CHARACTERIZING THE BIOCHEMICAL DETERMINANTS GOVERNING MERS-CORONAVIRUS HOST RANGE

Kayla M. Peck

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology.

Chapel Hill
2016

Approved by:

Christina Burch

Mark Heise

Ralph Baric

Todd Vision

Ronald Swanstrom

Joel Kingsolver

ABSTRACT

Kayla M. Peck: Characterizing the Biochemical Determinants Governing
MERS-Coronavirus Host Range
(Under the direction of Christina Burch and Mark Heise)

Coronaviruses are a diverse family of viruses that infect a wide range of hosts, including both mammalian and avian species. Within recent history, coronaviruses have expanded their host range into humans, with four emergence events resulting in infections that cause only mild disease. However, two additional emergence events resulted in outbreaks of severe disease, causing heightened concern for public health. The 2003 severe acute respiratory syndrome coronavirus (SARS-CoV) emerged in Southeast Asia and rapidly spread around the world with a 9 percent mortality rate before being controlled by public health intervention strategies. In 2012, Middle East respiratory syndrome coronavirus (MERS-CoV) emerged from its zoonotic reservoir. To date, it has infected over 1800 people with a 36 percent mortality rate and is still circulating in the population. Due to the emergence of coronaviruses with pandemic potential, it is important to understand how these lineages have been able to expand their host range to infect new species. One key determinant of viral host range is the interaction between the virus spike protein and the host cell receptor. For MERS-CoV specifically, the virus can infect bats, camels (the putative intermediate host species), and humans, but is unable to infect mice or other traditional small animal models due to receptor incompatibilities. The inability of MERS-CoV to infect any small animal model species leaves us unable to study pathogenesis or begin to develop potential vaccines or therapeutics. Here, I present work on the biochemical determinants that govern MERS-CoV host range. Specifically, I 1) characterize the interactions between the

MERS-CoV receptor binding domain and the mouse cell receptor; 2) investigate biochemical determinants that govern infection for other species; 3) attempt to generate a mouse-adapted MERS-CoV; and 4) present an approach to investigate potential evolutionary mechanisms of coronavirus host range expansion. This work has contributed to the development of a small animal model, allowing us to begin pathogenesis studies. Additionally, understanding the biochemical determinants and evolutionary mechanisms of coronavirus host range expansion can help evaluate the pandemic potential of currently circulating zoonotic strains and better prepare us for future pathogenic coronaviruses that may emerge.

ACKNOWLEDGMENTS

I have been extremely fortunate to be a part of three amazing labs. Christina Burch has been the best mentor a student could hope for, providing me with an environment of kindness, independence, and overwhelming faith in me. Mark Heise graciously accepted my eclectic skillset into his group and gave me the best confidence-boosting pep talks. Ralph Baric forged my dissertation project and gave me access to his space and resources that advanced my academic career. While most people had no idea where I belonged or who I was actually affiliated with, thank you to all of the labs who accepted me into their group, the hybrid that I was. This also includes Mark Tanaka who generously hosted me for two summers and provided me with invaluable computational biology experience.

Thanks to Adam Cockrell for training me in molecular virology, to Boyd Yount who taught me to be overly liberal with ethanol in the BSL3, and to Trevor Scobey who was always willing to help me with experiments, training, and finding things in the ever-rearranging Baric lab. Thanks to Jenny McGraw for being my desk partner and soup day companion, as well as my inspiration for working harder.

I could not have gotten through graduate school without the support of my amazing friends. Thank you to everyone who suffered through my Eevee, potato, raccoon, Louis Pasteur, zombie apocalypse, and Shrek obsessions. All the love in the world to my tortuga, Hailey.

Finally, a special dedication to Jack Weiss. His influence on generations of statisticians and R programmers will carry on, as well as his love of the R color “dodgerblue.” His passion for statistics and teaching will inspire me always.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACE2:	Angiotensin-converting enzyme 2
APN:	Aminopeptidase N
bDPP4:	Bat dipeptidyl peptidase IV
BHK:	Baby hamster kidney cells
BSL3:	Biosafety level 3
cDPP4:	Camel DPP4
DBT:	Delayed brain tumor cells
DNA:	Deoxyribonucleic acid
DPP4:	Dipeptidyl peptidase IV
DPP8:	Dipeptidyl peptidase VIII
DPP9:	Dipeptidyl peptidase IX
E:	Envelope protein
ExoN:	Exoribonuclease
FAP:	Fibroblast activation protein
fDPP4:	Ferret DPP4
FMDV:	Foot and mouth disease virus
GII.4:	Genogroup II genotype 4
gly:	Glycosylation site
gpDPP4:	Guinea pig DPP4
haDPP4:	Hamster DPP4
HCoV:	Human coronavirus
hDPP4:	Human dipeptidyl peptidase IV
HEK 293T:	Human embryonic kidney 293T cells
HIV:	Human immunodeficiency virus
hpi:	Hours post-infection
HRP:	Horseradish peroxidase
HSV-2:	Herpes simplex virus 2

Huh7:	Human hemochromatotic cells
IFA:	Immunofluorescence assay
kb:	Kilobases
kDa:	Kiladaltons
M:	Matrix protein
mDPP4:	Mouse dipeptidyl peptidase IV
MERS-CoV:	Middle East respiratory syndrome coronavirus
MHV:	Murine hepatitis virus
mL:	Milliliter
MOI:	Multiplicity of infection
N:	Nucleocapsid protein
NoV:	Norovirus
nsp10:	Nonstructural protein 10
nsp14:	Nonstructural protein 14
ORF:	Open reading frame
PCR:	Polymerase chain reaction
PFU:	Plaque-forming units
PNGase F:	Peptide-N-glycosidase F
poly-A:	Poly-adenosine
RBD:	Receptor binding domain
RdRp:	RNA-dependent RNA polymerase
RFP:	Red fluorescent protein
RIPA:	Radioimmunoprecipitation assay
rMERS:	Recombinant MERS-CoV
RMS:	Root mean square
RNA:	Ribonucleic acid
S:	Spike protein
SARS-CoV:	Severe acute respiratory syndrome coronavirus

s/n/c:	Substitutions per nucleotide site per cell infection
s/n/y:	Substitutions per nucleotide site per year
TBST:	Tris-buffered saline plus Tween
Tu:	Tunicamycin
µL:	Microliter
VSV:	Vesicular stomatitis virus

INTRODUCTION

0.1 Virus Evolution

Viruses are small infectious agents that replicate by infecting the cells of other organisms. Their general structure includes genomic material, RNA or DNA, a protein coat (capsid) that protects the genome, and sometimes an envelope that surrounds the capsid. Despite the small number of components, viruses accomplish a surprising amount of diversity, with breadth in what organisms they infect, their life-cycle strategies, and the impact they have on their hosts. In fact, estimates suggest that there are over 320,000 different viral species that infect mammals (Anthony *et al.* 2013a). If this number is extrapolated out to include all vertebrate species, the estimated number of viral species rises to over 3.6 million, representing an extraordinary amount of diversity. In addition to this remarkable diversity, viruses are also the most prevalent biological entity on the planet, with an estimated 10^{30} viruses in the ocean alone (Suttle 2007).

This impressive diversity comes from the unique characteristics that define viruses. Of particular note is the fact that they evolve rapidly. This rapid evolution is a result of their large population sizes, short generation times, and elevated mutation rates, particularly for RNA viruses. Compared to larger organisms, viruses evolve on a time frame that we can witness and document in order to understand what evolutionary processes shape microbial populations. Importantly, this rapid evolution has severe consequences for infectious diseases. For example, influenza A virus changes dramatically each year, requiring a new vaccine to be produced and distributed each outbreak season. Additionally, viruses can rapidly evolve to infect new species, or expand their host range; when viruses expand their host range to infect humans, it can have

severe consequences for public health. However, despite these negative aspects of rapid viral evolution, it also allows us to use viruses as model systems for investigating and understanding evolutionary processes.

One of the fundamental facets of evolution is the rate at which populations evolve. Unfortunately, the dynamics that govern the rate of molecular evolution for viruses are not quite straightforward. For other organisms, the rate of molecular evolution roughly follows neutral theory, which proposes that mutation rate, specifically the rate of neutral mutations, is the sole predictor of evolution rate (Kimura 1984). The success of this theory in estimating rates of evolution (Li *et al.* 1987, Kimura 1991, Bromham and Penny 2003) does not universally hold true for viruses, with molecular clock dynamics both supported (Gojobori *et al.* 1990, Leitner and Albert 1999) and refuted (Jenkins *et al.* 2002). The diversity and complexity of viral life cycles complicates our ability to understand the precise relationship between viral evolution rate and neutral theory.

As an important component of determining the evolution rate, viral mutation rates are noteworthy in that they span a broad range of values. DNA viruses can have rates as low as 10^{-8} substitutions per nucleotide site per cell infection (s/n/c) while RNA viruses can have rates as high as 10^{-3} (s/n/c) (Drake 1993, Sanjuan *et al.* 2010) (Figure 0.1). Previous studies have found that the relationship between evolution rate and mutation rate is linear for slowly mutating viruses (DNA and double-stranded RNA viruses) (Sanjuan 2012). However, this linear relationship breaks down for faster mutating viruses (single-strand RNA and retroviruses) (Sanjuan 2012) (Figure 0.1). This suggests that although faster mutating viruses have higher evolution rates, they cannot achieve indefinitely increasing evolution rates with increasing mutation rates; at some point, the mutation rate is so high that the virus population approaches

the error or extinction threshold (Eigen 1993, Bull *et al.* 2005). At this threshold, the population is overwhelmed with abundant transient deleterious mutations which arise more rapidly than can be removed by natural selection (Holmes 2003, Bull *et al.* 2007), causing the population to die out. Including the proportion of deleterious mutations into a model of evolution rate greatly improves the ability to explain empirical evolution rates seen for rapidly mutating viruses (Sanjuan 2012). Additionally, because of the diversity of virus life cycles, including within-host parameters, specifically the within-host reproductive ratio, also improves explanatory power (Peck *et al.* 2015a). With the increasing amount of available data on empirically estimated parameter values, we can improve our ability to understand what components influence the current evolution rates of viruses.

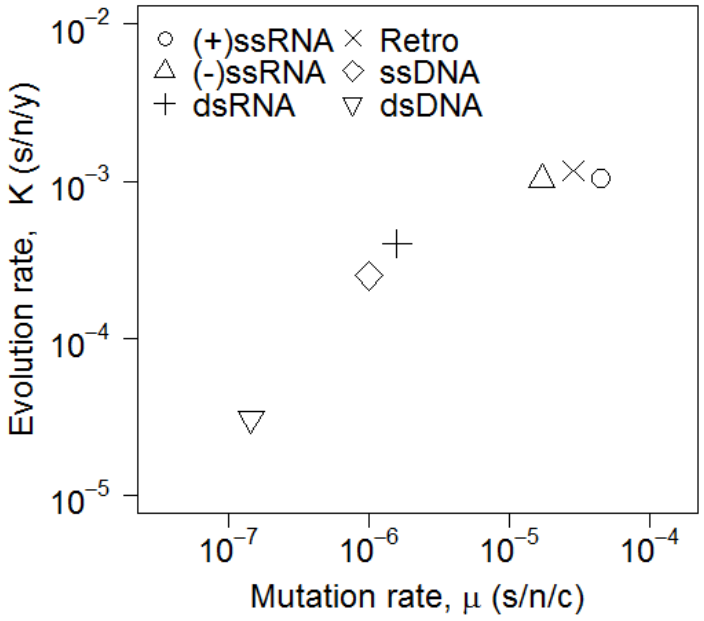


Figure 0.1: Virus evolution rate against mutation rate

Log-scale mean evolution rates (*K*) against mutation rates (μ) for each Baltimore class (data from Sanjuan (2012) and Sanjuan *et al.* (2010), respectively). (+)ssRNA, positive-sense single-stranded RNA; (-)ssRNA, negative-sense single-stranded RNA; dsRNA, double-stranded RNA; Retro, retroviruses; ssDNA, single-stranded DNA; dsDNA, double-stranded DNA.

One important component of virus evolution is host range expansion, or the ability to infect a new species. The infection of a new species, whether by adaptation or by exposure from geographic proximity, represents a new niche that a virus can explore, promoting adaptive radiation (Nichol *et al.* 1993, Rainey and Travisano 1998, MacLean and Bell 2002, Chow *et al.* 2004). Many viruses that infect the human population originated from a zoonotic source (Daszak *et al.* 2000, Taylor *et al.* 2001). The prevalence of zoonotic viruses that have gained the ability to infect humans is largely due to their rapid evolution rate; approximately one new emerging virus disease occurs each year (Howard and Fletcher 2012). RNA viruses, in particular, represent the majority of viral species that have emerged into humans (Howard and Fletcher 2012). This is often attributed to their high mutation rates, which result in a faster evolution rate (Figure 0.1). These higher rates of evolution enhance the ability with which viruses can expand into previously uninfected species. Unfortunately, the specific selective pressures that precede a host range expansion event are unknown; additionally, many of the evolutionary mechanisms that facilitate host range expansion for viruses still remain to be elucidated. Here, I examine virus evolution specifically in the context of viral host range expansion in order to better understand the dynamics that link mutation rate, evolution rate, emergence, and public health.

0.2 Host Range Expansion

Host range expansion is a complex process with many components determining which viruses can expand their host range and which species those hosts may comprise. Simply, there are three main steps for a pathogen to successfully expand its host range. The *first step is exposure*. In many cases, host jumps are caused by ecological factors that either change the proximity of species to each other or change the density of either species, increasing the likelihood of exposure (Holmes and Rambaut 2004).

The *second step is for the pathogen and host to be compatible* on a number of levels (e.g. by route of transmission, cell receptor, and/or infectious dose). Arguably, the most crucial barrier in this step for viruses is utilization of a new host species cell receptor by the viral spike protein, which is responsible for attachment and entry into the host cell. However, even compatibility between these two components may not be enough; for example, some viruses require specific host cell factors like proteases to facilitate viral entry (Gierer *et al.* 2013, Shirato *et al.* 2013, Millet and Whittaker 2015). Physiology can also play an important role; differences in the temperatures of human and avian airways are a crucial barrier in preventing avian influenza A viruses from replicating efficiently in humans (Scull *et al.* 2009).

The *third step is for the pathogen to be efficiently transmitted between individuals* of the new host species (Woolhouse *et al.* 2005). Most zoonotic pathogens are *not* effectively transmitted between humans (Woolhouse and Gowtage-Sequeria 2005), causing humans to act as a dead end host in these scenarios. However, if a zoonotic virus is compatible with human cells *and* is able to transmit efficiently between humans, it has the potential to result in a severe outbreak. For example, although the avian influenza A virus H5N1 has been transmitted multiple times to humans from avian species, its inability to transmit reliably between humans is a primary hypothesis for the absence of a H5N1 pandemic (Hayden and Croisier 2005).

Successful emergence of a virus into a new host species occurs when all of the above requirements have been met. After emergence, viruses can either become specialists by adapting specifically to the environment of the new host; or, they can become generalists by maintaining the ability to infect the ancestral species. Examples of each of these include the specialist HIV, which originated from a primate (Gao *et al.* 1999), but has changed such that HIV can no longer infect non-human primate species. Rabies virus, however, is a generalist in that it is capable of

efficiently infecting a wide range of mammalian species (Rupprecht *et al.* 2002). Note that while rabies virus is not commonly transmitted between humans, this is likely due to behavioral dynamics and not due to the biological inability of the virus to be transmitted. Previous studies have investigated what potential drivers can lead to the selection of a specialist or generalist virus population (Bono *et al.* 2012) with advantages and disadvantages of both infection strategies.

0.3 Coronaviruses

Here I focus on coronaviruses, a diverse family of viruses that has expanded its host range many times over the course of its evolutionary history. Coronaviruses are enveloped, single-stranded positive-sense RNA viruses with particularly large genomes (28-32 kb) that can infect a wide range of avian and mammalian hosts. Genomes contain a 5' cap and 3' poly-A tail, with the genome organization divided into nonstructural protein genes and structural and accessory genes. The core structural proteins include S (spike), E (envelope), M (matrix), and N (nucleocapsid) proteins. While accessory genes vary among coronaviruses, and may include some strain-specific structural glycoproteins, the order of the structural proteins is highly conserved as S, E, M, and N. The ~180 kDa spike glycoprotein (S) mediates entry into the host cell and surrounds the virus particle, yielding a crown-like appearance. Coronaviruses utilize a variety of cellular proteins as receptors (Perlman and Netland 2009, Graham *et al.* 2013), with cleavage of the spike protein crucial for mediating virus-host membrane fusion and subsequent entry into the cell.

Bats, rodents, and birds act as the natural reservoir species for many coronaviruses (Li *et al.* 2005a, Woo *et al.* 2009, Wang *et al.* 2015), but host range expansion has been prevalent, with over 20 species currently able to be infected by coronaviruses (Figure 0.2). To date, six human

coronaviruses (HCoV) have been identified, with each of these hypothesized to have originated as a zoonotic strain that underwent host range expansion. Of these six HCoVs, four are associated with mild respiratory disease: HCoV-229E, HCoV-HKU1, HCoV-NL63, and HCoV-OC43. Two of these (HCoV-229E and HCoV-NL63) are thought to have originated from bats. While the standing hypothesis for HCoV-OC43 was that it likely emerged from a bovine reservoir species (Vijgen *et al.* 2005), new analyses suggest that the original host of this lineage, and perhaps HCoV-HKU1 as well, may have been a murine species (Lau *et al.* 2015). In addition to HCoVs that cause mild symptoms, two strains have emerged to cause severe disease, including severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV). These two viruses combined have resulted in over 1,100 deaths, with MERS-CoV still circulating in the human population, causing heightened concern due to the lack of vaccines or therapeutics. The threat to public health caused by the emergence of these highly pathogenic strains into humans draws attention to the importance of understanding both the biochemical and evolutionary mechanisms of coronavirus host range expansion.

The selective pressures that drive coronavirus host range expansion are not yet understood. Additionally, a pattern of coronavirus host range characteristics or biases is not yet evident. Some coronaviruses appear to be generalists, capable of infecting many different orders of mammals. For example, *Betacoronavirus 1* has been detected in dogs, humans, and numerous ungulate species (Erles *et al.* 2003, Hasoksuz *et al.* 2007, Alekseev *et al.* 2008). Other coronaviruses have been detected in only a single mammalian order, such as the many SARS-like coronaviruses that have been detected only in bats (Tang *et al.* 2006, Drexler *et al.* 2010). Bats exist as the largest reservoir for coronaviruses, potentially due to their diversity,

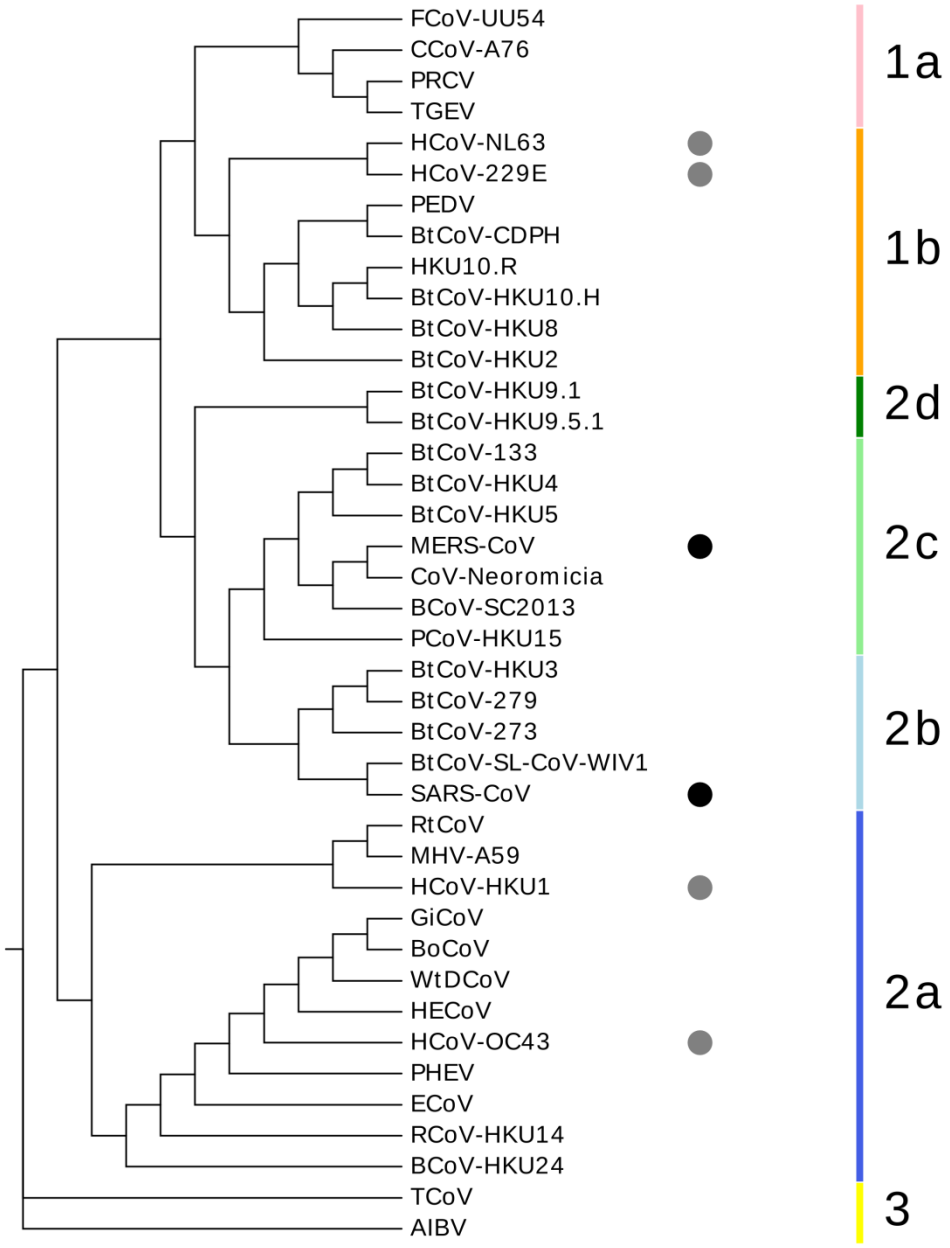


Figure 0.2: Phylogenetic tree of whole-genome length coronavirus sequences

Phylogenetic tree of a subset of coronaviruses representing groups 1, 2, and 3 (the alpha-, beta-, and gammacoronaviruses; colors represent each designated group/clade). The newly discovered deltacoronavirus clade is not depicted. Sequences were aligned using MAFFT (Katoh *et al.* 2002). The phylogenetic tree was generated using maximum likelihood with the PhyML package (Guindon *et al.* 2010) and visualized using EvolView (Zhang *et al.* 2012). Circles denote human coronavirus strains; gray indicates viruses that cause mild symptoms and black indicates viruses that have emerged to cause severe disease. Of note is that coronaviruses within the same clade infect many species. See Appendix A for strain abbreviation key and NCBI accession numbers.

immunology, physiology, and ability to traverse wide geographic regions during seasonal migrations (Dobson 2005, Calisher *et al.* 2006). Due to this, metagenomics analyses have focused on examining the diversity and prevalence of coronaviruses in bats. Studies have found varying levels of coronavirus diversity in bat populations in North America (Donaldson *et al.* 2010) and China (Ge *et al.* 2012, Wu *et al.* 2012b), as well as detecting individual strains in bat populations worldwide (reviewed in Drexler *et al.* 2014). Furthermore, novel coronaviruses continue to be discovered globally, with recent findings coming from Mexico (Anthony *et al.* 2013b), Brazil (Goes *et al.* 2013), and South Africa (Ithete *et al.* 2013). Continued efforts to sample bat populations worldwide will help us better estimate the diversity and prevalence of coronaviruses, as well as more thoroughly map their phylogenetic relationships.

The biggest threat to the public health comes from coronaviruses that emerge into the human population to cause severe disease. The first of these, SARS-CoV, emerged from bats into the human population in 2003 and infected over 8,000 people with a 9.6% mortality rate before it was controlled by public health measures (Cherry 2004). Despite the awareness raised for coronaviruses during this outbreak, reconstructing the evolutionary path that led to the host range expansion of SARS-CoV has been surprisingly difficult. SARS-CoV is closely related to bat coronaviruses, with up to 92% nucleotide sequence identity to its closest detected relative (Li *et al.* 2005a). However, no virus identical to SARS-CoV has ever been isolated from bats, raising suspicion on whether the virus did actually emerge from this reservoir species. Data in support of an origin from bats include an analysis of the SARS-CoV receptor, angiotensin-converting enzyme 2 (ACE2). The bat ACE2 gene was found to show the signature of recurrent positive selection ($d_N/d_S > 1$) (Figure 0.3A), suggesting that a SARS-like coronavirus was circulating in the bat population before emerging into humans (Demogines *et al.* 2012). Data refuting the bat

origin hypothesis include the fact that, initially, no bat coronaviruses were found to utilize ACE2 or any ACE2 ortholog (Ren *et al.* 2008). In fact, it took over a decade before researchers were able to identify a SARS-like coronavirus capable of utilizing human, civet, and Chinese horseshoe bat ACE2 for cell entry (Ge *et al.* 2012). The discovery of this virus (bat SL-CoV-WIV1) provides strong evidence that SARS-CoV originated from a bat reservoir (see Figure 0.2). Currently, the best supported hypothesis is that an intermediate host species, specifically the civet, played an important role in facilitating the host range expansion event, either by imposing a specific selective pressure or by putting the virus into close proximity with humans (reviewed in Wang and Eaton 2007, Plowright *et al.* 2015). However, the identification of bat SL-CoV-WIV1 suggests that an intermediate host may not have been required for emergence of SARS-CoV into humans. Further research is needed to determine the precise mutational and cross-species path of SARS-CoV in order to help reveal how new human pathogens emerge.

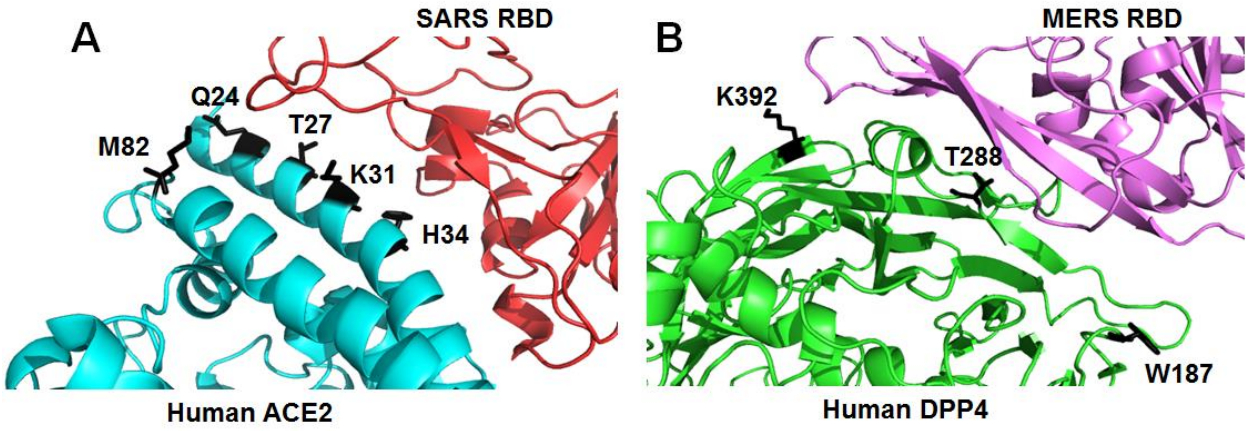


Figure 0.3: Positive selection on coronavirus host cell receptors

(A) Human ACE2 (light blue) bound to the SARS-CoV receptor-binding domain (red) (PDB 2AJF), with homologous residues under positive selection in bats labeled (Demogines *et al.* 2012). (B) Human DPP4 (green) bound to the MERS-CoV RBD (purple) (PDB 4L72), with homologous residues under positive selection in bats labeled (Cui *et al.* 2013). Structures visualized using PyMOL. From Peck *et al.* 2015c.

In 2012, MERS-CoV emerged into the human population in Saudi Arabia. As of November 2016, there were 1,813 confirmed cases with a 36% mortality rate (WHO 2016). MERS-CoV is grouped phylogenetically into the C betacoronavirus clade along with the bat coronaviruses BtCoV-HKU4 and BtCoV-HKU5 (Figure 0.2) (van Boheemen *et al.* 2012, Drexler *et al.* 2014, Peck *et al.* 2015c). Unlike SARS-CoV, MERS-CoV utilizes dipeptidyl peptidase 4 (DPP4) as an entry receptor (Raj *et al.* 2013). To date, only MERS-CoV and BtCoV-HKU4 have been found to utilize DPP4 (Wang *et al.* 2014, Yang *et al.* 2014); the presence of a closely related bat coronavirus that uses the same host cell receptor as MERS-CoV provides strong support for the emergence of MERS-CoV from a bat lineage. Adaptation of MERS-CoV to humans specifically is supported by the fact that MERS-CoV utilizes human DPP4 (hDPP4) more efficiently than bat DPP4 (bDPP4), yet BtCoV-HKU4 utilizes both with equivalent efficiency (Yang *et al.* 2014). Thus far, the host cell receptors of BtCoV-HKU5 and other closely related group 2c coronaviruses have not been identified.

The similarity between MERS-CoV and group 2c bat coronaviruses (Figure 0.2) strongly suggests that MERS-CoV originated from bats. Additionally, similar to the SARS-CoV story, bDPP4 has been found to be under strong positive selection (Figure 0.3), which may be due to the circulation of DPP4-utilizing coronaviruses in bats (Cui *et al.* 2013). It is not clear whether this signal is from the progenitor of MERS-CoV or from other coronaviruses that utilize DPP4, such as BtCoV-HKU4. As with SARS-CoV, the ancestral virus has been elusive, with no detection of a full-length MERS-CoV sequence in a bat population to date. The best example has been a coronavirus isolate from a bat population where a 190-nucleotide fragment of the RNA dependent RNA polymerase (RdRp) gene was 100% identical to MERS-CoV (Memish *et al.* 2013). Since this discovery, bat coronaviruses with high sequence similarity to MERS-CoV have

been detected in geographic regions very distant from its original emergence location in Saudi Arabia. RdRp gene sequences with 96.5% and 99.6% amino acid identity to MERS-CoV were detected from bat populations in Mexico (Anthony *et al.* 2013b) and South Africa (Ithete *et al.* 2013), respectively. These discoveries emphasize the importance of further metagenomics analyses of bat viromes in widespread geographic locations.

Looking at the phylogenetic tree of coronaviruses (Figure 0.2), the lineages that have emerged into humans occupy several clades. This suggests that there is not a single lineage of zoonotic coronaviruses that is capable of expanding its host range into humans. Instead, many lineages have been able to emerge into the human population. Additionally, these lineages utilize different host cell receptors for entry into the cell. SARS-CoV and HCoV-NL63 utilize ACE2 (Li *et al.* 2003, Hofmann *et al.* 2005), MERS-CoV utilizes DPP4 (Raj *et al.* 2013), and HCoV-229E utilizes a protein known as aminopeptidase N (APN) (Yeager *et al.* 1992). The receptor molecules of HCoV-HKU1 and HCoV-OC43 have not yet been identified, however the presence of O-acetylated sialic acid has been shown to serve as a receptor determinant (Krempl *et al.*1995, Huang *et al.* 2015). The diversity of receptors utilized, combined with the large number of species that coronaviruses have evolved to infect, poses the question of whether coronaviruses have an increased capacity for host range expansion relative to other RNA viruses. Perhaps there are unique characteristics of coronaviruses that increase their ability to emerge into new species, providing an interesting case study of the evolution of host range.

0.4 Overview of Chapters

This dissertation focuses on understanding the recent host range expansion of MERS-CoV from a biochemical and evolutionary perspective. *Chapter 1* investigates the biochemical determinants of permissivity to MERS-CoV in the context of the mouse host cell receptor. The

interactions between the MERS-CoV receptor binding domain (RBD) and the human and mouse DPP4 orthologs are analyzed both computationally and experimentally to provide data on what particular residues and residue properties are important for mediating MERS-CoV permissivity. I find that two residues are capable of mediating this interaction; two amino acid changes within the mouse DPP4 (mDPP4) backbone allow it to successfully support infection. The biochemical explanation behind these two changes is 1) the strengthening of a hydrophobic core in the MERS-CoV RBD and 2) the removal of a non-conserved glycosylation site. Data gathered from this chapter helped generate a transgenic mouse model that will be utilized to study MERS-CoV pathogenesis.

Chapter 2 investigates whether there is a broader signature of permissivity in additional DPP4 orthologs. Other traditional small animal model species, namely ferrets, guinea pigs, and hamsters, are also nonpermissive to MERS-CoV infection. Using the insight gained in my previous experiments, I determine whether the same biochemical mechanisms that conferred permissivity of mDPP4 to MERS-CoV are applicable to these other species. Many nonpermissive species have glycosylation sites near the location that aligns with the mDPP4 motif, suggesting that this may potentially be a broadly acting mechanism for blocking MERS-CoV infection. I find that although successful infection is never achieved without removal of glycosylation in these orthologs, the interaction dynamics are more complex, suggesting that there are other important key determinants that will require more work to reveal.

Chapter 3 turns to the field of evolutionary biology in order to determine how coronaviruses may expand their host range when receptor incompatibilities are the primary barrier. Using the toolkit established in Chapters 1 and 2, I attempt to adapt MERS-CoV to utilize the wildtype mDPP4 molecule. However, due to the likely requirement of multiple

mutations, including remodeling of the RBD to bind around the mDPP4 glycan, there was no adaptation of MERS-CoV to wildtype mDPP4. One potential hypothesis is that there was not enough genetic variation present in the population to achieve the multiple mutations required for successful binding (contingent on it being biologically possible). In *Chapter 4*, I present future directions for generating a MERS-CoV mutator that can be utilized in adaptation experiments to enhance the sequence space accessible by the viral population. Mutations in the nsp14 gene of coronaviruses have been found to generate elevated mutation rates, allowing mutators to be a possible characteristic that permits coronaviruses to expand their host range more frequently, and with more success, than other viruses.

The increasing prevalence of emerging pathogens heightens the need to understand the biochemical and evolutionary dynamics of host range expansion events. Coronaviruses, particularly, draw attention due to the recent emergence of highly pathogenic strains that lack effective vaccines or therapeutics. For many viruses, including SARS-CoV and MERS-CoV, the specific factors contributing to the host range expansion into humans are still unclear. Elucidating the selective pressures imposed on the virus populations will help reveal the specific path of emergence of SARS-CoV and MERS-CoV. Understanding the adaptation pathway of coronaviruses to new species will help us better catalogue the selective pressures driving host range expansion and will provide a framework for which we can set up ecological and/or geographical preventative measures. Although the limited number of outbreak events in coronaviruses reduces our ability to create a comprehensive picture of key host range expansion determinants, understanding the potential contribution of different biochemical and evolutionary mechanisms is a useful place to start.

CHAPTER 1: BIOCHEMICAL DETERMINANTS OF MOUSE DPP4 PERMISSIVITY TO MERS-CORONAVIRUS

1.1 Introduction

Coronaviruses have expanded their host range into humans multiple times over the course of their evolutionary history (reviewed in Peck *et al.* 2015c). In recent history, two of these events have resulted in viruses with pandemic potential. MERS-CoV jumped into humans in 2012, providing the second case of a coronavirus emergence event resulting in severe disease. As of November 2016, MERS-CoV has infected over 1,800 people, with a 36% percent mortality rate (WHO 2016). Despite being similar to SARS-CoV in that it causes severe disease, MERS-CoV shares only 57% sequence identity with SARS-CoV and utilizes a different host cell receptor. Whereas SARS-CoV utilizes ACE2 for entry (Li *et al.* 2003), MERS-CoV utilizes DPP4 (Raj *et al.* 2013). DPP4 is a cell-surface protease that has a catalytic role in selectively removing the N terminus from certain peptides (Lambeir *et al.* 2003). It has been well studied due to the role it plays in glucose metabolism, immune responses, adhesion, and apoptosis (Boonacker and Noorden 2003). MERS-CoV was the first coronavirus identified to utilize this receptor, although since then the bat coronavirus BtCoV-HKU4 has also been found to utilize DPP4 (Yang *et al.* 2014). To date, MERS-CoV is known to successfully enter cells utilizng DPP4 from humans, nonhuman primates, bats (de Wit *et al.* 2013b), rabbits (Haagmans *et al.* 2015), camels, horses, and to a lesser extent goats (Barlan *et al.* 2014, Eckerle *et al.* 2014). However, it is unable to enter cells using DPP4 from traditional small animal model species such as mice (Cockrell *et al.* 2014, Coleman *et al.* 2014), ferrets (Raj *et al.* 2014), and hamsters (de

Wit *et al.* 2013a, van Doremalen *et al.* 2014). This species restriction prevents the adequate study of MERS-CoV pathogenesis and limits the development of vaccine strategies or alternate therapeutics. Understanding the determinants that mediate MERS-CoV permissivity in these species will provide novel insights into interactions between the MERS-CoV RBD and DPP4, as well as help in the development of new small animal models.

1.2 Identify key mutations in mouse DPP4 that allow it to support MERS-CoV infection (modified from Cockrell *et al.* 2014)

In order to investigate the relationship between MERS-CoV and DPP4, we designed an ectopic expression system utilizing the 945ΔRRE expression vector, a lentiviral vector derived from pTK945, to constitutively express the human and mouse DPP4 genes in human embryonic kidney 293T (HEK 293T) cells. These cells lack detectable endogenous expression of hDPP4 (Zhao *et al.* 2013). Both orthologs were expressed either as full-length proteins or as fusions to the Venus protein at the carboxy terminus as a marker of green fluorescence. HEK 293T cells were transfected with 3 μg of the appropriate DPP4 expression plasmid. At ~24 hours post-transfection, cells were infected at a multiplicity of infection (MOI) of 5 with a recombinant MERS-CoV strain designed to express tomato red fluorescent protein (rMERS-CoV-RFP) (Figures 1.1A and B). This rMERS-CoV strain is derived from the original EMC2012 isolate (van Boheemen *et al.* 2012) and was previously demonstrated to infect and replicate in a manner similar to wildtype MERS-CoV (Scobey *et al.* 2013). Cells were imaged ~24 hours post-infection (hpi) by rMERS-CoV-RFP. High transfection efficiency was seen by visualizing the DPP4-Venus fusion constructs (Figures 1.1A and B).

Cells expressing hDPP4 were readily infected with rMERS-CoV-RFP virus (Figure 1.1A), whereas overexpressing mDPP4 did not result in infection (Figure 1.1B). In fact,

overexpression of mDPP4 showed levels of infection that were equivalent to control HEK 293T cells (Figure 1.1B). To confirm that mDPP4 was being expressed, and a negative infection

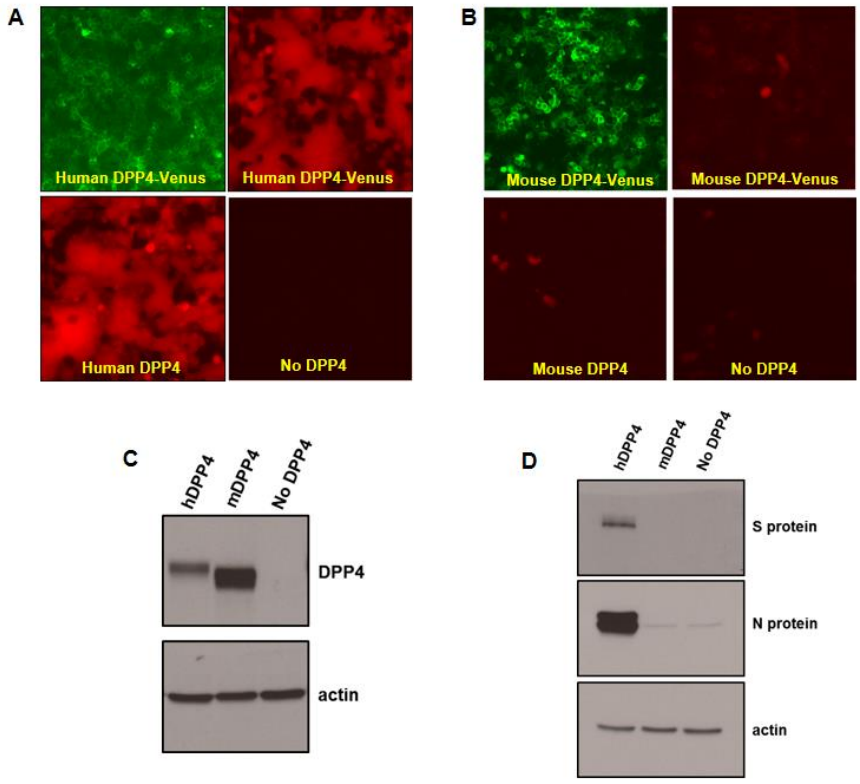


Figure 1.1: MERS-CoV infection utilizing hDPP4 and mDPP4

mDPP4 does not support MERS-CoV infection. HEK 293T cells were transfected with 3 µg of plasmid expressing (A) hDPP4 or hDPP4-Venus fusion and (B) mDPP4 or mDPP4-venus fusion. At ~24 hours post-transfection, cells were infected with rMERS-CoV-RFP virus at an MOI of 5. Venus fusion proteins were assessed by fluorescence microscopy at ~48 hours post-transfection. In independent experiments, infection with rMERS-CoV-RFP virus was assessed for red cells by fluorescence microscopy at ~18 hpi. (C) Western blot analysis demonstrates overexpression of mDPP4 and hDPP4. Extracts were prepared at ~48 hours post-transfection using AV lysis buffer (Agnihothram *et al.* 2014), and samples were heat inactivated for 60 min at 90 °C for removal from a BSL3 facility and resolved on an 8% SDS-PAGE gel. Blots were probed with primary goat-anti-DPP4 polyclonal antibody (R&D Systems) at 1:1,000 in 1X Tris TBST or goat anti-actin polyclonal antibody (Santa Cruz) and detected with a secondary rabbit anti-goat–HRP-conjugated antibody (Sigma) at 1:10,000 in 1X TBST in 5% milk. (D) Western blot analysis of MERS-CoV S and N proteins. Lysates were collected at ~18 hpi and treated as in panel C. Blots were probed with primary mouse polyclonal antiserum at 1:400, raised to S and N proteins as described previously (Agnihothram *et al.* 2014), and detected with a secondary goat anti-mouse–HRP (GE Healthcare) at 1:10,000 in 1X TBST in 5% milk. From Cockrell *et al.* 2014.

readout was not based on absence of the protein, we performed Western Blot analyses to detect the presence of mDPP4 after transfection into HEK 293T cells (Figure 1.1C). Bands present at ~110 kDa for both hDPP4 and mDPP4 indicate successful expression of both constructs. Additionally, Western blot analysis for MERS-CoV spike (S) and nucleocapsid (N) proteins show detection of both viral proteins in cells expressing hDPP4 but not mDPP4 (Figure 1.1D). These results support hDPP4, but not mDPP4, as a functional receptor for MERS-CoV (Cockrell *et al.* 2014).

The crystal structure of hDPP4 complexed with the MERS-CoV RBD (Lu *et al.* 2013, Wang *et al.* 2013) shows that blades IV and V of the β-propeller domain primarily interact with the MERS-CoV RBD (Figure 1.2C). Unfortunately, the crystal structure of mDPP4 has not yet been solved. This limits our ability to analyze the binding interactions between mDPP4 and the MERS-CoV RBD computationally. Due to this limitation, we first set out to test whether the mDPP4 backbone was capable of acting as an entry receptor for MERS-CoV. We generated a chimeric DPP4 molecule that introduced residues 273 through 340 (chDPP4 (273-340)) from hDPP4 into mDPP4 using overlap extension PCR (Figure 1.2A) (Cockrell *et al.* 2014). These residues compose the regions of blades IV and V that interact with the MERS-CoV RBD. Within this region, there are 22 different amino acid identities between hDPP4 and mDPP4 (Figure 1.2A), suggesting that there may be important differences within this region that prevent mDPP4 from acting as a functional receptor. Expressing chDPP4 (273-340) in HEK 293T cells shows that these cells are equivalently susceptible to cells expressing hDPP4 (Figure 1.2B). These results indicate 1) that the hDPP4 residues 273 to 340 confer MERS-CoV permissivity to mDPP4 and 2) that the mDPP4 molecule as a backbone is capable of supporting MERS-CoV

infection. Thus, there must be important differences between blades IV and V of hDPP4 and mDPP4 that determine permissivity.

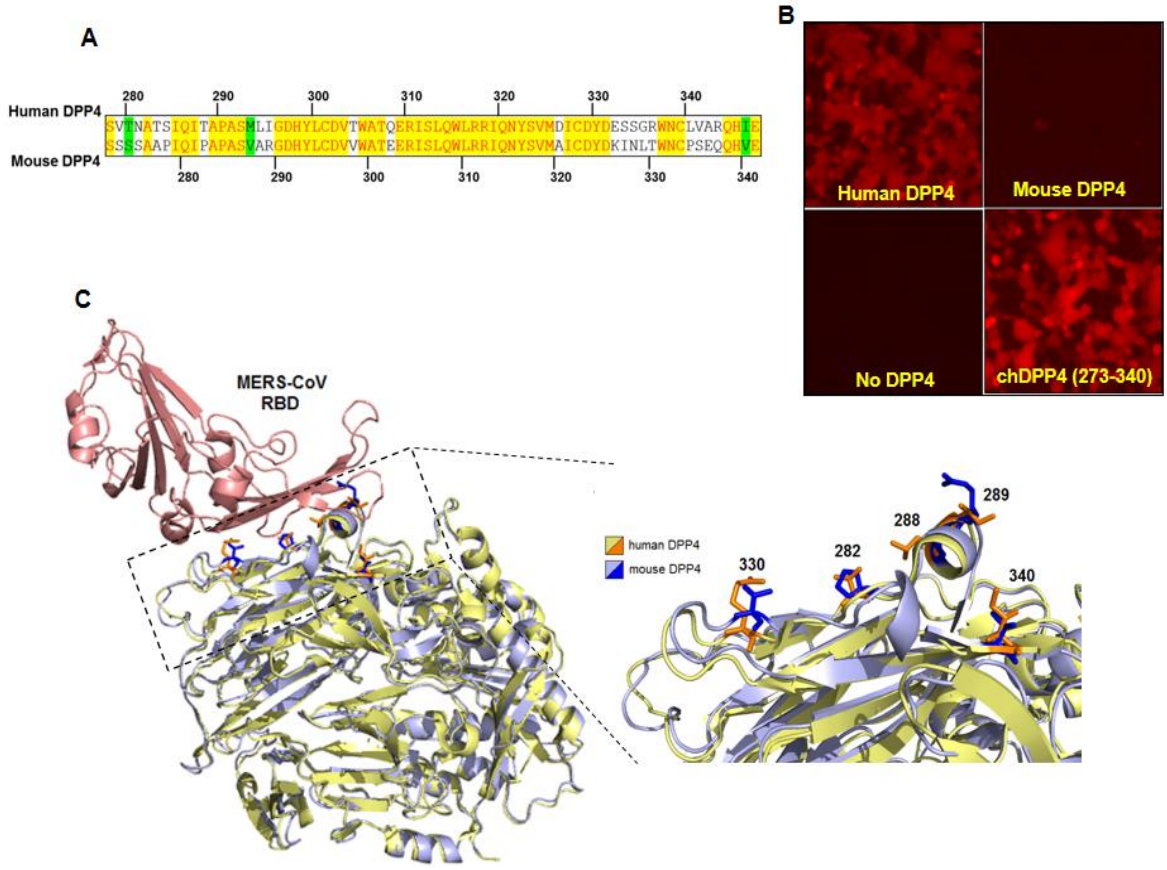


Figure 1.2: mDPP4 as a backbone can support MERS-CoV infection

Blades IV and V from the β-propeller of hDPP4 make mDPP4 permissible to MERS-CoV infection. (A) Vector NTI protein sequence alignment of human (top strand) and mouse (bottom strand) DPP4 molecules. Yellow highlighted regions indicate conserved amino acids, white regions signify amino acids that are functionally different (i.e., hydrophobic and hydrophilic), and green highlighting indicates amino acids that are different but functionally similar (i.e., the threonine and serine are both polar and uncharged). (B) HEK 293T cells were transfected with the indicated DPP4. At ~20 hours post-transfection, cells were infected with rMERS-CoV-RFP virus at an MOI of 5, and infection was assessed ~18 hpi by fluorescence microscopy. (C) 3D molecular PyMOL software was employed to visualize the mDPP4 structure overlaid onto the hDPP4 structure. The hDPP4 structure was based upon the crystal structure resolved in context with the MERS S RBD (PDB code 4L72). MERS S protein is displayed in red, hDPP4 in yellow, and mDPP4 in blue. The mDPP4 sequence was threaded using the I-TASSER software (Zhang 2008). The expanded view depicts the DPP4 region at the interaction surface. Numbered and highlighted are the specific amino acids chosen for mutation in the mDPP4 protein. From Cockrell *et al.* 2014.

In order to further explore the differences between hDPP4 and mDPP4 in this region, we threaded the mDPP4 molecule using I-TASSER (Zhang 2008) to produce a predicted protein structure. Overlaying the resulting molecule with hDPP4 and employing three-dimensional (3D) visualization using PyMOL (Molecular Graphics System, Version 1.6.0.0 Schrodinger, LLC) shows that overall, these two proteins are highly similar in structure (Figure 1.2C), keeping in mind that we are using a predicted structure for mDPP4. This supports our experimental findings that the mDPP4 molecule can act as a backbone to support infection and that there are important differences at the interface of the receptor and the virus RBD.

Due to a lack of obvious structural differences, we computationally interrogated the interface of mDPP4 and the MERS-CoV RBD to determine what amino acid differences between hDPP4 and mDPP4 might explain the differences in permissivity. Based on analysis in Rosetta (Rohl *et al*. 2004), we identified five functionally variant surface amino acids in this region of mDPP4 that may affect MERS-CoV binding. Using overlap extension PCR, we generated various chimeric mDPP4 molecules corresponding to the changes introduced based on the hDPP4 amino acid identities. These mutations include P282T, A288L, R289I, T330R, and V340I (residue numbering relative to mDPP4) (Figure 1.3A). HEK 293T cells were transfected with the indicated chimeric mDPP4 construct. At ~20 hours post-transfection, cells were infected with rMERS-CoV-RFP at an MOI of 1 and imaged ~18 hpi. A chimeric mDPP4 molecule with all five identified mutations (chDPP4 (5 mut)) showed MERS-CoV infection comparable to that of hDPP4 (Figure 1.3B). This suggests that one or more of these mutations may be sufficient for mediating MERS-CoV permissivity in the context of mDPP4. Previous structural studies of hDPP4 suggest that two distinct interactions are important for MERS-CoV binding, specifically one interaction on blade IV and one on blade V (Wang *et al.* 2013). Our set of identified

mutations includes two located on blade IV (T330R and V340I) and three on blade V (P282T, A288L, and R289I). These two sets of mutations were introduced independently into mDPP4 in

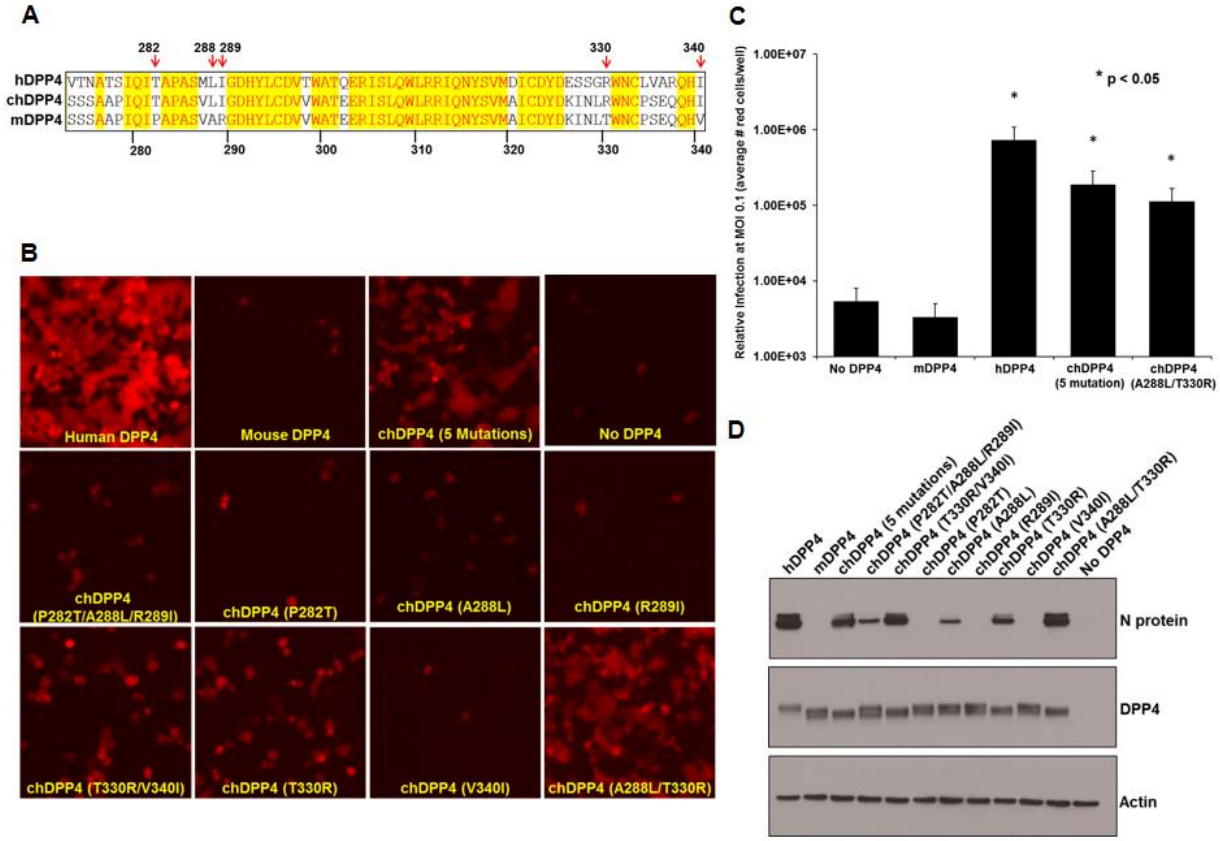


Figure 1.3: MERS-CoV infection is dependent upon specific amino acids in DPP4

(A) Vector NTI protein sequence alignment of hDPP4 (top strand) with chDPP4 (middle strand) and mDPP4 (bottom strand) indicating positions of introduced human mutations with red arrows. (B) HEK 293T cells were transfected with the indicated DPP4 molecule. At ~20 hours post-transfection, cells were infected with rMERS-CoV-red virus at MOI of 1, and infection was assessed ~18 hpi by fluorescence microscopy. (C) In an independent experiment, cells overexpressing the indicated DPP4 constructs were infected with rMERS-CoV-red virus at MOI of 0.1, 0.01, and 0.001 on six-well plates. At ~18 hpi, cells were scored at the following MOI: no DPP4 and mDPP4, 0.1; chDPP4 P282T A288L R289I T330R V340I ["chDPP4 (5 mutations)"] and chDPP4 A288L T330R, 0.01; and hDPP4, 0.001. Values were normalized to an MOI of 0.1 and expressed as relative infection at 0.1. Human DPP4, chDPP4 P282T A288L R289I T330R V340I, and chDPP4 A288L T330R showed a significant increase in infection over mDPP4 (*, $P < 0.05$, Student's t test). (D) Western blots demonstrating overexpression of hDPP4, mDPP4, and each chDPP4 molecule, N protein of infected cells, and β-actin as a loading control. Western blots were prepared and probed as described in Figure 1.1C and D. From Cockrell *et al.* 2014.

order to test their impact on permissivity to rMERS-CoV-RFP. Neither set recapitulated the levels of infection seen when all five mutations were included (Figure 1.3B). Further investigation led us to introduce each mutation singly and determine the impact on receptor permissivity. This experiment revealed that A288L and T330R were partly responsible for the observed increase in infection from each group (on blades V and IV, respectively) (Figure 1.3B). This conclusion is supported by Western blot analysis, which shows the detection of N protein in infected cells expressing the A288L or T330R construct, but not the other three individual mutants (Figure 1.3D). Combining these two mutations into a single construct (chDPP4 288, 330) results in high levels of infection by rMERS-CoV-RFP that are comparable to when all five mutations are included (Figure 1.3B). Quantifying the number of MERS-CoV-infected red cells reveals that chDPP4 288, 330 exhibits nearly a 1.5-log increase in infection compared to mDPP4 (Figure 1.3C). Western blot analysis also supports this conclusion (Figure 1.3D). Although no mDPP4 mutants achieved a level of infection quantitatively similar to hDPP4, the dramatic increase in infection offers strong support for the identified residues playing an important role in mediating permissivity.

Here, we focused on receptor compatibilities between mDPP4 and the MERS-CoV RBD; however, other factors in the replication cycle of MERS-CoV may be incompatible in the context of the mouse. For example, host proteases are required to cleave the MERS-CoV spike protein in order to allow entry into the cell (Gierer *et al.* 2013, Shirato *et al.* 2013), and these may differ between species. To determine whether additional restriction factors in rodent cells may affect permissivity to MERS-CoV, we transfected hDPP4 and our chimeric mDPP4 constructs into both mouse and hamster cell lines (NIH 3T3 and BHK, respectively). Receptors that were permissive when transfected into HEK 293T cells were also permissive in mouse and hamster

cells (Figure 1.4A and B, respectively), suggesting that additional rodent-specific factors are not responsible for preventing MERS-CoV infection, at least in an *in vitro* context.

Overall, our results indicate that successful infection by MERS-CoV requires a combination of at least two mutations in mDPP4 (A288L and T330R) (Cockrell *et al.* 2014), lying within blades V and IV, respectively. These results are in agreement with previous crystal

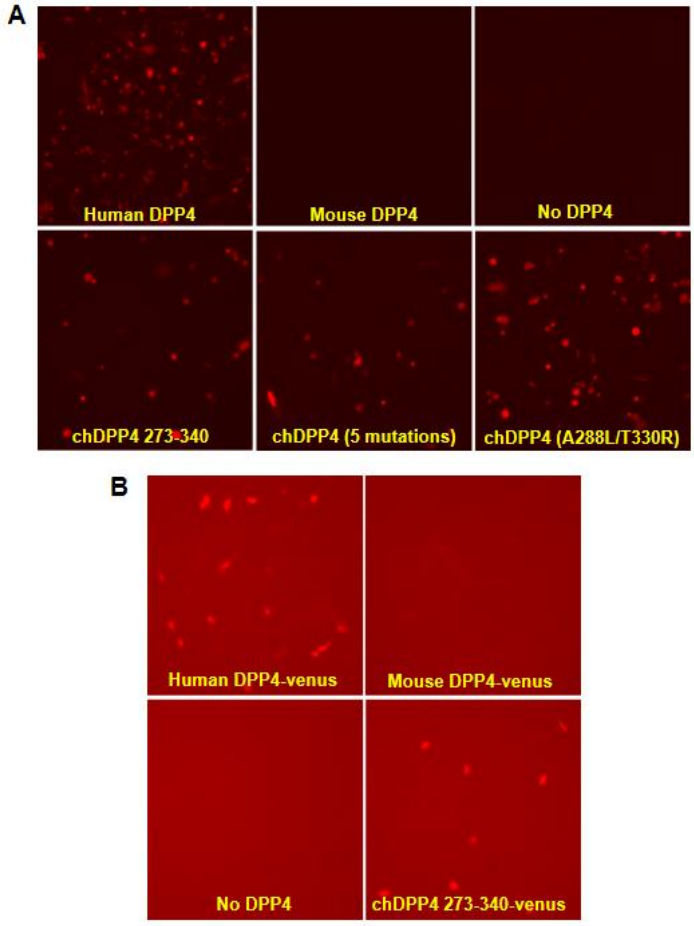


Figure 1.4: Human and chimeric DPP4 molecules can support MERS-CoV infection in hamster and mouse cells

(A) BHK cells were electroporated with the indicated DPP4 molecules. At ~20 hours post-transfection, cells were seeded in 6-well plates and infected with rMERS-CoV-RFP at an MOI of 2. (B) Mouse NIH 3T3 cells were transfected using Nucleofection (according to the Amaxa procedure) with the indicated DPP4 molecules. Cells were seeded into 12-well plates and infected with rMERS-CoV-RFP at an MOI of ~4 at 24 hours post-Nucleofection. All infections were assessed at ~24 hpi by fluorescence microscopy. From Cockrell *et al.* 2014.

structure data from hDPP4 and the MERS-CoV RBD, which suggest the hDPP4 equivalents (L294 and R336) are critical residues for binding and successful infection of MERS-CoV (Lu *et al.* 2013, Wang *et al.* 2013).

Using this information, we generated a transgenic mouse using the CRISPR/Cas genome editing technique that engineers these two mutations into mDPP4. Whereas other mouse models have been developed, each comes with its own caveat that reduces the effectiveness of pathogenesis research. The first transgenic mouse was generated by transient adenovirus-mediated hDPP4 expression and resulted in susceptibility to MERS-CoV (Zhao *et al.* 2014). However, the transient nature of this model, as well as immune responses that can be triggered by the adenovirus delivery system make this model less than ideal. The second mouse model produced showed global expression of hDPP4 and is also successfully infected by MERS-CoV (Agrawal *et al.* 2015). However, this model results in high viral titers in most organs including the brain, suggesting that additional improvements are needed to more faithfully phenocopy the human disease model. In addition, the enzymatic activity of DPP4 can have detrimental effects, particularly when the protein is overexpressed (Takasawa *et al.* 2010), and the impact of these effects in the transgenic mouse model should be explored. Our mouse model is still in the validation phase, and while it may come with its own set of caveats, the endogenous expression of mDPP4 behind its natural promoter suggests that we may overcome some of the previous complications. Overall, the production of an accurate mouse model will provide a new tool for studying pathogenesis and developing potential therapeutics.

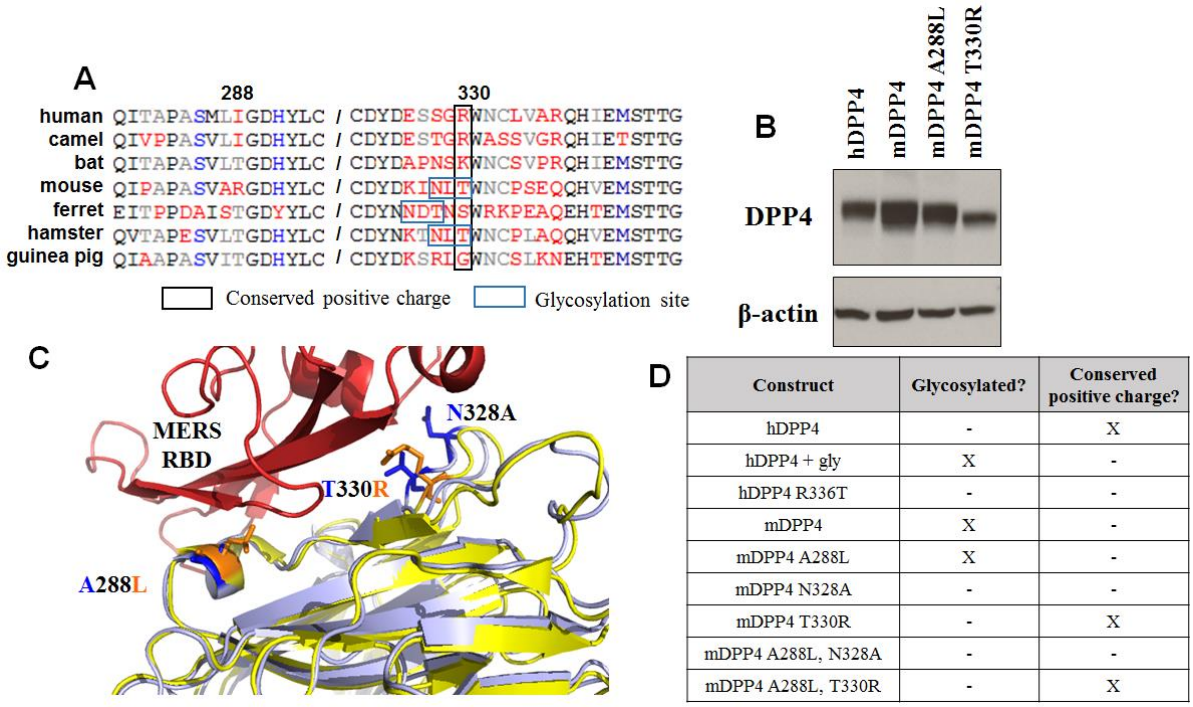
1.3 Identify key biochemical determinants that prevent mouse DPP4 from acting as a valid receptor (modified from Peck *et al.* 2015b)

The functional receptor for MERS-CoV was recently identified as DPP4 (Raj *et al.* 2013). Interestingly, while MERS-CoV can utilize human, bat, and camel DPP4, traditional

small animal models are nonpermissive, including mice (Cockrell *et al.* 2014, Coleman *et al.* 2014), ferrets (Raj *et al.* 2014), and hamsters (de Wit *et al.* 2013a). The relevance of MERS-CoV as an emerging pathogen and the importance of small animal models for studying pathogenesis and for developing vaccines and therapeutics led us to identify the determinants of interactions between the MERS-CoV RBD and mDPP4. Interactions between DPP4 and the MERS-CoV RBD are primarily restricted to blades IV and V of the DPP4 N-terminal β-propeller domain (Lu *et al.* 2013, Wang *et al.* 2013). Recently, we found that two key residues in mDPP4 (A288L and T330R) could permit infection by MERS-CoV when mutated to the hDPP4 amino acids (Cockrell *et al.* 2014). These residues lie within blades IV and V of the β-propeller domain (see Lu *et al.* 2013, Wang *et al.* 2013). The importance of A288L can be understood by recognizing that there is a strong hydrophobic region in the MERS-CoV RBD that engages the equivalent hDPP4 residue (L294) (Wang *et al.* 2013). In fact, all permissive DPP4 orthologs have a leucine residue at this site (i.e. bat, camel, human, marmoset). This interaction, however, is altered in mDPP4, potentially making this hydrophobic region less amenable to interacting with the MERS-CoV RBD.

On blade IV, the T330R substitution in mDPP4 regulates two potentially critical virus-host cell receptor interaction events. *First*, the 330 arginine provides a highly conserved charge that is present in all known permissive hosts, but missing from all known nonpermissive hosts (Figure 1.5A). In hDPP4, the interaction between this residue (R336 relative to hDPP4 numbering) and the MERS-CoV RBD Y499 has been previously noted as a key interaction (Lu *et al.* 2013, Wang *et al.* 2013). The absence of this interaction could be a primary factor behind the lack of permissivity of mDPP4, as well as other nonpermissive DPP4 orthologs. *Second*, the T330R mutation knocks out an NXT glycosylation motif in mDPP4 (Figure 1.5A). Western Blot

analysis is consistent with the loss of glycosylation at this site, as evidenced by a ~2.5 kDa downward shift in the mDPP4 T330R mutant (Figure 1.5B). Considering these two potentially important effects, we hypothesized that either the introduction of the conserved charge or the removal of glycosylation was crucial for regulating mDPP4 permissivity to MERS-CoV infection.



Construct	Glycosylated?	Conserved positive charge?
hDPP4	-	X
hDPP4 + gly	X	-
hDPP4 R336T	-	-
mDPP4	X	-
mDPP4 A288L	X	-
mDPP4 N328A	-	-
mDPP4 T330R	-	X
mDPP4 A288L, N328A	-	-
mDPP4 A288L, T330R	-	X

Figure 1.5: Is charge or glycosylation important for mediating mouse DPP4 permissivity?

(A) MEGA6 (Tamura *et al.* 2013) protein sequence alignment of DPP4 for various permissive (human, camel, bat) and nonpermissive (mouse, ferret, hamster, guinea pig) species, visualized in GeneDoc. Residue numbers are relative to mDPP4. The mutation T330R in mDPP4 introduces a conserved positive charge for permissive hosts, but also knocks out a glycosylation site. NCBI accession numbers: human, NP_001926.2; camel, AIG55259; bat, AGF80256.1; mouse, NP_034204.1; ferret, ABC72084.1; hamster, AIG55262.1; guinea pig, XP_003478612.2. (B) The downward shift in the mDPP4 T330R band is consistent with the removal of glycosylation. Western blot protocol follows Cockrell *et al*. 2014. (C) Structure of hDPP4 (yellow) complexed with the MERS-RBD (red) (PDB code 4L72) visualized using PyMOL. mDPP4 (blue), threaded using I-TASSER (Zhang 2008), is overlaid to show the key mutations: A288L, T330R, N328A. Blue indicates wildtype mDPP4 residues while orange indicates the human amino acid identity. (D) DPP4 constructs used and whether they are glycosylated at the 328 residue or whether the conserved positive arginine is present at the 330 residue (numbered relative to mDPP4). From Peck *et al.* 2015b.

To test the impact of glycosylation versus charge on the ability of mDPP4 to support infection by MERS-CoV, we generated a panel of DPP4 mutants (Figures 1.5C and 1.5D) contained within the 945ΔRRE expression vector, a lentiviral vector derived from pTK945. DPP4 constructs were expressed in HEK 293T cells that lack detectable expression of endogenous hDPP4 (Zhao *et al.* 2013). At ~18 hours post-transfection with 3 μg of the DPP4 expression plasmid, cells were infected with rMERS-CoV-RFP which encodes tomato red fluorescent protein in place of ORF5 (Scobey *et al.* 2013). Cells were imaged ~24 hours post-infection to assess the number of positive cells as a readout for MERS-CoV infection.

A set of hDPP4 mutants were generated and assayed for permissivity to MERS-CoV infection in order to first assess the importance of glycosylation versus charge in the human context. We generated two mutants: one that included a glycosylation site and one that removed the charge. First, we swapped the three residues of the NLT mDPP4 putative glycosylation site with residues 334 to 336 of hDPP4 (hDPP4 + gly). This addition shows a severe reduction in infection (Figures 1.6A and 1.6B), with an upward shift in the Western blot band consistent with successful introduction of the glycosylation site (Figure 1.6C). However, this mutation impacts both the glycosylation site and the charged 336 residue (aligning to residue 330 in mDPP4, Figure 1.5A). Therefore, our second mutant introduces the R336T mutation by itself, which removes the positive charge without introducing glycosylation. While we do observe a decrease in infection, it is not comparable in magnitude to the decrease seen when glycosylation is included (Figures 1.6A and 1.6B), suggesting that the presence of a positively charged residue at position 336 is not essential for hDPP4-mediated MERS-CoV infection. Additionally, the presence of glycosylation does not impact the ability of hDPP4 + gly to be expressed on the cell

surface (Figure 1.8). These results show that glycosylation can act to inhibit infection by MERS-CoV and that the positive charge is not a crucial interaction in the context of hDPP4.

In order to directly assess the relative contribution of charge versus glycosylation in the context of mDPP4, we evaluated whether the presence of glycosylation or charge at the 330 site regulates mDPP4 receptor activity. For these studies, mutations were evaluated singly and in the presence of the secondary mutation (A288L), which is essential for high levels of MERS-CoV receptor activity. Importantly, introduction of the charged residue at 330 simultaneously destroys the glycosylation site, preventing us from testing whether the presence of the charged residue at 330 can enhance mDPP4 receptor activity in the presence of a glycosylation site. However, we can remove the glycosylation site without introducing a charged residue with the mutation

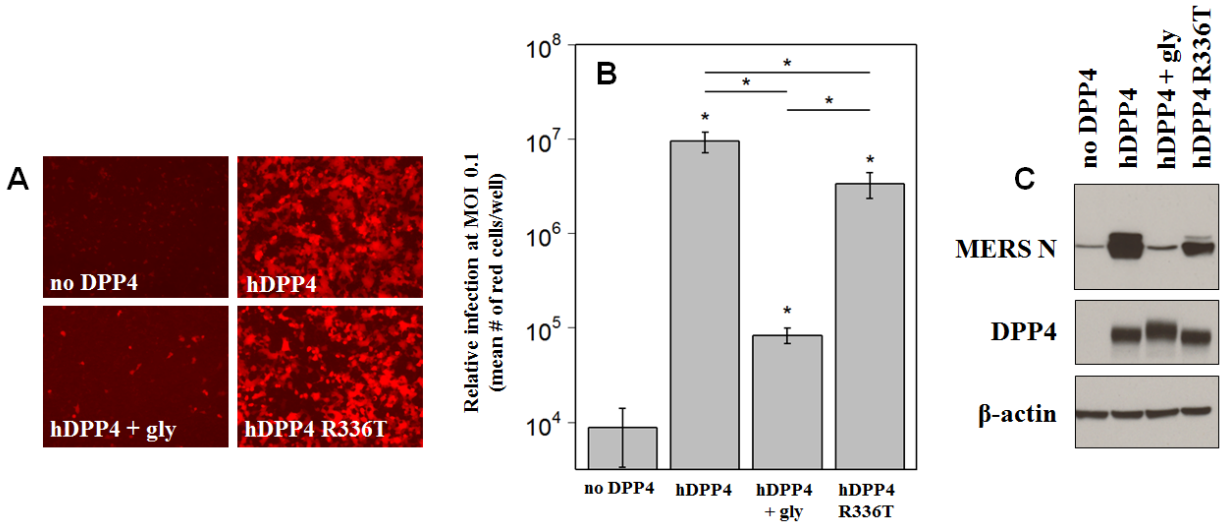


Figure 1.6: Glycosylation can act to dramatically reduce infection by MERS-CoV

(A) HEK 293T cells were transfected with each DPP4 construct and infected with rMERS-CoV-RFP at an MOI of 1 at ~18 h post-transfection. At ~24 hpi, cells were imaged. (B) Cells were transfected as in (A) and infected with rMERS-CoV-RFP at the following MOIs: hDPP4 and hDPP4 R336T, 0.001; no DPP4 and hDPP4 + gly, 0.1. At 24 hpi, cells were counted based on red fluorescence and values were normalized to an MOI of 0.1. Values represent 3 replicates. All mutants have levels that are statistically greater than no DPP4 and all other pairwise comparisons are also statistically significant (indicated by *, $p < 0.05$, Student's t-test). (C) Western blot analysis for MERS nucleocapsid (N) protein, DPP4, and β-actin as a loading control. Western blot protocol follows Cockrell *et al.* 2014. From Peck *et al.* 2015b.

N328A, which disrupts the N of the NXT motif (Figures 1.5A and 1.5D). When we assessed the N328A mutant in the context of the A288L background we observed high levels of infection (Figure 1.7A) that are not statistically different from mDPP4 A288L, T330R (Figure 1.7B). Both glycosylation knockout mutants have levels that are statistically greater than mDPP4 but statistically less than hDPP4 (Figure 1.7B). All mutants containing the T330R or N328A mutation show a ~2.5 kDa downward shift in the Western Blot, consistent with the loss of glycosylation (Figure 1.7C). Importantly, surface staining for mDPP4 and hDPP4 signifies that

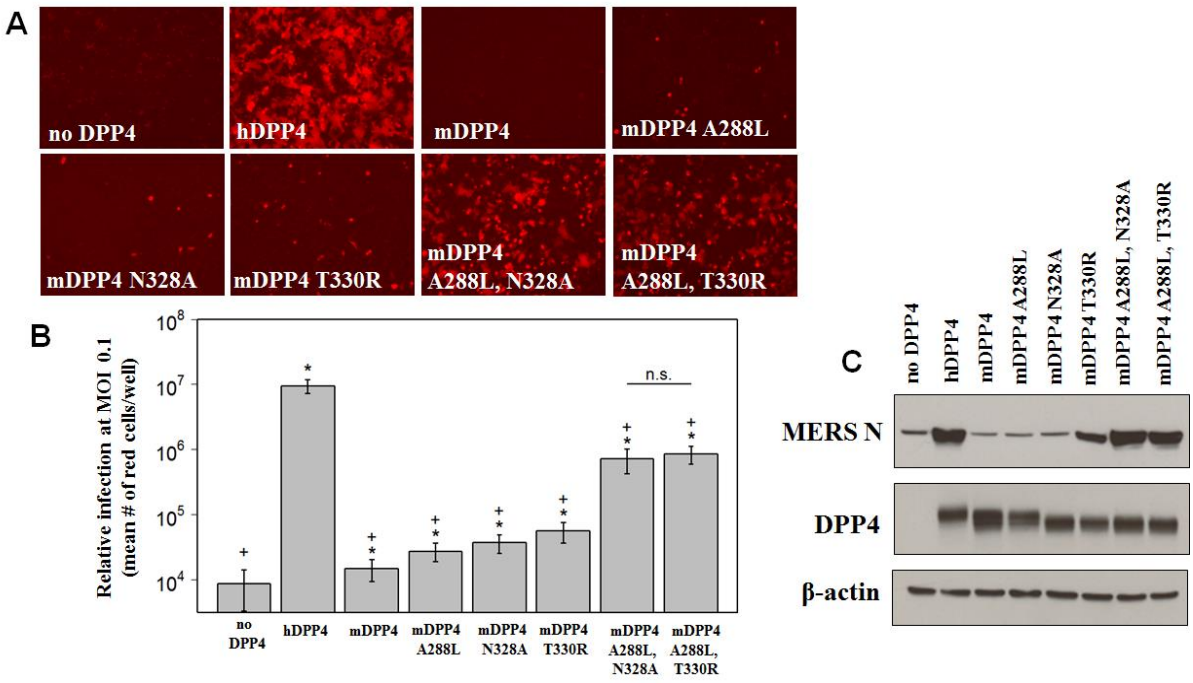


Figure 1.7: Glycosylation is more important than charge in mediating MERS-CoV infection

(A) Cells were transfected and infected following the protocol detailed in Figure 1.6A. Neither mDPP4 N328A, nor mDPP4 T330R can confer permissivity to MERS-CoV, however both result in strong levels of infection when coupled with A288L. (B) Red cell counts were calculated as in Figure 1.6B with the following MOIs: hDPP4, 0.001; mDPP4, mDPP4 288, mDPP4 328, mDPP4 330, no DPP4, 0.1; mDPP4 288, 328 and mDPP4 288, 330, 0.01. All DPP4 constructs are significantly greater than no DPP4 and mDPP4 (*, $p < 0.05$, Student's t-test) and significantly less than hDPP4 (+, $p < 0.05$, Student's t-test); however mDPP4 A288L, N328A and mDPP4 A288L, T330R are not statistically different from each other (n.s., $p < 0.05$, Student's t-test). (C) Western blot analysis for MERS nucleocapsid (N) protein, DPP4, and β-actin as a loading control. Western blot protocol follows Cockrell *et al.* 2014. From Peck *et al.* 2015b.

all derivatives of the DPP4 receptors are expressed at the cell surface and available to interact with the MERS-CoV RBD (Figure 1.8). Together, these results indicate that removal of the glycosylation site, rather than addition of the charged residue at position 330, is responsible for regulating the ability of MERS-CoV to utilize mDPP4 as a functional receptor. The secondary mutation, A288L, also plays an important role in MERS-CoV permissivity because high levels of infection are only seen when the glycosylation mutants are combined with the A288L substitution (Figures 1.7A and 1.7B). Together, this suggests that while glycosylation is an important barrier, its removal is not sufficient to permit infection in the absence of the A to L modification at position 288. Overall, glycosylation can act to block MERS-CoV infection, yet other determinants are also important for mediating permissivity. Further research will determine the relative contributions of these determinants and whether they act as a broader signature of permissivity among DPP4 orthologs.

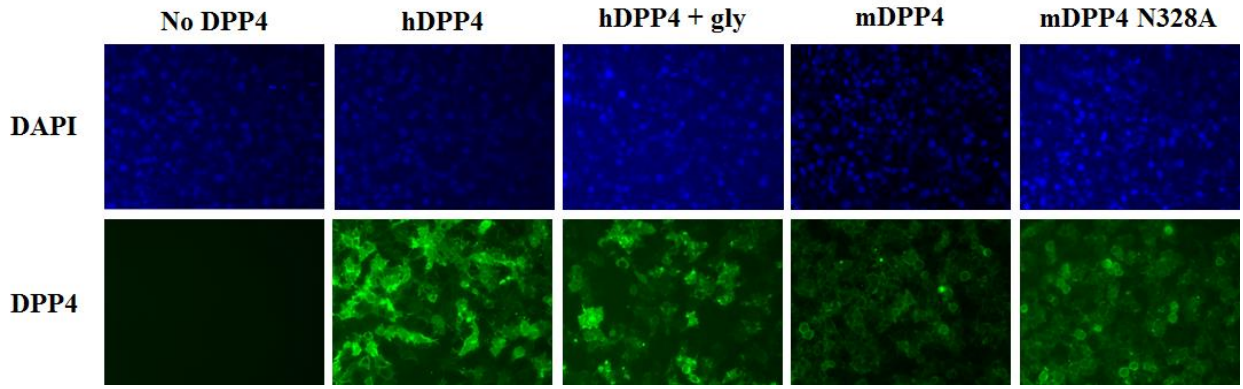


Figure 1.8: DPP4 construct expression in HEK 293T cells

DPP4 and mutant variants are expressed on the surface of cells, visible by immunofluorescence (IFA). Cells were transfected as described in Figure 1.6A, fixed, and probed with primary goat-anti-DPP4 polyclonal antibody (R&D Systems) at 1:50 and secondary donkey-anti-goat Alexa Fluor 488 (Life Technologies) at 1:500. Cells were imaged at 40X for DAPI (30 ms exposure) and DPP4 (160 ms exposure). From Peck *et al.* 2015b.

1.4 Removing glycosylation from mDPP4

The above experiments analyzed the removal of glycosylation from mDPP4 using PCR mutagenesis. Knockout of the glycosylation site was evaluated by Western blot, with a downward shift of the protein band by ~2.5 kDa indicating successful removal of the glycan. However, additional resources are available for removing (or preventing) the glycosylation of protein molecules. Tunicamycin acts by inhibiting the formation of N-linked glycans (Esko and Bertozzi 2009). PNGase F, on the other hand, acts to cleave the glycan from the amino terminus after formation (Mulloy *et al.* 2009). I tested both of these methods in order to determine whether the previously discovered phenotype, that removing glycosylation from mDPP4 can confer permissivity to MERS-CoV, could be recapitulated with various methods of glycan removal. One caveat to these experiments is that DPP4 is known to have many putative glycosylation sites. hDPP4, for example, has nine N-linked glycosylation sites throughout the molecule, eight of which are clustered to blades II through V of the β-propeller domain (Rasmussen *et al.* 2003). The presence of multiple glycosylation sites within DPP4 may cause the use of tunicamycin or PNGase F to be too disruptive for proper folding and function of the molecule.

HEK 293T cells were transfected with 3 µg of hDPP4 or mDPP4 using previous protocols. At ~2 hours post-transfection, 1 µg/mL of tunicamycin was added to the cells and incubated at 37 ºC for ~24 hours. Different amounts of tunicamycin are efficacious for various cell types (Esko and Bertozzi 2009); 1 ug/mL was chosen as a reasonable concentration after testing a range of 0 to 5 µg/mL. One set of cells was used for protein analysis; protein lysates were harvested using the previously detailed 1X RIPA buffer protocol and analyzed by Western blot (Figure 1.9A). A second set of cells was used for analysis of infection by rMERS-CoV-RFP (Figure 1.9B). The Western blot analysis shows a severe downward shift in the DPP4 bands after

tunicamycin treatment, suggesting successful inhibition of glycan formation (Figure 1.9A). However, the infection results show that mDPP4 does not gain permissivity to rMERS-CoV-RFP when glycosylation is inhibited and hDPP4 loses its ability to confer infection (Figure 1.9B). The ablation of permissivity in hDPP4 suggests that other glycosylation sites are critical to DPP4 folding and/or function, specifically for MERS-CoV RBD interactions. The reduction in molecule size of both mDPP4 and hDPP4 suggests that they have a similar number of glycosylation sites and that the additional glycosylation sites in mDPP4 may also be crucial for proper functionality. These results rule out my ability to use tunicamycin in future glycosylation analysis experiments for DPP4.

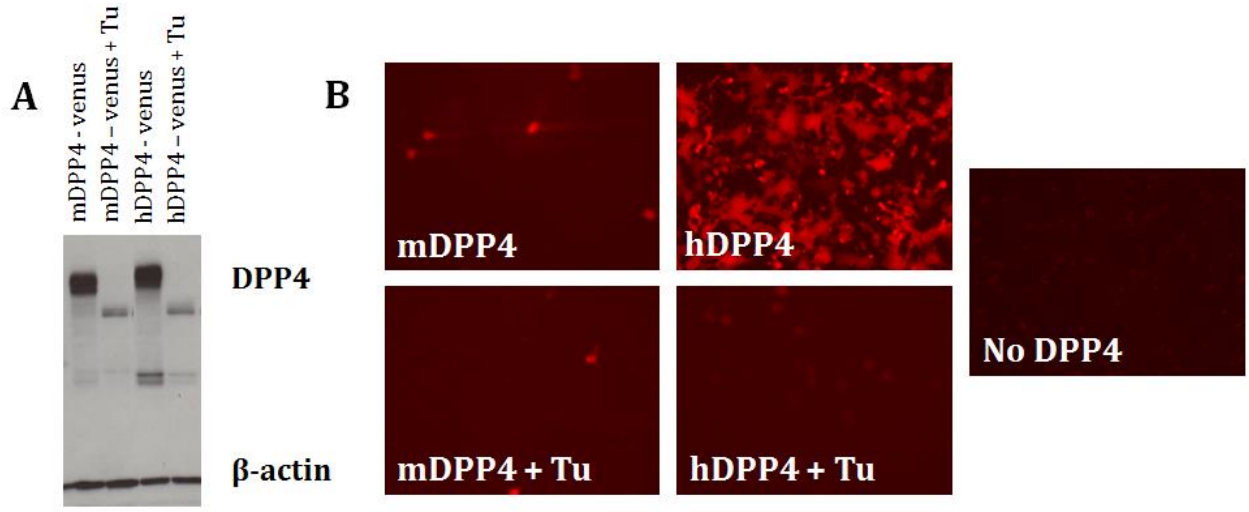


Figure 1.9: Tunicamycin treatment of DPP4 molecules

(A) Western blot analysis of HEK 293T cells transfected with either mDPP4 or hDPP4 (fused to venus) and treated with 1 ug/mL tunicamycin (Tu) ~2 hours post-transfection. Protein lysates were harvested ~24 hours post-treatment with Tu. Shifted bands indicate the successful inhibition of glycosylation. Western blot protocol follows Figure 1.1 (Cockrell *et al.* 2014). (B) HEK 293T cells transfected with mDPP4 or hDPP4 and treated with Tu ~2 hours post-transfection. Cells were infected with rMERS-CoV-RFP ~24 hours post-treatment with Tu and imaged ~24 hpi.

To test the efficacy of PNGase F, HEK 293T cells were transfected with 3 µg of hDPP4 or mDPP4 following previously detailed protocols. At ~24 hours post-transfection, cells were washed twice with 1X PBS. PNGase F (NEB) was diluted in 1X PBS to final concentrations of 0, 2500, and 5000 units/mL. Cells were incubated at 37 ºC for 1 hour and then rinsed twice with 1X PBS. Protein lysates were then harvested using the previously describe 1X RIPA buffer protocol and analyzed by Western blot (Figure 1.10). Unfortunately, even at a concentration using 5000 units/mL, no downward shift in protein size is visible (Figure 1.10). This indicates that 1) the concentration of PNGase F may need to be higher, although this puts it outside of the range found to be effective for other proteins, or 2) the surface glycans on DPP4 do not meet the requirements needed for PNGase F to be effective. Namely, the enzyme requires at least one amino acid at both the amino and carboxyl terminus of the asparagine (N) in order for proper cleavage (Mulloy *et al* 2009). Additionally, the traditional PNGase F protocol treats proteins

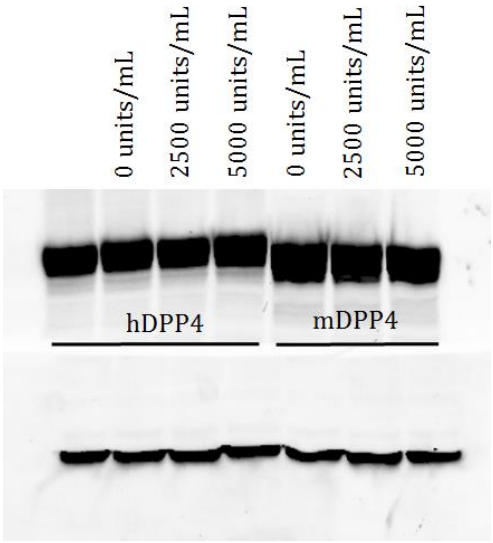


Figure 1.10: PNGase F treatment of DPP4 molecules

Western blot analysis of HEK 293T cells transfected with either hDPP4 or mDPP4. At ~24 hours post-transfection, cells were washed and treated with 0 to 2500 units/mL of PNGase F at 37 ºC for 1 hour. Protein lysates were harvested ~1 hours post-treatment with PNGase F. Western blot protocol follows Figure 1.1 (Cockrell et al. 2014) with the top bands indicating DPP4 and the bottom bands indicating β-actin.

after they have been harvested from the cells. However, because I was interested in treated monolayers and testing for infection, I treated the cells directly. This protocol adjustment could also explain the failure of PNGase F to remove glycans. Because of the lack of evidence for glycan removal (protein size shifting via Western blot) after treatment (Figure 1.10), PNGase F was not used in subsequent experiments on DPP4 glycosylation and MERS-CoV permissivity.

The importance of glycosylation in the interactions between coronaviruses and host-cell receptors has previously been recognized. For example, the introduction of a glycosylation site into human APN prevents HCoV-229E from utilizing it as a receptor (Wentworth and Holmes 2001). For MERS-CoV, it is possible that glycosylation can act as a broader determinant of DPP4-mediated host range, since other nonpermissive hosts (i.e. ferrets, hamsters) also have a non-conserved glycosylation site in the region of DPP4 that interacts with the MERS-RBD (Figure 1.5A). In the context of a small animal model, the presence of the glycosylation site at 330 may sterically hinder multiple interacting residues between the MERS-CoV RBD and mDPP4, complicating the generation of a mouse-adapted strain. Therefore, it may be necessary to partially, or fully, humanize mDPP4 to achieve *in vivo* MERS-CoV replication. Additionally, the finding that changes in both blades of mDPP4 are crucial for mediating permissivity to MERS-CoV (Figure 1.7A) has two major implications. *First*, it may help inform future studies in other nonpermissive hosts. *Second,* it suggests that circulating MERS-like coronaviruses cannot expand their host range into mice and possibly other rodent species with just one change. Rather, extensive remodeling of the MERS-CoV RBD is likely required for it to successfully utilize nonpermissive DPP4 orthologs as receptors, especially if glycosylation acts to block infection in these alternate species. Presumably, the modifications that would allow the MERS-CoV RBD to utilize mDPP4 and other orthologs would likely attenuate or even ablate its ability to utilize

hDPP4. Overall, by understanding the biochemical determinants that mediate MERS-CoV utilization of DPP4 orthologs, we can begin to characterize the selective pressures leading up to host-range expansion events, with the broader goal of being able to predict future emergences.

CHAPTER 2: BIOCHEMICAL DETERMINANTS OF DPP4 ORTHOLOG PERMISSIVITY TO MERS-CORONAVIRUS

2.1 Introduction

Among viruses for which the host receptor has been identified, there is an association between host range and phylogenetic conservation of that receptor (Woolhouse 2002). This result is consistent with previous species-level studies that have shown that the more phylogenetically related two species are, the more likely it is that a virus will be able to jump between them (deFilippis and Villarreal 2000). These observations confirm that the host receptor is a primary determinant of host range expansion and also that receptor conservation can potentially act as a screen to identify viruses that are likely to jump into humans.

The link between DPP4 sequence conservation across species and permissivity to MERS-CoV infection, however, is not obvious. While only a small subset of species have been tested *in vitro* (Muller *et al.* 2012, de Wit *et al.* 2013a, de Wit *et al.* 2013b, Ohnuma *et al.* 2013, Barlan *et al.* 2014, Cockrell *et al.* 2014, Coleman *et al.* 2014, Eckerle *et al.* 2014, Falzarano *et al.* 2014, Raj *et al.* 2014, van Dormalen *et al.* 2014), there is no clear phylogenetic clustering of permissive and nonpermissive hosts when analyzing the DPP4 gene tree (Cui *et al.* 2013, Peck *et al.* 2015c). The lack of a clear pattern of permissivity among closely-related DPP4 genes suggests that other aspects of DPP4 may be more important than simply linear amino acid sequence, such as structural similarity or conservation of post-translational modifications (e.g. glycosylation). Alternatively, or in addition, receptor-independent host restriction mechanisms may operate in some species.

To understand the interactions between the MERS-CoV RBD and DPP4 and to better characterize the host range expansion events of MERS-CoV, both past and potential, I examined the permissivity of various DPP4 orthologs. After establishing the characteristics and permissivity of these orthologs, I determined whether there was a detectable broad signature of permissivity. Primarily, do all nonpermissive species block MERS-CoV infection using the same mechanisms? Because phylogenetic relatedness did not yield a signal (Cui *et al.* 2013, Peck *et al.* 2015c), and analyzing the sequences at previously identified key RBD residues (Cockrell *et al.* 2014) did not yield any obvious hypotheses (Table 1), I turned to the results found in Chapter 1. Based on the discovery that glycosylation is an important barrier to MERS-CoV infection, I noticed that some nonpermissive species also have glycosylation sites near the site observed in mDPP4 (Figure 1.5A). Thus, I tested whether glycosylation acts to block MERS-CoV infection in other nonpermissive species before investigating alternate determinants.

2.2 Determine the role of glycosylation in additional DPP4 orthologs

First, I validated the panel of DPP4 constructs I had generated by confirming the permissive or nonpermissive nature of various species orthologs. While hDPP4 and mDPP4 were

	282	288	289	330	340
Human	T	L	I	R	I
Bat	T	L	T	K	I
Camel	V	L	I	R	I
Mouse	P	A	R	T	V
Guinea Pig	A	I	T	G	T
Ferret	T	S	T	S	T
Hamster	T	L	T	T	V

Table 1: Amino acid identities at five DPP4 ortholog residues important for MERS-CoV binding

Amino acid identities at the five residues identified to be important for mediating MERS-CoV permissivity in the context of mDPP4 (Cockrell *et al.* 2014). Residue numbering is relative to mDPP4. Permissive and nonpermissive species are indicated by black and red text, respectively.

obtained as a complete plasmid, I generated the remaining DPP4 orthologs using Gibson assembly (Gibson *et al.* 2009). This method employs an *in vitro* recombination method to allow for the rapid and financially tractable assembly of multiple DPP4 constructs.

Following previous methods, I transfected 3 µg of each DPP4 ortholog into HEK 293T cells. At ~24 hours post-transfection, I infected the cells with rMERS-CoV-RFP at an MOI of 1. At ~24 hours post-infection, I visualized the cells to determine whether they were permissive to MERS-CoV. As expected, human, bat, and camel DPP4 molecules are permissive to MERS-CoV infection while mouse, ferret, guinea pig, and hamster are not (Figure 2.1). Although previous work (Cockrell *et al.* 2014) generated data that was used to develop a mouse model for MERS-CoV, understanding why ferret, hamster, and guinea pig cannot be infected could help create a secondary option for studying MERS-CoV pathogenesis.

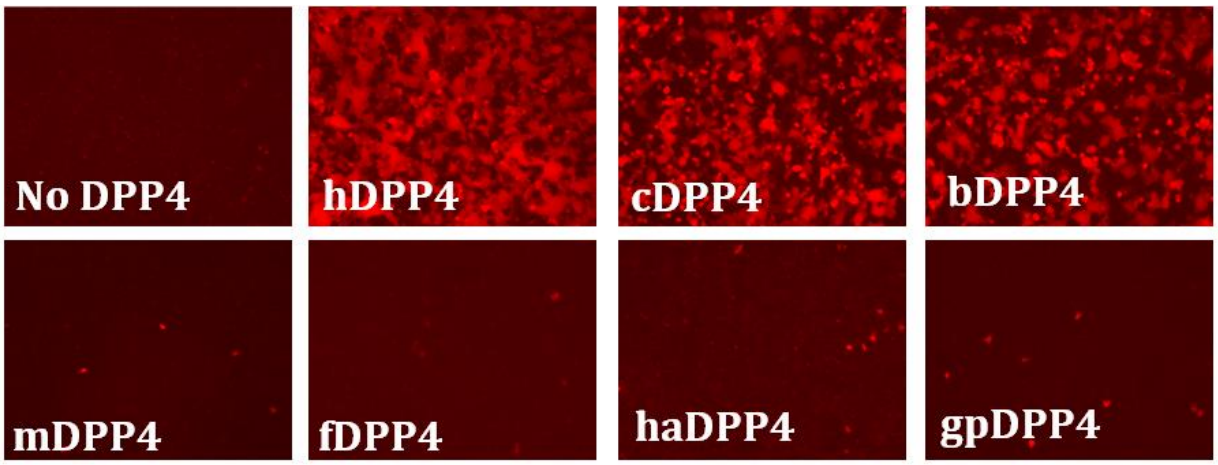


Figure 2.1: Permissivity of DPP4 orthologs to MERS-CoV

Seven DPP4 orthologs were tested for their ability to support infection by rMERS-CoV-RFP. DPP4 constructs were transfected into HEK 293T cells and infected at an MOI of 1 ~24 hours post-transfection. Cells were imaged for fluorescence ~24 hpi. hDPP4, human DPP4; cDPP4, camel DPP4; bDPP4, bat DPP4; mDPP4, mouse DPP4; fDPP4, ferret DPP4; haDPP4, hamster DPP4; gpDPP4, guinea pig DPP4.

Aligning the sequences of the permissive and nonpermissive species in a region crucial to MERS-CoV binding reveals the presence of a glycosylation site in hamster DPP4 (haDPP4), ferret DPP4 (fDPP4), and guinea pig DPP4 (gpDPP4) molecules (Figure 2.2A). The glycosylation site in hamster is identical to that in mouse, suggesting that it may interact with the MERS-CoV RBD in a similar way. The glycosylation site in fDPP4 is slightly upstream, and the glycosylation site in gpDPP4 is slightly downstream (Figure 2.2A). However, the latter is shared by bDPP4 (sequence from species *Pipistrellus pipistrelle*), suggesting that it might not play as important of a role in blocking infection unless there is a gpDPP4-specific structural effect.

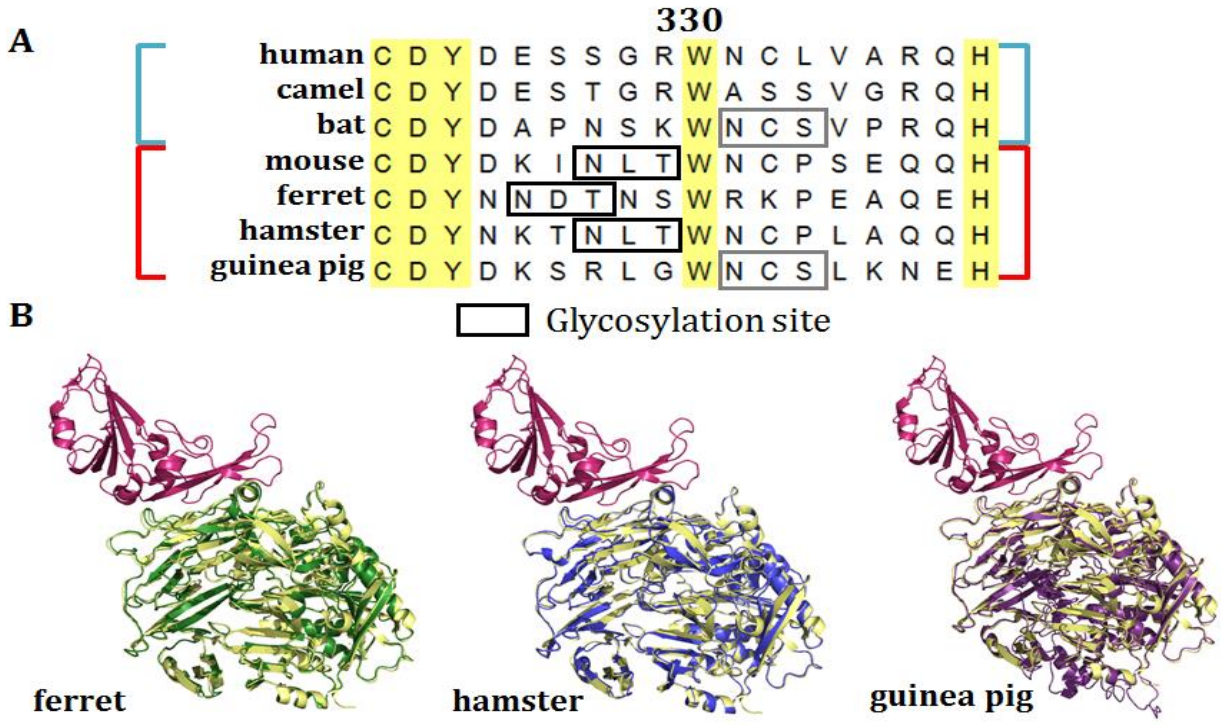


Figure 2.2: Sequence and structural comparison of nonpermissive DPP4 orthologs

(A) Sequence alignment of permissive (human, camel, bat; blue) and nonpermissive (mouse, ferret, hamster, guinea pig; red) DPP4 amino acid sequences. Residue 330 is numbered relative to mDPP4. Boxes represent glycosylation sites that are either unique to nonpermissive species (black) or shared with a permissive species (gray). (B) Structural comparison of threaded molecules (Zhang 2008) fDPP4 (green), haDPP4 (blue), and gpDPP4 (purple) overlaid on hDPP4 (yellow) complexed with the MERS-CoV RBD (red) (PDB code 4L72).

The crystal structures for fDPP4, haDPP4, and gpDPP4 have not yet been solved. Utilizing the same threading technique as in Chapter 1, I generated predicted structures for each of these proteins using I-TASSER (Zhang 2008). Overlaying the structures with hDPP4 shows that they are predicted to have highly similar structural backbones (Figure 2.2B). Again, maintaining awareness that the structures are predictions, the RMS scores obtained for fDPP4, haDPP4 and gpDPP4 overlaid with hDPP4 are 0.616, 0.378, and 0.604, respectively. This can be compared to amino acid sequence identity values of 88%, 85%, and 87%, respectively. From these computational results, haDPP4 is predicted to be the most structurally similar to hDPP4 and may be the best option for an alternate transgenic animal model.

Based on my previous discovery that glycosylation plays an important role in blocking MERS-CoV, I knocked out the glycosylation sites in each of these species using PCR mutagenesis. Each of these knockout mutations changed the N of the glycosylation NXT (or NXS) motif to an alanine and are designated as "-gly" in subsequent figures. Following previous methods, 3 µg of each DPP4 construct was transfected into HEK 293T cells. At ~24 hours post-transfection, cells were infected with rMERS-CoV-RFP at an MOI of 1. At ~24 hours post-infection, cells were imaged and red fluorescence analyzed as a readout of infection. My results show that removing glycosylation from these three DPP4 orthologs did not result in an increase in infection (Figure 2.3A). The amount of infection that was supported by all three glycosylation knockout molecules was equivalent to their respective wildtype molecules. The glycosylation, and subsequent removal, of each of these sites was confirmed by Western blot analysis (Figure 2.3B), following previous protocols (Peck *et al.* 2015b).

Upon further inspection of the haDPP4 sequence, I noticed a putative glycosylation site upstream of the previously identified one (Figure 2.2A). The motif NKT (starting at residue 329)

is directly upstream from the previously studied NLT (starting at residue 332) glycosylation site. As a test to determine whether this site affected permissivity, particularly because the equivalent 288 residue found to be important in mice shares the human amino acid identity in haDPP4, I generated constructs that mutated the N of the NKT motif using PCR mutagenesis, both singly and in combination with the NLT glycosylation knockout. Transfecting the haDPP4 glycosylation variants into HEK 293Ts and infecting with rMERS-CoV-RFP revealed that mutating this additional glycosylation site had no impact on infection (Figure 2.4). Furthermore, mutating both of the glycosylation sites together did not have any impact on infection levels

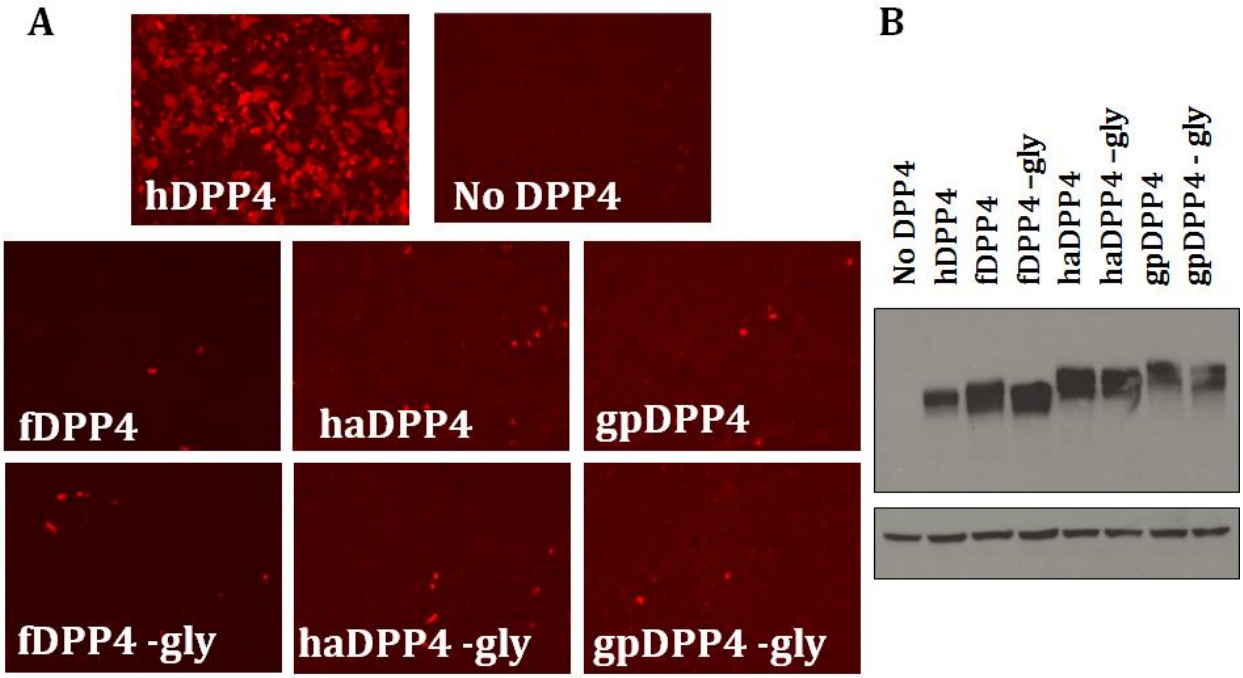


Figure 2.3: DPP4 ortholog glycosylation knockout mutants

(A) Neither wildtype nor glycosylation knockout DPP4 molecules for ferret (fDPP4), hamster (haDPP4), nor guinea pig (gpDPP4) support infection by MERS-CoV. (B) Successful removal of glycosylation is supported by a ~2.5 kDa downward shift in protein analysis via Western blot. Top blot represents DPP4 and the bottom blot represents β-actin as a control. Western blot protocol follows Peck *et al.* 2015b.

(Figure 2.4). It is possible that the NKT motif is not actually glycosylated; my next step is to validate glycosylation by Western blot before drawing further conclusions.

Whereas fDPP4 and haDPP4 did not share their glycosylation sites with permissive molecules, gpDPP4 shared its downstream glycosylation site with bDPP4. Due to this similarity, I knocked out the glycosylation site in bDPP4 to determine 1) whether it was actually glycosylated and 2) its impact on the permissivity of bDPP4. Using PCR mutagenesis to mutate the N of the NXS motif to an alanine, I followed previous protocols to express the bDPP4 mutant in HEK 293T cells and visualized the cells ~24 hours post-infection. My results show that removing glycosylation from bDPP4 has no detectable difference on its ability to support MERS-CoV infection (Figure 2.5). My next step is to perform a Western blot in order to

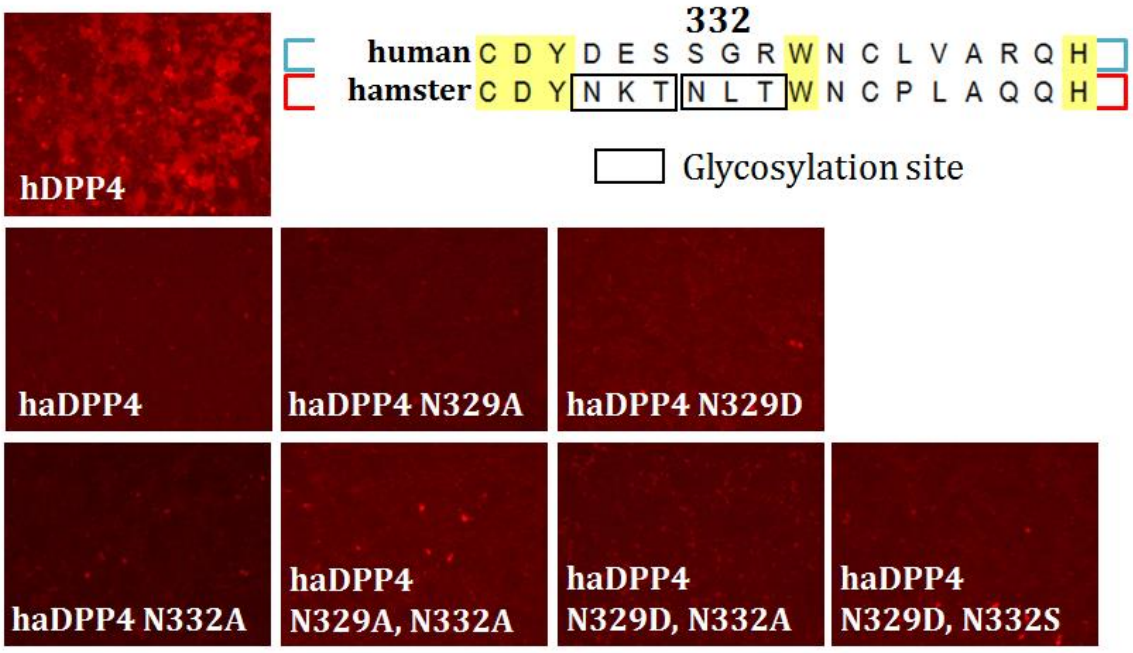


Figure 2.4: Glycosylation knockout panel in haDPP4

haDPP4 has two glycosylation sites near the site identified as important for mediating MERS-CoV permissivity in mDPP4. Knocking out the NKT (starting at residue 329) or the NLT (starting at residue 332) motif, or both in combination, has no impact on MERS-CoV permissivity. Mutating the residues to either alanines or the equivalent amino acid identity in hDPP4 (D or S, respectively) also shows no impact.

determine whether the bDPP4 molecule is actually glycosylated or whether the NXS represents a sequence motif that is not, in actuality, glycosylated in the final molecule.

Overall, glycosylation is not the only determinant that mediates MERS-CoV infection in nonpermissive DPP4 orthologs. This is particularly surprising in the case of haDPP4; this protein not only has a glycosylation site in the same location as mDPP4 (Figure 2.2A), but the secondary residue that was identified to be important for mDPP4 (residue 288) in haDPP4 is the same amino acid identity as hDPP4 (Table 1). Based on this observation, while removing glycosylation may be a substantial component of permitting infection in these DPP4 orthologs, other determinants clearly play an important role.

The importance of glycosylation in blocking MERS-CoV infection may vary between species. To gain a better intuition about the extent of glycosylation among DPP4 orthologs, I

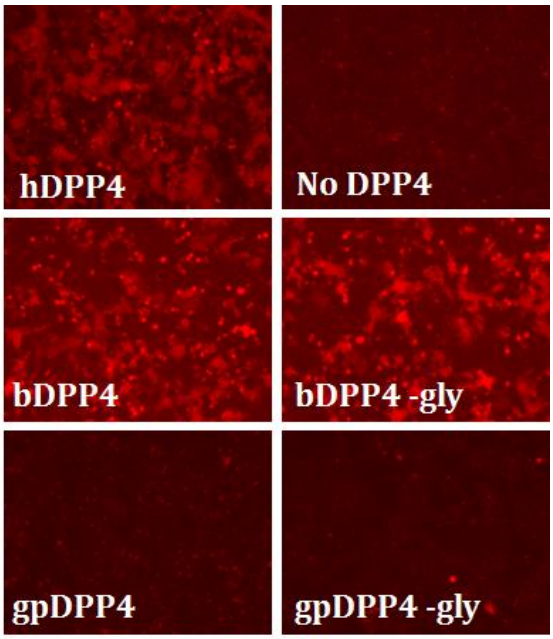


Figure 2.5 Guinea pig and bat DPP4 glycosylation knockout mutants

Bat and guinea pig DPP4 share the same glycosylation site downstream of the site identified to be important in mDPP4 (Figure 2.2A). Removing the glycosylation site from bDPP4 shows no decrease in infection while removing glycosylation from gpDPP4 shows no increase in infection.

constructed a phylogenetic tree of a subset of full-length DPP4 protein sequences (Figure 2.6). Retrieving the sequences from GenBank, I used MAFFT to align the amino acid sequences (Katoh *et al.* 2002). The phylogenetic tree was generated using maximum likelihood with the PhyML package (Guindon *et al.* 2010) and visualized using EvolView (Zhang *et al.* 2012). In the tree, shaded colors indicate the general organism group that each species belongs to - blue: reptiles and amphibians; green: avian species; orange: other mammals; red: Chiroptera (bats); purple: ungulates; gray: rodents; pink: primates (Figure 2.6). The DPP4 protein tree is slightly discordant with the species tree, notably with the horse and African savanna elephant DPP4 sequences not clustering with other ungulate (purple) orthologs.

Plotted adjacent to the phylogenetic tree are glycosylation sites that are either upstream (column 1), at the same site (column 2), or downstream (column 3) of the glycosylation site that is present in mDPP4 (residues 328-330). Based on this tree, only eight other species have putative glycosylation sites at the same location as in mDPP4; the majority of these are present in the rodent (gray) group, however two are present within the Chiroptera group (black flying fox and large flying fox) (Figure 2.6). Permissivity data is indicated in the far right column, with green squares indicating permissive species and red squares indicating nonpermissive species, based on either *in vitro* or *in vivo* data (de Wit *et al.* 2013a, Barlan *et al.* 2014, Eckerle *et al.* 2014, Cockrell *et al.* 2014, Coleman *et al.* 2014, Raj *et al.* 2014, van Doremalen *et al.* 2014). The small number of data points for species makes it difficult to map potential shifts in permissivity. However, a few interesting observations emerge. First, all non- human primates lack glycosylation sites near residue 330 (equivalent residue 336 in humans) (Figure 2.6). Current data would postulate that all non-human primates are permissive to MERS-CoV infection, but testing this hypothesis *in vitro* would help reveal whether any of these orthologs are

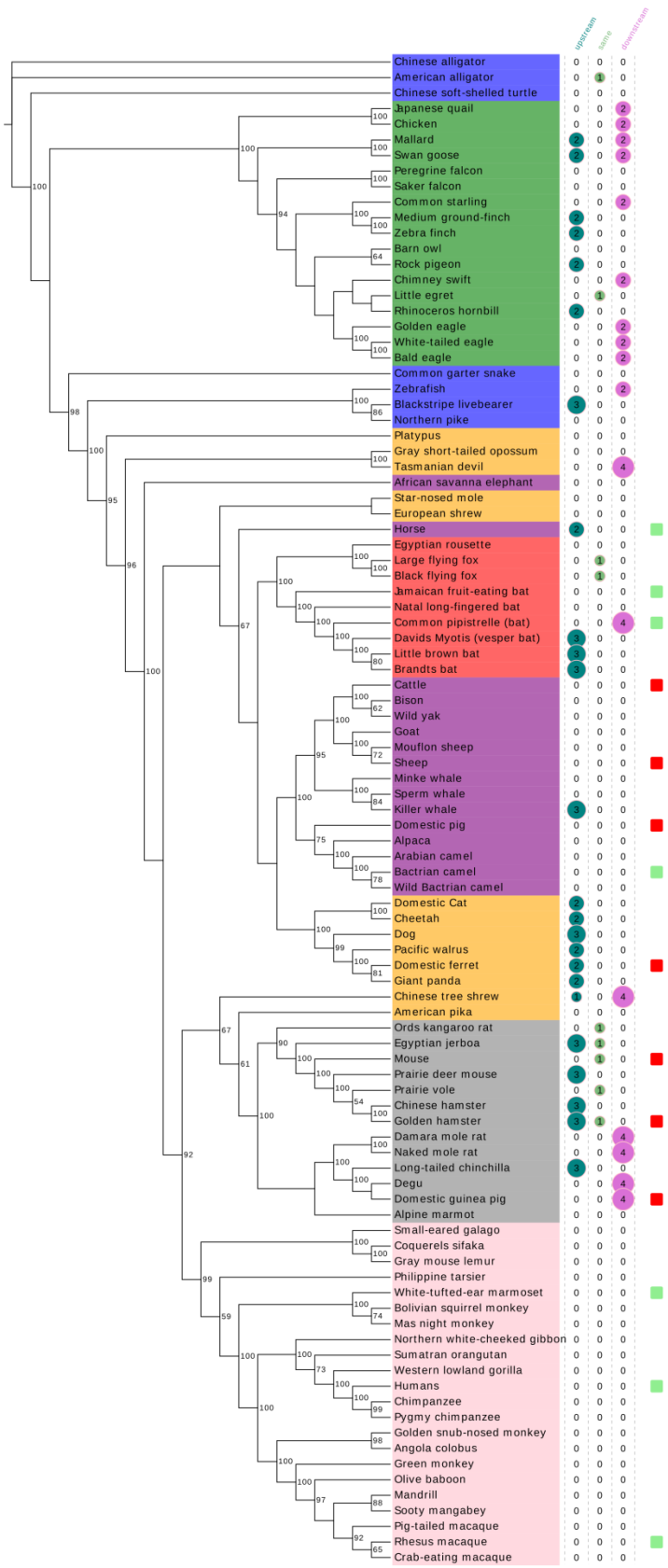


Figure 2.6: DPP4 protein phylogenetic tree

DPP4 phylogenetic tree based on amino acid sequences. Shaded colors indicate the group each species falls in. Blue: reptiles and amphibians; Green: avian species; Orange: other mammals; Red: Chiroptera (bats); Purple: ungulates; Gray: rodents; Pink: primates. Colored circles to the right of the species names indicate whether the sequence has a glycosylation site upstream (first column), at the same site (second column), or downstream (third column) of the NXT glycosylation site in mDPP4 (residues 332-334). Numbers inside the circle designate how many amino acids upstream (or downstream) the N of the NXT or NXS putative glycosylation site is. For the second column, a 1 indicates that there is a glycosylation site at the same location as in mDPP4. Squares in the rightmost column indicate permissive (green) or non-permissive (red) species, as determined from either *in vivo* or *in vitro* studies. See Appendix B for species Latin names and NCBI accession numbers.

nonpermissive. If so, a separate mechanism from glycosylation at our location of interest would be responsible for blocking MERS-CoV infection. Second, whereas other permissive DPP4 orthologs have glyocyslation sites in this region (horse and common pipistrelle DPP4), of note is the lack of a glycosylation site in the nonpermissive domestic pig, sheep, and cattle DPP4 molecules (Figure 2.6). This suggests that glycosylation in this region is not the primary explanation for why these species do not support MERS-CoV infection. Follow-up studies could explore the mechanism of nonpermissivity in this species specifically, and whether or not it lies at the level of the receptor. Third, glycosylation sites in the designated region are prevalent in rodents, other mammals, and avian species (Figure 2.6). The diversity of glycosylation profiles suggests a lack of strong conservation of the site seen in mDPP4. Further research to determine whether these other orthologs are permissive can help elucidate whether this region plays a broader role in the MERS-CoV infection phenotype. In general, while some trends are seen when looking at the glycosylation profile across many species (e.g. upstream glycosylation sites in the other mammal (orange) group), more data on species permissivity will help determine whether DPP4 phylogenetic relationships can help inform receptor-binding dynamics.

2.3 Identify key mutations and determinants of DPP4 ortholog permissivity to MERS-CoV infection

The mDPP4 data suggest that two mutations are necessary to support MERS-CoV infection, one on each blade of the DPP4 molecule (Cockrell *et al.* 2014, Peck *et al.* 2015b). The mechanisms of mDPP4 permissivity included stabilizing a hydrophobic core and removing a glycosylation site (Peck *et al.* 2015b). However, the simplicity of this determinant does not carry over; haDPP4 has the same amino acid as hDPP4 at the residue that stabilizes the MERS-CoV RBD hydrophobic core, and yet knocking out glycosylation does not confer infection (Figure

2.3A). This suggests other determinants are important for conferring permissivity in these other molecules.

Other groups have introduced single point mutations in both haDPP4 and fDPP4 to try to instigate permissivity (Raj *et al.* 2014, van Doremalen *et al.* 2016). For haDPP4, other studies found that a minimum of five amino acid mutations allowed the molecule to support MERS-CoV infection (van Doremalen *et al.* 2016). Focusing on the idea that changes in *both* blades IV and V of DPP4 are required to confer infection in the context of mDPP4, I constructed chimeric fDPP4 and haDPP4 molecules that made substantial changes on both blades. Using overlap PCR, I constructed a chimeric fDPP4 that swaps out 16 residues on blade V and 11 residues on blade IV to the equivalent human amino acid identities, indicated by their starting residues of 278 and 331, respectively. Transfecting 3 µg of the chimeric fDPP4 constructs into HEK 293T cells and infecting with rMERS-CoV-RFP (MOI 1) at ~24 hours post-transfection yielded results that were imaged ~24 hours post-infection. Results show that while the single glycosylation knockout mutant did not show an increase in infection, the combination of 27 amino acid changes on blades IV and V show a dramatic increase in infection (Figure 2.7A). Western blot analysis shows that each DPP4 variant is being highly expressed; however, because of the large number of amino acid changes present in each mutant, a downward shift in the protein band is visible for each construct (fDPP4 (278), fDPP4 –gly, fDPP4 (278)(331)) (Figure 2.7C), obscuring the usual signal representative of glycan removal.

To further investigate the contributions of the 27 swapped amino acids in fDPP4 that supported infection, I took a subset of amino acid changes on each blade and made subsets of point mutations inside the larger block (indicated by ‘snp’). These mutations were selected based on whether the amino acids were found in permissive or nonpermissive species; amino

acids in fDPP4 that could be found in a permissive species were not included in the fine-tuning mutation sets. 278snp includes the mutations E285Q, D290A, A291S, and S293L. 330snp includes the mutations N330D, N331E, and D332S. 334snp includes the mutations N334G, S335R, and R337N. 338snp includes the mutations K338C, P339L, and E340V (Figure 2.8). Each construct was transfected into HEK 293T cells following previous protocols and infected with rMERS-CoV-RDP at an MOI of 1. Results show that all combinations of mutants show

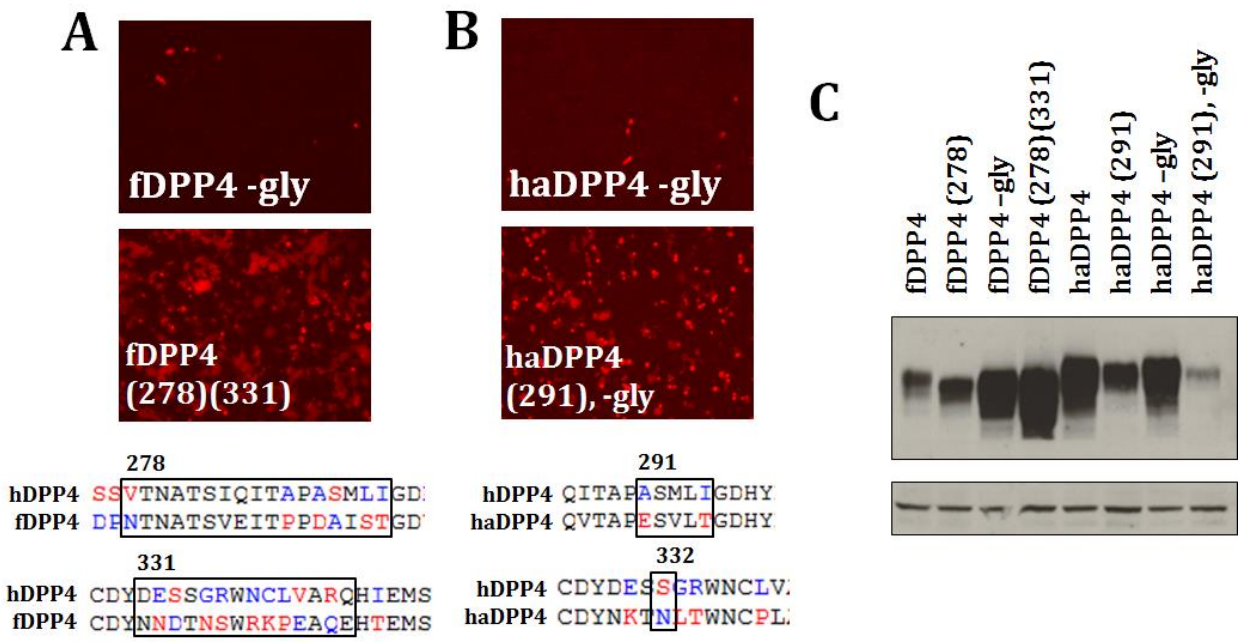
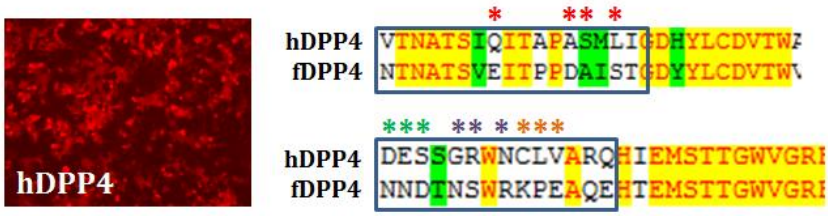


Figure 2.7: Many amino acid changes are required to make fDPP4 and haDPP4 permissive to MERS-CoV infection

(A) Removing glycosylation on its own does not confer permissivity to fDPP4. However, introducing a set of amino acid changes on blade IV (starting at residue 331) and blade V (starting at residue 278) allow fDPP4 to support MERS-CoV infection (fDPP4 (278)(331)). Sequences show the alignment between hDPP4 and fDPP4 with the blue boxes indicating the amino acids that were swapped from hDPP4 into fDPP4. Note that fDPP4 –gly is a negative control and only includes the single point mutation N332A. (B) Removing glycosylation on its own does not confer permissivity to haDPP4. However, combining three amino acid changes on blade V (starting at residue 291) with the glycosylation knockout mutant on blade IV (N332A) results in high levels of MERS-CoV infection. Sequences show the alignment between hDPP4 and haDPP4 with the blue boxes indicating the amino acids that were swapped from hDPP4 into haDPP4. (C) Western blot analysis of fDPP4 and haDPP4 and designated variants for DPP4 and β-actin expression. Western blot protocol follows Peck *et al.* 2015b.

lower levels of infection compared to when both blocks (278) and (331) are included together (Figure 2.8). Interestingly, the mutants fDPP4 (278), 334snp and fDPP4 (278), 338snp show increased levels of infection compared to all other combination sets (Figure 2.8). This suggests that 1) the (278) block is an important determinant and cannot be recapitulated with a subset of the four selected amino acids and 2) infection is more robust to amino acid changes on blade IV.



	None	(331)	330snp	334snp	338snp
None					
(278)					
278snp					

Figure 2.8: Fine-tune mapping of fDPP4 determinants for MERS-CoV permissivity

Sets of mutations were made in fDPP4 for fine-tune mapping of MERS-CoV permissivity determinants. The table indicates which combination of hDPP4 amino acid identities were introduced into fDPP4 on blade V (rows) or blade IV (columns). Blue boxes around the sequences indicate blocks of amino acid swaps designated by (278) and (331). 278snp includes four amino acid swaps on blade V designated by the red asterisks. 330snp, 334snp, and 338snp each include three amino acid swaps on blade IV indicated by the green, purple, and orange asterisks, respectively.

In fact, the lack of increased infection for fDPP4 (278), 330snp (Figure 2.8) suggests that our hypothesis surrounding glycosylation as an important barrier may not be as applicable to fDPP4. The 330snp subset removes the glycosylation site while the 334snp and 338snp amino acid changes do not. However, we cannot rule out that the 334snp and 338snp amino acid mutations alter the structure of fDPP4 in a way that shifts the glycosylation site and allows proper binding by MERS-CoV. Overall, the determinants that would allow fDPP4 to fully support MERS-CoV infection are complex and require further investigation to fully understand.

In the context of haDPP4, fewer amino acid changes can result in a dramatic increase in the infection levels (Figure 2.7B). Compared to fDPP4, which I found to require 27 mutations (possibly less), haDPP4 can support high levels of infection with just four mutations: E291A, V293M, T295I, and N332A. The first three are present on blade V while the fourth is on IV and removes the glycosylation site (Figure 2.7B). As with fDPP4, these residues were selected based on examining amino acids that were not present in permissive DPP4 orthologs. The four identified mutations overlap with previous studies that found five amino acid changes in haDPP4 were required for high levels of MERS-CoV infection (van Doremalen *et al.* 2016). To determine whether fewer mutations could recapitulate the results seen with all four, I used PCR mutagenesis to make a set of the mutations individually and in combination with the glycosylation knockout (N332A, designated as '-gly').

Following previous protocols, I transfected each haDPP4 construct into HEK 293T cells and infected with rMERS-CoV-RFP at an MOI of 1. Visualizing at ~24 hours post-infection reveals that no smaller set of amino acid mutations can fully recapitulate the increased levels of infection seen by all four combined (Figure 2.9). Although the construct haDPP4 293, -gly was not tested, it is unlikely that it would result in high levels of infection. This is supported by the

work by van Doremalen *et al.* (2016) which demonstrated that less than five amino acid changes were unable to confer high permissivity to MERS-CoV. Of these, residue 295 overlaps with the results found here, while residue 336 removes a glycosylation site, functionally achieving the same as the 334 mutation presented here (Table 2). Our conclusions show that while glycosylation is likely an important barrier to MERS-CoV infection, there are additional determinants on blade V that also play a crucial role.

Comparing the data we have so far on the residues that are important for mediating permissivity in hDPP4, mDPP4, haDPP4, it is difficult to discern an obvious pattern (Table 2). The only clear trend is that at least one change is required on *both* blades IV and V of DPP4. This indicates that there are two key points of interaction between DPP4 and the MERS-CoV RBD that are important for allowing the virus to utilize a new species receptor. Further understanding of the host range expansion of MERS-CoV will come with further data on which mutations can confer permissivity to DPP4 receptors of currently nonpermissive species.

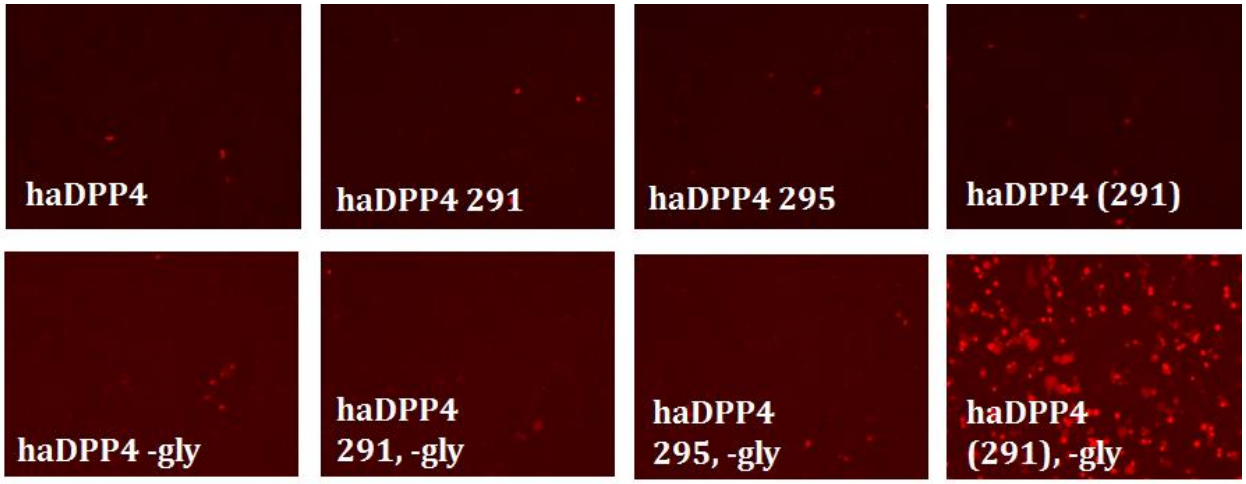


Figure 2.9: Fine-tune mapping of haDPP4 permissivity to MERS-CoV

Sets of mutations in haDPP4 in order to map determinants of permissivity to MERS-CoV infection. Mutations, indicated by residue numbers are as follows: E291A; T295I; (291): E291A, V293M, T295I; -gly: N332A (see Figure 2.7B).

Additionally, solving the crystal structures of nonpermissive DPP4 orthologs will reveal the specific interactions between these molecules and MERS-CoV, particularly what prevents MERS-CoV from successfully utilizing these molecules are functional receptors. The inability to identify a small number of changes that confer permissivity to DPP4 orthologs substantiates the earlier hypothesis that extensive remodeling of the MERS-CoV RBD would be required for the virus to infect these nonpermissive species, predominantly in order to overcome the barrier of glycosylation. While this result indicates that generating a virus that can infect these species will be difficult, it also suggests that the emergence of MERS-like coronaviruses into these species in a natural setting is unlikely. Additionally, the generation of a transgenic model in these other species is intractable due to the number of mutations each DPP4 would require.

2.4 Adaptation to alternate receptor molecules

In addition to considering the adaptation of MERS-CoV to nonpermissive DPP4 orthologs, we can also consider the possibility of MERS-CoV gaining the ability to utilize a new receptor.

Ortholog	**Important residues**										**Source**
Human DPP4	267			294	295			336			Song *et al.* 2014
Mouse DPP4				294				336*			Cockrell *et al.* 2014
Hamster DPP4		291			295			336*	341	346	van Doremalen *et al.* 2014, 2016
Hamster DPP4			293		295	297	334*				Peck *et al.* in prep

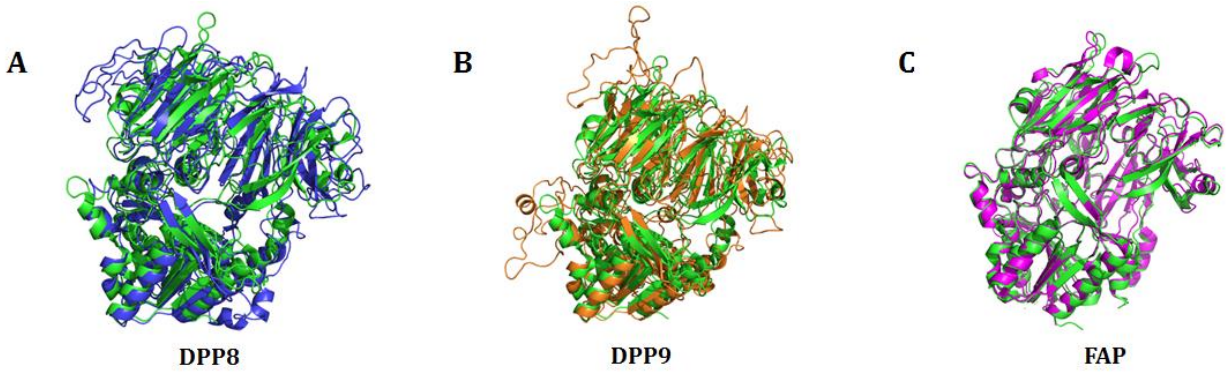
Table 2: Residues identified to be important for MERS-CoV permissivity

Experimental studies and the residues that have been identified as important for mediating permissivity to MERS-CoV among various DPP4 orthologs *in vitro*. All residues are relative to the aligning residue in hDPP4. Residues are either on blade IV (blue shading) or on blade V (gray shading). * indicates residues that knock out a glycosylation site in the mouse and hamster DPP4 molecules. Modified from Peck *et al.* 2015c.

Experimental evidence has shown that coronaviruses can adapt to an alternate receptor; MHV can shift from using its natural receptor CEACAM1a, to using heparan sulfate to enter the host cell in persistently infected cell cultures (de Haan *et al.* 2005). MERS-CoV could potentially evolve to utilize another dipeptidyl peptidase family member, such as DPP8, DPP9, or fibroblast activation protein (FAP). All three of these proteins are predicted to share high structural homology with DPP4 (Figure 2.10), despite low amino acid sequence identity (22%, 20%, and 52%, respectively). FAP appears to be the most likely candidate with the highest sequence and structural homology (Figure 2.10); FAP can also be expressed on the surface of cells, particularly at sites of tissue remodeling (Levy *et al.* 1999). While only some of the key residues identified in DPP4 studies match the amino acid identities and/or properties of FAP (Figure 2.10D), it could still act as a candidate for MERS-CoV adaptation. This shift in receptor usage could, for example, result from increased selective pressure following a reduction in expression of DPP4. Reduction in expression has been shown to occur *in vitro* where persistent MERS-CoV infection induces downregulation of DPP4 expression in bat cells (Cai *et al.* 2014). Additionally, selection for an alternate receptor could come from therapeutics that block DPP4 from being bound by the virus. While the use of antibodies against DPP4 is impractical due to the importance of DPP4 in other roles (Boonacker and van Noorden 2003), the administration of soluble DPP4 to prevent cell entry has been proposed (Xia *et al.* 2014) with results showing that soluble DPP4 can reduce and even block infection of cells by a pseudotyped MERS-CoV (Raj *et al.* 2013, Wang *et al.* 2013). Further experiments should be considered to determine whether selection imposed by reduced DPP4 expression or DPP4-based therapeutics would drive MERS-CoV to utilize an alternate receptor.

To investigate whether FAP can support MERS-CoV infection as a backbone molecule, I created a chimeric protein that swaps the region of hDPP4 that interacts with the MERS-CoV RBD into the FAP molecule (residues 279-343, numbering relative to FAP) (Figure 2.11B) using overlap extension PCR. The chimeric FAP molecule (FAP (279)) was then transfected into HEK 293T cells following previously detailed protocols. At ~24 hours post-transfection, cells were infected with rMERS-CoV-RFP at an MOI of 1. At ~24 hours post-infection, cells were imaged with red fluorescence as a readout of infection.

Results show that swapping the region of hDPP4 that interacts with MERS-CoV into the FAP backbone does allow for increased infection of the nonpermissive protein (Figure 2.11A).



D

Protein	267	291	294	295	336	341	346*
Human DPP4	K	A	L	I	R	V	I
Bat DPP4	K	A	L	T	K	V	I
Camel DPP4	K	A	L	I	R	V	I
FAP	R	A	A	S	T	K	I

*Residues numbered relative to human DPP4

Figure 2.10: DPP4 family members

(A) PyMOL visualization of crystal structures of human DPP4 (green, PDB code 4L72, chain A) aligned structurally to DPP8 (blue), (B) DPP9 (orange), or (C) FAP (magenta, PDB code 1Z68). DPP8 and DPP9 protein sequences were threaded using I-TASSER (Zhang 2008) to generate predicted structures. Alignment with DPP4 yields RMS values of 3.34 for DPP8, 3.41 for DPP9, and 0.782 for FAP calculated using PyMOL. (D) Key residues identified in mutagenesis studies with DPP4 orthologs (see Table 2) for permissive DPP4 orthologs and FAP. From Peck *et al.* 2015c.

However, the increase in infection is slight and requires quantification before any conclusions can be drawn. Additionally, during the cloning procedure, three amino acid identities in the final FAP (279) product did not faithfully match hDPP4 (Figure 2.11A, red box). Thus, the next step is to fix this error and determine whether the correct product yields higher levels of permissivity. Additionally, IFA should be performed to determine whether FAP is being successfully expressed on the surface of the cells. As noted above, FAP is expressed on the cell surface primarily during tissue remodeling (Levy et al. 1999). Thus, some additional activation component may be required to properly express FAP on the cell surface and evaluate the ability of MERS-CoV to utilize it as a functional receptor. Regardless of the troubleshooting that needs to take place to solidify these conclusions, the increase in infection seen when FAP is utilized as a backbone (Figure 2.11A), although small, raises the question of whether MERS-CoV could adapt to utilize FAP as an entry receptor. Extra care should be taken during the design of MERS-CoV therapeutics that are based on receptor targeting or interference.

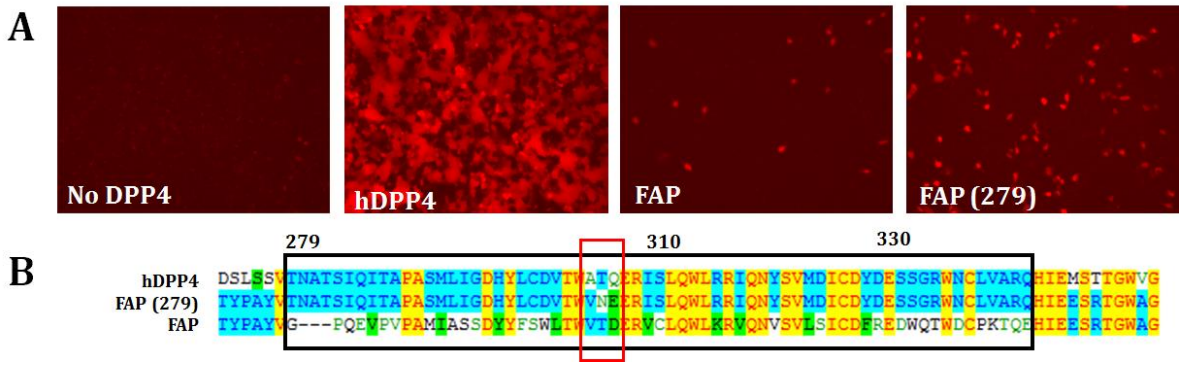


Figure 2.11: FAP as a backbone for supporting MERS-CoV infection

(A) HEK 293T cells were transfected with hDPP4, FAP, or FAP (279). At ~24 hours post-transfection, cells were infected with rMERS-CoV-RFP at an MOI of 1. Cells were imaged ~24 hpi. (B) Sequences of hDPP4, FAP (279), and FAP. FAP (279) uses the FAP sequence as a backbone, but swaps in the hDPP4 amino acid identities for residues 279-343 (residues numbered relative to FAP). This region interacts with the MERS-CoV RBD during infection. Three amino acids were not swapped during cloning (red box), making this an imperfect replica of the hDPP4 binding interface.

Because the bat coronavirus BtCoV-HKU4 also utilizes DPP4 as an entry receptor, we can compare its RBD to that of MERS-CoV to determine whether it might show the same infection profile (and concerns) as MERS-CoV. The spike proteins of MERS-CoV and BtCoV-HKU4 show 67% amino acid identity (accession numbers AHX00731.1 and YP_001039953.1, aligned using Vector NTI). MERS-CoV RBD residues important for facilitating the interaction between the virus and DPP4 include L506, W553, and V555 which form a hydrophobic core that interacts with hDPP4 residue L294 (Figure 2.12A), and RBD residue Y499 which engages the hDPP4 residue R336 (Wang *et al.* 2013). The amino acid identities at these locations in the RBD are partially conserved in BtCoV-HKU4 (Y503, L510, L558, I560) (Figures 2.12B, 2.12C), suggesting that utilization of hDPP4 and bDPP4 is robust to variation for some of these key interactions. Further studies can determine whether or not this characteristic is sufficient to allow BtCoV-HKU4 to utilize other host species receptors with the same efficiency as MERS-CoV, or whether BtCoV-HKU4 has the potential to emerge into the human population in the future. Based on the barrier of glycosylation in mDPP4 (Figure 1.7), it seems unlikely that BtCoV-HKU4 is able to utilize mDPP4 as a functional receptor. BtCoV-HKU4 and MERS-CoV share the conserved Y499 residue that, probably, introduces steric hindrance into the binding interaction if a glycan is present at the R336 residue in DPP4 (Figure 2.12A). Unfortunately, the lack of a tissue culture system for BtCoV-HKU4 prevents easily testing its ability to infect the panel of mDPP4 variants generated above. Previous work done on BtCoV-HKU4 entry has had to utilize a BtCoV-HKU4-spike pseudovirus system (Yang *et al.* 2014), with potential noise in how accurately this recapitulates a wildtype infection. Potential solutions include introducing the BtCoV-HKU4 spike protein into the background of a closely related virus (e.g. BtCoV-HKU5) in order to test its host range and receptor interaction dynamics. Additionally, it seems likely

that further metagenomics studies will reveal clusters of MERS-like group 2c strains that are more closely related to either BtCoV-HKU4 or MERS-CoV (Figure 0.2), which also utilize DPP4 receptors for entry. Thus, the structural mechanisms regulating DPP4 species-specificity and usage will become more clear over the next few years.

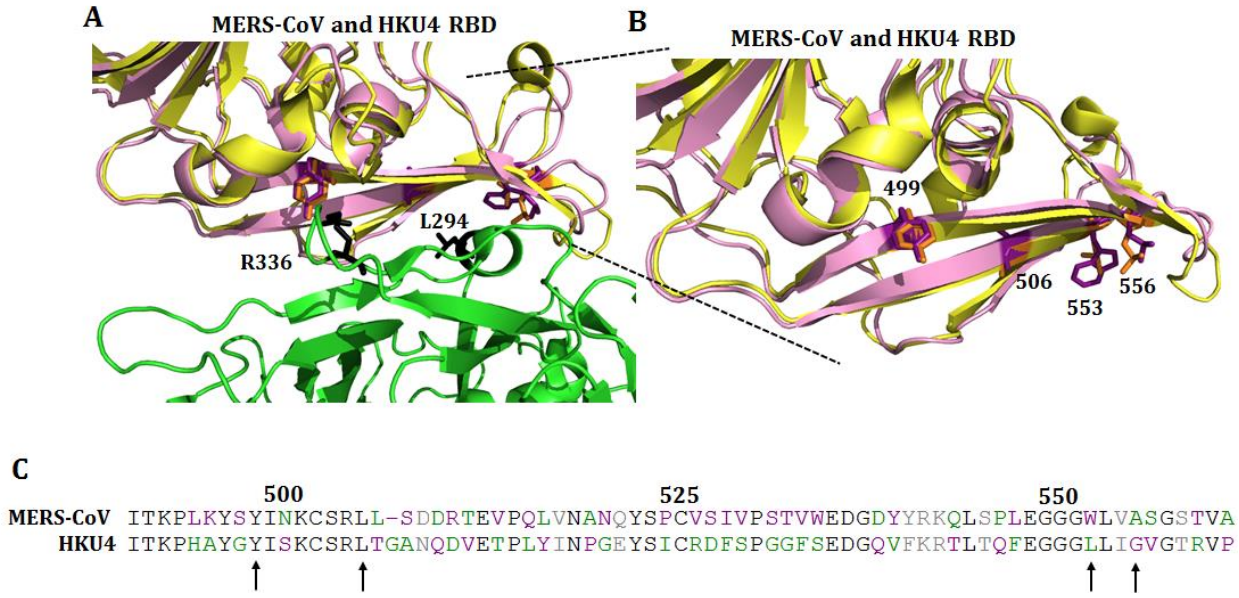


Figure 2.12: Comparison of BtCoV-HKU4 and MERS-CoV RBD

(A) Human DPP4 (green) bound to the MERS-CoV RBD (pink) (PDB code 4L72) overlaid by hDPP4 bound to the BtCoV-HKU4 RBD (yellow) (PDB code 4QZV). DPP4 residues 294 and 336 (black) have been identified as important in the interactions with MERS-CoV. (B) MERS-CoV (pink/purple) and BtCoV-HKU4 (yellow/orange) are partially conserved at the residues of the RBD that interact with hDPP4: 499, 506, 553, and 555. (C) GeneDoc alignment of part of the MERS-CoV and BtCoV-HKU4 RBDs. Arrows point to the RBD residues that interact with hDPP4. Numbers relative to MERS-CoV. Structures visualized using PyMOL. From Peck *et al.* 2015c.

CHAPTER 3: MOUSE-ADAPTATION OF MERS-CORONAVIRUS

3.1 Introduction

Coronaviruses can infect over 20 different species and have undergone cross-species transmission into new species multiple times over the course of their evolutionary history. For example, the first human coronavirus that we know of was introduced into humans ~800 years ago and five other host range expansion events have been documented since then. Presently, what factors drive viral host range expansion is not well known. Changing habitats and the geographic expansion of humans into zoonotic territories may increase the probability of viruses achieving the needed exposure to accomplish a host range expansion (deFillippis and Villareal 2000). However, because coronaviruses can infect such a diverse number of species, it may be possible that coronaviruses, specifically, have biological characteristics that make them more likely to infect new species compared to other viruses. No concrete evidence has surfaced to support this claim, but by understanding the recent expansion of SARS-CoV and MERS-CoV, we may gain some insight into whether coronaviruses are more evolvable in terms of cross-species transmission.

My previous work has demonstrated that one of the primary barriers to host range expansion is host cell receptor incompatibilities (Cockrell *et al.* 2014, Peck *et al.* 2015b). However, experiments in other coronaviruses have shown that receptor differences can be easily overcome. For example, studies in murine hepatitis virus (MHV) found the emergence of host range variants following persistent infection (Baric *et al.* 1999). Four amino acid substitutions in the MHV spike gene allowed the virus to successfully infect nonpermissive human and hamster

cells (McRoy and Baric 2008). Three mutations in a strain of SARS-CoV isolated from a civet were required for successful infection and replication in human cells (Sheahan *et al.* 2008), although this was accompanied by a tradeoff in binding affinity to either civet or human ACE2 (Wu *et al.* 2012a). Additionally, a single mutation in the spike protein was associated with increased mouse ACE2 receptor usage (Roberts *et al.* 2007). These data suggest that coronaviruses can easily adapt to utilize receptor orthologs and engage in cross-species transmission. As we improve our understanding of the DPP4 biochemical interface, we will have a better grasp on predicting which orthologs MERS-CoV may be most likely to adapt to, and which species may be the most likely candidates for MERS-CoV host range expansion.

The biochemical determinants of MERS-CoV permissivity can inform the potential evolution of host range by the virus. The knowledge that other DPP4 orthologs can act as backbones to support infection (Cockrell *et al.* 2014, Raj *et al.* 2014, van Dormelan *et al.* 2014) suggests that in theory, it is possible for MERS-CoV to expand into currently nonpermissive hosts. However, experimental results to date indicate that it may not be easy for the virus to adapt to these alternate receptors. First, the presence of glycans as barriers can dramatically disrupt the interaction between the MERS-CoV RBD and DPP4 (Peck *et al.* 2015b). Second, changing the interactions on both blades IV and V of nonpermissive DPP4 orthologs, as seen with mouse and hamster studies (Cockrell *et al.* 2014, van Dormelan *et al.* 2014, van Dormelan *et al.* 2016), is crucial. Both of these observations suggest that *multiple* mutations in the spike protein would be required for a virus to efficiently utilize these orthologous receptors. The probability of these mutations occurring simultaneously in the same genome may be unlikely, depending on the actual number of changes needed. If MERS-CoV does readily adapt to orthologous receptors, another question is what changes are needed to promote increased

transmission between species or between individuals within a species. Currently, no transmission model exists for coronaviruses, creating a gap in our ability to experimentally evaluate the evolution of enhanced transmissibility.

A powerful tool for evaluating whether MERS-CoV can evolve to utilize a nonpermissive host cell receptor is to perform adaptation experiments *in vitro.* Thus I performed passaging experiments of MERS-CoV on wildtype mDPP4 to determine whether MERS-CoV could adapt to a glycosylated DPP4 molecule. The benefit of this experiment is that it 1) allows me to document the mutations necessary for host range expansion at a receptor level, specifically following the adaptation pathway of mutations, and 2) produces a mouse-adapted MERS-CoV strain that can be used to study pathogenesis. While the developed mouse models provide a basis for beginning these studies, the production of a mouse-adapted MERS-CoV would broaden the span of possible research by allowing studies in different genetic backgrounds and without the caveats associated with transgenic mouse models (see Chapter 1). Of note is that adaptation of MERS-CoV to a glycosylated receptor will likely drive the virus away from high efficiency binding to hDPP4, potentially reducing the risk of this virus to human infection.

3.2 Adaptation of MERS-CoV to glycosylated DPP4 receptors

Developing a mouse model that can be used to study MERS-CoV pathogenesis is crucial for the development of vaccines and potential therapeutics. The previously produced mouse models, however, are not ideal, due to issues of non-specific (Agrawal *et al.* 2015) or transient (Zhao *et al.* 2014) expression. These issues make the development of a mouse-adapted MERS-CoV appealing. A virus that can infect wildtype mice not only permits more researchers to engage in pathogenesis studies, but also allows for investigating different mouse genetic backgrounds. Chapters 1 and 2 reveal that a crucial barrier to MERS-CoV infection is

glycosylation of the DPP4 receptor. This is evident when the removal of glycosylation allows mDPP4 to support MERS-CoV infection (Figure 1.7A) and the addition of glycosylation to hDPP4 restricts MERS-CoV infection (Figure 1.6A). Additionally, permissivity of additional DPP4 orthologs, primarily fDPP4 and haDPP4, seems to require the removal of glycosylation (Figures 2.7A and B), although other unknown determinants also play a role. Unfortunately, because the crystal structure of mDPP4 has not been solved, we cannot use computational modeling to engineer a virus that is capable of infecting mDPP4. However, we can try adapting MERS-CoV around the glycan barrier.

Adaptation of MERS-CoV to mDPP4 could reveal insights to the mutational pathways that coronaviruses take to gain compatibility with new host species receptors. To date, the mutational pathway that MERS-CoV took to expand from bats into humans is unknown. A camel intermediate host species is the primary hypothesis (Reusken *et al.* 2013, Gossner *et al.* 2014), but it is uncertain whether specific changes in the spike protein facilitated the infection of camels, and then the subsequent infection of humans. Without identification of the MERS-CoV progenitor, we are unable to retrace its adaptation trajectory to understand what changes were required and whether those changes were stochastic or driven by some selective pressure. Furthermore, these adaptation experiments can help reveal whether MERS-CoV is even capable of adapting around a glycan barrier. This will have implications for 1) potential mechanisms of host evolution to evade viral infection and 2) whether MERS-CoV (or MERS-like coronaviruses) can potentially emerge into these nonpermissive species in nature.

Previous studies have found that experimental adaptation of an RNA virus to a new host can be successful when a host invasion strategy is used (Morley *et al.* 2015). Briefly, the population starts in an environment that is composed of 100% permissive and 0% nonpermissive

hosts. The population is then passaged with a decreasing proportion of permissive to nonpermissive hosts until the latter makes up 100% of the population. Using this strategy, I performed two treatments of permissive-nonpermissive host combinations: the first uses hDPP4-expressing cells as the permissive host and hDPP4 + gly-expressing cells as the nonpermissive host (Figure 3.1); the second uses mDPP4 288, 330-expressing cells as the permissive host and mDPP4 288-expressing cells as the nonpermissive host (Figure 3.2). For these experiments, I created NIH 3T3 cell lines stably expressing the appropriate DPP4 variant. Using a mouse cell line for adaptation experiments better recapitulates the eventual environment the virus will encounter and prevents selection for human cell line-adapted mutations. One observation of note, however, is that human cell lines seem to more faithfully glycosylate putative sites compared to

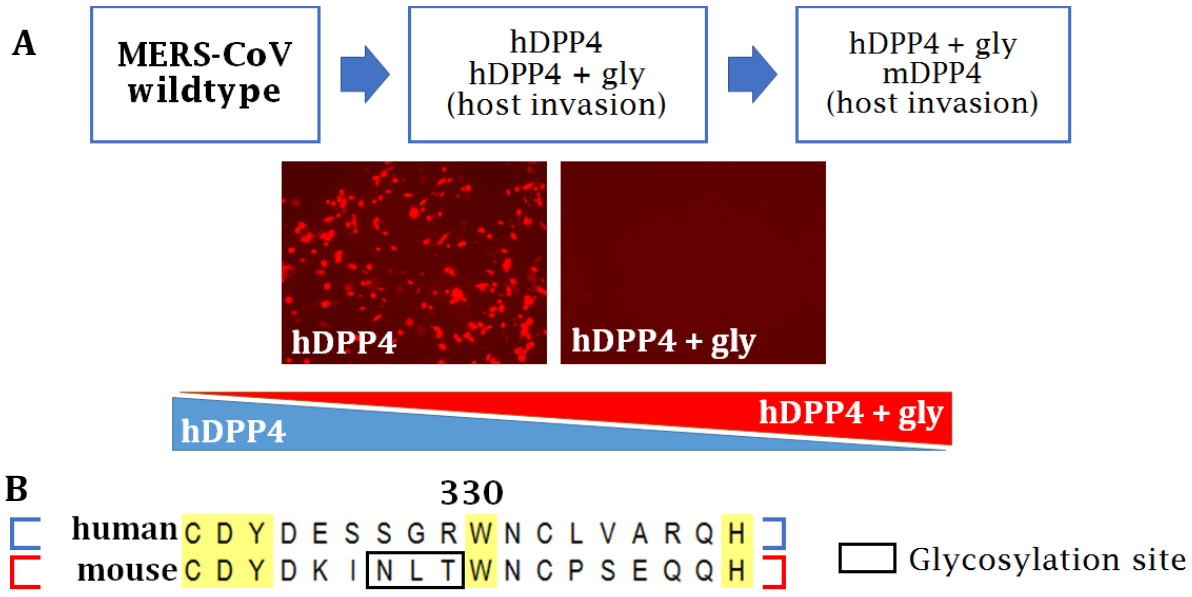


Figure 3.1: Host invasion adaptation strategy using hDPP4 as a backbone

Adaptation experiment design using a host invasion strategy. (A) Wildtype MERS-CoV is passaged on a mixture NIH 3T3 cells expressing either hDPP4 or hDPP4 + gly. Over time, the proportion of hDPP4 to hDPP4 + gly decreases until the cell population is composed entirely of hDPP4 + gly. (B) The hDPP4 + gly sequence uses hDPP4 as a backbone and swaps the NLT glycosylation site (black box) of mDPP4 in place of the aligning hDPP4 residues.

rodent cell lines (Figure 3.3). HEK 293T cells and Huh7 cells (both human) show intermediate MERS-CoV infection levels when hDPP4 + gly or mDPP4 330 are overexpressed (Figure 3.3). However, when these DPP4 constructs are overexpressed in mouse cells (NIH 3T3 and DBT) or hamster cells (BHK), there is no increase in infection (Figure 3.3). This trend holds true for other DPP4 constructs that are glycosylated, whether in the hDPP4 or mDPP4 backbone (data not shown). Other studies have found similar variation in levels of glycosylation between different cell types and this discrepancy has had an impact on coronavirus species specificity (Wentworth and Holmes 2001, Li *et al.* 2005b). However, more studies are needed to fully understand whether the difference seen here in the context of DPP4 truly is variation in the level of glycosylation fidelity within these cell lines. However, if this hypothesis proves true, it could suggest that glycosylation would not be a proper barrier to infection in hDPP4 (and/or other permissive non-glycosylated DPP4 orthologs) *in vivo*.

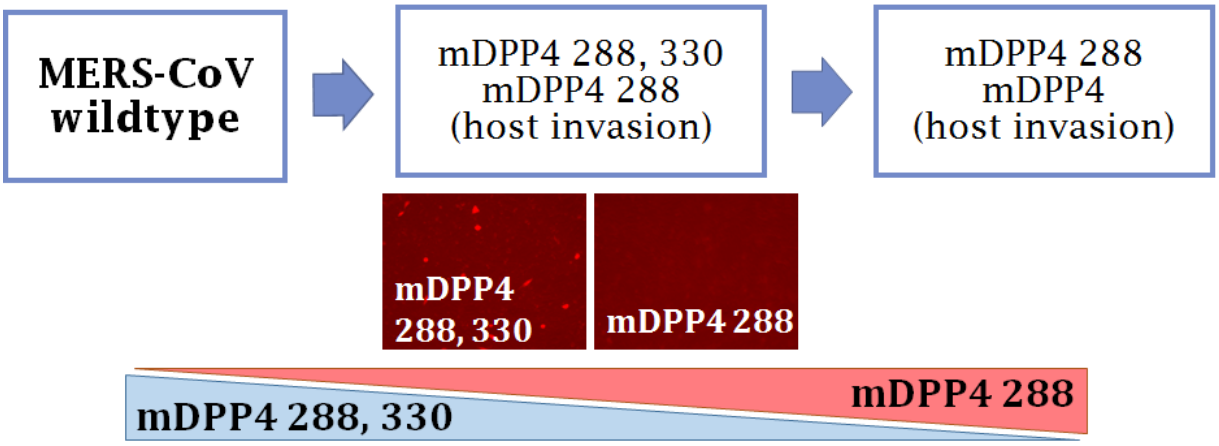


Figure 3.2: Host invasion adaptation strategy using mDPP4 as a backbone

Adaptation experiment design using a host invasion strategy. Wildtype MERS-CoV is passaged on a mixture NIH 3T3 cells expressing either mDPP4 288, 330 or mDPP4 288. Over time, the proportion of mDPP4 288, 330 to mDPP4 288 decreases until the cell population is composed entirely of mDPP4 288.

Each NIH 3T3 cell line used in the following adaptation experiments stably expresses the indicated DPP4 construct. NIH 3T3 cells were transfected with each DPP4 variant and then selected with Bleocin (EMD Millipore) in order to isolate those cells that had successfully integrated the plasmid into their genomes. Selection was maintained on the cells to prevent reversion of the population; however, infection experiments were carried out with fresh media (i.e., no Bleocin was present during passaging of the virus). Stable NIH 3T3 cell lines were established expressing the following DPP4 constructs: hDPP4, hDPP4 + gly, mDPP4, mDPP4 288, and mDPP4 288, 330. To test for successful selection of cells with integrated DPP4 genes, I

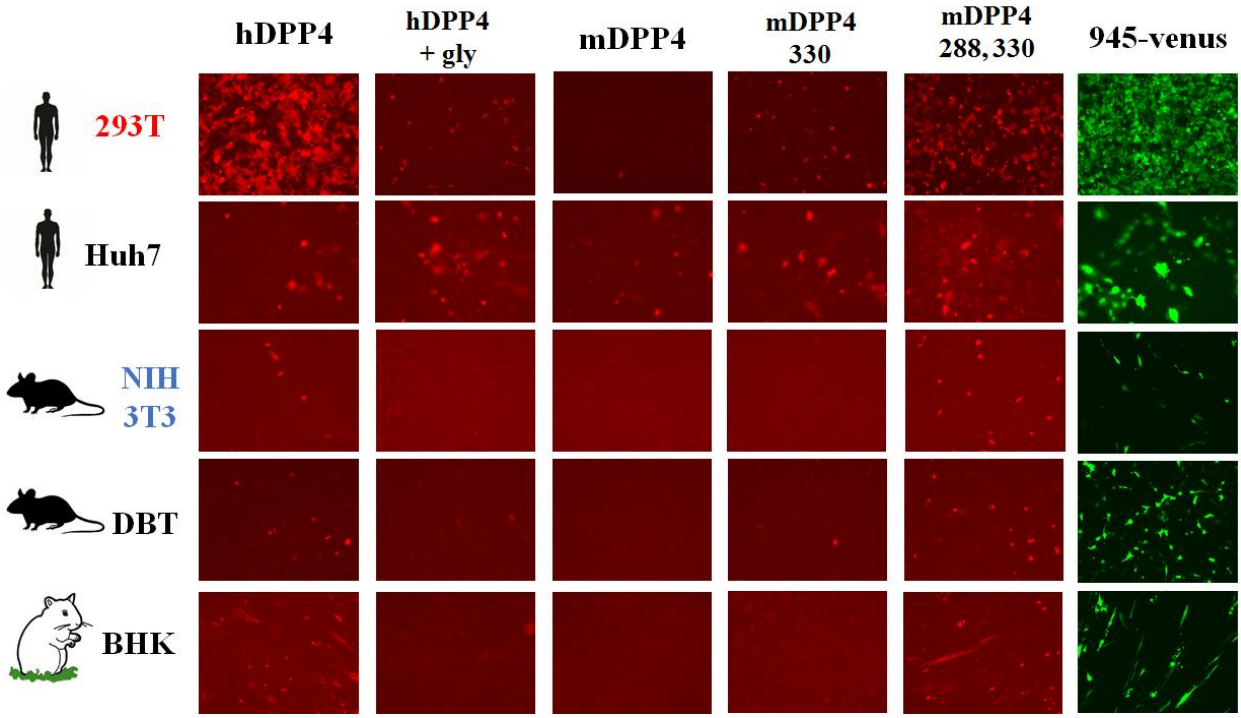


Figure 3.3: DPP4 panel tested on human, mouse, and hamster cell lines

Five DPP4 constructs were tested for permissivity using five different cell lines. 945-venus construct used as a control for transfection efficiency. Human cell lines include 293T and Huh7; mouse cell lines include NIH 3T3 and DBT; hamster cell lines include BHK. hDPP4, hDPP4 + gly, mDPP4, mDPP4 330, and mDPP4 288, 330 were overexpressed in each cell line and infected with MERS-CoV-RFP. 293T (red) and NIH 3T3 (blue) cell names indicate the two used in this dissertation.

analyzed three of the cell lines by immunofluorescence (Figure 3.4). Staining for DPP4 shows high levels of DPP4 expression on the surface of the cells, recapitulating the expected endogenous localization of DPP4.

After generating these cell lines, I characterized the growth curve for both of the permissive (starting) cell lines to be used: hDPP4 and mDPP4 288, 330 (Figure 3.5). T75 flasks of each cell type were infected with rMERS-CoV-RFP at an MOI of 0.01. Supernatant samples were taken at 0, 6, 12, 18, 24, 48, 72, and 96 hours post-infection and viral titers measured via plaque assay. Determining the growth curves informed the selection of an incubation period that ensured peak viral titer for both cell lines before passaging. Both stable cell lines reach their peak viral titer at 72 hours post-infection (Figure 3.5B). In fact, NIH 3T3 cells expressing hDPP4

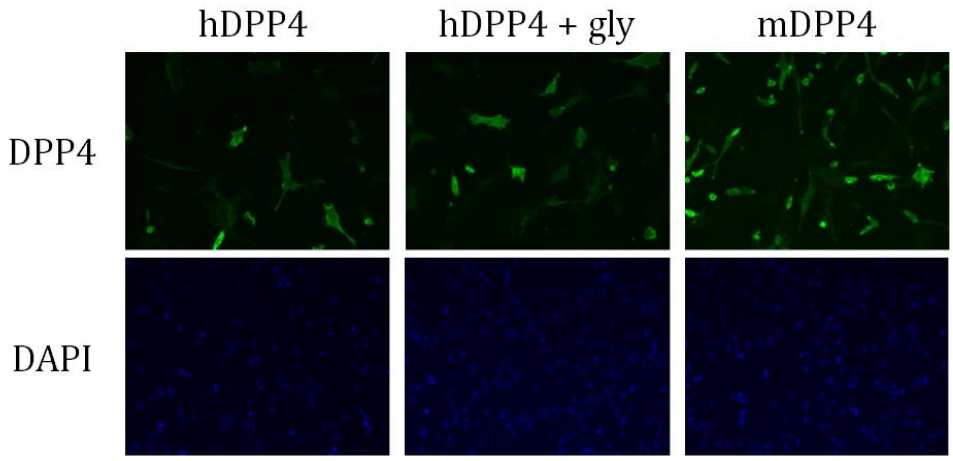


Figure 3.4: DPP4 construct expression in stably expressing NIH 3T3 cells

DPP4 and mutant variants are expressed on the surface of cells, visible by immunofluorescence (IFA). NIH 3T3 cells were transfected with the given DPP4 variant and selected for stable expression using Bleocin. After obtaining a population of Bleocin-resistant cells, cell lines were seeded, fixed, and probed with primary goat-anti-DPP4 polyclonal antibody (R&D Systems) at 1:50 and secondary donkey-anti-goat Alexa Fluor 488 (Life Technologies) at 1:500. Cells were imaged at 40X for DAPI (30 ms exposure) and DPP4 (160 ms exposure). The hDPP4 + gly variant has the hDPP4 backbone sequence with the mDPP4 glycosylation site swapped in (see Figure 3.1B)

reach viral titers that are equivalent to when MERS-CoV is grown in Veros (monkey cells; permissive) (Figure 3.5B). Due to these results, virus populations were transferred at 72 hours for each passage.

For the start of the adaptation experiments, only NIH 3T3 cells stably expressing hDPP4 or mDPP4 288, 330 were seeded in separate T75 flasks and infected with rMERS-CoV-RFP at an MOI of 1 (Passage 1). At 72 hours post-infection, the supernatant was harvested and 2 mL

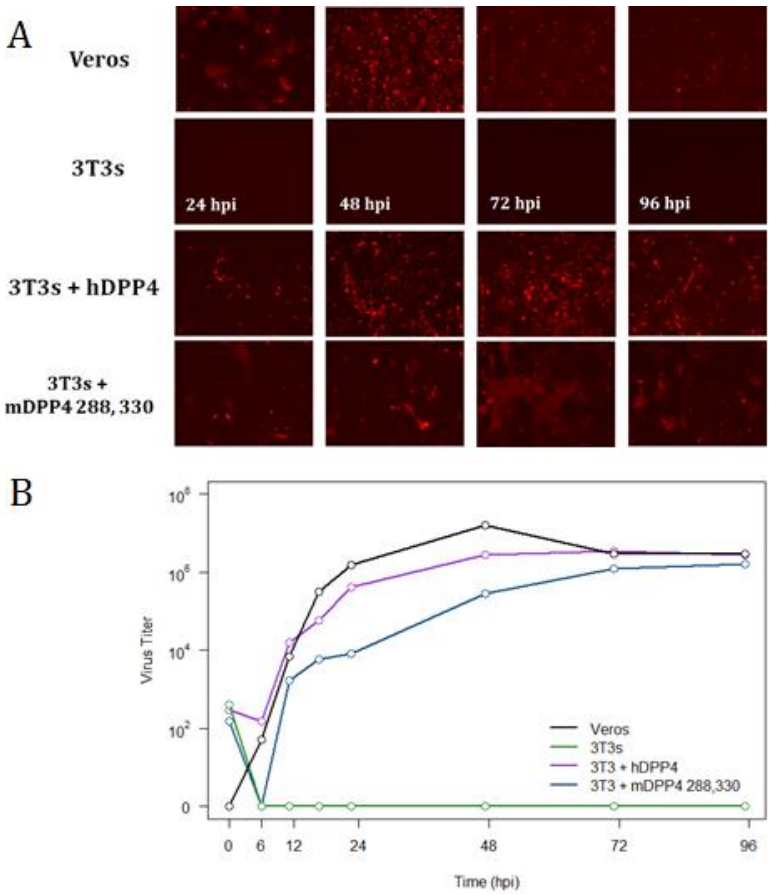


Figure 3.5: Growth curve of permissive NIH 3T3 stably-expressing cell lines

NIH 3T3 cell lines stably expressing hDPP4 and mDPP4 288, 330 were generated and rMERS-CoV-RFP infection (MOI 0.01) tracked over time. (A) Infections were imaged at 24, 48, 72, and 96 hpi. Veros are a positive control for infection and wildtype NIH 3T3s are a negative control. (B) Viral titers (PFU/mL) were measured at various time points for each cell type. Peak titers are reached by both stable cell lines by 72 hpi.

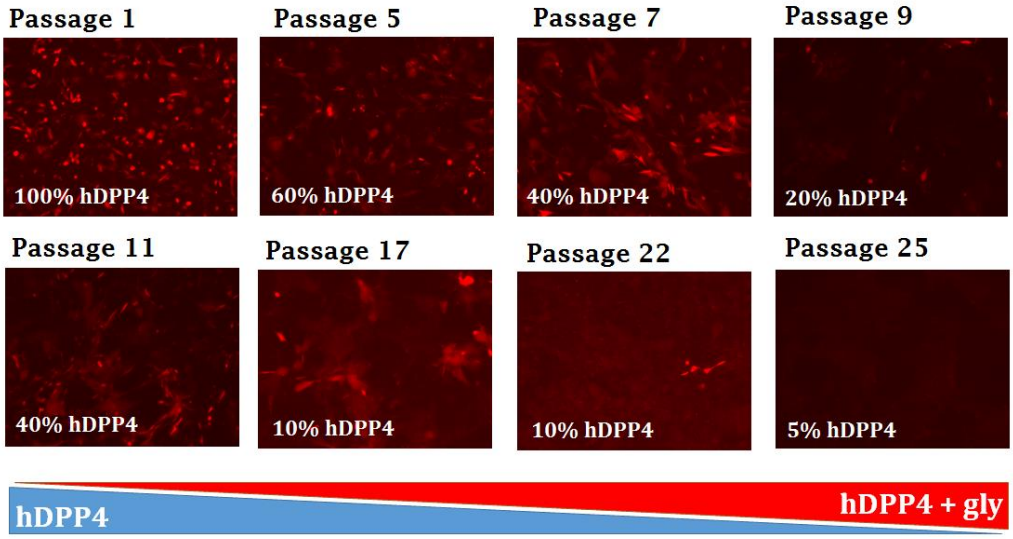


Figure 3.6: Host invasion adaptation of MERS-CoV to hDPP4 + gly

rMERS-CoV-RFP was passaged on NIH 3T3 cells stably expressing hDPP4 or hDPP4 + gly. Passage 1 included 100% hDPP4 cells and decreased by 10% hDPP4 each passage. At passage 10, the percentage of permissive cells was increased to 40% and decreased by 10% every other passage (see Figure 3.8). Each passage was at an MOI of 1, with a ~72 hour incubation period. Images were taken at ~72 hpi before the next round of infection.

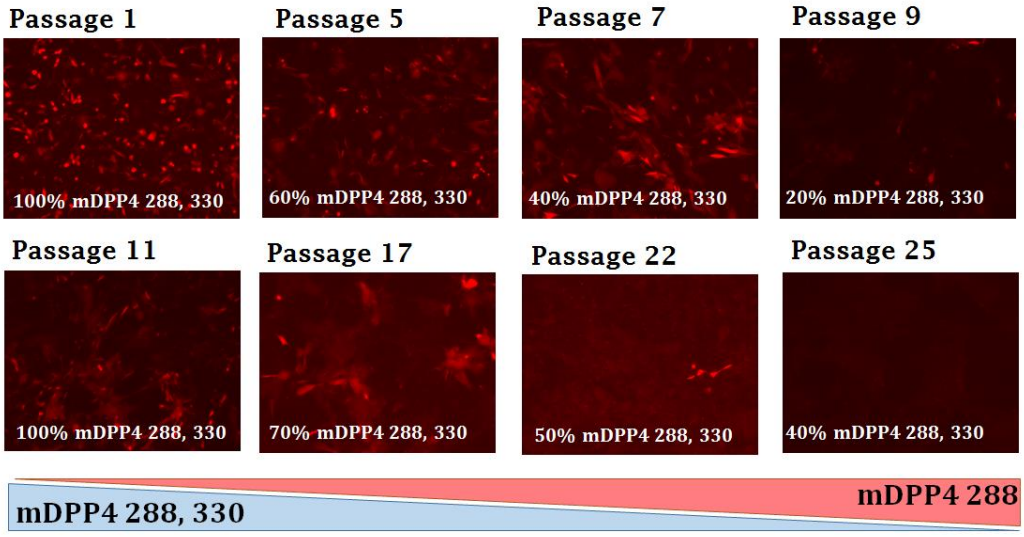


Figure 3.7: Host invasion adaptation of MERS-CoV to mDPP4 288

rMERS-CoV-RFP was passaged on NIH 3T3 cells stably expressing mDPP4 288, 330 or mDPP4 288. Passage 1 included 100% mDPP4 288, 330 cells and decreased by 10% each passage. At passage 10, the percentage of permissive cells was increased back to 100% and decreased by 10% every other passage (see Figure 3.8). Each passage was at an MOI of 1, with a ~72 hour incubation period. Images were taken at ~72 hpi before the next round of infection.

transferred to a fresh T75 flask of cells with 90% hDPP4 and 10% hDPP4 + gly, or 90% mDPP4 288, 330 and 10% mDPP4 288 (Passage 2) at an MOI of ~1. At each passage, the cells were visualized (Figures 3.5 and 3.6) and viral titer measured via plaque assay (Figure 3.8). Based on the earlier passages, it appeared as though changing the percent permissive host by 10% was too rapid of a change for the viral population to have the opportunity to allow beneficial mutations to increase in frequency (Figures 3.5 and 3.6). In the work done by Morley *et al.* (2015), the rate of change in the population composition did play an important role in the adaptation and endpoint fitness of the viral population. Adjusting my approach, at passage 10, I increased the permissive

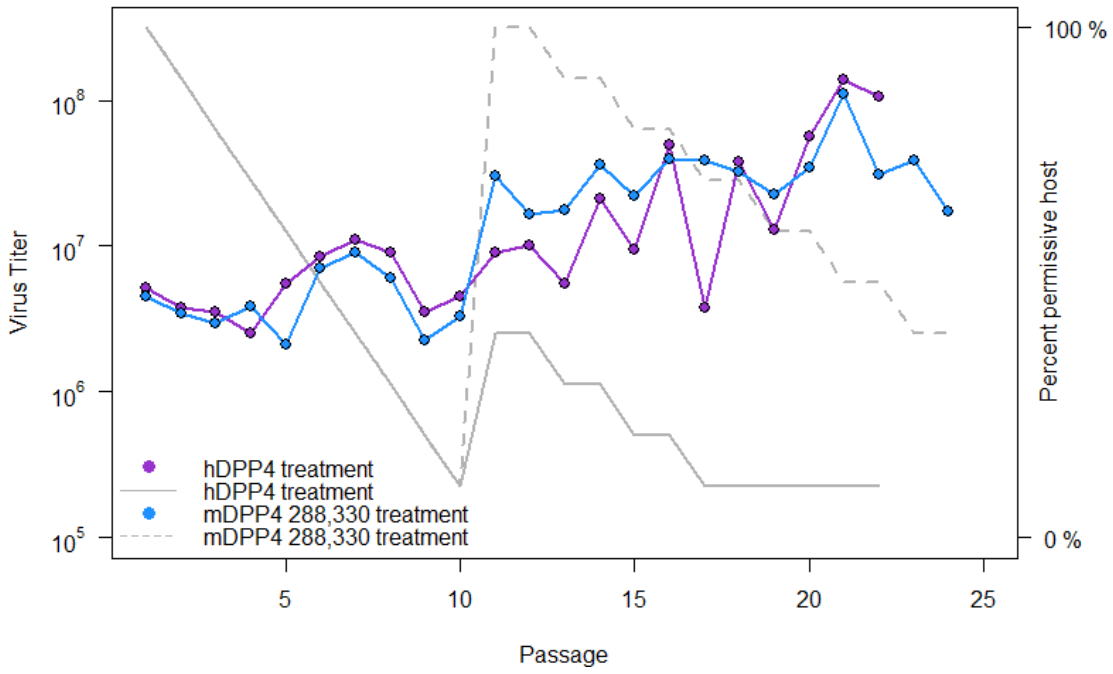


Figure 3.8: Virus titer and percent permissive host cells for host invasion adaptation experiments

Virus titer (left y-axis, PFU/mL) against passage number for both host invasion treatments. hDPP4 treatment includes hDPP4 stably expressing NIH 3T3 cells as the permissive host and hDPP4 + gly as the nonpermissive host (purple). mDPP4 288, 330 treatment includes mDPP4 288, 330 stably expressing cells as the permissive host and mDPP4 288 as the nonpermissive host (blue). Cell composition is represented by percent permissive (right y-axis) which indicates the percent of the culture that was composed of permissive host cells (hDPP4, solid gray line; mDPP4 288, 330, dashed gray line).

host percentage to 40% for the hDPP4 treatment and 100% for the mDPP4 288, 330 treatment (Figure 3.8, gray lines). Subsequent passages decreased the percent permissive host by 10% every other passage (instead of every passage). The titers of both populations seem to increase over time, albeit with a consistent decrease at the first drop in permissive host percentage. However, the detection of red fluorescence from imaging seems to decrease steadily over time (Figures 3.5 and 3.6), providing contradictory results.

One possible explanation is that the rMERS-CoV-RFP virus has kicked out the RFP tag over time. Because the RFP is not crucial for replication, a virus that has a deletion in this region would replicate more quickly and have a greater fitness than other viral variants. To test whether the RFP tag was still present in the adapting populations, and to evaluate whether any adaptation has occurred, I tested passage 25 virus populations from both the hDPP4 and mDPP4 288, 330

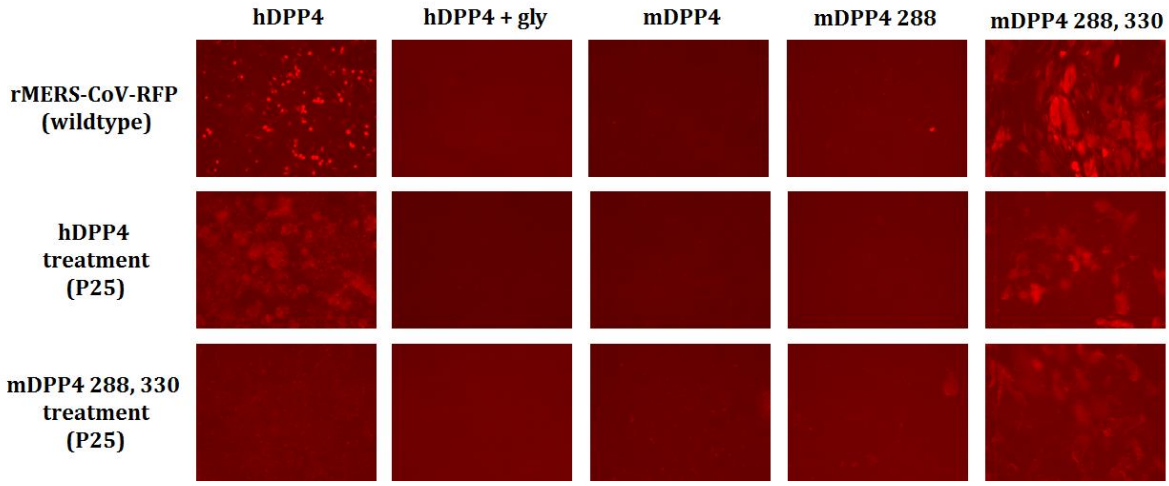


Figure 3.9: Infection panel of passage 25 virus from two adaptation experiment treatments

Three virus populations were tested for successful infection on NIH 3T3 cell lines stably expressing five different DPP4 constructs: hDPP4, hDPP4 + gly, mDPP4, mDPP4 288, and mDPP4 288, 330. Wildtype rMERS-CoV-RFP (top row); passage 25 from the hDPP4 treatment (middle row, see Figure 3.6); and passage 25 from the mDPP4 288, 330 treatment (bottow row, see Figure 3.7). Cells were infected at an MOI of 1 and imaged 48 hpi.

treatments on NIH 3T3 cells stably expressing: hDPP4, hDPP4 + gly, mDPP4, mDPP4 288, and mDPP4 288, 330. Wildtype rMERS-CoV-RFP virus was used as a positive control. Each cell type was infected with the three virus populations (hDPP4 P25; mDPP4 288, 330 P25; and wildtype rMERS-CoV-RFP) at an MOI of 1, and the cells imaged ~48 hpi (Figure 3.9). Results show that fluorescence can be seen when both virus populations are tested against mDPP4 288, 330. However, infection is only visible on the hDPP4 construct when the hDPP4 treatment population is tested (Figure 3.9). This indicates 1) that the populations did not lose RFP expression and 2) that the mDPP4 288, 330 treatment population may be evolving to better utilize the mDPP4 backbone and become less efficient at utilizing hDPP4. Additional analyses are needed to determine whether the fluorescence visualized in Figure 3.9 is truly indicative of each population's ability to utilize the panel of receptors. Sequencing of both populations at passage 25 to evaluate any mutations in the spike protein is currently in progress.

The inability of MERS-CoV to easily adapt around a blade V glycan suggests that generating a mouse-adapted MERS-CoV by passaging may be biologically intractable. Even when the virus is, presumably, well-adapted to the backbone molecule of hDPP4, the population is unable to restructure its RBD in order to infect utilizing the hDPP4 + gly receptor variant (Figures 3.6 and 3.8). Using mDPP4 as a backbone also did not result in successful adaptation (Figures 3.7 and 3.8). These results support the hypothesis that multiple mutations are required in order to overcome the barrier of glycosylation. With multiple mutations needed, the probability of sampling a successful variant decreases dramatically. Even within 25 passages, both populations were unable to sample enough sequence space to encounter such a variant, if one is biologically possible.

There are two evident solutions to the dilemma of multiple mutations being required to produce a mouse-adapted MERS-CoV. The first is to crystallize the mDPP4 molecule and computationally predict the RBD remodeling that would need to occur to bind around glycosylation. The identified mutations could then be reverse engineered into the MERS-CoV genome to produce a variant capable of infecting wildtype mice. The second is to enhance the genetic variation available to the population; this can be done by chemical mutagenesis or by production of a mutator allele. Due to the evolutionary implications of the latter, I propose to investigate whether a mutator allele can facilitate host range expansion by increasing the genetic variation of a population and producing sequence variants that could successfully infect a new species.

4. FUTURE DIRECTIONS

4.1 Development of a MERS-CoV mutator

The high mutation rates of RNA viruses give them the ability to evolve rapidly, particularly when combined with their large population sizes and short generation times. However, most mutations are lethal or deleterious (Sanjuan *et al.* 2004, Stern *et al.* 2014). The large number of deleterious mutations can cause high mutation rates to be disadvantageous by pushing the population closer to the extinction threshold – the mutation rate becomes so high that the deleterious mutation load causes the mean fitness of the population to decrease to such an extent that the population goes extinct (Bull *et al.* 2007). RNA viruses exist close to this threshold, as shown through lethal mutagenesis experiments in many different RNA viruses (FMDV, (Sierra *et al.* 2000); poliovirus, (Crotty *et al.* 2001)). In fact, it has been suggested that most RNA viruses replicate at an optimum mutation rate that maximizes fitness, virulence (Korboukh *et al.* 2014), and evolvability (Sanjuan 2012) without crossing the extinction threshold.

Despite the proximity of RNA viruses to the extinction threshold, mutator alleles have the potential to play a role in promoting host range expansion. A mutator allele increases the inherent mutation rate of a virus and thus enhances the genetic variation of the population; this variation may harbor mutations that are beneficial or required for expansion into a new host. If the benefit of a mutation that allows a virus to jump into a new host outweighs the cost of an increased deleterious mutation load compared to wildtype, the mutator allele has the potential to increase in frequency in the population. This increase comes in the form of hitchhiking – the

mutator allele hitchhikes along with the beneficial mutation that it produced. Note that hitchhiking is dependent on asexual reproduction (i.e., no recombination) so the mutator remains linked to the beneficial mutation it produced. This works best when a single beneficial mutation can confer an advantage; in more complex systems where adaptation requires multiple beneficial sites, recombination plays an important role in bringing these mutations onto the same background. The hitchhiking of mutator alleles has been well documented in other organisms, such as bacteria (Sniegowski *et al.* 2000, Shaver *et al.* 2002), but has yet to be experimentally demonstrated in a viral system. Still, theory supports that mutators will be favored by natural selection in the face of novel environments (Taddei *et al.* 1997, Tanaka *et al.* 2003), such as that presented by a new host species.

Mutator phenotypes have been isolated from natural populations of HIV-1 (Gutierrez-Rivas and Menendez-Arias 2001) and influenza A virus (Suarez *et al.* 1992). Additionally, targeted mutations have resulted in mutator strains for poliovirus (Liu *et al.* 2013, Korboukh *et al.* 2014), coxsackievirus B3 (Gnadig *et al.* 2012), chikungunya virus (Rozen-Gagnon *et al.* 2014), foot and mouth disease virus (Xie *et al.* 2014) and coronaviruses (Eckerle *et al.* 2007, Eckerle *et al.* 2010, Graham *et al.* 2012). However, coronaviruses are the only RNA viruses that have evolved a replication proofreading mechanism independent of the RNA polymerase. The nsp14 gene, known as ExoN, has 3'-to-5' exoribonuclease activity that is similar to the proofreading activity of DNA polymerases (Minskaia *et al.* 2006). This proofreading capability has several implications. *First*, it likely allowed coronaviruses to expand their genome size past that of other viruses (Gorbalenya *et al.* 2006, Lauber *et al.* 2012). *Second*, it allows the production of mutator alleles that are outside of the polymerase. When nsp14 is mutated in either MHV or SARS-CoV, the virus suffers from a 16 to 20-fold increase in mutation rate (Eckerle *et*

al. 2007, Eckerle *et al.* 2010), otherwise known as a mutator phenotype. Experiments with the SARS-CoV mutator show that in direct competition with wildtype, the mutators quickly die out when no selection is present (Graham *et al.* 2012). This suggests a cost to the mutator allele, whether directly through the mutation itself or indirectly through the accumulation of deleterious mutations. However, the fact that viable mutators are likely to easily arise through mutations in nsp14 suggests that they may be influential in shaping the evolution of coronaviruses when a selective pressure is present, such as that of a new host or immune system. One caveat, however, is that if the *in vivo* MOI is high, coinfection of cells can result in recombination of the mutator with wildtype genomes, thus uncoupling the mutator allele from the beneficial mutations that it produced. This is a particular concern for coronaviruses which have been shown to have up to 25% recombination rates *in vitro* (Baric *et al.* 1990). Improved *in vivo* recombination and coinfection rate estimates will help inform the potential for mutators to succeed in wildtype populations.

Mutators have been implicated as potential factors in previous host range expansion events. For example, phylogenetic data suggest that an avian influenza virus (H1N1) jumped into pigs and then into humans about 100 years ago (Webster *et al.* 1992), with a mutator allele as the primary hypothesis for how the virus could rapidly cross two species barriers (Ludwig *et al.* 1995). When the H1N1 avian influenza jumped into European swine again in 1979, a mutator allele was thought to be responsible. However, even though evolution rates were higher in the new influenza strain compared to the ancestral strain, its mutation rate was not (Stech *et al.* 1999). While the role of a mutator in this host range expansion event is highly speculative and controversial, mutators may still play a role in influenza evolution since clinical isolates have been readily identified with mutation rates 3-4 times greater than in ancestral strain (Suarez *et al.*

1992), suggesting that natural populations could contain high frequencies of mutators at any given time. Additionally, mutator strains of norovirus (NoV) have been suggested to play an important role in the recent NoV pandemics. Using *in vitro* RdRp assays, the mutation rates of various genogroup II genotype 4 (GII.4) strains in addition to less common GII.b/GII.3, GII.3, and GII.7 strains (Bull *et al.* 2010) were measured. The mutation rates of the predominant GII.4 strains were 5 to 36-fold higher than the less frequently detected lineages. This evidence suggests that increased inherent mutation rate can enhance the epidemiological fitness of circulating virus strains.

Mutators have not yet been implicated for host range expansion in coronaviruses, but this could simply be due to a lack of discovery. While the mutation rate of the original SARS-CoV isolate has been measured (9×10^{-7} substitutions per nucleotide per replication cycle, (Eckerle *et al.* 2007)), the mutation rates of closely related bat coronaviruses have not. Furthermore, the mutation rate of MERS-CoV has not yet been experimentally determined, although the evolution rate has been estimated at 1.12×10^{-3} substitutions per site per year (Cotten *et al.* 2014) which is comparable to the SARS-CoV estimate of 2.82×10^{-3} (Lau *et al.* 2010). It would be interesting to compare the MERS-CoV mutation rate with those of BtCoV-HKU4 and BtCoV-HKU5. If a mutator allele was responsible for producing the variation necessary for MERS-CoV to emerge in humans, we would expect it to have a higher inherent mutation rate than those of closely related lineages circulating in bats. Furthermore, mutators may play an important role in coronavirus vaccine development. The SARS-CoV mutator has been shown to be attenuated *in vivo* with no reversion to virulence following passaging or persistent infection (Graham *et al.* 2012), providing a promising starting candidate for further development. Additionally, while SARS-CoV is resistant to ribavirin (Chiou *et al.* 2005, Barnard *et al.* 2006), the mutator strain is

not (Smith *et al.* 2013). This suggests that a possible therapeutic strategy could be to combine mutagens with inhibitors of nsp14 activity. Further studies can help advance the use of mutators in the development of vaccines and therapeutics.

Whereas mutators have been readily engineered for murine hepatitis virus (MHV) and SARS-CoV (Eckerle *et al*., 2007, Eckerle *et al.* 2010), a successful mutator has not yet been produced for MERS-CoV. Generating a mutator for MERS-CoV to add to the panel of coronavirus mutators is beneficial for many reasons. First, it can *increase our understanding of the potential for MERS-CoV to expand its host range* to other species. While utilizing chemical mutagenesis would allow for similar tests of the impact of enhanced genetic variation, developing a mutator is crucial for directly testing whether mutators play a significant role in promoting host range expansion events. This outcome is directly applicable to naturally circulating coronavirus populations. Second, a MERS-CoV mutator can help us *characterize the range of potential antibody escape mutants*, having high relevance to pathogenesis and the future development of therapeutics. Third, the SARS-CoV mutator has been explored as a *potential player in vaccine design* due to its increased attenuation *in vivo* (Graham *et al.* 2012), implying that a MERS-CoV mutator may serve a similar application. Overall, mutator alleles have a wide range of applications, including being used to explore questions of polymerase fidelity (poliovirus, Pfeiffer and Kirkegaard 2003), antibody escape (Bayliss *et al.* 2008), drug resistance (HIV, Mansky and Bernard 2000; HSV-2, Nishiyama *et al.* 1985), attenuation of pathogenesis (coronavirus, Graham *et al.* 2012; coxsackievirus B3, Gnädig *et al.* 2012), and lethal mutagenesis (FMDV, Sierra *et al.* 2000).

My strategy for identifying a mutator allele makes use of a thorough understanding of mutator dynamics in DNA-based organisms, achieved through the combination of population

genetics and evolution experiments (Sniegowski *et al.*1997, Chao and Cox 1983). These studies show that mutators are favored under three primary conditions. First, *the population must be adapting*. The presence of a selective pressure ensures that beneficial mutations are possible, resulting in the mutator strain having an increased probability to produce the beneficial mutation before the wildtype strain. Second, *the population must reproduce asexually*. This ensures that the mutator will increase in frequency by hitchhiking along with the beneficial mutation that it produced. Third, *the cost of the mutator allele must not be too high*, or rather, the benefit of the beneficial mutation must be greater than both the direct cost of the mutator allele (the mutation itself) and the indirect cost of an accumulation of deleterious mutations elsewhere in the genome.

Previous attempts at generating a MERS-CoV mutator have been unsuccessful; namely, introducing the same mutations into MERS-CoV that produce a mutator in SARS-CoV and

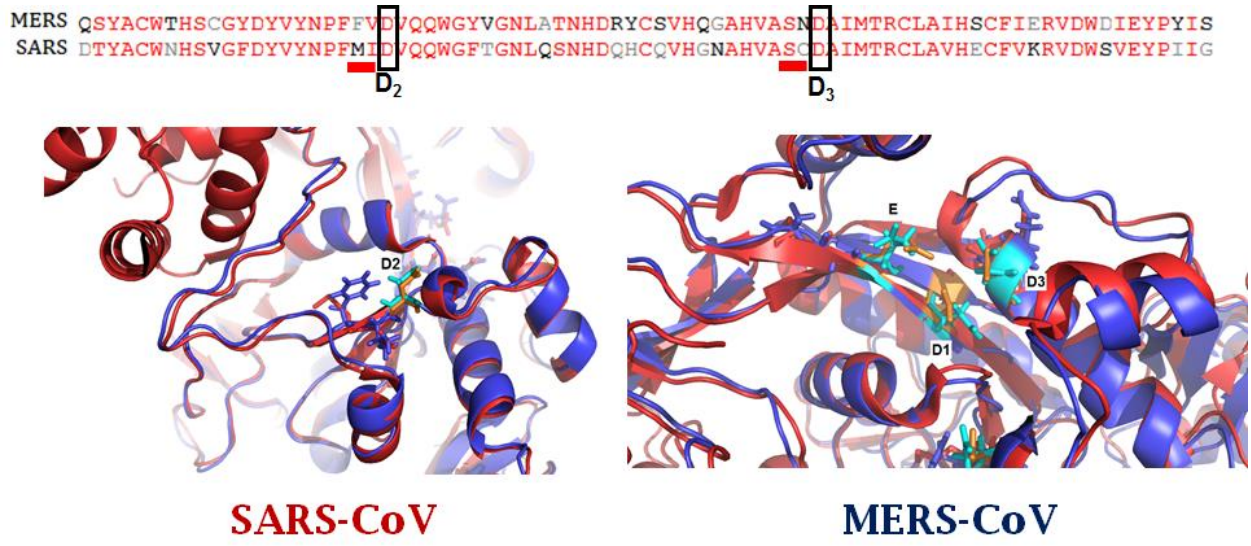


Figure 4.1: Sequence and structural alignment of SARS-CoV and MERS-CoV nsp14

(A) MERS-CoV and SARS-CoV nsp14 sequence alignments with the second and third D residue of the DEDD motif highlighted (black boxes). Red lines indicate which residues were chosen in MERS-CoV for generating a codon library. (B) Structural alignment of SARS-CoV (red; PDB code 5C8U) and MERS-CoV (blue; threaded using I-TASSER (Zhang 208)) nsp14. Amino acids representing the DEDD motif are shown in orange (SARS-CoV) and light blue (MERS-CoV), with the residues adjacent to the DEDD sites visible in red (SARS-CoV) and blue (MERS-CoV).

MHV (Eckerle *et al.* 2007, Eckerle *et al.* 2010) are inviable in the context of MERS-CoV (data not shown). These mutators were achieved by mutating one to two of the DEDD motif residues to alanines; additional attempts to change these four residues to other amino acid identities also failed (personal correspondence with Mark Denison). Thus, in order to identify a MERS-CoV mutator, specifically one with a low-fitness cost, I generated a pool of potential mutators by creating a library of nsp14 mutants. However, instead of making mutations within the DEDD motif residues, I created a codon library (i.e. all permutations of each codon) at residues *adjacent* to the DEDD active site residues (Figure 3.10A). Assuming that the active sites in the context of MERS-CoV are too sensitive to tolerate mutations, changing the amino acids adjacent to the active site residues may alter the structure enough to decrease fidelity without resulting in an inviable variant.

Residues for my codon library were chosen based on a sequence and structural comparison with SARS-CoV nsp14 (Figure 4.1). Two residues (six nucleotides) were chosen directly adjacent to a DEDD active site residue; this number of nucleotides is based on the number of variants that would be experimentally tractable. With six positions and four different nucleotides, there are over 4,000 possible sequence combinations. This reaches the threshold of the number of viruses that will be successfully transcribed and electroporated into cells to produce infectious virions during the reverse genetics process. Although redundant codons will cause some amino acid combinations to be more represented than others, including more residues in our library would only enhance this bias.

Another consideration for my selection criteria is that nsp14 is present across two of the seven plasmids that are used in the MERS-CoV infectious clone system (see Scobey *et al.* 2013). This has the benefit of generating codon libraries on two plasmids that can be used to produce

variants individually or in combination. Selecting active sites that are represented on both plasmids increases my ability to generate variant combinations and potentially identify a viable mutator strain. For example, each plasmid will have 20^2 amino acid combinations; when these two plasmids are combined, the number of variants increases to 20^4 (or 160,000). One note is that the nature of a codon library prevents me from being able to exclude wildtype sequences from the virus pools. Even if the exact wildtype sequence was eliminated (e.g., by some fortuitous restriction sites), alternate codons that result in the same amino acid sequence would also need to be eliminated. Thus my downstream experiments keep in mind that most populations will have some (or many) wildtype sequences that will outcompete mutator variants in environments with no selection (Graham *et al.* 2012).

I used sequence and structural information to choose the residues to compose my codon library. I noted both the conservation of amino acid identities at adjacent DEDD residues to SARS-CoV (Figure 4.1A) and the structure of those amino acids around the active sites (Figure 4.1B). Even though the MERS-CoV nsp14 structure used here is a predicted molecule threaded using I-TASSER (Zhang 2008), it can still provide useful visual information. After combining these analyses, I chose one set of two residues adjacent to the D_2 active site (the second D in the DEDD motif) that were nonconserved with SARS-CoV (Figure 4.1A). The second set I chose was adjacent to the D_3 active site and partially conserved with SARS-CoV (Figure 4.1A). Examining the structure of these amino acids shows extreme differences adjacent to D_2, and subtle differences adjacent to D_3 (Figure 4.1B). Creating libraries from both of these sequence and structural phenotypes will increase the probability of identifying a viable mutator variant, if one is biologically possible.

Codon libraries were ordered from Biobasic and amplified for virus production. Briefly, the infectious clone system utilizes seven plasmids that contain the entire genome (Scobey *et al.* 2013). Unique restriction sites (BglI) are located on both the 5' and 3' end of the genome fragment within each plasmid. Genome fragments are digested and purified before ligation of the fragments occurs to form a complete cDNA copy of the genome. Note that the ligation reaction and all subsequent virus production steps are carried out in a BSL3 facility. Using the complete cDNA copy of the genome as a template, a transcription reaction transcribes the genome into RNA. This full-length RNA is then electroporated into Vero cells. Because coronaviruses are positive-sense ssRNA viruses, they do not require extra machinery to be included along with the genome copies; once inside, the genome is translated into protein and the virus produces all of the components needed to complete its replication cycle. After ~72 hours, viral progeny are harvested from the culture supernatant and these virions represent fully infectious particles originating from the electroporated sequences. As noted above, however, this process is not highly efficient; the yield of each step decreases such that the final harvested virus is a fraction of the original input.

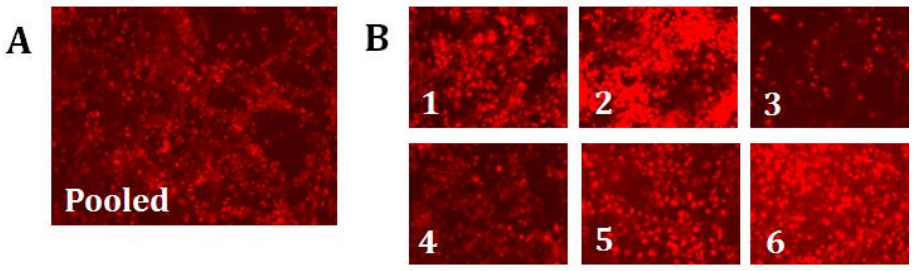


Figure 4.2: Pooled and plaque isolated rMERS-CoV-RFP D_2 mutator virus

rMERS-CoV-RFP D_2 mutator virus populations. (A) rMERS-CoV-RFP mutator viruses are electroporated into Vero cells and grown as a pooled population. Cells are imaged ~72 hpi. (B) rMERS-CoV-RFP mutator viruses are grown on an agarose plate, allowing individual plaques to be isolated. Plaques are grown in a 96-well plate of Vero cells and imaged at ~72 hpi. Viruses shown here are made from the plasmid with the codon library adjacent to D_2 (second D of the DEDD motif, see Figure 4.1).

Based on pilot experiments for mutator virus production, it seems as though the majority of mutant variants (not necessarily mutator phenotypes) are inviable. Pools of viruses were harvested from the initial electroporation products; imaging showed high levels of infection from the pooled samples (Figure 4.2A). However, sequencing of this pool suggests that the majority of sequences present are, in fact, wildtype (data not shown). Additionally, instead of electroporating and growing the potential mutator variants as a pool, I also grew them on an agarose plate. This removes competition and allows each virus to grow as an isolated plaque. Each plaque was then harvested and grown in a separate well of a 96-well plate. Imaging of each individual well shows high levels of variation in infection levels (Figure 4.2B). This suggests that even if I am not successful in the production of a fidelity variant (e.g., increase or decrease in mutation rate), there is variation in overall fitness.

Plaque-isolated viruses were also heavily biased toward the wildtype sequence. In the first replicate, 84 plaques were isolated, grown, and sequenced. Out of these, 39 were wildtype (46.4%), 42 had indistinguishable sequences (50%), and 3 were a clear non-wildtype sequence (3.6%). This is discouraging because it suggests that 1) wildtype is outcompeting variants rapidly, perhaps in a manner independent from growth rate and burst size in the context of a pooled population; or 2) most mutations in the nsp14 gene are inviable. The expected wildtype (sequence) frequency in our library is 0.20% for the D_2 mutator ($(2^1 \times 4^1)/4^6 = 0.002$), and 0.30% for the D_3 mutator ($(6^1 \times 2^1)/4^6 = 0.003$). However, this expected frequency assumes that each codon is represented equally when the cDNA is transcribed into RNA. Additionally, while the above values represent the expected wildtype *sequence*, the possibility of a mutation that still results in a wildtype *phenotype* cannot be excluded. Thus, more replicates and

sequencing are needed to understand the wildtype bias seen above and whether any mutator strains are present in the viruses generated.

Once I have generated a pool of viruses that I am confident has a proportion of non-wildtype sequence variants, the aforementioned population biology approach can help identify a low-fitness cost mutator. Pooled virus will undergo passaging (and hence, competition) under conditions where the population will be either adapting (with monoclonal antibody selection) (Figure 4.3) or not adapting (without monoclonal antibody selection). Recall that a mutator is expected to increase in frequency when: 1) the population is adapting; 2) the population is asexual; and 3) the cost of the mutator is not too high. The novel selective pressure of monoclonal antibodies meets the first condition of an adapting population. Several monoclonal antibodies have been identified for MERS-CoV (Agnihothram *et al.* 2014). These antibodies fall

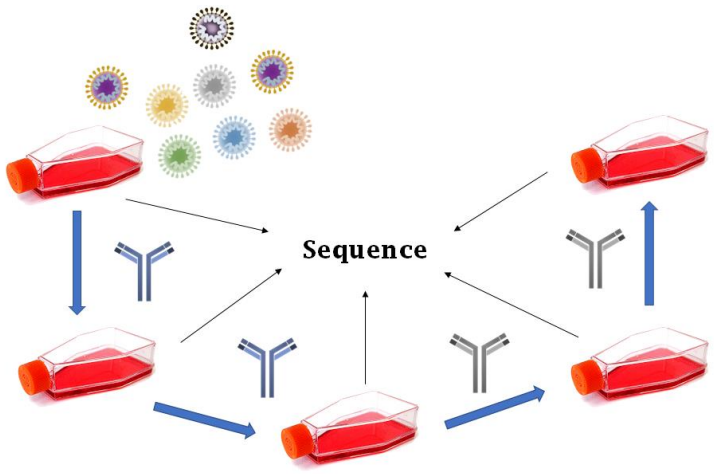


Figure 4.3: Schematic for passaging rMERS-CoV-RFP pooled variants under selection to identify a mutator allele

A codon library of variants (colored virus diagrams) is passaged in the presence of sequential monoclonal antibodies against MERS-CoV (blue and gray antibody diagrams). At each passage, the population is submitted for deep-sequencing. Mutator viruses are expected to increase in frequency over time and wildtype (sequence and phenotype) viruses are expected to decrease in frequency.

into three groups that neutralize different epitopes on the MERS-CoV spike protein, without cross-neutralization between groups. Additionally, escape mutants have been readily identified for all except one of these antibodies (Agnihothram *et al.* 2014). This confirms that escape mutants are capable of being produced, and also improves my ability to determine how much time and genetic variation would be required for a mutator to produce an escape mutant before the wildtype population. To meet the second condition of an asexually reproducing population, the MOI is kept very low (e.g., 0.001) to prevent two virions from infecting the same cell where their genomes could recombine. For the third condition, cost of the mutator cannot be controlled, but the use of a codon library improves the chance that there is a low-fitness cost mutator variant in the population.

I expect the wildtype virus to reach the highest frequency in an environment *without selection*. If a variant reaches a similar or greater frequency than wildtype, it is unlikely to be a mutator; mutators will produce more deleterious mutations than wildtype and will be unable to gain an advantage since beneficial mutations are not present in an environment without selection. I expect mutator viruses to reach the highest frequency in an environment *with selection*. Mutators will be more likely to produce escape variants compared to wildtype, which will be favored under the selective pressure of neutralizing monoclonal antibodies. Among mutators, strains with the lower fitness costs will be favored above other mutators and will reach higher frequencies. Here, deep sequencing allows me to not only identify a set of potential mutator alleles through comparison of mutant frequencies in the two population treatments, but also attribute fitness costs to those alleles. After potential mutators have been identified by deep sequencing, the actual mutation rate can be measured by passaging a population of each potential mutator - generated by reverse genetics using the infectious clone system - and sequencing at

regular intervals to count the number of mutations that occur for each approximate replication cycle (see Eckerle *et al.* 2010).

The importance of understanding mutators and the potential role they play in promoting host range expansion and viral emergence has not gone unnoticed (Mansky and Cunningham 2000, Smith *et al.* 2014). However, the tractability of testing these in an experimental setting has thus far been absent. Here, I present an approach for characterizing host range expansion and the role of mutators in an RNA virus system. First, MERS-CoV has a clearly defined spike-receptor relationship with various DPP4 orthologs (see Chapters 1 and 2), resulting in a panel of permissive and nonpermissive receptors that can be used to understand the mechanisms of host range expansion. This is given further power by the mutant mDPP4 molecules generated in Chapter 1, which will allow me to fine-tune the characterization of permissivity. Second, the convenience of the coronavirus infectious clone system allows us to test hypotheses by directly mutating or constructing specific viral variants. Third, the presence of an RdRp-independent proofreading mechanism allows us to generate viable mutator strains that can be used to explore various aspects of mutator dynamics in an RNA virus population. Based on our previous knowledge of the relationship between mutation rate and evolution rate (Figure 0.1), the presence of proofreading in coronaviruses presents an interesting case study. It is fathomable that a population would fix a mutator allele in a selective environment, thus causing the population to overcome a pressure such as a new host species, and then revert back to the wildtype mutation rate in order to reduce the number of deleterious mutations accumulating in the population. Whether this is a viable scenario in a wildtype population will be determined by further experiments and surveillance research.

The coronavirus system is significant because it allows for increases of as much as 20-fold in the genome-wide mutation rate (Eckerle *et al.* 2010), compared to the 3 to 4-fold increase seen for influenza (Suarez *et al.* 1992). Utilizing these innovative tools, we can expand future experiments to examine 1) the dynamics of species-specific restriction factors by transfecting DPP4s into their native host cell types; 2) the prevalence and possible strategies of compensatory and/or revertants after a mutator has reached fixation in the population and how this can impact host range expansion events. Whereas developing a mutator MERS-CoV strain would permit the use of an extensive panel of receptor variants (see Chapters 1 and 2), the mutator strains already identified for SARS-CoV and MHV can be used to study these questions if additional receptor resources were developed.

4.2 Potential evolutionary mechanisms for coronavirus host range expansion (modified from Peck *et al.* 2015c).

Understanding the evolutionary mechanisms facilitating viral host range expansion is crucial for prevention of, and preparation for, new emerging pathogens. Unfortunately, the forces that drive host range expansion events are still relatively unknown. Even for well-studied viruses such as influenza viruses, SARS-CoV, and MERS-CoV, the path of emergence and the selective pressures preceding emergence are often not clear. Here I discuss two potential evolutionary mechanisms, in addition to mutators, that may influence coronavirus host range expansion: recombination and mutational robustness (i.e. the ability to remain phenotypically constant or functional in the face of genetic perturbations). These topics can be explored in future studies to determine their potential impact on the emergence of MERS-CoV.

Recombination can act to create new viral variants and is a common event in the adaptive evolution of RNA viruses (Worobey and Holmes 1999). The ability to generate new genetic variants allows viral populations to explore the adaptive landscape at a faster rate than permitted

by mutation alone (Shackelton *et al.* 2005). Recombination has been implicated as a major evolutionary force for HIV-1 (Robertson *et al.* 1995), and has been suggested to contribute to host range expansion for many viruses, such as influenza A virus (Scholtissek *et al.* 1978), nuclear polyhedrosis virus (Maeda *et al.* 1993), and cauliflower mosaic virus (Schoelz and Wintermantel 1993). Coronaviruses can have *in vitro* recombination rates approaching 25 percent in progeny after a single round of coinfection at high multiplicity of infection (MOI) (Baric *et al.* 1990), likely due to the ability of the RdRp to switch templates during replication (Holmes 2003). In fact, recombination has been implicated in the sharing of homologous genes by distantly related members of the *Coronaviridae* family (Worobey and Holmes 1999). Phylogenetic analyses have revealed that hCoV-NL63 may have undergone many recombination events, including two sites of recombination in the S gene and potential recombination between hCoV-NL63 and porcine enteric disease virus (PEDV) in the M gene (Pyrc *et al* 2006). Additionally, evidence suggests that natural avian infectious bronchitis virus (IBV) isolates recombined with a vaccine strain (Holland 52), specifically in the nucleocapsid (N) gene (Jia *et al.* 1995). Not only has recombination been detected between coronaviruses, but traces of gene acquisition from highly distinct viruses have been detected, such as the transfer of influenza C-like hemagglutinin-esterase to coronaviruses in the 2a subgroup (Zeng *et al.* 2008). The potential for inter- and intraspecies recombination among coronaviruses gives them the ability to create a greater panel of novel and potentially pathogenic genomes.

The role of recombination in the SARS-CoV emergence is controversial. Initially, it was hypothesized that SARS-CoV was a product of recombination between two circulating bat coronaviruses, producing a hybrid virus capable of jumping into other species (Rest and Mindell 2003, Li *et al.* 2006). Studies suggest that the SARS-CoV M and N genes shared a common

ancestor with the lineage that eventually became the avian infectious bronchitis coronaviruses (Marra *et al.* 2003, Stavrinides and Guttman 2004). By contrast, the PP1ab polyprotein is most similar to murine-bovine coronaviruses, while the S gene is most similar to both the avian and group 1 (feline, canine, and porcine) coronaviruses (Stavrinides and Guttman 2004). The apparent homology of SARS-CoV with very different clades suggests that extensive recombination has occurred in its evolutionary history. However, others argue that SARS-CoV is distinct and not a product of a recombination at all (Holmes 2003, Demogines *et al.* 2012), with a false signal for recombination detected due to a diversity in evolution rates between the coronavirus lineages (Holmes and Rambaut 2004). Still, recombination occurs frequently in natural populations, with recent sequence analyses of SARS-like coronaviruses in horseshoe bats in China detecting high levels of recombination between coronavirus strains and even between strains from varying geographical locations (Lau *et al.* 2010). Whereas preliminary phylogenetic analyses suggest that MERS-CoV has undergone numerous recombination events in its evolutionary history (Dudas and Rambaut 2016), the exact role recombination may have played in the emergence of MERS-CoV should be investigated in more detail. Further sampling of bat coronaviruses will help to determine the contribution of this evolutionary mechanism.

Robustness, on the other hand, is the ability of a phenotype to remain constant or functional in the face of perturbation, whether environmental or genetic (i.e. mutations). Mutational robustness enables a population to explore more sequence space (i.e. sample more mutations) without losing viability, thus potentially yielding novel functions or lifecycle strategies. In terms of host range expansion, these novel functions could include changes within the spike protein that allow the virus to utilize new receptor molecules or evade a new host's immune defenses. While experimentally measuring robustness in virus populations can be

complex, competition experiments in vesicular stomatitis virus (VSV) have shown that a more mutationally robust virus lineage can have an advantage over a faster-replicating lineage when the mutation rate is increased via a mutagen (Sanjuan *et al.* 2007). This work, combined with theoretical studies, suggests that natural selection can favor mutational robustness in viral populations (Wilke *et al.* 2001) and even facilitate adaptation to a novel host under certain conditions (Stern *et al.* 2014).

Mutational robustness has been suggested to play a role in influenza A virus evolution. H3N2 isolates were found to accumulate genetic variation by moving through neutral networks (i.e. mutationally-connected genotypes that produce the same phenotype). These periods of phenotypic constancy were then disrupted by a sudden phenotypic change in the virus antigenic structure. The exploration of sequence space facilitated by a high mutational robustness allowed these changes to accumulate and result in epochal shifts that are well matched by epidemiological models (Koelle *et al.* 2006, Lauring *et al.* 2013). While the H3N2 study found a role for robustness in immune evasion specifically, the same process can apply to host range expansion. A virus population that could explore a neutral network within the context of the spike protein could potentially accumulate changes that increase the likelihood of producing a host range mutant. This process could be particularly applicable to MERS-CoV which may require multiple mutations to expand its host range to currently nonpermissive species, as previously discussed.

Mutational robustness among RNA viruses varies both among distantly related families (Elena *et al.* 2006) and among closely related strains (Sanjuan *et al.* 2007). Mutational robustness has not been measured in coronaviruses, however the higher resistance of SARS-CoV to the mutagen ribavirin (Barnard *et al.* 2006) compared to that of MERS-CoV (Falzarano *et al.*

2013) suggests a difference in mutational robustness between these viruses (Graci *et al.* 2012, Lauring *et al.* 2013). Even if coronaviruses are not robust to mutations during the infection of a host cell by a single virus, coinfection of host cells by two or more viruses can buffer individuals from deleterious mutations. During peak infection (i.e. high viral load), coinfection can occur frequently which enables complementation. When multiple genomes contribute to the same protein pool, mutant genomes can persist because their more fit coinfection partners compensate for their deleterious mutations at the protein level. Experimental evidence has shown that complementation results in increased mutational robustness in the RNA bacteriophage Φ6 (Montville *et al.* 2005, Joseph *et al.* 2014) and the maintenance of lethal mutations in several different types of viruses (Moreno *et al.* 1997, Cicin-Sain *et al.* 2005), including coronaviruses (Kim *et al.* 1997, Stalcup *et al.* 1998). By conferring mutational robustness, coinfection could enable the virus population to sample a broader sequence space, increasing the probability of producing a variant with the ability to infect a new host species.

At the moment, the role of mutational robustness in host range expansion is purely speculative. Although robustness may increase the likelihood of producing a variant with an expanded host range, the success of this variant relies on how well it can adapt to the new host. Since robustness can either hinder or facilitate adaptation depending on various conditions, the role of robustness in host range expansion is complex (Draghi *et al.* 2010, Stern *et al.* 2014). Further experimental studies are needed both to confirm the mutational robustness of coronaviruses and to determine its contribution to host range expansion.

CONCLUSION

The increasing prevalence of emerging pathogens heightens the need to understand the biochemical and evolutionary dynamics of host range expansion events. While coronaviruses have been circulating in the human population for over hundreds of years, the recent emergence of highly pathogenic strains and the lack of effective vaccines or therapeutics emphasizes the need to prioritize these studies. For SARS-CoV and MERS-CoV, the specific factors contributing to the host range expansion are still unclear. Elucidating the selective pressures imposed on the virus populations will help reveal the specific path of emergence for SARS-CoV and MERS-CoV. Additionally, the isolation of MERS-CoV and BtCoV-HKU4 reveals an increased variation in the receptors that coronaviruses can utilize, adding DPP4 to the list of already identified human receptors, APN (hCoV-229E) and ACE2 (SARS-CoV, hCoV-NL63). This discovery shows the enhanced variation and functionality of coronavirus spike proteins, but reduces the expected power of screening potential for emerging coronaviruses based on receptor utilization.

Currently, we know that MERS-CoV permissivity is primarily mediated by the interactions between the host-cell receptor and the MERS-CoV RBD. Other areas of research that require additional investigation include the role of host cell proteases and innate immune antiviral defense pathways, which for the most part are understudied in coronavirus host range expansion research. For MERS-CoV, permissive and nonpermissive DPP4 orthologs have been identified and tested experimentally (de Wit *et al.* 2013a, de Wit *et al.* 2013b, Cockrell *et al.* 2014, Coleman *et al.* 2014, Raj *et al.* 2014, Song *et al.* 2014, van Doremalen *et al.* 2014).

Together, this work has generated a useful toolkit for identifying the specific residues or biochemical characteristics that allow or prevent MERS-CoV infection. These studies can inform the development of small animal models, vaccines, and therapeutics, in addition to revealing the mutational path that allowed MERS-CoV to expand its host range into humans.

Here, I present work that investigated the biochemical determinants of MERS-CoV host range. Upon emergence, it was found that the MERS-CoV host cell receptor was DPP4 and that while it could utilize hDPP4, cDPP4, and bDPP4, it was unable to utilize the DPP4 molecules from traditional small animal models, including mice, ferrets, hamsters, and guinea pigs. With this landscape, understanding the biochemical determinants of infection could not only help in the generation of a small animal model, but also allow us to better understand how MERS-CoV was able to expand its host range into humans.

We found that two amino acids changes in mDPP4 could allow it to support MERS-CoV infection. These two mutations (A288L and T330R) stabilize an important hydrophobic core in the MERS-CoV RBD and remove a sterically hindering glycosylation site, respectively (Figures 1.5 and 1.7). This data were used to generate a transgenic mouse model using the CRISPR/Cas9 genome editing technique. Validation studies of the transgenic mouse model are currently underway. Additionally, these data were used to investigate whether there was a broader signature of permissivity among other nonpermissive DPP4 orthologs. Namely, fDPP4 and haDPP4 also share a nonconserved glycosylation site that could act to block MERS-CoV infection, whereas gpDPP4 has a downstream glycosylation site that is shared with bDPP4 (Figure 2.2A). Removing these glycosylation sites, while important, is not sufficient on its own to confer permissivity (Figure 2.3A). Rather, additional complex determinants mediate MERS-CoV permissivity in these DPP4 orthologs (Figure 2.7).

Although MERS-CoV mouse models have been developed, a more valuable resource would be a mouse-adapted MERS-CoV strain. However, due to the presence of a glycan as a barrier to infection by MERS-CoV, we hypothesize that extensive remodeling of the MERS-CoV RBD would need to take place before it could overcome the mDPP4 glycosylation site. Adaptation experiments using a host invasion strategy (Figures 3.1 and 3.2) did not result in a successful host range variant (Figures 3.6 and 3.7). Because of the population's inability to produce enough genetic variation to sample potential host range variants, I developed a framework for producing a MERS-CoV mutator allele (Figure 4.1). By combining population biology theory and high-throughput sequencing techniques, a low-fitness cost mutator allele can be identified (Figure 4.3) to be used for host range expansion experiments. Namely, can a mutator allele promote host range expansion in the context of coronaviruses? Although a MERS-CoV mutator appears to be more difficult to produce compared to SARS-CoV and MHV (Eckerle *et al.* 2007, Eckerle *et al.* 2010), the coronavirus system can still act as a useful system for studying the impact of mutation rate on host range expansion. Measuring the mutation rate of MERS-CoV and closely related viruses can reveal whether an increased production of genetic variation played a role in its emergence into humans.

While ecological factors play an important role in host range expansion, several inherent biological properties may increase the emergence probability of a viral strain. In addition to mutator alleles, these evolutionary mechanisms can include recombination and mutational robustness. Recombination occurs at high frequencies *in vitro* (Baric *et al.* 1990), suggesting that it has the potential to shape the evolutionary trajectories of coronavirus populations. Recombination has been implicated in the SARS-CoV jump (Rest and Mindell 2003, Li *et al.* 2006), but also see (Holmes 2003, Demogines *et al.* 2012). It has been suggested to play a role in

the evolution of MERS-CoV (Dudas and Rambaut 2016) and this hypothesis will become more resolved as the number of bat coronavirus sequence samples increases. Additionally, high mutational robustness has the potential to allow coronaviruses to tolerate the mutations that would allow them to expand their host range. Evaluating the mutational robustness for SARS-CoV and MERS-CoV can help detail whether robustness plays a role in coronavirus evolution.

While the genetics and biochemistry of coronaviruses have been well studied, we know very little about the underlying evolutionary mechanisms that result in cross-species transmission and emergence into humans. Coronaviruses can act as an important model system for addressing the fundamental principles of host range evolution. The SARS-CoV and MERS-CoV infectious clone systems (Yount *et al.* 2003, Scobey *et al.* 2013) allow robust reverse genetics experiments and the presence of a proofreading mechanism in coronaviruses allows for the production and study of mutator phenotypes. These mutator phenotypes also provide a unique tool for studying robustness, since strains with higher mutation rates are expected to evolve greater robustness than a wildtype strain (Wilke *et al.* 2001). Additionally, SARS-CoV and MERS-CoV have well characterized virus-host cell receptor interactions as well as known human monoclonal antibodies that target the RBDs of each virus. While experimental evolution has been scarce in the coronavirus world, the toolset is present for strong collaborations with evolutionary biologists to help better understand the evolutionary mechanisms that govern highly relevant pathogens. Ideally, understanding the host range expansion of emerging pathogens would enable us to screen circulating virus populations and identify strains that are most likely to emerge into the human population. Investigating the biochemical determinants and evolutionary mechanisms that govern coronavirus host range expansion can allow us to protect public health and better understand the processes which govern viral evolution.

APPENDIX A: PHYLOGENETIC TREE CORONAVIRUS STRAINS

Species	group	Accession number
Avian IBV (AIBV)	3	NC_001451.1
BCoV-SC2013	2c	KJ473821.1
Betacoronavirus HKU24 (BCoV-HKU24)	2a	NC_026011.1
Bovine CoV (BoCoV)	2a	NC_003045.1
BtCoV-133-2005	2c	NC_008315.1
BtCoV-273	2b	DQ648856
BtCoV-279-2005	2b	DQ648857.1
BtCoV-CDPHE15	1b	NC_022103.1
BtCoV-HKU10.H	1b	JQ989273.1
BtCoV-HKU10.R	1b	NC_018871.1
BtCoV-HKU2	1b	NC_009988.1
BtCoV-HKU3	2b	DQ084199
BtCoV-HKU4	2c	NC_009019
BtCoV-HKU5	2c	NC_009020
BtCoV-HKU8	1b	NC_010438.1
BtCoV-HKU9.1	2d	NC_009021
BtCoV-HKU9.5.1	2d	HM211098.1
BtCoV-SL-CoV-WIV1	2b	KF367457.1
Canine CoV A76 (CCoV-A76)	1a	JN856008.2
CoV-Neoromicia	2c	KC869678.4
Equine CoV (ECoV)	2a	EF446615
Feline CoV UU54 (FCoV-UU54)	1a	JN183883.1
Giraffe CoV (GiCoV)	2a	EF424623
HCoV-229E	1b	KF514433.1
HCoV-HKU1	2a	DQ339101
HCoV-NL63	1b	KF530114.1
HCoV-OC43	2a	AY903459
Human Enteric CoV (HECoV-4408)	2a	FJ415324
MERS-CoV	2c	JX869059.2
MHV-A59	2a	AF029248
PEDV	1b	NC_003436.1
PHEV	2a	DQ011855
PRCV	1a	DQ811787.1
Rabbit CoV HKU14 (RCoV-HKU14)	2a	NC_017083.1
Rat CoV (Rt-CoV)	2a	NC_012936
SARS-CoV	2b	AY278491

TGEV Purdue	1a	DQ811789.2
Turkey CoV (TCoV)	3	EU022526.1

APPENDIX B: PHYLOGENETIC TREE DPP4 ORTHOLOGS

Species	Latin name	Genbank	Length
African savanna elephant	Loxodonta africana	XM_010586282.1	2412
Alpaca	Vicugna pacos	XM_006196217.2	3494
Alpine marmot	Marmota marmota marmota	XM_015479253.1	3556
American alligator	Alligator mississippiensis	XM_014599323.1	2781
American pika	Ochotona princeps	XM_004577330.2	3300
Angola colobus	Colobus angolensis palliatus	XM_001947095.1	4058
Arabian camel	Camelus dromedarius	KF574263.1	2301
Bactrian camel	Camelus bactrianus	XM_010971448.1	3376
Bald eagle	Haliaeetus leucocephalus	XM_010569541.1	2982
Barn owl	Tyto alba	XM_009969592.1	1958
Bison	Bison bison bison	XM_010858966.1	3992
Black flying fox	Pteropus alecto	XM_006921123.2	3914
Blackstripe livebearer	Poeciliopsis prolifica	GBYX01453948	2860
Bolivian squirrel monkey	Saimiri boliviensis boliviensis	XM_010329649.1	3275
Brandts bat	Myotis brandtii	XM_0058593272.2	3467
Cattle	Bos taurus	BC102523.1	2497
Cheetah	Acinonyx jubatus	XM_015082793.1	2498
Chicken	Gallus gallus	NM_001031255.2	3300
Chimney swift	Chaetura pelagica	XM_009996287.1	2248
Chimpanzee	Pan troglodytes	XM_016949951.1	4074
Chinese alligator	Alligator sinensis	XM_014526682.1	2875
Chinese hamster	Cricetulus griseus	XM_007610182.2	3517
Chinese soft-shelled turtle	Pelodiscus sinensis	XM_006114841.2	3987
Chinese tree shrew	Tupaia chinensis	XM_006160271.1	2875
Common garter snake	Thamnophis sirtalis	XM_014060153.1	2857
Common Pipistrelle	Pipistrellus pipistrellus	KC249974	2283
Common starling	Sturnus vulgaris	XM_014880883.1	3414
Coquerels sifaka	Propithecus coquereli	XM_012655850.1	3257
Crab-eating macaque	Macaca fascicularis	XM_015432296.1	3245
Damara mole rat	Fukomys damarensis	XM_010606844.1	2718
Davids Myotis (vesper bat)	Myotis davidii	XM_006766490.2	3455
Degu	Octodon degus	XM_004629976.1	2385
Dog	Canis lupus familiaris	XM_014110860.1	3291
Domestic cat	Felis catus	NM_001009838.1	2573
Domestic ferret	Mustela putorius furo	KF574264.1	2298
Domestic guinea pig	Cavia porcellus	XM_013142395.1	3431
Egyptian jerboa	Jaculus jaculus	XM_004651712.1	3319

Egyptian rousette	Rousettus aegyptiacus	XM_016146442.1	3225
European shrew	Sorex araneus	XM_004608172.1	2298
Giant panda	Ailuropoda melanoleuca	XM_011231867.1	3558
Goat	Capra hircus	KF574265.1	2298
Golden eagle	Aquila chrysaetos canadensis	XM_011580051.1	3485
Golden hamster	Mesocricetus auratus	NM_001310571.1	2295
Golden snub-nosed monkey	Rhinopithecus roxellana	XM_010361854.1	3899
Gray mouse lemur	Microcebus murinus	XM_012735717.1	3568
Gray short-tailed opossum	Monodelphis domestica	XM_007494268.2	3764
Green monkey	Chlorocebus sabaeus	XM_007965152.1	3803
Horse	Equus caballus	XM_005601544.2	3432
Humans	Homo sapiens	NM_001935.3	3913
Jamaican fruit-eating bat	Artibeus jamaicensis	KF574262.1	2298
Japanese quail	Coturnix japonica	XR_001556650.1	2252
Killer whale	Orcinus orca	XM_012538174.1	3408
Large flying fox	Pteropus vampyrus	XM_011358549.1	3954
Little brown bat	Myotis lucifugus	XM_014445986.1	3553
Little egret	Egretta garzetta	XM_009640707.1	2324
Long-tailed chinchilla	Chinchilla lanigera	XM_005393319	3887
Mallard	Anas platyrhynchos	XR_001194974.1	1952
Mandrill	Mandrillus leucophaeus	XM_011991008.1	4060
Mas night monkey	Aotus nancymaae	XM_012440525.1	3585
Medium ground-finch	Geospiza fortis	XM_014306574.1	2274
Minke whale	Balaenoptera acutorostrata scammoni	XM_007183309.1	3605
Mouflon sheep	Ovis aries musimon	XM_012169318.2	3993
Mouse	Mus musculus	NM_001159543.1	3654
Naked mole rat	Heterocephalus glaber	XM_013076128.1	2288
Natal long-fingered bat	Miniopterus natalensis	XM_016196976.1	3403
Northern pike	Esox lucius	XM_010890872.1	2561
Northern white-cheeked gibbon	Nomascus leucogenys	XM_003266171.2	4062
Olive baboon	Papio anubis	XM_009182328.1	3641
Ords kangaroo rat	Dipodomys ordii	XM_013029133.1	3555
Pacific walrus	Odobenus rosmarus	XM_004410199.2	3412
Peregrine falcon	Falco peregrinus	XM_013296766.1	2202
Philippine tarsier	Tarsius syrichta	XM_008048884.1	3262
Pig-tailed macaque	Macaca nemestrina	XM_011744180.1	3541
Platypus	Ornithorhynchus anatinus	XM_001512829.2	3300
Prairie deer mouse	Peromyscus maniculatus bairdii	XM_006984654.2	3734
Prairie vole	Microtus ochrogaster	XM_013346055.1	3504

Pygmy chimpanzee	Pan paniscus	XM_008974411	3541
Rhesus macaque	Macaca mulatta	NM_001039190.2	3596
Rhinoceros hornbill	Buceros rhinoceros silvestris	XM_010133207.1	2276
Rock pigeon	Columba livia	XM_013367520.1	2444
Saker falcon	Falco cherrug	XM_014282916.1	2202
Sheep	Ovis aries	KF574268.1	2298
Small-eared galago	Otolemur garnettii	XM_012808448.1	2590
Sooty mangabey	Cercocebus atys	XM_012038558.1	3553
Sperm whale	Physeter catadon	XM_007105293.1	3407
Star-nosed mole	Condylura cristata	XM_004674874.2	3490
Sumatran orangutan	Pongo abelii	NM_001132869.1	3411
Swan goose	Anser cygnoides domesticus	XM_013171141.1	2859
Tasmanian devil	Sarcophilus harrisii	XM_003763903.2	2365
Western lowland gorilla	Gorilla gorilla gorilla	XM_004032706.1	4068
White-tailed eagle	Haliaeetus albicilla	XM_009918923.1	2441
White-tufted-ear marmoset	Callithris jacchus	XM_002749392.2	3838
Wild Bactrian camel	Camelus ferus	XM_006176809.2	3389
Wild boar	Sus scrofa	NM_214257.1	2400
Wild yak	Bos mutus	XM_005897282.2	3633
Zebra finch	Taeniopygia guttata	XM_012574811.1	3336
Zebrafish	Danio rerio	NM_00161337.1	2878

REFERENCES

Agnihothram S, Gopal R, Yount BL, Jr, Donaldson, EF, Menachery VD, Graham RL, Scobey TD, Gralinski LE, Dension MR, Zombon M, Baric, RS. 2014. Evaluation of serologic and antigenic relationships between Middle East respiratory syndrom coronavirus and other coronaviruses to develop vaccine platforms for the rapid response to emerging coronaviruses. *J Infect Dis* 209: 995-1006

Agrawal AS, Garron T, Tao X, Peng BH, Wakamiya M, et al. 2015. Generation of Transgenic Mouse Model of Middle East Respiratory Syndrome-Coronavirus Infection and Disease. *J Virol* 89: 3659-70

Alekseev KP, Vlasova AN, Jung K, Hasoksuz M, Zhang X, et al. 2008. Bovine-like coronaviruses isolated from four species of captive wild ruminants are homologous to bovine coronaviruses, based on complete genomic sequences. *J Virol* 82: 12422-31

Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrelio CM, et al. 2013a. A strategy to estimate unknown viral diversity in mammals. *MBio* 4: e00598-13

Anthony SJ, Ojeda-Flores R, Rico-Chavez O, Navarrete-Macias I, Zambrana-Torrelio CM, et al. 2013b. Coronaviruses in bats from Mexico. *J Gen Virol* 94: 1028-38

Baric RS, Fu K, Schaad MC, Stohlman SA. 1990. Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups. *Virology* 177: 646-56

Baric RS, Sullivan E, Hensley L, Yount B, Chen W. 1999. Persistent infection promotes cross-species transmissibility of mouse hepatitis virus. *J Virol* 73: 638-49

Barlan A, Zhao J, Sarkar MK, Li K, McCray PB, Jr., et al. 2014. Receptor variation and susceptibility to Middle East respiratory syndrome coronavirus infection. *J Virol* 88: 4953-61

Barnard DL, Day CW, Bailey K, Heiner M, Montgomery R, et al. 2006. Enhancement of the infectivity of SARS-CoV in BALB/c mice by IMP dehydrogenase inhibitors, including ribavirin. *Antiviral Res* 71: 53-63

Bayliss CD, Hoe JC, Makepeace K, Martin P, Hood DW, and Moxon ER. 2008. Neisseria meningitidis Escape from the Bactericidal Activity of a Monoclonal Antibody Is Mediated by Phase Variation of lgtG and Enhanced by a Mutator Phenotype. *Infect. Immun.* 76: 5038–5048

Bono LM, Gensel CL, Pfennig DW, Burch CL. 2012. Competition and the origins of novelty: experimental evolution of niche-width expansion in a virus. *Biology Letters* 9: 20120616

Boonacker E, Van Noorden CJ. 2003. The multifunctional or moonlighting protein CD26/DPPIV. *Eur J Cell Biol* 82: 53-73

Bromham L, Penny D. 2003. The modern molecular clock. *Nature Reviews Genetics* 4: 216-24

Bull JJ, Meyers LA, Lachmann M. 2005. Quasispecies made simple. *PLoS Computational Biology* 1: e61

Bull JJ, Sanjuan R, Wilke CO. 2007. Theory of lethal mutagenesis for viruses. *J Virol* 81: 2930-39

Bull RA, Eden JS, Rawlinson WD, White PA. 2010. Rapid evolution of pandemic noroviruses of the GII.4 lineage. *PLoS Pathog* 6: e1000831

Cai YY, Yu SQ, Postnikova EN, Mazur S, Bernbaum JG, et al. 2014. CD26/DPP4 Cell-Surface Expression in Bat Cells Correlates with Bat Cell Susceptibility to Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Infection and Evolution of Persistent Infection. *Plos One* 9: e112060

Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T. 2006. Bats: Important reservoir hosts of emerging viruses. *Clinical Microbiology Reviews* 19: 531-45

Chao L, Cox EC. 1983. Competition Between High and Low Mutating Strains of Escherichia coli. *Evolution* 37: 125–134

Cherry JD. 2004. The chronology of the 2002-2003 SARS mini pandemic. *Paediatr Respir Rev* 5: 262-9

Chiou HE, Liu CL, Buttrey MJ, Kuo HP, Liu HW, et al. 2005. Adverse effects of ribavirin and outcome in severe acute respiratory syndrome: experience in two medical centers. *Chest* 128: 263-72

Chow SS, Wilke CO, Ofria C, Lenski RE, Adami C. 2004. Adaptive radiation from resource competition in digital organisms. *Science* 305: 84-86

Cicin-Sain L, Podlech R, Messerle M, Reddehase MJ, Koszinowski UH. 2005. Frequent coinfection of cells explains functional in vivo complementation between cytomegalovirus variants in the multiply infected host. *Journal of Virology* 79: 9492-502

Cockrell AS, Peck KM, Yount BL, Agnihothram SS, Scobey T, et al. 2014. Mouse dipeptidyl peptidase 4 is not a functional receptor for Middle East respiratory syndrome coronavirus infection. *J Virol* 88: 5195-9

Coleman CM, Matthews KL, Goicochea L, Frieman MB. 2014. Wild-type and innate immune-deficient mice are not susceptible to the Middle East respiratory syndrome coronavirus. *J Gen Virol* 95: 408-12

Cotten M, Watson SJ, Zumla AI, Makhdoom HQ, Palser AL, et al. 2014. Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *MBio* 5

Crotty S, Cameron CE, Andino R. 2001. RNA virus error catastrophe: Direct molecular test by using ribavirin. *Proceedings of the National Academy of Sciences of the United States of America* 98: 6895-900

Cui J, Eden JS, Holmes EC, Wang LF. 2013. Adaptive evolution of bat dipeptidyl peptidase 4 (dpp4): implications for the origin and emergence of Middle East respiratory syndrome coronavirus. *Virology Journal* 10: 304

Daszak P, Cunningham AA, Hyatt AD. 2000. Emerging infectious disease of wildlife – threates to biodiversity and human health. *Science* 287: 443-449

de Haan CAM, Li Z, Lintelo ET, Bosch BJ, Haijema BJ, Rottier PJM. 2005. Murine coronavirus with an extended host range uses heparan sulfate as an entry receptor. *Journal of Virology* 79: 14451-56

de Wit E, Prescott J, Baseler L, Bushmaker T, Thomas T, et al. 2013a. The Middle East respiratory syndrome coronavirus (MERS-CoV) does not replicate in Syrian hamsters. *Plos One* 8: e69127

de Wit E, Rasmussen AL, Falzarano D, Bushmaker T, Feldmann F, et al. 2013b. Middle East respiratory syndrome coronavirus (MERS-CoV) causes transient lower respiratory tract infection in rhesus macaques. *Proceedings of the National Academy of Sciences of the United States of America* 110: 16598-603

deFilippis VR, Villarreal LP. 2000. An introduction to the evolutionary ecology of viruses. In *Viral Ecology*, ed. CJ Hurst, pp. 125-208. San Diego: Academic Press

Demogines A, Farzan M, Sawyer SL. 2012. Evidence for ACE2-utilizing coronaviruses (CoVs) related to severe acute respiratory syndrome CoV in bats. *J Virol* 86: 6350-53

Dobson AP. 2005. Virology. What links bats to emerging infectious diseases? *Science* 310: 628-9

Donaldson EF, Haskew AN, Gates JE, Huynh J, Moore CJ, Frieman MB. 2010. Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. *J Virol* 84: 13004-18

Draghi JA, Parsons TL, Wagner GP, Plotkin JB. 2010. Mutational robustness can facilitate adaptation. *Nature* 463: 353-5

Drake, JW. 1993. Rates of spontaneous mutation among RNA viruses. *Proceedings of the National Academy of Sciences* 90: 4171-5

Drexler JF, Gloza-Rausch F, Glende J, Corman VM, Muth D, et al. 2010. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J Virol* 84: 11336-49

Drexler JF, Corman VM, Drosten C. 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. *Antiviral Research* 101: 45-56

Dudas G, Rambaut A. 2016. MERS-CoV recombination: implications about the reservoir and potential for adaptation. *Virus Evolution* 2: vev023

Eckerle LD, Lu X, Sperry SM, Choi L, Denison MR. 2007. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J Virol* 81: 12135-44

Eckerle LD, Becker MM, Halpin RA, Li K, Venter E, et al. 2010. Infidelity of SARS-CoV Nsp14-Exonuclease Mutant Virus Replication Is Revealed by Complete Genome Sequencing. *Plos Pathogens* 6

Eckerle I, Corman VM, Muller MA, Lenk M, Ulrich RG, Drosten C. 2014. Replicative Capacity of MERS Coronavirus in Livestock Cell Lines. *Emerging Infectious Diseases* 20: 276-79

Eigen M. 1993. Viral quasispecies. *Scientific American* 269: 42-9

Elena SF, Carrasco P, Daros JA, Sanjuan R. 2006. Mechanisms of genetic robustness in RNA viruses. *Embo Reports* 7: 168-73

Erles K, Toomey C, Brooks HW, Brownlie J. 2003. Detection of a group 2 coronavirus in dogs with canine infectious respiratory disease. *Virology Journal* 310: 216-23

Esko JD, Bertozzi CR. 2009. Chemical tools for inhibiting glycosylation. In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of glycobiology. 2nd edition. *Cold Spring Harbor* (NY): Cold Spring Harbor Laboratory Press; Chapter 50.

Falzarano D, de Wit E, Rasmussen AL, Feldmann F, Okumura A, et al. 2013. Treatment with interferon-alpha 2b and ribavirin improves outcome in MERS-CoV-infected rhesus macaques. *Nature Medicine* 19: 1313-7

Falzarano D, de Wit E, Feldmann F, Rasmussen AL, Okumura A. 2014. Infection with MERS-CoV Causes Lethal Pneumonia in the Common Marmoset. *Plos Pathogens* 10: e1004250

Gao, F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters Martine, Shaw GM, Sharp PM, Hahn BH. 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* 397: 436-441

Ge XY, Li Y, Yang XL, Zhang HJ, Zhou P, et al. 2012. Metagenomic Analysis of Viruses from Bat Fecal Samples Reveals Many Novel Viruses in Insectivorous Bats in China. *J Virol* 86: 4620-30

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison III CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 6: 343-345

Gierer S, Bertram S, Kaup F, Wrensch F, Heurich A, et al. 2013. The Spike Protein of the Emerging Betacoronavirus EMC Uses a Novel Coronavirus Receptor for Entry, Can Be Activated by TMPRSS2, and Is Targeted by Neutralizing Antibodies. *Journal of Virology* 87: 5502-11

Gnadig NF, Beaucourt S, Campagnola G, Borderia AV, Sanz-Ramos M, et al. 2012. Coxsackievirus B3 mutator strains are attenuated in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 109: E2294-E303

Goes LGB, Ruvalcaba SG, Campos AA, Queiroz LH, de Carvalho C, et al. 2013. Novel Bat Coronaviruses, Brazil and Mexico. *Emerging Infectious Diseases* 19: 1711-13

Gojobori T, Moriyama EN, Kimura M. 1990. Molecular clock of viral evolution and the neutral theory. *Proceedings of the National Academy of Sciences* 87: 10015-8

Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ. 2006. Nidovirales: evolving the largest RNA virus genome. *Virus Research* 117: 17-37

Gossner C, Danielson N, Gervelmeyer A, Berthe F, Faye B, et al. 2014. Human-dromedary camel interactions and the risk of acquiring zoonotic Middle East respiratory syndrome coronavirus infection. *Zoonoses Public Health.* doi: 10.1111/zph.12171

Graci JD, Gnadig NF, Galarraga JE, Castro C, Vignuzzi M, Cameron CE. 2012. Mutational robustness of an RNA virus influences sensitivity to lethal mutagenesis. *J Virol* 86: 2869-73

Graham RL, Becker MM, Eckerle LD, Bolles M, Denison MR, Baric RS. 2012. A live, impaired-fidelity coronavirus vaccine protects in an aged, immunocompromised mouse model of lethal disease. *Nat Med* 18: 1820-6

Graham RL, Donaldson EF, Baric RS. 2013. A decade after SARS: strategies for controlling emerging coronaviruses. *Nat Rev Microbiol* 11: 836-48

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59: 307-21.

Gutierrez-Rivas M, Menendez-Arias L. 2001. A mutation in the primer grip region of HIV-1 reverse transcriptase that confers reduced fidelity of DNA synthesis. *Nucleic Acids Res* 29: 4963-72

Haagmans BL, van den Brand JMA, Provacia LB, Raj VS, Stittelaar KJ, et al. 2015. Asymptomatic Middle East respiratory syndrome coronavirus infection in rabbits. *J Virol.* 89: 6131-35

Hasoksuz M, Alekseev K, Vlasova A, Zhang X, Spiro D, et al. 2007. Biologic, antigenic, and full-length genomic characterization of a bovine-like coronavirus isolated from a giraffe. *J Virol* 81: 4981-90

Hayden F, Croisier A. 2005. Transmission of avian influenza viruses to and between humans. *J Infect Dis* 192: 1311-1314

Hoffman H, Pyrc K, van der Hoek L, Geier M, Berkhout B, Pohlmann S. 2005. Human coronavirus NL63 employs the severe acute respiratory syndrom coronavirus receptor for cellular entry. *Proceedings of the National Academy of Sciences* 102: 7988-93

Holmes KV. 2003. SARS-associated coronavirus. *N Engl J Med* 348: 1948-51

Holmes EC, Rambaut A. 2004. Viral evolution and the emergence of SARS coronavirus. *Philos Trans R Soc Lond B Biol Sci* 359: 1059-65

Howard CR, Fletcher NF. 2012. Emerging virus disease: can we ever expect the unexpected? *Emerging Microbes and Infections* 1: e46

Huang X, Dong W, Milewska A, Qi Y, Zhu QK, Marasco WA, Baric RS, Sims AC, Pyrc K, Li W, Sui J. 2015. Human coronavirus HKU1 spike protein uses O-acetylated sialic acid as an attachment receptor determinant and employs hemagglutinin-esterase protein as a receptor-destroying enzyme. *J Virol* 89: 7202-13

Ithete NL, Stoffberg S, Corman VM, Cottontail VM, Richards LR, et al. 2013. Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. *Emerg Infect Dis* 19: 1697-9

Jenkins GM, et al. 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *Journal of Molecular Evolution* 54: 156-65

Jia W, Karaca K, Parrish CR, Naqi SA. 1995. A Novel Variant of Avian Infectious-Bronchitis Virus Resulting from Recombination among 3 Different Strains. *Archives of Virology* 140: 259-71

Joseph SB, Peck KM, Burch CL. 2014. Dominance effects of deleterious and beneficial mutations in a single gene of the RNA virus ϕ6. *PLoS ONE* 9: e97717

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid mutliple sequence alighment based on fast Fourier transform. *Nucleic Acids Research* (30): 3059-3066

Kim KH, Narayanan K, Makino S. 1997. Assembled coronavirus from complementation of two defective interfering RNAs. *J Virol* 71: 3922-31

Kimura M. 1984. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Kimura, M. 1991. The neutral theory of molecular evolution: a review of recent evidence. *Japanese Journal of Genetics* 66: 367-86

Koelle K, Cobey S, Grenfell B, Pascual M. 2006. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* 314: 1898-903

Korboukh VK, Lee CA, Acevedo A, Vignuzzi M, Xiao Y, et al. 2014. RNA virus population diversity, an optimum for maximal fitness and virulence. *J Biol Chem* 289: 29531-44

Krempl C, Schultze B, Herrler G. 1995. Analysis of cellular receptors for human coronavirus OC43. *Adv Exp Med Biol* 380: 341-4

Lambeir A, Durinx C, Scharpe S, de Meester I. 2003. Dipeptidyl-peptidase IV from bench to bedside: an update on structural properties, functions, and clinical aspects of the enzyme DPPIV. *Critical Reviews in Clinical Laboratory Sciences* 40: 209-294

Lau SK, Li KS, Huang Y, Shek CT, Tse H, et al. 2010. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. *J Virol* 84: 2808-19

Lau SK, Woo PC, Li KS, Tsang AK, Fan RY, et al. 2015. Discovery of a novel coronavirus, China Rattus coronavirus HKU24, from Norway rats supports murine origin of Betacoronavirus 1 with implications on the ancestor of Betacoronavirus lineage A. J. Virol. 89:3076–92

Lauring AS, Frydman J, Andino R. 2013. The role of mutational robustness in RNA virus evolution. *Nat Rev Microbiol* 11: 327-36

Leitner T, Albert J. 1999. The molecular clock of HIV-1 unveiled through analysis of a known transmission history. *Proceedings of the National Academy of Sciences* 96: 10752-7.

Levy MT, McCaughan GW, Abbott CA, Park JE, Cunningham AM, et al. 1999. Fibroblast activation protein: a cell surface dipeptidyl peptidase and gelatinase expressed by stellate cells at the tissue remodelling interface in human cirrhosis. *Hepatology* 29: 1768-78

Li WH, Tanimura M, Sharp PM. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *Journal of Molecular Evolution* 25: 330-42

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: 450-4

Li W, Shi Z, Yu M, Ren W, Smith C, et al. 2005a. Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310: 676-9

Li W, Zhang, C, Sui J, Kuhn JH, Moore MJ, Luo S, Wong S, Huang I, Xu K, Vasilieva N, Murakami A, He Y, Marasco WA, Guan Y, Choe H, Farzan M. 2005b. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO Journal* 24: 1634-1643

Li W, Wong SK, Li F, Kuhn JH, Huang IC, et al. 2006. Animal origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein interactions. *J Virol* 80: 4211-9

Liu X, Yang X, Lee CA, Moustafa IM, Smidansky ED, et al. 2013. Vaccine-derived mutation in motif D of poliovirus RNA-dependent RNA polymerase lowers nucleotide incorporation fidelity. *J Biol Chem* 288: 32753-65

Lu G, Hu Y, Wang Q, Qi J, Gao F, et al. 2013. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* 500: 227-31

Ludwig S, Stitz L, Planz O, Van H, Fitch WM, Scholtissek C. 1995. European swine virus as a possible source for the next influenza pandemic? *Virology* 212: 555-61

MacLean RC, Bell G. 2002. Experimental adaptive radiation in *Psuedomonas*. *The American Naturalist* 160: 569-581

Maeda S, Kamita SG, Kondo A. 1993. Host range expansion of Autographa californica nuclear polyhedrosis virus (NPV) following recombination of a 0.6-kilobase-pair DNA fragment originating from Bombyx mori NPV. *J Virol* 67: 6234-8

Mansky LM, Bernard LC. 2000. 3'-Azido-3'-deoxythymidine (AZT) and AZT-resistant reverse transcriptase can increase the in vivo mutation rate of human immunodeficiency virus type 1. *J Virol.* 74: 9532–9539

Mansky, LM, and Cunningham, KS. 2000. Virus mutators and antimutators: roles in evolution, pathogenesis and emergence. *Trends Genet.* 16: 512–517.

Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, et al. 2003. The Genome sequence of the SARS-associated coronavirus. *Science* 300: 1399-404

McRoy WC, Baric RS. 2008. Amino acid substitutions in the S2 subunit of mouse hepatitis virus variant V51 encode determinants of host range expansion. *J Virol* 82: 1414-24

Memish ZA, Mishra N, Olival KJ, Fagbo SF, Kapoor V, et al. 2013. Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerg Infect Dis* 19: 1819-23

Millet JK, Whittaker GR. 2015. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Res* 202: 120-34

Minskaia E, Hertzig T, Gorbalenya AE, Campanacci V, Cambillau C, et al. 2006. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci U S A* 103: 5108-13

Montville R, Froissart R, Remold SK, Tenaillon O, Turner PE. 2005. Evolution of mutational robustness in an RNA virus. *PLoS Biol* 3: e381

Moreno IM, Malpica JM, Rodriguez-Cerezo E, Garcia-Arenal F. 1997. A mutation in tomato aspermy cucumovirus that abolishes cell-to-cell movement is maintained to high levels in the viral RNA population by complementation. *J Virol* 71: 9157-62

Morley VJ, Mendiola SY, Turner PE. 2015. Rate of novel host invasion affects adaptability of evolving RNA virus lineages. *Proc R Soc B* 282: 1813

Muller MA, Raj VS, Muth D, Meyer B, Kallies S, et al. 2012. Human coronavirus EMC does not require the SARS-coronavirus receptor and maintains broad replicative capability in mammalian cell lines. *MBio* 3

Mulloy B, Hart GW, Stanley P. Structural Analysis of Glycans. In: Varki A, Cummings RD, Esko JD, et al., editors. 2009. Essentials of Glycobiology. 2nd edition. *Cold Spring Harbor* (NY): Cold Spring Harbor Laboratory Press; Chapter 47.

Nichol ST, Rowe JE, Fitch WM. 1993. Punctuate equilibrium and positive Darwinian evolution in vesicular stomatitis virus. *Proc Natl Acad Sci U S A* 90: 10424-28

Nishiyama Y, Yoshida S, Tsurumi T, Yamamoto N, Maeno K. 1985. Drug-resistant mutants of herpes simplex virus type 2 with a mutator or antimutator phenotype. *Micro and Immuno* 29: 377-81

Ohnuma K, Haagmans BL, Hatano R, Raj VS, Mou H, et al. 2013. Inhibition of Middle East respiratory syndrome coronavirus infection by anti-CD26 monoclonal antibody. *J Virol* 87: 13892-9

Peck, KM, Chan CHS, Tanaka MM. 2015a. Connecting within-host dynamics to the rate of molecular evolution in viruses. *Virus Evolution* 1: vev013

Peck, KM, Cockrell AS, Yount BL, Scobey R, Baric RS, Heise MT. 2015b. Glycosylation plays a role in inhibiting MERS-Coronavirus infection for mouse DPP4. *J Virol* 89: 4696-4699.

Peck, KM, Burch CL, Heise MT, Baric RS. 2015c. Middle East respiratory syndrome coronavirus host range expansion: biochemical mechanisms and evolutionary perspectives. *Annual Review of Virology* 2: 95-117

Perlman S, Netland J. 2009. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol* 7: 439-50

Pfeiffer JK, Kirkegaard K. 2003. A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. *Proc. Natl. Acad. Sci*. 100: 7289–7294

Plowright RK, Eby P, Hudson PJ, Smith IL, Westcott D, et al. 2015. Ecological dynamics of emerging bat virus spillover. *Proc Biol Sci* 282: 20142124

Pyrc K, Dijkman R, Deng L, Jebbink MF, Ross HA, et al. 2006. Mosaic structure of human coronavirus NL63, one thousand years of evolution. *J Mol Biol* 364: 964-73

Rainey PB, Travisano M. 1998. Adaptive radiation in a hetergenous environment. *Nature* 394: 69-72

Raj VS, Mou H, Smits SL, Dekkers DH, Muller MA, et al. 2013. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 495: 251-4

Raj VS, Smits SL, Provacia LB, van den Brand JM, Wiersma L, et al. 2014. Adenosine deaminase acts as a natural antagonist for dipeptidyl peptidase 4-mediated entry of the Middle East respiratory syndrome coronavirus. *J Virol* 88: 1834-8

Rasmussen HB, Branner S, Wiberg FC, Wagtmann N. 2003. Crystal structure of human DPP-IV/CD26 in complex with a substrate analogue. *Nature Structural Biology* 10: 19-25

Ren W, Qu X, Li W, Han Z, Yu M, et al. 2008. Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. *J Virol* 82: 1899-907

Rest JS, Mindell DP. 2003. SARS associated coronavirus has a recombinant polymerase and coronaviruses have a history of host-shifting. *Infect Genet Evol* 3: 219-25

Reusken CBEM, Haagmans BL,M¨ ullerMA, Gutierrez C, Godeke GJ, et al. 2013. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *Lancet Infect. Dis.* 13:859–66

Roberts A, Deming D, Paddock CD, Cheng A, Yount B, et al. 2007. A mouse-adapted SARS-coronavirus causes disease and mortality in BALB/c mice. *PLoS Pathog* 3: e5

Robertson DL, Sharp PM, McCutchan FE, Hahn BH. 1995. Recombination in HIV-1. *Nature* 374: 124-6

Rohl CA, Strauss CEM, Misura KMS, Baker D. 2004. Protein structure prediction using Rosetta. *Methods in Enzymology* 383: 66–93

Rozen-Gagnon K, Stapleford KA, Mongelli V, Blanc H, Failloux AB, et al. 2014. Alphavirus mutator variants present host-specific defects and attenuation in mammalian and insect models. *PLoS Pathog* 10: e1003877

Rupprecht CE, Hanlon CA, Hemachudha T. 2002. Rabies re-examined. *Lancet Infect Dis* 2: 327-43

Sanjuan R, Moya A, Elena SF. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci U S A* 101: 8396-401

Sanjuan R, Cuevas JM, Furio V, Holmes EC, Moya A. 2007. Selection for robustness in mutagenized RNA viruses. *PLoS Genet* 3: e93

Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. 2010. Viral mutation rates. *J Virol* 84: 9733-48

Sanjuan R. 2012. From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. *PLoS Pathog* 8: e1002685

Schoelz JE, Wintermantel WM. 1993. Expansion of Viral Host Range through Complementation and Recombination in Transgenic Plants. *Plant Cell* 5: 1669-79

Scholtissek C, Koennecke I, Rott R. 1978. Host range recombinants of fowl plague (influenza A) virus. *Virology* 91: 79-85

Scobey T, Yount BL, Sims AC, Donaldson EF, Agnihothram SS, et al. 2013. Reverse genetics with a full-length infectious cDNA of the Middle East respiratory syndrome coronavirus. *Proc Natl Acad Sci U S A* 110: 16157-62

Scull MA, Gillim-Ross, Santos C, Roberts KL, Bordonali E, Subbarao K, Barclay WS, Pickles RJ. 2009. Avian influenza virus glycoproteins restrict virus replication and spread through human airway epithelium at temperatures of the proximal airways. *PLoS Pathogens* 5: e1000424

Shackelton LA, Parrish CR, Truyen U, Holmes EC. 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proc Natl Acad Sci U S A* 102: 379-84

Sheahan T, Rockx B, Donaldson E, Sims A, Pickles R, et al. 2008. Mechanisms of zoonotic severe acute respiratory syndrome coronavirus host range expansion in human airway epithelium. *J Virol* 82: 2274-85

Shaver AC, Dombrowski PG, Sweeney JY, Treis T, Zappala RM, Sniegowski PD. 2002. Fitness evolution and the rise of mutator alleles in experimental Escherichia coli populations. *Genetics* 162: 557-66

Shirato K, Kawase M, Matsuyama S. 2013. Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease TMPRSS2. *J Virol* 87: 12552-61

Sierra S, Davila M, Lowenstein PR, Domingo E. 2000. Response of foot-and-mouth disease virus to increased mutagenesis: influence of viral load and fitness in loss of infectivity. *J Virol* 74: 8316-23

Smith EC, Blanc H, Vignuzzi M, Denison MR. 2013. Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog* 9: e1003565

Smith, EC, Sexton, NR, and Denison, MR. 2014. Thinking outside of the triangle: replication fidelity of the largest RNA viruses. *Annu Rev of Virol*. 1: 111-32

Sniegowski PD, Gerrish PJ, Lenski RE. 1997. Evolution of high mutation rates in experimental populations of E. coli. *Nature* 387: 703–705

Sniegowski PD, Gerrish PJ, Johnson T, Shaver A. 2000. The evolution of mutation rates: separating causes from consequences. *Bioessays* 22: 1057-66

Song W, Wang Y, Wang N, Wang D, Guo J, et al. 2014. Identification of residues on human receptor DPP4 critical for MERS-CoV binding and entry. *Virology* 471-473: 49-53

Stalcup RP, Baric RS, Leibowitz JL. 1998. Genetic complementation among three panels of mouse hepatitis virus gene 1 mutants. *Virology* 241: 112-21

Stavrinides J, Guttman DS. 2004. Mosaic evolution of the severe acute respiratory syndrome coronavirus. *J Virol* 78: 76-82

Stech J, Xiong X, Scholtissek C, Webster RG. 1999. Independence of evolutionary and mutational rates after transmission of avian influenza viruses to swine. *J Virol* 73: 1878-84

Stern A, Bianco S, Yeh MT, Wright C, Butcher K, et al. 2014. Costs and benefits of mutational robustness in RNA viruses. *Cell Rep* 8: 1026-36

Suarez P, Valcarcel J, Ortin J. 1992. Heterogeneity of the mutation rates of influenza A viruses: isolation of mutator mutants. *J Virol* 66: 2491-4

Suttle, CA. 2007. Marine viruses – major players in the global ecosystem. *Nature Reviews Microbiology* 5: 801-812

Taddei F, Radman M, Maynard-Smith J, Toupance B, Gouyon PH, Godelle B. 1997. Role of mutator alleles in adaptive evolution. *Nature* 387: 700-2

Takasawa W, Ohnuma K, Hatano R, Endo Y, Dang NH, Morimoto C. 2010. Inhibition of dipeptidyl peptidase 4 regulates microvascular endothelial growth induced by inflammatory cytokines. *Biochem Biophys Res Commun* 401: 7-12

Tamura, K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30: 2725-29

Tanaka MM, Bergstrom CT, Levin BR. 2003. The evolution of mutator genes in bacterial populations: the roles of environmental change and timing. *Genetics* 164: 843-54

Tang XC, Zhang JX, Zhang SY, Wang P, Fan XH, et al. 2006. Prevalence and genetic diversity of coronaviruses in bats from China. *J Virol* 80: 7481-90

Taylor LH, Latham SM, Woolhouse MEJ. 2001. Risk factors for human disease emergence. *Phil Trans R Soc Lond* 356: 983-989

van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, et al. 2012. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio* 3: e00473-12

van Doremalen N, Miazgowicz KL, Milne-Price S, Bushmaker T, Robertson S, et al. 2014. Host species restriction of Middle East respiratory syndrome coronavirus through its receptor, dipeptidyl peptidase 4. *J Virol* 88: 9220-32

van Doremalen N, Miazgowicz KL, Munster VJ. 2016. Mapping the specific amino acid residues that make hamster DPP4 functional as a receptor for Middle East respiratory syndrome coronavirus. *J Virol* 90: 5499-5502

Vijgen L, Keyaerts E, Moes E, Thoelen I, Wollants E, et al. 2005. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *J Virol* 79:1595–604

Wang LF, Eaton BT. 2007. Bats, Civets and the emergence of SARS. *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission* 315: 325-44

Wang N, Shi X, Jiang L, Zhang S, Wang D, et al. 2013. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res* 23: 986-93

Wang Q, Qi J, Yuan Y, Xuan Y, Han P, et al. 2014. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* 16: 328-37

Wang W, Lin XD, Guo WP, Zhou RH, Wang MR, et al. 2015. Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. *Virology* 474: 19-27

Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. 1992. Evolution and ecology of influenza A viruses. *Microbiol Rev* 56: 152-79

Wentworth DE, and Holmes KV. 2001. Molecular determinants of species specificity in the coronavirus receptor aminopeptidase N (CD13): influence of N-linked glycosylation. *J Virol* 75: 9741–52

WHO (World Health Organization) 2016. *Middle East respiratory syndrome coronavirus (MERS-CoV).* Nov. 11, WHO, Geneva. http://www.who.int/emergencies/mers-cov/en/

Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C. 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* 412: 331-3

Woo PCY, Lau SKP, Huang Y, Yuen KY. 2009. Coronavirus Diversity, Phylogeny and Interspecies Jumping. *Experimental Biology and Medicine* 234: 1117-27

Woolhouse MEJ. 2002. Population biology of emerging and re-emerging pathogens. *Trends in Microbiology* 10: S3-S7

Woolhouse MEJ, Gowtage-Sequeria S. 2005. Host Range and Emerging and Reemerging Pathogens. *Emerging Infectious Diseases* 11: 1842–1847

Woolhouse MEJ, Haydon DT, Antia R. 2005. Emerging pathogens: the epidemiology and evolution of species jumps. *Trends in Ecology and Evolution* 20: 238-244

Worobey M, Holmes EC. 1999. Evolutionary aspects of recombination in RNA viruses. *Journal of General Virology* 80: 2535-43

Wu KL, Peng GQ, Wilken M, Geraghty RJ, Li F. 2012a. Mechanisms of Host Receptor Adaptation by Severe Acute Respiratory Syndrome Coronavirus. *J of Biol Chem* 287: 8904-11

Wu ZQ, Ren XW, Yang L, Hu YF, Yang J, et al. 2012b. Virome Analysis for Identification of Novel Mammalian Viruses in Bat Species from Chinese Provinces. *J Virol* 86: 10999-1012

Xia S, Liu Q, Wang Q, Sun Z, Su S, et al. 2014. Middle East respiratory syndrome coronavirus (MERS-CoV) entry inhibitors targeting spike protein. *Virus Res* 194: 200-10

Xie X, Wang H, Zeng J, Li C, Zhou G, et al. 2014. Foot-and-mouth disease virus low-fidelity polymerase mutants are attenuated. *Arch Virol* 159: 2641-50

Yang Y, Du LY, Liu C, Wang LL, Ma CQ, et al. 2014. Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. *Proceedings of the National Academy of Sciences of the United States of America* 111: 12516-21

Yeager CL, Ashmun RA, Williams RK, Cardellichio CB, Shapiro LH, Look AT, Holmes KV. 1992. Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* 357: 420-2

Yount B, Curtis KM, Fritz EA, Hensley LE, Jahrling PB, et al. 2003. Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus.

Proceedings of the National Academy of Sciences of the United States of America 100: 12995-3000

Zeng QH, Langereis MA, van Vliet ALW, Huizinga EG, de Groot RJ. 2008. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proceedings of the National Academy of Sciences of the United States of America* 105: 9065-69

Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9: 40.

Zhang H, Gao S, Lercher MJ, Hu S, Chen WH. 2012. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res.* 40: W569-72

Zhao G, Du L, Ma C, Li Y, Li L, Poon VKM, Wang L, et al. 2013. A safe and convenient pseudovirus-based inhibition assay to detect neutralizing antibodies and screen for viral entry inhibitors against the novel human coronavirus MERS-CoV. *Virology* 10: 266

Zhao JC, Li K, Wohlford-Lenane C, Agnihothram SS, Fett C, et al. 2014. Rapid generation of a mouse model for Middle East respiratory syndrome. *Proceedings of the National Academy of Sciences of the United States of America* 111: 4970-75