

Histo-blood group antigens as divergent factors of groups A and C rotaviruses circulating in humans and different animal species

Dandan Zhao^{a*}, Yang Liu^{b*}, Pengwei Huang^c, Ming Xia^c, Weiwei Li^c, Ming Tan^{c,d}, XuFu Zhang^a and Xi Jiang^{c,d}

^aSchool of Traditional Chinese Medicine, Southern Medical University, Guangdong, People's Republic of China; ^bTianjin Key Laboratory of Molecular Nuclear Medicine, Institute of Radiation Medicine, Chinese Academy of Medical Sciences and Peking Union Medical College, Tianjin, People's Republic of China; ^cDivision of Infectious Diseases, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; ^dDepartment of pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

ABSTRACT

Histo-blood group antigens (HBGAs) have been found to be important host susceptibility factors or receptors for human rotavirus (RVs) with genotype-specific host ranges, impacting the disease patterns, epidemiology, and strategy development against RV diseases in humans. However, how the glycan factors contribute to RV diversity and host ranges to different animal species remains unclear. In this study using recombinant VP8* proteins as probes to perform glycan array analyses of RVs, we observed a wide range of glycan-binding profiles, including those binding to sialic acid-containing glycans, among group A (RVA) and group C (RVC) RVs that mainly infect different animal species. A tri-saccharide glycan Gal α 1-3Gal β 1-4Glc containing a terminal α -Gal was recognised by multiple RVA/RVC genotypes, providing valuable information on RV evolution under selection of the step-wisely synthesised HBGA in many animals before they were introduced to humans to be human pathogens. Saliva binding studies of VP8* also revealed strain-specific host ranges or species barriers between humans and these animal RV genotypes, further improved our understanding on RV host ranges, disease burdens, epidemiology, and vaccine strategy against RVs.

ARTICLE HISTORY Received 9 March 2020; Revised 8 June 2020; Accepted 9 June 2020

KEYWORDS Rotavirus; HBGA; evolution; receptor; diarrhea

Introduction

Rotaviruses (RVs) are an important cause of severe diarrhea in young children with high morbidity and mortality, claiming approximately 200,000 lives each year worldwide [1]. They are double-stranded RNA viruses, consisting 5 serological species (A–E) and two additional tentative species (F and G) [2]. RVs belonging to species A, B and C (RVA, RVB and RVC) are known to infect humans and various animals, whereas RVs of species D, E, F and G (RVD, RVE, RVF and RVG) have only been identified in animals, mostly birds [2]. RVs contain an 11-segmented dsRNA genome that encode 6 structural proteins and 6 nonstructural proteins. The structural proteins VP4 and VP7 constitute the outermost layer of the triple layer RV capsid. Based on the sequences of VP4 and VP7 proteins, RVs are divided into different P (VP4) and G (VP7) genotypes, respectively. The spike protein formed by VP4 can be cleaved into two fragments, including VP5* corresponding to the basis and VP8* that forms the distal head of the spike protein. The

VP8* interacts with the glycan receptors on the host cell surface to initiate RV infection, while VP5* facilitates penetration of RVs into host cells [3–5].

Early studies showed that some animal RVAs recognise the terminal sialic acids (SAs) as host receptors for attachment and these RV strains are referred as SA-dependent based on neuraminidase-sensitive assays on RV replication in cell cultures [6,7]. However, most human RVAs were found neuraminidase-insensitive and thus are referred as SA-independent [6]. Recent discoveries that major human RVs recognise HBGA as receptors or host susceptible factors for RV attachment to host cells support this notion [8–12].

HBGAs are complex carbohydrates consisting of three major families, the ABH(O), secretor, and Lewis families that are high diverse and distribute differently in the world population. HBGA are synthesised step-wisely by glycosyl transferases that are specific for individual ABH and Lewis families [13]. The biosynthetic processes are genetically controlled, resulting in polymorphic HBGA types, which is evolutionarily regulated in humans and some animal species.

CONTACT Yang Liu  Jason.Jiang@cchmc.org  Tianjin Key Laboratory of Molecular Nuclear Medicine, Institute of Radiation Medicine, Chinese Academy of Medical Sciences and Peking Union Medical College, Tianjin, People's Republic of China; XuFu Zhang  xfzhang2002@126.com  School of Traditional Chinese Medicine, Southern Medical University, 1838 North Guangzhou Avenue, Guangdong, People's Republic of China; Xi Jiang  liuyang@irm-cams.ac.cn  Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA Department of pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA
*these individuals contributed to the study equally.

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group, on behalf of Shanghai Shangyixun Cultural Communication Co., Ltd
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Significant advancements have been made in understanding evolution of major human RVs following step wise adaptations from an animal host origin to humans under selection of strain-specific HBGA glycans, reflecting the step-wisely synthesised HBGAs in humans [8,14–16]. This new knowledge provides valuable information on disease burdens and epidemiology, as well as vaccine development against RVs. However, how the diverse HBGAs are involved in host ranges and evolution of RVs among different animal species before these RVs become human pathogens infecting different human populations remains unknown.

In this study, we performed *in vitro* binding studies using recombinant VP8* proteins of the RVA P[I] genogroup that mainly infect animals to explore the roles of HBGAs in binding and host ranges to different animal and human hosts. A wide range of binding profiles to different HBGA glycans was found among strains representing different genotypes infecting humans and different animal species. Comparison of the binding profiles identified a potential evolutionary lineage from a potential ancestor with an animal host origin under selection by the step-wisely synthesised HBGAs in humans and many animal species. These findings significantly improved our understanding on RV host ranges and epidemiology. We also performed similar studies of RVCs and found the same principles of RV evolution and host ranges that may apply to RVCs and other RV groups/species.

Material and methods

Phylogenetic analysis of the VP8* sequence. The RV VP8* amino acid sequences from thirty-eight RVA and three RVC contained the intact protease-resistant VP8* core (amino acid (aa) 46–231) from the GenBank were used to perform alignment by the Laser gene software followed to build phylogenetic trees by the MEGA, version 4.1, software. A reference VP8* sequence from an RVB was included in the phylogenetic trees to estimate genetic distance among RV groups.

Expression and purification of VP8* proteins in *E. coli*. The VP8* sequences (amino acids 46–231) from different RVAs and RVCs (EDIM: AF039219, EMcN: AY267006, WC3: GU565077, P[5]:GQ496266, P[7]: QAY29504.1, P[9]: AB077766, P[10]:EU791922, P[14]: GQ398013, P[15]:JQ013506, P[19]: DQ887060, P[23]: QAY29526.1, P[25]: GU199495, P[28]: EU805773.1, P[30]:EU486960) were expressed in *E. coli* BL21 and purified by GST-tag affinity purification, as described previously [9].

Glycan array analyses of recombinant VP8* proteins. Initial ligand screenings for different RVA and RVC VP8* were performed by the Protein-Glycan Interaction Core of CFG (Consortium for Functional Glycomics) (The glycan library information is available

from the website <http://www.functionalglycomics.org/>). The recombinant GST-VP8* proteins were applied to individual glycan arrays at a protein concentration of 50 and 200 µg/ml and the bound GST-VP8* were detected using a fluorescent-labeled anti-GST monoclonal antibody. Relative fluorescent units (RFU) of each glycan were calculated to rank the reactivity in the interaction with individual RV VP8*s.

Binding of recombinant RV VP8* to human saliva. A set of previously characterised adult saliva samples with known ABO, secretor, and Lewis phenotypes were used in binding assays with the GST-VP8* fusion proteins [9].

Results

The VP8* sequences of RVAs are genetically segregated from RVCs. 48 VP8* reference sequences representing 35 P genotypes in five genogroups (P [I] – P[V]) of RVAs [8] plus three new genotypes (P[36], P[37] and P[38]) [17–20] were analysed. The newly identified genotypes P[36] and P[37] were sorted into P[I] genogroup and the P[38] genotype into the P[V] genogroup. Genogroup P[I] contains the most P genotypes (25 genotypes) including P[1]–P[3], and P[7] that are known as sialic acid-dependent RVs, infecting monkeys, dogs, bovine, porcine, horses and occasionally humans, as well as sialic acid-independent RVs that infect a wide ranges of animal species, including bovine, porcine, equine, monkey, lamps, murine, rabbits and humans. P[36] infects sugar glider [19], while P[37] infects pheasant [17,18].

Significant advancements on the P[II] RV evolution from an ancestor in P[I] genogroup following a step wise evolution with an animal host origin and adaptation to humans and become the major human pathogens have been made [8,14,15]. P[III] contains three genotypes (P[9], P[14] and P[25]) that recognise the type A antigens and infect humans, cats and bovine with the A antigen as a potential cross-species transmission factor between humans and these animal species [8]. P[IV] has one genotype P[11] that recognises the type 2 HBGAs and can switch binding to type 1 HBGAs and infect bovines and humans. The P [V] genogroup included four P genotypes (P[17], P [30], P[31] and P[35]) that mainly infect avian and bovine species plus the new P[38] genotype that infects Turkey [20], the host receptor ligand recognised by P [V] RVs remain unknown.

We also performed phylogenetic analyses of VP8* sequences for RVCs for a comparison with that of RVAs. A total of 32 VP8* sequences (aa 46–231) of RVCs from GenBank have been grouped into multiple porcine RV lineages, as well as several other lineages, each for bovine, dog and human genogroups (Figure 1), which is similar to the genetic grouping results

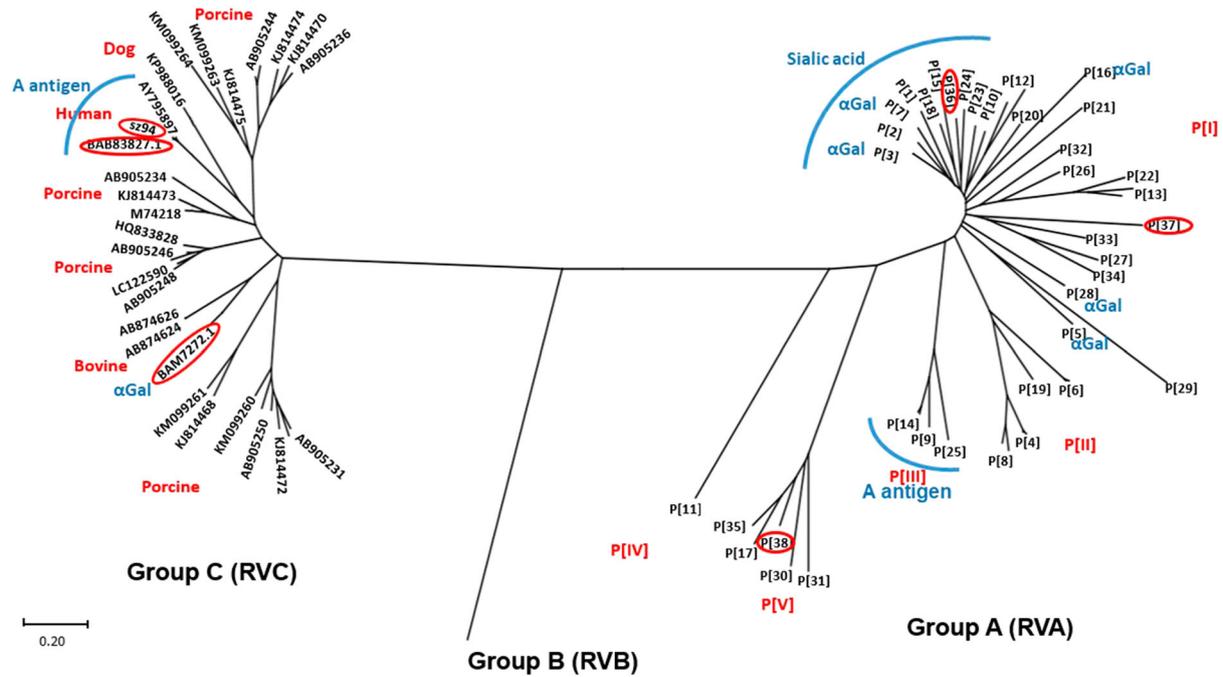


Figure 1. Phylogeny of group A rotaviruses. The group A RVs have been divided into 38 P genotypes (P[1] – P[38]) including 3 new genotypes (P[36]–P[38]) based on the major surface spike protein VP4/VP8* based on previous study [8]. A phylogeny dividing 38 group A RV genotypes into five P genogroups (P[I] – P[IV]) is shown. A total of 34 VP8* amino acid (aa) sequences (aa 46 to 231) of RVCs from GenBank have been grouped into multiple porcine and one for each of bovine and human genogroups, consistent with genetic grouping results reported by Sun et al. [21].

reported by Sun et al. [21]. The VP8* sequences of RVCs were segregated from those of RVAs in the phylogenetic tree (Figure 1), consistent with the VP6-based classification of RVA and RVC as distinct groups/species.

A trisaccharide glycan Gal α 1-3Gal β 1-4Glc containing a terminal α Gal is recognised by multiple RVA and RVC genotypes. To understand the diversity and evolution of major animal RVs, we performed glycan microarray analyses on selected RVs representing multiple RVA genotypes infecting different animal species and two RVC genotypes that infect human and bovine, respectively. Widely variable glycan binding profiles with top list glycans have been detected for individual genotypes. Interestingly, a trisaccharide glycan Gal α 1-3Gal β 1-4Glc (glycan number 117) containing an α Gal was found to be recognised by multiple strains representing one RVC and six RVA genotypes, indicating evolutionary connections among these genotypes with the α Gal glycan as a common receptor (Figure 1). Four strains exhibited strongest binding signals to the α Gal containing glycans, including two murine RVAs in the P[16] genotype (the prototype murine RV EDIM and a variant murine RV EMcN [22,23]), P[3] genotype in RVA that infects animals and humans, and a bovine strain in a RVC genotype. It is known that the α Gal transferase responsible for synthesis of α Gal glycan is produced in some animal species, but the gene encoding the specific α Gal transferase is inactive in humans [24]. Thus, these RVA and RVC genotypes may have a common ancestor with an

animal host origin, e.g. mouse and bovine for RVAs and RVCs, respectively.

Sialic acids are important host ligands for many animal RVAs. RVAs representing two P[I] genotypes (P[3] and P[7]) known as sialic acid (SA)-dependent and another two P[I] genotypes known as SA-independent (P[15] and P[23]) [7,25,26] have been found binding SA-glycans at high affinity (Figure 2). Interestingly, these four SA-binding genotypes are genetically closely related and form a cluster together with other four SA-dependent genotypes (P[1], P[2], P[16] and P[20]), excepting P[18], Figure 1 and Table 1) in the RVA P[I] genogroup. These findings indicated a new direction of RV evolution under selection of SA-containing glycans, consistent with the previous finding that many animal RVAs are SA-dependent [6,7].

Interestingly, a step-wise adaptation with a switch from an α Gal- to an SA-binding mode has been deduced among five SA-dependent genotypes in the order of P[3]→P[7] →(P[5]/P[15]/P[23]) (Figure 2) by analysing their glycan binding profiles. For example, the P[3] RVAs recognised α Gal as the top-list glycan at high affinity with a 3rd list SA-glycan at a low affinity (Figure 2), indicating that they are at an early stage of RV evolution that still relies on the α Gal glycan as their host receptor. In addition, P[3] RVAs also had a comparable high binding affinity to the 2nd list glycan [(3S)Gal β 1-3(Fuca1-4)GlcNAc β] with the backbone of the type 1 HBGA precursor, suggesting an intermediate stage of divergence with a possible flexible

Chart ID	RVA/P[3]	Average RFU	StDev	%CV
117	Gala1-3Galb1-4Glc-Sp10	527	23	4
487	(3S)Galb1-3(Fuca1-4)GlcNAc-Sp0	464	149	32
458	Neu5Aca2-3Galb1-4GlcNAcb1-6(Man1-6)(Neu5Aca2-3Galb1-4GlcNAcb1-2)(Man1-6)(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2)(Man1-3)(Manb1-4GlcNAcb1-4GlcNAcb-Sp21	184	59	32
Chart ID	RVA/P[7]	Average RFU	StDev	%CV
475	Neu5Aca2-3Galb1-4GlcNAcb1-2Man1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Man1-3)(Manb1-4GlcNAcb1-4)(Fuca1-6)GlcNAcb-Sp24	211	35	17
117	Gala1-3Galb1-4Glc-Sp10	140	27	19
320	Neu5Aca2-3Galb1-4GlcNAcb1-2Man1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Man1-3)(Manb1-4GlcNAcb1-4GlcNAcb-Sp12	39	25	63
Chart ID	RVA P[5]	Average RFU	StDev	%CV
594	GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAc-Sp14	447	65	14
117	Gala1-3Galb1-4Glc-Sp10	197	22	11
448	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glc-Sp0	86	3	3
23	6S(3S)Galb1-4GlcNAcb-Sp0	72	5	7
Chart ID	RVA P[15]	Average RFU	StDev	% CV
241	Neu5Aca2-3Galb1-4(Neu5Aca2-3Galb1-3)GlcNAcb-Sp8	1155	45	4
543	Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Man1-6(Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Man1-3)(Manb1-4GlcNAcb1-4GlcNAcb-Sp24	950	80	8
144	Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp0	896	69	8
Chart ID	RVA/P[23]	Average RFU	StDev	% CV
483	Neu5Aca2-3Galb1-4GlcNAcb1-2Man1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Man1-3)(Manb1-4GlcNAcb1-4)(Fuca1-6)GlcNAcb-Sp24	780	63	8
459	Neu5Aca2-3Galb1-4GlcNAcb1-2Man1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Man1-3)(Manb1-4GlcNAcb1-4GlcNAcb-Sp21	498	8	2
325	Neu5Aca2-3Galb1-4GlcNAcb1-2Man1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Man1-3)(Manb1-4GlcNAcb1-4GlcNAcb-Sp12	390	19	5
Chart ID	P[16]/EDIM	Average RFU	StDev	%CV
117	Gala1-3Galb1-4Glc-Sp10	502	36	7
23	6S(3S)Galb1-4GlcNAcb-Sp0	169	16	10
339	GlcNAcb1-4Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	145	39	27
528	Fuca1-4(Galb1-3)GlcNAcb1-2Mana-Sp0	93	17	18
Chart ID	P[16]/EMcN	Average RFU	StDev	%CV
117	Gala1-3Galb1-4Glc-Sp10	444	29	7
375	GalNAcb1-4GlcNAcb1-2Man1-6(GalNAcb1-4GlcNAcb1-2Man1-3)(Manb1-4GlcNAcb1-4GlcNAcb-Sp12	161	16	10
528	Fuca1-4(Galb1-3)GlcNAcb1-2Mana-Sp0	109	27	25
Chart ID	RVC/bovine	Average RFU	StDev	%CV
117	Gala1-3Galb1-4Glc-Sp10	829	57	7
186	GlcNAcb1-4-MDPLys	330	10	3
506	GalNAcb1-4(6S)GlcNAc-Sp8	246	57	23
87	GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0	159	40	25
Chart ID	RVA/P[28]	Average RFU	StDev	%CV
65	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glc-Sp8	3169	361	11
66	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glc-Sp10	3056	482	16
117	Gala1-3Galb1-4Glc-Sp10	451	77	17
375	GalNAcb1-4GlcNAcb1-2Man1-6(GalNAcb1-4GlcNAcb1-2Man1-3)(Manb1-4GlcNAcb1-4GlcNAcb-Sp12	295	64	22

Figure 2. Glycan array results of selected RV strains representing different RVA and RVC genotypes. The top list glycans from an array library contains 610 glycans recognised by each RV strain representing different RVA and RVC genotypes are shown. The sialic acid-containing glycans are highlighted red for individual strains. The αGal trisaccharide was highlighted in yellow. All nine strains gave perfect matched relative fluorescent unites (RFUs) of αGal vs. other top-list glycans between two concentrations (50 and 5 μg) of VP8* proteins tested for each strain except two strains (P[16]/EMcN and RVA/P[5]) that had one glycan position changes at the 3rd and/or 4th glycans in the top-lists, respectively, but the ranks of αGal did not change (data not shown).

binding interface to both αGal- and SA-epitopes, possibly responsible for their broad host ranges to many animal species (canine, feline, simian, bovine and caprine) and humans [8]. In the case of P[7], however, αGal was ranked 2nd with a comparable binding affinity to that of the No. 1 SA-containing glycans (Figure 2), therefore represents a later stage of divergence with a narrower host ranges, infecting porcine,

bovine and humans [8]. Finally, P[15], P[5], and P[23] seemed have further shifted to bind SA-containing glycans, explaining their narrow host ranges to lamb and bovine, bovine, and porcine only, respectively. The dual recognition of αGal and SA-glycans by the P[5] RVs has been demonstrated by Alfajaro et al., by studying the bovine RV G6P[5] WC3 and its mono-reassortant G4P[5] RotaTeq vaccine strains [27].

Table 1. Summary of in vitro binding results of selected group A rotaviruses.

Genotype	Genogroup	Host species	SA-depend	Glycans arrays	Mucin arrays	Saliva binding
P[1]	P[1]	Bovine/Huan/porcine	Yes	ND	ND	ND
P[2]	P[1]	Simian/Human	Yes	ND	ND	ND
P[3]	P[1]	Canine/Human/Feline/simian//bovine/caprine	Yes	ND	ND	+
P[5]	P[1]	Bovine	No	ND	+	+
P[7]	P[1]	Porcine/Bovine/Human	Yes	SA+	+	+
P[10]	P[1]	Human	No	SA-	+	+
P[13]	P[1]	Porcine	No	SA-	ND	ND
P[15]	P[1]	Lamb/Bovine	No	SA+	+	+
P[16]	P[1]	Murine (EDIM)	Yes	ND	ND	+
P[16]	P[1]	Murine (EMcN)	Yes	ND	ND	+
P[18]	P[1]	Equine	No	ND	ND	ND
P[20]	P[1]	Bovine	Yes	ND	ND	ND
P[23]	P[1]	Porcine	No	SA+	ND	ND
P[28]	P[1]	Human	Unknown	SA-	ND	+
P[36]	P[1]	Sugar Glider	Unknown	ND	ND	ND
P[37]	P[1]	Pheasant	Unknown	ND	ND	ND
P[6]	P[11]	bovine	unknown	SA-	ND	+
P[19]	P[11]	Porcine/Human	No	SA-	ND	+
P[9]	P[111]	Human/Feline/Canine	No	ND	+	+
P[14]	P[111]	Human/Bovine/Lupine	Yes	ND	+	+
P[25]	P[111]	Human	No	SA-	ND	+
P[30]	P[1V]	Avian	No	SA-	ND	ND
P[38]	P[1V]	Turkey	Unknown	ND	ND	ND

Animal RVAs also recognise a wide range of non-sialic acid glycans. P[28] is the only genotype characterised in this study in the RVA P[I] genogroup that infects humans exclusively. Glycan array analyses showed that the P[28] VP8* recognised the H type 1 HBGA precursors (Figure 2), a property very similar to that of the P[6] and P[19] human strains in RVA P[II] genogroup, explaining why P[28] also causes sporadic cases in humans similar to that of P[19] and P[6][14].

The A antigen may serve as a convergent factor in host ranges and evolution between RVAs and RVCs circulating in humans and animals sharing the common A antigen. Two RVCs were studied for glycan-binding properties by glycan array analysis. One strain (BAB83827.1) that caused several human epidemics in China [21] revealed significantly binding signals to the type A HBGAs (Figure 3). Saliva binding assays of VP8* of this RVC strain also showed a typical binding pattern to A and AB positive human saliva, consistent with the report that human RVCs recognise the A antigens [21]. In fact, the A antigen binding-human RVs also have been found in three P genotypes (P[9], P[14] and P[25]) in the P[III] genogroup of RVAs [8]. Interestingly, sequence alignment showed that the amino acid compositions of the A antigen-binding interfaces were highly conserved within but not between the two RV groups (Figure 3), suggesting an evolutionary relationship between these RVAs and RVCs by the A antigen as a convergent evolutionary factor.

Prediction of host ranges and species barriers between animals and humans by saliva binding assays. Unlike the A antigen-binding results of human saliva for the human RVAs and RVCs described above, a wide diverse binding profiles of human saliva samples have been observed among strains representing eight RVA and RVC genotypes mainly infecting animals (Figure 4). These results suggested additional unidentified polymorphisms of HBGAs or glycans that exist in humans and shared with many animal species. On the other hand, the low binding rates to only small subsets of the saliva donors observed for individual genotypes suggested strong species barriers between humans and these animal RV genotypes, consistent with the fact that these genotypes mainly infect different animal species with only a few genotypes occasionally infect humans (Table 1).

Discussion

In our previous studies, we have demonstrated an evolutionary pathway of the P[II] RVAs that mainly infect humans responsible for ~95% global RV epidemics. These P[II] RVAs were originated from an ancestor with a potential animal host origin in P[I] genogroup

that mainly infecting different animals [14,15]. Genetic analyses also revealed a co-evolutionary relationship of P[II] RVs with humans under selection of the step-wisely synthesised H type 1 HBGAs in humans that are developmentally regulated in newborn infants and evolutionarily conserved with some animals (Figure 5) [8,14–16]. Structural/functional analyses of VP8* of P[II] RVs also elucidated the molecular details of receptor binding property changes in each step of RV co-evolution with humans reflecting the step wise synthesis of HBGAs in humans vs. some animal species, explaining the genotype-specific host ranges, disease burden, epidemiology and vaccine development against RVs [15,16].

By performing genetic and *in vitro* binding experiments of the P[I] RVAs that mainly infect different animal species, we further extended our understanding on RV evolution under selection of the step-wisely synthesised HBGAs in different animal species before they entered humans and became the major P[II] human RVs. The α Gal transferase that catalyses the synthesis of the α Gal glycan is found in many animal species, but it is inactivated in humans and other primate species [24]. Analysing the glycan array binding profiles of individual genotypes also identified several P[I] clusters (sub-lineages) from a common α Gal-ancestor under selection of step wise synthesis HBGAs in their animal hosts before introduction to humans, further improved our understanding on RV diversity, host ranges and cross-species transmission which should facilitate future RV strain and epidemic prediction and vaccine development against RV diseases.

The A antigen-binding human RVCs described in this study represent a typical host range change from bovine to humans by switching binding properties from α Gal(Gal α 1-3Gal β 1-4GlcNAc) to the A antigens (GalNAc-Gal β 1-GlcNAc) by structural adaptation from the terminal α Gal to β GalNAc of the receptor binding interfaces. Both of the α Gal- and the β GalNAc-glycan transferases are produced in bovine [24,28], supporting such binding specificity switch and the bovine-to-human host range changes. Similar receptor binding changes and animal-to-human host range switch has also been described in our previous studies of the three P[III] genotypes (P[9], P[14] and P[25]) of RVAs [8]. Thus, the α Gal-to-GalNAc binding switch could be a common mechanism of RV evolution with host range changes under selection of stepwise synthesis of HBGAs in many animal species, explaining the P[III] RVAs commonly infect a subset of A antigen-positive humans and many antelope animals because the A enzyme is known to be produced in many antelope species [28] in addition to humans.

The terminal sialic acid residues could be another example of binding property change resulting in host ranges changes of the SA-dependent RVA genotypes

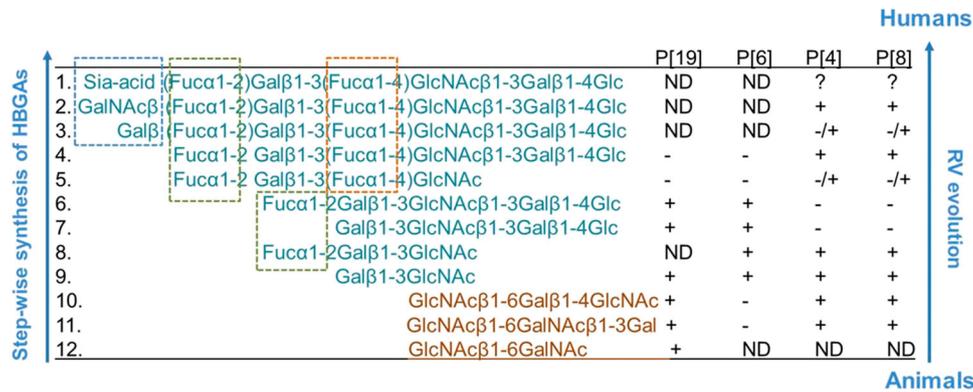


Figure 5. Summary of P[II] RV evolution under selection of type 1 HBGAs. The P[II] genogroup of RVs (P[4], P[6], P[8] and P[19]) that mainly infect humans has been deduced originating from an animal RV ancestor in P[I] genogroup that mainly infects different animals. P[19] represents an early stage as an intermediate jumping from animals to humans via recognising shared structures of the type 1 HBGA precursors commonly seen in humans and the mucin cores commonly seen in animals. P[6] is more advanced and recognises the type 1 precursors without the Lewis fucose (brown box) and has lost binding property to mucin cores. The P[4] and P[8] RVs are much further advanced and have broad host ranges to the secretor and Lewis positive individuals expressing the type 1 HBGAs with both Lewis and secretor fucoses, in which the Lewis fucose has been considered as a positive selection factors in P[4] and P[8] evolution. Three mucin cores are listed in brown colour while the rest type 1 HBGAs are in light blue colour. The brown dashed line box shows Lewis fucoses; the green boxes are secretor fucoses, and the blue box are sialic acid (top), the A antigen (middle), and the B antigen (bottom). The binding activities of individual P[II] RVs are listed as “-”, no binding; “+”, positive binding; “ND”, test not done; and “?”, unknown.

significant for disease control such as vaccine development against RVs. For example, the finding of the well adapted sialic acid binding properties of P[15] and P[5] RVAs and their low rates of human saliva binding suggests that neither P[15] lamb nor P[5] bovine RVs are suitable candidate for developing live attenuated human vaccines because they may not replicate efficiently in humans because of the species barriers.

The saliva binding assays of many animal RVs performed in this study also provide further information on the host ranges and species barrier between humans and many animal RV genotypes in the early stages of RV evolution. For example, the P[I] P[28] RVAs recognised only a small subset (5 of 81, 6%) of human saliva samples, explaining why P[28] RVs only cause sporadic cases in humans because they recognise the H type 1 HBGA precursors similar to that of the P[II] P[6] and P[19] RVAs only cause sporadic cases in young children [14,15]. Similarly, the other seven animal RVA and RVC genotypes are also at an early stage of RV evolution recognising less advanced HBGAs compared with the P[II] P[4] and P[8] RVs recognising much matured HBGAs containing the major ABH and Lewis determinants (Figure 5), explaining why these animal RV genotypes mainly infect different animal species and only occasionally infect humans.

In this study, we also included studies of RVCs to further extend our understanding of host HBGAs as an important receptor or host susceptible factor. We deduce that the same principle of co-evolution with different animal species and humans determined by their stepwise synthesis of HBGAs in parallel with RVAs. In comparison, however, the RVAs seem much more advanced than RVCs in term of recognising

more mature HBGAs in humans involving the major A, H and Lewis epitopes of the human HBGAs, while RVCs recognise only the A epitope, consistent with the fact that RVAs are major human pathogens.

Our study also has limitations. For example, the assessment of the roles of α Gal in RV binding and host ranges was based on comparisons of the ranks and binding signals of α Gal in relative to other top-list glycans of individual RVs and the binding signals varied significantly among RV strains tested, including strains had only marginal binding signals, which need to be verified in future studies. In addition, the glycan array library used in this study contained only 610 glycan ligands which may miss important host receptor ligands. Therefore future studies with alternative glycan array libraries such as the “Mucin Array” may be used [14,29]. Furthermore, the saliva binding results revealed a widely diverse binding profiles with low binding rates to human saliva samples by many RVAs and RVCs tested, including strains exclusively infecting animals, indicating unidentified polymorphism of HBGAs in humans and shared with some animals. Future studies to verify these binding signals responsible for RV infection and replication are necessary.

Acknowledgements

We would like to acknowledge the Protein-Glycan Interaction Core of the Consortium for Functional Glycomics for their analysis of samples by glycan microarray.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was partially supported by NIH grant R56AI114831 to X. J. This study was also supported by grants from the National Natural Science Foundation of China (81772243) and the Natural Science Foundation of Tianjin City (17JCYBJC42700) for Y. L. and National Natural Science Foundation of China (81773975) for X. F. Z.

References

- [1] Walker CLF, Rudan I, Liu L, et al. Global burden of childhood pneumonia and diarrhoea. *Lancet*. 2013;381(9875):1405–1416. PubMed PMID: Medline: 23582727.
- [2] Matthijnsens J, Ciarlet M, McDonald SM, et al. Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). *Arch virol*. 2011;156(8):1397–1413. PubMed PMID: Medline:21597953.
- [3] Nava P, Lopez S, Arias CF, et al. The rotavirus surface protein VP8 modulates the gate and fence function of tight junctions in epithelial cells. *J cell sci*. 2004;117(Pt 23):5509–5519. PubMed PMID: Medline:15494377.
- [4] Yu X, Guillon A, Szyzew AJ, et al. Crystallization and preliminary X-ray diffraction analysis of the carbohydrate-recognizing domain (VP8*) of bovine rotavirus strain NCDV. *Acta crystallogr B, Struct Biol Crystallogr*. 2008;64(Pt 6):509–511. PubMed PMID: Medline:18540062.
- [5] Yoder JD, Trask SD, Vo TP, et al. VP5* rearranges when rotavirus uncoats. *J virol*. 2009;83(21):11372–11377. PubMed PMID: Medline:19692464.
- [6] Ciarlet M, Estes MK. Human and most animal rotavirus strains do not require the presence of sialic acid on the cell surface for efficient infectivity. *J gen virol*. 1999;80(Pt 4):943–948. PubMed PMID: Medline: 10211964.
- [7] Dormitzer PR, Sun Z-YJ, Wagner G, et al. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *Embo J*. 2002;21(5):885–897. PubMed PMID: Medline: 11867517.
- [8] Liu Y, Huang P, Tan M, et al. Rotavirus VP8*: phylogeny, host range, and interaction with histo-blood group antigens. *J virol*. 2012;86(18):9899–9910. PubMed PMID: Medline:22761376.
- [9] Huang P, Xia M, Tan M, et al. Spike protein VP8* of human rotavirus recognizes histo-blood group antigens in a type-specific manner. *J virol*. 2012;86(9):4833–4843. PubMed PMID: Medline:22345472.
- [10] Hu L, Sankaran B, Laucirica DR, et al. Glycan recognition in globally dominant human rotaviruses. *Nat Commun*. 2018;9(1):2631, 1–12. PubMed PMID: Medline:29980685.
- [11] Ramani S, Cortes-Penfield NW, Hu L, et al. The VP8* domain of neonatal rotavirus strain G10P[11] binds to type II precursor glycans. *J virol*. 2013;87(13):7255–7264. PubMed PMID: Medline:23616650.
- [12] Hu L, Crawford SE, Czako R, et al. Cell attachment protein VP8* of a human rotavirus specifically interacts with A-type histo-blood group antigen. *Nature*. 2012;485(7397):256–259. PubMed PMID: Medline: 22504179.
- [13] Le Pendu J, Nystrom K, Ruvoen-Clouet N. Host-pathogen co-evolution and glycan interactions. *Curr Opin Virol*. 2014;7:88–94. PubMed PMID: Medline: 25000207.
- [14] Liu Y, Ramelot TA, Huang P, et al. Glycan specificity of P[19] rotavirus and comparison with those of related P genotypes. *J virol*. 2016;90(21):9983–9996. PubMed PMID: Medline:27558427.
- [15] Liu Y, Xu S, Woodruff AL, et al. Structural basis of glycan specificity of P[19] VP8*: Implications for rotavirus zoonosis and evolution. *PLoS Pathog*. 2017;13(11): e1006707, 1–19. PubMed PMID: Medline:29136651.
- [16] Xu S, Ahmed LU, Stuckert MR, et al. Molecular basis of P[11] major human rotavirus VP8* domain recognition of histo-blood group antigens. *PLoS Pathog*. 2020;16(3):e1008386, 1–31. Epub 2020/03/26. doi:10.1371/journal.ppat.1008386. PubMed PMID: 32208455; PubMed Central PMCID: PMC7122821.
- [17] Trojnar E, Sachsenroder J, Twardziok S, et al. Identification of an avian group A rotavirus containing a novel VP4 gene with a close relationship to those of mammalian rotaviruses. *J gen virol*. 2013;94(Pt 1):136–142. PubMed PMID: Medline:23052396.
- [18] Gal J, Marton S, Ihasz K, et al. Complete genome sequence of a genotype G23P[37] pheasant rotavirus strain identified in Hungary. *Genome Announc*. 2016;4(2):1–2. PubMed PMID: Medline:27034484.
- [19] Okadera K, Abe M, Ito N, et al. Isolation and characterization of a novel type of rotavirus species A in sugar gliders (*Petaurus breviceps*). *J gen virol*. 2016;97(5):1158–1167. PubMed PMID: Medline: 26902884.
- [20] Fujii Y, Mitake H, Yamada D, et al. Genome sequences of rotavirus A strains Ty-1 and Ty-3, isolated from Turkeys in Ireland in 1979. *Genome Announc*. 2016;4(1):1–2. PubMed PMID: Medline:26769943.
- [21] Sun X, Wang L, Qi J, et al. Human group C rotavirus VP8*s recognize type A Histo-blood group antigens as ligands. *J virol*. 2018;92(11):1–17. PubMed PMID: Medline:29593033.
- [22] McNeal MM, Belli J, Basu M, et al. Discovery of a new strain of murine rotavirus that is consistently shed in large quantities after oral inoculation of adult mice. *Virology*. 2004;320(1):1–11. PubMed PMID: Medline: 15003858.
- [23] Macartney KK, Baumgart DC, Carding SR, et al. Primary murine small intestinal epithelial cells, maintained in long-term culture, are susceptible to rotavirus infection. *J virol*. 2000;74(12):5597–5603. PubMed PMID: Medline:10823867.
- [24] Zakhour M, Ruvoen-Clouet N, Charpilienne A, et al. The alphaGal epitope of the histo-blood group antigen family is a ligand for bovine norovirus Newbury2 expected to prevent cross-species transmission. *PLoS Pathog*. 2009;5(7):e1000504, 1–14. doi:10.1371/journal.ppat.1000504. PubMed PMID: 19578439; PubMed Central PMCID: PMC2699481.
- [25] Haselhorst T, Fleming FE, Dyason JC, et al. Sialic acid dependence in rotavirus host cell invasion. *Nat Chem Biol*. 2009;5(2):91–93. PubMed PMID: Medline: 19109595.
- [26] Kraschnefski MJ, Bugarcic A, Fleming FE, et al. Effects on sialic acid recognition of amino acid mutations in the carbohydrate-binding cleft of the rotavirus spike protein. *Glycobiology*. 2009;19(3):194–200. PubMed PMID: Medline:18974199.
- [27] Alfajaro MM, Kim JY, Barbe L, et al. Dual recognition of sialic acid and alphaGal epitopes by the VP8* Domains of the bovine rotavirus G6P[5] WC3 and of

- Its mono-reassortant G4P[5] RotaTeg vaccine strains. *J virol.* **2019**;93(18):1–25. Epub 2019/06/28. doi:10.1128/JVI.00941-19. PubMed PMID: 31243129; PubMed Central PMCID: PMC6714814.
- [28] Matthijnsens J, Potgieter CA, Ciarlet M, et al. Are human P[14] rotavirus strains the result of interspecies transmissions from sheep or other ungulates that belong to the mammalian order Artiodactyla? *J virol.* **2009**;83(7):2917–2929. PubMed PMID: 19153225. Medline:19153225.
- [29] Li Z, Gao C, Zhang Y, et al. O-Glycome Beam Search arrays for Carbohydrate ligand Discovery. *Mol cell proteomics.* **2018**;17(1):121–133. PubMed PMID: 29183914. Medline:29183914.