

Sequencing of clinical samples reveals that adaptation keeps establishing during H7N9 virus infection in humans

Liqiang Li^{1*}, Jinmin Ma^{1,4#}, Jiandong Li^{1#}, Jianying Yuan^{1#}, Wei Su^{1,6#}, Tao jin^{1#}, Xinfu Wang², Renli Zhang³, Rongrong Zou², Lei Li¹, Jianming Li², Shisong Fang³, Jing Yuan², Chentao Yang¹, Yanwei Qi^{1,6}, Qi Gao^{1,6}, Jingkai, Ji^{1,5,6}, Kailong Ma¹, Guangyi Fan¹, Na pei¹, Yong Deng², Yang Zhou², Dechun Lin^{1,5,6}, Fei Li^{1,5,6}, Wenjie Ouyang¹, Huijue Jia¹, Xin Liu¹, Hui Jiang¹, Huanming Yang^{1,7}, Xun Xu¹, Hui Wang^{1,2,8*}, Yingxia Liu^{2*}

¹ BGI-Shenzhen, Shenzhen 518083, China; ² Shenzhen Third People's Hospital, Shenzhen, China; ³ Major Infectious Disease Control Key Laboratory, the Shenzhen Center for Disease Control and Prevention, Shenzhen; ⁴ Department of Biology, University of Copenhagen, Copenhagen, Denmark; ⁵ College of Life Science, University of Chinese Academy of Sciences, Beijing, China; ⁶ BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China; ⁷ James D. Watson Institute of Genome Science, Hangzhou, China; ⁸ Department of Zoology, University of Oxford, Oxford, UK.

These authors contributed equally to this work.

* *Correspondence and requests for materials should be addressed to* Liqiang Li:

liliqiang@genomics.cn, Yingxia Liu: yingxialiu@hotmail.com and Hui Wang: wanghui1@genomics.cn.

1 **The H7 subtype avian influenza viruses (AIV) have a much longer history**
2 **and their adaptation through evolution pose continuous threat to humans ¹.**
3 **Since 2013 March, the novel reassorted H7N9 subtype have transmitted to**
4 **humans through their repeated assertion in the poultry market. Through**
5 **repeated transmission, H7N9 gradually became the second AIV subtype posing**
6 **greater public health risk after H5N1 ^{2,3}. After infection, how the virus tunes its**
7 **genome to adapt and evolve in humans remains unknown. Through direct**
8 **amplification of H7N9 and high throughput (HT) sequencing of full genomes**
9 **from the swabs and lower respiratory tract samples collected from infected**
10 **patients in Shenzhen, China, we have analyzed the *in vivo* H7N9 mutations at the**
11 **level of whole genomes and have compared with the genomes derived by *in vitro***
12 **cultures. These comparisons and frequency analysis against the H7N9 genomes**
13 **in the public database, 40 amino acids were identified that play potential roles in**
14 **virus adaptation during H7N9 infection in humans. Various synonymous**
15 **mutations were also identified that might be crucial to H7N9 adaptation in**
16 **humans. The mechanism of these mutations occurred in a single infection are**
17 **discussed in this study.**

18

19 Earlier study have shown that after just one or two normal passages
20 propagation in chick embryos, human Influenza virus A differs sharply from the *in*
21 *vivo* ones ⁴. Not only were tissue tropism and binding properties altered, but the

1 mutations responsible also gave rise to detectable changes to their antigenic
2 characteristics. Influenza A virus culture-based mutation phenomenon were also
3 reported by several studies ⁵⁻⁸. Since the first outbreak of H7N9, full genomes of
4 various isolates derived human have been sequenced. Genomes of most of the human
5 H7N9 viruses were obtained after *in vitro* propagation in embryonated chicken eggs.
6 Ren *et al.* also found that the consensus genomes from the original human patients
7 and the embryonated eggs culture show differences¹². Previous studies have shown
8 that H7N9 viruses show mutations during infecting the ferret ⁹⁻¹¹. Over time, ranging
9 from 12.6-40 days ¹³, H7N9 virus in infected human host should have tuned their
10 genetic information for adapting to novel mammalian host. We hypothesize that
11 characterizing the genomes of H7N9 viruses *in vivo*, other than *in vitro* cultures can
12 help to find out the mutations that might be playing roles during adaption in the host,
13 and observing the intra-host characteristics..

14 To verified this thought, we analyzed samples derived from H7N9 infected
15 patients admitted to the Third People's Hospital of Shenzhen (TPH-SZ). Nasal and
16 pharyngeal swabs and phlegm samples from four patients (one mild and three severe
17 patients) were collected and subjected to HT sequencing with and without
18 embryonated chicken egg propagation (Supplementary Table 1, 2; Supplementary
19 discussion) (Phase I). Surprisingly, pairwise alignment of the above 4 pair H7N9
20 isolates showed that consensus sequences changed for the same sample before and
21 after *in vitro* culture, with identities varying from 0% to 4.8% on each segment

1 (Figure 1a). The polymerase encoding segments, segment 1-3, maintained high
2 identity following culture, when compared to the other five segments (5-8) (Figure 1a;
3 supplementary table 3). The genetic variation in the NP and NS was the largest,
4 showing rapid diversification, consistent with the previous findings in human at
5 macro-evolution level ¹⁴. One pair (2014S4/cul.2014S4) maintained all 8 consensus
6 genome segments well, with the exception of PA and M segments, which showed
7 changing only 0.1% and 0.6% variation, respectively (Figure 1a; supplementary table
8 3). The phylogenetic relationship based on the later five genome segments were also
9 changed when compared against the isolated derived from *in vitro* egg culture
10 (Supplementary figure 1). These sharply changes from *in vitro* propagation were also
11 verified by compared our data in this study with cultured Shenzhen human H7N9 and
12 avian H7N9 isolates report in Lam *et al.* (Supplementary discussion).

13 For the three high-mutated pairs, on average about 1 out of 4 mutations were
14 non-synonymous, 3 out of 4 were synonymous (Supplementary table 4). Surprisingly,
15 most of these mutations were convergent (appeared at least in two isolates)
16 (Supplementary table 4). Through depth frequency analysis, we found that most of
17 these mutations were largely dominant (depth frequency >99%, supplementary table
18 6,7), suggesting that these mutations can be established following just one passage (or
19 once infection in human). From neutral evolution theory and by comparing with
20 egg-cultured isolates in parallel, we found that the H7N9 NP, NS1/2, M1, and HA2
21 suffered the strong purification effect, while the polymerase complex components

1 PB2, PB1, PA, and the two surface proteins HA1 and M2 are in a steady neutral
2 evolution during host shift (supplementary table 4,5).

3 Further, non-synonymous mutations, e.g. amino-acid level changes were
4 observed in 75 residues compared with the egg cultures ones. 33 of them (41.3%)
5 showed convergent (Figure 1b). We postulate that these 33 residues might be more
6 adaptable in humans and play a vital role during establishment and evolution of
7 infection in humans. To verify this, additional 35 isolates were further analyzed by
8 HT sequencing from the H7N9-infecting patients collected from the 2014 and 2015
9 pandemic season (Phase II). Data of 18 isolates directly from patients, 6 isolates of
10 egg cultured human H7N9 (the 2013-2014 season) were obtained (Supplementary
11 table 2). Overall, all 33 residues occurred in human *in vivo* H7N9 viruses
12 (SZ_*in-vivo*-H7N9, n=22) at a higher frequency (>30%) than in cultured human
13 H7N9 viruses (SZ_*in-vitro*-H7N9, n=10), with 28 residues showing statistics
14 significance (Figure 2a). Between 2013-2014 and 2014-2015 endemic seasons (11 vs
15 11) of SZ_*in-vivo*-H7N9, occurring frequencies showed no statistics differences
16 except at 89S (M2) but all significant higher than SZ_*in-vitro*-H7N9, ruling out the
17 influence of genetic background (Figure 2a). We also compared the occurrence
18 frequency of these 33 shift residues between these SZ_*in-vivo*-H7N9 with the avian
19 and human H7N9 sequences deposited in the GenBank (GB_*avian*-H7N9 and
20 GB_*human*-H7N9) since 2013 in China. 22 residues in SZ_*in-vivo* H7N9 showed
21 higher occurrence frequencies than in GB_*avian*-H7N9 (n=~400), and 19, included in

the above 22 than GB_human-H7N9 (n~80), and 21, included in the above 22 than SZ_*in-vitro*-H7N9 with statistical significance (Figure 2b). Unexpectedly, most (22~23) of these 33 residues occurring frequencies showed no significant differences among the GB_avian-H7N9, GB_human-H7N9 and SZ_*in-vitro*-H7N9, except the residues in NP and NS1 segments (Figure 2b).

Besides the above convergent mutations in four paired samples, we further identified other variable sites between SZ_*in-vivo*-H7N9 and SZ_*in-vitro*-H7N9, and then compared the mutations of these sites with GB_avian-H7N9 and GB_human-H7N9. Among the 55 variable residue sites, though only 1 site (57Q of PA) in SZ_*in-vivo*-H7N9 showed significant higher frequency than SZ_*in-vitro*-H7N9 (Figure 2c), mount to 18 of the 55 residue sites showed higher occurrence frequencies than in GB_avian-H7N9 and 13 of them were higher than GB_human-H7N9 (Figure 2d). Simliar to the convergent mutations identified in phase I, most (9-12) of 13 residues showed no differences among GB_avian-H7N9 and GB_human-H7N9 and SZ_*in-vitro*-H7N9. Because most viruses of GB_human-H7N9 were derived through *in vitro* cultures, we postulate that human H7N9 have reversed to avian H7N9 status after *in vitro* culture, a trait not explained before. This sequence reversal could be attributed to the viral adaptation in the embryonated eggs.

Collectively, the above results reveal that the *in vitro* culture has considerable influence on the H7N9 at the genome, viral proteins. 40 residue sites (Supplementary table 8) were identified in a high frequency (27-100%, Figure 2) in human *in vivo*

1 H7N9 virus. All these mutations were also verified by Illumina Hiseq 2500
2 sequencing (data not shown). Most of these mutations are not identified playing
3 certain function in previous studies. Primary structure mapping and tetra-structure
4 remodeling analysis were done to assess the functional influences of these mutations.
5 Two surface glycoproteins, HA and NA contain the most mutations, amount to 7 on
6 HA, and 6 on NA. In HA, all mutations are located in HA1: 3 (114K, 177I and 186A)
7 were located near the receptor-binding pocket (Figure 3), 2 (88V, 273I) in the esterase
8 subdomain , and 2 (276D, 300K) in the fusion domain. Due to their functional
9 locations, these mutations could play a significant biological role in HA stability, cell
10 fusion and/or receptor-binding specificity or affinity. In NA, all 6 (78K, 178A, 320S,
11 345I, 396T, 431G) mutations were located on the head domain and close to the
12 C-terminal, with 78K located near the stalk and trans-membrane region. These
13 mutations might play important roles in HA-NA interaction and virion stability.

14 When AIV jumps to a new host, sustainable and stable replication is its first key
15 steps to survival, adaptation and establishment of infection ¹⁵. For the virion
16 replication complex, ribonucleoproteins (RNP), mutations mostly occurred at the
17 domains involved in RNP protein interaction (Figure 2; supplementary figure 3),
18 might playing an important role in effective and stable replication establishment in
19 human. It is know that PB2 plays key roles during host adaption. 570M mutation was
20 identified located in the CAP binding domain and near the C-terminal NP and PB1
21 interaction domain. Especially, at residue 627 and 701, besides the well-known host

1 adaption mutation 627K, 701N, two novel mutations, 627V and 701E, were also
2 identified occurring at high frequency in *in-vivo* H7N9 respectively (Figure 2d),
3 suggesting might be adapted in human host. In PB1, 2 mutations were identified, one
4 located near the N terminal and the other in the PB2 interaction domain near the
5 C-terminus (supplementary figure 3). As for the PA, 4 mutations were identified, 57Q
6 localized at the N terminal, two in (237K)/near (262K) the NLS domain, and 444D at
7 the middle of the PA. As for the NP, 4 mutations were identified, all localized in the
8 NP-NP interaction domain, with 3 of them belonging to the NP-PB2 interaction
9 domain implying roles in NP-NP complex stability in virus replication
10 (supplementary figure 3).

11 It is believed that the viral matrix protein M1 plays an important role during RNP
12 nuclear transportation ¹⁶, NS1 is the main viral antagonist of the innate immune
13 response during influenza virus infection to overcome the first barrier the host
14 presents to halt the viral infection ¹⁷. 5 mutational changes were identified in the
15 matrix protein, 2 (232N, 248M) were located at the RNP interaction domain on M1, 2
16 (18K and 24E) were closely located at the N terminal of M2 with 24E belonging to
17 the RNP interaction domain, and 1 (89S) was located close to the N terminal of M2
18 (Supplementary figure 3). The distributions of these mutations are consistent with the
19 possible role they may have in viral RNP complex transport, viral packaging and
20 budding. 3 mutations were identified in the NS1/NEP, 2 of which resided in the NS1
21 with 27M resident at N-terminal dsRNA/PABP1/RIGI/EIB-AP5 binding domain and

1 145V in the short intergenic region of CPSF30 and NES domain, the third mutation
2 (92S) located in C-terminal of NEP (Supplementary figure 3). These mutations might
3 play roles in RNP transportation or in escaping the host immune response.

4 In recent years, synonymous mutations, correlating with the codon usage bias
5 issues, were believed to influence virus fitness and pathogenicity through translation
6 efficiency regulation and CpG and TpA dinucleotides frequency changes ¹⁸. G+C
7 content, CpG and ApT dinucleotide frequency, GC3 number and observed ENcs
8 changed during host shift (supplementary figure 5). Through phase I four paired
9 sample comparison, we found that, unlike the non-synonymous mutations in which
10 A-C transversions (tv) occurred at a much higher frequency besides purine (A/G) to
11 purine transitions (ts), the synonymous mutations showed that that (A/G), or
12 pyrimidine (C/T) to pyrimidine tv occurred at a frequency higher than the purine to
13 pyrimidine or pyrimidine to purine tv similar to other organisms ¹⁹(Supplementary
14 figure 4). Further, most of the synonymous mutations are convergent across different
15 pairs and not distributed randomly and uniformly on the genome segments, especially
16 one to four synonymous mutations cluster found on the HA, NA, M1 and NS1
17 segments (Supplementary figure 9). Collectively, these changes suggest H7N9 virus
18 might be tune their genome to adapt to human, and might play important roles in viral
19 replication and immune escape during H7N9 infecting human.

20 To analyze how these viruses has been influenced by these accompanying
21 mutations, we compared their intra-host diversity and show that *in vitro* H7N9 viruses

1 shows less diversity than *in vivo* human H7N9 viruses suggesting bottle-neck effect to
2 have occurred during *in vitro* culture (Supplementary figure 7a), and display a
3 different diversity profile (Supplementary figure 6). Most human adapted mutations
4 might have been filtered during inoculation to the egg, the second host jump, and the
5 minority egg suitable ones quickly arise during *in vitro* culture. Intra-host diversity
6 correlation analysis showed that in human the NP and NS1/2, MP segments are
7 closely linked, otherwise in embryo chicken, the HA are closely related with MP and
8 NS1/2 (Supplementary figure 8), suggesting different intra-host fluctuation
9 mechanisms. Thus, H7N9 might have utilized the unknown molecular strategy for
10 adaptation, efficient replication and establishment in a new host.

11 In summary, we found that *in vivo* H7N9 virus in the human host are enduring a
12 genetic tune to adaption and in this context, we, identified various amino acids and
13 synonymous mutations that might play a critical role in infection. This gives us a
14 different aspect to recognize AIV infecting human or other mammal hosts and provide
15 vital clues for further experimental and functional validation *in vitro*. We also found
16 *in vitro* culture can change the intra-host diversity of *in vivo* H7N9 virus and produce
17 various mutations during a single passage. These findings provide further molecular
18 insights into functional identification of virus *in vivo* and design next generation of
19 therapeutic strategies to halt virus infection at critical stages of life cycle in humans.

20

21 **Acknowledgements**

We thank Professor Nitin K. Saxena for his constructive work on manuscript revising (Sydney Medical School, the University of Sydney). This study was supported by Shenzhen Science and Technology Research and Development projects (Grant No. JCYJ20150402111430617 and JCYJ20151029151932602), Key specialized fund for new infectious diseases in Shenzhen City (No. 201161).

Author contributions

The manuscript was written by L.Q.L., H.W., J.M.M., Y.X.L., G.F.G. and W.-C.C. Samples were collected by R.R.Z., T.J., R.L.Z., Y.M.L., S.S.F., X.F.W. Experiment and data analysis were performed by L.Q.L., J.D.L., J.Y.Y., W.S., J.M.M., H.W., C.T.Y., Y.W.Q., Y.X.L., X.L., H.J.J. The study was designed by Y.X.L., L.Q.L., H.W., X.X.

Competing financial interests

The authors declare that they have no competing financial interests.

Reference:

- 1 Belser, J. A. *et al.* Contemporary North American influenza H7 viruses possess human receptor specificity: Implications for virus transmissibility. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 7558-7563, doi:10.1073/pnas.0801259105 (2008).
- 2 Lam, T. T. *et al.* Dissemination, divergence and establishment of H7N9 influenza viruses in China. *Nature* **522**, 102-105, doi:10.1038/nature14348 (2015).
- 3 Lam, T. T. *et al.* The genesis and source of the H7N9 influenza viruses causing human

1 infections in China. *Nature* **502**, 241-244, doi:10.1038/nature12515 (2013).

2 4 Burnet, F. M. & Bull, D. H. CHANGES IN INFLUENZA VIRUS ASSOCIATED

3 WITH ADAPTATION TO PASSAGE IN CHICK EMBRYOS. *Aust J Exp Biol Med* **21**,

4 55-69 (1943).

5 5 Lin, Y. P. *et al.* Adaptation of egg-grown and transfectant influenza viruses for growth

6 in mammalian cells: selection of hemagglutinin mutants with elevated pH of membrane fusion.

7 *Virology* **233**, 402-410 (1997).

8 6 Widjaja, L., Ilyushina, N., Webster, R. G. & Webby, R. J. Molecular changes

9 associated with adaptation of human influenza A virus in embryonated chicken eggs. *Virology*

10 **350**, 137-145, doi:10.1016/j.virol.2006.02.020 (2006).

11 7 Stevens, J. *et al.* Receptor specificity of influenza A H3N2 viruses isolated in

12 mammalian cells and embryonated chicken eggs. *Journal of virology* **84**, 8287-8299,

13 doi:10.1128/JVI.00058-10 (2010).

14 8 Said, A. W. *et al.* Molecular changes associated with adaptation of equine influenza

15 H3N8 virus in embryonated chicken eggs. *The Journal of veterinary medical science / the*

16 *Japanese Society of Veterinary Science* **73**, 545-548 (2011).

17 9 Richard, M. *et al.* Limited airborne transmission of H7N9 influenza A virus between

18 ferrets. *Nature* **501**, 560-563, doi:10.1038/nature12476 (2013).

19 10 Zhang, Q. *et al.* H7N9 influenza viruses are transmissible in ferrets by respiratory

20 droplet. *Science* **341**, 410-414, doi:10.1126/science.1240532 (2013).

21 11 Watanabe, T. *et al.* Characterization of H7N9 influenza A viruses isolated from

1 humans. *Nature* **501**, 551-555, doi:10.1038/nature12392 (2013).

2 12 Ren, X. *et al.* Full genome of influenza A (H7N9) virus derived by direct
3 sequencing without culture. *Emerging infectious diseases* **19**, 1881-1884,
4 doi:10.3201/eid1911.130664 (2013).

5 13 Li, Q. *et al.* Epidemiology of human infections with avian influenza A(H7N9)
6 virus in China. *The New England journal of medicine* **370**, 520-532,
7 doi:10.1056/NEJMoa1304617 (2014).

8 14 Gorman, O. T., Bean, W. J. & Webster, R. G. Evolutionary processes in influenza
9 viruses: divergence, rapid evolution, and stasis. *Current topics in microbiology and*
10 *immunology* **176**, 75-97 (1992).

11 15 Kuiken, T. *et al.* Host species barriers to influenza virus infections. *Science* **312**,
12 394-397, doi:10.1126/science.1122818 (2006).

13 16 Eisfeld, A. J., Neumann, G. & Kawaoka, Y. At the centre: influenza A virus
14 ribonucleoproteins. *Nature reviews. Microbiology* **13**, 28-41, doi:10.1038/nrmicro3367
15 (2015).

16 17 Ayllon, J. & Garcia-Sastre, A. The NS1 protein: a multitasking virulence factor.
17 *Current topics in microbiology and immunology* **386**, 73-107, doi:10.1007/82_2014_400
18 (2015).

19 18 Martinez, M. A., Jordan-Paiz, A., Franco, S. & Nevot, M. Synonymous Virus
20 Genome Recoding as a Tool to Impact Viral Fitness. *Trends in microbiology*,
21 doi:10.1016/j.tim.2015.11.002 (2015).

1 19 Wakeley, J. The excess of transitions among nucleotide substitutions: new
2 methods of estimating transition bias underscore its significance. *Trends in ecology &*
3 *evolution* **11**, 158-162 (1996).

4 **Methods**

5 **Patients and Sample collection**

6 From Dec 18th, 2013 to April 1st, 2014, 25 H7N9 infection patients with
7 Influenza-like symptoms in Shenzhen City, Guangdong Province were admitted to the
8 Third People's Hospital of Shenzhen (TPH-SZ). In the second waves of H7N9
9 outbreak in Shenzhen, i.e. third wave in China, 13 more patients infected with H7N9
10 were confirm in Shenzhen and were admitted to TPH-SZ from Jan. 4th, 2015 to April
11 2nd, 2015. In all these cases, diagnostic and treatment decisions were made by
12 consortia of more than three panel members of clinical specialists. After admitted, we
13 immediately collected the respiratory sample for clinical confirmation of H7N9
14 infecting. Standard Real-time reverse transcription polymerase chain reaction
15 (RT-PCR) assay for H7N9 confirmation were done in the Influenza Reference
16 Laboratory of the Shenzhen municipal center for disease control (CDC) and
17 Guangdong province key laboratory of emerging infectious diseases., C_T value
18 smaller than 38 were judged as H7N9 positive, according to the Guidelines authorized
19 by China National Influenza Center of China CDC. Mild and severe cases were
20 distinguished according to IDSA/ATS criteria.¹ Severe cases met at least one of major
21 criteria (Invasive mechanical ventilation/Septic shock with the need for vasopressors)

1 or more than three minor criteria¹. Analysis of the patient's clinical samples for the
2 identification of potential pathogens was approved by the ethics committee of
3 TPH-SZ.

4 **Viral RNA preparation and high throughout Sequencing**

5 Nasopharyngeal swab, phlegm samples and some other respiratory specimens
6 were collected for diagnosis and egg culture. Nasal, oropharyngeal swabs and/or
7 tracheal aspirate samples from each patient were collected into transport medium and
8 separated into two aliquots within two hours, one for diagnostic labs and the other for
9 long term storage. For *in vivo* analysis, viral RNAs were extracted directly from
10 pharyngeal and nasal swabs or phlegm samples using Qiagen viral RNA mini kit. For
11 *in vitro* virus isolation, samples testing positive for both the H7 and N9 genes were
12 propagated in 10-day-old embryonated chicken eggs and then cultured at 37 °C for 2
13 days in a bio-safety level 3 laboratory. Viral RNA were extracted using Qiagen viral
14 RNA mini kit from allantoic fluid, and tested for H7 and N9 genes by RT-PCR assays
15 and, if positive, used for further genomic sequencing.

16 To obtain whole H7N9 genome, reserves transcriptions were done from viral
17 RNA samples using primer pair uniR_RT (5'-AGTAGAAACAAGG-3') and uniF_RT
18 (AGCGAAAGCAGG) to obtain viral full genomic cDNAs. Then PCR application
19 using specific primer pairs (Supplemental Table 2) designed for H7N9 virus and
20 Takara One-step RT-PCR kit (Takara, China) segment by segment. PCR products of 8
21 segments were quantified by 1% gel, and mixed with equal mole and then subject to

1 library construction and HT sequencing using BGI seq100 (Ion proton, Life
2 technology, USA).

3 **Consensus genome construction, comparison and phylogenetic analysis**

4 To obtain the consensus genome, we mapping all reads to a reference H7N9
5 virus with the threshold match rate at 0.80, and at each site, we choose the dominant
6 bases as final consensus genome. Sequences were aligned using MUSCLE v3.5.
7 Phylogenies were inferred on the basis of Neighbor-Joining method, by using the
8 maximum composite likelihood model in MEGA 6.0^{2,3}. Topological robustness was
9 assessed by bootstrap test. For analysis the novel mutations in human H7N9, we
10 retrieved all protein sequences of avian and human H7N9 genome segments with full
11 length ORF from GenBank (until 08 Sep. 2015).

12 **Entropy, mutation rate, amino acid mutation rate and viral fitness**

13 Intra-host diversity implication viral fitness were assessed by the average
14 Entropy or mutation rate of one genome segment or whole genome. Entropy were
15 calculated by using equation, $S = -100 * \sum (P_i * \log P_i)$ where P_i is the frequency
16 of the i th allele⁵. Rate of minor nucleotide variant allele, other than the base with the
17 largest depth, at each nucleic acid site were calculated by the minor allele depth
18 deduced the total depth of this site. Mutation rate of each nucleic acid site were
19 designate the ratio of the minor variant allele, that is adding all Rate of all Minor
20 variant allele (MR_i). Minor alleles changing the amino acid coding were designated as
21 a no-synonymous SNV. At specific codon, the sum value of each no-synonymous

1 nucleotide mutations rate compared to consensus codon were designated as the at one
2 amino acid site were add together were designated as amino acid mutation rate.

3 Codon-based test of neutrality were done using Nei-Gojobori method in MEGA
4 6³. G+C content, CpG and ApT dinucleotide frequency, GC3 number and observed
5 ENcs were calculated using SSE version 1.2⁴.

6 **Linear mapping and tertiary structure of amino acid mutations**

7 The linear primary structural maps of HA, NA, PB2, PB1, PA, NP, M1/2 and
8 NS1/NEP were derived and modified from the previous maps of Ping, J., et al.⁵.
9 Tertiary structures were generated using the PyMOL viewer. For H7, Structural maps
10 were generated using the structural files of two mature proteins, HA1 (PDB ID
11 4ln6.1.A) and HA2 (PDB ID 4bsa.1.B).

13 **References:**

- 14 1 Mandell, L. A. *et al.* Infectious Diseases Society of America/American Thoracic
15 Society consensus guidelines on the management of community-acquired pneumonia in adults.
16 *Clinical infectious diseases : an official publication of the Infectious Diseases Society of*
17 *America* **44 Suppl 2**, S27-72, doi:10.1086/511159 (2007).
- 18 2 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high
19 throughput. *Nucleic acids research* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).
- 20 3 Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular
21 Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* **30**, 2725-2729,

1 doi:10.1093/molbev/mst197 (2013).

2 4 Simmonds, P. SSE: a nucleotide and amino acid sequence analysis platform. *BMC*

3 *research notes* **5**, 50, doi:10.1186/1756-0500-5-50 (2012).

4 5 Ping, J. *et al.* Genomic and protein structural maps of adaptive evolution of human

5 influenza A virus to increased virulence in the mouse. *PLoS One* **6**, e21740,

6 doi:10.1371/journal.pone.0021740 (2011).

7

8 **Figure and table Legend**

9 **Figure 1. Comparison of the human H7N9 consensus sequence obtained from**

10 **direct clinical samples and embryonated chicken egg cultures in phase I. a),**

11 Identity of consensus nucleotide sequences between four *in vivo* H7N9 viruses and

12 their *in vitro* cultured counterparts; 8 genomic segments are list in x-axis sequentially,

13 y-axis are nucleotide sequence identities of each paired samples. b) Amino acids

14 mutations identified by the four cultured and uncultured pair human H7N9 virus

15 samples. Residues mutations occurred in more than two pairs are shaded in pink,

16 singleton mutations are shaded in blue. *, T76A, the residue T, before the number 76

17 is putative specific in avian host, and the residue A after 76 is putative specific in

18 avian host.

19 **Figure 2. Occurring frequency comparison of mutation residues among SZ_in**

20 **vivo-H7N9, SZ_in-vitro-H7N9, and GB_avian-H7N9 and GB_human-H7N9**

21 **deposit in GenBank. a), comparison of convergent mutation residues (phase I)**

1 between *SZ_in-vivo*-H7N9 (n=~22), *SZ_in-vitro* H7N9 (n=10); b), comparison of
 2 convergent mutation residues among *SZ_in-vivo*-H7N9, *SZ_in-vitro*-H7N9, and
 3 GB_avian-H7N9 (n=~99-411) and GB_human-H7N9 (n=~80-100); c), comparison of
 4 residues showing population bias between *SZ_in-vivo*-H7N9 and *SZ_in-vitro*-H7N9;
 5 d), comparison of residues showing population bias (from phase II) among
 6 *SZ_in-vivo*-H7N9, *SZ_in-vitro*-H7N9, GB_avian-H7N9 and GB_human-H7N9.
 7 Y-axis is occurring frequency (percentage) of mutation residues. Occurring frequency
 8 differences among different groups were compared using fisher exact test. Alignment
 9 files can be provided upon request.

10 **Figure 3. Structure modeling of mutation sites in HA of human H7N9.** Structural
 11 modeling were generated using the structural files of two mature proteins, HA1 (PDB
 12 ID 4ln6.1.A) and HA2 (4bsa.1.B) in PyMOL viewer.

13

14 **Supplementary material legend**

15 **Supplementary figures**

16 **Supplementary figure 1. Phylogenies relationship of hemagglutinin (a),**
 17 **neuraminidase and PB2 (b) genes of 4 H7N9 clinical-culture pairs.** Bootstrap
 18 support values (%) from 1,000 replicates are shown for selected lineages. The scale
 19 bar to the left of each tree represents the substitutions per site. Four phase I in-vivo
 20 and in vitro H7N9 pairs were indicated by colored solid circle.

21 **Supplementary figure 2. Sequence Comparison of *SZ_in-vivo*-H7N9 and**

1 **SZ_***in-vitro*-H7N9 with other Shenzhen human and avian H7N9 virus reported in

2 **Lam *et al.*** a), heatmap and cluster analysis of pairwise identity data; b), with-in

3 genetic diversity and between-group distance analysis of SZ_in-vivo-H7N9 and

4 SZ_in-vitro-H7N9 with other Shenzhen human and avian H7N9 virus reported in

5 Lam *et al.*

6 **Supplementary figure 3. Linear mapping of potential functional adaptive sites to**

7 **primary structure of H7N9 structural and functional proteins.** Boxes represent

8 viral proteins are not drawn according to the actual length ratio.

9 **Supplementary figure. 4. Comparison of transition and transversion mutation**

10 **frequency of four *in vivo* H7N9 and their *in vitro* cultured counterparts.**

11 **Supplementary figure 5. Codon usage differences of four paired *in vivo* H7N9**

12 **compared with *in vitro* H7N9 virus.** Cells indicated with red, green and grey colors

13 represent up, down and no significant change respectively, compared with *in vitro*

14 H7N9 virus (fisher exact test).

15 **Supplementary figure 6. Whole genome intra-host diversity profile of Shenzhen**

16 **H7N9 influenza viruses.** Heatmap showing nucleotide mutation frequencies of

17 genomic segment identified in SZ_in-vivo-H7N9 and SZ_in-vitro-H7N9. At each site,

18 bases other than the base with the largest depth were designated as minor alleles.

19 Nucleotide mutation frequencies were calculated by the minor allele depth deduced

20 the total depth. 8 genomic segments are ordered sequentially, corresponding to PB2,

21 PB1, PA, HA, NA, NP, M1/2, NS1/2 respectively.

Supplementary figure 7. Intra-host diversity of Shenzhen H7N9 influenza viruses.

a, b, Comparison of mean nucleotide mutation frequency of whole genome. Mean nucleotide mutation frequency for each sample was calculated by adding all nucleotide mutation frequencies together then deduced the total length sequenced by Ion proton. Mean nucleotide mutation frequency was expressed as the negative log. Virus isolated were classified into three different patient prognosis groups, SZ_ *in-vivo*-H7N9.S groups H7N9 represented patients manifesting severe symptoms but cured finally; and SZ_ *in-vivo*-H7N9.M groups represented patients manifesting mild symptoms through the H7N9 infection and recovered; SZ_ *in-vivo*-H7N9.D represented patients manifesting severe symptoms and finally dead. * $P < 0.05$, two-tailed Mann–Whitney *U*-test. c, d, Comparison of mean nucleotide mutation frequency of whole genome segment by segment.

Supplementary figure 8. Correlation analysis among different viral genomic segments of *in vivo* and *in vitro* H7N9 viruses. Pearson correlation coefficient were calculated using average entropy of each genomic segment of every sample from *in vivo* or *in vitro*.

Supplementary figure 9. Synonymous mutation distributions analysis of four paired *in vivo* and *in vitro* H7N9 viruses. On each separate figure, above slide windows analysis shows mutation distribution density, low panel shows each mutation concretely.

1 **Supplementary tables**

2 **Supplementary table 1.** Sample information of patients investigated in this study.

3 **Supplementary table 2.** High throughput sequencing profiles of H7N9 viruses in this
4 study.

5 **Supplementary table 3.** Identity of consensus sequence Human H7N9 between *in*
6 *vivo* samples and from egg cultured ones.

7 **Supplementary table 4.** Synonymous and non-synonymous mutation statistics of 4
8 paired H7N9 viruses in phase I.

9 **Supplementary table 5.** Codon-based test of neutrality for analysis of 4 paired H7N9
10 viruses in phase I.

11 **Supplementary table 6.** Base composition of no-synonymous mutated sites in 4
12 paired H7N9 viruses in phase I.

13 **Supplementary table 7.** Base composition of synonymous mutated sites in 4 paired
14 H7N9 viruses in phase I.

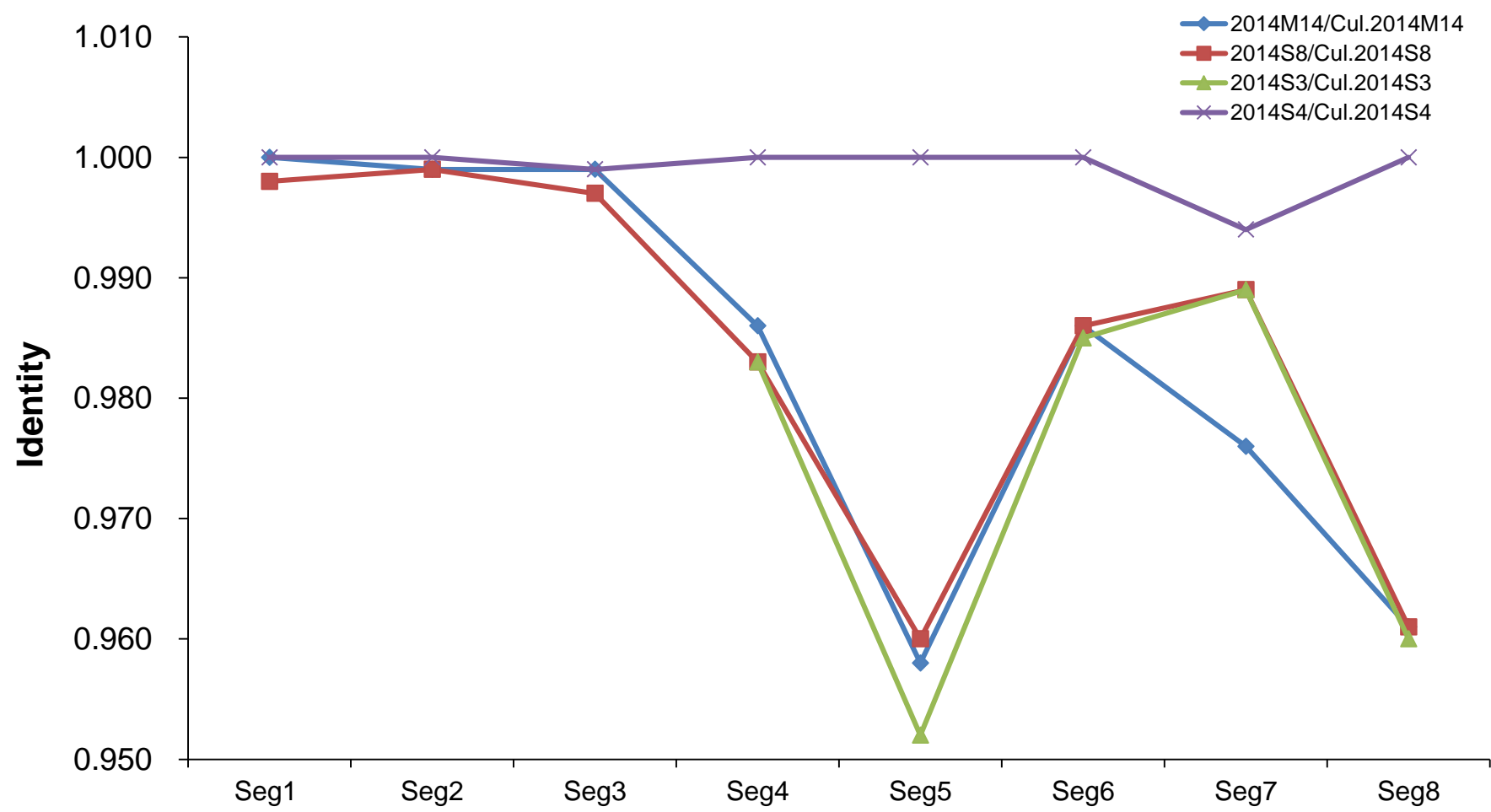
15 **Supplementary table 8.** Functional residues list of H7N9 viral proteins identified in
16 this study.

17 **Supplementary table 9.** Correlation analysis among different viral genomic segments
18 of *in vivo* and *in vitro* H7N9 viruses.

19 **Supplementary table 10.** Estimates of average evolutionary divergence within and
20 between H7N9 of different hosts.

21 **Supplementary table 11.** Primers for H7N9 genome amplification in this study.

Figure 1a



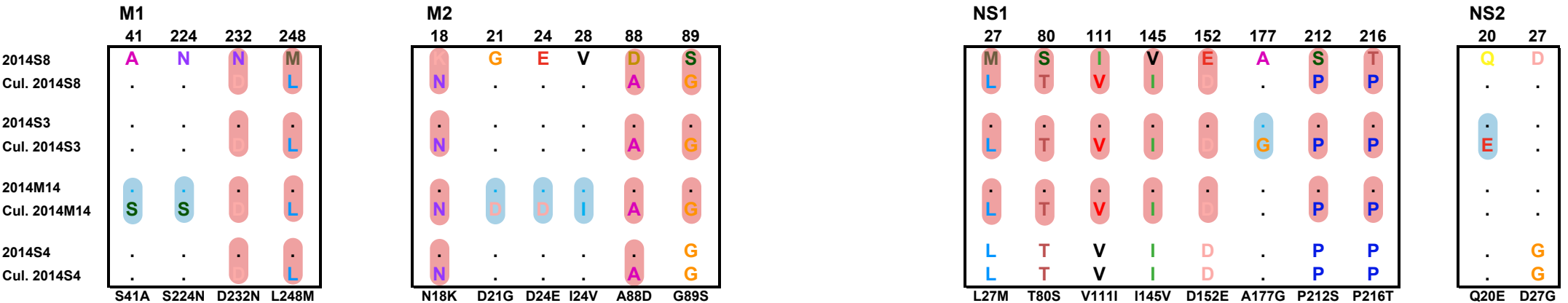
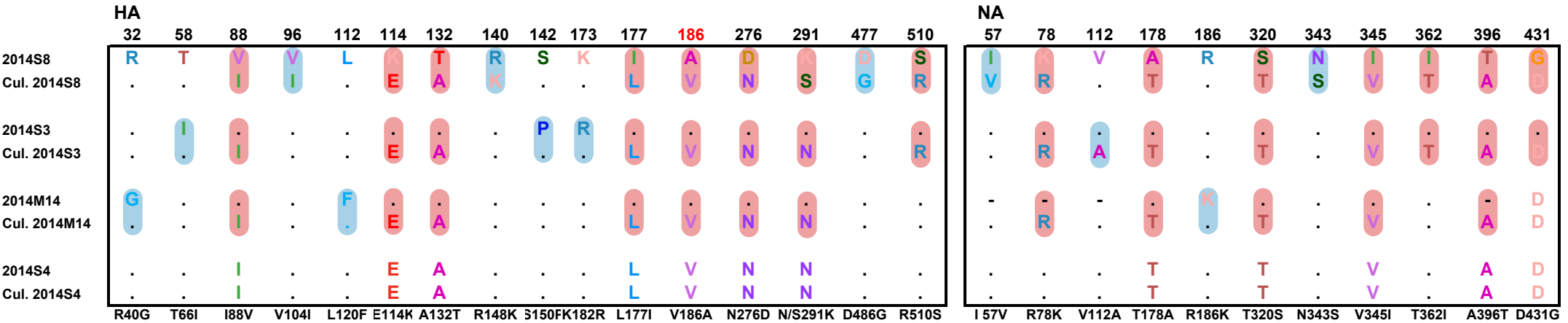
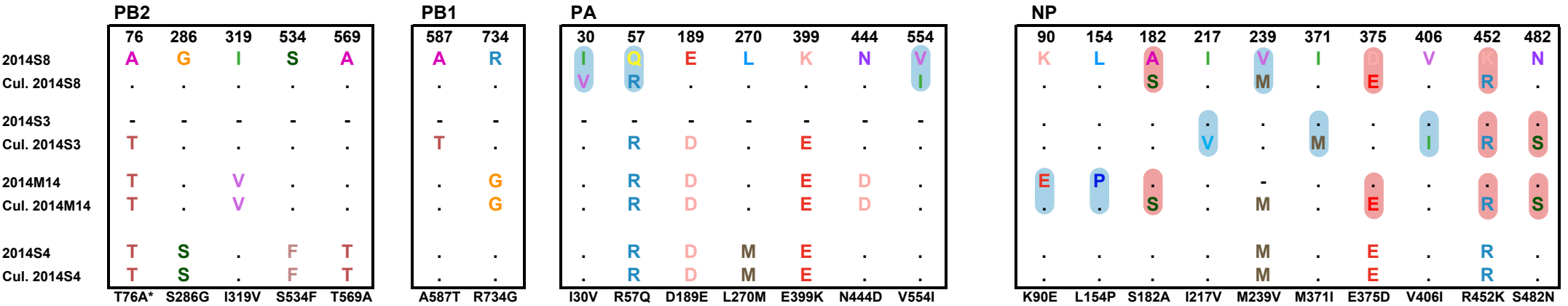


Figure 2

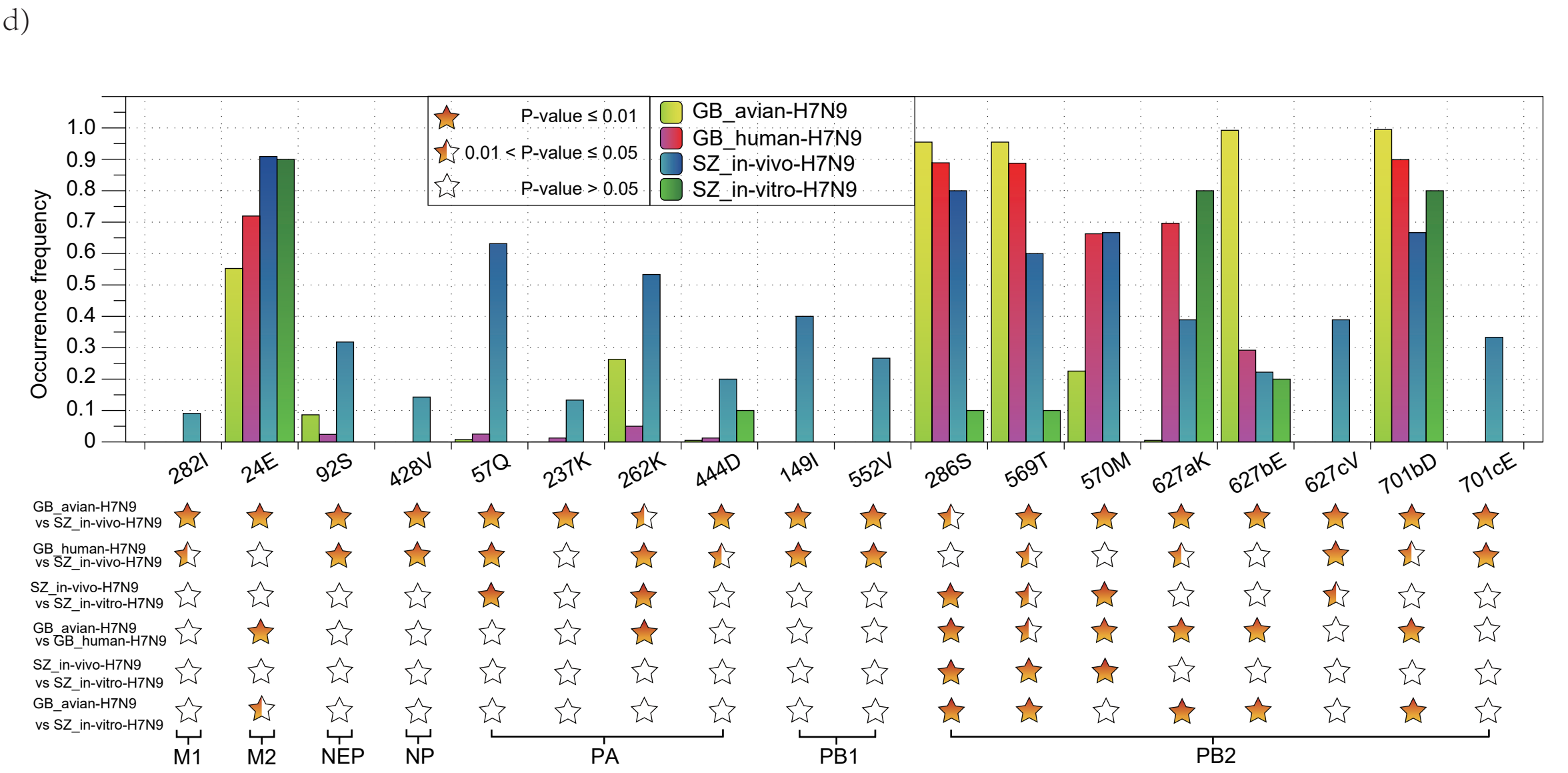
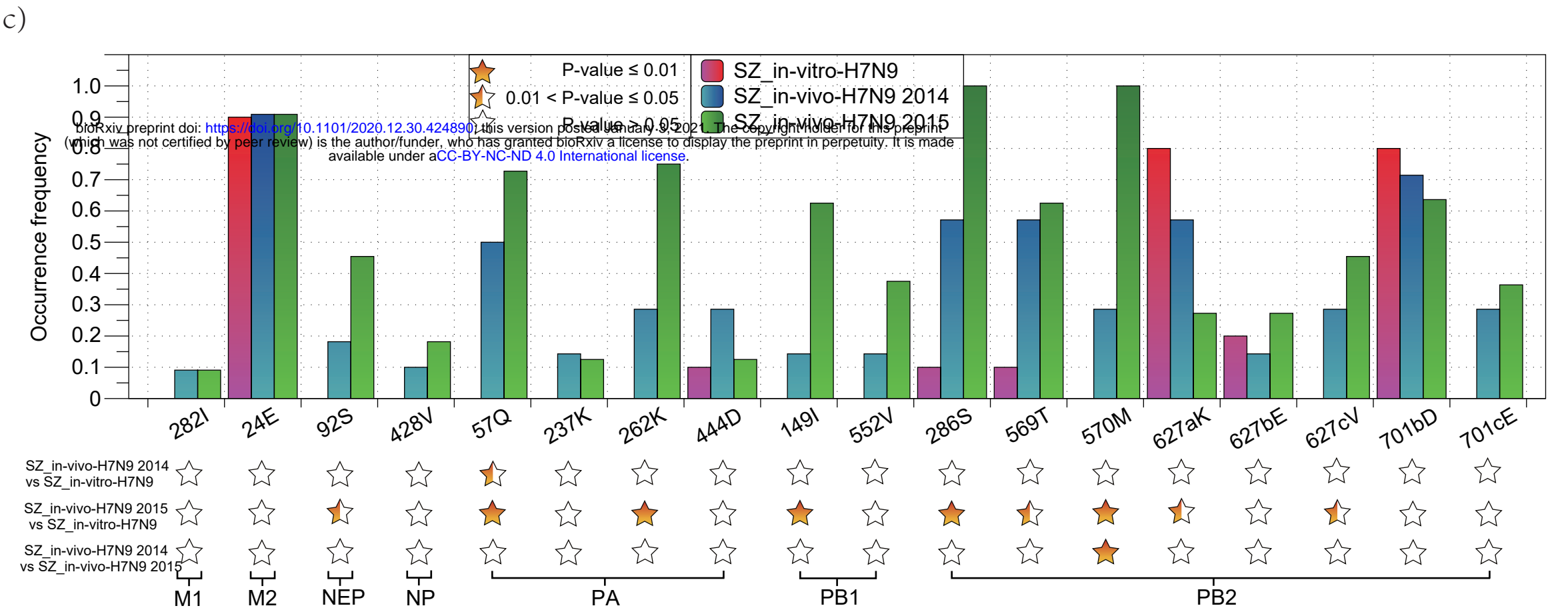
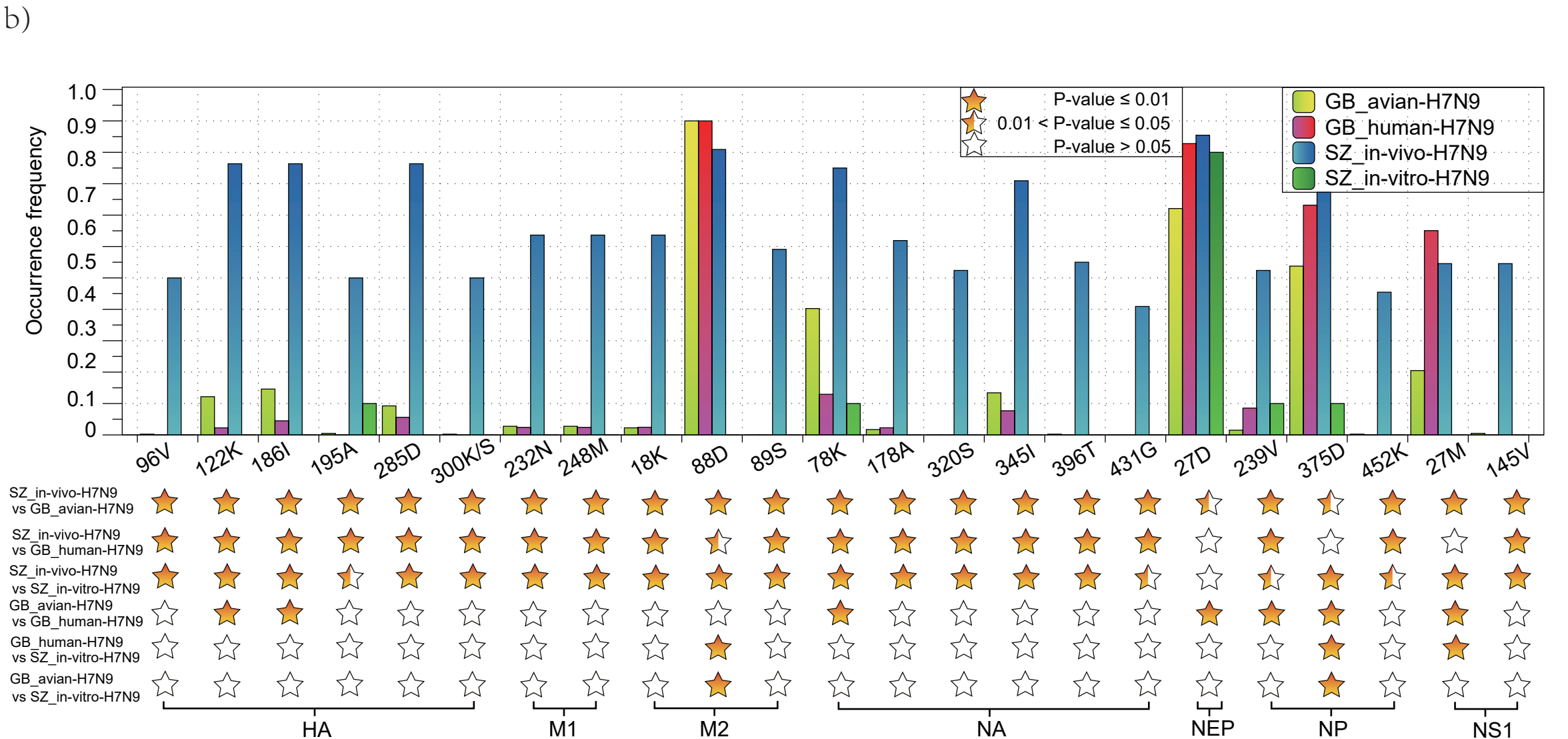
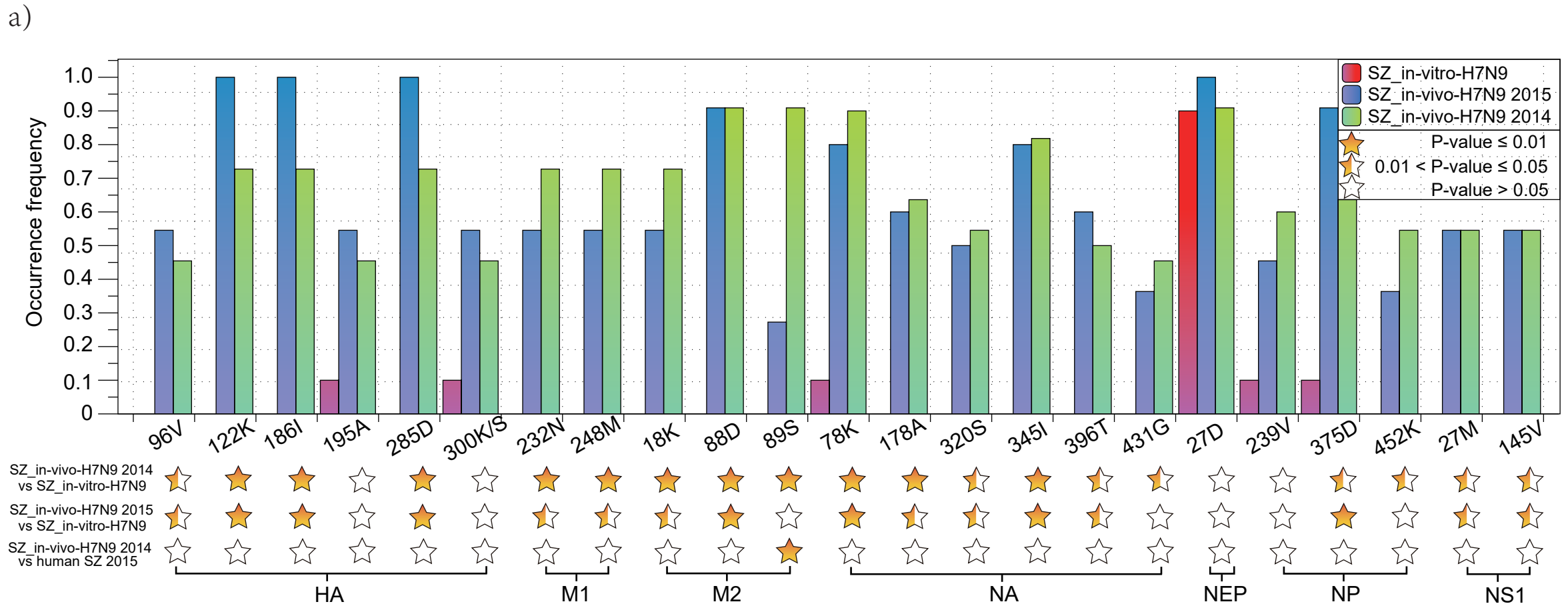


Figure 3

