Matrix metalloproteinases outside vertebrates

Laura Marino-Puertas, Theodoros Goulas* and F. Xavier Gomis-Rüth*

Proteolysis Lab; Structural Biology Unit; "María-de-Maeztu" Unit of Excellence; Molecular Biology Institute of Barcelona (CSIC); Barcelona Science Park; c/Baldiri Reixac, 15-21; 08028 Barcelona (Spain).

* Corresponding authors: e-mail: thgcri@ibmb.csic.es, phone: (+34) 934 020 187 or e-mail: xgrcri@ibmb.csic.es, phone: (+34) 934 020 186.

ABSTRACT

The matrix metalloproteinase (MMP) family belongs to the metzincin clan of zinc-dependent metallopeptidases. Due to their enormous implications in physiology and disease, MMPs have mainly been studied in vertebrates. They are engaged in extracellular protein processing and degradation, and present extensive paralogy, with 23 forms in humans. One characteristic of MMPs is a ~165-residue catalytic domain (CD), which has been structurally studied for 14 MMPs from human, mouse, rat, pig and the oral-microbiome bacterium *Tannerella forsythia*. These studies revealed close overall coincidence and characteristic structural features, which distinguish MMPs from other metzincins and give rise to a sequence pattern for their identification. Here, we reviewed the literature available on MMPs outside vertebrates and performed database searches for potential MMP CDs in invertebrates, plants, fungi, viruses, protists, archaea and bacteria. These and previous results revealed that MMPs are widely present in several copies in Eumetazoa and higher plants (*Tracheophyta*), but have just token presence in eukaryotic algae. A few dozen sequences were found in *Ascomycota* (within fungi) and in double-stranded DNA viruses infecting invertebrates (within viruses). In contrast, a few hundred sequences were found in archaea and >1000 in bacteria, with several copies for some species. Most of the archaeal and bacterial phyla containing potential MMPs are present in human oral and gut microbiomes. Overall, MMP-like sequences are present across all kingdoms of life, but their asymmetric distribution contradicts the vertical descent model from a eubacterial or archaeal ancestor.

Keywords: zinc metalloproteinase; metzincin; MMP; invertebrates; catalytic domain; structure-based sequence motif

---

1. Molecular characteristics of matrix metalloproteinases — The matrix metalloproteinases (MMPs) are a widespread family of zinc-dependent metallopeptidases (MPs), which either broadly degrade extracellular matrix components or selectively activate or inactivate other proteins through limited proteolysis [1-3]. MMPs were discovered in 1962 as an active principle in frog metamorphosis [4], and they contain a central zinc-dependent catalytic domain (CD) of ~165 residues, which is mostly furnished at its N-terminus with a ~20-residue signal peptide for secretion and an ~80-residue zymogenic pro-domain (PD). Some MMPs possess a "furin recognition motif" (R-X-R/K-R; [5]) for proteolytic activation between the PD and the CD. Into this minimal configuration, distinct MMPs have inserted extra segments and domains, such as fibronectin-type-II inserts within the CD and C-terminal unstructured linker regions, ~200-residue hemopexin domains (HDs) and other domains, as well as glycosyl phosphatidylinositol (GPI) anchors and transmembrane segments [1, 3, 6]. Overall, this variability divides human MMPs into subfamilies: archetypal MMPs, gelatinases, matrilysins and furin-activatable MMPs, which include membrane-type MMPs [6, 7].

MMPs belong to the metzincin clan of MPs, which currently comprises twelve structurally characterized families with common features [8-12]. However, this structural similarity is not reflected at sequence level, because

pairwise identities are typically <20%, which is below the twilight zone of relevant sequence similarity (25–35%; [13]). Overall, metzincins share a globular CD of ~130—270 residues, which is split into a structurally conserved N-terminal sub-domain (NTS) and a diverging C-terminal sub-domain (CTS) by a horizontal active-site cleft. This cleft contains the catalytic zinc ion and accommodates protein or peptide substrates for catalysis (Figure 1A). NTSs contain a mostly five-stranded β-sheet, whose strands (βI-βV) parallel the active-site cleft. The lowermost strand βIV shapes the upper rim of the cleft and runs antiparallel to the other strands and to a bound substrate, thus establishing inter-main-chain interactions to fix it. NTSs also possess a "backing helix" (αA) with structural functions and an "active-site helix" (αB) with an extended "zinc-binding motif", **H**-E-X-X-**H**-X-X-G/N-X-X-**H**/**D**-Φ. This motif includes the general base/acid glutamate required for catalysis and three metal-binding protein residues (in **bold**), as well as a "family-specific residue" (Φ; [8-10]). Some families further comprise an additional "adamalysin helix", which was first identified in the adamalysin/ADAM family [12, 14]. In contrast to this structural conservation, CTSs largely deviate in length and structure but share the "Met-turn" [15], a loop with a methionine placed just below the catalytic zinc, and a "C-terminal helix" (αC) that terminates the CD. Into this common minimal scaffold, specific molecular elements decorate each metzincin family in the form of loops, ion-binding sites, additional regular secondary structure elements, extra domains, etc. [8-12].

To date, structural studies on MMP CDs are restricted to mammals (human, mouse, rat and pig), with the notable exception of bacterial karilysin ([12]; see also Section 6). These structures revealed that MMPs are very similar (Figure 1B), have much longer NTSs (~127 residues) than CTSs (~37 residues), and include an "S-shaped double loop", which connects the third and fourth strands of the NTS β-sheet (Figure 1A). This loop binds essential structural zinc and calcium cations, after which the polypeptide forms a prominent bulge ("bulge-edge segment") that protrudes into the active-site groove and assists in substrate specificity. A further hallmark of MMPs is a second calcium site within the NTS and that the family-specific residue is a serine or threonine [3, 10]. This residue makes a strong hydrogen bond between side chains with the first of two consecutive aspartates at the beginning of helix αC (Figure 1C). This aspartate is also engaged in fixing the N-terminus of the mature CD upon proteolytic activation. In turn, the second aspartate participates in a double electrostatic interaction with main-chain atoms of the Met-turn (Figure 1A,C). This intricate electrostatic network is structured around the "connecting segment", which links the third zinc-binding histidine with the Met-turn methionine and spans only seven residues; thus, it is the shortest of the metzincins [10]. Further typical of MMPs are two tyrosines, respectively four positions downstream of the Met-turn methionine and featuring the penultimate residue of the CD. Finally, the Met-turn and helix αC are connected by a loop subdivided into the "S_1'-wall-forming segment" and the downstream "specificity loop" (Figure 1A). Taken together, all these characteristic structural features yield an extended sequence pattern, H-E-2X-H-2X-G-2X-H-S/T-6X-M-3X-Y-9X-D-D-7X-Y-X (Figure 1D), which distinguishes the C-terminal segment of MMP CDs from other metzincin families and MPs in general [12].

In addition to the CD, a further hallmark of MMPs is that latency is maintained through the PD [3], which together with transcriptional regulation and dedicated protein inhibitors regulates physiological MMP activity [16]. In most MMPs, zymogenicity is exerted by a "cysteine-switch" mechanism centered on a cysteine within a consensus motif at the end of the PD, P-R-C-G-X-P-D [17], which is found in all human MMPs except MMP-23B. Structurally, the polypeptide chain spanning these residues shields the active-site cleft by adopting a U-shaped conformation mediated by a double salt bridge between the arginine-aspartate pair and by the glycine, whose missing side chain prevents collision with upper-rim strand βIV [3]. In addition, the cysteine Sγ atom coordinates and thus blocks the catalytic zinc (see Fig. 3B in [3]). MMP activation entails PD removal to free access to the cleft.

2. Matrix metalloproteinases in invertebrates — MMPs have been extensively studied in vertebrates, particularly in mammals, due to their widespread implications in human health and disease [6, 16, 18, 19]. However, they are also widely present in invertebrates, where they participate in tissue turnover processes such as dendritic remodeling, regeneration, tracheal growth, axon guidance, histolysis and matrix degradation during hatching. They also participate in cancer processes [20]. Studied organisms belong to the phyla Arthropoda, including insects (dipterans, lepidopterans, hymenopterans and coleopterans) and crustaceans; Echinoderma, including sea urchins and sea cucumbers; Mollusca, including mussels and snails; Nematoda, Annelida and Planaria; invertebrate Chordata, including sea squirt and lancelet; and Cnidaria, including hydra and jellyfish (see Table 1). In addition, we found MMP-like sequences in the genomes of organisms as primitive as starlet sea anemone (*Nematostella vectensis*; UniProt database [www.uniprot.org] access code [UP] A7RJ22), which is also from the phylum Cnidaria, and sea gooseberry (*Pleurobrachia pileus;* GenBank database [www.ncbi.nlm.nih.gov/genbank] access code [GB] FQ005948) from the phylum Ctenophora, i.e. MMPs are present across the subkingdom Eumetazoa. In contrast, MMPs are apparently absent from the genomes of the sponge *Amphimedon queenslandica* (phylum Porifera; [21]) and of *Trichoplax adherens* (phylum Placozoa; [22]), which belong to the basalmost clades of metazoans.

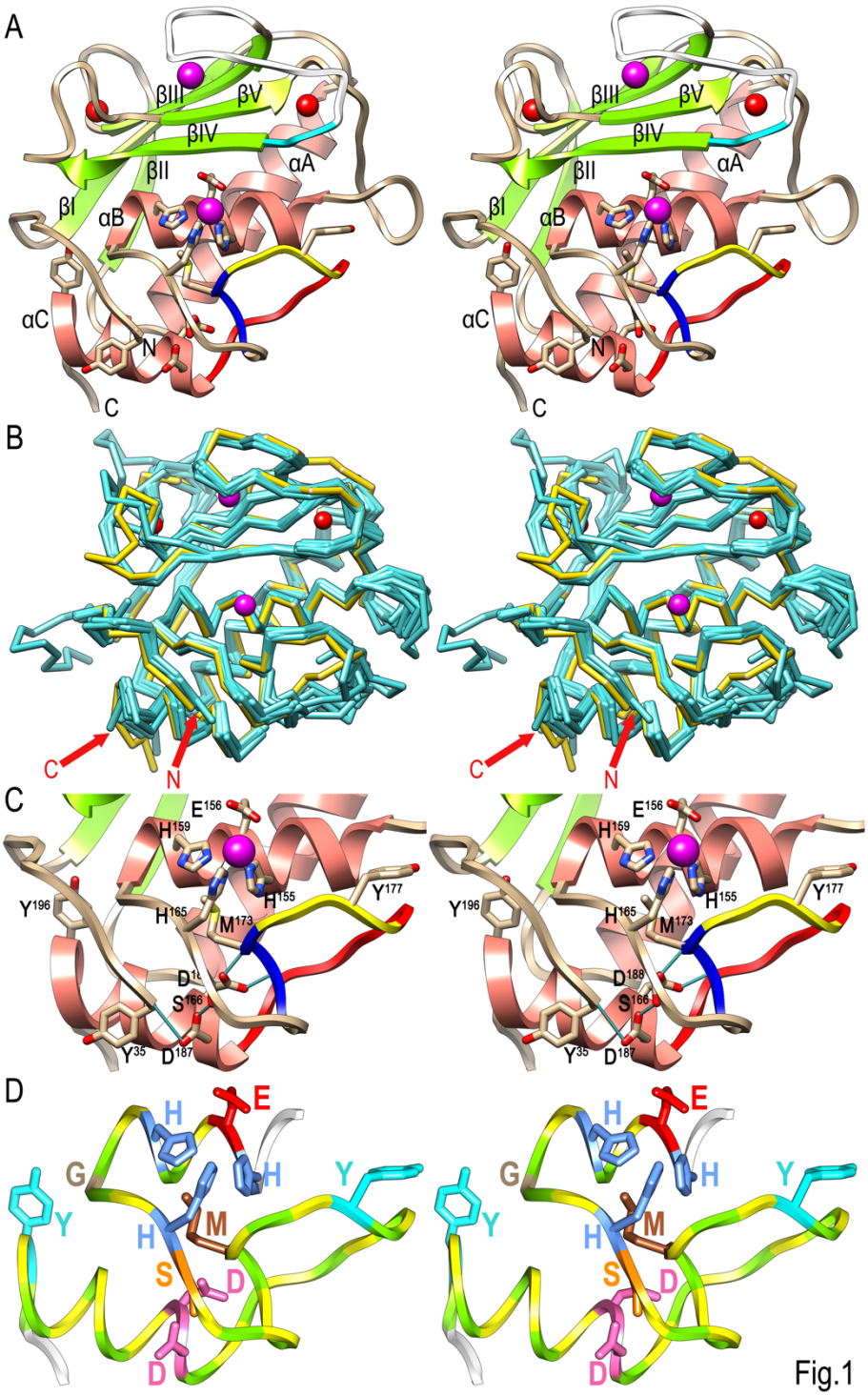


Fig.1

Figure 1 — MMP catalytic domain structure. **(A)** Cross-eye stereographic Richardson-plot of the catalytic domain of *T. forsythia* karilysin (Y^{35}-P^{200}; PDB 2XS3; [23]) in standard orientation [24] as a representative of MMP catalytic domains (see [3]). Green arrows represent β-strands (labeled βI-βV), brown ribbons stand for α-helices (αA-αC). The two zinc cations found across MMPs are shown as magenta spheres, the calcium cations found in most MMPs but missing in karilysin are depicted in red. Relevant chain segments are shown in distinct colors and labeled (Met-turn in blue, specificity loop in red, S_1'-wall-forming segment in yellow, S-loop in white, and bulge-edge segment in cyan). The side chains of the zinc-binding histidines, the general base/acid glutamate, the Met-turn methionine, the family-specific serine, and the two tyrosines and aspartates further included in the extended signature characteristic of MMPs are shown as stick models. **(B)** Superposition of the Cα-traces of the catalytic domains of bacterial karilysin (in yellow) and human MMP-1 (PDB 966C; [25]), MMP-3 (PDB 1CIZ; [26]), MMP-7 (PDB 1MMQ; [27]), MMP-8 (PDB 1JAN; [28]), MMP-9 (4H3X; [29]), MMP-10 (PDB 1Q3A; [30]), MMP-11 (PDB 1HV5; [31]), MMP-12 (PDB 1Y93; [32]), MMP-13 (PDB 2D1N; [33]) and MMP-16 (PDB 1RM8; [34]), all in cyan. The two consensus zinc and calcium cations are shown as magenta and red spheres, respectively; the consensus N- and C-termini are indicated by red arrows. **(C)** Close-up view of (A) highlighting the structural features of the C-terminal sub-domain characteristic for MMPs. The hallmark electrostatic interactions within the CTS are shown as cyan lines, residues included in the extended sequence pattern are depicted for their side chains and labeled. **(D)** Color-coded mapping of the MMP extended sequence pattern (H-E-2X-H-2X-G-2X-H-S/T-6X-M-3X-Y-9X-D-D-7X-Y-X) onto the karilysin polypeptide chain. Each residue type is shown in one color and labeled, random residues (X) are alternatively in light green and yellow.

Eumetazoans are animals comprised of tissues organized into germ layers, with a gastrula stage during embryogenesis. Primitive animals of this lineage up to the emergence of the clade Bilateria likely had few MMP copies within their genomes, as currently found in several *Drosophila* species (two copies) and other insects such as red flour beetle, silkworm or malaria mosquito (three copies) [6, 20, 35-38]. In contrast, higher animals show gene polyplication, which likely arose during early vertebrate evolution and led to the substantial paralogy that is currently found, e.g., in humans (24 genes), mice (23 genes) and zebrafish (26 genes). However, development of this complexity was not linear: while seven genes are found in the sea squirt *Ciona intestinalis* and nine in the mosquito *Aedes aegypti*, the sea urchin *Strongylocentrotus purpuratus* has 26 [6, 20, 39-44].

Table 1 — MMPs reported from invertebrates

Phylum *Annelida*	
Class *Polychaeta*	
Boneworm (*Osedax japonicus*)	[45]
Phylum *Arthropoda*	
Class *Branchiopoda*	
Water flea (*Daphnia pulex*)	[46]
Class *Insecta*	
Yellow-fever mosquito (*Aedes aegypti*)	[44]
Malaria mosquito (*Anopheles gambiae*)	[47]
Honeybee (*Apis mellifera*)	[48]
Silkworm (*Bombyx mori*)	[49]
Fruitfly (*Drosophila melanogaster*)	[20, 35, 36])
Greater wax moth (*Galleria mellonella*)	[50]
Tobacco hornworm (*Manduca sexta*)	[51]
Red flour beetle (*Tribolium castaneum*)	[37]
Cabbage looper (*Trichoplusia ni*)	[52]
Phylum *Chordata*	
Class *Ascidiacea*	
Sea squirt (*Ciona intestinalis*)	[41]
Class *Leptocardii*	
Lancelet (*Branchiostoma japonicum*)	[53]
Phylum *Cnidaria*	
Class *Hydrozoa*	
Common hydra (*Hydra vulgaris* / *Hydra magnipapillata*)	[54, 55]
Class *Scyphozoa*	
Nomura's jellyfish (*Nemopilema nomurai*)	[56]
Phylum *Echinodermata*	
Class *Echinoidea*	
Asian sea urchin (*Hemicentrotus pulcherrimus*)	[57]
Mediterranean purple sea urchin (*Paracentrotus lividus*)	[58]
Pacific purple sea urchin (*Strongylocentrotus purpuratus*)	[40]
Class *Holothuroidea*	
Japanese sea cucumber (*Apostichopus japonicus*)	[59]
Rock sea cucumber (*Holothuria glaberrima*)	[60]
Phylum *Mollusca*	
Class *Bivalvia*	
American oyster (*Crassostrea virginica*)	[61]
Mediterranean mussel (*Mytilus galloprovincialis*)	[62]
Class *Gasteropoda*	
Giant African snail (*Achatina fulica*)	[63]
Bloodfluke planorb (*Biomphalaria glabrata*)	[64]
Many-colored abalone (*Haliotis diversicolor*)	[65]
Red abalone (*Haliotis rufescens*)	[66]
Korean common dogwhelk (*Thais clavigera*)	[67]
Phylum *Nematoda*	
Class *Secernentea*	
Rat lungworm (*Angiostrongylus cantonensis*)	[68]
Roundworm (*Caenorhabditis elegans*)	[69-71]
Yellow potato cyst nematode (*Globodera rostochiensis*)	[72]
Parasitic nematode (*Gnathostoma spinigerum*)	[73]
Soybean cyst nematode (*Heterodera glycines*)	[72]
Phylum *Platyhelminthes*	
Class *Rhabditophora*	
Freshwater planarian flatworm (*Dugesia japonica*)	[74]
Freshwater planarian flatworm (*Schmidtea mediterranea*)	[74]

Taxonomy according to the Catalogue of Life (http://www.catalogueoflife.org/col; [75]).

The complex evolution of MMPs in Eumetazoa is also reflected by the presence of several ancillary domains in the full-length enzymes (see Section 1 and Fig. 1 in [6]). It can be speculated that primitive MMPs consisted of standalone CDs or possibly PD+CD tandems, which underwent duplication, gene fusion and exon shuffling to result in multi-domain architectures [6, 39, 76, 77]. However, in some instances evolution progressed in the opposite direction, i.e. multi-domain enzymes underwent truncation to yield proteins of fewer domains, even in mammals. This is the case for matrilysins, which span only a signal peptide, a PD with a cysteine-switch motif and a CD [6, 78, 79]. This minimal architecture is also predominant across invertebrates [6], although many MMP sequences in insects—e.g. in *Drosophila* [20]—further comprise a furin recognition sequence, an HD and a GPI anchor.

Table 2 — MMPs referenced from plants and selected sequences within genomes

Phylum *Tracheophyta*	
Class *Liliopsida*	
Barley (*Hordeum vulgare*): F2DC11, F2D2W1, F2DGF5, F2D593	[81]
Rice (*Oryza sativa*): A2X9G3, A2YB49, A2ZA53, A2ZA54, A2ZA55, A2ZA56, A2ZA53, A2ZA53	[81]
Sugar cane (*Saccharum hybrid cultivar*): A0A059Q041	[82]
Corn (*Zea mays*): B6U4D1, B4FZU1	[81]
Eelgrass (*Zostera marina*): A0A0K9P9J4*, A0A0K9PG73*, A0A0K9PXQ4*, A0A0K9PK10*, A0A0K9PLB4*, A0A0K9PY38*, A0A0K9PBG9*, A0A0K9P805*	
Class *Magnoliopsida*	
Lyre-leaved rock cress (*Arabidopsis lyrata*): D7LBU9*	
Mouse-ear cress (*Arabidopsis thaliana*): At1/2/3/4/5-MMP (O23507, O04529, Q5XF51, Q8GWW6, Q9ZUJ5)	[83-88]
Sugar beet (*Beta vulgaris*): A0A0J8F9H7, A0A0J8B8R0, GB XP_010692448, GB XP_010695895	[89]
Cucumber (*Cucumis sativus*): Cs1-MMP (Q9LEL9)	[90]
Soybean (*Glycine max*): Gm1/2-MMP, Gm-Slti114-MMP (C6TNN5, Q93Z89, B6CAM2) and C6TNN5	[81, 91-96]
Tree cotton (*Gossypium arboreum*): A0A0B0ME77*	
Barrel medic (*Medicago truncatula*): Mt1-MMP (Q9ZR44)	[97]
Tobacco (*Nicotiana tabacum*): Nt1-MMP (C3PTL6)	[98, 99]
Western balsam poplar (*Populus trichocarpa*): B9I3X8*	
Castor bean (*Ricinus communis*): B9RUG7*	
Tomato (*Solanum lycopersicum*): Sl1/2/3/4/5-MMP (I7JCM3, I7KJ40, K4BWG3, K4CNL4, A0A0G3ZAU2, K4CYZ6)	[81]
Cocoa (*Theobroma cacao*): A0A061DVV2*, S1S448*	
Class *Pinopsida*	
Loblolly pine (*Pinus taeda*): Pta1-MMP (B7TVN4)	[100]

All sequence codes are from UniProt (UP; www.uniprot.org) or GenBank (GB; www.ncbi.nlm.nih.gov/genbank).
Taxonomy according to the Catalogue of Life (http://www.catalogueoflife.org/col; [75]).
Nomenclature of validated plant MMPs according to [101].
* Sequences annotated in UniProt as "matrix metalloproteinases" and manually curated.
Searches completed on 10 January, 2017.

3. Matrix metalloproteinases in plants and algae — In addition to Eumetazoa, MMPs have been reported from higher plants (phylum Tracheophyta), where they are generally present in lower copy numbers than in animals. Enzymes were described from soybean, mouse-ear cress, cucumber, barrel medic, tobacco, loblolly pine and tomato (see Table 2). In addition, sequences were referenced from sugar beet, rice, corn, sugar cane and barley (Table 2). Moreover, MP activity attributable to an MMP was also described for the jack bean *Canavalia ensiformis*, though validation is still pending [80]. Finally, current sequence similarity searches identified several hundreds of potential hits in higher plants, Table 2 provides a curated selection of them.

Plant MMPs localize to the plasma membrane or the extracellular space and have been found to be involved in remodeling of the extracellular matrix during plant growth and development processes, such as germination, programmed cell death and senescence, as well as in biotic and abiotic stress responses [86, 101-103]. Sequence analyses revealed that plant MMPs have a homogenous domain distribution and mostly comprise a signal peptide, a PD with a cysteine-switch motif that occasionally diverges from the consensus (see Section 1 and [102]), and a CD. Within the latter, some sequences have the general base/acid glutamate mutated to glutamine [102], a change that in mammalian MMPs leads to ablation or strong reduction of proteolytic activity [104]. Uniquely, plant MMPs

encompass a specific consensus sequence, D-L-E-S/T, two residues upstream of the zinc-binding motif [81, 84, 101]. In addition, the loop that connects strand βV with helix αB is up to 10 residues longer in plant MMPs than in mammalian counterparts. In the absence of experimental structures, these two combined features point to a specific structural element of plant MMPs, putatively an extra cation-binding site. Downstream of the CD, plant MMPs only contain ~40-residue C-terminal GPI anchors or transmembrane segments for localization to the plasma membrane [81].

The unicellular green alga *Chlamydomonas reinhardtii* encodes two MPs dubbed gametolysins (also known as gamete lytic enzymes, MMP1 and MMP2, and autolysins), which are engaged in cell-wall turnover and have been recurrently associated with the MMP family [101, 102, 105-108]. Similarly to MMPs, these ~635-residue enzymes comprise a signal peptide and a putative PD with a cysteine-switch-like motif. These enzymes are similar to four enzymes from the multicellular green alga *Volvox carteri* (VMP1-VMP4), in which the first zinc-binding histidine is replaced with glutamine [109-111]. Current sequence similarity searches identified new potential paralogues of these MPs in *C. reinhardtii* and *V. carteri*, as well as in the green algae *Gonium pectorale* and *Chlorella variabilis.* However, in the absence of three-dimensional structures, the primary sequences of these MPs deviate from the extended sequence pattern of MMPs (see Section 1). Consequently, they belong to a separate metzincin family, which was tentatively dubbed gametolysins in the past [10, 12, 108]. This is consistent with their adscription to a family separate from MMPs within the MEROPS peptidase database (M11 *vs.* M10; see merops.sanger.ac.uk; [112]). Of note, searches for true MMPs within eukaryotic algae revealed six potential relatives, two in phytoplankton *Emiliania huxleyi* (UP R1DV14 and R1E2E0/R1EHE7) and one each in *G. pectorale* (UP A0A150GPR5), *Symbiodinium minutum* (KEGG Gene symbB.v1.2.029330.t2; see www.genome.jp), *Aureococcus anophagefferens* (UP F0Y382) and *V. carteri* (UP D8UD00). This restricted presence of MMPs within algae is consistent with recent genome-wide analyses of the secretomes of nine brown algae belonging to the phyllum Phaeophyceae, which revealed no potential MMP ortholog [113]. Furthermore, no other sequences were presently found in other Protista/Chromista.

4. Matrix metalloproteinases in fungi — Although some MPs were described from fungi [114-116], none has yet been confirmed to be an MMP. We thus performed a database search, which revealed several potential sequences in the phylum *Ascomycota* but not in other phyla (Table 3). These hypothetical proteins span between 255 and 659 residues, some have their catalytic glutamate replaced with glutamine (see Sections 1 and 3), and most comprise N-terminal extensions with a potential cysteine-switch-like motif. Those lacking this motif possess potential NTSs that are significantly shorter than in standard MMPs, so they may correspond to incomplete sequences or truncated variants. Several sequences also contain large N- and/or C-terminal extensions that could correspond to additional domains. Hit fungal species contain only one sequence each, with the exception of *Arthrobotrys oligospora* and *Dactylellina haptotyla*, with five sequences each, and all but two species are fungal pathogens of higher plants, nematodes, insects and animals—including humans—or endophytic fungi. Accordingly, their lifestyle entails very intimate contact with organisms that contain MMPs.

5. Matrix metalloproteinases in viruses — In 2000, the functional characterization of an MMP from *Xestia c-nigrum* granulovirus was reported [117]. In the absence of structural information, analysis of its amino acid sequence reveals an MMP CD. The upstream N-terminal segment lacks the canonical cysteine-switch motif, but shows sequence stretch C^{42}-G-G-G-N-H-R-R-T-K-R^{52} immediately before the predicted mature N-terminus, which includes a cysteine-glycine pair reminiscent of those motifs, as well as a recognition sequence characteristic of furin-activatable MMPs (see [6] and Section 1). This notwithstanding, the full-length protein was active, thus suggesting it was in an at least partially competent state without hypothetical activation [117]. Threading calculations (see the legend to Figure 2A) with the protein segment downstream of the CD of *Xestia* MMP indicated that it most likely contained an HD (Figure 2A), preceded by an intermediate "threonine-rich region". More recently, *Cydia pomonella* granulovirus was also shown to express a functional MMP, the only other viral family member studied to date [38]. Like the *Xestia* ortholog, it contained a CD preceded by a potential furin-recognition sequence. However, it lacked any cysteine in the N-terminal fragment, so its function and/or activation mechanism might diverge from *Xestia* MMP. In contrast to previous hypotheses [38] but in agreement with the prediction for *Xestia* MMP, threading calculations indicated that a threonine-rich region and an HD are present in the C-terminal part of the protein (Figure 2B).

Granuloviruses belong to the genus *Betabaculovirus* within the family Baculoviridae and all species sequenced to date within this genus contain putative MMP homologs (see Suppl. Table 1 and [38]). In contrast, MMPs are absent from the other Baculoviridae geni, *viz.* Alphabaculovirus, Deltabaculovirus and Gammabaculovirus. In addition to Baculoviridae, we also found MMP sequences in the families Iridoviridae, Nudiviridae, Poxviridae (subfamily Entomopoxvirinae) and Ascoviridae (Suppl. Table 1) but not in any other viruses or viroids. Collectively, all these virus families are double-stranded DNA viruses with no RNA stage, which infect invertebrates (arthropods, lepidopters, hymenopters, diptera and decapods), i.e. organisms that contain MMPs. The potential viral MMPs vary in length, some contain a signal peptide, a potential cysteine-switch-like motif, a

putative furin-cleavage site, a predicted HD and a threonine-rich region as in *Xestia* and *Cydia* MMPs, but others do not.

Table 3 — Fungal MMP sequences

Phylum *Ascomycota*		
Class *Dothideomycetes*		
Bipolaris oryzae	388 residues	GB XP_0077684757 / UP W6ZMI0
Bipolaris victoriae	388 residues	GB XP_014556326 / UP W7EIS0
Bipolaris zeicola	388 residues	GB XP_007709037 / UP W6YG21
Cochliobulus heterostrophus (Bipolaris maydis)	388 residues	GB XP_014075121 / UP N4X7I4, UP M2V381 **
Cochliobulus sativus (Bipolaris sorokiniana)	320 residues	GB XP_007696738 / UP M2TCL8
Paraphaeosphaeria sporulosa	411 residues	GB XP_018039150 / UP A0A177CMM7
Phaeosphaeria nodorum (Parastagonospora nodorum) (Septoria nodorum)	326 residues	GB XP_001794978 / UP Q0UUK1
Pyrenochaeta sp.	565 residues	UP A0A178DJW7
Stagonospora sp.	393 residues	UP A0A178BDB5
Stemphylium lycopersici	563 residues	UP A0A0L1HKC2
Class *Eurotiomycetes*		
Aspergillus calidoustus	290 residues	UP A0A0U5G9D2
Aspergillus flavus	274 residues	GB XP_002379978 / UP B8NI13, UP A0A0D9MRC7 **
Aspergillus lentulus	311 residues	UP A0A0S7DKQ5
Aspergillus udagawae	282 residues	UP A0A0K8L198
Endocarpon pusillum	306 residues	GB XP_007803399 / UP U1HKZ8
Exophiala aquamarina	282 residues	GB XP_013264742 / UP A0A072PQX4
Neosartorya fischeri (Aspergillus fischerianus)	616 residues	GB XP_001261124 / UP A1DIM0
Class *Leotiomycetes*		
Pseudogymnoascus sp.	255 residues	UP A0A094DQI1, UP A0A094ITE6 *
Class *Orbiliomycetes*		
Arthrobotrys oligospora	609 residues	GB XP_011127847 / UP G1XU64
Arthrobotrys oligospora	599 residues	GB XP_011127846 / UP G1XU63
Arthrobotrys oligospora	290 residues	GB XP_011126672/ UP G1XQB0
Arthrobotrys oligospora	315 residues	GB XP_011126207 / UP G1XNZ5
Arthrobotrys oligospora	232 residues	GB XP_011120912 / UP G1X8V0
Drechslerella stenobrocha	210 residues	UP W7IFU4
Class *Sordariomycetes*		
Dactylellina haptotyla	659 residues	GB XP_011116834 / UP S7ZXB5
Dactylellina haptotyla	308 residues	GB XP_011114406 / UP S8A464
Dactylellina haptotyla	292 residues	GB XP_011112095 / UP S8AFF5
Dactylellina haptotyla	268 residues	GB XP_011113422 / UP S8BTL0
Dactylellina haptotyla	246 residues	GB XP_011107496 / UP S8AU63
Metarhizium anisopliae	619 residues	UP A0A0B4G9D2
Metarhizium anisopliae	607 residues	UP A0A0D9NZH8
Metarhizium brunneum	619 residues	GB XP_014544223 / UP A0A0B4FXN2
Metarhizium robertsii	619 residues	GB XP_007825075 / UP E9F9D7, A0A014N9M3 **
Pochonia chlamydosporia	636 residues	GB XP_018148437 / UP A0A179G3B6
Purpureocillium lilacinum	627/629 residues	GB XP_018174930 / UP A0A179GT13, A0A179GDA3 *

All sequence codes are from UniProt (UP; www.uniprot.org) or GenBank (GB; www.ncbi.nlm.nih.gov/genbank).
Sequence searches were performed with structurally validated MMP CDs within UniProt (www.uniprot.org) or the National Center for Biotechnology Information (blast.ncbi.nlm.nih.gov/Blast.cgi) using standard parameters.
Taxonomy according to the Catalogue of Life (http://www.catalogueoflife.org/col; [75]).
* These sequences have only minimal differences.
** These entries are identical.
Searches completed on 19 January, 2017.

Moreover and consistent with the host specificity of the harboring viruses, sequences cluster closely with insect MMPs [38], which possess a similar domain architecture (see Section 2 and [20]). In insects, housekeeping MMPs participate in physiological remodeling of the basal lamina, which lines the midgut to prevent systemic infections [126]. In turn, as part of the infective process, ingested viral particles reach the midgut of target insects and breach the basal lamina, a task that might be carried out by viral MMPs [126].

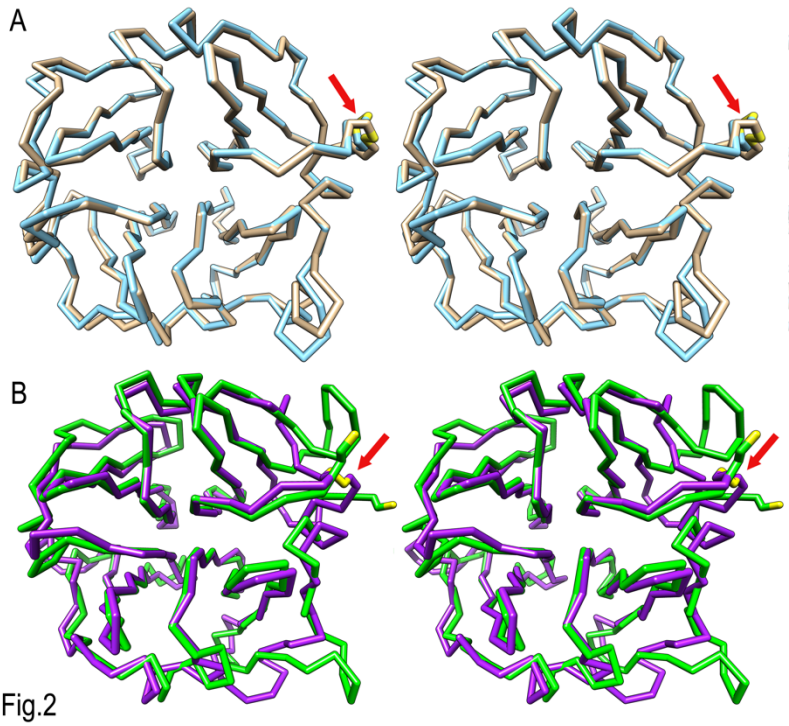


Fig.2

Figure 2 — Predicted structure of baculoviral MMP C-terminal domains. **(A)** Superposition in cross-eye stereo of the two top-ranked homology models (in tan and cyan respectively) of *Xestia c-nigrum* granulovirus MMP segment C^{280}-C^{469} (UP Q9PZ03) by threading calculations with the LOMETS meta-server using programs cdPPAS (Z-score 16.9; [118]) and SP3 (Z-score 36.6; [119]), respectively. Full-length modeling was automatically carried out with the program MODELLER [120]. The corresponding Cα-traces and the critical disulfide bond linking the terminal β-leaflets (C^{280}-C^{469}; red arrow), which is usually found in fourfold β-propeller structures such as MMP hemopexin domains [121], are depicted. In both cases, the automatically selected template structure was that of the hemopexin domain of human MMP-14 (PDB 3C7X; [122]). **(B)** Same as (A) but for *Cydia pomonella* granulovirus MMP (UP Q91F09) segment C^{359}-C^{545}. The two top predictions (in purple and green, respectively) were obtained with the program HHSEARCH (Z-scores 29.6 and 35.9; [123]). Selected templates were the hemopexin domains of human MMP-2 (PDB 1GEN; [124]) and MMP-1 (PDB 1SU3; [125]). In the second automatic homology model, the two cysteines are not linked but close to each other in space.

6. Matrix metalloproteinases in archaea and bacteria — The only prokaryotic MMP characterized at the functional and structural level to date is karilysin from *Tannerella forsythia*. This is a human oral-microbiome bacterium from the phylum Bacteroidetes that is engaged in odontopathogenic infections [23, 127-135]. Karilysin comprises a CD flanked upstream by a signal peptide for secretion and a 14-residue PD, which does not proceed over a cysteine-switch mechanism but an "aspartate-switch" mechanism, as found in the otherwise unrelated astacin family within the metzincin clan [23, 128, 136, 137]. The CD is followed by two domains of unknown structure and function, collectively spanning 275 residues, which comprise the C-terminal residues K-L-I-K-K. A C-terminal domain similar to karilysin is found in other unrelated peptidases within *T. forsythia*, *Bacteroides* sp. and *Prevotella* sp., which have been collectively termed KLIKK proteases [138]

The CD of karilysin was characterized at the structural level (see Figure 1A), which revealed that it fulfilled all the structural criteria of MMP CDs of mammals [23] (Figure 1B). These studies also suggested that karilysin is evolutionary closer to forms from mosquitoes that are insect vectors of malaria (*Anopheles gambiae*), dengue fever, Chikungunya and yellow fever (*Aedes aegypti*), and West Nile virus and Zika virus infections (*Culex quinquefasciatus*) than to bacterial counterparts [23]. The lifecycle of these mosquitoes entails feeding on human blood and they are mostly found in poor countries, which have the highest incidence of odontopathogenic bacterial infections [139].

Further to karilysin in bacteria, a potential MMP ortholog in *Bacillus anthracis*, MmpZ, was shown to participate in the extracellular degradation of anthrax toxin components and anthrolysin O at the onset of the stationary growth phase of the bacterium [140]. However, detailed inspection of its protein sequence (UP Q81NM7) reveals that it deviates from the extended sequence pattern of MMP CTSs. Therefore, it cannot be assigned unambiguously to MMPs until its three-dimensional structure has been resolved. Moreover, there have been other reports postulating the existence of bacterial MPs, which were hailed as ancestral forms of MMPs [43]. In particular, *Bacteroides fragilis* toxins *alias* fragilysins were thought to accomplish this role [39, 141-143]. However, when the structures of profragilysin-3 and the closely-related metalloproteinase II were reported, it became obvious that fragilysins, which are present in enterotoxigenic *B. fragilis* strains but not commensal ones, represent a metzincin family on their own, which is closer to adamalysin/ADAMs than MMPs, if at all [12, 144-146].

To complete the picture of MMP distribution in prokaryotes, we conducted sequence similarity searches and identified several hundred potential MMP orthologs across archaeal and bacterial genomes, some of them with several copies. A representative selection of manually curated sequences is provided in Suppl. Tables 2 and 3. Inspection of the archaeal sequences reveals that they cluster to phyla Euryarchaeota and Thaumarchaeota, which populate the human digestive tract together with Crenarchaeota [147]. The healthy gut microbiome is also dominated in humans by the bacterial phyla Firmicutes and Bacteroidetes, with Actinobacteria, Proteobacteria and Verrucomicrobia present in smaller proportions [148]. Upstream in the gastrointestinal tract, six major bacterial phyla populate the oral microbiome: Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria, Spirochaetes, and Fusobacteria [149]. Bacterial species that potentially contain MMPs also belong to these phyla, as does karilysin, with the notable exception of a few bacterial sequences from the phyla Planctomycetes, an aquatic phylum present in brackish and fresh marine waters, and Cyanobacteria, which also inhabit waters and moist soils (Suppl. Table 3). Accordingly, the distribution of archaeal and bacterial sequences is patchy and they are almost exclusively found in species of phyla highly represented in human microbiomes.

7. Conclusion — MMPs are arguably the best studied MPs at the molecular, functional, physiological and structural levels, but most reports are restricted to humans and a few other animals. However, a comprehensive review of the literature and current sequence similarity searches revealed that validated and potential MMP CDs are widespread. In general, most proteins possess common ancestors in Eubacteria or Archaea, so their presence within the latter indicates that inheritance follows the Darwinian tree-based pathway or vertical descent model [77, 150]. This is the case for ~60% of human protein domains, which have their origins in these kingdoms and eukaryotic nodes before the metazoan era [77]. However, some proteins originate at nodes that appeared later in evolution [151], as reported for the large, multi-domain pan-peptidase inhibitors of the α_2-macroglobulin (α_2M) family [152]. These >1,000-residue proteins are widely distributed across metazoans, but missing in all non-metazoan eukaryotic lineages. Unexpectedly, homologous proteins were found in several bacterial proteomes [153, 154], but their distribution was patchy and incompatible with vertical descent from a common ancestral eubacterium. As most of these bacterial species encoding α_2Ms exploited higher eukaryotes as hosts, either as pathogenic invaders or commensal colonizers, it was proposed that they were acquired by eukaryotic-to-prokaryotic horizontal gene transfer [152], similarly to previously suggested for some metabolic enzymes [155, 156].

MMPs are likewise widespread, perhaps across all kingdoms of life, where they are possibly involved in extracellular processing of proteins. However, they only show a homogenous gene distribution that is probably consistent with a vertical descent model within animals of the subkingdom Eumetazoa, as they are absent from more primitive metazoans. Within plants, they have only been found in higher plants. Here, the domain architecture is reminiscent of invertebrate MMPs, which suggests that plant and invertebrate MMPs are more closely related to each other than to vertebrate MMPs. This, in turn, hints that they could be modern representatives of an ancient MMP ancestor, common to the three groups [6, 39]. In fungi, protists, viruses, bacteria and archaea, the presence of MMP sequences is reduced and patchy, which violates the vertical descent model. Generally, distribution is restricted to phyla with a lifecycle entailing intimate, direct or indirect, pathogenic or commensal, interaction with members of the subkingdom Eumetazoa. This suggests that MMPs in those kingdoms, like bacterial α_2Ms, are xenologs coopted several times during evolution from eumetazoan hosts by independent horizontal gene transfer events (see e.g. Suppl. Fig. 1), which would include uptake of mRNA by competence or abiotic mechanisms [157], followed by subsequent spreading and polyplication within phyla.

ACKNOWLEDGMENTS

This study was supported in part by grants from European, Spanish, and Catalan agencies (grant references FP7-HEALTH-2012-306029-2 “TRIGGER”; BFU2015-64487R, MDM-2014-0435; BIO2013-49320-EXP; BIO2015-64216-P; BIO2013-49604-EXP; and 2014SGR9). The Structural Biology Unit (www.sbu.csic.es) of IBMB is a “María de Maeztu” Unit of Excellence from the Spanish Ministry of Economy, Industry and Competitiveness.

REFERENCES

[1] H. Nagase, R. Visse, G. Murphy, Structure and function of matrix metalloproteinases and TIMPs., Cardiovasc. Res., 69 (2006) 562-573.

[2] C.J. Morrison, G.S. Butler, D. Rodríguez, C.M. Overall, Matrix metalloproteinase proteomics: substrates, targets, and therapy., Curr. Opin. Cell Biol., 21 (2009) 645-653.

[3] C. Tallant, A. Marrero, F.X. Gomis-Rüth, Matrix metalloproteinases: fold and function of their catalytic domains., Biochim. Biophys. Acta - Mol. Cell Res., 1803 (2010) 20-28.

[4] J. Gross, C.M. Lapière, Collagenolytic activity in amphibian tissues: a tissue culture assay., Proc. Natl. Acad. Sci. USA, 48 (1962) 1014-1022.

[5] D.J. Krysan, N.C. Rockwell, R.S. Fuller, Quantitative characterization of furin specificity. Energetics of substrate discrimination using an internally consistent set of hexapeptidyl methylcoumarinamides., J. Biol. Chem., 274 (1999) 23229-23234.

[6] M. Fanjul-Fernández, A.R. Folgueras, S. Cabrera, C. López-Otín, Matrix metalloproteinases: evolution, gene regulation and functional analysis in mouse models., Biochim. Biophys. Acta, 1803 (2010) 3-19.

[7] H. Nagase, J.F. Woessner Jr., Matrix metalloproteinases., J. Biol. Chem., 274 (1999) 21491-21494.

[8] W. Bode, F.X. Gomis-Rüth, W. Stöcker, Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the ´metzincins´. FEBS Lett., 331 (1993) 134-140.

[9] W. Stöcker, F. Grams, U. Baumann, P. Reinemer, F.X. Gomis-Rüth, D.B. McKay, W. Bode, The metzincins - Topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases, Prot. Sci., 4 (1995) 823-840.

[10] F.X. Gomis-Rüth, Structural aspects of the *metzincin* clan of metalloendopeptidases., Mol. Biotech., 24 (2003) 157-202.

[11] F.X. Gomis-Rüth, Catalytic domain architecture of metzincin metalloproteases., J. Biol. Chem., 284 (2009) 15353-15357.

[12] N. Cerdà-Costa, F.X. Gomis-Rüth, Architecture and function of metallopeptidase catalytic domains., Prot. Sci., 23 (2014) 123-144.

[13] B. Rost, Twilight zone of protein sequence alignments., Prot. Eng., 12 (1999) 85-94.

[14] F.X. Gomis-Rüth, L.F. Kress, W. Bode, First structure of a snake venom metalloproteinase : a prototype for matrix metalloproteinases/ collagenases, EMBO J., 12 (1993) 4151-4157.

[15] C. Tallant, R. García-Castellanos, U. Baumann, F.X. Gomis-Rüth, On the relevance of the Met-turn methionine in metzincins., J. Biol. Chem., 285 (2010) 13951-13957.

[16] R. Mittal, A.P. Patel, L.H. Debs, D. Nguyen, K. Patel, M. Grati, J. Mittal, D. Yan, P. Chapagain, X.Z. Liu, Intricate Functions of Matrix Metalloproteinases in Physiological and Pathological Conditions, Journal of cellular physiology, 231 (2016) 2599-2621.

[17] H.E. van Wart, H. Birkedal-Hansen, The cysteine switch : a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family., Proc. Natl. Acad. Sci. USA, 87 (1990) 5578-5582.

[18] A. Page-McCaw, A.J. Ewald, Z. Werb, Matrix metalloproteinases and the regulation of tissue remodelling, Nat. Rev. Mol. Cell Biol., 8 (2007) 221-233.

[19] C. López-Otín, L.H. Palavalli, Y. Samuels, Protective roles of matrix metalloproteinases: from mouse models to human cancer., Cell cycle, 8 (2009) 3657-3662.

[20] A. Page-McCaw, Remodeling the model organism: matrix metalloproteinase functions in invertebrates., Semin. Cell Dev. Biol., 19 (2008) 14-23.

[21] D. Pisani, W. Pett, M. Dohrmann, R. Feuda, O. Rota-Stabelli, H. Philippe, N. Lartillot, G. Worheide, Genomic data do not support comb jellies as the sister group to all other animals., Proc. Natl. Acad. Sci. USA, 112 (2015) 15402-15407.

[22] M. Srivastava, E. Begovic, J. Chapman, N.H. Putnam, U. Hellsten, T. Kawashima, A. Kuo, T. Mitros, A. Salamov, M.L. Carpenter, A.Y. Signorovitch, M.A. Moreno, K. Kamm, J. Grimwood, J. Schmutz, H. Shapiro, I.V. Grigoriev, L.W. Buss, B. Schierwater, S.L. Dellaporta, D.S. Rokhsar, The *Trichoplax* genome and the nature of placozoans., Nature, 454 (2008) 955-960.

[23] N. Cerdà-Costa, T. Guevara, A.Y. Karim, M. Ksiazek, K.A. Nguyen, J.L. Arolas, J. Potempa, F.X. Gomis-Rüth, The structure of the catalytic domain of *Tannerella forsythia* karilysin reveals it is a bacterial xenologue of animal matrix metalloproteinases., Mol. Microbiol., 79 (2011) 119-132.

[24] F.X. Gomis-Rüth, T.O. Botelho, W. Bode, A standard orientation for metallopeptidases., Biochim. Biophys. Acta, 1824 (2012) 157-163.

[25] B. Lovejoy, A.R. Welch, S. Carr, C. Luong, C. Broka, R.T. Hendricks, J.A. Campbell, K.A. Walker, R. Martin, H. Van Wart, M.F. Browner, Crystal structures of MMP-1 and -13 reveal the structural basis for selectivity of collagenase inhibitors., Nat. Struct. Biol., 6 (1999) 217-221.

[26] A.G. Pavlovsky, M.G. Williams, Q.Z. Ye, D.F. Ortwine, C.F. Purchase II, A.D. White, V. Dhanaraj, B.D. Roth, L.L. Johnson, D. Hupe, C. Humblet, T.L. Blundell, X-ray structure of human stromelysin catalytic domain complexed with nonpeptide inhibitors: implications for inhibitor selectivity., Protein Sci., 8 (1999) 1455-1462.

[27] M.F. Browner, W.W. Smith, A.L. Castelhano, Matrilysin-inhibitor complexes: common themes among metalloproteases., Biochemistry, 34 (1995) 6602-6610.

[28] P. Reinemer, F. Grams, R. Huber, T. Kleine, S. Schnierer, M. Piper, H. Tschesche, W. Bode, Structural implications for the role of the N terminus in the 'superactivation' of collagenases. A crystallographic study., FEBS Lett., 338 (1994) 227-233.
[29] C. Antoni, L. Vera, L. Devel, M.P. Catalani, B. Czarny, E. Cassar-Lajeunesse, E. Nuti, A. Rossello, V. Dive, E.A. Stura, Crystallization of bi-functional ligand protein complexes., J. Struct. Biol., 182 (2013) 246-254.
[30] I. Bertini, V. Calderone, M. Fragai, C. Luchinat, S. Mangani, B. Terni, Crystal structure of the catalytic domain of human matrix metalloproteinase 10., J. Mol. Biol., 336 (2004) 707-716.
[31] A.L. Gall, M. Ruff, R. Kannan, P. Cuniasse, A. Yiotakis, V. Dive, M.C. Rio, P. Basset, D. Moras, Crystal structure of the stromelysin-3 (MMP-11) catalytic domain complexed with a phosphinic inhibitor mimicking the transition-state., J. Mol. Biol., 307 (2001) 577-586.
[32] I. Bertini, V. Calderone, M. Cosenza, M. Fragai, Y.M. Lee, C. Luchinat, S. Mangani, B. Terni, P. Turano, Conformational variability of matrix metalloproteinases: beyond a single 3D structure., Proc. Natl. Acad. Sci. USA, 102 (2005) 5334-5339.
[33] T. Kohno, H. Hochigai, E. Yamashita, T. Tsukihara, M. Kanaoka, Crystal structures of the catalytic domain of human stromelysin-1 (MMP-3) and collagenase-3 (MMP-13) with a hydroxamic acid inhibitor SM-25453., Biochem. Biophys. Res. Commun., 344 (2006) 315-322.
[34] R. Lang, M. Braun, N.E. Sounni, A. Noel, F. Frankenne, J.M. Foidart, W. Bode, K. Maskos, Crystal structure of the catalytic domain of MMP-16/MT3-MMP: characterization of MT-MMP specific features., J. Mol. Biol., 336 (2004) 213-225.
[35] E. Llano, G. Adam, A.M. Pendas, V. Quesada, L.M. Sánchez, I. Santamaria, S. Noselli, C. López-Otín, Structural and enzymatic characterization of *Drosophila* Dm2-MMP, a membrane-bound matrix metalloproteinase with tissue-specific expression., J. Biol. Chem., 277 (2002) 23321-23329.
[36] E. Llano, A.M. Pendas, P. Aza-Blanc, T.B. Kornberg, C. López-Otín, Dm1-MMP, a matrix metalloproteinase from *Drosophila* with a potential role in extracellular matrix remodeling during neural development., J. Biol. Chem., 275 (2000) 35978-35985.
[37] E. Knorr, H. Schmidtberg, A. Vilcinskas, B. Altincicek, MMPs regulate both development and immunity in the tribolium model insect, PloS one, 4 (2009) e4751.
[38] E. Ishimwe, J.J. Hodgson, A.L. Passarelli, Expression of the *Cydia pomonella* granulovirus matrix metalloprotease enhances *Autographa californica* multiple nucleopolyhedrovirus virulence and can partially substitute for viral cathepsin., Virology, 481 (2015) 166-178.
[39] I. Massova, L.P. Kotra, R. Fridman, S. Mobashery, Matrix metalloproeinases : structures, evolution, and diversification., FASEB J., 12 (1998) 1075-1095.
[40] L. Angerer, S. Hussain, Z. Wei, B.T. Livingston, Sea urchin metalloproteases: a genomic survey of the BMP-1/tolloid-like, MMP and ADAM families, Dev. Biol., 300 (2006) 267-281.
[41] J. Huxley-Jones, T.K. Clarke, C. Beck, G. Toubaris, D.L. Robertson, R.P. Boot-Handford, The evolution of the vertebrate metzincins; insights from *Ciona intestinalis* and *Danio rerio*., BMC Evol. Biol., 7 (2007) 63.
[42] V. Quesada, G. Velasco, X.S. Puente, W.C. Warren, C. López-Otín, Comparative genomic analysis of the zebra finch degradome provides new insights into evolution of proteases in birds and mammals., BMC genomics, 11 (2010) 220.
[43] C.D. Small, B.D. Crawford, Matrix metalloproteinases in neural development: a phylogenetically diverse perspective., Neural Regen. Res., 11 (2016) 357-362.
[44] A.M. Kantor, S. Dong, N.L. Held, E. Ishimwe, A.L. Passarelli, R.J. Clem, A.W.E. Franz, Identification and initial characterization of matrix metalloproteinases in the yellow fever mosquito, *Aedes aegypti*., Insect Mol. Biol., 26 (2017) 113-126.
[45] N. Miyamoto, M.A. Yoshida, H. Koga, Y. Fujiwara, Genetic mechanisms of bone digestion and nutrient absorption in the bone-eating worm *Osedax japonicus* inferred from transcriptome and gene expression analyses., BMC Evol. Biol., 17 (2017) 17.
[46] K.I. Spanier, F. Leese, C. Mayer, J.K. Colbourne, D. Gilbert, M.E. Pfrender, R. Tollrian, Predator-induced defences in *Daphnia pulex*: selection and evaluation of internal reference genes for gene expression studies with real-time PCR., BMC Mol. Biol., 11 (2010) 50.
[47] E. Goulielmaki, I. Siden-Kiamos, T.G. Loukeris, Functional characterization of *Anopheles* matrix metalloprotease 1 reveals its agonistic role during sporogonic development of malaria parasites., Infect. Immun., 82 (2014) 4865-4877.
[48] T. Ueno, T. Nakaoka, H. Takeuchi, T. Kubo, Differential gene expression in the hypopharyngeal glands of worker honeybees (*Apis mellifera* L.) associated with an age-dependent role change., Zoolog. Sci., 26 (2009) 557-563.
[49] J.M. Guan, L. Bing, D. Wang, C.L. Liu, W.D. Shen, Cloning, sequence analysis and expression of a matrix metalloproteinase gene (Bm-MMP) in the silkworm, *Bombyx mori*., Acta Entomol. Sinica, 52 (2009) 353-362.
[50] B. Altincicek, A. Vilcinskas, Identification of a lepidopteran matrix metalloproteinase with dual roles in metamorphosis and innate immunity, Dev. Comp. Immunol., 32 (2008) 400-409.
[51] S. Vishnuvardhan, R. Ahsan, K. Jackson, R. Iwanicki, J. Boe, J. Haring, K.J. Greenlee, Identification of a novel metalloproteinase and its role in juvenile development of the tobacco hornworm, *Manduca sexta* (Linnaeus). J. Exp. Zool. B Mol. Dev. Evol., 320 (2013) 105-117.
[52] J.C. Means, A.L. Passarelli, Viral fibroblast growth factor, matrix metalloproteases, and caspases are associated with enhancing systemic infection by baculoviruses., Proc. Natl. Acad. Sci. USA, 107 (2010) 9825-9830.
[53] Y. Zhang, H. Zhang, Y. Kong, L. Feng, Identification and characterization of an amphioxus matrix metalloproteinase homolog BbMMPL2 responding to bacteria challenge., Dev. Comp. Immunol., 37 (2012) 371-380.

[54] A.A. Leontovich, J. Zhang, K. Shimokawa, H. Nagase, M.P. Sarras Jr., A novel hydra matrix metalloproteinase (HMMP) functions in extracellular matrix degradation, morphogenesis and the maintenance of differentiated cells in the foot process., Development, 127 (2000) 907-920.
[55] H. Shimizu, X. Zhang, J. Zhang, A. Leontovich, K. Fei, L. Yan, M.P. Sarras Jr., Epithelial morphogenesis in hydra requires *de novo* expression of extracellular matrix components and matrix metalloproteinases., Development, 129 (2002) 1521-1532.
[56] C. Kang, D.Y. Han, K.I. Park, M.J. Pyo, Y. Heo, H. Lee, G.S. Kim, E. Kim, Characterization and neutralization of *Nemopilema nomurai* (Scyphozoa: Rhizostomeae) jellyfish venom using polyclonal antibody., Toxicon : official journal of the International Society on Toxinology, 86 (2014) 116-125.
[57] K. Nomura, T. Shimizu, H. Kinoh, Y. Sendai, M. Inomata, N. Suzuki, Sea urchin hatching enzyme (envelysin): cDNA cloning and deprivation of protein substrate specificity by autolytic degradation., Biochemistry, 36 (1997) 7225-7238.
[58] C. Ghiglione, G. Lhomond, T. Lepage, C. Gache, Structure of the sea urchin hatching enzyme gene., Eur. J. Biochem., 219 (1994) 845-854.
[59] L. Sun, M. Chen, H. Yang, T. Wang, B. Liu, C. Shu, D.M. Gardiner, Large scale gene expression profiling during intestine and body wall regeneration in the sea cucumber *Apostichopus japonicus*., Comp. Biochem. Physiol. Part D Genomics Proteomics, 6 (2011) 195-205.
[60] J.L. Quiñones, R. Rosa, D.L. Ruíz, J.E. García-Arrarás, Extracellular matrix remodeling and metalloproteinase involvement during intestine regeneration in the sea cucumber *Holothuria glaberrima*., Dev. Biol., 250 (2002) 181-197.
[61] C. Nikapitiya, I.C. McDowell, L. Villamil, P. Muñoz, S. Sohn, M. Gómez-Chiarri, Identification of potential general markers of disease resistance in American oysters, *Crassostrea virginica* through gene expression studies, Fish Shellfish Immunol., 41 (2014) 27-36.
[62] F. Mannello, L. Canesi, G. Gazzanelli, G. Gallo, Biochemical properties of metalloproteinases from the hemolymph of the mussel Mytilus galloprovincialis Lam., Comp. Biochem. Physiol. B Biochem. Mol. Biol., 128 (2001) 507-515.
[63] D. Indra, K. Ramalingam, M. Babu, Isolation, purification and characterization of collagenase from hepatopancreas of the land snail *Achatina fulica*., Comp. Biochem. Physiol. B Biochem. Mol. Biol., 142 (2005) 1-7.
[64] T.P. Yoshino, M. Brown, X.J. Wu, C.J. Jackson, R. Ocadiz-Ruíz, I.W. Chalmers, M. Kolb, C.H. Hokke, K.F. Hoffmann, Excreted/secreted *Schistosoma mansoni* venom allergen-like 9 (SmVAL9) modulates host extracellular matrix remodelling gene expression., Int. J. Parasitol., 44 (2014) 551-563.
[65] K.J. Wang, H.L. Ren, D.D. Xu, L. Cai, M. Yang, Identification of the up-regulated expression genes in hemocytes of variously colored abalone (*Haliotis diversicolor* Reeve, 1846) challenged with bacteria., Dev. Comp. Immunol., 32 (2008) 1326-1347.
[66] O. Chovar-Vera, V. Valenzuela-Muñoz, C. Gallardo-Escarate, Molecular characterization of collagen IV evidences early transcription expression related to the immune response against bacterial infection in the red abalone (*Haliotis rufescens*). Fish Shellfish Immunol., 42 (2015) 241-248.
[67] J.S. Rhee, B.M. Kim, C.B. Jeong, T. Horiguchi, Y.M. Lee, I.C. Kim, J.S. Lee, Immune gene mining by pyrosequencing in the rockshell, *Thais clavigera*., Fish Shellfish Immunol., 32 (2012) 700-710.
[68] R. Sun, Z.Y. Li, H.J. He, J. Wei, J. Wang, Q.X. Zhang, J. Zhao, X.M. Zhan, Z.D. Wu, Molecular cloning and characterization of a matrix metalloproteinase, from *Caenorhabditis elegans*: employed to identify homologous protein from *Angiostrongylus cantonensis*., Parasitol. Res., 110 (2012) 2001-2012.
[69] K. Wada, H. Sato, H. Kinoh, M. Kajita, H. Yamamoto, M. Seiki, Cloning of three *Caenorhabditis elegans* genes potentially encoding novel matrix metalloproteinases., Gene, 211 (1998) 57-62.
[70] D. Coates, R. Siviter, R.E. Isaac, Exploring the *Caenorhabditis elegans* and *Drosophila melanogaster* genomes to understand neuropeptide and peptidase function., Biochem. Soc. Trans., 28 (2000) 464-469.
[71] B. Altincicek, M. Fischer, M. Fischer, K. Luersen, M. Boll, U. Wenzel, A. Vilcinskas, Role of matrix metalloproteinase ZMP-2 in pathogen resistance and development in *Caenorhabditis elegans*., Dev. Comp. Immunol., 34 (2010) 1160-1169.
[72] E.S. Kovaleva, E.P. Masler, A.M. Skantar, D.J. Chitwood, Novel matrix metalloproteinase from the cyst nematodes *Heterodera glycines* and *Globodera rostochiensis*., Mol. Biochem. Parasitol., 136 (2004) 109-112.
[73] P. Uparanukraw, N. Morakote, T. Harnnoi, A. Dantrakool, Molecular cloning of a gene encoding matrix metalloproteinase-like protein from *Gnathostoma spinigerum*., Parasitol. Res., 87 (2001) 751-757.
[74] M.E. Isolani, J.F. Abril, E. Salo, P. Deri, A.M. Bianucci, R. Batistoni, Planarians as a model to assess in vivo the role of matrix metalloproteinase genes during homeostasis and regeneration., PloS one, 8 (2013) e55649.
[75] Y. Roskov, L. Abucay, T. Orrell, D. Nicolson, N. Bailly, P. Kirk, T. Bourgoin, R.E. de Walt, W. Decock, A. de Wever, E. van Nieukerken, Species 2000 & ITIS Catalogue of Life., Species 2000, Leiden (The Netherlands), 2017.
[76] G.J. Murphy, G. Murphy, J.J. Reynolds, The origin of matrix metalloproteinases and their familial relationships., FEBS Lett., 289 (1991) 4-7.
[77] L.R. Pal, C. Guda, Tracing the origin of functional and conserved domains in the human proteome: implications for protein evolution at the modular level., BMC Evol. Biol., 6 (2006) 91.
[78] I. Massova, S. Mobashery, Kinship and diversification of bacterial penicillin-binding proteins and β-lactamases., Antimicrob. Agents Chemother., 42 (1998) 1-17.
[79] X.S. Puente, L.M. Sánchez, C.M. Overall, C. López-Otín, Human and mouse proteases: a comparative genomic approach., Nat. Rev. Genet., 4 (2003) 544-558.
[80] R.N. Gonçalves, S.D. Gozzini Barbosa, R.E. da Silva-López, Proteases from *Canavalia ensiformis*: active and thermostable enzymes with potential of application in biotechnology., Biotechnol. Res. Int., 2016 (2016) 3427098.

[81] D. Li, H. Zhang, Q. Song, L. Wang, S. Liu, Y. Hong, L. Huang, F. Song, Tomato Sl3-MMP, a member of the matrix metalloproteinase family, is required for disease resistance against *Botrytis cinerea* and *Pseudomonas syringae* pv. *tomato* DC3000, BMC Plant Biol., 15 (2015) 143.
[82] O.H.P. Ramos, H.S. Selistre-de-Araujo, Identification of metalloprotease gene families in sugarcane., Genet. Mol. Biol., 24 (2001) 285-290.
[83] C.Y. Liu, J.S. Graham, Cloning and characterisation of an *Arabidopsis thaliana* cDNA homologous to the matrix metalloproteinases (The Electronic Plant Gene register PGR 98–130). Plant Physiol., 117 (1998) 1127-1127.
[84] J.M. Maidment, D. Moore, G.P. Murphy, G. Murphy, I.M. Clark, Matrix metalloproteinase homologues from *Arabidopsis thaliana*. Expression and activity., J. Biol. Chem., 274 (1999) 34706-34710.
[85] D. Golldack, O.V. Popova, K.J. Dietz, Mutation of the matrix metalloproteinase At2-MMP inhibits growth and causes late flowering and early senescence in *Arabidopsis*., J. Biol. Chem., 277 (2002) 5541-5547.
[86] B.S. Flinn, Plant extracellular matrix metalloproteinases., Funct. Plant Biol., 35 (2008) 1183-1193.
[87] G. Marino, P.F. Huesgen, U. Eckhard, C.M. Overall, W.P. Schroder, C. Funk, Family-wide characterization of matrix metalloproteinases from *Arabidopsis thaliana* reveals their distinct proteolytic activity and cleavage site specificity., Biochem. J., 457 (2014) 335-346.
[88] R.A.L. van der Hoorn, T. Colby, S. Nickel, K.H. Richau, J. Schmidt, M. Kaiser, Mining the active proteome of *Arabidopsis thaliana*., Front. Plant Sci., 2 (2011) 89.
[89] C. Broccanello, P. Stevanato, F. Biscarini, D. Cantu, M. Saccomani, A new polymorphism on chromosome 6 associated with bolting tendency in sugar beet., BMC Genet., 16 (2015) 142.
[90] V.G.R. Delorme, P.F. McCabe, D.J. Kim, C.J. Leaver, A matrix metalloproteinase gene is expressed at the boundary of senescence and programmed cell death in cucumber., Plant Physiol., 123 (2000) 917-927.
[91] L.V. Ragster, M.J. Chrispeels, Azocoll-digesting proteinases in soybean leaves: characteristics and changes during leaf maturation and senescence., Plant Physiol., 64 (1979) 857-862.
[92] J.S. Graham, J. Xiong, J.W. Gillikin, Purification and developmental analysis of a metalloendoproteinase from the leaves of *Glycine max*., Plant Physiol., 97 (1991) 786-792.
[93] G. McGeehan, W. Burkhart, R. Anderegg, J.D. Becherer, J.W. Gillikin, J.S. Graham, Sequencing and characterization of the soybean leaf metalloproteinase : structural and functional similarity to the matrix metalloproteinase family., Plant Physiol., 99 (1992) 1179-1183.
[94] J.H. Pak, C.Y. Liu, J. Huangpu, J.S. Graham, Construction and characterization of the soybean leaf metalloproteinase cDNA., FEBS Lett., 404 (1997) 283-288.
[95] Y. Liu, C. Dammann, M.K. Bhattacharyya, The matrix metalloproteinase gene GmMMP2 is activated in response to pathogenic infections in soybean., Plant Physiol., 127 (2001) 1788-1797.
[96] C.-W. Cho, E. Chung, K. Kim, H.-A. Soh, Y.K. Jeong, S.-W. Lee, Y.-C. Lee, K.-S. Kim, Y.-S. Chung, J.-H. Lee, Plasma membrane localization of soybean matrix metalloproteinase differentially induced by senescence and abiotic stress., Biol. Plant, 53 (2009) 461-467.
[97] J.P. Combier, T. Vernie, F. de Billy, F. El Yahyaoui, R. Mathis, P. Gamas, The MtMMPL1 early nodulin is a novel member of the matrix metalloendoproteinase family with a role in *Medicago truncatula* infection by *Sinorhizobium meliloti*., Plant Physiol., 144 (2007) 703-716.
[98] A. Schiermeyer, H. Hartenstein, M.K. Mandal, B. Otte, V. Wahner, S. Schillberg, A membrane-bound matrix-metalloproteinase from *Nicotiana tabacum* cv. BY-2 is induced by bacterial pathogens., BMC Plant Biol., 9 (2009) 83.
[99] M.K. Mandal, R. Fischer, S. Schillberg, A. Schiermeyer, Biochemical properties of the matrix metalloproteinase NtMMP1 from *Nicotiana tabacum* cv. BY-2 suspension cells., Planta, 232 (2010) 899-910.
[100] S.M. Ratnaparkhe, E.M. Egertsdotter, B.S. Flinn, Identification and characterization of a matrix metalloproteinase (Pta1-MMP) expressed during Loblolly pine (*Pinus taeda*) seed development, germination completion, and early seedling establishment., Planta, 230 (2009) 339-354.
[101] I.M. Clark, M.D. Thomas, S. de Vos, Chapter 177 — Plant matrixins., in: N.D. Rawlings, G. Salvesen (Eds.) Handbook of Proteolytic Enzymes, vol. 1, Academic Press, Oxford, 2013, pp. 854-856.
[102] G. Marino, C. Funk, Matrix metalloproteinases in plants: a brief overview., Physiol. Plant., 145 (2012) 196-202.
[103] D. Zimmermann, J.A. Gómez-Barrera, C. Pasule, U.B. Brack-Frick, E. Sieferer, T.M. Nicholson, J. Pfannstiel, A. Stintzi, A. Schaller, Cell death control by matrix metalloproteinases., Plant Physiol., 171 (2016) 1456-1469.
[104] S. Rowsell, P. Hawtin, C.A. Minshull, H. Jepson, S.M. Brockbank, D.G. Barratt, A.M. Slater, W.L. McPheat, D. Waterson, A.M. Henney, R.A. Pauptit, Crystal structure of human MMP9 in complex with a reverse hydroxamate inhibitor., J. Mol. Biol., 319 (2002) 173-181.
[105] Y. Matsuda, Chapter 187 - Gametolysin, in: N.D. Rawlings, G. Salvesen (Eds.) Handbook of Proteolytic Enzymes, vol. 1, Elsevier Ltd., 2013, pp. 891-895.
[106] H. Claes, Autolysis of the cell wall of gametes of *Chlamydomonas reinhardii* (GERMAN). Arch. Mikrobiol., 78 (1971) 180-188.
[107] T. Kinoshita, H. Fukuzawa, T. Shimada, T. Saito, Y. Matsuda, Primary structure and expression of a gamete lytic enzyme in *Chlamydomonas reinhardtii* : similarity of functional domains to matrix metalloproteases., Proc. Natl. Acad. Sci. USA, 89 (1992) 4693-4697.
[108] T. Kubo, T. Saito, H. Fukuzawa, Y. Matsuda, Two tandemly-located matrix metalloprotease genes with different expression patterns in the chlamydomonas sexual cell cycle., Curr. Genet., 40 (2001) 136-143.

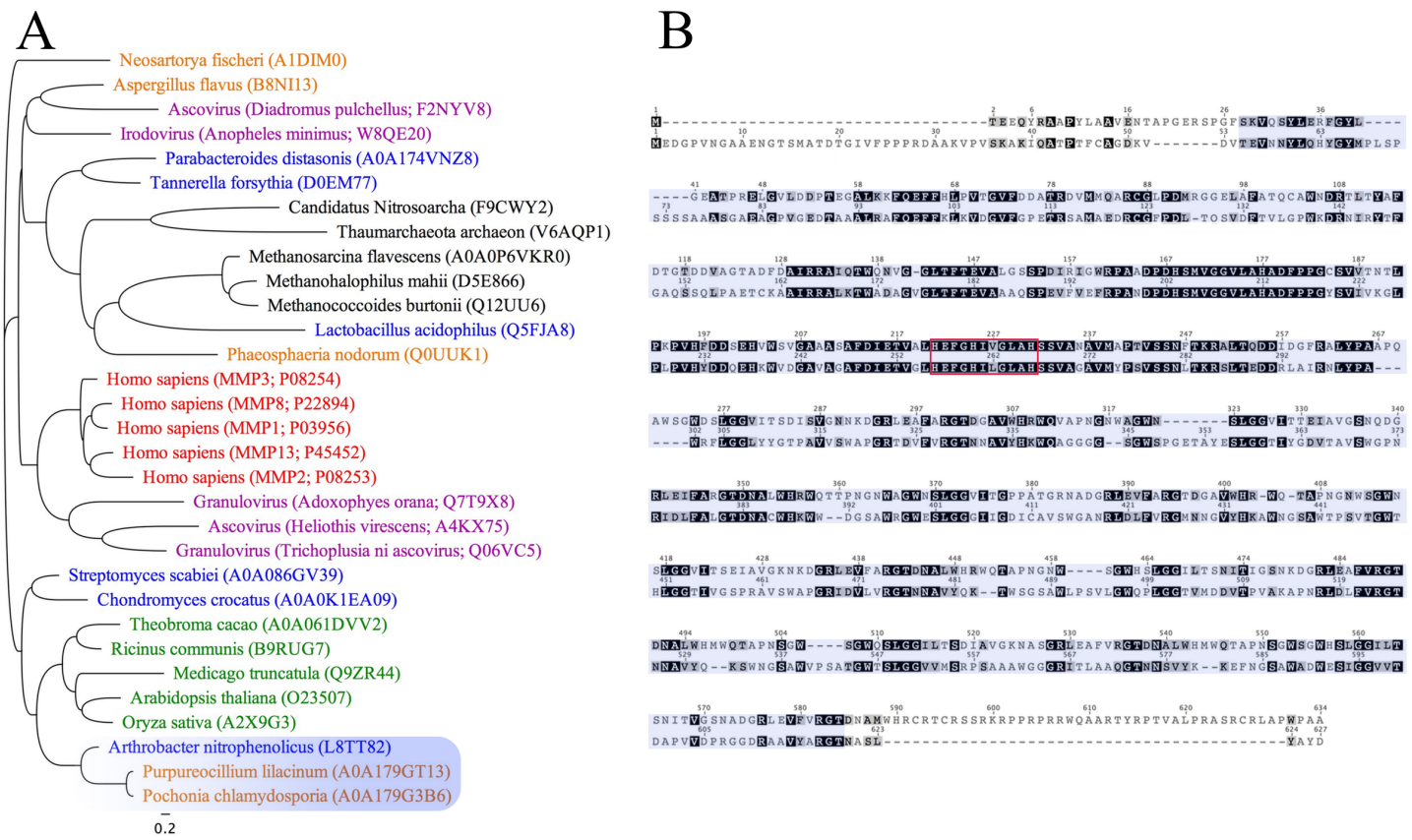
[109] A. Hallmann, P. Amon, K. Godl, M. Heitzer, M. Sumper, Transcriptional activation by the sexual pheromone and wounding : a new gene family from *Volvox* encoding modular proteins with (hydroxy)proline-rich and metalloproteinase homology domains., Plant J., 26 (2001) 583-593.
[110] T. Shimizu, T. Inoue, H. Shiraishi, Cloning and characterization of novel extensin-like cDNAs that are expressed during late somatic cell phase in the green alga *Volvox carteri*., Gene, 284 (2002) 179-187.
[111] M. Heitzer, A. Hallmann, An extracellular matrix-localized metalloproteinase with an exceptional QEXXH metal binding site prefers copper for catalytic activity., J. Biol. Chem., 277 (2002) 28280-28286.
[112] N.D. Rawlings, A.J. Barrett, R. Finn, Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors., Nucleic Acids Res., 44 (2016) D343-D350.
[113] M. Terauchi, T. Yamagishi, T. Hanyuda, H. Kawai, Genome-wide computational analysis of the secretome of brown algae (Phaeophyceae). Mar. Genomics, XX (2017) in press.
[114] T. Terashita, K. Oda, M. Kono, S. Murao, Purification and some properties of metal proteinases from *Lentinus edodes*., Agric. Biol. Chem., 49 (1985) 2293-2300.
[115] J.Z. McHenry, J.T. Christeller, E.A. Slade, W.A. Laing, The major extracellular proteinases of the silverleaf fungus, *Chondrostereum purpureum*, are metalloproteinases., Plant Pathol., 45 (1996) 552-563.
[116] E.B. Pavlukova, M.A. Belozersky, Y.E. Dunaevsky, Extracellular proteolytic enzymes of filamentous fungi, Biochemistry. Biokhimiia, 63 (1998) 899-928.
[117] R. Ko, K. Okano, S. Maeda, Structural and functional analysis of the *Xestia c-nigrum* granulovirus matrix metalloproteinase., J. Virol., 74 (2000) 11240-11246.
[118] S. Wu, Y. Zhang, LOMETS: a local meta-threading-server for protein structure prediction., Nucleic Acids Res., 35 (2007) 3375-3382.
[119] H. Zhou, Y. Zhou, Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments., Proteins, 58 (2005) 321-328.
[120] N. Eswar, B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.Y. Shen, U. Pieper, A. Sali, Comparative protein structure modeling using MODELLER., Curr. Protoc. Bioinformatics, Suppl. 15 (2006) 5.6.1-5.6.30.
[121] F.X. Gomis-Rüth, Hemopexin domains., in: A. Messerschmidt, W. Bode, M. Cygler (Eds.) Handbook of metalloproteins., vol. 3, John Wiley & Sons, Ltd., Chichester (UK), 2004, pp. 631-646.
[122] A. Tochowicz, P. Goettig, R. Evans, R. Visse, Y. Shitomi, R. Palmisano, N. Ito, K. Richter, K. Maskos, D. Franke, D. Svergun, H. Nagase, W. Bode, Y. Itoh, The dimer interface of the membrane type 1 matrix metalloproteinase hemopexin domain: crystal structure and biological functions., J. Biol. Chem., 286 (2011) 7587-7600.
[123] J. Söding, Protein homology detection by HMM-HMM comparison., Bioinformatics, 21 (2005) 951-960.
[124] A.M. Libson, A.G. Gittis, I.E. Collier, B.L. Marmer, G.I. Goldberg, E.E. Lattman, Crystal structure of the haemopexin-like C-terminal domain of gelatinase A., Nat. Struct. Biol., 2 (1995) 938-942.
[125] D. Jozic, G. Bourenkov, N.H. Lim, R. Visse, H. Nagase, W. Bode, K. Maskos, X-ray structure of human proMMP-1: new insights into procollagenase activation and collagen binding, J. Biol. Chem., 280 (2005) 9578-9585.
[126] A.L. Passarelli, Barriers to success: how baculoviruses establish efficient systemic infections., Virology, 411 (2011) 383-392.
[127] A. Siddiqi, T. Milne, M.P. Cullinan, G.J. Seymour, Analysis of *P. gingivalis*, *T. forsythia* and *S. aureus* levels in edentulous mouths prior to and 6 months after placement of one-piece zirconia and titanium implants., Clin. Oral Implants. Res., 27 (2016) 288-294.
[128] M. López-Pelegrín, M. Ksiazek, A.Y. Karim, T. Guevara, J.L. Arolas, J. Potempa, F.X. Gomis-Rüth, A novel mechanism of latency in matrix metalloproteinases., J. Biol. Chem., 290 (2015) 4728-4740.
[129] D. Bryzek, M. Ksiazek, E. Bielecka, A.Y. Karim, B. Potempa, D. Staniec, J. Koziel, J. Potempa, A pathogenic trace of *Tannerella forsythia* - shedding of soluble fully active tumor necrosis factor alpha from the macrophage surface by karilysin., Mol. Oral Microbiol., 29 (2014) 294-306.
[130] J. Potempa, F.X. Gomis-Rüth, A.Y. Karim, 185. Karilysin., in: N.D. Rawlings, G.S. Salvesen (Eds.) Handbook of proteolytic enzymes., vol. 1, Academic Press, Oxford, Great Britain, 2013, pp. 883-886.
[131] T. Guevara, M. Ksiazek, P.D. Skottrup, N. Cerda-Costa, S. Trillo-Muyo, I. de Diego, E. Riise, J. Potempa, F.X. Gomis-Rüth, Structure of the catalytic domain of the *Tannerella forsythia* matrix metallopeptidase karilysin in complex with a tetrapeptidic inhibitor., Acta Crystallogr. sect. F, 69 (2013) 472-476.
[132] P.D. Skottrup, G. Sorensen, M. Ksiazek, J. Potempa, E. Riise, A phage display selected 7-mer peptide inhibitor of the *Tannerella forsythia* metalloprotease-like enzyme karilysin can be truncated to Ser-Trp-Phe-Pro., PloS one, 7 (2012) e48537.
[133] M. Jusko, J. Potempa, A.Y. Karim, M. Ksiazek, K. Riesbeck, P. Garred, S. Eick, A.M. Blom, A metalloproteinase karilysin present in the majority of *Tannerella forsythia* isolates inhibits all pathways of the complement system., J. Immunol., 188 (2012) 2338-2349.
[134] J. Koziel, A.Y. Karim, K. Przybyszewska, M. Ksiazek, M. Rapala-Kozik, K.A. Nguyen, J. Potempa, Proteolytic inactivation of LL-37 by karilysin, a novel virulence mechanism of *Tannerella forsythia*., J. Innate Immun., 2 (2010) 288-293.
[135] A.Y. Karim, M. Kulczycka, T. Kantyka, G. Dubin, A. Jabaiah, P.S. Daugherty, I.B. Thogersen, J.J. Enghild, K.A. Nguyen, J. Potempa, A novel matrix metalloprotease-like enzyme (karilysin) of the periodontal pathogen *Tannerella forsythia* ATCC 43037., Biol. Chem., 391 (2010) 105-117.
[136] W. Stöcker, F.X. Gomis-Rüth, Astacins: proteases in development and tissue differentiation., in: K. Brix, W. Stöcker (Eds.) Proteases: structure and function., Springer Verlag, Vienna, Austria, 2013, pp. 235-263.

[137] F.X. Gomis-Rüth, S. Trillo-Muyo, W. Stöcker, Functional and structural insights into astacin metallopeptidases., Biol. Chem., 393 (2012) 1027-1041.
[138] M. Ksiazek, D. Mizgalska, S. Eick, I.B. Thøgersen, J.J. Enghild, J. Potempa, KLIKK proteases of *Tannerella forsythia*: putative virulence factors with a unique domain structure., Front. Microbiol., 6 (2015) 312.
[139] P. Axelsson, J.M. Albandar, T.E. Rams, Prevention and control of periodontal diseases in developing and industrialized nations., Periodontol. 2000, 29 (2002) 235-246.
[140] A.P. Pomerantsev, O.M. Pomerantseva, M. Moayeri, R. Fattah, C. Tallant, S.H. Leppla, A *Bacillus anthracis* strain deleted for six proteases serves as an effective host for production of recombinant proteins., Prot. Expr. Purif., 80 (2011) 80-90.
[141] J.S. Moncrief, R.J. Obiso, L.A. Barroso, J.J. Kling, R.L. Wright, R.L. Van Tassell, D.M. Lyerly, T.D. Wilkins, The enterotoxin of *Bacteroides fragilis* is a metalloprotease., Infect. Immun., 63 (1995) 175-181.
[142] R.J.J. Obiso, D.R. Bevan, T.D. Wilkins, Molecular modeling and analysis of fragilysin, the *Bacteroides fragilis* toxin., Clin. Infect. Dis., 25(Suppl. 2) (1997) S153-S155.
[143] C.L. Sears, Enterotoxigenic *Bacteroides fragilis*: a rogue among symbiotes., Clin. Microbiol. Rev., 22 (2009) 349-369.
[144] T. Goulas, J.L. Arolas, F.X. Gomis-Rüth, Structure, function and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog., Proc. Natl. Acad. Sci. USA, 108 (2011) 1856-1861.
[145] S.A. Shiryaev, A.E. Aleshin, N. Muranaka, M. Kukreja, D.A. Routenberg, A.G. Remacle, R.C. Liddington, P. Cieplak, I.A. Kozlov, A.Y. Strongin, Structural and functional diversity of metalloproteinases encoded by the *Bacteroides fragilis* pathogenicity island., FEBS J., 281 (2014) 2487-2502.
[146] T. Goulas, F.X. Gomis-Rüth, 186. Fragilysin., in: N.D. Rawlings, G.S. Salvesen (Eds.) Handbook of proteolytic enzymes., Academic Press, Oxford, Great Britain, 2013, pp. 887-891.
[147] N. Gaci, G. Borrel, W. Tottey, P.W. O'Toole, J.F. Brugere, Archaea and the human gut: new beginning of an old story., World J. Gastroenterol., 20 (2014) 16062-16078.
[148] S.V. Lynch, O. Pedersen, The human intestinal microbiome in health and disease., N. Engl. J. Med., 375 (2016) 2369-2379.
[149] F.E. Dewhirst, T. Chen, J. Izard, B.J. Paster, A.C. Tanner, W.H. Yu, A. Lakshmanan, W.G. Wade, The human oral microbiome., J. Bacteriol., 192 (2010) 5002-5017.
[150] G. Apic, J. Gough, S.A. Teichmann, Domain combinations in archaeal, eubacterial and eukaryotic proteomes., J. Mol. Biol., 310 (2001) 311-325.
[151] L. Aravind, G. Subramanian, Origin of multicellular eukaryotes - insights from proteome comparisons., Curr. Opin. Genet. Dev., 9 (1999) 688-694.
[152] A. Budd, S. Blandin, E.A. Levashina, T.J. Gibson, Bacterial α_2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome?, Genome Biol., 5 (2004) R38.
[153] I. Garcia-Ferrer, P. Arêde, J. Gómez-Blanco, D. Luque, S. Duquerroy, J.R. Castón, T. Goulas, F.X. Gomis-Rüth, Structural and functional insights into *Escherichia coli* α_2-macroglobulin endopeptidase snap-trap inhibition., Proc. Natl. Acad. Sci. USA, 112 (2015) published online on June 22, 2015, doi:2010.1073/pnas.1506538112.
[154] T. Goulas, I. Garcia-Ferrer, A. Marrero, L. Marino-Puertas, S. Duquerroy, F.X. Gomis-Rüth, Structural and functional insight into pan-endopeptidase inhibition by α_2-macroglobulins., Biol. Chem., 398 (2017) in press.
[155] W. Martin, R. Cerff, Prokaryotic features of a nucleus-encoded enzyme. cDNA sequences for chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases from mustard (*Sinapis alba*). Eur. J. Biochem., 159 (1986) 323-331.
[156] R.F. Doolittle, D.F. Feng, K.L. Anderson, M.R. Alberro, A naturally occurring horizontal gene transfer from a eukaryote to a prokaryote., J. Mol. Evol., 31 (1990) 383-388.
[157] T. Kotnik, J.C. Weaver, Abiotic gene transfer: rare or rampant?, J. Membr. Biol., 249 (2016) 623-631.

Supplementary Information
of
Matrix metalloproteinases outside vertebrates

Laura Marino-Puertas, Theodoros Goulas and F. Xavier Gomis-Rüth

Supplementary Figures



Suppl. Fig. 1: Phylogenetic studies. (**A**) Rooted phylogenetic tree reflecting evolutionary distances among selected protein sequences (in parenthesis the UniProt codes) of organisms from different kingdoms of life. Names in *red* are mammals, in *green* plants, in *orange* fungus, in *lilac* viruses, in *blue* bacteria and in *black* archaea. In *blue* background are highlighted proteins of different kingdoms (bacterial and fungal) that display high sequence similarity (see in B). This could be an indication of an eukaryotic-to-prokaryotic horizontal gene transfer event. The bar represents 0.2 substitutions per site. **(B)** Alignment of protein sequences from *Arthrobacter nitrophenolicus* (UniProt code: L8TT82) and *Purpureocillium lilacinum* (UniProt code: A0A179GT13) displaying more than 42% sequence identity in the *blue* backgrounded area. In *red rectangle* the zinc-binding motif (H-E-X-X-H-X-X-G/N-X-X-H/D) found in MMPs.

Supplementary Tables

Suppl. Table 1 — Viral MMP sequences

Family *Ascoviridae*		
Genus *Ascovirus*		
Heliothis virescens ascovirus 3e	474 residues	GB YP_001110872 / UP A4KX75
Heliothis virescens ascovirus 3f	439 residues	GB AJP08985 / UP A0A171PVB2
Heliothis virescens ascovirus 3g	439 residues	GB AFV50272 / UP K4NY21
Spodoptera frugiperda ascovirus 1a	386 residues	GB YP_762369 / UP Q0E587
Trichoplusia ni ascovirus 2c	501 residues	GB YP_803380 / UP Q06VC5
Genus *Toursvirus*		
Diadromus pulchellus ascovirus 4a	223 residues	GB YP_009220674 /UP F2NYV8
Family *Baculoviridae*		
Genus *Betabaculovirus*		
Adoxophyes orana granulovirus	395 residues	GB NP_872491 / UP Q7T9X8
Cryptophlebia leucotreta granulovirus	486 residues	GB NP_891890 / UP Q7T5P6
Cydia pomonella granulovirus #	546 /545 residues	GB AIU36692 / UP A0A097P0M6 // GB NP_148830 / UP Q91F09 *
Helicoverpa armigera granulovirus	596 residues	GB YP_001649020 / UP A9YMN0
Phthorimaea operculella granulovirus	469 residues	GB NP_663206 / UP Q8JS18
Plodia interpunctella granulovirus	620 residues	GB YP_009330170 / UP
Plutella xylostella granulovirus	402 / 403 residues	GB NP_068254 / UP Q9DVZ7 // GB ANY57554 / UP A0A1B2CSG4 *
Pseudalatia unipuncta granulovirus	593 residues	GB YP_003422377 / UP B6S6Q7
Trichoplusia ni granulovirus LBIV-12	593 residues	GB AOW41373 / UP A0A1D8QL47
Xestia c-nigrum granulovirus #	469 residues	GB NP_059188 / UP Q9PZ03
Unclassified *Betabaculoviridae*		
Agrotis segetum granulovirus	481 residues	GB YP_006303 / UP Q6QXG0
Choristoneura occidentalis granulovirus	488 residues	GB YP_654454 / UP Q1A4R3
Clostera anachoreta granulovirus	412 residues	GB YP_004376233 / UP F4ZKQ2
Clostera anastomosis granulovirus	412 residues	GB YP_008719974 / UP U5KB82
Clostera anastomosis granulovirus	486 residues	GB AKS25377 / UP A0A0K0WSE3
Cnaphalocrocis medinalis granulovirus	441 residues	GB ALN41975 / UP A0A0X9FQ45 // GB YP_009229958 / UP A0A109WW48 *
Diatraea saccharalis granulovirus	370 residues	GB YP_009182238 / UP A0A0R7EZ65
Epinotia aporema granulovirus	359 residues	GB YP_006908552 / UP K4ERV0
Erinnyis ello granulovirus	462 residues	GB YP_009091878 / UP A0A097DAI6
Mocis sp. granulovirus	580 residues	GB YP_009249873 / UP A0A162GWM9
Pieris rapae granulovirus	432 residues	GB YP_003429361 / UP D2J4K4
Pieris rapae granulovirus	444 residues	GB ADO85463 / UP E7BN22
Spodoptera frugiperda granulovirus	539 residues	GB YP_009121819 / UP A0A0C5B309
Spodoptera litura granulovirus	464 residues	GB YP_001256988 / UP A5IZN9
Family *Iridoviridae*		
Genus *Iridovirus*		
Invertebrate iridescent virus 6 (Chilo iridescent virus)	264 residues	GB NP_149628 / UP O55761
Invertebrate iridovirus 22	352 residues	GB YP_008357315 / UP S6DDP6
Invertebrate iridescent virus 22	365 residues	GB YP_009010779 / UP W8W1A0
Invertebrate iridovirus 25	362 residues	GB YP_009010550 / UP W8W2D9
Invertebrate iridescent virus 30	367 residues	GB YP_009010310 / UP W8W249
Wiseana iridescent virus (Insect iridescent virus type 9)	346 residues	GB YP_004732919 / UP G0T5G2
Genus *Chloriridovirus*		
Aedes taeniorhyncus iridescent virus	363 residues **	GB ABF82125
Invertebrate iridescent virus 3 (IIV-3)	363 residues **	GB YP_654667 / UP Q196W5

(Mosquito iridescent virus)		
Unclassified *Iridoviridae*		
Anopheles minimus irodovirus	366 residues	GB YP_009021109 / UP W8QE20
Family *Nudiviridae*		
Genus *Betanudivirus*		
Helicoverpa zea nudivirus 2 (HzNV-2)	789 residues	GB YP_004956816 / UP G9I094
Heliothis zea nudivirus	792 residues	GB AAN04364 / UP Q8JKP2
Family *Poxviridae* (Subfamily *Entomopoxvirinae*)		
Genus *Alphaentomopoxvirus*		
Anomala cuprea entomopoxvirus	268 residues	GB YP_009001641 / UP W6JIV8
Genus *Betaentomopoxvirus*		
Amsacta moorei entomopoxvirus (AmEPV)	252 residues	GB AAG02776, GB NP_064852 / UP Q9EMX9 **
Mythimna separata entomopoxvirus 'L'	411 residues	GB YP_008003705 / UP R4ZFL1
Unclassified dsDNA viruses, no RNA stage		
Apis mellifera filamentous virus	830 residues	GB AKY03287, GB YP_009165969 / UP A0A0K1Y874 **
Apis mellifera filamentous virus	1354 residues	GB AKY03074, GB YP_009165756 / UP A0A0K1Y866 **

All sequence codes are from UniProt (UP; www.uniprot.org) or GenBank (GB; www.ncbi.nlm.nih.gov/genbank).
Sequence searches were performed with MMP CDs within UniProt (www.uniprot.org) or the National Center for Biotechnology Information (blast.ncbi.nlm.nih.gov/Blast.cgi) using standard parameters.
Taxonomy according to the International Committee on Taxonomy of Viruses (ictvonline.org/virusTaxonomy.asp), NCBI (www.ncbi.nlm.nih.gov/taxonomy) or the Catalogue of Life (http://www.catalogueoflife.org/col; [1]).
[#] *Xestia* and *Cydia* MMPs are the only viral enzymes studied.
* These entries have only minimal differences.
** These entries are identical.
Searches completed on 10 January, 2017.

Suppl. Table 2 — Selected archaeal MMP sequences

Phylum *Euryarchaeota*		
Class *Halobacteria*		
Haladaptatus sp.	344 residues	GB WP_066144928 / UP A0A166SWC5
Halomicrobium mukohataei *(Haloarcula mukohataei)*	392 residues	UP C7NZX4
Class *Methanomicrobia*		
Methanocalculus sp.	177 residues	UP A0A101H2S1 *
Methanococcoides burtonii	231 residues	GB WP_011499923 / UP Q12UU6
Methanococcoides methylutens	233 residues	UP A0A099T1M7
Methanoculleus sp.	194 residues	UP A0A0Q1AIF2
Methanohalobium evestigatum	240 residues	GB WP_013195212 / UP D7EBD0
Methanohalophilus mahii	231 residues	GB WP_013038296 / UP D5E866
Methanolobus tindarius	241 residues	GB WP_023845349 / UP W9DPH7
Methanolobus psychrophilus	235 residues	GB WP_015053499 / UP K4MC96
Methanomicrobiales archaeon	177 residues	UP A0A117MHP9 *
Methanomethylovorans hollandica	242 residues	GB WP_015323622 / UP L0KWK3
Methanoregula formicica	179 residues	GB WP_015286272 / UP L0HJR6
Methanosalsum zhilinae *(Methanohalophilus zhilinae)*	227 residues	GB WP_013899163 / UP F7XME3
Methanosarcina acetivorans	211 residues	UP Q8TIZ9
Methanosarcina barkerii	217 residues	UP A0A0E3R7C9
Methanosarcina flavescens	217 residues	GB WP_054299817 / UP A0A0P6VKR0
Methanosarcina horonobensis	210 residues	UP A0A0E3SB37
Methanosarcina lacustris	215 residues	GB WP_048124635 / UP A0A0E3RZU1
Methanosarcina mazei *(Methanosarcina frisia)*	220 residues	GB KKG06863 / UP A0A0F8EZR8
Methanosarcina siciliae	211 residues	UP A0A0E3PQS9
Methanosarcina vacuolata	217 residues	UP A0A0E3LI77
Phylum *Thaumarchaeota*		
Class — / Order *Cenarchaeales*		
Cenarchaeum symbiosum	345 residues	UP A0RWX1
Class — / Order *Nitrosopumilales*		
Candidatus Nitrosoarchaeum koreensis	253 residues	UP F9CWY2
Candidatus Nitrosoarchaeum limnia	283 residues	GB WP_010190351 / UP S2E6F9
Candidatus Nitrosopumilus adriaticus	296 residues	UP A0A0D5C2T1
Candidatus Nitrosopumilus koreensis	238 residues	UP K0B4I4
Candidatus Nitrosopumilus salaria	269 residues	GB WP_008298808 / UP I3D381
Class — / Order *Nitrososphaeria*		
Candidatus Nitrososphaera evergladensis	251 residues	UP A0A075MRP9
Unclassified *Thaumarcheota*		
Candidatus Nitrosotalea devanaterra	423 residues	UP A0A128A209
Thaumarchaeota archaeon	540 residues	GB WP_048194249 / UP V6AQP1
Candidatus Nitrosopelagicus brevis	505 residues	UP A0A0A7V069
Unclassified *Archaea*		
Candidatus Pacearchaeota archaeon	301 residues	UP A0A1F6ZFA3

All sequence codes are from UniProt (UP; www.uniprot.org) or GenBank (GB; www.ncbi.nlm.nih.gov/genbank).

Sequence similarity searches were performed with karilysin CD (UP D0EM77) within UniProt using standard parameters. Only one representative from each species or genus (when sp.) was chosen. All sequences were manually curated.

Taxonomy according to the Catalogue of Life (http://www.catalogueoflife.org/col; [1]) or, when absent, UniProt (www.uniprot.org).
* Identical sequences.
Searches completed on 23 January, 2017.

Suppl. Table 3 — Selected bacterial MMP sequences

Phylum *Acidobacteria*		
Class *Solibacteres*		
Solibacter usitatus	442 residues	GB WP_011688765 / UP Q02908
Phylum *Actinobacteria*		
Class *Acidimicrobiia*		
Acidithrix ferrooxidans	358 residues	UP A0A0D8HKH4
Class *Actinobacteria*		
Pseudarthrobacter phenanthrenivorans (Arthrobacter phenanthrenivorans)	687 residues	GB WP_013599449 / UP F0M7V4
Pseudarthrobacter siccitolerans	686 residues	UP A0A024GZ92
Streptomyces lincolnensis	610 residues	GB WP_067438548 / UP A0A1B1MG04
Streptomyces scabiei	359 residues	UP A0A086GV39
Class - / Order *Actinomycetales*		
Alloactinosynnema sp.	556 residues	GB WP_054055428 / UP A0A0H5CVK8
Arthrobacter nitrophenolicus	634 residues	GB WP_009356397 / UP L8TT82
Knoellia aerolata	271 residues	GB WP_052113283 / UP A0A0A0JTC1
Pseudonocardia dioxanivorans	500 residues	GB WP_013678450 / UP F4CV22
Phylum *Bacteroides*		
Class *Bacteroidia*		
Parabacteroides distasonis	446 residues	UP A0A174VNZ8
Tannerella forsythia (Bacteroides forsythus) *	472 residues	UP D0EM77
Class *Cytophagia*		
Cyclobacterium marinum (Flectobacillus marinus)	482 residues	GB WP_014022474 / UP G0J0B7
Phylum *Cyanobacteria*		
Class *Cyanophyceae*		
Microcystis aeruginosa	268 residues	GB WP_052734169 / UP A0A0F6U407
Synechococcus sp.	439 residues	UP B4WJ70
Class - / Order *Synechococcales*		
Acaryochloris marina	580 residues	GB WP_012164578 / UP B0CCT5
Phormidesmis priestleyi Ana	516 residues	UP A0A0P7ZK60
Phylum *Firmicutes*		
Class *Bacilli*		
Bacillus cereus	258 residues	GB WP_000775479 / UP R8XLT0
Bacillus mycoides	253 residues	UP A0A090ZAN4
Halobacillus halophilus (Sporosarcina halophila)	390 residues	GB WP_014642607 / UP I0JKN6
Lactobacillus acidifarinae	228 residues	GB WP_057801390 / UP A0A0R1LQY6
Lactobacillus acidophilus	222 residues	GB WP_003548046, GB YP_194247 / UP Q5FJA8
Lactobacillus apodemi	193 residues	GB WP_035459587 / UP A0A0R1TU85
Lactobacillus brevis	223 residues	GB WP_011666964 / UP Q03TZ3
Lactobacillus plantarum	266 residues	GB WP_003643352 / UP M4KPC6
Lactobacillus nodensis	240 residues	UP A0A0R1KCP7
Lactobacillus senmaizukei	238 residues	UP A0A0R2DS50

Lactobacillus sunkii	233 residues	GB WP_057825587 / UP A0A0R1KWI9
Leuconostoc kimchii	226 residues	GB WP_013103335 / UP D5T391
Paenibacillus terrae	280 residues	GB WP_044649112 / UP A0A0D7WTR0
Streptococcus suis	231 residues	GB WP_044681537 / UP A0A0Z8JY51
Phylum *Nitrospirae*		
Class *Nitrospira*		
Candidatus Magnetoovum chiemensis	426 residues	UP A0A0F2J1S3
Phylum *Plantomycetes*		
Class *Plantomycetia*		
Candidatus Brocadia fulgida	418 residues	UP A0A0M2UXB6
Candidatus Jettenia caeni	420 residues	GB WP_007220894 / UP I3IJ81
Candidatus Scalindua brodae	382 residues	UP A0A0B0EQJ8
Class *Plantomycetacia*		
Isosphaera pallida	435 residues	GB WP_013565370 / UP E8QY87
Pirellula staleyi (Pirella staleyi)	281 residues	GB WP_012911897 / UP D2R985
Planctomyces sp.	499 residues	GB WP_068412701 / UP A0A142YDV0
Singulisphaera acidiphila	538 residues	UP L0DHV3
Phylum *Proteobacteria*		
Class *Alphaproteobacteria*		
Aurantimonas sp.	321 residues	GB WP_055845852 / UP A0A0Q6EYY6
Aureimonas sp.	484 residues	GB WP_056689492 / UP A0A0Q6EFI8
Bradyrhizobium elkanii	454 residues	UP A0A1E3EYM1
Bradyrhizobium jicamae	318 residues	GB WP_057837259 / UP A0A0R3LGN7
Bradyrhizobium lablabi	324 residues	GB WP_057857189 / UP A0A0R3N9Q1
Hyphomicrobium nitrativorans	291 residues	GB WP_023788197 / UP V5SJS8
Labrenzia alexandrii (Stappia alexandrii)	378 residues	UP B9QV57
Methylobacterium sp.	470 residues	UP B0UC88
Puniceibacterium sp.	271 residues	GB WP_053078859 / UP A0A0J5QD38
Rhizobium sp.	404 residues	GB WP_062553422 / UP A0A0Q7YFP6
Rhodomicrobium udaipurense	248 residues	GB WP_037237908 / UP A0A037UYG6
Rhodospirillaceae bacterium	357 residues	UP A0A0F2S3S3
Sphingomonas changbaiensis	556 residues	UP A0A0E9MSC9
Sulfitobacter geojensis	501 residues	GB WP_064223228 / UP A0A196QSA6
Class *Betaproteobacteria*		
Burkholderia gladioli (Phytomonas marginata)	463 residues	UP A0A0M2QDA5
Burkholderia vietnamiensis	1065 residues	GB WP_059890570 / UP A0A132D6D9
Paraburkholderia glathei	1101 residues	UP A0A0J1D520
Rubrivivax gelatinosus	633 residues	UP I0HQ73
Class *Deltaproteobacteria*		
Chondromyces apiculatus	463 residues	UP A0A017SZ72
Chondromyces crocatus	433 residues	UP A0A0K1EA09
Labilithrix luteola	314 residues	UP A0A0K1Q883
Sandaracinus amylolyticus	588 residues	GB WP_053234042 / UP A0A0F6W3W9

Sorangium cellulosum (Polyangium cellulosum)	247 residues	UP A0A150T1A1
Stigmatella aurantiaca	477 residues	GB WP_002612142 / UP Q098Y1
Vulgatibacter incomptus	602 residues	GB WP_050725892 / UP A0A0K1PDK5
Class *Gammaproteobacteria*		
Acinetobacter sp.	205 residues	GB WP_005215616 / UP N9NAE2
Methylophaga aminisulfidivorans	237 residues	GB WP_007146704 / UP F5T331
Methylothermaceae bacteria	432 residues	UP A0A148N5W8
Pseudoalteromonas luteoviolacea	1466 residues	GB WP_065788537 / UP A0A1C0TT69
Candidatus Tenderia electrophaga	261 residues	UP A0A0S2TI48
Thiolapillus brandeum	387 residues	UP W0TNH9
Unclassified *Bacteria*		
Dadabacterium bacterium	302 residues	UP A0A0T5ZUN1
Latescibacteria bacterium	337 residues	UP A0A0S7XD53
Parcubacteria bacterium	310 residues	UP A0A0L0LDW4

All sequence codes are from UniProt (UP; www.uniprot.org) or GenBank (GB; www.ncbi.nlm.nih.gov/genbank).

Sequence similarity searches were performed with karilysin CD (UP D0EM77) within UniProt using standard parameters. Only one representative from each species or genus (when sp.) was chosen. All sequences were manually curated.

Taxonomy according to the Catalogue of Life (http://www.catalogueoflife.org/col; [1]) or, when absent, UniProt (www.uniprot.org).

* *T. forsythia* karilysin is the only bacterial MMP studied.

Searches completed on 27 January, 2017.

Supplementary References

[1] Y. Roskov, L. Abucay, T. Orrell, D. Nicolson, N. Bailly, P. Kirk, T. Bourgoin, R.E. de Walt, W. Decock, A. de Wever, E. van Nieukerken, Species 2000 & ITIS Catalogue of Life., Species 2000, Leiden (The Netherlands), 2017.