

Review

Expanding Repertoire of Plant Positive-Strand RNA Virus Proteases

Krin S. Mann [†] and H el ene Sanfa on ^{*,*} 

Summerland Research and Development Centre, Agriculture and Agri-Food Canada, Summerland, BC V0H 1Z0, Canada; mann.krin@gmail.com

* Correspondence: Helene.Sanfacon@canada.ca; Tel.: +1-250-494-6393

[†] Current Address: University of British Columbia Okanagan, Kelowna, BC V1V 1V7, Canada.

Received: 21 December 2018; Accepted: 12 January 2019; Published: 15 January 2019



Abstract: Many plant viruses express their proteins through a polyprotein strategy, requiring the acquisition of protease domains to regulate the release of functional mature proteins and/or intermediate polyproteins. Positive-strand RNA viruses constitute the vast majority of plant viruses and they are diverse in their genomic organization and protein expression strategies. Until recently, proteases encoded by positive-strand RNA viruses were described as belonging to two categories: (1) chymotrypsin-like cysteine and serine proteases and (2) papain-like cysteine protease. However, the functional characterization of plant virus cysteine and serine proteases has highlighted their diversity in terms of biological activities, cleavage site specificities, regulatory mechanisms, and three-dimensional structures. The recent discovery of a plant picorna-like virus glutamic protease with possible structural similarities with fungal and bacterial glutamic proteases also revealed new unexpected sources of protease domains. We discuss the variety of plant positive-strand RNA virus protease domains. We also highlight possible evolution scenarios of these viral proteases, including evidence for the exchange of protease domains amongst unrelated viruses.

Keywords: proteolytic processing; viral proteases; protease specificity; protease structure; virus evolution

1. Introduction

Eukaryotic RNA viruses have a long evolution history, which is driven by their necessary adaptation to their hosts [1]. Viruses have likely evolved from capsid-less genetic parasites that later acquired various types of capsid proteins [2,3]. Thus, many viral replication proteins, notably RNA-dependent RNA polymerases (RdRps), helicases, and some proteases, have ancient origins, traced back to eukaryogenesis or even earlier [1,4]. Virus evolution has been described as modular [5]. Indeed, phylogenetic analyses of hallmark genes (capsid, RdRp, helicase) have revealed multiple examples of protein domain exchanges between viruses of different types. In addition, horizontal virus transfer between lower eukaryotes, invertebrates, plants, and vertebrates have also contributed to the evolution of eukaryotic viruses [6]. Plant viruses exemplify the modular evolution of RNA viruses. Progressive protein domain acquisitions by plant viruses can be partly attributed to the mixed infections that are typically observed in plants [7,8]. Most plant viruses depend on insect, nematode, or fungal vectors for plant-to-plant transmission, and in some cases, also replicate in these vectors [9–11]. Thus, selection pressure from both plant and insect hosts along with the heterogeneous population structure of RNA viruses have likely played a major role in the gain and/or loss of protein domains [11–13].

Positive-strand [(+)-strand] RNA viruses have a constrained genome size and have adapted to this limitation by encoding multifunctional proteins and/or by expressing different forms of their proteins using sophisticated expression strategies [14,15]. One strategy that is employed by many (+)-strand

RNA viruses is to express large polyproteins that are subsequently processed into smaller functional gene products by viral and/or cellular proteases [16–21]. Polyprotein processing allows for the controlled and timely release of mature functional gene products or partially processed intermediate polypeptides that can differ in their biological activities. Thus, viruses have also developed various mechanisms to regulate the activity of viral proteases and/or the efficiency of cleavage of polyproteins at specific sites.

Positive-strand RNA viruses have been divided into two large groups, alpha-like or picorna-like based on the phylogeny of their RdRps [1]. Picorna-like viruses typically encode a main protease that cleaves at multiple sites in the large polyproteins [22,23]. They may also encode additional accessory proteases. In contrast, alpha-like viruses do not always encode a protease, and when they do, they are often leader proteases, auto-catalytically cleaving the polyprotein at a single site to promote their own release from the polyprotein [24,25].

Viral proteases are often multi-functional, playing key roles in multiple steps of the infection cycle. Indeed, they have been described to counteract antiviral defense responses by suppressing RNA silencing or having deubiquitinase activity, facilitate symplastic and systemic long distance movement, aid in viral replication and virion maturity, and promote insect host transmission and retention. For an in-depth discussion of the multifunctional activities of plant virus proteases, we refer the readers to a recent review [16]. In many cases, the diverse biological functions of viral proteases are directly related to their proteolytic activity. For example, the controlled cleavage of polyproteins encoding viral coat proteins, replication proteins or suppressors of silencing regulates virion maturation, viral RNA replication, and viral counter-defense responses. In addition, many animal and human virus proteases do not only cleave the viral polyproteins but also host proteins to facilitate the various steps of the virus infection cycle (translation, replication, suppression of host defense responses) [17,18]. In other cases, viral proteases have additional functions that are separate from their proteolytic activities. For example, different domains of the plant potyvirus HC-Pro protease have been implicated in proteolytic cleavage, suppression of antiviral RNA silencing, and aphid transmission [26]. Finally, the multifunctional activities of proteases can also overlap in their three-dimensional structure (3D structure), as shown for the dual protease-deubiquitinase of turnip yellow mosaic virus [27]. As will be described below, the multifunctional activities of viral proteases have influenced their evolution, in some cases resulting in the adoption of simplified folds for their proteolytic domains as compared to their cellular counterparts.

In this review, we discuss the diversity, evolution, and possible origin of various protease domains found in plant viruses. We also provide examples of possible exchange of protease domains among unrelated plant viruses. Please note that we did not attempt to list every protease signature motifs that were identified in virus genome annotations, rather we have focused the text on well-characterized proteases with known activities, specificities, and/or structures. Throughout the text, we have adopted the terminology of the MEROPS database (available at www.ebi.ac.uk/merops) for protease clans and families [28]. For virus taxonomy (genera, family, order), we use the 2018 taxonomy release of the International Committee for the Taxonomy of Viruses (available at <https://talk.ictvonline.org/taxonomy/>) [29].

2. Functional Protease Types

Proteases are classified in clans and families, which are based on their catalytic types (serine, threonine, cysteine, aspartic, glutamic, asparagine, and metallo), phylogenies, and molecular structures [28]. The catalytic mechanisms are best understood for the prevalent serine, cysteine, aspartic, and metalloproteases. Serine and cysteine proteases use the hydroxyl group (serine, threonine) or sulfhydryl group (cysteine) of the catalytic residue side chain as the catalytic nucleophile and a histidine as a base residue. Metallo and aspartic proteases require a water molecule as a nucleophile. Aspartic and glutamic proteases are considered acid proteases and normally function optimally at low

pH. The vast majority of plant cellular proteases are cysteine, serine, aspartic, and metalloproteases, many of which have been implicated in plant defense responses [30,31].

Until recently, proteases that are encoded by plant (+)-strand RNA viruses were reported to belong to two clans: chymotrypsin-like cysteine or serine proteases (clan PA) and papain-like cysteine proteases (clan CA) [16,28]. In addition, plant reverse-transcribing viruses encode pepsin-like aspartic proteases (clan AA). As will be described below, the structural and functional characterizations of cysteine and serine proteases from various (+)-strand RNA viruses have led to new insights in their diversity and specialization. Many plant virus proteases have ancient origins, for example, the chymotrypsin-like cysteine or serine proteases of picorna-like viruses [4]. Although related in structures to their cellular counterparts, these viral proteases show little sequence identities other than a few conserved catalytic residues and they display more stringent cleavage site specificities [32–34]. In contrast, other plant virus proteases are likely to be more recent acquisitions, such as the newly described glutamic protease of strawberry mottle virus [35].

3. Chymotrypsin-Like Serine and Cysteine Proteases

Many plant (+)-strand RNA virus proteases are cysteine or serine proteases that share structural homologies with the cellular chymotrypsin serine proteases. These chymotrypsin-like proteases are encoded by viruses of the “picorna-like supergroup”, which, although infecting a variety of eukaryotic hosts, are grouped together based on the common origin of their RdRps [1,4,36]. Notable chymotrypsin-like viral proteases include the archetype 3C cysteine proteases (3C-Pro) of animal and human viruses in the family *Picornaviridae* [36,37] and the related 3C-like proteases (3CL-Pros) of plant viruses in the families *Secoviridae* [38] and *Potyviridae* [39] (Figure 1A). Serine proteases of plant viruses in the family *Solemoviridae*, and in the genera *Polerovirus* and *Enamovirus* also adopt a chymotrypsin-like fold. Relationships between these viral proteases and the cellular chymotrypsin serine protease were identified more than 30 years ago based on sequence alignments and mutagenesis studies [40–45], and these were confirmed almost a decade later with the elucidation of the crystal structures of several 3C-Pros [46–49]. Viral chymotrypsin-like proteases are very diverse. This is likely due to their ancient origin. Indeed, phylogenetic relationships with bacterial serine membrane proteases led to the suggestion that a chymotrypsin-like protease was acquired from a bacterial endosymbiont by an ancestral picorna-like virus during eukaryogenesis [4].

The structure of chymotrypsin is characterized by a double-barrel fold, which is shared by the viral chymotrypsin-like proteases (Figure 1B). Chymotrypsin activity is dependent on a catalytic triad consisting of His⁵⁷, Asp¹⁰², and the nucleophile Ser¹⁹⁵, which are brought together in the 3D structure. The base His residue is highly conserved in all viral chymotrypsin-like proteases. The nucleophile Ser residue is conserved in the viral serine proteases, but it is replaced by Cys in the 3C- and 3CL-Pros [36,37]. The Asp residue is substituted by Glu in some viral proteases [36,50]. In contrast to the relaxed specificity of chymotrypsin, viral chymotrypsin-like proteases have stringent cleavage site specificities, which are determined by the shape and depth of their substrate-binding pocket (SBP) [36,37,51]. A conserved His in the S1 position of the SBP of most chymotrypsin-like viral proteases is responsible for the recognition of cleavage sites with a Gln (or Glu) at the P1 position. Other residues in the P6 to P2' positions of the cleavage sites also contribute to the protease specificity. Finally, the folding of the polyprotein substrate determines the cleavage site accessibility to the protease [52,53].

3.1. The Archetype Picornavirus 3C-Pro

Proteolytic processing was shown to be a prerequisite for the formation of mature proteins of human and animal picornaviruses in the late 1960s [54]. The protease activity was assigned to the 3C protein ten years later [55–57]. In the case of the 3C-Pro of poliovirus (PV, species *Human enterovirus C*, genus *Enterovirus*), the catalytic triad consists of His⁴⁰, Glu⁷¹, and Cys¹⁴⁷. The spacing of the residues forming the catalytic triad is tighter than that of chymotrypsin, and this was attributed to deletions in exposed surface loops. The overall 3D structure of picornavirus 3C-Pro is otherwise similar to that of chymotrypsin [47,48] (Figure 1B). Although other proteases are encoded by picornaviruses, the 3C-Pro is responsible for the majority of the cleavage events in the viral polyproteins [36,58] (Figure 1A). The picornavirus 3C-Pro is active in the context of large polyproteins and cleave at several sites with different efficiencies. Primary cleavage occurs rapidly, allowing for the accumulation of intermediate polyproteins, which are then processed sequentially at secondary cleavage sites [18]. In addition to their essential role in viral polyprotein processing, the picornavirus 3C-Pro binds the viral RNA and enhance viral RNA replication [18], a property that is shared by the 3CL-Pro of the plant potyviruses [59] and possibly other plant picorna-like viruses. RNA binding is facilitated by conserved residues that are located on a surface opposite of the catalytic site in the 3D structure [47,48]. Picornavirus 3C-Pro also facilitate viral infection by cleaving various host proteins, notably translation and transcription factors, nuclear pore proteins, and proteins implicated in the innate immune response [17–19]. Whether similar host protein cleavage is also directed by the plant virus 3CL-Pro has so far not been investigated.

3.2. The 3CL-Pro of Viruses in the Family Secoviridae and Their Diverse Cleavage Site Specificities

The family *Secoviridae* is a large family of plant viruses that mostly infects horticultural crops [38]. Viruses in the family *Secoviridae* (secovirids) share capsid structures and signature replication proteins with picornaviruses and they are included in the order *Picornavirales* [22,23,38,60]. Secovirids can have monopartite or bipartite genome and they encode large polyproteins (Figure 1A). Until very recently, secovirids were only known to encode a single protease, the 3CL-Pro, which processes the viral polyproteins *in cis* and *in trans*. All secovirid 3CL-Pro share the conserved catalytic triad of the picornavirus 3C-Pro. Most have the signature replacement of the chymotrypsin nucleophile Ser by Cys, with the notable exception of the 3CL-Pro of blueberry latent spherical virus (genus *Nepovirus*), which has retained the Ser [61].

The cowpea mosaic virus (CPMV, genus *Comovirus*) 3CL-Pro, often referred to as the 24K protease, was the first secovirid 3CL-Pro identified [62]. Other well-characterized secovirid 3CL-Pro include those of grapevine fanleaf virus (GFLV) and tomato ringspot virus (ToRSV), two members of the genus *Nepovirus* [63,64] (Figure 1A). Comoviruses and nepoviruses have a bipartite genome. Each RNA encodes a single large polyprotein, with the protease domain being contained in the RNA1 polyprotein. Cleavage of comovirus and nepovirus polyproteins is sequential, leading to the accumulation of multiple intermediates and mature proteins [38,65,66]. The proteolytic activity of the comovirus and nepovirus 3CL-Pro is regulated by the presence of other viral protein domains. In the case of CPMV, a viral co-factor protein (the 32K protein, also encoded by RNA1) slows down the *cis*-cleavage of the RNA1 polyprotein but it enhances the *trans*-cleavage of the RNA2 polyprotein [67]. This allows the accumulation of several intermediate RNA1 polyproteins containing two or more protein domains. Although a corresponding protein domain is present in nepovirus RNA1 polyproteins, it does not impact the activity of nepovirus 3CL-Pro [68,69]. Rather, the activity of these proteases is influenced by the presence of the genome-linked viral protein (VPg) domain on VPg-Pro intermediate polyproteins. Cleavage at nepovirus VPg-Pro cleavage sites is inefficient, leading to the accumulation of the VPg-Pro or larger precursors, from which the mature Pro is slowly released [70,71]. The mature Pro of two nepoviruses (ToRSV and GFLV) is more active than the VPg-Pro in cleaving the RNA2 polyprotein *in trans* to release the coat proteins [71,72], adding a regulatory step to delay virus encapsidation.

The consensus sequence of CPMV cleavage sites is Ax(A,P)Q↓(S,G or M) and includes a strictly conserved Gln at the P1 position and preferred residues at the P1', P2, and P4 positions with the

down-arrow indicating the exact position of the cleavage. This consensus sequence is similar to that of the 3C-Pro of many picornaviruses, owing to the presence of the conserved His in the protease SBP [73]. Systematic mutagenesis of the ToRSV cleavage sites also showed a preference for *trans*-cleavage at sites conforming to the consensus sequence (C,V)Q↓(G,S) [74]. Interestingly, the specificities of the CPMV and ToRSV proteases were shown to be more relaxed for cleavage sites that were recognized *in cis* [53,74]. While many secovirid 3CL-Pro share a preference for (Q,E) ↓(S,G) cleavage sites, this is not universally observed in this family [38,65]. In fact, secovirid 3CL-Pro show surprising differences in their specificity. Some lack the conserved His in their SBP, which is replaced by Leu (nepoviruses of subgroup A and B, sequiviruses), Val (stocky prune virus), or Cys (sadvaviruses). These proteases recognize a variety of atypical cleavage sites with Arg, Lys, Gly, Cys, Thr, or Ala at the P1 position. Additionally, some proteases with the conserved His present in the SBP have relaxed specificities recognizing D/S, N/S, or even C/S cleavage sites [38,65]. Elucidation of the crystal structures of secovirid 3CL-Pro would be required to understand how their SBPs accommodate these unusual cleavage site specificities. The evolutionary constraints that drove these diverging protease specificities are not well understood. For example, although many nepoviruses share similar infection cycles and host ranges, they differ widely in their protease specificity [65]. If nepovirus proteases target similar host proteins, then they must have adapted to cleave these proteins at different sites. This question should be a fruitful area of research in the future.

3.3. The well-characterized potyvirus NIa proteases

The family *Potyviridae* is the largest and most economically important family of plant (+)-strand RNA viruses [39]. Although often referred to as “picorna-like” viruses because of the signature RdRps and 3CL-Pro that they encode, viruses in the family *Potyviridae* (potyvirids) differ from members of the order *Picornavirales* by their capsid structure and by the nature of the helicase and VPg [22,39] (Figure 1A). The 3CL-Pro of potyvirids is referred to as the NIa protease, an abbreviation for nuclear inclusion protein a. Although potyvirids encode other proteases (Figure 1A), the NIa protease is responsible for most of the viral polyprotein cleavage events and it can act *in cis* or *in trans*, depending on the cleavage site [39,51]. The NIa protease of tobacco etch virus (TEV, genus *Potyvirus*) is well characterized. It has a stringent specificity, recognizing cleavage sites with the consensus sequence ExxYxQ↓(S or G). The requirement for a Gln at P1 position and for small residues at the P1' position is consistent with that of most other 3CL-Pro, while the preference for specific residues at the P3 and P6 position is more unique. The overall 3D structure of the TEV NIa protease was found to be similar to that of the 3C-Pro of human and animal picornaviruses [75] (Figure 1B). The specificity of the TEV NIa protease was confirmed by mutagenesis studies and later explained by the conformation of the SBP in the crystal structure [75–77]. The TEV NIa protease has been developed into a versatile biotechnology tool, facilitating the removal of fusion tags following affinity purification of recombinant proteins or protein complexes and allowing the simultaneous expression of multiple proteins in plant, mammalian, or bacterial cells [78–85].

Other well-characterized potyvirus NIa proteases include those of plum pox virus (PPV), tobacco vein mottling virus (TVMV), and turnip mosaic virus (TuMV) [86–90]. Like the TEV NIa protease, they recognize cleavage sites with a Gln in the P1 position and small residues at the P1' position. However, they require a Val at the P4 position, and either a His (PPV, TuMV) or Phe (TVMV) at the P2 position of the cleavage sites. Structure comparisons revealed differences in the SBPs of the TEV and TVMV NIa proteases that explain their contrasting specificities [91]. Most other NIa proteases have similar specificities, with the notable exception of the sweet potato mild mottle virus (SPMMV, genus *Ipomovirus*) NIa protease that lacks the conserved His in the SBP (replaced by Asn) and recognizes cleavage sites with a His at the P1 position [51]. Experiments that were designed to alter protease specificities by exchanging fragments or modifying specific amino acids in the SBP resulted in overall poor protease activities, revealing that long-distance interactions contribute to the proper folding of these highly adapted proteases [90,92].

The *trans*-processing activity of the NIa protease is impacted by its maturation stage. Indeed, the VPg-Pro intermediate has been shown to be more active than the mature Pro [93,94]. The cleavage site between the VPg and protease domains is deliberately sub-optimal (with Glu instead of Gln at the P1 position), allowing the accumulation of VPg-Pro intermediate polyproteins in infected plants [95]. The VPg is naturally unstructured [94,96]. The N-terminal 22 amino acids of the VPg are essential to maintain the disordered state of the protein and to enhance the processing activity of the protease [94]. Given that the VPg is also known to interact with a large number of host proteins [97,98], it is possible that its presence in VPg-Pro intermediates facilitates the cleavage of host proteins. However, this has not yet been investigated.

3.4. The Sobemovirus Serine Protease

A putative serine protease domain with homology to cellular and viral chymotrypsin-like proteases was first identified in the genome of southern bean mosaic virus (genus *Sobemovirus*, family *Solemoviridae*) [42]. The serine protease domain is conserved in all members of the family *Solemoviridae*, which also includes the genus *Polemovirus* [99]. A closely related serine protease is also found in the genome of viruses in the genera *Enamovirus* and *Polerovirus*, which are currently classified in the family *Luteoviridae* but also share the Pro-VPg-RdRp replication module with the family *Solemoviridae* [99]. The serine proteases from this group of viruses show the characteristic catalytic triad (His¹⁸¹, Asp²¹⁶, and Ser²⁸⁴ for sesbania mosaic virus, SeMV) [100–102]. Cleavage sites with the consensus sequence E↓(S,T) are recognized by the polerovirus and sobemovirus serine proteases [100,102,103], which is similar to the cleavage site specificity of most 3CL-ProS. Accordingly, the conserved His of the 3CL-Pro SBPs is also present in this group of serine proteases. The structure of the SeMV protease was solved, which confirmed a double-barrel domain fold that is typical of chymotrypsin-like proteases and the role of the conserved His in the SBP [104] (Figure 1B).

The regulation of SeMV proteolytic cleavage is interesting. Presence of the intrinsically unfolded VPg domain is strictly required for the SeMV protease activity [105]. In contrast to other picorna-like viruses, the VPg domain is C-terminal to the serine protease domain in the polyprotein (Figure 1A). Exposed aromatic residues in the protease domain interact with a tryptophane in the VPg domain to facilitate the interaction and the activation of the protease [104,106]. SeMV encodes two polyproteins 2a and 2ab, which share identical N-terminal domains (including the Pro domain and several cleavage sites) but differ in their C-terminal domains [99]. This is due to an inefficient frameshift in the RdRp domain, which allows the formation of the longer 2ab polyprotein (Figure 1A). Interestingly, the processing efficiency at cleavage sites that is shared by the two polyproteins differ considerably between polyprotein 2a or 2ab, suggesting that the presentation of the cleavage sites to the protease is influenced by the conformation of the polyproteins [102]. These results highlight the unique mechanisms regulating the activity of the SeMV serine protease.

3.5. The Compact and Diverse Potyvirus P1 Serine Proteases and Their Intricate Regulatory Mechanisms

The P1 protein is a minor protease that is present in some but not all members of the family *Potyviridae* [39,51] (Figure 1A). Located at the N-terminus of the polyprotein, it directs a single *cis*-cleavage to release itself from the polyprotein [107]. The protease activity was mapped to the C-terminal region of the P1 protein domain. His²¹⁴, Asp²²³, and Ser²⁵⁶ were identified as catalytic residues for the TEV P1 protease [108]. Although these residues are typical of serine protease catalytic triads, they are more closely spaced than any other viral cysteine or serine chymotrypsin-like proteases. Analysis of other potyvirus P1 proteases revealed similar tight arrangements of the putative catalytic triad, with only seven to nine residues separating the His and Asp (or Glu) residues [50]. The cleavage site consensus sequence is (V,I,L,M)xx(Y,F)↓S, which is conserved for all viruses in the family *Potyviridae* that have a P1 protease [51]. The protease structure has not yet been determined and the conformations of the catalytic site or of the SBP are not well understood.

Potyvirus P1 proteases are diverse, varying in their size and amino acid sequence [50]. Some viruses even have two copies of the P1 proteases, for example, cucumber vein yellowing virus (CVYV, genus *Ipomovirus*) [50,109,110] (Figure 1A). Most of the P1 diversity was attributed to the variable N-terminal region, which has been implicated in host specificity and virus virulence [111–113]. Complex schemes of gene duplication and recombination have been proposed to explain the evolution and diversity of P1 proteases [50]. P1 proteases have been divided into two main types based on sequences and regulatory mechanisms [50,110,114]. Type A includes most potyvirus P1 proteases and the first of the two CVYV P1 proteases (termed P1a). The second CVYV P1 protease (P1b) and the Ugandan cassava brown streak virus P1 protease are examples of type B. Mechanistically, the two types differ in that type B proteases are fully functional, while type A proteases depend on a plant factor for their activity [110,114]. Indeed, it was noted early that, while the activity of the full-length TEV P1 protease is easily detected in wheat germ extracts or upon expression in plants, the protease is not active in rabbit reticulocyte extracts unless a heat-labile plant factor present in wheat germ extracts is supplied to the reaction [108]. Later work demonstrated that the N-terminal region of P1 is antagonistic to the protease activity. This antagonism is relieved by binding to an as-yet unidentified plant factor in a host dependent manner or by the deletion of the N-terminal region [110,114]. The antagonistic regulation of type A P1 proteases has been linked to host range and viral virulence [112–115]. Strikingly, introducing deletions of the P1 N-terminal region in PPV infectious clones resulted in increased virulence and host range [112,114]. Efficient cleavage by the P1 and P1a proteases is necessary to activate the silencing suppression activity of downstream protein domains (HC-Pro and P1b, respectively), thereby counteracting plant antiviral defenses. This sophisticated protease regulatory mechanism demonstrates a remarkable level of adaptation of potyvirids to their hosts.

4. The Diverse “Papain-Like” Cysteine Proteases

The cellular papain cysteine protease is a globular protein with two interacting domains: an N-terminal helical domain (termed R domain) that includes the nucleophile Cys¹⁵⁸ residue and a C-terminal domain mostly composed of β -sheets (termed L domain) that encompasses the base His²⁹² residue, 134 amino acid downstream of the Cys¹⁵⁸ residue (Figure 2B) [116]. The two catalytic residues are brought in close proximity in the 3D structure. A third residue (Asn¹⁷⁵) forms a hydrogen bond with Cys¹⁵⁸. However, mutagenesis studies did not confirm a strict requirement of Asn¹⁷⁵ for proteolytic activity [117], and the catalytic site is generally considered to function as a dyad. Papain-like proteases represent a large and diverse group of proteases (clan CA) [118].

Possible relationships between papain and some (+)-strand RNA virus proteases were identified in the early 1990s and they were based on sequence alignments that confirmed the presence of conserved cysteine and histidine residues [41,119]. Elucidation of the structure of the leader protease (L-Pro) of foot-and-mouth disease virus (FMDV, genus *Aphthovirus*, family *Picornaviridae*) revealed a papain-like fold with Cys⁵¹, His¹⁴⁸, and Asp¹⁶³ (replacing the Asn of papain) forming the catalytic cleft [120] (Figure 2B). The FMDV L-Pro is an accessory protease. It is not conserved in all members of the family and it processes the polyprotein *in cis* at a single KLK↓GAG cleavage site to release itself (Figure 2A). In contrast, the nsP2 cysteine protease of Venezuelan equine encephalitis virus (VEEV, genus *Alphavirus*, family *Togaviridae*) is the main protease and it cleaves the nonstructural polyprotein at four sites (Figure 2A). The structure of the VEEV cysteine protease revealed major divergence from the papain fold [121] (Figure 2B). The C-terminal region of the nsP2 protein (which contains the protease domain) includes the catalytic protease domain and a C-terminal methyl-transferase like domain. The protease domain differs from a classical papain fold, in that it is mainly helical. The N-terminal region shares some similarities with the papain helical R domain, while a simplified L domain contains only two short β -sheets. The two catalytic residues Cys⁴⁷⁷ and His⁵⁴⁶ are more closely spaced than in papain, with only 69 residues separating them. However, they are brought together in the catalytic site to adopt a conformation that is similar to that of the papain fold. An equivalent to the papain Asn¹⁷⁵ residue was not found.

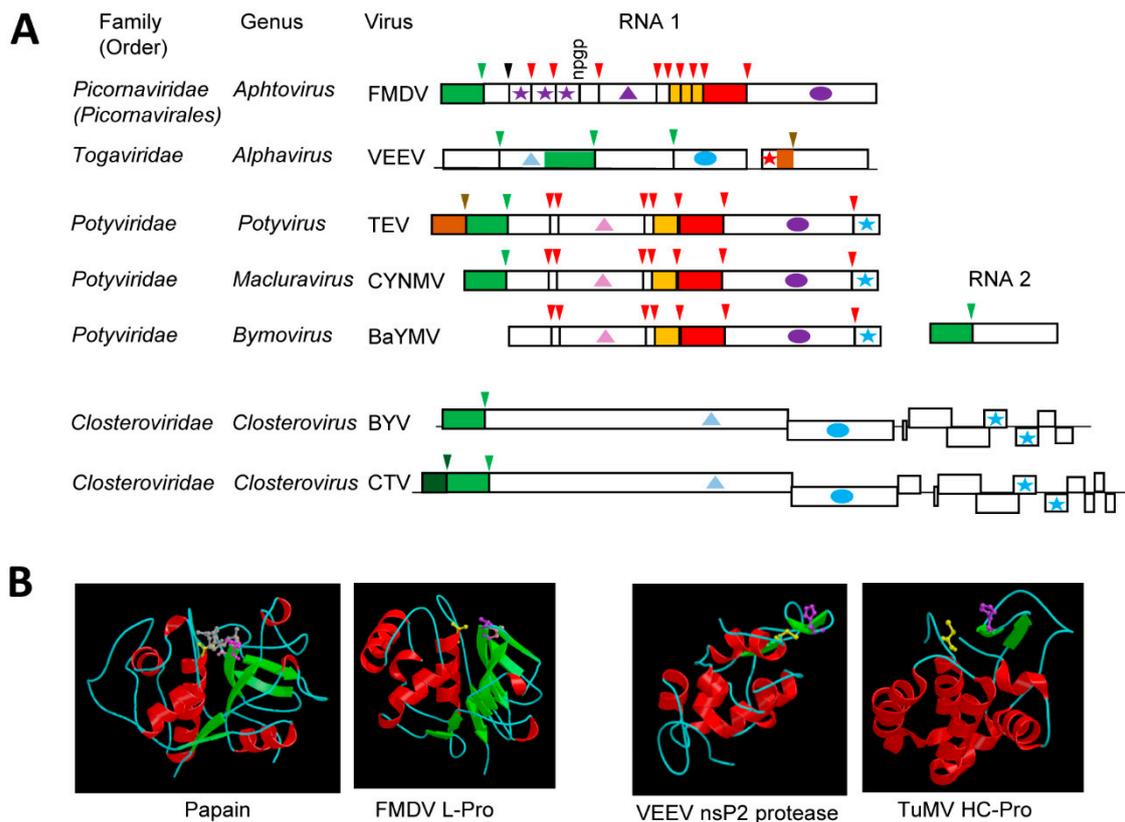


Figure 2. Papain-like cysteine proteases. **(A)** Genomic organization and proteolytic cleavages of representative viruses. Polyproteins are shown by boxes with cleavage sites indicated by vertical lines. Protease domains are shown in green for papain-like cysteine proteases, red for chymotrypsin-like 3CL cysteine proteases and brown for chymotrypsin-like serine proteases. The same color code is used for arrows above each cleavage site indicating the protease responsible for the cleavage. A black arrow in the foot-and-mouth disease virus (FMDV) polyprotein indicates an autocatalytic maturation cleavage event of the capsid protein. The “ngpg” sequence represent the 2A translational stop-go sequence of FMDV. Orange boxes indicate the VPg proteins. Ovals represent RdRp domains: purple for picorna-like RdRp and blue for alpha-like RdRp. Stars indicate the coat protein domains: purple for picorna-like type 1 jelly-roll domains icosahedral coat protein, red for unrelated icosahedral CP of togaviruses, blue for filamentous coat proteins. Helicase domains are shown by the triangles: light blue, pink and purple for helicase superfamilies 1, 2, and 3, respectively. Please note that cleavage sites processed only by viral proteases are shown. The VEEV cleavage sites processed by the cellular furin protease are not shown. **(B)** Structure of representative proteases. Catalytic residues are represented, as follows; Cellular protease: papain from *Carica papaya* (pdb: 1PE6, in complex with E-64 inhibitor, which is shown in grey): Cys¹⁵⁸ (yellow), His²⁹² (purple), Asn³⁰⁸ (pink), animal virus proteases: FMDV L-Pro (pdb: 1QMY_A: mutant protease with Cys⁵¹ mutated to Ala), Cys⁵¹ (yellow, mut. to Ala), His¹⁴⁸ (purple), Asp¹⁶³ (pink), and VEEV nsP2 protease (pdb: 2HWK, only the protease catalytic domain is shown): Cys⁴⁷⁷ (yellow), His⁵⁴⁶ (purple), and plant virus protease: TuMV HC-Pro (pdb: 3RNV, protease catalytic domain): Cys⁷⁰⁶ (yellow), His⁷⁷⁹ (purple). Images of protein structures are reprinted with permission from the MEROPS database (www.ebi.ac.uk/merops). FMDV: foot-and-mouth disease virus, VEEV: Venezuelan equine encephalitis virus, TEV: tobacco etch virus, CYNMV: Chinese yam necrotic mosaic virus, BaYMV: barley yellow mosaic virus, BYV: beet yellows virus, CTV: citrus tristeza virus, TuMV: turnip mosaic virus.

Not surprisingly, several plant (+)-strand RNA viruses also encode cysteine proteases. A putative papain-like protease was initially identified in the polyprotein of potyviruses [119]. Since then, other plant viruses were predicted to encode cysteine proteases, which have been referred to as papain-like

proteases. However, as will be discussed below, the structure of characterized plant virus cysteine proteases differ from that of papain in many ways and reflect the unique adaptations and/or diverse origins of these enzymes.

4.1. The Multifunctional Potyvirus HC-Pro Protease with a Minimalistic Papain-Like Fold

Like the P1 protease, the helper component protease (HC-Pro) of potyvirids is a minor protease that cleaves the polyprotein *in cis* at a single site [122,123] (Figure 2A). Also similar to the P1 protease, not all viruses in the family *Potyviridae* encode an HC-Pro domain [26,39]. The cleavage site consensus sequence, YxG↓G, is strictly conserved amongst the monopartite potyvirids [51]. Viruses in the genus *Bymovirus* have a bipartite genome and also encode a related cysteine protease, which is termed P2-1. The P2-1 protease is located in the N-terminal region of the RNA2 polyprotein and it cleaves at a single related cleavage site with the consensus sequence of (Y,F, G)xG↓(A, N, S) [51] (Figure 2A).

The multi-functional HC-Pro plays roles in vector transmission and in the suppression of antiviral plant defense responses, notably RNA silencing [26]. HC-Pro is a multi-domain protein, with the protease catalytic region present in the C-terminus of the protein [124]. While the presence of the protease domain is conserved in all potyvirids that encode the HC-Pro protein, other domains that are involved in RNA binding, silencing suppression, and vector transmission vary widely within the family [26]. Even though HC-Pro was dubbed a papain-like protease, it was noted early on that the spacing between the catalytic residues (Cys⁷⁰⁶ and His⁷⁷⁹ for TuMV) is shorter than that observed in papain or in the leader proteases of animal and human picornaviruses [119]. The crystal structure of the C-terminal protease domain of the TuMV HC-Pro protein confirmed a simplified papain-like fold with some similarities to that of the VEEV nsP2 protein, including the presence of two short β-sheets forming the corresponding papain R domain [125] (Figure 2B). As with the animal alphavirus nsP2 protein, the two catalytic residues of the plant potyvirus HC-Pro are brought together in a papain-like topology. Interestingly, the C-terminus of HC-Pro was found locked into the active site cleft, preventing further cleavage and providing an explanation for the strict *cis*-cleavage mechanism [125]. The structure of other regions of the protein has not been determined with confidence, but phylogenetic analyses of separate domains coupled with structure modelling suggested that the multi-functionality of the protein drove the co-evolution of its domains [126]. It is possible that these overlapping selection pressures contributed to the evolution of a minimalistic papain-like fold for the protease domain.

4.2. The single or Tandem Closterovirus Leader Proteases

Viruses in the family *Closteroviridae* have the longest (+)-strand RNA genome (between 15 and 20 kb) amongst plant viruses [25,127]. The replication proteins are expressed as a large polyprotein that is cleaved at a single site by the leader protease (L-Pro), which is located at the N-terminus of the polyprotein (Figure 2A). The protease of beet yellows virus (BYV, genus *Closterovirus*) was first characterized in 1994 [128]. It was shown to have limited sequence similarities with other viral papain-like proteases, notably the potyvirus HC-Pro, and to require catalytic residues Cys⁵⁰⁹ and His⁵⁶⁹ to cleave at the single G↓G cleavage site [128,129]. Similar to HC-Pro, L-Pro is a multifunctional and multi-domain protein that has been implicated in the regulation of virus accumulation, long-distance transport, and host adaptation [130,131]. Also similar to HC-Pro, the protease catalytic domain is contained in the highly conserved C-terminal region of L-Pro, while the more variable N-terminal region orchestrates the other biological activities. The structure of the protease has not been determined and it not known whether it adopts a simplified papain-like fold that is similar to that of the plant potyvirus HC-Pro and the animal alphavirus nsP2 protease.

It is interesting to note that some members of the family *Closteroviridae* encode two L-Pro, notably citrus tristeza virus (CTV) and grapevine leafroll-associated virus 2 (GLRaV-2). The tandem proteases probably arose by domain duplication [132–134]. In model herbaceous hosts, the first L-Pro copy of either GLRaV-2 or CTV is strictly required for achieving infection, while the second copy plays accessory roles [132,134]. In contrast, both copies are strictly required for infection in the natural

perennial hosts (grapevine and citrus, for GLRaV-2 and CTV, respectively). Cleavage after the first GLRaV-2 L-Pro is dispensable, as long as cleavage occurred at the second site to release the viral replication proteins [132]. It has been suggested that the acquisition of a second protease domain may assist in host range expansion [132,134].

4.3. The Tymovirus Cysteine Protease with a Compact Ovarian-Tumor (OTU) Domain-Like Fold Driven by Its Dual Function as a Protease and Deubiquitinase

A new superfamily of cysteine proteases was proposed in 2000 that shares the catalytic Cys and His residues with papain, but has sequence relationships closer to the Ovarian Tumor Domain (OTU) proteins than to papain [135]. Similar to papain, the OTU-like proteins include a conserved Asp or Asn residue to form a catalytic triad. This OTU-like superfamily includes sequences from several viruses that are currently classified in the order *Tymovirales*, *Nidovirales* and *Bunyavirales* (an order encompassing a diverse group of negative-strand RNA viruses). It is now well-established that many OTU-like cysteine proteases have deubiquitinase (DUB) activity and cleave lysine-bound ubiquitin units at GG↓K cleavage sites [136].

The cysteine protease of turnip yellow mosaic virus (TYMV, genus *Tymovirus*, family *Tymoviridae*, order *Tymovirales*) was initially referred to as papain-like, following the identification of the catalytic Cys⁷⁸³ and His⁸⁶⁹ residues in the mid-1990s [137,138]. However, it was noted early on that the protease differs from papain in many aspects, notably in the sequence of residues surrounding the catalytic residues [138]. For example, Cys⁷⁸³ was followed by Leu rather than by the aromatic residue normally found in papain-like proteases. Determination of the TYMV protease structure confirmed a fold with more similarities to the yeast OTU DUB than to papain [139] (Figure 3B). The TYMV protease does not only process the polyprotein at two sites (S↓Q and A↓T, suggesting a relatively relaxed specificity) [140] (Figure 3A), it also functions as a DUB to regulate the ubiquitination status and stability of the viral RdRp [141]. The structures of the TYMV OTU-like Pro-DUB, the OTU-like nsp2 Pro-DUB of equine arteritis virus (EAV, genus *Alphaarterivirus*, family *Arteriviridae*, order *Nidovirales*) and the OTU-DUB of Crimean-Congo hemorrhagic fever virus (CCHMV, a negative-strand RNA virus from the genus *Orthonairovirus*, family *Nairoviridae*, order *Bunyavirales*) are strikingly similar, although they also display interesting differences [136,139,142–145] (Figure 3B). In contrast to the TYMV and EAV proteases, the CCHMV DUB does not have proteolytic activity on viral proteins. The DUB activity of the TYMV OTU-like protease is very specific, preferring a subset of ubiquitinated substrates, notably the TYMV RdRp [141]. The TYMV Pro-DUB differs from related viral or cellular OTU-like DUBs in that it is more compact and that it also lacks the third catalytic Asp (or Asn) residue [136,139]. The TYMV protease has developed a clever mechanism to regulate its dual activities through reversible conformation changes [27]. The catalytic site can adopt open or closed conformations that favor the protease or DUB activities, respectively. This is regulated by a mobile loop located near the catalytic site that can form a rigid flap against the catalytic cleft. Mutations that affect the mobility of the loop and prevent the close conformation resulted in a loss of DUB activity while conserving the protease function [27]. Thus, the detailed functional and structural characterization of the TYMV protease has provided a fine example of the strong selection pressures at work to regulate the multiple activities of viral proteases.

The order *Tymovirales* encompasses a large group of viruses related by common signatures of their replication enzymes but are otherwise quite diverse, for example, having different capsid structures and/or different movement protein modules [24]. The acquisition and evolution of protease and DUB domains to regulate the processing and stability of the replication proteins is also complex. Although members of the family *Alphaflexiviridae* lack a protease domain, most members of the families *Betaflexiviridae* encode either a single cysteine protease or tandem OTU-like and papain-like motifs (Figure 3A). Indeed, the presence of duplicated putative Cys-His dyads was already noted in the polyprotein of blueberry scorch virus (BBScV, genus *Carlavirus*, family *Betaflexiviridae*) in the early 1990s with only the C-terminal dyad implicated in polyprotein cleavage [146]. The N-terminal dyad is consistent with an OTU-like motif and may be specialized in DUB activity although this will need to be

confirmed experimentally [24] (Figure 3A). It has been hypothesized that the dual Pro-DUB function of the single TYMV protease may have been derived from an original OTU-like and papain-like dyad tandem, although alternative evolution scenarios are also possible [139]. Functional and structural characterization of additional proteases from the order *Tymovirales* will be necessary to gain a better understanding of their evolution history.

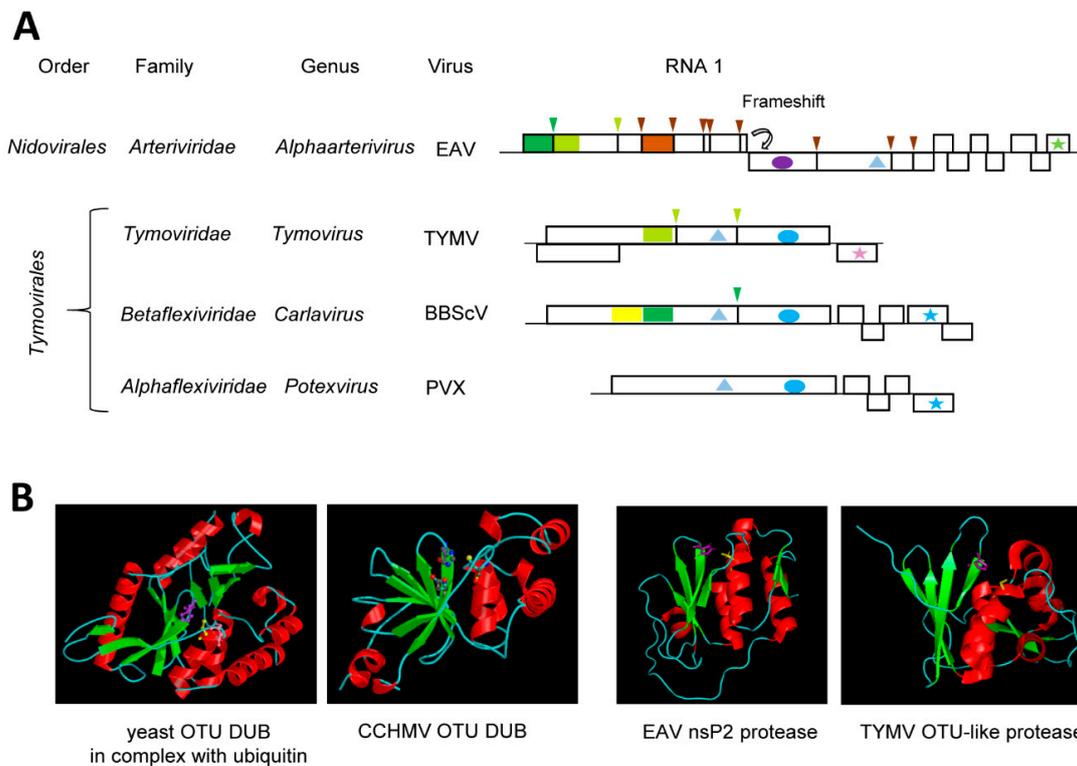


Figure 3. Ovarian Tumor Domain (OTU)-like cysteine proteases. **(A)** Genomic organization and proteolytic cleavages of representative (+)-strand RNA viruses. Polyproteins are shown by boxes with cleavage sites indicated by vertical lines. Protease domains are shown in dark green for papain-like cysteine proteases, lime green for OTU-like cysteine proteases and brown for chymotrypsin-like serine proteases. The same color code is used for arrows above each cleavage site indicating the protease responsible for the cleavage. The yellow square represents an OTU-like domain that does not orchestrate viral polyprotein cleavage. Ovals represent RdRp domains: purple for picorna-like RdRp and blue for alpha-like RdRp. Stars indicate the coat protein domains: green for the arteriviridae nucleocapsid, pink for type 2 jelly-roll domains and blue for filamentous coat proteins. Helicase domains are shown by the triangles: light blue for superfamily 1. **(B)** Structure of representative deubiquitinases (DUB) and/or proteases. Catalytic residues are represented as follows: Yeast OTU DUB (pdb: 3C0R, in complex with ubiquitin, only one monomer of the trimer is shown): Asp¹⁷⁷ (pink), Cys¹²⁰ (yellow), and His²²² (purple). Animal negative-strand RNA virus DUB: CCHMV OTU DUB (pdb: 3PT2_A): Cys⁴⁰ (yellow), His¹⁵¹ (purple), and Asp¹⁵³ (pink). (Image of the yeast and CCHMV OTU DUB structures are reprinted with permission from the MEROPS database (www.ebi.ac.uk/merops)). Animal (+)-strand RNA virus protease-DUB: EAV nsP2 protease (pdb: 4IUM) Cys²⁷⁰ (yellow) and His³³² (purple) and plant (+)-strand RNA virus protease-DUB: TYMV OTU-like protease (pdb: 4A5U): Cys⁷⁸³ (yellow) and His⁸⁶⁹ (purple). Image for the EAV and TYMV protease structure were generated using PyMol. EAV: equine arteritis virus, TYMV: turnip yellow mosaic virus, BBScV: blueberry scorch virus, PVX: potato virus X, CCHMV: Crimean-Congo hemorrhagic fever virus.

Homology-based modeling of the SMoV Pro2-Glu catalytic region implied a beta-sandwich structure that is typical of a concanavalin A-like lectin/glucanase fold [35] (Figure 4B). Although placed on different β -sheets, residues Glu¹¹⁹², Glu¹²⁷⁴, and Gln¹¹⁸⁰ are brought in close proximity in this model. Concanavalin A-like lectin/glucanases are a superfamily of proteins that share very low sequence identity and are found in a wide range of species, including microorganisms, insects, plants, and animals [148]. Most members of the superfamily have carbohydrate-binding activities and they influence an array of complex biological processes. Of interest, a group of bacterial and fungal glutamic proteases, which are collectively referred to as eqolisins (family G1, clan GA), also adopt the concanavalin A-like lectin/glucanase fold. The first described eqolisins was isolated from the plant pathogenic fungi *Scytalidium lignicolum* [149]. Structural and functional studies of eqolisins revealed a catalytic dyad consisting of glutamic acid (Glu¹³⁶) and glutamine (Gln⁵³) residues that are arranged on opposing beta-sheets [149–153]. Although the model of the SMoV Pro2-Glu protease catalytic domain shows some structural similarities with eqolisins, their amino acid sequence, the spacing between the catalytic residues, and the arrangements of the β -sheets differ significantly [35]. High-resolution structure analysis will be required to confirm the structure of the SMoV Pro2-Glu domain and its relationship with eqolisins or with other proteins with the concanavalin A-like lectin/glucanase fold. To date, there are no other viral proteases reported to adopt a similar fold. Whether the SMoV Pro2-Glu evolved from a divergent protease domain acquired from a plant pathogenic fungus or whether the protease catalytic activity developed from a host lectin protein (as suggested for eqolisins), is not clear. Either way, the unique proposed concanavalin A-like lectin/glucanases fold of Pro2-Glu, combined with the absence of this domain in most members of the family *Secoviridae*, suggest that it was relatively recently acquired.

A putative Pro2-Glu domain was detected in the minor coat protein (CPm) of several criniviruses and velariviruses (family *Closteroviridae*) [35]. Sequence motifs around the SMoV Pro2-Glu catalytic glutamic acid and glutamine residues were highly conserved in the CPms. Similar β -sheets structures were also predicted around the conserved sequences, suggesting a possible concanavalin A-like lectin/glucanases fold. Members of the family *Closteroviridae* incorporate the CPm in the short tail of the virion filamentous particle and the CPm is required for virus transmission by its aphid vector [25,154–156]. The CPm functions by binding to chitin sugar moieties on the cuticular surface of the aphid cibarium, thus anchoring citrus tristeza virions [157]. Competitive assays using monosaccharides or lectins reduced virion binding. Similar results were reported for the CPm of lettuce infectious yellows virus and its role in whitefly transmission [158]. These observations raise several interesting questions: (1) Does the SMoV Pro2-Glu domain have a similar lectin activity? If so, is it also involved in the transmission of SMoV by its aphid vector and does the protease activity contributes to regulating vector transmission? (2) Do the crinivirus and velarivirus CPms have glutamic protease activity? If so, what is the biological significance of this activity in vector transmission or in other aspects of the virus infection cycle?

6. Aspartic Proteases Encoded by Reverse-Transcribing Viruses and by a Plant Negative-Strand RNA Virus, but Not (Yet?) by (+)-Strand RNA Viruses

Reverse-transcribing viruses belonging to the order *Ortervirales* encode aspartic proteases that are related to the cellular pepsin and they are often referred to as retropepsin [159]. Retroviruses are restricted to vertebrate hosts and they are estimated to be the oldest known viral group [160]. The aspartic protease of human immunodeficiency virus 1 (HIV-1) is the best characterized and it has been shown to be essential for virion maturation [161]. The HIV-1 protease activity requires the formation of a homodimer, whereby each monomer confers an aspartate residue to form the active site (Asp²⁵ and Asp^{25'}) [162,163]. The family *Caulimoviridae* (order *Ortervirales*) is a family of plant-infecting pararetroviruses. Pararetroviruses share genomic organization, replication strategy, and phylogenetic relatedness with ancestral retroviruses but lack an integrase gene and package DNA instead of RNA into virions [159]. Cauliflower mosaic virus (CaMV, genus *Caulimovirus*) was the first pararetrovirus

identified to encode an aspartic protease, which has also been implicated in virion maturation and regulation of the viral coat protein nuclear targeting [164–166]. Although the tertiary structure of caulimovirus proteases has not been determined, catalytic Asp residues that are found in the conserved D(T/S)G motif are assumed to adopt a homodimer configuration similar to that of retroviral proteases.

Recent findings revealed that aspartic proteases are not restricted to members of the order *Ortervirales*. Indeed, a retropepsin-like aspartic protease domain with a conserved DTG sequence was found to be encoded by a plant negative-strand RNA virus [167]. Citrus psorosis ophiovirus (CPsV, genus *Ophiovirus*, family *Aspiviridae*) is a tripartite virus that contains a protease domain within its movement protein. Mutation of the catalytic Asp residue prevented not only the maturation cleavage of the MP, but also the formation of tubular structures that are necessary for cell-to-cell movement of the virus. Whether the CPsV 20K protease was acquired from a plant reverse-transcribing element or from a cellular protease by parallel evolution is not known. To date, aspartic protease domains have not been identified with certainty in association with positive-strand RNA viruses. Although an early study identified a possible retropepsin-like DSG motif in the polyprotein of a closterovirus (BYV) [128], a proteolytic activity that is associated with this motif has not been confirmed experimentally. However, given the recent identification of a glutamic protease in a plant (+)-strand RNA virus [35], the possibility that they could also encode aspartic proteases cannot be excluded.

7. Conclusions

Detailed functional and structural analyses of viral proteases have provided new insights into the recurrent acquisition of proteolytic enzymes by viruses and into the adaptation of these protease domains to accommodate and facilitate virus infection cycles. The origin of some viral proteases are ancient, as exemplified by the main chymotrypsin-like serine and cysteine proteases of animal, plant, and lower eukaryotes picorna-like viruses that have limited sequence identities but share conserved catalytic residues and a similar general fold. As detailed in Section 3, this long history has driven diverse and intricate adaptations to facilitate and regulate the multi-functional activities of these viral proteases within the constraints of viral RNA genome size limitations. When compared to the cellular chymotrypsin, viral proteases have developed more stringent cleavage site specificities, which vary from one virus to another and are determined by the structure of their SBPs. Although specificity determinants are relatively well understood for the 3C-Pro of animal picornaviruses and 3CL-Pro of plant potyviruses and sobemoviruses, further work is required to understand the determinants and biological relevance of the divergent specificities of the 3CL-Pro of many plant viruses in the family *Secoviridae*. Understanding these protease specificities is critical not only for accurate prediction of viral polyprotein cleavage sites in emerging viruses, but also to identify host protein targets of viral proteases, an area that remains unexplored for plant viruses.

While the large family of chymotrypsin-like viral proteases share a common and ancient origin, other viral proteases have more complex evolutionary histories, which is reflected by the sequential acquisition of diverse protein domains, duplication and/or deletion of protein domains, and in some cases adaptation of non-proteolytic enzymes to attain protease activities (Sections 4 and 5). It is clear that the evolution of these diverse proteases has been influenced by their multi-functional activities. Some viral proteases adopt multi-domain structures, where the protease activity is restricted to a small domain that is linked to other functional domains by flexible loops. These small protease domains often adopt a simplified fold when compared to their cellular counterparts, as exemplified by the minimalistic papain-like domains of the plant potyvirus HC-Pro or animal arterivirus nsP2 proteases (Sections 4 and 4.1). Other viral proteases have adopted compact but flexible structures to achieve and regulate dual activities, for example the tymovirus OTU-like protease/deubiquitinase (Section 4.3). The catalytic mechanisms, biological functions and structures of many viral proteases remain to be examined. These include cysteine proteases from plant viruses that have been annotated as papain-like or OTU-like, such as the closterovirus leader proteases (Section 4.2) and the putative proteases of benyviruses [168] and cileviruses [169].

The recent discovery of an atypical glutamic protease encoded by a plant picorna-like virus (Section 5) implies that we have not yet fully explored the diversity of plant (+)-strand virus proteases. The SMOV glutamic protease is predicted to adopt a concanavalin A-like lectin/glucanase fold, suggesting that it has other as of yet unexplored activities. Structural studies will be required to confirm the predicted fold and to understand the catalytic mechanism of this novel viral protease. The discovery of an accessory glutamic protease encoded by a secovirid was unexpected, as other viruses in the family are not known to encode accessory proteases in addition to the main 3C-like protease. The SMOV glutamic protease presents little sequence identity with other proteins in available databases, and the prediction of its proteolytic activity would have been difficult without functional analysis. There are many other viral protein domains of “unknown functions” in sequence annotations, some of which may represent novel protease activities. Thus, it is likely that the repertoire of proteases encoded by plant (+)-strand RNA viruses will continue to expand in the years to come.

Author Contributions: K.S.M. and H.S. wrote the manuscript jointly.

Funding: “The APC was funded by Agriculture and Agri-Food Canada Core Funding”.

Conflicts of Interest: “The authors declare no conflict of interest.”

References

1. Wolf, Y.I.; Kazlauskas, D.; Iranzo, J.; Lucia-Sanz, A.; Kuhn, J.H.; Krupovic, M.; Dolja, V.V.; Koonin, E.V. Origins and Evolution of the Global RNA Virome. *MBio* **2018**, *9*. [[CrossRef](#)] [[PubMed](#)]
2. Krupovic, M.; Koonin, E.V. Multiple origins of viral capsid proteins from cellular ancestors. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E2401–E2410. [[CrossRef](#)] [[PubMed](#)]
3. Koonin, E.V.; Dolja, V.V. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiol. Mol. Biol. Rev.* **2014**, *78*, 278–303. [[CrossRef](#)] [[PubMed](#)]
4. Koonin, E.V.; Wolf, Y.I.; Nagasaki, K.; Dolja, V.V. The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nat. Rev. Microbiol.* **2008**, *6*, 925–939. [[CrossRef](#)] [[PubMed](#)]
5. Koonin, E.V.; Dolja, V.V.; Krupovic, M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* **2015**, *479*, 2–25. [[CrossRef](#)]
6. Dolja, V.V.; Koonin, E.V. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Res.* **2018**, *244*, 36–52. [[CrossRef](#)] [[PubMed](#)]
7. Mascia, T.; Gallitelli, D. Synergies and antagonisms in virus interactions. *Plant Sci.* **2016**, *252*, 176–192. [[CrossRef](#)]
8. Roossinck, M.J. Lifestyles of plant viruses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2010**, *365*, 1899–1905. [[CrossRef](#)]
9. Dietzgen, R.G.; Mann, K.S.; Johnson, K.N. Plant Virus-Insect Vector Interactions: Current and Potential Future Research Directions. *Viruses* **2016**, *8*, 303. [[CrossRef](#)]
10. Whitfield, A.E.; Falk, B.W.; Rotenberg, D. Insect vector-mediated transmission of plant viruses. *Virology* **2015**, *479–480*, 278–289. [[CrossRef](#)]
11. Elena, S.F.; Fraile, A.; Garcia-Arenal, F. Evolution and emergence of plant viruses. *Adv. Virus Res.* **2014**, *88*, 161–191. [[PubMed](#)]
12. Borderia, A.V.; Rozen-Gagnon, K.; Vignuzzi, M. Fidelity Variants and RNA Quasispecies. *Curr. Top. Microbiol. Immunol.* **2016**, *392*, 303–322. [[PubMed](#)]
13. Borderia, A.V.; Stapleford, K.A.; Vignuzzi, M. RNA virus population diversity: Implications for inter-species transmission. *Curr. Opin. Virol.* **2011**, *1*, 643–648. [[CrossRef](#)] [[PubMed](#)]
14. Holmes, E.C. Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol.* **2003**, *11*, 543–546. [[CrossRef](#)] [[PubMed](#)]
15. Belshaw, R.; Gardner, A.; Rambaut, A.; Pybus, O.G. Pacing a small cage: Mutation and RNA viruses. *Trends Ecol. Evol.* **2008**, *23*, 188–193. [[CrossRef](#)] [[PubMed](#)]
16. Rodamilans, B.; Shan, H.; Pasin, F.; Garcia, J.A. Plant Viral Proteases: Beyond the Role of Peptide Cutters. *Front. Plant Sci.* **2018**, *9*, 666. [[CrossRef](#)] [[PubMed](#)]

17. Laitinen, O.H.; Svedin, E.; Kapell, S.; Nurminen, A.; Hytonen, V.P.; Flodstrom-Tullberg, M. Enteroviral proteases: Structure, host interactions and pathogenicity. *Rev. Med. Virol.* **2016**, *26*, 251–267. [[CrossRef](#)] [[PubMed](#)]
18. Sun, D.; Chen, S.; Cheng, A.; Wang, M. Roles of the Picornaviral 3C Proteinase in the Viral Life Cycle and Host Cells. *Viruses* **2016**, *8*, 82. [[CrossRef](#)]
19. Jensen, L.M.; Walker, E.J.; Jans, D.A.; Ghildyal, R. Proteases of human rhinovirus: Role in infection. *Meth. Mol. Biol.* **2015**, *1221*, 129–141.
20. Mielech, A.M.; Chen, Y.; Mesecar, A.D.; Baker, S.C. Nidovirus papain-like proteases: Multifunctional enzymes with protease, deubiquitinating and deISGylating activities. *Virus Res.* **2014**, *194*, 184–190. [[CrossRef](#)]
21. Yost, S.A.; Marcotrigiano, J. Viral precursor polyproteins: Keys of regulation from replication to maturation. *Curr. Opin. Virol.* **2013**, *3*, 137–142. [[CrossRef](#)] [[PubMed](#)]
22. Le Gall, O.; Christian, P.; Fauquet, C.M.; King, A.M.; Knowles, N.J.; Nakashima, N.; Stanway, G.; Gorbalenya, A.E. Picornavirales, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-T = 3 virion architecture. *Arch. Virol.* **2008**, *153*, 715–727. [[CrossRef](#)] [[PubMed](#)]
23. Sanfacon, H.; Gorbalenya, A.E.; Knowles, N.J.; Chen, Y. Order Picornavirales. In *Virus Taxonomy: Classification and Nomenclature of Viruses; Ninth Report of the International Committee on the Taxonomy of Viruses*; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier: San Diego, CA, USA, 2011; pp. 835–839.
24. Martelli, G.P.; Adams, M.J.; Kreuze, J.F.; Dolja, V.V. Family Flexiviridae: A Case Study in Virion and Genome Plasticity. *Annu. Rev. Phytopathol.* **2007**, *45*, 73–100. [[CrossRef](#)] [[PubMed](#)]
25. Dolja, V.V.; Kreuze, J.F.; Valkonen, J.P. Comparative and functional genomics of closteroviruses. *Virus Res.* **2006**, *117*, 38–51. [[CrossRef](#)] [[PubMed](#)]
26. Valli, A.A.; Gallo, A.; Rodamilans, B.; Lopez-Moya, J.J.; Garcia, J.A. The HCPro from the Potyviridae family: An enviable multitasking Helper Component that every virus would like to have. *Mol. Plant Pathol.* **2018**, *19*, 744–763. [[CrossRef](#)] [[PubMed](#)]
27. Jupin, I.; Ayach, M.; Jomat, L.; Fieulaine, S.; Bressanelli, S. A mobile loop near the active site acts as a switch between the dual activities of a viral protease/deubiquitinase. *PLoS Pathog.* **2017**, *13*, e1006714. [[CrossRef](#)]
28. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **2018**, *46*, D624–D632. [[CrossRef](#)]
29. King, A.M.Q.; Lefkowitz, E.J.; Mushegian, A.R.; Adams, M.J.; Dutilh, B.E.; Gorbalenya, A.E.; Harrach, B.; Harrison, R.L.; Junglen, S.; Knowles, N.J.; et al. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2018). *Arch. Virol.* **2018**, *163*, 2601–2631. [[CrossRef](#)]
30. Thomas, E.L.; van der Hoorn, R.A.L. Ten Prominent Host Proteases in Plant-Pathogen Interactions. *Int. J. Mol. Sci.* **2018**, *19*, 639. [[CrossRef](#)]
31. Van der Hoorn, R.A. Plant proteases: From phenotypes to molecular mechanisms. *Annu. Rev. Plant Biol.* **2008**, *59*, 191–223. [[CrossRef](#)]
32. Tong, L. Viral proteases. *Chem. Rev.* **2002**, *102*, 4609–4626. [[CrossRef](#)] [[PubMed](#)]
33. Verdaguer, N.; Ferrero, D.; Murthy, M. Viruses and viral proteins. *IUCrJ* **2014**, *1*, 492–504. [[CrossRef](#)] [[PubMed](#)]
34. Babe, L.M.; Craik, C.S. Viral proteases: Evolution of diverse structural motifs to optimize function. *Cell* **1997**, *91*, 427–430. [[CrossRef](#)]
35. Mann, K.S.; Chisholm, J.; Sanfacon, H. Strawberry mottle virus (family Secoviridae, order Picornavirales) encodes a novel glutamic protease to process the RNA2 polyprotein at two cleavage sites. *J. Virol.* **2018**, in press. [[CrossRef](#)] [[PubMed](#)]
36. Ryan, M.D.; Flint, M. Virus-encoded proteinases of the picornavirus super-group. *J. Gen. Virol.* **1997**, *78* (Pt 4), 699–723. [[CrossRef](#)] [[PubMed](#)]
37. Seipelt, J.; Guarne, A.; Bergmann, E.; James, M.; Sommergruber, W.; Fita, I.; Skern, T. The structures of picornaviral proteinases. *Virus Res.* **1999**, *62*, 159–168. [[CrossRef](#)]
38. Thompson, J.R.; Dasgupta, I.; Fuchs, M.; Iwanami, T.; Karasev, A.V.; Petrzik, K.; Sanfacon, H.; Tzanetakis, I.; van der Vlugt, R.; Wetzel, T.; et al. ICTV Virus Taxonomy Profile: Secoviridae. *J. Gen. Virol.* **2017**, *98*, 529–531. [[CrossRef](#)]

39. Wylie, S.J.; Adams, M.; Chalam, C.; Kreuze, J.; Lopez-Moya, J.J.; Ohshima, K.; Praveen, S.; Rabenstein, F.; Stenger, D.; Wang, A.; et al. ICTV Virus Taxonomy Profile: Potyviridae. *J. Gen. Virol.* **2017**, *98*, 352–354. [[CrossRef](#)]
40. Wellink, J.; van Kammen, A. Proteases involved in the processing of viral polyproteins. Brief review. *Arch. Virol.* **1988**, *98*, 1–26. [[CrossRef](#)]
41. Dougherty, W.G.; Semler, B.L. Expression of virus-encoded proteinases: Functional and structural similarities with cellular enzymes. *Microbiol. Rev.* **1993**, *57*, 781–822.
42. Gorbalenya, A.E.; Koonin, E.V.; Blinov, V.M.; Donchenko, A.P. Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Lett.* **1988**, *236*, 287–290. [[CrossRef](#)]
43. Gorbalenya, A.E.; Donchenko, A.P.; Blinov, V.M.; Koonin, E.V. Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. *FEBS Lett.* **1989**, *243*, 103–114. [[CrossRef](#)]
44. Gorbalenya, A.E.; Blinov, V.M.; Donchenko, A.P. Poliovirus-encoded proteinase 3C: A possible evolutionary link between cellular serine and cysteine proteinase families. *FEBS Lett.* **1986**, *194*, 253–257. [[CrossRef](#)]
45. Bazan, J.F.; Fletterick, R.J. Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: Structural and functional implications. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 7872–7876. [[CrossRef](#)]
46. Allaire, M.; Chernaia, M.M.; Malcolm, B.A.; James, M.N. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* **1994**, *369*, 72–76. [[CrossRef](#)] [[PubMed](#)]
47. Bergmann, E.M.; Mosimann, S.C.; Chernaia, M.M.; Malcolm, B.A.; James, M.N. The refined crystal structure of the 3C gene product from hepatitis A virus: Specific proteinase activity and RNA recognition. *J. Virol.* **1997**, *71*, 2436–2448. [[PubMed](#)]
48. Mosimann, S.C.; Cherney, M.M.; Sia, S.; Plotch, S.; James, M.N. Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.* **1997**, *273*, 1032–1047. [[CrossRef](#)] [[PubMed](#)]
49. Matthews, D.A.; Smith, W.W.; Ferre, R.A.; Condon, B.; Budahazi, G.; Sisson, W.; Villafranca, J.E.; Janson, C.A.; McElroy, H.E.; Gribskov, C.L.; et al. Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell* **1994**, *77*, 761–771. [[CrossRef](#)]
50. Valli, A.; Lopez-Moya, J.J.; Garcia, J.A. Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *J. Gen. Virol.* **2007**, *88*, 1016–1028. [[CrossRef](#)]
51. Adams, M.J.; Antoniw, J.F.; Beaudoin, F. Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Mol. Plant Pathol.* **2005**, *6*, 471–487. [[CrossRef](#)]
52. Ypma-Wong, M.F.; Filman, D.J.; Hogle, J.M.; Semler, B.L. Structural domains of the poliovirus polyprotein are major determinants for proteolytic cleavage at Gln-Gly pairs. *J. Biol. Chem.* **1988**, *263*, 17846–17856. [[PubMed](#)]
53. Clark, A.J.; Bertens, P.; Wellink, J.; Shanks, M.; Lomonossoff, G.P. Studies on hybrid comoviruses reveal the importance of three-dimensional structure for processing of the viral coat proteins and show that the specificity of cleavage is greater in trans than in cis. *Virology* **1999**, *263*, 184–194. [[CrossRef](#)] [[PubMed](#)]
54. Jacobson, M.F.; Baltimore, D. Polypeptide cleavages in the formation of poliovirus proteins. *Proc. Natl. Acad. Sci. USA* **1968**, *61*, 77–84. [[CrossRef](#)] [[PubMed](#)]
55. Gorbalenya, A.E.; Svitkin, Y.V.; Kazachkov, Y.A.; Agol, V.I. Encephalomyocarditis virus-specific polypeptide p22 is involved in the processing of the viral precursor polypeptides. *FEBS Lett.* **1979**, *108*, 1–5. [[CrossRef](#)]
56. Shih, D.S.; Shih, C.T.; Zimmern, D.; Rueckert, R.R.; Kaesberg, P. Translation of encephalomyocarditis virus RNA in reticulocyte lysates: Kinetic analysis of the formation of virion proteins and a protein required for processing. *J. Virol.* **1979**, *30*, 472–480.
57. Palmenberg, A.C.; Pallansch, M.A.; Rueckert, R.R. Protease required for processing picornaviral coat protein resides in the viral replicase gene. *J. Virol.* **1979**, *32*, 770–778.
58. Zell, R.; Delwart, E.; Gorbalenya, A.E.; Hovi, T.; King, A.M.Q.; Knowles, N.J.; Lindberg, A.M.; Pallansch, M.A.; Palmenberg, A.C.; Reuter, G.; et al. ICTV Virus Taxonomy Profile: Picornaviridae. *J. Gen. Virol.* **2017**, *98*, 2421–2422. [[CrossRef](#)]
59. Daros, J.A.; Carrington, J.C. RNA binding activity of NIa proteinase of tobacco etch potyvirus. *Virology* **1997**, *237*, 327–336. [[CrossRef](#)]

60. Sanfacon, H.; Wellink, J.; Le Gall, O.; Karasev, A.; van der Vlugt, R.; Wetzel, T. Secoviridae: A proposed family of plant viruses within the order Picornvirales that combines the families Sequiviridae and Comoviridae, the unassigned genera Cheravirus and Sadwavirus, and the proposed genus Torradovirus. *Arch. Virol.* **2009**, *154*, 899–907. [[CrossRef](#)]
61. Isogai, M.; Tatuto, N.; Ujiie, C.; Watanabe, M.; Yoshikawa, N. Identification and characterization of blueberry latent spherical virus, a new member of subgroup C in the genus Nepovirus. *Arch. Virol.* **2012**, *157*, 297–303. [[CrossRef](#)]
62. Verver, J.; Goldbach, R.; Garcia, J.A.; Vos, P. In vitro expression of a full-length DNA copy of cowpea mosaic virus B RNA: Identification of the B RNA encoded 24-kd protein as a viral protease. *EMBO J.* **1987**, *6*, 549–554. [[CrossRef](#)] [[PubMed](#)]
63. Hans, F.; Sanfacon, H. Tomato ringspot nepovirus protease: Characterization and cleavage site specificity. *J. Gen. Virol.* **1995**, *76*, 917–927. [[CrossRef](#)] [[PubMed](#)]
64. Margis, R.; Viry, M.; Pinck, M.; Pinck, L. Cloning and in vitro characterization of the grapevine fanleaf virus proteinase cistron. *Virology* **1991**, *185*, 779–787. [[CrossRef](#)]
65. Fuchs, M.; Schmitt-Keichinger, C.; Sanfacon, H. A Renaissance in Nepovirus Research Provides New Insights Into Their Molecular Interface With Hosts and Vectors. *Adv. Virus Res.* **2017**, *97*, 61–105. [[PubMed](#)]
66. Pouwels, J.; Carette, J.E.; Van Lent, J.; Wellink, J. Cowpea mosaic virus: Effects on host cell processes. *Mol. Plant Pathol.* **2002**, *3*, 411–418. [[CrossRef](#)] [[PubMed](#)]
67. Peters, S.A.; Voorhorst, W.G.; Wery, J.; Wellink, J.; van Kammen, A. A regulatory role for the 32K protein in proteolytic processing of cowpea mosaic virus polyproteins. *Virology* **1992**, *191*, 81–89. [[CrossRef](#)]
68. Wetzel, T.; Chisholm, J.; Bassler, A.; Sanfacon, H. Characterization of proteinase cleavage sites in the N-terminal region of the RNA1-encoded polyprotein from Arabis mosaic virus (subgroup A nepovirus). *Virology* **2008**, *375*, 159–169. [[CrossRef](#)] [[PubMed](#)]
69. Wang, A.; Sanfacon, H. Proteolytic processing at a novel cleavage site in the N-terminal region of the tomato ringspot nepovirus RNA-1-encoded polyprotein in vitro. *J. Gen. Virol.* **2000**, *81*, 2771–2781. [[CrossRef](#)] [[PubMed](#)]
70. Hemmer, O.; Greif, C.; Dufourcq, P.; Reinbolt, J.; Fritsch, C. Functional characterization of the proteolytic activity of the tomato black ring nepovirus RNA-1-encoded polyprotein. *Virology* **1995**, *206*, 362–371. [[CrossRef](#)]
71. Margis, R.; Viry, M.; Pinck, M.; Bardonnnet, N.; Pinck, L. Differential proteolytic activities of precursor and mature forms of the 24K proteinase of grapevine fanleaf nepovirus. *Virology* **1994**, *200*, 79–86. [[CrossRef](#)]
72. Chisholm, J.; Wiczorek, A.; Sanfacon, H. Expression and partial purification of recombinant tomato ringspot nepovirus 3C-like proteinase: Comparison of the activity of the mature proteinase and the VPg-proteinase precursor. *Virus Res.* **2001**, *79*, 153–164. [[CrossRef](#)]
73. Wellink, J.; Rezelman, G.; Goldbach, R.; Beyreuther, K. Determination of the proteolytic processing sites in the polyprotein encoded by the bottom-component RNA of Cowpea mosaic virus. *J. Virol.* **1986**, *59*, 50–58. [[PubMed](#)]
74. Carrier, K.; Hans, F.; Sanfacon, H. Mutagenesis of amino acids at two tomato ringspot nepovirus cleavage sites: Effect on proteolytic processing in cis and in trans by the 3C-like protease. *Virology* **1999**, *258*, 161–175. [[CrossRef](#)] [[PubMed](#)]
75. Phan, J.; Zdanov, A.; Evdokimov, A.G.; Tropea, J.E.; Peters, H.K., 3rd; Kapust, R.B.; Li, M.; Wlodawer, A.; Waugh, D.S. Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* **2002**, *277*, 50564–50572. [[CrossRef](#)]
76. Dougherty, W.G.; Carrington, J.C.; Cary, S.M.; Parks, T.D. Biochemical and mutational analysis of a plant virus polyprotein cleavage site. *EMBO J.* **1988**, *7*, 1281–1287. [[CrossRef](#)] [[PubMed](#)]
77. Kapust, R.B.; Tozser, J.; Copeland, T.D.; Waugh, D.S. The P1' specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* **2002**, *294*, 949–955. [[CrossRef](#)]
78. Nautiyal, K.; Kuroda, Y. A SEP tag enhances the expression, solubility and yield of recombinant TEV protease without altering its activity. *New Biotechnol.* **2018**, *42*, 77–84. [[CrossRef](#)]
79. Raran-Kurussi, S.; Cherry, S.; Zhang, D.; Waugh, D.S. Removal of Affinity Tags with TEV Protease. *Meth. Mol. Biol.* **2017**, *1586*, 221–230.
80. Taxis, C.; Knop, M. TIPI: TEV protease-mediated induction of protein instability. *Meth. Mol. Biol.* **2012**, *832*, 611–626.

81. Rubio, V.; Shen, Y.; Saijo, Y.; Liu, Y.; Gusmaroli, G.; Dinesh-Kumar, S.P.; Deng, X.W. An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation. *Plant J.* **2005**, *41*, 767–778. [[CrossRef](#)]
82. Kapust, R.B.; Waugh, D.S. Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr. Purif.* **2000**, *19*, 312–318. [[CrossRef](#)] [[PubMed](#)]
83. Ceriani, M.F.; Marcos, J.F.; Hopp, H.E.; Beachy, R.N. Simultaneous accumulation of multiple viral coat proteins from a TEV-NIa based expression vector. *Plant Mol. Biol.* **1998**, *36*, 239–248. [[CrossRef](#)] [[PubMed](#)]
84. Marcos, J.F.; Beachy, R.N. In vitro characterization of a cassette to accumulate multiple proteins through synthesis of a self-processing polypeptide. *Plant Mol. Biol.* **1994**, *24*, 495–503. [[CrossRef](#)] [[PubMed](#)]
85. Puckette, M.; Smith, J.D.; Gabbert, L.; Schutta, C.; Barrera, J.; Clark, B.A.; Neilan, J.G.; Rasmussen, M. Production of foot-and-mouth disease virus capsid proteins by the TEV protease. *J. Biotechnol.* **2018**, *275*, 7–12. [[CrossRef](#)] [[PubMed](#)]
86. Garcia, J.A.; Riechmann, J.L.; Martin, M.T.; Lain, S. Proteolytic activity of the plum pox potyvirus NIa-protein on excess of natural and artificial substrates in Escherichia coli. *FEBS Lett.* **1989**, *257*, 269–273. [[CrossRef](#)]
87. Garcia, J.A.; Lain, S.; Cervera, M.T.; Riechmann, J.L.; Martin, M.T. Mutational analysis of plum pox potyvirus polyprotein processing by the NIa protease in Escherichia coli. *J. Gen. Virol.* **1990**, *71* (Pt 12), 2773–2779. [[CrossRef](#)] [[PubMed](#)]
88. Martin, M.T.; Lopez Otin, C.; Lain, S.; Garcia, J.A. Determination of polyprotein processing sites by amino terminal sequencing of nonstructural proteins encoded by plum pox potyvirus. *Virus Res.* **1990**, *15*, 97–106. [[CrossRef](#)]
89. Kang, H.; Lee, Y.J.; Goo, J.H.; Park, W.J. Determination of the substrate specificity of turnip mosaic virus NIa protease using a genetic method. *J. Gen. Virol.* **2001**, *82*, 3115–3117. [[CrossRef](#)]
90. Tozser, J.; Tropea, J.E.; Cherry, S.; Bagossi, P.; Copeland, T.D.; Wlodawer, A.; Waugh, D.S. Comparison of the substrate specificity of two potyvirus proteases. *FEBS J.* **2005**, *272*, 514–523. [[CrossRef](#)]
91. Sun, P.; Austin, B.P.; Tozser, J.; Waugh, D.S. Structural determinants of tobacco vein mottling virus protease substrate specificity. *Protein Soc.* **2010**, *19*, 2240–2251. [[CrossRef](#)]
92. Garcia, J.A.; Lain, S. Proteolytic activity of plum pox virus-tobacco etch virus chimeric NIa proteases. *FEBS Lett.* **1991**, *281*, 67–72. [[CrossRef](#)]
93. Mathur, C.; Jimsheena, V.K.; Banerjee, S.; Makinen, K.; Gowda, L.R.; Savithri, H.S. Functional regulation of PVBV Nuclear Inclusion protein-a protease activity upon interaction with Viral Protein genome-linked and phosphorylation. *Virology* **2012**, *422*, 254–264. [[CrossRef](#)] [[PubMed](#)]
94. Sabharwal, P.; Srinivas, S.; Savithri, H.S. Mapping the domain of interaction of PVBV VPg with NIa-Pro: Role of N-terminal disordered region of VPg in the modulation of structure and function. *Virology* **2018**, *524*, 18–31. [[CrossRef](#)] [[PubMed](#)]
95. Schaad, M.C.; Haldeman-Cahill, R.; Cronin, S.; Carrington, J.C. Analysis of the VPg-proteinase (NIa) encoded by tobacco etch potyvirus: Effects of mutations on subcellular transport, proteolytic processing, and genome amplification. *J. Virol.* **1996**, *70*, 7039–7048. [[PubMed](#)]
96. Rantalainen, K.I.; Uversky, V.N.; Permi, P.; Kalkkinen, N.; Dunker, A.K.; Makinen, K. Potato virus A genome-linked protein VPg is an intrinsically disordered molten globule-like protein with a hydrophobic core. *Virology* **2008**, *377*, 280–288. [[CrossRef](#)] [[PubMed](#)]
97. Martinez, F.; Rodrigo, G.; Aragonés, V.; Ruiz, M.; Lodewijk, I.; Fernandez, U.; Elena, S.F.; Daros, J.A. Interaction network of tobacco etch potyvirus NIa protein with the host proteome during infection. *BMC Genom.* **2016**, *17*, 87. [[CrossRef](#)] [[PubMed](#)]
98. Jiang, J.; Laliberte, J.F. The genome-linked protein VPg of plant viruses—a protein with many partners. *Curr. Opin. Virol.* **2011**, *1*, 347–354. [[CrossRef](#)]
99. Somera, M.; Sarmiento, C.; Truve, E. Overview on Sobemoviruses and a Proposal for the Creation of the Family Sobemoviridae. *Viruses* **2015**, *7*, 3076–3115. [[CrossRef](#)]
100. Li, X.; Ryan, M.D.; Lamb, J.W. Potato leafroll virus protein P1 contains a serine proteinase domain. *J. Gen. Virol.* **2000**, *81*, 1857–1864. [[CrossRef](#)]
101. Sadowy, E.; Juszczuk, M.; David, C.; Gronenborn, B.; Hulanicka, M.D. Mutational analysis of the proteinase function of Potato leafroll virus. *J. Gen. Virol.* **2001**, *82*, 1517–1527. [[CrossRef](#)]
102. Nair, S.; Savithri, H.S. Processing of SeMV polyproteins revisited. *Virology* **2010**, *396*, 106–117. [[CrossRef](#)] [[PubMed](#)]

103. Li, X.; Halpin, C.; Ryan, M.D. A novel cleavage site within the potato leafroll virus P1 polyprotein. *J. Gen. Virol.* **2007**, *88*, 1620–1623. [[CrossRef](#)] [[PubMed](#)]
104. Gayathri, P.; Satheshkumar, P.S.; Prasad, K.; Nair, S.; Savithri, H.S.; Murthy, M.R. Crystal structure of the serine protease domain of Sesbania mosaic virus polyprotein and mutational analysis of residues forming the S1-binding pocket. *Virology* **2006**, *346*, 440–451. [[CrossRef](#)] [[PubMed](#)]
105. Satheshkumar, P.S.; Gayathri, P.; Prasad, K.; Savithri, H.S. “Natively unfolded” VPg is essential for Sesbania mosaic virus serine protease activity. *J. Biol. Chem.* **2005**, *280*, 30291–30300. [[CrossRef](#)] [[PubMed](#)]
106. Nair, S.; Gayathri, P.; Murthy, M.R.; Savithri, H.S. Stacking interactions of W271 and H275 of SeMV serine protease with W43 of natively unfolded VPg confer catalytic activity to protease. *Virology* **2008**, *382*, 83–90. [[CrossRef](#)] [[PubMed](#)]
107. Verchot, J.; Koonin, E.V.; Carrington, J.C. The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. *Virology* **1991**, *185*, 527–535. [[CrossRef](#)]
108. Verchot, J.; Herndon, K.L.; Carrington, J.C. Mutational analysis of the tobacco etch potyviral 35-kDa proteinase: Identification of essential residues and requirements for autoproteolysis. *Virology* **1992**, *190*, 298–306. [[CrossRef](#)]
109. Valli, A.; Martin-Hernandez, A.M.; Lopez-Moya, J.J.; Garcia, J.A. RNA silencing suppression by a second copy of the P1 serine protease of Cucumber vein yellowing ipomovirus, a member of the family Potyviridae that lacks the cysteine protease HCPro. *J. Virol.* **2006**, *80*, 10055–10063. [[CrossRef](#)]
110. Rodamilans, B.; Valli, A.; Garcia, J.A. Mechanistic divergence between P1 proteases of the family Potyviridae. *J. Gen. Virol.* **2013**, *94*, 1407–1414. [[CrossRef](#)]
111. Maliogka, V.I.; Salvador, B.; Carbonell, A.; Saenz, P.; Leon, D.S.; Oliveros, J.C.; Delgadillo, M.O.; Garcia, J.A.; Simon-Mateo, C. Virus variants with differences in the P1 protein coexist in a Plum pox virus population and display particular host-dependent pathogenicity features. *Mol. Plant Pathol.* **2012**, *13*, 877–886. [[CrossRef](#)]
112. Pasin, F.; Simon-Mateo, C.; Garcia, J.A. The hypervariable amino-terminus of P1 protease modulates potyviral replication and host defense responses. *PLoS Pathog.* **2014**, *10*, e1003985. [[CrossRef](#)] [[PubMed](#)]
113. Shan, H.; Pasin, F.; Valli, A.; Castillo, C.; Rajulu, C.; Carbonell, A.; Simon-Mateo, C.; Garcia, J.A.; Rodamilans, B. The Potyviridae P1a leader protease contributes to host range specificity. *Virology* **2015**, *476*, 264–270. [[CrossRef](#)] [[PubMed](#)]
114. Shan, H.; Pasin, F.; Tzanetakis, I.E.; Simon-Mateo, C.; Garcia, J.A.; Rodamilans, B. Truncation of a P1 leader proteinase facilitates potyvirus replication in a non-permissive host. *Mol. Plant Pathol.* **2018**, *19*, 1504–1510. [[CrossRef](#)] [[PubMed](#)]
115. Garcia, J.A.; Schrijvers, L.; Tan, A.; Vos, P.; Wellink, J.; Goldbach, R. Proteolytic activity of the cowpea mosaic virus encoded 24K protein synthesized in Escherichia coli. *Virology* **1987**, *159*, 67–75. [[CrossRef](#)]
116. Drenth, J.; Jansonius, J.N.; Koekoek, R.; Swen, H.M.; Wolthers, B.G. Structure of papain. *Nature* **1968**, *218*, 929–932. [[CrossRef](#)]
117. Vernet, T.; Tessier, D.C.; Chatellier, J.; Plouffe, C.; Lee, T.S.; Thomas, D.Y.; Storer, A.C.; Menard, R. Structural and functional roles of asparagine 175 in the cysteine protease papain. *J. Biol. Chem.* **1995**, *270*, 16645–16652. [[CrossRef](#)]
118. Novinec, M.; Lenarcic, B. Papain-like peptidases: Structure, function, and evolution. *Biomol. Concepts* **2013**, *4*, 287–308. [[CrossRef](#)]
119. Gorbalenya, A.E.; Koonin, E.V.; Lai, M.M. Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. *FEBS Lett.* **1991**, *288*, 201–205. [[CrossRef](#)]
120. Guarne, A.; Tormo, J.; Kirchweger, R.; Pfistermueller, D.; Fita, I.; Skern, T. Structure of the foot-and-mouth disease virus leader protease: A papain-like fold adapted for self-processing and eIF4G recognition. *EMBO J.* **1998**, *17*, 7469–7479. [[CrossRef](#)]
121. Russo, A.T.; White, M.A.; Watowich, S.J. The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease. *Structure* **2006**, *14*, 1449–1458. [[CrossRef](#)]
122. Carrington, J.C.; Cary, S.M.; Parks, T.D.; Dougherty, W.G. A second proteinase encoded by a plant potyvirus genome. *EMBO J.* **1989**, *8*, 365–370. [[CrossRef](#)] [[PubMed](#)]
123. Carrington, J.C.; Freed, D.D.; Sanders, T.C. Autocatalytic processing of the potyvirus helper component proteinase in Escherichia coli and in vitro. *J. Virol.* **1989**, *63*, 4459–4463. [[PubMed](#)]

124. Plisson, C.; Drucker, M.; Blanc, S.; German-Retana, S.; Le Gall, O.; Thomas, D.; Bron, P. Structural characterization of HC-Pro, a plant virus multifunctional protein. *J. Biol. Chem.* **2003**, *278*, 23753–23761. [[CrossRef](#)] [[PubMed](#)]
125. Guo, B.; Lin, J.; Ye, K. Structure of the autocatalytic cysteine protease domain of potyvirus helper-component proteinase. *J. Biol. Chem.* **2011**, *286*, 21937–21943. [[CrossRef](#)] [[PubMed](#)]
126. Hasiow-Jaroszewska, B.; Fares, M.A.; Elena, S.F. Molecular evolution of viral multifunctional proteins: The case of potyvirus HC-Pro. *J. Mol. Evol.* **2014**, *78*, 75–86. [[CrossRef](#)] [[PubMed](#)]
127. Agranovsky, A. Closteroviruses: Molecular biology, evolution and interactions with cells. In *Plant Viruses: Evolution and Management*; Springer: Singapore, 2016; pp. 231–252.
128. Agranovsky, A.A.; Koonin, E.V.; Boyko, V.P.; Maiss, E.; Frotschl, R.; Lunina, N.A.; Atabekov, J.G. Beet yellows closterovirus: Complete genome structure and identification of a leader papain-like thiol protease. *Virology* **1994**, *198*, 311–324. [[CrossRef](#)]
129. Zinovkin, R.A.; Erokhina, T.N.; Lesemann, D.E.; Jelkmann, W.; Agranovsky, A.A. Processing and subcellular localization of the leader papain-like proteinase of Beet yellows closterovirus. *J. Gen. Virol.* **2003**, *84*, 2265–2270. [[CrossRef](#)]
130. Peng, C.W.; Napuli, A.J.; Dolja, V.V. Leader proteinase of beet yellows virus functions in long-distance transport. *J. Virol.* **2003**, *77*, 2843–2849. [[CrossRef](#)]
131. Peremyslov, V.V.; Hagiwara, Y.; Dolja, V.V. Genes required for replication of the 15.5-kilobase RNA genome of a plant closterovirus. *J. Virol.* **1998**, *72*, 5870–5876.
132. Liu, Y.P.; Peremyslov, V.V.; Medina, V.; Dolja, V.V. Tandem leader proteases of Grapevine leafroll-associated virus-2: Host-specific functions in the infection cycle. *Virology* **2009**, *383*, 291–299. [[CrossRef](#)]
133. Peng, C.W.; Peremyslov, V.V.; Mushegian, A.R.; Dawson, W.O.; Dolja, V.V. Functional specialization and evolution of leader proteinases in the family Closteroviridae. *J. Virol.* **2001**, *75*, 12153–12160. [[CrossRef](#)]
134. Kang, S.H.; Atallah, O.O.; Sun, Y.D.; Folimonova, S.Y. Functional diversification upon leader protease domain duplication in the Citrus tristeza virus genome: Role of RNA sequences and the encoded proteins. *Virology* **2018**, *514*, 192–202. [[CrossRef](#)] [[PubMed](#)]
135. Makarova, K.S.; Aravind, L.; Koonin, E.V. A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and Chlamydia pneumoniae. *Trends Biochem. Sci.* **2000**, *25*, 50–52. [[CrossRef](#)]
136. Bailey-Elkin, B.A.; van Kasteren, P.B.; Snijder, E.J.; Kikkert, M.; Mark, B.L. Viral OTU deubiquitinases: A structural and functional comparison. *PLoS Pathog.* **2014**, *10*, e1003894. [[CrossRef](#)] [[PubMed](#)]
137. Bransom, K.L.; Dreher, T.W. Identification of the essential cysteine and histidine residues of the turnip yellow mosaic virus protease. *Virology* **1994**, *198*, 148–154. [[CrossRef](#)] [[PubMed](#)]
138. Rozanov, M.N.; Drugeon, G.; Haenni, A.L. Papain-like proteinase of turnip yellow mosaic virus: A prototype of a new viral proteinase group. *Arch. Virol.* **1995**, *140*, 273–288. [[CrossRef](#)] [[PubMed](#)]
139. Lombardi, C.; Ayach, M.; Beaurepaire, L.; Chenon, M.; Andreani, J.; Guerois, R.; Jupin, I.; Bressanelli, S. A compact viral processing proteinase/ubiquitin hydrolase from the OTU family. *PLoS Pathog.* **2013**, *9*, e1003560. [[CrossRef](#)]
140. Jakubiec, A.; Drugeon, G.; Camborde, L.; Jupin, I. Proteolytic processing of Turnip yellow mosaic virus replication proteins and functional impact on infectivity. *J. Virol.* **2007**, *81*, 11402–11412. [[CrossRef](#)]
141. Chenon, M.; Camborde, L.; Cheminant, S.; Jupin, I. A viral deubiquitylating enzyme targets viral RNA-dependent RNA polymerase and affects viral infectivity. *EMBO J.* **2012**, *31*, 741–753. [[CrossRef](#)]
142. Van Kasteren, P.B.; Bailey-Elkin, B.A.; James, T.W.; Ninaber, D.K.; Beugeling, C.; Khajehpour, M.; Snijder, E.J.; Mark, B.L.; Kikkert, M. Deubiquitinase function of arterivirus papain-like protease 2 suppresses the innate immune response in infected host cells. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E838–E847. [[CrossRef](#)]
143. Akutsu, M.; Ye, Y.; Virdee, S.; Chin, J.W.; Komander, D. Molecular basis for ubiquitin and ISG15 cross-reactivity in viral ovarian tumor domains. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2228–2233. [[CrossRef](#)] [[PubMed](#)]
144. Capodagli, G.C.; McKercher, M.A.; Baker, E.A.; Masters, E.M.; Brunzelle, J.S.; Pegan, S.D. Structural analysis of a viral ovarian tumor domain protease from the Crimean-Congo hemorrhagic fever virus in complex with covalently bonded ubiquitin. *J. Virol.* **2011**, *85*, 3621–3630. [[CrossRef](#)] [[PubMed](#)]
145. James, T.W.; Frias-Staheli, N.; Bacik, J.P.; Levingston Macleod, J.M.; Khajehpour, M.; Garcia-Sastre, A.; Mark, B.L. Structural basis for the removal of ubiquitin and interferon-stimulated gene 15 by a viral ovarian tumor domain-containing protease. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2222–2227. [[CrossRef](#)] [[PubMed](#)]

146. Lawrence, D.M.; Rozanov, M.N.; Hillman, B.I. Autocatalytic processing of the 223-kDa protein of blueberry scorch carlavirus by a papain-like proteinase. *Virology* **1995**, *207*, 127–135. [[CrossRef](#)] [[PubMed](#)]
147. Mann, K.S.; Walker, M.; Sanfaçon, H. Identification of cleavage sites recognized by the 3C-like cysteine protease within the two polyproteins of strawberry mottle virus. *Front. Microbiol.* **2017**, *8*, 745. [[CrossRef](#)] [[PubMed](#)]
148. Chandra, N.R.; Prabu, M.M.; Suguna, K.; Vijayan, M. Structural similarity and functional diversity in proteins containing the legume lectin fold. *Protein Eng.* **2001**, *14*, 857–866. [[CrossRef](#)]
149. Fujinaga, M.; Cherney, M.M.; Oyama, H.; Oda, K.; James, M.N. The molecular structure and catalytic mechanism of a novel carboxyl peptidase from *Scytalidium lignicolum*. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 3364–3369. [[CrossRef](#)]
150. Kataoka, Y.; Takada, K.; Oyama, H.; Tsunemi, M.; James, M.N.; Oda, K. Catalytic residues and substrate specificity of scytalidoglutamic peptidase, the first member of the eqolisin in family (G1) of peptidases. *FEBS Lett.* **2005**, *579*, 2991–2994. [[CrossRef](#)]
151. Kondo, M.Y.; Okamoto, D.N.; Santos, J.A.; Juliano, M.A.; Oda, K.; Pillai, B.; James, M.N.; Juliano, L.; Gouvea, I.E. Studies on the catalytic mechanism of a glutamic peptidase. *J. Biol. Chem.* **2010**, *285*, 21437–21445. [[CrossRef](#)]
152. Pillai, B.; Cherney, M.M.; Hiraga, K.; Takada, K.; Oda, K.; James, M.N. Crystal structure of scytalidoglutamic peptidase with its first potent inhibitor provides insights into substrate specificity and catalysis. *J. Mol. Biol.* **2007**, *365*, 343–361. [[CrossRef](#)]
153. Sasaki, H.; Kubota, K.; Lee, W.C.; Ohtsuka, J.; Kojima, M.; Iwata, S.; Nakagawa, A.; Takahashi, K.; Tanokura, M. The crystal structure of an intermediate dimer of aspergilloglutamic peptidase that mimics the enzyme-activation product complex produced upon autoproteolysis. *J. Biochem.* **2012**, *152*, 45–52. [[CrossRef](#)]
154. Wintermantel, W.M. Transmission efficiency and epidemiology of criniviruses. In *Bemisia: Bionomics and Management of a Global Pest*; Springer: Dordrecht, The Netherlands; Heidelberg, Germany; London, UK; New York, NY, USA, 2009; pp. 319–331.
155. Kiss, Z.A.; Medina, V.; Falk, B. Crinivirus replication and host interactions. *Front. Microbiol.* **2013**, *4*, 99. [[CrossRef](#)] [[PubMed](#)]
156. Agranovsky, A.A.; Lesemann, D.E.; Maiss, E.; Hull, R.; Atabekov, J.G. “Rattlesnake” structure of a filamentous plant RNA virus built of two capsid proteins. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2470–2473. [[CrossRef](#)]
157. Killiny, N.; Harper, S.; Alfaress, S.; El Mohtar, C.; Dawson, W. Minor coat and heat-shock proteins are involved in binding of citrus tristeza virus to the foregut of its aphid vector, *Toxoptera citricida*. *Appl. Environ. Microbiol.* **2016**. [[CrossRef](#)]
158. Stewart, L.R.; Medina, V.; Tian, T.; Turina, M.; Falk, B.W.; Ng, J.C. A mutation in the Lettuce infectious yellows virus minor coat protein disrupts whitefly transmission but not in planta systemic movement. *J. Virol.* **2010**, *84*, 12165–12173. [[CrossRef](#)]
159. Krupovic, M.; Blomberg, J.; Coffin, J.M.; Dasgupta, I.; Fan, H.; Geering, A.D.; Gifford, R.; Harrach, B.; Hull, R.; Johnson, W.; et al. Ortervirales: New Virus Order Unifying Five Families of Reverse-Transcribing Viruses. *J. Virol.* **2018**, *92*, e00515-18. [[CrossRef](#)] [[PubMed](#)]
160. Hayward, A. Origin of the retroviruses: When, where, and how? *Curr. Opin. Virol.* **2017**, *25*, 23–27. [[CrossRef](#)] [[PubMed](#)]
161. Dunn, B.M.; Goodenow, M.M.; Gustchina, A.; Wlodawer, A. Retroviral proteases. *Genome Biol.* **2002**, *3*, 3006.1–3006.7. [[CrossRef](#)]
162. Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B.K.; Baldwin, E.; Weber, I.T.; Selk, L.M.; Clawson, L.; Schneider, J.; Kent, S. Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease. *Science* **1989**, *245*, 616–621. [[CrossRef](#)] [[PubMed](#)]
163. Navia, M.A.; Fitzgerald, P.M.; McKeever, B.M.; Leu, C.T.; Heimbach, J.C.; Herber, W.K.; Sigal, I.S.; Darke, P.L.; Springer, J.P. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **1989**, *337*, 615–620. [[CrossRef](#)]
164. Torruella, M.; Gordon, K.; Hohn, T. Cauliflower mosaic virus produces an aspartic proteinase to cleave its polyproteins. *EMBO J.* **1989**, *8*, 2819–2825. [[CrossRef](#)] [[PubMed](#)]
165. Champagne, J.; Benhamou, N.; Leclerc, D. Localization of the N-terminal domain of cauliflower mosaic virus coat protein precursor. *Virology* **2004**, *324*, 257–262. [[CrossRef](#)] [[PubMed](#)]

166. Karsies, A.; Merkle, T.; Szurek, B.; Bonas, U.; Hohn, T.; Leclerc, D. Regulated nuclear targeting of cauliflower mosaic virus. *J. Gen. Virol.* **2002**, *83*, 1783–1790. [[CrossRef](#)]
167. Luna, G.R.; Peña, E.J.; Borniego, M.B.; Heinlein, M.; García, M.L. Citrus psorosis virus movement protein contains an aspartic protease required for autocleavage and the formation of tubule-like structures at plasmodesmata. *J. Virol.* **2018**, *92*, e00355-18. [[CrossRef](#)] [[PubMed](#)]
168. Kondo, H.; Hirano, S.; Chiba, S.; Andika, I.B.; Hirai, M.; Maeda, T.; Tamada, T. Characterization of burdock mottle virus, a novel member of the genus Benyvirus, and the identification of benyvirus-related sequences in the plant and insect genomes. *Virus Res.* **2013**, *177*, 75–86. [[CrossRef](#)] [[PubMed](#)]
169. Pascon, R.C.; Kitajima, J.P.; Breton, M.C.; Assumpcao, L.; Greggio, C.; Zanca, A.S.; Okura, V.K.; Alegria, M.C.; Camargo, M.E.; Silva, G.G.; et al. The complete nucleotide sequence and genomic organization of Citrus Leprosis associated Virus, Cytoplasmatic type (CiLV-C). *Virus Genes* **2006**, *32*, 289–298. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the by Her Majesty the Queen in Right of Canada as Represented by the Minister of Agriculture and Agri-Food Canada. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).