Into the deep (sequence) of the foot-and-mouth disease virus gene pool: bottlenecks and adaptation during infection in naïve and vaccinated cattle

Ian Fish[1,2], Carolina Stenfeldt[1,3], Rachel M. Palinski[1], Steven J. Pauszek[1], Jonathan Arzt[1, #]

[1] Foreign Animal Disease Research Unit, Plum Island Animal Disease Center, ARS, USDA, NY, USA

[2] Oak Ridge Institute for Science and Education, PIADC Research Participation Program, Oak Ridge, TN, USA

[3] College of Veterinary Medicine, Kansas State University, Manhattan, KS, USA

[#]Correspondence: Jonathan.Arzt@ars.usda.gov

Abstract

Foot-and-mouth disease virus (FMDV), like many RNA viruses, infects hosts as a population of closely related viruses referred to as a quasispecies. The behavior of this quasispecies has not been described in detail over the full course of infection in a natural host species. In this study, virus samples taken from vaccinated and non-vaccinated cattle up to 35 days post experimental infection with FMDV A24-Cruzeiro were analyzed by deep-sequencing. Vaccination induced significant differences compared to viruses from non-vaccinated cattle. in virus substitution rates, entropy, and evidence for adaptation. Genomic variation detected during early infection was found to reflect the diversity inherited from the source virus (inoculum), whereas by 12 days post infection (dpi) dominant viruses were defined by newly acquired mutations. In most serially sampled cattle, mutations conferring recognized fitness gain occurred within numerous genetic backgrounds, often associated with selective sweeps. Persistent infections always included multiple FMDV subpopulations, suggesting independently maintained foci of infection within the nasopharyngeal mucosa. Although vaccination prevented disease, subclinical infection in this group was associated with very early bottlenecks which subsequently reduced the diversity within the virus population. This implies an added consequence of vaccination in the control of foot-and-mouth disease. Viruses sampled from both animal cohorts contained putative antigenic escape mutations. However, these mutations occurred during later stages of infection, at which time transmission between animals is less likely to occur.

Importance

Preparedness and control of foot-and-mouth disease virus have substantial, yet distinct implications in endemic and free regions. Viral evolution and emergence of novel strains are of critical concern in both settings. The factors that contribute to the asymptomatic carrier state, a common form of long-term FMDV infection in cattle and other species, are important but not well-understood. This experimental study of foot-and-mouth disease virus in cattle explored the evolution of the pathogen through detailed sampling and analytical methods in both vaccinated and non-vaccinated hosts. Significant differences were identified between the viruses subclinically infecting vaccinated animals and those causing clinical disease in the non-vaccinated cohort. These results can benefit vaccination programs and contribute to the understanding of persistent infection of cattle.

Introduction

Foot-and-mouth disease (FMD) is a viral disease affecting even-toed ungulates, causing economically devastating effects on animal production and international trade [1-3]. Though Europe and the Americas have largely eliminated FMD, the disease remains a substantial concern for livestock farmers in much of the world [4]. The clinical form of FMD manifests with fever, lameness and characteristic vesicular lesions on the feet, oral cavities and teats, thereby negatively impacting animal welfare and herd productivity [2, 5]. The economic impacts associated with decreased production, disease surveillance, trade restrictions and vaccination campaigns comprise a large burden to FMD-endemic and neighboring countries [3].

The etiologic agent of FMD, foot-and-mouth disease virus (FMDV) (family: *Picornaviridae,* genus: *Aphthovirus*) is a single-stranded positive-sense RNA virus with a particularly high mutation rate [6]. The approximately 8.3 kb genome includes a 7 kb open reading frame encoding a polyprotein that is post-translationally processed into four structural and eleven non-structural proteins including two forms of Leader protease (Lpro) and three copies of VPg [5, 7]. High levels of variation, often associated with positive selection and antigenic escape, are commonly detected in the capsid protein coding regions [8-10]. The GH loop in capsid protein VP1 is particularly important for host cell entry and antibody-mediated neutralization [11, 12]. Specifically, the Arg-Gly-Asp (RGD) motif within this GH loop interacts directly with the host cell integrin receptors (e.g. αvβ6) [Reviewed in 13].

Like many RNA viruses, FMDV naturally exists as a population with a complex depth of genetic variation, i.e. as a quasispecies. In recent years, next-generation sequencing technology has enabled subconsensus-level characterization of this complex viral population [14-17]. Viral population diversity and availability of low frequency mutations have been demonstrated to mediate viral swarm adaptability, fitness and ultimately virulence [18-20]. The relevance of the FMDV quasispecies swarm has been extensively investigated in cell culture [reviewed in 21]. Those studies have shown that events that disrupt the quasispecies can affect virus adaptation and virulence. Relating understanding of the FMDV quasispecies *in vitro* to swarm behavior in natural host species is an important goal as it relates to infection progression, differential virulence, and mechanisms mediating immunity in vaccinated and non-vaccinated hosts.

The path to improved FMD countermeasures may depend upon elucidation of viral and host determinants of FMDV evolution. Delineating differences in viral genetic change in response to both primed (vaccinated) and unprimed host humoral immunity are of particular interest as conventional FMD vaccines do not prevent subclinical or persistent infection. In cattle, both primary and persistent infection have been localized to the nasopharyngeal mucosa [22-25]. In non-vaccinated animals, acute FMDV infection lasts approximately one to two weeks, and involves systemic dissemination of virus and transient viremia [26]. The clinical phase of disease is followed by a transitional period during which the virus is completely cleared in a subset of animals that do not maintain persistent infection (herein referred to as terminators) [25]. Persistent FMDV infection is defined by the presence of infectious FMDV in oropharyngeal fluid samples four or more weeks after infection [27]. Appropriately vaccinated animals are generally protected against systemic generalization of virus and clinical FMD. However, these vaccinated animals often become subclinically infected and traverse the corresponding phases of early, transitional, and persistent infection, during which viral replication is fully restricted to the nasopharynx (upper respiratory tract).

The present study investigated the evolution of FMDV populations within natural hosts by examining the deep sequences of viruses sampled through 35 days following experimental infection of naïve and vaccinated cattle [17, 24, 25]. This study identified changes in the viral swarm, the influence of bottlenecks and adaptive immunity and the respective roles of novel subconsensus mutations. The results of this study suggest that vaccination causes earlier bottlenecks in FMDV populations as compared to the viruses present in their non-vaccinated counterparts and that antigenic escape along with other novel mutations, occur during the

105 persistent phase of infection. These findings enhance understanding of FMDV evolution *in vivo*
106 and may contribute to development of improved FMD vaccines.
107
108
109 Results
110
111 *Animal experiments and clinical outcomes*
112 The 10 non-vaccinated cattle included in the study all developed fulminant clinical FMD after
113 virus exposure. The 10 vaccinated animals were protected from clinical disease but were
114 subclinically infected as demonstrated by repeated recovery of virus from clinical samples.
115 Details of clinical symptoms, infection dynamics, and tissue distribution of FMDV in these cattle
116 have been published previously [24, 25]. Half of the animals (5 vaccinated and 5 non-vaccinated)
117 were euthanized between 1 and 10 days post inoculation (dpi) for harvest of tissue samples
118 (Figure S1) while the remainder of the animals were sampled through 35 dpi. Among the 10
119 cattle followed through study-end, seven were determined to be persistently infected with FMDV
120 (carriers) while three individuals (animal IDs 14-108, 14-111, and 14-57) fully cleared infection
121 (terminators) during the transitional phase of infection.
122
123
124 *Effects of vaccination on FMDV populations*
125
126 The rate of FMDV genomic change over time was compared between vaccinated and non-
127 vaccinated cohorts in order to investigate if vaccination induced selective pressures upon the

inoculating virus population which were distinct from those which occurred in naïve (non-vaccinated) animals. In order to quantitate this effect, pairwise differences over time were first calculated between each FMDV sample's consensus sequence and the preceding sequence obtained from the same animal (inoculum used as 0 dpi). Viruses in non-vaccinated animals had higher substitution rates than the viruses in the vaccinated cohort during early- (0.188 vs. 0.131 substitutions/site/year (subs/st/yr)) and transitional (0.127 vs. 0.089 subs/st/yr) phases of infection; however, these differences between groups were not statistically significant (Table 1). While the substitution rates decreased with advancing phase of infection for synonymous and nonsynonymous sites (Table 1, Figure S2), this change was only statistically significant for the non-vaccinated cohort between early infection (0.188 subs/st/yr) and persistent infection (0.080 subs/st/yr, $P < 0.05$). Specifically, nonsynonymous changes had the highest observed substitution rate change between early and persistent phases of infection (Figure S2, $P < 0.005$). These data suggest that the non-vaccinated cattle maintain a large, diverse and dynamic virus population during early infection which evolves more slowly during later stages. During the persistent phase of infection, FMDV genomic rates of change in non-vaccinated and vaccinated cattle were similar.

Table 1. Nucleotide substitution rates and Shannon entropy for FMDV samples

	substitution rate (subs/st/yr)			Shannon entropy			
	early	transitional	persistent	CDS	nonstructural	capsid	capsid / CDS
inoculum	-	-	-	0.0156	0.0147	0.0181	1.16
non-vaccin. cattle	0.188[a]	0.127	0.080[a]	0.0171[b]	0.0160[b]	0.0200[b]	1.17[b]
vaccinated cattle	0.131	0.089	0.079	0.0120[b]	0.0115[b]	0.0130[b]	1.08[b]

[a] = $P < 0.05$; [b] = $P < 0.001$; subs/st/yr: substitutions / site / year. CDS: coding region.

148
149

150 *Entropy*

151 Deep sequence analyses of the sampled FMDV populations allowed for estimation of Shannon
152 entropy. This provided a site-specific quantitation of nucleotide variation for each sampled virus,
153 calculated from aligned sequencing reads and averaged across distinct genomic regions. Entropy
154 for the protein coding region (CDS) of the inoculum was 0.0156, which was greater than that of
155 nearly all vaccinated cattle samples, indicating higher diversity (mean 0.0120, median 0.0110,
156 Table 1 and Figure 1). This evidence aligns with a previous analysis of the multi-haplotypic
157 composition of this inoculum [17]. The CDS average entropy across all samples from non-
158 vaccinated cattle (mean 0.0171, median 0.0169), was similar to the inoculum. The average dpi-
159 matched non-vaccinated sample entropies were significantly ($P < 0.001$) higher than samples
160 from vaccinated cattle across the full CDS, as well as for the capsid and nonstructural coding
161 regions separately (Table 1). Since coding region entropies did not significantly change within
162 either cohort over time (Figure 1), non-vaccinated cattle maintained significantly higher average
163 entropies than vaccinated cattle through all phases of infection (Table 1). These data suggest a
164 strong, early and enduring reduction in FMDV population diversity, i.e. an early bottleneck in
165 the vaccinated animals. In contrast, there was little or no evidence of reduction in diversity or
166 effective population size during initial infection of non-vaccinated cattle (Tables 1 and S2).
167

168 Average site-wise entropy across the capsid coding regions for VP1, VP2, and VP3 was
169 calculated in proportion to the overall CDS entropy (capsid / CDS entropy, Table 1); this allowed
170 for comparison between samples of capsid entropy proportional to each sample's global (CDS)
171 diversity. Averaged across all phases of infection, capsid/CDS entropy values of the viral

	animal ID	DPI	CDS	nonstruc.	capsid	capsid/CDS	avg. coverage
non-vaccinated	14-33	1_dpi	0.0158	0.0150	0.0177	1.13	3174
	14-33	2_dpi	0.0154	0.0142	0.0186	1.21	2034
	14-33	3_dpi	0.0176	0.0165	0.0204	1.16	1298
	14-33	4_dpi	0.0151	0.0138	0.0183	1.22	3280
	14-33	6_dpi	0.0152	0.0140	0.0182	1.20	2730
	14-33	10_dpi_r	0.0133	0.0129	0.0143	1.08	1069
	14-33	21_dpi	0.0154	0.0142	0.0185	1.20	3574
	14-33	28_dpi	0.0173	0.0161	0.0205	1.18	5378
	14-34	1_dpi	0.0151	0.0141	0.0177	1.17	4043
	14-34	2_dpi	0.0146	0.0133	0.0178	1.22	3430
	14-34	3_dpi	0.0183	0.0178	0.0198	1.08	1614
	14-34	4_dpi	0.0145	0.0137	0.0166	1.14	3602
	14-34	5_dpi	0.0160	0.0147	0.0194	1.21	98100
	14-34	6_dpi	0.0169	0.0162	0.0189	1.12	119659
	14-34	7_dpi	0.0175	0.0171	0.0185	1.06	91384
	14-34	8_dpi	0.0160	0.0154	0.0177	1.10	96647
	14-34	9_dpi	0.0176	0.0171	0.0189	1.07	152331
	14-34	28_dpi	0.0179	0.0163	0.0221	1.23	1133
	14-34	31_dpi	0.0171	0.0154	0.0214	1.25	4557
	14-49	2_dpi	0.0261	0.0239	0.0317	1.22	99169
	14-49	6_dpi	0.0215	0.0211	0.0223	1.04	154576
	14-49	10_dpi	0.0210	0.0203	0.0228	1.08	120842
	14-108	0.25_dpi	0.0192	0.0177	0.0230	1.20	3332
	14-108	0.5_dpi	0.0172	0.0155	0.0215	1.25	3554
	14-108	0.75_dpi	0.0149	0.0139	0.0175	1.18	4413
	14-108	1_dpi	0.0203	0.0185	0.0249	1.23	2993
	14-108	3_dpi	0.0200	0.0190	0.0227	1.13	3795
	14-108	5_dpi	0.0122	0.0126	0.0112	0.92	368
	14-108	8_dpi	0.0143	0.0132	0.0172	1.20	3410
	14-108	10_dpi	0.0135	0.0121	0.0172	1.27	715
	14-110	0.17_dpi	0.0172	0.0155	0.0216	1.26	6031
	14-110	0.42_dpi	0.0165	0.0146	0.0213	1.30	826
	14-110	0.75_dpi	0.0195	0.0185	0.0218	1.12	2551
	14-110	0.88_dpi	0.0199	0.0184	0.0238	1.19	1255
	14-110	2_dpi	0.0151	0.0140	0.0177	1.18	4298
	14-110	3_dpi	0.0161	0.0149	0.0191	1.19	3380
	14-110	8_dpi	0.0188	0.0181	0.0206	1.10	4919458
	14-110	14_dpi	0.0118	0.0113	0.0132	1.12	129333
	14-110	17_dpi	0.0159	0.0138	0.0214	1.35	118564
	14-110	28_dpi_r	0.0184	0.0181	0.0191	1.04	1331
	14-110	35_dpi	0.0157	0.0145	0.0188	1.20	2589
	14-111	0.17_dpi	0.0174	0.0159	0.0215	1.23	4149
	14-111	0.42_dpi	0.0182	0.0167	0.0222	1.22	5601
	14-111	0.75_dpi	0.0198	0.0183	0.0234	1.19	3209
	14-111	0.88_dpi	0.0181	0.0169	0.0210	1.16	338
	14-111	2_dpi	0.0195	0.0187	0.0218	1.11	3153
	14-111	3_dpi	0.0224	0.0208	0.0266	1.19	3770
	14-111	7_dpi	0.0164	0.0152	0.0195	1.19	4092
	14-111	8_dpi	0.0162	0.0151	0.0189	1.17	3747
	14-111	9_dpi	0.0148	0.0137	0.0178	1.20	1702
	14-111	10_dpi	0.0162	0.0148	0.0199	1.23	587
		non-vax avg:	**0.0171**	0.0160	0.0200	**1.17**	
vaccinated	14-50	2_dpi	0.0083	0.0080	0.0090	1.09	8634
	14-51	2_dpi	0.0087	0.0085	0.0093	1.06	12933
	14-57	2_dpi	0.0134	0.0126	0.0156	1.17	4839
	14-57	3_dpi	0.0152	0.0144	0.0173	1.14	6938
	15-12	1_dpi	0.0153	0.0143	0.0178	1.16	2275
	15-12	2_dpi	0.0146	0.0134	0.0176	1.21	4373
	15-12	3_dpi	0.0153	0.0148	0.0166	1.09	137552
	15-12	28_dpi	0.0170	0.0166	0.0180	1.06	117464
	15-12	35_dpi	0.0140	0.0135	0.0153	1.10	146728
	15-13	7_dpi	0.0093	0.0092	0.0096	1.02	154975
	15-13	14_dpi	0.0095	0.0093	0.0100	1.05	128923
	15-14	5_dpi	0.0093	0.0093	0.0092	0.99	74267
	15-14	17_dpi	0.0089	0.0088	0.0091	1.03	138526
	15-14	28_dpi	0.0110	0.0107	0.0117	1.06	113413
	1551	8_dpi	0.0095	0.0096	0.0095	0.99	2238
		vax avg:	**0.0120**	0.0115	0.0130	**1.08**	
		all avg:	**0.0159**	0.0149	0.0184	**1.15**	
inoculum			**0.0156**	0.0147	0.0181	**1.16**	10443755
		non-carriers*:	0.0148	0.0138	0.0174	1.17	
		carriers*:	0.0169	0.0162	0.0186	1.11	

Figure 1. Average Shannon entropy in sample FMDV coding regions. Abbreviations: dpi: days post infection, CDS: coding sequence, nonstruc.: nonstructural gene coding regions (including VP4), capsid: VP1, VP2, and VP3 coding regions. Heat coloring indicates relative values with red denoting higher and blue denoting lower. *Non-vaccinated only, dpi-matched (5-14 dpi) samples averaged for non-carriers and carriers. Animals with undetermined carrier status (14-49, 14-50 and 14-51) were sacrificed at 10 dpi or earlier.

172 populations were significantly higher in non-vaccinated animals (1.17, $P < 0.001$) than in
173 vaccinated animals (1.08, Table 1). This suggests that the early immune response, which had
174 been primed in the vaccinated animals, more strongly reduced relative capsid diversity in these
175 hosts and that this took place rapidly following infection; this was also consistent with
176 population bottlenecks. Notably, there was no significant difference in entropy or relative capsid
177 entropy based upon the phase of infection (time) within either cohort ($P > 0.05$). This suggested
178 that reduced FMDV diversity in vaccinated animals was sustained through the phases of
179 infection examined herein. For non-vaccinated animals, this suggests that substantial population
180 diversity is maintained despite the reduction in total virus load associated with the clearance of
181 generalized infection and virus restriction to the nasopharynx during persistent infection. This is
182 consistent with previous reports of FMDV population diversity detected during persistent
183 infection [28, 29].
184
185 *Haplotypic population structure*
186
187 The inoculum used to infect the animals in this study was derived from pooled vesicular lesion
188 samples from multiple cattle which resulted in a highly heterogeneous virus population. As
189 previously reported for a subset of the current samples, multiple haplogroups originating in the
190 inoculum were detected at consensus level within samples derived from different animals at
191 different times after infection [17]. The phylogenetic relationship between these viruses was
192 assessed by maximum likelihood (Figure 2) and six haplogroups (A through F) were assigned
193 based on phylogenetic clustering and inferred ancestral relationships.
194

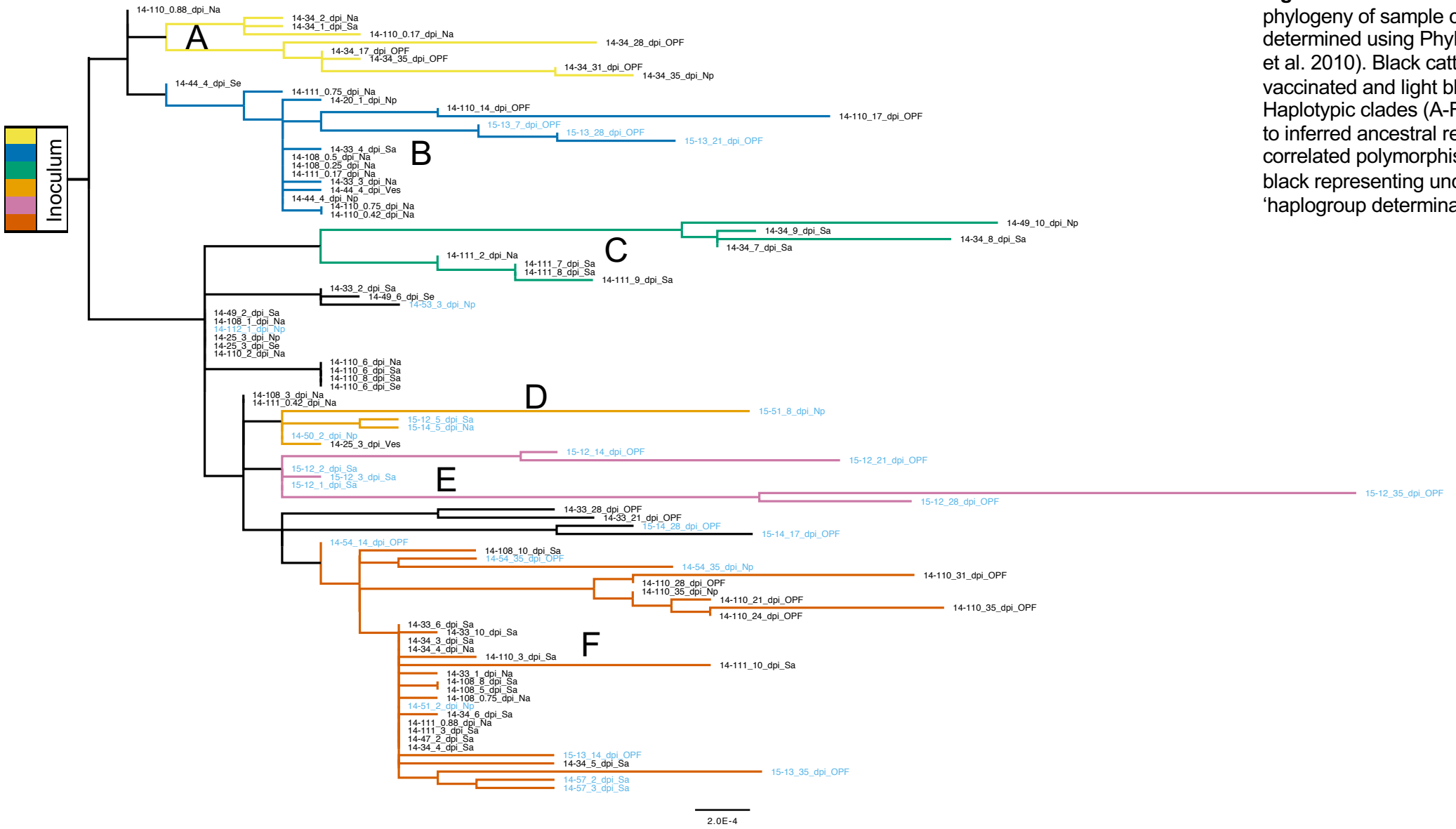


Figure 2. Inoculum-rooted maximum likelihood phylogeny of sample consensus sequences determined using PhyML (Guindon, Dufayard et al. 2010). Black cattle IDs are non-vaccinated and light blue are vaccinated. Haplotypic clades (A-F) are colored according to inferred ancestral relationships and correlated polymorphism frequencies with black representing unclassified (see 'haplogroup determination' in methods).

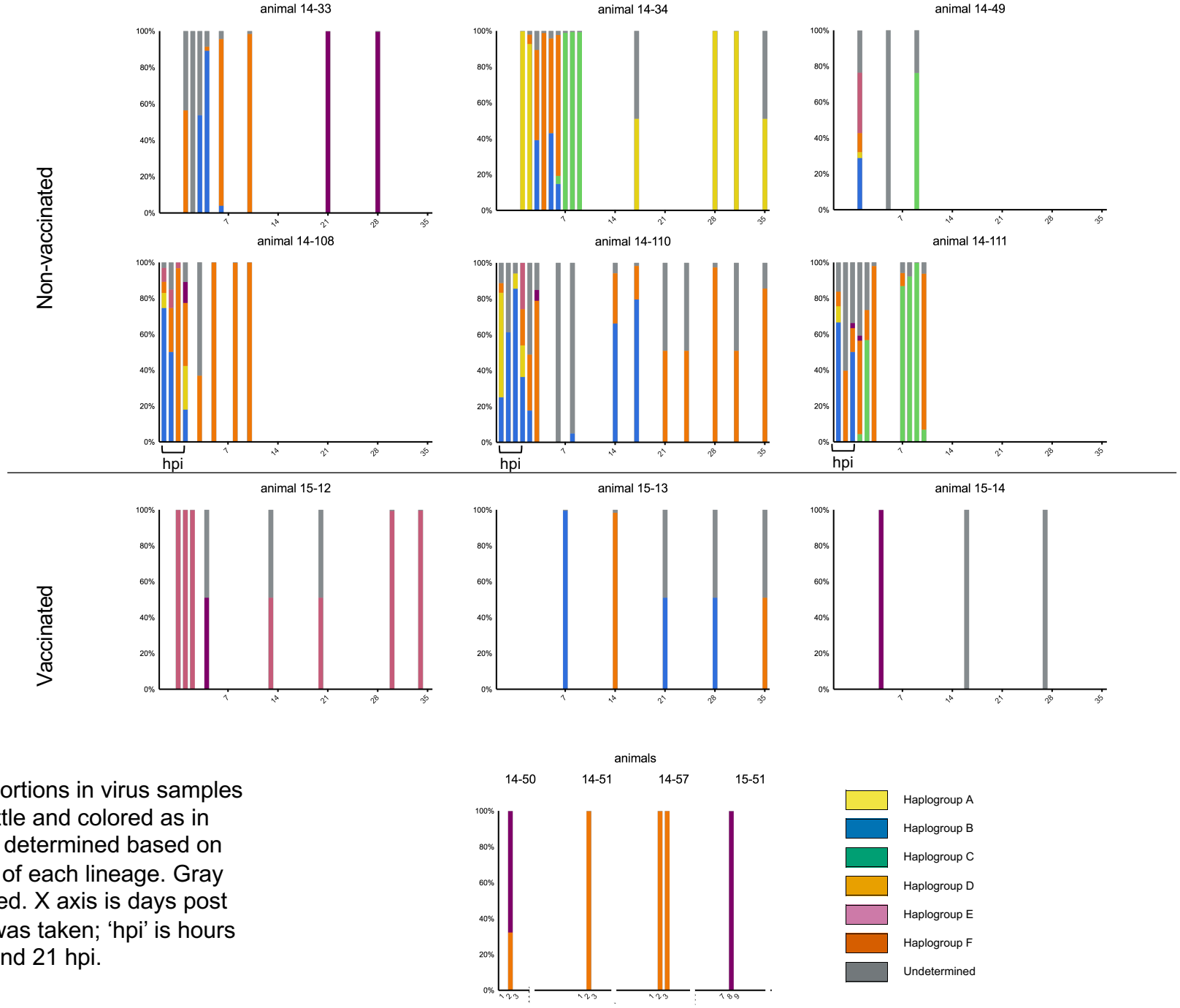


Figure 3. Haplogroup proportions in virus samples from 9 serially-sampled cattle and colored as in Figure 2. Proportions were determined based on SNP profiles characteristic of each lineage. Gray indicates lineage unidentified. X axis is days post infection at which sample was taken; 'hpi' is hours post infection, between 4 and 21 hpi.

Polymorphisms present at ≥ 2% (Table S1) were assessed for all deep-sequenced samples (75 of 103 total samples; see Figure S1). Sets of single nucleotide polymorphisms (SNPs) characteristic of haplogroups A-F were used to classify subpopulations present within samples (Figure 3). Non-vaccinated cattle's acute phase virus populations were highly haplotypically diverse, while vaccinated host samples tended to include only single haplotypes. In addition to haplotypic polymorphism, abundant variation ≥ 2% was regularly detected in virus populations (Table S1). Overall, substantially less genetic diversity was detected in samples derived from vaccinated- as compared to non-vaccinated animals. Samples isolated from non-vaccinated hosts were in daily flux through the early days of infection and regularly included viruses belonging to multiple haplogroups. For example, in animal 14-34, group A viruses dominated samples from 1 and 2 dpi, group B and F viruses co-dominated at 3 dpi and group F viruses dominated at 4 dpi (Figure 3). This is consistent with the high entropy and elevated rates of substitution measured for these animals (Table 1, Figure 1 and Figure S2). In contrast, virus populations in vaccinated cattle typically contained a single haplogroup and less polymorphism overall (Figure 3 and Table S1).

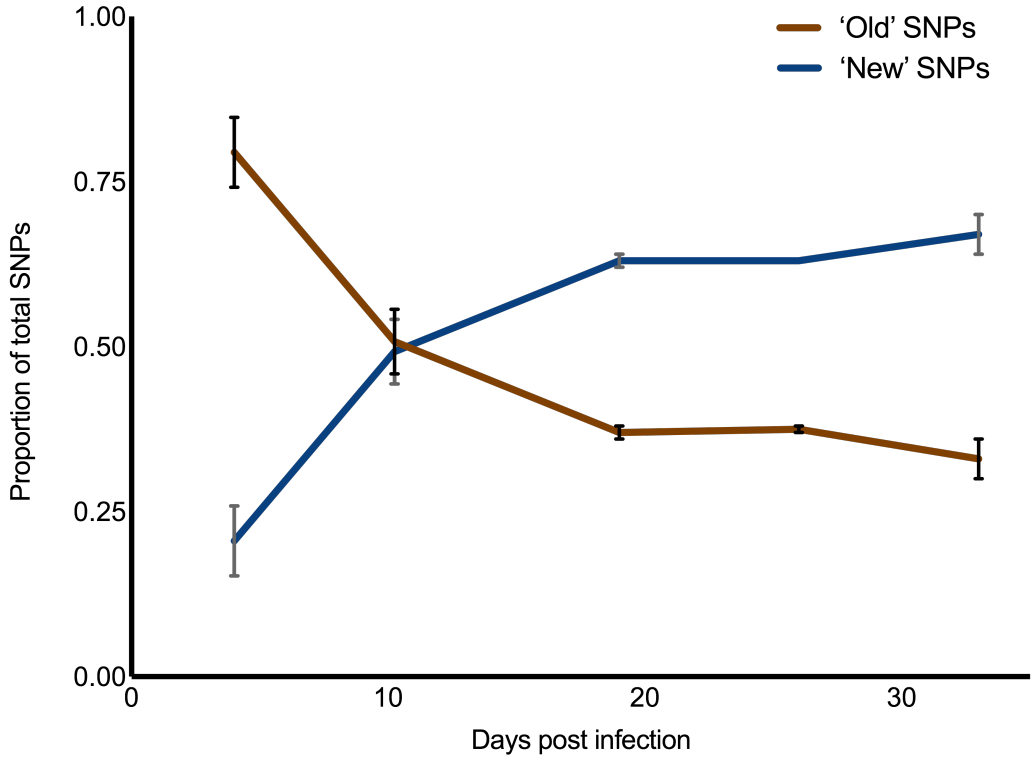
As both animal cohorts (vaccinated and non-vaccinated) progressed to the persistent phase of infection, within-host haplotypic diversity decreased, with each persistent phase sample containing viruses belonging to only a single haplogroup (Figure 3). The one exception to this was animal 15-13's persistent phase samples that sequentially included viruses belonging to either groups B or F. Across animals, no specific haplogroup dominated for any particular time range or cohort, thus no objective fitness advantage was detected between haplogroups. Despite haplotypic stabilization in later stages of infection, population diversity as indicated by subconsensus polymorphism did not decrease over time (Table S1). Rather, population diversity

218 was either maintained or increased through the persistent phase of infection. Specifically, OPF
219 samples at 28 dpi from animals 14-34, 14-110, 15-12, and 15-14 all included substantial
220 assortments of SNPs at frequencies between 10 and 49% (11-24 SNPs each, Table S1).
221
222 *Genomic evolution of the viral swarm*
223
224 The site-specific heterogeneity within the initial infecting population (inoculum) was assessed
225 and compared to consensus sequences of samples collected from the infected animals. Across the
226 inoculum CDS, ultra-deep sequencing (10.4 million reads) indicated 217 variable sites encoding
227 220 single nucleotide variants (SNVs) at frequencies ≥ 0.5%, three of which had multiple
228 polymorphisms at the same site (Table S2). Twenty-seven (12.3%) of these SNVs were
229 identified at the consensus level in cattle samples (Figure 4). All SNPs that were present at high
230 frequency (>10%) in the inoculum, 14 in total, were detected at the consensus level in multiple
231 samples (Table S2); this is consistent with genetic drift contributing to the dominance of specific
232 genotypes. As a gross means of measuring the influence of ancestral variation compared to novel
233 mutation over time, inoculum minority variants (≥ 0.5%) present in sample consensus sequences
234 (ancestral SNPs) on each sampling date were measured proportionally to those SNPs not
235 detected in the inoculum (novel SNPs) (Figure 5). Within the first week of infection, the majority
236 of SNPs in virus consensus sequences were ancestral, i.e. already present in the founding
237 populations. The rate of loss of ancestral SNPs and gain of novel SNPs were comparable
238 between non-vaccinated and vaccinated cattle (not shown). At approximately 10 days after
239 infection, novel SNPs became equally common to ancestral SNPs.
240

Figure 4. Variable FMDV capsid amino acids and positively-selected sites combined. [a] Presence of positively-selected sites identified in individual serially-sampled cattle using mixed effects model of evolution (MEME) analysis (P ≤ 0.10) (Murrell et al. 2012). [b] Amino acids identified as variable between viruses at the consensus level in VP1, VP2, and VP3 capsid coding regions.

	Gene Position	Inoculum Consensus	Variant 1	Variant 2	Variant 3	Variant % in inoculum	MEME[a] N = 9	Naïve[b] N = 10	Vaccinated[b] N = 10
Lpro	11	V	A				1	1	
	22	L	P			4.4	1		
	81	E	K			1.2	1	1	
	107	I	P				1		
	122	H	Y				1		
VP2	44	A	T				1	1	
	65	Y	C	S			1		
	78	L	M						1
	82	E	K						1
	88	H	N				1		1
	131	E	G						1
VP3	65	T	N				1		
	70	D	A	N	G		2	2	
	99	T	A				1	1	
	111	F	V				1		
	131	E	K	G		23 (K)		7	1
	175	T	K						2
	220	Q	R					2	
VP1	96	S	T					1	
	131	N	S				1		
	142	G	R				1	1	
	144	S	R	G			3	5	2
	147	T	M			48.2	1	9	9
	149	S	A	Y	P		4		
	155	V	A	M		3.2 (A)		2	2
	160	A	V				2	1	
	172	I	V				1		
	196	S	F					1	
	197	S	L					1	
	199	D	G					1	1
2B	85	L	V				1	1	
2C	59	D	A				1		
	136	D	E				1	2	
	282	D	H				1		
	283	V	M			49.8	1	8	1
	288	P	L				1	1	
3A	94+	Insertion	Insertion				1		
	136	N	D			35.9	6	7	8
	146	A	V				1		
	147	E	A				1	1	
3B	3	Y	H			0.6	1		
	7	L	R				1		
3C	66	D	D			11.4	1		1
3D	3	I	V				1		1
	138	T	I				1		2

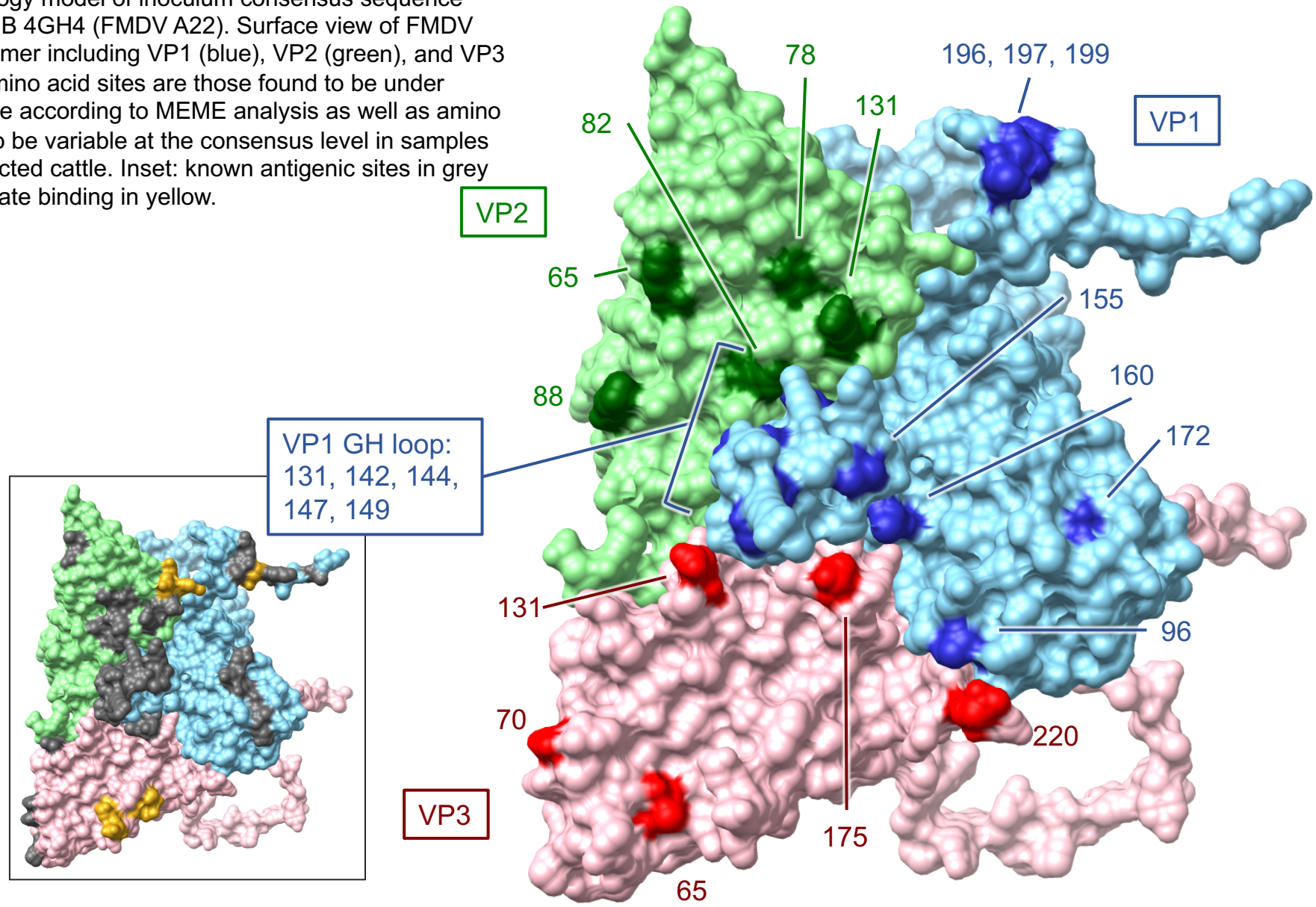
Figure 5. Proportion of inherited polymorphism



241 *Nonsynonymous substitutions*

242 Genetic variations that resulted in amino acid substitutions in sampled viruses were examined at
243 the consensus and subconsensus level. Focusing on the regions encoding the capsid proteins
244 VP2, VP3 and VP1 at the consensus level, 31 sites encoded amino acid changes yet only 3 of
245 these were polymorphic in the inoculum (Figure 4 and Table S2). Homology modeling
246 implicated 15 of these sites as putative antigenic targets based on published data [30-36]: VP2-
247 78, 82, 88, and 131; VP3-70, 131, and 175; VP1-96, 142, 144, 147, 155, 196, 197, and 199
248 (Figure 6). In order to detect residues with evidence of positive selective pressure, mixed effects
249 model of evolution (MEME) analysis was run collectively (per host) on FMDV haplotypes
250 reconstructed from serially-sampled deep sequence [37]. Haplotypes present in each sample were
251 resolved at relative frequencies > 0.5% or > 2.0% with ViQuaS [38]. The presence of positively-
252 selected amino acid changes predicted my MEME ($P \leq 0.10$) are tallied for each host (Figure 4).
253 The majority of consensus-level nonsynonymous substitutions and those detected in MEME
254 analysis were detected only transiently (did not become fixed) and none were detected
255 consistently in either vaccinated or naïve groups.
256
257 The greatest quantity of residues with evidence of positive selective pressure was in VP1,
258 specifically in the GH loop (Figure 4, Figure 6). The canonical FMDV receptor in cattle is
259 integrin αvβ6, which binds to a conserved RGD motif within the GH loop of VP1 [12, 39]. The
260 vaccine used in this study was an adenovirus-vectored recombinant which encoded RGD at the
261 anti-receptor motif. However, the inoculum encoded an SGD motif at this locus (residues 144-
262 146) with no evidence of an RGD virus ≥ 0.5% in the population (Table S2). The earliest
263 detection of RGD at this site was in animal 14-34 at 6 dpi, in which 4.4% of the-virus population

Figure 6. Homology model of inoculum consensus sequence with template PDB 4GH4 (FMDV A22). Surface view of FMDV A24 capsid protomer including VP1 (blue), VP2 (green), and VP3 (red). Labeled amino acid sites are those found to be under selective pressure according to MEME analysis as well as amino acids identified to be variable at the consensus level in samples derived from infected cattle. Inset: known antigenic sites in grey and heparan sulfate binding in yellow.



had the arginine substitution. The following day, 99.3% of the virus population sampled in this host encoded RGD at the GH loop. Ultimately, VP1-S144R came to fixation in all cattle sampled in the persistent phase except for two vaccinated individuals (15-12 and 15-13); this substitution was also commonly observed in persistent-phase cattle samples in a previous study that used the same inoculum [40]. The latest initial detection of a dominant RGD genome was in animal 14-33, in which it emerged between 10 dpi (≤ 2%) and 21 dpi (99.6%). The VP1-S144R substitution was separately encoded by 2 of the 3 possible S>R codon changes (AGU>CGU and AGU>AGA) in 5 of the 6 characterized haplotypic backgrounds and one uncategorized genotype (15-14 at 17 and 28 dpi). There was also evidence of multiple independent RGD subpopulations co-infecting hosts. The subconsensus variants in animal 14-110 samples at 14 and 17 dpi included dozens of intermediate-frequency (10-50%) SNPs indicative of many different viruses alongside a fixed (99.8%) VP1-S144R substitution (Table S1). At the consensus level, a distinct shift in dominance from one haplogroup (group B) to another (group F) was evident between 17 dpi and 21 dpi, differing by 23 SNPs, (Figure 2).

There was strong evidence suggesting that the selection for RGD viruses resulted in global reductions in population diversity and introduction of novelty through genetic draft (i.e. resultant selective sweeps). The clearest evidence of this was the dominance of haplogroup C viruses in animals 14-34 and 14-49, in which a series of changes in the coding region for 2C - T43M…H84N…D136E…I248T co-emerged with VP1-S144R at corresponding proportions (Tables S1 and Figure S3). Selective sweeps associated with RGD genome emergences were also evident in animal 15-14 at 17 dpi (VP3-A75V and VP1-I35V) as well as animal 14-33 at 21 dpi (VP1-G33S) (Table S1 and Figure S3). In contrast, animals 15-12 and 15-13 maintained SGD viruses through study end. Interestingly, this was associated with consensus-level changes

287 indicative of antigenic escape. In 15-12, 21 dpi FMDV samples had a qualitatively divergent
288 VP2-E82K capsid substitution and subsequent 28 and 35 dpi samples had VP2-H88N and VP1-
289 V155A substitutions. In 15-13, 21 and 28 dpi viruses had VP3-E131K capsid substitution and
290 were followed at 35 dpi by variants with dominant VP2-E131G and VP3-E131G substitutions.
291 Each of these amino acid changes involved electrostatic shifts on the capsid surface in known
292 antigenic regions (Figure 6).
293 In the present study, there were relatively few sites with evidence of adaptation within
294 nonstructural proteins. The most common replacement in nonstructural regions found across all
295 animals was 3A-N136D, which was present in 15 of 20 animals. Although this replacement was
296 relatively common in the inoculum at 35.9%, its presence as fixed or in the final sample
297 consensus of 5/7 persistently infected animals, suggested an adaptive advantage. MEME analysis
298 identified several sites under selective pressure in coding region for the C-terminus of 3A, with
299 3A-N136D identified in more cattle than any other substitution (Figure 4). In contrast, other
300 commonly variable consensus-level amino acid changes such as 2C-V283M and 3A-S117N, had
301 no evidence of adaptive value in that there was little or no predilection for fixation.
302
303 *Persistent infection & the nasopharynx*
304
305 The hypothesis that viral population diversity correlated with the establishment or maintenance
306 of persistent FMDV infection was tested. Viruses characterized from terminators (14-108, 14-
307 111, and 14-57) did not significantly differ from dpi-matched viruses sampled from persistently
308 infected carriers in substitution rate (4.29 vs. 3.47 subs/day, Figure S2) or global entropy (Figure
309 1). Entropy was lower in viruses sampled from terminators compared to dpi-matched samples

from carriers (Figure 1). This difference was only statistically significant within the nonstructural protein coding regions, indicating more conservation (purifying selection) in these regions of the FMDV genomes in terminators than in carriers. While very few consensus-level amino acid changes were identified in non-carrier viruses, the mutation VP3-Q220R, located at the VP3-VP1 cleavage site and on the capsid surface, was uniquely identified in samples derived from these cattle at 10 dpi.

Discussion

Although the quasispecies character of RNA viruses have been known for many years, investigation during infection of natural hosts considering the inter-relationships with hosts' biological processes have received less attention. In this study, samples collected from FMDV-vaccinated and non-vaccinated cattle over 5 weeks were analyzed by deep-sequencing in order to investigate viral population dynamics in a natural host over all phases of infection. Host vaccination status was associated with significant differences in virus substitution rates, entropy, and evidence for adaptation. While both vaccinated and non-vaccinated cohorts established and maintained FMDV infection at similar prevalence [25, 41], deep sequence evaluation of the sampled viruses clearly demonstrated a narrow population bottleneck during early infection of vaccinated animals and contrasted the absence of population contraction in the non-vaccinated hosts. Specifically, measures of global Shannon entropy and rates of consensus-level nucleotide substitution were consistently lower through the early phase of infection in virus populations sampled from vaccinated versus non-vaccinated cattle (Table 1, Table S1, Figure S2).

333 Additionally, virus populations in samples from the early phase of infection in non-vaccinated
334 animals were composed of multiple haplogroups while those in samples from vaccinated cattle
335 belonged to single lineages. However, detection of endurance of the multiple founding lineages
336 became increasingly uncommon in non-vaccinated hosts over time and by 21 dpi, nearly all
337 samples included only single haplogroups. This pattern is consistent with bottlenecks observed in
338 the transition from acute to chronic stages of infection with hepatitis C virus [42] and human
339 immunodeficiency virus 1 [43]. Those studies reported that acute infection ended with a
340 bottleneck of the multiple lineages that had thus far co-existed, leaving only viruses of a single
341 lineage during chronic infection. In the present study, this decrease in haplogroup heterogeneity
342 during persistent infection was accompanied by intra-haplotypic diversification as evidenced by
343 abundant subconsensus polymorphism and sustained entropy measures.
344
345 The early phase of infection in both non-vaccinated and vaccinated hosts are situations in which
346 virus adaptation is hypothesized to be low [reviewed in 44]. For viruses infecting non-vaccinated
347 cattle, large population sizes combined with a naïve (non-primed) immune response [24, 41]
348 creates a context in which fitness differences between variants are minimized, leading to
349 relatively unconstrained virus propagation. In the current study, these conditions contributed to
350 relaxed selective pressure which resulted in high virus entropy and maintenance of inherited
351 variation within virus populations. In vaccinated cattle, a strong primed immune response [24]
352 provided efficient restriction of most of the variants within the inoculated virus as illustrated by
353 low entropy and low global variation as compared to non-vaccinates. Importantly, this restriction
354 eliminated low-frequency mutants with potential adaptive value. Thus, there was no evidence for
355 viral adaptation in either cohort during the early stages of infection. This is consistent with a

356 previous analysis of FMDV minority variance within serial samples taken from three FMDV-
357 infected cattle during the acute phase of infection which suggested genetic drift as the primary
358 mechanism of FMDV evolution both within and between hosts [45]. Of importance to FMDV
359 epidemiology, these findings suggest that variation detected in field samples from FMDV
360 outbreaks (acute phase) is most likely the result of stochastic processes e.g. genetic drift and
361 transmission bottlenecks.
362
363 Nonetheless, virus evolution did take place within the scope of this study, most notably in
364 genomic regions encoding capsid proteins. The canonical FMDV receptor in cattle is integrin
365 αvβ6, which the virus utilizes for host cell entry by binding to a conserved RGD motif within the
366 GH loop of capsid protein VP1 (residues 144-146) [12, 39]. The inoculum used in the current set
367 of experiments was known to instead encode an SGD motif at this locus [17] as a result of prior
368 passages in bovine tongue epithelium (see methods for details). Mutation of the SGD motif to
369 RGD is fitness-enhancing as it facilitates integrin binding and host cell entry in all cattle.
370 Specifically, experimental work has demonstrated that VP1-144 serine to arginine substitution
371 allows for improved cell to cell transmission [46]. In the current study, RGD fixation took place
372 in every non-vaccinated animal sampled after 10 dpi (5/5) yet in only half of vaccinated animals
373 through persistent-phase to study end (2/4). This suggests that even though vaccination does not
374 prevent subclinical or persistent infection, it can provide improved protection from critical viral
375 adaptations (in this study, SGD->RGD). This represents an important benefit not typically
376 attributed to vaccines, i.e. impeding the accrual of mutations which might be beneficial to the
377 virus.
378

379 Experimental works have shown that escape mutations were likely to arise at GH loop residues
380 near to a conserved RGD motif, including under GH loop-specific monoclonal antibody
381 neutralization [47-51]. In the current study, MEME analysis identified numerous sites in
382 antigenic capsid regions suggestive of escape mutations. The majority of the sites identified by
383 MEME were detected at low subconsensus frequencies and were not shared across animals. This
384 type of low-frequency variation, independent of vaccination status or phase of infection in this
385 study, is consistent with mutant swarm character described as central to the concept of FMDV
386 quasispecies [6, 52].
387
388 In six serially sampled cattle, MEME identified residue 136 in the 3A protein to have evolved
389 under positive selective pressure. 3A is a membrane-integrated protein that interacts with FMDV
390 RNA polymerase and, although not fully understood, has been implicated in intracellular
391 transport [53, 54]. The C-terminus of 3A has been associated with host-specific adaptation of
392 FMDV. Specific deletions within this region have been associated with clinical attenuation in
393 cattle [55-57] while virulence in pigs is maintained [58, 59].
394
395 In order to test the hypothesis that FMDV clearance was associated with specific virus
396 population characteristics, samples from terminators were examined for consistent trends.
397 Viruses isolated from these hosts during their later stages of infection had reduced global entropy
398 when compared to both earlier samples from those same animals and comparable carrier viruses.
399 Specifically, virus sampled from animal 14-108 at 10 dpi encoded a fixed haplotype (group F)
400 with low entropy and no polymorphism; these population characteristics are consistent with a
401 swarm that was nearing extinction. Viruses from animal 14-111 also had reduced global entropy

without ever acquiring the VP1-144R that was detected in all other viruses isolated from non-vaccinated cattle, possibly making the virus more vulnerable to clearance despite its two divergent lineages (groups B and F). A single amino acid replacement shared exclusively by terminators was VP3 Q220R. While this site has been shown to be prone to variation, including glutamine and arginine [30, 60], and thus unlikely deleterious, it merits further investigation. These results suggest that reduced diversity of viral populations may contribute to termination of infection, consistent with the concept that the mutant spectrum is important to viral fitness [21, 61]. Specific host immunological profiles [41] were not associated with observed viral genomic changes. Although the transitional phase is the period during which persistence-determining events are hypothesized to take place [25], the mechanisms responsible for FMDV persistence were not clearly established herein. Future investigation focused on the FMDV quasispecies and host factors during the transitional period in terminators may elucidate the critical factors that determine viral clearance.

The particularly rapid mutation rates of RNA viruses can facilitate responses to changing adaptive host immunity though the selection for escape mutants at antigenic epitopes [62-65]. We hypothesized that as FMDV evolved within each host, the virus would acquire escape mutations as a result of actuated humoral and cellular immune responses. However, based upon comparative literature- and homology model-based SNP analysis, there was limited expansion of antigenic diversity observed in the majority (5/7) of animals that were sampled through the persistent phase of infection. Each of these five cattle acquired populations with fixed (> 98%) RGD genomes associated with selective sweeps. Such sweeps include the clearance of ancestral and novel low-frequency variation as well as genetic draft, both of which were observed in these

425 cattle. Loss of low-frequency variation provides a plausible explanation for the limited detection
426 of antigenic variation in these cattle while genetic draft may explain the acquisition of irregular
427 changes (e.g. 2C substitutions in haplogroup C members) that occurred synchronously with RGD
428 replacement. Although previous studies have demonstrated that continual changes to FMDV
429 capsids occur over longer time courses of persistent infection [66, 67], this could not be
430 addressed in the current study.
431
432 Notably, there is some suggestion from the current findings that selective pressure on the virus
433 may also be reduced during FMDV persistence. Specifically, capsid/CDS entropy did not
434 significantly differ between phases of infection (in any cohort), nor did subconsensus or
435 consensus amino acid replacements indicate strong adaptation. This is consistent with previous
436 findings in Cape buffalo that autologous antibody neutralization of FMDV does not change
437 throughout persistent infection [29] and other reports demonstrating that the nasopharyngeal
438 mucosa may function as an immunoprivileged or immunosuppressed site, supported by gene
439 expression patterns suggesting a down-regulated anti-viral response [68, 69]. This privileged
440 state may in effect relax selective pressure, thus further limiting persistent virus escape
441 adaptation. A goal of our ongoing research is to integrate sub-anatomic host tissue features and
442 signaling patterns with FMDV subconsensus variation.
443
444 In contrast to the SGD-RGD transformation which occurred in most animals, viruses isolated
445 from vaccinated animals 15-12 and 15-13 never acquired the RGD motif nor were these
446 populations affected by associated selective sweeps. Viruses in these two hosts were thus more
447 capable of acquiring predicted escape substitutions featuring substantial electrostatic changes

448 within the time frame of the study. Specifically, persistent phase viruses included replacements
449 at VP2-82 and -88, VP3-131 and VP1-155. An alternative hypothesis for capsid mutation
450 particular to SGD viruses is adaptation to an alternative host cell receptor, such as heparan
451 sulfate [70]. However, evaluation of amino acid changes in capsid structural models did not
452 support any of these lying in the heparan sulfate binding site (Figure 6 inset) [31]; nonetheless,
453 efficiency to bind other integrins or alternate receptors may be involved [46, 71]. The extent to
454 which the recombinant RGD vaccine prevented emergence of the RGD motif in 2 vaccinated
455 animals could not be verified within the current study design.
456
457 Persistent FMDV infection in cattle is restricted to distinct epithelial foci within the
458 nasopharyngeal mucosa [23, 25, 72]. If these foci represent distinct viral subpopulations, this
459 should be reflected in the deep sequence of oropharyngeal fluid samples which are retrieved with
460 a probang cup, which harvests cells from multiple regions of the pharynx epithelium. Each OPF
461 sample included at least 2 subpopulations in all hosts. These subpopulations may represent sub-
462 anatomic vicariance at distinct epithelial foci, i.e. viruses infect cells and replicate in isolated
463 groups, leading to genomically identifiable subgroups. Coexistent persistent-phase FMDVs
464 belonging to different lineages has previously been reported for cattle [73, 74]. Genomic RNA
465 and in some cases, infectious virus, belonging to multiple FMDV serotypes have been detected
466 in subclinically infected Cape buffalo (*Syncerus caffer*) and water buffalo (*Bubalus bubalis*) [28,
467 75, 76]. Coexistent viruses are a prerequisite for recombination, which has been demonstrated to
468 play a role in the evolution of FMDV [28, 77, 78]. The divergence among the viruses sampled
469 herein was inadequate for recombination detection. Nonetheless, inability to determine

haplogroups in late-stage samples from hosts 15-14 and 14-33 was due to ambiguity of clade-informative SNPs, which may be a result of recombination.

These findings have important implications for the inter-relationship between FMDV within-host evolution and transmission. In both vaccinated and non-vaccinated hosts, variation detected during the first few days of infection appears not to be driven by selective (immunological) pressures. Novel mutations, while highly-adaptive, took at least one week to reach consensus level. Because most transmission of FMDV is believed to occur within the first few days of infection [79, 80], these novel mutations would have a narrow chance of being passed on within this window of transmission. If this course of evolution is typical throughout chains of transmission, it follows that nearly all FMDV genomic change observed in field isolates is the result of purifying and neutral evolution, as has been suggested for the virus [45, 74, 81, 82]. For those low-frequency adaptive SNPs that are successfully transmitted, neutral or weak-purifying selection within this window would not favor them through extended chains of transmission.

Summary and Conclusion

Subconsensus variation in FMDV populations were investigated in vaccinated and naïve cattle for 35 days following simulated natural infections. FMDV genomic change detected during early infection was consistent with neutral evolution all cattle. A critical capsid adaptation at the site of host cell entry, VP1 S144R, came to fixation in most animals and was associated with selective sweeps; putative antigenic escape mutations only arose in vaccinated animals within the time frame of the study. Furthermore, during early infection, vaccination caused virus population

494 bottlenecks which did not occur in the naïve cattle. This differential quasispecies behavior in
495 vaccinated hosts may provide insights into further enhancement of countermeasures to impede
496 viral propagation at the individual animal level. Additionally, multiple subpopulations were
497 present in viruses recovered during the persistent phase, consistent with distinct foci of FMDV
498 infection in nasopharyngeal epithelial cells, furthering our understanding of the nature of
499 persistent infection. These findings contribute novel insights to the evolution of FMDV in natural
500 host species.
501
502 **Methods**
503
504 *Animal studies*
505 The animal experiments were part of a multi-study analysis of the FMDV carrier state described
506 in previous publications [24, 25, 41]. All studies were carried out at Plum Island Animal Disease
507 Center, New York under BSL-3Ag conditions and with approval from the Plum Island Animal
508 Disease Center Institutional Animal Care and Use Committee (protocol 209-13). Briefly, a group
509 of steers were vaccinated with a recombinant adenovirus-vectored FMDV A vaccine 2-weeks
510 prior to intra-nasopharygeal inoculation with FMDV A24 Cruzeiro. In parallel, a group of non-
511 vaccinated animals were inoculated with the same FMDV A24 Cruzeiro inoculum. Animals
512 were sacrificed at predetermined time points, up to 35 days post inoculation and tissues were
513 harvested for analysis.
514 Three distinct phases, namely, the early, transitional and persistent phases, define FMDV
515 infection in livestock [25, 83]. The phases vary between animals of different immune statuses i.e.
516 vaccinated or non-vaccinated animals. Non-vaccinated cattle undergo clinical and systemic

517 disease in the early (acute) period lasting approximately 1-9 dpi while vaccinated cattle remain
518 subclinically infected yet shedding virus between 1-7 dpi [84]. The transitional phase is
519 associated with a reduction of clinical signs (if present) and either completely clearing of
520 infection or 'transitioning' from early to persistent infection. The transitional phase in vaccinated
521 and non-vaccinated animals occurs approximately between 7-14 dpi and 10-21 dpi, respectively.
522 Entry into the persistent phase of infection is associated with subclinical FMDV replication in
523 the nasopharygeal mucosa if the infection was not cleared in the transitional phase.
524 Antemortem samples collected from these animals included oral swabs, nasal swabs, serum, and
525 oropharyngeal fluid (OPF) harvested using a probang cup [83]. Postmortem vesicular lesions
526 (Ves) or nasopharygeal mucosa (Np) were collected at necropsy. Host factors including
527 immunoglobulin and transcriptomic data were analyzed in prior publications [24, 25, 41]. The
528 inoculum and a total of 103 samples from 20 animals were included in the present analysis; 77
529 samples originated from 10 non-vaccinated animals and 26 from 10 vaccinated animals.
530
531 *Inoculum*
532 The FMDV A24 Cruzeiro (GenBank # SRP149342) inoculum was derived from a field strain
533 passaged once in BHK-21 cells and twice in cattle as previously described [17]. The first bovine
534 passage consisted of harvested vesicular epithelium and vesicular fluid obtained at 48 hours post
535 tongue inoculation of two animals. The filtered suspension generated from the harvested material
536 was subsequently used to inoculate a second cohort of three cattle. Vesicular fluid and
537 epithelium were again harvested at 48 hours post inoculation and processed (macerated and
538 filtered) to generate the virus suspension that was used to infect all animals in the present work.
539 The inoculum was aliquoted and stored at -70ºC until use, at which time 10^5 $BTID_{50}$ (50%

infectious does titrated in bovine tongue epithelium) [27] was used for inoculation in the current study.

Sequencing

Illumina-derived deep sequence was examined for 75 of the total 103 virus samples (NCBI PRJNA473786). The consensus sequences of 52 of these samples were previously published, (GenBank MH426523-74) [17]. Nine samples were not passaged, noted as 'raw' in Figure S1 and Figure S3, and 94 samples were passaged once in LFBK- αvβ6 cells [85]. Viral RNA was extracted using the MagMAX RNA Isolation Kit (Thermo Fisher Scientific), reverse-transcribed and amplified, generating three overlapping amplicons covering the full CDS. Sequencing libraries were prepared with the Nextera XT DNA Library Prep Kit (Illumina, USA) and sequenced on the Illumina NextSeq 500 platform. All reads were quality-filtered, primer-trimmed and mapped to the inoculum consensus sequence in CLC Genomics Workbench v. 10 (www.qiagenbioinformatics.com). Read coverage of samples for which deep sequence was included in this study ranged from 338 – 155,000 (mean 30,000) averaged across the CDS. The inoculum deep sequence run totaled 10.4 million reads with a minimum coverage of 72,700 across the CDS.

Consensus-level sequence analysis

Alignments, pairwise distances and the maximum likelihood phylogeny (PhyML 3.2 [86]) were evaluated in MEGA 7.0 [87] and Geneious 7.1 (www.geneious.com [88]). Substitution rates were calculated by tabulating pairwise nucleotide differences between each consensus sequence and the preceding sample sequence as a function of elapsed time between the two sample

563 acquisitions. In cases for which there were multiple samples from the same animal on the same
564 date (differing only by sample type or passage history), values were averaged. Statistical
565 significance of differences between rates was calculated via T-test and Rank-sum, with the
566 higher value of the two used as P in associated text, figures and tables.
567
568 *Subconsensus sequence analysis*
569 The Low Frequency Variant Detection tool in CLC Genomics Workbench was utilized to
570 determine variant sites present within each deep-sequenced sample present in > 2% of mapped
571 reads with a minimum coverage of 20 reads and .75 strand-bias filter. For the inoculum, variants
572 present ≥ 0.5% were determined. Consensus-level sample substitutions that matched SNVs
573 present at ≥ 0.5% in the inoculum deep sequence were categorized as ancestral SNPs, having
574 most likely been present in an ancestral genome (i.e. transmitted in the inoculum gene pool). The
575 remaining substitutions, those not detected in the inoculum ≥ 0.5%, were classified as novel
576 SNPs, more likely to have resulted from within-host *de novo* mutation. Shannon entropy was
577 calculated from quality-filtered and primer-trimmed reads in natural log units with a custom
578 script.
579
580 *Test of diversifying selection*
581 Sites with evidence of having evolved under positive selective pressure in FMDV populations
582 within hosts over time were determined with mixed effects model of evolution (MEME) analysis
583 in the HyPhy package [89]. MEME analysis of each individual host identified sites encoding
584 amino acid changes that significantly deviate from those that occur under neutral models of
585 evolution; positively selected sites of statistical significance (P ≤ 0.10) are included in Figure 4.

586 In order to incorporate subconsensus variation and linkage between low-frequency variants in
587 MEME, haplotypes were reconstructed with the ViQuaS pipeline [38] with *SSAKE* [90]
588 parameters o = 5, r = 0.75. This pipeline reconstructs the haplotypic composition present within
589 each sample present either > 0.5% for cattle 14-33, 14-49, 14-108, 14-111, 15-12, 15-13, and 15-
590 14 or > 2% for cattle 14-34 and 14-110 thresholds from quality-filtered, primer-trimmed deep
591 sequence reads.
592
593 *Haplotypic composition of sample populations*
594 In order to characterize FMDV lineages that made up each sample population, haplogroup-
595 specific SNPs were first inferred from consensus sequence-derived phylogenetic relationships.
596 The presence of these characteristic (haplotypic) SNPs dictated the subpopulation in which the
597 sample was classified (Figure 3). Idealized criteria for these SNPs: i) shared with all members of
598 a lineage, ii) present in identical consensus sequenced derived from different animals, iii)
599 detected at proportionate frequencies at the subconsensus level and iv) present in at least one
600 homogeneous sample. For most samples, the dominant (majority) virus was identified by the
601 sample consensus and location in the phylogeny. This approach is exemplified with sample 14-
602 34_6_dpi (Table S1ex.), where haplogroup F characteristic SNPs (orange) are represented by
603 approximately 79% of reads, group A (blue) includes SNPs ranging from 14.5 – 20.3% and
604 group C (green) by ~4.5% of reads. Relative haplotype frequencies within each sample were
605 established using the lowest SNP frequency among each haplotype's characteristic SNPs. These
606 frequencies were used to construct proportional stacked bar graphs (Figure 3); samples lacking
607 deep sequence data had 49% of the population designated undetermined.
608

609 *Protein structure*

610 The inoculum capsid protomer homology modeling was executed in SWISS-MODEL
611 (swissmodel.expasy.org) with an FMDV A22 (PDB 4GH4) template. UCSF Chimera 1.13
612 (www.cgl.ucsf.edu/chimera [91]) was used for annotation and imaging. Annotation of antigenic
613 regions and heparan sulfate binding site are based upon published works [31, 34-36].

614

615 *Data availability*

616 All new sequence data have been made available as sequence read archive (SRA) files at the
617 National Center for Biotechnology Information (NCBI) under SAMN10280742-861. Previously
618 published sequence data included in the present work is also available at NCBI, GenBank
619 MH426523-74.

620

621 *Acknowledgements*

622 This research was funded in part by United States Department of Agriculture (USDA) ARS-
623 CRIS Project 1940-32000-061-00D. IF is the recipient of a Plum Island Animal Disease Center
624 Research Participation Program fellowship, administered by the Oak Ridge Institute for Science
625 and Education (ORISE) through an interagency agreement between the United States
626 Department of Energy (DOE) and the USDA. All opinions expressed in this paper are the
627 authors' and do not necessarily reflect the policies and views of the USDA, DOE, or ORISE. The
628 authors would like to thank GR Smoliga and EJ Hartwig for support with sample processing, G
629 Belsham for manuscript review and valuable insights, WM Fischer for his Shannon entropy
630 script, and D Rai, E Rieder, T Goldberg, BP Brito and LL Rodriguez for insightful discussions
631 and suggestions towards the foundations of this project.

632

633 Figure / Table Captions:

634

635 Table 1. ^{a} = $P < 0.05$; ^{b} = $P < 0.001$; subs/st/yr: substitutions / site / year. CDS: coding region.

636

637 Figure 1. Average Shannon entropy in sample FMDV coding regions. DPI: days post infection,
638 CDS: coding sequence, nonstruc.: nonstructural protein coding regions (including VP4), capsid:
639 VP1, VP2, and VP3 coding regions. Heat coloring indicates relative values with red denoting
640 higher and blue denoting lower. *Non-vaccinated only, dpi-matched (5-14 dpi) samples
641 averaged for non-carriers and carriers. Animals with undetermined carrier status (14-49, 14-50
642 and 14-51) were euthanized at 10 dpi or earlier.

643

644 Figure 2. Inoculum-rooted maximum likelihood phylogeny of sample consensus sequences
645 determined using PhyML [86]. Non-vaccinated cattle IDs are black text and vaccinated are light
646 blue. Haplotypic clades (A-F) are colored according to inferred ancestral relationships and
647 correlated polymorphism frequencies (see 'haplotypic composition of sample populations' in
648 methods).

649

650 Figure 3. Haplogroup proportions in virus samples from 9 serially-sampled cattle and colored as
651 in Figure 2. Proportions were determined based on SNP profiles characteristic of each lineage.
652 Gray indicates lineage unidentified. X axis is days post infection at which sample was taken;
653 'hpi' is hours post infection, between 4 and 21 hpi.

654

Figure 4. Variable FMDV capsid amino acids and positively-selected sites combined. [a] Presence of positively-selected sites identified in individual serially-sampled cattle using mixed effects model of evolution (MEME) analysis (P ≤ 0.10) (Murrell et al. 2012). [b] Amino acids identified as variable between viruses at the consensus level in VP1, VP2, and VP3 capsid coding regions.

Figure 5. Proportion of inherited polymorphism. Substitutions present at the consensus level across samples also present at the subconsensus level in the inoculum (≥ 0.5%) are defined as ancestral SNPs. The remaining nucleotide changes observed in sample consensuses (not detected in the inoculum) are defined as novel.

Figure 6. Homology model of inoculum consensus sequence with template PDB 4GH4 (FMDV A22). Surface view of FMDV A24 capsid protomer including VP1 (blue), VP2 (green), and VP3 (red). Labeled amino acid sites are those found to be under selective pressure according to MEME analysis as well as amino acids identified to be variable at the consensus level in samples derived from infected cattle. Inset: known antigenic sites in grey and heparan sulfate binding in yellow.

Supplemental Figures and Tables

Figure S1. Experimental design: sequenced sample sources and times.

Figure S2. Pairwise differences: Vaccinated vs. Naïve, Carriers vs. Terminators.

Table S1. Table of all variant nucleotides ≥ 2% in deep-sequenced samples.

Table S2. Table of all variant nucleotides ≥ 0.5% in inoculum.

Figure S3. Amino acid alignment of all sample consensus sequences.

678

679 1. Alexandersen, S., et al., *The pathogenesis and diagnosis of foot-and-mouth disease.* J
680 Comp Pathol, 2003. **129**(1): p. 1-36.
681 2. Arzt, J., et al., *The pathogenesis of foot-and-mouth disease II: viral pathways in swine,*
682 *small ruminants, and wildlife; myotropism, chronic syndromes, and molecular virus-host*
683 *interactions.* Transbound Emerg Dis, 2011. **58**(4): p. 305-26.
684 3. Knight-Jones, T.J. and J. Rushton, *The economic impacts of foot and mouth disease -*
685 *what are they, how big are they and where do they occur?* Prev Vet Med, 2013. **112**(3-
686 4): p. 161-73.
687 4. Brito, B.P., et al., *Review of the Global Distribution of Foot-and-Mouth Disease Virus*
688 *from 2007 to 2014.* Transbound Emerg Dis, 2015.
689 5. Grubman, M.J. and B. Baxt, *Foot-and-mouth disease.* Clin Microbiol Rev, 2004. **17**(2): p.
690 465-93.
691 6. Domingo, E., et al., *Evolution of foot-and-mouth disease virus.* Virus Res., 2003. **91**(1): p.
692 47-63.
693 7. Belsham, G.J., *Influence of the Leader protein coding region of foot-and-mouth disease*
694 *virus on virus replication.* J Gen Virol, 2013. **94**(Pt 7): p. 1486-95.
695 8. Mittal, M., et al., *Phylogeny, genome evolution, and antigenic variability among endemic*
696 *foot-and-mouth disease virus type A isolates from India.* Arch Virol, 2005. **150**(5): p. 911-
697 28.
698 9. Carrillo, C., et al., *Genetic and phenotypic variation of foot-and-mouth disease virus*
699 *during serial passages in a natural host.* J Virol, 2007. **81**(20): p. 11341-51.
700 10. Tully, D.C. and M.A. Fares, *The tale of a modern animal plague: tracing the evolutionary*
701 *history and determining the time-scale for foot and mouth disease virus.* Virology, 2008.
702 **382**(2): p. 250-6.
703 11. Strohmaier, K., R. Franze, and K.H. Adam, *Location and characterization of the antigenic*
704 *portion of the FMDV immunizing protein.* J Gen Virol, 1982. **59**(Pt 2): p. 295-306.
705 12. Logan, D., et al., *Structure of a major immunogenic site on foot-and-mouth disease virus.*
706 Nature, 1993. **362**(6420): p. 566-8.
707 13. Wang, G., et al., *How foot-and-mouth disease virus receptor mediates foot-and-mouth*
708 *disease virus infection.* Virol J, 2015. **12**: p. 9.
709 14. Acevedo, A., L. Brodsky, and R. Andino, *Mutational and fitness landscapes of an RNA*
710 *virus revealed through population sequencing.* Nature, 2014. **505**(7485): p. 686-90.
711 15. Borderia, A.V., et al., *Group Selection and Contribution of Minority Variants during Virus*
712 *Adaptation Determines Virus Fitness and Phenotype.* PLoS Pathog, 2015. **11**(5): p.
713 e1004838.
714 16. Wei, H., et al., *Deep-sequencing of Marburg virus genome during sequential mouse*
715 *passaging and cell-culture adaptation reveals extensive changes over time.* Sci Rep,
716 2017. **7**(1): p. 3390.
717 17. Arzt, J., et al., *The evolution of a super-swarm of foot-and-mouth disease virus in cattle.*
718 PLoS One, 2019. **14**(4): p. e0210847.
719 18. Domingo, E., et al., *Evolution of foot-and-mouth disease virus.* Virus Res, 2003. **91**(1): p.
720 47-63.

19. Zeng, J., et al., *Ribavirin-resistant variants of foot-and-mouth disease virus: the effect of restricted quasispecies diversity on viral virulence.* J Virol, 2014. **88**(8): p. 4008-20.
20. Rai, D.K., et al., *Attenuation of Foot-and-Mouth Disease Virus by Engineered Viral Polymerase Fidelity.* Journal of Virology, 2017. **91**(15).
21. Domingo, E., J. Sheldon, and C. Perales, *Viral quasispecies evolution.* Microbiol Mol Biol Rev, 2012. **76**(2): p. 159-216.
22. Arzt, J., J.M. Pacheco, and L.L. Rodriguez, *The early pathogenesis of foot-and-mouth disease in cattle after aerosol inoculation. Identification of the nasopharynx as the primary site of infection.* Vet Pathol, 2010. **47**(6): p. 1048-63.
23. Pacheco, J.M., et al., *Persistent Foot-and-Mouth Disease Virus Infection in the Nasopharynx of Cattle; Tissue-Specific Distribution and Local Cytokine Expression.* PLoS One, 2015. **10**(5): p. e0125698.
24. Stenfeldt, C., et al., *Pathogenesis of Primary Foot-and-Mouth Disease Virus Infection in the Nasopharynx of Vaccinated and Non-Vaccinated Cattle.* PLoS One, 2015. **10**(11): p. e0143666.
25. Stenfeldt, C., et al., *The Foot-and-Mouth Disease Carrier State Divergence in Cattle.* J Virol, 2016. **90**(14): p. 6344-64.
26. Arzt, J., et al., *The pathogenesis of foot-and-mouth disease I: viral pathways in cattle.* Transbound Emerg Dis, 2011. **58**(4): p. 291-304.
27. OIE. *Terresterrial Animal Health Code, Chapter 8.8 (Infection with foot and mouth disease virus)*. 2016 [cited 2016 July 14, 2016]; Available from: http://www.oie.int/fileadmin/Home/eng/Health_standards/tahc/current/chapitre_fmd.pdf.
28. Ferretti, L., et al., *Within-Host Recombination in the Foot-and-Mouth Disease Virus Genome.* Viruses, 2018. **10**(5).
29. Cortey, M., et al., *Persistent infection of African buffalo (Syncerus caffer) with Foot-and-Mouth Disease Virus: limited viral evolution and no evidence of antibody neutralization escape.* J Virol, 2019.
30. Mateu, M.G., et al., *Antigenic heterogeneity of a foot-and-mouth disease virus serotype in the field is mediated by very limited sequence variation at several antigenic sites.* J Virol, 1994. **68**(3): p. 1407-17.
31. Fry, E.E., et al., *The structure and function of a foot-and-mouth disease virus-oligosaccharide receptor complex.* Embo J, 1999. **18**(3): p. 543-54.
32. Perez Filgueira, M., et al., *Detection and characterization of functional T-cell epitopes on the structural proteins VP2, VP3, and VP4 of foot and mouth disease virus O1 campos [In Process Citation].* Virology, 2000. **271**(2): p. 234-9.
33. Alam, S.M., et al., *Antigenic heterogeneity of capsid protein VP1 in foot-and-mouth disease virus (FMDV) serotype Asia 1.* Adv Appl Bioinform Chem, 2013. **6**: p. 37-46.
34. Fry, E.E., et al., *Structure of Foot-and-mouth disease virus serotype A10 61 alone and complexed with oligosaccharide receptor: receptor conservation in the face of antigenic variation.* J Gen Virol, 2005. **86**(Pt 7): p. 1909-20.
35. Kitson, J.D., D. McCahon, and G.J. Belsham, *Sequence analysis of monoclonal antibody resistant mutants of type O foot and mouth disease virus: evidence for the involvement*

of the three surface exposed capsid proteins in four antigenic sites. Virology, 1990. **179**(1): p. 26-34.

36. Thomas, A.A., et al., *Antigenic sites on foot-and-mouth disease virus type A10.* J Virol, 1988. **62**(8): p. 2782-9.
37. Murrell, B., et al., *Detecting individual sites subject to episodic diversifying selection.* PLoS Genet, 2012. **8**(7): p. e1002764.
38. Jayasundara, D., et al., *ViQuaS: an improved reconstruction pipeline for viral quasispecies spectra generated by next-generation sequencing.* Bioinformatics, 2015. **31**(6): p. 886-96.
39. Kotecha, A., et al., *Rules of engagement between alphavbeta6 integrin and foot-and-mouth disease virus.* Nat Commun, 2017. **8**: p. 15408.
40. Arzt, J., et al., *Transmission of Foot-and-Mouth Disease from Persistently Infected Carrier Cattle to Naive Cattle via Transfer of Oropharyngeal Fluid.* mSphere, 2018. **3**(5).
41. Eschbaumer, M., et al., *Systemic immune response and virus persistence after foot-and-mouth disease virus infection of naive cattle and cattle vaccinated with a homologous adenovirus-vectored vaccine.* BMC Vet Res, 2016. **12**: p. 205.
42. Bull, R.A., et al., *Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection.* PLoS Pathog, 2011. **7**(9): p. e1002243.
43. Kijak, G.H., et al., *Rare HIV-1 transmitted/founder lineages identified by deep viral sequencing contribute to rapid shifts in dominant quasispecies during acute and early infection.* PLoS Pathog, 2017. **13**(7): p. e1006510.
44. Grenfell, B.T., et al., *Unifying the epidemiological and evolutionary dynamics of pathogens.* Science, 2004. **303**(5656): p. 327-32.
45. Morelli, M.J., et al., *Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts.* Vet Res, 2013. **44**: p. 12.
46. Rieder, E., et al., *Analysis of a Foot-and-Mouth Disease Virus Type A24 Isolate Containing an SGD Receptor Recognition Site In Vitro and Its Pathogenesis in Cattle.* J. Virol., 2005. **79**(20): p. 12989-12998.
47. Mateu, M.G., et al., *Implications of a quasispecies genome structure: effect of frequent, naturally occurring amino acid substitutions on the antigenicity of foot-and-mouth disease virus.* Proc Natl Acad Sci U S A, 1989. **86**(15): p. 5883-7.
48. Carrillo, C., et al., *Comparison of vaccine strains and the virus causing the 1986 foot-and-mouth disease outbreak in Spain: epizootiological analysis.* Virus Res, 1990. **15**(1): p. 45-55.
49. Martinez, M.A., et al., *Evolution subverting essentiality: dispensability of the cell attachment Arg-Gly-Asp motif in multiply passaged foot-and-mouth disease virus.* Proc Natl Acad Sci U S A, 1997. **94**(13): p. 6798-802.
50. Baranowski, E., et al., *Cell recognition by foot-and-mouth disease virus that lacks the RGD integrin-binding motif: flexibility in aphthovirus receptor usage.* J Virol, 2000. **74**(4): p. 1641-7.
51. Perales, C., et al., *Monitoring sequence space as a test for the target of selection in viruses.* J Mol Biol, 2005. **345**(3): p. 451-9.
52. Ruiz-Jarabo, C.M., et al., *Memory in viral quasispecies.* J Virol, 2000. **74**(8): p. 3543-7.

53. Gladue, D.P., et al., *Interaction of foot-and-mouth disease virus nonstructural protein 3A with host protein DCTN3 is important for viral virulence in cattle.* J Virol, 2014. **88**(5): p. 2737-47.
54. Gonzalez-Magaldi, M., et al., *Membrane topology and cellular dynamics of foot-and-mouth disease virus 3A protein.* PLoS One, 2014. **9**(9): p. e106685.
55. Stenfeldt, C., et al., *A partial deletion within foot-and-mouth disease virus non-structural protein 3A causes clinical attenuation in cattle but does not prevent subclinical infection.* Virology, 2018. **516**: p. 115-126.
56. Pacheco, J.M., et al., *A partial deletion in non-structural protein 3A can attenuate foot-and-mouth disease virus in cattle.* Virology, 2013. **446**(1-2): p. 260-7.
57. Nunez, J.I., et al., *A single amino acid substitution in nonstructural protein 3A can mediate adaptation of foot-and-mouth disease virus to the guinea pig.* J Virol, 2001. **75**(8): p. 3977-83.
58. Beard, C.W. and P.W. Mason, *Genetic determinants of altered virulence of Taiwanese foot-and-mouth disease virus.* J Virol, 2000. **74**(2): p. 987-91.
59. Pacheco, J.M., et al., *Role of Nonstructural Proteins 3A and 3B in Host Range and Pathogenicity of Foot-and-Mouth Disease Virus.* Journal of Virology, 2003. **77**(24): p. 13017-13027.
60. Carrillo, C., et al., *Comparative genomics of foot-and-mouth disease virus.* J Virol, 2005. **79**(10): p. 6487-504.
61. Lauring, A.S. and R. Andino, *Quasispecies theory and the behavior of RNA viruses.* PLoS Pathog, 2010. **6**(7): p. e1001005.
62. Pawlotsky, J.M., *Hepatitis C virus population dynamics during infection.* Quasispecies: Concept and Implications for Virology, 2006. **299**: p. 261-284.
63. Gebauer, F., et al., *Rapid selection of genetic and antigenic variants of foot-and-mouth disease virus during persistence in cattle.* J Virol, 1988. **62**(6): p. 2041-9.
64. Ciurea, A., et al., *Viral persistence in vivo through selection of neutralizing antibody-escape variants.* Proceedings of the National Academy of Sciences, 2000. **97**(6): p. 2749-2754.
65. Fischer, W., et al., *Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing.* PLoS One, 2010. **5**(8): p. e12303.
66. Bertram, M.R., et al., *Lack of Transmission of Foot-and-Mouth Disease Virus From Persistently Infected Cattle to Naive Cattle Under Field Conditions in Vietnam.* Front Vet Sci, 2018. **5**: p. 174.
67. Biswal, J.K., et al., *Genetic and antigenic variation of foot-and-mouth disease virus during persistent infection in naturally infected cattle and Asian buffalo in India.* PLoS One, 2019. **14**(6): p. e0214832.
68. Eschbaumer, M., et al., *Transcriptomic Analysis of Persistent Infection with Foot-and-Mouth Disease Virus in Cattle Suggests Impairment of Apoptosis and Cell-Mediated Immunity in the Nasopharynx.* PLoS One, 2016. **11**(9): p. e0162750.
69. Stenfeldt, C., et al., *Clearance of a persistent picornavirus infection is associated with enhanced pro-apoptotic and cellular immune responses.* Sci Rep, 2017. **7**(1): p. 17800.
70. Jackson, T., et al., *Efficient infection of cells in culture by type O foot-and-mouth disease virus requires binding to cell surface heparan sulfate.* J Virol, 1996. **70**(8): p. 5282-7.

71. Lawrence, P., et al., *Role of Jumonji C-domain containing protein 6 (JMJD6) in infectivity of foot-and-mouth disease virus.* Virology, 2016. **492**: p. 38-52.
72. Stenfeldt, C., et al., *Contact Challenge of Cattle with Foot-and-Mouth Disease Virus Validates the Role of the Nasopharyngeal Epithelium as the Site of Primary and Persistent Infection.* mSphere, 2018. **3**(6).
73. Juleff, N., et al., *Accumulation of nucleotide substitutions occurring during experimental transmission of foot-and-mouth disease virus.* J Gen Virol, 2013. **94**(Pt 1): p. 108-19.
74. King, D.J., et al., *Investigating intra-host and intra-herd sequence diversity of foot-and-mouth disease virus.* Infect Genet Evol, 2016. **44**: p. 286-92.
75. Maree, F., et al., *Differential Persistence of Foot-and-Mouth Disease Virus in African Buffalo Is Related to Virus Virulence.* J Virol, 2016. **90**(10): p. 5132-40.
76. Farooq, U., et al., *Characterization of naturally occurring, new and persistent subclinical foot-and-mouth disease virus infection in vaccinated Asian buffalo in Islamabad Capital Territory, Pakistan.* Transbound Emerg Dis, 2018. **65**(6): p. 1836-1850.
77. Lewis-Rogers, N., D.A. McClellan, and K.A. Crandall, *The evolution of foot-and-mouth disease virus: impacts of recombination and selection.* Infect Genet Evol, 2008. **8**(6): p. 786-98.
78. 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.* Scientific Reports, 2018. **8**(1).
79. Charleston, B., et al., *Relationship between clinical signs and transmission of an infectious disease and the implications for control.* Science, 2011. **332**(6030): p. 726-9.
80. Stenfeldt, C., et al., *Transmission of Foot-and-Mouth Disease Virus during the Incubation Period in Pigs.* Front Vet Sci, 2016. **3**: p. 105.
81. Tully, D.C. and M.A. Fares, *Shifts in the selection-drift balance drive the evolution and epidemiology of foot-and-mouth disease virus.* J Virol, 2009. **83**(2): p. 781-90.
82. Cottam, E.M., et al., *Analysis of Foot-and-mouth disease virus nucleotide sequence variation within naturally infected epithelium.* Virus Res, 2009. **140**(1-2): p. 199-204.
83. Sutmoller, P. and A. Gaggero, *Foot-and mouth diseases carriers.* Vet Rec, 1965. **77**(33): p. 968-9.
84. Paton, D.J., S. Gubbins, and D.P. King, *Understanding the transmission of foot-and-mouth disease virus at different scales.* Curr Opin Virol, 2018. **28**: p. 85-91.
85. LaRocco, M., et al., *A continuous bovine kidney cell line constitutively expressing bovine alphavbeta6 integrin has increased susceptibility to foot-and-mouth disease virus.* J Clin Microbiol, 2013. **51**(6): p. 1714-20.
86. Guindon, S., et al., *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.* Syst Biol, 2010. **59**(3): p. 307-21.
87. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.* Mol Biol Evol, 2016. **33**(7): p. 1870-4.
88. Kearse, M., et al., *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.* Bioinformatics, 2012. **28**(12): p. 1647-9.
89. Pond, S.L., S.D. Frost, and S.V. Muse, *HyPhy: hypothesis testing using phylogenies.* Bioinformatics, 2005. **21**(5): p. 676-9.

90. Warren, R.L., et al., *Assembling millions of short DNA sequences using SSAKE.* Bioinformatics, 2007. **23**(4): p. 500-1.
91. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

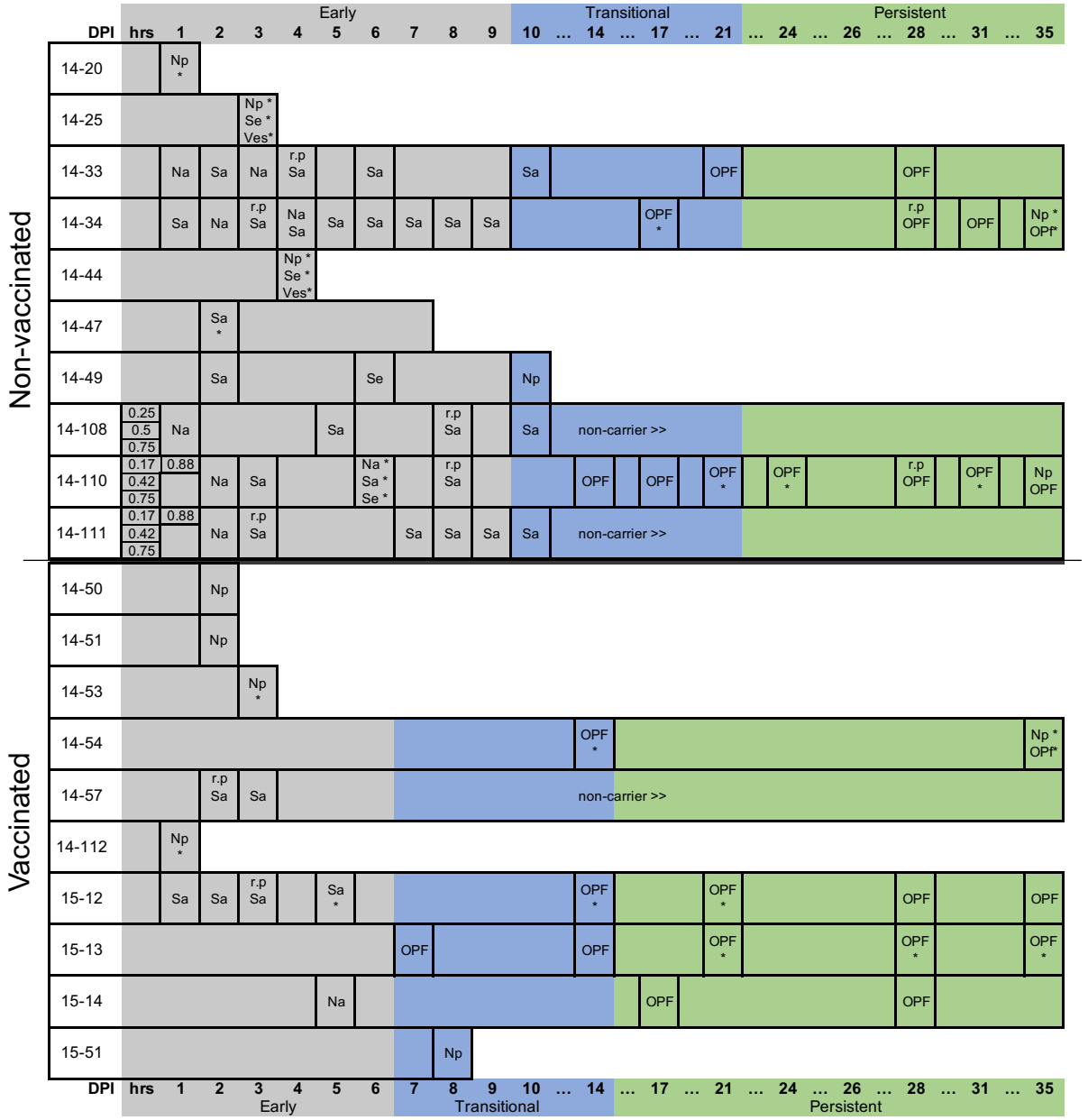


Figure S1. Experimental design: sequenced sample sources and times. Naïve cattle above center black line and vaccinated below. Abbreviations: hrs - fraction of first 24 hours (all nasal secretions), dpi – days post infection, Na – nasal secretion, Np – nasopharyngeal tissue (necropsy), OPF – oropharyngeal fluid, Sa – saliva, Se – serum, Ves – epithelial vesicle, r.p – both passaged and un-passaged samples were sequenced. * - consensus sequence available only (no deep sequencing data).

	Early Subs Rate - Naïve		
	All Subs	Nonsyn Subs	Syn Subs
subs/day	3.61[a]	1.52[b]	2.14
subs/site/year	0.188[a]	0.079[b]	0.111
	Early Subs Rate - Vaccinated		
subs/day	2.51	1.07	1.43
subs/site/year	0.131	0.056	0.075

	Transitional Subs Rate - Naïve		
	All Subs	Nonsyn Subs	Syn Subs
subs/day	2.44	0.70	1.74
subs/site/year	0.127	0.037	0.091
	Transitional Subs Rate - Vaccinated		
subs/day	1.70	0.73	0.97
subs/site/year	0.089	0.038	0.050

	Persistent Subs Rate - Naïve		
	All Subs	Nonsyn Subs	Syn Subs
subs/day	1.54[a]	0.28[b]	1.27
subs/site/year	0.080[a]	0.014[b]	0.066
	Persistent Subs Rate - Vaccinated		
subs/day	1.52	0.49	1.03
subs/site/year	0.079	0.026	0.054

	Trans-Persistent Subs Rate - Naïve		
	All Subs	Nonsyn Subs	Syn Subs
subs/day	1.90	0.46	1.44
subs/site/year	0.099	0.024	0.075
	Trans-Persistent Subs Rate - Vaccinated		
subs/day	1.53	0.54	0.99
subs/site/year	0.080	0.028	0.051

	1-10 dpi Subs Rates - Terminators		
	All Subs	Nonsyn Subs	Syn Subs
subs/day	4.29	1.92	2.50
subs/site/year	0.224	0.100	0.131
	1-10 dpi Subs Rates - Carriers		
subs/day	3.47	1.45	2.02
subs/site/year	0.181	0.076	0.105

Figure S2. Pairwise differences: Vaccinated vs. Naïve, Carriers vs. Terminators. The number of consensus-level pairwise differences between each sample and the preceding sample (inoculum = 0 dpi) divided by intervening time within each animal. Pairwise values from 0.25 - 0.88 dpi and animals of unknown carrier status omitted. For instances of samples with same dpi - same animal (different tissue) and where both raw and passaged were sequenced, pairwise differences were averaged. [a] = $P < 0.05$ and [b] = $P < 0.005$.

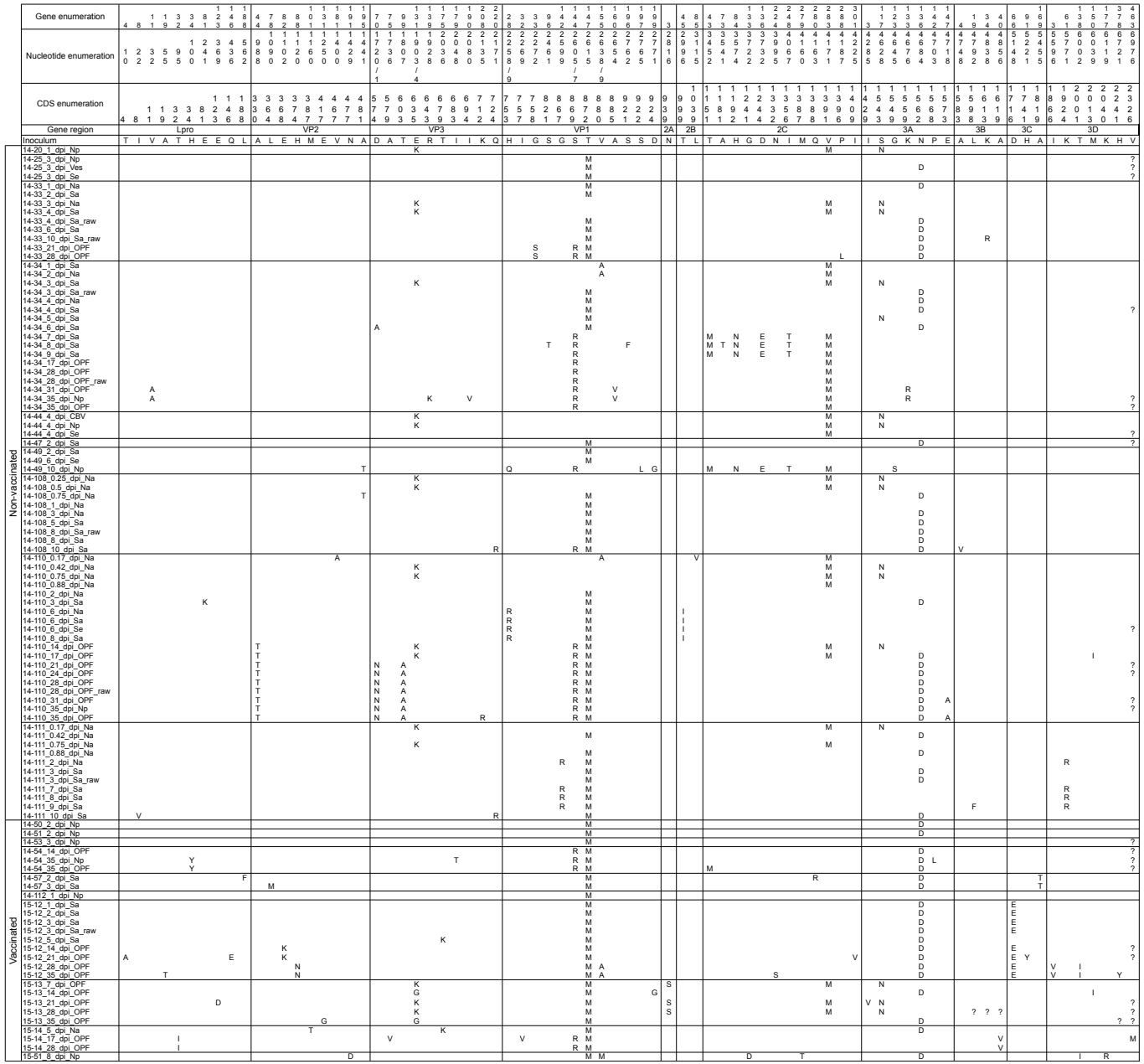


Figure S3. Amino acid alignment of all sample consensus sequences. Inoculum used as reference.