

Title:

1 Sequence and phylogenetic analysis identifies a putative novel gyrovirus 3 genotype in ferret feces

Authors:

2
3
4
5 Enikő Fehér¹, Péter Pazár², György Lengyel³, Tung Gia Phan^{4,5}, Krisztián Bányai¹

Author affiliations:

6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

¹ Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Hungária krt. 21, H-1143 Budapest, Hungary

² Department and Clinic of Internal Medicine, Faculty of Veterinary Sciences, Szent István University, István utca 2, H-1078 Budapest, Hungary

³ Hungarian Defence Forces, Military Medical Centre, Róbert Károly krt. 44, H-1134 Budapest, Hungary

⁴ Blood Systems Research Institute, San Francisco, CA 94118, USA

⁵ Department of Laboratory Medicine, University of California at San Francisco, San Francisco, CA 94118, USA

Corresponding author:

Enikő Fehér, Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Hungária krt. 21, H-1143 Budapest, Hungary; e-mail: feher.eniko@agrar.mta.hu, Tel.: +36 1 467 4060, Fax: +36 1 467 4076

Keywords: semiconductor sequencing; whole genome sequencing; Hungary; *Circoviridae*

Abstract

The genomic sequence of a novel gyrovirus (GyV) 3 strain was detected from the fecal sample of a pet ferret. The length (2359 nt) and the basic genomic structure of this strain was very similar to that of the single known GyV3 reference strain, whereas the genome sequence identity between the two strains was only 76%. Similarly, moderate sequence homology was found within the predicted protein coding regions, VP1 (nt, 72%; aa, 76%), VP2 (nt, 84%; aa, 85%) and VP3 (nt, 85%; aa, 73%). Sequence identities were lower when comparing our strain with other GyV species (48-65% genome-wide nt identity). Phylogenetic analysis of the coding regions clustered the ferret origin GyV3 strain within Clade A. Although the available whole genomic sequence of novel GyVs permits limited conclusions to be drawn regarding the classification of the Hungarian GyV3 strain, our data indicate that this novel strain may be considered as a new genotype within GyV3. Further investigations are needed to reveal the genetic diversity and biological properties of newly described members of the *Gyrovirus* genus.

Introduction

Gyroviruses (GyVs) comprise a genus (*Gyrovirus*) within the family *Circoviridae*. The prototype GyV species, *Chicken anaemia virus* (CAV), causes significant losses in young chickens. Candidate new GyV species, including human GyV/avian GyV 2 (HGyV/AGV2), GyV3 to GyV6, have been first identified from chicken serum and/or human skin and fecal samples [1–5]. So far, the etiologic role of novel GyVs in any disease has not been formally demonstrated [1–7]. The ssDNA genome of GyVs encodes two or three potential proteins and based on phylogenetic analysis they have been grouped into two major clades; members of Clade A GyVs possess a larger genome (2315-2383 nt) and include HGyV/AGV2, GyV3 and GyV6 along with the CAV, whereas Clade B includes GyV4 and GyV5, both with a smaller genome (2020-2034 nt) [1–5].

In a previous study traces of CAV, HGyV/AGV2, GyV3 and GyV4 were detected in the fecal samples of pet ferrets (*Mustela putorius furo*) by viral metagenomics and by diagnostic PCRs [8]; however, discrepancy in the positivity rate for GyV3 and GyV4 was seen when the results obtained by the two methods were merged, probably due to the extensive sequence variability. A partial fragment of a distant relative of GyV3 was also detected by deep sequencing, but could not be fully analysed because of issues in whole genome amplification and sequencing. Since then the whole genomic sequence of the novel GyV3 strain was completed and in this study we describe the striking genomic and genetic features of this strain.

Results and discussion

Back-to-back PCR primers were designed to amplify the genome of the novel GyV strain, G19. However, contrary to our previous study [8], this approach worked only when the modified nucleotide, 7-deaza-dGTP, was incorporated in the back-to-back PCR assay. The PCR mixture contained 50 μ M of dATP, dCTP, dGTP and dTTP, 50 μ M of 7-deaza-dGTP (New England Biolabs), 200 nM of each primers, 1X Phusion Green HF Buffer and 0.3 U Phusion DNA polymerase (Thermo Scientific). The cycling profile was basically the same as described elsewhere for other GyVs [8], although we used 54 °C primer annealing temperature in this experiment. The ~2.3 kb long PCR products obtained by various combinations of the back-to-back primers were purified (Geneaid) from gel slices and then directly sequenced on an ABI PRISM® 3100-Avant Genetic Analyzer using the BigDye Terminator v1.1 Cycle Sequencing Kit (Life Technologies), with the addition of 2.5 μ M of 7-deaza-dGTP to the sequencing reactions. The genome sequence of G19 has been deposited in GenBank (accession number KM348009).

The genome length of the novel GyV strain was the same (2359 nt) as that of the single GyV3 reference sequence (JQ308210). The overall nt identity was 76% between the two genomes and was lower with other members of the genus (HGyV/AGV2 and GyV6, 65%; CAV, 63-64%; GyV4 and GyV5, 48%). The location and structure of the three hypothetical genes (VP1, VP2 and VP3) was comparable, only with 1-3 aa differences between the two GyV3 strains (Online Resource Fig. OR1). The intergenic region between the VP3 and VP1 consisted of one cytosine that was shared among GyV3, HGyV/AGV2 and GyV6 reference strains, whereas CAV had a guanosine in this position. The putative VP1 (1398 nt), VP2 (711 nt) and VP3 (372 nt) genes of G19 shared, respectively, 72%, 84% and 85% nt and 76%, 85% and 73% aa identities with the appropriate genomic

1 regions of the reference GyV3 strain (Online resource Fig. OR2). The aa based phylogenies of the VP1, VP2 and
2 VP3 genes using representative strains of the major GyV groups confirmed the previously described
3 relationships among all currently known GyVs (Fig. 1) [5]. The aa composition of the genes of GyV4 and GyV5
4 markedly differed from other GyVs [5].
5

6 In addition to the distinctive features, the aa sequence alignments revealed some conserved motifs
7 across different GyV species. The GyV3 genomes encoded the longest VP1 in the genus (465 aa for strain G19
8 and 463 aa for the reference GyV3 strain vs. 352-460 for other GyVs). The motif ¹⁷³WWRWA¹⁷⁷ within the VP1
9 appears to be conserved for all GyVs, whereas other previously described conserved regions had some apparent
10 aa changes [5]. GyV3 strains code for the longest VP2 (236 aa for the G19 and 239 for the reference GyV3 strain
11 vs. 216-232 for other GyVs) as well with the aa motif ⁴²LKKHDSPY⁴⁹, which is shared by all Clade A GyVs.
12 Conserved motifs with up to 5 aa were common among Clade A viruses, but only the motif ¹²³HWFQ¹²⁶ was
13 detected in all sequences, including GyV4 and GyV5. Beside shorter conserved sequences the VP3 of GyV3 and
14 HGyV/AGV2 contained the motif ³⁷EIQIGXGSTITXSLPGXASVRVLTTRSA⁶⁵ with only three substitutions
15 by homologous aa with similar polarity. In GyV6, this motif had some additional mutations, whereas CAV was
16 only distantly related.
17
18
19
20
21

22 The nt identity in the non-translated region (NTR) of the two GyV3 strains was 88%. This region of the
23 G19 strain contained two GC-rich region at nt 120-155 and nt 2296-2359 with 94.4% and 85.9% GC-content,
24 respectively, that may be a strong loop-forming segment of the genome (Online Resource Fig. OR3 and OR4).
25 The 36 nt long GC-rich region of the G19 was identical with the corresponding region of the reference GyV3
26 and GyV6 strains and differed only in one base from CAV. The 64 nt long GC-rich sequence showed 92%
27 similarity with the 67 nt long GC-rich region of the reference GyV3 strain. Furthermore, this 5' GC-rich region
28 of the G19 GyV3 NTR well aligned with the GC-rich region of HGyV/AGV2 (61-69 nt long, 88-89% identity),
29 GyV6 (partial, 51 nt long sequence, 89% identity) and CAV (66-67 nt long, 84-92% identity) 5' NTRs. Both
30 GyV3 NTRs contained the CAV promoter TGTACAGGGGGGTACGTC and two repeats of the reverse
31 complement of this sequence (TGACGTACCCCCCTGTACA) just upstream of the 19 nt long CAV promoter
32 (Online Resource Fig. OR3 and OR4). A short insert, GGGCGG, between the reverse complement and forward
33 CAV promoter in the G19 strain genome might serve as a SP1 transcription factor binding site. An insert
34 between the CAV direct repeats showed high affinity to SP1 as well [9]; homologous direct repeats of CAV
35 containing the oestrogen responsive element-like sequence AGCTCA may have a role in viral transcription and
36 replication [9,10].
37
38
39
40
41
42
43
44

45 The classification of GyV strains is not elaborated. So far, only CAV strains have been classified into
46 groups and subgroups (designated A1-A3, B, C, and D1-D2) on the basis of nt phylogenies of the coding region
47 [11]. The greatest difference among CAV sequences was 5.2% in the coding region [11]. Similarly, high
48 genome-wide identities were described for HGyV/AGV2 strains of distinct geographic and host-species origin
49 [8]. Consequently, the >20% nt and/or aa difference in the VP1 and VP3 regions between the two GyV3 strains
50 represents a sharp contrast in the intra-species diversity of GyVs and might be challenging to elaborate a uniform
51 classification scheme of the genus. In more recent studies circo- and cyclovirus genomes were analysed in order
52 to set up the species demarcation criteria. As a result, 75% and 76% species-specific sequence cut-off values
53 were proposed for circoviruses and cycloviruses, respectively [12]. A proposal (assigned code 2014.006a-gV)
54 submitted to the International Committee on Taxonomy of Viruses for consideration suggested that >20%
55
56
57
58
59
60
61
62
63
64
65

1 genome wide sequence divergence may define different circo- and cyclovirus species. Although the *Gyrovirus*
2 genus currently belongs to the *Circoviridae* family, these demarcation criteria may not be extended automatically
3 to this virus group as only limited sequence information is available for the novel GyV species. Among various
4 GyVs, at present only the GyV3 species seems to include virus variants with <80% genome wide sequence
5 identity, albeit this moderate sequence homology is accompanied with great gene and genome structure
6 similarity. Thus, we tentatively may classify the strain G19 as a novel genotype within the GyV3 species, which
7 would be consistent with the proposal that assigns a new species only when the genome wide nt identity to an
8 extant reference strain is lower than 75% [13]. Similarly, the genome of GyV4 strain G14 showed a genome-
9 wide identity of 90.7% with the only GyV4 reference sequence (JX310702) that was more divergent than the
10 sequences of the HGyV/AGV2 species [8] and may also represent a new genotype within the GyV4 species.

11
12
13
14
15 In conclusion, in this study we reported the genome of a putative new GyV3 genotype. So far GyV3 has
16 been described only from a child with unexplained diarrhea [3]; however, it should be noted that only few
17 studies investigated the prevalence of new GyVs in host species other than humans [4, 14-15]. The GyV strains
18 we analysed in this and a previous study were detected in the fecal samples of pet ferrets where chicken meat as
19 their food may be the source of GyVs. Further investigations are needed to provide more precise estimations of
20 the genetic diversity, and to describe the biological properties and the host spectrum of the novel members of
21 GyVs.
22
23
24
25
26
27

28 **Statement of author contributions**

29 E.F. and K.B. conceived the study, E.F., P.P., G.L. and K.B. performed experiments, E.F., P.P., G.L., T.G.P and
30 K.B. analysed and interpreted data; E.F. prepared the first manuscript draft, all authors edited an author draft and
31 approved the submitted manuscript.
32
33
34
35

36 **Acknowledgments**

37 This study was supported by the Momentum Program of the Hungarian Academy of Sciences. Enikő Fehér was
38 further supported by the European Union and the State of Hungary, co-financed by the European Social Fund in
39 the framework of TÁMOP 4.2.4. A/2-11-1-2012-0001 ‘National Excellence Program’.
40
41
42
43
44
45

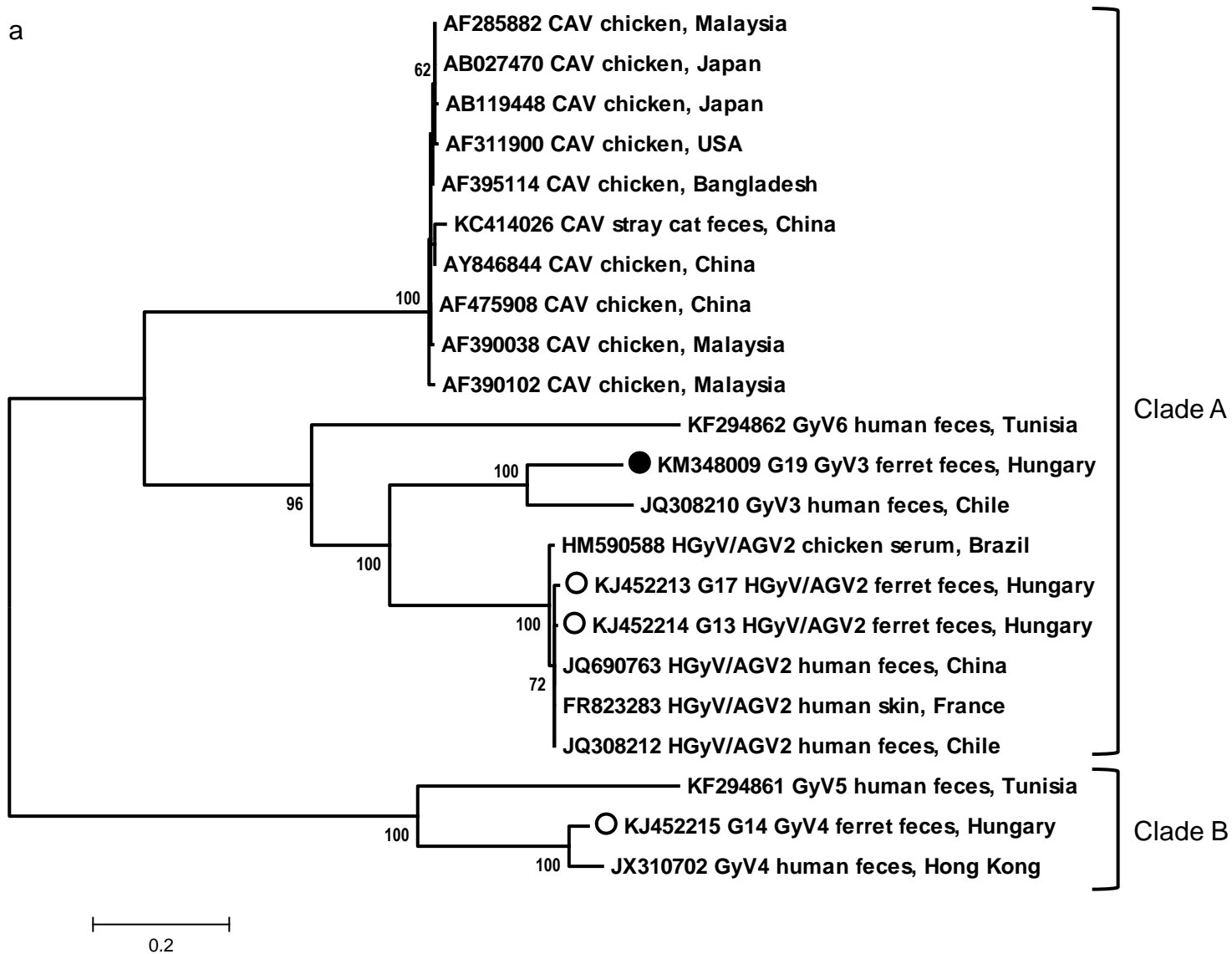
46 **Figure legends**

47 **Fig. 1**

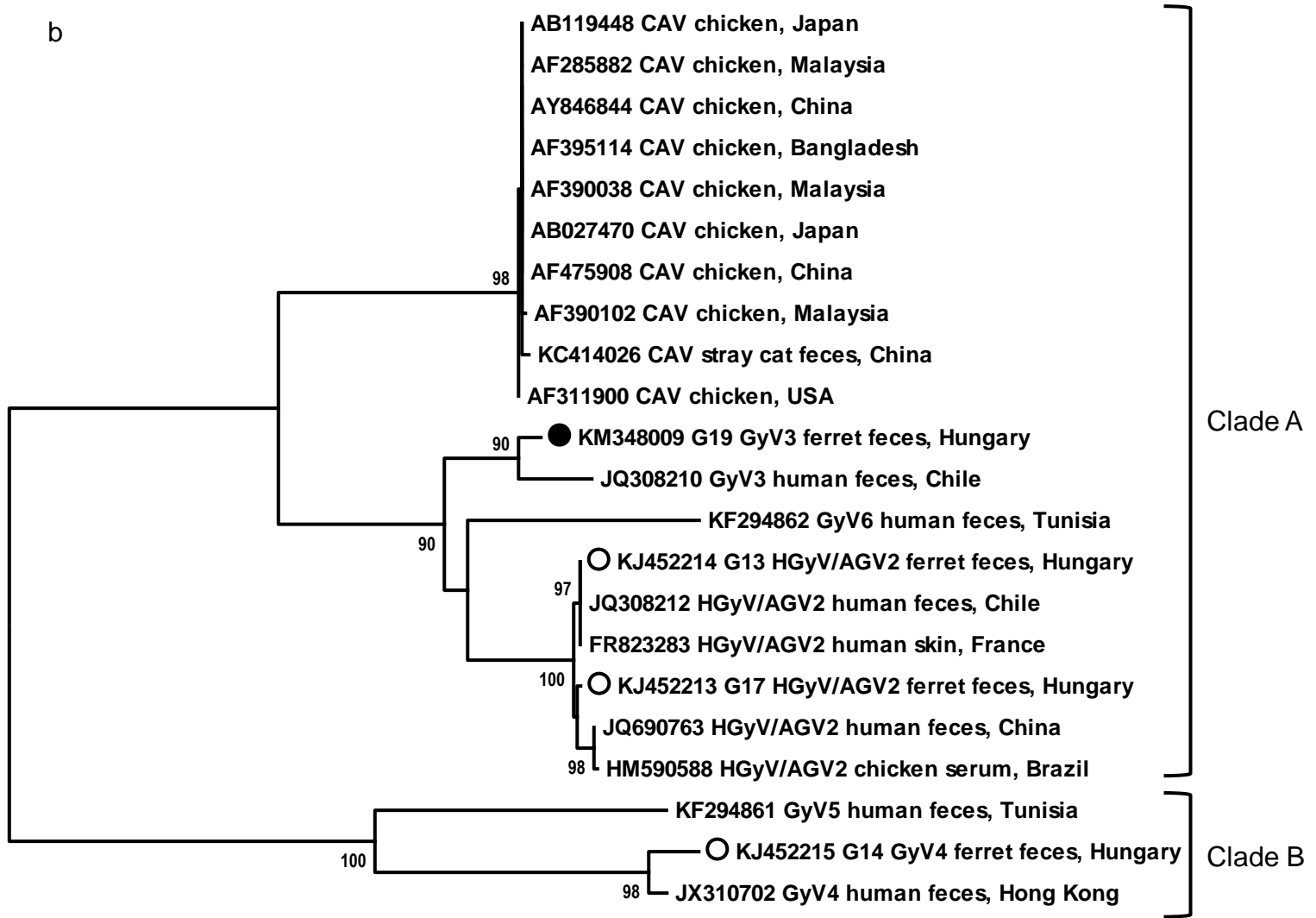
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
Maximum likelihood phylogenetic trees of VP1 (a), VP2 (b), VP3 (c) generated from the aa sequences of
gyroviruses (GyV) using the LG substitution model for VP1 and VP3, and JTT substitution model for VP2 as
implemented in MEGA 6 (<http://www.megasoftware.net/>). Bootstrap values >60% are shown. Scale bars are
proportional to the genetic distance. Circles indicate the Hungarian GyV strains. GenBank accession numbers of
the reference human gyrovirus/avian gyrovirus 2 (HGyV/AGV2), GyV1-6 and *Chicken Anaemia Virus* (CAV)
strains are indicated in the phylogenetic trees

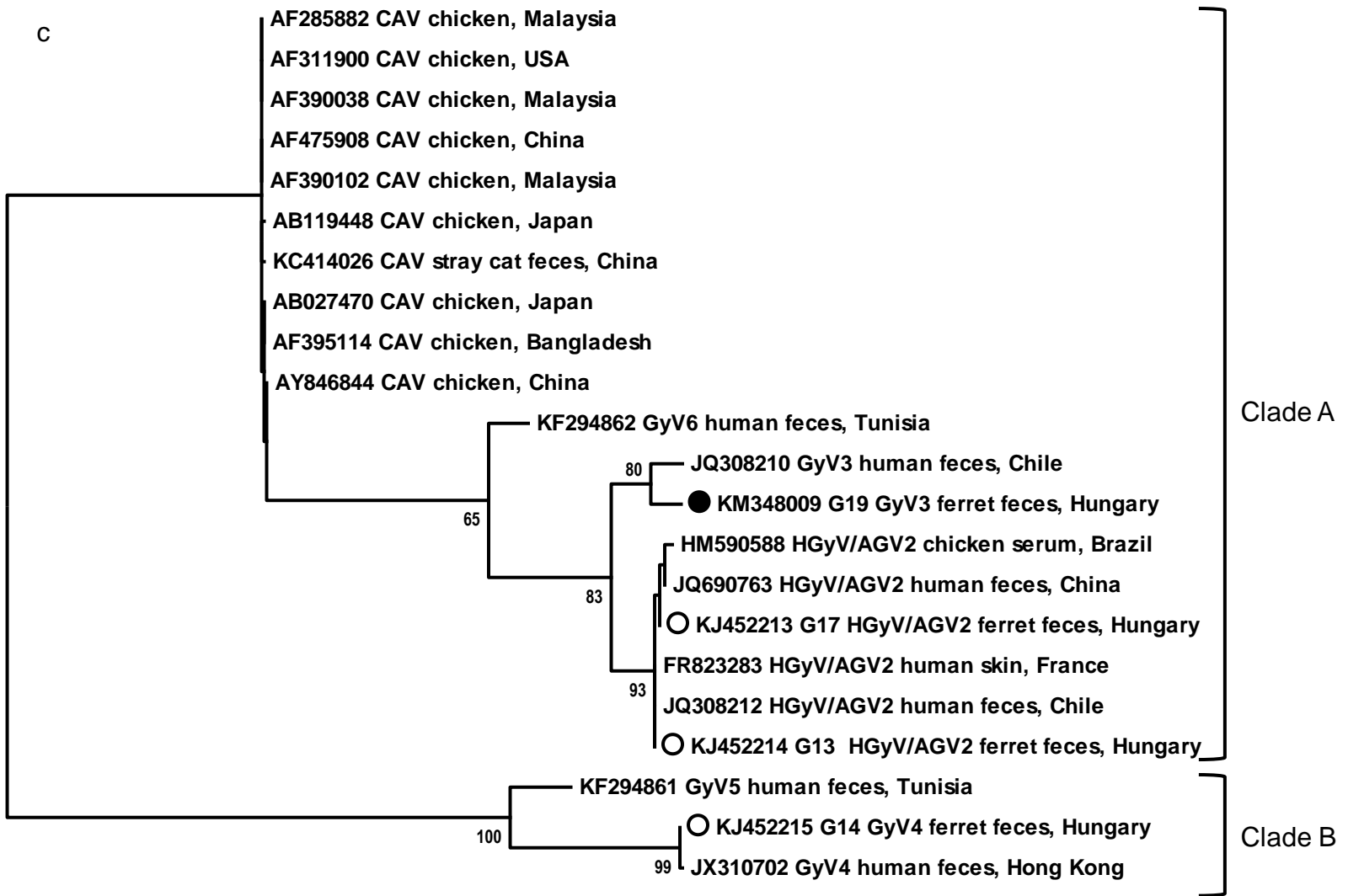
References

1. F.A. Rijsewijk, H.F. Dos Santos, T.F. Teixeira, S.P. Cibulski, A.P. Varela, D. Dezen, A.C. Franco, P.M. Roehe, *Arch. Virol.* 156, 1097 (2011)
2. V. Sauvage, J. Cheval, V. Foulongne, M.A. Gouilh, K. Pariente, J.C. Manuguerra, J. Richardson, O. Dereure, M. Lecuit, A. Burguiere, V. Caro, M. Eloit, *J. Virol.* 85, 7948 (2011)
3. T.G. Phan, L. Li, M.G. O’Ryan, H. Cortes, N. Mamani, I.J. Bonkougou, C. Wang, C.M. Leutenegger, E. Delwart, *J. Gen. Virol.* 93, 1356 (2012)
4. D.K. Chu, L. L. Poon, S.S. Chiu, K.H. Chan, E.M. Ng, I. Bauer, T.K. Cheung, I.H. Ng, Y. Guan, D. Wang, J. S. Peiris, *J. Clin. Virol.* 55, 209 (2012)
5. T.G. Phan, N. Phung Vo, K. Sdiri-Loulizi, M. Aouni, P. Pothier, K. Ambert-Balay, X. Deng, E. Delwart, *Virology* 446, 346 (2013)
6. F. Maggi, L. Macera, D. Focosi, M.L. Vatteroni, U. Boggi, G. Antonelli, M. Eloit, M. Pistello, *Emerg. Infect. Dis.* 18, 956 (2012)
7. P. Biagini, S. Bédarida, M. Touinssi, V. Galicher, P. de Micco, *Emerg. Infect. Dis.* 19, 1014 (2013)
8. E. Fehér, P. Pazár, E. Kovács, S.L. Farkas, G. Lengyel, F. Jakab, V. Martella, K. Bányai, *Arch. Virol.* (2014)
doi:10.1007/s00705-014-2203-3
9. M.H. Noteborn, C. A. Verschuere, A. Zantema, G. Koch, A.J. van der Eb, *Gene* 150, 313 (1994)
10. M.M. Miller, K.W. Jarosinski, K.A. Schat, *J. Virol.* 79, 2859 (2005)
11. Y.M. Eltahir, K. Qian, W. Jin, P. Wang, A. Qin, *Virol. J.* 8, 145 (2011)
12. A. Dayaram, K.A. Potter, A.B. Moline, D.D. Rosenstein, M. Marinov, J.E. Thomas, M. Breitbart, K. Rosario, G.R. Argüello-Astorga, A. Varsani, *J. Gen. Virol.* 94, 1827 (2013)
13. K. Rosario, S. Duffy, M. Breitbart, *Arch. Virol.* 157, 1851 (2012)
14. H.E. Smuts, *Adv. Virol.* 2014, 321284 (2014)
15. H.F. dos Santos, M.B. Knak, F.L. de Castro, J. Slongo, G.A. Ritterbusch, T.A. Klein, P.A. Esteves, A.D. Silva, I. M. Trevisol, E.A. Claassen, L.A. Cornelissen, M. Lovato, A.C. Franco, P.M. Roehe, F.A. Rijsewijk, *Vet. Microbiol.* 155, 230 (2012)



b





0.5

Supplementary Material

[Click here to download Supplementary Material: ESM.pdf](#)