Composition and Dynamics of H1N1 and H7N9 Influenza A Virus Quasispecies in a Co-infected Patient Analyzed by Single Molecule Sequencing Technology

Peng Lin1,2, Tao Jin2,3, Xinfen Yu4, Lifeng Liang2, Dragomirka Jovic3, Zhou Sun4, Zhe Yu3, Jingcao Pan4* and Guangyi Fan2,3*

1College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, 2BGI-Qingdao, BGI-Shenzhen, Qingdao, China, 3BGI-Shenzhen, Shenzhen, China, 4Hangzhou Center for Disease Control and Prevention, Hangzhou, China

A human co-infected with H1N1 and H7N9 subtypes influenza A virus (IAV) causes a complex infectious disease. The identification of molecular-level variations in composition and dynamics of IAV quasispecies will help to understand the pathogenesis and provide guidance for precision medicine treatment. In this study, using single-molecule real-time sequencing (SMRT) technology, we successfully acquired full-length IAV genomic sequences and quantified their genotypes abundance in serial samples from an 81-year-old male co-infected with H1N1 and H7N9 subtypes IAV. A total of 26 high diversity nucleotide loci was detected, in which the A-G base transversion was the most abundant substitution type (67 and 64%, in H1N1 and H7N9, respectively). Seven significant amino acid variations were detected, such as NA:H275Y and HA: R222K in H1N1 as well as PB2: E627K and NA: K432E in H7N9, which are related to viral drug-resistance or mammalian adaptation. Furtherly, we retrieved 25 H1N1 and 22 H7N9 genomic segment haplotypes from the eight samples based on combining high-diversity nucleotide loci, which provided a more concise overview of viral quasispecies composition and dynamics. Our approach promotes the popularization of viral quasispecies analysis in a complex infectious disease, which will boost the understanding of viral infections, pathogenesis, evolution, and precision medicine.

Keywords: H1N1 and H7N9, quasispecies, composition and dynamics, SMRT, precision medicine

INTRODUCTION

Influenza A virus (IAV) is a contagious pathogen that constantly infects many hosts, including but not limited to humans, birds, and pigs (Medina and García-Sastre, 2011). Annual influenza virus infections have significant health and economic burdens to mankind and livestock (Gordon and Reingold, 2018). IAV is a member of the Orthomyxoviridae family, and its genome contains eight negative-sense single-stranded RNA segments, ranging from 850 to 2,350 bp (Pleschka, 2013; Hutchinson, 2018). IAV can be subtyped as HxNy by viral surface antigens hemagglutinin (HA) and neuraminidase (NA) proteins, which govern the viral lifecycle at cellular entry and release of virions (Dou et al., 2018). So far, eighteen different HA subtypes (H1-H18) and eleven different NA...
subtypes (N1-11) have been observed (Boktor and Hafner, 2019). There are two common IAV cellular receptors: α-2,3-Sialic acid (α-2,3-SA) and α-2,6-Sialic acid (α-2,6-SA) in hosts (Nelli et al., 2010; França et al., 2013; Byrd-Leotis et al., 2017; Chen et al., 2018a; Xu et al., 2019). The avian influenza viruses such H7N9 preferentially recognize α-2,3-SA receptors, while human influenza viruses such H1N1 has a priority to α-2,6-SA receptors (Xiong et al., 2013; de Graaf and Fouchier, 2014). The α-2,6-SA receptors are dominant in the upper respiratory tract (URT) of humans, while α-2,3-SA receptors are relatively more abundant than α-2,6-SA receptors found in the lower respiratory tract (LRT) of humans (Walther et al., 2013; Lakdawala et al., 2015; Long et al., 2019).

Influenza A viruses in a host exist as a population including thousands of virions containing closely related (but nonidentical) genomes, also called quasispecies (Lauring and Andino, 2010; Martínez et al., 2012; Watanabe et al., 2018; Bonomo et al., 2019; Domingo and Perales, 2019). These closely related genomes result from error-prone replication and frequent reassortment of influenza virus genomes (Steel and Lowen, 2014; Pauly et al., 2017). The complicated interactions (cooperativity or interference) among genomes and their productions collectively determine the biological or medical implications of a viral population such as fitness, virulence, pathogenesis, immune escape, or drug resistance (Vignuzzi et al., 2006; Sanz-Ramos et al., 2008; Aragri et al., 2016; Schuster, 2016; Perales, 2020). Therefore, it is primary to reveal the composition and dynamic of viral quasispecies to better understand viral infection, adaptation, and evolution at the level of population (Xue et al., 2017; Donohue et al., 2019; Jary et al., 2020).

Achieving thousands of full-length genomes in a viral population is decisive for quasispecies composition. Previous short-read massively parallel sequencing (MPS) projects have collected abundant consensus genomic sequences (CGSs) and single nucleotide variants (SNVs) to explore influenza virus quasispecies (Van den Hoek et al., 2015; Ali et al., 2016; McGinnis et al., 2016). Nevertheless, the quasispecies composition is still unclear because of degenerated CGSs and scattered SNVs rooted in short reads from MPS (Schadt et al., 2010; Beerenwinkel et al., 2012; Chen et al., 2018b). The long-read single-molecule real-time sequencing (SMRT) provides an access to full-length influenza virus genomes even with a low frequency in a viral population (Ardui et al., 2018; Lui et al., 2019). The circular consensus sequencing (CCS) reads (average length 13.5 kb) produced by SMRT are 5–15 times as long as the genomic RNAs of influenza A virus, avoiding the fragmentation and assembly of genomes before and after sequencing (Wenger et al., 2019a; Van Poelvoorde et al., 2020).

The sample co-infected with two IAV subtypes is a very good case to understand viral infection, adaptation, and evolution at the level of population (Xue et al., 2017; Donohue et al., 2019; Jary et al., 2020). Therefore, it is primary to reveal the composition and dynamic of viral quasispecies to better understand viral infection, adaptation, and evolution at the level of population (Xue et al., 2017; Donohue et al., 2019; Jary et al., 2020).

MATERIALS AND METHODS

Patient, Symptoms and Therapies
An 81-year-old male had a slight cough and chest distress on 1/12/2016 at his home in Xihu District, Hangzhou City, Zhejiang Province, China. On the morning of 1/15/2016, the symptoms worsened with a nasty cough, chest distress and fever. That afternoon, the man went to the local community hospital, where the temperature was 38.5 C, and then he was sent to the Hangzhou First People’s Hospital for medical treatment. The examination showed that the white blood cell count (WBC) was 13.4 × 10^9/L, the percentage of neutrophils (N%) was 92.2%, and the C-reactive protein (CRP) was 110 mg/L. The chest radiographs showed an infection of the right lower lung. The merocillin sodium and sulbutamn sodium were given for intravenous injection as an anti-infective therapy. The patient was sent to the respiratory department for hospital treatment and pneumonia was confirmed on 1/16/2016.

The next day, a PCR result from a throat swab was positive on influenza A virus and the patient was sent to infection ward for further treatment which included oseltamivir (75mg/bid) and meropenem drugs. On 1/19/2016, patient’s symptoms worsened further and chest radiographs confirmed that infections spread on both lungs. The patient received endotracheal intubation and then admitted to intensive care unit (ICU). The treatment was continued, the dose of oseltamivir was doubled (150mg/bid) and meropenem was changed to imipenem. On the same day, the patient’s RT-PCR test taken from the throat was positive on the M gene, H7 gene, N9 gene and H1 gene of influenza A virus. Next day, the patient was transferred to Hangzhou Xixi Hospital for treatment in isolation where the anti-viral oseltamivir (150mg/bid) continued to be given until 2/20/2016. Although, the symptomatic treatments such as diuresis, analgesia, vasodilation and nutritional support has been given in the Xixi hospital, the patient did not show signs of improvement, and passed away on 2/28/2016.

In the patient’s anamnesis it is stated that the patient went to local live poultry market and bought a live duck at a merchant’s site about a week before the first symptoms appeared on 1/12/2016. The live duck was slaughtered, depilated, and bellied by the merchant at his site. After returning home, the patient salted the duck. The patient had a history of hypertension and denied the history of diabetes, viral hepatitis, tuberculosis, and other diseases. There is no history of trauma, surgery, or blood transfusion. Denying any history of drug or food allergies.
Samples and RT-PCR
Ten serial samples were collected from this patient from 1/19/2016 to 2/19/2016, including seven throat swabs and three sputa. The swabs and sputa were placed into 1 ml viral transport medium, transported to the laboratory within 24 h at 4°C, and then frozen at −80°C. Viral RNAs were extracted from samples using a RNAeasy Mini Kit (QIAGEN, Germany). Identification of influenza A virus was achieved by RT-PCR using specific primers targeting the M, H7, N9, and H1 gene according to the protocol provided by WHO manual (Organization, W.H, 2002). This study was approved by the Institutional Review Board of BGI (NO.BGI-IRB 16008).

Single-Molecule Real-Time Sequencing
Top eight samples were taken to perform single-molecule real-time sequencing (SMRT). The cDNAs were synthesized from viral RNAs by reverse transcription using Uni12 and Uni13 primers (Bi et al., 2016). The PCR was performed using a Phusion High-Fidelity PCR Kit (New England Biolabs) utilizing the barcoded influenza A virus general primers (Supplementary Table S1) (Mei et al., 2016). The concentration of PCR product was quantified by the Agilent Technologies 2,100 bioanalyzer. The two corresponding volumes of PCR products (containing equal mass of dsDNA) were mixed into one sample and quantified in the bioanalyzer again. About 2–3 µg mixed sample was used to SMRTbell library construction following the 2 kb template preparation protocol (Roberts et al., 2013b). The sequencing was performed on a PacBio RS II instrument (Pacific Biosciences, USA) with one SMRT Cell used for each library, using P6/C4 chemistry with a 4 h movie (Bull et al., 2016). SMRTbell adapter sequences were removed and circular consensus sequence (CCS) reads were achieved with SMRT Analysis v2.3 (Roberts et al., 2013a).

Sequence Quality Control
The raw CCS reads were filtered by removing low quality reads (length<800bp, passes<5 or estimated accuracy<99.9%). The 800 bp length near the lower limit of influenza A virus genomic RNAs was used to exclude non-full-length genomic sequences. The other two criteria ensured reads with at least 99.9% estimated accuracy and necessary passes (Korlach, 2015). The sequencing error bases (frequency<0.3%) were additionally corrected to improve sequence reliability as follows: First, the remaining sequences after filtration were split into corresponding samples by 100% base match with barcodes. Then, the sequences of one sample were grouped by subtypes and genomic segments according to the sequence annotation result against an influenza virus genomes database downloaded from NCBI (https://ftp.ncbi.nih.gov/genomes/INFLUENZA) using BLASR (v5.1 with options: -bestn 1) (Chaisson and Tesler, 2012). All full-length genomic sequences of one group were aligned end to end using MUSCLE (v3.8.31 with default options) (Edgar, 2004). Following, the number and percentage of base A/C/T/G in the same nucleotide locus of genomic sequences were stated. Finally, the very low frequency base which percentage was less than 0.3% in the number of four type bases of the same nucleotide locus, was replaced with the dominant type of base with the largest proportion in this nucleotide locus.

Diversity Index of Genomic Sequences
The diversity index (Shannon entropy) of one group sequences is calculated by the formula (Crooks and Brenner, 2004):

$$S = -100 + \sum_{i=1}^{n} P_i \cdot \log_2 P_i$$

In which $S$ is the Shannon entropy and $P_i$ is the ration of the number of one type of sequence to the number of total types of sequence in one group.

Nucleotide Loci With High-Diversity Base Composition
In order to screen out nucleotide loci with high-diversity base composition, we stated the number and percentage of base A/C/T/G in one nucleotide locus and screened out the loci in which the percentages of at least two types of bases were more than 10%.

RESULTS
Sequences Quality Control
The clinical symptoms and therapeutic schedule of this patient were recorded in Table 1. Ten samples were collected from this patient on different days, including seven throat swabs (S1-4, S7-10) and three sputa (S5-7). The collected date, sample types and Ct value of RT-PCR for H1N1 and H7N9 of each sample were listed in Table 2. The top eight samples (S1-8) were performed using SMRT with four SMRT cells (S9 and S10 were RT-PCR negative for influenza A virus). A total of 142,496 CCS reads (221.72 Mb) were generated from four SMRT cells, of which 82,471 high quality reads (≥99.9% estimated accuracy) were selected for further analysis according to strict filtering criteria. The IAV mutation rate was about 0.018–0.025%, which means that each replicated influenza genome (~13 kb) contained an average of 2–3 mutations (Pauly et al., 2017). However, the estimated base sequencing error rate of high-quality CCS reads (~0.1%) in this study, was notable higher than the normal replication mutation rate (0.018–0.025%) of IAV genomes. Therefore, it was necessary to correct sequencing error bases in the prevention of taking them for real mutations. Finally, 69 group sequences were clustered from four SMRT cells according to samples, subtypes, and genomic segments (Figure 1A). Among them, 58 group sequences were with a satisfactory abundance (the sequences number ≥20). In case that the base type (A/C/G/T) in one nucleotide locus of one group sequences was less than 0.3%, it was considered as sequencing error base. The base sequencing error rates of 58 group sequences ranged from 0.13 to 0.21%, approximate to the estimated sequencing error rate of 0.1%, which are composed of 78.37% mismatches, 13.78% deletions and 7.16% insertions (Figures 1B,C and Supplementary Table S2). Thus, we set 0.3% as the cutoff value to screen out sequencing error base types in nucleotide loci of each group sequences and correct them with dominant base types in these loci.
had transferred to LRT environment from URT (Gao et al., 2013). Meanwhile, there was an obvious increase of H1N1 viral load in the URT samples from S1 to S2 (Figure 2A). The reason for this increase might be due to the transfer of H7N9 from URT to LRT that contributes freeing up more cellular resources for H1N1 growth in LRT environment.

**High Diversity Nucleotide Loci in H1N1 and H7N9 Genomes**

Fifteen and eleven high diversity nucleotide loci were detected in H1N1 and H7N9 genomes, respectively, in which the A-G transversion was the most abundant substitution type (67% in H1N1 and 64% in H7N9) (Table 3). In H1N1, another two substitutions were C-T and A-C transversion (20 and 13% respectively). In H7N9, the C-T, A-C and G-T took the same 0.09% proportion respectively (Table 3). In terms of amino acid, the percentage of non-synonymous mutation was greater than synonymous mutation, especially the percentage of non-synonymous mutation was up to 91% (10/11) in H7N9, comparing with the non-synonymous mutation of 66% (10/15) in H1N1 (Table 3). In H1N1, all five synonymous mutations are in the internal genes of IAV, including two in NS gene (R78R and L69L), one in PB2 gene (T25T), one in NP gene (E64E) and one in M gene (R134R). In H7N9, the only one synonymous mutation was R753R in internal PB2 gene (Table 3). The H7N9 with a high frequency of non-synonymous mutation might indicate that H7N9 as avian-origin influenza virus was subjected to higher selection pressures in the human host (Wei et al., 2012; Xu et al., 2019). It is noteworthy to mention that among these high-diversity loci, four mutations in H1N1 were involved in the evolution or viral drug-resistance, such as the R222K in the HA protein and the H275Y, A204T and K207R in NA protein (Table 3). In H7N9, three mutations were related to host adaption or viral drug-resistance, including the G685R and E627K in PB2 protein, and K432E in NA protein (Table 3).

**The Haplotype Analysis of H1N1 and H7N9 Virus Quasispecies**

We achieved 25 and 22 haplotypes for H1N1 and H7N9 genomic segments in all eight samples based on combining high diversity nucleotide loci in the same genomic sequences, respectively. For
instance, we obtained seven haplotypes of NA in H1N1 based on four high diversity nucleotide loci (Table 4). The bases on high diversity nucleotide loci of each haplotype and its abundance in each sample were listed in Supplementary Table S4. The composition and dynamics of viral quasispecies in this patient co-infected with H1N1 and H7N9 IAV were displayed along the clinical treatment process (Figure 3). In NA gene, the haplotype Hap_2 replaced Hap_1 as the dominant haplotype with more than half proportion (53.11%, 1,119/2,107) in sample S2, which contain H275Y drug-resistant mutation. However, Hap_2 in NA failed to be continuously dominant and Hap_1 return the dominant haplotype in following samples (S3 to S6) (Figure 3A). Similarly, we found Hap_3 of HA with antigenic drift mutation R222K transiently become dominant in sample S6 (56.19%, 127/226) (Figure 3A). Compared with the genomic segment integrity of H1N1 viral quasispecies among almost all samples, The H7N9 quasispecies had a poor integrity except for the first sample S1 (Figure 3B). In conclusion, the haplotype provides a more concise overview of viral quasispecies composition and dynamics than whole genomic sequences (Figure 2 and Figure 3).

FIGURE 1 | The process of sequences grouping and the SMRT sequencing error rate. (A) There are 69 group sequences clustered from four SMRT cells by samples, subtypes, and genomic segments. The sequences in one group indicates that they are with the same sample resource, the same influenza subtype, and the same genomic RNA segments. The light yellow and pink figures around each sample indicate the number of groups of H1N1 and H7N9. (B) The base sequencing error rates of 58 group sequences. In each group, the number of sequences is more than 20. The rate is the ratio of the number of sequencing error bases to the number of total bases in one group sequences. (C) The three types and their percentage of sequencing error bases. Red indicates mismatch, blue indicates deletion, and green indicates insertion. The percentage is the ratio of the number of one type of sequencing error bases to the number of total sequencing error bases in one group sequences. The detailed values about sequencing error bases in each group are listed in Supplementary Table S2.
Figure 2] The abundance and diversity of H1N1 and H7N9 genomic sequences. In H1N1 (A) and H7N9 (B), from left to right, three subgraphs respectively indicate the number of sequence reads, the number of sequence types and the sequence diversity index in each group. In each subgraph, eight colors represent eight genomic segments of influenza A virus. One sequence type is defined that there are one or more different bases from any other sequence. The sequence diversity index is calculated as the Shannon entropy formula described in the “Materials and Methods". The detailed values about abundance and diversity of genomic sequences are listed in Supplementary Table S3.

Table 3 | The high-diversity nucleotide loci in H1N1 and H7N9 genomes.

| Subtype | Segment | Nucleotide loci | Base composition a | Amino acid b |
|---------|---------|----------------|------------------|-------------|
|         |         |                | A    | C    | G    | T    |        |
| H1N1    | PB2     | 2045           | 124  | 0    | 608  | 0    | G673D  |
|         | PB1     | 99             | 0    | 38   | 0    | 138  | T25T   |
|         | PA      | 182            | 0    | 58   | 0    | 484  | F53S   |
|         | HA      | 86             | 148  | 3,071| 0    | 0    | D18E   |
|         | HA      | 697            | 127  | 0    | 3,092| 0    | R222K  |
|         | NP      | 133            | 155  | 0    | 6,391| 0    | G30R   |
|         | NP      | 237            | 101  | 0    | 6,445| 0    | E64E   |
|         | NA      | 147            | 127  | 5,288| 0    | 0    | G43K   |
|         | NA      | 630            | 107  | 0    | 5,308| 0    | A204T  |
|         | NA      | 640            | 5,280| 0    | 135  | 0    | K207R  |
|         | NA      | 843            | 0    | 4,217| 0    | 1,198| H275Y  |
|         | M       | 427            | 663  | 0    | 11,809| 0   | R134R in M1 |
|         | M       | 835            | 107  | 0    | 12,385| 0   | W41* in M2 |
|         | NS      | 260            | 9,591| 0    | 114  | 0    | R78R in NS1 |
|         | NS      | 705            | 100  | 0    | 9,605| 0    | L69L in NEP |
|         | NS      | 705            | 100  | 0    | 9,605| 0    | L69L in NEP |
| H7N9    | PB2     | 1906           | 35   | 0    | 31   | 0    | E627K  |
|         | PB2     | 2080           | 10   | 0    | 56   | 0    | G685R  |
|         | PB2     | 2,286          | 52   | 0    | 14   | 0    | R753R  |
|         | HA      | 722            | 43   | 0    | 269  | 0    | G234D  |
|         | HA      | 1,163          | 228  | 0    | 104  | 0    | G331R  |
|         | NA      | 661            | 76   | 0    | 0    | 385  | F206I  |
|         | NA      | 339            | 165  | 76   | 0    | 0    | R107S  |
|         | NA      | 448            | 215  | 0    | 26   | 0    | T144A  |
|         | NA      | 1,312          | 168  | 0    | 73   | 0    | K432E  |
|         | NS      | 523            | 0    | 101  | 0    | 457  | L166P in NEP |
|         | NS      | 707            | 0    | 0    | 497  | 61   | S70I in NEP |

A: The number of base A/C/G/T in one nucleotide locus.
B: The site and types of amino acid corresponding to base composition.
Synonymous mutations.
*: Termination codon.
### TABLE 4 | The forming and abundance of NA haplotypes in H1N1.

| Haplotype | Loci and bases | Abundance in each samplea |
|-----------|----------------|----------------------------|
|           | 147 630 640 843 | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | Totalb |
| hap_1     | C G A C         | 282 (100%) | 980 (46.51%) | 1,228 (91.99%) | 460 (73.23%) | 710 (90.91%) | 322 (100%) | 0 | 0 | 3,982 |
| hap_2     | C G A T         | 0 | 1,119 (53.11%) | 0 | 0 | 126 (21.43%) | 0 | 0 | 0 | 1,190 |
| hap_3     | A G G C         | 0 | 0 | 0 | 107 (8.01%) | 0 | 0 | 0 | 0 | 126 |
| hap_4     | C A A C         | 0 | 0 | 0 | 0 | 1 (0.17%) | 0 | 0 | 0 | 1 |
| hap_5     | C G G T         | 0 | 8 (0.38%) | 0 | 0 | 0 | 0 | 0 | 0 | 8 |
| hap_6     | C G G C         | 0 | 0 | 0 | 1 (0.17%) | 0 | 0 | 0 | 0 | 1 |
| hap_7     | A G A C         | 0 | 0 | 0 | 1 (0.17%) | 0 | 0 | 0 | 0 | 1 |

aThe number and percentage of haplotype in each sample.
bThe number of haplotypes in all samples.

### FIGURE 3 | The composition and dynamics of H1N1 and H7N9 virus quasispecies. In H1N1 (A) or H7N9 (B), eight subgraphs respectively eight genomic segments of influenza A virus. In each subgraph, different colors indicate different haplotypes. In the same subgraph, the same color represents the same haplotype, while in different subgraphs, same color is irrelevant. The forming and abundance of each haplotype are listed in Supplementary Table S4.


DISCUSSION

Four mutations in H1N1 and three in H7N9 were related to viral drug-resistance, host adaption or evolution. The H275Y in NA protein of H1N1 was the widely investigated drug-resistant mutation against oseltamivir as the commonly used first-line drug for the treatment or prophylaxis of influenza (Hurt et al., 2012; Vidaña et al., 2020). Another two mutations A204T and K207R in NA protein of H1N1 were recently reported to have effects on drug-resistance and vaccine efficacy (Nandihini and Sista, 2020; Skowronski et al., 2020). The antigenic drift mutation R222K in the HA protein was believed to play a role in virus evolution by altering receptor binding specificity (Al Khatib et al., 2019). For H7N9, the mutation E627K on PB2 is a well-characterized host adaption mutation from the avian signature Glu (E) to the mammalian-adapted signature Lys (K), which have been associated with enhanced polymerase activity, high virus replication and pathogenicity in humans (Baccam et al., 2001). Besides, the mutation G685R on PB2 also help to promotes the mammalian adaptation of avian influenza virus (Baccam et al., 2003; Capitán et al., 2011). Interestingly, the mutation K432E on NA, alone or together with mutation H275Y on NA, had a significant impact on the binding pattern and affinity of oseltamivir for neuraminidase, rendering neuraminidase less susceptible (Aguirre and Manrubia, 2008).

The viral quasispecies as a viral population plays a very important role in the process of viral infection, adaption, and evolution through complex cooperative or competitive interactions (Domingo et al., 2012). A population of viruses can be partitioned into subpopulations by the genetic similarities (Baccam et al., 2001; Baccam et al., 2003). The spatial interactions of subpopulations have effects on host cell availability and defense responses (Aguirre and Manrubia, 2008; Capitán et al., 2011). A specific cooperative interaction is that mixed populations of D151 and G151 variants in H3N2 viruses grow better than pure populations of either variant, in which one subpopulation is good at entering new cells, while the other is better at exiting cells to spread the infection (Xue et al., 2016). In our case, the co-existent viral population of H1N1 and H7N9 might help to H7N9 subpopulation migration from URT to LRT and growth in LRT. The H1N1 subpopulation can grow in both URT and LRT of this patient, while H7N9 grow better in LRT than URT, which is related to the different distribution of α-2,3-SA and α-2,6-SA receptors in URT and LRT, as well as the preferential recognition of H1N1 and H7N9 with two receptors (de Graaf and Fouchier, 2014; Byrd-Leotis et al., 2017). The H7N9 subpopulation was easier to transfer to the LRT with the assistance of H1N1 subpopulation in the co-existence of H1N1 and H7N9 than that only in H7N9.

Besides, the competitive interactions among viral quasispecies are also reported (Andino and Domingo, 2015). An example is that in co-infected cells with wild-type polioviruses at a high multiplicity of infection (MOI) and drug-resistant virus at a much lower MOI, the yield of drug-resistant virus was significantly reduced to 3–7% of the output from a single infection due to the interference of chimeric capsid formation (Crowder and Kirkegaard, 2005). We detected the drug-resistant mutation H275Y in NA protein and antigenic drift mutation R222K in HA protein. But both failed to be continuously dominant in the subsequent viral quasispecies composition. A possible explanation is that the forming of the viral particles containing mutations were interfered by normal strains with a similar mechanism illustrated in above poliovirus.

The composition, complexity and dynamic of a viral quasispecies determined its biological or medical implications, such as host range, pathogenesis, and coping with selection pressure (Roedig et al., 2011; Gregori et al., 2016; Barbezange et al., 2018). In this work, the composition and dynamics of viral quasispecies in a patient co-infected with H1N1 and H7N9 influenza A virus were clearly revealed along its treatment process using the single-molecule real-time sequencing (SMRT). Compared with the flaws of consensus genomic sequences (CGSs) and single nucleotide variants (SNVs) by short-read massively parallel sequencing (MPS), the SMRT embody the obvious advantage in investigating the complex haplotype distribution of an IAV population, especially a population coexisting with two subtype IAV. Because of a human co-infected with two subtype IAV is very rare, the single patient in this study restricted the conclusions. Fortunately, the co-existence of two or more subtype IAV in wildfowls is not rare. These works studying the composition and dynamics of viral quasispecies in wildfowls co-infected with multi subtype IAV will be conducted in future. One scarcity of SMRT is the limit of detection (LOD) not good enough to detect the low abundant IAV sequences. For example, when the Ct values of RT-PCR for IAV in samples of this study were less than 30, SMRT was hard to generate IAV genomic sequences. This is also the reason why several types of sequences were not detected in some samples and only 69 rather than 128 groups to be conducted in the study. Besides, the expensive sequencing costs is another limitation. With the optimization and upgrade of SMRT in terms of limit of detection and sequencing cost, using SRMT to reveal the composition and dynamic of influenza A virus quasispecies will become a necessary method for studying viral biological behaviors and medical implications, which will boost the understanding of viral infections, pathogenesis, evolution, and precision medicine (Nakano et al., 2017; Wenger et al., 2019b; Beaulaurier et al., 2020).

Accession Number

The data that support the findings of this study have been deposited into CNGB Sequence Archive (Guo et al., 2020)of CNGBdb (Chen et al., 2020)with accession number CNP0001131.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://db.cnbg.org/search/project/CNP0001131/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of BGI. The patients/
participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XY and ZS collected samples and performed experimental detection. PL, TJ, and LL performed single molecule real-time sequencing. PL, LL and GL wrote bioinformatics pipeline. PL, TJ, JP, and GF designed study and analyzed data. PL, GL, ZY, DJ, JP, and GF contributed to writing this study. All authors had full access to the final version of the manuscript and agreed to its submission.

FUNDING

This study was supported by Shenzhen Science and Technology Research and Development Project (No. JCYJ20151029151932602).

REFERENCES

Aguirre, I., and Manrubia, S. C. (2008). Effects of Spatial Competition on the Diversity of a Quasispecies. Phys. Rev. Lett. 100 (3), 038106. doi:10.1103/PhysRevLett.100.038106

Al Khatib, H. A., Al Thami, A. A., Gallouzi, L., and Yassine, H. M. (2019). Epidemiological and Genetic Characterization of pH1N1 and H3N2 Influenza Viruses Circulated in MENA Region during 2009-2017. BMC Infect. Dis. 19 (1), 314. doi:10.1186/s12879-019-3930-6

Ali, R., Blackburn, R. M., and Kozlakidis, Z. (2016). Next-Generation Sequencing and Influenza Virus: A Short Review of the Published Implementation Attempts. HAYATI J. Biosciences 23 (4), 155–159. doi:10.1016/j.hjb.2016.12.007

Andino, R., and Domingo, E. (2015). Viral Quasispecies. Virology 479-480, 46–51. doi:10.1016/j.virology.2015.03.022

Argris, M., Alteri, C., Battisti, A., Di Carlo, D., Minichini, C., Sagnelli, C., et al. (2016). Multiple Hepatitis B Virus (HBV) Quasispecies and Immune-Escape Mutations Are Present in HBV Surface Antigen and Reverse Transcriptase of Patients with Acute Hepatitis B. J. Infect. Dis. 213 (12), 1897–1905. doi:10.1093/infdis/jiw049

Ardui, S., Ameer, A., Vermesch, J. R., and Hestand, M. S. (2018). Single Molecule Real-Time (SMRT) Sequencing Comes of Age: Applications and Utilities for Medical Diagnostics. Nucleic Acids Res. 46 (5), 2159–2168. doi:10.1093/nar/gky066

Baccam, P., Thompson, R. J., Fedrigo, O., Carpenter, S., and Cornette, J. L. (2001). PAQ: Partition Analysis of Quasispecies. Bioinformatics 17 (1), 16–22. doi:10.1093/bioinformatics/17.1.16

Baccam, P., Thompson, R. J., Li, Y., Sparks, W. O., Behshan, M., Dorman, K. S., et al. (2003). Subpopulations of Equine Infectious Anemia Virus Rev Coexist In Vivo and Differ in Phenotype. J. Virol. 77 (22), 12122–12131. doi:10.1128/JVI.77.22.12122-12131.2003

Barbezange, C., Jones, L., Blanc, H., Isakov, O., Celniker, G., Enouf, V., et al. (2018). Free Single-Molecule Sequencing Recover Complete Virus Genomes from Natural Microbial Communities. Genome Res. 30 (3), 437–446. doi:10.1101/gr.251686.119

Beerenwinkel, N., Günthard, H. F., Roth, V., and Metzner, K. J. (2012). Challenges and Opportunities in Estimating Viral Genetic Diversity from Next-Generation Sequencing Data. Front. Microbiol. 3, 329. doi:10.3389/fmicb.2012.00329

Bi, Y., Chen, Q., Wang, Q., Chen, J., Jin, T., Wong, G., et al. (2016). Genesis, Evolution and Prevalence of H5N6 Avian Influenza Viruses in China. Cell Host & Microbe 20 (6), 810–821. doi:10.1016/j.chom.2016.10.022

Bi, Y., Li, L., Li, G., Chen, Q., Jin, T., Zhang, C., et al. (2020). Dominant Subtype Switch in Avian Influenza Viruses during 2016–2019 in China. Nat. Commun. 11 (1), 5909. doi:10.1038/s41467-020-19671-3

The funders had no role in the study design, data collection and interpretation, or the decision to submit the study for publication.

ACKNOWLEDGMENTS

We thank staff of Hangzhou First People’s Hospital and Hangzhou Xixi Hospital for their clinical care given to the patient and facilitating access to the relevant medical records. We also thank China National GeneBank for storing the data from this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.754445/full#supplementary-material

Boktor, S. W., and Hafner, J. W. (2019). Influenza. Available from https://www.ncbi.nlm.nih.gov/books/NBK459363/.

Bonomo, M. E., Kim, R. Y., and Deem, M. W. (2019). Modular Epitope Binding Predicts Influenza Quasispecies Dominance and Vaccine Effectiveness: Application to 2018/19 Season. Vaccine 37 (24), 3154–3158. doi:10.1016/j.vaccine.2019.03.068

Bull, R. A., Eltahlil, A. A., Rodrigo, C., Koekkoek, S. M., Walker, M., Pirzoyan, M. R., et al. (2016). A Method for Near Full Length Amplification and Sequencing for Six Hepatitis C Virus Genotypes. BMC Genomics 17, 247. doi:10.1186/s12864-016-2575-8

Byrd-Leotis, L., Cummings, R. D., and Steinhauser, D. A. (2017). The Interplay between the Host Receptor and Influenza Virus Hemagglutinin and Neuraminidase. Int. J. Mol. Sci. 18 (7). doi:10.3390/ijms18071541

Capitán, J. A., Cuesta, J. A., Manrubia, S. C., and Aguirre, J. (2011). Severe Hindrance of Viral Infection Propagation in Spatially Extended Hosts. PLoS One 6 (8), e23358. doi:10.1371/journal.pone.0023358

Chaisson, M. J., and Tesler, G. (2012). Mapping Single Molecule Sequencing Reads Using Basic Local Alignment with Successive Refinement (BLASR): Application and Theory. BMC Bioinformatics 13, 238. doi:10.1186/1471-2105-13-238

Chen, F. Z., You, L. J., Yang, F., Wang, L. N., Guo, X. Q., Gao, F., et al. (2020). CNGBdb: China National GeneBank DataBase. Yi Chuan 42 (8), 799–809. doi:10.16288/j.yczz.2020.00320

Chen, J., Zhao, Y., and Sun, Y. (2018). De Novo haplotype Reconstruction in Viral Quasispecies Using Paired-End Read Guided Path Finding. Bioinformatics 34 (17), 2927–2935. doi:10.1093/bioinformatics/bty202

Chen, X., Liu, S., Goraya, M. U., Maarouf, M., Huang, S., and Chen, J.-L. (2018). Host Immune Response to Influenza A Virus Infection. Front. Immunol. 9, 320. doi:10.3389/fimmu.2018.00320

Chen, Y., Liang, W., Yang, S., Wu, N., Gao, H., Sheng, J., et al. (2013). Human Infections with the Emerging Avian Influenza A H7N9 Virus from Wet Market Poultry: Clinical Analysis and Characterisation of Viral Genome. The Lancet 381 (9881), 1916–1925. doi:10.1016/s0140-6736(13)60903-4

Crooks, G. E., and Brenner, S. E. (2004). Protein Secondary Structure: Entropy, Correlations and Prediction. Bioinformatics 20 (10), 1603–1611. doi:10.1093/bioinformatics/bth132

Crowder, S., and Kirkgaard, K. (2005). Trans-dominant Inhibition of RNA Viral Replication Can Slow Growth of Drug-Resistant Viruses. Nat. Genet. 37 (7), 701–709. doi:10.1038/ng1583

de Graaf, M., and Fouchier, R. A. M. (2014). Role of Receptor Binding Specificity in Influenza A Virus Transmission and Pathogenesis. EMBO J. 33 (8), 823–841. doi:10.1002/embo.201387442

Domingo, E., and Perales, C. (2019). Viral Quasispecies. PloS Genet. 15 (10), e1008271. doi:10.1371/journal.pgen.1008271
Wenger, A. M., Peluso, P., and Rowell, W. J. (2019a). Highly-accurate Long-Read Sequencing Improves Variant Detection and Assembly of a Human Genome. Nat. Biotechnol. 37, 1155. doi:10.1038/s41587-019-0217-9

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., et al. (2019b). Accurate Circular Consensus Long-Read Sequencing Improves Variant Detection and Assembly of a Human Genome. Nat. Biotechnol. 37 (10), 1155–1162. doi:10.1038/s41587-019-0217-9

Xiong, X., Martin, S. R., Haire, L. F., Wharton, S. A., Daniels, R. S., Bennett, M. S., et al. (2013). Receptor Binding by an H7N9 Influenza Virus from Humans. Nature 499 (7459), 496–499. doi:10.1038/nature12372

Xu, Y., Peng, R., Zhang, W., Qi, J., Song, H., Liu, S., et al. (2019). Avian-to-Human Receptor-Binding Adaptation of Avian H7N9 Influenza Virus Hemagglutinin. Cell Rep. 29 (8), 2217–2228. doi:10.1016/j.celrep.2019.10.047

Xue, K. S., Hooper, K. A., Ollodart, A. R., Dingens, A. S., and Bloom, J. D. (2016). Cooperation between Distinct Viral Variants Promotes Growth of H3N2 Influenza in Cell Culture. Elife 5, e13974. doi:10.7554/eLife.13974

Xue, Y., Wang, M. J., Yang, Z. T., Yu, D. M., Han, Y., Huang, D., et al. (2017). Clinical Features and Viral Quasispecies Characteristics Associated with Infection by the Hepatitis B Virus G145R Immune Escape Mutant. Emerg. Microbes Infect. 6 (3), e15. doi:10.1038/emi.2017.2

Zhu, Y., Qi, X., Cui, L., Zhou, M., and Wang, H. (2013). Human Co-infection with Novel Avian Influenza A H7N9 and Influenza A H3N2 Viruses in Jiangsu Province, China. Lancet 381 (9883), 2134. doi:10.1016/S0140-6736(13)61135-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MG declared a past co-authorship with one of the authors GF to the handling editor.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Lin, Jin, Yu, Liang, Liu, Jovic, Sun, Yu, Pan and Fan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.