# Spread of SARS-CoV-2 Genomes on Genomic Index Maps of Hierarchy

Jeffrey Zheng, Yang Zhou, Minghan Zhu, Mu Qiao, Zhigang Zhang

**Abstract** Using visual technologies, COVID-19 patients worldwide are conveniently described by position information to collect sample, modern GIS maps are useful to show influenced flows and numbers of patients on various regions of a pendamic. 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. However, there are structural difficulties to project phylogenic information into 2D distributions as GIS maps naturally.

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 the 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. Visual effects are complementarily provided with Phylogenic tree technology. Sample regions and various projections show spread effects of five thousand SARS-CoV-2 genomes in 72 countries and special four countries are selected on genomic index maps.

**Keyword** genomic index, visual maps, phylogeny, projection, information entropy, global invariant, hierarchical projection

Jeffrey Zheng[1,2]
[1] Key Laboratory of Quantum Information of Yunnan
[2] Key Laboratory of Software Engineering of Yunnan
Yunnan University, Kunming, e-mail: conjugatelogic@yahoo.com

Yang Zhou, Minghan Zhu, Mu Qiao
Yunnan University

Zhigan Zhang
School of Life Sciences & Technology, Yunnan University

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

# Introduction

The outbreak of SARS-CoV-2 causes COVID-19 earlier started from Dec. 2019 and now is a pendamic. To the date of 25 May 2020, there are more than 5.33 million confirmed and 0.34 million dead worldwide. A understanding of 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 over the global into GISAID genetic database [?] for infected viruses. Based on this effective activity, Nextstrain provides Phylogenic tree [2] to organize sample datasets from different places to category them as clusters under the maximal likelihood relationship to view intrinsic variations among SARS-CoV-2 genomes. Based on Phylogenic information, a dynamic simulation system provides flexible illustrations on selected branches [3] to support medical doctors, virological experts, biomedical specialists and psychologic doctors for detailed treatments on COVID-19 patients.

## *Weakness of Phylogenic Tree*

Phylogeny tree on 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 relationship among various clusters in a phylogenic tree shown in Fig 1. Using GIS maps, it is useful to see various genomes distributed worldwide.

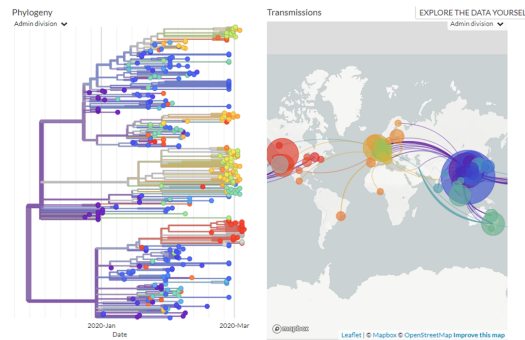



**Fig. 1** Phylogeny of real cases over global on Nextstrain

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 phylogenic tree. Since Phylogenic tree corresponds neither 1D nor 2D structure, it is hard to rearrange various subtrees [**?**] as visual objects.

In other words, effective projections for a subset of phylogenic tree provides a natural projection, other forms of visual representations could not be directly generated.

## *Combination, Matrix and Thermodynamics*

In modern mathematics and physics, there are many theoretical constructions to handle invariant and variation problems for entropy issues [5]-[35] such as combinatorial mathematics, combinatorial theory, combinatorics, multiple variable complex theory, statistical physics, thermodynamics, thermostatistics, statistical mechanics et al.

## *Variant Construction*

In this direction, vectors, matrices and invariant measurements are described relevant to wider applications [37]-[40] on variant construction [41]-[44].

Genomic index provides unique identification for each genome to be an invariant in given conditions. Based on this type of global quantitative characteristics, it is convenient for huge numbers of genomes to be located in a certain geometric region to be collected as clusters.

Different entropy quantities were discussed in separated papers: Visualization of SARS-CoV-2 Genomes on Genomic Index Maps [45], Visualizations of Topological Entropy on SARS-CoV-2 Genomes in Multiple Regions [46], Cluster Analysis of Visual Differences on Pairs of SARS-CoV-2 Genomes [47], Visualizations of Combinatorial Entropy Index on Whole SARS-CoV-2 Genomes [48].

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 from four meta symbols: {A,C,G,T}.

### *The First Order of Combinations*

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

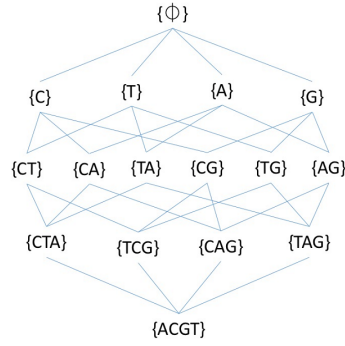



**Fig. 2** Sixteen combination of four meta-symbols in a hierarchy of a lattice

The sixteen states $SS = \{\emptyset, A, C, G, T, AC, AG, AT, CG, CT, GT, CGT, AGT, ACT, ACG, ACGT\}$ can be mapped into the sixteen numbers $SI = \{0, 1, 2, \cdots, 15\}$ respectively to represent 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: $\{m_i\}, 0 \leq i \leq 15$ 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 symbol s and $p_s$ be a probability measure. We have 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 $\{p_i\}_{i=0}^{15}, 0 \le p_i \le 1, i \in SI$ respectively.

## *Two Workflows from Input to Output*

Two workflows (1) and (2) can be identified by the type of output.
    (1) Vector of Genome $\rightarrow$ Probability   $\rightarrow$ Sixteen Probability Vectors
                                  $\rightarrow$ Entropy      $\rightarrow$ Sixteen Indexes
                                  $\rightarrow$ Selection   $\rightarrow$ An Index
    (2) $\{$Pair of Indexes$\}$ $\rightarrow$ Mapping    $\rightarrow$ A Genomic Index Map

## *Genomic Index Projection and Genomic Index Map*

Three workflows are described in three parts as Input, Output and Process respectively.
    In Step (1), one index of 16 Combinatorial Entropies can be generated.

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

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

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

### Combinatorial Entropy Measurement

Let a vector $Z$ with $(m+1)$ elements, $Z = (Z_0, Z_1, \cdots, Z_j, \cdots, Z_m), 0 \leq Z_j \leq M$ and $M = \sum_{j=0}^{m} Z_j$. Under this condition, let $P_j = \frac{Z_j}{M}$ 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.

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

$$1 = \sum_{j=0}^{m} P_j, \ 0 \leq j \leq m \tag{2}$$

For sixteen combinations of the first order, sixteen entropy measurements of $eZ$ are corresponding to $\{eZ_i\}, 0 \leq i \leq 15$.

### 2D Combinatorial Entropies

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

Under this condition for a segment with $m$ elements on a genome $Z$ with $N = m \times M$ elements, sixteen entropies $\{eZ_i\}, 0 \leq i \leq 15, ZE_i \in [0, \log_2(m+1)]$ 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 is corresponding to : $eZ_{i,j} = (eZ_i, eZ_j), 0 \leq i, j \leq 15$. There are a total of 256 pairs of 2D positions determined by the genome $Z$ in the square on $[0, \log_2(m+1)] \times [0, \log_2(m+1)]$ region.

### Multiple Genomes

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

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

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

### Genomic Index Maps

Different from a genome, it has a relative position in a phylogenic tree on the maximal 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.

### Datasets

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

### Visual Tool - Plotly

Plotly is a visual tool [36] 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 possible to apply 1D or 2D distributions to represent complicated clusters for multiple genomes in hierarchical structures.

Two distinct schemes are shown in Fig 3 for both Phylogeny 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

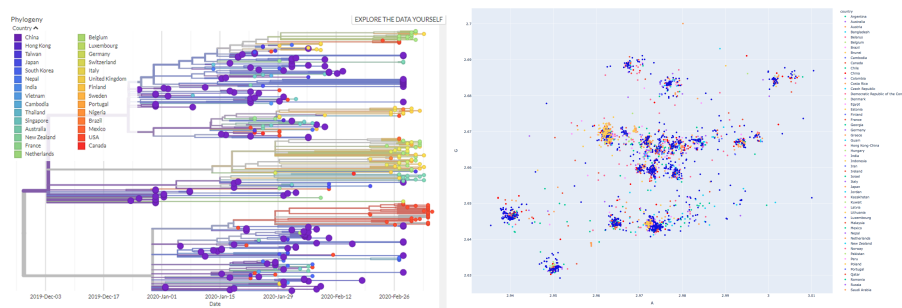



**Fig. 3** Phylogeny of real cases over global on Nextstrain and Global Genomic Index Map

are clearly visualized by distinct color points for relevant countries on the genomic index map. Refined maps are showing in the next section.

## Results

Relevant results are included in a separated file: Human-16-A-G.html that can be visualized by a HTML browser in the newest version for Plotly libraries.

For five thousand genomes, global genomic index map and various projection maps for four countries: Australia, China, Netherlands, USA were selected to show relevant projections of results in Fig 4 (a)-(b) and Fig 5(a)-(f) respectively.

For an enlarged region, regional genomic index map and various projection maps for four countries: Australia, China, Netherlands, USA were selected to show relevant projections of results in Fig 6 (a)-(b) and Fig 7(a)-(f) respectively.

Since there are more functions included in the package, it is interesting to control this program to show specific country's distribution double (single) click the entry extracted (removed) from the 72 countries or to draw a box on the region to investigate a special genome on its clustering neighborhood. When the console is pointed to a genome index position, there is a display box to show regular No. of the genome and the values of two entropy indexes.

Under this package, many complicated explorations could be carried out.

## Discussion

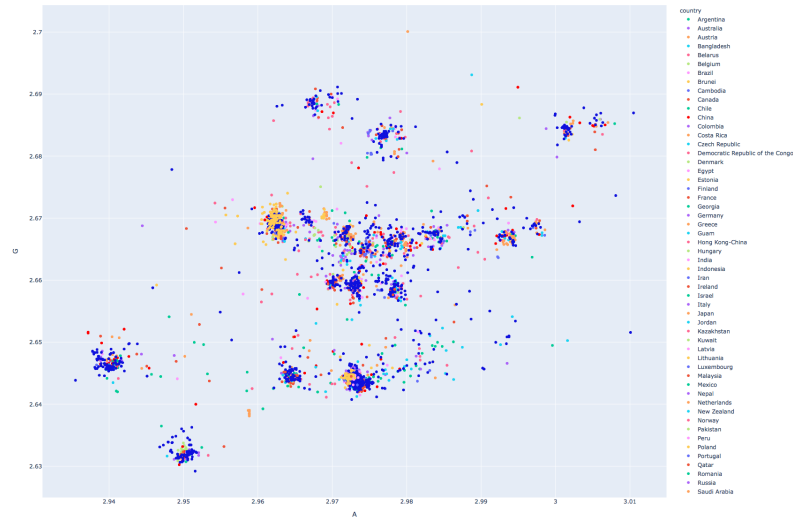
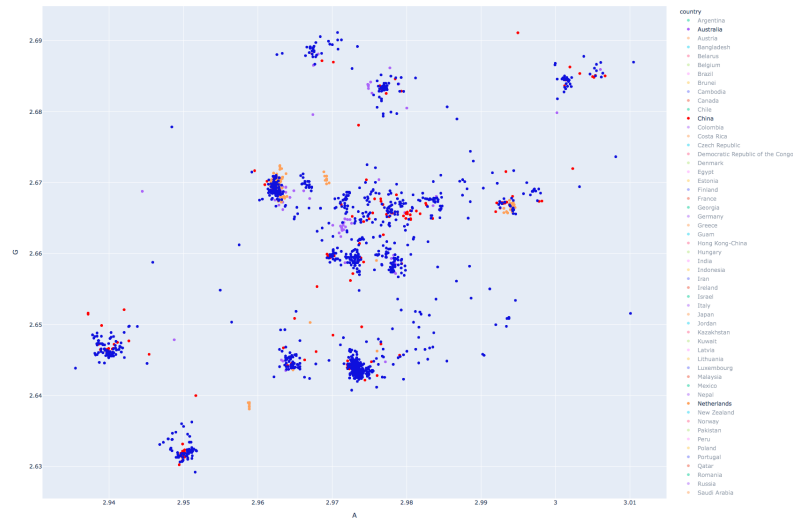
Since there is an autoscale function in Plotly package, visual regions for selected datasets may not be a fixed one with slight differences for each selected region.

(a)

(b)

**Fig. 4** Five thousands of genomes on genomic index maps (a) Global (b) Four Countries: Australia + China + Netherland + USA

## Global Projections for Four Countries

In Fig 4 and Fig 5, six genomic index maps are represented for all genomes of 72 countries and selected four countries: Australia, China, Netherlands and USA respectively.

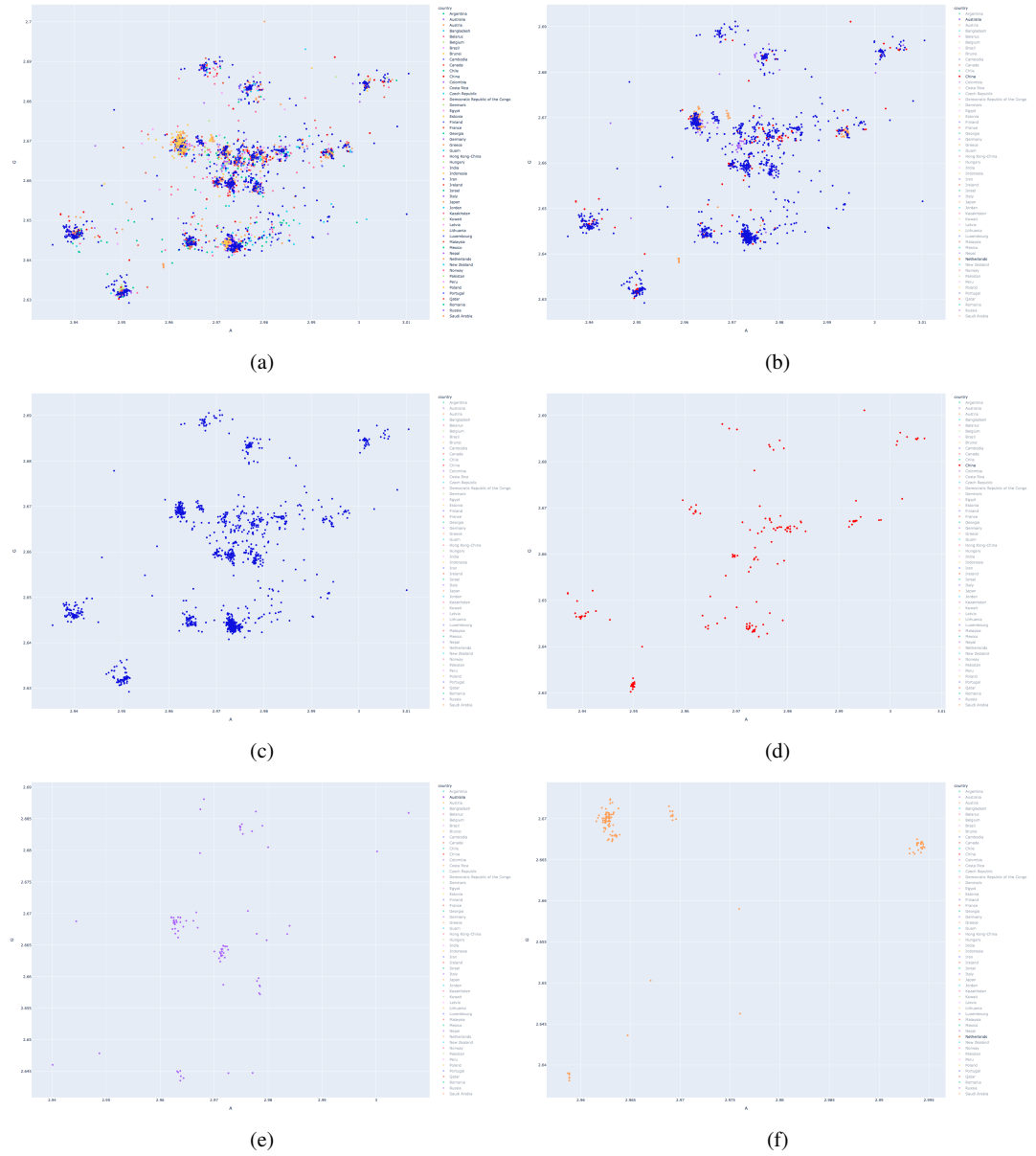



**Fig. 5** Five thousands of genomes on genomic index maps (a) Global (b) Four Countries: Australia + China + Netherlands + USA (c) USA (d) China (e) Australia (f) Netherlands

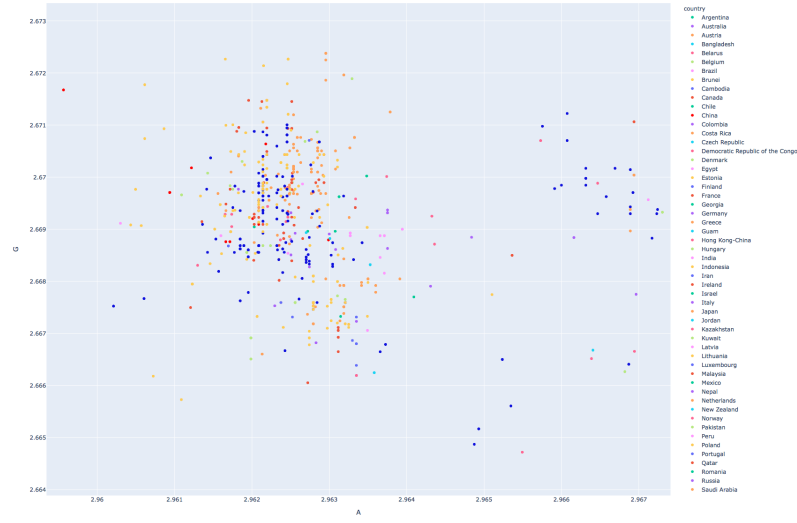
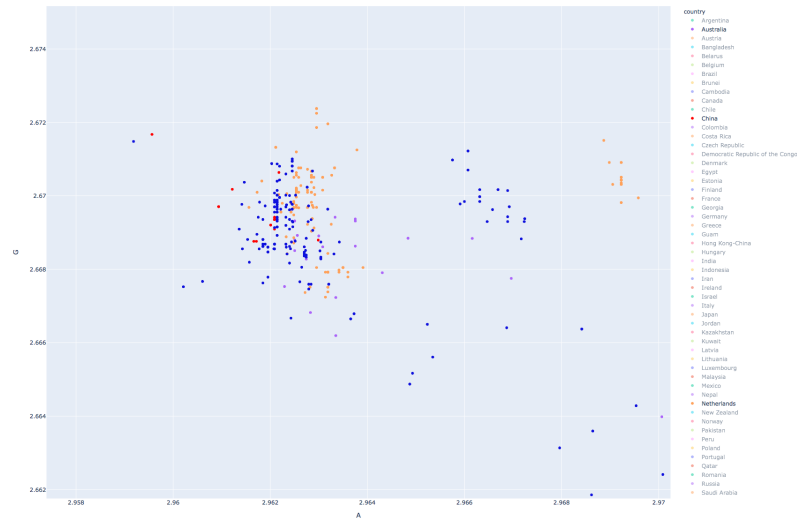
In Fig 4(a) and Fig 5(a), all genomic indexes of 72 countries are restricted in a region of $[2.93, 3.02] \times [2.62, 2.71]$ and more than sixteen clusters could be identi-

(a)

(b)

**Fig. 6** An Enlarged Region (a) Global (b) Four Countries: Australia + China + Netherland + USA

fied by visual clustering technologies. Centers of two larger clusters are located on (2.963,2.669) and (2.973,2.642) respectively.

In Fig 4(b) and Fig 5(b), all genomic indexes of four countries: Australia, China, Netherlands, USA are selected in a region of $[2.93, 3.01] \times [2.62, 2.70]$ and the sim-

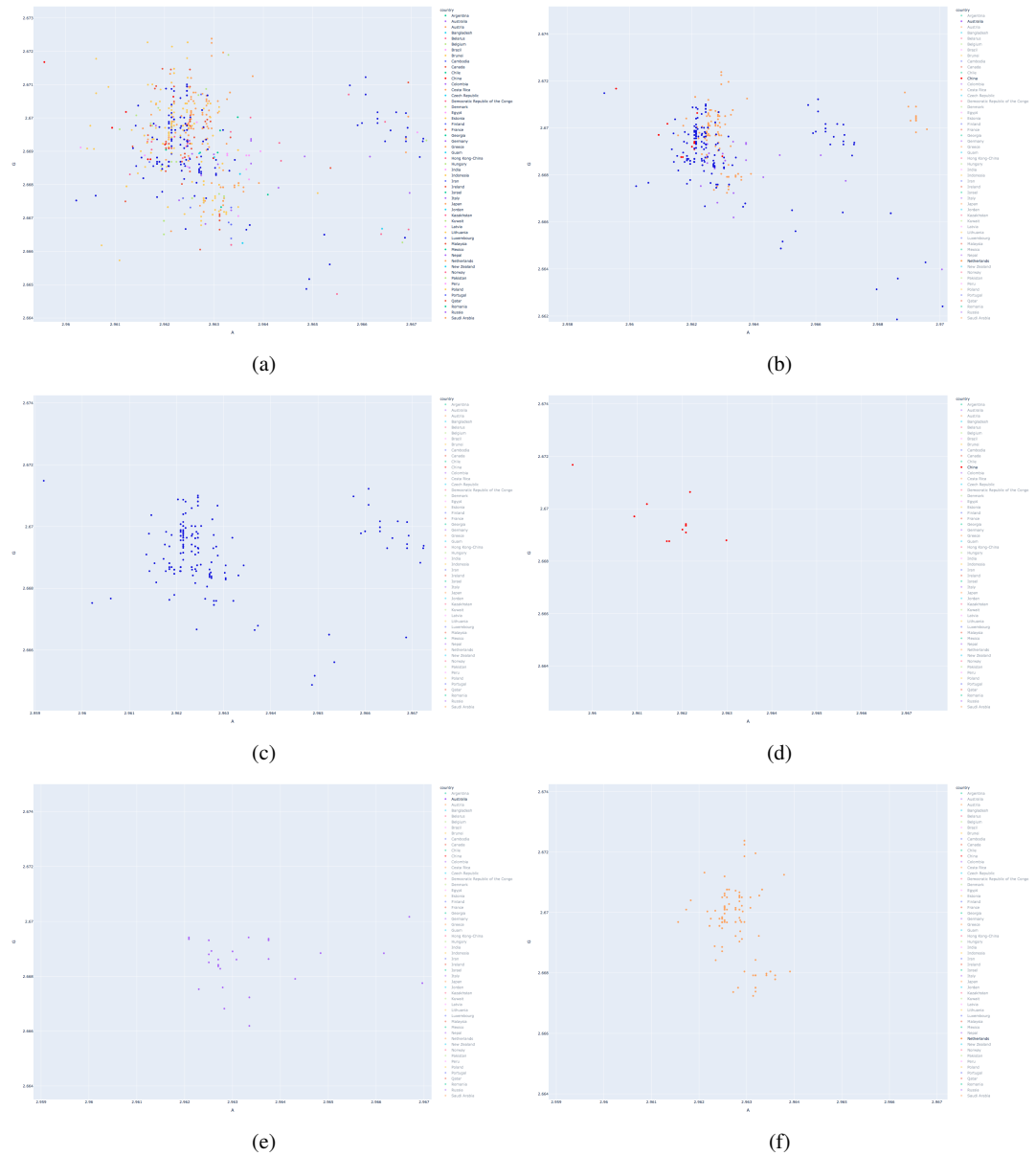



**Fig. 7** An Enlarged Region (a) Global (b) Four Countries: Australia + China + Netherlands + USA (c) USA (d) China (e) Australia (f) Netherlands

ilar number of clusters could be identified as same as in Fig 4(a). Centers of two larger clusters are located on (2.963,2.669) and (2.973,2.643) respectively.

In Fig 5(c), all genomic indexes of USA in blue color are selected from a region of $[2.93, 3.01] \times [2.62, 2.70]$ with the similar number of clusters identified in Fig 4(a). Centers of two larger clusters are located on (2.963,2.669) and (2.973,2.643) respectively.

In Fig 5(d), all genomic indexes of China in red color are selected from a region of $[2.93, 3.01] \times [2.63, 2.70]$ with lighter clusters identified in Fig 4(a) or Fig 4(b). The centers of two larger clusters are located on (2.963,2.670) and (2.994,2.643) respectively.

In Fig 5(e), all genomic indexes of Australia in light blue color are selected from a region of $[2.94, 3.01] \times [2.64, 2.69]$ with much lighter clusters identified in Fig 5(a) or Fig 5(b). The centers of two larger clusters are located on (2.963,2.668) and (2.972,2.664) respectively.

In Fig 5(f), all genomic indexes of Netherlands in yellow color are selected from a region of $[2.955, 2.995] \times [2.635, 2.675]$ with much less clusters identified in Fig 5(a) or Fig 5(b). The centers of two larger clusters are located on (2.963,2.668) and (2.972,2.667) respectively.

From this group of comparison, centers of the larger clusters have significantly differences on selected four countries. Four countries: Australia, Netherlands, USA and China are shared with the cluster on (2.963,2.668), (2.963,2.668), (2.963,2.669) and (2.963,2.670) with slightly different centers. The second larger clusters are completely different for the four countries. This could be the intrinsic facts to indicate basic genomic differences on various countries in medical treatments for COVID-19 patients.

### Enlarged Region of Global Projections for Four Countries

In Fig 6 and Fig 7, six genomic index maps are represented on enlarged regions for all genomes of 72 countries and selected four countries: Australia, China, Netherlands and USA respectively.

In Fig 6(a) and Fig 7(a), all genomic indexes of 72 countries are restricted in a region of $[2.959, 2.968] \times [2.664, 2.673]$ and more than sixteen clusters could be identified by visual clustering technologies. Center of larger cluster is located on (2.9624,2.669).

In Fig 6(b) and Fig 7(b), all genomic indexes of four countries: Australia, China, Netherlands, USA are selected in a region of $[2.957, 2.971] \times [2.662, 2.674]$. The center of larger cluster is located on (2.963,2.669).

In Fig 7(c), all genomic indexes of USA in blue color are selected from a region of $[2.959, 2.968] \times [2.664, 2.674]$. The centers of larger cluster is located on (2.9625,2.6691). The most genomic indexes may be located around left-center of the region.

In Fig 7(d), all genomic indexes of China in red color are selected from a region of $[2.959, 2.968] \times [2.663, 2.674]$. The center of larger cluster may be located on (2.962,2.669). The most genomic indexes are located on left-top part of the region.

In Fig 7(e), all genomic indexes of Australia in light blue color are selected from a region of $[2.959, 2.967] \times [2.664, 2.674]$. The center of larger cluster is located on (2.963,2.668). The most genomic index may be located on center-bottom part of the region.

In Fig 7(f), all genomic indexes of Netherlands in yellow color are selected from a region of $[2.959, 2.967] \times [2.664, 2.674]$. The center of larger cluster is located on (2.9625,2.670). The most genomic indexes may be located on center area of the region.

In this region, four countries provide genomic indexes as main clusters with slightly different positions. Compared with four centers: USA (2.9625,2.6691), China (2.962,2.669), Australia (2.963,2.668) and Netherlands (2.9625,2.670). There are around $(\pm 0.001, \pm 0.0002)$ differences among four centers respectively.

In principle, the enlarged operations can be repeatedly applied to selected regions to refine detailed location in the finest value. In the most condition, two genomic indexes could be separated when a larger fold magnification has applied.This may provide conveniently classified effects for medical doctors to treatment 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. Applying five thousand genomes of SARS-CoV-2 on 72 countries and special selections on four countries based on Plotly libraries, two specific sets of six genomic index maps are shown in significantly different distributions on each country to illustrate complicated contagiousness patterns among various regions.

Using genomic index maps, further refined classifications and categories of genomes can be visually explored and look forward this visual tool useful in further medical treatments for COVID-19 patients worldwide in coming future.

## Conflict Interest

No conflict of interest has claimed.

**Acknowledgements** The authors would like to thank NCBI, GISAID, CNGBdb, Nextstrain to provide invaluable information on the newest dataset collections of SARS-CoV-2 & other coronavirus genomes to support this project working smoothly.

# References

1. GISAID: Open access to influenza virus data https://gisaid.org
2. Nextstrain Real time tracking of pathogen evolution https://nextstrain.org
3. HP Yao, XY Lu, ···, LJ Li, Patient-drived mutations impact pathogenicity of SARS-CoV-2, DOI: https://doi.org/10.1101/2020.04.14.20060160 https://www.medrxiv.org/10.1101/2020.04.14.20060160v2
4. Li C, Yang Y, Ren L. Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species. Infect Genet Evol. 2020 Mar 10;82:104285. doi: 10.1016/j.meegid.2020.104285. [Epub ahead of print] PMID: 32169673
5. P. J. Cameron, *Combinatorics: Topics, Techniques, Algorithms*, Cambridge University Press, 1994.
6. J. R. Chen, *Combinatorial Mathematics*, Harbin Institute of Technology Press, 2012 (in Chinese).
7. H. W. Gould, Some Generalizations of Vandermonde's Convolution, *The American Mathematical Monthly*, Vol. 63, No.2 84-91, 1956.
8. H. W. Gould, *Combinatorial identities*, Morganton, 1972.
9. M. Hall, *Combinatorial Theory*, 2nd edition, Blaisdell, 1986.
10. L. K. Hua, *Loo-Keng Hua Selected Papers*, Springer, 1982.
11. L. K. Hua, *Selected Work of Hua Loo-Keng on Popular Sciences*, Shanghai Education Press, 1984 (in Chinese).
12. D. E. Knuth, *The Art of Computer Programming*, Vol. 1, 3rd edition, Addison-Wesley, 1998.
13. D. E. Knuth, *The Art of Computer Programming*, Vol. 4A: Combinatorial Algorithms, Part 1, Addison-Wesley, 2011.
14. F. Morgan, *Geometric Measure Theory*, 4th edition, Elsevier 2009.
15. G. Polya, R. Tarjan and D. Woods, *Notes on Introductory Combinatorics*, Birkhauser, 1983.
16. R. P. Stanley, *Enumerative Combinatorics*, Vol. 1, 2nd edition, Cambridge University Press, 1997.
17. D. Stanton, R. Stanton and D. White, *Constructive Combinatorics*, Springer-Verlag, 1986.
18. G.Z. Tu, *Combinatorial Enumeration Methods & Applications*, Science Press,1981 (in Chinese).
19. A. Tucker, *Applied Combinatorics*, John Wiley & Sons, 2007.
20. J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*, 2nd edition, Cambridge University Press, 2001.
21. L. X. Wang, *An Elementary Treatise on Combinations*, Harbin Institute of Technology Press, 2012 (in Chinese).
22. L. Z. Xu, M. S. Jiang and Z. Q. Zhu, *Combinatorial Mathematics of Computation*, Shanghai Science & Technology Press, 1983 (in Chinese).
23. L D Landau, E M Lifshitz. Statistical Physics, 3rd edition, Part 1, Pergamon Press 1986.
24. HB Callen. Thermodynamics and an Introduction to Thermostatistics, 2nd Edition. John Wiley & Sons 1985.
25. W Greiner, L Neise, H Stocker. Thermodynamics and Statistical Mechanics. Springer-Verlag 1995.
26. T L Hill, Introduction to Statistical Thermodynamics, Addison-Wesley, Reading Mass. 1960.
27. R Kubo. Thermodynamics. North-Holland Pub. Co. 1968.
28. R P Freynman, Statistical Mechanics, Benjamin Reading Mass. 1972.
29. B Widom, in Foundamental Problems in Statistical Mechanics, Vol. III edi. by Cohen, Horth-Holland, 1975 1-45.
30. K G Wilson, Rev., Mod. Phys. 55, 583 1983
31. J Beatte, I Oppenheim. Thermodynamics, Elsevier Scientific 1979.
32. S K Ma, Statistical Mechanics, World Scientific 1985.
33. P W Atkins, The Second Law, Scientific American Books and W H Freeman and Co. 1984.
34. D Chandler. Introduction to Modern Statistical Mechanics, 1st edition. Oxford University Press Inc. 1987.

35. K Huang. Statistical Mechanics, 2nd edition John Wiley & Sons 1987.

36. Plotly: The front-end for ML and data science models. https://plotly.com

37. Z. J. Zheng, A. Maeder, The The conjugate classification of the kernel form of the hexagonal grid, *Modern Geometric Computing for Visualization*, Springer-Verlag, 73-89, 1992.

38. Z. J. Zheng,*Conjugate transformation of regular plan lattices for binary images*, PhD Thesis, Monash University, 1994.

39. Jeffrey Z. J. Zheng, Christian H. H. Zheng, A framework to express variant and invariant functional spaces for binary logic, *Frontiers of Electrical and Electronic Engineering in China*, 5(2):163-172, Higher Educational Press and Springer-Verlag, 2010.

40. Jeffrey Z.J. Zheng, Christian H.H. Zheng and Tosiyasu L. Kunii. A Framework of Variant Logic Construction for Cellular Automata, *Cellular Automata - Innovative Modeling for Science and Engineering*, Dr. Alejandro Salcido (Ed.), InTech Press, 2011.

41. Jeffrey Zheng, Variant Construction from Theoretical Foundation to Applications, Springer Nature 2019 https://www.springer.com/in/book/9789811322815

42. Jeffrey Zheng, Variant Construction Theory and Applications, Vol. 1: Theoretical Foundation and Applications, Science Press 2020 (Chinese, Formal Publishing Soon).

43. Jeffrey Zheng, ResearchGate: http://researchgate.net/pprofile/Jeffrey_Zheng

44. Jeffrey Zheng, Chris Zheng, Biometrics and Knowledge Management Information Systems, Chapter 11: Variant Construction from Theoretical Foundation to Applications, Springer Nature 2019, 193-202 https://link.springer.com/chapter/10.1007/978-981-13-2282-2_11
Selected in Research of COVID-19 for PubMed Central PMC and the World Health Organization WHO by Springer-Nature on Free Access of all scientific researchers worldwide.

45. Jeffrey Zheng, Minghan Zhu, Mu Qiao, Yang Zhou. Visualizations of SARS-CoV-2 Genomes on Genomic Index Maps, submitted to Electric Current Neurology, in review process.

46. Mu Qiao, Renyang Liu, Zhenhui Wang, Xinmei Li, Jeffrey Zheng. Visualizations of Topologic Entropy on SARS-CoV-2 Genomes in Multiple Regions, submitted to Electric Current Neurology, in review process.

47. Minghan Zhu, Jeffrey Zheng. Cluster Analysis of Visual Differences on Pairs of SARS-CoV-2 Genomes, submitted to Electric Current Neurology, in review process.

48. Yang Zhou, Jeffrey Zheng. Visualizations of Combinatorial Entropy Index on Whole SARS-CoV-2 Genomes，submitted to Electric Current Neurology, in review process.