2019 novel coronavirus is undergoing active recombination

Huiguang Yi*

Department of Biology, South University of Science and Technology of China, Shenzhen, China

*Correspondence: Huiguang Yi: yhg926@gmail.com

TO THE EDITOR:

The 2019 novel coronavirus (2019-nCoV) outbreak in Wuhan since December 2019 [1] has quickly spread to twenty-five countries [2], caused more than 44,000 cases and 1,000 deaths so far [3]. The fast sharing of 2019-nCoV genomes in GISAID (https://www.gisaid.org) provides a valuable dataset for 2019-nCoV haplotype network analysis, which is crucial for transmission and evolutionary track surveillance and secondary outbreak prevention.

We called single-nucleotide variations (SNVs) for all the 84 2019-nCov genomes in GISAID using MN908947 as the reference. Genomes with same SNVs are grouped into a haplotype (shown as a pie chart node in Fig. 1, see Tab. S1 for the accessions), then the network is constructed using the median join [4] method in popART [5]. We found the 2019-nCoV haplotype network has obvious characteristics of single origin from haplotype hap_011: first, the network is star-like, centralized on the haplotype hap_011; second, hap_011 has the largest sample size and majority of the samples are from Hubei province—where outbreak originated (Tab. S1); third, most of satellite haplotypes are also from Hubei (Fig. 1); fourth, the average collection dates of hap_011 (has 0 mutation relative to MN908947) is earlier than all other mutation groups (Fig. S1). The single origin of 2019-nCov indicates a persistent animal to human transmission is unlikely, otherwise, multiple nodes with above characteristics should be observed.

We found five haplotypes (hap_009, hap_017, hap_023, hap_048 and hap_050) forming loops (Fig. 1), which typically indicate existing of genetic recombination. However, in rare cases, loops can also be formed by recurring sequencing error or parallel/back mutations. Further examination of the differences pattern among the five haplotypes found three sets of recurrent differences: C29095T, G11083T and C8782T/T28144C (Fig. 1), most of which are nonsynonymous changes except C29095T (Tab. S2). Such highly frequent recurrent differences looks extremely unlikely raised by rare events like recurring sequencing error/mutations. To quantify the significance of recombination over recurring sequencing error/mutations, Phi [6] and Max Chi-squared tests [7] in the software PhiPack [8] were performed. It showed p-values of 0.03 and 0.005 for Phi and Max Chi-squared, respectively, indicating the presence of recombinants in 2019-nCoV population.

Since all the haplotypes originate from hap_011, the formation mechanisms of the five haplotypes could be inferred as follow: First, hap_048 and hap_050 mutated from the common ancestry hap_011 independently, hap_009 then mutated from hap_050; next, hap_048 donated the genome region harboring 11083T to hap_050 generated the recombinant hap_017; similarly, hap_048 or hap_017 donated the same region to hap_009 generated the recombinant hap_023 (Fig. 2).

Recent studies have found some ancient recombination events around region 11655-20953 (relative to MN908947, containing gene orf1b) in bat coronavirus CoVZC45/CoVZXC21 (a sister group of 2019-nCoV), but shown no evidence of recombination for 2019-nCoV [9-11]. It is anticipated since all these studies adopted region-wise phylogenetics analysis which cannot efficiently detect recent recombination events like haplotype network does. However, the recombination region they found in bat coronavirus is quite close to the 2019-nCoV recombination site 11083 (harbored by gene ORF1a) found in this report. This consistency of the ancient and the recent recombination analysis solidifies both our findings and prompts potential recombination hotspots existing around region 11083-20953 or gene ORF1a/ORF1b.

This report provides the first evidence for genetic recombination—a new way of evolution besides mutation in 2019-nCoV. The existing of genetic recombination has the following implications: two different 2019-nCoV strains (here, hap_048 and hap_050) should have co-infected the same cell; a 2019-nCoV strain might acquire new traits like virulence and drug susceptibility directly from other strains [12, 13]; the adaptability of 2019-nCoV to human immune system might be significantly strengthened through genetic recombination; the accuracy of diagnosis based on serologic and molecular biology assays might be compromised [14]; and the transmission tracking based on phylogenetics tree could be misleading since the topology of mutation route is a network rather than a tree.

We think the communities of infectious disease physicians and disease control specialists should arouse particular concerns about this finding and its potential implications. And we expect this

findings may provide a valuable perspective for strategies design for secondary transmission prevention and drug treatments for 2019-nCoV pneumonia.

Acknowledgments

We acknowledge the authors and the originating and submitting laboratories of the nucleotide sequences from BetaCoV2019-2020 of GISAID EpiFluTM Database (9 Feb 2020, 84 isolates).

Potential conflicts of interest

All authors: No reported conflicts.

References

1. Li, Q., et al., *Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia.* 2020.
2. *Confirmed 2019-nCoV Cases Globally*

Global Map. 2020 2020-02-12]; Available from: https://www.cdc.gov/coronavirus/2019-ncov/locations-confirmed-cases.html#map.

3. Quick, J., et al., *Real-time, portable genome sequencing for Ebola surveillance.* Nature, 2016. **530**(7589): p. 228-232.
4. Bandelt, H.-J., P. Forster, and A. Röhl, *Median-joining networks for inferring intraspecific phylogenies.* Molecular Biology and Evolution, 1999. **16**(1): p. 37-48.
5. Leigh, J., Bryant D *PopART: Full-feature software for haplotype network construction.* . Methods Ecol Evol 2015. **6**(9): p. 1110–1116.
6. Bruen, T.C., H. Philippe, and D. Bryant, *A Simple and Robust Statistical Test for Detecting the Presence of Recombination.* Genetics, 2006. **172**(4): p. 2665-2681.
7. Smith, J.M., *Analyzing the mosaic structure of genes.* J Mol Evol, 1992. **34**(2): p. 126-9.
8. Bruen, T. *PhiPack: PHI test and other tests of recombination*. 2005.
9. david.l.robertson. *nCoV's relationship to bat coronaviruses & recombination signals (no snakes) - no evidence the 2019-nCoV lineage is recombinant*. 2020; Available from: http://virological.org/t/ncovs-relationship-to-bat-coronaviruses-recombination-signals-no-snakes-no-evidence-the-2019-ncov-lineage-is-recombinant/331.
10. Paraskevis, D., et al., *Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event.* Infect Genet Evol, 2020. **79**: p. 104212.
11. Lu, R., et al., *Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding.* Lancet, 2020.
12. Gu, Z., et al., *Possible involvement of cell fusion and viral recombination in generation of human immunodeficiency virus variants that display dual resistance to AZT and 3TC.* J Gen Virol, 1995. **76 (Pt 10)**: p. 2601-5.
13. Moutouh, L., J. Corbeil, and D.D. Richman, *Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure.* Proc Natl Acad Sci U S A, 1996. **93**(12): p. 6106-11.
14. Blackard, J.T., D.E. Cohen, and K.H. Mayer, *Human immunodeficiency virus superinfection and recombination: current state of knowledge and potential clinical consequences.* Clin Infect Dis, 2002. **34**(8): p. 1108-14.

Figures

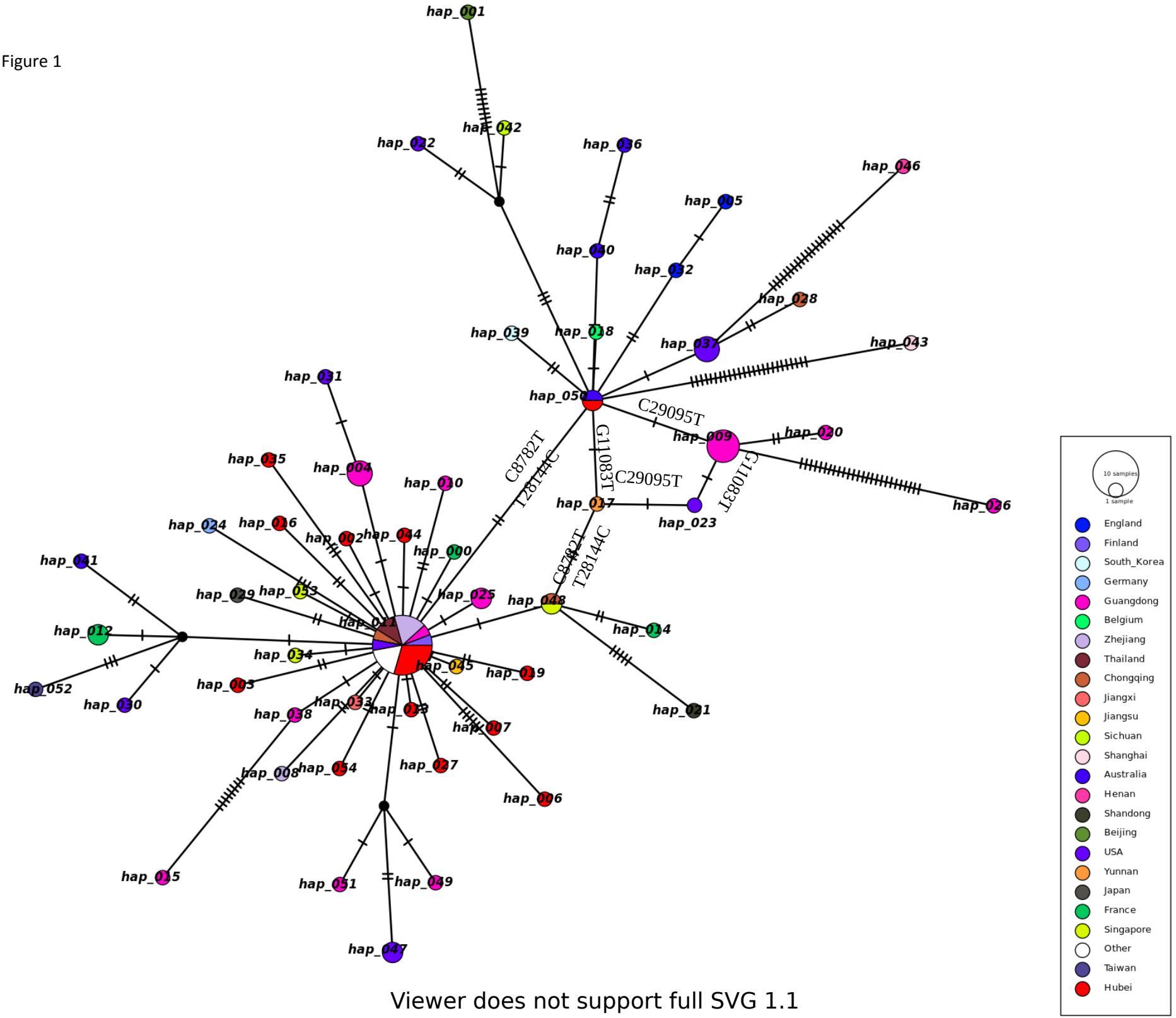
Figure 1. The Median-Joining haplotype network of 2019-nCoV.

Each haplotype node is represent by a pie chart, with size indicating the sample size (Tab S1), and colors indicating collection regions. Each hatch mark along the edge corresponds to a mutation between the two haplotypes linked. The mutation notation along a loop edge shows the position and nucleotide status change between the two haplotypes linked.

Figure 2. The putative evolutionary mechanisms of the five haplotypes forming the loops.

Genome is represented by a long rectangle, which consists of four regions represented by short rectangles. Each region contain a SNV with coordinate at bottom. The haplotype of a genome is represented by the combined SNVs from each of the region. SNVs of reference status are font colored as black and SNVs of alternative status are font colored by others, where mutants sourced from hap_048, hap_050 and hap_009 are font colored as green, blue and red, respectively. Regions sourced from hap_011/ref, hap_048 and hap_050 are filled with white, blue and pink, respectively.

Figure 1



Viewer does not support full SVG 1.1

Figure 2