than that between RaTG13 and SARS-CoV-2 (96.2%), we suspected that the sequence of RmYN02 might have been fabricated by modifying the sequence of RaTG13. Such an approach could easily ensure that the evolutionary distance between RmYN02 and SARS-CoV-2 is greater than that between RaTG13 and SARS-CoV-2. It also ensures that RmYN02 and RaTG13 would appear to be evolutionarily close, consistent with the claim that they both infect bats although of different species.

We therefore compared the *spike* genes of RmYN02 and RaTG13 on the quantity and distribution of

syn and non-syn mutations. The severe divergence at the S1 portion between the two viral sequences did not allow the *S1* sequences to be properly codon-aligned. Therefore, only the *S2* half was analyzed (Figure 10). For the beginning 200 codons of S2, both types of mutations accumulate steadily and gradually. However, for the final 378 codons, once again, the non-syn curve flattens and the concerted growth of the two curves has disappeared. In this region, there are 57 syn mutations and only one non-syn mutation. The syn/non-syn ratio of 57:1 for a region as wide as 378 codons (1,344 nucleotides) is severely inconsistent with what is observed naturally (Figure 4A left and Figure 8B)[41].

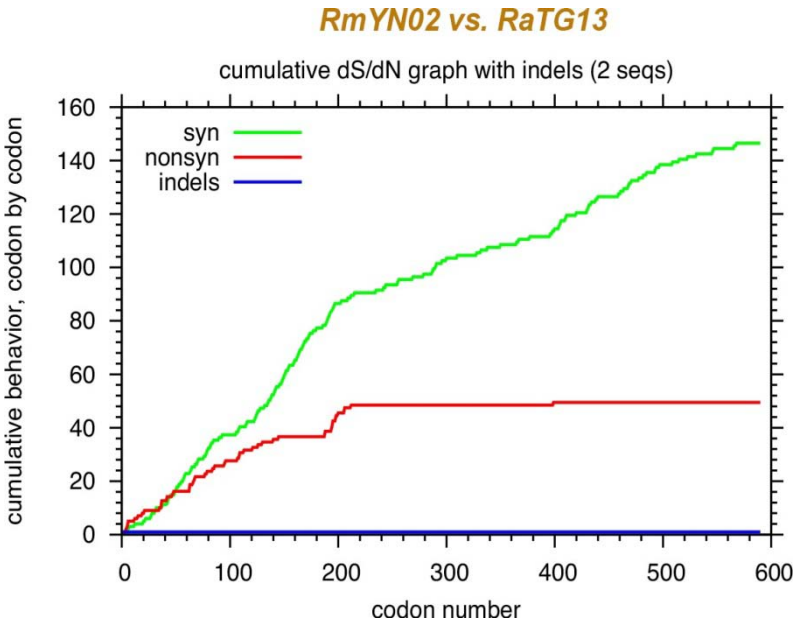


Figure 10: Analysis of synonymous and non-synonymous mutations in S2 between RmYN02 and RaTG13. The abrupt change of trajectory of the non-synonymous mutation (red) curve and its subsequent flattening are observed.

Logically, between RaTG13 and RmYN02, at least one must be artificial. Here, however, we are convinced that both viruses are artificial. As shown in part 1, the sequence of RaTG13 must have been fabricated. Therefore, the fact that the last 378 codons of RmYN02's S2 are identical, with the exception of one, to that of RaTG13 proves that the RmYN02 sequence must be artificial as well. This also proves our earlier suspicion that the RaTG13 sequence should have been used as the template for the fabrication of the RmYN02 sequence. RaTG13 was published in late January[4], while RmYN02 was published in early June (manuscript submitted in April)[9]. Therefore, enough time is in between for the sequence fabrication to be carried out.

While introducing nucleotide changes to create the apparent divergence between the two viruses, the expert(s) may have overly restricted amino acid changes in this part of Spike. Again, the abrupt change of trajectory of the non-syn curve and its excessive flattening later in the sequence likely reflect their overestimation of the purifying selection pressure on S2. The fact that this abnormal pattern has been observed in all three cases (Figure 4A right, 8A, and 10) reiterates the point raised in section 2.5 that all sequence fabrications may have been carried out by the same person or group.

IV. Final Discussion and Remarks

a) All fabricated coronaviruses share a 100% amino acid sequence identity on the E protein with ZC45 and ZXC21

Evidence herein clearly indicates that the novel coronaviruses recently published by the CCP controlled laboratories are all fraudulent and do not exist in nature. One final proof of this conclusion is the fact that all of these viruses share a 100% amino acid sequence identity on the E protein with bat coronaviruses ZC45 and ZXC21, which, as revealed in our earlier report[1], should be the template/backbone used for the creation of SARS-CoV-2 (Figure 11). Despite its conserved function in the viral replication cycle, the E protein is tolerant and permissive of amino acid mutations[1]. It is therefore impossible for the amino acid sequence of the E protein to remain unchanged when the virus has allegedly crossed species barrier multiple times (between different bat species, from bats to pangolins, and from pangolins to humans). The 100% identity observed here, therefore, further proves that the sequences of these recently published novel coronaviruses have been fabricated.

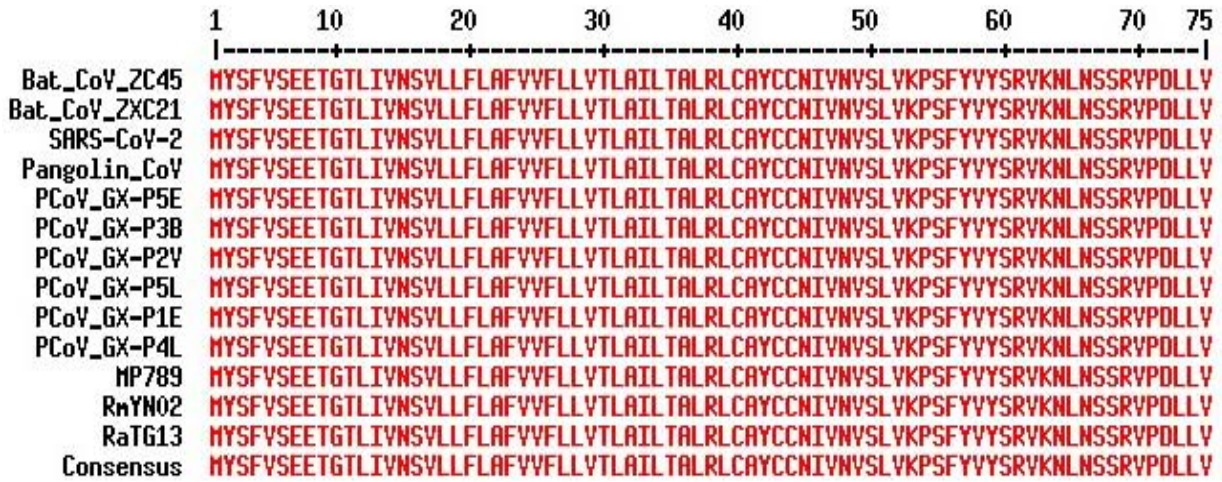


Figure 11: All novel coronaviruses recently published by the CCP-controlled laboratories share a 100% amino acid sequence identity on the E protein with ZC45 and ZXC21. Additional accession numbers of viruses: SARSCoV-2 (NC_045512.2), Pangolin-CoV (EPI_ISL_410721), P5E (MT040336.1), P3B (MT072865.1), P2V (MT072864.1), P5L (MT040335.1), and P1E (MT040334.1).

A main goal of these fabrications was to obscure the connection between SARS-CoV-2 and ZC45/ZXC21. Therefore, from their perspective, the fabricated viruses should resemble SARS-CoV-2 more than ZC45 and ZXC21 do. Because ZC45 and ZXC21 already share a 100% identity with SARSCoV-2 on the E protein, the fabricated viruses therefore were made to adopt this sequence completely as well.

b) Important implications of this large-scale, organized scientific fraud

If SARS-CoV-2 is of a natural origin, no fabrications would be needed to suggest so. The current paper, therefore, corroborates our earlier paper and further proves that SARS-CoV-2 is a laboratory product[1]. As revealed[1], the creation of SARS-CoV-2 is convenient by following established concepts and techniques, some of which (for example, restriction enzyme digestion) are considered classic and yet still preferred widely including by experts of the field[35,73]. A key component of the creation, the template virus ZC45/ZXC21, is owned by military research laboratories[3].

Importantly, as revealed here, multiple research laboratories and institutions have engaged in the fabrication and cover-up[4-9,59]. It is clear that this was an operation orchestrated by the CCP government.

In addition, raw sequencing reads for RaTG13, which were integral parts of the fabrication, were obtained in 2017 and 2018[24,33]. Furthermore, manuscript reporting the falsified coronavirus infections of Malayan pangolins was submitted for publication in September 2019[59]. Evidently, the cover-up had been planned and initiated before the COVID-19 outbreak. Therefore, the unleashing of the virus must be a planned execution rather than an accident.

c) SARS-CoV-2 is an Unrestricted Bioweapon

Although it is not easy for the public to accept SARS-CoV-2 as a bioweapon due to its relatively low lethality, this virus indeed meets the criteria of a bioweapon as described by Dr. Ruifu Yang. Aside from his appointment in the AMMS, Dr. Yang is also a key member of China's *National and Military Bioterrorism Response Consultant Group* and had participated in the investigation of the Iraqi bioweapon program as a member of the *United Nations Special Commission (UNSCOM)* in 1998. In 2005, Dr. Yang specified the criteria for a pathogen to qualify as a bioweapon[74]:

1. It is significantly virulent and can cause large scale casualty.
2. It is highly contagious and transmits easily, often through respiratory routes in the form of aerosols. The most dangerous scenario would be that it allows human-to-human transmission.
3. It is relatively resistant to environmental changes, can sustain transportation, and is capable of supporting targeted release.

All of the above have been met by SARS-CoV-2: it has taken millions of lives, led to numerous hospitalizations, and left many with sequela and various complications; it spreads easily by contact, droplets, and aerosols *via* respiratory routes and is capable of transmitting from human to human[75-77], the latter of which was initially covered up by the CCP government and the WHO and was first revealed by Dr. Li-Meng Yan on January 19th, 2020 on *Lude Press*[78]; it is temperature-insensitive (unlike seasonal flu) and remains viable for a long period of time on many surfaces and at 4°C (e.g. the ice/water mixture)[79,80].

Adding to the above properties is its high rate of asymptomatic transmission, which renders the control of

SARS-CoV-2 extremely challenging. In addition, the transmissibility, morbidity, and mortality of SARS-CoV-2 also resulted in panic in the global community, disruption of social orders, and decimation of the world's economy. The range and destructive power of SARS-CoV-2 are both unprecedented.

Clearly, SARS-CoV-2 not only meets but also surpasses the standards of a traditional bioweapon. Therefore, it should be defined as an Unrestricted Bioweapon.

d) The current pandemic is an attack on humanity

The scientific evidence and records indicate that the current pandemic is not a result of accidental release of a gain-of-function product but a planned attack using an Unrestricted Bioweapon. The current pandemic therefore should be correspondingly considered as a result of Unrestricted Biowarfare.

Under such circumstances, the infected population are being used, unconsciously, as the vectors of the disease to facilitate the spread of the infection. The first victims of the attack were the Chinese people, especially those in the city of Wuhan. At the initial stage, the hidden spread in Wuhan could have also served another purpose: the final verification of the bioweapon's functionality, an important aspect of which is the human-to-human transmission efficiency. Upon the success of this last step, targeted release of the pathogen might have been enabled.

Given the global presence of SARS-CoV-2 and the likelihood of its long-term persistence, it is appropriate to say that this attack was on the humanity as a whole and has put its fate at risk.

e) Actions need to be taken to combat the current pandemic and save the future of humanity

Given the CCP's role here, it is of paramount importance that the CCP is held accountable for its actions. In addition, the world needs to find out what other variants of SARS-CoV-2 exist in the CCP-controlled laboratories, whether or not SARS-CoV-2 or its variant(s) are still being actively released, whether or not re-infection of SARS-CoV-2 leads to worsened outcomes due to inefficient immunity and/or antibody dependent enhancement (ADE)[81-83], and whether other weaponized pathogens are owned by the CCP as a result of their excessive, state-stimulated efforts in collecting novel animal pathogens and studying their potentials in zoonosis [3,25,26,28,32,36,37,84-114].

It is also of paramount importance that all the hidden knowledge of SARS-CoV-2 be brought out as soon as possible. As illustrated in our earlier paper, although a template virus was used, the creation of SARS-CoV-2 must have involved introducing changes to the template sequence through DNA synthesis (steps 1 and 4 in part 2 of our earlier paper)[1]. Such a practice can be safely guided by multi-sequence alignment of available SARS and SARS-like coronavirus sequences. The process of this practice has been illustrated[115], and both syn mutations and amino acid (non-syn) mutations at variable positions/regions would be introduced. From the perspective of the responsible scientists, these changes are necessary because, otherwise, the engineered nature of the virus and its connection to its template would be evident. However, importantly, the introduced changes might have also altered the functions of the various viral components, which could be either by design or unintended. Nonetheless, it remains to be answered whether or how the introduced changes might be responsible for the various lasting complications that many COVID-19 patients experience and what barriers these changes might pose to the development of effective vaccines and other antiviral therapeutics. It is reasonable to believe that the responsible laboratories under the control of the CCP have been engaged in this research for a long period of time and therefore keep in possession a considerable amount of concealed knowledge of SARS-CoV-2. Some of the knowledge may provide answers to questions that need to be addressed urgently in the global combat against COVID-19. Such hidden knowledge ought to be made available to the world immediately.

What also need to be held accountable are the individuals and groups within certain organizations and institutions in the fields of public health and academic research, who knowingly and collaboratively facilitated the CCP's misinformation campaign and misled the world. On January 18th and 19th, 2020, Dr. Li-Meng Yan, then anonymously, first revealed that SARS-CoV-2 is of a laboratory origin[78,116]. Immediately afterwards, on January 20th, Dr. Zhengli Shi submitted her manuscript to *Nature* and reported the first fabricated virus, RaTG13[4]. Since then, many virus fabrications have taken place and all of them were published as peer-reviewed articles on top scientific journals[4-9]. Subsequently, based on such reports, influential opinion articles promoting the natural origin theory have then been published by prominent scientists and international organizations on such and other high-profile platforms[10,117-120].

In contrast to the rigorous promotion of the natural origin theory, strict censorship has been placed by these and other journals on manuscripts discussing a possible laboratory origin of SARS-CoV-2[18,121]. Our earlier report[1], which was one of such manuscripts and published as a preprint article, also faced unfounded criticisms dressed as unbiased peer reviews from two groups of scientists led by Drs. Robert Gallo and Nancy Connell, respectively[122,123] (*our point-to-point responses are being prepared and will be published soon*). As a result of this collaborative efforts, the public has been largely removed from the truth about COVID-19 and SARS-CoV-2, which has led to misjudgments, delayed actions, and greater sufferings of the global community. It is imperative to investigate the scientists, laboratories, institutions, and relevant collaborators responsible for

the creation of SARS-CoV-2 and for the fabrications/cover-up. It is also imperative to investigate the relevant individuals in the WHO, at the relevant scientific journals, in the relevant funding agencies, and in other relevant bodies, which have facilitated the creation of SARSCoV-2 and the scientific cover-up of its true origin while under full awareness of the nature of these operations. Finally, it also needs to be investigated which ones of the scientists engaged in the promotion of the natural origin theory were purely misled by the scientific fraud and which ones were colluding with the CCP government.

The time has come that the world faces the truth of COVID-19 and takes actions to save the future of humanity.

Acknowledgements

We thank Daoyu Zhang for sharing with us the observation of abnormal distribution of nonsynonymous mutations between RaTG13 Spike and SARS-CoV-2. We thank Francisco de Asis for revealing the filenames of the raw sequencing reads for RaTG13. We also thank other individuals, including anonymous scientists, for uncovering various facts associated with the origin of SARS-CoV-2.

References Références Referencias

1. Yan, L.-M., Kang, S., Guan, J. & Hu, S. Unusual Features of the SARS-CoV-2 Genome Suggesting Sophisticated Laboratory Modification Rather Than Natural Evolution and Delineation of Its Probable Synthetic Route. *Zenodo.org (preprint)*, http://doi.org/10.5281/zenodo.4028830 (2020).
2. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature*579, 265-269 (2020).
3. Hu, D. et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg Microbes Infect*7, 154 (2018).
4. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*579, 270–273 (2020).
5. Lam, T.T. et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* (2020).
6. Liu, P. et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoSPathog*16, e1008421 (2020).
7. Xiao, K. et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* (2020).
8. Zhang, T., Wu, Q. & Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID19 Outbreak. *Curr Biol*30, 1578 (2020).
9. Zhou, H. et al. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Curr Biol*30, 2196-2203 e3 (2020).
10. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C. & Garry, R.F. The proximal origin of SARSCoV-2. *Nat Med*26, 450-452 (2020).
11. Boni, M.F. et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol* (2020).
12. Hu, B., Guo, H., Zhou, P. & Shi, Z. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*, https://doi.org/10.1038/s41579-020-00459-7 (2020).
13. Chan, Y.A. & Zhan, S.H. Single source of pangolin CoVs with a near identical Spike RBD to SARSCoV-2. *bioRxiv*, https: //doi.org/10.1101/2020.07.07.184374 (2020).
14. Hassanin, A. The SARS-CoV-2-like virus found in captive pangolins from Guangdong should be better sequenced. *bioRxiv*, https: //doi.org/10.1101/2020.05.07.077016 (2020).
15. Lin, X. & Chen, S. Major Concerns on the Identification of Bat Coronavirus Strain RaTG13 and Quality of Related Nature Paper. *Preprints*, 2020060044 (2020).
16. Rahalkar, M. & Bahulikar, R. The Abnormal Nature of the Fecal Swab Sample used for NGS Analysis of RaTG13 Genome Sequence Imposes a Question on the Correctness of the RaTG13 Sequence. *Preprints.org*, 2020080205 (2020).
17. Rahalkar, M.C. & Bahulikar, R.A. Understanding the Origin of 'BatCoVRaTG13', a Virus Closest to SARS-CoV-2. *Preprints*, 2020050322 (2020).
18. Segreto, R. & Deigin, Y. Is considering a genetic-manipulation origin for SARS-CoV-2 a conspiracy theory that must be censored? *Preprint (Researchgate)* DOI: 10.13140/ RG.2.2.31358.13129/1 (2020).
19. Seyran, M. et al. Questions concerning the proximal origin of SARS-CoV-2. *J Med Virol* (2020).
20. Zhang, D. The Pan-SL-CoV/GD sequences may be from contamination. *Preprint (zenodo.org)*, DOI: 10.5281/zenodo.3885333 (2020).
21. Zhang, D. Anomalies in BatCoV/RaTG13 sequencing and provenance. *Preprint (zenodo.org)*, https://zenodo.org/record/3987503#.Xz9GzC-z3GI (2020).
22. Lab That First Shared Novel Coronavirus Genome Still Shut Down by Chinese Government. *Global Biodefense*, https://globalbiodefense.com/headlines/chinese-lab-that-first-shared-novel-coronavirusgenome-shut-down/ (2020).
23. CGTN Exclusive: Director of Wuhan Institute of Virology says 'let science speak'. *CGTN*, https://news.cgtn.com/news/2020-05-23/Exclusive-with-head-of-Wuhan-Institute-of-Virology-Letscience-speak-QJeOjOZt4Y/index.html (2020).
24. Cohen, J. Wuhan coronavirus hunter Shi Zhengli speaks out. *Science*, https://science.sciencemag.org/content/369/6503/487 (2020).

25. Hu, B. et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoSPathog*13, e1006698 (2017).
26. Ge, X.Y. et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. *Virol Sin*31, 31-40 (2016).
27. Courtney-Guy, S. Chinese scientists 'found closest relative of coronavirus seven years ago'. https://metro.co.uk/2020/07/05/chinese-scientists-found-closest-relative-coronavirus-seven-years-ago12948668/ (2020).
28. Wu, Z. et al. Novel Henipa-like virus, Mojiang Paramyxovirus, in rats, China, 2012. *Emerg Infect Dis*20, 1064-6 (2014).
29. Qiu, J. How China's 'Bat Woman' Hunted Down Viruses from SARS to the New Coronavirus. *Scientific American*, https://www.scientificamerican.com/article/how-chinas-bat-woman-hunted-down-virusesfrom-sars-to-the-new-coronavirus1/ (2020).
30. Li, X. The Analysis of Six Patients With Severe Pneumonia Caused By Unknown Viruses. *Master's Thesis*, https://www.documentcloud.org/documents/6981198-Analysis-of-Six-Patients-With-Unknown Viruses.html (2013).
31. Huang, C. Novel Virus Discovery in Bat and the Exploration of Receptor of Bat Coronavirus HKU9. *PhD Dissertation (in Chinese)*, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention (2016).
32. Ge, X.Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*503, 535-8 (2013).
33. Names of the RaTG13 Amplicon Sequences. https://graph.org/RaTG13-Amplicon-Names-07-03 (2020).
34. Zeng, L.P. et al. Bat Severe Acute Respiratory Syndrome-Like Coronavirus WIV1 Encodes an Extra Accessory Protein, ORFX, Involved in Modulation of the Host Immune Response. *J Virol*90, 6573-6582 (2016).
35. Shang, J. et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature*581, 221-224 (2020).
36. Ren, W. et al. Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. *J Virol*82, 1899-907 (2008).
37. Yang, X.L. et al. Isolation and Characterization of a Novel Bat Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus. *J Virol*90, 3253-6 (2015).
38. Luo, C.M. et al. Discovery of Novel Bat Coronaviruses in South China That Use the Same Receptor as Middle East Respiratory Syndrome Coronavirus. *J Virol*92, DOI: 10.1128/JVI.00116-18 (2018).
39. Wang, N. et al. Characterization of a New Member of Alphacoronavirus with Unique Genomic Features in Rhinolophus Bats. *Viruses*11, https://doi.org/10.3390/v11040379 (2019).
40. Korber, B. HIV Signature and Sequence Variation Analysis. *Computational Analysis of HIV Molecular Sequences* Chapter 4, 55-72 (2000).
41. Shijulal Nelson-Sathi, P.U., E Sreekumar, R Radhakrishnan Nair, Iype Joseph, Sai Ravi Chandra Nori, Jamiema Sara Philip, Roshny Prasad, KV Navyasree, Shikha Ramesh, Heera Pillai, Sanu Ghosh, TR Santosh Kumar, M. Radhakrishna Pillai. Structural and Functional Implications of Non-synonymous Mutations in the Spike protein of 2,954 SARS-CoV-2 Genomes. *bioRxiv*, https://doi.org/10.1101/2020.05. 02.071811 (2020).
42. Li, X. et al. Emergence of SARS-CoV-2 through Recombination and Strong Purifying Selection. *Science Advances*6, eabb9153 (2020).
43. Wang, H., Pipes, L. & Nielsen, R. Synonymous mutations and the molecular evolution of SARS-Cov-2 origins. *bioRxiv*, https://doi.org/10.1101/2020.04.20.052019 (2020).
44. Tang, X. et al. On the origin and continuing evolution of SARS-CoV-2. *National Science Review*7, 1012–1023 (2020).
45. Mou, H. et al. Mutations from bat ACE2 orthologs markedly enhance ACE2-Fc neutralization of SARSCoV-2. *bioRxiv*, https://doi.org/10.1101/2020.06.29.178459 (2020).
46. Latham, J. & Wilson, A. A Proposed Origin for SARS-CoV-2 and the COVID-19 Pandemic. *Independent Science News*, https://www.Independentsciencenews.org/commentaries/a-proposed-origin-for-sars-cov-2and-the-covid-19-pandemic/ (2020).
47. Liu, Z.L. et al. Antibody Profiles in Mild and Severe Cases of COVID-19. *Clin Chem*66, 1102-1104 (2020).
48. Hicks, J. et al. Serologic cross-reactivity of SARS-CoV-2 with endemic and seasonal Beta coronaviruses. *medXriv*, https://doi.org/10.1101/2020.06.22.20137695 (2020).
49. Anand, P., Puranik, A., Aravamudan, M., Venkatakrishnan, A.J. & Soundararajan, V. SARS-CoV-2 strategically mimics proteolytic activation of human ENaC. *Elife*9, e58603 (2020).
50. Vallet, V., Chraibi, A., Gaeggeler, H.P., Horisberger, J.D. & Rossier, B.C. An epithelial serine protease activates the amiloride-sensitive sodium channel. *Nature*389, 607-10 (1997).
51. Claas, E.C. et al. Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet*351, 472-7 (1998).
52. Ito, T. et al. Generation of a highly pathogenic avian influenza A virus from an avirulent field isolate by passaging in chickens. *J Virol*75, 4439-43 (2001).

53. Watanabe, R. et al. Entry from the cell surface of severe acute respiratory syndrome coronavirus with cleaved S protein as revealed by pseudotype virus bearing cleaved S protein. *J Virol*82, 11985-91 (2008).
54. Belouzard, S., Chu, V.C. & Whittaker, G.R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci U S A*106, 5871-6 (2009).
55. Sun, X., Tse, L.V., Ferguson, A.D. & Whittaker, G.R. Modifications to the hemagglutinin cleavage site control the virulence of a neurotropic H1N1 influenza virus. *J Virol*84, 8683-90 (2010).
56. Kido, H. et al. Role of host cellular proteases in the pathogenesis of influenza and influenza-induced multiple organ failure. *BiochimBiophys Acta*1824, 186-94 (2012).
57. Cheng, J. et al. The S2 Subunit of QX-type Infectious Bronchitis Coronavirus Spike Protein Is an Essential Determinant of Neurotropism. *Viruses*11, doi: 10.3390/v11100972 (2019).
58. Hoffmann, M., Kleine-Weber, H. & Pohlmann, S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol Cell*78, 779-784 e5 (2020).
59. Liu, P., Chen, W. & Chen, J.P. Viral Metagenomics Revealed Sendai Virus and Coronavirus Infection of Malayan Pangolins (Manis javanica). *Viruses*11, doi: 10.3390/v11110979 (2019).
60. Tommy Tsan-Yuk Lam, M.H.-H.S., Hua-Chen Zhu, Yi-Gang Tong, Xue-Bing Ni, Yun-Shi Liao, Wei Wei, William Yiu-Man Cheung, Wen-Juan Li, Lian-Feng Li, Gabriel M Leung, Edward C. Holmes, YanLing Hu, Yi Guan. Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. *bioRxiv*, doi.org/10.1101/2020.02.13.945485 (2020).
61. University, S.C.A. 华南农大发现穿山甲为新型冠状病毒潜在中间宿主 (South China Agricultural University Found that Pangolins Are The Possible Intermediate Host of SARS-CoV-2). *www.edu.cn*, http://www.edu.cn/ke_yan_yu_fa_zhan/gao_xiao_cheng_guo/gao_xiao_zi_xun/202002/t20200207_1710427.shtml (2020).
62. 华南农业大学：穿山甲为新型冠状病毒潜在中间宿主 (South China Agricultural University: Pangolins Are The Possible Intermediate Host of SARS-CoV-2). *IFENG NEWS*, https://news.ifeng.com/c/7tr8u2sAQFc (2020).
63. Kangpeng Xiao, J.Z., Yaoyu Feng, Niu Zhou, Xu Zhang, Jie-Jian Zou, Na Li, Yaqiong Guo, Xiaobing Li, Xuejuan Shen, Zhipeng Zhang, Fanfan Shu, Wanyi Huang, Yu Li, Ziding Zhang, Rui-Ai Chen, Ya-Jiang Wu, Shi-Ming Peng, Mian Huang, Wei-Jun Xie, Qin-Hui Cai, Fang-Hui Hou, Yahong Liu, Wu Chen, Lihua Xiao, Yongyi Shen. Isolation and Characterization of 2019-nCoV-like Coronavirus from Malayan Pangolins. *bioRxiv*, doi.org/10.1101/2020.02.17.951335 (2020).
64. Bi, S. et al. Complete genome sequences of the SARS-CoV: the BJ Group (Isolates BJ01-BJ04). *Genomics Proteomics Bioinformatics*1, 180-92 (2003).
65. Qin, E. et al. A genome sequence of novel SARS-CoV isolates: the genotype, GD-Ins29, leads to a hypothesis of viral transmission in South China. *Genomics Proteomics Bioinformatics*1, 101-7 (2003).
66. Qin, E. et al. A complete sequence and comparative analysis of a SARS-associated virus (Isolate BJ01). *Chin Sci Bull*48, 941-948 (2003).
67. Zhu, X. et al. Genetic variation of the human alpha-2-Heremans-Schmid glycoprotein (AHSG) gene associated with the risk of SARS-CoV infection. *PLoS One*6, e23730 (2011).
68. Lee, J. et al. No evidence of coronaviruses or other potentially zoonotic viruses in Sunda pangolins (Manis javanica) entering the wildlife trade via Malaysia. *bioRxiv*, https://doi.org/10.1101/2020.06.19.158717 (2020).
69. Piplani, S., Singh, P.K., Winkler, D.A. & Petrovsky, N. In silico comparison of spike protein-ACE2 binding affinities across species; significance for the possible origin of the SARS-CoV-2 virus. *arXiv*, arXiv:2005.06199 (2020).
70. Wrobel, A.G. et al. Structure and binding properties of Pangolin-CoV Spike glycoprotein inform the evolution of SARS-CoV-2. *Research Square*, DOI: 10.21203/rs.3.rs-83072/v1 (2020).
71. Drexler, J.F. et al. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J Virol*84, 11336-49 (2010).
72. Tao, Y. & Tong, S. Complete Genome Sequence of a Severe Acute Respiratory Syndrome-Related Coronavirus from Kenyan Bats. *Microbiol Resour Announc*8(2019).
73. Hou, Y.J. et al. SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell*182, 429-446 e14 (2020).
74. 生物恐怖袭击——并非杞人忧天 (Bioterrarism Attack — Not An Unfounded Concern). 保健时报 *(Health Times)*, http://www.bjsbnet.com/Article/InArticle/BJSB200505120031 (2005).
75. COVID-19 Overview and Infection Prevention and Control Priorities in non-US Healthcare Settings. *cdc.org*, https://www.cdc.gov/coronavirus/2019-ncov/hcp/non-ussettings/overview/index.html#:~:text=COVID%2D19%20is%20primarily,inhaled%20into%20the%20lun gs (2020).

76. Sia, S.F. et al. Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. *Nature* (2020).
77. Scientific Brief: SARS-CoV-2 and Potential Airborne Transmission. *cdc.org*, https://www.cdc.gov/coronavirus/2019-ncov/more/scientific-brief-sars-cov-2.html (2020).
78. The Morning Show (with An Hong and Ai Li) on Jan 19th, 2020. *Lude Press (YouTube)*, https://www.youtube.com/watch?v=CLTjg03CPEs (2020).
79. Chin, A.W.H. et al. Stability of SARS-CoV-2 in different environmental conditions. *Lancet Microbe*1, e10 (2020).
80. Van Doremalen, N. et al. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *N Engl J Med*382, 1564-1567 (2020).
81. Eroshenko, N. et al. Implications of antibody-dependent enhancement of infection for SARS-CoV-2 countermeasures. *Nat Biotechnol*38, 789-791 (2020).
82. Bowen, T. Nevada State Public Health Lab-led team studying COVID-19 reinfection. *News & Events (University of Nevada, Reno, School of Medicine)*, https://med.unr.edu/news/archive/2020/covid-19reinfection (2020).
83. Kam, Y.W. et al. Antibodies against trimeric S glycoprotein protect hamsters against SARS-CoV challenge despite their capacity to mediate FcgammaRII-dependent entry into B cells in vitro. *Vaccine*25, 729-40 (2007).
84. Ren, W. et al. Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis. *J Gen Virol*87, 3355-9 (2006).
85. Yuan, J. et al. Intraspecies diversity of SARS-like coronaviruses in Rhinolophus sinicus and its implications for the origin of SARS coronaviruses in humans. *J Gen Virol*91, 1058-62 (2010).
86. Ge, X.Y. et al. Detection of alpha- and beta coronaviruses in rodents from Yunnan, China. *Virol J*14, 98 (2017).
87. Luo, Y. et al. Longitudinal Surveillance of Beta coronaviruses in Fruit Bats in Yunnan Province, China During 2009-2016. *Virol Sin*33, 87-95 (2018).
88. Wang, Y. Preliminary investigation of viruses carried by bats on the southeast coastal area (东南沿海地区蝙蝠携带病毒的初步调查研究). *Master's Thesis* (2017).
89. Wu, Z. et al. Detection of Hantaviruses and Arenaviruzses in three-toed jerboas from the Inner Mongolia Autonomous Region, China. *Emerg Microbes Infect*7, 35 (2018).
90. Menachery, V.D. et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med*21, 1508-13 (2015).
91. Yang, X.L. et al. Characterization of a filovirus (Mengla virus) from Rousettus bats in China. *Nat Microbiol*4, 390-395 (2019).
92. Waruhiu, C. et al. Molecular detection of viruses in Kenyan bats and discovery of novel astroviruses, caliciviruses and rotaviruses. *Virol Sin*32, 101-114 (2017).
93. Ge, X.Y. et al. Fugong virus, a novel hantavirus harbored by the small oriental vole (Eothenomyseleusis) in China. *Virol J*13, 27 (2016).
94. Yang, X.L. et al. Isolation and identification of bat viruses closely related to human, porcine and mink orthoreoviruses. *J Gen Virol*96, 3525-3531 (2015).
95. Hu, B. et al. Detection of diverse novel astroviruses from small mammals in China. *J Gen Virol*95, 24422449 (2014).
96. Yang, X., Zhang, Y., Ge, X., Yuan, J. & Shi, Z. A novel totivirus-like virus isolated from bat guano. *Arch Virol*157, 1093-9 (2012).
97. Zhao, L. et al. Characterization of a Novel Tanay Virus Isolated from Anopheles sinensis Mosquitoes in Yunnan, China. *Front Microbiol*10, 1963 (2019).
98. Xia, H. et al. First Isolation and Characterization of a Group C Banna Virus (BAV) from Anopheles sinensis Mosquitoes in Hubei, China. *Viruses*10, doi: 10.3390/v10100555 (2018).
99. Wang, Y., Xia, H., Zhang, B., Liu, X. & Yuan, Z. Isolation and characterization of a novel mesonivirus from Culex mosquitoes in China. *Virus Res*240, 130-139 (2017).
100. Li, L.L. et al. Detection and characterization of a novel hepacivirus in long-tailed ground squirrels (Spermophilus undulatus) in China. *Arch Virol*164, 2401-2410 (2019).
101. Zhou, Z. et al. Complete genome sequences of two crimean-congohemorrhagic Fever viruses isolated in china. *Genome Announc*1, doi: 10.1128/genomeA.00571-13 (2013).
102. Lau, S.K. et al. Complete genome analysis of three novel picornaviruses from diverse bat species. *J Virol*85, 8819-28 (2011).
103. Woo, P.C. et al. Discovery of seven novel Mammalian and avian coronaviruses in the genus delta coronavirus supports bat coronaviruses as the gene source of alpha coronavirus and beta coronavirus and avian coronaviruses as the gene source of gamma coronavirus and delta coronavirus. *J Virol*86, 39954008 (2012).
104. Zuo, S.Q. et al. A new hantavirus from the stripe-backed shrew (Sorex cylindricauda) in the People's Republic of China. *Virus Res*184, 82-6 (2014).
105. Zuo, S. et al. Detection of Quang Binh virus from mosquitoes in China. *Virus Res*180, 31-8 (2014).
106. He, B. et al. Group A Rotaviruses in Chinese Bats: Genetic Composition, Serology, and Evidence for Batto-Human Transmission and Reassortment. *J Virol*91, doi: 10.1128/JVI.02493-16 (2017).
107. Feng, Y. et al. Isolation and full-length genome analysis of mosquito-borne Manzanilla virus from

Yunnan Province, China. *BMC Res Notes*8, 255 (2015).
108. Xu, L. et al. Novel hantavirus identified in black-bearded tomb bats, China. *Infect Genet Evol*31, 158-60 (2015).
109. He, B. et al. Hepatitis virus in long-fingered bats, Myanmar. *Emerg Infect Dis*19, 638-40 (2013).
110. Sun, H. et al. Prevalence and genetic characterization of Toxoplasma gondii in bats in Myanmar. *Appl Environ Microbiol*79, 3526-8 (2013).
111. Yang, R. Progress in research concerning Yersinia pestis and its significance in military medicine. in *semanticscholar.org* https://www.semanticscholar.org/paper/Progress-in-research-concerning-Yersiniapestis-and-Rui-fu/69c3be6d683dce8e992086d8c92c8119c039260c (2012).
112. He, B. et al. Characterization of a novel G3P[3] rotavirus isolated from a lesser horseshoe bat: a distant relative of feline/canine rotaviruses. *J Virol*87, 12357-66 (2013).
113. He, B. et al. Identification of a novel Orthohepadnavirus in pomona roundleaf bats in China. *Arch Virol*160, 335-7 (2015).
114. Hu, T. et al. Characterization of a novel orthoreovirus isolated from fruit bat, China. *BMC Microbiol*14, 293 (2014).
115. Becker, M.M. et al. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc Natl Acad Sci U S A*105, 19944-9 (2008).
116. LuDe. *Twitter*, https://twitter.com/ding_gang/status/1218547052084441088 (2020).
117. Calisher, C. et al. Statement in support of the scientists, public health professionals, and medical professionals of China combatting COVID-19. *Lancet*395, e42-e43 (2020).
118. Liu, S.L., Saif, L.J., Weiss, S.R. &Su, L. No credible evidence supporting claims of the laboratory engineering of SARS-CoV-2. *Emerg Microbes Infect*9, 505-507 (2020).
119. Collins, F. Genomic Study Points to Natural Origin of COVID-19. *NIH Director's Blog*, https://directorsblog.nih.gov/2020/03/26/genomic-research-points-to-natural-origin-of-covid-19/ (2020).
120. WHO. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-finalreport.pdf (2020).
121. Robinson, C. Journals censor lab origin theory for SARS-CoV-2. (https://www.gmwatch.org/en/news/latest-news/19475-journals-censor-lab-origin-theory-for-sars-cov-2, 2020).
122. Warmbrod, K.L., West, R.M., Connell, N.D. & Gronvall, G.K. In Response: Yan et al Preprint Examinations of the Origin of SARS-CoV-2. *John Hopkins Center for Health Security*, https://www.centerforhealthsecurity.org/our-work/pubs_archive/pubs-pdfs/2020/200921-in-responseyan.pdf (2020).
123. Koyama, T., Lauring, A., Gallo, R. & Reitz, M. Reviews of "Unusual Features of the SARS-CoV-2 Genome Suggesting Sophisticated Laboratory Modification Rather Than Natural Evolution and Delineation of Its Probable Synthetic Route". *Rapid Reviews COVID-19*, https://rapidreviewscovid19.mitpress.mit.edu/pub/78we86rp/release/2 (2020).

Global Journals Guidelines Handbook 2021

www.GlobalJournals.org

Memberships

Fellows/Associates of Computer Science Research Council

FCSRC/ACSRC Memberships

Introduction

FCSRC/ACSRC is the most prestigious membership of Global Journals accredited by Open Association of Research Society, U.S.A (OARS). The credentials of Fellow and Associate designations signify that the researcher has gained the knowledge of the fundamental and high-level concepts, and is a subject matter expert, proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice. The credentials are designated only to the researchers, scientists, and professionals that have been selected by a rigorous process by our Editorial Board and Management Board.

Associates of FCSRC/ACSRC are scientists and researchers from around the world are working on projects/researches that have huge potentials. Members support Global Journals' mission to advance technology for humanity and the profession.

FCSRC

Fellow of Computer Science Research Council

FELLOW OF COMPUTER SCIENCE RESEARCH COUNCIL is the most prestigious membership of Global Journals. It is an award and membership granted to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Fellows are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Fellow Members.

Benefit

To the institution

Get letter of appreciation

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.

Exclusive Network

Get access to a closed network

A FCSRC member gets access to a closed network of Tier 1 researchers and scientists with direct communication channel through our website. Fellows can reach out to other members or researchers directly.They should also be open to reaching out by other.

Career | Credibility | Exclusive | Reputation

Certificate

Certificate, LoR and Laser-Momento

Fellows receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career | Credibility | Exclusive | Reputation

Designation

Get honored title of membership

Fellows can use the honored title of membership. The “FCSRC” is an honored title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., FCSRC or William Walldroff, M.S., FCSRC.

Career | Credibility | Exclusive | Reputation

Recognition on the Platform

Better visibility and citation

All the Fellow members of FCSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation. All fellows get a dedicated page on the website with their biography.

Career | Credibility | Reputation

Future Work

Get discounts on the future publications

Fellows receive discounts on future publications with Global Journals up to 60%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career | Financial

GJ Account

Unlimited forward of Emails

Fellows get secure and fast GJ work emails with unlimited forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career | Credibility | Reputation

Premium Tools

Access to all the premium tools

To take future researches to the zenith, fellows receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

Conferences & Events

Organize seminar/conference

Fellows are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career | Credibility | Financial

Early Invitations

Early invitations to all the symposiums, seminars, conferences

All fellows receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive

Publishing Articles & Books

Earn 60% of sales proceeds

Fellows can publish articles (limited) without any fees. Also, they can earn up to 70% of sales proceeds from the sale of reference/review books/literature/publishing of research paper. The FCSRC member can decide its price and we can help in making the right decision.

Exclusive | Financial

Reviewers

Get a remuneration of 15% of author fees

Fellow members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

Access to Editorial Board

Become a member of the Editorial Board

Fellows may join as a member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. Additionally, Fellows get a chance to nominate other members for Editorial Board.

Career | Credibility | Exclusive | Reputation

And Much More

Get access to scientific museums and observatories across the globe

All members get access to 5 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 10 GB free secure cloud access for storing research files.

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL is the membership of Global Journals awarded to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Associate membership can later be promoted to Fellow Membership. Associates are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Associate Members.

To the institution

Get letter of appreciation

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.

Exclusive Network

Get access to a closed network

A ACSRC member gets access to a closed network of Tier 2 researchers and scientists with direct communication channel through our website. Associates can reach out to other members or researchers directly.They should also be open to reaching out by other.

Career | Credibility | Exclusive | Reputation

Certificate

Certificate, LoR and Laser-Momento

Associates receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career | Credibility | Exclusive | Reputation

Designation

Get honored title of membership

Associates can use the honored title of membership. The “ACSRC” is an honored title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., ACSRC or William Walldroff, M.S., ACSRC.

Career | Credibility | Exclusive | Reputation

Recognition on the Platform

Better visibility and citation

All the Associate members of ACSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation.

Career | Credibility | Reputation

Future Work

Get discounts on the future publications

Associates receive discounts on future publications with Global Journals up to 30%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career | Financial

GJ Account

Unlimited forward of Emails

Associates get secure and fast GJ work emails with 5GB forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career | Credibility | Reputation

Premium Tools

Access to all the premium tools

To take future researches to the zenith, associates receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

Conferences & Events

Organize seminar/conference

Associates are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career | Credibility | Financial

Early Invitations

Early invitations to all the symposiums, seminars, conferences

All associates receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive

Publishing Articles & Books

Earn 30-40% of sales proceeds

Associates can publish articles (limited) without any fees. Also, they can earn up to 30-40% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.

Exclusive | Financial

Reviewers

Get a remuneration of 15% of author fees

Associate members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

And Much More

Get access to scientific museums and observatories across the globe

All members get access to 2 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 5 GB free secure cloud access for storing research files.

Associate	Fellow	Research Group	Basic
$4800 **lifetime designation**	$6800 **lifetime designation**	$12500.00 **organizational**	APC **per article**
Certificate, LoR and Momento **2** discounted publishing/year **Gradation** of Research **10** research contacts/day **1 GB** Cloud Storage **GJ** Community Access	**Certificate,** LoR and Momento **Unlimited** discounted publishing/year **Gradation** of Research **Unlimited** research contacts/day **5 GB** Cloud Storage **Online Presense** Assistance **GJ** Community Access	**Certificates,** LoRs and Momentos **Unlimited** free publishing/year **Gradation** of Research **Unlimited** research contacts/day **Unlimited** Cloud Storage **Online Presense** Assistance **GJ** Community Access	**GJ** Community Access

Preferred Author Guidelines

We accept the manuscript submissions in any standard (generic) format.

We typeset manuscripts using advanced typesetting tools like Adobe In Design, CorelDraw, TeXnicCenter, and TeXStudio. We usually recommend authors submit their research using any standard format they are comfortable with, and let Global Journals do the rest.

Alternatively, you can download our basic template from https://globaljournals.org/Template.zip

Authors should submit their complete paper/article, including text illustrations, graphics, conclusions, artwork, and tables. Authors who are not able to submit manuscript using the form above can email the manuscript department at submit@globaljournals.org or get in touch with chiefeditor@globaljournals.org if they wish to send the abstract before submission.

Before and during Submission

Authors must ensure the information provided during the submission of a paper is authentic. Please go through the following checklist before submitting:

1. Authors must go through the complete author guideline and understand and *agree to Global Journals' ethics and code of conduct,* along with author responsibilities.
2. Authors must accept the privacy policy, terms, and conditions of Global Journals.
3. Ensure corresponding author's email address and postal address are accurate and reachable.
4. Manuscript to be submitted must include keywords, an abstract, a paper title, co-author(s') names and details (email address, name, phone number, and institution), figures and illustrations in vector format including appropriate captions, tables, including titles and footnotes, a conclusion, results, acknowledgments and references.
5. Authors should submit paper in a ZIP archive if any supplementary files are required along with the paper.
6. Proper permissions must be acquired for the use of any copyrighted material.
7. Manuscript submitted *must not have been submitted or published elsewhere* and all authors must be aware of the submission.

Declaration of Conflicts of Interest

It is required for authors to declare all financial, institutional, and personal relationships with other individuals and organizations that could influence (bias) their research.

Policy on Plagiarism

Plagiarism is not acceptable in Global Journals submissions at all.

Plagiarized content will not be considered for publication. We reserve the right to inform authors' institutions about plagiarism detected either before or after publication. If plagiarism is identified, we will follow COPE guidelines:

Authors are solely responsible for all the plagiarism that is found. The author must not fabricate, falsify or plagiarize existing research data. The following, if copied, will be considered plagiarism:

- Words (language)
- Ideas
- Findings
- Writings
- Diagrams
- Graphs
- Illustrations
- Lectures

- Printed material
- Graphic representations
- Computer programs
- Electronic material
- Any other original work

Authorship Policies

Global Journals follows the definition of authorship set up by the Open Association of Research Society, USA. According to its guidelines, authorship criteria must be based on:

1. Substantial contributions to the conception and acquisition of data, analysis, and interpretation of findings.
2. Drafting the paper and revising it critically regarding important academic content.
3. Final approval of the version of the paper to be published.

Changes in Authorship

The corresponding author should mention the name and complete details of all co-authors during submission and in manuscript. We support addition, rearrangement, manipulation, and deletions in authors list till the early view publication of the journal. We expect that corresponding author will notify all co-authors of submission. We follow COPE guidelines for changes in authorship.

Copyright

During submission of the manuscript, the author is confirming an exclusive license agreement with Global Journals which gives Global Journals the authority to reproduce, reuse, and republish authors' research. We also believe in flexible copyright terms where copyright may remain with authors/employers/institutions as well. Contact your editor after acceptance to choose your copyright policy. You may follow this form for copyright transfers.

Appealing Decisions

Unless specified in the notification, the Editorial Board's decision on publication of the paper is final and cannot be appealed before making the major change in the manuscript.

Acknowledgments

Contributors to the research other than authors credited should be mentioned in Acknowledgments. The source of funding for the research can be included. Suppliers of resources may be mentioned along with their addresses.

Declaration of funding sources

Global Journals is in partnership with various universities, laboratories, and other institutions worldwide in the research domain. Authors are requested to disclose their source of funding during every stage of their research, such as making analysis, performing laboratory operations, computing data, and using institutional resources, from writing an article to its submission. This will also help authors to get reimbursements by requesting an open access publication letter from Global Journals and submitting to the respective funding source.

Preparing your Manuscript

Authors can submit papers and articles in an acceptable file format: MS Word (doc, docx), LaTeX (.tex, .zip or .rar including all of your files), Adobe PDF (.pdf), rich text format (.rtf), simple text document (.txt), Open Document Text (.odt), and Apple Pages (.pages). Our professional layout editors will format the entire paper according to our official guidelines. This is one of the highlights of publishing with Global Journals—authors should not be concerned about the formatting of their paper. Global Journals accepts articles and manuscripts in every major language, be it Spanish, Chinese, Japanese, Portuguese, Russian, French, German, Dutch, Italian, Greek, or any other national language, but the title, subtitle, and abstract should be in English. This will facilitate indexing and the pre-peer review process.

The following is the official style and template developed for publication of a research paper. Authors are not required to follow this style during the submission of the paper. It is just for reference purposes.

Manuscript Style Instruction (Optional)

- Microsoft Word Document Setting Instructions.
- Font type of all text should be Swis721 Lt BT.
- Page size: 8.27" x 11'", left margin: 0.65, right margin: 0.65, bottom margin: 0.75.
- Paper title should be in one column of font size 24.
- Author name in font size of 11 in one column.
- Abstract: font size 9 with the word "Abstract" in bold italics.
- Main text: font size 10 with two justified columns.
- Two columns with equal column width of 3.38 and spacing of 0.2.
- First character must be three lines drop-capped.
- The paragraph before spacing of 1 pt and after of 0 pt.
- Line spacing of 1 pt.
- Large images must be in one column.
- The names of first main headings (Heading 1) must be in Roman font, capital letters, and font size of 10.
- The names of second main headings (Heading 2) must not include numbers and must be in italics with a font size of 10.

Structure and Format of Manuscript

The recommended size of an original research paper is under 15,000 words and review papers under 7,000 words. Research articles should be less than 10,000 words. Research papers are usually longer than review papers. Review papers are reports of significant research (typically less than 7,000 words, including tables, figures, and references)

A research paper must include:

a) A title which should be relevant to the theme of the paper.
b) A summary, known as an abstract (less than 150 words), containing the major results and conclusions.
c) Up to 10 keywords that precisely identify the paper's subject, purpose, and focus.
d) An introduction, giving fundamental background objectives.
e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition, sources of information must be given, and numerical methods must be specified by reference.
f) Results which should be presented concisely by well-designed tables and figures.
g) Suitable statistical data should also be given.
h) All data must have been gathered with attention to numerical detail in the planning stage.

Design has been recognized to be essential to experiments for a considerable time, and the editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned unrefereed.

i) Discussion should cover implications and consequences and not just recapitulate the results; conclusions should also be summarized.
j) There should be brief acknowledgments.
k) There ought to be references in the conventional format. Global Journals recommends APA format.

Authors should carefully consider the preparation of papers to ensure that they communicate effectively. Papers are much more likely to be accepted if they are carefully designed and laid out, contain few or no errors, are summarizing, and follow instructions. They will also be published with much fewer delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and suggestions to improve brevity.

Format Structure

It is necessary that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

All manuscripts submitted to Global Journals should include:

Title

The title page must carry an informative title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) where the work was carried out.

Author details

The full postal address of any related author(s) must be specified.

Abstract

The abstract is the foundation of the research paper. It should be clear and concise and must contain the objective of the paper and inferences drawn. It is advised to not include big mathematical equations or complicated jargon.

Many researchers searching for information online will use search engines such as Google, Yahoo or others. By optimizing your paper for search engines, you will amplify the chance of someone finding it. In turn, this will make it more likely to be viewed and cited in further works. Global Journals has compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Keywords

A major lynchpin of research work for the writing of research papers is the keyword search, which one will employ to find both library and internet resources. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining, and indexing.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy: planning of a list of possible keywords and phrases to try.

Choice of the main keywords is the first tool of writing a research paper. Research paper writing is an art. Keyword search should be as strategic as possible.

One should start brainstorming lists of potential keywords before even beginning searching. Think about the most important concepts related to research work. Ask, “What words would a source have to include to be truly valuable in a research paper?” Then consider synonyms for the important words.

It may take the discovery of only one important paper to steer in the right keyword direction because, in most databases, the keywords under which a research paper is abstracted are listed with the paper.

Numerical Methods

Numerical methods used should be transparent and, where appropriate, supported by references.

Abbreviations

Authors must list all the abbreviations used in the paper at the end of the paper or in a separate table before using them.

Formulas and equations

Authors are advised to submit any mathematical equation using either MathJax, KaTeX, or LaTeX, or in a very high-quality image.

Tables, Figures, and Figure Legends

Tables: Tables should be cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g., Table 4, a self-explanatory caption, and be on a separate sheet. Authors must submit tables in an editable format and not as images. References to these tables (if any) must be mentioned accurately.

Figures

Figures are supposed to be submitted as separate files. Always include a citation in the text for each figure using Arabic numbers, e.g., Fig. 4. Artwork must be submitted online in vector electronic form or by emailing it.

Preparation of Eletronic Figures for Publication

Although low-quality images are sufficient for review purposes, print publication requires high-quality images to prevent the final product being blurred or fuzzy. Submit (possibly by e-mail) EPS (line art) or TIFF (halftone/ photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Avoid using pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings). Please give the data for figures in black and white or submit a Color Work Agreement form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution at final image size ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs): >350 dpi; figures containing both halftone and line images: >650 dpi.

Color charges: Authors are advised to pay the full cost for the reproduction of their color artwork. Hence, please note that if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a Color Work Agreement form before your paper can be published. Also, you can email your editor to remove the color fee after acceptance of the paper.

Tips for writing a good quality Computer Science Research Paper

Techniques for writing a good quality computer science research paper:

1. Choosing the topic: In most cases, the topic is selected by the interests of the author, but it can also be suggested by the guides. You can have several topics, and then judge which you are most comfortable with. This may be done by asking several questions of yourself, like "Will I be able to carry out a search in this area? Will I find all necessary resources to accomplish the search? Will I be able to find all information in this field area?" If the answer to this type of question is "yes," then you ought to choose that topic. In most cases, you may have to conduct surveys and visit several places. Also, you might have to do a lot of work to find all the rises and falls of the various data on that subject. Sometimes, detailed information plays a vital role, instead of short information. Evaluators are human: The first thing to remember is that evaluators are also human beings. They are not only meant for rejecting a paper. They are here to evaluate your paper. So present your best aspect.

2. Think like evaluators: If you are in confusion or getting demotivated because your paper may not be accepted by the evaluators, then think, and try to evaluate your paper like an evaluator. Try to understand what an evaluator wants in your research paper, and you will automatically have your answer. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

3. Ask your guides: If you are having any difficulty with your research, then do not hesitate to share your difficulty with your guide (if you have one). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work, then ask your supervisor to help you with an alternative. He or she might also provide you with a list of essential readings.

4. Use of computer is recommended: As you are doing research in the field of computer science then this point is quite obvious. Use right software: Always use good quality software packages. If you are not capable of judging good software, then you can lose the quality of your paper unknowingly. There are various programs available to help you which you can get through the internet.

5. Use the internet for help: An excellent start for your paper is using Google. It is a wondrous search engine, where you can have your doubts resolved. You may also read some answers for the frequent question of how to write your research paper or find a model research paper. You can download books from the internet. If you have all the required books, place importance on reading, selecting, and analyzing the specified information. Then sketch out your research paper. Use big pictures: You may use encyclopedias like Wikipedia to get pictures with the best resolution. At Global Journals, you should strictly follow here.

6. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right? It is a good habit which helps to not lose your continuity. You should always use bookmarks while searching on the internet also, which will make your search easier.

7. Revise what you wrote: When you write anything, always read it, summarize it, and then finalize it.

8. Make every effort: Make every effort to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in the introduction—what is the need for a particular research paper. Polish your work with good writing skills and always give an evaluator what he wants. Make backups: When you are going to do any important thing like making a research paper, you should always have backup copies of it either on your computer or on paper. This protects you from losing any portion of your important data.

9. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several unnecessary diagrams will degrade the quality of your paper by creating a hodgepodge. So always try to include diagrams which were made by you to improve the readability of your paper. Use of direct quotes: When you do research relevant to literature, history, or current affairs, then use of quotes becomes essential, but if the study is relevant to science, use of quotes is not preferable.

10.Use proper verb tense: Use proper verb tenses in your paper. Use past tense to present those events that have happened. Use present tense to indicate events that are going on. Use future tense to indicate events that will happen in the future. Use of wrong tenses will confuse the evaluator. Avoid sentences that are incomplete.

11. Pick a good study spot: Always try to pick a spot for your research which is quiet. Not every spot is good for studying.

12. Know what you know: Always try to know what you know by making objectives, otherwise you will be confused and unable to achieve your target.

13. Use good grammar: Always use good grammar and words that will have a positive impact on the evaluator; use of good vocabulary does not mean using tough words which the evaluator has to find in a dictionary. Do not fragment sentences. Eliminate one-word sentences. Do not ever use a big word when a smaller one would suffice.

Verbs have to be in agreement with their subjects. In a research paper, do not start sentences with conjunctions or finish them with prepositions. When writing formally, it is advisable to never split an infinitive because someone will (wrongly) complain. Avoid clichés like a disease. Always shun irritating alliteration. Use language which is simple and straightforward. Put together a neat summary.

14. Arrangement of information: Each section of the main body should start with an opening sentence, and there should be a changeover at the end of the section. Give only valid and powerful arguments for your topic. You may also maintain your arguments with records.

15. Never start at the last minute: Always allow enough time for research work. Leaving everything to the last minute will degrade your paper and spoil your work.

16. Multitasking in research is not good: Doing several things at the same time is a bad habit in the case of research activity. Research is an area where everything has a particular time slot. Divide your research work into parts, and do a particular part in a particular time slot.

17. Never copy others' work: Never copy others' work and give it your name because if the evaluator has seen it anywhere, you will be in trouble. Take proper rest and food: No matter how many hours you spend on your research activity, if you are not taking care of your health, then all your efforts will have been in vain. For quality research, take proper rest and food.

18. Go to seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

19. Refresh your mind after intervals: Try to give your mind a rest by listening to soft music or sleeping in intervals. This will also improve your memory. Acquire colleagues: Always try to acquire colleagues. No matter how sharp you are, if you acquire colleagues, they can give you ideas which will be helpful to your research.

20. Think technically: Always think technically. If anything happens, search for its reasons, benefits, and demerits. Think and then print: When you go to print your paper, check that tables are not split, headings are not detached from their descriptions, and page sequence is maintained.

21. Adding unnecessary information: Do not add unnecessary information like "I have used MS Excel to draw graphs." Irrelevant and inappropriate material is superfluous. Foreign terminology and phrases are not apropos. One should never take a broad view. Analogy is like feathers on a snake. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Never oversimplify: When adding material to your research paper, never go for oversimplification; this will definitely irritate the evaluator. Be specific. Never use rhythmic redundancies. Contractions shouldn't be used in a research paper. Comparisons are as terrible as clichés. Give up ampersands, abbreviations, and so on. Remove commas that are not necessary. Parenthetical words should be between brackets or commas. Understatement is always the best way to put forward earth-shaking thoughts. Give a detailed literary review.

22. Report concluded results: Use concluded results. From raw data, filter the results, and then conclude your studies based on measurements and observations taken. An appropriate number of decimal places should be used. Parenthetical remarks are prohibited here. Proofread carefully at the final stage. At the end, give an outline to your arguments. Spot perspectives of further study of the subject. Justify your conclusion at the bottom sufficiently, which will probably include examples.

23. Upon conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print for the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects of your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form which is presented in the guidelines using the template.
- Please note the criteria peer reviewers will use for grading the final paper.

Final points:

One purpose of organizing a research paper is to let people interpret your efforts selectively. The journal requires the following sections, submitted in the order listed, with each section starting on a new page:

The introduction: This will be compiled from reference matter and reflect the design processes or outline of basis that directed you to make a study. As you carry out the process of study, the method and process section will be constructed like that. The results segment will show related statistics in nearly sequential order and direct reviewers to similar intellectual paths throughout the data that you gathered to carry out your study.

The discussion section:

This will provide understanding of the data and projections as to the implications of the results. The use of good quality references throughout the paper will give the effort trustworthiness by representing an alertness to prior workings.

Writing a research paper is not an easy job, no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record-keeping are the only means to make straightforward progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear: Adhere to recommended page limits.

Mistakes to avoid:

- Insertion of a title at the foot of a page with subsequent text on the next page.
- Separating a table, chart, or figure—confine each to a single page.
- Submitting a manuscript with pages out of sequence.
- In every section of your document, use standard writing style, including articles ("a" and "the").
- Keep paying attention to the topic of the paper.
- Use paragraphs to split each significant point (excluding the abstract).
- Align the primary line of each section.
- Present your points in sound order.
- Use present tense to report well-accepted matters.
- Use past tense to describe specific results.
- Do not use familiar wording; don't address the reviewer directly. Don't use slang or superlatives.
- Avoid use of extra pictures—include only those figures essential to presenting results.

Title page:

Choose a revealing title. It should be short and include the name(s) and address(es) of all authors. It should not have acronyms or abbreviations or exceed two printed lines.

Abstract: This summary should be two hundred words or less. It should clearly and briefly explain the key findings reported in the manuscript and must have precise statistics. It should not have acronyms or abbreviations. It should be logical in itself. Do not cite references at this point.

An abstract is a brief, distinct paragraph summary of finished work or work in development. In a minute or less, a reviewer can be taught the foundation behind the study, common approaches to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Use comprehensive sentences, and do not sacrifice readability for brevity; you can maintain it succinctly by phrasing sentences so that they provide more than a lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study with the subsequent elements in any summary. Try to limit the initial two items to no more than one line each.

Reason for writing the article—theory, overall issue, purpose.

- Fundamental goal.
- To-the-point depiction of the research.
- Consequences, including definite statistics—if the consequences are quantitative in nature, account for this; results of any numerical analysis should be reported. Significant conclusions or questions that emerge from the research.

Approach:

- Single section and succinct.
- An outline of the job done is always written in past tense.
- Concentrate on shortening results—limit background information to a verdict or two.
- Exact spelling, clarity of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else.

Introduction:

The introduction should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable of comprehending and calculating the purpose of your study without having to refer to other works. The basis for the study should be offered. Give the most important references, but avoid making a comprehensive appraisal of the topic. Describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will give no attention to your results. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here.

The following approach can create a valuable beginning:

- Explain the value (significance) of the study.
- Defend the model—why did you employ this particular system or method? What is its compensation? Remark upon its appropriateness from an abstract point of view as well as pointing out sensible reasons for using it.
- Present a justification. State your particular theory(-ies) or aim(s), and describe the logic that led you to choose them.
- Briefly explain the study's tentative purpose and how it meets the declared objectives.

Approach:

Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done. Sort out your thoughts; manufacture one key point for every section. If you make the four points listed above, you will need at least four paragraphs. Present surrounding information only when it is necessary to support a situation. The reviewer does not desire to read everything you know about a topic. Shape the theory specifically—do not take a broad view.

As always, give awareness to spelling, simplicity, and correctness of sentences and phrases.

Procedures (methods and materials):

This part is supposed to be the easiest to carve if you have good skills. A soundly written procedures segment allows a capable scientist to replicate your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order, but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt to give the least amount of information that would permit another capable scientist to replicate your outcome, but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section.

When a technique is used that has been well-described in another section, mention the specific item describing the way, but draw the basic principle while stating the situation. The purpose is to show all particular resources and broad procedures so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step-by-step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

Materials may be reported in part of a section or else they may be recognized along with your measures.

Methods:

- Report the method and not the particulars of each process that engaged the same methodology.
- Describe the method entirely.
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures.
- Simplify—detail how procedures were completed, not how they were performed on a particular day.
- If well-known procedures were used, account for the procedure by name, possibly with a reference, and that's all.

Approach:

It is embarrassing to use vigorous voice when documenting methods without using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result, when writing up the methods, most authors use third person passive voice.

Use standard style in this and every other part of the paper—avoid familiar lists, and use full sentences.

What to keep away from:

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings—save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part as entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Use statistics and tables, if suitable, to present consequences most efficiently.

You must clearly differentiate material which would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matters should not be submitted at all except if requested by the instructor.

Content:

- Sum up your conclusions in text and demonstrate them, if suitable, with figures and tables.
- In the manuscript, explain each of your consequences, and point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation of an exacting study.
- Explain results of control experiments and give remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or manuscript.

What to stay away from:

- Do not discuss or infer your outcome, report surrounding information, or try to explain anything.
- Do not include raw data or intermediate calculations in a research manuscript.
- Do not present similar data more than once.
- A manuscript should complement any figures or tables, not duplicate information.
- Never confuse figures with tables—there is a difference.

Approach:

As always, use past tense when you submit your results, and put the whole thing in a reasonable order.

Put figures and tables, appropriately numbered, in order at the end of the report.

If you desire, you may place your figures and tables properly within the text of your results section.

Figures and tables:

If you put figures and tables at the end of some details, make certain that they are visibly distinguished from any attached appendix materials, such as raw facts. Whatever the position, each table must be titled, numbered one after the other, and include a heading. All figures and tables must be divided from the text.

Discussion:

The discussion is expected to be the trickiest segment to write. A lot of papers submitted to the journal are discarded based on problems with the discussion. There is no rule for how long an argument should be.

Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implications of the study. The purpose here is to offer an understanding of your results and support all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of results should be fully described.

Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact, you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved the prospect, and let it drop at that. Make a decision as to whether each premise is supported or discarded or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."

Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work.

- You may propose future guidelines, such as how an experiment might be personalized to accomplish a new idea.
- Give details of all of your remarks as much as possible, focusing on mechanisms.
- Make a decision as to whether the tentative design sufficiently addressed the theory and whether or not it was correctly restricted. Try to present substitute explanations if they are sensible alternatives.
- One piece of research will not counter an overall question, so maintain the large picture in mind. Where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

When you refer to information, differentiate data generated by your own studies from other available information. Present work done by specific persons (including you) in past tense.

Describe generally acknowledged facts and main beliefs in present tense.

The Administration Rules

Administration Rules to Be Strictly Followed before Submitting Your Research Paper to Global Journals Inc.

Please read the following rules and regulations carefully before submitting your research paper to Global Journals Inc. to avoid rejection.

Segment draft and final research paper: You have to strictly follow the template of a research paper, failing which your paper may get rejected. You are expected to write each part of the paper wholly on your own. The peer reviewers need to identify your own perspective of the concepts in your own terms. Please do not extract straight from any other source, and do not rephrase someone else's analysis. Do not allow anyone else to proofread your manuscript.

Written material: You may discuss this with your guides and key sources. Do not copy anyone else's paper, even if this is only imitation, otherwise it will be rejected on the grounds of plagiarism, which is illegal. Various methods to avoid plagiarism are strictly applied by us to every paper, and, if found guilty, you may be blacklisted, which could affect your career adversely. To guard yourself and others from possible illegal use, please do not permit anyone to use or even read your paper and file.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

INDEX

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350