ECRONICON OPEN ACCESS

EC NEUROLOGY
Research Article

Spread of SARS-CoV-2 Genomes on Genomic Index Maps of Hierarchy - Compared with B.1.1.7 Lineage on BLAST

Jeffrey Zheng[1,2]*, Yang Zhou[3], Minghan Zhu[3], Mu Qiao[3] and Zhigang Zhang[4]

[1]*Key Laboratory of Quantum Information of Yunnan, China*

[2]*Key Laboratory of Software Engineering of Yunnan, Yunnan University, Kunming, China*

[3]*Yunnan University, Kunming, China*

[4]*School of Life Sciences and Technology, Yunnan University, Kunming, China*

***Corresponding Author:** Jeffrey Zheng, Key Laboratory of Quantum Information of Yunnan and Key Laboratory of Software Engineering of Yunnan, Yunnan University, Kunming, China.

Received: April 20, 2020; **Published:** June 22, 2021

Abstract

COVID-19 patients worldwide are conveniently described by position in-formation to collect samples, and modern GIS maps are useful to show influenced flows and numbers of patients on various regions of a pandemic. From an analysis viewpoint, it is more interesting to organize genomic information into a phylogenic tree with multiple branches and leaves in representations. Clusters of genomes are organized as phylogenic trees to represent intrinsic information of genomes. How-ever, there are structural difficulties in projecting phylogenetic information into 2D distributions as GIS maps naturally.

Considering advanced generating schemes of phylogenetic trees, information entropy provides ultra optimal properties in the minimum computational complexity, superior flexibility, better stability, improved reliability and higher quality on global constructions. This super technology may play a key role in future development of advanced neurology, neuroscience and brain researches.

In this paper, a novel projection is proposed to arrange SARS-CoV-2 genomes by genomic indexes to make a structural organization as 2D GIS maps. For any genome, there is a unique invariant under certain conditions to provide an absolute position on a specific region. In this hierarchical framework, it is possible to use a visual tool to represent any selected region for clustering genomes on refined effects. Applied diversity measure to a given set of genomes, equivalent clusters and complementary visual effects are provided between genomic index maps and phylogenetic trees. Sample genomes of three UK new lineages are aligned by BLAST as a basis on both RNA-dependent RNA polymerase RDRP segments and whole genomes. Selected regions and various projections show spread effects of five thousand SARS-CoV-2 genomes in 72 countries on both RDRP and whole genomes, and six special countries/regions are selected on genomic index maps.

Based on genomic index maps, one SNV of two genomes on B.1.1.7 lineage can be identified from a unit of 10^{-4} probability measure to a unit of 10^{-6} difference for genomic indexes on a special 'G' projection to extract the finest variation.

Further exploration on optimal classification and phylogenetic analysis of genomic index maps and phylogenetic trees on SARS-CoV-2 genomes worldwide are dis-cussed.

***Keywords**: Genomic Index; Visual Maps; Phylogeny; Projection; Information Entropy; Diversity Measure; Global Invariant; Hierarchical Projection; Optimization*

Introduction

The outbreak of SARS-CoV-2 caused COVID-19 to start in Dec. 2019 and is now pandemic. To the date of 26 January 2021, there are more than 100 million con-firmed cases and 2.15 million deaths worldwide. An understanding of the prevalence and contagiousness of the disease and of whether the strategies used to contain it to date have been successful is important for understanding future containment strategies.

One excellent strategy for containment of SARS-CoV-2 is to collect sample genomes globally into the GISAID genetic database [1] for infected viruses. Based on this effective activity, Next strain provides Phylogenetic tree [2] to organize sample datasets from different places to categorize them as clusters under the maximal likelihood relationship to view intrinsic variations among SARS-CoV-2 genomes. Based on phylogenetic information, a dynamic simulation system provides flexible illustrations on selected branches [4] to support medical doctors, virological experts, biomedical specialists and psychologic doctors for detailed treatments on COVID-19 patients.

Advanced researches in phylogenetic analysis

The NCBI developed Basic Local Alignment Search Tool BLAST [3] in 1990s to provide powerful software tools for generating phylogenetic trees under a list of optimal inference conditions [6-28]: maximum likelihood [6,7,23], probability [8], Bayers [9], stochastic search [10], unalignable sequences [11], best fit model [12], tradition and Bayers [13], reconstruction [14], multiple alternative phylogenies [15], phylogenetic diversity measures [16], entropy approach [17], Shannon entropy and mutual information [18], viral phylogenomics [19], phylogenetic tree building [20], IQ-TREE [21], neural network and deep learning classifier [28].

Viral phylogenetic using an alignment-free method [19] provide optimal length of k-mer on N genomes of a phylogenetic tree to have computational complexity applying cumulative relative entropy and Shannon entropy on O(N), significantly faster than minimum likelihood or Bayers alignment on $O(N^2)$.

Useful technologies to build phylogenetic trees are viewed in [20]. Special problems in data collection of the world for SARS-CoV-2 are discussed. The key difficulties of phylogenetic analysis of SARS-CoV-2 are described in [22]. On wider researches on SARS-CoV-2 of phylogenetic analysis, a list of researches are carried out: phylogenetic supertree [24], informative subtype marker ISM [25], CG dinucleotide [26], CpG deficiency [27], classification and geographical analysis, light-weight classifier [28] and phylogenetic structure, S protein and stability [29].

In [25], informative subtype marker ISM applied entropy analysis and ISM ex-traction to simulate Next strain through GISAID clades of SARS-CoV-2 genomes in details. In this scheme, key positions of relevant open reader frames ORFs associated with probability measures on time variations to describe viral evolutionary information from historic datasets.

For N number of unique sequences, L width of alignment, a size of alphabet, three executable complexity as follows.

Scheme	Complexity	Description
Minimum likelihood	$O(N^2 L a)$	
Bayers alignment		Common optimal schemes
Fasttree	$O(N^{1:25} \log(N) L a)$	Next strain's phylogenetic trees
Information entropy	O(N L a)	Fastest optimal scheme

Limitations of phylogenetic representations

Further arrangement may not be a direct approach. Regular zoom operators in GIS could be simulated along deeper or upper movement along branching nodes in a phylogenetic tree. Since phylogenetic trees correspond to neither 1D nor 2D structures, it is difficult to rearrange various subtrees [24] as visual objects. Using BLAST or MEGA packages, expensive computational complexity may be required to process N genomes to handle a set of phylogenetic trees.

In general, effective projections for a subset of phylogenetic trees provide a natural projection, and other forms of visual representations could not be directly supported.

Phylogenetic trees in Nextstrain

The phylogenetic tree of Nextstrain is based on the maximal likelihood relationship to organize genomic datasets as hierarchical clusters under differential information. After a sample genome of SARS-CoV-2 compared with root node and following branch nodes recursively, it is possible to push it into the most likelihood node that contains the most similar genomes to be a target group. Since a genome contains a long sequence, there are multiple relationships among various clusters in the phylogenetic tree shown in figure 1. Using GIS maps, it is useful to see various genomes distributed worldwide.

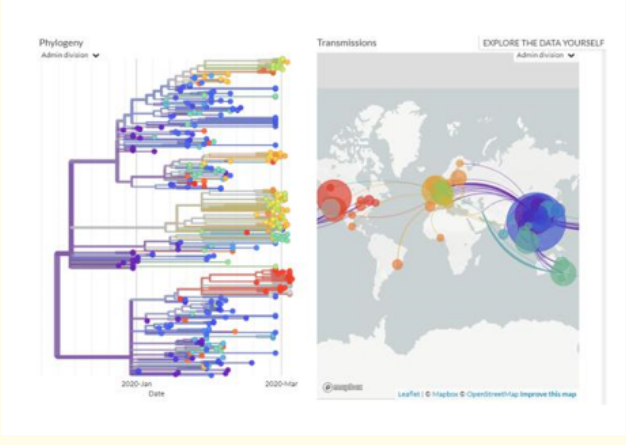


Figure 1: *The phylogenetic tree of real cases over global on Next strain.*

Difficulties in phylogenetic analysis of SARS-CoV-2 data

A list of difficulties are discussed in [22], phylogenetic analysis of SARS-CoV-2 data is challenging due to numeric difficulties and the rugged likelihood surface.

Larger taxa on a low number of distinct site patterns have large topologic variability.

Signal is weak, it is difficult for standard phylogenetic significant tests. Bayersian tree interferences use a plausible tree set for computing summary statistics on trees.

Since old phylogenetic trees were generated from original genomes, there may not contain invariant structures to support new variations and mutations emergent from larger numbers of genomes everywhere.

In general, huge number of new genomes collected over the world makes ex-tensive structural difficulties to use phylogenetic trees constructed for update and extensions.

Combination, matrix and thermodynamics

In modern mathematics and physics, there are many theoretical constructions to handle invariant and variation problems for entropy issues [31-51] such as combinatorial mathematics, combinatorial theory, combinatorics, multiple variable complex theory, statistical physics, thermodynamics, thermostatistics, statistical mechanics.

Neurology, neuroscience and brain science

The foundations of advanced brain sciences are strongly related to modern scientific research developments on modern neurology and neurosciences [53-63].

Hierarchy of visual cells

Visual receptors and retinal interaction opened new ways to explore neural signals on different visual disorders and frog eyes in details. Three levels of hierarchy are developed on simple, complex and hypercomplex cells of multiple neural network functionals [64-67].

Imaging and reconstruction on cellular level

Following cellular dynamics, powerful microscope tools are developed from gastrulation to early organogenesis at the cellular level. Microscopy and computational methods are explored. Multiple visual attentions are developed [68-75]. Memory process on cell levels were developed [76-80].

AI problems, sensor technologies and computational theory

Various hard problems in AI were discussed [81-85]. Multiple sensors and neuron-dynamics were investigated [86-92]. Computational theory of cognition were proposed [93-97].

Integrated information theory and natural intelligence

Integrated information theory is an advanced theoretical construction developed to share five basic properties: existing, composition, informative, integration, exclusion [98-103].

Further developments were focused on multiple entropy schemes [104-107] and natural intelligence NI approaches with biological neural network BNN and minimum free energy MFE [108].

Variant construction

In this direction, vectors, matrices and invariant measurements are described relevant to wider applications [109-112] on variant construction [113-116].

The genomic index provides unique identification for each genome to be an in-variant under given conditions. Based on these types of global quantitative characteristics, it is convenient for large numbers of genomes to be located in a certain geometric region to be collected as clusters.

Different entropy quantities were discussed in separate papers: Visualization of SARS-CoV-2 Genomes on Genomic Index Maps [117], Visualizations of Topological Entropy on SARS-CoV-2 Genomes in Multiple Regions [118], Visual Variations between Pairs of SARS-CoV-2 Genomes on Integrated Density Matrix [119], Visualizations of Combinatorial Entropy Index on Whole SARS-CoV-2 Genomes [120].

Considering this is an extremely important research direction, it is necessary to handle this topic from a foundation level to provide additional information to explore hidden structures among this type of multiple levels of hierarchical constructions from a visual representation viewpoint.

Materials and Methods

Input on four meta symbols

For genomes, each element of input sequences is composed of four meta symbols states SS: {∅; A;C; G; T}.

The first order of combinations

From a combinatorial viewpoint, the first order of combinations from the four symbols is composed of sixteen states as a lattice of hierarchy, as shown in figure 2.

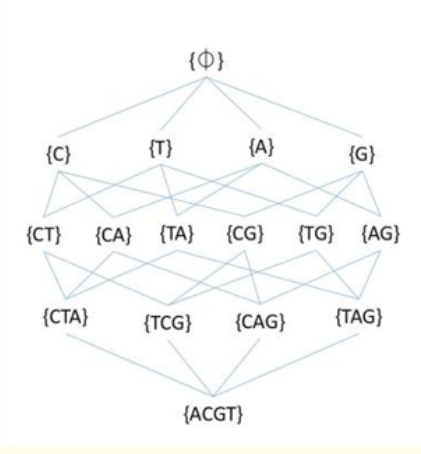


Figure 2: *Sixteen combinations of four meta-symbols in a hierarchy of a lattice.*

The sixteen states SS = {0/; A;C; G; T; AC; AG; AT; CG; CT; GT; CGT; AGT; ACT; ACG; ACGT} can be mapped into the sixteen numbers SI = {0; 1; 2; … ; 15} to rep-resent a 1D linear structure with 16 distinct positions. For a segment of a genome with m elements, there are four meta measures: $\{m_A; m_C; m_G; m_T\} = \{m_1; m_2; m_3; m_4\}$ and sixteen combinatorial measures: $\{m_i\}$; $0 \le i \le 15$ to correspond a meta measuring vector with four elements and a combinatorial measuring vector with sixteen elements, respectively.

Multiple probability measures

When a genome contains m elements, the numbers of four Meta symbols can be counted. Let m_s; $s \in SS$ be a number of symbols s and p_s be a probability measure. We have the following equations for multiple probability measures:

$$m = m_A + m_C + m_G + m_T$$
$$p_s = \frac{m_s}{m}, s \in SS$$
$$1 = p_A + p_C + p_G + p_T$$

Under multiple probability conditions, there are sixteen distinct probability measures $\{p_i\}^{15}_{i=0}$; $0 \le p_i \le 1$, $I \in SI$ respectively.

Two workflows from input to output

Two workflows (1) and (2) can be identified by the type of output:

(1) Vector of Genome → Probability → Sixteen Probability Vectors
→ Entropy → Sixteen Indexes
→ Selection → An Index
(2) {Pair of Indexes} → Mapping → A Genomic Index Map

Genomic index projection and genomic index map

Three workflows are described in three parts as input, output and process.

In Step (1), one index of 16 Combinatorial Entropies can be generated.

$$\begin{aligned} Input: &\ N \text{ elements in a genome, } N = m \times M \\ Output: &\ 1 \text{ (an index} \in [0,\ \log_2(m+1)]) \\ Process: &\ N \overset{\text{Segment}}{\rightarrow} m \times M \overset{\text{Meta-Measure}}{\rightarrow} 4 \times M \overset{\text{Combination}}{\rightarrow} 16 \times (m+1) \\ &\overset{\text{Entropy}}{\rightarrow} 16 \overset{\text{Selection}}{\rightarrow} 1 \\ CF: &\ 16 \text{ (Total number of selections)} \end{aligned}$$

In Step (2) a genomic index map can be generated from multiple sets of sixteen indexes.

$$\begin{aligned} Input: &\ \forall(x,y) \in \text{Multiple sets of sixteen indexes}, x, y \in [0,\ \log_2(m+1)] \\ Output: &\ \text{An Index Map on } [0,\ \log_2(m+1)] \times [0,\ \log_2(m+1)] \text{ Region} \\ Process: &\ \forall(x,y) \overset{Projection}{\rightarrow} [0,\ \log_2(m+1)] \times [0,\ \log_2(m+1)] \\ CF: &\ 256 \text{ (Total number of selections)} \end{aligned}$$

Combinatorial entropy measurement

Let a vector Z with (m + 1) elements, Z = (Z_0; Z_1; ... ; Z_j; ; Z_m); 0 ≤ Z_j ≤ M and M=$\sum^{m}_{j=0} Z_j$. Under this condition, let Pj=Z_j/M be the j-th probability measurement, and a relevant information entropy eZ can be determined and restricted in a [0; $\log_2$(m + 1)] region:

$$eZ = -\sum_{j=0}^{m} P_j \log_2(P_j),\ eZ \in [0,\ \log_2(m+1)] \quad (1)$$

$$1 = \sum_{j=0}^{m} P_j,\ 0 \le j \le m \quad (2)$$

For sixteen combinations of the first order, sixteen entropy measurements of eZ correspond to {eZ_i},0 ≤ i ≤ 15.

2D combinatorial entropies

Extending this construction to higher orders, the second order of combinations are composed of 2D 16 16 pairs of states or a 2D square with 256 positions.

Under this condition for a segment with m elements on a genome Z with N = m x M elements, sixteen entropies {eZ_i},0 ≤ i ≤ 15, ZE_i ∈ [0; $\log_2$(m + 1)] are determined:

$$(eZ_{i,j}) = \begin{pmatrix} eZ_{0,0} & \cdots & eZ_{i,0} & \cdots & eZ_{15,0} \\ \cdots & & \cdots & & \cdots \\ eZ_{0,j} & \cdots & eZ_{i,j} & \cdots & eZ_{15,j} \\ \cdots & & \cdots & & \cdots \\ eZ_{0,15} & \cdots & eZ_{i,15} & \cdots & eZ_{15,15} \end{pmatrix} \quad i, j \in SI$$

A pair of indexes corresponds to: $eZ_{i,j}$ = (eZ_i, eZ_j); 0 ≤ i,j ≤ 15. There are a total of 256 pairs of 2D positions determined by the genome Z in the square on the [0; $\log_2$(m + 1)] [0; $\log_2$(m + 1)] region.

Multiple genomes

For multiple genomes $\{Z^t\}$; $1 \le t \le T$ on maximal T members of each (i; j) projection, a total number of T positions can be collected on 2D square of $\forall(eZ_i^t ; eZ_j^t)$; $1 \le t \le T$. This provides a special distribution for whole genomes of T members on (i; j) projection based on combinatorial entropy measurements.

$$EZ_{i,j} = \sum_{t=1}^{T} eZ_{i,j}^t = \sum_{t=1}^{T} (eZ_i^t, eZ_j^t), 0 \le i, j \le 15$$

Each $EZ_{i; j}$ represents an index map corresponding to a $[0; \log_2(m + 1)]$ $[0; \log_2(m + 1)]$ region.

Genomic index maps

Different from a genome, it has a relative position in a phylogenic tree on the maxi-mal likelihood relationship. A genomic index is an absolute invariant to correspond a genome into a quantitative measurement under information entropy based on variant construction. Visual representations of multiple projections are illustrated.

Diversity measures between BLAST phylogenetics and genomic index maps

Differences between phylogenetic trees on given levels and whole genomes on genomic index maps can be systematically measured by diversity measure for N genomes. In [16], a list of phylogenetic diversity measures are discussed on phylogenetic trees. Using information theory, this type of diversity measures is restricted in $[0; \log_2 N]$.

For a selected segment R such as an ORF area, if all R segments of N genomes are transferred as a genomic index map with at most M distinguished positions 1 M N (or a certain level of a relevant phylogenetic tree with M branches), let $E_R(N)$ be a diversity entropy of genomic index maps on R areas, then the diversity measure is defined as

$$E_R(N) = \log_2(M) \qquad (3)$$

Difference and error margin

For two genomes Z_1; Z_2, let s; $s \in SS$ be a projection on s direction, and a difference $\Delta (eZ_s(Z_1); eZ_s(Z_2))$ of two genomic indexes is

$$\Delta(eZ_s(Z_1), eZ_s(Z_2)) = \max(eZ_s(Z_1), eZ_s(Z_2)) - \min(eZ_s(Z_1), eZ_s(Z_2)) \qquad (4)$$

Let Δe be a given error margin, e.g. Δe = 0:001. If $\Delta (eZ_s(Z_1); eZ_s(Z_2)) > De$ is true, then two genomes can be distinguished in a genomic index map. Otherwise, two genomes cannot be separated in a cluster on the genomic index map.

If all R segments of N genomes contain in the same content as the same genomic index, then there is $E_R(N) = \log_2(1) = 0$ to be the mini-malist diversity measure for the system configuration. However, if all N genomic indexes of R segments can be distinguished without any equal genomic index, then there is $E_R(N) = \log_2(N)$ to be the maximalist diversity measure for the R segments.

Equivalent condition between phylogenetic trees and genomic index maps

From diversity measures for N genomes, there is a natural correspondence via equivalent diversity measures between clusters of a certain level of a phylogenetic subtree and an enlarged region of a genomic index map in general.

Datasets

From a collection of more than 30K genomes from the GISAID genetic database before July 2020, more than 5K genomes were selected without any uncertain element of 'N' in whole sequences. Approximately 25K genomes contain at least one 'N'. There are 72 countries involved that contain more than one genome.

Based on COG-UK's report of SARS-CoV-2 spike mutations [30], three groups of datasets on UK new variations {B.1.1.7, B.1.177, B.1.258} are selected, and each group contains 10 genomes.

Selecting a total of 5336 genomes, both RNA-dependent RNA polymerase RDRP and whole genomes are involved for further analysis in corresponding genomic index maps.

RDRP and S protein alignments, and phylogenetic trees

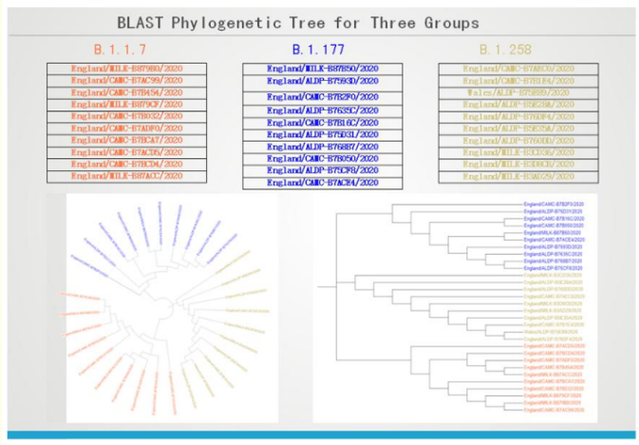
Three selected datasets of 30 variation genomes and their processed results from BLAST are shown in figure 3a-3c. Two phylogenetic trees for three groups are constructed by BLAST using neighbor-joining (NJ) for DNA on maximal likelihood (ML) shown in figure 3a. Three groups are separated as three branches. Variations on RDRP are listed in figure 3b and variations on S protein are listed in figure 3c.

Variations of BLAST on RDRP

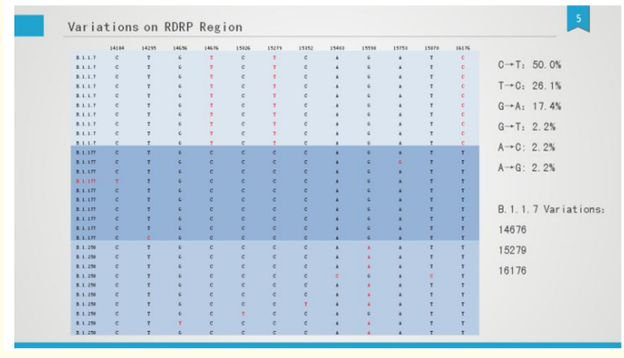
From figure 3b, all variations of B.1.1.7 RDRP sequences have the same contents with {C ↔ T} exchanged positions, two positions with {A → C, C → T} are in B.1.177 sequences, and there are multiple variations on {C → A, G → A, C ↔ T,⋯} in B.1.258 respectively.

Variations of BLAST on S protein

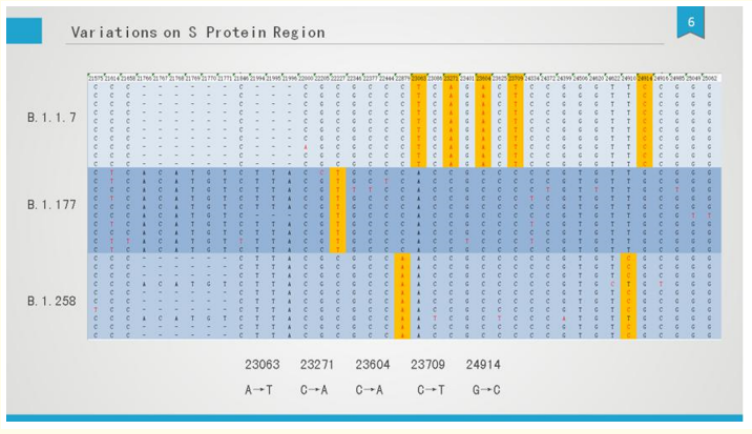
From figure 3c, B.1.1.7 S protein sequences have at least five significant variations on {A → T, C → A, C → T, G → C}, B.1.177 sequences have multiple variations on {C → T, G → T, G → C} and B.1.258 sequences contain multiple variations on {C → A, C ↔ T, G → A} respectively.



a



b



c

Figure 3a-3c: *Three groups of variation genomes on BLAST (a) Selected genomes and two Phylogenetic trees (b) Variations in RDRP (c) Variations in S Protein.*

Visual tool-plotly

Plotly is a visual tool [52] of open-source visualization libraries for R, Python and JavaScript. In this project, we use this visual tool to illustrate hierarchical distributions for multiple genomes on selected regions of $EZ_{i;\,j}$ maps.

Clustering on genomic index maps

Since all genomic indexes are associated with absolute invariants, this makes it possible to apply 1D or 2D distributions to represent complicated clusters for multiple genomes in hierarchical structures.

Two distinct schemes are shown in figure 4 for both the phylogenetic tree of Nextstrain and a global genomic index map on five thousand genomes in 72 countries. Different colors are applied to distinguish relevant countries. Various clusters of genomes are clearly visualized by distinct color points for relevant countries on the genomic index map. Refined maps are shown in the next section.

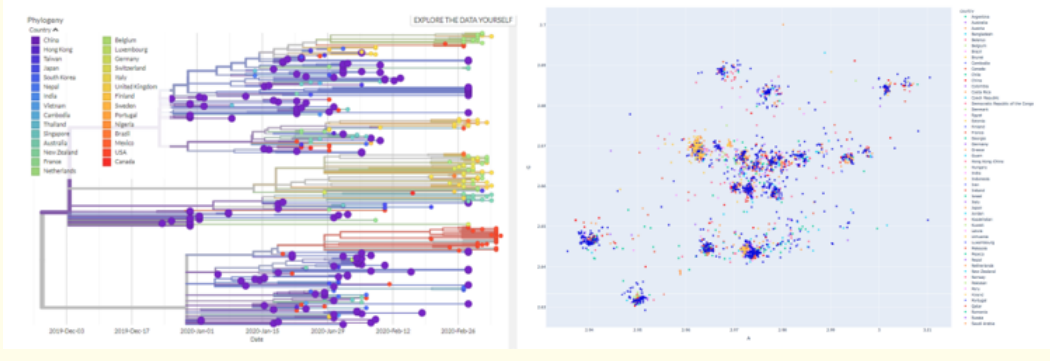


Figure 4: *The phylogenetic tree of real cases over global on Next strain and global genomic index map.*

Results

Relevant results are included in two separated files: 5306-RDRP16-A-G.html (for 30 + 5306 RDRP segments) and 5306-Whole16-A-G.html (30 + 5306 Whole genomes) that can be visualized by an HTML browser in the newest version for Plotly libraries.

For RDRP sequences of five thousand genomes, a global genomic index map and various projection maps for three variations and six regions: {Australia, Chile, China, Taiwan, UK, USA} were selected to show relevant projections of results in figure 5a-5c and enlarged parts of selected regions are shown in figure 6a-6h. Two special projections are shown in figure 7a and 7b to illustrate six selected regions and three variation on RDRP.

a

b

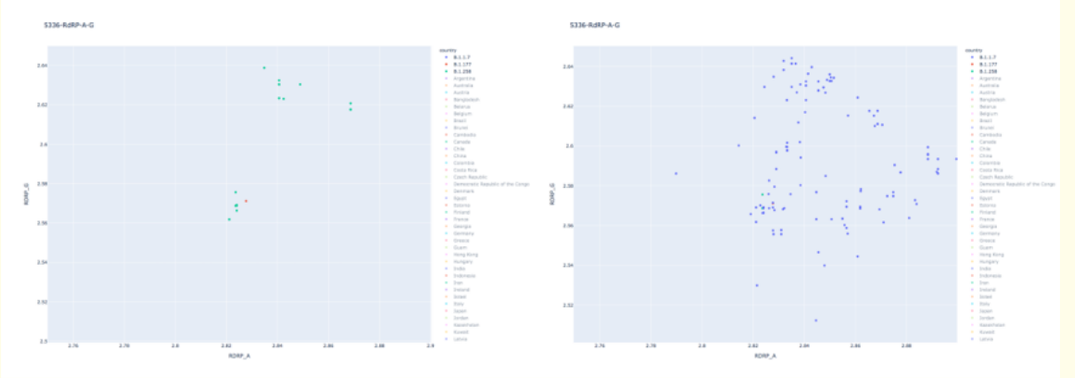
c

Figure 5a-5c: *Five thousands of RDRP genomes on genomic index maps (a) Global (b) Six selected regions: Australia + Chile + China + Taiwan + UK + USA (c) An enlarged region selected from (b).*

a ***b***

c ***d***

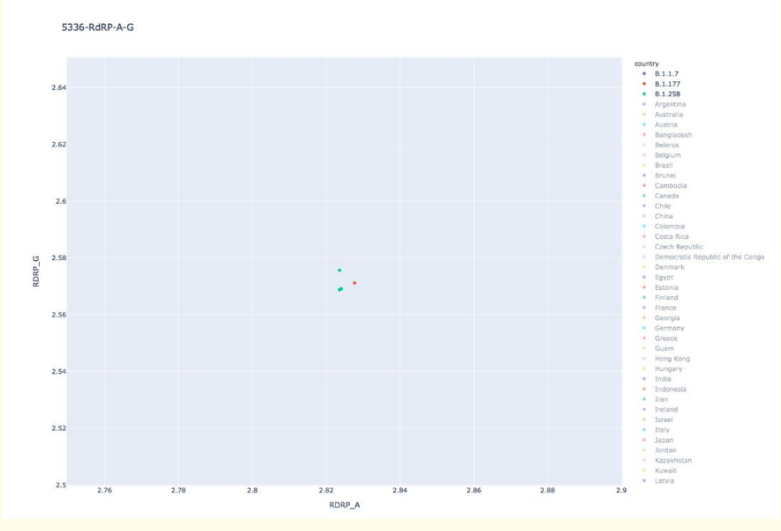
e ***f***



g ***h***

Figure 6a-6h: *FAn enlarged region of RDRP on genomic index maps with three groups of variations. (a) Six regions: Australia + Chile + China + Taiwan + UK + USA (b) Three groups: B.1.1.7 + B.1.177+B.1.258 (c) Australia (d) Chile (e) China (f) Taiwan (g) UK (h) USA.*

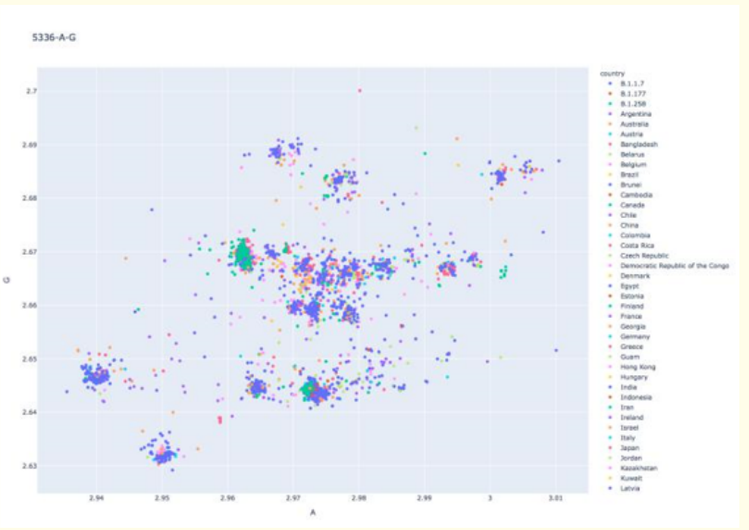
a



b

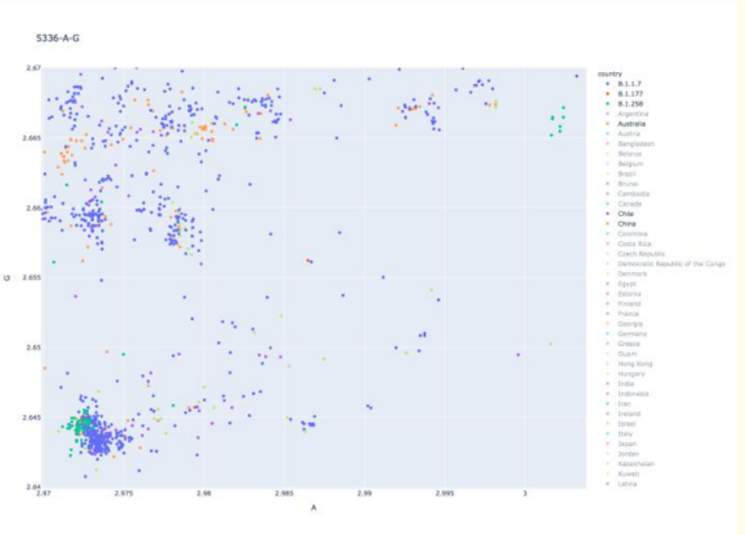
Figure 7a and 7b: *Enlarged Region of RDRP genomes on genomic index maps (a) Six selected regions: Australia + Chile + China + Taiwan + UK + USA (b) Three variations.*

For whole sequences of five thousand genomes, a global genomic index map and various projection maps for three variations and six regions: {Australia, Chile, China, Taiwan, UK, USA} were selected to show relevant projections of results in figure 8a-8c and enlarged parts of selected regions are shown in figure 9a-9h. Two special projections are shown in figure 10a and 10b to illustrate six selected regions and three variation on whole genomes.



a

b

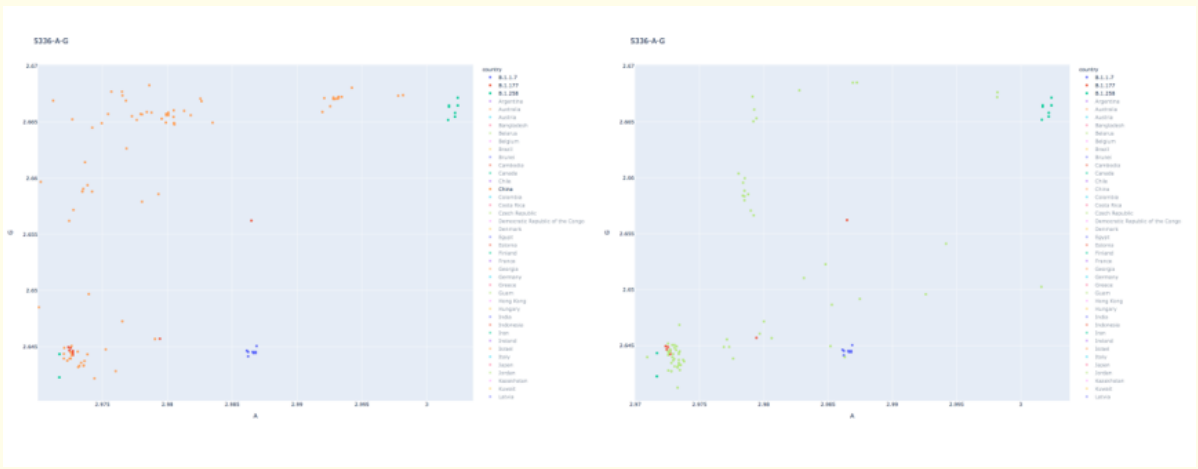


c

Figure 8a-8c *Five thousands of whole genomes on genomic index maps (a) Global (b) Six selected regions: Australia + Chile + China + Taiwan + UK + USA (c) An enlarged region selected from (b).*

a ***b***

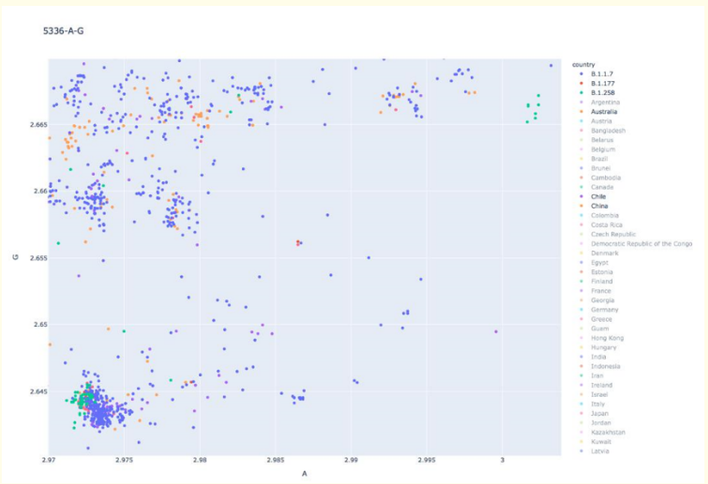
c ***d***



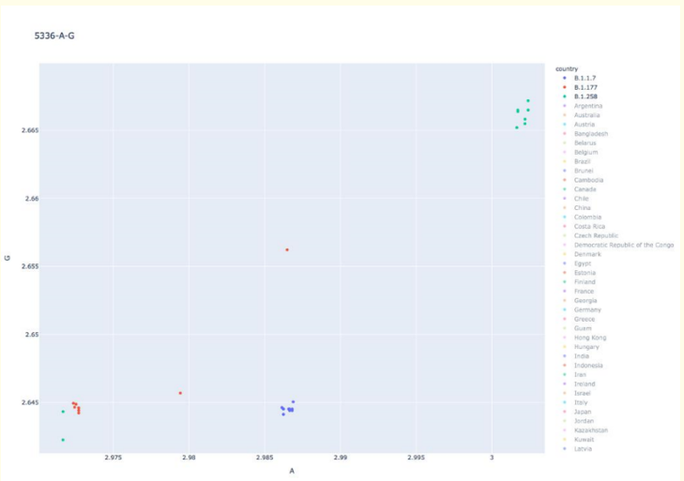
e ***f***

g ***h***

Figure 9a-9h: *An enlarged region of whole genomes on genomic index maps with three groups of varia-tions (a) Six regions: Australia + Chile + China + Taiwan + UK + USA (b) Three groups: B.1.1.7 + B.1.177+B.1.258 (c) Australia (d) Chile (e) China (f) Taiwan (g) UK (h) USA.*

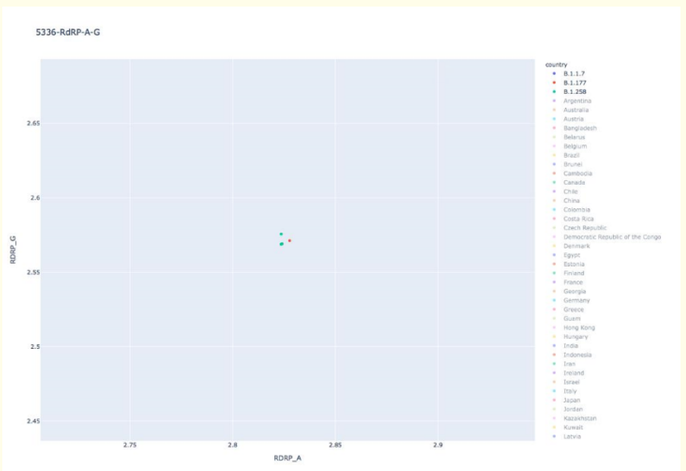


a

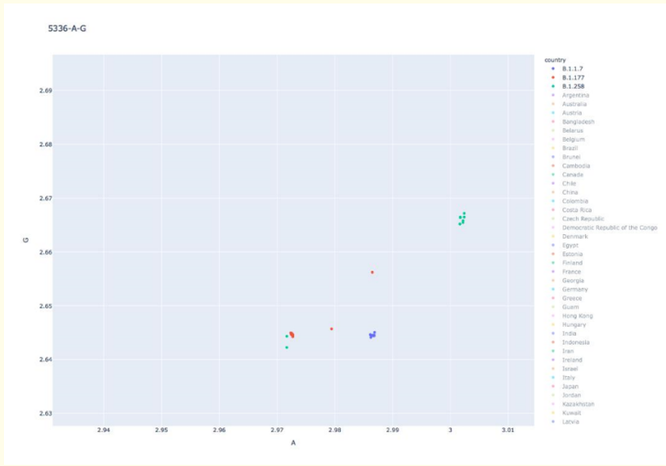


b

Figure 10a and 10b: *Enlarged Region of whole genomes on genomic index maps (a) Six selected regions: Australia + Chile + China + Taiwan + UK + USA (b) Three variations.*

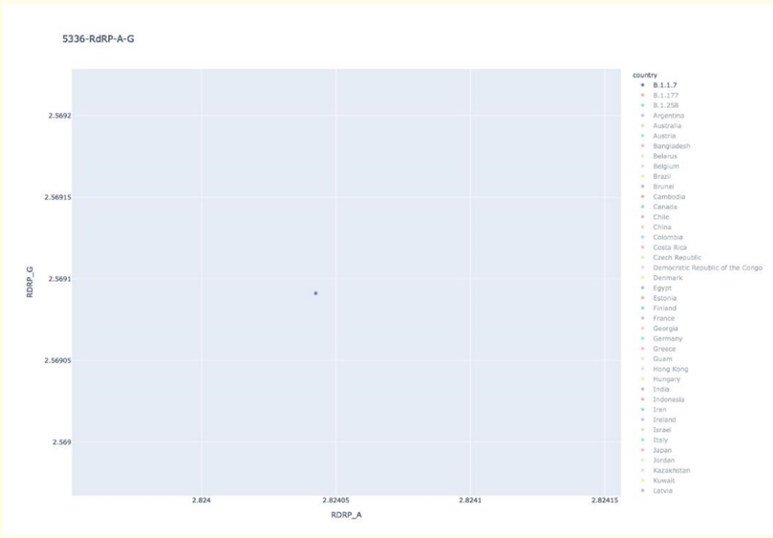


a

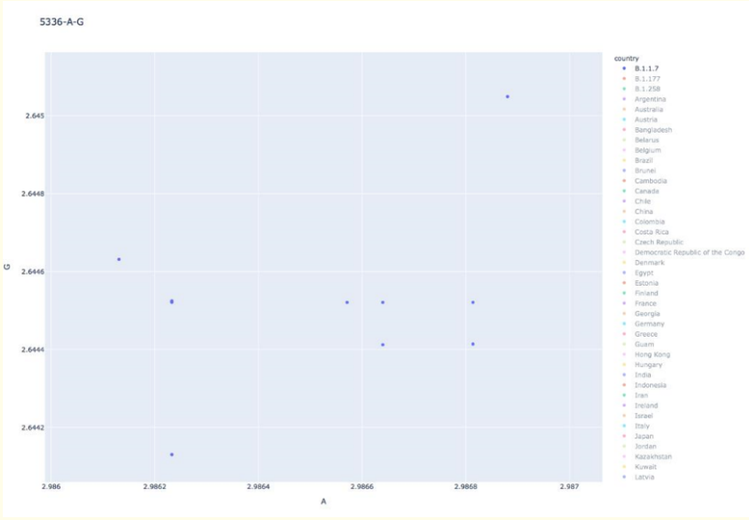


b

Figure 11a and 11b: *Three variations of RDRP and whole genomes on genomic index maps (a) RDRP (b) Whole genomes.*



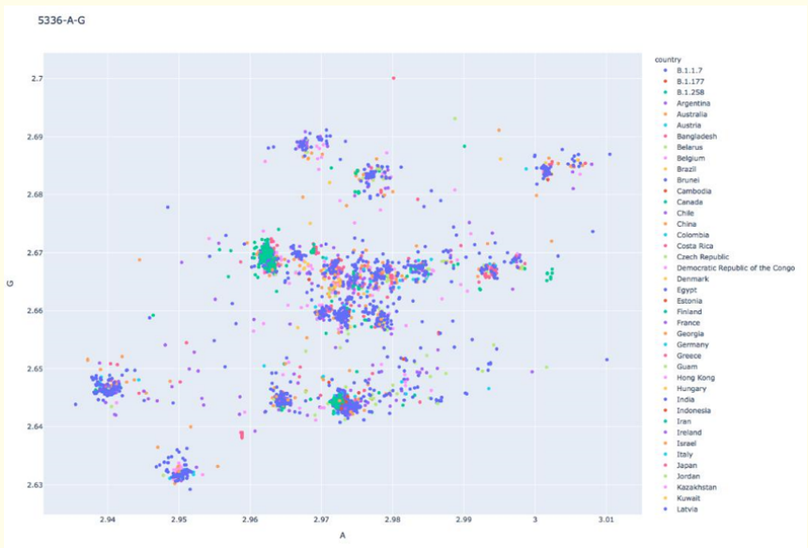
a



b

Figure 12a and 12b: *Three variations of RDRP and whole genomes on 100 times of enlarged genomic index maps (a) RDRP (b) Whole genomes.*

a



b

Figure 13a and 13b: *Three variations and six selected regions of RDRP and whole genomes on genomic index maps (a) RDRP (b) Whole genomes.*

Projections of three variations from RDRP to whole genomes are illustrated in figure 11a and 11b, 100 times of enlarged projections of ten B.1.1.7 genomes from RDRP to whole genomes are illustrated in figure 12a and 12b, and projections of six regions and three variations from RDRP to whole genomes are illustrated in figure 13a and 13b.

Discussion

Since there is an autoscale function in the Plotly package, visual regions for selected datasets may not be a fixed one with slight differences for each selected region.

Projections for RDRP

In figure 5a-5c, three genomic index maps are represented for all genomes of three variations, 72 countries and selected six regions: Australia, Chile, China, Taiwan, the UK and the USA.

Initial maps for RDRP

In figure 5a, all genomic indexes of three variations and 72 countries are restricted to a region of A ∈ [2.70,2.95] × G ∈ [2.40,2.75], (ΔA = 0.25, ΔG = 0.35) with an error margin Δe=0.01 and multiple clusters could be identified by visual clustering technologies. The central point of this map is located on (A = 2.825, G = 2.675).

In figure 5b, all genomic indexes of three variations and six regions: Australia, Chile, China, Taiwan, the UK and the USA are selected in a region of [2.70, 2.95] × [2.43, 2.72], (ΔA = 0.25, ΔG = 0.29) and a selected region of [2.75, 2.90] × [2.50, 2.65], (ΔA = 0.25, ΔG = 0.15) with an error margin Δe = 0.01.

In figure 5c, the selected region of figure 5b has expanded as a full frame restricted in the region of [2.75, 2.90] × [2.50, 2.65], (ΔA = 0.25, ΔG = 0.15). Three variations and six selected regions will be projected as further selection in an enlarged map.

Selected areas for RDRP

Eight maps are shown in figure 6a-6h in the region of [2.75, 2.90] × [2.50, 2.65], (ΔA = 0.25, ΔG = 0.15).

In figure 6a, RDRP of three variations and six selected regions are illustrated.

In figure 6b, RDRP of three variations are illustrated around the center part of the frame. There are three color points distinguished one in red (B.1.177), three (two connected) in green (B.1.258), and one in green is a common point to be shared with red (B.1.177) and blue (B.1.1.7) overlapped at the same position: (A = 2:824042; G = 2:569091). Three color points in a triangle can be restricted in a rectangle on [2:8236; 2:8277] [2:5687; 2:5756] of (ΔA = 0:0041; ΔG = 0:0069) differences with an error margin Δe = 0:0001.

In figure 6c, RDRP collected from Australia and three variations are illustrated. One pink point is covered on the same position of B.1.1.7, other pink points are located on the east and north-east direction far away from the three variations.

In figure 6d, RDRP collected from Chile and three variations are illustrated. One purple point is covered on the same position of B.1.1.7, other purple points are located on from the north, north-east, east to south-east directions far away from the three variations.

In figure 6e, RDRP collected from China and three variations are illustrated. One yellow point is covered on the same position of B.1.1.7, other yellow points are located on from north, north-east, east, south-east to south directions far away from the three variations.

In figure 6f, RDRP collected from Taiwan and three variations are illustrated. One green point is covered on the same position of B.1.1.7, other green points are located on from the north, north-east, east, south-east to south directions far away from the three variations.

In figure 6g, RDRP collected from the UK and three variations are illustrated. One light-blue point is covered on the same position of B.1.1.7, other light-blue points are located on from the north to north-east directions far away from the three variations.

In figure 6h, RDRP collected from the USA and three variations are illustrated. One blue point is covered on the same position of B.1.1.7, other blue points are located on from north-west, north, north-east, east, south-east to south directions far away from the three variations.

It is interesting to notice that at least one genome in each region has covered RDRP of B.1.1.7 on A-G projections.

Enlarged maps for RDRP

In figure 7a and 7b, two enlarged genomic index maps are represented for six regions and three variations on RDRP. Figure 7a is an enlarged map of figure 6a to show refined genomic indexes for the six regions selected. Figure 7b is an enlarged map of figure 6b to show refined genomic indexes for three variations. There is a clear triangle shape in the middle area of the map.

Projections for whole genomes

In figure 8a-8h, eight genomic index maps are represented for all whole genomes of three variations, 72 countries and selected six regions: Australia, Chile, China, Taiwan, the UK and the USA.

Initial maps for whole genomes

In figure 8a, all genomic indexes of three variations and 72 countries are restricted to a region of A ∈ [2.93, 3.02] × G ∈ [2.62, 2.70], (ΔA=0.09, ΔG=0.08) and multiple clusters could be identified by visual clustering technologies. The center of this map is located on (A = 2:97; G = 2:66).

In figure 8b, all genomic indexes of three variations and six regions: Australia, Chile, China, Taiwan, the UK and the USA are selected in a region of [2:93; 3:02] x [2:62; 2:70]; (ΔA = 0:09; ΔG = 0:08) and a selected region of [2:97; 3:01] x [2:64; 2:67], (ΔA = 0:04; DG = 0:03).

In figure 8c, the selected region of figure 8b has expanded as a full frame restricted in the region of [2:97; 3:01] x [2:64; 2:67]; (ΔA = 0:04; ΔG = 0:03). Three variations and six selected regions will be projected as further selection in an enlarged map.

Selected areas for whole genomes

Eight maps are shown in figure 6a-6h in the region of [2:97; 3:01] x [2:64; 2:67]; (ΔA = 0:04; ΔG = 0:03).

In figure 9a, whole genomes of three variations and six selected regions are illustrated. At least, three clusters in green (B.1.258) and blue (B.1.1.7) are located on the right-top and middle-bottom positions as two edge parts. A larger cluster with multiple color points is located on the left-bottom of the map.

In figure 9b, whole genomes of three variations are illustrated on the bottom and north-east parts of the map. There are three types of color points distinguished. One cluster in red (B.1.177) is located on left-bottom corner and a single one is located on the center far away from 1/2 map in a region of [2:9723; 2:9865] x [2:6442; 2:6563], (ΔA = 0:0142; ΔG = 0:0121). Eight clusters in green (B.1.258) are located on left-bottom and right-top of the frame in diagonal directions in a region of [2:9716; 3:0024] x [2:6422; 2:6672], (ΔA = 0:0308; ΔG = 0:0250). And one cluster in blue (B.1.177) is located on middle-bottom part of the map in a region of [2:9861; 2:9869] x [2:6440; 2:6451], (ΔA = 0:0008; ΔG = 0:0006) with an error margin Δe = 0:0001. Three variations are restricted in three areas with 18 ~ 40 times respectively. For the most points compared with RDRP maps, each genomic index can be identified with less overlaps.

In figure 9c, whole genomes collected from Australia and three variations are illustrated. One pink point is located between two clusters of B.1.177, other pink points are located on the right-top direction far away from the three variations.

In figure 9d, whole genomes collected from Chile and three variations are illustrated. Multiple purple points are located between clusters of B.1.177 and B.1.1.7, other purple points are located on from the north-west, north, north-east to south-west directions far away from the three variations.

In figure 9e, whole genomes collected from China and three variations are illustrated. Multiple yellow points are located between clusters of B.1.177 and B.1.1.7, other yellow points are located on from the north-west, north, north-east to south-west directions far away from the three variations.

In figure 9f, whole genomes collected from Taiwan and three variations are illustrated. Multiple green points are closed to clusters of B.1.258, B.1.177 and B.1.1.7, other green points are located on from the north-west, north, north-east to south-west directions far away from the three variations.

In figure 9g whole genomes collected from the UK and three variations are illustrated. Multiple light-blue points are closed to clusters of B.1.177 and B.1.258, other light-blue points are separated from the north-west, north to south-west directions far away from the three variations.

In figure 9h, whole genomes collected from the USA and three variations are illustrated. Main cluster of blue points are located on B.1.177 and a few points are closed to B.1.1.7, other blue points are located mainly from the north-west, north, north-east, south-east, south to south-west directions far away from the three variations.

Enlarged maps for whole genomes

In figure 10a and 10b, two enlarged genomic index maps are represented for six regions and three variations on whole genomes. Figure 10a is an enlarged map of figure 10a to show refined genomic indexes for the six regions selected. Figure 10b is an enlarged map of figure 10b to show refined genomic indexes for three variations. It is interesting to see B.1.1.7 located as an edge cluster in the middle bottom of the map.

Projections from RDRP to whole genomes

Pair of genomic index maps for three variations from RDRP to whole genomes are compared in figure 11a and 11b. In figure 11a clusters of 30 variations are expanded from a triangle shape in a smaller area of RDRP to at least seven clusters of whole genomes from east, south-east and south directions as brushes. For one point of B.1.1.7 on RDRP, a unique blue cluster of whole genomes was developed on the edge part of south-east direction. Ten genomes of B.1.1.7 could be separated under enlarged genomic index maps.

Ten genomes of B.1.1.7

Pair of genomic index maps for ten B.1.1.7 genomes from RDRP to whole genomes are compared in figure 12a and 12b. In figure 12a a single blue cluster of 10 genomes on (A = 2:824042; G = 2:569091) position is still as a point, even both vertical and horizontal axises have been magnified more than 100 times. Nine clusters of B.1.1.7 can be clearly observed under enlarged genomic index map in figure 12b. Only the second cluster on the west direction is composed of two separable points on (A = 2:986233; G = 2:644524) and (A = 2:986233; G = 2:644521) with (ΔA = 0; ΔG = 0:000003) differences with an error margin Δe = 0:000001 to show at least one G variation between the two whole genomes. Further 1000 times of enlarged operation can effectively separate two points on the vertical direction.

Verification on two SNVs of two whole genomes for B.1.1.7 lineage

Based on the 5336-Whole16-A-G.html package, it is convenient to identify the two genomes {England/CAMC-B7B454/2020, England/MILK-B87ACC/2020} from enlarged genomic index maps on two visual screens. The two genomes are aligned by BLAST to extract the finest variations as follows.

For the two whole genomes, only two SNV sites can be identified different at 18252 and 25437 positions. Two SNVs of the England/CAMC-B7B454/2020 genome contain two 'T' symbols, but two SNVs of the England/MILK-B87ACC/2020 genome change 'T' to 'C' and 'G' symbols respectively. The pair of differences on bi-pairs of genomic indexes is (ΔT = 0:000234814; ΔG = 0:000003557).

From this pair of differences, the statement in previous section has been verified. There is merely one 'G' partial variation to be a SNV at the 25437 site from T → G shown in figure 12b.

Since a SARS-CoV-2 genome has 30K nucleotides, a unit of probability measure on its nucleotides is $\sim O(10^{-4})$. In the above 'G' SNV projection, the unit of genomic indexes is DG = 0:000003557 $\sim O(10^{-6})$ significantly enlarged visual maps at least 100 times than original region. Selected an error margin properly, different topologic configurations can be illustrated similar to separate distinct numbers of branches on given levels of a phylogenetic tree.

5336 genomes

Pair of genomic index maps for 5306 genomes and three variations from RDRP to whole genomes are compared in figure 13a and 13b. In figure 13a, compact clusters of 5306 genomes and three variations on RDRP are developed mainly in the middle areas and two larger clusters in north-east and west directions. Compared with figure 11a, it is feasible to identify the three UK variation locations on the map as a reference.

In figure 13b, expanded clusters of 5306 genomes and three variations on whole genomes are distributed in the middle areas and at least four larger clusters are distributed in north-east and south-west directions. Compared with figure 11b, it is feasible to identify the three UK variation locations on the map as a reference, especially for the B.1.1.7 cluster.

Differences between RDRP and whole genomes in genomic index maps

The corresponding relationships of three variations are transformed from RDRP in figure 11a to whole genomes in figure 11b to illustrate characteristic distributions with significantly visual diffusion.

For genomic index maps of whole genomes in figure 11b, B.1.1.7 retains one cluster, both B.1.177 and B.1.258 separated as three clusters with larger distances more than 0:01 ~ 0:1 differences among clusters on genomic index maps.

Since the 5306 genomes were collected before July 2020 from GISAID over the world, no B.1.1.7 variations were identified on this dataset.

From listed comparisons on genomic index maps, larger clusters have significant differences in the six selected regions shown in figure 6a-6h and figure 9a-9h.

Different from RDRP maps at least one genome has covered B.1.1.7 position. In relation to whole genomes, there are only two regions (Taiwan, USA) contained a few genomes located nearby the B.1.1.7 cluster shown in figure 9f and 9h. Other four regions of selected whole genomes were far away from the B.1.1.7 cluster.

Larger clusters of whole genomes

For all 5336 genomes on RDRP, multiple clusters may have higher compacted degrees. Hundreds of distinguished color points can be identified in figure 13a. This density indicates at least 30 genomes with the same RDRP content may be mapped in one position.

For all 5336 genomes on whole genomes, different types of distributions were shown in figure 13b. More than 20 ~ 50 larger clusters can be identified with multiple color points connected as distinguished areas, and many separated single points on the map. Thousands of distinguished color points could be visualized in figure 13b as larger connected areas. From a statistical viewpoint, each cluster could be distributed as Gaussian normal distributions with central symmetry. This type of clusters could collect huge number of genomes especially in central areas to be generated as multiple normal distributions of the statistical probability for larger number of whole genomes.

Optimal properties of BLAST results and genomic index maps

Significant differences between RDRP and whole genomes on genomic index maps can be systematically compared by diversity measures for N genomes. Using diversity measures, this type of diversity measures is restricted in $[0; \log_2(N)]$.

Diversity measures between RDRP segments and whole genomes

Using BLAST operations, multiple RDRP segments are processed to make alignments one by one on selected N genomes.

Let $E_{RDRP}(N)$ be a diversity measure of genomic index maps on RDRP, and $E_{WG}(N)$ be a diversity measure of genomic index map on whole genomes.

If all RDRP segments of N genomes contain in the same content being the same genomic index, then there is $E_{RDRP}(N) = \log_2(1) = 0$ to provide the minimalist diversity measure for the system configuration. However, if all RDRP segments of N genomes can be distinguished without any equal genomic index, then there is $E_{RDRP}(N) = E_{WG}(N) = \log_2(N)$ to provide the maximalist diversity measure for the system configuration.

In figure 12a and 12b, N = 10 only a single cluster of ten B.1.1.7 RDRP segments can be identified in figure 12a, and so the diversity measure of figure 12a is $E_{RDRP}(10) = 0$. However, nine clusters of ten B.1.1.7 genomes are separated with an error margin Δe = 0:00001 on (A; P) genomic index map in figure 12b and the diversity measure of figure 12b is $E_{WG}(10) = \log_2(9)$. If further enlargement has performed and an error margin Δe 0:000001, then ten clusters can be distinguished and $E_{WG}(10) = \log_2(10)$.

Under those conditions, both the minimal and maximal borders of diversity measures can be obtained.

If no BLAST operations were performed to align RDRP segments, then the diversity measure satisfies $E_{RDRP}(N) > 0$. It is extremely hard for anyone to obtain a better result if at least two distinguished genomes are selected from different countries over the world.

In general, N genomes collected from different places, a diversity measure on $0 < E_{RDRP}(N)\ E_{WG}(N) < \log_2(N)$ will be observed.

Due to this structural restriction, traditional BLAST operations provide a necessary condition for genomic index in system optimization. Under BLAST supports, genomic index maps provide an optimal scheme in genomic analysis to visualize multiple genomes in one genomic index map.

Without BLAST operations, genomic index maps for multiple genomes cannot have the minimal configuration of the diversity measure systematically at all.

Optimal solution for multiple ORFs

Twenty nine ORFs are identified from SARS-CoV-2 genomes, in a natural condition, each ORF may bring some random variations. The local alignment can be effectively performed on one selected ORF. It is difficult to make alignment same time more than one ORFs in general.

It is necessary for multiple ORFs to make multiple alignments of relevant ORF first, aligned ORF segments can be processed in further calculation.

If each aligned ORF segment has transferred into a genomic index, multiple aligned ORF recombination will provide the minimal diversity measures smaller than directly calculated from whole unalignable genomes.

From an optimal viewpoint, neither RDRP nor whole genomes provides an optimal solution to explore complex-inner structures of whole SARS-CoV-2 genomes. A better solution is to apply multiple ORFs of alignments separately to create an optimal solution of the diversity measure for future explorations.

Phylogenetic trees and genomic index maps

Using diversity measures, it is convenient for both phylogenetic trees and genomic index maps to be compared consistently. This measurable mechanism is confirmed on the equivalent diversity measures between viral genomes under certain levels of a phylogenetic tree and various enlarged regions of genomic index maps.

In principle on any genomic index map, the enlarging operations can be repeatedly applied to selected regions to recursively detailed regions via a series of proper error margins $e \in \{1, 0.1, 0.01, \cdots, 0.0\cdots01\}$ from a rough gap to the finest margin respectively.

In the most conditions, if two genomic indexes are different, then two positions can be visually separated when a larger fold magnification has been applied and proper error margin selected.

In application levels, the diversity measure provides conveniently classified effects for medical doctors and researchers to treat COVID-19 patients with similar genomic indexes as one group of genomes.

Conclusion

Using combinatorial entropy as 2D genomic index maps, there are 256 projections to support multiple genomes in representations. Various computational measurements are described to cover from local to global statistics properties. Richness of both phylogenetic trees and genomic index maps can be measured on diversity measures with equivalent effects.

Applying thirty genomes of UK new variations, and five thousand genomes of SARS-CoV-2 on 72 countries and special selections on six countries based on Plotly libraries, a list of genomic index maps selected for both RDRP segments and whole genomes are shown in significant different distributions on each country to illustrate complicated contagiousness patterns among various regions.

From the ten genomes of B.1.1.7 lineage, it is feasible to distinguish one SNV from genomic index maps, and different magnifications on selected areas to provide better effects of visualization on selected samples for the finest analysis under the minimum optimal condition. Diversity measures provide numeric quantities as clade information to be compared with both RDRP and whole genomes on the ten whole genomes of B.1.1.7 samples consistently.

It is a challenging task to generate optimal phylogenetic trees of SARS-CoV-2 genomes accurate with stability to support huge number of update genomes over the global, to make twenty nine ORFs such as {S protein, M, N, E} gradually in optimal conditions to simulate GISAID clades and Nextstrain phylogenetic trees in higher stability under huge updates of mutations and variations of SARS-CoV-2 genomes worldwide.

Using genomic index maps, further refined classifications and categories of genomes could be visually and numerically explored, and this powerful optimal-measure tool would be useful in refined medical treatments for COVID-19 patients worldwide in near future.

From further requirements from neurology, neurosciences and brain research directions, the new construction provides the first evidence to support integrated in-formation theory IIT and natural intelligence NI based on genomic index maps GIM as global invariants in hierarchy. It is an interesting topic to explore relevant correspondences and restrictions in proper complex neurons conditions. Further investigations are required.

Conflict of Interest

No conflict of interest has been claimed.

Acknowledgements

The authors would like to thank NCBI, GISAID, CNGBdb, and Nextstrain for providing invaluable information on the newest dataset collections of SARS-CoV-2 and other coronavirus genomes to support this project working smoothly. The NSFC (62041213) provided financial support for the project.

This work was supported by the NSFC (62041213), the Key Project on Electric Information and Next Generation IT Technology of Yunnan (2018ZI002).

5336 genomes and genomic indexes on (A,G)

Two information files in two formats contain detailed information for each selected genome: (Genome Name, Location, Time, Type, Clade, ..., Genomic Indexes on whole genome and RDRP).

55306 genome information: 5306-Genome-Information.xlsx, 5306-Genome-Information.tsv

30 UK genome information: 30-UK-Genome-Information.xlsx, 30-UK-Genome-Information.tsv

Two executable packages

Two interactive visual packages: RDRP and whole genomes.

5336-RDRP-Genomes-16-A-G.html (30 + 5306 RDRP segments in 72 countries/regions).

5336-Whole-Genomes-16-A-G.html (30 + 5306 Whole genomes in 72 countries/regions).

Appendix

Follow the below links

30-UK-Genome-Information

5306-Genome-Information

30-UK-Genome-Information

5306-Genome-Information

5336-RDRP-Genomes-16-A-G

5336-Whole-Genomes-16-A-G

Bibliography

1. GISAID: Open access to influenza virus data.
2. Nextstrain Real time tracking of pathogen evolution.
3. Basic Local Alignment Search Tool BLAST.
4. HP Yao., *et al*. "Patient-derived mutations impact pathogenicity of SARS-CoV-2" (2020).
5. Li C., *et al*. "Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species". *Infection, Genetics and Evolution* 82 (2020): 104285.
6. Thorne JL., *et al*. "An evolutionary model for maximum likelihood alignment of DNA sequences". *Molecular Biology and Evolution* 33 (1991): 114124.
7. Lewis PO. "A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data". *Molecular Biology and Evolution* 15 (1998): 277283.
8. Mitchison GJ. "A probabilistic treatment of phylogeny and sequence alignment". *Journal of Molecular Evolution* 49 (1999): 1122.
9. Holmes I and Bruno WJ. "Evolutionary HMMs: a Bayesian approach to multiple alignment". *Bioinformatics* 17 (2001): 803820.
10. Salter LA and Pearl DK. "Stochastic search strategy for estimation of maximum likelihood phylogenetic trees". *Systematic Biology* 50 (2001): 717.

11. Lee MSY. "Unalignable sequences and molecular evolution". *Trends in Ecology and Evolution* 16 (2001): 681-685.

12. Posada D and Crandall KA. "Selecting the best-fit model of nucleotide substitution". *Systematic Biology* 50 (2001): 580601 (2001).

13. Holder M and Lewis P. "Phylogeny estimation: traditional and Bayesian approaches". *Nature Reviews Genetics* 4 (2003): 275284.

14. Delsuc F., *et al*. "Phylogenomics and the reconstruction of the tree of life". *Nature Reviews Genetics* 6 (2005): 361375.

15. Nye TMW. "Trees of Trees: An Approach to Comparing Multiple Alternative Phylogenies". *Systematic Biology* (2008): 785794.

16. Chao A., *et al*. "Phylogenetic diversity measures based on Hill numbers. Philosophical Transactions of the Royal Society of London". *Series B, Biological Sciences* 365.1558 (2010): 3599-3609.

17. Batista MVA., *et al*. "An entropy-based approach for the identification of phylogenetically informative genomic regions of Papillomavirus". *Infection, Genetics and Evolution* 11.8 (2011): 2026-2033.

18. Arellano-Valle RB., *et al*. "Shannon entropy and mutual in-formation for multivariate skew-elliptical distributions". *Scandinavian Journal of Statistics* 40 (2013): 4262.

19. Zhang Q., *et al*. "Viral Phylogenomics Using an Alignment-Free Method: A Three-Step Approach to Determine Optimal Length of k-mer". *Scientific Reports* 7 (2017): 40712.

20. Kapli P., *et al*. "Phylogenetic tree building in the genomic age". *Nature Reviews Genetics* 21 (2020): 428444.

21. Minh BQ., *et al*. "IQ-TREE 2:New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era". *Molecular Biology and Evolution* 37 (2020): 15301534.

22. Morel B., *et al*. "Phylogenetic analysis of SARS-CoV-2 data is difficult". *Molecular Biology and Evolution* (2020).

23. Shen XX., *et al*. "An investigation of irreproducibility in maximum like-lihood phylogenetic inference". *Nature Communications* 11 (2020): 6096.

24. Li T., *et al*. "Phylogenetic supertree reveals detailed evolution of SARS-CoV-2". *Scientific Reports* 10 (2020): 22366.

25. Zhao Z., *et al*. "Genetic grouping of SARS-CoV-2 coronavirus sequences using informative subtype markers for pandemic spread visualization". *PLOS Computational Biology* 16.9 (2020): e1008269.

26. Wang Y., *et al*. "Human SARS-CoV-2 has evolved to reduce CG dinucleotide in its open reading frames (2021).

27. Xia X. "Extreme genomic CpG deficiency in SARS-CoV-2 and evasion of host antiviral defense". *Molecular Biology and Evolution* (2020).

28. Acera Mateos P., *et al*. "PACIFIC: a lightweight deep-learning classifier of SARS-CoV-2 and co-infecting RNA viruses". *Scientific Reports* 11.1 (2021): 3209.

29. Turakhia Y., *et al*. "Stability of SARS-CoV-2 phylogenies". *Plos Genetics* 11 (2020): e1009175.

30. COG-UK. "COG-UK Update on SARS-CoV-2 Spike Mutations of Special Interest Re-port 1 (2020).

31. PJ Cameron. "Combinatorics: Topics, Techniques, Algorithms, Cambridge University Press (1994).

32. JR Chen. "Combinatorial Mathematics, Harbin Institute of Technology Press (2012).

33. HW Gould. "Some Generalizations of Vandermonde's Convolution". *The American Mathematical Monthly* 63.2 (1956): 84-91.

34. HW Gould. "Combinatorial identities". Morganton (1972).

35. M Hall. "Combinatorial Theory". 2nd edition, Blaisdell (1986).

36. LK Hua. "Loo-Keng Hua Selected Papers". Springer (1982).

37. LK Hua. "Selected Work of Hua Loo-Keng on Popular Sciences". Shanghai Education Press (1984).

38. DE Knuth. "The Art of Computer Programming". 3rd edition, Addison-Wesley 1 (1998).

39. DE Knuth. "The Art of Computer Programming, A: Combinatorial Algorithms, Part 1". Addison-Wesley 4 (2011).

40. F. Morgan. "Geometric Measure Theory, 4th edition". Elsevier (2009).

41. G Polya., *et al*. "Notes on Introductory Combinatorics". Birkhauser (1983).

42. RP Stanley. "Enumerative Combinatorics, 2nd edition". Cambridge University Press 1 (1997).

43. GZ Tu. "Combinatorial Enumeration Methods and Applications". Science Press (1981).

44. LZ Xu., *et al*. "Combinatorial Mathematics of Computation". Shanghai Science and Technology Press (1983).

45. L D Landau and E M Lifshitz. "Statistical Physics, 3rd edition". Part 1, Pergamon Press (1986).

46. W Greiner., *et al*. "Thermodynamics and Statistical Mechanics". Springer-Verlag (1995).

47. R P Freynman. "Statistical Mechanics". Benjamin Reading Mass (1972).

48. J Beatte. "I Oppenheim. Thermodynamics". Elsevier Scientific (1979).

49. S K Ma. "Statistical Mechanics". World Scientific (1985).

50. D Chandler. "Introduction to Modern Statistical Mechanics, 1st edition". Oxford University Press Inc (1987).

51. K Huang. "Statistical Mechanics, 2nd edition". John Wiley and Sons (1987).

52. Plotly: The front-end for ML and data science models.

53. History of Neuroscience.

54. Schrdinger Erwin. "What Is Life?: With Mind and Matter and Autobiographical Sketches. Cambridge University Press (1992).

55. Neuroscience Vs Philosophy: Taking Aim at Free Will: Nature News, n.d.

56. McGinn Colin. "The Mysterious Flame: Conscious Minds in a Material World". Basic Books (2000).

57. Lichtman JW and W Denk. "The Big and the Small: Challenges of Imaging the Brain s Circuits". *Science* 334.6056 (2011): 618623.

58. Harnish RM. "Minds, brains, computers: An historical introduction to the foundations of cognitive science". *Blackwell Publishers* (2002).

59. Li A., *et al*. "Micro-optical sectioning tomography to obtain a high-resolution atlas of the mouse brain". *Science* 330.6009 (2010): 1404.

60. Buxhoeveden D P and Casanova MF. "The minicolumn hypothesis in neuroscience". *Brain* 125.5 (2002): 935-951.

61. DeFelipe J. "From the Connectome to the Synaptome: An Epic Love Story". *Science* 330.6008 (2010): 1198.

62. Miller Greg. "Blue Brain Founder Responds to Critics, Clarifies His Goals". *Science* 334.6057 (2011): 748 749.

63. Owen MM. "Freud in the Scanner" (2017).

64. Sporns O. "Cerebral cartography and connectomics". *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1668 (2015): 20140173.

65. Kragel PA., *et al*. "Representation, Pattern Information, and Brain Signatures: From Neurons to Neuroimaging". *Neuron* 99.2 (2018): 257-273.

66. Ashourvan A., *et al*. "Multi-scale detection of hierarchical community architecture in structural and functional brain networks". *PLoS One* 14.5 (2019): e0215520.

67. Kastanenka KV., *et al*. "A roadmap to integrate astrocytes into Systems Neuroscience". *Glia* 68.1 (2020): 5-26.

68. Maunsell JHR. "Neuronal Mechanisms of Visual Attention". *Annual Review of Vision Science* 1 (2015): 373-391.

69. Jensen O., *et al*. "Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing". *Trends in Neurosciences* 37.7 (2014): 357-369.

70. Jia X., *et al*. "High-density extracellular probes reveal dendritic backpropagation and facilitate neuron classification". *Journal of Neurophysiology* 121.5 (2019): 1831-1847.

71. Huang L., *et al*. "Measuring the interrelations among multiple paradigms of visual attention: an individual differences approach". *Journal of Experimental Psychology. Human Perception and Performance* 38.2 (2012): 414-428.

72. Matthews J., *et al*. "Conscious access in the near absence of attention: critical extensions on the dual-task paradigm". *Philosophical Transactions of the Royal Society, B: Biological Sciences* 373.1755 (2018): 20170352.

73. Fiebelkorn IC and Kastner S. "A Rhythmic Theory of Attention". *Trends in Cognitive Sciences* 23.2 (2019): 87-101.

74. Cha O and Blake R. "Evidence for neural rhythms embedded within binocular rivalry". *Proceedings of the National Academy of Sciences of the United States of America* 116.30 (2019): 14811-14812.

75. Solms M. "The hard problem of consciousness and the free energy principle". *Frontiers in Psychology* 9 (2019): 2714.

76. Eichenbaum H. "The cognitive neuroscience of memory: an introduction". OUP USA (2012).

77. McGaugh and James L. "Memory–a Century of Consolidation". *Science* 287.5451 (2000): 248251.

78. Horton JC and DL Adams. "The Cortical Column: a Structure Without a Function". *Philosophical Transactions of the Royal Society B: Biological Sciences* 360.1456 (2005): 837.

79. Eliceiri Kevin W., *et al*. "Biological Imaging Software Tools". *Nature Methods* 9.7 (2012): 697710.

80. Edelman GM and G Tononi. "A Universe of Consciousness: How Matter Becomes Imagination". Basic books (2001).

81. Farhadi A. "There is no I in AI". *AI and Soc* (2021).

82. Schmid MC and Maier A. "To see or not to see - thalamo-cortical net-works during blindsight and perceptual suppression". *Progress in Neurobiology* 126 (2015): 36-48.

83. Sanchez-Lopez, J., *et al*. "Neural Correlates of Visuospatial Attention to Unseen Stimuli in Hemianopic Pa-tients. A Steady-State Visual Evoked Potential Study". *Frontiers in Psychology* 10 (2019): 198.

84. Musk E. "An integrated brain-machine interface platform with thousands of channels". *Journal of Medical Internet Research* 21.10 (2019): e16194.

85. Dehaene S., *et al*. "A neuronal model of a global workspace in effortful cognitive tasks". *Proceedings of the National Academy of Sciences of the United States of America* 95 (1998): 1452914534.

86. Logothetis Nikos K. "What We Can Do and What We Cannot Do with fMRI". *Nature* 453.7197 (2008): 869878.

87. Attwell D., *et al*. "Glial and Neuronal Control of Brain Blood Flow". *Nature* 468.7321 (2010): 232243.

88. Heeger David J and David Ress. "What Does fMRI Tell Us About Neuronal Activity?" *Nature Reviews Neuroscience* 3.2 (2002): 142151.

89. Smith Kerri. "Brain Imaging: fMRI 2.0". *Nature* 484.7392 (2012): 2426.

90. Webb TW., *et al*. "Cortical net-works involved in visual awareness independent of visual attention". *Proceedings of the National Academy of Sciences of the United States of America* 113.48 (2016): 13923-13928.

91. Van Boxtel J. "Different Signal Enhancement Pathways of Attention and Consciousness Underlie Perception in Humans". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 37.24 (2017): 59125922.

92. Van Rullen R. "Perceptual cycles". *Trends in Cognitive Sciences* 20.10 (2016): 723-735.

93. Butz MV. "Toward a Unified Sub-symbolic Computational Theory of Cognition". *Frontiers in Psychology* 7 (2016): 925.

94. Fitch WT. "Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition". *Physics of Life Reviews* 11.3 (2014): 329-364.

95. Stringer C., *et al*. "High-dimensional geometry of population responses in visual cortex". *Nature* 571.7765 (2019): 361-365.

96. Steinmetz NA., *et al*. "Distributed cod-ing of choice, action and engagement across the mouse brain". *Nature* 576.7786 (2019): 266-273.

97. Song C., *et al*. "Plasticity in the structure of visual space". *Eneuro* 4.3 (2017).

98. Tononi G. "An information integration theory of consciousness". *BMC Neuroscience* 5 (2004): 42.

99. Oizumi M., *et al*. "From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0". *PLOS Computational Biology* 10.5 (2014): e1003588.

100. Tononi G., *et al*. "Integrated information theory: from consciousness to its physical substrate". *Nature Reviews Neuroscience* 17 (2016): 450461.

101. Tegmark M. "Improved Measures of Integrated Information". *PLoS Computational Biology* 12.11 (2016): e1005123.

102. Mayner WGP., *et al*. "PyPhi: A toolbox for integrated information theory". *PLoS Computational Biology* 14.7 (2018): e1006343.

103. Gomez JD., *et al*. "Computing Integrated Information (ʃ) in Discrete Dynamical Systems with Multi-Valued Elements". *En-tropy* 23 (2021): 6.

104. Alireza Namdari and Zhaojun (Steven) Li. "A review of entropy measures for uncertainty quantification of stochastic processes". *Advances in Mechanical Engineering* 11.6 (2019).

105. Kalita P., *et al*. "Informational Structures and Informational Fields as a Prototype for the Description of Postulates of the Integrated Information Theory". *Entropy* 21 (2019): 493.

106. Hanson JR and Walker SI. "Integrated Information Theory and Isomorphic Feed-Forward Philosophical Zombies". *Entropy* 21 (2019): 1073.

107. Bolognesi T. "Integrated Information in Process-Algebraic Compositions". *Entropy* (2019).

108. Harold Szu., *et al*. "Theory of Glial Cells and Neurons Emulating Biological Neural Networks (BNN) for Natural Intelligence (NI) Operated effortlessly at a Minimum Free Energy (MFE)". *MOJ Applied Bionics and Biomechanics* 1.1 (2017): 8-29.

109. ZJ Zheng and A Maeder. "The conjugate classification of the kernel form of the hexagonal grid, Modern Geometric Computing for Visualization, Springer-Verlag (1992): 73-89.

110. ZJ Zheng. "Conjugate transformation of regular plan lattices for binary images, PhD Thesis, Monash University (1994).

111. Jeffrey ZJ Zheng., *et al*. "A framework to express variant and invariant functional spaces for binary logic". *Frontiers of Electrical and Electronic Engineering in China* 5 (2010): 163-172.

112. Jeffrey ZJ Zheng., *et al*. "A Framework of Variant Logic Construction for Cellular Automata, Cellular Automata - Innovative Modeling for Science and Engineering, Dr. Alejandro Salcido (Ed.), InTech Press (2011).

113. Jeffrey Zheng. "Variant Construction from Theoretical Foundation to Applications, Springer". *Nature* (2019).

114. Jeffrey Zheng. "Variant Construction Theory and Applications, Vol. 1: Theoretical Foundation and Applications". 1 Science Press (2021).

115. Jeffrey Zheng. ResearchGate.

116. Jeffrey Zheng and Chris Zheng. "Biometrics and Knowledge Management Information Systems, Chapter 11: Variant Construction from Theoretical Foundation to Applications, Springer". *Nature* (2019): 193-202.

117. Jeffrey Zheng., *et al*. "Visualizations of SARS-CoV-2 Genomes on Genomic Index Maps". *EC Neurology* SI-02 (2021): 206-221.

118. Mu Qiao., *et al*. "Visualizations of Topo-logic Entropy on SARS-CoV-2 Genomes in Multiple Regions". *EC Neurology* SI-02 (2021): 86-93.

119. Minghan Zhu and Jeffrey Zheng. "Visual Variations between Pairs of SARS-CoV-2 Genomes on Integrated Density Matrix". *EC Neurology* SI-02(2021): 94-100.

120. Yang Zhou and Jeffrey Zheng. "Visualizations of Combinatorial Entropy Index on Whole SARS-CoV-2 Genomes". *EC Neurology* SI-02 (2021): 101-109.

Volume 13 Issue 7 July 2021