Supplemental Data
Networks of Genomic Co-occurrence Capture Characteristics of Human Influenza A Evolution

Xiangjun Du, Zhuo Wang, Aiping Wu, Lin Song, Yang Cao, Haiying Hang, Tajiao Jiang*

Figure S1. Evolutionary relationships of influenza A MP, NS, NP, PA, PB1 and PB2 from influenza A viruses sampled from season 1997 to 2006.

The clades for strains from 2000 to 2006 were labeled using the season from which majority of the isolates come. MP evolves very slowly in K from season 1998 to 2004 and its K only changed remarkably from 2004 to 2005 (Figure 3A), which is consistent with the existence of a long side lineage for 7 seasons from 1998 to 2004 in its phylogenetic tree analysis. Evolution of NS appears more complicated during 1999-2000 season (Figure 3A), where H3N2 isolates of 2000-2001 split into two distant clades: the majority clade has a closer phylogenetic relationship with those in 2005-2006 rather than those in 2002-2004; the minor clade came from those in 1999 (shown in the phylogenetic tree analysis of NS). NP, PA, PB1 and PB2 have similar evolving topology patterns as NA has (Figure 3B). In addition to a significant network topology change from season 2004 to 2005, they have another one from season 2001 to 2002, suggesting the emergence of new side lineages in season 2002 and 2005 in their phylogenetic tree analysis.

Figure S2. Heat map of the distribution of transient and transition modules from 1968 to 2006 subject to hierarchical clustering based on the strain–module profile.

Rows correspond to transient and transition modules, and columns correspond to the indicated season. Transient modules are labelled as predominant (green) or not predominant (gray). Transition modules are labelled according to which of the two member modules is predominant (first member, green; second member, red). The conserved predominant modules not shown in the figures are 0,1,2,5,6,8,13,16,18,21,22,26,29,31,34,38,41,43,45,48,50,52,53,54,56,58,59,65,66,69,71,72,75,77,79,80,81,86,89,91,92,93,94,95,97,98,99,100,101,103,104,107,110,111,112,113,114,117,118,119,120,121,122,124,126,128,129,131,132,133,136,137,138,139,141,143,147,148,149,154,155,156,158,159,160,164,165,167,169,170,176,178,181,183,184,186,187,189,190,191,194,197,198,199,201,202,206,208,210,212,213,216,217,219,227,228,229,230,231,232,233,234,236,237,238,240,242,244,245,247,249,250,251,252,253,254,255,257,258,259,261,263,264,268,269,270,271,274,276,277,278,279,280.

Figure S3. Module-based amino acid changes corresponding to the nucleotide changes in transient and transition modules from 1968-2006.

Each row represents a single amino position in a viral protein that is labelled on the right of the figure. Amino acids (single-letter abbreviations) are also color-coded so

that mutations can be seen as changes in both amino acid identity and color.

Figure S4. **Comparison of connectivity and HA sequence clusters**

Sequence clusters for 1032 HA sequences are obtained using method developed by Plotkin et al (Plotkin et al. 2002). Different threshold d (d=2, 3, 4, 5, 6, 8) were attempted. For each threshold d, we first exclude those clusters consist of a single sequence. Then we assign remaining clusters to seasons using season information of their member sequences. Finally, we find sequence clusters, which match the most of corresponding seasons for each antigenic cluster, and assign same color with antigenic clusters (Smith et al. 2004). Grey areas indicate seasons out of best-matched clusters', while blank spaces indicate absence of sequences in those seasons among our dataset. Through counting right assignment according to antigenic clusters for both connectivity and sequence clusters, we can see that sequence clusters with threshold d=4(star bold, 18 out of 32 total clusters contain only single sequence and 9 of the remaining 14 clusters match most for corresponding antigenic clusters), which match antigenic clusters best (68%, 21 out 32 season), has a remarkable correspondence (66%, 23 out of 35 season) with connectivity clusters. In contrast, the connectivity clusters matches the antigenic clusters up to 84%.

Figure S5. Comparison of the season-wise average connectivity changes (ΔK) and the rate of network changes (S) of simulated H3N2 networks of nucleotide co-occurrence from 1968-2006 over time.

The ΔK from season m to n is calculated by subtracting the average connectivity of simulated H3N2 networks of nucleotide of co-occurrence in season n to that in season m. Similar to R, the S from season m to n, is computed as:

$$S(m,n) = \sum_{i=m+1}^{n} X_i \; / \sum_{i=m}^{n-1} Y_i$$

where i is time in seasons, X_i is the number of nucleotide pairs changed when a network evolves from season i-1 to season i, and Y_i is the number of total edges of a network in season i.

References

Plotkin, J.B., J. Dushoff, and S.A. Levin. 2002. Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. *Proc Natl Acad Sci U S A* **99:** 6263-6268.

Smith, D.J., A.S. Lapedes, J.C. de Jong, T.M. Bestebroer, G.F. Rimmelzwaan, A.D. Osterhaus, and R.A. Fouchier. 2004. Mapping the antigenic and genetic evolution of influenza virus. *Science* **305:** 371-376.

Figure S1

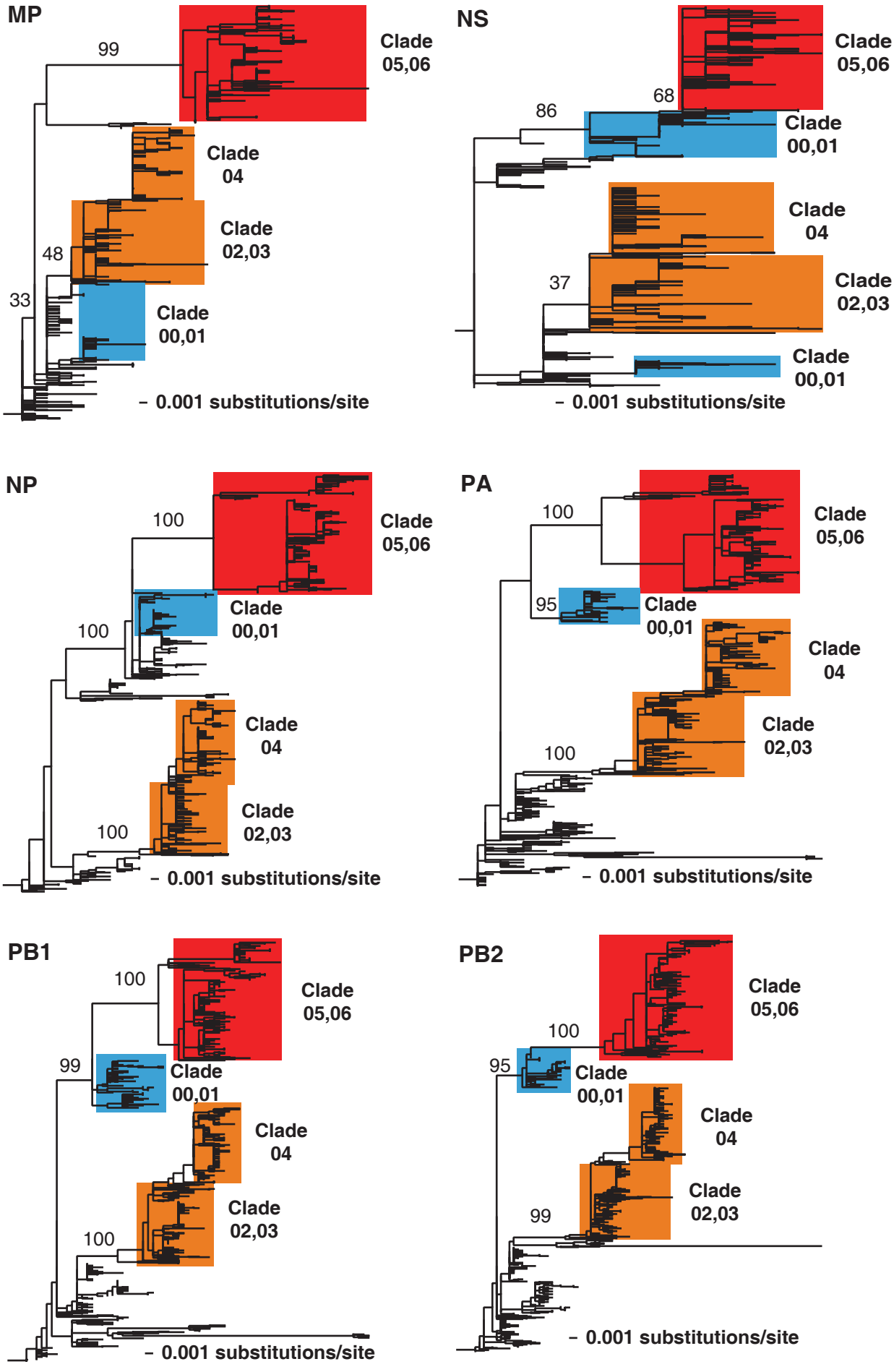
MP
99
Clade 05,06
Clade 04
Clade 02,03
48
33
Clade 00,01
0.001 substitutions/site
NS
68
86
Clade 05,06
Clade 00,01
Clade 04
37
Clade 02,03
Clade 00,01
0.001 substitutions/site
NP
100
Clade 05,06
Clade 00,01
100
Clade 04
Clade 02,03
100
0.001 substitutions/site
PA
100
Clade 05,06
95
Clade 00,01
Clade 04
Clade 02,03
100
0.001 substitutions/site
PB1
100
Clade 05,06
99
Clade 00,01
Clade 04
Clade 02,03
100
0.001 substitutions/site
PB2
100
Clade 05,06
95
Clade 00,01
Clade 04
Clade 02,03
99
0.001 substitutions/site

Figure S2

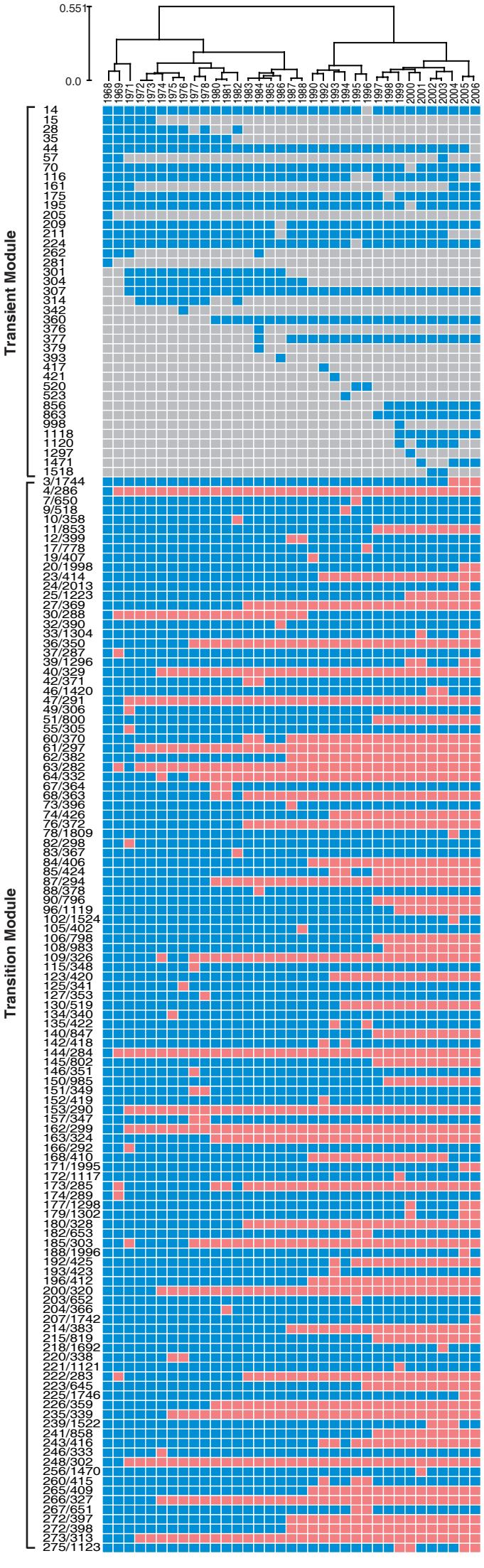
0.551
0.0
1968 1969 1990 1971 1972 1973 1974 1975 1976 1977 1978 1980 1981 1982 1983 1984 1985 1986 1987 1988 1990 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006
Transient Module
14 15 28 35 44 57 70 116 161 175 195 205 209 211 224 262 281 301 304 307 314 342 360 376 377 379 393 417 421 520 523 856 863 998 1118 1120 1297 1471 1518
Transition Module
3/1744 4/286 7/650 9/518 10/358 11/853 12/399 17/778 19/407 20/1998 23/414 24/2013 25/1223 27/369 30/288 32/390 33/1304 36/350 37/287 39/1296 40/329 42/371 46/1420 47/291 49/306 51/800 55/305 60/370 61/297 62/382 63/282 64/332 67/364 68/363 73/396 74/426 76/372 78/1809 82/298 83/367 84/406 85/424 87/294 88/378 90/796 96/1119 102/1524 105/402 106/798 108/983 109/326 115/348 123/420 125/341 127/353 130/519 134/340 135/422 140/847 142/418 144/284 145/802 146/351 150/985 151/349 152/419 153/290 157/347 162/299 163/324 166/292 168/410 171/1995 172/1117 173/285 174/289 177/1298 179/1302 180/328 182/653 185/303 188/1996 192/425 193/423 196/412 200/320 203/652 204/366 207/1742 214/383 215/819 218/1692 220/338 221/1121 222/283 223/645 225/1746 226/359 235/339 239/1522 241/858 243/416 246/333 248/302 256/1470 260/415 265/409 266/327 267/651 272/397 272/398 273/313 275/1123

Figure S3

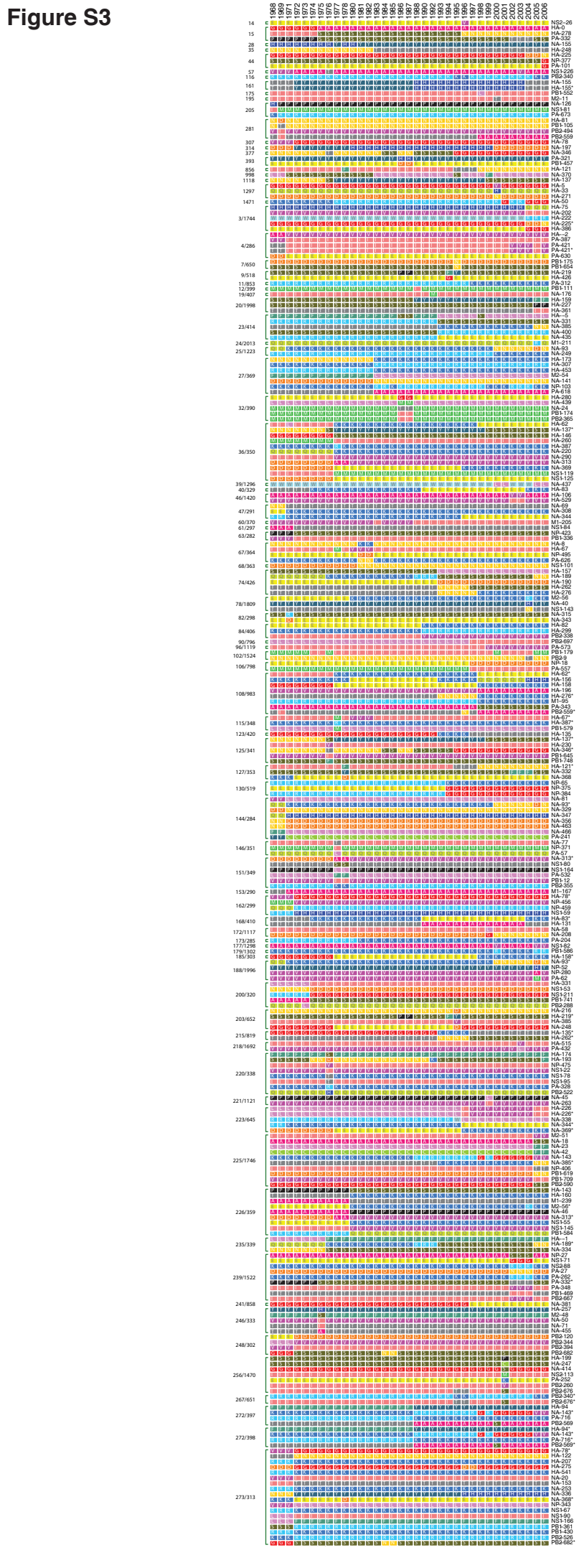
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
14
15
28
35
44
57
116
161
175
195
205
281
307
314
377
393
856
998
1118
1297
1471
3/1744
4/286
7/650
9/518
11/853
12/399
19/407
20/1998
23/414
24/2013
25/1223
27/369
32/390
36/350
39/1296
40/329
46/1420
47/291
60/370
61/297
63/282
67/364
68/363
74/426
78/1809
82/298
84/406
90/796
96/1119
102/1524
106/798
108/983
115/348
123/420
125/341
127/353
130/519
144/284
146/351
151/349
153/290
162/299
168/410
172/1117
173/285
177/1298
179/1302
185/303
188/1996
200/320
203/652
215/819
218/1692
220/338
221/1121
223/645
225/1746
226/359
235/339
239/1522
241/858
246/333
248/302
256/1470
267/651
272/397
272/398
273/313
NS2-26
HA-6
HA-278
HA-332
NA-155
HA-248
HA-225
NP-377
PA-101
NS1-226
PB2-340
HA-156
HA-155*
PB1-552
M2-11
NA-126
NS1-81
PA-673
HA-81
PB1-105
PB2-494
PB2-559
HA-78
NA-197
NA-346
PA-321
PB1-457
HA-121
NA-370
NA-137
HA-5
NA-33
HA-271
HA-50
HA-75
HA-200
HA-222
HA-225*
NA-386
HA--2
PA-387
PA-421
PA-421*
PA-630
PB1-175
PB1-654
HA-219
HA-428
PA-212
PB1-111
NA-176
HA-159
HA-227
NA-361
HA--5
NA-331
NA-385
NA-400
NA-435
M1-211
NA-93
NA-249
HA-173
HA-307
HA-453
M2-54
NA-141
NP-103
PA-618
HA-280
HA-439
NA-24
PB1-174
PB2-365
HA-62
HA-137*
HA-146
HA-260
HA-387
HA-220
NA-290
NA-313
NA-369
NS1-119
NS1-125
NA-437
HA-63
HA-106
HA-529
NA-69
NA-308
NA-344
M1-205
NS1-84
NP-423
PB1-336
HA-9
HA-67
NP-495
PA-626
NS1-101
HA-157
HA-189
HA-190
HA-262
HA-276
M2-56
NA-40
NS1-143
NA-315
NA-343
HA-82
NA-299
PB2-338
PB2-697
PA-573
PB1-179
PB2-9
NP-18
PA-557
HA-62*
HA-156
HA-158
HA-196
HA-276*
M1-95
PA-343
PB2-559*
HA-67*
HA-387*
PB1-579
HA-135
HA-137*
HA-230
HA-346*
PB1-545
PB1-748
HA-121*
NA-332
NA-368
NP-85
NP-375
NP-384
NA-81
NA-93*
NA-329
NA-347
NA-356
NA-463
NA-466
PA-241
NA-77
NP-371
PA-57
NA-313*
NS1-80
NS1-164
PA-532
PB1-12
PB2-305
M1-167
HA-78*
NP-456
NP-459
NS1-59
HA-83*
HA-131
NA-58
NA-208
PA-204
NS1-82
PB1-586
HA-158*
HA-93*
NP-52
NP-280
PA-62
NA-331
NS1-53
NS1-211
PB1-741
PB2-288
HA-216
HA-219*
NA-385
NA-248
HA-135*
HA-262*
PA-515
PA-432
HA-174
HA-193
NP-475
NS1-22
NS1-78
NS1-195
NA-326
PB2-522
NA-49
NA-283
HA-226
HA-226*
NA-338
NA-344*
NA-369*
M2-51
NA-18
NA-23
NA-42
NA-143
NA-385*
NP-406
PB1-619
PB1-709
PB2-590
HA-143
NA-160
M1-239
M2-56*
NA-46
NA-312*
NS1-155
NS1-145
PB1-584
HA--1
NA-189*
NA-334
NP-27
NS1-171
NS2-88
PA-27
PA-262
PA-332*
PA-348
PB1-469
PB2-667
NA-381
HA-257
M2-48
NA-50
NA-71
NA-455
PB2-100
PB2-344
PB2-394
PB2-682
NA-199
NA-247
NA-414
NS2-113
PA-252
PB2-260
PB2-676
PB2-340*
PB2-676*
HA-94
NA-143*
PA-716
PB2-569
HA-94*
NA-143*
PA-716*
PB2-569*
HA-78*
HA-122
HA-207
HA-275
HA-541
NA-20
NA-153
NA-253
NA-336
NA-368*
NA-343
NS1-67
NS1-90
NS1-166
PB1-361
PB1-430
PB2-526
PB2-682*

Figure S4

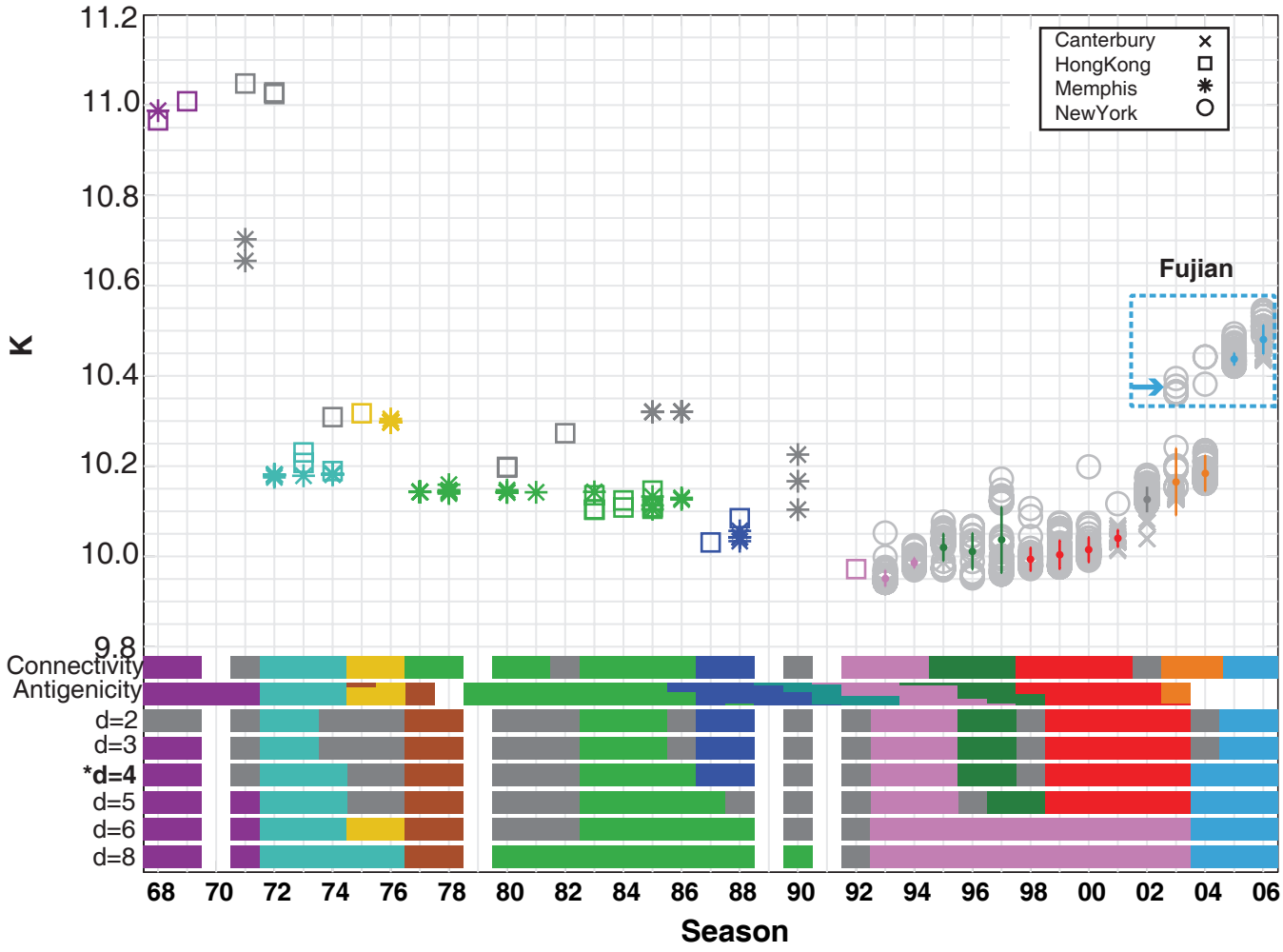

Figure S5