

## Supplementary Material

### 1. Deep Sequencing methods

Viral RNA was extracted (59) without the addition of carrier RNA and either (i) directly sequenced or (ii) PCR-amplified across the CS region. For (i), double-stranded cDNA was prepared using the cDNA synthesis system, according to manufacturer's instructions (Roche, UK). For (ii), RNA was reverse-transcribed into double-stranded cDNA using Uni 12 primer 5'-AGCAAAAGCAGG-3' at a final concentration 5 $\mu$ M and MMLV-RT enzyme (Promega, UK), followed by the FastStart High Fidelity PCR system (Roche, UK) with H7 CS-specific primers (59) modified to contain the Illumina specific overhang sequencing adapters at a final concentration 0.4 $\mu$ M each:

Forward: 5'-

**TCGTCCGGCAGCGTCAGATGTGTATAAGAGACAGCGTGCAAGTTTTCTGA**

GAGG-3' and reverse: 5'-

**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACCTTCCCATCCA**

TTTTCA-3'.

PCR cycling conditions: 95°C for 2 minutes; followed by 35 cycles of: 95°C for 30 seconds, 50°C for 30 seconds, 72°C for 2 minutes; with a final extension at 72°C for 5 minutes.

A tailored Python script was written to select sequence reads from raw fastq files starting with 'CCCGAAATCCCAAAG' (PEIPK) and ending with

'AGAGGCCTATTT' (RGLF). The resulting output report recorded the total number of reads sequenced and the total number of reads that spanned the entire CS based on the defined motifs. A breakdown of all the nucleotide sequences found between the two defined starting and ending motifs were given with the number of reads for each variant and the proportions of each variant found.

## **2. Establishing background sequencing limitations**

RG viruses and pol I HA plasmids were sequenced to establish the frequency of background variants generated by experimental and technical error. Firstly, separate sequencing was done using (i) the H7N7-RG viruses' RNA, sequenced multiple times, and (ii) from amplicons of the CS region to obtain independent replicates for comparison. The percentage of correct CS reads obtained from sequencing RNA from the RG viruses directly and from sequencing the CS PCR products ranged from 99.08%-100% and 98.79%-99.69% respectively (Table S1). These frequencies were compared to error levels derived from sequencing pol I HA plasmid DNA that was used to generate the RG viruses to determine an overall background error level. The percentage of correct CS reads obtained from sequencing the plasmids directly ranged from 98.08%-99.82% (Table S2). The percentage of correct CS reads obtained from sequencing the CS PCR targeted products had a narrower range between 99.33%-99.75% (Table S2). The maximum error level obtained from sequencing RNA was 0.92% compared to 1.21% from the PCR products. More variation was obtained when sequenced directly from the plasmids (max. 1.92%) compared to sequencing from PCR products derived

from the plasmids (max. 0.67%). The most conservative error level derived, 1.92%, was applied to sequenced samples derived from *in vitro* and *in ovo* passage experiments. The greater depth offered by enriching for the CS region by PCR, without introducing additional sequencing errors above that seen for direct sequencing from plasmids, was preferred for this CS variant analysis.

**Table S1. Deep sequence results from H7N7 RG viruses sequenced from RNA or from DNA following CS amplification by PCR.** Samples were sequenced up to three times independently.

| RG virus sequenced     | RNA         |           |                      | CS PCR      |           |                       |
|------------------------|-------------|-----------|----------------------|-------------|-----------|-----------------------|
|                        | Total reads | CS reads  | Correct CS reads (%) | Total reads | CS reads  | Correct CS reads (%)  |
| H7N7 <sub>SB</sub> (1) | 2,211,399   | 168       | 167<br>(99.40%)      | 270,140     | 254,759   | 253,688<br>(99.58%)   |
| H7N7 <sub>SB</sub> (2) | 5,899,294   | 1102      | 1098<br>(99.64%)     | 224,607     | 213,992   | 213,319<br>(99.69%)   |
| H7N7 <sub>DB</sub> (1) | 4,128,249   | 537       | 537<br>(100%)        | 86,514      | 81,115    | 80,542<br>(99.29%)    |
| H7N7 <sub>DB</sub> (2) | 8,631,446   | 3,030     | 3,013<br>(99.44%)    | 78,011      | 73,306    | 72,418<br>(98.79%)    |
| H7N7 <sub>DB</sub> (3) | <i>nd</i>   | <i>nd</i> | <i>nd</i>            | 3,242,056   | 2,898,069 | 2,884,787<br>(99.54%) |
| H7N7 <sub>MB</sub> (1) | 6,800,855   | 8,797     | 8,716<br>(99.08%)    | <i>nd</i>   | <i>nd</i> | <i>nd</i>             |

*nd; not done*

**Table S2. Deep sequence results from pol I HA plasmids.** Plasmids

directly sequenced three or four times, or amplified by PCR targeting the CS

sequenced in two independent runs.

| Sample                              | DNA         |          |                      | CS PCR      |           |                      |
|-------------------------------------|-------------|----------|----------------------|-------------|-----------|----------------------|
|                                     | Total reads | CS reads | Correct CS reads (%) | Total reads | CS reads  | Correct CS reads (%) |
| pPol I H7N7<br>HA <sub>SB</sub> (1) | 18,606,348  | 893,333  | 888,843<br>(99.50%)  | 213,630     | 204,587   | 203,834<br>(99.63%)  |
| pPol I H7N7<br>HA <sub>SB</sub> (2) | 371,387     | 19,506   | 19,457<br>(99.75%)   | 254,220     | 246,307   | 245,701<br>(99.75%)  |
| pPol I H7N7<br>HA <sub>SB</sub> (3) | 118,685     | 6,330    | 6,252<br>(98.77%)    | <i>nd</i>   | <i>nd</i> | <i>nd</i>            |
| pPol I H7N7<br>HA <sub>SB</sub> (4) | 238,977     | 12,715   | 12,512<br>(98.40%)   | <i>nd</i>   | <i>nd</i> | <i>nd</i>            |
| pPol I H7N7<br>HA <sub>DB</sub> (1) | 292,564     | 14,425   | 14,346<br>(99.45%)   | 108,129     | 104,406   | 104,026<br>(99.64%)  |
| pPol I H7N7<br>HA <sub>DB</sub> (2) | 297,650     | 14,815   | 14,788<br>(99.82%)   | 83,053      | 78,633    | 78,107<br>(99.33%)   |
| pPol I H7N7<br>HA <sub>DB</sub> (3) | 247,054     | 12,608   | 12,495<br>(99.10%)   | <i>nd</i>   | <i>nd</i> | <i>nd</i>            |
| pPol I H7N7<br>HA <sub>MB</sub> (1) | 168,902     | 8,230    | 8,072<br>(98.08%)    | 256,161     | 247,045   | 245,986<br>(99.57%)  |
| pPol I H7N7<br>HA <sub>MB</sub> (2) | 184,109     | 8,556    | 8,493<br>(99.26%)    | 267,024     | 257,429   | 256,150<br>(99.51%)  |
| pPol I H7N7<br>HA <sub>MB</sub> (3) | 233,390     | 10,809   | 10,719<br>(99.17%)   | <i>nd</i>   | <i>nd</i> | <i>nd</i>            |

*nd; not done*

**Table S3. Cleavage site variant analysis results from H7N7<sub>DB</sub> E3, E10 and E14 embryo passaged viruses.** All samples sequenced from PCR products unless specified.

| Sample                                          | Total reads | CS reads  | MB reads (%)       | DB reads (%)       | SB reads (%)  | Other reads (%) |
|-------------------------------------------------|-------------|-----------|--------------------|--------------------|---------------|-----------------|
| <b>(A) H7N7<sub>DB</sub> E3</b>                 |             |           |                    |                    |               |                 |
| Tissue homogenate                               | 770,590     | 644,580   | 341,971 (53.10%)   | 280,420 (43.50%)   | 8,068 (1.25%) | 14,121 (2.15%)  |
| CEF cell supernatant                            | 710,736     | 587,308   | 349,490 (59.51%)   | 221,071 (37.64%)   | 4,813 (0.82%) | 11,934 (2.03%)  |
| Allantoic fluid <sup>a</sup>                    | 816,760     | 687,744   | 222,003 (32.28%)   | 440,088 (63.99%)   | 9,680 (1.41%) | 15,973 (2.32%)  |
| <b>(B) H7N7<sub>DB</sub> E10</b>                |             |           |                    |                    |               |                 |
| Tissue homogenate                               | 7,904,568   | 6,541,266 | 4,529,484 (69.24%) | 1,935,289 (29.59%) | 0 (0.00%)     | 76,493 (1.17%)  |
| CEF cell supernatant                            | 721,272     | 613,267   | 598,803 (97.64%)   | 3,147 (0.51%)      | 0 (0.00%)     | 11,317 (1.85%)  |
| Allantoic fluid <sup>a</sup>                    | 1,743,746   | 1,447,041 | 1,368,801 (94.59%) | 48,985 (3.39%)     | 0 (0.00%)     | 29,255 (2.02%)  |
| Allantoic fluid <sup>a,b</sup>                  | 4,820,626   | 1,920     | 1,794 (93.44%)     | 58 (3.02%)         | 0 (0.00%)     | 68 (3.54%)      |
| <b>(C) H7N7<sub>DB</sub> E14</b>                |             |           |                    |                    |               |                 |
| FFPE head and membranes                         | 463,370     | 412,148   | 109,297 (26.52%)   | 299,934 (72.77%)   | 1 (0.00%)     | 2,916 (0.71%)   |
| FFPE body                                       | 1,669,997   | 1,469,994 | 1,454,394 (98.94%) | 245 (0.02%)        | 0 (0.00%)     | 15,355 (1.04%)  |
| <b>H7N7<sub>DB</sub> E3</b><br>Allantoic fluid  | 441,963     | 378,665   | 522 (0.14%)        | 361,119 (95.37%)   | 7,449 (1.97%) | 9,575 (2.52%)   |
| <b>H7N7<sub>DB</sub> E10</b><br>Allantoic fluid | 5,745,780   | 5,078,863 | 0 (0.00%)          | 5,057,280 (99.58%) | 0 (0.00%)     | 21,583 (0.42%)  |

<sup>a</sup>Allantoic fluid after inoculation with CEF supernatant, <sup>b</sup> Sequenced from RNA

**Table S4. Cleavage site variant analysis results from H7N7<sub>SB</sub> embryo passaged viruses.** All samples sequenced from PCR products unless specified.

| Sample                                               | Total reads | CS reads | MB reads (%)   | DB reads (%)    | SB reads (%)        | Other reads (%) |
|------------------------------------------------------|-------------|----------|----------------|-----------------|---------------------|-----------------|
| <b>Passage 1:<br/>Tissue homogenates</b>             |             |          |                |                 |                     |                 |
| <b>A</b> H7N7 <sub>SB</sub> E6                       | 323,943     | 272,940  | 65<br>(0.02%)  | 1452<br>(0.53%) | 268,777<br>(98.47%) | 2646<br>(0.98%) |
| <b>B</b> H7N7 <sub>SB</sub> E12                      | 532,669     | 457,607  | 169<br>(0.04%) | 1124<br>(0.25%) | 452,935<br>(98.98%) | 3379<br>(0.73%) |
| <b>C</b> H7N7 <sub>SB</sub> E14                      | 821,784     | 703,135  | 413<br>(0.06%) | 1992<br>(0.28%) | 695,175<br>(98.87%) | 5555<br>(0.79%) |
| <b>D</b> H7N7 <sub>SB</sub> E15                      | 486,025     | 404,222  | 126<br>(0.03%) | 1796<br>(0.44%) | 397,986<br>(98.46%) | 4314<br>(1.07%) |
| <b>Passage 2:<br/>H7N7<sub>SB</sub> E14, E1</b>      |             |          |                |                 |                     |                 |
| <b>E</b> Tissue homogenate                           | 818,251     | 726,213  | 54<br>(0.01%)  | 1295<br>(0.18%) | 719,701<br>(99.10%) | 5163<br>(0.71%) |
| <b>F</b> Tissue homogenate*                          | 3,310,812   | 2,248    | 0<br>(0.00%)   | 0<br>(0.00%)    | 2,239<br>(99.60%)   | 9<br>(0.40%)    |
| <b>G</b> FFPE tissue: body                           | 885,890     | 770,455  | 79<br>(0.01%)  | 1334<br>(0.17%) | 764,110<br>(99.18%) | 4932<br>(0.64%) |
| <b>Passage 2:<br/>H7N7<sub>SB</sub> E14, E11</b>     |             |          |                |                 |                     |                 |
| <b>H</b> Tissue homogenate                           | 726,953     | 404,301  | 24<br>(0.01%)  | 0<br>(0.00%)    | 402,886<br>(99.65%) | 1391<br>(0.34%) |
| <b>I</b> Tissue homogenate*                          | 4,361,339   | 444      | 0<br>(0.00%)   | 2<br>(0.45%)    | 441<br>(99.32%)     | 1<br>(0.23%)    |
| <b>Passage 3:<br/>H7N7<sub>SB</sub> E14, E11, E9</b> |             |          |                |                 |                     |                 |
| <b>J</b> Tissue homogenate                           | 745,233     | 650,514  | 212<br>(0.03%) | 2053<br>(0.32%) | 643,260<br>(98.88%) | 4989<br>(0.77%) |
| <b>K</b> Tissue homogenate*                          | 4,663,085   | 499      | 0<br>(0.00%)   | 2<br>(0.40%)    | 496<br>(99.40%)     | 1<br>(0.20%)    |
| <b>L</b> FFPE tissue: head and membranes             | 213,403     | 184,356  | 18<br>(0.01%)  | 1222<br>(0.66%) | 181,477<br>(98.44%) | 1639<br>(0.89%) |
| <b>M</b> FFPE tissue: body                           | 789,557     | 671,886  | 284<br>(0.04%) | 1878<br>(0.28%) | 664,775<br>(98.94%) | 4949<br>(0.74%) |

\*Samples sequenced from RNA