



OPEN

Early origin and global colonisation of foot-and-mouth disease virus

Pakorn Aiewsakun^{1,2}✉, Nakarin Pamornchainavakul³ & Chaidate Inchai³✉

In this study, we compiled 84-year worth (1934–2017) of genomic and epidemiological data of foot-and-mouth disease virus (FMDV), and performed comprehensive analyses to determine its early origin and transmission route. We found that recombination is a key feature of FMDV, and that the genomic regions coding for structural and non-structural proteins have markedly different evolutionary histories, and evolve at different rates. Despite all of these differences, analyses of both structural and non-structural protein coding regions consistently suggested that the most recent common ancestor of FMDV could be dated back to the Middle Age, ~200 to 300 years earlier than previously thought. The ancestors of the Euro-Asiatic and SAT strains could be dated back to the mid-seventeenth century, and to the mid-fifteenth to mid-sixteenth century, respectively. Our results implicated Mediterranean counties as an early geographical origin of FMDV before spreading to Europe and subsequently to Asia and South America.

Foot-and-mouth disease virus (FMDV) is a pathogen infecting a wide range of cloven-hoofed animals, many of which are domesticated and economically important animals, including pigs, cattle, sheep, and goats^{1,2}. FMDV causes a disease called foot-and-mouth disease (FMD); although the disease itself is usually nonfatal in adult animals, it is highly contagious and can cause significant farming production losses, animal abortions, and high mortality rates in young infected animals^{1,2}. Indeed, FMD outbreaks have caused severe economic losses worldwide^{3,4}, making it a major threat to the global livestock industry. The global annual impact of FMDV has been estimated to be at least between 8–22.5 billion US dollars⁴. Despite this enormous global impact, where and when the disease originated, and how the disease spread globally still remind largely poorly characterised.

FMDV is a single-stranded, positive-sense RNA virus, belonging to the genus *Aphthovirus*, family *Picornaviridae*⁵. Its genome contains a single open reading frame (ORF), coding for a large polyprotein, which, after the translation, is processed into several mature proteins. The ORF can be divided into four regions based on the mature proteins' functions, namely L, P1, P2 and P3. The P1 region codes for four structural proteins involved in capsid formation called VP1–VP4, while the L, P2, and P3 regions code for several non-structural proteins involved in virus genome replication and protein cleavage activities (see Ref.⁶ for review).

FMDV is currently classified into seven serotypes, namely A, O, Asia1, C, SAT1, SAT2, and SAT3. The former four are also known as Euro-Asiatic serotypes, circulating mainly in Europe, Asia, and South America, while the latter three are collectively known as South African Territories (SAT) serotypes, found mainly in Africa. This classification scheme was established based on immunological reactions⁷, but studies have shown that analyses of VP1 molecular sequences can reproduce this classification scheme well^{8–10}. This is consistent with the fact that the VP1 protein is a major antigen which elicits protective humoral immunity^{11,12}. Consequently, VP1 was quickly adopted and widely used by the scientific community for the purpose of FMDV identification and classification, outbreak assessment, and vaccination program design and evaluation (e.g. Refs.^{13,14}).

Recent studies, however, showed that recombination is a common feature of FMDV, and different parts of its genome can have different evolutionary histories^{15–19}. Most importantly, one of the major recombination regions could be mapped to the P1 region, containing the VP1 protein coding region. This means that analyses of VP1 alone cannot fully reveal how the virus evolves. Furthermore, most FMDV studies so far either focused on specific serotypes^{20–23}, or those restricted to particular geographical regions^{13,24,25}; thus, our understanding

¹Department of Microbiology, Faculty of Science, Mahidol University, 272, Rama VI Road, Ratchathewi, Bangkok 10400, Thailand. ²Pornchai Matangkasombut Center for Microbial Genomics (CENMIG), Department of Microbiology, Faculty of Science, Mahidol University, 272, Rama VI Road, Ratchathewi, Bangkok 10400, Thailand. ³Veterinary Epidemiology and Economics Group, Department of Veterinary Medicine, Faculty of Veterinary Science, Chulalongkorn University, Henri Dunant Road, Patumwan, Bangkok 10330, Thailand. ✉email: pakorn.aie@mahidol.ac.th; chaidate.i@chula.ac.th

| Serotypes | O | A | Asia1 | C | SAT1 | SAT2 | SAT3 | Unclassified | Total |
|----------------------|--------------|--------------|------------|------------|------------|------------|------------|--------------|-------|
| Before data curation | 545 (56.77%) | 200 (20.83%) | 63 (6.56%) | 30 (3.13%) | 27 (2.81%) | 30 (3.13%) | 24 (2.50%) | 41 (4.27%) | 960 |
| After data curation | 476 (61.90%) | 158 (20.55%) | 48 (6.24%) | 12 (1.56%) | 27 (3.51%) | 25 (3.25%) | 23 (2.99%) | – | 769 |

Table 1. Summary of the FMDV sequences retrieved from the NCBI database.

of FDMV evolution still remains largely fragmented and incomplete. Here, we set to investigate the global evolutionary history of FMDV across seven serotypes. Our results provided new insights into its origins and how it spread globally.

Materials and methods

Data collection. All publicly available whole genomes and complete coding sequences of FMDVs were retrieved from the National Center for Biotechnology Information (NCBI) database, using the search terms “(foot-and-mouth disease virus [Organism]) AND 6,000:9,000 [Sequence Length]”. Metadata associating with the sequences were compiled (Table S1), and examined to exclude experimental isolates. These included vaccine seed viruses, viruses from cell culture, and those with genomes modified for various research purposes. Duplicated sequences, sequences with significant portions of undetermined nucleotide sequences (≥ 200 ‘N’), and those without sampling dates were also removed from the dataset.

Recombination detection. A manually-curated alignment of FMDVs’ ORFs was prepared using MAFFT²⁶. Potential recombination events within the ORF were checked using RDP, GENECONV, Chimaera, MaxChi, BootScan, SiScan, and 3Seq, all implemented in Recombination Detection Program 4²⁷. Only those that were detected by more than four programs were considered. The distribution of recombination breakpoints was manually examined, and the alignment was split according to the identified frequent breakpoints. The process was repeated until no frequent breakpoints were found. If small and sporadic recombination events were detected without any consistent breakpoint patterns, they were deleted to generate “recombination-free” alignments. Five recombination-free nucleotide alignments were constructed in total.

Phylogeny reconstruction. Phylogenies were estimated from the five alignments under the maximum likelihood framework using RAxML V8²⁸, all with the GTRCAT substitution model. The number of rate categories was set to 25. Bootstrap clade support values were calculated using Booster v.0.1.2²⁹ from 1,000 bootstrap trees.

Molecular dating. We regressed root-to-tip genetic divergence against sampling date to evaluate if the alignments had sufficient temporal signals using TempEst v.1.5.1³⁰. The root placements were determined by minimising the heuristic residual mean squared errors. Given the estimated root placements, the root-to-tip regression analyses were repeated in R v.3.5.2 using the ‘*lm*’ function, and one-tail t-tests were used to assess the significance of the slopes (i.e. if the slopes were significantly positive). Examination revealed that there were several outliers (i.e. viruses with either too low or too high genetic distances given their sampling dates), and analyses were thus repeated again without them. For those with a significant temporal signal, i.e. the p-values of the estimated slopes were less than 0.05, the estimated models were used to calculate the dates of other divergent events by using the ‘*predict*’ function, also implemented in R v.3.5.2.

Results and discussion

Data. 960 genomic sequences of FMDVs were retrieved from the NCBI database, 41 of which lacked serotype records. Out of the 919 sequences with serotype records, 91.19% of the sequences belonged to the Euro-Asiatic serotypes (O, A, Asia1, and C), and 8.81% belonged to the South African Territories strains (SAT1–3). All SAT strains had an African origin, except for one (AY593839.1; SAT1; collected in 1970), of which the origin was unclear. Serotype A, Asia1, C, and O FMDVs were from the rest of the world, with a few being from the African continent.

Upon manual inspection, we found a number of these sequences to be duplicates. Some belonged to vaccine or experimental strains, while some contained large strings of undetermined or missing nucleotides. Including these sequences could potentially bias our analyses, and thus they were removed from our dataset. In addition, we also excluded sequences without sampling dates as they were required for evolutionary dating analysis (see below). Incidentally, all of the sequences lacking the serotype records were also either obtained from experimental virus isolates, or did not have sampling dates, and thus were removed from the dataset. The curated dataset contained 769 sequences (Table 1). The proportion of the serotypes in the dataset did not change significantly (Chi-squared test: $\chi^2 = 6.53$, degree of freedom = 6, *p* value = 0.37), suggesting that the curation did not significantly change the data structure. Our data comprised epidemiological data and genomic sequences of FMDVs collected over a time span of 84 years (1934–2017).

Recombination detection. An alignment of the sequences was made and checked for potential recombination. The alignment was manually curated, and was 6,999 nucleotide (nt) long after the curation (Data S1).

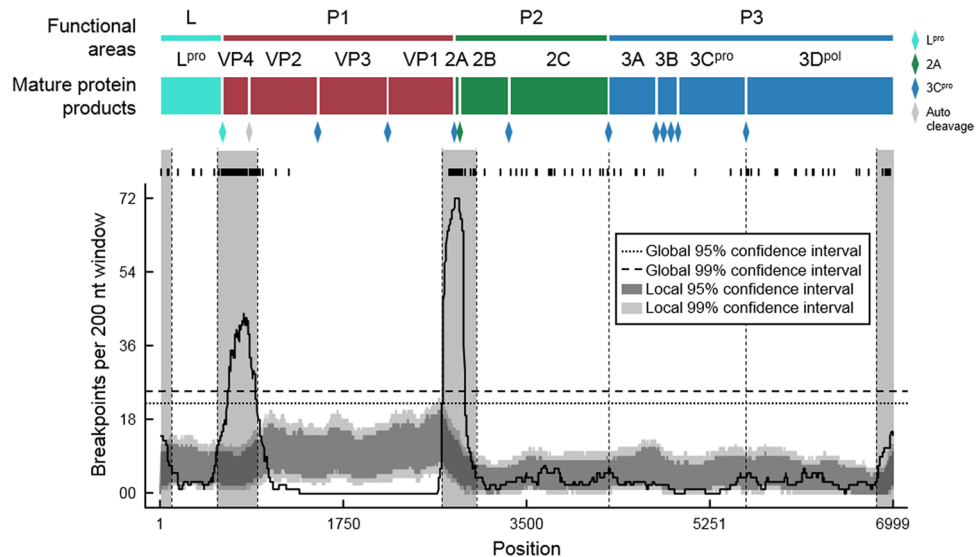


Figure 1. Recombination analysis. (Top) schematic diagrams of FMDV's ORF. The polyprotein encoded by the ORF is processed into several protein products by proteases (Diamonds; see Keys for the names of proteases). Based on the functions of mature proteins, the ORF can be divided into four main functional areas (L: turquoise; P1: red; P2: green; and P3: blue). L region encodes a leader protease (L^{pro}). The P1 region encodes four structural capsid proteins, VP1–4. The P2 and P3 regions encode several non-structure proteins, including 2A–C, and 3A–D. 3C and 3D are a viral protease ($3C^{pro}$) and an RNA-dependent RNA polymerase ($3D^{pol}$), respectively. (Bottom) number of inferred recombination breakpoints estimated per 200-nucleotide window. The tick marks above the plot show the locations of individual breakpoints inferred. Global and local 99% and 95% confidence intervals of the estimated recombination breakpoints are shown (see Keys). Regions with the inferred number of recombination breakpoints more than the 99% local confidence intervals are considered recombination hotspots. Four recombination hotspots were detected in total, indicated by vertical transparent strips. Vertical dotted lines are the locations at which the ORF alignment was split (Data S2–6).

Our analyses detected four major recombination hotspots (Fig. 1); two were located at the extreme ends of the ORF (nt 1–99, and nt 6,841–6,999), and the other two were mapped to the VP4 (nt 557–939), and 2A coding regions (nt 2,698–3,033).

These hotspots separated the ORF into three distinct regions, corresponding well to the established functional areas⁶. This suggested that recombination in FMDV is constrained by biological functions, and further supported the idea of mosaic structure of FMDV genome^{17–19}. The genome of other viruses has also been observed to have a mosaic and modular nature, such as foamy virus³¹, and enteroviruses³², for example. Excluding the inferred recombination hotspots, the first region from the 5' end contained the leader protease coding region (nt 100–556; L^{pro} alignment). The second region comprised almost the entire P1 area, covering the VP1–3 coding regions (nt 940–2,697; P1 alignment). The third region mapped to the P2 and P3 regions (nt 3,034–6,840; P2–3 alignment), excluding the 2A coding region. As previously noted, only a few recombination events were detected in the P1 region, which codes for capsid protein subunits^{15–19}. A previous study of serotype O, A, and Asia-1 FMDVs circulating in West Eurasia also showed that, while the VP1–3 phylogenies were the same, a phylogeny estimated from the VP4 coding region was different from the rest³³. This finding was consistent and further supported that the VP4 region is a recombination hotspot, and that the VP1–3 regions share the same evolutionary history. This finding suggested that the interaction between protein subunits required to form a functional capsid protein shell, especially the VP1–3 proteins, is highly constrained, such that recombinants involving exchanging genomic materials in this region are strongly unfavoured by natural selection¹⁹. Potential recombination was then checked again within these three alignments. While we could not detect additional recombination events within the L^{pro} , and P1 alignments, a number of recombination could still be found in the P2–3 alignment.

Inspection of the recombination breakpoints in the P2–3 alignment revealed that, although they were not hotspots, there were some aggregations of recombination breakpoints at the P2–P3 junction and the 3C–3D junction (Fig. 1). Again, this separated the ORFs into genomic regions coding for functionally related proteins, suggesting a role of epistatic interaction between them. We thus further split the P2–3 alignment into three alignments: (1) the P2 region (nt 3,034–4,275; P2 alignment), (2) the 3A–C region (nt 4,276–5,586; 3A–C alignment), and (3) the 3D region (nt 5,587–6,840; 3D alignment). Recombination detection analyses were performed again on these three alignments. A number of recombination events could still be detected, but there were only a few of them, and they were small and sporadic. The detected recombination regions were thus removed to create “recombination-free” alignments. Five recombination-free alignments were made in total, the L^{pro} , P1, P2, 3A–C and 3D regions (Data S2–S6).

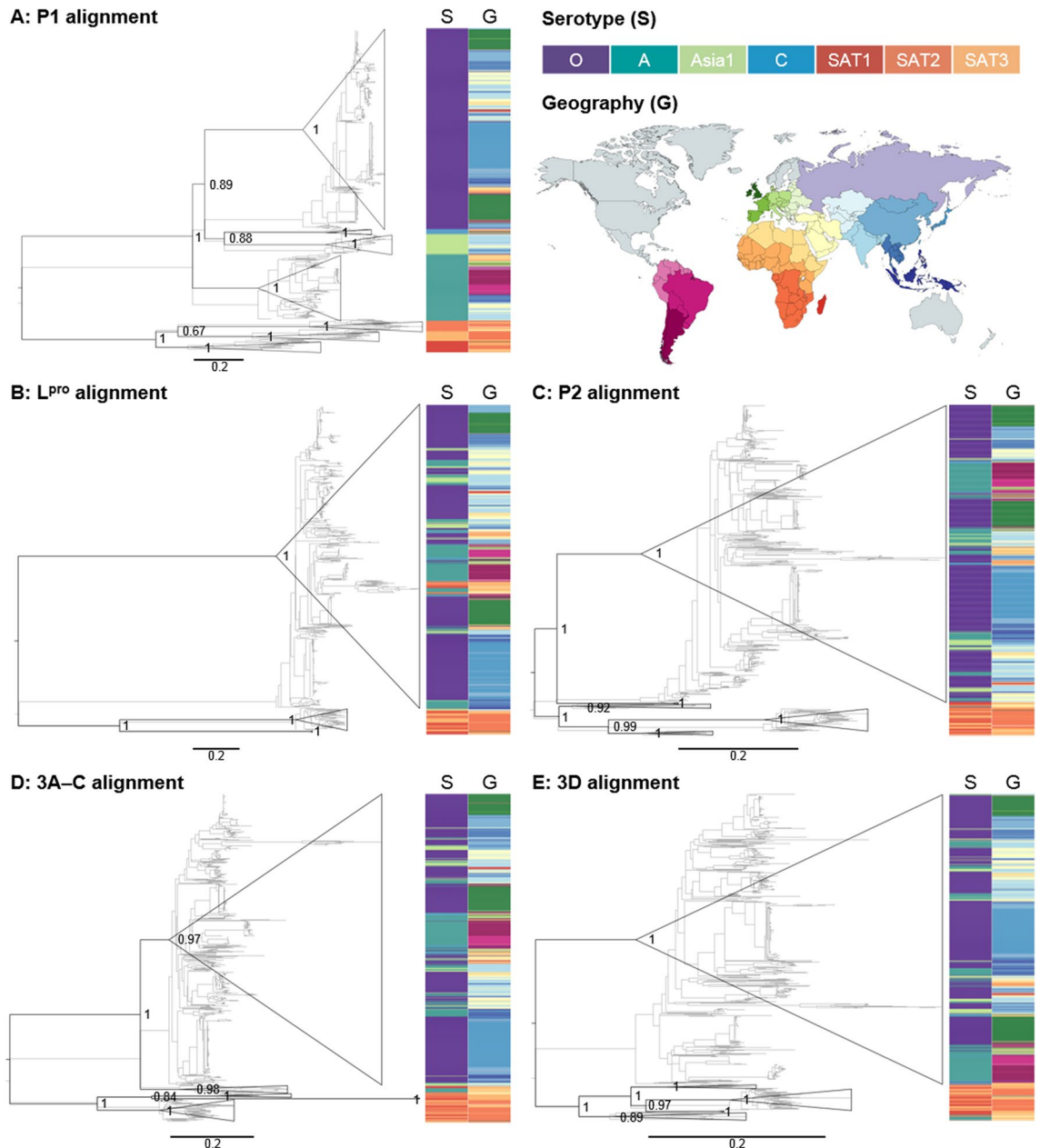


Figure 2. FMDV phylogenies. Five phylogenies were estimated from the five recombination-free alignments (A: P1; B: L^{PRO}; C: P2; D: 3A-C; and E: 3D) under the maximum likelihood framework. The numbers on nodes are bootstrap clade support values, shown for various distinct clades. Serotype (S) and geographical (G) information were mapped onto the trees (see Keys). Scale bars are in the units of nucleotide substitutions per site. The map (licence CC BY-SA 4.0) was created with mapchart.net.

Phylogenetic analyses. Five maximum likelihood phylogenies were estimated from the curated alignments and were rooted by maximising the phylogenetic temporal signals. The results are shown in Fig. 2 (see Tree S1–5 for the trees in the Newick format).

The current classification of FMDVs divides the virus into seven serotypes⁷, and the P1 phylogeny (Fig. 2A) showed that FMDVs segregated into seven well-supported clades (100% bootstrap support), perfectly mirroring their conventional classification. This was expected however, given that the P1 region contains the VP1 coding region commonly used to define the serotype in the first place. A clear separation between the Euro-Asiatic serotypes and the SAT serotypes could also be seen, each forming a clade with 100% bootstrap support. Our analysis suggested that serotypes C and Asia1 were sisters (88% bootstrap support), forming a sister clade to serotype O (89% bootstrap support), while serotype A was the most basal serotype of all Euro-Asiatic serotypes. For the SAT serotypes, our analysis suggested that SAT1 was the most basal one, and SAT2 and SAT3 formed a clade, although not very well-supported (67% bootstrap support).

Four trees were estimated from non-structural protein coding regions (Fig. 2B–E). Again, we found that, although not absolute, Euro-Asiatic FMDVs tended to cluster together separately from the SAT stains (>97% bootstrap support). However, in stark contrast to what we observed in the P1 tree, serotypes A, O, C and Asia1 appeared to intermix with one another in all of these trees, consistent with previous results³³. Furthermore, a similar pattern was also found for the SAT1–3 serotypes. This finding strongly suggested that recombination among Euro-Asiatic serotypes and among SAT serotypes is very common. Recombination involving SAT and Euro-Asiatic serotypes could also be observed although rarer. This can be seen pictorially as clusters of green and orange colours in the S(erotype) columns in Fig. 2.

In terms of geographical separation, we found that, while FMDV sequences from the same country tended to cluster together, this phylogenetic pattern was not absolute. This finding was consistent with results previously reported, indicating that the spread of FMDV transcends country boundaries, perhaps as a consequence of international livestock trade. Nonetheless, we observed that African FMDVs tended to aggregate into a few large clades (Fig. 2; orange colour in the G(geography) columns). The same was observed for Asian FMDVs (Fig. 2; blue and yellow in G columns). South American FMDVs (Fig. 2; pink in G columns) also consistently appeared to be closely related to those from middle Europe in all trees (Fig. 2; light green in G columns). Together with our date estimates, these phylogenetic patterns gave new insight into the past transmission history of FMDVs.

Molecular dating. Temporal signal within each phylogeny was evaluated using root-to-tip regression analysis (Fig. 3A). The P1, 3A–C and 3D regions had statistically significant temporal signals while the L^{P1} and P2 region did not. Their substitution rates (i.e. the slopes of the regressions) were estimated to be 1.71×10^{-3} , 3.01×10^{-4} , and 3.85×10^{-4} substitutions/site/year (s/n/y), respectively. Manual inspection, however, showed that some virus isolates appeared to evolve much faster or slower than average (Fig. 3A, red dots), including FMDVs from Taiwan (1997–1999), China (2000–2002, 2005, 2013), and Nigeria (2015). We thus repeated the analyses again but without these “outliers”, and found that the 3A–C region no longer had a significant temporal signal, suggesting that the initially detected signal was an artefact, driven by a small group of viruses. The P1 and 3D regions, in contrast, retained the signals. The rate (i.e. the slope) estimates also changed only slightly for these two regions (P1: 1.81×10^{-3} s/n/y; 3D: 3.81×10^{-4} s/n/y; Fig. 3A), suggesting the signals were genuine and robust. The rate estimate of the P1 region, in particular, was also comparable to those previously reported (Fig. 3). We noted that the P1 region evolved about 1 order of magnitude faster than the 3D region. This marked difference in the rates could be explained by the fact that the P1 region codes for capsid proteins, which are known to be immunogenic and therefore are likely under a strong positive selection pressure, while the 3D region codes for a polymerase, which is generally known to be highly conserved.

The P1 and 3D regions were further analysed to compute the timings of the origins of various FMDV lineages, since they were the only two regions that had statistically significant and robust temporal signals for molecular dating. The root-to-tip regression analysis of the P1 region (Fig. 3B) suggested that the most recent common ancestor (MRCA) of FMDVs was ~750 (95% CI: 598–978) years old (the x-axis intercept), diversifying into the SAT and Euro-Asiatic lineages around 1,265 AD (95% CI: 1,041–1,421 AD). Thereafter, the ancestral SAT lineage diverged into the SAT1 lineage, and subsequently the SAT2 and SAT3 lineages in 1561 AD (95% CI: 1,402–1679 AD) and 1,610 AD (95% CI: 1,460–1723 AD), respectively. Their MRCAs started to diversify ~388 (1631 AD; 95% CI: 1,486–1743 AD), ~118 (1901 AD; 95% CI: 1794–1998 AD), and ~245 (1774 AD; 95% CI: 1652–1875 AD) years ago in their respective order. For the Euro-Asiatic serotypes, their MRCA was dated back to 1642 AD (95% CI: 1,499–1752 AD), with serotype A being the first lineage to branch out. Serotype O subsequently branched out into a separate lineage in 1669 AD (95% CI 1,530–1777 AD), and the last two serotypes, Asia1 and C, diverged from one another in 1713 AD (95% CI 1582–1817 AD). Furthermore, we estimated the time to MRCA (tMRCA) of serotypes A, O, Asia1 and C to be ~232 (1787 AD; 95% CI: 1668–1887 AD), ~132 (1887 AD; 95% CI: 1779–1984 AD), ~84 (1935 AD; 95% CI: 1831–2032 AD), and ~81 (1938 AD; 95% CI: 1835–2035 AD) years ago, respectively.

Analysis of the 3D region (Fig. 3C) suggested that the tMRCA of FMDV could be dated back to 1,330 AD (95% CI: 1,039–1514), comparable to the one yielded from the P1 analysis. Since the analysis could not recover a clear division among the seven serotypes of FMDV, we could not estimate the timescale for each one of them individually. Nevertheless, we could estimate the clade predominantly Euro-Asiatic to be ~392 years old, dated back to ~1627 AD (95% CI: 1,426–1768), and the clade dominated by SAT isolates to be ~560 years old, dated back to ~1,458 AD (95% CI: 1,207–1621). Remarkably, these results were also comparable to those obtained from the P1 analysis.

One of the earliest observations of FMD could be dated back to 1514 in Italy³⁴. Coinciding with this record, Tully et al.²⁰ and Yoon et al.⁹ dated the MRCA of FMDVs to early 1500 s based on analyses of the VP1 coding regions and virus whole genomes, respectively (Fig. 3). Based on this date estimate, which fell within the Age of Discovery, Tully et al.²⁰ hypothesised that the origin of FMDVs was likely in Europe, and that the European exploration played a key role in the emergence of FMDV in Africa, Asia and the South America. In contrast, our analyses dated that the MRCA of FMDVs to the mid-thirteenth to mid-fourteenth century, being ~700–750 years old, which is ~200–300 years older than the previous age estimates^{9,20} (Fig. 3B, C). Our date estimates instead coincided with the increase in goods and livestock transportation from Mediterranean countries to Northern Europe in the Middle Age³⁵, fitting very well with the early record of FMD in a Mediterranean country like Italy.

Regarding the Euro-Asiatic serotypes, our analyses consistently dated their MRCA to the mid-seventeenth century (Fig. 3). Our date estimates were comparable to that estimated by Tully et al.²⁰ (~1695 AD), but were considerably older than that estimated by Brito et al.¹⁵ (~1889 AD). Our dates matched well with the end of the Age of Discovery, which, by the end of the era, the foundation for the European colonial empires and global trade between Europe and the rest of the world had been laid. This suggested that, following the introduction of

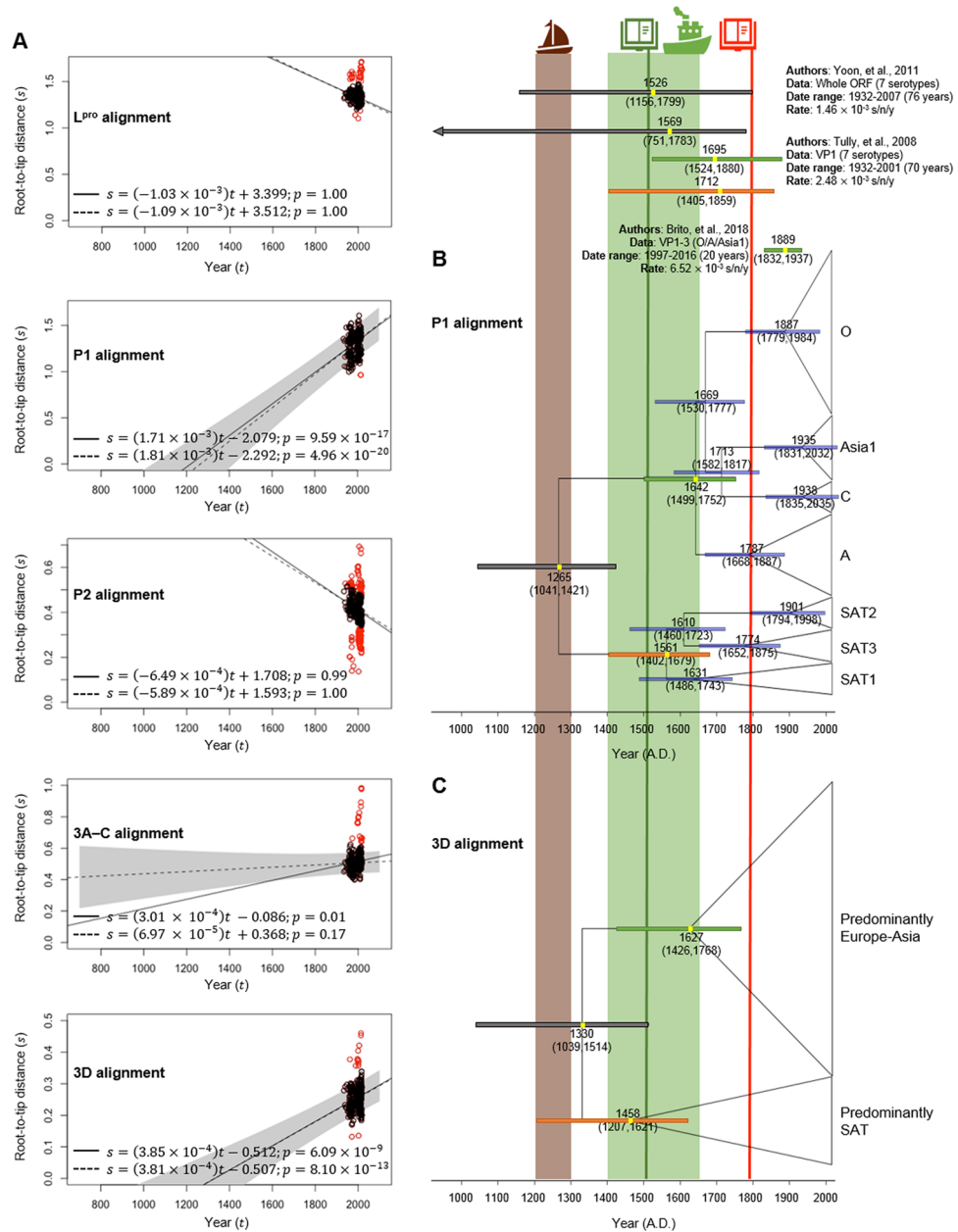


Figure 3. Molecular dating of FMDVs. Regression analyses of root-to-tip genetic distance (s) and sampling dates (t) were performed for the five genomic regions to evaluate their temporal signals (A). The solid lines are models estimated based on the entire dataset. The dotted lines are models estimated without outliers (red dots). P-values are reported next to the equations. Analyses suggested only the P1 and 3D regions had sufficient temporal signals for molecular dating (p values of the slopes < 0.05). Evolutionary timescales of FMDVs were computed by using the relationships between s and t estimated based on the P1 (B) and 3D (C) regions. Confidence intervals of the estimated dates are shown by the horizontal bars; grey: tMRCA of all FMDVs, green: tMRCA of Euro-Asiatic serotypes, orange: tMRCA of SAT serotypes, purple; others. Corresponding date estimates by other studies are shown for comparison (Top). The numbers on nodes are median date estimates and their associated confidence intervals. Coloured background and vertical line indicate historical events that might be relevant in the global transmission of FMDVs. The “sailboat” symbol: the rise of livestock transportation from the Mediterranean region to Northern Europe in the thirteenth century³⁵; the green “book” symbol: early symptomatic record of FMD in Italy in 1514³⁴; the orange “book” symbol: early symptomatic record of FMD in Africa in 1795⁴³; the “ship” symbol: the age of discovery starting around early fifteenth century to mid-seventeenth century. Ship and book drawings were Microsoft PowerPoint ClipArts.

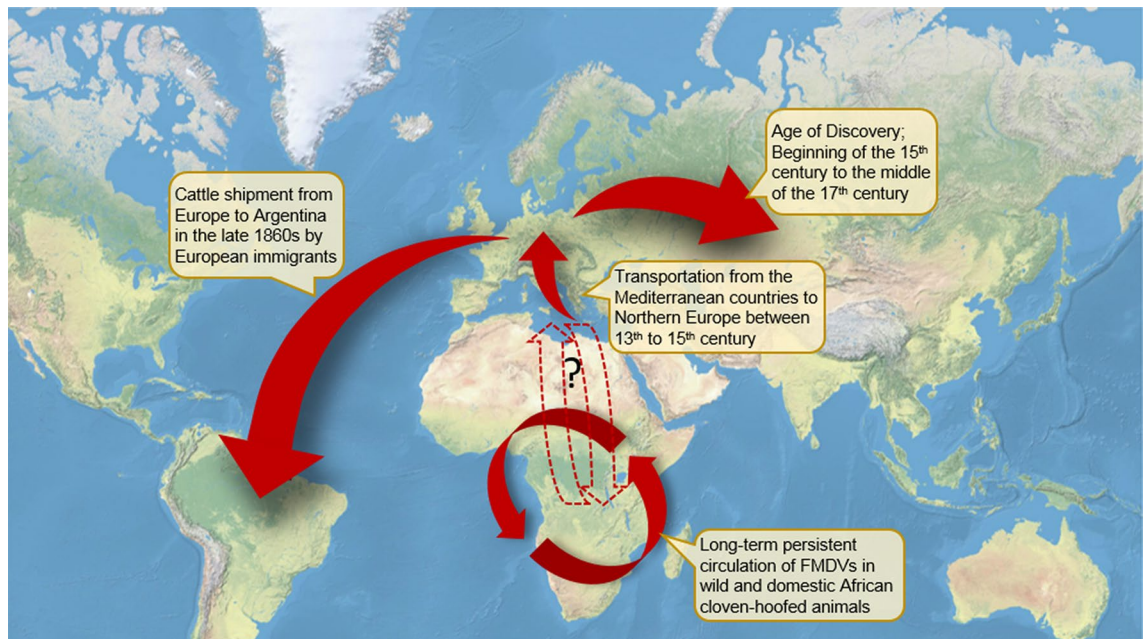


Figure 4. An updated hypothesis on the global colonisation of FMDV. The ancestor of Euro-Asiatic FMDV strains likely originated in the Mediterranean region, and spread to Europe, facilitated by the raise in goods and livestock transportation from the Mediterranean countries to Northern Europe between the thirteenth and fifteenth century. The foundation for the European global trade between Europe and the rest of the world was laid during the Age of Discovery (fifteenth–seventeenth century), and this likely helped the disease to disperse from Europe to Asia. The cattle shipment by European immigrants from Europe to Argentina in the late 1860s was likely responsible for the emergence of FMD in the South America continent. The SAT strains, on the other hand, are very well confined to the African continent. They likely have been circulating in wild and farm African animals for a long time, which play an important part in maintenance of the disease in the African continent. It is unclear, however, if the disease itself was first developed in the African continent then spread to Mediterranean countries (dotted upward arrow) or the other way around (dotted downward arrow). Further studies are required to distinguish between these two competing hypotheses. The world map was created by using *OpenStreetMap* V 0.3.4 R package (<https://cran.r-project.org/package=OpenStreetMap>).

FMD from Mediterranean countries, Europe then acted as a hub to spread the disease to Asia via exploration and trading. In addition, our analyses showed that South American serotype A FMDVs were closely related to European strains in all of the five trees, currently classified as belonging to the same topology, namely the Europe–South American topotype. Their basal diversification could be dated back to only ~1722–1858 AD. This phylogenetic pattern and date estimates were in line with the first record of FMD outbreak in the South America continent in 1,870, and could be explained by the historical record of cattle shipment from Europe to Argentina in the late 1860s by European immigrants. Again, this result implicated Europe as a major player in the spread of the disease^{20,36,37}.

Regarding the SAT serotypes, they are very well confined to the African continent, and their MRCA could be dated back to the mid-1400 s and 1500 s, pre-dating and thus consistent with an early record of FMD in Africa in 1795 (Fig. 3). These date estimates were considerably earlier than that suggested by Tully et al.²⁰, placing the MRCA of the SAT serotypes in early 1700s. Our older age estimates, however, agreed with the results from other studies, which, although did not estimate the tMRCA of SAT serotypes directly, calculated the ages of SAT1 and 2 to be at least ~400 to 540 years old^{25,38}. These results further corroborated a well-received hypothesis of long-term persistent circulation of FMDVs in African wild animals, such as African buffalo and impala, which may have been acting as FMDV natural reservoirs, sporadically passing the disease to domesticated animals^{20,25,39,40}.

As discussed above, Tully et al.²⁰ suggested that the MRCA of FMDVs likely originated in Europe and later spread to the rest of world via exploration and trading routes. Our analyses, on the other hand, located the MRCA of FMDV to be the Mediterranean region, which lies between the Africa and Europe continents. As such, it is unclear whether the disease was first originally developed in the African or European continent. Indeed, it was originally generally believed that the disease might originate in the Africa due to the long-term FMDV infection of African wildlife and the greater genetic diversity of the SAT serotypes⁴¹. Analyses of additional FMDVs from wild animals or their virus relatives may help distinguish between these two competing hypotheses.

Conclusion and final remarks

By integrating the genomic data with epidemiological information collected over 84 years, we performed comprehensive phylogenetic analyses to characterise how FMDV evolved and spread globally. Our analyses showed that recombination can readily occur in FMDV, especially among Euro-Asiatic serotypes and among SAT serotypes,

but the process is likely largely constrained by epistatic interaction between regions coding for functionally related proteins, resulting in relatively clear mosaic genomic blocks of low recombination-rate regions as previously noted^{17–19}. This phenomenon has in fact also been observed in other viruses^{31,32}. Given how common it is, recombination likely plays an important role in survival and adaptation of viruses in the face of new environments. In FMD endemic areas, especially those with high incidence rates of FMDV persistent infection, co-infection of FMDVs can indeed occur⁴², and intra-host recombination could play a significant role in increasing the genetic diversity of the virus, generating new virus variants^{18,19}, and potentially altering the virus biology³³. Although not directly shown in this study, it has been noted that recombination in the capsid protein coding region could occur, and this could potentially reduce the success of FMD vaccination programs³³. Thus, antigenic diversities must be considered in the development of FMDV vaccine.

Our evolutionary dating analyses suggested that FMDV are older than previously thought, dated back to the mid-thirteenth or mid-fourteenth century. By mapping our date estimates to historical events, we hypothesised that Euro-Asiatic strains had their early origin in the Mediterranean countries and subsequently spread to Europe and the rest of the world via European livestock trading and migration (Fig. 4). The SAT strains are also older than previously estimated. We calculated them to be ~450 to 550 years old, consistent with the hypothesis that FMDV has been circulating in African wild animals for a long time, playing an important role in maintaining FMD in the African continent^{20,25,39,40}. The results from this study further our understanding and provided new insights into the early diversification, emergence, and global colonisation of FMDV.

Data availability

The data that supports the findings of this study are available in the supplementary material of this article.

Received: 9 May 2020; Accepted: 26 August 2020

Published online: 17 September 2020

References

- Arzt, J., Juleff, N., Zhang, Z. & Rodriguez, L. L. The pathogenesis of foot-and-mouth disease I: Viral pathways in cattle. *Transbound. Emerg. Dis.* **58**, 291–304 (2011).
- Arzt, J. *et al.* The pathogenesis of foot-and-mouth disease II: Viral pathways in swine, small ruminants, and wildlife: Myotropism, chronic syndromes, and molecular virus-host interactions. *Transbound. Emerg. Dis.* **58**, 305–326 (2011).
- Onono, J. O., Wieland, B. & Rushton, J. Constraints to cattle production in a semi-arid pastoral system in Kenya. *Trop. Anim. Health Prod.* **45**, 1415–1422 (2013).
- Knight-Jones, T. J. D. & Rushton, J. The economic impacts of foot and mouth disease—What are they, how big are they and where do they occur?. *Prev. Vet. Med.* **112**, 161–173 (2013).
- Zell, R. *et al.* ICTV virus taxonomy profile: *Picornaviridae*. *J. Gen. Virol.* **98**, 2421–2422 (2017).
- Gao, Y., Sun, S. Q. & Guo, H. C. Biological function of foot-and-mouth disease virus non-structural proteins and non-coding elements. *Virol. J.* **13**, 107 (2016).
- Bachrach, H. L. Foot-and-mouth disease. *Annu. Rev. Microbiol.* **22**, 201–244 (1968).
- Marquardt, O. & Adam, K. H. Foot-and-mouth disease virus subtyping by sequencing VP1 genes. *Vet. Microbiol.* **23**, 175–183 (1990).
- Yoon, S. H., Park, W., King, D. P. & Kim, H. Phylogenomics and molecular evolution of foot-and-mouth disease virus. *Mol. Cells* **31**, 413–421 (2011).
- Stram, Y., Molad, T., Chai, D., Gelman, B. & Yadin, H. Detection and subtyping of foot-and-mouth disease virus in infected cattle by polymerase chain reaction and amplified VP1 sequencing. *J. Vet. Diagn. Invest.* **7**, 52–55 (1995).
- Baxt, B., Morgan, D. O., Robertson, B. H. & Timponi, C. A. Epitopes on foot-and-mouth disease virus outer capsid protein VP1 involved in neutralization and cell attachment. *J. Virol.* **51**, 298–305 (1984).
- Alam, S. M. S., Amin, R., Rahman, M. Z., Anwar Hossain, M. & Sultana, M. Antigenic heterogeneity of capsid protein VP1 in foot-and-mouth disease virus (FMDV) serotype asia1. *Adv. Appl. Bioinform. Chem.* **13**, 37–46 (2013).
- Feng, Q. *et al.* Serotype and VP1 gene sequence of a foot-and-mouth disease virus from Hong Kong (2002). *Biochem. Biophys. Res. Commun.* **302**, 715–721 (2003).
- Gomaa Hemida, M., Rizk EL-Ghareeb, W., Al-Hizab, F. & Ibrahim, A. Foot-and-mouth disease virus O/ME-SA/Ind 2001 lineage outbreak in vaccinated Holstein Friesian cattle in Saudi Arabia in 2016. *Vet. Q.* **38**, 88–98 (2018).
- Brito, B. *et al.* A traditional evolutionary history of foot-and-mouth disease viruses in Southeast Asia challenged by analyses of non-structural protein coding sequences. *Sci. Rep.* **8**, 6472 (2018).
- Heath, L., van der Walt, E., Varsani, A. & Martin, D. P. Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J. Virol.* **80**, 11827–11832 (2006).
- Jackson, A. L. *et al.* Mosaic structure of foot-and-mouth disease virus genomes. *J. Gen. Virol.* **88**, 487–492 (2007).
- Ferretti, L. *et al.* Within-host recombination in the foot-and-mouth disease virus genome. *Viruses* **10**, 221 (2018).
- Ferretti, L. *et al.* Pervasive within-host recombination and epistasis as major determinants of the molecular evolution of the foot-and-mouth disease virus capsid. *PLoS Pathog.* **16**, e1008235 (2020).
- Tully, D. C. & Fares, M. A. The tale of a modern animal plague: Tracing the evolutionary history and determining the time-scale for foot and mouth disease virus. *Virology* **382**, 250–256 (2008).
- Ranaweera, L. T. *et al.* Characterization of the FMDV-serotype-O isolates collected during 1962 and 1997 discloses new topotypes, CEY-1 and WCSA-1, and six new lineages. *Sci. Rep.* **9**, 14526 (2019).
- Zhang, Q. *et al.* Evolutionary analysis of structural protein gene VP1 of foot-and-mouth disease virus serotype asia 1. *Sci. World J.* **2015**, 734253 (2015).
- Abeyratne, S. A. E. *et al.* The phylogenetic analysis of VP1 genomic region in foot-and-mouth disease virus serotype O isolates in Sri Lanka reveals the existence of ‘Srl-97’, a newly named endemic lineage. *PLoS ONE* **13**, e0194077 (2018).
- Diab, E. *et al.* Foot-and-mouth disease outbreaks in Egypt during 2013–2014: Molecular characterization of serotypes A, O, and SAT2. *Vet. World* **12**, 190–197 (2019).
- Duchatel, F., Bronsvoort, B. M. C. & Lycett, S. Phylogeographic analysis and identification of factors impacting the diffusion of foot-and-mouth disease virus in Africa. *Front. Ecol. Evol.* **7**, 371 (2019).
- Katoh, K. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
- Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015).

28. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
29. Lemoine, F. *et al.* Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452–456 (2018).
30. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vev007 (2016).
31. Aiewsakun, P. *et al.* Modular nature of simian foamy virus genomes and their evolutionary history. *Virus Evol.* **5**, vev007 (2019).
32. Muslin, C., Kain, A. M., Bessaud, M., Blondel, B. & Delpeyroux, F. Recombination in enteroviruses, a multi-step modular evolutionary process. *Viruses* **11**, 859 (2019).
33. Jamal, S. M. *et al.* Evidence for multiple recombination events within foot-and-mouth disease viruses circulating in West Eurasia. *Transbound. Emerg. Dis.* **67**, 979–993 (2020).
34. Blank, A. De Sympathia et Antipathia Rerum, Liber I./Simpattia, fantasia e contagio: Il pensiero medico e il pensiero filosofico de Girolamo Fracastoro. *Renaiss. Q.* **62**, 996–998 (2009).
35. Blancou, J. & Parsonson, I. Historical perspectives on long distance transport of animals. *Vet. Ital.* **44**, 19–30 (2008).
36. Naranjo, J. & Cosivi, O. Elimination of foot-and-mouth disease in South America: Lessons and challenges. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120381 (2013).
37. Saraiva, V. Foot-and-mouth disease in the Americas: Epidemiology and ecologic changes affecting distribution. *Ann. NY Acad. Sci.* **1026**, 73–78 (2004).
38. Sangula, A. K. *et al.* Evolutionary analysis of foot-and-mouth disease virus serotype SAT 1 isolates from East Africa suggests two independent introductions from southern africa. *BMC Evol. Biol.* **10**, 371 (2010).
39. Hall, M. D., Knowles, N. J., Wadsworth, J., Rambaut, A. & Woolhouse, M. E. J. Reconstructing geographical movements and host species transitions of foot-and-mouth disease virus serotype SAT 2. *MBio* **4**, e00591-e613 (2013).
40. Vosloo, W., Bastos, A. D. S., Sangare, O., Hargreaves, S. K. & Thomson, G. R. Review of the status and control of foot and mouth disease in sub-Saharan Africa. *Rev. Sci. Technol.* **21**, 437–449 (2002).
41. Casey, M. B. *et al.* Patterns of foot-and-mouth disease virus distribution in Africa: The role of livestock and wildlife in virus emergence. In *The Role of Animals in Emerging Viral Diseases* (ed. Johnson, N.) 21–38 (Academic Press, London, 2014).
42. Al-Hosary, A. A. *et al.* Co-infection with different serotypes of FMDV in vaccinated cattle in Southern Egypt. *Virus Genes* **55**, 304–313 (2019).
43. Knowles, N. J. FMD history. *The Pirbright Institute The Picornavirus* (1990). https://www.picornaviridae.com/apthovirus/fmdv/fmd_history.htm.

Acknowledgements

This research project is supported by Mahidol University (A31/2562) and partly funded by Agricultural Research Development Agency (Public Organization), Thailand.

Author contributions

PA, and NP conceived the study; PA designed, and supervised the study; PA, and NP analysed the data; PA, NP, and CI interpreted the results; PA, and NP wrote the manuscript; all reviewed and approved of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-72246-6>.

Correspondence and requests for materials should be addressed to P.A. or C.I.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020