Guinea Fowl Coronavirus Diversity has Phenotypic Consequences for Glycan and Tissue Binding

Kim M. Bouwman[1], Mattias Delpont[2], Frederik Broszeit[3], Renaud Berger[2], Erik A.W.S. Weerts[1], Marie-Noëlle Lucas[2], Maxence Delverdier[2], Sakhia Belkasmi[2], Andreas Papanikolaou[1], Geert-Jan Boons[3], Jean-Luc Guérin[2], Robert P. de Vries[3], Mariette F. Ducatez[2]*, Monique H. Verheije[1]*

[1] Department of Pathobiology, Faculty of Veterinary Medicine, Utrecht University, 3584 CL Utrecht, The Netherlands

[2] IHAP, Université de Toulouse, INRA, ENVT, 23 Chemin des Capelles, 31076 Toulouse, France

[3] Department of Chemical Biology & Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CG Utrecht, The Netherlands

* corresponding authors M. Verheije: m.h.verheije@uu.nl

M. Ducatez: m.ducatez@envt.fr

ABSTRACT

Guinea fowl coronavirus (GfCoV) causes fulminating enteritis that can result in a daily death rate of 20% in guinea fowl flocks. Here we studied GfCoV diversity and evaluated its phenotypic consequences. Over the period 2014-2016, affected guinea fowl flocks were sampled in France and avian coronavirus presence was confirmed by PCR on intestinal content and immunohistochemistry of intestinal tissue. Sequencing revealed 89% amino acid identity between the viral attachment protein S1 of GfCoV/2014 and the previously identified GfCoV/2011. To study the receptor interactions as a determinant for tropism and pathogenicity, recombinant S1 proteins were produced and analyzed by glycan and tissue arrays. Glycan array analysis revealed that viral attachment S1 proteins from GfCoV/2014 and GfCoV/2011 can, in addition to the previously elucidated biantennary diLacNAc receptor, bind to glycans capped with alpha 2,6-linked sialic acids. Interestingly, recombinant GfCoV/2014-S1 has an increased affinity for these glycans compared to GfCoV/2011-S1, which was in agreement with the increased avidity of GfCoV/2014-S1 for gastrointestinal tract tissues. Enzymatic removal of receptors from tissues before applying spike proteins confirmed the specificity of S1 tissue binding. Overall, we demonstrate that diversity in GfCoV S1 proteins results in differences in glycan and tissue binding properties.

IMPORTANCE Avian coronaviruses cause major global problems in the poultry industry. As causative agents of huge economical losses, the detection and understanding of the molecular determinants of viral tropism is of ultimate

42 importance. Here we set out to study those parameters and obtained in-depth insight
43 in the virus-host interactions of guinea fowl coronavirus (GfCoV). Our data indicate
44 that diversity in GfCoV viral attachment proteins result in differences in affinity for
45 glycan receptors, as well as altered avidity for intestinal tract tissues, which might
46 have consequences for its tissue tropism and pathogenesis in guinea fowls.
47

48 **INTRODUCTION**

49 Avian coronaviruses (AvCoV) pose a major threat to poultry health, production and
50 welfare worldwide. AvCoVs are highly infectious, remain endemic in poultry
51 populations and, due to their high mutation rate, frequently produce new antigenic
52 variants (1, 2). The best-known AvCoV is infectious bronchitis virus (IBV), causing
53 mainly respiratory disease in chickens. In addition, IBV-like viruses have been
54 detected in other domestic poultry, including turkey and quail (3-5). In guinea fowl,
55 coronaviruses have been identified for the first time in 2011 as the causative agent
56 for fulminating enteritis(6). Full genome sequencing revealed that guinea fowl
57 coronavirus GfCoV/FR/2011 is closely associated with turkey coronavirus (TCoV) (7),
58 both causing gastrointestinal tract infections in their respective host (6, 8). Clinical
59 signs related to GfCoV infection in guinea fowl include prostration, ruffled feathers,
60 decreased water and feed consumption, and have resulted in a daily death rate up to
61 20% in several farms in France (6). Upon necropsy, whitish and enlarged pancreases
62 were consistently reported. Histopathological analyses revealed pancreatic necrosis
63 and lesions of various intensities in the intestinal epithelium, with most severe lesions
64 found in the duodenum of affected animals (6).
65

Genetic classification of AvCoVs is based on phylogenetic analysis of the S1 domain of its viral attachment protein spike (2). The spike protein is the main determinant for tropism (9), and the N-terminal part of the S1 of IBV has been shown to contain the receptor-binding domain (RBD) (10). Studies using recombinant IBV-S1 and/or -RBD proteins have demonstrated that the viral tropism is reflected by tissue binding of such proteins (11). Mutations in the spike proteins of IBV might either result in decreased (10) or increased (12) avidity for its receptor present on epithelial cells of the chicken trachea. In contrast to IBV, GfCoV and TCoV target the epithelial cells of the gastrointestinal tract (4, 6), and recombinant protein binding of their S1 proteins reflects this viral tropism, with predominant staining of epithelial cells of the small intestine (4). Glycan array analysis identified elongated LacNAcs on branched N-glycans as the host receptor for enteric AvCoVs, which are abundantly expressed on intestinal tissues (4).

Clinical symptoms in guinea fowl similar to those reported in 2011 are continuously reported by veterinarians in France (personal communication). However, studies on newly emerging GfCoVs are particularly hampered by the lack of models to grow the virus. More specifically, susceptible cell lines have not yet been identified, inoculation of embryonated guinea fowl eggs did not result in GfCoV production (data not shown), and SPF guinea fowls are not available for experimental infection.

Here we set out to study the consequences of GfCoV genetic diversity for glycan and tissue interactions. We revealed that the GfCoV spike gene from the 2014-2016 outbreak in guinea fowl flocks in France was 89% identical to that of GfCoV/2011 (7). Glycan and tissue binding analyses of GfCoV/2011 and GfCoV/2014 recombinant

spike S1 revealed that, while both proteins had the same specificity, GfCoV/2014-S1 had a much higher affinity toward glycan receptors and tissues of the lower gastrointestinal tract, in agreement with the observed replication of the virus in these tissues from field cases. Taken together we demonstrate GfCoV diversity results in phenotypically different receptor binding properties.

RESULTS

Lesions and coronaviral protein expression in the gastrointestinal tract of diseased guinea fowls between 2014-2016.

Fulminating disease (peracute enteritis) in guinea flocks continued to be reported after the initial outbreak of GfCoV infection in 2011 (6). Between February 2014 and November 2016, duodena from 29 diseased guinea fowls were collected and analyzed for lesions and coronaviral protein expression. Histological analysis of tissues by H&E staining revealed lesions in all duodena, with clear infiltration of inflammatory cells in remnants of the villi (Fig. 1, black arrowheads). For seven animals the entire gastrointestinal tract was available for histological analysis, showing lesions across the entire length of the intestinal tract, including the colon (Fig 1, black arrowheads). Viral protein expression using antibodies against the M protein of avian coronaviruses was observed in all duodena and four out of the seven lower intestinal tracts by immunohistochemistry (Fig. 1, white arrowheads). In the colons devoid of expression of viral proteins, the infiltration of inflammatory cells was noted, suggestive of a previous exposure to a virologic agent.

In contrast to what we observed, virus replication of GfCoV/2011 appeared to be restricted to the duodenum (6). Unfortunately, we were unable to confirm the lack of

116 infection of lower gastrointestinal tract samples in the previous outbreak due to
117 unavailability of samples. Nevertheless, we here hypothesize that genetically
118 divergent GfCoVs might have caused phenotypic differences in guinea fowls over the
119 years.
120
121 **Circulation of genetically diverse GfCoV.**
122 Gastrointestinal content collected from twenty affected animals between February
123 2014 and November 2016 were analyzed for the presence of gammacoronavirus
124 genetic material by one-step real-time RT-PCR using pan-gammacoronavirus
125 primers (13). For all samples, Ct values obtained were below 35 (data not shown),
126 confirming the presence of coronaviral RNA in all tested samples (Table 1). Next,
127 overlapping conventional PCRs were performed with primers based on the spike
128 gene of the GfCoV/2011 virus (sequences available upon request). Partial S1
129 sequences could be obtained from ten out of twenty RT-PCR positive samples (893-
130 1841nt/ 3624nt for complete S, Table1), the quality and/or quantity of the remaining
131 ten samples was too low to generate PCR products. Sanger sequencing of the
132 obtained fragments confirmed the presence of GfCoV in the intestinal content of all
133 ten birds, confirming continuous GfCoV circulation in France.
134
135 Phylogenetic analysis was performed to investigate the genetic diversity of the
136 obtained partial S1 sequences using Maximum likelihood analyses (Fig. 2). The
137 results showed that the 2014/2016 sequences clearly clustered with the S1 reference
138 gene from GfCoV/2011 (NCBI HF544506) supported by a bootstrap value of 100,
139 while they were genetically more distantly related to TCoV. Each of the GfCoV-

2014/2016 partial S1 sequences shared 84-90% nt identity with GfCoV/2011 and between the 2014-2016 partial S1 sequences the variation was 0.1 to 8.0%.

Only from one sample a full spike sequence could be obtained (γCoV/AvCoV/guinea fowl/France/14032/2014; NCBI MG765535), while for the others the amount and/or quality of the viral RNA samples were too low for further analyses. Comparison of the S1 gene of GfCoV/2014 with that of GfCoV/2011 using the Kimura 2-parameter distance model indicated that the genes had an 85% nucleotide and 89% amino acid sequence identity. Alignment of the amino acid sequences did not indicate clear mutation hotspots (data not shown) and the huge sequence diversity with IBV-M41-S1 (the only avian coronavirus for which a cryo-EM structure has been elucidated (14)) impairs further suggestions on the implications of each of the mutations.

GfCoV/2014-S1 recognizes the enteric coronavirus diLacNAc glycan receptor with higher affinity than GfCoV/2011-S1.

Using the glycan array of the Consortium for Functional Glycomics, we previously determined that S1 from GfCoV/2011 specifically binds to the diLacNAc glycan receptors (Gal_1,4GlcNAc_1,3Gal_1,4GlcNAc) (4). To study whether the observed changes in the spike of GfCoV/2014 resulted in differences in recognition of this glycan receptor, we recombinantly produced GfCoV/2014-S1 and GfCoV/2011-S1 and applied both proteins to diLacNAc-PAA conjugates in an ELISA as previously described (4). At similar protein concentrations GfCoV/2014-S1 showed improved binding to this receptor (Fig. 3), indicating that the mutations in S1 did not affect the specificity, but resulted in significant higher affinity, for this particular receptor.

The genetic differences between GfCoV/2014 and /2011 did not alter glycan specificity.

Next, we investigated whether the mutations in S1 resulted in recognition of additional N-linked glycans. To this end, both S1 proteins were applied to a novel glycan array containing N-glycan structures with their linear counterparts, either with terminal galactose or two differently linked sialic acid moieties (F. Broszeit and R.P. de Vries, submitted for publication). Schematic representations of each of the glycans are given in Fig. 4A. The data revealed that both GfCoV-S1 proteins bind to longer biantennary LacNAc structures (Fig. 4B, structures #**3-4**), including the diLacNAc structure used in the ELISA (Fig. 3). Furthermore, both GfCoV-S1 proteins bound to longer linear LacNAc repeats (Fig. 4B, structure #**1**), which were not included in the previous array (4). Finally, both GfCoV-S1 proteins bound longer linear and biantennary LacNAc repeats with terminal alpha 2,6 sialic acid (Fig. 4B, structures #**9-12**), but not those capped with alpha 2,3 linked sialic acids (Fig. 4B, structures #**5-8**). *Erythrina cristagalli* lectin (ECA), *Sambucus nigra* lectin (SNA) and *Maackia Amurensis* Lectin I (MAL1) were taken along as controls. We observed as expected specific binding to galactose, alpha 2,6 linked and alpha 2,3 linked sialic acid terminal glycans, respectively (Fig. 4C). In conclusion, both GfCoV-S1 proteins show specificity for the same glycans, ending with either galactose or alpha 2,6 linked sialic acids on the glycan array. However, the relative fluorescence observed for GfCoV/2014-S1 was consistently higher when compared to GfCoV/2011-S1, which is suggestive for differences in affinity for glycan receptors, as was observed for diLacNAcs in Fig. 3.

GfCoV/2014-S1 has a higher affinity for glycan receptors compared to /2011-S1. To allow comparison of the binding affinities of both proteins for each glycan, we applied fivefold serial S1 protein dilutions onto the glycan array and compared binding intensities at various scan powers. At each concentration, for all glycans shown in Figure 4A, binding signals of GfCoV/2014-S1 (Fig. 5A) were consistently higher than those of GfCoV/2011-S1 (Fig. 5B). Detection of linear glycan binding (glycan **#1** and **#9**) required higher concentrations and scan powers compared to the detection of biantennary LacNAc structures (glycans #**3-4** and #**11-12**) for both proteins. Interestingly, binding intensity of GfCoV/2011-S1 to glycans with terminal alpha 2,6 sialic acids was less compared to binding to glycans with terminal galactose (Fig. 5B compare structures #**3-4** to #**11-12** in 100 µg/mL to 20 µg/mL). This difference in preference for galactose-terminal glycans was not observed for GfCoV/2014-S1, since binding to glycan structures **#3-4** and **#11-12** was similar in each dilution applied to the array (Fig. 5A). Taken together, the data indicate that GfCoV/2014-S1 has a higher affinity for all glycans bound on the array compared to /2011-S1.

GfCoV/2014-S1 has a broader gastrointestinal tract tropism. To reveal whether the observed differences in glycan binding properties of the S1 proteins have biological consequences for tissue tropism, we first determined whether the identified glycans are indeed present on gastrointestinal tract tissues of healthy, uninfected guinea fowl. Both SNA and ECA lectins stained the epithelial lining of the duodenum, jejunum and caecum intensely, while intermediate staining of the proventriculus and colon was observed. In the pancreas only limited binding of SNA was observed, with no staining by ECA; in contrast, in the ileum ECA strongly bound whereas SNA

bound only to a limited extend. In conclusion, all tissues of the gastrointestinal tract, except cloaca, express GfCoV glycan receptors (Table 2) (15).

Next, we investigated the binding patterns of GfCoV-S1 proteins to gastrointestinal tissues. Both proteins stained the epithelial cells of almost the entire gastrointestinal tract (duodenum and colon in Fig. 6 1st column; summary of results in Table 2), indicating that receptors present on the tissues allow binding of S1. Interestingly, staining intensities of the lower intestinal tract (ileum, caecum, colon) were much more apparent for GfCoV/2014-S1 than for GfCoV/2011-S1. This prompted us to analyze avidity and specificity to glycan receptors in the guinea fowl gastrointestinal tissues by GfCoV-S1 proteins. We therefore pre-treated tissue slides with *Arthrobacter ureafaciens* neuraminidase (AUNA) and/or galactosidase to cleave off terminal sialic acids and galactose residues from host glycans, respectively. Treatment of the tissues with AUNA had only a minor effect on the binding of both GfCoV-S1, with a slight decrease in binding intensity to the duodenum for GfCoV/2014-S1 (Fig. 6A 2nd column; Table 2). SNA lectin binding was completely abolished after pre-treatment with AUNA, confirming that the treatment did effectively cleave off all sialic acids from the host glycans (Table 2).

When galactose residues were removed from the tissues by treatment with galactosidase prior to applying ECA, binding was severely reduced or totally absent (Table 2). Binding of GfCoV/2011-S1 to the tissue was completely abolished (Fig. 6 3rd column; Table 2), indicating that GfCoV tissue engagement is almost exclusively dependent on the presence of galactose-terminating glycans. On the other hand, GfCoV/2014-S1 still clearly bound to the epithelial cells of the intestinal tract,

indicating a significant difference in receptor binding avidity (Fig. 6 3rd column; Table 2).

Finally, tissues were simultaneously pre-treated with AUNA and galactosidase to remove both galactose and sialic acids from the glycans of the host. Indeed binding of both ECA and SNA were strongly reduced (Table 2). Tissue binding of GfCoV/2011-S1 was completely prevented, while GfCoV/2014-S1 still clearly bound to the epithelial cells of the gastrointestinal tract (except pancreas) (Fig. 6 4th column; Table 2). These results suggest that either a minor amount of receptors is still present, or that yet an additional (glycan) receptor is involved in tissue binding of GfCoV/2014-S1.

DISCUSSION

In this study we demonstrated ongoing GfCoV circulation in guinea fowl flocks in France. The sequence diversity between the viral attachment proteins of GfCoV circulating in 2011 and 2014 resulted in differences in receptor binding properties with profound phenotypic consequences. This relationship between these findings and in vivo pathogenesis can, however, only be elucidated in detail when new models to study this virus have been developed.

An amino acid sequence identity of 89% between viruses circulating only several years apart might indicate suggest that either a novel GfCoV strain was introduced in France from a yet unidentified source, or that there was high evolutionary pressure on the 2011 GfCoV strain. High mutation rates for avian coronaviruses are not uncommon (based on full genome sequences around 1.2×10^{-3} substitution/site/year

(16, 17)). When comparing GfCoV/2011 and /2014-S1 sequences, the calculated mutation rate was $5x10^{-2}$ substitution/site/year with a dN/dS ratio of 0.45. Similar mutation rates of the spike have been reported for IBV (18) and are believed to be driven by selective pressure after vaccination (19, 20). However, no vaccine is available against GfCoV, nor against the closely related turkey coronavirus, TCoV. Another driver for genetic diversity is the population size (21), however, this is unlikely to explain the observed fast mutation rate of GfCoV since flocks are considerably smaller compared to chicken flocks. It might well be that circulating antibodies against field strains of GfCoV are main drivers of the observed sequence diversity. Unfortunately, retrospective studies to further elucidate the contribution of virus evolution, the circulation of other virus populations in the last years, or introduction of novel strains via for example trade of birds between farms, are impossible due to the lack of archive material.

Here we revealed a novel glycan receptor for GfCoV, the first coronavirus that binds N-glycans capped with alpha 2,6 linked sialic acids. Alpha 2,6 sialic acid presence has been reported previously in guinea fowl large intestine (15), as well as the previously elucidated poly-LacNAc expressed in guinea fowl small intestine (4). Together, their expression patterns can explain in large part the tropism of GfCoV, but it does not exclude, together with the results presented in this manuscript, that yet another host factor plays a role in GfCoV/2014 infection. Initial attempts to show whether protein receptors, required for infection of many other coronaviruses (22-24), are required were yet unsuccessful (data not shown).

While spike protein binding analyses suggest phenotypic differences between these viruses *in vivo*, the reported gross clinical signs in field cases between 2011 and 2014 were not markedly different. Attempts to study the pathogenesis of GfCoV/2014 by inoculating commercial guinea fowls with GfCoV-containing fecal samples did, unfortunately and in contrast to a previous study (6), not result in manifestations of clinical signs or convincing detection of viral RNA by RT-QPCR (data not shown). Whether this was due to previous exposure of commercial birds to GfCoV and hence circulating antibodies preventing the infection remains to be investigated.

Here, we have demonstrated that GfCoV/2014-S1 has higher affinity for glycan receptors and increased avidity for the lower gastrointestinal tract compared to GfCoV/2011-S1. The viral genetic diversity between these spikes and the implications for receptor recognition further add to our understanding of this virus for which models are basically lacking.

MATERIAL AND METHODS

Collection of field samples. Samples were collected from guinea fowls showing enteritis and concomitant high mortality (>10%) in flocks in five regions in France (Bretagne, Pays de Loire, Nouvelle-Aquitaine, Occitanie, and Auvergne-Rhône-Alpes) from February 2014 through November 2016. Gastrointestinal content was collected and stored at -80^{0}C for viral RNA isolation. Tissues (duodenum, pancreas, airsac, lung, 'small intestine', large intestine, kidney, cloaca, trachea and bursa) were collected during necropsy, fixed for 24h in 4% buffered formaldehyde (m/v) and stored in 70% ethanol.

Immunohistochemistry. Paraffin-embedded tissues were sliced at 4µm and deparaffinized in xylene and rehydrated in an ethanol gradient from 100%-70%. Antigen retrieval was carried out in Tris-EDTA pH 9,0 (preheated) before applying 1% H_2O_2 in methanol. After washing twice in Normal antibody diluent (Immunologic) mAb mouse anti IBV M protein 25.1 (Prionics, Lelystad, The Netherlands), cross reacting with TCoV and GfCoV (5) was applied for 1 hour at room temperature (RT). Slides were washed in PBS-0,1%Tween and EnVision kit (cat. no. K4001; Dako) was used for anti-mouse secondary antibody staining according to the manufacturers protocol. Slides were washed three times in PBS and viral M-protein presence was visualized with AEC. The tissues were counterstained with hematoxylin and mounted with AquaMount (Merck).

Molecular characterization of GfCoV. The gastrointestinal content collected from affected guinea fowl was clarified by centrifugation (30 sec at 11.000xg), and RNA was extracted using a Qiagen Viral RNA extraction kit following the instructions of the manufacturer. A one-step real-time RT-PCR targeting the avian coronavirus N-gene was carried out to confirm the presence of coronavirus RNA as previously described (13). Subsequently, the isolated RNA was reverse transcribed using the Revertaid kit with random hexamers (Thermo Fisher, Waltham, MA), and overlapping conventional PCRs were performed to amplify the guinea fowl S-gene (primer sequences available upon request). Sanger sequencing of the resulting fragments was performed using PCR primers. Contigs were generated with BioEdit (version 7.0.8.0) (25) and submitted to NCBI. Muscle (26) was used for the alignment, and Mega (version 6.06) with bootstrap value of 1000 for the phylogeny (27). Selective pressure was

337 calculated as dN/dS, and the dN=dS hypothesis was tested using Pamilo-Bianchi-Li
338 method (28) with a $p<0.05$ considered statistically significant.
339
340 **Construction of the expression vector.** The codon-optimized sequence for
341 GfCoV/2014-S1 (γCoV/AvCoV/guinea fowl/France/14032/2014; NCBI MG765535),
342 containing an upstream *Nhe*I and downstream *Pac*I restriction site, was obtained
343 from GenScript and cloned into the pCD5 expression vector by restriction digestion
344 (as previously described (29)). The S1 sequence is in frame with a C-terminal GCN4
345 trimerization motif and Strep-Tag. The expression vector encoding GfCoV/2011-S1
346 was generated previously (29).
347
348 **Production of recombinant proteins.** Recombinant S1 proteins were expressed by
349 transfection of human embryonic kidney (HEK293T) cells with pCD5-expression
350 vectors using polyethylenimine (PEI) at a 1:12 (w/w) ratio. Cell culture supernatants
351 were harvested after six days. The recombinant proteins were purified using Strep-
352 Tactin sepharose beads as previously described (29).
353
354 **ELISA.** Gal_1,4GlcNAc_1,3Gal_1,4GlcNAc (Consortium for Functional Glycomics),
355 was coated in a 96-well maxisorp plate (NUNC, Sigma-Aldrich) at 0.5 µg/well
356 overnight at 4˚C, followed by blocking with 3% BSA (Sigma) in PBS-0.1% Tween. S1
357 proteins were pre-incubated with Strep-Tactin HRPO (1:200) for 30 minutes on ice.
358 For each protein, 2-fold dilutions were made in triplicate in PBS, and applied onto the
359 coated well, followed by incubation for 2 hours at room temperature. TMB (3,3',5,5'-
360 tetramethylbenzidine, Thermo Scientific) substrate was used to visualize binding,
361 after which the reaction was terminated using 1M H2SO4. Optical densities

(OD450nm) were measured in FLUOstar Omega (BMG Labtech), and MARS Data Analysis Software was used for data analysis. Statistical analysis was performed using a 2-way ANOVA.

Glycan array. Glycan structures were printed in six replicates on glass slides (NEXTERION® Slide H, Schott Inc.). Prelabeled S1-proteins with Alexa647-linked anti-Strep-tag mouse antibody and with Alexa647-linked anti-mouse IgG (4:2:1 molar ratio) were applied to the slides (concentrations in figure legends) and incubated for 90 minutes, after which the slides were washed with PBS and deionized water, dried and imaged immediately.

As controls different lectins were applied: *Erythrina cristagalli* agglutinin *(*ECA), which is specific for glycans with terminal galactose, N-acetylgalactosamine, or lactose and *Sambuca nigra* agglutinin (SNA) and *Maackia Amurensis* Lectin I (MAL1) which are specific for alpha 2,6 linked and alpha 2,3 linked sialic acids attached to terminal galactose respectively. Of the six replicates, the highest and lowest value were removed, and of the remaining four the total signal and SD values were calculated and plotted in bar graphs or heatmaps.

Spike histochemistry. Spike histochemistry was performed as previously described (29). S1 proteins pre-complexed with Streptactin-HRPO were applied onto 4 µm sections of formalin-fixed paraffin embedded healthy guinea fowl tissues and binding was visualized using 3-amino-9-ethyl-carbazole (AEC; Sigma-Aldrich). Proteins were applied onto slides at 5 µg/ml. Where indicated the tissues were treated per slide with 40U β-galactosidase (Gal; Megazyme, USA) or 2 mU Neuraminidase (Sialidase)

from *Arthrobacter ureafaciens* (AUNA, Sigma, Germany) in 10 mM potassium acetate, 2,5 mg/ml TritonX100, pH 4.2 at 40^0C O/N before protein application.

Lectin histochemistry. Lectin histochemistry was performed as previously described (4). Biotinylated-*Erythrina cristagalli* lectin or Biotinylated-*Sambucus nigra* lectin (both Vector Laboratories) were diluted in PBS to a final concentration of 2 µg/ml (ECA) or 6 µg/ml (SNA) and applied to healthy guinea fowl tissue sections for 30 min. After washing with PBS the signal was visualized by an Avidin-Biotin complex (ABC kit; Vector Laboratories) and counterstained with hematoxylin.

Data availability. Contigs are available in GenBank under accession numbers MG765535 to MG765542 and MK290733 to MK290734.

ACKNOWLEDGMENTS

R.P.dV is a recipient of a VENI grant from the Netherlands Organization for Scientific Research (NWO); M.H.V is a recipient of a MEERVOUD grant from the NWO; M.H.V. and M.F.D. are recipients of van Gogh collaboration grant from Nuffic.

REFERENCES

1. **Duraes-Carvalho R, Caserta LC, Barnabe AC, Martini MC, Simas PV, Santos MM, Salemi M, Arns CW.** 2015. Phylogenetic and phylogeographic mapping of the avian coronavirus spike protein-encoding gene in wild and synanthropic birds. Virus Res **201:**101-112.
2. **Valastro V, Holmes EC, Britton P, Fusaro A, Jackwood MW, Cattoli G, Monne I.** 2016. S1 gene-based phylogeny of infectious bronchitis virus: An attempt to harmonize virus classification. Infect Genet Evol **39:**349-364.
3. **Circella E, Camarda A, Martella V, Bruni G, Lavazza A, Buonavoglia C.** 2007. Coronavirus associated with an enteric syndrome on a quail farm. Avian Pathol **36:**251-258.

4. **Ambepitiya Wickramasinghe IN, de Vries RP, Weerts EA, van Beurden SJ, Peng W, McBride R, Ducatez M, Guy J, Brown P, Eterradossi N, Grone A, Paulson JC, Verheije MH.** 2015. Novel Receptor Specificity of Avian Gammacoronaviruses That Cause Enteritis. J Virol **89:**8783-8792.
5. **Brown PA, Courtillon C, Weerts E, Andraud M, Allee C, Vendembeuche A, Amelot M, Rose N, Verheije MH, Eterradossi N.** 2018. Transmission Kinetics and histopathology induced by European Turkey Coronavirus during experimental infection of specific pathogen free turkeys. Transbound Emerg Dis doi:10.1111/tbed.13006.
6. **Liais E, Croville G, Mariette J, Delverdier M, Lucas MN, Klopp C, Lluch J, Donnadieu C, Guy JS, Corrand L, Ducatez MF, Guerin JL.** 2014. Novel avian coronavirus and fulminating disease in guinea fowl, France. Emerg Infect Dis **20:**105-108.
7. **Le VB, Schneider JG, Boergeling Y, Berri F, Ducatez M, Guerin JL, Adrian I, Errazuriz-Cerda E, Frasquilho S, Antunes L, Lina B, Bordet JC, Jandrot-Perrus M, Ludwig S, Riteau B.** 2015. Platelet activation and aggregation promote lung inflammation and influenza virus pathogenesis. Am J Respir Crit Care Med **191:**804-819.
8. **Jindal N, Mor SK, Goyal SM.** 2014. Enteric viruses in turkey enteritis. Virusdisease **25:**173-185.
9. **Casais R, Dove B, Cavanagh D, Britton P.** 2003. Recombinant avian infectious bronchitis virus expressing a heterologous spike gene demonstrates that the spike protein is a determinant of cell tropism. J Virol **77:**9084-9089.
10. **Promkuntod N, van Eijndhoven RE, de Vrieze G, Grone A, Verheije MH.** 2014. Mapping of the receptor-binding domain and amino acids critical for attachment in the spike protein of avian coronavirus infectious bronchitis virus. Virology **448:**26-32.
11. **Promkuntod N, Wickramasinghe IN, de Vrieze G, Grone A, Verheije MH.** 2013. Contributions of the S2 spike ectodomain to attachment and host range of infectious bronchitis virus. Virus Res **177:**127-137.
12. **Leyson C, Franca M, Jackwood M, Jordan B.** 2016. Polymorphisms in the S1 spike glycoprotein of Arkansas-type infectious bronchitis virus (IBV) show differential binding to host tissues and altered antigenicity. Virology **498:**218-225.
13. **Maurel S, Toquin D, Briand FX, Queguiner M, Allee C, Bertin J, Ravillion L, Retaux C, Turblin V, Morvan H, Eterradossi N.** 2011. First full-length sequences of the S gene of European isolates reveal further diversity among turkey coronaviruses. Avian Pathol **40:**179-189.
14. **Shang J, Zheng Y, Yang Y, Liu C, Geng Q, Luo C, Zhang W, Li F.** 2018. Cryo-EM structure of infectious bronchitis coronavirus spike protein reveals structural and functional evolution of coronavirus spike proteins. PLoS Pathog **14:**e1007009.
15. **Kimble B, Nieto GR, Perez DR.** 2010. Characterization of influenza virus sialic acid receptors in minor poultry species. Virol J **7:**365.
16. **Hanada K, Suzuki Y, Gojobori T.** 2004. A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. Mol Biol Evol **21:**1074-1080.
17. **Mahy BWJ.** 2010. The Evolution and Emergence of RNA Viruses. Emerging Infectious Diseases **16:**899-899.

18. **Lee CW, Jackwood MW.** 2001. Origin and evolution of Georgia 98 (GA98), a new serotype of avian infectious bronchitis virus. Virus Res **80:**33-39.
19. **Kant A, Koch G, van Roozelaar DJ, Kusters JG, Poelwijk FA, van der Zeijst BA.** 1992. Location of antigenic sites defined by neutralizing monoclonal antibodies on the S1 avian infectious bronchitis virus glycopolypeptide. J Gen Virol **73 (Pt 3):**591-596.
20. **Zou N, Wang F, Duan Z, Xia J, Wen X, Yan Q, Liu P, Cao S, Huang Y.** 2015. Development and characterization of neutralizing monoclonal antibodies against the S1 subunit protein of QX-like avian infectious bronchitis virus strain Sczy3. Monoclon Antib Immunodiagn Immunother **34:**17-24.
21. **Toro H, van Santen VL, Jackwood MW.** 2012. Genetic diversity and selection regulates evolution of infectious bronchitis virus. Avian Dis **56:**449-455.
22. **Williams RK, Jiang GS, Holmes KV.** 1991. Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. Proc Natl Acad Sci U S A **88:**5533-5536.
23. **Delmas B, Gelfi J, L'Haridon R, Vogel LK, Sjostrom H, Noren O, Laude H.** 1992. Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. Nature **357:**417-420.
24. **Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC, Choe H, Farzan M.** 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature **426:**450-454.
25. **Alzohairy A.** 2011. BioEdit: An important software for molecular biology, vol 2.
26. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-1797.
27. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30:**2725-2729.
28. **Pamilo P, Bianchi NO.** 1993. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol Biol Evol **10:**271-281.
29. **Wickramasinghe IN, de Vries RP, Grone A, de Haan CA, Verheije MH.** 2011. Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. J Virol **85:**8903-8912.

FIGURE LEGENDS

Figure 1. (Immuno)histological analyses of guinea fowl intestinal tract. Representative images of duodenum and colon from a guinea fowl presented with peracute enteritis in 2014 after staining with H&E (left) or antibodies against the M protein of infectious bronchitis virus, known to cross react with GfCoV-M protein in immunohistochemistry (IHC, right). Black arrowheads indicate inflammatory cells and white arrowheads indicate viral protein expression.

506

507 **Figure 2. Molecular phylogenetic analysis by Maximum Likelihood method**
508 **comparing GfCoV (partial) spike sequences.** Phylogenetic tree based on the
509 Kimura 2-parameter model, in which bootstrap values are shown next to the
510 branches. The analysis involved 22 nucleotide sequences. All positions containing
511 gaps and missing data were eliminated. There were a total of 893 nucleotide
512 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. *
513 indicate partial S1 sequences of GfCoV.

514

515 **Figure 3. Binding of GfCoV-S1 to the enteric coronavirus glycan receptor**
516 **diLacNAc.** Concentration-dependent binding of GfCoV-S1 proteins to
517 Gal_1,4GlcNAc_1,3Gal_1,4GlcNAc in ELISA. As negative control, IBV-M41-NTD
518 was taken along (10); 1: significant difference between GfCoV-S1 and IBV-M41, 2:
519 significant difference between GfCoV/2014-S1 and GfCoV/2011-S1 ($p<0.001$).

520

521 **Figure 4. Glycan binding specificity of guinea fowl S1 proteins.** Schematic
522 representation of selected glycan structures present on the glycan array; numbers
523 correspond to those shown in the graphs (A). Number 1-4 represent glycans ending
524 with galactose, number 5-8 glycans capped with alpha 2,3 linked sialic acids, number
525 9-12, glycans capped with alpha 2,6 linked sialic acids. Glycan receptor specificity of
526 GfCoV-S1 proteins (B) and lectins ECA, MAL1 and SNA (C) in glycan array assay (F.
527 Broszeit and R.P. de Vries, submitted for publication); RFU: relative fluorescent units;
528 yellow circle: galactose, blue square: GlcNAc, green circle: mannose, pink diamond:
529 NeuAc.

530

Figure 5. Glycan binding affinity of guinea fowl S1 proteins. Glycan binding of GfCoV/2014-S1 (A) and GfCoV/2011-S1 (B) are shown as heatmaps with 5-fold dilutions (100 µg/mL to 4 µg/mL) of the proteins applied to glycan array slides that are scanned with different laser intensities. RFU: relative fluorescent units; glycan numbers correspond to schematic representations shown in Figure 4A.

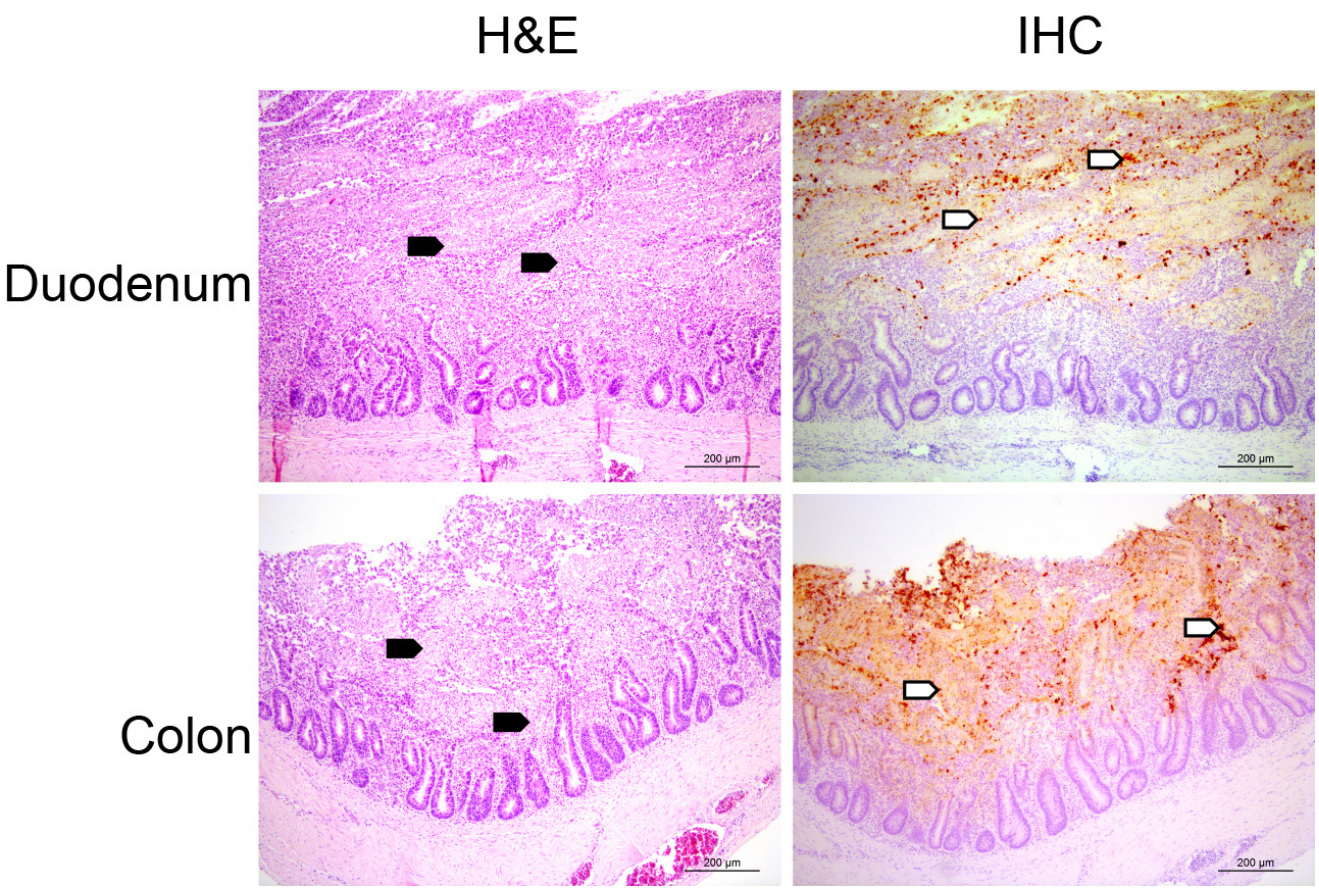
Figure 6. Binding of GfCoV-S1 proteins to guinea fowl duodenum and colon without and with enzymatic pretreatment of the tissues. Spike histochemistry was performed on uninfected, healthy duodenum (A) and colon (B) tissues without and with pre-treatment of enzymes (AUNA and/or galactosidase) before applying GfCoV/2014-S1 and GfCoV/2011-S1. Binding of proteins was visualized by red staining.

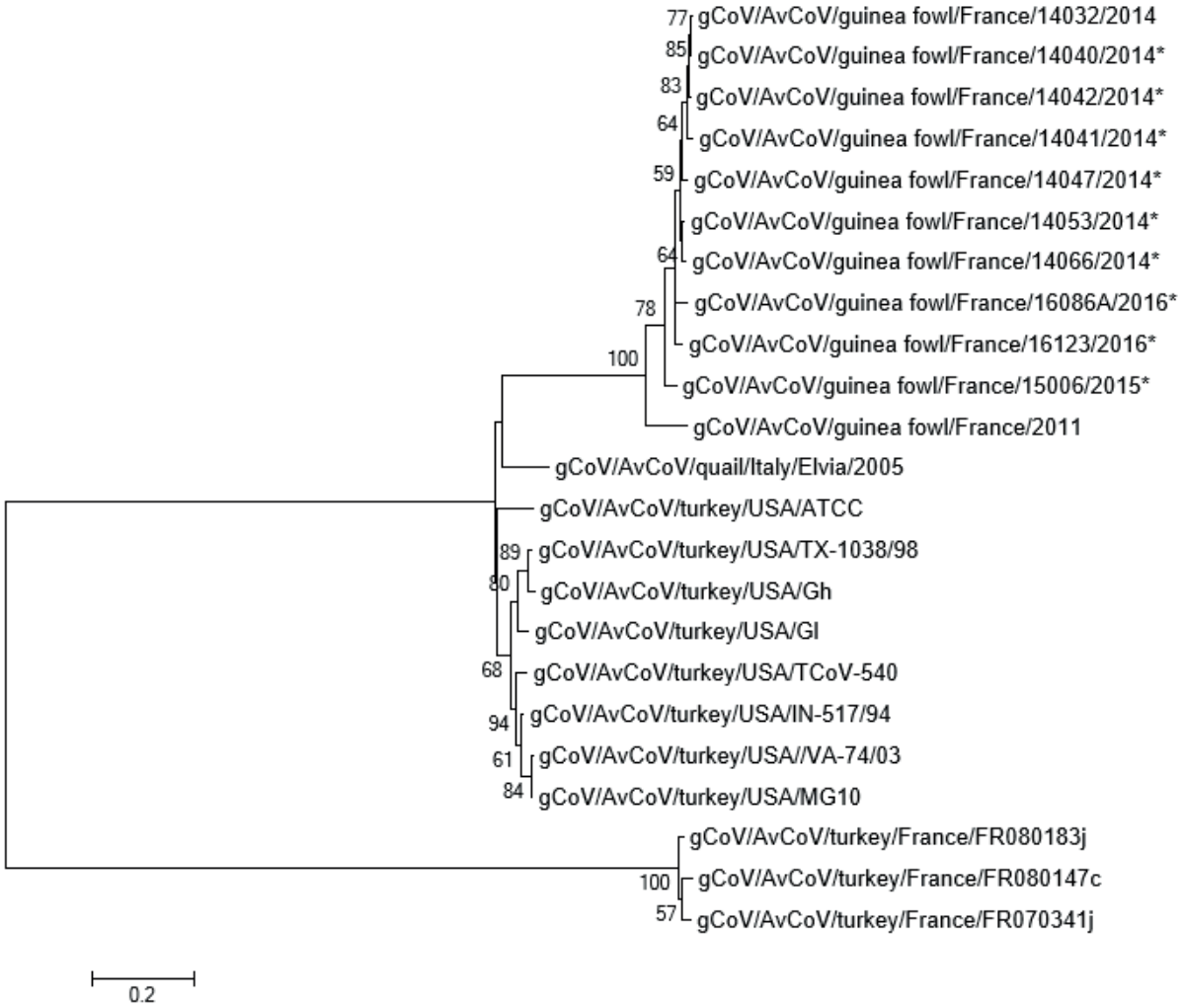
TABLE 1 Overview of selected guinea fowls and obtained GfCoV spike sequences.

Animals with bold animal numbers were included for immunohistological examination as well. *ND = unknown

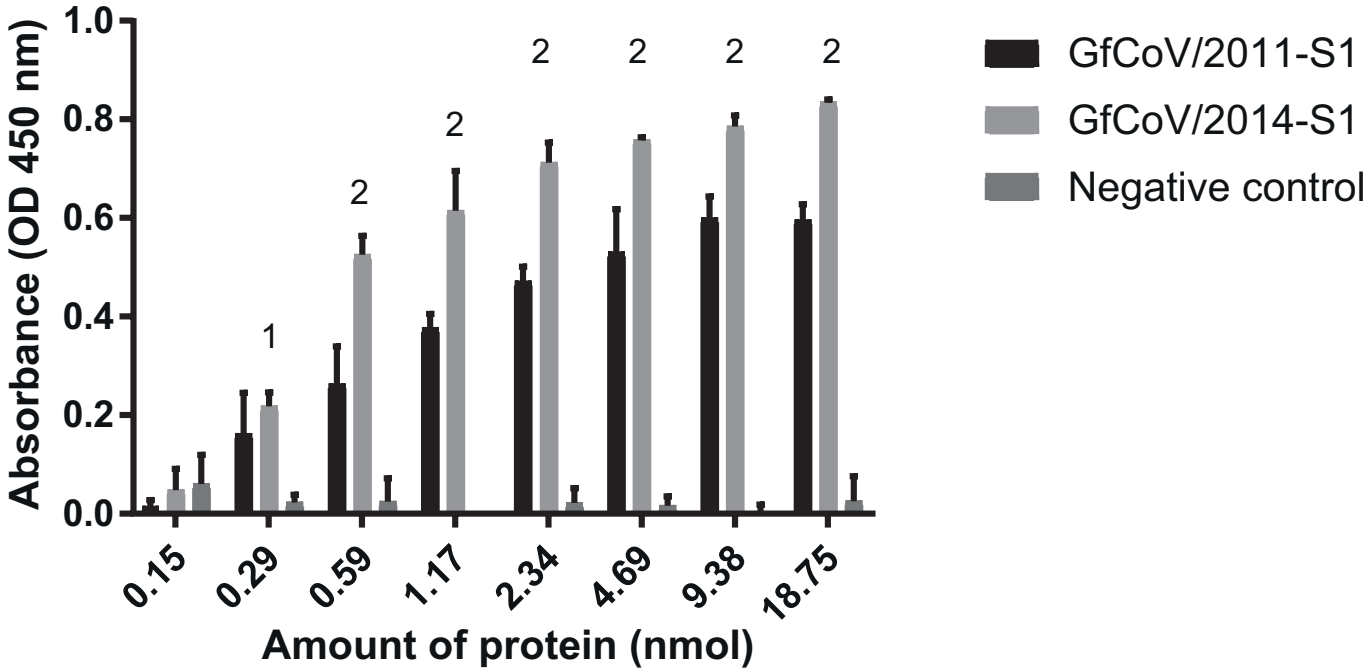
TABLE 2 Relative binding of viral proteins and lectins on guinea fowl intestinal tissues.

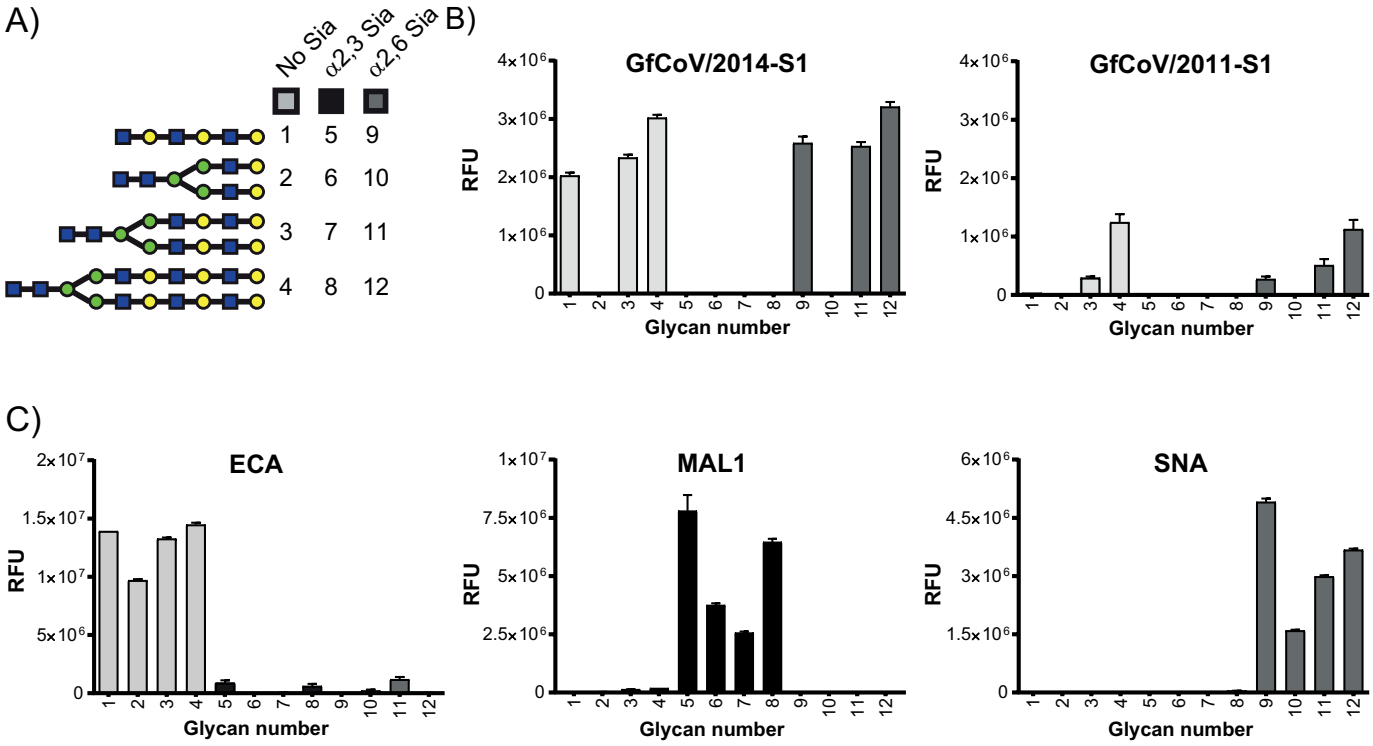
White box indicates no visible staining, light grey box indicates light to mild staining and/or not all epithelial cells show staining, dark grey indicates intense staining, most of the epithelial cells are showing positive signal. na = not analyzed

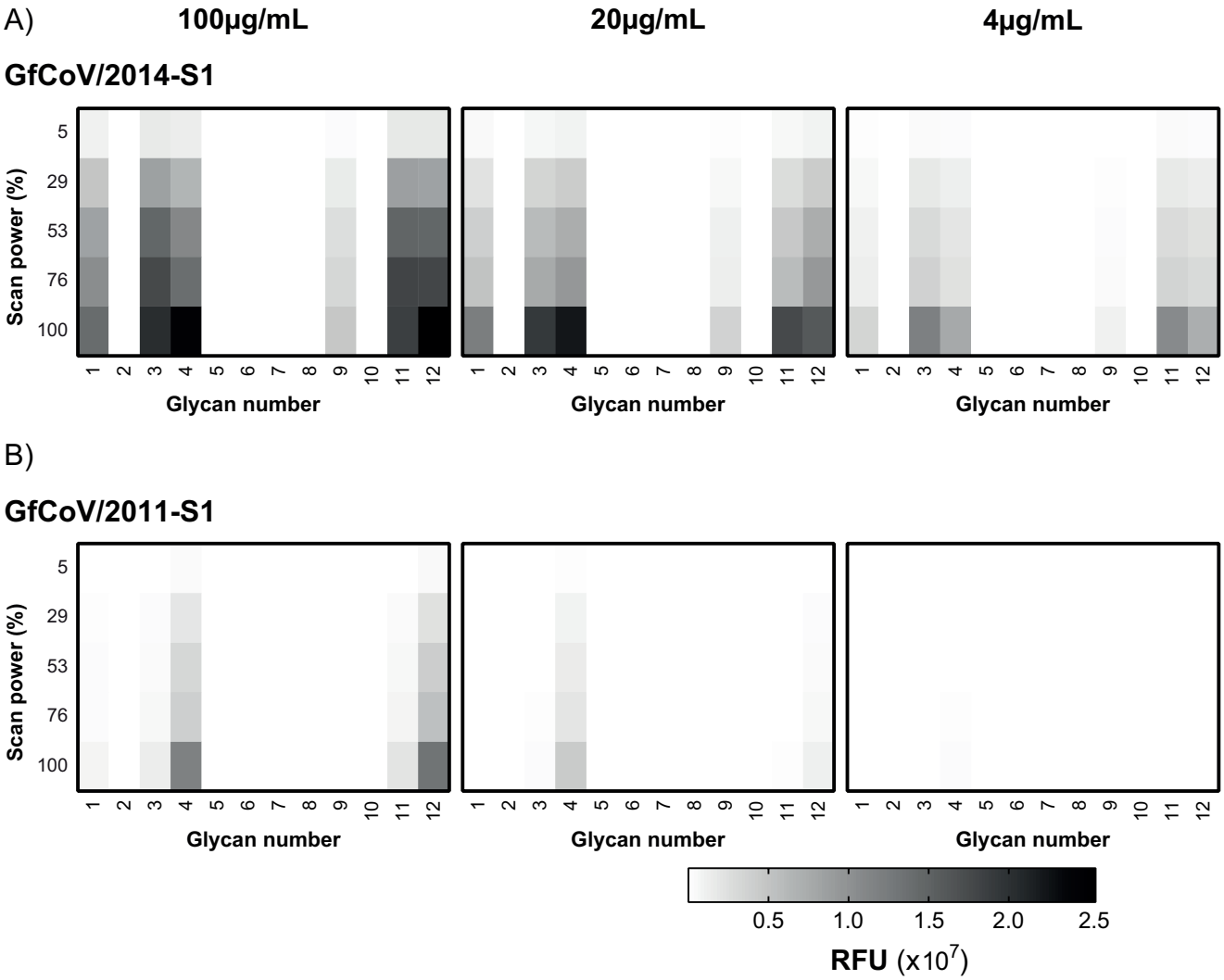




Journal of Virology







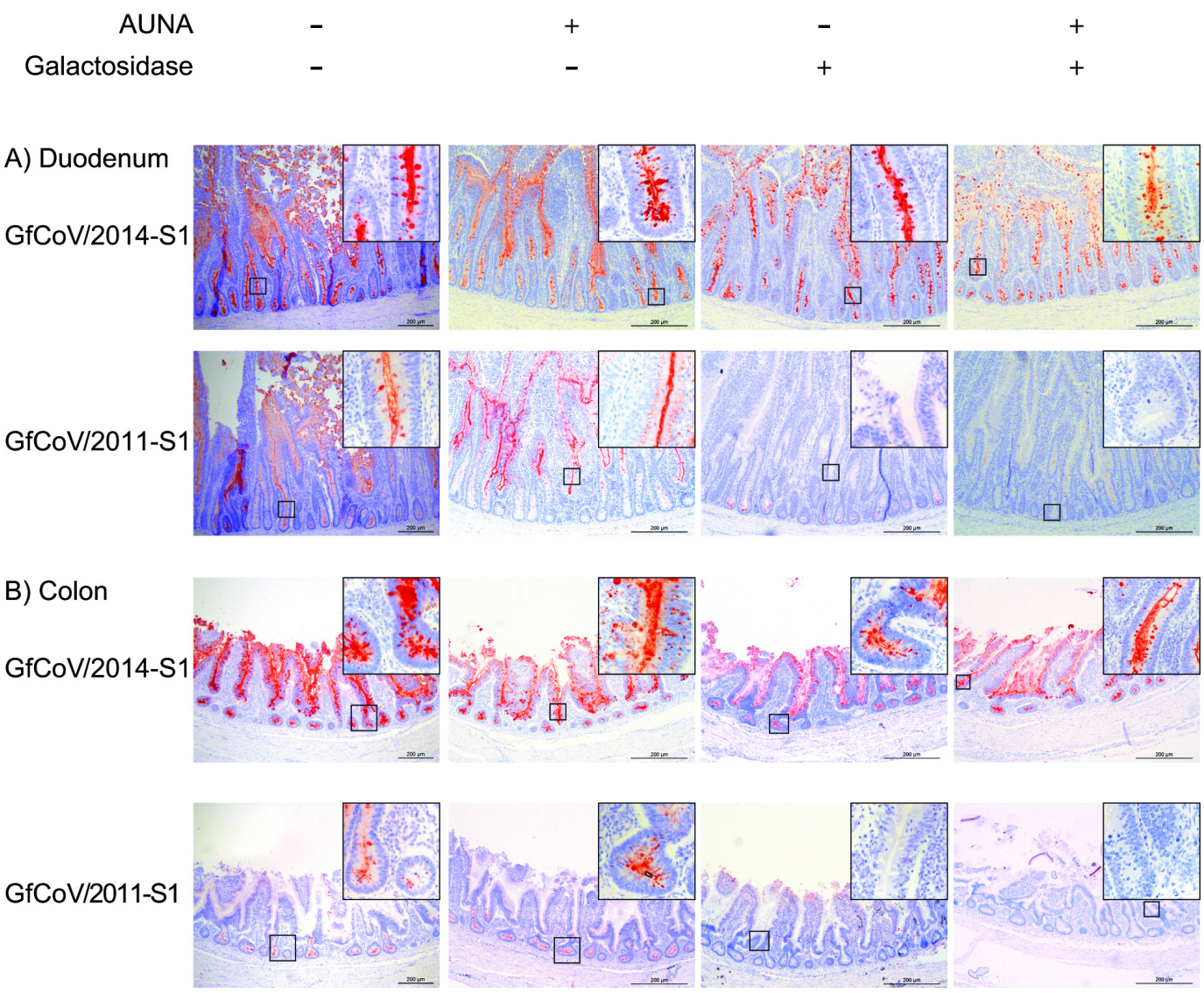


TABLE 1 Overview of selected guinea fowls and obtained GfCoV spike sequences. Animals with bold animal numbers were included for immunohistological examination as well

Animal number	Date of sample collection (week / year)	Age at sampling time (weeks)	Accession number: (Spike sequence)	Nt identity (%) with GfCoV/2011-S1
	2011		LN610099.1 nt: 1-3708	100
14-002	6 / 2014	10		
14-013	15 / 2014	8		
14-032	22 / 2014	7	MG765535 nt: 1-3669	85
14-036	24 / 2014	7		
14-037	25 / 2014	7		
14-039	26 / 2014	5.5		
14-040	23 / 2014	ND*	MG765536 nt: 1-1392	88
14-041	23 / 2014	ND*	MG765537 nt: 1-1771	88
14-042	23 / 2014	ND*	MG765538 nt: 1-1392	88
14-047	33 / 2014	3	MG765539 nt: 1-1378	88
14-053	37 / 2014	9	MG765540 nt: 1-1393	88
14-065	44 / 2014	12		
14-066	45 / 2014	4	MG765541 nt: 1-1384	88
15-006	3 / 2015	ND*	MG765542 nt: 1-980	87
15-116	46 / 2015	7		
15-118	47 / 2015	8		
16-086	38 / 2016	ND*	MK290733 nt: 1-2465	85
16-115	45 / 2016	4		
16-123	47 / 2016	ND*	MK290734 nt: 571-1895	86

*ND = unknown

TABLE 2 Relative binding of viral proteins and lectins on guinea fowl intestinal tissues.

		GfCoV/2014-S1				GfCoV/2011-S1				ECA			SNA		
Treatment	AUNA	-	+	-	+	-	+	-	+	-	-	+	-	+	+
	Galactosidase	-	-	+	+	-	-	+	+	-	+	+	-	-	+
Tissue	proventiculus														
	duodenum														
	pancreas														
	jejunum														
	ileum														
	ceacum								na						
	colon														
	cloaca														

White box indicates no visible staining, light grey box indicates light to mild staining and/or not all epithelial cells show staining, dark grey indicates intense staining, most of the epithelial cells are showing positive signal. na = not analyzed