UNIVERSITY OF CALIFORNIA SAN DIEGO

Evolution-guided Discovery of Species-specific Viral Protease Targets

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biology

by

Brian Vay Tsu

Committee in charge:

Professor Matthew D. Daugherty, Chair
Professor Justin R. Meyer, Co-Chair
Professor Pascal Gagneux
Professor Emily R. Troemel
Professor Joel O. Wertheim

2021

The dissertation of Brian Vay Tsu is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my family:

My parents, Kathy and Ronald Tsu,

my siblings, Richard Tsu, Sally Tran and Amy Tsu,

my siblings-in-law Clarice Tsu and Dennis Tran, Ryan Nishioka

To my dearest friends and colleagues:

Andrew Ryan, Tim Shaw, Hannah Tsunemoto,

Christopher Beierschmitt and Corrina Elder

And lastly, in memory of my role model:

Patricia Hadley

I am truly grateful for all their support and all their sacrifices that have allowed me to be the person and scientist I am today.

TABLE OF CONTENTS

LIST OF SUPPLEMENTARY FILES

Supplemental-Files.zip. Contains Supplementary files 3.1 through 4.1.

Supplementary file 3.1. Training set of enteroviral polyproteins.

Supplementary file 3.2. Enteroviral polyproteins with unique 8mer 3Cpro cleavage site concatenations.

Supplementary file 3.3. Un-optimized 3Cpro cleavage motif scores for true positive, false positive and human sites within the enteroviral polyprotein and human training sets.

Supplementary file 3.4. Optimized 3Cpro cleavage motif scores for true positive, false positive and human sites within the enteroviral polyprotein and human training sets.

Supplementary file 3.5. List of primers and gBlocks used.

Supplementary file 3.6. List of antibodies used for immunoblots.

Supplementary file 4.1. Coronaviral polyproteins with unique 8mer 3CLpro cleavage site concatenations.

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Matt D. Daugherty, who supervised my Ph.D. research at UC San Diego. I arrived with some background in computational biology but had little knowledge and experience with molecular techniques. Matt has supported me unequivocally, inspiring me to consider new biological perspectives and maintaining charisma and encouragement throughout my journey to become a scientist.

For building my initial interest into molecular biology, I would like to thank Milton H. Saier, who supervised my undergraduate research. He helped to nurture my curiosity toward the analysis of large sets of sequence data, and highlighted the need to continue building and improving such tools in the field. I would also like to thank Zhongge Zhang and Jonathan Monk of the Bernhard O. Palsson lab for providing me the opportunity to apply molecular and bioinformatics techniques taught in classes to small-scale projects.

Additionally, I would like to thank Joel O. Wertheim at the UC San Diego School of Medicine and Benjamin Murrell, now at the Karolinska Institutet, for providing guidance and computational resources to address novel, large-scale bioinformatics questions, and Patrick S. Mitchell for sharing his knowledge and insight on inflammasome biology, as well as numerous plasmids, reagents and cell lines.

I also would like to thank members of the Daugherty laboratory, members of the San Diego Program in Immunology, and my Ph.D. committee members for helpful suggestions, Julie Pfeiffer for reagents and protocols for enterovirus virus generation, and Russell Vance, Andrew Sandstrom, and members of the Daugherty laboratory for critical reading of the NLRP1 and CARD8 manuscripts. This work was supported by the National Institutes of Health (R35 GM133633), Pew Biomedical Scholars Program and

Hellman Fellows Program to Matthew D. Daugherty, T32 GM007240 to myself, Christopher Beierschmitt and Andrew P. Ryan, a National Science Foundation graduate research fellowship (2019284620) to CB, and a Jane Coffin Childs Memorial Fund for Medical Research postdoctoral fellowship to Patrick S. Mitchell.

Chapter 1 in part is currently being prepared for submission for publication of the material to Frontiers 2021, including co-authors Elizabeth J. Fay, Katelyn T. Nguyen, Miles R. Corley, Bindhu Hosuru, Viviana A. Dominguez and Matthew D. Daugherty. I, Brian V. Tsu, am the co-first author of this paper, alongside Elizabeth J. Fay.

Chapter 2 includes unpublished material co-authored with Ben Murrell, Joel O. Wertheim and Matthew D. Daugherty. I, Brian V. Tsu, am the primary author of this material.

Chapter 3 is published and can be found available in eLife in Immunology and Inflammation Microbiology and Infectious Disease 2021, including co-authors Christopher Beierschmitt, Andrew P. Ryan, Rimjhim Agarwal, Patrick S. Mitchell and Matthew D. Daugherty. I, Brian V. Tsu, am the co-first author of this paper, alongside Christopher Beierschmitt.

Chapter 4 in part is currently being prepared for submission for publication of the material, including co-authors Rimjhim Agarwal, Elizabeth J. Fay, Christopher Beierschmitt, Andrew P. Ryan, Patrick S. Mitchell and Matthew D. Daugherty. I, Brian V. Tsu, am the first author of this paper.

In closing, I would like to thank my family, friends, Shishigami Leona, Inugami Korone and Vanessa Carlton for providing me with more than enough support and inspiration to walk *A Thousand Miles*.

VITA

2014 Bachelor of Science, University of California San Diego

2015 Master of Science, University of California San Diego

2015-2021 Research Assistant, University of California San Diego

2021 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Tsu BV†, Beierschmitt CM†, Ryan AP, Agarwal R, Mitchell PS, Daugherty MD. “Diverse viral proteases activate the NLRP1 inflammasome.” eLife, Jan 2021.

Garrigues JM, Tsu BV, Daugherty MD, Pasquinelli AE. “Diversification of the Caenorhabditis heat shock response by helitron transposable elements.” eLife, Dec 2019.

Saier MH Jr, Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. “The Transporter Classification Database (TCDB): recent advances.” Nucleic acids research, Jan 2016.

Tsu BV, Saier MH Jr. “The LysE Superfamily of Transport Proteins Involved in Cell Physiology and Pathogenesis.” PLoS One, Oct 2015.

PUBLICATIONS IN PREPARATION

Tsu BV†; Fay EJ†; Nguyen KT; Corley MR; Hosuru B; Dominguez VA; Daugherty MD. “Running with scissors: evolutionary conflicts between viral proteases and the host immune system.” Frontiers, 2021.

Tsu BV; Agarwal R, Fay EJ; Beierschmitt CM, Ryan AP, Mitchell PS, Daugherty MD. “Diverse viral proteases activate the CARD8 inflammasome.”

†These authors contributed equally to the indicated work.

FIELDS OF STUDY

Major Field: Biology

Studies in Molecular Biology
Professor Matthew D. Daugherty

ABSTRACT OF THE DISSERTATION

Evolution-guided Discovery of Species-specific Viral Protease Targets

by

Brian Vay Tsu

Doctor of Philosophy in Biology

University of California San Diego, 2021

Professor Matthew Daugherty, Chair
Professor Justin Meyer, Co-Chair

Many pathogens encode proteases that serve to antagonize the host immune system. In particular, viruses with a positive-sense single-stranded RNA genome ((+)ssRNA), including picornaviruses, flaviviruses, and coronaviruses, encode proteases that are not only required for processing the viral polyprotein into functional units but also manipulate crucial host cellular processes through their proteolytic activity.

Because these proteases must coordinate both the cleavage of numerous polyprotein sites and subversion of host immunity, evolution of viral proteases is expected to be highly constrained. Despite this strong evolutionary constraint, these viral proteins and host immune factors are engaged in “evolutionary arms races” that results in diverse protease-host interactions even within closely related species. In some cases, rapid host gene evolution can result in avoidance of cleavage by viral proteases. In other, more recently described cases, hosts can evolve to bait viral proteases into cleaving them using a “tripwire” strategy of immune activation. Such data provide an explanation for why viral polyprotein sites evolve despite such a strong evolutionary constraint and highlight the importance of identifying and characterizing host proteins that are targeted by viral proteases. Moreover, such an evolutionary model provides insight into the changes in molecular functions between viral proteases and host factors, and underscores the role of viral proteases in viral host range, zoonosis and host immune gene evolution. Here, I describe a combined computational and functional approach that guide the discovery of new host targets of viral proteases, including the characterization of two new host innate immunity tripwires that trigger inflammation in response to viral protease activity.

Chapter 1: Running with scissors: evolutionary conflicts between viral proteases and the host immune system

Abstract

Many pathogens encode proteases that serve to antagonize the host immune system. In particular, viruses with a positive-sense single-stranded RNA genome ((+)ssRNA), including picornaviruses, flaviviruses, and coronaviruses, encode proteases that are not only required for processing viral polyproteins into functional units but also manipulate crucial host cellular processes through their proteolytic activity. Because these proteases must cleave numerous polyprotein sites as well as diverse host targets, evolution of these viral proteases is expected to be highly constrained. However, despite this strong evolutionary constraint, mounting evidence suggests that viral proteases such as picornavirus 3C, flavivirus NS3, and coronavirus 3CL, are engaged in molecular 'arms races' with their targeted host factors, resulting in host- and virus-specific determinants of protease cleavage. In cases where protease-mediated cleavage results in host immune inactivation, recurrent host gene evolution can result in avoidance of cleavage by viral proteases. In other cases, such as recently described examples in NLRP1 and CARD8, hosts have evolved 'tripwire' sequences that mimic protease cleavage sites and activate an immune response upon cleavage. In both cases, host evolution may be responsible for driving viral protease evolution, helping explain why viral proteases and polyprotein sites are divergent among related viruses despite such strong evolutionary constraint. Importantly, these evolutionary conflicts result in diverse protease-host interactions even within closely related host and viral species, thereby contributing to host range, zoonotic potential, and pathogenicity of viral infection. Such examples highlight the importance of examining viral protease-host interactions through an evolutionary lens.

Introduction

Positive-sense single-stranded RNA ((+)ssRNA) viruses represent the largest group of RNA viruses, spanning 30 divergent viral families that include important human pathogens in Flaviviridae, Picornaviridae, and Coronaviridae such as dengue virus, poliovirus, and SARS-CoV-2 (Walker et al., 2020). Despite their diversity, many viruses in this group share a common replication strategy: their (+)ssRNA viral genomes are delivered to host cells as a translation-ready mRNA that encodes a multidomain viral polyprotein. Following translation of the viral polyprotein by host ribosomes, one or more embedded viral proteases cleave the polyprotein into individual, functional proteins at numerous sequence-specific positions (Figure 1.1A). Polyprotein cleavage at these specific sites is necessary for sustained virus replication and propagation, making viral proteases an attractive target for development of antiviral therapeutics (Anirudhan, Lee, Cheng, Cooper, & Rong, 2021; Steuber & Hilgenfeld, 2010).

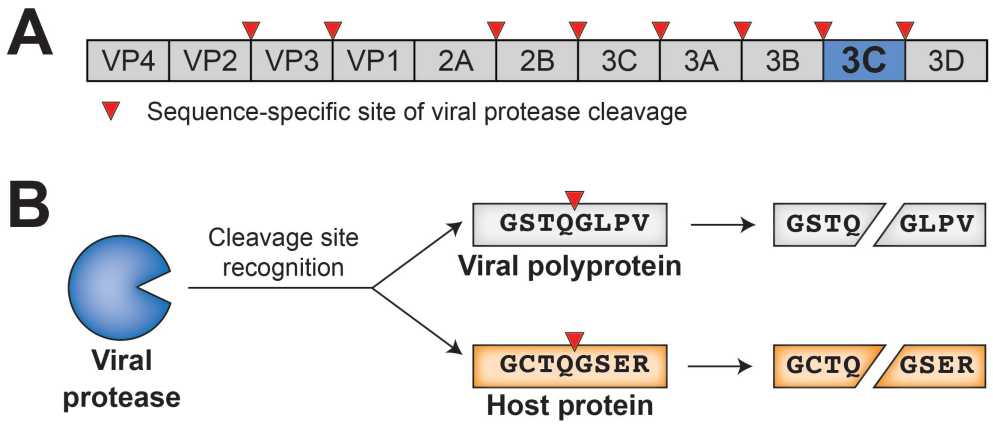


Figure 1.1. Viral proteases cleave specific sites within the viral polyprotein and host proteins. Schematic of an enterovirus (family: Picornaviridae) polyprotein, with the position of the 3C protease and sites of 3C-mediated cleavage shown. (B) 3C protease recognizes and cleaves viral polyprotein sites and host proteins with the same sequence specificity.

In addition to their essential role in the viral life cycle, (+)ssRNA viral proteases also cleave host proteins to manipulate host processes, including the host innate antiviral immune response (Lei & Hilgenfeld, 2017). Importantly, host targets are cleaved with the same sequence specificity as sites within the viral polyprotein (Figure 1.1B). These dual roles place viral proteases at the intersection of two opposing selective pressures. On one side, the virus and its polyprotein site targets are under strong pressure to be conserved, as any changes to the protease sequence specificity or protease sites without concomitant changes to the others would be deleterious for viral fitness. On the other side, viral fitness may be expected to benefit from a protease's ability to adapt to and cleave new host targets, newly evolved sequences in the same host, or divergent sequences in a different host to facilitate cross-species transmission. This type of direct engagement between viral proteases and host factors thus generates an evolutionary conflict where both sides may be driven to adapt in a type of escalating molecular 'arms race' (Daugherty & Malik, 2012; Duggal & Emerman, 2012; Meyerson & Sawyer, 2011; Rothenburg & Brennan, 2020). Such genetic conflicts between viruses and their hosts, particularly those interactions that contribute to potentiation or inhibition of virus replication, lead to immense evolutionary pressure on both parties: hosts are driven to both maintain interactions that activate or carry out antiviral defenses and evade virus interactions that prevent these responses, and viruses are driven to do the opposite. The result is cyclical adaptation of both virus and host to promote either virus replication or host antiviral mechanisms, respectively (Figure 1.2A) (Daugherty & Malik, 2012; Duggal & Emerman, 2012; Meyerson & Sawyer, 2011; Rothenburg & Brennan, 2020). Due to the fact that the direct molecular

host-virus interfaces are those that are being remodeled during such molecular arms races, single amino acid changes can change the outcome of these conflicts (Daugherty & Malik, 2012; Meyerson & Sawyer, 2011). Indeed, traces of these host-virus conflicts can be detected in host genomes by identifying gene codons that show evolutionary signatures of recurrent diversifying (positive) selection (Daugherty & Malik, 2012; Meyerson & Sawyer, 2011; Sironi, Cagliani, Forni, & Clerici, 2015). Importantly, the resulting genetic and molecular changes determine the host range and pathogenesis of viruses, including influencing the ability of viruses to zoonotically transmit into the human population (Meyerson & Sawyer, 2011; Rothenburg & Brennan, 2020; Sawyer & Elde, 2012).

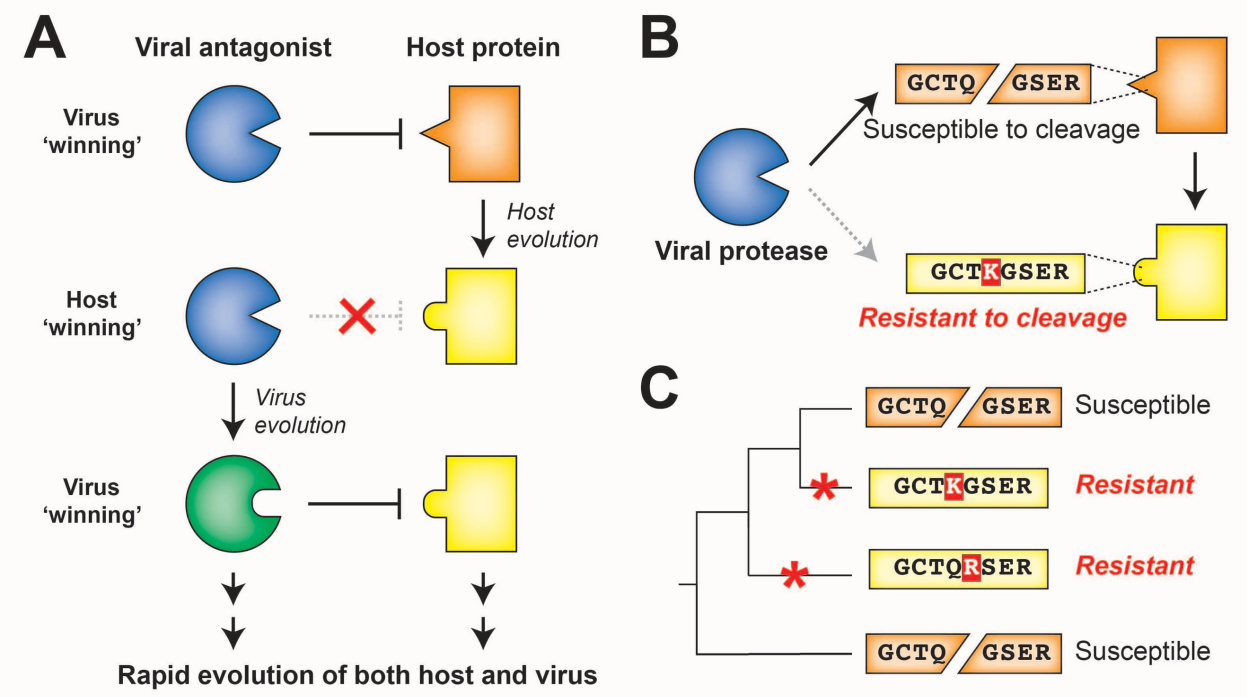


Figure 1.2. Host-virus evolutionary arms races can be driven by protease-target interactions. (A) Host-virus arms races occur when there is direct interaction between host and viral factors, which places evolutionary pressure to select for variants. In this scenario, a viral antagonist recognizes and inactivates a host protein, driving host evolution away from this interaction. The necessity of host target cleavage for virus replication in turn drives evolution of the viral antagonist to reestablish host target recognition. (B) Single amino acid changes in the sequence-specific cleavage motif can eliminate cleavage by a viral protease. (C) Across a phylogenetic tree, changes can occur recurrently resulting in differential susceptibility between even closely related species.

Due to the importance of sequence specificity to protease-host interactions, evolutionary arms races at the interfaces of proteases and their targets would be expected to exist. For instance, a single amino acid change in a targeted host protein at a position that is important for sequence-specific protease cleavage could completely reverse cleavage susceptibility (Figure 1.2B). As a result, single lineage-specific changes at any number of positions in the cleavage motif would be expected to result in rapid toggling of cleavage susceptibility even among closely related hosts, establishing species-specific host-virus interactions that could drive viral host range (Figure 1.2C). Indeed, while a great deal of research on viral proteases has focused on conserved elements of protease function, emerging evidence suggests that both hosts and viruses are evolving in ways that can impact the host- and virus-specificity of cleavage. Here, we review the host-viral molecular conflicts engaged by the main proteases of flaviviruses, picornaviruses, and coronaviruses to emphasize how proteases of (+)ssRNA viruses act as evolutionary drivers of host innate immunity, and how viral proteases are being shaped by these same molecular conflicts. This evolutionary perspective highlights the importance of viral proteases and the host targets they target as being an important determinant of viral host range, tissue tropism and pathogenesis, and zoonotic potential of (+)ssRNA viruses.

Despite evolutionary constraints, main proteases of (+)ssRNA viruses press on

Virus-encoded proteases are essential to the life cycle of numerous (+)ssRNA viruses. Newly synthesized viral polyproteins mature into individual, functional proteins via a series of cleavage events carried out by virus-encoded and host proteases. For

Picornaviridae and Coronaviridae, the cysteine proteases 3C and 3C-Like (3CL), are responsible for the majority of polyprotein processing events (Figure 1.3A-B). Most picornaviruses have six or more 3C cleavage sites throughout the polyprotein (Figure 1.1A), and there is a preference to cleave between a glutamine (Q) in the P1 position and a small residue (e.g. alanine (A) or serine (S)) in the P1' position (Figure 1.3A) (Tsu et al., 2021). Likewise, coronaviruses (CoVs) have ten or more cleavage sites for the 3CL protease (also known as MPro or nsp5 in several CoVs including SARS-CoV-2) (Figure 1.3B) (Roe, Junod, Young, Beachboard, & Stobart, 2021). Numerous other viral families, including members of the Caliciviridae (e.g. norovirus) (Clarke & Lambden, 2000) and Potyviridae (e.g. tobacco etch virus) (Valli, Gallo, Rodamilans, Lopez-Moya, & Garcia, 2018) encode a cysteine protease with a similar specificity for cleavage between a Q and a small residue, whereas members of the Togaviridae (e.g. Chikungunya virus) use a cysteine protease with different cleavage specificity (Ten Dam, Flint, & Ryan, 1999). Other viral families use a serine protease, including Flaviviridae, where the serine protease NS3 processes at least four polyprotein cleavage sites (Figure 1.3C) (S. Chen, Wu, Wang, & Cheng, 2017). Here and in subsequent sections, we will predominantly discuss activities of the 3C, 3CL, and NS3 proteases of Picornaviridae, Coronaviridae, and Flaviviridae, respectively, due to their known roles in cleaving mammalian host factors. It is important to point out that because the polyprotein is sequentially processed, and not always to completion, protease activity may also be carried out when 3C, 3CL, or NS3 remains fused or associated with additional viral proteins (Lindenbach & Rice, 2003; Winston & Boehr, 2021). This is especially true in the Flaviviridae, where the NS3 protease usually

functions in association with NS2B (in the case of flaviviruses such as dengue and Zika viruses) or NS4A (in the case of hepatitis C virus (HCV)). However, for the sake of clarity, we will subsequently only refer to the protease domains of 3C, 3CL, or NS3. Moreover, many (+)ssRNA viruses encode additional proteases involved in both polyprotein processing and host antagonism, including the 2A protease in some picornaviruses and the PLP (papain-like protease) in coronaviruses (Baggen, Thibaut, Strating, & van Kuppeveld, 2018; V'Kovski, Kratzel, Steiner, Stalder, & Thiel, 2021). Finally, viruses other than (+)ssRNA viruses can encode proteases that are important for polyprotein processing, most notably, the retrovirally-encoded aspartyl protease (Navia & McKeever, 1990). While all of these additional proteases from (+)ssRNA viruses and retroviruses play important host antagonism roles, and likely shape host and viral evolution, they will not be extensively explored here.

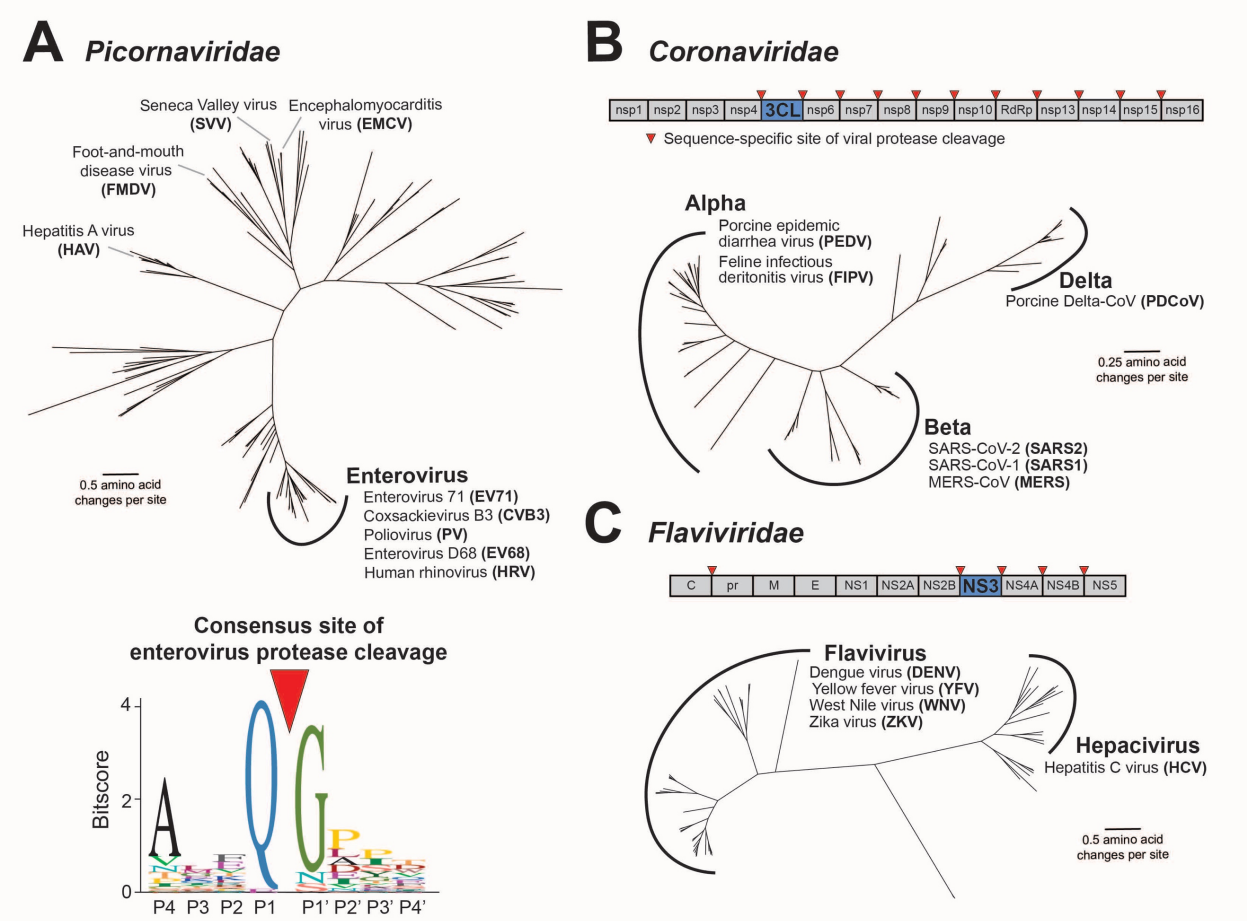


Figure 1.3. Main proteases in *Picornaviridae*, *Coronaviridae*, and *Flaviviridae*. (A) Phylogenetic tree of available RefSeq Picornaviridae 3C protease protein sequences (151 total, top). Names of viruses with human relevance or referenced throughout the text are listed next to their respective genus or singular node. The consensus enterovirus 3C cleavage motif (bottom) as was generated previously (Tsu et al., 2021). The cleavage site is shown flanked by four amino acids upstream (labeled P4 through P1) and four amino acids downstream (labeled P1' through P4'). (B) Schematic of the SARS-CoV-2 (family: Coronaviridae) nonstructural (ORF1ab) polyprotein, with the position of the 3CL protease and sites of 3CL-mediated cleavage shown. Phylogenetic tree of available RefSeq Coronaviridae 3CL protease protein sequences (64 total). Names of viruses with human relevance or referenced throughout the text are listed next to their respective genus. (C) Schematic of the dengue virus (family: Flaviviridae) polyprotein, with the position of the NS3 protease and sites of NS3-mediated cleavage shown. Phylogenetic tree of available RefSeq Flaviviridae NS3 protease protein sequences (68 total). Names of viruses with human relevance or referenced throughout the text are listed next to their respective genus.

The functions of the (+)ssRNA viral proteases described above are, by definition of being required for completion of the viral life cycle, well conserved. In addition to homology between the proteases themselves, the positions and sequences of the polyprotein cleavage motifs are often similar between members of the same viral family.

Indeed, this conservation of polyprotein cleavage motifs has made it possible to compile sequences surrounding the cleavage site from genome sequences alone to generate a consensus motif for the viral protease that can be used to predict host and viral targets (Blom, Hansen, Blaas, & Brunak, 1996; Kiemer, Lund, Brunak, & Blom, 2004; Stanley et al., 2020; Tsu et al., 2021) (Figure 1.3A). These consensus motifs are often generated using many diverse viruses, relying on the assumption that protease sequence specificity is well conserved among virus species. Interestingly, despite the evolutionary constraint to maintain cleavage across multiple sites in the polyprotein, virus-encoded proteases are substantially divergent across viruses (Figure 1.3). For instance, picornavirus 3C proteases can share less than 20% amino acid sequence identity, despite sharing an overall similar fold and many homologous cleavage sites (Tsu et al., 2021). Similar evolutionary distances are observed with other families of proteases, including Coronaviridae 3CL and Flaviviridae NS3 (Figure 1.3).

Even with the divergence of protease sequences, protease sequence specificity is expected to be well conserved within closely related viruses given the essentiality of cleaving multiple site-specific polyprotein sites. Surprisingly, there is mounting evidence that this is not the case. For instance, among closely related serotypes of dengue virus (DENV), biochemical substrate profiling has revealed a subtle but clear shift in the NS3 protease cleavage sequence specificity profile (J. Li et al., 2005). This type of in-depth comparative biochemical analysis of other (+)ssRNA proteases has not been conducted, but assays on model substrates have revealed differences in cleavage specificity even among 3C proteases within the Enterovirus genus of picornaviruses (O'Donoghue et al., 2012). In addition, some of the best evidence that protease

sequence specificity is changing between related viruses has come from studies using chimeric viruses in which the protease of one virus species is inserted into the backbone of another virus. If this results in insufficient or improper cleavage of the polyprotein and reduced viral replication, it would suggest divergence in protease sequence specificity between the parental viruses. For example, within the Enterovirus genus of Picornaviridae (Figure 1.3A), replacing the poliovirus (PV) 3C protease with 3C proteases from human rhinovirus 14 or coxsackievirus B3 (CVB3) resulted in reduced, changed, or loss of cleavage products (Dewalt, Lawson, Colonno, & Semler, 1989). Likewise, within the Flavivirus genus (Figure 1.3C), swapping the protease domain of DENV NS3 for the protease domain of yellow fever virus (YFV) ablates processing of polyproteins containing DENV cleavage sites (Preugschat, Lenches, & Strauss, 1991). Additionally, West Nile virus NS3 can cleave a polyprotein site in only one of two closely related DENV2 strains, where the only difference is in the residue in the P1' position (VanBlargan et al., 2015). While some of these differences may be attributed to the requirement for NS3 proteases to bind to lineage-specific activating cofactors to further augment cleavage specificity (Butkiewicz et al., 2000; Cahour, Falgout, & Lai, 1992; Chambers, Grakoui, & Rice, 1991; Falgout, Miller, & Lai, 1993; Falgout, Pethel, Zhang, & Lai, 1991; Jan, Yang, Trent, Falgout, & Lai, 1995; Sbardellati et al., 2000; Wiskerchen & Collett, 1991; Xu et al., 1997), it is also likely that these changes in cleavage specificity are dependent on non-conserved residues in the binding pocket of the NS3 protease (Junaid et al., 2012). Similarly, within Coronaviridae, replication competent chimeric murine hepatitis virus (MHV) could not be recovered when the 3CL protease was replaced with one of many related alpha- or beta-coronavirus proteases including

SARS-CoV, hCoV-229E and bat CoV-HKU4 (Figure 1.3B). Only when the MHV 3CL was replaced with the two most closely related betacoronaviruses, hCoV-OC43 and hCoV-HKU1, could virus be recovered, but with a substantial fitness cost (Stobart et al., 2013). Altogether, these biochemical and chimeric virus studies illustrate that (+)ssRNA viruses have undergone lineage-specific evolution in both their protease sequence specificity as well as their many polyprotein cleavage sites.

(+)ssRNA viral proteases have evolved in conflict with the host

The above-described changes in protease sequence specificity do not require invocation of adaptation. Indeed, evolutionary drift could result in changes to the viral protease and its cleavage sites, including those that result in loss of fitness for chimeric viruses. However, there is another selective pressure that likely shapes viral protease evolution: the advantage that viruses gain by cleaving host targets. Proteins in multiple cellular processes have been identified as targets of viral proteases, many of which are involved in the host antiviral immune response (Lei & Hilgenfeld, 2017). Many of these host targets are divergent between species, potentially establishing molecular barriers to cross-species transmission. Indeed, pathogenicity of a mouse-adapted SARS coronavirus required two mutations in 3CL to facilitate rapid, robust virus replication (Roberts et al., 2007). Although it has not been established whether these 3CL changes result in changes in host target cleavage, these data indicate that protease evolution may be required for successful adaptation to a novel host species.

Several excellent reviews have been written describing the diverse host targets that are cleaved by viral proteases (Laitinen et al., 2016; Lei & Hilgenfeld, 2017; Ng,

Stobart, & Luo, 2021; Yi et al., 2021). In many cases, the described host-virus interaction has focused on a single or a small number of related viral proteases and only a single host species, often humans. The importance of host and virus diversity in these interactions is often poorly understood. However, evidence is accumulating that viral proteases and their host targets are engaged in species-specific interactions. Below, we highlight such cases in which host and viral diversity alter the outcome of the interaction between host pathways and proteases of picornaviruses, flaviviruses and coronaviruses, illustrating this ongoing molecular arms race.

Viral proteases target essential host processes in a virus-specific manner

Some of the best studied targets of viral proteases, especially from picornaviruses, are involved in well conserved processes such as translation initiation or translation control (Lloyd, 2016; McCormick & Khaperskyy, 2017). In many cases, the functional outcome is similar: viral proteases antagonize a host molecular function in a way that benefits the virus. However, the specific host protein or specific site within that host protein is in some cases divergent between different viruses, highlighting differences in protease cleavage specificity between related viruses, as well as the convergence of viral protease cleavage on the same host pathways. Thus, even for host functions that are ‘well-conserved’ targets of protease cleavage, there is surprising mechanistic diversity.

Translation of picornavirus mRNAs occurs via an internal ribosome entry site (IRES) (Lee, Chen, & Shih, 2017). This bypasses the need to engage with host cap-dependent translation machinery and offers the opportunity to induce a ‘host-shutoff’ of

translation of host antiviral proteins while maintaining production of viral proteins. Many picornaviruses inhibit host translation in a protease-dependent manner via cleavage of subunits of the eIF4F cap-binding complex, which binds to host mRNA cap structures to establish the initiation complex, or poly-A binding protein (PABP), which binds the 3' polyA tail of mRNAs and eIF4G to circularize mRNAs for optimal translation initiation (Jackson, Hellen, & Pestova, 2010) (Figure 1.4A). For instance, Foot-and-mouth disease virus (FMDV) 3C cleaves the eIF4G and eIF4A subunits of eIF4E (Belsham, McInerney, & Ross-Smith, 2000). Interestingly, neither hepatitis A (HepA) virus nor encephalomyocarditis virus (EMCV) 3C target eIF4G for cleavage, but both target PABP (Kobayashi, Arias, Garabedian, Palmenberg, & Mohr, 2012; Zhang, Morace, Gauss-Muller, & Kusov, 2007). Convergently, PV also targets PABP, but at a site that 100 residues away from the cleavage site of EMCV (Bonderoff, Larey, & Lloyd, 2008) and additionally uses its 2A protease to cleave eIF4G (Gradi, Svitkin, Imataka, & Sonenberg, 1998). Despite cleaving different host targets and/or host sites, these interactions all result in host translation shut-off. These data highlight functional conservation, rather than molecular conservation, of picornavirus 3C-mediated inhibition of host translation and suggests that even among related viruses, there are important differences in the viral specificity of host target cleavage.

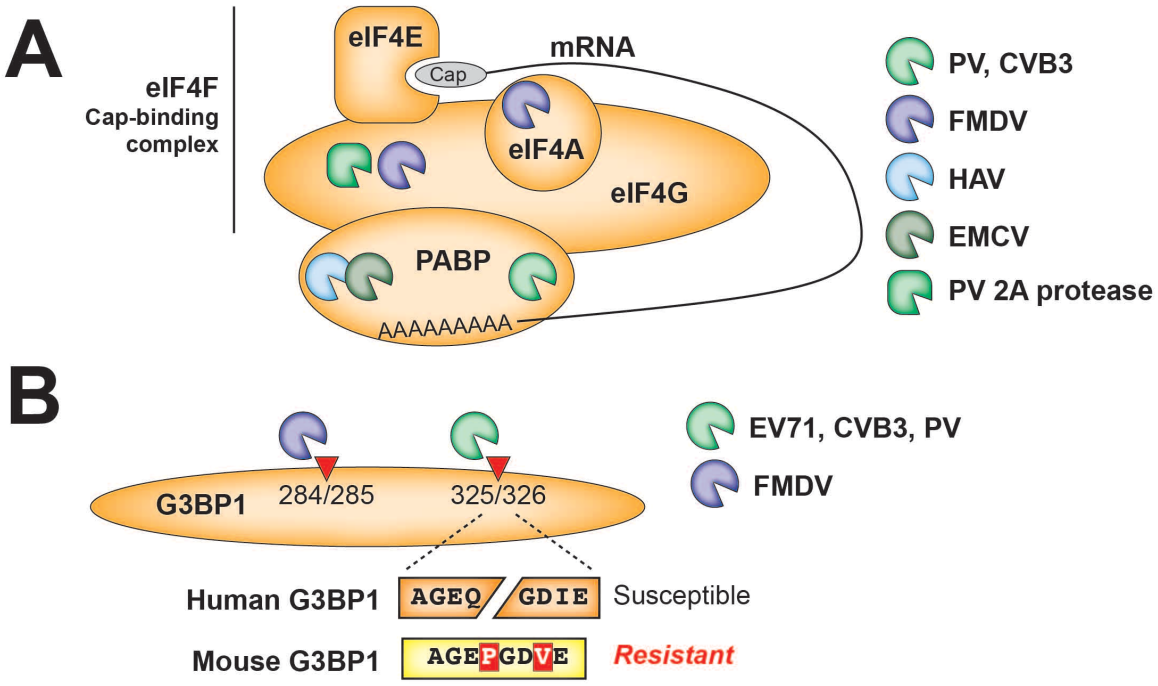


Figure 1.4. Antagonism of host cellular processes by viral proteases. (A) Diverse viral proteases inhibit translation of host mRNA through cleavage of initiation factors and/or poly(A)-binding protein. (B) Host species-specific cleavage of the stress granule protein G3BP1 by picornavirus proteases.

A similar phenomenon is observed in another well-established target of picornavirus 3C proteases, the stress granule protein G3BP1 (Figure 1.4B). Numerous viruses manipulate stress granule formation for their benefit, as this is a major intersection point between translation control and cellular stress responses (McCormick & Khaperskyy, 2017; White & Lloyd, 2012). Among the mapped cleavage sites in G3BP1, PV 3C cleaves at Q326 (White, Cardenas, Marissen, & Lloyd, 2007) while FMDV 3C cleaves at E284 (Ye et al., 2018), but both of these cleavage events benefit the virus by manipulating stress granule formation. These findings further demonstrate the convergence of 3C cleavage onto the same host target, while highlighting that subtle differences in cleavage specificity can impact viral targeting of host factors. Of

note, the P1 positions of both of these cleavage sites are different in ways that would prevent cleavage in the mouse protein, which is otherwise >90% identical to the human protein (Figure 1.4B). Whether host G3BP1 is cleaved by picornaviruses that infect rodents, and at what sites, has yet to be determined.

Proteins in the innate antiviral immune response are common targets of viral proteases

In addition to essential cellular processes, proteins in the innate antiviral immune response are common targets of diverse viral proteases. The host antiviral response is initiated when cells detect viral products (Figure 1.5A, left). Following entry into a host cell, viral nucleic acids can be detected by host pattern recognition receptors (PRRs) such as RIG-I, MDA5, and cGAS. While RIG-I and MDA5 directly detect viral dsRNA as a product of (+)ssRNA virus replication (Deddouche et al., 2014; Kato et al., 2006; Loo et al., 2008; Luo et al., 2008; van Kasteren et al., 2012), the cytosolic DNA sensor cGAS can be indirectly activated via virus-induced mitochondrial damage and subsequent release of mitochondrial DNA (West et al., 2015). After ligand binding, PRRs recruit a series of adaptor proteins, ultimately resulting in the production and secretion of type I and III interferons (IFN-I and IFN-III) (Hoffmann, Schneider, & Rice, 2015). IFN-I and IFN-III are antiviral cytokines that signal in an autocrine or paracrine manner to induce expression of interferon-stimulated genes (ISGs), which act to directly and indirectly inhibit virus replication and establish an antiviral state in the host (Schoggins, 2014) (Figure 1.5A, right).

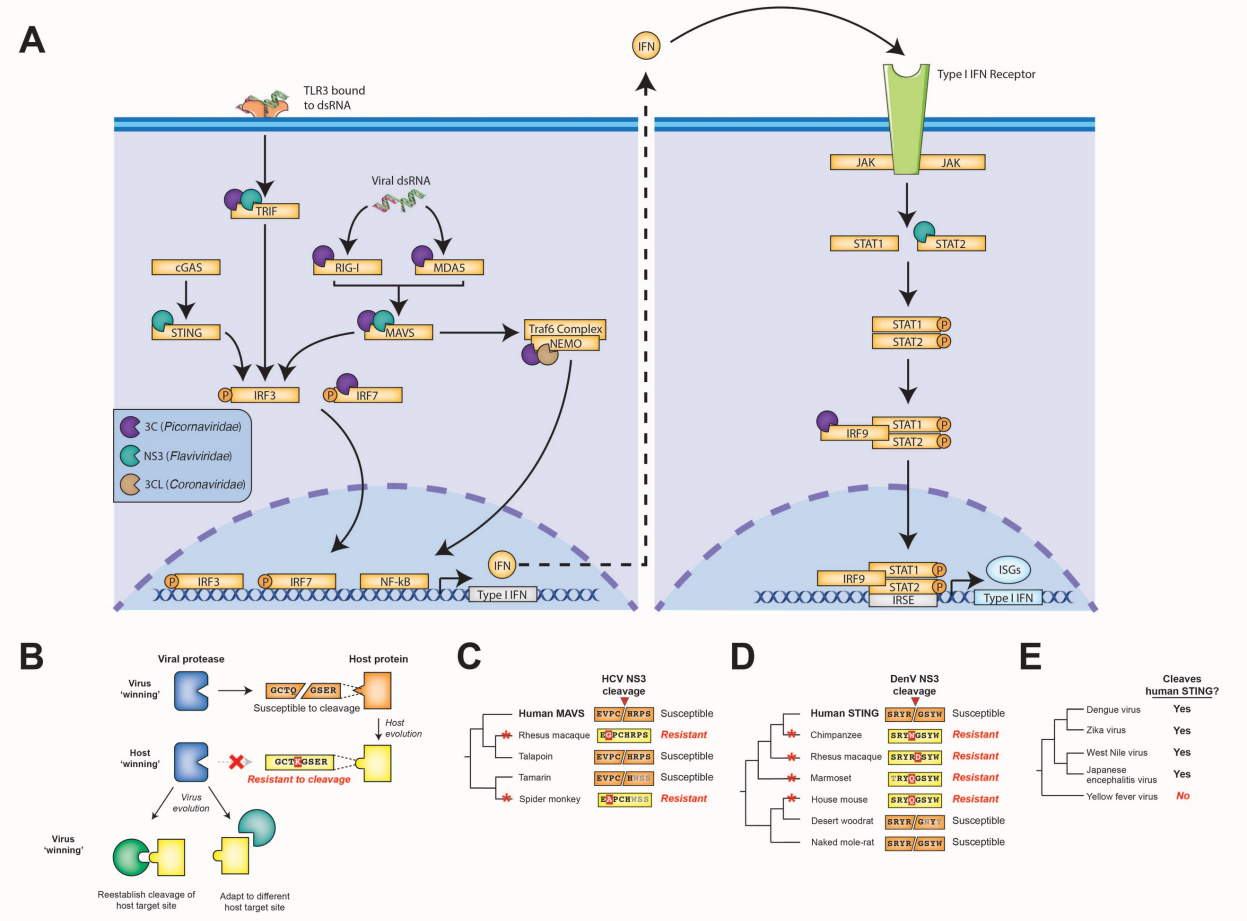


Figure 1.5. Protease antagonism of IFN induction and signaling pathways. (A) Viral protease antagonism of induction (left) or signaling (right) of IFN. (B) Model for how protease sequence specificity may be driven to evolve by conflicts with host factors. Following host evolution, or cross-species transmission, viral proteases may no longer be able to antagonize a given host factor. To re-establish host antagonism, the protease can evolve to cut a different sequence at same host site (left) or may evolve to cut a new site elsewhere in the host protein (right). (C-D) Evolution of MAVS (Patel et al., 2012) (C) and STING (Stabell et al., 2018) (D) across primates and other mammals confers resistance or susceptibility to flaviviral protease cleavage. (E) Human STING cleavage by flavivirus NS3 proteases is virus species-specific. Data adapted from (Ding et al., 2018).

Induction of IFN and subsequent upregulation of ISGs is critical to the host antiviral defense. Therefore, proteins involved in these pathways are common targets of viral antagonism (Hoffmann et al., 2015), including several that are cleaved by (+)ssRNA viral proteases (Table 1.1, Figure 1.5A). For instance, NS3 from dengue virus and other flaviviruses can cleave and inactivate STING to prevent sensing of cytoplasmic mitochondrial DNA (Aguirre et al., 2012; Stabell et al., 2018), whereas PV

and possibly other 3C proteases cleave RIG-I during infection (Barral, Sarkar, Fisher, & Racaniello, 2009; Papon et al., 2009). Tellingly, many proteases convergently cleave the same host targets. For instance, CVB3 3C and HCV NS3 are both able to cleave MAVS (Meylan et al., 2005; Mukherjee et al., 2011), a critical innate immune adaptor for both MDA5 and RIG-I. 3CL proteases from Porcine Epidemic Diarrhea Virus (PEDV), porcine deltacoronavirus (PDCoV), and feline infectious peritonitis virus (FIPV), as well as 3C proteases from FMDV and HepA can also inhibit RIG-I/MDA5 pathways by cleaving nuclear transcription factor κB (NF-κB) essential modulator (NEMO), a bridging adaptor protein involved in activating both NF-κB and interferon-regulatory factor signaling pathway (J. Chen et al., 2019; S. Chen et al., 2019; D. Wang et al., 2012; D. Wang et al., 2016; D. Wang et al., 2014; Zhu, Fang, et al., 2017). Finally, STAT2, one of the critical transcription factors that transmits the signaling of IFN to ISG production, is cleaved by the 3CL from PDCoV (Zhu, Wang, et al., 2017), although whether other 3CL's cleave this protein is unknown. Altogether these data show that viral protease-mediated cleavage of innate immune signaling proteins is a common strategy across (+)ssRNA viruses to prevent the antiviral response and promote virus replication.

Table 1.1. List of known IFN pathway-related targets of (+)ssRNA virus proteases.

Host target	Viral protease	References
STING	NS2B3 (ZIKV, JEV, WNV, YFV, DENV)	(Aguirre et al., 2012; Ding et al., 2018; Stabell et al., 2018)
RIG-I	3C (PV)	(Barral et al., 2009)
MAVS	NS3-4A (HCV, GBV-B) 3C (CVB3, SVV)	(Anggakusuma et al., 2016; Z. Chen et al., 2007; Ferreira et al., 2016; Meylan et al., 2005; Mukherjee et al., 2011; Patel et al., 2012; Qian et al., 2017)
Riplet	NS3-4A (HCV)	(Vazquez, Tan, & Horner, 2019)
MDA5	3C (FMDV)	(Kim et al., 2021)
STAT2	3CL (PDCoV)	(Zhu, Wang, et al., 2017)
TRIF	NS3-4A (HCV) 3C (CVB3, SVV, EV68) 3CD (HAV)	(K. Li et al., 2005; Mukherjee et al., 2011; Qian et al., 2017; Qu et al., 2011; Xiang et al., 2014)
NEMO	3CL (PEDV, FIPV, PDCoV) 3C (FMDV, HAV)	(S. Chen et al., 2019; D. Wang et al., 2012; D. Wang et al., 2016; D. Wang et al., 2014; Zhu, Fang, et al., 2017)
IRF7	3C (EV68)	(Xiang et al., 2016)
IRF9	3C (EV71)	(Hung et al., 2011)

Many proteins in the innate antiviral immune response are rapidly evolving within and between host populations (Enard, Cai, Gwennap, & Petrov, 2016; Judd, Gilchrist, Meyerson, & Sawyer, 2021; Shultz & Sackton, 2019). One consequence of these host changes is the potential that a cleavage site for a viral protease may be present in one host but not another. If there is strong selection for the virus to restore antagonism of that host function, there would be selection for viral proteases that would change the sequence specificity of host target cleavage to either restore cleavage of the original

site, cleave another site on the host protein, or cleave another protein in the host pathway (Figure 1.5B). Such an evolutionary model can be used to understand the genetic bases for host- and viral-specificity of protease cleavage. For many of the known interactions between host immunity proteins and viral proteases, there is little information on how host and viral evolution shapes the outcome. However, analyses on two host targets, described in more detail below, provide evidence for an arms races between host immunity proteins and viral proteases.

MAVS and STING have evolved in conflict with viral proteases

One well-characterized instance of viral proteases shaping host gene evolution is in HCV NS3 protease antagonism of the host protein MAVS. MAVS serves as a critical signaling node to integrate signals from the nucleic acid sensors RIG-I and MDA5 to downstream IFN production (Figure 1.5A). Early observations indicated that MAVS cleavage by HCV NS3 was site specific and important for viral evasion of the immune system (X. D. Li, Sun, Seth, Pineda, & Chen, 2005). Subsequent evolutionary analyses revealed that one residue within the HCV cleavage site in MAVS has evolved under recurrent positive selection, suggestive that MAVS evolution has been shaped by NS3 antagonism (Patel, Loo, Horner, Gale, & Malik, 2012). Indeed, variation at this site across primates affects susceptibility to cleavage by HCV NS3. Importantly, primate MAVS proteins that have evolved resistance to cleavage retain a functional IFN response during HCV infection, providing a potential explanation for the restricted host range of HCV (Figure 1.5C) (Patel et al., 2012). Interestingly, this work also identified a site evolving under positive selection that is known to be antagonized by the CVB3 3C

protease, and variation at this site across primates could also alter protease-mediated antagonism and antiviral signaling through MAVS (Mukherjee et al., 2011; Patel et al., 2012) (Figure 1.5C).

Another adaptor protein that connects nucleic acid sensing to the IFN response is STING, which operates downstream of the cytoplasmic DNA sensor cGAS. Originally described as a species-specific target of dengue virus NS3 cleavage (Aguirre et al., 2012; Stabell et al., 2018), STING has evolved under positive selection in primates and the NS3 cleavage site within STING contains several amino acid differences across primates that alter the outcome of cleavage (Figure 1.5D) (Stabell et al., 2018). Expanding this analysis to a broader panel of mammals, site of cleavage in human STING has evolved to be cleavage resistant in mice, pigs, and ground squirrels, whereas naked mole rat and desert woodrat are susceptible to cleavage (Stabell et al., 2018) (Figure 1.5D). Interestingly, differences in protein sequences that affect cleavage do not just occur between host species; polymorphisms within a host can also alter the ability of a viral protease to cleave a given target. Evidence of this process can be observed in human STING polymorphisms, where the three most common human STING haplotypes are differentially cleaved by DENV NS3 (Su et al., 2020). Not only is host diversity important, but viral diversity is as well. For instance, ZIKV, DENV, JEV, and WNV NS2B3 can cleave human but not mouse STING, whereas YFV NS3 cannot cleave STING from either species (Ding et al., 2018) (Figure 1.5E). Additional work to identify more divergent flaviviral protease interactions will further define evolution of STING antagonism.

Immune sensors of viral protease activity: who is chasing whom?

Cleavage of host proteins by viral proteases often inactivates the host protein and results in a fitness advantage for the virus. In these cases, host evolutionary signatures reveal adaptations that are presumed to evade cleavage. However, another possibility exists, in which the host protein can sense the presence of the viral protease in the cytoplasm through an evolved sequence that mimics the viral polyprotein cleavage site. Sensing of pathogen-encoded activities such as toxins and effector enzymes, known as effector-trigged immunity (ETI), is well-described in plants but also emerging as an important immune mechanism in animals (Cui, Tsuda, & Parker, 2015; Fischer, Naseer, Shin, & Brodsky, 2020; Jones, Vance, & Dangl, 2016; Vance, Isberg, & Portnoy, 2009). Three such signaling pathways, described below, are known to detect the main protease activity of (+)ssRNA viruses in mammals.

NLRP1 mimics diverse picornaviral 3C cleavage to trigger inflammation

One of the best described cases of mammalian ETI involves NLRP1 (NACHT, LRR, and PYD domains-containing protein 1; Figure 1.6A), a critical sensor for the innate immune complex known as the inflammasome. Mouse NLRP1B was identified in a genetic screen as a determinant of differential susceptibility between mouse strains to Lethal Toxin, a virulence factor responsible for the major pathologies seen during infection by the bacterial pathogen Bacillus anthracis (Boyden & Dietrich, 2006). Further research identified that NLRP1B was a target of cleavage by the secreted bacterial protease component of Lethal Toxin, termed Lethal Factor (LF). Interestingly, mice with a cleavage-susceptible variant of NLRP1B were protected from B. anthracis, indicating

that cleavage of NLRP1B was immunologically protective (Chavarria-Smith, Mitchell, Ho, Daugherty, & Vance, 2016; Moayeri et al., 2010; Terra et al., 2010). The mechanism by which this occurs, termed ‘functional degradation’, depends on proteasomal-mediated degradation of the newly cleaved N-terminal region of NLRP1B and subsequent release of the bioactive C-terminal fragment that can assemble into an active inflammasome (Figure 1.6B) (Chui et al., 2019; Sandstrom et al., 2019). Importantly, for functional degradation to occur, the bioactive C-terminal CARD domain needs to be separated from the N-terminus by a FIIND domain, which is auto-processed to cleave the polypeptide backbone between the N- and C-terminal parts of NRLP1B. This unusual domain architecture thus facilitates the mounting of the inflammasome response upon proteolytic cleavage of the N-terminus.

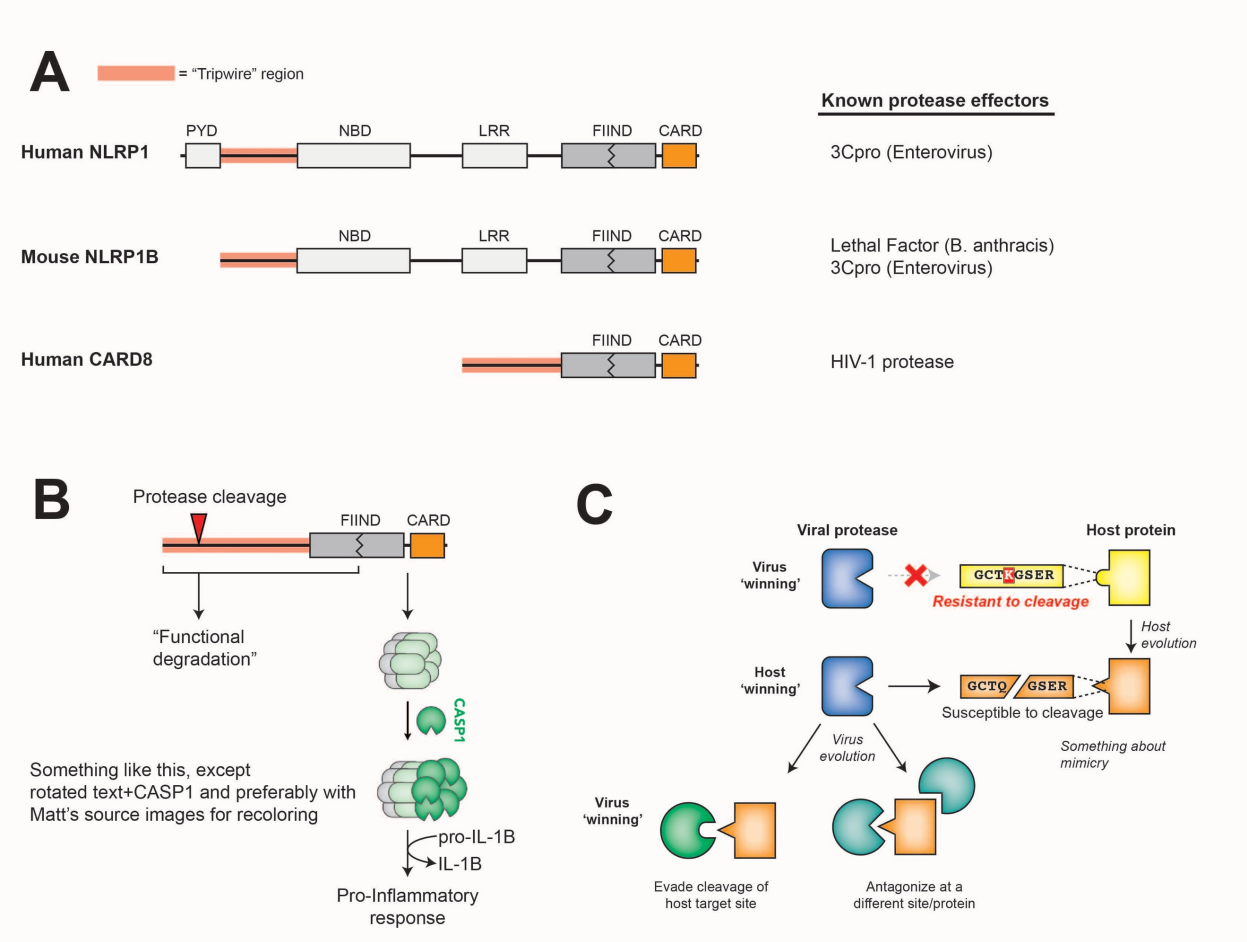


Figure 1.6. Sensing of pathogen-encoded protease activities by host 'tripwires'. (A) NLRP1 and CARD8 serve as effector-triggered immunity (ETI) sensors to detect cleavage by viral proteases. Schematic of human NLRP1, mouse NLRP1B, and human CARD8, highlighting the tripwire region (left) and the known protease effectors (right). (B) Model for how protease cleavage initiates functional degradation of the N-terminal region of inflammasome activators. Activation recruits and activates caspase-1, which cleaves multiple host proteins, including processing of the proinflammatory cytokine, IL-1β, into its mature, bioactive form. (C) Model for how evolution of host resistance to cleavage can drive protease evolution in multiple ways.

LF cleaves within the rapidly evolving 'tripwire' region of mouse NLRP1B but fails to cleave or activate human NLRP1. Interestingly, human NLRP1 has an analogous rapidly-evolving 'tripwire' region (Figure 1.6A), and cleavage of human NLRP1 via an engineered TEV (tobacco etch virus) protease cleavage site can activate the inflammasome (Chavarria-Smith et al., 2016). These data suggested that human NLRP1 may also detect pathogen-encoded proteases via a functional degradation

mechanism. Indeed, we and others recently identified that human NLRP1 recognizes picornavirus 3C protease activity and serves as a tripwire for inflammatory cell death and downstream inflammatory signaling (Robinson et al., 2020; Tsu et al., 2021). During enterovirus infection, 3C cleavage of NLRP1 results in assembly of the active inflammasome and subsequent pro-inflammatory cytokine release (Robinson et al., 2020; Tsu et al., 2021). Interestingly, based on phylogenetic analyses, the 3C-protease site mimic in this specific region of NLRP1 only evolved in the primate lineage, and is only cleavable in some primates (Tsu et al., 2021). Differences across simian primates and a SNP within the human population prevent cleavage and inflammasome activation (Tsu et al., 2021). Although mice lack this human-aligned cleavage site, we discovered a similar phenomenon where picornavirus 3C proteases cleave NLRP1B at different sites to activate the inflammasome in a strain-specific manner (Tsu et al., 2021).

In addition to host diversity, viral diversity also determines NLRP1 cleavage. While all enteroviruses cleave the same site within NLRP1 and activate the inflammasome, other picornaviruses cleave NLRP1 at different sites within the tripwire, and some do not cleave NLRP1 (Tsu et al., 2021). For instance, the 3C protease of EMCV does not cleave NLRP1, and resultingly no activation of the inflammasome was observed (Tsu et al., 2021). As numerous sites in the protease-sensing N-terminal region of NLRP1 are evolving under positive selection (Chavarria-Smith et al., 2016), other independently evolved tripwire mimics may sense divergent 3C or other viral proteases. Finally, 3C proteases from some picornaviruses can antagonize activation. Although further investigation is needed to determine the mechanism by which this

occurs, this suggests that some 3C proteases have evolved to evade detection by NLRP1 by antagonizing NLRP1 function elsewhere.

Intracellular HIV-1 protease activity triggers inflammation via CARD8

Another inflammasome mediator, CARD8, is known to share the unusual C-terminal domain structure critical for the sensing mechanism of NLRP1 – the FIIND domain followed by a CARD domain (Figure 1.6A) (Mitchell, Sandstrom, & Vance, 2019; Taabazuing, Griswold, & Bachovchin, 2020). Bearing further similarities, CARD8 inflammasome assembly can also be activated by the same small molecules as NLRP1 (Taabazuing et al., 2020). Such similarities initially suggested that CARD8 could also be activated using a 'functional degradation' model to act as a tripwire sensor of pathogen-encoded activities (Mitchell et al., 2019). Indeed, the protease of human immunodeficiency virus 1 (HIV-1) can cleave and activate CARD8 inflammasome in an activation mechanism that resembles NLRP1 (Q. Wang et al., 2021). While HIV-1 protease is normally important for cleaving viral polyproteins in the maturing capsid, treatment with specific non-nucleoside reverse transcriptase inhibitors (NNRTIs) can result in protease activity in the cytoplasm (Figueiredo et al., 2006). Under these NNRTI treatment conditions, HIV-1 proteases from four prevalent HIV-1 subtypes cleave CARD8 and activate the inflammasome, results in pro-inflammatory cytokine release and influencing clearance of latent HIV-1 in primary CD4+ T cells (Q. Wang et al., 2021). While the extent to which host evolution or evolution of other viruses influences the activation this same system remains unknown, these findings reveal a broader role

for host encoded ‘tripwires’ for viral proteases that can activate a robust immune response using mimicry of viral protease cleavage sites.

3C-mediated cleavage of a regulator of NF-kB triggers apoptosis

‘Tripwire’ mechanisms such as NLRP1 and CARD8 rely on a specific elegant, but rare, domain architecture that allows for coupling of a cleavage event to generation of a specific bioactive signaling molecule. An additional mechanism for sensing of viral proteases arises from the intricate ways that the innate immune response is negatively regulated. For instance, downstream of NLRP1 and CARD8, mature inflammatory cytokines are detected by the IL-1 receptor to activate the transcription factor NF-κB, which can amplify the inflammatory response (X. Li et al., 2002). NF-κB is an essential transcription factor involved in many innate immune pathways and can mediate a variety of downstream responses depending on the input stimuli (Oeckinghaus, Hayden, & Ghosh, 2011), including both pro- or anti-apoptotic responses (Radhakrishnan & Kamalakaran, 2006). Within the cytoplasm, the NF-κB heterodimer, composed of the Rel family proteins p65 and p50, remains bound and inactive by members of the inhibitor of κB (IκB) family. In response to cytokines such as IL-1β, IκB kinase (IKK) family proteins phosphorylate IκB proteins, releasing the active transcription factor to translocate into the nucleus (X. Li et al., 2002). A previous study demonstrated that the inhibitor of κB alpha (IκBα) senses CVB3 3C protease activity (Zaragoza et al., 2006). The 3C protease was shown to cleave IκBα, producing a fragment that stably complexes with p65 and translocates to the nucleus. This stable complex blocks NF-κB transcriptional activation, resulting in increased cell apoptosis and decreased viral

replication (Zaragoza et al., 2006). Thus, cleavage of IκBα may have evolved as another way to sense viral protease activity and induce cell death to prevent further virus propagation. Many viral proteases are known to cleave proteins in the NF-κB pathway (Table 1). Additional characterization of these virus-host interactions may reveal additional antiviral mechanisms associated with this critical immune pathway.

Evolutionary advantages of ETI

In the continual evolutionary conflict between viruses and their hosts, cleavage mimicry encoded in NLRP1, CARD8 and NF-kB serve as an example of a successful strategy emerging in host organisms to exploit highly constrained pathogenic processes. Rather than mimicry of entire proteins or protein domains, mimicry of these cleavage sites as 'short linear motifs' (SLIMs) require only a small number of amino acids to hijack the highly conserved protease activity. In order to avoid these 'tripwires' and negative regulators of the immune response, these viruses must either evolve their respective main proteases along with all affiliate cleavage sites or antagonize the process some other way (Figure 1.6C). Supporting this idea, we previously discussed that some 3C proteases do not activate the NLRP1 inflammasome despite still cleaving NLRP1, suggesting that 3C proteases have evolved to evade detection by NLRP1 by antagonizing NLRP1 function elsewhere (Tsu et al., 2021). We expect that this work may lead to the discovery that protease-driven ETI strategies may have evolved more broadly at other sites of host-pathogen conflicts.

Conclusion

The proteases of (+)ssRNA viruses have multiple roles in establishing and maintaining virus infection within a host. First and foremost, virally-encoded proteases cleave numerous sequence-specific sites within the viral polyprotein, which is essential for completion of the viral replication cycle. As a consequence of this essential activity, the ability of proteases to evolve novel sequence-specificity is highly constrained. However, viral proteases also serve to manipulate numerous host processes in the infected cell through site-specific cleavage of host targets. In this context, changes in protease sequence-specificity would allow the virus to cleave new host targets that might benefit the virus, or avoid cleaving host targets that are detrimental to the virus. It is at this intersection that viral proteases are engaged in evolutionary 'arms races' with the host, resulting in varied interactions across viral and host species and across evolutionary time. Several examples, including virus-specific cleavage of essential mRNA translation machinery, and host-specific evasion of cleavage of innate antiviral immune components, highlight the consequences of these evolutionary conflicts. More recently, the discovery of host-encoded effector-triggered immunity (ETI) sensors such as NLRP1 and CARD8 suggest that host mimicry of viral protease cleavage sites is an efficient strategy to detect the cellular activity of viral proteases.

The extent to which viral protease evolution, and host target diversity, shape viral host range and pathogenesis remains unknown and is an exciting area of future research. The majority of characterized protease-host interactions have been described for a single virus against a single host, leaving open the opportunity for more detailed exploration of the evolutionary dynamics of these interactions. Indeed, examples such as cleavage of host proteins such as MAVS, STING, and NLRP1 highlight the insights

that can be gained from additional analyses of host and viral diversity in these interactions. Likewise, future studies aiming to discover additional host targets of viral proteases, especially those that may be cleaved in a virus-specific manner, will advance our knowledge of the ways that protease-host interactions shape viral phenotypes. Finally, ETI sensors such as NLRP1 and CARD8 may represent just the start of host proteins that mimic viral protease cleavage sites to induce an immune response. Further studies aimed to identify ETI mechanisms against both viral and other pathogen-encoded proteases will likely continue to reveal novel mechanisms and evolutionary principles of the host innate immune response.

Chapter 1 in part is currently being prepared for submission for publication of the material to Frontiers 2021, including co-authors Elizabeth J. Fay, Katelyn T. Nguyen, Miles R. Corley, Bindhu Hosuru, Viviana A. Dominguez and Matthew D. Daugherty. I, Brian V. Tsu, am the co-first author of this paper, alongside Elizabeth J. Fay.

References

Aguirre, S., Maestre, A. M., Pagni, S., Patel, J. R., Savage, T., Gutman, D., Maringer, K., Bernal-Rubio, D., Shabman, R. S., Simon, V., Rodriguez-Madoz, J. R., Mulder, L. C., Barber, G. N., & Fernandez-Sesma, A. (2012). DENV inhibits type I IFN production in infected cells by cleaving human STING. *PLoS Pathog, 8*(10), e1002934. doi:10.1371/journal.ppat.1002934

Anggakusuma, Brown, R. J. P., Banda, D. H., Todt, D., Vieyres, G., Steinmann, E., & Pietschmann, T. (2016). Hepacivirus NS3/4A Proteases Interfere with MAVS Signaling in both Their Cognate Animal Hosts and Humans: Implications for Zoonotic Transmission. *J Virol, 90*(23), 10670-10681. doi:10.1128/JVI.01634-16

Anirudhan, V., Lee, H., Cheng, H., Cooper, L., & Rong, L. (2021). Targeting SARS-CoV-2 viral proteases as a therapeutic strategy to treat COVID-19. *J Med Virol, 93*(5), 2722-2734. doi:10.1002/jmv.26814

Baggen, J., Thibaut, H. J., Strating, J., & van Kuppeveld, F. J. M. (2018). The life cycle of non-polio enteroviruses and how to target it. *Nat Rev Microbiol, 16*(6), 368-381. doi:10.1038/s41579-018-0005-4

Barral, P. M., Sarkar, D., Fisher, P. B., & Racaniello, V. R. (2009). RIG-I is cleaved during picornavirus infection. *Virology, 391*(2), 171-176. doi:10.1016/j.virol.2009.06.045

Belsham, G. J., McInerney, G. M., & Ross-Smith, N. (2000). Foot-and-mouth disease virus 3C protease induces cleavage of translation initiation factors eIF4A and eIF4G within infected cells. *J Virol, 74*(1), 272-280. doi:10.1128/jvi.74.1.272-280.2000

Blom, N., Hansen, J., Blaas, D., & Brunak, S. (1996). Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks. *Protein Sci, 5*(11), 2203-2216. doi:10.1002/pro.5560051107

Bonderoff, J. M., Larey, J. L., & Lloyd, R. E. (2008). Cleavage of poly(A)-binding protein by poliovirus 3C proteinase inhibits viral internal ribosome entry site-mediated translation. *J Virol, 82*(19), 9389-9399. doi:10.1128/JVI.00006-08

Boyden, E. D., & Dietrich, W. F. (2006). Nalp1b controls mouse macrophage susceptibility to anthrax lethal toxin. *Nat Genet, 38*(2), 240-244. doi:10.1038/ng1724

Butkiewicz, N., Yao, N., Zhong, W., Wright-Minogue, J., Ingravallo, P., Zhang, R., Durkin, J., Standring, D. N., Baroudy, B. M., Sangar, D. V., Lemon, S. M., Lau, J. Y., & Hong, Z. (2000). Virus-specific cofactor requirement and chimeric hepatitis C virus/GB virus B nonstructural protein 3. *J Virol, 74*(9), 4291-4301. doi:10.1128/jvi.74.9.4291-4301.2000

Cahour, A., Falgout, B., & Lai, C. J. (1992). Cleavage of the dengue virus polyprotein at the NS3/NS4A and NS4B/NS5 junctions is mediated by viral protease NS2B-NS3, whereas NS4A/NS4B may be processed by a cellular protease. *J Virol, 66*(3), 1535-1542. doi:10.1128/JVI.66.3.1535-1542.1992

Chambers, T. J., Grakoui, A., & Rice, C. M. (1991). Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. *J Virol, 65*(11), 6042-6050. doi:10.1128/JVI.65.11.6042-6050.1991

Chavarria-Smith, J., Mitchell, P. S., Ho, A. M., Daugherty, M. D., & Vance, R. E. (2016). Functional and Evolutionary Analyses Identify Proteolysis as a General Mechanism for NLRP1 Inflammasome Activation. *PLoS Pathog, 12*(12), e1006052. doi:10.1371/journal.ppat.1006052

Chen, J., Wang, D., Sun, Z., Gao, L., Zhu, X., Guo, J., Xu, S., Fang, L., Li, K., & Xiao, S. (2019). Arterivirus nsp4 Antagonizes Interferon Beta Production by Proteolytically Cleaving NEMO at Multiple Sites. *J Virol, 93*(12). doi:10.1128/JVI.00385-19

Chen, S., Tian, J., Li, Z., Kang, H., Zhang, J., Huang, J., Yin, H., Hu, X., & Qu, L. (2019). Feline Infectious Peritonitis Virus Nsp5 Inhibits Type I Interferon Production by Cleaving NEMO at Multiple Sites. *Viruses, 12*(1). doi:10.3390/v12010043

Chen, S., Wu, Z., Wang, M., & Cheng, A. (2017). Innate Immune Evasion Mediated by Flaviviridae Non-Structural Proteins. *Viruses, 9*(10). doi:10.3390/v9100291

Chen, Z., Benureau, Y., Rijnbrand, R., Yi, J., Wang, T., Warter, L., Lanford, R. E., Weinman, S. A., Lemon, S. M., Martin, A., & Li, K. (2007). GB virus B disrupts RIG-I signaling by NS3/4A-mediated cleavage of the adaptor protein MAVS. *J Virol, 81*(2), 964-976. doi:10.1128/JVI.02076-06

Chui, A. J., Okondo, M. C., Rao, S. D., Gai, K., Griswold, A. R., Johnson, D. C., Ball, D. P., Taabazuing, C. Y., Orth, E. L., Vittimberga, B. A., & Bachovchin, D. A. (2019). N-terminal degradation activates the NLRP1B inflammasome. *Science, 364*(6435), 82-85. doi:10.1126/science.aau1208

Clarke, I. N., & Lambden, P. R. (2000). Organization and expression of calicivirus genes. *J Infect Dis, 181 Suppl 2*, S309-316. doi:10.1086/315575

Cui, H., Tsuda, K., & Parker, J. E. (2015). Effector-triggered immunity: from pathogen perception to robust defense. *Annu Rev Plant Biol, 66*, 487-511. doi:10.1146/annurev-arplant-050213-040012

Daugherty, M. D., & Malik, H. S. (2012). Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet, 46*, 677-700. doi:10.1146/annurev-genet-110711-155522

Deddouche, S., Goubau, D., Rehwinkel, J., Chakravarty, P., Begum, S., Maillard, P. V., Borg, A., Matthews, N., Feng, Q., van Kuppeveld, F. J., & Reis e Sousa, C. (2014). Identification of an LGP2-associated MDA5 agonist in picornavirus-infected cells. *Elife, 3*, e01535. doi:10.7554/eLife.01535

Dewalt, P. G., Lawson, M. A., Colonno, R. J., & Semler, B. L. (1989). Chimeric picornavirus polyproteins demonstrate a common 3C proteinase substrate specificity. *J Virol, 63*(8), 3444-3452. doi:10.1128/JVI.63.8.3444-3452.1989

Ding, Q., Gaska, J. M., Douam, F., Wei, L., Kim, D., Balev, M., Heller, B., & Ploss, A. (2018). Species-specific disruption of STING-dependent antiviral cellular defenses by the Zika virus NS2B3 protease. *Proc Natl Acad Sci U S A, 115*(27), E6310-E6318. doi:10.1073/pnas.1803406115

Duggal, N. K., & Emerman, M. (2012). Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat Rev Immunol, 12*(10), 687-695. doi:10.1038/nri3295

Enard, D., Cai, L., Gwennap, C., & Petrov, D. A. (2016). Viruses are a dominant driver of protein adaptation in mammals. *Elife, 5*. doi:10.7554/eLife.12469

Falgout, B., Miller, R. H., & Lai, C. J. (1993). Deletion analysis of dengue virus type 4 nonstructural protein NS2B: identification of a domain required for NS2B-NS3 protease activity. *J Virol, 67*(4), 2034-2042. doi:10.1128/JVI.67.4.2034-2042.1993

Falgout, B., Pethel, M., Zhang, Y. M., & Lai, C. J. (1991). Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *J Virol, 65*(5), 2467-2475. doi:10.1128/JVI.65.5.2467-2475.1991

Ferreira, A. R., Magalhaes, A. C., Camoes, F., Gouveia, A., Vieira, M., Kagan, J. C., & Ribeiro, D. (2016). Hepatitis C virus NS3-4A inhibits the peroxisomal MAVS-dependent antiviral signalling response. *J Cell Mol Med, 20*(4), 750-757. doi:10.1111/jcmm.12801

Figueiredo, A., Moore, K. L., Mak, J., Sluis-Cremer, N., de Bethune, M. P., & Tachedjian, G. (2006). Potent nonnucleoside reverse transcriptase inhibitors target HIV-1 Gag-Pol. *PLoS Pathog, 2*(11), e119. doi:10.1371/journal.ppat.0020119

Fischer, N. L., Naseer, N., Shin, S., & Brodsky, I. E. (2020). Effector-triggered immunity and pathogen sensing in metazoans. *Nat Microbiol, 5*(3), 528. doi:10.1038/s41564-020-0682-4

Gradi, A., Svitkin, Y. V., Imataka, H., & Sonenberg, N. (1998). Proteolysis of human eukaryotic translation initiation factor eIF4GII, but not eIF4GI, coincides with the shutoff of host protein synthesis after poliovirus infection. *Proc Natl Acad Sci U S A, 95*(19), 11089-11094. doi:10.1073/pnas.95.19.11089

Hoffmann, H. H., Schneider, W. M., & Rice, C. M. (2015). Interferons and viruses: an evolutionary arms race of molecular interactions. *Trends Immunol, 36*(3), 124-138. doi:10.1016/j.it.2015.01.004

Hung, H. C., Wang, H. C., Shih, S. R., Teng, I. F., Tseng, C. P., & Hsu, J. T. (2011). Synergistic inhibition of enterovirus 71 replication by interferon and rupintrivir. *J Infect Dis, 203*(12), 1784-1790. doi:10.1093/infdis/jir174

Jackson, R. J., Hellen, C. U., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol, 11*(2), 113-127. doi:10.1038/nrm2838

Jan, L. R., Yang, C. S., Trent, D. W., Falgout, B., & Lai, C. J. (1995). Processing of Japanese encephalitis virus non-structural proteins: NS2B-NS3 complex and heterologous proteases. *J Gen Virol, 76 (Pt 3)*, 573-580. doi:10.1099/0022-1317-76-3-573

Jones, J. D., Vance, R. E., & Dangl, J. L. (2016). Intracellular innate immune surveillance devices in plants and animals. *Science, 354*(6316). doi:10.1126/science.aaf6395

Judd, E. N., Gilchrist, A. R., Meyerson, N. R., & Sawyer, S. L. (2021). Positive natural selection in primate genes of the type I interferon response. *BMC Ecol Evol, 21*(1), 65. doi:10.1186/s12862-021-01783-z

Junaid, M., Chalayut, C., Sehgelmeble Torrejon, A., Angsuthanasombat, C., Shutava, I., Lapins, M., Wikberg, J. E., & Katzenmeier, G. (2012). Enzymatic analysis of recombinant Japanese encephalitis virus NS2B(H)-NS3pro protease with fluorogenic model peptide substrates. *PLoS One, 7*(5), e36872. doi:10.1371/journal.pone.0036872

Kato, H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., Uematsu, S., Jung, A., Kawai, T., Ishii, K. J., Yamaguchi, O., Otsu, K., Tsujimura, T., Koh, C. S., Reis e Sousa, C., Matsuura, Y., Fujita, T., & Akira, S. (2006). Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature, 441*(7089), 101-105. doi:10.1038/nature04734

Kiemer, L., Lund, O., Brunak, S., & Blom, N. (2004). Coronavirus 3CLpro proteinase cleavage sites: possible relevance to SARS virus pathology. *BMC Bioinformatics, 5*, 72. doi:10.1186/1471-2105-5-72

Kim, H., Kim, A. Y., Choi, J., Park, S. Y., Park, S. H., Kim, J. S., Lee, S. I., Park, J. H., Park, C. K., & Ko, Y. J. (2021). Foot-and-Mouth Disease Virus Evades Innate Immune Response by 3C-Targeting of MDA5. *Cells, 10*(2). doi:10.3390/cells10020271

Kobayashi, M., Arias, C., Garabedian, A., Palmenberg, A. C., & Mohr, I. (2012). Site-specific cleavage of the host poly(A) binding protein by the encephalomyocarditis virus 3C proteinase stimulates viral replication. *J Virol, 86*(19), 10686-10694. doi:10.1128/JVI.00896-12

Laitinen, O. H., Svedin, E., Kapell, S., Nurminen, A., Hytonen, V. P., & Flodstrom-Tullberg, M. (2016). Enteroviral proteases: structure, host interactions and pathogenicity. *Rev Med Virol, 26*(4), 251-267. doi:10.1002/rmv.1883

Lee, K. M., Chen, C. J., & Shih, S. R. (2017). Regulation Mechanisms of Viral IRES-Driven Translation. *Trends Microbiol, 25*(7), 546-561. doi:10.1016/j.tim.2017.01.010

Lei, J., & Hilgenfeld, R. (2017). RNA-virus proteases counteracting host innate immunity. *FEBS Lett, 591*(20), 3190-3210. doi:10.1002/1873-3468.12827

Li, J., Lim, S. P., Beer, D., Patel, V., Wen, D., Tumanut, C., Tully, D. C., Williams, J. A., Jiricek, J., Priestle, J. P., Harris, J. L., & Vasudevan, S. G. (2005). Functional profiling of recombinant NS3 proteases from all four serotypes of dengue virus using tetrapeptide and octapeptide substrate libraries. *J Biol Chem, 280*(31), 28766-28774. doi:10.1074/jbc.M500588200

Li, K., Foy, E., Ferreon, J. C., Nakamura, M., Ferreon, A. C., Ikeda, M., Ray, S. C., Gale, M., Jr., & Lemon, S. M. (2005). Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proc Natl Acad Sci U S A, 102*(8), 2992-2997. doi:10.1073/pnas.0408824102

Li, X., Massa, P. E., Hanidu, A., Peet, G. W., Aro, P., Savitt, A., Mische, S., Li, J., & Marcu, K. B. (2002). IKKalpha, IKKbeta, and NEMO/IKKgamma are each required for the NF-kappa B-mediated inflammatory response program. *J Biol Chem, 277*(47), 45129-45140. doi:10.1074/jbc.M205165200

Li, X. D., Sun, L., Seth, R. B., Pineda, G., & Chen, Z. J. (2005). Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. *Proc Natl Acad Sci U S A, 102*(49), 17717-17722. doi:10.1073/pnas.0508531102

Lindenbach, B. D., & Rice, C. M. (2003). Evasive maneuvers by hepatitis C virus. *Hepatology, 38*(3), 769-771. doi:10.1002/hep.510380327

Lloyd, R. E. (2016). Enterovirus Control of Translation and RNA Granule Stress Responses. *Viruses, 8*(4), 93. doi:10.3390/v8040093

Loo, Y. M., Fornek, J., Crochet, N., Bajwa, G., Perwitasari, O., Martinez-Sobrido, L., Akira, S., Gill, M. A., Garcia-Sastre, A., Katze, M. G., & Gale, M., Jr. (2008). Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *J Virol, 82*(1), 335-345. doi:10.1128/JVI.01080-07

Luo, R., Xiao, S., Jiang, Y., Jin, H., Wang, D., Liu, M., Chen, H., & Fang, L. (2008). Porcine reproductive and respiratory syndrome virus (PRRSV) suppresses interferon-beta production by interfering with the RIG-I signaling pathway. *Mol Immunol, 45*(10), 2839-2846. doi:10.1016/j.molimm.2008.01.028

McCormick, C., & Khaperskyy, D. A. (2017). Translation inhibition and stress granules in the antiviral immune response. *Nat Rev Immunol, 17*(10), 647-660. doi:10.1038/nri.2017.63

Meyerson, N. R., & Sawyer, S. L. (2011). Two-stepping through time: mammals and viruses. *Trends Microbiol, 19*(6), 286-294. doi:10.1016/j.tim.2011.03.006

Meylan, E., Curran, J., Hofmann, K., Moradpour, D., Binder, M., Bartenschlager, R., & Tschopp, J. (2005). Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature, 437*(7062), 1167-1172. doi:10.1038/nature04193

Mitchell, P. S., Sandstrom, A., & Vance, R. E. (2019). The NLRP1 inflammasome: new mechanistic insights and unresolved mysteries. *Curr Opin Immunol, 60*, 37-45. doi:10.1016/j.coi.2019.04.015

Moayeri, M., Crown, D., Newman, Z. L., Okugawa, S., Eckhaus, M., Cataisson, C., Liu, S., Sastalla, I., & Leppla, S. H. (2010). Inflammasome sensor Nlrp1b-dependent resistance to anthrax is mediated by caspase-1, IL-1 signaling and neutrophil recruitment. *PLoS Pathog, 6*(12), e1001222. doi:10.1371/journal.ppat.1001222

Mukherjee, A., Morosky, S. A., Delorme-Axford, E., Dybdahl-Sissoko, N., Oberste, M. S., Wang, T., & Coyne, C. B. (2011). The coxsackievirus B 3C protease cleaves MAVS and TRIF to attenuate host type I interferon and apoptotic signaling. *PLoS Pathog, 7*(3), e1001311. doi:10.1371/journal.ppat.1001311

Navia, M. A., & McKeever, B. M. (1990). A role for the aspartyl protease from the human immunodeficiency virus type 1 (HIV-1) in the orchestration of virus assembly. *Ann N Y Acad Sci, 616*, 73-85. doi:10.1111/j.1749-6632.1990.tb17829.x

Ng, C. S., Stobart, C. C., & Luo, H. (2021). Innate immune evasion mediated by picornaviral 3C protease: Possible lessons for coronaviral 3C-like protease? *Rev Med Virol*, e2206. doi:10.1002/rmv.2206

O'Donoghue, A. J., Eroy-Reveles, A. A., Knudsen, G. M., Ingram, J., Zhou, M., Statnekov, J. B., Greninger, A. L., Hostetter, D. R., Qu, G., Maltby, D. A., Anderson, M. O., Derisi, J. L., McKerrow, J. H., Burlingame, A. L., & Craik, C. S. (2012). Global identification of peptidase specificity by multiplex substrate profiling. *Nat Methods, 9*(11), 1095-1100. doi:10.1038/nmeth.2182

Oeckinghaus, A., Hayden, M. S., & Ghosh, S. (2011). Crosstalk in NF-kappaB signaling pathways. *Nat Immunol, 12*(8), 695-708. doi:10.1038/ni.2065

Papon, L., Oteiza, A., Imaizumi, T., Kato, H., Brocchi, E., Lawson, T. G., Akira, S., & Mechti, N. (2009). The viral RNA recognition sensor RIG-I is degraded during encephalomyocarditis virus (EMCV) infection. *Virology, 393*(2), 311-318. doi:10.1016/j.virol.2009.08.009

Patel, M. R., Loo, Y. M., Horner, S. M., Gale, M., Jr., & Malik, H. S. (2012). Convergent evolution of escape from hepaciviral antagonism in primates. *PLoS Biol, 10*(3), e1001282. doi:10.1371/journal.pbio.1001282

Preugschat, F., Lenches, E. M., & Strauss, J. H. (1991). Flavivirus enzyme-substrate interactions studied with chimeric proteinases: identification of an intragenic locus

important for substrate recognition. *J Virol, 65*(9), 4749-4758. doi:10.1128/JVI.65.9.4749-4758.1991

Qian, S., Fan, W., Liu, T., Wu, M., Zhang, H., Cui, X., Zhou, Y., Hu, J., Wei, S., Chen, H., Li, X., & Qian, P. (2017). Seneca Valley Virus Suppresses Host Type I Interferon Production by Targeting Adaptor Proteins MAVS, TRIF, and TANK for Cleavage. *J Virol, 91*(16). doi:10.1128/JVI.00823-17

Qu, L., Feng, Z., Yamane, D., Liang, Y., Lanford, R. E., Li, K., & Lemon, S. M. (2011). Disruption of TLR3 signaling due to cleavage of TRIF by the hepatitis A virus protease-polymerase processing intermediate, 3CD. *PLoS Pathog, 7*(9), e1002169. doi:10.1371/journal.ppat.1002169

Radhakrishnan, S. K., & Kamalakaran, S. (2006). Pro-apoptotic role of NF-kappaB: implications for cancer therapy. *Biochim Biophys Acta, 1766*(1), 53-62. doi:10.1016/j.bbcan.2006.02.001

Roberts, A., Deming, D., Paddock, C. D., Cheng, A., Yount, B., Vogel, L., Herman, B. D., Sheahan, T., Heise, M., Genrich, G. L., Zaki, S. R., Baric, R., & Subbarao, K. (2007). A mouse-adapted SARS-coronavirus causes disease and mortality in BALB/c mice. *PLoS Pathog, 3*(1), e5. doi:10.1371/journal.ppat.0030005

Robinson, K. S., Teo, D. E. T., Tan, K. S., Toh, G. A., Ong, H. H., Lim, C. K., Lay, K., Au, B. V., Lew, T. S., Chu, J. J. H., Chow, V. T. K., Wang, Y., Zhong, F. L., & Reversade, B. (2020). Enteroviral 3C protease activates the human NLRP1 inflammasome in airway epithelia. *Science*. doi:10.1126/science.aay2002

Roe, M. K., Junod, N. A., Young, A. R., Beachboard, D. C., & Stobart, C. C. (2021). Targeting novel structural and functional features of coronavirus protease nsp5 (3CL(pro), M(pro)) in the age of COVID-19. *J Gen Virol, 102*(3). doi:10.1099/jgv.0.001558

Rothenburg, S., & Brennan, G. (2020). Species-Specific Host-Virus Interactions: Implications for Viral Host Range and Virulence. *Trends Microbiol, 28*(1), 46-56. doi:10.1016/j.tim.2019.08.007

Sandstrom, A., Mitchell, P. S., Goers, L., Mu, E. W., Lesser, C. F., & Vance, R. E. (2019). Functional degradation: A mechanism of NLRP1 inflammasome activation by diverse pathogen enzymes. *Science, 364*(6435). doi:10.1126/science.aau1330

Sawyer, S. L., & Elde, N. C. (2012). A cross-species view on viruses. *Curr Opin Virol, 2*(5), 561-568. doi:10.1016/j.coviro.2012.07.003

Sbardellati, A., Scarselli, E., Amati, V., Falcinelli, S., Kekule, A. S., & Traboni, C. (2000). Processing of GB virus B non-structural proteins in cultured cells requires both NS3 protease and NS4A cofactor. *J Gen Virol, 81*(Pt 9), 2183-2188. doi:10.1099/0022-1317-81-9-2183

Schoggins, J. W. (2014). Interferon-stimulated genes: roles in viral pathogenesis. *Curr Opin Virol, 6*, 40-46. doi:10.1016/j.coviro.2014.03.006

Shultz, A. J., & Sackton, T. B. (2019). Immune genes are hotspots of shared positive selection across birds and mammals. *Elife, 8*. doi:10.7554/eLife.41815

Sironi, M., Cagliani, R., Forni, D., & Clerici, M. (2015). Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nat Rev Genet, 16*(4), 224-236. doi:10.1038/nrg3905

Stabell, A. C., Meyerson, N. R., Gullberg, R. C., Gilchrist, A. R., Webb, K. J., Old, W. M., Perera, R., & Sawyer, S. L. (2018). Dengue viruses cleave STING in humans but not in nonhuman primates, their presumed natural reservoir. *Elife, 7*. doi:10.7554/eLife.31919

Stanley, J. T., Gilchrist, A. R., Stabell, A. C., Allen, M. A., Sawyer, S. L., & Dowell, R. D. (2020). Two-stage ML Classifier for Identifying Host Protein Targets of the Dengue Protease. *Pac Symp Biocomput, 25*, 487-498.

Steuber, H., & Hilgenfeld, R. (2010). Recent advances in targeting viral proteases for the discovery of novel antivirals. *Curr Top Med Chem, 10*(3), 323-345. doi:10.2174/156802610790725470

Stobart, C. C., Sexton, N. R., Munjal, H., Lu, X., Molland, K. L., Tomar, S., Mesecar, A. D., & Denison, M. R. (2013). Chimeric exchange of coronavirus nsp5 proteases (3CLpro) identifies common and divergent regulatory determinants of protease activity. *J Virol, 87*(23), 12611-12618. doi:10.1128/JVI.02050-13

Su, C. I., Kao, Y. T., Chang, C. C., Chang, Y., Ho, T. S., Sun, H. S., Lin, Y. L., Lai, M. M. C., Liu, Y. H., & Yu, C. Y. (2020). DNA-induced 2'3'-cGAMP enhances haplotype-specific human STING cleavage by dengue protease. *Proc Natl Acad Sci U S A, 117*(27), 15947-15954. doi:10.1073/pnas.1922243117

Taabazuing, C. Y., Griswold, A. R., & Bachovchin, D. A. (2020). The NLRP1 and CARD8 inflammasomes. *Immunol Rev*. doi:10.1111/imr.12884

Ten Dam, E., Flint, M., & Ryan, M. D. (1999). Virus-encoded proteinases of the Togaviridae. *J Gen Virol, 80 (Pt 8)*, 1879-1888. doi:10.1099/0022-1317-80-8-1879

Terra, J. K., Cote, C. K., France, B., Jenkins, A. L., Bozue, J. A., Welkos, S. L., LeVine, S. M., & Bradley, K. A. (2010). Cutting edge: resistance to Bacillus anthracis infection mediated by a lethal toxin sensitive allele of Nalp1b/Nlrp1b. *J Immunol, 184*(1), 17-20. doi:10.4049/jimmunol.0903114

Tsu, B. V., Beierschmitt, C., Ryan, A. P., Agarwal, R., Mitchell, P. S., & Daugherty, M. D. (2021). Diverse viral proteases activate the NLRP1 inflammasome. *Elife, 10*. doi:10.7554/eLife.60609

V'Kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol, 19*(3), 155-170. doi:10.1038/s41579-020-00468-6

Valli, A. A., Gallo, A., Rodamilans, B., Lopez-Moya, J. J., & Garcia, J. A. (2018). The HCPro from the Potyviridae family: an enviable multitasking Helper Component that every virus would like to have. *Mol Plant Pathol, 19*(3), 744-763. doi:10.1111/mpp.12553

van Kasteren, P. B., Beugeling, C., Ninaber, D. K., Frias-Staheli, N., van Boheemen, S., Garcia-Sastre, A., Snijder, E. J., & Kikkert, M. (2012). Arterivirus and nairovirus ovarian tumor domain-containing Deubiquitinases target activated RIG-I to control innate immune signaling. *J Virol, 86*(2), 773-785. doi:10.1128/JVI.06277-11

VanBlargan, L. A., Davis, K. A., Dowd, K. A., Akey, D. L., Smith, J. L., & Pierson, T. C. (2015). Context-Dependent Cleavage of the Capsid Protein by the West Nile Virus Protease Modulates the Efficiency of Virus Assembly. *J Virol, 89*(16), 8632-8642. doi:10.1128/JVI.01253-15

Vance, R. E., Isberg, R. R., & Portnoy, D. A. (2009). Patterns of pathogenesis: discrimination of pathogenic and nonpathogenic microbes by the innate immune system. *Cell Host Microbe, 6*(1), 10-21. doi:10.1016/j.chom.2009.06.007

Vazquez, C., Tan, C. Y., & Horner, S. M. (2019). Hepatitis C Virus Infection Is Inhibited by a Noncanonical Antiviral Signaling Pathway Targeted by NS3-NS4A. *J Virol, 93*(23). doi:10.1128/JVI.00725-19

Walker, P. J., Siddell, S. G., Lefkowitz, E. J., Mushegian, A. R., Adriaenssens, E. M., Dempsey, D. M., Dutilh, B. E., Harrach, B., Harrison, R. L., Hendrickson, R. C., Junglen, S., Knowles, N. J., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Nibert, M., Orton, R. J., Rubino, L., Sabanadzovic, S., Simmonds, P., Smith, D. B., Varsani, A., Zerbini, F. M., & Davison, A. J. (2020). Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses (2020). *Arch Virol, 165*(11), 2737-2748. doi:10.1007/s00705-020-04752-x

Wang, D., Fang, L., Li, K., Zhong, H., Fan, J., Ouyang, C., Zhang, H., Duan, E., Luo, R., Zhang, Z., Liu, X., Chen, H., & Xiao, S. (2012). Foot-and-mouth disease virus 3C protease cleaves NEMO to impair innate immune signaling. *J Virol, 86*(17), 9311-9322. doi:10.1128/JVI.00722-12

Wang, D., Fang, L., Shi, Y., Zhang, H., Gao, L., Peng, G., Chen, H., Li, K., & Xiao, S. (2016). Porcine Epidemic Diarrhea Virus 3C-Like Protease Regulates Its Interferon Antagonism by Cleaving NEMO. *J Virol, 90*(4), 2090-2101. doi:10.1128/JVI.02514-15

Wang, D., Fang, L., Wei, D., Zhang, H., Luo, R., Chen, H., Li, K., & Xiao, S. (2014). Hepatitis A virus 3C protease cleaves NEMO to impair induction of beta interferon. *J Virol, 88*(17), 10252-10258. doi:10.1128/JVI.00869-14

Wang, Q., Gao, H., Clark, K. M., Mugisha, C. S., Davis, K., Tang, J. P., Harlan, G. H., DeSelm, C. J., Presti, R. M., Kutluay, S. B., & Shan, L. (2021). CARD8 is an inflammasome sensor for HIV-1 protease activity. *Science, 371*(6535). doi:10.1126/science.abe1707

West, A. P., Khoury-Hanold, W., Staron, M., Tal, M. C., Pineda, C. M., Lang, S. M., Bestwick, M., Duguay, B. A., Raimundo, N., MacDuff, D. A., Kaech, S. M., Smiley, J. R., Means, R. E., Iwasaki, A., & Shadel, G. S. (2015). Mitochondrial DNA stress primes the antiviral innate immune response. *Nature, 520*(7548), 553-557. doi:10.1038/nature14156

White, J. P., Cardenas, A. M., Marissen, W. E., & Lloyd, R. E. (2007). Inhibition of cytoplasmic mRNA stress granule formation by a viral proteinase. *Cell Host Microbe, 2*(5), 295-305. doi:10.1016/j.chom.2007.08.006

White, J. P., & Lloyd, R. E. (2012). Regulation of stress granules in virus systems. *Trends Microbiol, 20*(4), 175-183. doi:10.1016/j.tim.2012.02.001

Winston, D. S., & Boehr, D. D. (2021). The Picornavirus Precursor 3CD Has Different Conformational Dynamics Compared to 3C(pro) and 3D(pol) in Functionally Relevant Regions. *Viruses, 13*(3). doi:10.3390/v13030442

Wiskerchen, M., & Collett, M. S. (1991). Pestivirus gene expression: protein p80 of bovine viral diarrhea virus is a proteinase involved in polyprotein processing. *Virology, 184*(1), 341-350. doi:10.1016/0042-6822(91)90850-b

Xiang, Z., Li, L., Lei, X., Zhou, H., Zhou, Z., He, B., & Wang, J. (2014). Enterovirus 68 3C protease cleaves TRIF to attenuate antiviral responses mediated by Toll-like receptor 3. *J Virol, 88*(12), 6650-6659. doi:10.1128/JVI.03138-13

Xiang, Z., Liu, L., Lei, X., Zhou, Z., He, B., & Wang, J. (2016). 3C Protease of Enterovirus D68 Inhibits Cellular Defense Mediated by Interferon Regulatory Factor 7. *J Virol, 90*(3), 1613-1621. doi:10.1128/JVI.02395-15

Xu, J., Mendez, E., Caron, P. R., Lin, C., Murcko, M. A., Collett, M. S., & Rice, C. M. (1997). Bovine viral diarrhea virus NS3 serine proteinase: polyprotein cleavage sites, cofactor requirements, and molecular model of an enzyme essential for pestivirus replication. *J Virol, 71*(7), 5312-5322. doi:10.1128/JVI.71.7.5312-5322.1997

Ye, X., Pan, T., Wang, D., Fang, L., Ma, J., Zhu, X., Shi, Y., Zhang, K., Zheng, H., Chen, H., Li, K., & Xiao, S. (2018). Foot-and-Mouth Disease Virus Counteracts on Internal Ribosome Entry Site Suppression by G3BP1 and Inhibits G3BP1-

Mediated Stress Granule Assembly via Post-Translational Mechanisms. *Front Immunol, 9*, 1142. doi:10.3389/fimmu.2018.01142

Yi, J., Peng, J., Yang, W., Zhu, G., Ren, J., Li, D., & Zheng, H. (2021). Picornavirus 3C - a protease ensuring virus replication and subverting host responses. *J Cell Sci, 134*(5). doi:10.1242/jcs.253237

Zaragoza, C., Saura, M., Padalko, E. Y., Lopez-Rivera, E., Lizarbe, T. R., Lamas, S., & Lowenstein, C. J. (2006). Viral protease cleavage of inhibitor of kappaBalpha triggers host cell apoptosis. *Proc Natl Acad Sci U S A, 103*(50), 19051-19056. doi:10.1073/pnas.0606019103

Zhang, B., Morace, G., Gauss-Muller, V., & Kusov, Y. (2007). Poly(A) binding protein, C-terminally truncated by the hepatitis A virus proteinase 3C, inhibits viral translation. *Nucleic Acids Res, 35*(17), 5975-5984. doi:10.1093/nar/gkm645

Zhu, X., Fang, L., Wang, D., Yang, Y., Chen, J., Ye, X., Foda, M. F., & Xiao, S. (2017). Porcine deltacoronavirus nsp5 inhibits interferon-beta production through the cleavage of NEMO. *Virology, 502*, 33-38. doi:10.1016/j.virol.2016.12.005

Zhu, X., Wang, D., Zhou, J., Pan, T., Chen, J., Yang, Y., Lv, M., Ye, X., Peng, G., Fang, L., & Xiao, S. (2017). Porcine Deltacoronavirus nsp5 Antagonizes Type I Interferon Signaling by Cleaving STAT2. *J Virol, 91*(10). doi:10.1128/JVI.00003-17

Chapter 2: Evolution-guided discovery of new host-viral interactions

Abstract

Viruses are embroiled in arms races that have profoundly shaped the evolution of their hosts. These arms races leave behind distinct evolutionary signature of rapid amino acid site evolution in relevant host genes. Genes saturated with these signatures are suggested to be performing roles in numerous arms race conflicts and therefore play critical roles in host innate immunity. Here I identify, in an unbiased whole-genome approach, which genes elicit strong signatures of arms races. By identifying these genes, we are now poised to discover where viral proteases are driving this evolution, and the functional outcomes of these interactions.

Introduction

Most mammalian proteins are highly conserved in sequence due to their evolutionary constraint to perform key cellular housekeeping functions. This constraint largely prevents these genes from playing additional roles to protect the host organism from ever-changing factors from the external world. However, in every genome there are a subset of genes that do evolve under strong selection pressure by these external factors, and this evolution can be observed within the physical interface encoded by these genes. Such changes have been observed in numerous genes classes, including those involved in reproduction, diet, appearance, sensory systems, behavior, brain development and innate immunity (Vallender & Lahn, 2004). Understanding how viruses have profoundly shaped the evolution of these subsets of genes within their hosts is an important step in understanding how viruses have adapted to new hosts and inflict varying levels of disease severity.

As discussed in Chapter 1, in recurring host-virus conflicts, repeated counter-

adaptation shifts the replicative advantage back-and-forth between host and pathogen, resembling an evolutionary arms race (Daugherty & Malik, 2012; Warren & Sawyer, 2019). Retrospective analyses focused on known sites of virus conflicts within primate innate immunity genes identified high amino acid site diversification as a common signature of arms races (Duggal, Malik, & Emerman, 2011; Elde, Child, Geballe, & Malik, 2009; Han, Lou, & Sawyer, 2011; Hancks, Hartley, Hagan, Clark, & Elde, 2015; Lim, Malik, & Emerman, 2010; Liu, Chen, Wang, & Zhang, 2010; Patel, Loo, Horner, Gale, & Malik, 2012; Sawyer, Emerman, & Malik, 2004; Sawyer, Wu, Emerman, & Malik, 2005). This degree of amino acid diversification can be identified by calculating dN/dS across a given gene sequence, where the accumulation of nonsynonymous substitutions (dN) is compared to that of synonymous substitutions (dS) (Murrell et al., 2013; Yang, 2007). An emerging question from these analyses was whether these signatures of high amino acid site diversification could be used to predict which subsets of genes are in evolutionary arms races with viruses.

Results

Filtering alignment errors across the primate exome

To discover and characterize new ways that viruses have driven host genetic innovation, I collaborated with Ben Murrell and Joel Wertheim at the UCSD Hillcrest campus to determine, from a whole genome, unbiased perspective, the human genes that display evolutionary signatures consistent with a host-virus arms race. Using FUBAR, they initially discovered that among a subset of 18 well-aligned genes, the four genes previously reported to be in arms races exhibited a distinct evolutionary gene fingerprint (Murrell et al., 2013; Murrell, Vollbrecht, Guatelli, & Wertheim, 2016). In

adapting this work to the whole genome level, Ben and Joel first obtained exon alignments corresponding to 41,735 gene isoform alignments between humans and 11 additional primates from the UCSC Genome Browser. I then developed a computational method to identify poorly aligned regions within these alignments to be discarded from further analyses (Figure 2.1 and 2.2). After Joel removed the flagged sequences, Ben used FUBAR to generate unique evolutionary fingerprints for each gene alignment. Together, we asked whether a principal component analysis of these fingerprints would yield easily separable clusters for conflict-driven genes.

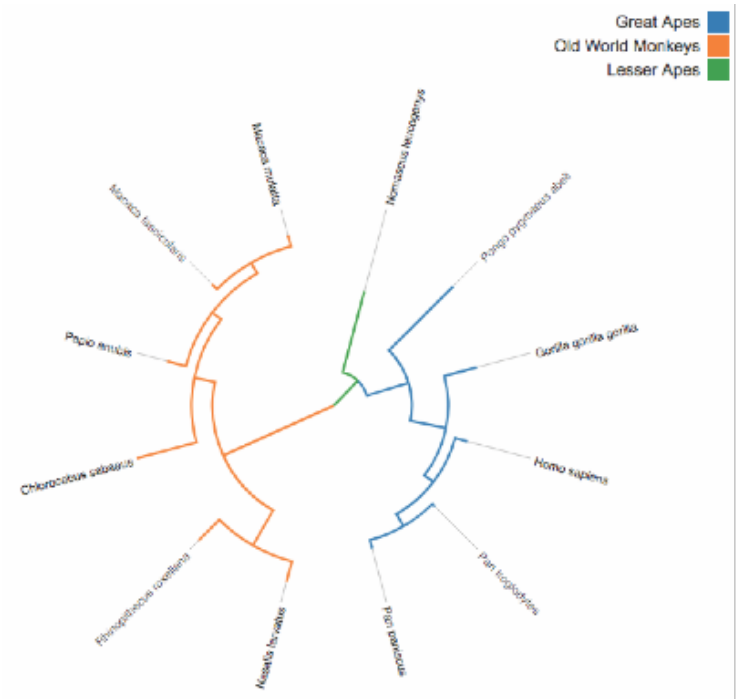
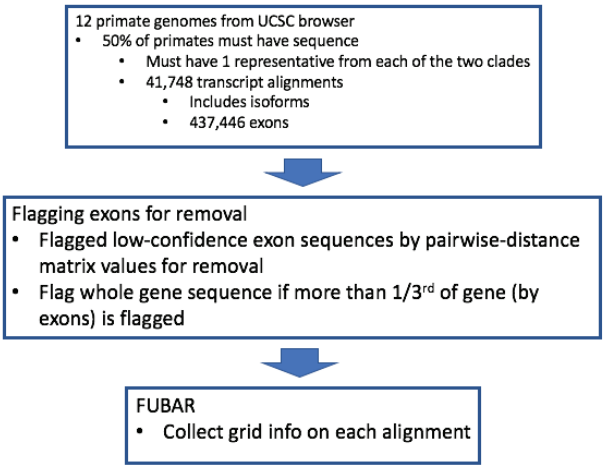


Figure 2.1. Phylogenetic tree of twelve primates in study. Each taxonomic family is labeled by colors corresponding to the legend in the top right.

A



B

			Species-specific sequences removed from alignment											
			nomLeu3	ponAbe2	gorGor3	hg38	panTro4	panPan1	nasLar1	rhiRox1	chlSab2	papAnu2	macFas5	rheMac3
	Total	Complete alignment removal	Gibbon	Orangutan	Gorilla	Human	Chimp	Bonobo	Proboscis	Golden Snub-nosed	AGM	Baboon	Crab-eating macaque	Rhesus Macaque
Whole gene alignments	41742	6	86	3	17	0	5	4	41	9	6	21	6	0
Individual exon	437441	5	528	48	126	0	66	8	387	17	36	93	51	7

Figure 2.2. Sequence alignments flagged for removal and FUBAR analysis. (A) Pipeline to remove sequences based on pairwise-distance values. (B) Number of low-quality species-specific sequences or complete alignments removed by the pipeline at either the whole gene or individual exon level.

Characterization of evolutionary fingerprints across the genome

The resulting distribution across the strongest principal component, PC1, strongly correlates to a trend of increasingly rapid evolution (Figure 2.3A). Whether this original approach using FUBAR evolutionary fingerprints captured additional biological insight over the more classical approach of gene-wide dN/dS analyses was unclear (Figure 2.3B). Still, this approach allowed us to visualize the evolutionary differences for each gene within the primate genome.

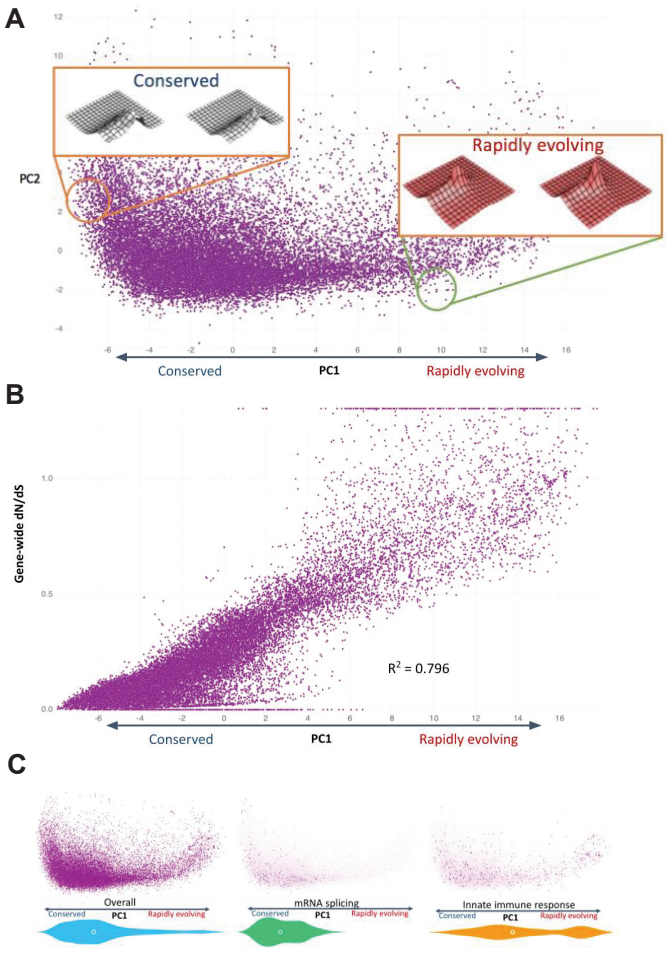


Figure 2.3. Known evolutionary profiles of gene functions are captured across the whole genome. (A) Of the 41,735 isoform alignments, 21337 alignments representative of canonical human isoforms are shown here. Fingerprints of genes with increasing PC1 values approach fingerprints of genes in arms races. (B) PC1 and gene-wide dN/dS are correlated. (C) Compared to the overall distribution, genes with housekeeping roles such as mRNA splicing are largely globally conserved, whereas genes that are reported to directly interface with pathogens are more strongly enriched on the right.

Regardless of whether we used PC1 or this gene-wide dN/dS metric, an analysis querying gene ontology terms confirms the evolutionary profiles of genes described in literature (Figure 2.3C). With GO terms, genes with known housekeeping function such as the mRNA splicing genes are not rapidly evolving. On the opposite side of the spectrum, a group known to directly interact with pathogens, the innate immunity factors, exhibit greater enrichment for rapidly evolving genes. Interestingly, innate

immunity factors largely exist in two clusters, where the second group is largely not rapidly evolving. This suggests that we can prioritize which subset of genes might be more critically involved in genetic conflicts. By following this trend, we can also make similar hypotheses regarding non-associated or lesser-associated immunity genes within the same evolutionary bounds, where evolutionary profiles like that of the rapidly evolving group of innate immunity factors hint at potential undiscovered roles in host-virus conflicts.

Discussion

Here we showed that we can use canonical isoforms to represent each gene to compare how fast different genes are evolving across the primate genome. In this work, we observed that the GO category of innate immune factors (Figure 2.3) was bimodal, and that this clear delineation could be used to delineate the restrictive search space to direct experimental hypotheses.

Similar, recent genome-wide evolutionary studies focused on identifying genes under positive selection across whole exomes of bats and primates (Hawkins et al., 2019; van der Lee, Wiel, van Dam, & Huynen, 2017). These findings capture previously identified arms race cases while also re-iterating our challenge of using genome-scale exon alignments. With advances in sequencing technology and more availability of transcriptome assemblies for diverse host organisms, much of these concerns will be reduced as better isoform alignments are produced with less dependency on exon-intron junction predictions.

By defining the search space of genes suggested to be in arms race conflicts, the next step is to discover which of these genes are specifically driven by viruses. To

further home in on virus-driven conflict, I focus on the common viral antagonism strategy of protease cleavage, as discussed in Chapter 1. In this work, I discuss the 3C and 3CL proteases of the *Picornaviridae* and *Coronaviridae* families, which encompass a diverse set of human enteric and respiratory pathogens. By overlapping genome-wide evolutionary profiles with the 3C and 3CL protease cleavage motifs, I identified two rapidly evolving genes, NLRP1 and CARD8, as host targets in conflicts with viruses.

Chapter 2 includes unpublished material co-authored with Ben Murrell, Joel O. Wertheim and Matthew D. Daugherty. I, Brian V. Tsu, am the primary author of this material.

References

Daugherty, M. D., & Malik, H. S. (2012). Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet, 46*, 677-700. doi:10.1146/annurev-genet-110711-155522

Duggal, N. K., Malik, H. S., & Emerman, M. (2011). The breadth of antiviral activity of Apobec3DE in chimpanzees has been driven by positive selection. *J Virol, 85*(21), 11361-11371. doi:10.1128/JVI.05046-11

Elde, N. C., Child, S. J., Geballe, A. P., & Malik, H. S. (2009). Protein kinase R reveals an evolutionary model for defeating viral mimicry. *Nature, 457*(7228), 485-489. doi:10.1038/nature07529

Han, K., Lou, D. I., & Sawyer, S. L. (2011). Identification of a genomic reservoir for new TRIM genes in primate genomes. *PLoS Genet, 7*(12), e1002388. doi:10.1371/journal.pgen.1002388

Hancks, D. C., Hartley, M. K., Hagan, C., Clark, N. L., & Elde, N. C. (2015). Overlapping Patterns of Rapid Evolution in the Nucleic Acid Sensors cGAS and OAS1 Suggest a Common Mechanism of Pathogen Antagonism and Escape. *PLoS Genet, 11*(5), e1005203. doi:10.1371/journal.pgen.1005203

Hawkins, J. A., Kaczmarek, M. E., Muller, M. A., Drosten, C., Press, W. H., & Sawyer, S. L. (2019). A metaanalysis of bat phylogenetics and positive selection based on

genomes and transcriptomes from 18 species. *Proc Natl Acad Sci U S A, 116*(23), 11351-11360. doi:10.1073/pnas.1814995116

Lim, E. S., Malik, H. S., & Emerman, M. (2010). Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses. *J Virol, 84*(14), 7124-7134. doi:10.1128/JVI.00468-10

Liu, J., Chen, K., Wang, J. H., & Zhang, C. (2010). Molecular evolution of the primate antiviral restriction factor tetherin. *PLoS One, 5*(7), e11904. doi:10.1371/journal.pone.0011904

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol, 30*(5), 1196-1205. doi:10.1093/molbev/mst030

Murrell, B., Vollbrecht, T., Guatelli, J., & Wertheim, J. O. (2016). The Evolutionary Histories of Antiretroviral Proteins SERINC3 and SERINC5 Do Not Support an Evolutionary Arms Race in Primates. *J Virol, 90*(18), 8085-8089. doi:10.1128/JVI.00972-16

Patel, M. R., Loo, Y. M., Horner, S. M., Gale, M., Jr., & Malik, H. S. (2012). Convergent evolution of escape from hepaciviral antagonism in primates. *PLoS Biol, 10*(3), e1001282. doi:10.1371/journal.pbio.1001282

Sawyer, S. L., Emerman, M., & Malik, H. S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol, 2*(9), E275. doi:10.1371/journal.pbio.0020275

Sawyer, S. L., Wu, L. I., Emerman, M., & Malik, H. S. (2005). Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc Natl Acad Sci U S A, 102*(8), 2832-2837. doi:10.1073/pnas.0409853102

Vallender, E. J., & Lahn, B. T. (2004). Positive selection on the human genome. *Hum Mol Genet, 13 Spec No 2*, R245-254. doi:10.1093/hmg/ddh253

van der Lee, R., Wiel, L., van Dam, T. J. P., & Huynen, M. A. (2017). Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. *Nucleic Acids Res, 45*(18), 10634-10648. doi:10.1093/nar/gkx704

Warren, C. J., & Sawyer, S. L. (2019). How host genetics dictates successful viral zoonosis. *PLoS Biol, 17*(4), e3000217. doi:10.1371/journal.pbio.3000217

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol, 24*(8), 1586-1591. doi:10.1093/molbev/msm088

Chapter 3: Diverse Viral Proteases Activate the NLRP1 Inflammasome

Abstract

The NLRP1 inflammasome is a multiprotein complex that is a potent activator of inflammation. Mouse NLRP1B can be activated through proteolytic cleavage by the bacterial Lethal Toxin (LeTx) protease, resulting in degradation of the N-terminal domains of NLRP1B and liberation of the bioactive C-terminal domain, which includes the caspase activation and recruitment domain (CARD). However, natural pathogen-derived effectors that can activate human NLRP1 have remained unknown. Here, we use an evolutionary model to identify several proteases from diverse picornaviruses that cleave human NLRP1 within a rapidly evolving region of the protein, leading to host-specific and virus-specific activation of the NLRP1 inflammasome. Our work demonstrates that NLRP1 acts as a 'tripwire' to recognize the enzymatic function of a wide range of viral proteases and suggests that host mimicry of viral polyprotein cleavage sites can be an evolutionary strategy to activate a robust inflammatory immune response.

Introduction

The ability to sense and respond to pathogens is central to the mammalian immune system. However, immune activation needs to be properly calibrated, as an overactive immune response can at times be as pathogenic as the pathogen itself. To ensure accurate discrimination of self and non-self, innate immune sensors detect broadly conserved microbial molecules such as bacterial flagellin or double-stranded RNA (Janeway, 1989). However, such microbial patterns can be found on harmless and pathogenic microbes alike. More recently, pathogen-specific activities such as toxins or

effector enzymes have also been shown to be targets of innate immune recognition (Jones et al., 2016; Mitchell et al., 2019; Vance et al., 2009). Such a system for detection, termed effector-trigged immunity (ETI), has been well-established in plants (Cui et al., 2015; Jones et al., 2016) and is emerging as an important means to allow the immune system to distinguish pathogens from harmless microbes in animals (Fischer et al., 2020; Jones et al., 2016).

Complicating the success of host detection systems, innate immune sensors are under constant selective pressure to adapt due to pathogen evasion or antagonism of immune detection. Such evolutionary dynamics, termed host-pathogen arms races, result from genetic conflicts where both host and pathogen are continually driven to adapt to maintain a fitness advantage. The antagonistic nature of these conflicts can be distinguished via signatures of rapid molecular evolution at the exact sites where host and pathogen interact (Daugherty and Malik, 2012; Meyerson and Sawyer, 2011; Sironi et al., 2015). Consistent with their role as the first line of cellular defense against incoming pathogens, innate immune sensors of both broad molecular patterns as well as specific pathogen-associated effectors have been shown to be engaged in genetic conflicts with pathogens (Cagliani et al., 2014; Chavarría-Smith et al., 2016; Hancks et al., 2015; Tenthorey et al., 2014; Tian et al., 2009).

Inflammasomes are one such group of rapidly evolving cytosolic immune sensor complexes (Broz and Dixit, 2016; Chavarría-Smith et al., 2016; Evavold and Kagan, 2019; Rathinam and Fitzgerald, 2016; Tenthorey et al., 2014; Tian et al., 2009). Upon detection of microbial molecules or pathogen-encoded activities, inflammasome-forming sensor proteins serve as a platform for the recruitment and activation of

proinflammatory caspases including caspase-1 (CASP1) through either a pyrin domain (PYD) or a caspase activation and recruitment domain (CARD) (Broz and Dixit, 2016; Rathinam and Fitzgerald, 2016). Active CASP1 mediates the maturation and release of the proinflammatory cytokines interleukin (IL)−1β and IL-18 (Broz and Dixit, 2016; Rathinam et al., 2012). CASP1 also initiates a form of cell death known as pyroptosis (Broz and Dixit, 2016; Rathinam et al., 2012). Together, these outputs provide robust defense against a wide array of eukaryotic, bacterial, and viral pathogens (Broz and Dixit, 2016; Evavold and Kagan, 2019; Rathinam and Fitzgerald, 2016).

The first described inflammasome is scaffolded by the sensor protein NLRP1, a member of the nucleotide-binding domain (NBD), leucine-rich repeat (LRR)-containing (NLR) superfamily (Martinon et al., 2002; Ting et al., 2008). NLRP1 has an unusual domain architecture, containing a CARD at its C-terminus rather than the N-terminus like all other inflammasome sensor NLRs, and a function-to-find domain (FIIND), which is located between the LRRs and CARD (Ting et al., 2008). The FIIND undergoes a constitutive self-cleavage event, such that NLRP1 exists in its non-activated state as two, noncovalently associated polypeptides (D'Osualdo et al., 2011; Finger et al., 2012; Frew et al., 2012), the N-terminal domains and the C-terminal CARD-containing fragment.

The importance of the unusual domain architecture of NLRP1 for mounting a pathogen-specific inflammasome response has been elucidated over the last several decades (Evavold and Kagan, 2019; Mitchell et al., 2019; Taabazuing et al., 2020). The first hint that NLRP1 does not detect broadly conserved microbial molecules came from the discovery that the Bacillus anthracis Lethal Toxin (LeTx) is required to elicit a

protective inflammatory response against B. anthracis infection via one of the mouse NLRP1 homologs, NLRP1B (Boyden and Dietrich, 2006; Greaney et al., 2020; Moayeri et al., 2010; Terra et al., 2010). Paradoxically, inflammasome activation is the result of site-specific cleavage in the N-terminus of mouse NLRP1B by the Lethal Factor (LF) protease subunit of LeTx, indicating that protease-mediated cleavage of NLRP1 does not disable its function but instead results in its activation (Chavarría-Smith and Vance, 2013; Levinsohn et al., 2012). More recently, the mechanism by which LF-mediated proteolytic cleavage results in direct NLRP1B inflammasome activation has been detailed (Chui et al., 2019; Sandstrom et al., 2019). These studies revealed that proteolysis of mouse NLRP1B by LF results in exposure of a novel N-terminus, which is then targeted for proteasomal degradation by a protein quality control mechanism called the 'N-end rule' pathway (Chui et al., 2019; Sandstrom et al., 2019; Wickliffe et al., 2008; Xu et al., 2019). Since the proteasome is a processive protease, it progressively degrades the N-terminal domains of NLRP1B but is disengaged upon arriving at the self-cleavage site within the FIIND domain. Degradation of the N-terminal domains thus releases the bioactive C-terminal CARD-containing fragment of NLRP1B from its non-covalent association with the N-terminal domains, which is sufficient to initiate downstream inflammasome activation (Chui et al., 2019; Sandstrom et al., 2019). By directly coupling NLRP1 inflammasome activation to cleavage by a pathogen-encoded protease, NLRP1 can directly sense and respond to the activity of a pathogen effector. Such a model indicates that the N-terminal domains are not required for NLRP1 activation per se, but rather serve a pathogen-sensing function. Interestingly, the N-terminal 'linker' region in mouse NLRP1B that is cleaved by LF is rapidly evolving in

rodents, and the analogous linker region is likewise rapidly evolving in primate species (Chavarría-Smith et al., 2016). These data suggest that selection from pathogens has been driving diversification of this protease target region of NLRP1, possibly serving to bait diverse pathogenic proteases into cleaving NLRP1 and activating the inflammasome responses.

Consistent with the rapid evolution in NLRP1 at the site of proteolytic cleavage, LF neither cleaves nor activates human NLRP1 (Mitchell et al., 2019; Moayeri et al., 2012; Taabazuing et al., 2020). Despite this, human NLRP1 can also be activated by proteolysis when a tobacco etch virus (TEV) protease site is engineered into the rapidly evolving linker region of human NLRP1 that is analogous to the site of LF cleavage in mouse NLRP1B (Chavarría-Smith et al., 2016). Thus, like mouse NLRP1B, it has been predicted that human NLRP1 may serve as a 'tripwire' sensor for pathogen-encoded proteases (Mitchell et al., 2019).

Here, we investigate the hypothesis that viral proteases cleave and activate human NLRP1. We reasoned that human viruses, many of which encode proteases as necessary enzymes for their replication cycle, may be triggers for NLRP1 activation. To pursue this hypothesis, we focused on viruses in the *Picornaviridae* family, which encompass a diverse set of human enteric and respiratory pathogens including coxsackieviruses, polioviruses, and rhinoviruses (Zell, 2018). These viruses all translate their genome as a single polyprotein, which is then cleaved into mature proteins in at least six sites in a sequence-specific manner by a virally encoded 3C protease, termed 3Cpro (Laitinen et al., 2016; Solomon et al., 2010; Sun et al., 2016; Zell, 2018). 3Cpro is also known to proteolytically target numerous host factors, many of which are

associated with immune modulation (Croft et al., 2018; Huang et al., 2015; Lei et al., 2017; Mukherjee et al., 2011; Qian et al., 2017; Wang et al., 2019; Wang et al., 2012; Wang et al., 2014; Wang et al., 2015; Wen et al., 2019; Xiang et al., 2014; Xiang et al., 2016; Zaragoza et al., 2006). Because 3Cpro are evolutionarily constrained to cleave several specific polyprotein sites and host targets for replicative success, we reasoned that human NLRP1 may have evolved to sense viral infection by mimicking viral polyprotein cleavage sites, leading to NLRP1 cleavage and inflammasome activation. Using an evolution-guided approach, we now show that NLRP1 is cleaved in its rapidly evolving linker region by several 3Cpro from picornaviruses, resulting in inflammasome activation. These results are consistent with the recent discovery that human rhinovirus (HRV) 3Cpro cleaves and activates NLRP1 in airway epithelia (Robinson et al., 2020). We also find that variation in the cleavage site among primates, and even within the human population, leads to changes in cleavage susceptibility and inflammasome activation. Interestingly, we observe that proteases from multiple genera of viruses cleave and activate human NLRP1 and mouse NLRP1B at different sites, supporting a role for an evolutionary conflict between viral proteases and NLRP1. Taken together, our work highlights the role of NLRP1 in sensing and responding to diverse viral proteases by evolving cleavage motifs that mimic natural sites of proteolytic cleavage in the viral polyprotein.

Results

Human NLRP1 contains mimics of viral protease cleavage sites

Our hypothesis that NLRP1 can sense viral proteases is based on two prior observations. First, both human NLRP1 and mouse NLRP1B can be activated by N-terminal proteolysis (Chavarría-Smith et al., 2016). Second, the linker in primate NLRP1, which is analogous to the N-terminal disordered region of NLRP1B that is cleaved by LF protease, has undergone recurrent positive selection (Chavarría-Smith et al., 2016), or an excess of non-synonymous amino acid substitutions over what would be expected by neutral evolution (Kimura, 1983). We reasoned that this rapid protein sequence evolution may reflect a history of pathogen-driven selection, wherein primate NLRP1 has evolved to sense pathogen-encoded proteases such as those encoded by picornaviruses. To test this hypothesis, we first generated a predictive model for 3Cpro cleavage site specificity. We chose to focus on the enterovirus genus of picornaviruses, as there is a deep and diverse collection of publicly available viral sequences within this genus due to their importance as human pathogens including coxsackieviruses, polioviruses, enterovirus D68, and HRV (Blom et al., 1996; Pickett et al., 2012). We first compiled complete enterovirus polyprotein sequences from the Viral Pathogen Resource (ViPR) database (Pickett et al., 2012) and extracted and concatenated sequences for each of the cleavage sites within the polyproteins (Figure 3.1A and B, Supplementary files 3.1 and 3.2). After removing redundant sequences, we used the MEME Suite (Bailey et al., 2009) to create the following 3Cpro cleavage motif: [A/Φ]XXQGXXX (where Φ denotes a hydrophobic residue and X denotes any amino acid), which is broadly consistent with previous studies that have experimentally profiled the substrate specificity of enterovirus 3Cpros (Blom et al., 1996; Fan et al., 2020; Jagdeo et al., 2018; O'Donoghue et al., 2012; Figure 3.1C).

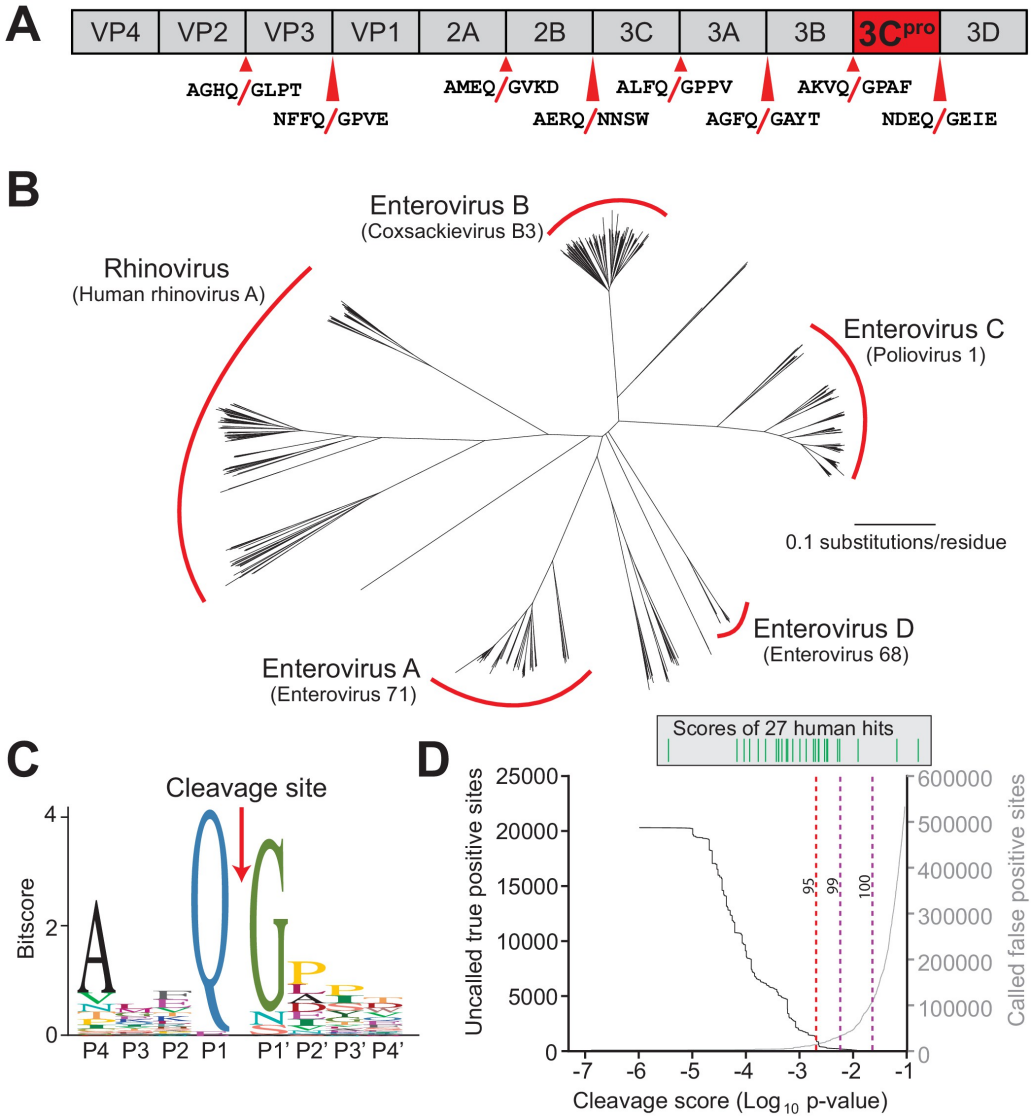


Figure 3.1. Conserved polyprotein cleavage sites across enteroviruses inform substrate specificity of the enteroviral 3Cpro. (A) Schematic of 3Cpro cleavage sites (red arrows) within the polyprotein of coxsackievirus B3 Nancy (CVB3), a model enterovirus. Shown are the eight amino acids flanking each cleavage site within the polyprotein. (B) Phylogenetic tree of 796 enteroviral polyprotein coding sequences depicting the major clades of enteroviruses sampled in this study with representative viruses from each clade in parentheses (Supplementary file 3.2). (C) Eight amino acid polyprotein cleavage motif for enteroviruses (labeled as positions P4 to P4') generated from the 796 enteroviral polyprotein sequences in (B) using the MEME Suite (Supplementary file 3.2). (D) Training set data used to determine the motif search threshold for FIMO (Supplementary files 3.1, 3.3 and 3.4). The X-axis represents a log10 of the p-value reported by FIMO as an indicator for the strength of the cleavage motif hit (cleavage score). (Left) The Y-axis depicts the number of uncalled true positives, or motif hits that overlap with the initial set of 8mer polyprotein cleavage sites used to generate the motif, in the training set of enteroviral polyprotein sequences (black). (Right) The Y-axis depicts the number of called false positive sites, or any motif hits found in the polyprotein that are not known to be cleaved by 3Cpro, in the training set of enteroviral polyprotein sequences (gray). (Above) Each line depicts a single, experimentally validated case of enteroviral 3Cpro cleavage site within a human protein as reported in Laitinen et al., 2016 and is ordered along the X-axis by its resulting cleavage score. A vertical dotted line is used to represent the decided threshold that captures 95% of true positive hits and 16 out of 27 reported human hits (Figure 3.2).

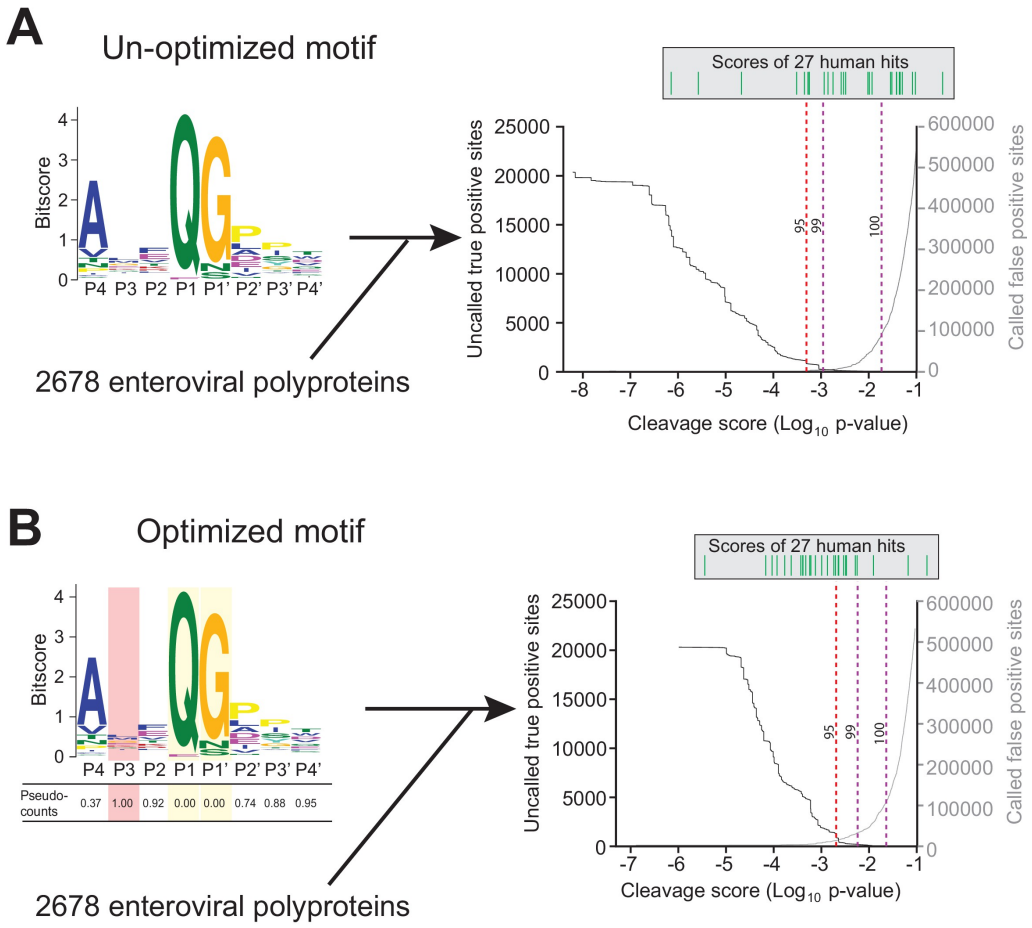


Figure 3.2. Motif optimization enhances capture of known human targets of enteroviral 3Cpro. (A) As described in Figure 3.1B and C and Materials and Methods, the 8mer (P4–P4') 3Cpro polyprotein cleavage motif was initially generated from unique, concatenated 8mer cleavage sites across 796 enteroviral polyprotein sequences. To assess the capture capability of the motif on both virus and host targets, the motif was then used to conduct a low threshold (p-value=0.1) FIMO (MEME Suite) search across training set of 2678 nonredundant enteroviral polyproteins from ViPR and 27 experimentally validated human targets of 3Cpro (Laitinen et al., 2016). In the graph, the X-axis represents a log10 of the p-value reported by FIMO as an indicator for the strength of the cleavage motif hit, or cleavage score. The left Y-axis depicts the number of uncalled 'true positives', or motif hits within the enteroviral polyprotein training set that overlap with the initial set of 8mer polyprotein cleavage sites used to generate the motif (black). The right Y-axis depicts the number of called false positive sites, or any motif hits that are not true positives, in the training set of enteroviral polyprotein sequences (gray). (Above) Each line depicts a single, experimentally validated case of enteroviral 3Cpro cleavage site within a human protein as reported in Laitinen et al., 2016 and is ordered along the x-axis by its corresponding cleavage score. Vertical dotted lines are used to represent the decided thresholds for comparison of capture capability. Capture of human targets at 95%, 99%, or 100% capture of true positives in the polyprotein dataset corresponds to capture of 4, 7, and 16 human hits. (B) Pseudo-counts to the position-specific scoring matrix of the motif shown in (A) were adjusted by total information content where the two most information-dense positions P1 and P1' are assigned pseudocount = 0 and the least information-dense position P3 pseudocount = 1, and the remaining positions are assigned a pseudocount value relative to the most information-dense position P1. This optimized motif is then used to FIMO search against the same training set as described in (A). Capture of human targets at 95%, 99%, or 100% capture of true positives in the polyprotein dataset corresponds to capture of 16, 23, and 24 human hits.

We next optimized our 3Cpro cleavage site motif prediction by querying against predicted viral polyprotein and experimentally validated host cleavage sites (Laitinen et al., 2016), allowing us to set thresholds for predicting new cleavage sites (Supplementary files 3.3 and 3.4). Due to the low-information content of the polyprotein motif (Figure 3.1C), such predictions are necessarily a compromise between stringency and capturing the most known cleavage sites. In particular, we wished to make sure that the model was able to capture a majority of experimentally validated human hits (compiled in Laitinen et al., 2016) in addition to the known sites of polyprotein cleavage ('true positives'), while minimizing the prediction of sites outside of known polyprotein cleavage sites ('false positives'). By adjusting the model to allow greater flexibility for amino acids not sampled in the viral polyprotein (see Materials and methods and Figure 3.2 and Supplementary file 3.4), we were able to capture 95% of known viral sites and the majority of the known human hits, while limiting the number of false negative hits within the viral polyprotein (Figure 3.1D).

The coxsackievirus B3 3Cpro cleaves human NLRP1 at a predicted site within the linker region

We next used our refined model to conduct a motif search for 3Cpro cleavage sites in NLRP1 using Find Individual Motif Occurrences (FIMO) (Grant et al., 2011). We identified three occurrences of the motif across the full-length human NLRP1 protein (Figure 3.3A). Of these sites, one in particular, 127-GCTQGSER-134, fell within the previously described rapidly evolving linker (Chavarría-Smith et al., 2016) and

demonstrates the lowest percent conservation across mammalian species at each of the predicted P4-P4' positions (Figure 3.3B).

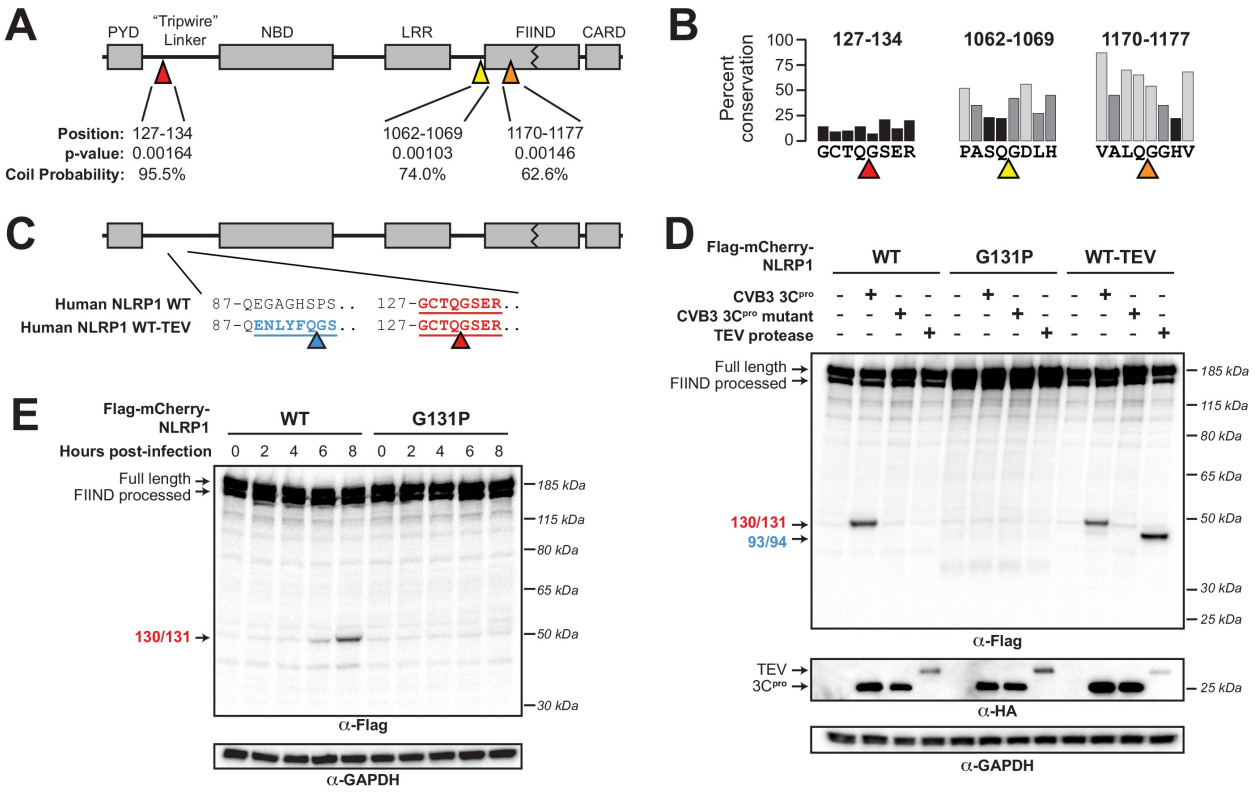


Figure 3.3. Enterovirus 3Cpro cleaves human NLRP1 at the predicted site of mimicry. (A) Schematic of the domain structure of NLRP1, with predicted cleavage sites (triangles). FIMO-reported p-values and average NetsurfP-reported coil probabilities are described at the predicted sites. (B) Percent conservation across 100 mammalian species at each position of each predicted 8mer cleavage site within human NLRP1. (C) Schematic of the human NLRP1 sequence used to assess enteroviral cleavage and activation. The predicted enteroviral cleavage site found in the linker region (127-GCTQGSER-134) is shown in red. Human NLRP1 WT-TEV contains an engineered TEV cleavage site between residues 93 and 94 (underlined green) in human NLRP1 WT. (D) Immunoblot depicting human NLRP1 cleavage by CVB3 3Cpro and TEV protease. HEK293T cells were co-transfected using 100 ng of the indicated Flag-tagged mCherry-NLRP1 fusion plasmid constructs with 250 ng of the indicated protease construct and immunoblotted with the indicated antibodies. (E) Immunoblot depicting human NLRP1 cleavage at the indicated timepoints after infection with 250,000 PFU (MOI = ~1) CVB3. HEK293T cells were transfected using 100 ng of either WT NLRP1 or NLRP1 G131P and infected 24–30 hr later. All samples were harvested 32 hr post-transfection and immunoblotted with the indicated antibodies.

To assess if human NLRP1 is cleaved by enteroviral 3Cpro, we co-expressed a N-terminal mCherry-tagged wild-type (WT) human NLRP1 with the 3Cpro from the model enterovirus, coxsackievirus B3 (CVB3) in HEK293T cells (Figure 3.3C). The

mCherry tag stabilizes and allows visualization of putative N-terminal cleavage products, similar to prior studies (Chavarría-Smith et al., 2016). We observed that the WT but not catalytically inactive (C147A) CVB3 3Cpro cleaved NLRP1, resulting in a cleavage product with a molecular weight consistent with our predicted 3Cpro cleavage at the predicted 127-GCTQGSER-134 site (44 kDa) (Figure 3.3D). Based on the presence of a single cleavage product, we assume that the other predicted sites are either poor substrates for 3Cpro or less accessible to the protease as would be predicted from their NetSurfP-reported (Klausen et al., 2019) coil probability within structured domains of the protein (Figure 3.3A). To determine if the cleavage occurs between residues 130 and 131, we mutated the P1' glycine to a proline (G131P), which abolished 3Cpro cleavage of NLRP1 (Figure 3.3D). CVB3 3Cpro cleavage of NLRP1 resulted in a similarly intense cleavage product when compared to the previously described system in which a TEV protease site was introduced into the linker region of NLRP1 (Chavarría-Smith et al., 2016; Figure 3.3D). Taken together, these results indicate that cleavage of WT NLRP1 by a protease from a natural human pathogen is robust and specific.

During a viral infection, 3Cpro is generated in the host cell cytoplasm after translation of the viral mRNA to the polyprotein and subsequent processing of the viral polyprotein into constituent pieces (Laitinen et al., 2016). To confirm that virally-produced 3Cpro, or the 3 CD precursor that can also carry out proteolytic cleavage during infection (Laitinen et al., 2016), is able to cleave NLRP1, we virally infected cells expressing either WT NLRP1 or the uncleavable (G131P) mutant. We observed accumulation of the expected cleavage product beginning at 6 hr post-infection when

we infected cells expressing WT NLRP1 and no cleavage product when we infected cells expressing the 131P mutant (Figure 3.3E). These results validate that CVB3 infection can result in rapid and specific cleavage of human NLRP1.

The CVB3 3Cpro activates human NLRP1 by cleaving within the linker region

Previous results with a TEV-cleavable human NLRP1 showed that cleavage by TEV protease was sufficient to activate the human NLRP1 inflammasome in a reconstituted inflammasome assay (Chavarría-Smith et al., 2016). Using the same assay, in which plasmids-encoding human NLRP1, CASP1, ASC, and IL-1β are transfected into HEK293T cells, we tested if the CVB3 3Cpro activates the NLRP1 inflammasome. We observed that the CVB3 3Cpro results in robust NLRP1 inflammasome activation, as measured by CASP1-dependent processing of pro-IL-1β to the active p17 form (Figure 3.4A). As expected, CVB3 3Cpro activation of the NLRP1 inflammasome was prevented by introduction of a mutation in the NLRP1 FIIND (S1213A) (D'Osualdo et al., 2011; Finger et al., 2012; Frew et al., 2012; Figure 3.5 – panel A), which prevents FIIND auto-processing and the release of the bioactive C-terminal UPA–CARD (Chui et al., 2019; Sandstrom et al., 2019). Consistent with recent results (Robinson et al., 2020), we also observed that chemical inhibitors of the proteasome (MG132) or the Cullin-RING E3 ubiquitin ligases that are required for the degradation of proteins with a novel N-terminal glycine (MLN4924) (Timms et al., 2019), also blocked CVB3 3Cpro activation of NLRP1 (Figure 3.5 – panel B). To confirm that 3Cpro-induced inflammasome activation resulted in release of bioactive IL-1β from cells, we measured active IL-1β levels in the culture supernatant using cells engineered

to express a reporter gene in response to soluble, active IL-1β. When compared to a standard curve (Figure 3.6), we found that 3Cpro treatment resulted in release of >4 ng/ml of active IL-1β into the culture supernatant (Figure 3.4B). Importantly, in both western blot and cell culture assays, 3Cpro-induced inflammasome activation was comparable to TEV-induced activation and was ablated when position 131 was mutated, validating that CVB3 3Cpro cleavage at a single site is both necessary and sufficient to activate NLRP1 (Figure 3.4A and B). Taken together, our results are consistent with CVB3 3Cpro activating the NLRP1 inflammasome via site-specific cleavage and subsequent 'functional degradation' (Chui et al., 2019; Sandstrom et al., 2019).

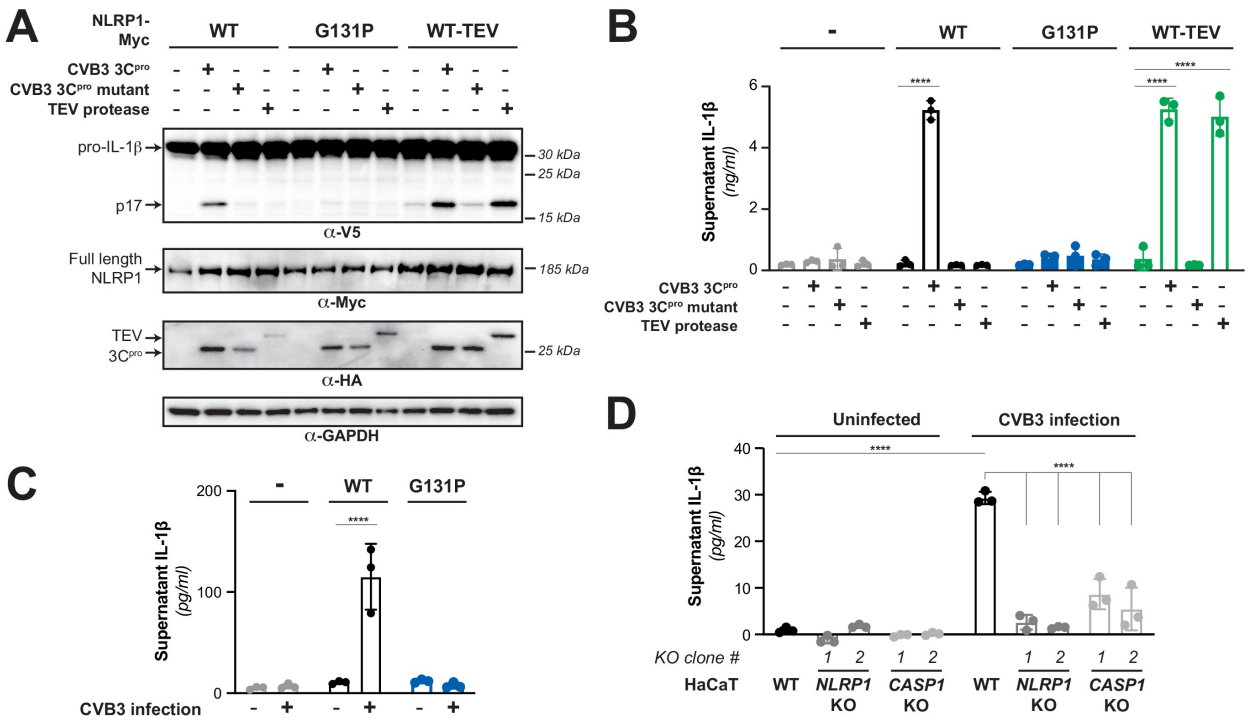


Figure 3.4. Enterovirus 3Cpro cleavage of human NLRP1 promotes pro-inflammatory cytokine release. (A) Immunoblot depicting human NLRP1 activation (maturation of IL-1β) by CVB3 3Cpro and TEV protease. HEK293T cells were co-transfected using 100 ng of the indicated protease, 50 ng V5-IL-1β, 100 ng CASP1, 5 ng ASC, and 4 ng of the indicated Myc-tagged NLRP1, and immunoblotted with the indicated antibodies. Appearance of the mature p17 band of IL-1β indicates successful assembly of the NLRP1 inflammasome and activation of CASP1. (B) Bioactive IL-1β in the culture supernatant was measured using HEK-Blue IL-1β reporter cells, which express secreted embryonic alkaline phosphatase (SEAP) in response to extracellular IL-1β. Supernatant from cells transfected as in (A) was added to HEK-Blue IL-1β reporter cells and SEAP levels in the culture supernatant from HEK-Blue IL-1β reporter cells were quantified by the QUANTI-Blue colorimetric substrate. Transfections were performed in triplicate and compared to the standard curve generated from concurrent treatment of HEK-Blue IL-1β reporter cells with purified human IL-1β (Figure 3.6). Data were analyzed using two-way ANOVA with Sidak's post-test. **** = p<0.0001. (C) CVB3 infection of inflammasome-reconstituted HEK293T cells results in IL-1β release when NLRP1 can be cleaved by 3Cpro. Cells were transfected with the indicated NLRP1 construct and other NLRP1 inflammasome components as in (B). Sixteen hours post-transfection, cells were mock infected or infected with 250,000 PFU (MOI = ~1) CVB3. Eight hours post-infection, culture supernatant was collected and bioactive IL-1β was measured as in (B). (D) CVB3 infection of an immortalized human keratinocyte cell line, HaCaT, activates the NLRP1 inflammasome. WT or knockout (Figure 3.7) HaCaT cell lines were mock infected or infected with 100,000 PFU (MOI = ~0.4) CVB3. Forty-eight hours post-infection, culture supernatant was collected and bioactive IL-1β was measured as in (B).

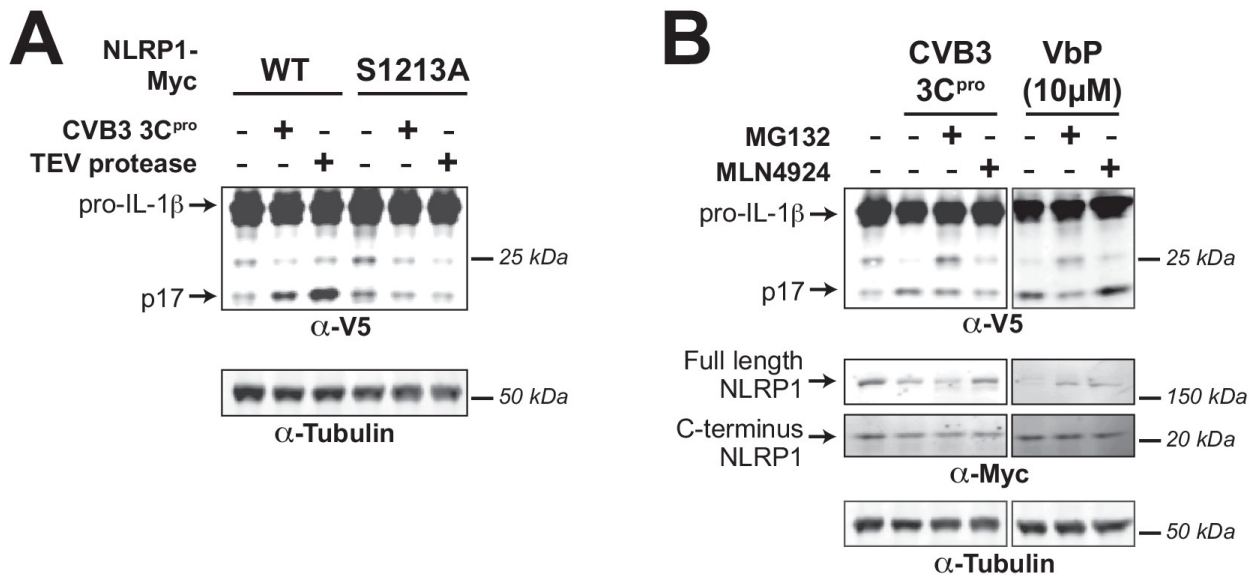


Figure 3.5. 3Cpro-mediated activation of the human NLRP1 inflammasome depends on FIIND autoprocessing and proteosomal degradation. (A) HEK293T cells were transfected with either WT NLRP1 or a FIIND auto-processed defective mutant (S1213A) along with other components of the NLRP1 inflammasome (CASP1, ASC, and IL-1β) as in Figure 3.4A. Only cells transfected with WT NLRP1 can produce mature IL-1β upon co-transfection with CVB3 3Cpro or TEV protease as indicated by the appearance of the p17 band. (B) HEK293T cells were transfected as in (A), and then treated with the indicated inhibitors of proteasomal-mediated degradation (0.5 μM MG132) or the N-glycine degron pathway (1.0 μM MLN4924) for 6 hr prior to harvest. Addition of 10 μM VbP, an inhibitor of the NLRP1 inhibitors DPP8/9 (Okondo et al., 2018), was used as a control for protease-independent activation of the inflammasome and is thus unaffected by MLN4924.

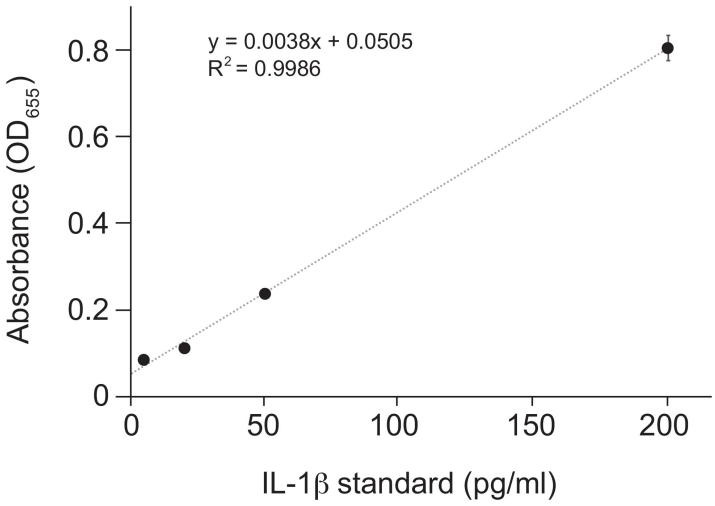


Figure 3.6. Standard curve for Figure 3.4B. Purified human IL-1β was added in duplicate to the indicated final concentration to HEK-Blue IL-1β reporter cells and SEAP activity was measured by increased absorbance at OD655. The indicated linear fit was used to calculate absolute concentrations of bioactive IL-1β from culture supernatants shown in Figure 3.4B. Note that supernatants from inflammasome-transfected cells was diluted 10-fold before addition to HEK-Blue IL-1β reporter cells to ensure that levels fell within the linear range of the indicated standard curve. Standard curves were generated in an identical manner for each panel of HEK-Blue data shown.

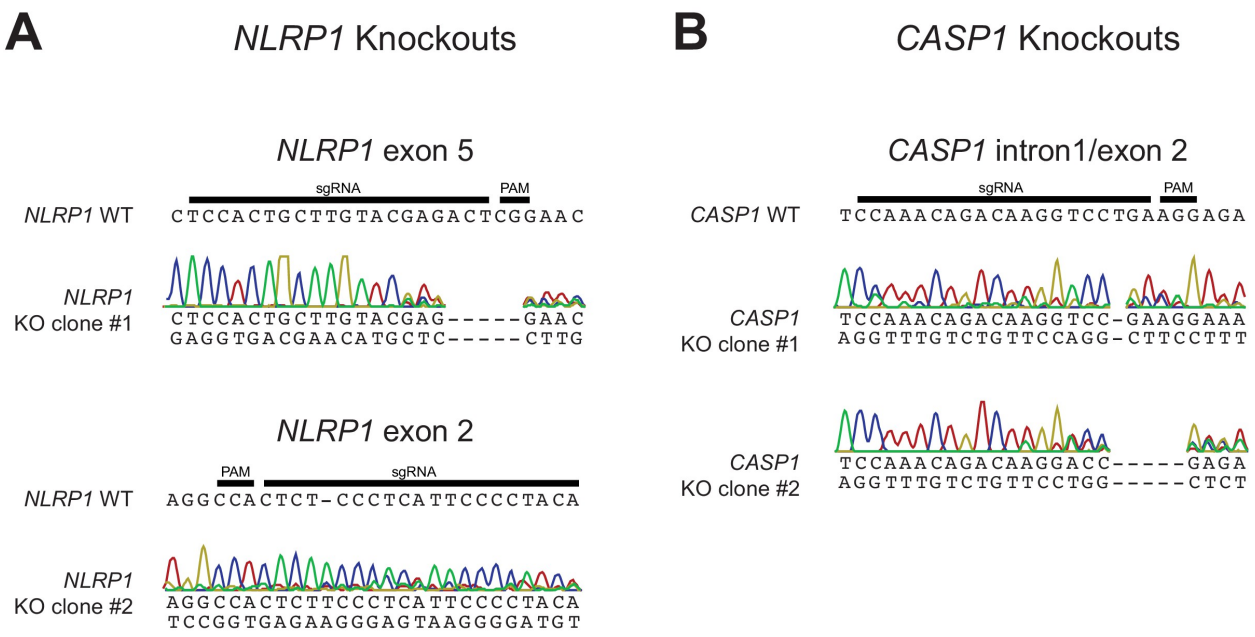


Figure 3.7. Validation of CRISPR/Cas9-editing of NLRP1 or CASP1 in HaCaT cells by Sanger sequencing. (A) The WT NLRP1 (exon 5 or exon 2) sequence is shown with CRISPR-targeting sgRNA and PAM sequences indicated. Below each WT sequence is shown the Sanger sequencing chromatogram and associated mutant sequence for each indicated clone. (B) Same as panel A, except indicating WT CASP1 intron 1/exon two sequence and sequencing data from two independently isolated knockout clones.

We next wished to test whether CVB3 infection, through the site-specific cleavage of NLRP1 by 3Cpro, can activate the NLRP1 inflammasome. Consistent with our prediction, recent work has revealed that HRV infection can cleave and activate human NLRP1 in airway epithelia (Robinson et al., 2020). However, prior work has also implicated a role for the NLRP3 inflammasome in enterovirus infection (Kuriakose and Kanneganti, 2019; Xiao et al., 2019), including activation of the NLRP3 inflammasome during CVB3 infection in mice and human cell lines (Wang et al., 2019; Wang et al., 2018). NLRP1 and NLRP3 have distinct expression patterns (Robinson et al., 2020; Zhong et al., 2016) including in epithelial cells, which are important targets of enterovirus infection. NLRP3 is activated in response to various noxious stimuli or

damage signals associated with pathogen infection (Evavold and Kagan, 2019; Spel and Martinon, 2021). In contrast, NLRP1 is activated by direct proteolytic cleavage of its N-terminal ‘tripwire’ region by viral proteases. We therefore wished to confirm that specific 3Cpro cleavage of NLRP1 during CVB3 infection is able to activate the NLRP1 inflammasome. We first virally infected 293 T cells, which do not express either NLRP1 or NLRP3, that were co-transfected with either WT NLRP1 or the uncleavable (G131P) mutant in our reconstituted inflammasome assay and measured active IL-1β in the culture supernatant. Eight hours after infection with CVB3, we observe robust release of active IL-1β into the culture supernatant when cells were transfected with WT NLRP1 but not the uncleavable mutant NLRP1 (Figure 3.4C). To test whether CVB3 infection can activate the inflammasome in an NLRP1-dependent fashion in cells that naturally express an intact NLRP1 inflammasome, we took advantage of the fact that NLRP1 has been described as the primary inflammasome in human keratinocytes (Zhong et al., 2016). We therefore infected WT, NLRP1, or CASP1 KO (Figure 3.7) immortalized HaCaT human keratinocytes with CVB3 and measured release of active IL-1β in the culture supernatant. Consistent with our model that CVB3 infection cleaves and activates the NLRP1 inflammasome, we observe a significant increase in supernatant IL-1β after CVB3 infection that is reduced in cells that lack either NLRP1 or CASP1 (Figure 3.4D). Together, these results indicate that CVB3 infection, through 3Cpro cleavage of the tripwire region of NLRP1, activates the NLRP1 inflammasome.

NLRP1 diversification across primates and within humans confers host differences in susceptibility to viral 3Cpro cleavage and inflammasome activation

Our evolutionary model in which NLRP1 is evolving in conflict with 3Cpro suggests that changes in the NLRP1 linker region, both among primates and within the human population (Figure 3.8A), would confer host-specific differences to NLRP1 cleavage and inflammasome activation. To test this hypothesis, we aligned the linker regions from NLRP1 from diverse mammals and human population sampling and compared the sequences around the site of CVB3 3Cpro cleavage (Figure 3.8B and C and Figure 3.9). We noted that while a majority of primate NLRP1s are predicted to be cleaved similarly to the human ortholog, several primate proteins would be predicted to not be cleaved by enteroviral 3Cpro as a result of changes to either the P4, P1 or P1' residues. To confirm these predictions, we made the human NLRP1 mutants G127E or G131R, which reflect the Old World monkey or marmoset residues at each position, respectively. As predicted, both primate NLRP1 variants prevented 3Cpro cleavage of NLRP1 (Figure 3.8D). These results indicate that multiple viral 3Cpro activate host NLRP1 in a host specific manner and suggest that single changes within a short linear motif can substantially alter cleavage susceptibility and inflammasome activation.

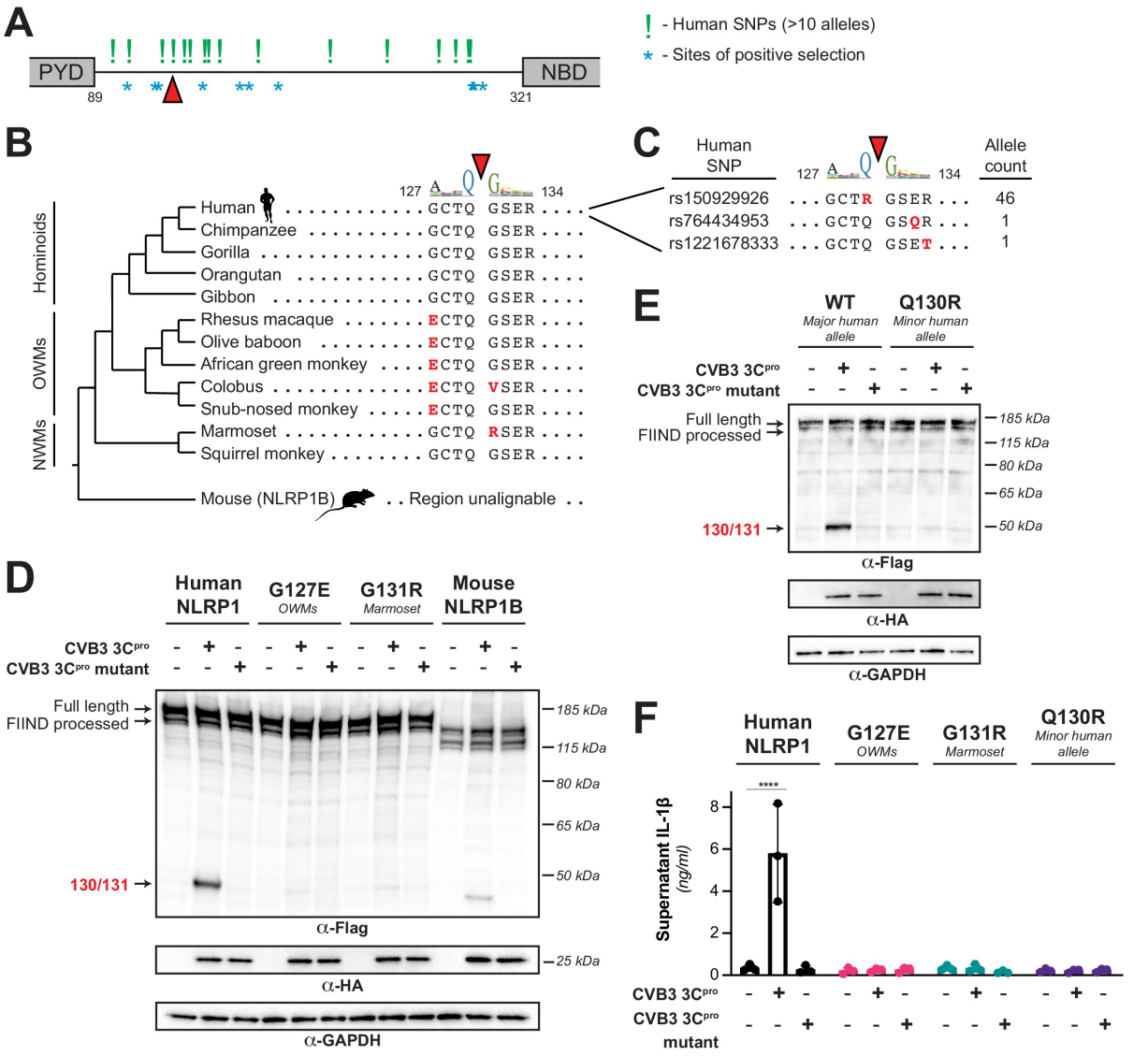


Figure 3.8. Naturally occurring cleavage site variants alter NLRP1 susceptibility to enteroviral 3Cpro. (A) Schematic of sites found to be evolving under positive selection (marked as *, from Chavarría-Smith and Vance, 2013) and human SNPs with at least 10 reported instances in the Genome Aggregation Database (GnomAD, Karczewski et al., 2020) (marked as !) within the linker region between the pyrin domain (PYD) and nucleotide binding domain (NBD) of NLRP1. The enteroviral 3Cpro cleavage site between position 130 and 131 is indicated by a red triangle. (B) Phylogenetic tree depicting the enteroviral 3Cpro cleavage site (red triangle) within NLRP1 across three clades of primates – hominoids, Old World monkeys (OWMs), and New World monkeys (NWMs). Mouse NLRP1B lacks any sequence that is alignable to this region of primate NLRP1 (see also Figure 3.9). Amino acid differences to the human NLRP1 reference sequence are highlighted in red. Above the alignment is the enterovirus 3Cpro sequence logo shown in Figure 3.1. (C) GnomAD-derived allele counts of each missense human SNP (by reference SNP #) within the 8mer of the determined enteroviral 3Cpro cleavage site. (D–E) Immunoblot depicting CVB3 3Cpro cleavage susceptibility of the indicated 8mer site variants introduced into human NLRP1 or full-length wild-type mouse NLRP1B (129 allele) (D) or the cleavage susceptibility of human NLRP1 Q130R, a naturally occurring human population variant (E). (F) Release of bioactive IL-1β into the culture supernatant was measured using HEK-Blue IL-1β reporter cells as in Figure 3.4B. Data were analyzed using two-way ANOVA with Sidak's post-test. **** = p<0.0001.

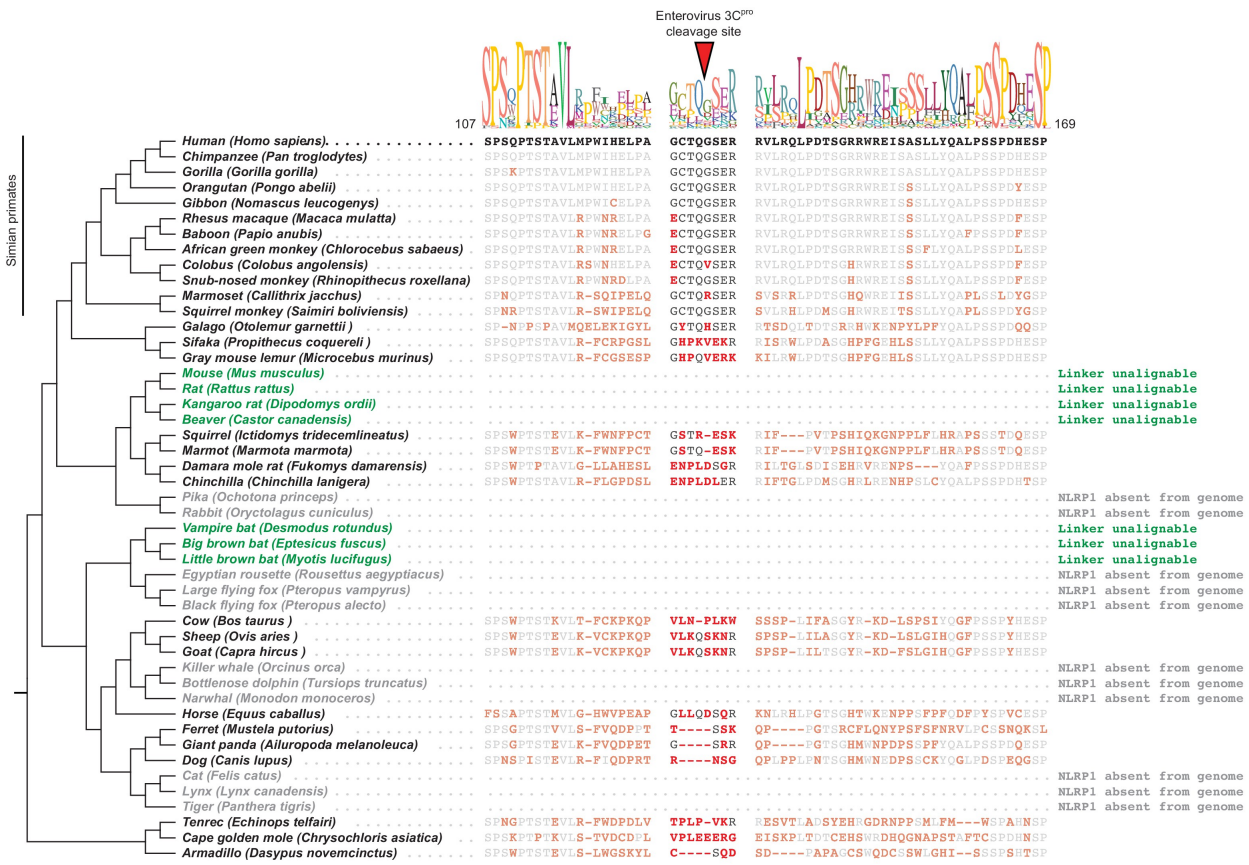


Figure 3.9. Mammalian NLRP1 phylogenomics and alignment of linker region. The indicated mammalian NLRP1 sequences were aligned and the region corresponding to residues 107–169 from human NLRP1 were extracted, which is anchored on both ends by well-conserved proline and serine-rich motifs. A consensus sequence generated from alignable sequences in this region is shown above the human sequence. The position of the CVB3 3Cpro cleavage site in human NLRP1 is shown, flanked by four amino acids on both sides (P4->P4'). In other mammals, residues that differ from the human sequence are shown in red. Within the aligned region that corresponds to the CVB3 3Cpro cleavage site, only simian primates have P4, P1 and P1' residues that would allow cleavage. The only other species that have a plausible cleavage site in this position are sheep and goats (P4 = Val, P1 = Gln, P1' = Ser), although those residues appear to have evolved independently at those positions. Two clades of species (the 'mouse-related' clade of rodents and the microbat clade, marked as green) have NLRP1 protein sequences with N-terminal linkers that are unalignable to human NLRP1 in this region. Four additional clades (lagomorphs, megabats, cetaceans, and felines, marked as gray) lack the NLRP1 gene altogether.

We further observed that this cleavage site is largely absent in non-primate species (Figure 3.9), suggesting that a 3Cpro cleavage site mimic emerged in simian primates 30–40 million years ago. While many other mammalian species have a region that is alignable to the primate linker, we noted that this region is unalignable to any sequence in the linker region of NLRP1 proteins from rodents or bats (Figure 3.8B and Figure 3.9). Despite this, we found that there was weak cleavage of mouse NLRP1B at a site closer to the N-terminus than the 127-GCTQGSER-134 site found in human NLRP1 (Figure 3.8D and Figure 3.15A), suggesting that an independent cleavage site could have arisen elsewhere in mouse NLRP1B. These data suggest that NLRP1 in other mammals may have convergently evolved cleavage sites in the linker region despite not having a cleavable sequence in the precise position that human NLRP1 is cleaved.

Differential host susceptibility to NLRP1 cleavage and activation extends to the human population level. Using GnomAD (Karczewski et al., 2020), we sampled the alternative alleles within the direct cleavage site (Figure 3.8C). While this region does not appear to be highly polymorphic in humans, we note that one alternative allele (rs150929926) results in a Q130R mutation and is present in >1 in every 1000 African alleles sampled. Introducing this mutation into NLRP1, we find the Q130R mutation eliminates NLRP1 cleavage susceptibility to CVB3 3Cpro (Figure 3.8E). In the case of primate and human diversity alleles at the site of 3Cpro cleavage, we also find that loss of cleavage susceptibility results in a loss of inflammasome activation in response to 3Cpro (Figure 3.8F), supporting the aforementioned notion that single changes in the

linker region can have drastic impacts on the ability of different hosts to respond to the presence of cytoplasmic 3Cpro.

3Cpro from diverse picornaviruses cleave and activate human NLRP1

Our evolutionary model predicted that NLRP1 would be cleaved by a broad range of 3Cpro from viruses in the enterovirus genus (Figure 3.1B). To test this hypothesis, we cloned 3Cpro from representative viruses from four additional major species of human enteroviruses: enterovirus 71 (EV71, species: Enterovirus A), poliovirus 1 (PV1, species: Enterovirus C), enterovirus D68 (EV68, species: Enterovirus D), human rhinovirus A (HRV, species: Rhinovirus A), in order to compare them to the 3Cpro from CVB3 (species: Enterovirus B) (Figure 3.10A). Despite <50% amino acid identity between some of these proteases (Figure 3.11), the overall structures of these proteases are similar (Figure 3.12) and the cleavage motifs are closely related (Figure 3.10A). Consistent with this predicted target similarity and prior data with HRV (Robinson et al., 2020), we found that every tested member of enterovirus 3Cpro was able to cleave NLRP1 between residues 130 and 131 (Figure 3.10B). Moreover, expression of every tested enterovirus 3Cpro resulted in activation of the inflammasome in a manner that was dependent on cleavage at the 127-GCTQGSER-134 site (Figure 3.10C).

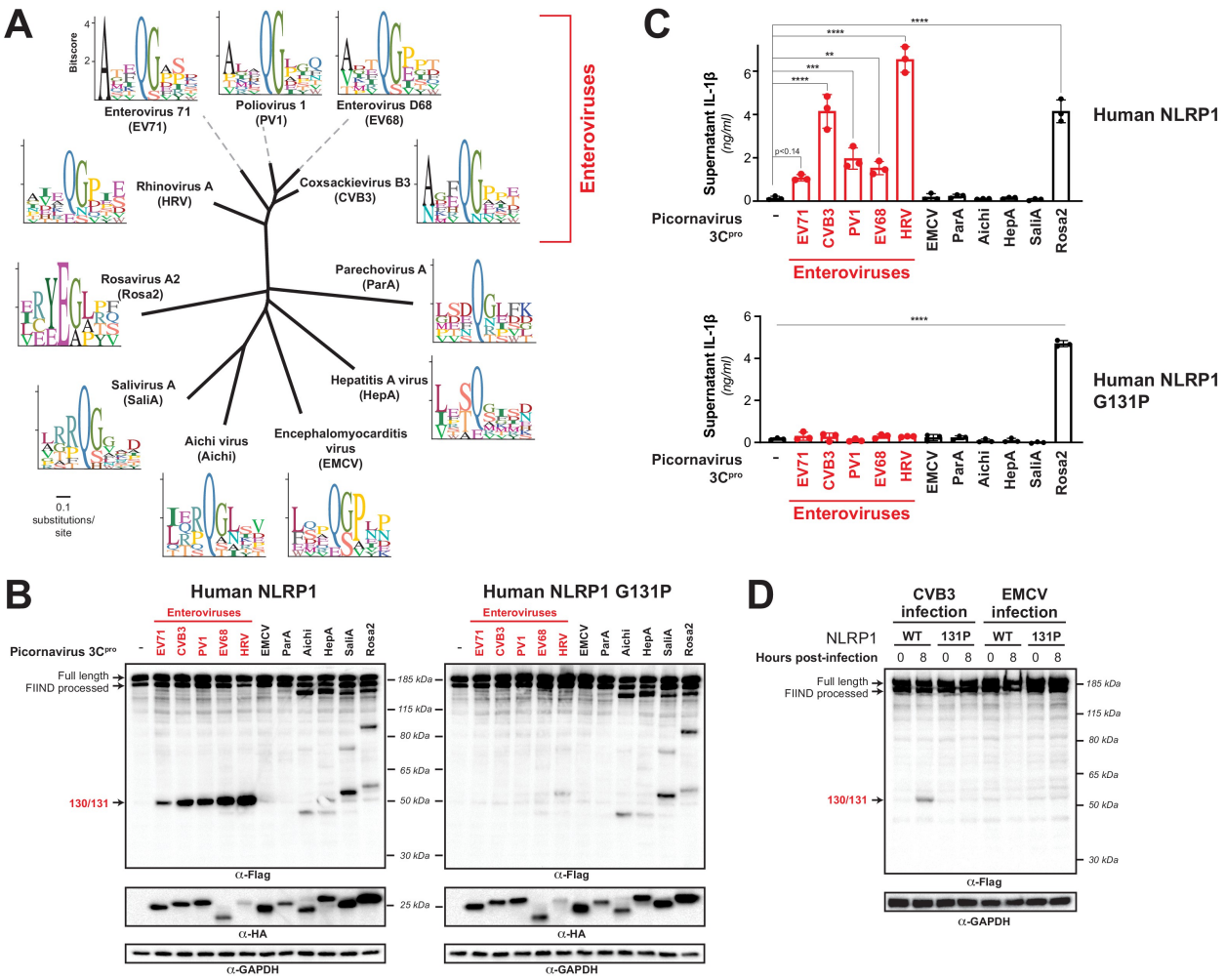


Figure 3.10. Diverse picornavirus 3Cpros cleave and activate NLRP1 at independently evolved sites. (A) Phylogenetic tree of 3Cpro protein sequences for the indicated picornaviruses (Figure 3.11). Shown next to the virus name is the sequence motif generated from the known sites of 3Cpro polyprotein cleavage in that specific virus. (B) Immunoblot depicting human NLRP1 cleavage by the indicated picornaviral 3Cpro. Abbreviations are as in (A). Assays were performed as in Figure 3.3D. (left) Cleavage assays against WT NLRP1. (right) Human NLRP1 G131P mutant used in Figure 3.3. (C) Release of bioactive IL-1β into the culture supernatant was measured using HEK-Blue IL-1β reporter cells as in Figure 3.4B. Data were analyzed using one-way ANOVA with Tukey's post-test. ** = p<0.01, *** = p<0.001, **** = p<0.0001. (D) Immunoblot depicting human NLRP1 cleavage at the indicated timepoints after infection with 250,000 PFU (MOI = ~1) CVB3 or EMCV. HEK293T cells were transfected using 100 ng of either WT NLRP1 or NLRP1 G131P and, 24 hr later, either mock infected (0 hr timepoint) or infected with CVB3 or EMCV as indicated (8 hr timepoint). All samples were harvested 32 hr post-transfection and immunoblotted with the indicated antibodies.

```
Coxsackievirus B3 3Cpro     1   GP--AFEFAVAMMK---RNSSTVKTEYG-------EFTMLGIYDRWAVLPRHAKPGP--------TILMN
Enterovirus D68 3Cpro       1   GP--GFDFAQAIMK---KNTVIARTEKG-------EFTMLGVYDRVAVIPTHASVGE--------IIYIN
Poliovirus 1 3Cpro          1   GP--GFDYAVAMAK---RNIVTATTSKG-------EFTMLGVHDNVAILPTHASPGE--------SIVID
Rhinovirus A 3Cpro          1   GP--EEEFGRSILK---NNTCVITTDNG-------KFTGLGIYDKTLIIPTHADPGR--------EVQVN
Enterovirus 71 3Cpro        1   GP--SLDFALSLLR---RNIRQAQTDQG-------HFTMLGVRDRLAILPRHSQPGK--------TIWVE
Hepatitis A virus 3Cpro     1   S---TLEIAGLVRK--NLVQFGVGEKNGCVR---WVMNALGVKDDWLLVPSHAYKFEKDYEMMEFYFNRG
Rosavirus A2 3Cpro          1   GL--PQIYRPVVANCFPINFYDCDPRDNARSGGVFTLTAVGMYDRTYICNAHGFKDA-------THIGLR
EMCV 3Cpro                  1   GPNPTMDFEKFVAK-FVTAPIGFVYPTGV-----STQTCLLVKGRTLAVNRHMAESD------WTSIVVR
Aichi virus 3Cpro           1   G------ISPAVPG-ISNNVVHVESGNGLNK---NVMSGFYIFSRFLLVPTHLREPH------HTTLTVG
Salivirus A 3Cpro           1   G------FDPAVMK-IMGNVDSFVTLSGTKP--IWTMSCLWIGGRNLIAPSHAFVSD---EYEITHIRVG
Parechovirus A 3Cpro        1   AP-----YDGQLEHIISQMAYITGSTTG------HITHCAGYQHDEIILHGHSIKYL--EQEEELTLHYK
                                                                                *

Coxsackievirus B3 3Cpro    51   DQEVGVLDAKEL-VDKDGTNLELTLLKLNRNEKFRDIRGFLAKEEV----EVNEAVLAINTS-KFPNMYI
Enterovirus D68 3Cpro      51   DVETRVLDACAL-RDLTDTNLEITIVKLDRNQKFRDIRHFLPRCED----DYNDAVLSVHTS-KFPNMYI
Poliovirus 1 3Cpro         51   GKEVEILDAKAL-EDQAGTNLEITIITLKRNEKFRDIRPHIPTQIT----ETNDGVLIVNTS-KYPNMYV
Rhinovirus A 3Cpro         51   GIHTKVLDSYDL-YNRDGVKLEITVIKLDRNEKFRDIRKYIPETED----DYPECNLALSAN-QVEPTII
Enterovirus 71 3Cpro       51   HKLINVLDAVEL-VDEQGVNLELTLVTLDTNEKFRDITKFIPEVIT----GASDATLVINTE-HMPSMFV
Hepatitis A virus 3Cpro    63   GTYYSISAGNVVIQSLDVGFQDVVLMKVPTIPKFRDITQHFIKKGDV-PRALNRLATLVTTV-NGTPMLI
Rosavirus A2 3Cpro         62   GRVYPISEINKKHVRRNHRRTDLMIFQIPDGDCVRNILKYFRKSPE----EAPSRSPAVMAVRGKFNIDV
EMCV 3Cpro                 59   GVSHTRSSVKIIAIAKAGKETDVSFIRLSSGPLFRDNTSKFVKASDVLP---HSSSPLIGIMNVDIPMMY
Aichi virus 3Cpro          55   ADTYDWATLQTQ------EFGEITIVHTPTSRQYKDMRRFIGAHP-------HPTGLLVSQF-KAAPLYV
Salivirus A 3Cpro          59   SRTLDVSRVTRV------DDGELSLLSVPDGPEHKSLIRYIRSAS-------PKSGILASKF-SDTPVFV
Parechovirus A 3Cpro       58   NKVFPIEQPSVTQVTLGGKPMDLAIVKCKLPFRFKKNSKYYTNKI-------GTESMLIWMTEQGIITKE

Coxsackievirus B3 3Cpro   114   ---PVGQVTEYGFLNLGGTPTK------RMLMYNFPTRAGQCGGVLMS-----TGKVLGIHVGG-NGHQ-
Enterovirus D68 3Cpro     114   ---PVGQVTNYGFLNLGGTPTH------RILMYNFPTRAGQCGGVVTT-----TGKVIGIHVGG-NGAQ-
Poliovirus 1 3Cpro        114   ---PVGAVTEQGYLNLGGRQTA------RTLMYNFPTRAGQCGGVITC-----TGKVIGMHVGG-NGSH-
Rhinovirus A 3Cpro        114   ---KVGDVVSYGNILLSGNQTA------RMLKYNYPTKSGYCGGVLYK-----IGQILGIHVGG-NGRD-
Enterovirus 71 3Cpro      114   ---PVGDVVQYGFLNLSGKPTH------RTMMYNFPTKAGQCGGVVTS-----VGKIIGIHIGG-NGRQ-
Hepatitis A virus 3Cpro   131   SEGPLKMEEKATYVHKKNDGTTVDLTVDQAWRGKGEGLPGMCGGALVSSNQSIQNAILGIHVAG--GNSI
Rosavirus A2 3Cpro        127   ---LATCVESFAFVQMSGDVNY------GALRYHAMTMPGYCGAPLISNDKA-AEKVLGIHMAS-NGAGI
EMCV 3Cpro                126   T--GTFLKAGVSVPVETGQTFN------HCIHYKANTRKGWCGSAILADLGG-SKKILGFHSAG---SM-
Aichi virus 3Cpro         110   ---RISDNRILDLDFPGVVVCK------QAYGYRAATFEGLCGSPLVTDDPS-GVKILGLHVAGVAGTS-
Salivirus A 3Cpro         114   ---SFWNGKSHSTPLPGVVDEK------DSFTYRCSSFQGLCGSPMIATDPG-GLGILGIHVAGVAGYN-
Parechovirus A 3Cpro      120   ---VQRVHHSGGIKTREGTEST------KTISYTVKSCKGMCGGLLISKVEG-NFKILGMHIAG-NGEM-
                                                             * **              * *

Coxsackievirus B3 3Cpro   169 GFSAA--LLKHYFNDE--------
Enterovirus D68 3Cpro     169 GFAAM--LLHSYFTDTQ-------
Poliovirus 1 3Cpro        169 GFAAA--LKRSYFTQSQ-------
Rhinovirus A 3Cpro        169 GFSAM--LLRSYFTDTQ-------
Enterovirus 71 3Cpro      169 GFCAG--LKRSYFASE--------
Hepatitis A virus 3Cpro   199 LVAKL--VTQEMFQNIDKKIE-SQ
Rosavirus A2 3Cpro        187 AYGTS--VYQSDFENLE-----YE
EMCV 3Cpro                183 GVAAASIISQEMIDAVVQAFE-PQ
Aichi virus 3Cpro         170 GFSAP--IHP-ILGQITQFATTQQ
Salivirus A 3Cpro         174 GFSAR--LTPERVQAFLSHLATPQ
Parechovirus A 3Cpro      179 GVAIPFNFLKNDMSD--------Q
```

Figure 3.11. Alignment of 3Cpros used in this study. Sequences were aligned using MAFFT (Katoh and Standley, 2013) and used to generate the phylogenetic tree shown in Figure 3.10A. Asterisks indicate residues 100% conserved in all sequences. The position of the catalytic cysteine, analogous to C147 in CVB3 3Cpro, is highlighted in yellow.

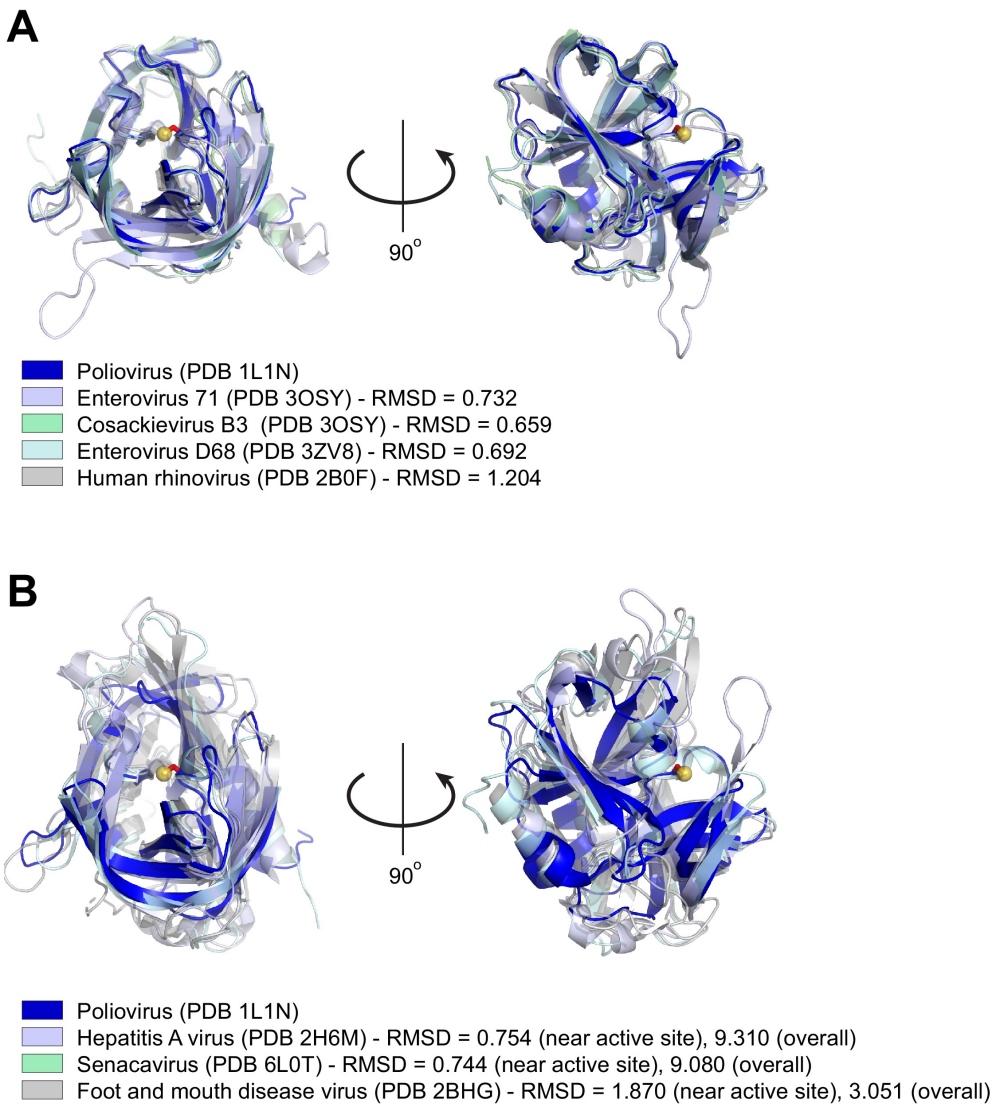


Figure 3.12. Structural similarity of picornavirus 3Cpros. (A) Structures of the indicated enterovirus 3Cpros were aligned to poliovirus 3Cpro using Pymol (https://pymol.org/2/ - The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.). For each enterovirus species used in this study, PDB codes are indicated in parentheses and Pymol-calculated RMSD values are shown. The catalytic cysteine, C147, in the poliovirus 3Cpro structure is shown as a ball-and-stick representation (red stick and yellow ball for the catalytic S atom). Despite substantial divergence in amino acid identity, the overall fold and active site configuration of these enzymes is similar. (B) Structures of the indicated picornavirus 3Cpros were aligned to poliovirus 3Cpro using Pymol, with a similar representation of the catalytic C147 residue in poliovirus 3Cpro as in (A). RMSD values are shown for alignments to either the whole poliovirus 3Cpro structure ('overall' RMSD) or only the 50 amino acids flanking the catalytic cysteine, C147 ('near active site' RMSD). The displayed aligned structures are from the 'near active site' alignment. Neither senecavirus nor foot and mouth disease virus (FDMV) were used in this study, but are included as additional representatives of non-enterovirus 3Cpro enzymes for which there are experimentally determined molecular structures. Compared to proteases aligned in (A), divergent picornavirus proteases show greater divergence in overall fold as indicated by the larger 'overall' RMSD values. However, comparison of the regions around the active site still show high structural similarity, consistent with the constraint on protease evolution and similarity of cleavage specificity shown in Figure 3.10A.

Enteroviruses are only one genus within the broad *Picornaviridae* family of viruses. We next asked if viruses in other *Picornaviridae* genera that infect humans are also able to cleave and activate human NLRP1. We were unable to generate a robust sequence motif for every genera of picornavirus due to lower depth of publicly available sequences. Instead, we cloned a 3Cpro from a representative of every genus of picornavirus that are known to infect humans: encephalomyocarditis virus (EMCV, genus: Cardiovirus), parechovirus A virus (ParA, genus: Parechovirus), Aichi virus (Aichi, genus: Kobuvirus), hepatitis A virus (HepA, genus: Hepatovirus), salivirus A virus (SaliA, genus: Salivirus), and rosavirus A2 (Rosa2, genus: Rosavirus). Each of these viral proteases is <20% identical to CVB3 3Cpro. Despite this, the sequence motif built from cleavage sites within the polyprotein of these individual viruses is broadly consistent with the motif seen in enteroviruses (Figure 3.10A), reflective of the strong evolutionary constraint on evolution of the sequence specificity of these proteases and overall structural conservation of the active sites of these proteases (Figure 3.12). Interestingly, we found that there was substantial variation in NLRP1 cleavage sites across these diverse 3Cpro even though most picornavirus proteases cleaved human NLRP1 to some degree (Figure 3.10B). For instance, while 3Cpro from EMCV and ParA did not cleave NLRP1, we observed distinct cleavage sites for 3Cpro from Aichi, HepA, SaliA and Rosa2 (Figure 3.10B), all of which have at least one cleavage site predicted to occur in the linker region (expected size between 40 kDa and 67 kDa). Confirming that these proteases cleave at a site that is distinct from that of enteroviruses, the G131P NLRP1 mutant is still cleaved by the non-enteroviral proteases (Figure 3.10B).

Surprisingly, when we interrogated NLRP1 inflammasome activation by 3Cpros from Aichi, HepA, SaliA, and Rosa2, all of which robustly cleave NLRP1 at a site in the linker region, we found that only Rosa2 was able to activate the NLRP1 inflammasome (Figure 3.10C). While it is possible that NLRP1 cleavage by 3Cpro from these other viruses is too weak or in a region that may be inconsistent with activation, we also noted that there are obvious cleavage sites in NLRP1 that are outside of the linker region and closer to the FIIND autocleavage site. Cleavage at these sites in NLRP1, or cleavage of other host genes, may interfere with activation that may have otherwise been induced by 3Cpro cleavage in the linker region. Indeed, we find that co-expression of 3Cpro from Aichi, HepA, SaliA can attenuate NLRP1 activation by TEV protease (Figure 3.13), consistent with the idea that these three proteases can actively block NLRP1 activation. Further investigation will be needed to determine the exact mechanism by which this occurs. Nevertheless, our data demonstrate that non-enteroviral 3Cpros can cleave NLRP1 at independent sites in the rapidly evolving linker region and can, in at least one case, activate the human NLRP1 inflammasome.

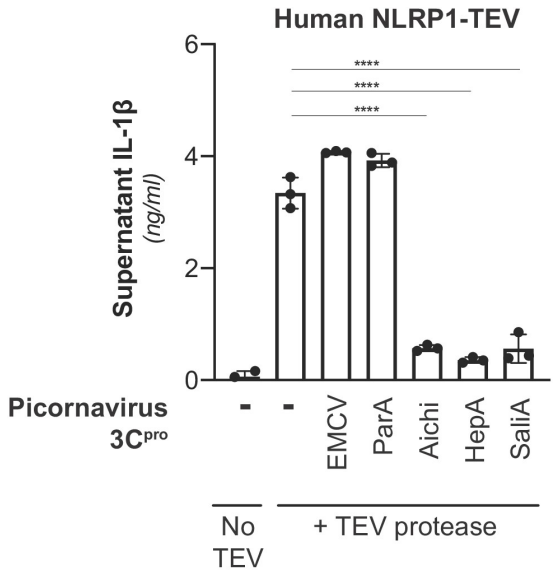


Figure 3.13. Inhibition of NLRP1 activation by non-enteroviral 3Cpro. HEK293T cells were transfected with inflammasome components as in Figure 3A using 100 ng TEV protease or pQCXIP empty vector, but with the additional inclusion of a non-enterovirus 3Cpro or empty vector (100 ng). Release of bioactive IL-1β into the culture supernatant was measured using HEK-Blue IL-1β reporter cells as in Figure 3.4B. Data were analyzed using one-way ANOVA with Tukey's post-test comparing all conditions containing TEV protease. **** = $p<0.0001$.

To further confirm that 3Cpro cleavage (or lack thereof) of NLRP1 is reflective of 3Cpro during viral infection, we infected cells expressing WT or 131P NLRP1 with EMCV. Consistent with our co-transfection experiments, we see no cleavage of NLRP1 when we infect with EMCV, despite seeing robust cleavage when we infect with CVB3 (Figure 3.10D). Likewise, we see no IL-1β release when we infect either inflammasome-reconstituted HEK293T cells or inflammasome-competent HaCaT cells with EMCV (Figure 3.14). These data indicate that evolution of viral 3Cpro cleavage specificity alters whether a virus can be sensed by the NLRP1 tripwire or not.

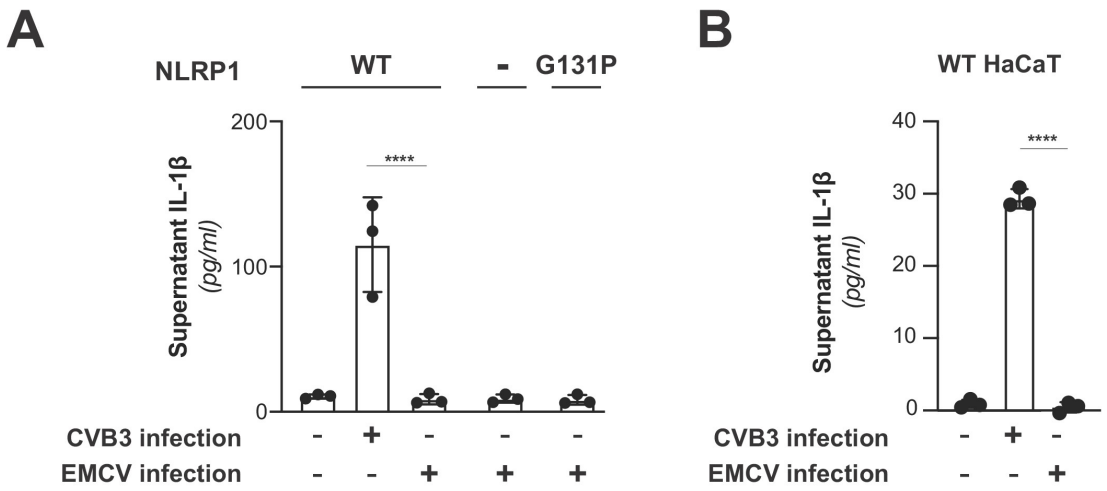


Figure 3.14. EMCV infection does not activate the NLRP1 inflammasome. (A) HEK293T cells were transfected with inflammasome components as in Figure 3.4C and mock infected or infected with 250,000 PFU (MOI = ~1) CVB3 or EMCV. Eight hours post-infection, culture supernatant was collected and bioactive IL-1β was measured as in Figure 3.4C. Data from the mock and CVB3 infections are reproduced from Figure 3.4C and included as a point of reference as they were done in parallel. Data were analyzed using two-way ANOVA with Sidak's post-test. **** = p<0.0001. (B) WT or knockout (Figure 3.7) HaCaT cell lines were mock infected or infected with 100,000 PFU (MOI = ~0.4) CVB3 of EMCV as in Figure 3D. Forty-eight hours post-infection, culture supernatant was collected and bioactive IL-1β was measured as in Figure 3.4D. Data from the mock and CVB3 infections are reproduced from Figure 3.4D and included as a point of reference as they were done in parallel. Data were analyzed using two-way ANOVA with Sidak's post-test. **** = p<0.0001.

Enterovirus 3Cpro cleaves and activates mouse NLRP1B in a virus- and host allele-specific manner

Two bacterial pathogen effectors are known to activate mouse NLRP1B, the LF protease from B. anthracis (Boyden and Dietrich, 2006; Greaney et al., 2020; Moayeri et al., 2010; Terra et al., 2010); (Chavarría-Smith and Vance, 2013; Levinsohn et al., 2012) and the IpaH7.8 E3 ubiquitin ligase from Shigella flexneri (Sandstrom et al., 2019). Interestingly, in both of these cases, activation is specific to the 129 allele of mouse NLRP1B, whereas the B6 allele of NLRP1B is not activated by these pathogenic effectors. Given the power of mouse models for understanding inflammasome biology, we wished to determine if 3Cpros cleave and activate mouse NLRP1B.

Strikingly, when we co-transfected NLRP1B from either the 129 or the B6 strains with diverse enterovirus 3Cpros, we observed allele-specific cleavage products (Figure 3.15A). Consistent with data in Figure 3.8D, we observed weak cleavage of 129 NLRP1B by CVB3 3Cpro. In addition, we found that 3Cpro from other enteroviruses varied substantially in their ability to cleave 129 NLRP1B, including no detectable cleavage with EV71 3Cpro and a different dominant position of cleavage by HRV 3Cpro. Despite this variation, we only observed weak cleavage (Figure 3.15A, left) and little to no inflammasome activation (Figure 3.15B, left) by any enterovirus 3Cpros tested against 129 NLRP1B. In contrast, enterovirus 3Cpro cleavage of B6 NLRP1B resulted in a consistent-sized cleavage product across all enterovirus 3Cpros that ranged in intensities between the different viral proteases, more similar to our observations with human NLRP1. Most interestingly, we observed that co-transfection with HRV 3Cpro resulted in the appearance of a very strong cleavage product (Figure 3.15A, right), almost complete loss of full length B6 NLRP1B (Figure 3.15A and B, right) and very strong activation of the inflammasome (Figure 3.15B, right). These data indicate that mouse NLRP1B can also be cleaved and activated by viral proteases, which suggests that the evolution of the N-terminus of NLRP1B between closely related mouse strains (Figure 3.16) is not only shaping susceptibility to tripwire cleavage by the bacterial LF protease, but also impacts tripwire cleavage by viral 3Cpros. Taken together, these data further support the model in which both host and viral evolution, even within closely related host and viral species, shape the outcome of the interaction between NLRP1 and 3Cpro.

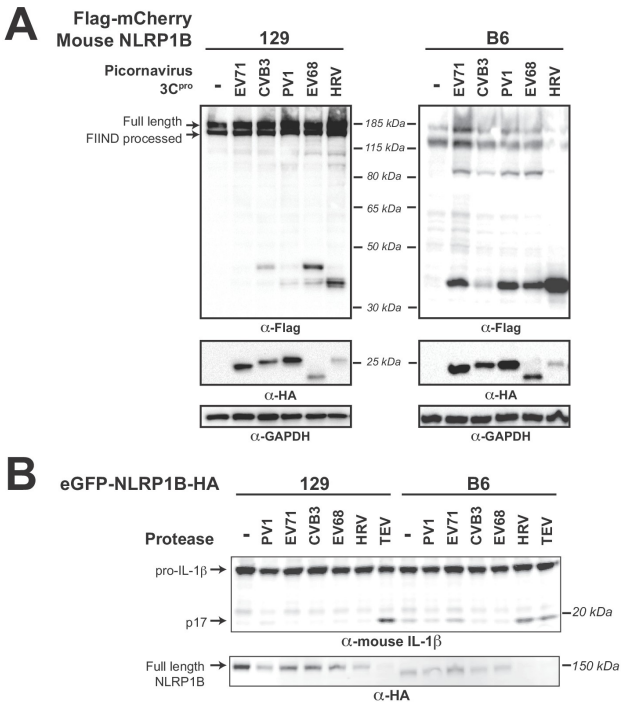


Figure 3.15. Diverse picornavirus 3Cpros cleave and activate mouse NLRP1B at independently evolved sites. (A) Immunoblot depicting CVB3 3Cpro cleavage susceptibility of two alleles (129 and B6) of mouse NLRP1B. Assays were performed as in Figure 3.3D. (B) Immunoblot depicting mouse NLRP1B activation (maturation of IL-1β) by enterovirus 3Cpro and TEV protease. HEK293T cells were co-transfected using 100 ng of the indicated protease, 50 ng mouse IL-1β, 50 ng mouse CASP1, and either 4 ng of 129 NLRP1B or 2.5 ng of B6 NLRP1B constructs, and immunoblotted with the indicated antibodies. Appearance of the mature p17 band of IL-1β indicates successful assembly of the NLRP1B inflammasome and activation of CASP1.

```
NLRP1B (129)    1    MEESPPKQKSNTKVAQHEGQQDLNTT----RHMNVEL-KHRPKLERHLKLGMIPVVYMKQ
NLRP1B (B6)     1    MEQSQPKKKSRTKVAQHEGQLNLNPTFKTRKRKEVELMKRRPKPEGHLKLGTIPKVHIKQ
                     ** * ** ** *********  ** *        *** * *** * ***** ** *  **

NLRP1B (129)    56   GEEILYPAQSLREENLIQNFTSLLLLQKLCPKDPENMIRKSWASCVPEEGGHMINIQDLF
NLRP1B (B6)     61   KGETLDPTWSRKRKNLVQKLTNLLLFEKLCSRGSENLIRKSWYSCEEEERGHMIEIQDLF
                       * * *  *    ** *  * ***  ***    ** ***** **  ** **** *****

NLRP1B (129)    116  GPNIGTQKEPQLVIIEGAA
NLRP1B (B6)     121  GPNRGTHKKPQLVIIEGAA
                     *** ** * **********
                              Start of NBD
```

Figure 3.16. Alignment of N-termini of mouse NLRP1B 129 and B6 alleles. Sequences were aligned using MAFFT (Katoh and Standley, 2013). Asterisks indicate conserved residues and the start of the nucleotide-binding domain (NBD) is shown.

Discussion

Pathogens and their hosts are locked in a continual evolutionary conflict in which each side is attempting to exploit the others' weakness. One particularly successful strategy that pathogens have adopted is to exploit host processes that are highly constrained, leaving the host little room to evolutionarily adapt to overcome the pathogen. For instance, molecular mimicry of host proteins is commonly deployed by pathogens to antagonize host defenses, as it limits the evolutionary options for the host to counter-evolve (Elde and Malik, 2009). Beyond mimicry of entire proteins or protein domains, pathogens can also mimic so-called 'short linear motifs' (SLIMs) through evolution of only a small number of amino acids to hijack highly conserved host processes such as post-translational modifications or binding by small protein domains (Chemes et al., 2015; Hagai et al., 2014). Although these strategies are generally described as taking advantage of host evolutionary constraint, pathogens also have potential weak points of evolutionary constraint. In particular, proteases from positive-sense RNA viruses, such as picornaviruses, need to specifically cleave numerous sites within the viral polyprotein in order to reproduce. Thus, changing protease specificity requires concomitant changes to several independent cleavage sites, which is difficult to accomplish in a single evolutionary step. On top of that, protease cleavage motifs often only span a small number of amino acids (Schechter and Berger, 1967), potentially facilitating the independent evolution of these SLIMs in host proteins.

Here, we show that the inflammasome protein, NLRP1, serves as a sensor for diverse proteases from the *Picornaviridae* family of human pathogens by mimicking the highly conserved protease cleavage sites found within the viral polyproteins. By exploiting a constrained feature of viral evolution and tying it to a pro-inflammatory

immune response, such a system allows the immune system to recognize and respond to a wide range of viral proteases expressed in the host cytoplasm. NLRP1 represents one of the few known cases of mammalian ETI (Cui et al., 2015; Fischer et al., 2020; Jones et al., 2016), where pathogen-mediated cleavage of NLRP1 promotes its activation. By holding the small C-terminal CARD-containing fragment in a non-covalent association with the larger N-terminal fragment, the majority of the protein can serve as a sensor for pathogen-encoded effectors (Mitchell et al., 2019; Taabazuing et al., 2020). This presents an opportunity to allow NLRP1 to evolve to be recognized by pathogenic effectors, ultimately leading to degradation of the N-terminal fragment. Indeed, mouse NLRP1B has been shown to be specifically cleaved by the protease-containing secreted effector from B. anthracis (LF) as well as being ubiquitylated by an E3-ubiquitin ligase from S. flexneri (IpaH7.8) (Sandstrom et al., 2019). While these two examples provide evidence that the mouse NLRP1B inflammasome operates by a ‘functional degradation’ model, a direct pathogen-encoded activator of human NLRP1 had remained elusive. We now show, using an evolution-guided approach, that proteases from diverse picornaviruses, including human pathogens such as coxsackievirus B3 (CVB3), human rhinovirus A (HRV), enterovirus D68 (EV68) and poliovirus 1 (PV1) and rosavirus A2 (Rosa2), specifically cleave several independently evolved sites in human NLRP1, leading to activation of the NLRP1 inflammasome and release of pro-inflammatory cytokines such as IL-1β. Together with recent findings (Robinson et al., 2020), our work has thus identified proteases from a diverse range of picornaviruses as pathogen-encoded activators of human NLRP1.

We previously speculated that the unique domain architecture of NLRP1 would allow the N-terminal linker of human NLRP1 to freely evolve to be recognized by pathogenic effectors. Indeed, by harvesting publicly available enterovirus polyprotein sequences for known 3Cpro cleavage sites, we created a 3Cpro cleavage motif that was used to successfully predict the site of enterovirus 3Cpro cleavage at position 130–131 within the rapidly-evolving linker NLRP1. Additionally, our finding that numerous enteroviruses also cleave at the Q130-G131 site and activate pro-inflammatory cytokine release suggests that human NLRP1 serves as a general enteroviral protease sensor by encoding a polyprotein cleavage site mimic. Our phylogenetic assessment of the Q130-G131 3Cpro cleavage site in NLRP1 suggests that NLRP1 sensing of enteroviruses at this specific site is an innovation in the primate lineage, and is largely absent in all other mammalian lineages with exception of a possible independent acquisition by members within the Caprinae subfamily of mammals (e.g. goats, sheep) (Figure 3.5). Interestingly, even within the primate lineage and a small fraction of the human population, some primate orthologs and human variants are cleavage-resistant and therefore do not activate the inflammasome upon cytoplasmic expression of 3Cpro. Such data may hint at three different possible explanations for these changes. First, evolutionary drift in the absence of pressure from pathogenic enteroviruses may account for loss of enterovirus 3Cpro responsiveness in these genes. Second, selection to sense another viral protease may shape the same region of the linker. Finally, while the ETI model of NLRP1 suggests that enteroviral cleavage of NLRP1 has evolved to activate a beneficial immune response in certain contexts, the effects of NLRP1 overactivation may be detrimental in other contexts. In human skin keratinocytes, where

NLRP1 is regarded as the key inflammasome, all components of the NLRP1 inflammasome are basally expressed and thus poised to elicit an inflammatory response (Zhong et al., 2016). Here, germline mutations in NLRP1 that result in overactivation can cause growth of warts in the upper airway in a condition known as recurrent respiratory papillomatosis (JRRP) (Drutman et al., 2019) and an increase in skin cancer susceptibility and skin disorders such as multiple self-healing palmoplantar carcinoma (MSPC), familial keratosis lichenoides chronica (FKLC) and auto-inflammation with arthritis and dyskeratosis (AIADK) (Grandemange et al., 2017; Herlin et al., 2020; Soler et al., 2013; Zhong et al., 2016; Zhong et al., 2018). Additional recent work has indicated that dsRNA can also activate the NLRP1 inflammasome in human keratinocytes (Bauernfried et al., 2020), adding to the role that NLRP1 may play in the inflammatory response. Beyond the skin, NLRP1 is also basally expressed in tissues such as the gut and brain (D'Osualdo et al., 2015; Kaushal et al., 2015; Kummer et al., 2007), which are sites of picornavirus replication where overactivation upon infection may result in immunopathology. Further in vivo studies will help determine the role of NLRP1 in antiviral immunity and/or immunopathology during viral infection. Facilitating these studies, our discovery that 3Cpro from HRV potently cleaves and activates NLRP1B from B6 but not 129 mice suggests that rhinovirus infection of B6 mice may be a good model for studying the in vivo consequences of viral-mediated NLRP1 inflammasome activation.

Intriguingly, 3Cpros from nearly every genus of human-infecting picornavirus can cleave NLRP1 somewhere in the rapidly evolving linker region between the PYD and NLR domain, although only enteroviruses cleave at the specific site between position

130 and 131. These data suggest that this extended linker, which we previously found showed widespread signatures of positive selection (Chavarría-Smith et al., 2016), may be convergently evolving to mimic cleavage sites from a diverse range of viruses at multiple independent sites. Supporting that model, we observe a similar phenomenon in mouse NLRP1B, where multiple viral proteases cleave at different sites within NLRP1 in a strain-specific manner. These data highlight the important functional differences in cleavage specificity between even closely related 3Cpro that are not accounted for by predictive models. Further studies will be required to understand the precise relationships between sites within NLRP1 and individual protease specificity. Intriguingly, not all these cleavage events lead to inflammasome activation in the same way that enteroviral cleavage does, and we find evidence for antagonism of NLRP1 activation by some 3Cpros, suggesting that additional activities of 3Cpro may be the next step in the arms race, serving to prevent inflammasome activation even after the tripwire has been tripped.

Taken together, our work suggests that host mimicry of viral polyprotein cleavage motifs could be an important evolutionary strategy in the ongoing arms race between host and viruses. Indeed, one explanation for the somewhat surprising observation that the specificity of viral proteases changes at all within a viral family such as the picornaviruses is that there is evolutionary pressure from the host to evolve cleavage sites and protease specificity. Prior work has highlighted the roles that viral proteases can play in antagonizing host immune factors and driving host evolution to avoid being cleaved (Patel et al., 2012; Stabell et al., 2018). In that case, the viral proteases would evolve to antagonize new factors while maintaining polyprotein cleavage. However,

mimicry coupled with cleavage-activating immunity as seen with NLRP1 could be an even stronger pressure to shape the protease specificity. By turning the tables, these host processes may drive the type of functional diversification of viral protease specificity that we observe in order to avoid cleaving NLRP1 and other similar ETI factors. We expect that this work may lead to the discovery that such an evolutionary strategy may be more broadly deployed at other sites of host-pathogen conflicts.

Materials and methods

Motif generation and search

To build the motif, 2658 nonredundant enteroviral polyprotein sequences were collected from the Viral Pathogen Resource (ViPR) and aligned with 20 well-annotated reference enteroviral polyprotein sequences from RefSeq (Supplementary file 3.1). P1 and P1' of the eight annotated cleavage sites across the RefSeq sequences served as reference points for putative cleavage sites across the 2658 ViPR sequences, with the exception of enterovirus D polyproteins. The 3Cpro cleavage site for VP3-VP1 within polyproteins from the clade of enterovirus D have been described to be undetectable and have thus been removed (Tan et al., 2013). Four amino acyl residues upstream (P4-P1) and downstream (P1'-P4') of each cleavage site were extracted from every MAFFT-aligned polyprotein sequence, resulting in 2678 sets of cleavage sites (RefSeq sites included). Each set of cleavage sites representative of each polyprotein was then concatenated. Next, 1884 duplicates were removed from the 2678 concatenated cleavage sites. The remaining 796 nonredundant, concatenated cleavage sites were then split into individual 8-mer cleavage sites and the 6333 8-mers were aligned using

MAFFT to generate Geneious-defined sequence logo information at each aligned position. Pseudo-counts to the position-specific scoring matrix were adjusted by total information content within each position relative to the two most information-dense position P1 and P1' (pseudocount = 0) and the least information-dense position P3 (pseudocount = 1). The 0.002 p-value threshold for FIMO motif searching against human NLRP1 was determined to optimize the capture of 95% of initial input cleavage sites within the set of 2678 whole enteroviral polyproteins and a majority sites within a previously described dataset of enteroviral 3Cpro targets (Laitinen et al., 2016).

NetSurfP

Prediction of the coil probability across human NLRP1 (NCBI accession NP_127497.1) was conducted using the protein FASTA as the input for the NetSurfP web server (http://www.cbs.dtu.dk/services/NetSurfP/).

Sequence alignments, phylogenetic trees, and NLRP1 phylogenomics

Complete polyprotein sequences from 796 picornaviruses with non-redundant 3Cpro cleavage sites (see 'Motif generation and search' section above) were downloaded from ViPR. Sequences were aligned using MAFFT (Katoh and Standley, 2013) and a neighbor-joining phylogenetic tree was generated using Geneious software (Kearse et al., 2012). An alignment and phylogenetic tree of all the 3Cpro sequences used in this study was generated similarly.

To identify mammalian NLRP1 homologs, and species that lack NLRP1, the human NLRP1 protein sequence was used to query the RefSeq protein sequence

database, a curated collection of the most well-assembled genomes, using BLASTp (Altschul et al., 1997). Sequences were downloaded and aligned using MAFFT implemented in Geneious software. Consensus sequence logos shown were generated using Geneious software. We determined that NLRP1 was 'absent' from a clade of species using the following criteria: (1) when searching with human NLRP1, we found an obvious homolog of another NLRP protein (generally NLRP3, NLRP12 or NLRP14) but no complete or partial homolog of NLRP1 and (2) this absence was apparent in every member of the clade of species (>2 species) in the RefSeq database.

Plasmids and constructs

For NLRP1 cleavage assays, the coding sequences of human NLRP1 WT (NCBI accession NP_127497.1), human NLRP1 mutants (G131P, G131R, Q130R, G127E), human NLRP1 TEV or mouse NLRP1B (129 allele, NCBI accession AAZ40510.1; B6 allele, NCBI accession XM_017314698.2) were cloned into the pcDNA5/FRT/TO backbone (Invitrogen, Carlsbad, CA) with an N-terminal 3xFlag and mCherry tag. For NLRP1 activation, the same sequences were cloned into the pQCXIP vector backbone (Takara Bio, Mountain View, CA) with a C-terminal Myc tag (human NLRP1 sequences) or N-terminal EGFP and C-terminal HA (mouse NLRP1B sequences). Vectors containing the coding sequences of human NLRP1 TEV (NLRP1-TEV2), ASC, human and mouse CASP1, human IL-1β-V5, mouse IL-1β, and TEV protease (Chavarría-Smith et al., 2016) were generous gifts from Dr. Russell Vance, UC Berkeley. Single point mutations were made using overlapping stitch PCR. A list of primers used to generate the wild-type and mutant NLRP1 constructs are described in Supplementary file 3.5.

CVB3 3Cpro and EMCV 3Cpro were cloned from CVB3-Nancy and EMCV-Mengo plasmids (generous gifts from Dr. Julie Pfeiffer, UT Southwestern). Remaining 3Cpro sequences were ordered as gBlocks (Integrated DNA Technologies, San Diego, CA). Each 3Cpro was cloned with an N-terminal HA tag into the QCXIP vector backbone. Catalytic mutations were made using overlapping stitch PCR. A list of primers and gBlocks used to generate the protease constructs are described in Supplementary file 3.5.

Following cloning, all plasmid stocks were sequenced across the entire inserted region to verify that no mutations were introduced during the cloning process.

Cell culture and transient transfection

All cell lines (HEK293T, HEK-Blue-IL-1β, and HaCaT) are routinely tested for mycoplasma by PCR kit (ATCC, Manassas, VA) and kept a low passage number to maintain less than one year since purchase, acquisition or generation. HEK293T cells were obtained from ATCC (catalog # CRL-3216), HEK-Blue-IL-1β cells were obtained from Invivogen (catalog # hkb-il1b) and HaCaT cells were obtained from the UC Berkeley Cell Culture Facility (https://bds.berkeley.edu/facilities/cell-culture) and all lines were verified by those sources. All cells were grown in complete media containing DMEM (Gibco, Carlsbad, CA), 10% FBS (Peak Serum, Wellington, CO), and appropriate antibiotics (Gibco, Carlsbad, CA). For transient transfections, HEK293T cells were seeded the day prior to transfection in a 24-well plate (Genesee, El Cajon, CA) with 500 µl complete media. Cells were transiently transfected with 500 ng of total DNA and 1.5 µl of Transit X2 (Mirus Bio, Madison, WI) following the manufacturer's

protocol. HEK-Blue IL-1β reporter cells (Invivogen, San Diego, CA) were grown and assayed in 96-well plates (Genesee, El Cajon, CA).

HaCaT knockouts

To establish NLRP1 and CASP1 knockouts in human immortalized keratinocyte HaCaT cells, lentivirus-like particles were made by transfecting HEK293T cells with the plasmids psPAX2 (gift from Didier Trono, Addgene plasmid # 12260) and pMD2.G (gift from Didier Trono, Addgene plasmid # 12259) and either the pLB-Cas9 (gift from Feng Zhang, Addgene plasmid # 52962) (Sanjana et al., 2014) or a plentiGuide-Puro, which was adapted for ligation-independent cloning (kindly gifted by Moritz Gaidt) (Schmidt et al., 2015). Guide sequences are shown in Supplementary file 3.5. Conditioned supernatant was harvested 48 and 72 hr post-transfection and used for spinfection of HaCaT cells at 1200 x g for 90 min at 32°C. Forty-eight hours post-transduction, cells with stable expression of Cas9 were selected in media containing 100 µg/ml blasticidin. Blasticidin-resistant cells were then transduced with sgRNA-encoding lentivirus-like particles, and selected in media containing 1.3 µg/ml puromycin. Cells resistant to blasticidin and puromycin were single cell cloned by limiting dilution in 96-well plates, and confirmed as knockouts by Sanger sequencing (Figure 3.7).

NLRP1 cleavage assays

100 ng of epitope-tagged human NLRP1 WT, human NLRP1 mutants (G131P, G131R, Q130R, G127E), human NLRP1 TEV or mouse NLRP1B was co-transfected with 250 ng of HA-tagged protease-producing constructs. Twenty-four hours post-

transfection, the cells were harvested, lysed in 1x NuPAGE LDS sample buffer (Invitrogen, Carlsbad, CA) containing 5% β-mercaptoethanol (Fisher Scientific, Pittsburg, PA) and immunoblotted with antibodies described below.

NLRP1 activity assays

For human NLRP1 activation assays, 5 ng of ASC, 100 ng of CASP1, 50 ng of IL-1β-V5, and 100 ng of various protease-producing constructs were co-transfected with 4 ng of either pQCXIP empty vector, wild-type or mutant pQCXIP-NLRP1-Myc constructs. For inhibitor treatments, cells were treated with either 0.5 μM MG132 or 1.0 μM MLN4924 18 hr after transfection. Twenty-four hours post-transfection, cells were harvested and lysed in 1x NuPAGE LDS sample buffer containing 5% β-mercaptoethanol or in 1x RIPA lysis buffer with protease inhibitor cocktail (Roche) and immunoblotted with antibodies described below or culture media was harvested for quantification of IL-1β levels by HEK-Blue assays (see below).

For mouse NLRP1B activation assays, 50 ng of mouse CASP1, 50 ng of mouse IL-1β, and 100 ng of various protease-producing constructs were co-transfected with either 4 ng of 129 NLRP1B or 2.5 ng B6 NLRP1B constructs. Twenty-four hours post-transfection, cells were harvested in 1x RIPA lysis buffer with protease inhibitor cocktail (Roche) and immunoblotted with antibodies described below.

Viral stocks and viral infections

CVB3 and EMCV viral stocks were generated by co-transfection of CVB3-Nancy or EMCV-Mengo infectious clone plasmids with a plasmid expressing T7 RNA polymerase (generous gifts from Dr. Julie Pfeiffer, UT Southwestern) as previously described (McCune et al., 2020). Supernatant was harvested, quantified by plaque assay or TCID50 on HEK293T cells, and frozen in aliquots at −80°C.

For viral infections of HEK293T cells, cells were transfected in 24-well plates and infected with 250,000 PFU (MOI = ~1) CVB3 or EMCV or mock infected for the indicated times. For cleavage assays, cells were transfected with 100 ng of the indicated NLRP1 construct and, 24 hr after transfection, cells were harvested and lysed in 1x NuPAGE LDS sample buffer containing 5% β-mercaptoethanol and immunoblotted with antibodies described below. For activation assays, cells were transfected with 4 ng of the indicated NLRP1 construct and 5 ng of ASC, 100 ng of CASP1, 50 ng of IL-1β-V5. Twenty-four hours after transfections, cells were infected with virus (or mock infected) and culture supernatant was collected 8 hr later (32 hr post-transfection). Culture supernatant was filtered through a 100,000 MWCO centrifugal spin filter (MilliporeSigma, Burlington, MA) for 10 min at 12,000xg to remove infectious virus and IL-1β levels were quantified by HEK-Blue assays (see below).

For viral infections of HaCaT cells, cells were plated in 24-well plates. The next day, cells were infected with 100,000 PFU/well (MOI = ~0.4) CVB3 or EMCV or mock infected. Forty-eight hours after infection, culture supernatant was collected, spin filtered as described above to remove infectious virus, and IL-1β levels were quantified by HEK-Blue assays (see below).

HEK-Blue IL-1β assay

To quantify the levels of bioactive IL-1β released from cells, we employed HEK-Blue IL-1β reporter cells (Invivogen, San Diego, CA). In these cells, binding to IL-1β to the surface receptor IL-1R1 results in the downstream activation of NF-kB and subsequent production of secreted embryonic alkaline phosphatase (SEAP) in a dose-dependent manner (Figure 3.5). SEAP levels are detected using a colorimetric substrate assay, QUANTI-Blue (Invivogen, San Diego, CA) by measuring an increase in absorbance at OD655.

Culture supernatant from inflammasome-reconstituted HEK293T cells or HaCaT cells that had been transfected with 3Cpro or virally infected (see above) was added to HEK-Blue IL-1β reporter cells plated in 96-well format in a total volume of 200 µl per well. On the same plate, serial dilutions of recombinant human IL-1β (Invivogen, San Diego, CA) were added in order to generate a standard curve for each assay. Twenty-four hours later, SEAP levels were assayed by taking 20 µl of the supernatant from HEK-Blue IL-1β reporter cells and adding to 180 µl of QUANTI-Blue colorimetric substrate following the manufacturer's protocol. After incubation at 37°C for 30–60 min, absorbance at OD655 was measured on a BioTek Cytation five plate reader (BioTek Instruments, Winooski, VT) and absolute levels of IL-1β were calculated relative to the standard curve. All assays, beginning with independent transfections or infections, were performed in triplicate.

Immunoblotting and antibodies

Harvested cell pellets were washed with 1X PBS, and lysed with 1x NuPAGE LDS sample buffer containing 5% β-mercaptoethanol at 98C for 10 min. The lysed samples were spun down at 15000 RPM for two minutes, followed by loading into a 4–12% Bis-Tris SDS-PAGE gel (Life Technologies, San Diego, CA) with 1X MOPS buffer (Life Technologies, San Diego, CA) and wet transfer onto a nitrocellulose membrane (Life Technologies, San Diego, CA). Membranes were blocked with PBS-T containing 5% bovine serum albumin (BSA) (Spectrum, New Brunswick, NJ), followed by incubation with primary antibodies for V5 (IL-1β), FLAG (mCherry-fused NLRP1 for protease assays), Myc (NLRP1-Myc for activation assays), HA (viral protease or mouse NLRP1B), β-tubulin, or GAPDH. Membranes were rinsed three times in PBS-T then incubated with the appropriate HRP-conjugated secondary antibodies. Membranes were rinsed again three times in PBS-T and developed with SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific, Carlsbad, CA). The specifications, source, and clone info for antibodies are described in Supplementary file 3.6.

Table 3.1. Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Gene (*Homo sapiens*)	NLRP1	NCBI	NCBI: NP_127497.1	
Gene (*Mus musculus*)	NLRP1B (129)	NCBI	NCBI: AAZ40510.1	
Gene (*Mus musculus*)	NLRP1B (B6)	NCBI	NCBI: XM_017314698.2	
Cell line (*Homo sapiens*)	HEK293T	ATCC	CRL-3216	
Cell line (*Homo sapiens*)	HEK-Blue IL-1β cells	Invivogen	HKB-IL1B	
Cell line (*Homo sapiens*)	HaCaT (parental)	UC Berkeley Cell Culture Facility		
Cell line (*Homo sapiens*)	HaCaT Cas9 (WT)	This paper		
Cell line (*Homo sapiens*)	HaCaT Cas9 ΔNLRP1 #1 (NLRP1 KO clone #1)	This paper		Exon 5 target (TCCACTGCTTGTACGAGACT)
Cell line (*Homo sapiens*)	HaCaT Cas9 ΔNLRP1 #2 (NLRP1 KO clone #2)	This paper		Exon 2 target (TGTAGGGGAATGAGGGAGAG)
Cell line (*Homo sapiens*)	HaCaT Cas9 ΔCASP1 #1 (CASP1 KO clone #1)	This paper		Exon 2 target (CCAAACAGACAAGGTCCTGA)
Cell line (*Homo sapiens*)	HaCaT Cas9 ΔCASP1 #2 (CASP1 KO clone #2)	This paper		Exon 2 target (CCAAACAGACAAGGTCCTGA)
Recombinant DNA reagent	pcDNA5/FRT/TO (plasmid)	Invitrogen	V652020	
Recombinant DNA reagent	pQCXIP (plasmid)	Takara Bio	631516	
Recombinant DNA reagent	psPAX2 (plasmid)	Addgene	12260	Gift from Dr. Didier Trono
Recombinant DNA reagent	pMD2.G (plasmid)	Addgene	12259	Gift from Dr. Didier Trono
Recombinant DNA reagent	pLB-Cas9 (plasmid)	Addgene	52962	Gift from Dr. Feng Zhang

Table 3.1. Key resources table (continued)

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Recombinant DNA reagent	pLentiGuide-Puro (plasmid)	Other		Gift from Dr. Mortiz Gaidt
Recombinant DNA reagent	Inflammasome reconstitution plasmids	PMID: 27926929		Gifts from Dr. Russell Vance: human NLRP1 TEV (NLRP1-TEV2), human ASC, human and mouse CASP1, human IL-1B-V5, mouse IL-1B, and TEV protease
Recombinant DNA reagent	CVB3-Nancy infectious clone plasmid	PMID: 2410905		Gift from Dr. Julie Pfeiffer
Recombinant DNA reagent	EMCV-Mengo infectious clone plasmid	PMID: 2538661		Gift from Dr. Julie Pfeiffer
Commercial assay or kit	QUANTI-Blue assay reagent (for HEK-Blue IL-1b cells)	Invivogen	REP-QBS	Includes necessary reagents for measuring IL-1b release from HEK-Blue-IL-1B reporter cell line
Chemical compound, drug	TransIT-X2	Mirus	MIR 6000	
Chemical compound, drug	MG132	Sigma-Aldrich	M7449	
Chemical compound, drug	MLN4924	APExBIO	B1036	
Antibody	V5-Tag Rabbit mAb	Cell Signaling Technology	13202S	Dilution ratio 1:1000
Antibody	Flag-Tag Mouse mAb	Sigma-Aldrich	F1804	Dilution ratio 1:2000
Antibody	Myc-Tag Rabbit mAb	Cell Signaling Technology	2278S	Dilution ratio 1:1000
Antibody	HA-Tag Rat mAb	Roche	11867423001	Dilution ratio 1:1000
Antibody	GAPDH Rabbit mAb	Cell Signaling Technology	2118S	Dilution ratio 1:2000
Antibody	Goat anti-Rat IgG (H+L) Secondary Antibody, HRP	Thermo Fisher Scientific	31470	Dilution ratio 1:10000
Antibody	Goat anti-Rabbit IgG (H+L) Secondary Antibody, HRP	Biorad	1706515	Dilution ratio 1:10000
Antibody	Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP	Biorad	1706516	Dilution ratio 1:10000

Table 3.1. Key resources table (continued)

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Antibody	β-Tubulin Mouse mAb	Sigma-Aldrich	T4026	Dilution ratio 1:2000
Antibody	Goat anti-mouse IL-1β antibody	R&D Systems	AF401SP	Dilution ratio 1:1000
Sequence-based reagent	Oligonucleotides	Other		See Supplementary file 5 for list of oligonucleotides used in this study
Software, algorithm	MEME v5.0.3	PMID: 25953851		Motif finder (FIMO)
Software, algorithm	MAFFT 7.309	PMID: 23329690		
Software, algorithm	NetSurfP	PMID: 30785653		http://www.cbs.dtu.dk/services/NetSurfP/ (Original); https://services.healthtech.dtu.dk/service.php?NetSurfP-2.0 (Alternate)
Software, algorithm	Geneious	PMID: 22543367		Neighbor-joining phylogenetic tree
Software, algorithm	BLASTp	PMID: 9254694		

Chapter 3 is published and can be found available in eLife in Immunology and Inflammation Microbiology and Infectious Disease 2021, including co-authors Christopher Beierschmitt, Andrew P. Ryan, Rimjhim Agarwal, Patrick S. Mitchell, Matthew D. Daugherty. I, Brian V. Tsu, am the co-first author of this paper, alongside Christopher Beierschmitt.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*(17), 3389-3402. doi:10.1093/nar/25.17.3389

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., . . . Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res, 37*(Web Server issue), W202-208. doi:10.1093/nar/gkp335

Bauernfried, S., Scherr, M. J., Pichlmair, A., Duderstadt, K. E., & Hornung, V. (2020). Human NLRP1 is a sensor for double-stranded RNA. *Science*. doi:10.1126/science.abd0811

Blom, N., Hansen, J., Blaas, D., & Brunak, S. (1996). Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks. *Protein Sci, 5*(11), 2203-2216. doi:10.1002/pro.5560051107

Boyden, E. D., & Dietrich, W. F. (2006). Nalp1b controls mouse macrophage susceptibility to anthrax lethal toxin. *Nat Genet, 38*(2), 240-244. doi:10.1038/ng1724

Broz, P., & Dixit, V. M. (2016). Inflammasomes: mechanism of assembly, regulation and signalling. *Nat Rev Immunol, 16*(7), 407-420. doi:10.1038/nri.2016.58

Cagliani, R., Forni, D., Tresoldi, C., Pozzoli, U., Filippi, G., Rainone, V., . . . Sironi, M. (2014). RIG-I-like receptors evolved adaptively in mammals, with parallel evolution at LGP2 and RIG-I. *J Mol Biol, 426*(6), 1351-1365. doi:10.1016/j.jmb.2013.10.040

Chavarria-Smith, J., Mitchell, P. S., Ho, A. M., Daugherty, M. D., & Vance, R. E. (2016). Functional and Evolutionary Analyses Identify Proteolysis as a General Mechanism for NLRP1 Inflammasome Activation. *PLoS Pathog, 12*(12), e1006052. doi:10.1371/journal.ppat.1006052

Chavarria-Smith, J., & Vance, R. E. (2013). Direct proteolytic cleavage of NLRP1B is necessary and sufficient for inflammasome activation by anthrax lethal factor. *PLoS Pathog, 9*(6), e1003452. doi:10.1371/journal.ppat.1003452

Chemes, L. B., de Prat-Gay, G., & Sanchez, I. E. (2015). Convergent evolution and mimicry of protein linear motifs in host-pathogen interactions. *Curr Opin Struct Biol, 32*, 91-101. doi:10.1016/j.sbi.2015.03.004

Chui, A. J., Okondo, M. C., Rao, S. D., Gai, K., Griswold, A. R., Johnson, D. C., . . . Bachovchin, D. A. (2019). N-terminal degradation activates the NLRP1B inflammasome. *Science, 364*(6435), 82-85. doi:10.1126/science.aau1208

Croft, S. N., Walker, E. J., & Ghildyal, R. (2018). Human Rhinovirus 3C protease cleaves RIPK1, concurrent with caspase 8 activation. *Sci Rep, 8*(1), 1569. doi:10.1038/s41598-018-19839-4

Cui, H., Tsuda, K., & Parker, J. E. (2015). Effector-triggered immunity: from pathogen perception to robust defense. *Annu Rev Plant Biol, 66*, 487-511. doi:10.1146/annurev-arplant-050213-040012

D'Osualdo, A., Anania, V. G., Yu, K., Lill, J. R., Kaufman, R. J., Matsuzawa, S., & Reed, J. C. (2015). Transcription Factor ATF4 Induces NLRP1 Inflammasome Expression during Endoplasmic Reticulum Stress. *PLoS One, 10*(6), e0130635. doi:10.1371/journal.pone.0130635

D'Osualdo, A., Weichenberger, C. X., Wagner, R. N., Godzik, A., Wooley, J., & Reed, J. C. (2011). CARD8 and NLRP1 undergo autoproteolytic processing through a ZU5-like domain. *PLoS One, 6*(11), e27396. doi:10.1371/journal.pone.0027396

Daugherty, M. D., & Malik, H. S. (2012). Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet, 46*, 677-700. doi:10.1146/annurev-genet-110711-155522

Drutman, S. B., Haerynck, F., Zhong, F. L., Hum, D., Hernandez, N. J., Belkaya, S., . . . Casanova, J. L. (2019). Homozygous NLRP1 gain-of-function mutation in siblings with a syndromic form of recurrent respiratory papillomatosis. *Proc Natl Acad Sci U S A, 116*(38), 19055-19063. doi:10.1073/pnas.1906184116

Elde, N. C., & Malik, H. S. (2009). The evolutionary conundrum of pathogen mimicry. *Nat Rev Microbiol, 7*(11), 787-797. doi:10.1038/nrmicro2222

Evavold, C. L., & Kagan, J. C. (2019). Inflammasomes: Threat-Assessment Organelles of the Innate Immune System. *Immunity, 51*(4), 609-624. doi:10.1016/j.immuni.2019.08.005

Fan, X., Li, X., Zhou, Y., Mei, M., Liu, P., Zhao, J., . . . Yi, L. (2020). Quantitative Analysis of the Substrate Specificity of Human Rhinovirus 3C Protease and Exploration of Its Substrate Recognition Mechanisms. *ACS Chem Biol, 15*(1), 63-73. doi:10.1021/acschembio.9b00539

Finger, J. N., Lich, J. D., Dare, L. C., Cook, M. N., Brown, K. K., Duraiswami, C., . . . Gough, P. J. (2012). Autolytic proteolysis within the function to find domain (FIIND) is required for NLRP1 inflammasome activity. *J Biol Chem, 287*(30), 25030-25037. doi:10.1074/jbc.M112.378323

Fischer, N. L., Naseer, N., Shin, S., & Brodsky, I. E. (2020). Publisher Correction: Effector-triggered immunity and pathogen sensing in metazoans. *Nat Microbiol, 5*(3), 528. doi:10.1038/s41564-020-0682-4

Frew, B. C., Joag, V. R., & Mogridge, J. (2012). Proteolytic processing of Nlrp1b is required for inflammasome activity. *PLoS Pathog, 8*(4), e1002659. doi:10.1371/journal.ppat.1002659

Grandemange, S., Sanchez, E., Louis-Plence, P., Tran Mau-Them, F., Bessis, D., Coubes, C., . . . Genevieve, D. (2017). A new autoinflammatory and autoimmune syndrome associated with NLRP1 mutations: NAIAD (NLRP1-associated autoinflammation with arthritis and dyskeratosis). *Ann Rheum Dis, 76*(7), 1191-1198. doi:10.1136/annrheumdis-2016-210021

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics, 27*(7), 1017-1018. doi:10.1093/bioinformatics/btr064

Greaney, A. J., Portley, M. K., O'Mard, D., Crown, D., Maier, N. K., Mendenhall, M. A., . . . Moayeri, M. (2020). Frontline Science: Anthrax lethal toxin-induced, NLRP1-mediated IL-1beta release is a neutrophil and PAD4-dependent event. *J Leukoc Biol*. doi:10.1002/JLB.4HI0320-028R

Hagai, T., Azia, A., Babu, M. M., & Andino, R. (2014). Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. *Cell Rep, 7*(5), 1729-1739. doi:10.1016/j.celrep.2014.04.052

Hancks, D. C., Hartley, M. K., Hagan, C., Clark, N. L., & Elde, N. C. (2015). Overlapping Patterns of Rapid Evolution in the Nucleic Acid Sensors cGAS and OAS1 Suggest a Common Mechanism of Pathogen Antagonism and Escape. *PLoS Genet, 11*(5), e1005203. doi:10.1371/journal.pgen.1005203

Herlin, T., Jorgensen, S. E., Host, C., Mitchell, P. S., Christensen, M. H., Laustsen, M., . . . Mogensen, T. H. (2019). Autoinflammatory disease with corneal and mucosal dyskeratosis caused by a novel NLRP1 variant. *Rheumatology (Oxford)*. doi:10.1093/rheumatology/kez612

Huang, L., Liu, Q., Zhang, L., Zhang, Q., Hu, L., Li, C., . . . Weng, C. (2015). Encephalomyocarditis Virus 3C Protease Relieves TRAF Family Member-associated NF-kappaB Activator (TANK) Inhibitory Effect on TRAF6-mediated NF-kappaB Signaling through Cleavage of TANK. *J Biol Chem, 290*(46), 27618-27632. doi:10.1074/jbc.M115.660761

Jagdeo, J. M., Dufour, A., Klein, T., Solis, N., Kleifeld, O., Kizhakkedathu, J., . . . Jan, E. (2018). N-Terminomics TAILS Identifies Host Cell Substrates of Poliovirus and Coxsackievirus B3 3C Proteinases That Modulate Virus Infection. *J Virol, 92*(8). doi:10.1128/JVI.02211-17

Janeway, C. A., Jr. (1989). Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol, 54 Pt 1*, 1-13. doi:10.1101/sqb.1989.054.01.003

Jones, J. D., Vance, R. E., & Dangl, J. L. (2016). Intracellular innate immune surveillance devices in plants and animals. *Science, 354*(6316). doi:10.1126/science.aaf6395

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., . . . MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature, 581*(7809), 434-443. doi:10.1038/s41586-020-2308-7

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol, 30*(4), 772-780. doi:10.1093/molbev/mst010

Kaushal, V., Dye, R., Pakavathkumar, P., Foveau, B., Flores, J., Hyman, B., . . . LeBlanc, A. C. (2015). Neuronal NLRP1 inflammasome activation of Caspase-1 coordinately regulates inflammatory interleukin-1-beta production and axonal degeneration-associated Caspase-6 activation. *Cell Death Differ, 22*(10), 1676-1686. doi:10.1038/cdd.2015.16

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649. doi:10.1093/bioinformatics/bts199

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge Cambridgeshire ; New York: Cambridge University Press.

Klausen, M. S., Jespersen, M. C., Nielsen, H., Jensen, K. K., Jurtz, V. I., Sonderby, C. K., . . . Marcatili, P. (2019). NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins, 87*(6), 520-527. doi:10.1002/prot.25674

Kummer, J. A., Broekhuizen, R., Everett, H., Agostini, L., Kuijk, L., Martinon, F., . . . Tschopp, J. (2007). Inflammasome components NALP 1 and 3 show distinct but separate expression profiles in human tissues suggesting a site-specific role in the inflammatory response. *J Histochem Cytochem, 55*(5), 443-452. doi:10.1369/jhc.6A7101.2006

Kuriakose, T., & Kanneganti, T. D. (2019). Pyroptosis in Antiviral Immunity. *Curr Top Microbiol Immunol*. doi:10.1007/82_2019_189

Laitinen, O. H., Svedin, E., Kapell, S., Nurminen, A., Hytonen, V. P., & Flodstrom-Tullberg, M. (2016). Enteroviral proteases: structure, host interactions and pathogenicity. *Rev Med Virol, 26*(4), 251-267. doi:10.1002/rmv.1883

Lei, X., Zhang, Z., Xiao, X., Qi, J., He, B., & Wang, J. (2017). Enterovirus 71 Inhibits Pyroptosis through Cleavage of Gasdermin D. *J Virol, 91*(18). doi:10.1128/JVI.01069-17

Levinsohn, J. L., Newman, Z. L., Hellmich, K. A., Fattah, R., Getz, M. A., Liu, S., . . . Moayeri, M. (2012). Anthrax lethal factor cleavage of Nlrp1 is required for activation of the inflammasome. *PLoS Pathog, 8*(3), e1002638. doi:10.1371/journal.ppat.1002638

Martinon, F., Burns, K., & Tschopp, J. (2002). The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta. *Mol Cell, 10*(2), 417-426. doi:10.1016/s1097-2765(02)00599-3

McCune, B. T., Lanahan, M. R., tenOever, B. R., & Pfeiffer, J. K. (2020). Rapid Dissemination and Monopolization of Viral Populations in Mice Revealed Using a Panel of Barcoded Viruses. *J Virol, 94*(2). doi:10.1128/JVI.01590-19

Meyerson, N. R., & Sawyer, S. L. (2011). Two-stepping through time: mammals and viruses. *Trends Microbiol, 19*(6), 286-294. doi:10.1016/j.tim.2011.03.006

Mitchell, P. S., Sandstrom, A., & Vance, R. E. (2019). The NLRP1 inflammasome: new mechanistic insights and unresolved mysteries. *Curr Opin Immunol, 60*, 37-45. doi:10.1016/j.coi.2019.04.015

Moayeri, M., Crown, D., Newman, Z. L., Okugawa, S., Eckhaus, M., Cataisson, C., . . . Leppla, S. H. (2010). Inflammasome sensor Nlrp1b-dependent resistance to anthrax is mediated by caspase-1, IL-1 signaling and neutrophil recruitment. *PLoS Pathog, 6*(12), e1001222. doi:10.1371/journal.ppat.1001222

Moayeri, M., Sastalla, I., & Leppla, S. H. (2012). Anthrax and the inflammasome. *Microbes Infect, 14*(5), 392-400. doi:10.1016/j.micinf.2011.12.005

Mukherjee, A., Morosky, S. A., Delorme-Axford, E., Dybdahl-Sissoko, N., Oberste, M. S., Wang, T., & Coyne, C. B. (2011). The coxsackievirus B 3C protease cleaves MAVS and TRIF to attenuate host type I interferon and apoptotic signaling. *PLoS Pathog, 7*(3), e1001311. doi:10.1371/journal.ppat.1001311

O'Donoghue, A. J., Eroy-Reveles, A. A., Knudsen, G. M., Ingram, J., Zhou, M., Statnekov, J. B., . . . Craik, C. S. (2012). Global identification of peptidase specificity by multiplex substrate profiling. *Nat Methods, 9*(11), 1095-1100. doi:10.1038/nmeth.2182

Okondo, M. C., Rao, S. D., Taabazuing, C. Y., Chui, A. J., Poplawski, S. E., Johnson, D. C., & Bachovchin, D. A. (2018). Inhibition of Dpp8/9 Activates the Nlrp1b Inflammasome. *Cell Chem Biol, 25*(3), 262-267 e265. doi:10.1016/j.chembiol.2017.12.013

Patel, M. R., Loo, Y. M., Horner, S. M., Gale, M., Jr., & Malik, H. S. (2012). Convergent evolution of escape from hepaciviral antagonism in primates. *PLoS Biol, 10*(3), e1001282. doi:10.1371/journal.pbio.1001282

Pickett, B. E., Greer, D. S., Zhang, Y., Stewart, L., Zhou, L., Sun, G., . . . Scheuermann, R. H. (2012). Virus pathogen database and analysis resource (ViPR): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. *Viruses, 4*(11), 3209-3226. doi:10.3390/v4113209

Qian, S., Fan, W., Liu, T., Wu, M., Zhang, H., Cui, X., . . . Qian, P. (2017). Seneca Valley Virus Suppresses Host Type I Interferon Production by Targeting Adaptor Proteins MAVS, TRIF, and TANK for Cleavage. *J Virol, 91*(16). doi:10.1128/JVI.00823-17

Rathinam, V. A., & Fitzgerald, K. A. (2016). Inflammasome Complexes: Emerging Mechanisms and Effector Functions. *Cell, 165*(4), 792-800. doi:10.1016/j.cell.2016.03.046

Rathinam, V. A., Vanaja, S. K., & Fitzgerald, K. A. (2012). Regulation of inflammasome signaling. *Nat Immunol, 13*(4), 333-342. doi:10.1038/ni.2237

Robinson, K. S., Teo, D. E. T., Tan, K. S., Toh, G. A., Ong, H. H., Lim, C. K., . . . Reversade, B. (2020). Enteroviral 3C protease activates the human NLRP1 inflammasome in airway epithelia. *Science, 370*(6521). doi:10.1126/science.aay2002

Sandstrom, A., Mitchell, P. S., Goers, L., Mu, E. W., Lesser, C. F., & Vance, R. E. (2019). Functional degradation: A mechanism of NLRP1 inflammasome activation by diverse pathogen enzymes. *Science, 364*(6435). doi:10.1126/science.aau1330

Sanjana, N. E., Shalem, O., & Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods, 11*(8), 783-784. doi:10.1038/nmeth.3047

Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun, 27*(2), 157-162. doi:10.1016/s0006-291x(67)80055-x

Schmidt, T., Schmid-Burgk, J. L., & Hornung, V. (2015). Synthesis of an arrayed sgRNA library targeting the human genome. *Sci Rep, 5*, 14987. doi:10.1038/srep14987

Sironi, M., Cagliani, R., Forni, D., & Clerici, M. (2015). Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nat Rev Genet, 16*(4), 224-236. doi:10.1038/nrg3905

Soler, V. J., Tran-Viet, K. N., Galiacy, S. D., Limviphuvadh, V., Klemm, T. P., St Germain, E., . . . Young, T. L. (2013). Whole exome sequencing identifies a mutation for a novel form of corneal intraepithelial dyskeratosis. *J Med Genet, 50*(4), 246-254. doi:10.1136/jmedgenet-2012-101325

Solomon, T., Lewthwaite, P., Perera, D., Cardosa, M. J., McMinn, P., & Ooi, M. H. (2010). Virology, epidemiology, pathogenesis, and control of enterovirus 71. *Lancet Infect Dis, 10*(11), 778-790. doi:10.1016/S1473-3099(10)70194-8

Spel, L., & Martinon, F. (2020). Detection of viruses by inflammasomes. *Curr Opin Virol, 46*, 59-64. doi:10.1016/j.coviro.2020.10.001

Stabell, A. C., Meyerson, N. R., Gullberg, R. C., Gilchrist, A. R., Webb, K. J., Old, W. M., . . . Sawyer, S. L. (2018). Dengue viruses cleave STING in humans but not in nonhuman primates, their presumed natural reservoir. *Elife, 7*. doi:10.7554/eLife.31919

Sun, D., Chen, S., Cheng, A., & Wang, M. (2016). Roles of the Picornaviral 3C Proteinase in the Viral Life Cycle and Host Cells. *Viruses, 8*(3), 82. doi:10.3390/v8030082

Taabazuing, C. Y., Griswold, A. R., & Bachovchin, D. A. (2020). The NLRP1 and CARD8 inflammasomes. *Immunol Rev*. doi:10.1111/imr.12884

Tan, J., George, S., Kusov, Y., Perbandt, M., Anemuller, S., Mesters, J. R., . . . Hilgenfeld, R. (2013). 3C protease of enterovirus 68: structure-based design of Michael acceptor inhibitors and their broad-spectrum antiviral effects against picornaviruses. *J Virol, 87*(8), 4339-4351. doi:10.1128/JVI.01123-12

Tenthorey, J. L., Kofoed, E. M., Daugherty, M. D., Malik, H. S., & Vance, R. E. (2014). Molecular basis for specific recognition of bacterial ligands by NAIP/NLRC4 inflammasomes. *Mol Cell, 54*(1), 17-29. doi:10.1016/j.molcel.2014.02.018

Terra, J. K., Cote, C. K., France, B., Jenkins, A. L., Bozue, J. A., Welkos, S. L., . . . Bradley, K. A. (2010). Cutting edge: resistance to Bacillus anthracis infection mediated by a lethal toxin sensitive allele of Nalp1b/Nlrp1b. *J Immunol, 184*(1), 17-20. doi:10.4049/jimmunol.0903114

Tian, X., Pascal, G., & Monget, P. (2009). Evolution and functional divergence of NLRP genes in mammalian reproductive systems. *BMC Evol Biol, 9*, 202. doi:10.1186/1471-2148-9-202

Timms, R. T., Zhang, Z., Rhee, D. Y., Harper, J. W., Koren, I., & Elledge, S. J. (2019). A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. *Science, 365*(6448). doi:10.1126/science.aaw4912

Ting, J. P., Lovering, R. C., Alnemri, E. S., Bertin, J., Boss, J. M., Davis, B. K., . . . Ward, P. A. (2008). The NLR gene family: a standard nomenclature. *Immunity, 28*(3), 285-287. doi:10.1016/j.immuni.2008.02.005

Vance, R. E., Isberg, R. R., & Portnoy, D. A. (2009). Patterns of pathogenesis: discrimination of pathogenic and nonpathogenic microbes by the innate immune system. *Cell Host Microbe, 6*(1), 10-21. doi:10.1016/j.chom.2009.06.007

Wang, C., Fung, G., Deng, H., Jagdeo, J., Mohamud, Y., Xue, Y. C., . . . Luo, H. (2019). NLRP3 deficiency exacerbates enterovirus infection in mice. *FASEB J, 33*(1), 942-952. doi:10.1096/fj.201800301RRR

Wang, D., Fang, L., Li, K., Zhong, H., Fan, J., Ouyang, C., . . . Xiao, S. (2012). Foot-and-mouth disease virus 3C protease cleaves NEMO to impair innate immune signaling. *J Virol, 86*(17), 9311-9322. doi:10.1128/JVI.00722-12

Wang, D., Fang, L., Wei, D., Zhang, H., Luo, R., Chen, H., . . . Xiao, S. (2014). Hepatitis A virus 3C protease cleaves NEMO to impair induction of beta interferon. *J Virol, 88*(17), 10252-10258. doi:10.1128/JVI.00869-14

Wang, H., Lei, X., Xiao, X., Yang, C., Lu, W., Huang, Z., . . . Wang, J. (2015). Reciprocal Regulation between Enterovirus 71 and the NLRP3 Inflammasome. *Cell Rep, 12*(1), 42-48. doi:10.1016/j.celrep.2015.05.047

Wang, Y., Qin, Y., Wang, T., Chen, Y., Lang, X., Zheng, J., . . . Zhong, Z. (2018). Pyroptosis induced by enterovirus 71 and coxsackievirus B3 infection affects viral replication and host response. *Sci Rep, 8*(1), 2887. doi:10.1038/s41598-018-20958-1

Wen, W., Yin, M., Zhang, H., Liu, T., Chen, H., Qian, P., . . . Li, X. (2019). Seneca Valley virus 2C and 3C inhibit type I interferon production by inducing the degradation of RIG-I. *Virology, 535*, 122-129. doi:10.1016/j.virol.2019.06.017

Wickliffe, K. E., Leppla, S. H., & Moayeri, M. (2008). Killing of macrophages by anthrax lethal toxin: involvement of the N-end rule pathway. *Cell Microbiol, 10*(6), 1352-1362. doi:10.1111/j.1462-5822.2008.01131.x

Xiang, Z., Li, L., Lei, X., Zhou, H., Zhou, Z., He, B., & Wang, J. (2014). Enterovirus 68 3C protease cleaves TRIF to attenuate antiviral responses mediated by Toll-like receptor 3. *J Virol, 88*(12), 6650-6659. doi:10.1128/JVI.03138-13

Xiang, Z., Liu, L., Lei, X., Zhou, Z., He, B., & Wang, J. (2016). 3C Protease of Enterovirus D68 Inhibits Cellular Defense Mediated by Interferon Regulatory Factor 7. *J Virol, 90*(3), 1613-1621. doi:10.1128/JVI.02395-15

Xiao, X., Qi, J., Lei, X., & Wang, J. (2019). Interactions Between Enteroviruses and the Inflammasome: New Insights Into Viral Pathogenesis. *Front Microbiol, 10*, 321. doi:10.3389/fmicb.2019.00321

Xu, H., Shi, J., Gao, H., Liu, Y., Yang, Z., Shao, F., & Dong, N. (2019). The N-end rule ubiquitin ligase UBR2 mediates NLRP1B inflammasome activation by anthrax lethal toxin. *EMBO J, 38*(13), e101996. doi:10.15252/embj.2019101996

Zaragoza, C., Saura, M., Padalko, E. Y., Lopez-Rivera, E., Lizarbe, T. R., Lamas, S., & Lowenstein, C. J. (2006). Viral protease cleavage of inhibitor of kappaBalpha triggers host cell apoptosis. *Proc Natl Acad Sci U S A, 103*(50), 19051-19056. doi:10.1073/pnas.0606019103

Zell, R. (2018). *Picornaviridae*-the ever-growing virus family. *Arch Virol, 163*(2), 299-317. doi:10.1007/s00705-017-3614-8

Zhong, F. L., Mamai, O., Sborgi, L., Boussofara, L., Hopkins, R., Robinson, K., . . . Reversade, B. (2016). Germline NLRP1 Mutations Cause Skin Inflammatory and Cancer Susceptibility Syndromes via Inflammasome Activation. *Cell, 167*(1), 187-202 e117. doi:10.1016/j.cell.2016.09.001

Zhong, F. L., Robinson, K., Teo, D. E. T., Tan, K. Y., Lim, C., Harapas, C. R., . . . Reversade, B. (2018). Human DPP9 represses NLRP1 inflammasome and

protects against autoinflammatory diseases via both peptidase activity and FIIND domain binding. *J Biol Chem, 293*(49), 18864-18878. doi:10.1074/jbc.RA118.004350

Chapter 4: Diverse Viral Proteases Activate the CARD8 Inflammasome

Abstract

In the ongoing Coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), SARS-CoV-2 infection potently activates inflammatory cytokines such as interleukin-1B (IL-1B) and interleukin-18 (IL-18). The mechanisms behind SARS-CoV-2-associated inflammation have yet to be uncovered. Mouse NLRP1B and human NLRP1 are among the inflammasome mediators known to mature pro-IL-1B and pro-IL-18 upon activation through proteolytic cleavage by the bacterial Lethal Toxin (LeTx) protease and picornaviral 3C proteases, respectively, resulting in degradation of the N-terminal domains of NLRP1 and liberation of the bioactive C-terminal domain, which includes the caspase activation and recruitment domain (CARD). However, natural pathogen-derived effectors that can activate human CARD8, a similar inflammasome mediator, have remained unknown. Here, we use an evolutionary model to identify several proteases from diverse coronaviruses that cleave both human NLRP1 and human CARD8, demonstrating that only cleavage of human CARD8 is sufficient to host-specific and virus-specific activation of the inflammasome. Our work demonstrates that CARD8 acts as a “tripwire” to recognize the enzymatic function of a wide range of viral proteases and suggests that host mimicry of viral polyprotein cleavage sites can be an evolutionary strategy to activate a robust inflammatory immune response.

Introduction

As the COVID-19 pandemic continues, understanding how the causative agent, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), antagonizes the

human host to induce disease is key in developing clinical leads for treatment of COVID-19 and other human-relevant coronavirus-mediated diseases. The life-threatening nature of many human-relevant coronaviruses, including Severe Acute Respiratory Syndrome Coronavirus 1 (SARS-CoV-1) and Middle Eastern Respiratory Syndrome Coronavirus (MERS-CoV), typically involves an elevated inflammatory innate immune response in the lower respiratory tract (Alosaimi et al., 2020), where patients with severe COVID-19 have reported increases in inflammatory monocytes and neutrophils, occurring alongside the presence of inflammatory cytokines including IL-1β, IL-6, IL-18, and TNF (G. Chen et al., 2020; N. Chen et al., 2020; Giamarellos-Bourboulis et al., 2020; Lucas et al., 2020; Mathew et al., 2020; Zhou et al., 2020). Despite the growing wealth of literature aimed toward understanding the pathogenesis of SARS-CoV-2, the mechanism by which SARS-CoV-2 infection triggers the inflammatory innate immune response remains unclear.

Inflammasomes are one such group of cytosolic immune sensor complexes responsible for the direct detection of microbial molecules or pathogen-encoded activities and subsequent propagation of inflammatory cytokines. Upon detection of microbial molecules or pathogen-encoded activities, inflammasome-forming sensor proteins serve as a platform for the recruitment and activation of proinflammatory caspases including caspase-1 (CASP1) through either a pyrin domain (PYD) or a caspase activation and recruitment domain (CARD) (Broz & Dixit, 2016; Rathinam & Fitzgerald, 2016). Active CASP1 mediates the maturation and release of the proinflammatory cytokines interleukin (IL)-1B and IL-18 (Broz & Dixit, 2016; Rathinam, Vanaja, & Fitzgerald, 2012). CASP1 also initiates a form of cell death known as

pyroptosis (Broz & Dixit, 2016; Rathinam et al., 2012). Together, these outputs provide robust defense against a wide array of eukaryotic, bacterial and viral pathogens (Broz & Dixit, 2016; Evavold & Kagan, 2019; Rathinam & Fitzgerald, 2016).

Recent studies suggest that NLRP1 is one such inflammasome mediator that serves as a sensor for a diverse range of positive-sense RNA viruses (Robinson et al., 2020; Tsu et al., 2021). Specifically, NLRP1 has been shown to detect the pathogen-encoded activities of the *Picornaviridae* family of viruses by mimicking the highly conserved 3C protease ($3C^{pro}$) cleavage sites found within the viral polyproteins (Tsu et al., 2021). Consequently, these virally encoded 3C proteases will cleave both the viral polyprotein sites as a part of their life cycle and NLRP1 to trigger activation of the inflammasome and downstream maturation of pro-inflammatory cytokines (Tsu et al., 2021). Because members of *Coronaviridae* similarly encode a 3C-like protease ($3CL^{pro}$) that necessarily cleaves numerous highly conserved protease cleavage sites within the viral polyprotein and antagonizes critical modulators of the host immune response as a part of their lifecycle (Chen et al., 2019; Moustaqil et al., 2021; V'Kovski, Kratzel, Steiner, Stalder, & Thiel, 2021; Zhu, Fang, et al., 2017; Zhu, Wang, et al., 2017), NLRP1 may also serve as a tripwire for coronaviral $3CL^{pro}$ activity.

Additional work has identified CARD8 to be yet another candidate tripwire sensor for coronaviral $3CL^{pro}$ activity (Wang et al., 2021). CARD8 has a nearly identical exon-intron organization of the FIIND-CARD module that serves as the platform for NLRP1 inflammasome formation (Finger et al., 2012; Frew, Joag, & Mogridge, 2012; Johnson et al., 2018), suggesting a common ancestral origin of these genes and a similar mechanism of action (D'Osualdo et al., 2011). Expanding on this similarity, CARD8 has

been shown to activate formation of the inflammasome upon cleavage by the HIV protease in the presence of non-nucleoside reverse transcriptase inhibitors, demonstrating that CARD8 also detects pathogenic protease activity (Wang et al., 2021).

Here we investigate the hypothesis that SARS-CoV-2 $3CL^{pro}$ cleaves and activates human NLRP1 and CARD8. To pursue this hypothesis, we focused on SARS-CoV-2 and other human-relevant viruses in the *Coronaviridae* family, which encompass a diverse set of human respiratory pathogens including 229E, NL63, HKU1, SARS-CoV-1, MERS and MHV. We show that many of these coronaviral $3CL^{pros}$ can strongly activate CARD8, but not NLRP1, via cleavage. We also find that variation in the cleavage site within the human population, leads to changes in cleavage susceptibility and inflammasome activation. Interestingly, we observe that proteases from multiple human-relevant coronaviruses cleave and activate human CARD8, but not CARD8 of a megabat*, Rousettus aegyptiacus*, supporting a role for an evolutionary conflict driven by viral proteases toward different host inflammatory disease trajectories. Taken together, our work highlights the role of NLRP1 and CARD8 in sensing and responding to coronaviral proteases by evolving cleavage motifs that mimic natural sites of proteolytic cleavage in the viral polyprotein.

Results

SARS-CoV-2 3CLpro cleaves human NLRP1 and CARD8 at predicted sites within the exposed N-terminal region.

Our hypothesis that NLRP1 and CARD8 can sense coronaviral 3CL protease cleavage is based on two prior observations. First, both human NLRP1 and mouse NLRP1B have recently been shown to be activated by N-terminal proteolysis via picornaviral 3C proteases (Tsu et al., 2021). Second, recent work by Wang et al., 2021 demonstrates that proteolytic cleavage by drug-induced active HIV protease will result in activation of the CARD8 inflammasome. To test this hypothesis, we first generated a predictive model for $3CL^{pro}$ cleavage site specificity. We chose to focus on betacoronaviruses, as there is a deep and diverse collection of publicly available viral sequences within this genus due to their importance as human pathogens including SARS-CoV-2, 229E, NL63, HKU1, SARS-CoV-1, MERS, MHV. We first compiled complete enterovirus polyprotein sequences from the Viral Pathogen Resource (ViPR) database (Pickett et al., 2012) and extracted and concatenated sequences for each of the cleavage sites within the polyproteins (Supplementary file 4.1). After removing redundant sequences, we used the MEME Suite (Bailey, Johnson, Grant, & Noble, 2015) to create the following $3CL^{pro}$ cleavage motif: XX[A/Φ]Q[S/A]XXX (where Φ denotes a hydrophobic residue and X denotes any amino acid), which is broadly consistent with previous studies that have experimentally profiled the substrate specificity of coronaviral $3CL^{pros}$ (Figure 4.1A).

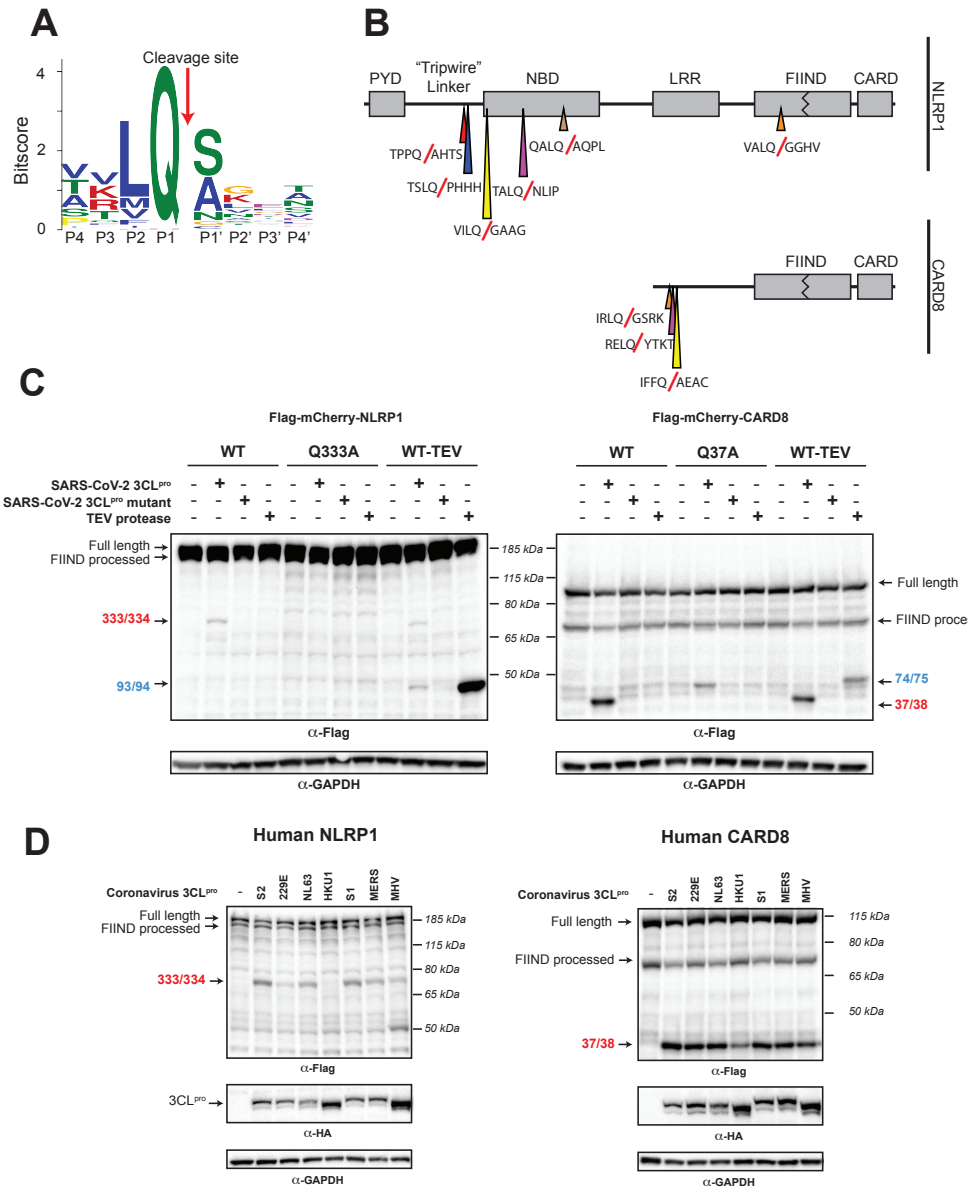


Figure 4.1. Identification of SARS-CoV-2 3CL protease cleavage motifs in NLRP1 and CARD8. (A) Eight amino acid polyprotein cleavage motif for betacoronaviruses (labeled as positions P4 to P4') generated from 60 coronaviral polyprotein sequences using the MEME Suite (Figure 4.2). (B) Schematic of the domain structures of NLRP1 and CARD8, with predicted cleavage sites (triangles). (C) Immunoblot depicting human NLRP1 (left) or CARD8 (right) cleavage by S2 3CLpro and TEV protease. HEK293T cells were co-transfected using 100 ng of the indicated Flag-tagged mCherry-NLRP1 or CARD8 fusion plasmid constructs with HA-tagged protease constructs containing either no protease (EV), 5 ng of active S2 protease (S2), 5 ng of catalytically inactive S2 protease (S2*), or 250 ng of TEV protease (TEV) and immunoblotted with the indicated antibodies. (D) Immunoblot depicting human NLRP1 (left) and CARD8 (right) cleavage by 5 ng of SARS-CoV-2 (S2), SARS-CoV-1 (S1), MERS-CoV (MERS), HCoV-229E (229E), HCoV-NL63 (NL63), HCoV-HKU1 (HKU1), or M-CoV strain Mouse hepatitis virus (MHV) 3CLpro.

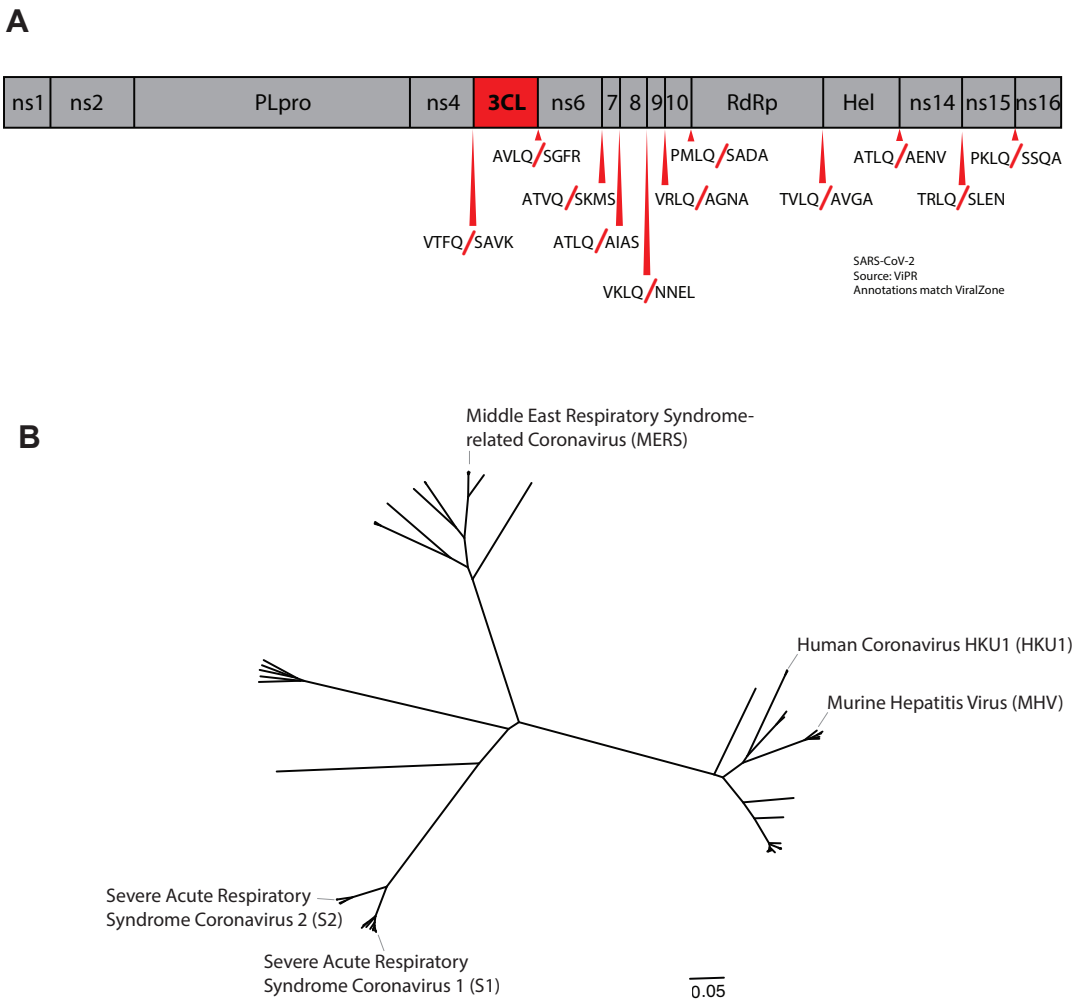


Figure 4.2. Motif capture of known human targets of coronaviral 3CLpro. Schematic of 3CLpro cleavage sites (red triangles) within the polyprotein of SARS-CoV-2 (S2), the causative agent of COVID-19. Shown are the eight amino acids flanking each cleavage site within the polyprotein. (B) Phylogenetic tree of 60 coronaviral polyprotein sequences depicting members of the betacoronavirus genus sampled in this study with human relevant coronaviruses in parentheses (Supplementary file 4.1).

We next used our refined model to conduct a motif search for 3CLpro cleavage sites in NLRP1 and CARD8 using Find Individual Motif Occurrences (FIMO) (Grant, Bailey, & Noble, 2011; Tsu et al., 2021). We identified five occurrences of the motif across the full-length human NLRP1 protein and three occurrences of the motif across the full-length human CARD8 protein (Figure 4.1B).

To assess whether human NLRP1 or human CARD8 is cleaved by coronaviral $3CL^{pro}$, we co-expressed a N-terminal mCherry-tagged wild-type (WT) human NLRP1 or human CARD8 with the $3CL^{pro}$ from the COVID-19 agent, SARS-CoV-2 (S2) in HEK293T cells (Figure 4.1C). The mCherry tag stabilizes and allows visualization of putative N-terminal cleavage products, similar to prior studies (Chavarria-Smith, Mitchell, Ho, Daugherty, & Vance, 2016). We observed that the WT but not catalytically inactive (C145A) S2 $3CL^{pro}$ cleaved NLRP1, resulting in a cleavage product with a molecular weight consistent with our predicted $3CL^{pro}$ cleavage at the predicted 330-GCTQGSER-337 site (Figure 4.1C). Based on the presence of a single cleavage product, we assume that the other predicted sites are poor substrates for $3CL^{pro}$. To determine if the cleavage occurs between residues 333 and 334, we mutated the P1 glutamine to an alanine (Q333A), which abolished $3CL^{pro}$ cleavage of NLRP1 (Figure 4.1C). S2 $3CL^{pro}$ cleavage of WT NLRP1 resulted in a weak cleavage product when compared to the previously described system in which a TEV protease site was introduced into the linker region of NLRP1 (Chavarria-Smith et al., 2016) (Figure 4.1C). Likewise, we observed that WT S2 $3CL^{pro}$ also cleaved CARD8, resulting in a cleavage product with a molecular weight that corresponds to our predicted 34-IRLQGSRK-41 site. Interestingly, S2 $3CL^{pro}$ cleavage of WT CARD8 resulted in a stronger cleavage

product when compared to the TEV protease cleavage of WT-TEV CARD8. Additionally, in determining the site of cleavage between residues 37 and 38, we mutated the P1 glutamine to an alanine (Q37A) which abolished the initial CARD8 site of 3CL^{pro} but revealed a secondary site of cleavage at site Q61 (Figure 4.1C, Figure 4.3). Taken together, these results indicate that cleavage of WT NLRP1 and CARD8 by the SARS-CoV-2 3CL^{pro} is robust and specific.

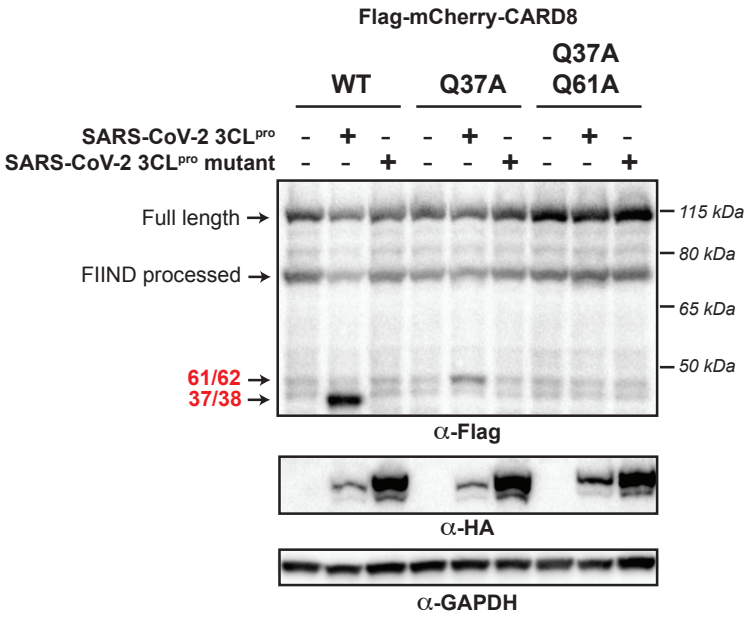


Figure 4.3. Mutations at the primary and secondary cleavage sites ablate CARD8 cleavage. Immunoblot depicting human CARD8 cleavage by S2 3CLpro. HEK293T cells were co-transfected using 100 ng of the indicated Flag-tagged mCherry-CARD8 fusion plasmid constructs with HA-tagged protease constructs containing either no protease (EV), 5 ng of active S2 protease (S2), or 5 ng of catalytically inactive S2 protease (S2) and immunoblotted with the indicated antibodies.

Our evolutionary model predicted that NLRP1 and CARD8 would be cleaved by a broad range of 3CL^{pro} from the betacoronavirus genus (Figure 4.1A). To better understand the interaction between coronaviruses and these related inflammasome mediators, we examined whether 3CL^{pro} from other human-relevant coronaviruses were

also able to cleave NLRP1 and CARD8. To test this hypothesis, we cloned 3CL^{pro} from Human Coronavirus 229E (229E, genus: *Alphacoronavirus*), Human Coronavirus NL63 (NL63, genus: *Alphacoronavirus*), Human Coronavirus HKU1 (HKU1, genus: *Betacoronavirus*), SARS-CoV-1 (S1, genus: *Betacoronavirus*), MERS-CoV (MERS, genus: *Betacoronavirus*) and hepatitis-related strain of Murine Coronavirus (MHV, genus: *Betacoronavirus*) to compare them to the 3CL^{pro} from S2 (species: *Betacoronavirus*). The overall structures of many of these proteases are similar and the cleavage motifs are closely related (Roe, Junod, Young, Beachboard, & Stobart, 2021). Consistent with this predicted target similarity and prior data, we found that nearly every tested member of *Coronaviridae* 3CL^{pro} was able to cleave both human NLRP1 and CARD8 between the same residues targeted by the S2 3CL^{pro} (Figure 4.1D).

The SARS-CoV-2 3CLpro-mediated cleavage activates human CARD8, but not NLRP1

Previous studies with human NLRP1 and human CARD8 showed that cleavage was sufficient to activate the inflammasome in a reconstituted inflammasome assay (Chavarria-Smith et al., 2016; Robinson et al., 2020; Tsu et al., 2021; Wang et al., 2021). Using the same assay, in which plasmids-encoding human NLRP1, CASP1, ASC and IL-1β or human CARD8, CASP1 and IL-1β are transfected into HEK293T cells, we tested whether S2 3CL^{pro} activates the NLRP1 or the CARD8 inflammasome. To our surprise, we discovered that HEK293T cells endogenously expressed CARD8 and upon transfection with only ASC, IL-1β and S2 3CL^{pro} we observed robust activation as measured by CASP1-dependent processing of pro-IL-1β to the active p17 form, where knocking out CARD8 in these cells prevented pro-IL-1β processing (Figure 4.4A

– Top). To confirm that S2 $3CL^{pro}$-induced inflammasome activation resulted in release of bioactive IL-1β from cells, we measured active IL-1β levels in the culture supernatant using cells engineered to express a reporter gene in response to soluble, active IL-1β. When compared to a standard curve (Figure 4.5), we found that S2 $3CL^{pro}$ treatment resulted in release of >3 ng/ml of active IL-1β into the culture supernatant (Figure 4.4A – Bottom). To avoid confounding activation by endogenous CARD8, we reconstituted inflammasomes in CARD8 KO cells using TEV-engineered NLRP1 and CARD8 with TEV protease as positive controls and observed that the S2 $3CL^{pro}$ again results in robust activation of CARD8 inflammasomes, but not NLRP1 inflammasomes Importantly, we found that in both western blot and cell culture assays, $3CL^{pro}$-induced inflammasome activation was ablated when position 37 was mutated, validating that S2 $3CL^{pro}$ cleavage at a single site is both necessary and sufficient to activate CARD8 (Figure 4.4B, Figure 4.6). As expected, S2 $3CL^{pro}$ activation of the CARD8 inflammasome was prevented by introduction of a mutation in the CARD8 FIIND (S297A) (Figure 4.7), which prevents FIIND auto-processing and the release of the bioactive C-terminal UPA–CARD (D'Osualdo et al., 2011; Wang et al., 2021). Taken together, our results are consistent with S2 $3CL^{pro}$ activating the CARD8, but not NLRP1, inflammasome that is dependent on FIIND auto-processing and $3CL^{pro}$ site-specific cleavage.

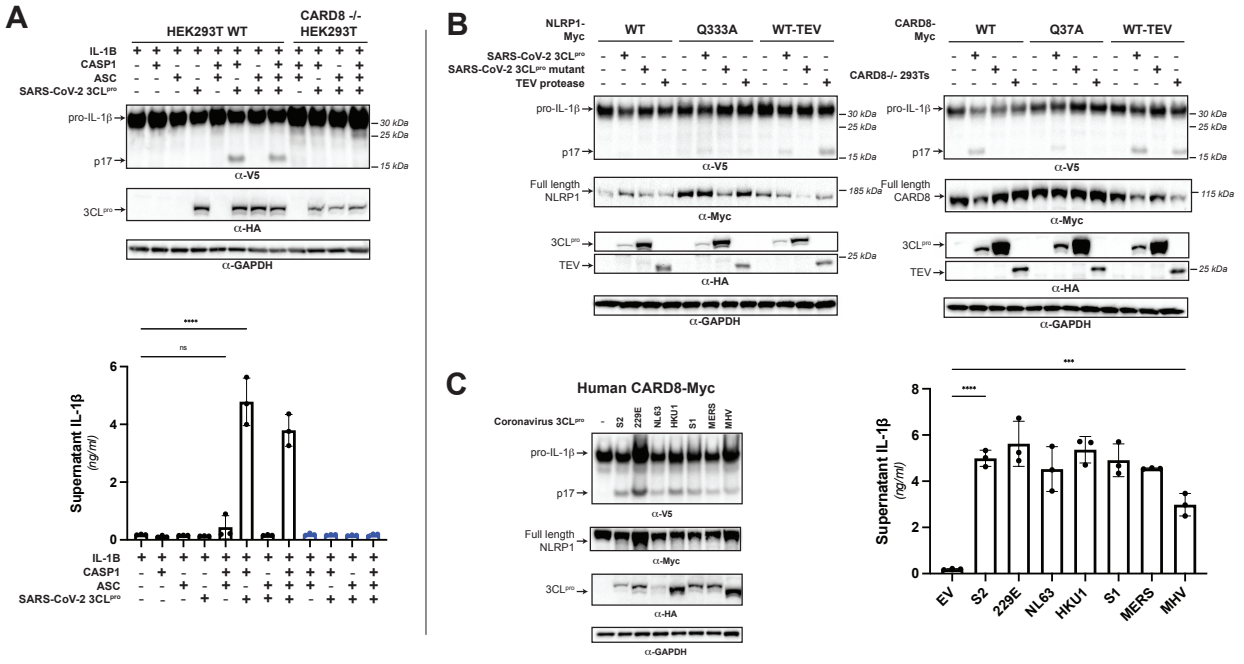


Figure 4.4. SARS2 3CL activates CARD8 but not NLRP1. (A) (Left) Immunoblot depicting endogenous human CARD8 activation (maturation of IL-1β) by S2 3CLpro in either WT or CARD8 -/- HEK293T cells. HEK293T cells were co-transfected using different combinations of 5 ng of S2 3CLpro, 50 ng V5-IL-1β, 100 ng CASP1, and 5 ng ASC, and immunoblotted with the indicated antibodies. Appearance of the mature p17 band of IL-1β indicates successful assembly of the CASP1- and CARD8-dependent inflammasome. (Right) Bioactive IL-1β in the culture supernatant was measured using HEK-Blue IL-1β reporter cells, which express secreted embryonic alkaline phosphatase (SEAP) in response to extracellular IL-1β. Supernatant from cells transfected as in (A) was added to HEK-Blue IL-1β reporter cells and SEAP levels in the culture supernatant from HEK-Blue IL-1β reporter cells were quantified by the QUANTI-Blue colorimetric substrate. Transfections were performed in triplicate and compared to the standard curve generated from concurrent treatment of HEK-Blue IL-1β reporter cells with purified human IL-1β (Figure 4.5). Data corresponding to HEK293T WT conditions were analyzed using one-way ANOVA with Tukey's post-test. **** = p<0.0001, ns = not significant. (B) Immunoblot depicting human NLRP1 (left) or CARD8 (right) activation (maturation of IL-1β) by S2 3CLpro and TEV protease. (left) CARD8 -/- HEK293T cells were co-transfected using 50 ng V5-IL-1β, 100 ng CASP1, 5 ng ASC, and 4 ng of the indicated Myc-tagged NLRP1 with HA-tagged protease constructs containing either no protease (EV), 5 ng of active S2 protease (S2), 5 ng of catalytically inactive S2 protease (S2*), or 250 ng of TEV protease (TEV). (right) CARD8 -/- HEK293T cells were co-transfected using 50 ng V5-IL-1β, 100 ng CASP1, and 50 ng of the indicated Myc-tagged CARD8 with HA-tagged protease constructs containing either no protease (EV), 5 ng of active S2 protease (S2), 5 ng of catalytically inactive S2 protease (S2*), or 250 ng of TEV protease (TEV). Both were immunoblotted with the indicated antibodies. Appearance of the mature p17 band of IL-1β indicates successful assembly of either the NLRP1 or CARD8 inflammasome and activation of CASP1. (C) (Left) Immunoblot depicting human CARD8 activation (maturation of IL-1β) by 5 ng of S2, S1, MERS, 229E, NL63, HKU1, or MHV 3CLpro. (Right) Release of bioactive IL-1β into the culture supernatant was measured using HEK-Blue IL-1β reporter cells as in Figure 4.4A. Data were analyzed using one-way ANOVA with Tukey's post-test. *** = p<0.001, **** = p<0.0001.

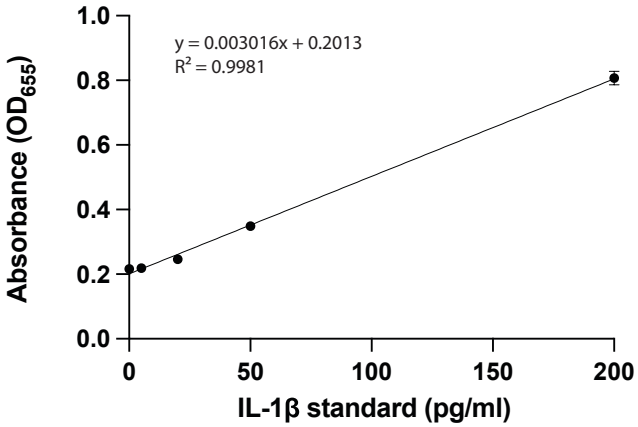


Figure 4.5. Standard curve for Figure 4.4A. Purified human IL-1β was added in duplicate to the indicated final concentration to HEK-Blue IL-1β reporter cells and SEAP activity was measured by increased absorbance at OD655. The indicated linear fit was used to calculate absolute concentrations of bioactive IL-1β from culture supernatants shown in Figure 4.4A. Note that supernatants from inflammasome-transfected cells was diluted 10-fold before addition to HEK-Blue IL-1β reporter cells to ensure that levels fell within the linear range of the indicated standard curve. Standard curves were generated in an identical manner for each panel of HEK-Blue data shown.

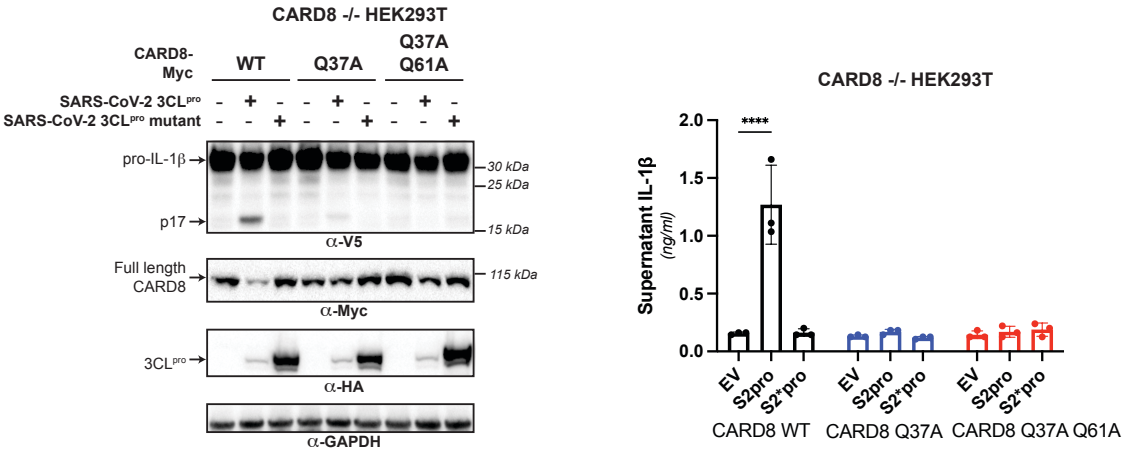


Figure 4.6. Mutations at the primary and secondary cleavage sites ablate CARD8 activation. (A) (Left) Immunoblot depicting human CARD8 activation (maturation of IL-1β) by S2 3CLpro. CARD8 -/- HEK293T cells were co-transfected using 50 ng V5-IL-1β, 100 ng CASP1, and 50 ng of the indicated Myc-tagged CARD8 with HA-tagged protease constructs containing either no protease (EV), 5 ng of active S2 protease (S2), or 5 ng of catalytically inactive S2 protease (S2*) and immunoblotted with the indicated antibodies. (Right) Release of bioactive IL-1β into the culture supernatant was measured using HEK-Blue IL-1β reporter cells as in Figure 4.4A. Data were analyzed using two-way ANOVA with Sidak's post-test. **** = p<0.0001. (B) Immunoblot depicting activation of the human CARD8 Q37A and Q37A-Q61A mutants using a TEV-engineered site at position #. CARD8 -/- HEK293T cells were co-transfected using 50 ng V5-IL-1β, 100 ng CASP1, and 50 ng of the indicated Myc-tagged CARD8-TEV with HA-tagged protease constructs containing either no protease (EV) or 250 ng of TEV protease.

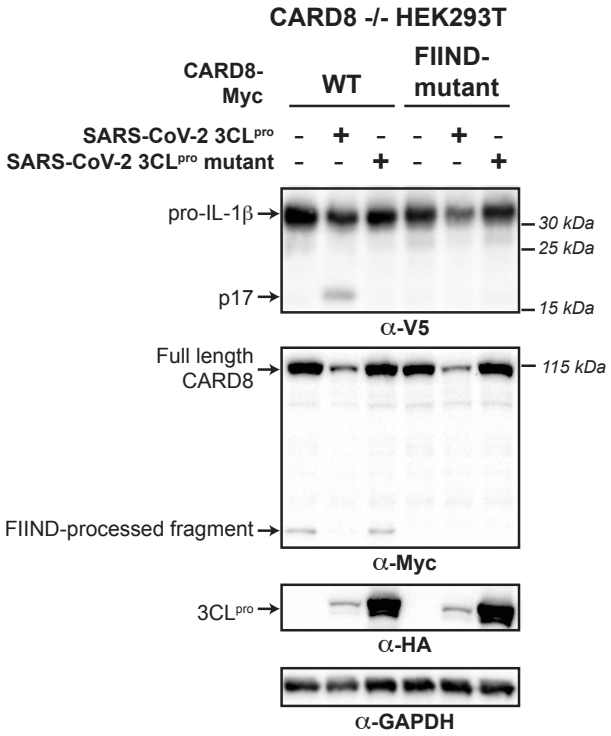


Figure 4.7. 3CLpro-mediated activation of the human CARD8 inflammasome depends on FIIND auto-processing. CARD8 -/- HEK293T cells were transfected with either WT CARD8 or a FIIND auto-processed defective mutant (S297A) along with other components of the CARD8 inflammasome (CASP1 and IL-1β) as in Figure 4.4B. Only cells transfected with WT CARD8 can produce mature IL-1β upon co-transfection with S2 3CLpro or TEV protease as indicated by the appearance of the mature p17 band.

Because we observed that nearly all tested human-relevant coronaviral 3CL^{pros} were able to cleave CARD8, we sought to better understand whether activation of the CARD8 inflammasome could similarly explain the mechanism behind broad coronaviral inflammatory immunopathology. Revisiting the 3CL^{pros} from S2, 229E, NL63, HKU1, S1, MERS and MHV, we found that expression of every tested coronavirus 3CL^{pro} resulted in activation of the inflammasome (Figure 4.4C).

CARD8 diversification confers host differences in susceptibility to coronaviral 3CLpro cleavage and inflammasome activation

Our evolutionary model in which CARD8 is evolving in conflict with coronaviral 3CL^{pro} suggests that changes in CARD8 within the human population would confer host-specific differences to CARD8 cleavage and inflammasome activation. To test this hypothesis, we aligned CARD8 from human population sampling and compared the sequences around the site of S2 3CL^{pro} cleavage. Using GnomAD (Karczewski et al., 2020), we sampled the alternative alleles within the direct cleavage site (Figure 4.8A). While this region does not appear to be highly polymorphic in humans, we note two alternative alleles (rs12463023 and rs138177358) that result in a S39P mutation and is present in ~1 in every 62 alleles sampled and a R40W mutation and is present in ~1 in every 1488 alleles sampled, respectively. Introducing these mutations into CARD8, we find the S39P mutation reduces CARD8 cleavage susceptibility to S2 3CL^{pro}, resulting in similar S2 3CL^{pro}-mediated secondary cleavage product to the Q37A point mutant at around position 61, but R40W results in similar cleavage to WT CARD8 (Figure 4.1C, Figure 4.3, Figure 4.8A). We find that reduced cleavage of this S39P human allele is

reflected in loss of inflammasome activation in response to S2 3CL^{pro} (Figure 4.8B), supporting the aforementioned notion that single changes in cleavage site can have drastic impacts on the ability of different hosts to respond to the presence of cytoplasmic 3CL^{pro}.

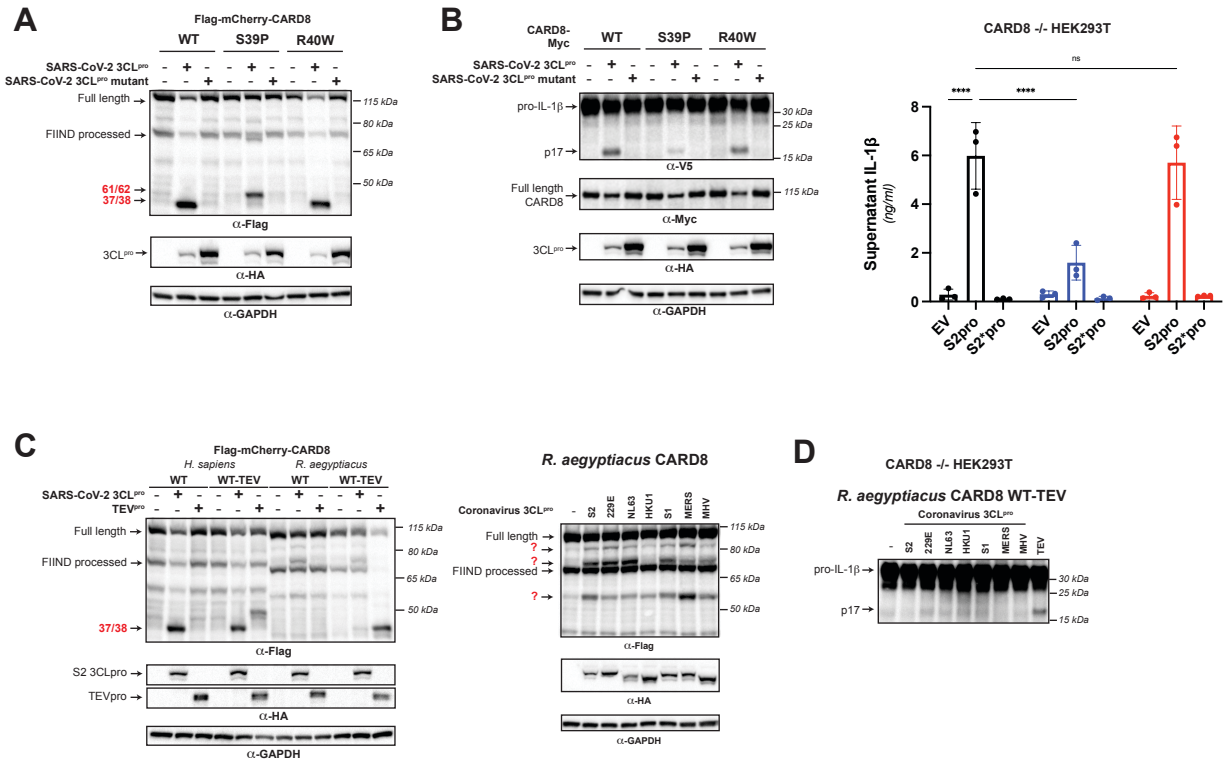


Figure 4.8. Host differences within the human population and bats alter susceptibility to the CARD8 inflammasome. (A) Immunoblot depicting S2 3CLpro cleavage susceptibility of the indicated naturally occurring human population 8mer site variants introduced into human CARD8, S39P and R40W. (B) (Left) Immunoblot depicting S2 3CLpro-mediated activation of the Myc-tagged human CARD8 variants, S39P and R40W. (Right) Release of bioactive IL-1β into the culture supernatant was measured using HEK-Blue IL-1β reporter cells as in Figure 4.4A. Data were analyzed using two-way ANOVA with Sidak's post-test. **** = p<0.0001, ns = not significant. (C) (Left) Immunoblot depicting S2 3CLpro cleavage susceptibility of *R. aegyptiacus* (megabat) CARD8 compared to the human version. (Right) Immunoblot depicting *R. aegyptiacus* (megabat) CARD8 cleavage susceptibility by 5 ng of S2, S1, MERS, 229E, NL63, HKU1, or MHV 3CLpro. (D) Immunoblot depicting 5ng of 3CLpro- or 250ng TEV protease-mediated activation of 2.5ng of *R. aegyptiacus* TEV-engineered CARD8.

We further observed that this cleavage site is absent in bat species, reservoir hosts of emerging viruses including SARS-related viruses. To assess whether bat

CARD8 were functional and capable of sensing S2 3CL^{pro}, we transfected both human and *R. aegyptiacus* (megabat) versions of CARD8 WT or TEV-engineered CARD8 into HEK293T cells and found that there was weak cleavage of megabat CARD8 at additional sites not found in human CARD8 (Figure 4.8C). Reconstituting the megabat inflammasome with human CASP1 and IL-1B, we demonstrated TEV-engineered megabat CARD8 activation upon cleavage by TEV protease, whereas activation was undetectable in the presence of S2 3CL^{pro} (Figure 4.8D). Taken together, these data suggest that while megabat CARD8 can be activated by site-specific cleavage, cleavage by human-relevant coronaviral 3CL^{pros} do not induce megabat CARD8 activation, possibly offering clues as to why bats do not suffer inflammation-related immunopathology.

Discussion

Earlier work demonstrated that the inflammasome protein, NLRP1, serves as a sensor for diverse 3C proteases from the *Picornaviridae* family of human pathogens by mimicking highly conserved protease cleavage sites found within viral polyproteins (Robinson et al., 2020; Tsu et al., 2021). Much like the 3C protease of picornaviruses, the 3CL protease of coronaviruses need to cleave numerous sites within the viral polyprotein to reproduce, and thus changing 3CL protease specificity requires concomitant changes to several independent cleavage sites. This evolutionary constraint potentially facilitates host protein mimicry of the 3CL protease cleavage motif.

Here we show that another inflammasome protein, CARD8, serves as a sensor for proteases from the *Coronaviridae* family of human pathogens by exploiting the evolutionary constraint of the 3CL protease. Much like NLRP1, the C-terminal CARD-

containing fragment of CARD8 non-covalently associates with the N-terminal fragment and upon cleavage, the N-terminal region degrades and C-terminal fragment releases to promote assembly of the active inflammasome (Wang et al., 2021). This allows the N-terminal region of CARD8 to evolve to be recognized by pathogenic effectors. In this study, we show that 3CL proteases from human-relevant coronaviruses such as SARS-CoV-2 (S2), 229E-CoV (229E), NL63-CoV (NL63), HKU1-CoV (HKU1), SARS-CoV-1 (S1), MERS-CoV (MERS) and murine hepatitis virus (MHV) specifically cleave at the same site within human CARD8, leading to activation of the CARD8 inflammasome and release of pro-inflammatory cytokines such as pro-IL-1β. Together with work by Wang et al., 2021, our work identifies CARD8 as a sensor for a diverse range of viral proteases as pathogen-encoded activators of human CARD8.

Because both NLRP1 and CARD8 share the unique domain architecture to allow for N-terminal degradation, we initially speculated that both could recognize 3CL protease cleavage. By harvesting publicly available sequences for known 3CL cleavage sites, we created a 3CL cleavage motif that successfully predicted the site of coronaviral 3CL cleavage at position 333-334 within NLRP1 and 37-38 within CARD8. Interestingly, when we reconstituted the inflammasome, only CARD8 elicited a detectable activation signal in the presence of 3CL proteases. Thus, we identify CARD8 as a general coronaviral 3CL protease sensor. Sampling the human population, we also identify a cleavage-resistant human variant that does not activate the inflammasome upon cytoplasmic expression of 3CL. To further explore the role of CARD8 in other mammals, we show that megabats (family *Pteropodidae*), the presumed reservoir for many human-relevant betacoronaviruses (Latinne et al., 2020), encode a functional CARD8 that can

be engineered with a TEV cleavage site to activate the inflammasome in the presence of human CASP1 and TEV protease. When in the presence of the 3CL proteases used in this study, megabat CARD8 does not activate the inflammasome. This lack of activation may be indicative of differences in immunopathologies between coronavirus-infected humans and megabats and expand upon the role of megabats as coronaviral reservoirs.

Taken together, our work expands upon earlier work demonstrating that host mimicry of viral polyprotein cleavage motifs is an evolutionary strategy in the ongoing arms race between host and viruses. NLRP1 and CARD8 represent two cases where this mimicry is coupled with cleavage-activating immunity. Additional studies expanding the range of viral proteases, the diversity at the host CARD8 and NLRP1 cleavage sites and infection-level studies will be needed to further elucidate how this functional consequence of inflammasome activation has evolved. Beyond NLRP1 and CARD8, we expect that this work will inspire discovery of other novel mammalian ETIs involved in broadly recognizing host-pathogen conflicts.

Materials and Methods

Motif generation and search

To build the motif, 995 nonredundant betacoronaviral polyprotein sequences were collected from the Viral Pathogen Resource (ViPR) and aligned with 5 well-annotated reference enteroviral polyprotein sequences from RefSeq. P1 and P1’ of the annotated cleavage sites across the RefSeq sequences served as reference points for putative cleavage sites across the 995 ViPR sequences. Four amino acyl residues upstream (P4-P1) and downstream (P1’-P4’) of each cleavage site were extracted from

every MAFFT-aligned polyprotein sequence, resulting in 1000 sets of cleavage sites (RefSeq sites included). Each set of cleavage sites representative of each polyprotein was then concatenated. Next, duplicates were removed from the concatenated cleavage sites. The remaining 60 nonredundant, concatenated cleavage sites were then split into individual 8-mer cleavage sites and were aligned using MAFFT to generate Geneious-defined sequence logo information at each aligned position. Pseudo-counts to the position-specific scoring matrix were adjusted as described previously (Tsu et al., 2021).

Sequence alignments and phylogenetic trees

Complete polyprotein sequences from 60 betacoronaviruses with non-redundant cleavage sites (see 'Motif generation and search' section above) were downloaded from ViPR. Sequences were aligned using MAFFT (Katoh & Standley, 2013) and a neighbor-joining phylogenetic tree was generated using Geneious software (Kearse et al., 2012).

Plasmids and constructs

For NLRP1 and CARD8 cleavage assays, the coding sequences of human NLRP1 WT (NCBI accession NP_127497.1), human NLRP1 mutants (Q333A), human NLRP1 TEV, human CARD8 (NCBI accession NP_001171829.1), human CARD8 mutants (Q37A, Q37A Q61A, S39P, R40W), human CARD8 TEV, *Rousettus aegyptiacus* (megabat) CARD8 (NCBI accession XP_016010896), and megabat CARD8 TEV were cloned into the pcDNA5/FRT/TO backbone (Invitrogen, Carlsbad, CA) with an N-terminal 3xFlag and mCherry tag. For NLRP1 and CARD8 activation, the

same sequences were cloned into the pQCXIP vector backbone (Takara Bio, Mountain View, CA) with a C-terminal Myc tag. Vectors containing the coding sequences of human NLRP1 TEV (NLRP1-TEV2), human CARD8, ASC, human CASP1, human IL-1β-V5, and TEV protease (Chavarria-Smith et al., 2016) were generous gifts from Dr. Russell Vance, UC Berkeley. Single point mutations were made using overlapping stitch PCR.

3CLpro sequences were ordered as gBlocks (Integrated DNA Technologies, San Diego, CA). Each 3CLpro was cloned with an N-terminal HA tag into the QCXIP vector backbone, flanked by polyprotein cleavage sites fused to N-terminal eGFP and C-terminal mCherry. Catalytic mutations were made using overlapping stitch PCR.

Following cloning, all plasmid stocks were sequenced across the entire inserted region to verify that no mutations were introduced during the cloning process.

Cell culture and transient transfection

All cell lines (HEK293T, HEK-Blue-IL-1β) are routinely tested for mycoplasma by PCR kit (ATCC, Manassas, VA) and kept a low passage number to maintain less than one year since purchase, acquisition or generation. HEK293T cells were obtained from ATCC (catalog # CRL-3216) and HEK-Blue-IL-1β cells were obtained from Invivogen (catalog # hkb-il1b) and all lines were verified by those sources. All cells were grown in complete media containing DMEM (Gibco, Carlsbad, CA), 10% FBS (Peak Serum, Wellington, CO), and appropriate antibiotics (Gibco, Carlsbad, CA). For transient transfections, HEK293T cells were seeded the day prior to transfection in a 24-well plate (Genesee, El Cajon, CA) with 500 µl complete media. Cells were transiently

transfected with 500 ng of total DNA and 1.5 µl of Transit X2 (Mirus Bio, Madison, WI) following the manufacturer's protocol. HEK-Blue IL-1β reporter cells (Invivogen, San Diego, CA) were grown and assayed in 96-well plates (Genesee, El Cajon, CA).

293T knockouts

293T knockouts were generous gifts from Dr. Russell Vance, UC Berkeley.

NLRP1 and CARD8 cleavage assays

100 ng of epitope-tagged human NLRP1 WT, human NLRP1 mutant (Q333A), human NLRP1 TEV, human CARD8 WT, human CARD8 NLRP1 mutants (Q37A, Q37A Q61A, S39P, R40W), human CARD8 TEV, megabat CARD8 WT or megabat CARD8 TEV was co-transfected with 5 ng of HA-tagged protease-producing constructs. Twenty-four hours post-transfection, the cells were harvested, lysed in 1x NuPAGE LDS sample buffer (Invitrogen, Carlsbad, CA) containing 5% β-mercaptoethanol (Fisher Scientific, Pittsburg, PA) and immunoblotted with antibodies described in Table 3.1.

NLRP1 and CARD8 activity assays

For human NLRP1 activation assays, 5 ng of ASC, 100 ng of human CASP1, 50 ng of human IL-1β-V5, and 100 ng of various protease-producing constructs were co-transfected with 4 ng of either pQCXIP empty vector, wild-type or mutant pQCXIP-NLRP1-Myc constructs. For human CARD8 activation assays, 100 ng of human CASP1, 50 ng of human IL-1β-V5, and 100 ng of various protease-producing constructs were co-transfected with 50 ng of either pQCXIP empty vector, wild-type or mutant

pQCXIP-CARD8-Myc constructs. For megabat CARD8 activation assays, 100 ng of human CASP1, 50 ng of human IL-1β-V5, and 5 ng of various protease-producing constructs were co-transfected with 2.5 ng megabat CARD8 constructs. Twenty-four hours post-transfection, cells were harvested and lysed in 1x NuPAGE LDS sample buffer containing 5% β-mercaptoethanol or in 1x RIPA lysis buffer with protease inhibitor cocktail (Roche) and immunoblotted with antibodies described in Table 3.1 or culture media was harvested for quantification of IL-1β levels by HEK-Blue assays (see below).

HEK-Blue IL-1β assay

To quantify the levels of bioactive IL-1β released from cells, we employed HEK-Blue IL-1β reporter cells (Invivogen, San Diego, CA). In these cells, binding to IL-1β to the surface receptor IL-1R1 results in the downstream activation of NF-kB and subsequent production of secreted embryonic alkaline phosphatase (SEAP) in a dose-dependent manner (Figure 4.5). SEAP levels are detected using a colorimetric substrate assay, QUANTI-Blue (Invivogen, San Diego, CA) by measuring an increase in absorbance at OD655.

Culture supernatant from inflammasome-reconstituted HEK293T cells or HEK293T CARD8 -/- cells that had been transfected with 3CL pro was added to HEK-Blue IL-1β reporter cells plated in 96-well format in a total volume of 200 µl per well. On the same plate, serial dilutions of recombinant human IL-1β (Invivogen, San Diego, CA) were added in order to generate a standard curve for each assay. Twenty-four hours later, SEAP levels were assayed by taking 20 µl of the supernatant from HEK-Blue IL-1β reporter cells and adding to 180 µl of QUANTI-Blue colorimetric substrate following

the manufacturer's protocol. After incubation at 37°C for 30–60 min, absorbance at OD655 was measured on a BioTek Cytation five plate reader (BioTek Instruments, Winooski, VT) and absolute levels of IL-1β were calculated relative to the standard curve. All assays, beginning with independent transfections or infections, were performed in triplicate.

Immunoblotting and antibodies

Harvested cell pellets were washed with 1X PBS, and lysed with 1x NuPAGE LDS sample buffer containing 5% β-mercaptoethanol at 98C for 10 min. The lysed samples were spun down at 15000 RPM for two minutes, followed by loading into a 4–12% Bis-Tris SDS-PAGE gel (Life Technologies, San Diego, CA) with 1X MOPS buffer (Life Technologies, San Diego, CA) and wet transfer onto a nitrocellulose membrane (Life Technologies, San Diego, CA). Membranes were blocked with PBS-T containing 5% bovine serum albumin (BSA) (Spectrum, New Brunswick, NJ), followed by incubation with primary antibodies for V5 (IL-1β), FLAG (mCherry-fused NLRP1 for protease assays), Myc (NLRP1-Myc for activation assays), HA (viral protease or mouse NLRP1B), β-tubulin, or GAPDH. Membranes were rinsed three times in PBS-T then incubated with the appropriate HRP-conjugated secondary antibodies. Membranes were rinsed again three times in PBS-T and developed with SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific, Carlsbad, CA). The specifications, source, and clone info for antibodies are described in Supplementary file 3.6.

Chapter 4 in part is currently being prepared for submission for publication of the material, including co-authors Rimjhim Agarwal, Elizabeth J. Fay, Christopher Beierschmitt, Andrew P. Ryan, Patrick S. Mitchell, Matthew D. Daugherty. I, Brian V. Tsu, am the first author of this paper.

References

Alosaimi, B., Hamed, M. E., Naeem, A., Alsharef, A. A., AlQahtani, S. Y., AlDosari, K. M., Alamri, A. A., Al-Eisa, K., Khojah, T., Assiri, A. M., & Enani, M. A. (2020). MERS-CoV infection is associated with downregulation of genes encoding Th1 and Th2 cytokines/chemokines and elevated inflammatory innate immune response in the lower respiratory tract. *Cytokine, 126*, 154895. doi:10.1016/j.cyto.2019.154895

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Res, 43*(W1), W39-49. doi:10.1093/nar/gkv416

Broz, P., & Dixit, V. M. (2016). Inflammasomes: mechanism of assembly, regulation and signalling. *Nat Rev Immunol, 16*(7), 407-420. doi:10.1038/nri.2016.58

Chavarria-Smith, J., Mitchell, P. S., Ho, A. M., Daugherty, M. D., & Vance, R. E. (2016). Functional and Evolutionary Analyses Identify Proteolysis as a General Mechanism for NLRP1 Inflammasome Activation. *PLoS Pathog, 12*(12), e1006052. doi:10.1371/journal.ppat.1006052

Chen, G., Wu, D., Guo, W., Cao, Y., Huang, D., Wang, H., Wang, T., Zhang, X., Chen, H., Yu, H., Zhang, X., Zhang, M., Wu, S., Song, J., Chen, T., Han, M., Li, S., Luo, X., Zhao, J., & Ning, Q. (2020). Clinical and immunological features of severe and moderate coronavirus disease 2019. *J Clin Invest, 130*(5), 2620-2629. doi:10.1172/JCI137244

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X., & Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet, 395*(10223), 507-513. doi:10.1016/S0140-6736(20)30211-7

Chen, S., Tian, J., Li, Z., Kang, H., Zhang, J., Huang, J., Yin, H., Hu, X., & Qu, L. (2019). Feline Infectious Peritonitis Virus Nsp5 Inhibits Type I Interferon Production by Cleaving NEMO at Multiple Sites. *Viruses, 12*(1). doi:10.3390/v12010043

D'Osualdo, A., Weichenberger, C. X., Wagner, R. N., Godzik, A., Wooley, J., & Reed, J. C. (2011). CARD8 and NLRP1 undergo autoproteolytic processing through a ZU5-like domain. *PLoS One, 6*(11), e27396. doi:10.1371/journal.pone.0027396

Evavold, C. L., & Kagan, J. C. (2019). Inflammasomes: Threat-Assessment Organelles of the Innate Immune System. *Immunity, 51*(4), 609-624. doi:10.1016/j.immuni.2019.08.005

Finger, J. N., Lich, J. D., Dare, L. C., Cook, M. N., Brown, K. K., Duraiswami, C., Bertin, J., & Gough, P. J. (2012). Autolytic proteolysis within the function to find domain (FIIND) is required for NLRP1 inflammasome activity. *J Biol Chem, 287*(30), 25030-25037. doi:10.1074/jbc.M112.378323

Frew, B. C., Joag, V. R., & Mogridge, J. (2012). Proteolytic processing of Nlrp1b is required for inflammasome activity. *PLoS Pathog, 8*(4), e1002659. doi:10.1371/journal.ppat.1002659

Giamarellos-Bourboulis, E. J., Netea, M. G., Rovina, N., Akinosoglou, K., Antoniadou, A., Antonakos, N., Damoraki, G., Gkavogianni, T., Adami, M. E., Katsaounou, P., Ntaganou, M., Kyriakopoulou, M., Dimopoulos, G., Koutsodimitropoulos, I., Velissaris, D., Koufargyris, P., Karageorgos, A., Katrini, K., Lekakis, V., Lupse, M., Kotsaki, A., Renieris, G., Theodoulou, D., Panou, V., Koukaki, E., Koulouris, N., Gogos, C., & Koutsoukou, A. (2020). Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. *Cell Host Microbe, 27*(6), 992-1000 e1003. doi:10.1016/j.chom.2020.04.009

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics, 27*(7), 1017-1018. doi:10.1093/bioinformatics/btr064

Johnson, D. C., Taabazuing, C. Y., Okondo, M. C., Chui, A. J., Rao, S. D., Brown, F. C., Reed, C., Peguero, E., de Stanchina, E., Kentsis, A., & Bachovchin, D. A. (2018). DPP8/DPP9 inhibitor-induced pyroptosis for treatment of acute myeloid leukemia. *Nat Med, 24*(8), 1151-1156. doi:10.1038/s41591-018-0082-y

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., Walters, R. K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J. X., Samocha, K. E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A. H., Minikel, E. V., Weisburd, B., Lek, M., Ware, J. S., Vittal, C., Armean, I. M., Bergelson, L., Cibulskis, K., Connolly, K. M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M. E., Genome Aggregation Database, C., Neale, B. M., Daly, M. J., & MacArthur, D. G.

(2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature, 581*(7809), 434-443. doi:10.1038/s41586-020-2308-7

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol, 30*(4), 772-780. doi:10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649. doi:10.1093/bioinformatics/bts199

Latinne, A., Hu, B., Olival, K. J., Zhu, G., Zhang, L., Li, H., Chmura, A. A., Field, H. E., Zambrana-Torrelio, C., Epstein, J. H., Li, B., Zhang, W., Wang, L. F., Shi, Z. L., & Daszak, P. (2020). Origin and cross-species transmission of bat coronaviruses in China. *bioRxiv*. doi:10.1101/2020.05.31.116061

Lucas, C., Wong, P., Klein, J., Castro, T. B. R., Silva, J., Sundaram, M., Ellingson, M. K., Mao, T., Oh, J. E., Israelow, B., Takahashi, T., Tokuyama, M., Lu, P., Venkataraman, A., Park, A., Mohanty, S., Wang, H., Wyllie, A. L., Vogels, C. B. F., Earnest, R., Lapidus, S., Ott, I. M., Moore, A. J., Muenker, M. C., Fournier, J. B., Campbell, M., Odio, C. D., Casanovas-Massana, A., Yale, I. T., Herbst, R., Shaw, A. C., Medzhitov, R., Schulz, W. L., Grubaugh, N. D., Dela Cruz, C., Farhadian, S., Ko, A. I., Omer, S. B., & Iwasaki, A. (2020). Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature, 584*(7821), 463-469. doi:10.1038/s41586-020-2588-y

Mathew, D., Giles, J. R., Baxter, A. E., Oldridge, D. A., Greenplate, A. R., Wu, J. E., Alanio, C., Kuri-Cervantes, L., Pampena, M. B., D'Andrea, K., Manne, S., Chen, Z., Huang, Y. J., Reilly, J. P., Weisman, A. R., Ittner, C. A. G., Kuthuru, O., Dougherty, J., Nzingha, K., Han, N., Kim, J., Pattekar, A., Goodwin, E. C., Anderson, E. M., Weirick, M. E., Gouma, S., Arevalo, C. P., Bolton, M. J., Chen, F., Lacey, S. F., Ramage, H., Cherry, S., Hensley, S. E., Apostolidis, S. A., Huang, A. C., Vella, L. A., Unit, U. P. C. P., Betts, M. R., Meyer, N. J., & Wherry, E. J. (2020). Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science, 369*(6508). doi:10.1126/science.abc8511

Moustaqil, M., Ollivier, E., Chiu, H. P., Van Tol, S., Rudolffi-Soto, P., Stevens, C., Bhumkar, A., Hunter, D. J. B., Freiberg, A. N., Jacques, D., Lee, B., Sierecki, E., & Gambin, Y. (2021). SARS-CoV-2 proteases PLpro and 3CLpro cleave IRF3 and critical modulators of inflammatory pathways (NLRP12 and TAB1): implications for disease presentation across species. *Emerg Microbes Infect, 10*(1), 178-195. doi:10.1080/22221751.2020.1870414

Pickett, B. E., Greer, D. S., Zhang, Y., Stewart, L., Zhou, L., Sun, G., Gu, Z., Kumar, S., Zaremba, S., Larsen, C. N., Jen, W., Klem, E. B., & Scheuermann, R. H. (2012). Virus pathogen database and analysis resource (ViPR): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. *Viruses, 4*(11), 3209-3226. doi:10.3390/v4113209

Rathinam, V. A., & Fitzgerald, K. A. (2016). Inflammasome Complexes: Emerging Mechanisms and Effector Functions. *Cell, 165*(4), 792-800. doi:10.1016/j.cell.2016.03.046

Rathinam, V. A., Vanaja, S. K., & Fitzgerald, K. A. (2012). Regulation of inflammasome signaling. *Nat Immunol, 13*(4), 333-342. doi:10.1038/ni.2237

Robinson, K. S., Teo, D. E. T., Tan, K. S., Toh, G. A., Ong, H. H., Lim, C. K., Lay, K., Au, B. V., Lew, T. S., Chu, J. J. H., Chow, V. T. K., Wang, Y., Zhong, F. L., & Reversade, B. (2020). Enteroviral 3C protease activates the human NLRP1 inflammasome in airway epithelia. *Science, 370*(6521). doi:10.1126/science.aay2002

Roe, M. K., Junod, N. A., Young, A. R., Beachboard, D. C., & Stobart, C. C. (2021). Targeting novel structural and functional features of coronavirus protease nsp5 (3CL(pro), M(pro)) in the age of COVID-19. *J Gen Virol, 102*(3). doi:10.1099/jgv.0.001558

Tsu, B. V., Beierschmitt, C., Ryan, A. P., Agarwal, R., Mitchell, P. S., & Daugherty, M. D. (2021). Diverse viral proteases activate the NLRP1 inflammasome. *Elife, 10*. doi:10.7554/eLife.60609

V'Kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol, 19*(3), 155-170. doi:10.1038/s41579-020-00468-6

Wang, Q., Gao, H., Clark, K. M., Mugisha, C. S., Davis, K., Tang, J. P., Harlan, G. H., DeSelm, C. J., Presti, R. M., Kutluay, S. B., & Shan, L. (2021). CARD8 is an inflammasome sensor for HIV-1 protease activity. *Science, 371*(6535). doi:10.1126/science.abe1707

Zhou, Z., Ren, L., Zhang, L., Zhong, J., Xiao, Y., Jia, Z., Guo, L., Yang, J., Wang, C., Jiang, S., Yang, D., Zhang, G., Li, H., Chen, F., Xu, Y., Chen, M., Gao, Z., Yang, J., Dong, J., Liu, B., Zhang, X., Wang, W., He, K., Jin, Q., Li, M., & Wang, J. (2020). Heightened Innate Immune Responses in the Respiratory Tract of COVID-19 Patients. *Cell Host Microbe, 27*(6), 883-890 e882. doi:10.1016/j.chom.2020.04.017

Zhu, X., Fang, L., Wang, D., Yang, Y., Chen, J., Ye, X., Foda, M. F., & Xiao, S. (2017). Porcine deltacoronavirus nsp5 inhibits interferon-beta production through the cleavage of NEMO. *Virology, 502*, 33-38. doi:10.1016/j.virol.2016.12.005

Zhu, X., Wang, D., Zhou, J., Pan, T., Chen, J., Yang, Y., Lv, M., Ye, X., Peng, G., Fang, L., & Xiao, S. (2017). Porcine Deltacoronavirus nsp5 Antagonizes Type I Interferon Signaling by Cleaving STAT2. *J Virol, 91*(10). doi:10.1128/JVI.00003-17

Chapter 5: Summary and future directions

Concluding perspectives

Analyses described in Chapter 1 and 2 demonstrate that gene targets engaging in host-viral arms races tend to be rapidly evolving across the whole gene. In studying the role of two rapidly evolving genes, NLRP1 and CARD8, as novel host innate immunity sensors for viral protease activity, I showed that these evolutionary signatures of host-virus arms races can be used to predict which host genes are in genetic conflicts with viral protease antagonism and how the functional consequences of those genetic conflicts have been altered throughout host and viral evolution. In addition to highlighting these inflammatory pathways for further study, these findings underscore the importance of examining viral protease-host interactions through an evolutionary lens more broadly across whole host genomic datasets, across diverse lineages of host organisms and across diverse viral pathogens.

Characterizing other pathogenic protease ETIs in mammals

Although we had expanded on the roles of NLRP1 and CARD8, there are still many avenues of study even among these two proteins. Our work focused on the canonical isoforms, or the longest isoforms, encoded by each gene. In some instances, we may not see an effect of inflammasome activation by proteases due differences in regions encoded by different exons. As we learn more about tissue-specific expression of isoforms, knowledge of whether these other isoforms encode other viral protease cleavage sites will help us better understand how viruses cause pathology to different tissues.

As discussed in Chapter 1, the host ETI response includes examples of protease triggered ETIs tied to cell death pathways beyond NLRP1 and CARD8. One approach

to uncovering new tripwires for viral proteases is to examine tripwires for other microbial proteases. The discovery that human NLRP1 senses viral proteases was guided by earlier work demonstrating that the mouse homolog NLRP1B senses the lethal factor protease of the bacterial pathogen, Bacillus anthracis. Two additional studies demonstrate that pro-IL-1B also exists as a tripwire downstream of the inflammasome to sense the protease activity of pathogenic bacteria (LaRock et al., 2016; Sun et al., 2020). Interestingly, many of these cell death pathways become functionally active through the activity of host proteases such as caspases (Kesavardhana, Malireddi, & Kanneganti, 2020). Another approach then, may be to comprehensively screen for adjacent or overlapping sites known to be cleaved by host proteases and predicted to be cleaved by viral proteases.

Expanding evolution-guided discovery of new host-viral protease targets

In Chapter 2, I discussed an approach to categorize evolutionary fingerprints across the whole genome. By focusing on rapidly evolving genes in the genome, we can saturate for genes that have a greater importance in arms races and use viral antagonism strategies such as protease cleavage to uncover novel mechanisms of innate immunity, as we did with NLRP1 and CARD8. In the process of pursing this work, I have identified three new avenues for future genome-wide evolution-guided discovery of protease targets.

First, rather than focusing on rapid evolution across the whole gene, we can identify individual rapidly evolving sites that directly overlap with predicted sites of cleavage (Murrell et al., 2013; Yang, 2007). Regardless of whether the host genes are

more highly conserved or rapidly evolving across the whole gene, presence of this site could suggest that proteases are directly driving the evolution of the site to change. These individual site signatures may hold greater importance within host genes that are more conserved, suggestive of greater evolutionary constraint to maintain critical biological functions.

Second, while canonical isoforms were the focus of this earlier work, use of the entire dataset of isoforms would allow us to investigate whether certain isoforms encoded by a single gene may be more involved in arms races than others. Despite canonical isoforms being the longest isoform encoded by the gene, the encoded amino acids gained or lost in alternative exon usage may result in changes in protein folding and, ultimately, changes in the protein function.

Third, nonsynonymous allelic variation across genes in the human population may yield additional evolutionary insights. By analyzing the relationship between human population diversity and broader primate diversity, we can begin to highlight three new categories of genes in the context of susceptibility to pathogens: 1) genes with low primate diversity, but are highly polymorphic in the human population (human population-specific role), 2) genes with high primate diversity and are highly polymorphic in the human population (continued utilization and importance in arms races), 3) genes with high primate diversity, but are lowly polymorphic (de-emphasis of arms race role in humans or greater population risk to pathogens that can take

advantage of the low diversity within arms race-relevant genes).

Predicting new host targets for other families of (+)ssRNA viruses

Although I have focused on computationally predicting enteroviral and coronaviral protease cleavage in the previous chapters, related work [Ref] and earlier work in the lab demonstrates that this approach is broadly applicable to other (+)ssRNA viruses including the families *Flaviviridae*, other genera of *Picornaviridae*, *Arteriviridae*, *Caliciviridae*, and *Togaviridae*. These other families of (+)ssRNA viruses also encode proteases that serve roles in polyprotein processing and antagonism of host innate immune factors. As such, cleavage motifs can be generated from the collection of publicly available polyprotein sequences. By continuing to overlap these cleavage signatures with rapidly evolving host targets, we can further elucidate how viral proteases have shaped the primate genome, including the discovery of lineage-specific or broad-spanning antiviral pathways.

References

Kesavardhana, S., Malireddi, R. K. S., & Kanneganti, T. D. (2020). Caspases in Cell Death, Inflammation, and Pyroptosis. *Annu Rev Immunol, 38*, 567-595. doi:10.1146/annurev-immunol-073119-095439

LaRock, C. N., Todd, J., LaRock, D. L., Olson, J., O'Donoghue, A. J., Robertson, A. A., Cooper, M. A., Hoffman, H. M., & Nizet, V. (2016). IL-1beta is an innate immune sensor of microbial proteolysis. *Sci Immunol, 1*(2). doi:10.1126/sciimmunol.aah3539

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol, 30*(5), 1196-1205. doi:10.1093/molbev/mst030

Sun, J., LaRock, D. L., Skowronski, E. A., Kimmey, J. M., Olson, J., Jiang, Z., O'Donoghue, A. J., Nizet, V., & LaRock, C. N. (2020). The Pseudomonas

aeruginosa protease LasB directly activates IL-1beta. *EBioMedicine, 60*, 102984. doi:10.1016/j.ebiom.2020.102984

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol, 24*(8), 1586-1591. doi:10.1093/molbev/msm088