



**HAL**  
open science

## Identification and characterization of a novel family of cysteine-rich peptides (MgCRP-I) from *Mytilus galloprovincialis*

Marco Gerdol, Nicolas Puillandre, Gianluca de Moro, Corrado Guarnaccia, Marianna Lucafò, Monica Benincasa, Ventislav Zlatev, Chiara Manfrin, Valentina Torboli, Piero Giulio Giulianini, et al.

### ► To cite this version:

Marco Gerdol, Nicolas Puillandre, Gianluca de Moro, Corrado Guarnaccia, Marianna Lucafò, et al.. Identification and characterization of a novel family of cysteine-rich peptides (MgCRP-I) from *Mytilus galloprovincialis*. *Genome Biology and Evolution*, 2015, 7 (8), pp.2203-2219. <10.1093/gbe/evv133>. <hal-03922650>

**HAL Id: hal-03922650**

**<https://hal.science/hal-03922650v1>**

Submitted on 4 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

**Identification and characterization of a novel family of  
cysteine-rich peptides (MgCRP-I) from *Mytilus  
galloprovincialis***

Journal:	<i>Genome Biology and Evolution</i>
Manuscript ID:	GBE-150335
Manuscript Type:	Research Article
Date Submitted by the Author:	27-Mar-2015
Complete List of Authors:	<p>Gerdol, Marco; University of Trieste, Department of Life Sciences          Puillandre, Nicolas; 2Muséum National d'Histoire Naturelle,, Département          Systématique et Evolution          De Moro, Gianluca; University of Trieste, Department of Life Sciences          Guarnaccia, Corrado; International Centre for Genetic Engineering and          Biotechnology,          Lucafò, Marianna; University of Trieste, Department of Life Sciences          Benincasa, Monica; University of Trieste, Department of Life Sciences          Zlatev, Ventislav; International Centre for Genetic Engineering and          Biotechnology,          Manfrin, Chiara; University of Trieste, Department of Life Sciences          Torboli, Valentina; University of Trieste, Department of Life Sciences          Giulianini, Piero; University of Trieste, Department of Life Sciences          Sava, Gianni; University of Trieste, Department of Life Sciences          Venier, Paola; University of Padova, Department of Biology          Pallavicini, Alberto; University of Trieste, Department of Life Sciences</p>
Keywords:	toxin, antimicrobial peptide, bivalve mollusc, mussel, transcriptome

SCHOLARONE™  
Manuscripts

**RESEARCH ARTICLE****Identification and characterization of a novel family of cysteine-rich peptides  
(MgCRP-I) from *Mytilus galloprovincialis***

Marco Gerdol<sup>1</sup>, Nicolas Puillandre<sup>2</sup>, Gianluca De Moro<sup>1</sup>, Corrado Guarnaccia<sup>3</sup>, Marianna Lucafò<sup>1</sup>,  
Monica Benincasa<sup>1</sup>, Ventislav Zlatev<sup>3</sup>, Chiara Manfrin<sup>1</sup>, Valentina Torboli<sup>1</sup>, Piero Giulio  
Giulianini<sup>1</sup>, Gianni Sava<sup>1</sup>, Paola Venier<sup>4</sup>, Alberto Pallavicini<sup>1\*</sup>

<sup>1</sup> Department of Life Sciences, University of Trieste, Via Giorgieri 5, 34127 Trieste, Italy

<sup>2</sup> Muséum National d'Histoire Naturelle, Département Systématique et Evolution, ISyEB Institut  
(UMR 7205 CNRS/UPMC/MNHN/EPHE), 43, Rue Cuvier, 75231 Paris, France

<sup>3</sup> Protein Structure and Bioinformatics Group, International Centre for Genetic Engineering and  
Biotechnology (ICGEB), Padriciano 99, 34149 Trieste, Italy

<sup>4</sup> Department of Biology, University of Padova, Via Ugo Bassi 58/B, 35131 Padova, Italy

\*Author for Correspondence: Alberto Pallavicini, Department of Life Science, University of  
Trieste, Via Giorgieri 5, 34127 Trieste, Italy, telephone number: +39 0405588736, FAX: +39, e-  
mail address: pallavic@units.it

**Data deposition**

MgCRP-I nucleotide sequences were deposited at GenBank under the accession IDs KJ002647-  
KJ002676 and KR017759-KR017770.

## Abstract

We report the identification of a novel gene family (named MgCRP-I) encoding short secreted cysteine-rich peptides in the Mediterranean mussel *Mytilus galloprovincialis*. These peptides display a highly conserved pre-pro region and a hypervariable mature peptide comprising six invariant cysteine residues arranged in three intramolecular disulfide bridges. Although their cysteine pattern is similar to cysteines-rich neurotoxic peptides of distantly related protostomes such as cone snails and arachnids, the different organization of the disulfide bridges and phylogenetic analyses revealed MgCRP-I as a novel protein family. Genome- and transcriptome-wide searches for orthologs in other bivalve species indicated the unique presence of this gene family in *Mytilus* spp.

Like other bioactive cysteine-rich peptides, such as antimicrobial peptides and neurotoxins, MgCRP-I peptides are produced as pre-propeptides, have a net positive charge and likely derive from similar evolutionary mechanisms, i.e. gene duplication and positive selection within the mature peptide region; however, synthetic MgCRP-I peptides did not display significant toxicity in cultured mammalian cells, insecticidal, antimicrobial or antifungal activities. Even though the functional role of MgCRP-I peptides remains puzzling, we discuss their possible involvement in mussel immunity.

## Keywords

Toxin; antimicrobial peptide; bivalve mollusc; mussel; transcriptome.

## Introduction

Marine ecosystems are characterized by an astonishing species diversity, with over 2 million different eukaryotic species belonging to various *phyla* estimated to compose the marine fauna (Mora, et al. 2011). Thus, marine organisms and environments can be regarded as a virtually unlimited source of bioactive compounds, either produced through complex biochemical synthetic reactions or gene-encoded peptides (Mayer, et al. 2011).

Nowadays, computer-assisted data mining coupled with the advent of next-generation sequencing (NGS) technologies allows the *in silico* identification of bioactive molecules also in non-model marine organisms (Li, et al. 2011; Sperstad, et al. 2011). The quick increase of bivalve transcriptome datasets (Suárez-Ulloa, et al. 2013b) and the recent genome sequencing of the oysters *Crassostrea gigas* (Zhang, et al. 2012) and *Pinctada fucata* (Takeuchi, et al. 2012) further broadens the horizons of genetic and genomic studies in bivalve mollusks. Due to their relevance as sea food and sentinel organisms, significant RNA sequencing efforts, both with 454 and Illumina technologies, have been performed on *Mytilus* spp. (Bassim, et al. 2014; Craft, et al. 2010; Freer, et al. 2014; Gerdol, et al. 2014; González, et al. 2015; Philipp, et al. 2012; Romiguiet, et al. 2014; Suárez-Ulloa, et al. 2013a). Moreover, an unrefined genome of the Mediterranean mussel *Mytilus galloprovincialis* recently released, further extends the molecular data available for this species (Nguyen, et al. 2014). The bioinformatic analysis of the mussel data has already contributed to the discovery of important immune-related molecules, including pathogen-recognition receptors, signaling intermediates and antimicrobial peptides (AMPs) (Gerdol, et al. 2012; Gerdol, et al. 2011; Gerdol and Venier ; Rosani, et al. 2011; Toubiana, et al. 2013; Toubiana, et al. 2014). As reported in this paper, large-scale bioinformatic analyses can also drive the discovery of novel gene families encoding peptides with unique chemico-physical properties and/or sequence patterns.

Actually, cysteine-rich peptides (CRPs) encompass a large and widespread group of secreted bioactive molecules, heterogeneous in primary sequence and structural arrangement, with

1  
2  
3 different functional roles and present in almost all living organisms, from bacteria to fungi, animals  
4  
5 and plants (Gruber, et al. 2007; Marshall, et al. 2011; Taylor, et al. 2008). Invertebrate CRPs are  
6  
7 particularly abundant and they have been frequently related to the immune defense against potential  
8  
9 pathogens (Mitta, et al. 2000b). According to the number of cysteine residues and their arrangement  
10  
11 in the tridimensional space, many families of cysteine-rich AMPs have been described in  
12  
13 invertebrates, as for instance in crustaceans (Bartlett, et al. 2002; Destoumieux, et al. 1997), insects  
14  
15 (Bulet and Stöcklin 2005) and arachnids (Ehret-Sabatier, et al. 1996; Fogaça, et al. 2004). In  
16  
17 *Mytilus* spp., different families of CRPs have been progressively discovered: defensins (Charlet, et  
18  
19 al. 1996), mytilins (Mitta, et al. 2000a), myticins (Mitta, et al. 1999), mytimycins (Charlet, et al.  
20  
21 1996), macins, big defensins (Gerdol, et al. 2012) and myticusins (Liao, et al. 2013).

22  
23  
24  
25 Other evolutionarily related CRPs are animal venom components which possess neurotoxic  
26  
27 properties, since they can selectively block various types of ion channels for predation or defense  
28  
29 (Froy and Gurevitz 2004; Rodríguez de la Vega and Possani 2005). Notably, spider and scorpion  
30  
31 venoms contain an extraordinary mixture of cysteine-rich peptides whose complexity has been only  
32  
33 recently fully appreciated by 'omics approaches (Ma, et al. 2009; Zhang, et al. 2010). Even within  
34  
35 the Mollusca phylum some species have developed a lethal venom arsenal to be used for predation:  
36  
37 marine gastropods of the genus *Conus* indeed use a modified radula as a sting to inject and paralyze  
38  
39 their prey with a powerful venom cocktail, mostly of peptidic nature (Olivera, et al. 2012). Due to  
40  
41 their biological properties, many CRPs have been studied to guide the development of new drugs  
42  
43 for therapeutic applications in both human and veterinary medicine (Adams, et al. 1999; Otero-  
44  
45 González, et al. 2010; Saez, et al. 2010).

46  
47  
48  
49 Despite having physico-chemical properties similar to cysteine-rich AMPs and toxins,  
50  
51 certain animal CRPs lack the expected activities and are instead involved in diverse functions:  
52  
53 among these, Kunitz-type (Ranasinghe and McManus 2013) and Kazal-type (Rimphanitchayakit  
54  
55 and Tassanakajon 2010) proteinase inhibitors represent two widespread groups.

1  
2  
3 The abundance and diversity of the CRPs described in protostomes is remarkable and, given  
4 the poor genomic knowledge of many taxonomic groups, a large part of these peptides probably still  
5 remain to be uncovered. In this paper, we report the application of a genome- and transcriptome-  
6 scale approach to the identification of novel cysteine-rich peptides the Mediterranean mussel *M.*  
7 *galloprovincialis*. In agreement to nomenclature criteria reported elsewhere (Gerdol and Venier),  
8 we present the new MgCRP-I family, characterized by a conserved pre-pro region and an highly  
9 variable mature peptide with six conserved cysteine residues organized in the consensus C(X<sub>3</sub>-  
10 <sub>6</sub>)C(X<sub>1-7</sub>)CC(X<sub>3-4</sub>)C(X<sub>3-5</sub>)C. We investigated the organization and evolution of mussel MgCRP-I  
11 genes, the main features and possible functional roles of the encoded peptides, and finally we  
12 discuss their possible involvement in innate immune defense.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Materials and methods

### Identification of MgCRP-I sequences in mussel transcriptomes

The *M. galloprovincialis* Illumina transcriptomes available at the NCBI Sequence Reads Archive (retrieved in February, 2015) were assembled with Trinity v.2014-07-17 (Grabherr, et al. 2011) and with the CLC Genomics Workbench 7.5 (CLC Bio, Aarhus, Denmark), using default parameters. Following translation of the assembled contigs into the 6 possible reading frames with EMBOSS TranSeq (Rice, et al. 2000) we investigated the virtual mussel proteins for the presence of the C(X<sub>1-10</sub>)C(X<sub>1-10</sub>)CC(X<sub>1-10</sub>)C(X<sub>1-10</sub>)C signature, using a Perl script developed in-house (available upon request to the corresponding author). Matching sequences were aligned with MUSCLE (Edgar 2004) to generate a HMMER v3.0 profile (Eddy 2011) which was then used to retrieve partial-matching cases within the assembly (e-value cutoff 1x10<sup>-5</sup>). The procedure was re-iterated until no additional matches could be retrieved. Sequences showing an identity higher than 95% at the nucleotide level were considered as redundant and collapsed in a single consensus sequence, unless they were confirmed by genomic evidence (see section below).

### Identification of MgCRP-I genes in the mussel genome

The *M. galloprovincialis* genomic contigs (Nguyen, et al. 2014) were downloaded from GenBank and scanned for the presence of MgCRP-I genes as follows: (i) genes were identified based on BLASTn identity (Altschul, et al. 1990) to the previously identified MgCRP-I transcripts (e-value threshold of 1x10<sup>-5</sup>); (ii) genomic scaffolds were translated into the six possible reading frames with the EMBOSS Transeq tool (Rice, et al. 2000) and novel MgCRP-I loci were identified with HMMERv 3.0 (Eddy 2011).

The genes identified were manually annotated with mRNA and CDS traces, based on: (i) MUSCLE alignment between genomic contigs and the corresponding assembled transcripts, whenever available; (ii) mapping of the available *M. galloprovincialis* sequencing reads (see

1  
2  
3 above), with the CLC Genomics Workbench *large gap mapping* tool; (iii) refinement of splice site  
4 positions with Genie (Reese, et al. 1997). An example of the results of the annotation procedure is  
5 shown in **Supplementary Figure 1 (Supplementary material online)**. Results obtained from the  
6 genome and transcriptome analyses were compared and redundant results (identity percentage  
7 higher than 95%) were removed, unless multiple gene copies were confirmed in the mussel genome  
8 (e.g. the presence of paralogous genes was tolerated).  
9  
10  
11  
12  
13  
14  
15  
16  
17

### 18 **MgCRP-I protein sequence analysis**

19  
20 Protein translations of mussel genes and transcripts matching the sequence pattern  
21 mentioned above were further analyzed as follows: the presence of a signal peptide was detected  
22 with SignalP 4.0 (Petersen, et al. 2011), and discriminated from transmembrane domains with  
23 Phobius (Käll, et al. 2004). Potential sites of post-translational proprotein convertase cleavage were  
24 identified with ProP 1.0 (Duckert, et al. 2004). Possible post-translational C-terminal cleavage sites  
25 by carboxypeptidase E or by peptidylglycine,  $\alpha$ -amidating monooxygenase were detected with  
26 ELM (Dinkel, et al. 2011). The subcellular localization was predicted (for full-length peptides only)  
27 with TargetP 1.1 (Emanuelsson, et al. 2007). Isoelectric point and molecular weight of the predicted  
28 mature peptides was calculated at ExPASy ([http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/)). Structural  
29 homologies with proteins deposited in the RCSB Protein DataBank database were investigated by  
30 Phyre2 (Kelley and Sternberg 2009).  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44

45 The probabilities of codon bias for the six conserved cysteines and for the arginine residue  
46 responsible of the post-translational pro region cleavage were calculated assuming a binomial  
47 distribution, based on the codon usage inferred from the *M. galloprovincialis* transcriptome  
48 (Gerdol, et al. 2014) and using the tool *cusp* included in the EMBOSS package (Rice, et al. 2000).  
49

50 We used PAML 4.7 (Yang 2007) and the graphical interface PAMLX 1.2 (Xu and Yang 2013) to  
51 test whether some sites in the codon-based alignments of MgCRP-I nucleotide sequences were  
52 under positive selection. In detail, only full-length coding sequences were processed and two site  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 models were compared: M1, which assumes that the dN/dS ratio along the sequence ranges from 0  
4 to 1 (purifying selection to neutral drift), and M2, which assumes that a few sites have a dN/dS ratio  
5 (i.e.,  $\omega > 1$ ; positive selection). The likelihoods of the two models were compared using a likelihood  
6 ratio test (LRT) with a  $\chi^2$  distribution, with 2 degrees of freedom. The Empirical Bayes approach  
7 was used to calculate the posterior probabilities (PP) for site classes. Positive selection was  
8 concluded at  $PP > 0.95$ .  
9  
10  
11  
12  
13  
14  
15  
16  
17

### 18 **Comparative genomics analyses**

19  
20 The NGS Illumina transcriptome data available for 70 bivalve species were downloaded  
21 from the Sequence Read Archive (SRA). The full list and the corresponding Bioproject accession  
22 IDs are shown in **Supplementary Table S1, supplementary material online**. The bivalve  
23 sequence datasets were independently *de novo* assembled with the CLC Genomic Workbench 7.5  
24 (CLC Bio, Aarhus, Denmark). All transcriptomes were translated into the six possible open reading  
25 frames with EMBOSS Transeq (Rice, et al. 2000) and significant similarity with MgCRP-I proteins  
26 was assessed with BLASTp (e-value threshold 0.01) and HMMER v 3.0 using the protein profile  
27 mentioned above (p-value threshold 0.01).  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37

38 The complete UniProtKB/Swiss-Prot protein sequence database and the whole set of  
39 peptides predicted from the fully sequenced genomes of *C. gigas* (release 9) (Wang, et al. 2012) and  
40 *P. fucata* (v 1.0) (Takeuchi, et al. 2012) were screened for the presence of the C-C-CC-C-C pattern  
41 with a custom Perl script, without any constraint about the spacing between cysteine residues. Only  
42 full-length sequences shorter than 100 amino acids and showing a signal peptide by SignalP 4.0  
43 (Petersen, et al. 2011)) were selected. Sequences bearing more than 7 cysteine residues within the  
44 mature region were considered as characterized by more complex disulfide arrays and discarded.  
45 The possible presence of mis-annotated CRP-I-like genes in *C. gigas* and *P. fucata* was evaluated  
46 by performing the same analyses on the genomic scaffolds translated into the 6 possible reading  
47 frames with EMBOSS TranSeq.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### Phylogeny and evolutionary tests

We used all the available MgCRP-I proteins, their orthologous sequences identified in *Mytilus edulis* (MeCRP-I), *Mytilus californianus* (McCRP-I) and *Mytilus trossulus* (MtCRP-I), and the positive hits resulting from the data mining (see Materials and methods), to infer the phylogenetic relationships among sequences bearing a similar cysteine signature. Given the high sequence diversity, only the signal peptides, as predicted using SignalP 4.0 (Petersen, et al. 2011), were retained to facilitate sequence alignment. Following alignment with Muscle (Edgar 2004) and manual refinement, maximum likelihood analyses were performed with RaxML (Stamatakis 2006) as implemented on the CYPRES Portal ([www.phylo.org/portal2](http://www.phylo.org/portal2)), using the RAxML-HPC2 on TG Tool. The robustness of the nodes was assessed with a bootstrapping procedure of 100 replicates. Because the evolutionary relationships of the peptides included in the analysis with other peptides were unknown, no outgroup was considered and a mid-point rooting strategy was applied. A similar analysis was performed with a dataset that included only the *M. galloprovincialis* CRP-I peptides. Partial MgCRP-I peptides with an incomplete signal peptide were excluded from the analyses.

### Peptide synthesis, oxidative folding and disulfide mapping

MgCRP-I 7 (26 aa) and MgCRP-I 9 (26 aa) peptides were synthesized according standard solid-phase Fmoc chemistry using 4 equivalents of HCTU/Fmoc-Xaa-OH/DIEA (0.95/1.00/1.90) with respect to the resin loading (Tentagel S-Trt, 0.2mmol/g (Sigma-Aldrich, St. Louis, MO)). The synthesis was semi-automatically performed with a customized Gilson Aspec XL peptide synthesizer (Middleton, WI) on a 0.05 mmol scale. Cysteines were manually added as N- $\alpha$ -Fmoc-S-trityl-L-cysteine pentafluorophenyl ester in order to minimize racemization. After cleavage from the resin the peptides were precipitated with diethylether, washed and freeze-dried. The peptides reduced by TCEP treatment were purified by RP-HPLC on a semipreparative Zorbax 300SB-C18

1  
2  
3 9.4x250mm column (Agilent, Santa Clara, CA) using a gradient from water/0.1% TFA to  
4  
5 acetonitrile/0.1% TFA.  
6

7 Pure peptide fractions from RP-HPLC were checked by electrospray mass spectrometry  
8 (amaZonSL iontrap, Bruker (Billerica, MA)) and immediately diluted at 0.1 mg/ml in either of the  
9 following refolding buffers: (i) RefoldA: 0.2 M Tris-HCl, 2 mM EDTA, 10 mM GSH, 1 mM  
10 GSSG, pH 8, previously degassed with argon bubbling; (ii) RefoldB: 50 mM NaOAc, 1 mM  
11 EDTA, 1 mM GSH, 0.1 mM GSSG, 2M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, pH 7.7. The oxidative refolding proceeded for  
12 18 h at 4°C, was quenched by TFA addition and finally checked by LC-MS analysis.  
13  
14  
15  
16  
17  
18  
19

20 All proteolysis reactions were carried out at 37 °C for 18-48 h in sodium acetate buffer (100  
21 mM, pH 5.5) containing 1 M GuHCl and 5 mM CaCl<sub>2</sub>. The purified MgCRP-I 7 peptide (60 µg)  
22 was dissolved in 90 µl of buffer and trypsin (3 µg) was added. A second aliquot of MgCRP-I 7  
23 (60µg in 90µl) was incubated for 48 h at 37 °C in the presence of chymotrypsin (6 µg). Digestions  
24 of MgCRP-I 9 with trypsin and chymotrypsin were carried out in the same conditions. The  
25 digestions were quenched using formic acid (1% final) and the proteolytic fragments were  
26 fractionated by RP-HPLC (column Jupiter C18, 1x50mm, Phenomenex (Torrance, CA) using a  
27 gradient from water/0.1% formic acid to 60% acetonitrile and analyzed by LC-MS/MS (amaZonSL,  
28 Bruker (Billerica, MA)).  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

### 43 **Gene expression analysis**

44 The expression levels of selected MgCRP-I genes were evaluated in samples representing  
45 hemolymph, digestive gland, inner mantle, mantle rim, gills, foot and posterior adductor muscle.  
46 Total RNA was extracted from the tissues of 30 adult specimens (5-7 cm shell length) collected  
47 from the Gulf of Trieste, Italy, homogenized in equal quantity in Trizol® (Life Technologies,  
48 Carlsbad, CA) according to the manufacturer's protocol. RNA quality was assessed by  
49 electrophoresis on denaturing agarose gel and its quantity was estimated by UV-spectrophotometry.  
50 cDNAs were prepared using a qScript™ cDNA Synthesis Kit (Quanta BioSciences Inc.,  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Gaithersburg, MD) according to the manufacturer's instructions. Primer pairs were designed to  
4  
5 obtain the specific PCR amplicons (**Table 1**), with the exception of the primer pairs co-amplifying  
6  
7 the paralogous sequences MgCRP-I 3/25 and MgCRP-I 10/26. The 15  $\mu$ L PCR reaction mix  
8  
9 comprised 7.5  $\mu$ L of SsoAdvanced™ SYBR® Green Supermix (Bio-rad, Hercules, CA), 0.3  $\mu$ L of  
10  
11 each of the two 10  $\mu$ M primers and 2  $\mu$ L of a 1:20 cDNA dilution.

12  
13  
14 The following thermal profile was used for qPCR amplification in a C1000 thermal cycler  
15  
16 (Bio-Rad, Hercules, CA): an initial denaturation step at 95° C for 3', followed by 40 cycles at 95°  
17  
18 for 5" and 55° for 30". The products of amplification were analyzed with a 65°/95° C melting  
19  
20 curve. The expression of the selected genes was calculated with the delta Ct method; Ct values were  
21  
22 corrected based on primer pairs PCR efficiencies using Lin-RegPCR (Ramakers, et al. 2003) and  
23  
24 expression values were normalized using the elongation factor EF-1 as a housekeeping gene.  
25  
26 Results are shown as the mean with standard deviation of three technical replicates.  
27  
28  
29  
30  
31

### 32 **Cytotoxicity assays**

33  
34 Human colorectal carcinoma (HT-29), human neuroblastoma (SHSY5Y) and breast cancer  
35  
36 (MDAMB231) cell lines were used for the cytotoxicity assays. HT-29 was maintained in RPMI-  
37  
38 1640 and MDAMB231 was maintained in Dulbecco's Modified Eagle's Medium (DMEM): the  
39  
40 culture medium was supplemented with 10% (v/v) fetal bovine serum (FBS), penicillin (100  
41  
42 U/mL), streptomycin (100  $\mu$ g/mL), and L-glutamine 2 mM. SHSY5Y was cultured in DMEM  
43  
44 medium supplemented with penicillin (100 U/mL), streptomycin (100  $\mu$ g/mL), L-glutamine 2 mM  
45  
46 and with 10% heat-inactivated fetal bovine serum. Cells were grown at 37 °C in a 95 % air and 5 %  
47  
48 CO<sub>2</sub> humidified incubator.  
49  
50

51  
52 HT-29, SHSY5Y and MDAMB231 were harvested by trypsinization and plated into 96-well  
53  
54 culture plates at a density of approximately  $1.5 \times 10^4$  cells per well. After 24 hrs of incubation,  
55  
56 different concentrations of MgCRP-I 7 and 9 (10, 1, 0.1 and 0.01  $\mu$ M) dissolved in culture medium  
57  
58  
59  
60

1  
2  
3 were added to each well. Then, the samples were incubated 24 hrs at 37 °C in the humidified  
4 atmosphere (5 % CO<sub>2</sub>). The colorimetric 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium  
5 bromide (MTT) assay was performed to assess the metabolic activity of cells treated as described  
6 above. Aliquots of 20 µl stock MTT (5 mg/mL) were added to each well, and cells were then  
7 incubated for 4 hrs at 37 °C. Cells were lysed with isopropanol HCl 0.04 N. Absorbance was  
8 measured at 540 and 630 nm using a microplate reader (Automated Microplate Reader EL311,  
9 BIOTEK® Instruments, Vermont, USA). All measurements were done in six technical replicates,  
10 and three independent experiments were carried out.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22

### 23 **Insecticidal test**

24  
25 The synthetic peptides MgCRP-I 7 and 9 were dissolved in PBS and injected with a sterile  
26 syringe in *Zophobas morio* larvae (~50 mm long, weighting ~500 mg). The control group (n = 10)  
27 was injected with a volume of 50 µl PBS. Two experimental groups of larvae for each peptide (n =  
28 10) were injected with 30 and 300 µg peptide/Kg body weight respectively, for an injection volume  
29 of 50 µl. Larvae were monitored for signs of neurotoxicity for 48 hours, including lack of  
30 movement, twitching and death. During the experimental time course, larvae were not fed and kept  
31 at room temperature. The median lethal dose (LD50) for the two mussel peptides was calculated  
32 according to Tedford and colleagues (2001).  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

### 46 **Bacterial/fungal strains and minimum inhibitory concentration (MIC assay)**

47  
48 The growth inhibitory effect of MgCRP-I 7 and 9 was tested on *Escherichia coli* ATCC  
49 25922, *Staphylococcus aureus* ATCC 25923, two strains of *Candida albicans* (ATCC 90029 and a  
50 clinical isolate), and on four strains of *Cryptococcus neoformans* (ATCC 90112, ATCC 52816,  
51 ATCC 52817 and a clinical isolate). The bacterial inoculum was incubated overnight in Mueller-  
52 Hinton Broth (MHB, Difco) at 37°C with shaking. For the assays, the overnight bacterial cultures  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 were diluted 1:30 in fresh MHB and incubated at 37°C with shaking for approximately two hours to  
4  
5 obtain a mid-logarithmic phase bacterial culture. Fungi were grown on Sabouraud agar (Difco)  
6  
7 plates at 30°C for 48 hours. Fungal suspensions were prepared by picking and suspending five  
8  
9 colonies in 5 ml of sterile PBS. The turbidity of the bacterial or fungal suspensions was measured at  
10  
11 600 nm and was adjusted to obtain the appropriate inoculum according to previously derived curves  
12  
13 relating the number of colony forming units (CFUs) with absorbance.  
14  
15

16 The antimicrobial activity was evaluated by the broth micro-dilution susceptibility assay  
17  
18 performed according to the guidelines of the Clinical and Laboratory Standards Institute (CLSI) and  
19  
20 previously described (Benincasa, et al. 2010; Benincasa, et al. 2004). Briefly, two-fold serial  
21  
22 dilutions of MgCRP-I 7 and MgCRP-I 9 were prepared in 96-well microplates in the appropriate  
23  
24 medium, to a final volume of 50 µl. Fifty µl of bacterial suspension in MHB, or fungal suspension  
25  
26 in Sabouraud, were added to each well to a final concentration of  $1-5 \times 10^5$  cells/ml for bacteria, and  
27  
28  $5 \times 10^4$  cells/ml for fungi. Bacterial and fungal samples were then incubated at 37°C for 24h or 30°C  
29  
30 for 48h, respectively. The MIC (Minimum Inhibitory Concentration) was taken as the lowest  
31  
32 concentration of peptide resulting in the complete inhibition of visible growth after incubation. All  
33  
34 tests were performed in triplicate.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Results and discussion

### MgCRP-I sequence features

Overall, we identified 67 different MgCRP-I sequences (**Table 2**). More in detail, 48 sequences could be identified in over 2 million genomic contig sequences, which provide a preliminary view of the mussel genome (Nguyen, et al. 2014). Twenty-three of them were also detected as expressed transcripts in publicly available RNA-seq data. In addition, 19 expressed sequences with no match in the genomic contigs were identified, likely corresponding to genes located in genomic regions which are still not covered by the assembly. Overall, 41 sequences can be considered of full length, as the entire CDS from the initial ATG to the STOP codon was represented. The remaining partial sequences (**Table 2**) were lacking either the 5' or the 3' end, due to truncated genomic contigs or low read coverage from RNA-seq data.

The MgCRP-I peptides are secreted pro-peptides characterized by the highly conserved cysteine array C-C-CC-C-C: all the members of this family display an unambiguous N-terminal signal peptide cleavage site, followed by a ~15 residues long pro-region ending with a highly conserved dibasic cleavage site for proprotein convertases (KR or, more rarely, RR). While the signal peptide and pro-peptide regions show low sequence variability, the C-terminal region of MgCRP-I corresponding to the putative mature peptide, appears hypervariable (**Figure 1**). The six invariant cysteine residues involved into the formation of intramolecular disulfide bridges are embedded within this highly variable region. The two central cysteine residues (Cys3 and Cys4) are directly linked with a peptidic bond. As a result, the basic consensus of these peptides can be defined as C(X<sub>3-6</sub>)C(X<sub>1-7</sub>)CC(X<sub>3-4</sub>)C(X<sub>3-5</sub>)C (Figure 1 and 2).

Most MgCRP-I peptides present a short C-terminal extension and, after the 6<sup>th</sup> cysteine, they often display dibasic amino-acidic motifs which might be the target of post-translational cleavage by carboxypeptidase E, one of the most common modifications observed in neurotoxic peptides

1  
2  
3 from invertebrates such as scorpion venoms (Xiong, et al. 1997) and conotoxins (Fan, et al. 2003;  
4 Wang, et al. 2003). Secreted cysteine-rich peptides are among the molecules undergoing the largest  
5 amount of different post-translational modifications, as demonstrated by the case of conotoxins  
6 (Bergeron, et al. 2013; Craig, et al. 1999) and the case of MgCRP-I peptides could be similar; given  
7 the difficulty of obtaining purified peptides from mussel tissues due to their low expression levels,  
8 in the absence of proteomic studies we had to rely on *in silico* prediction of the most likely  
9 modification sites.  
10  
11  
12  
13  
14  
15  
16  
17

18 Based on the predicted proteolytic cleavages, and with a few exceptions, the virtual  
19 MgCRP-I mature peptides are 25-38 amino acids long with estimated molecular weight of 2.5-4  
20 KDa. Almost invariably, the mature peptides have a basic isoelectric point (mostly between 8 and  
21 9.5), indicative of positive net charge at physiological pH, which might be balanced by the presence  
22 of conserved negatively charged residues in the pro-region (**Figure 1**); this feature might be  
23 important for the biological activity of MgCRP-I peptides, as it is maintained in all sequences  
24 despite their remarkable sequence variability.  
25  
26  
27  
28  
29  
30  
31  
32  
33

34 In addition to the standard pre-pro-peptide organization described above, several peculiar  
35 transcripts, named “multi-MgCRP-I” encoding precursors characterized by multiple cysteine-rich  
36 modules were also identified (**Table 2**). The modules, ranging from 2 to 4 in number, are  
37 structurally close to each other, as each Cys6 of the N-terminal module is separated by just 2-4  
38 residues from the Cys1 of the following one. Cysteine-rich domains of multi-MgCRP-I do not show  
39 any peculiarity compared to those of regular, mono-domain peptides (see **supplementary figure**  
40 **S3, supplementary material online**) and the different domains within the same sequence are likely  
41 derived by the duplication of a single original module. Although the functional significance of the  
42 multi-MgCRP-I peptides is still unknown, the maintenance of cysteine pattern and isoelectric points  
43 within each single domain (data not shown) suggest that these long precursors might be post-  
44 translationally cleaved into smaller functional peptides. In this case, such a process would be an  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 interesting strategy adopted to achieve the co-expression of different variants, in a similar fashion to  
4  
5 other invertebrate AMPs (Herbst, et al. 2004; Ratzka, et al. 2012).  
6  
7

### 8 9 10 **Structure of MgCRP-I genes**

11  
12 Despite the low average size of the assembled genomics contigs of *M. galloprovincialis*, we  
13  
14 could identify eleven MgCRP-I gene regions corresponding to a full-length coding sequence, from  
15  
16 the initial ATG to the STOP codon (MgCRP-I 13, 14, 18, 24, 28, 40, 45, multi-MgCRP-I 2, 11, 12  
17  
18 and 13) and several partial matches, either to the 3' or to the 5' region of the CDS (**Table 2**).  
19

20  
21 The structure of MgCRP-I genes is conserved, with four exons and three introns (**Figure 2**). The  
22  
23 first exon, which could only be annotated in five sequences thanks to the alignment with RNA-seq  
24  
25 data, partially includes the 5'UTR region. In most cases, the second exon encodes the first 17 amino  
26  
27 acids of the precursor protein, thus comprising most of the signal peptide. A phase-1 intron  
28  
29 separates the second and the third exon. The third exon is ~100 nt long and covers the signal  
30  
31 peptide cleavage site and most of the propeptide region. The open reading frame is interrupted by a  
32  
33 phase-2 intron, which separates the third and the fourth exon. The last exon invariably comprises  
34  
35 the final 10 nucleotides of the pro- region, including the highly conserved dibasic cleavage site, the  
36  
37 complete cysteine-rich C-terminal region and the entire 3'UTR region. No additional exons are  
38  
39 present in multi-MgCRP-I genes, whose multiple cysteine-rich arrays are encoded by the fourth  
40  
41 exon. A schematic representation of the structural organization of MgCRP-I genes and encoded  
42  
43 precursor proteins is shown in **Figure 2**.  
44  
45  
46  
47  
48

### 49 50 **Gene duplication and positive selection are driving the evolution of the the MgCRP-I gene** 51 52 **family**

53  
54 The combined genomic/transcriptomic analyses indicate the presence of at least 67 different  
55  
56 potentially functional CRP-I loci in the genome of the Mediterranean mussel. Owing to the  
57  
58 preliminary nature of the released mussel genome (Nguyen, et al. 2014), this has to be considered as  
59  
60

1  
2  
3 a conservative estimate. The phylogenetic analysis of the CRP-I signal peptide regions (**Figure 3**),  
4  
5 evidenced the existence of several highly similar paralogous genes, highlighting the important role  
6  
7 of gene duplication events in the evolution of the MgCRP-I gene family, which in some cases  
8  
9 appear to have occurred very recently (**supplementary figure S4, supplementary material**  
10  
11 **online**). A number of MgCRP-I pseudogenes with frame-shift or missense mutations were  
12  
13 identified in the mussel genome (data not shown) and, at the same time, the sequence truncation  
14  
15 caused by the small size of the genomic contigs makes impossible to infer how many of the  
16  
17 incomplete MgCRP-I loci are fully functional (see **Table 2**). For the same reason, the presence of  
18  
19 common regulatory regions and transposable elements which could have driven the expansion of  
20  
21 this gene family is matter of future investigations.  
22  
23

24  
25 Gene duplication is not sufficient by itself to explain some peculiar features of MgCRP-I  
26  
27 genes: indeed, the amino acid diversity of the peptide precursors is strikingly higher within the  
28  
29 mature peptide region compared to the signal peptide and pro-region which are, in turn, highly  
30  
31 conserved (**Figure 1**). This observation suggests an increased rate of mutations within the 4<sup>th</sup> exon,  
32  
33 and an accelerated evolutionary rate. The likelihood ratio tests we performed to assess this  
34  
35 hypothesis strongly support positive selection of the MgCRP-I genes (p-value =  $1.717 \times 10^{-8}$ ). In  
36  
37 fact, we could identify nine positively selected sites (PP > 0.95), all located within the mature  
38  
39 peptide region, after the cleavage site of the pro-peptide (**Figure 1**). Five out of the six invariable  
40  
41 cysteines engaged in disulfide bridges, buried within this hypervariable and positively selected  
42  
43 region, undergo site-specific codon preservation (**Figure 4**). This peculiar phenomenon, which was  
44  
45 also observed for the arginine responsible of the pro-peptide cleavage, is likely driven by the unique  
46  
47 properties of these residues for the maintenance of the tridimensional structure and an efficient  
48  
49 biosynthesis and folding of the mature peptide, as suggested for many other protein families,  
50  
51 including conotoxins (Conticello, et al. 2001; Steiner, et al. 2013).  
52  
53  
54

55  
56 Although structurally important codons appear to be somehow protected from variation, the  
57  
58 high selective pressure acting on the fourth exon (encoding the entire mature peptide) of MgCRP-I  
59  
60

1  
2  
3 genes in some cases introduced loss-of-function mutations, and resulted in the disruption of  
4  
5 disulfide bonds (**supplementary figure S5, supplementary material online**). MgCRP-I 12 and  
6  
7 MgCRP-I 23 represent two instructive examples, since the deduced proteins maintain the highly  
8  
9 conserved signal peptide and pro-peptide regions, features clearly identifying them as CRP-I related  
10  
11 sequences in a phylogenetic analysis though lacking the canonical cysteine array. Indeed, MgCRP-I  
12  
13 12 lacks four out of the six conserved cysteine residues and just the two adjacent residues Cys3 and  
14  
15 Cys4 are retained; on the other hand, MgCRP-I 23 is an even more extreme case, as completely  
16  
17 devoid of cysteines.  
18  
19

### 20 21 22 **CRP-I sequences are only found in the order Mytiloida**

23  
24  
25 Following BLAST searches, the MgCRP-I peptides did not show significant sequence  
26  
27 similarity with any other sequence deposited in public databases and the prediction of their  
28  
29 tridimensional structure was considered unreliable by Phyre 2, due to the absence of models sharing  
30  
31 sufficient sequence affinity within the PDB database. For this reason, we investigated the presence  
32  
33 of MgCRP-I-like sequences in genomic and transcriptomic datasets which are increasingly  
34  
35 available also for bivalve molluscs (Suárez-Ulloa, et al. 2013b).  
36  
37

38  
39 Looking for short secreted CRP-I-like peptides in the transcriptomes of 70 different species  
40  
41 (see Materials and methods **and supplementary table S1, supplementary material online**), we  
42  
43 could find the CRP-I signature only in the mussels *M. edulis*, *M. trossulus*, and *M. californianus*.  
44  
45 The first two species and *M. galloprovincialis* are widespread and genetically close to each other, as  
46  
47 evidenced by the presence of natural hybrid populations (Beaumont, et al. 2004) whereas *M.*  
48  
49 *californianus* is distributed in the Pacific coast of North America and is more distantly related to the  
50  
51 other mussel ecotypes (Hilbish, et al. 2000). The full length CRP-I sequences identified, here named  
52  
53 McCRP-I, MeCRP-I or MtCRP-I on the basis of the species name, are reported in **supplementary**  
54  
55 **material online**.  
56  
57  
58  
59  
60

1  
2  
3 No CRP-I-like sequence was detected in any of the other bivalve transcriptomes and we  
4 hypothesize their complete absence or no detectable expression in the analyzed bivalve species. Due  
5 to the high depth of NGS technologies, the latter hypothesis is unlikely and the genomic analysis of  
6 the oysters *C. gigas* and *P. fucata* strongly supports the absence of CRP-I-like genes, thus ruling out  
7 the possibility of a missed detection of poorly expressed transcripts. In addition, no evidence of  
8 CRP-I-like peptides was found in *Limnoperna fortunei* and *Perna viridis*, the only two species  
9 among the over 50 different genera of the order Mytiloida, beside *Mytilus* spp., which have been  
10 subjected to Illumina RNA-sequencing so far.

11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21 Based on the available data, CRP-Is are certainly present in *Mytilus* spp. and appear to be  
22 absent in other mytiloids. Since the C-C-CC-C-C array is not common in bivalves and nothing  
23 similar was found in the fully sequenced genomes of *C. gigas* (order Ostreoida) and *P. fucata* (order  
24 Pterioidea), the CRP-I-like sequences, appear to have a narrow taxonomical distribution, comparable  
25 to that of other mussel cysteine-rich peptides (i.e. mytilins, myticins and mytimycins which cannot  
26 be found outside Mytiloida).

### 27 28 29 30 31 32 33 34 35 36 **Occurrence of the C-C-CC-C-C array in protein databases and relationship with CRP-I** 37 **peptides**

38  
39  
40 Large-scale bioinformatic analyses revealed the presence of a CRP-I-like cysteine pattern  
41 mostly in animals and, among them, almost exclusively in invertebrates (see **Table 3**). More in  
42 detail, within the UniProtKB/Swiss-Prot database we found 452 peptides (279 if considering non-  
43 redundant peptides based on a 95% sequence identity criterion) mostly belonging to invertebrate  
44 animals: cone snails, turrids and terebrids (grouped as Conoidea in **Figure 5**), spiders, scorpions,  
45 pancrustaceans (with a single horseshoe crab and 35 insect sequences) and nematodes. Almost the  
46 totality of these peptides display a neurotoxic activity due to their high affinity to ion channels, like  
47 in the case of conotoxins, turrtoxins and augertoxins (Aguilar, et al. 2009; Imperial, et al. 2003;  
48 Terlau and Olivera 2004) and peptides produced in the venom gland of spiders (Zhang, et al. 2010)

1  
2  
3 and scorpions (Ma, et al. 2009). With the exception of a wasp toxin, all the entries from insects  
4  
5 were related to bombyxin, a prothoracicotropic hormone involved in morphogenesis (Nijhout and  
6  
7 Grunert 2002).  
8

9  
10 No CRP-I-like cysteine pattern was found in chordates, with the exception of 6 peptides (5  
11  
12 intestinal trefoil factors, with unclear function, and veswaprin, a snake antimicrobial peptide). A  
13  
14 limited number of such peptides were found in Fungi, with the positive hits corresponding to  
15  
16 uncharacterized proteins, and in green plants, with all cases representing antimicrobial peptides.  
17  
18 The CRP-I-like array was also found in Baculoviruses (in a viral family of conotoxin-like peptides)  
19  
20 (Eldridge, et al. 1992). The full list of these peptides with their taxonomic origin and reported  
21  
22 function are shown in **supplementary table 2, supplementary material online**.  
23

24  
25 Given the great sequence diversity, likely dependent on a fast evolutionary rate of molecular  
26  
27 substitutions, the relationships among the MgCRP-I peptides (**Figure 3**) and with other CRPs  
28  
29 (**Figure 5**) remain unresolved. Nevertheless, several features could be underlined by the  
30  
31 phylogenetic analyses. All MgCRP-I sequences and the orthologous sequences from *M. edulis*, *M.*  
32  
33 *trossulus* and *M. californianus* clustered together in a single clade. Most of the other CRPs  
34  
35 sequences clustered in distinct groups which included peptides from the same taxon (Pancrustacea,  
36  
37 Conoidea, Spiders, Scorpions, Nematodes, Chordates, Plants, Fungi and Viruses), even if: (i)  
38  
39 peptides from a single taxon were found in different clusters (e.g. conopeptides typically cluster in  
40  
41 different groups that correspond to different superfamilies (Kaas, et al. 2010; Puillandre, et al.  
42  
43 2012)); (ii) some peptides, characterized by a long branch in the tree, clustered in a group mostly  
44  
45 composed by peptides from another taxon (as an example, two conopeptides clustered within the  
46  
47 group of mussel CRP-I peptides, with long branches, suggesting that they may belong to completely  
48  
49 different structural classes).  
50  
51  
52

53  
54 Overall, no distinct traits can unequivocally link the evolutionary history of CRP-I peptides  
55  
56 with those of other protein families characterized by the same cysteine array. As CRP-I-like  
57  
58 peptides are absent in bivalves other than mussels, the only other molluscan group where the C-C-  
59  
60

1  
2  
3 CC-C-C array is present is represented by Conoidea (Gastropoda). Nevertheless the absence of this  
4  
5 molecular motif in the fully sequenced genomes of other gastropods (*Aplysia californica*, *Lottia*  
6  
7 *gigantea* and *Biomphalaria glabrata*) suggests that it might have been acquired independently in  
8  
9 these two molluscan groups. Even though the large sequence divergence prevents conclusions, the  
10  
11 study of the disulfide bonds topology in the MgCRP-I peptides has provided further support to this  
12  
13 hypothesis, as reported in the next section.  
14  
15

### 16 17 18 **Oxidative refolding and disulfide bond topology of synthetic MgCRP-I peptides** 19

20  
21 Optimization of the oxidative folding yield for peptides with disulfide bridges is still an  
22  
23 empiric exercise. Various parameters affect folding yields such as temperature, additives, redox  
24  
25 couples, peptide concentration and duration of the folding reaction (Bulaj 2005; Bulaj and Olivera  
26  
27 2008). We synthesized the MgCRP-I 7 and MgCRP 9 peptides and their purified fractions were  
28  
29 subjected to oxidative folding reactions in the presence of redox reagents GSH and GSSG (10:1) at  
30  
31 4°C buffered at pH 8.8 (RefoldA, see Materials and methods). This protocol, with slight  
32  
33 modifications, has been commonly used in the refolding of disulfide rich proteins and, for example,  
34  
35 it yielded good amounts (~90%) of the cystine knot peptide Huwentoxin-IV in its native structure  
36  
37 (Deng, et al. 2013) and also in the synthesis of  $\omega$ -conotoxin MVIIC (~50%). We tested also a high  
38  
39 salt refolding mixture (RefoldB) which had shown improved yields of the same conotoxin (Kubo, et  
40  
41 al. 1996).  
42  
43

44  
45 In our hands each peptide refolding mixture (RefoldA) produced one major component, with  
46  
47 yields of 85% and 79% for MgCRP-I 7 and MgCRP-I 9, respectively, as determined by RP-HPLC  
48  
49 peak area integration (area of the purified reduced peptide before folding = 100%) (see  
50  
51 **supplementary figure S6 and S7, supplementary material online**). Refold B gave similar results  
52  
53 but with slightly lower yields (data not shown). According to ESI-MS, the molecular mass of folded  
54  
55 MgCRP-I 7 and folded MgCRP-I 9 were in good agreement with those of the fully oxidized  
56  
57 products (see supplementary Figure S3 and S4, Supplementary Material online): 3,160.0 (calc.  
58  
59  
60

1  
2  
3 mono. 3,160.2) and 3,090.1 (calc. mono. 3,090.2), respectively. The evidence of one dominant  
4  
5 product in the refolding of both peptides is consistent with the assumption of native conformation  
6  
7 but, on the other hand, no comparison is currently possible between the synthetic products and the  
8  
9 native counterparts eventually present in *Mytilus*.

10  
11 We performed a disulfide connectivity prediction using the DiANNA (DiAminoacid Neural  
12  
13 Network Application) Web Server of the Boston college (Ferrè and Clote 2005); the algorithm  
14  
15 predicted a 1-2, 3-4, 5-6 topology for MgCRP-I 9 and a 1-4, 2-6, 3-5 topology for MgCRP-I 7 but  
16  
17 with a very high score (0.76 on a maximum of 1) in favor of a Cys14-Cys15 disulfide. At this point  
18  
19 we experimentally determined the disulfide bond geometry using enzymatic fragmentation (trypsin  
20  
21 and chymotrypsin) and LC/MS/MS. The analysis of both peptides revealed that the cysteine  
22  
23 connectivity follows the nearest-neighbor pattern (1-2, 3-4, 5-6), namely Cys3-Cys8, Cys14-Cys15  
24  
25 and Cys20-Cys25 (see **supplementary tables S3 and S4, supplementary material online**). In a  
26  
27 published detailed disulfide classification based on SwissProt and Pfam databases the topology 1-2,  
28  
29 3-4, 5-6 is largely represented and contains a very heterogeneous ensemble of protein families  
30  
31 (Gupta, et al. 2004); notably, the vicinal disulfide bond present in our MgCRP-I peptides and  
32  
33 formed between the side chains of adjacent cysteines (Cys14-Cys15) represents a rare structural  
34  
35 element. The vicinal disulfide, due to its intrinsic constrained nature, is usually described to be  
36  
37 accompanied by the formation of a tight turn of the protein backbone (Carugo, et al. 2003);  
38  
39 additionally, the oxidized and reduced states of this bond present very different structural features  
40  
41 suggesting a possible role as conformational switch (Carugo, et al. 2003). At the present time, we  
42  
43 do not know the significance that this vicinal bond could have on the activity of the MgCRP-I  
44  
45 peptides but the observed 1-2, 3-4, 5-6 topology is distinctively different from the 1-4, 2-5, 3-6  
46  
47 topology common to knottins, which comprise conopeptides and most of the peptides represented in  
48  
49 **Figure 5** with an experimentally determined tridimensional structure (Hartig, et al. 2005).

## 58 MgCRP-I transcript levels

1  
2  
3 The number of sequences related to the MgCRP-I family identified in the many mussel  
4 transcriptome datasets analyzed was extremely low (**Table 4**) and suggests a very limited basal  
5 expression of these genes in different tissues under physiological conditions. More in detail, no  
6 evidence of MgCRP-Is was found in Sanger sequencing-based EST collections, with the exception  
7 of a single *M. edulis* sequence detected in a SSH library (digestive gland of mussels exposed to  
8 styrene). The number of CRP-I sequences detected in the pyrosequencing-based datasets increased,  
9 even though in many cases MgCRP-I transcripts could not be detected. Finally, the analysis of  
10 Illumina sequencing-based transcriptomes clearly pointed out the high sequencing depth necessary  
11 to detect MgCRP-I messenger RNAs, which can be estimated to cumulatively contribute to less  
12 than 0.01% (but often to even less than 0.001%) of the total gene expression in most tissues.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

24  
25 To better evaluate the MgCRP-I tissue-specificity, we analyzed the expression levels of 17  
26 MgCRP-I transcripts by quantitative PCR (qPCR) in different tissues (hemolymph, digestive gland,  
27 inner mantle, mantle rim, foot, posterior adductor muscle and gills) of a pool of 30 naïve adult  
28 mussels (*M. galloprovincialis*). We found large variability among the expression profiles of  
29 individual CRP-I sequences, with the overall expression levels being almost invariably very low in  
30 all tissues. However, three tissues emerged as main sites of MgCRP-I expression, namely the  
31 digestive gland, the inner mantle and the mantle rim. Several MgCRP-I displayed, at least to some  
32 extent, a certain degree of tissue specificity (**Figure 6**). In most cases these genes were not  
33 expressed at all in hemolymph, foot, gills and posterior adductor muscle, indicating that these are  
34 not the primary sites of production of MgCRP-I peptides, which is consistent with RNA-seq data  
35 (**Table 4**).  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

49 Overall, the gene expression data leave room to different hypotheses which need to be tested  
50 in future experiments: (i) the expression of these peptides may be induced by still unknown specific  
51 stimuli; (ii) MgCRP-I are expressed by a low number of highly specialized cells and therefore the  
52 global contribution to mRNAs extracted from macro-tissues is low; (iii) MgCRP-I are not expressed  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 in adult individuals, but they play an important role in the early developmental stages (but this  
4 hypothesis seems to be disproved by the analysis of *M. edulis* larvae RNA-seq data, see **Table 4**).  
5  
6  
7  
8

9  
10 **MgCRP-I synthetic peptides do not show any significant cytotoxic, insecticidal, antifungal and**  
11 **antimicrobial activity**  
12

13  
14 In an attempt to characterize the biological activity of the synthetic peptides MgCRP-I 7 and  
15 MgCRP-I 9, we evaluated their cytotoxicity on human tumor cell lines and insect larvae, and their  
16 antimicrobial activity on the bacteria *E. coli* and *S. aureus*, and on fungal strains of *C. albicans* and  
17 *C. neoformans* (see Materials and methods). The MTT assays indicated that both synthetic peptides  
18 were not cytotoxic on the HT-29, SHSY5Y and MDAMB231 cell lines up to 10  $\mu$ M concentration.  
19 Similarly, no insecticidal effect was observed in *Z. morio* larvae 48 hours after the injection of 300  
20  $\mu$ g peptide/Kg body weight, a quantity much higher than those determining visible neurotoxic  
21 effects, or even death, for other invertebrate toxins (Yang, et al. 2012; Zhong, et al. 2014). Finally,  
22 the antimicrobial activity assay evidenced that both MgCRP-I 7 and 9 had a MIC > 32  $\mu$ M on the  
23 selected bacterial and fungal strains.  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35

36 Although these results indicate that MgCRP-I synthetic peptides did not display  
37 antimicrobial or cytotoxic activity in the tested conditions, their involvement in defense processes  
38 cannot be ruled out. In fact, post-translational modifications might occur in mussel cells but this can  
39 hardly be investigated due to the low expression levels of MgCRP-I genes which, in turn, makes the  
40 purification of native peptides difficult. Hence, the absence of biological effects could depend on a  
41 variety of modifications not present in the synthetic MgCRP-Is but often reported as fundamental  
42 for the antimicrobial or toxic activity of short cysteine-rich peptides (Bergeron, et al. 2013; Buczek,  
43 et al. 2005a; Buczek, et al. 2005b; Guder, et al. 2000).  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53

54 Other factors could also play a role in the evaluation of MgCRP-Is, such as the possible peptide  
55 interactions with the culture medium, and in particular with the serum components. Although these  
56  
57  
58  
59  
60

1  
2  
3 considerations are important in perspective, a detailed characterization of the MgCRP-Is is beyond  
4  
5 the scope of this paper.  
6  
7

## 8 9 **Conclusions**

10  
11 Thanks to an exploratory bioinformatics approach applied to the NGS sequencing data, we  
12  
13 could identify a novel family of cysteine rich peptides, named MgCRP-I, which appears to be  
14  
15 exclusively present in Mytiloidea, an order of marine filter-feeding mussels. The MgCRP-I gene  
16  
17 family and the encoded peptides share a number of structural and evolutionary traits in common  
18  
19 with other families of CRPs, which almost invariably have an antimicrobial or toxic function. These  
20  
21 marked similarities initially suggested that MgCRP-I peptides could have similar biological  
22  
23 functions, thus making them intriguing targets for possible future biotechnological and  
24  
25 pharmacological applications. However, all the tests performed on two synthetic MgCRP-I peptides  
26  
27 led to inconclusive results, leaving their biological role still puzzling. In addition, the biological  
28  
29 targets (both at the molecular and at the species level) of MgCRP-I peptides are still unknown and  
30  
31 the events triggering the expression of these molecules are still elusive. In absence of further  
32  
33 indications, these questions remain unsolved.  
34  
35  
36  
37

38  
39 Overall, we have provided a preliminary overview on MgCRP-I peptides, which is intended  
40  
41 as starting point for further investigations on their possible action on prokaryotic or eukaryotic cells.  
42  
43 Our work also highlights the possibility of identifying previously uncharacterized, potentially  
44  
45 bioactive, peptides from whole genomes and transcriptomes of non-model organisms without any  
46  
47 previous knowledge about their primary sequence, an experimental approach which could speed up  
48  
49 the discovery or the design of novel molecules with potential biotechnological applications. Due to  
50  
51 their still limited genomic knowledge, marine invertebrates in particular represent a virtually  
52  
53 unlimited and almost unexplored source of novel bioactive compounds.  
54  
55  
56  
57  
58  
59  
60

## Acknowledgements

This work was supported by BIVALIFE (FP7-KBBE-2010-4) and PRIN 2010-11 (20109XZEPR).

## Figure legends

**Figure 1:** Sequence variability of MgCRP-I sequences; variability index  $W$  is plotted in the upper panel, while the sequence consensus, obtained with Weblogo (<http://weblogo.berkeley.edu>) is shown in the lower panel. Only positions covered by at least 50% sequences in the global alignment of MgCRP-I peptides are shown. Sites under positive selection are indicated by an asterisk.

**Figure 2.** Exon/intron structure of the complete coding regions of the MgCRP-I 13, 14, 28, 45 and multi-MgCRP-I 2 genes (panel A) and corresponding organization of the encoded peptide precursors (panel B). The positions of the signal peptide, pro-region and mature peptide regions are highlighted and each cysteine-rich module is marked by a box.

**Figure 3:** Maximum likelihood tree obtained with the MgCRP-I peptides based on the alignment of the signal peptide only. Only Booststraps values  $> 75$  are shown. Some sequences were not considered in this analysis as their N-terminal region was incomplete (see **Table 2**). Arrows indicate two MgCRP-I peptides with a disrupted cysteine array (12 and 23) marking an unconventional mature region.

**Figure 4:** codon usage for the Arg residue responsible of the pro-peptide cleavage site and for the 6 cysteine residues involved in the disulfide bridges, calculated on *M. galloprovincialis* MgCRP-I peptides. The probabilities of finding the observed codon biases were calculated assuming a binomial distribution and the codon usage (a priori probabilities) inferred from the transcriptome published by Gerdol et al. 2014 (75.1%-24.9% for TGT-TGC, encoding Cys, 49.3%-16.3%-13.3%-

1  
2  
3 12.5%-4.7%-3.8% for AGA-AGG-CGA-CGT-CGG-CGC, encoding Arg). Significant ( $p < 0.05$ ) and  
4  
5 highly significant ( $p < 0.01$ ) deviations from the expected distributions ( $p < 0.01$ ) are marked by \* and  
6  
7 \*\* respectively. NS = not significant.  
8  
9

10  
11 **Figure 5:** Maximum likelihood tree obtained with the MgCRP-I, the orthologous sequences from  
12  
13 other mussel species and all the CRPs mined from UniProtKB/Swiss-Prot (see Materials and  
14  
15 methods). Peptides from pancrustaceans, conoideans, spiders, scorpions, nematodes, chordates,  
16  
17 plants, fungi, viruses and mussels are shown. Mussel CRP-I peptides are highlighted in a grey  
18  
19 background.  
20  
21

22  
23  
24 **Figure 6:** gene expression of 17 selected MgCRP-I genes in six tissues (H: hemolymph; DG:  
25  
26 digestive gland; IM: inner mantle; MR: mantle rim; F: foot; G: gills; AM: posterior adductor  
27  
28 muscle); primers for MgCRP-I 3 also target MgCRP-I 25, primers for MgCRP-I 10 also target  
29  
30 MgCRP-I 26. Bars represent the expression relative to EF-1 alpha; results are mean  $\pm$  standard  
31  
32 deviation of three replicates. MgCRP-I sequences are divided into three panels based on their  
33  
34 expression level: panel A – genes with maximum relative expression value comprised between 0.05  
35  
36 and 0.25; panel B – genes with maximum relative expression value comprised between 0.004 and  
37  
38 0.01; panel C - with maximum relative expression value lower than 0.003. Panel D shows a  
39  
40 schematic representation of a *M. galloprovincialis* anatomical features, highlighting the sampled  
41  
42 tissues.  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

- Adams DJ, Alewood PF, Craik DJ, Drinkwater RD, Lewis RJ 1999. Conotoxins and their potential pharmaceutical applications. *Drug Develop Res.* 46: 219-234.
- Aguilar MB, de la Rosa RAC, Falcón A, Olivera BM, Heimer de la Cotera EP 2009. Peptide pal9a from the venom of the turrid snail *Polystira albida* from the Gulf of Mexico: Purification, characterization, and comparison with P-conotoxin-like (framework IX) conoidean peptides. *Peptides.* 30: 467-476.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ 1990. Basic local alignment search tool. *J Mol Biol.* 215: 403-410.
- Bartlett TC, et al. 2002. Crustins, homologues of an 11.5-kDa antibacterial peptide, from two species of penaeid shrimp, *Litopenaeus vannamei* and *Litopenaeus setiferus*. *Mar Biotechnol.* 4: 278-293.
- Bassim S, Genard B, Gauthier-Clerc S, Moraga D, Tremblay R 2014. Ontogeny of bivalve immunity: assessing the potential of next-generation sequencing techniques. *Reviews in Aquaculture.* 6: 1-21. doi: 10.1111/raq.12064
- Beaumont AR, Turner G, Wood AR, Skibinski DOF 2004. Hybridisations between *Mytilus edulis* and *Mytilus galloprovincialis* and performance of pure species and hybrid veliger larvae at different temperatures. *J Exp Mar Biol Ecol.* 302: 177-188.
- Benincasa M, et al. 2010. Antifungal activity of Amphotericin B conjugated to carbon nanotubes. *ACS Nano.* 5: 199-208. doi: 10.1021/nn1023522
- Benincasa M, et al. 2004. Antimicrobial activity of Bac7 fragments against drug-resistant clinical isolates. *Peptides.* 25: 2055-2061.
- Bergeron ZL, et al. 2013. A 'conovenomic' analysis of the milked venom from the mollusk-hunting cone snail *Conus textile*—The pharmacological importance of post-translational modifications. *Peptides.* 49: 145-158. doi: <http://dx.doi.org/10.1016/j.peptides.2013.09.004>
- Buczek O, Bulaj G, Olivera BM 2005a. Conotoxins and the posttranslational modification of secreted gene products. *Cell Mol Life Sci.* 62: 3067-3079.
- Buczek O, Yoshikami D, Bulaj G, Jimenez EC, Olivera BM 2005b. Post-translational amino acid isomerization: a functionally important d-amino acid in an excitatory peptide. *J Biol Chem.* 280: 4247-4253. doi: 10.1074/jbc.M405835200
- Bulaj G 2005. Formation of disulfide bonds in proteins and peptides. *Biotechnology Advances* 23: 87-92.
- Bulaj G, Olivera BM 2008. Folding of conotoxins: Formation of the native disulfide bridges during chemical synthesis and biosynthesis of *Conus* peptides. *Antioxid Redox Sign.* 10: 141-155.
- Bulet P, Stöcklin R 2005. Insect antimicrobial peptides: Structures, properties and gene regulation. *Protein Peptide Lett.* 12: 3-11.
- Carugo O, et al. 2003. Vicinal disulfide turns. *Protein Eng.* 16: 637-639.
- Charlet M, et al. 1996. Innate immunity: Isolation of several cysteine-rich antimicrobial peptides from the blood of a mollusc, *Mytilus edulis*. *J Biol Chem.* 271: 21808-21813.
- Conticello SG, et al. 2001. Mechanisms for evolving hypervariability: The case of conopeptides. *Mol Biol Evol.* 18: 120-131.

- 1  
2  
3 Craft JA, et al. 2010. Pyrosequencing of *Mytilus galloprovincialis* cDNAs: Tissue-specific expression  
4 patterns. PLoS ONE. 5.  
5  
6 Craig AG, Bandyopadhyay P, Olivera BM 1999. Post-translationally modified neuropeptides from *Conus*  
7 venoms. Eur J Biochem. 264: 271-275. doi: 10.1046/j.1432-1327.1999.00624.x  
8  
9 Deng M, et al. 2013. Synthesis and biological characterization of synthetic analogs of Huwentoxin-IV (Mu-  
10 theraphotoxin-Hh2a), a neuronal tetrodotoxin-sensitive sodium channel inhibitor. Toxicon. 71: 57-65.  
11  
12 Destoumieux D, et al. 1997. Penaeidins, a new family of antimicrobial peptides isolated from the shrimp  
13 *Penaeus vannamei* (Decapoda). J Biol Chem. 272: 28398-28406.  
14  
15 Dinkel H, et al. 2011. ELM—the database of eukaryotic linear motifs. Nucleic Acids Res. 40: D242-51. doi:  
16 10.1093/nar/gkr1064  
17  
18 Duckert P, Brunak S, Blom N 2004. Prediction of proprotein convertase cleavage sites. Protein Eng Des Sel.  
19 17: 107-112. doi: 10.1093/protein/gzh013  
20  
21 Eddy SR 2011. Accelerated profile HMM searches. PLoS Comput Biol. 7: e1002195. doi:  
22 10.1371/journal.pcbi.1002195  
23  
24 Edgar RC 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic  
25 Acids Res. 32: 1792-1797.  
26  
27 Ehret-Sabatier L, et al. 1996. Characterization of novel cysteine-rich antimicrobial peptides from scorpion  
28 blood. J Biol Chem. 271: 29537-29544  
29  
30 Eldridge R, Li Y, Miller LK 1992. Characterization of a baculovirus gene encoding a small conotoxinlike  
31 polypeptide. J Virol. 66: 6563-6571.  
32  
33 Emanuelsson O, Brunak S, von Heijne G, Nielsen H 2007. Locating proteins in the cell using TargetP,  
34 SignalP and related tools. Nat Protoc. 2: 953-971.  
35  
36 Fan C-X, et al. 2003. A novel conotoxin from *Conus betulinus*,  $\kappa$ -BtX, unique in cysteine pattern and in  
37 function as a specific BK channel modulator. J Biol Chem. 278: 12624-12633. doi:  
38 10.1074/jbc.M210200200  
39  
40 Ferrè F, Clote P 2005. DiANNA: A web server for disulfide connectivity prediction. Nucleic Acids Res.33:  
41 W230-W232.  
42  
43 Fogaça AC, et al. 2004. Cysteine-rich antimicrobial peptides of the cattle tick *Boophilus microplus*:  
44 Isolation, structural characterization and tissue expression profile. Dev Comp Immunol. 28: 191-200.  
45  
46 Freer A, Bridgett S, Jiang J, Cusack M 2014. Biomineral proteins from *Mytilus edulis* mantle tissue  
47 transcriptome. Mar Biotechnol. 16: 34-45. doi: 10.1007/s10126-013-9516-1  
48  
49 Froy O, Gurevitz M 2004. Arthropod defensins illuminate the divergence of scorpion neurotoxins. J Pept Sci.  
50 10: 714-718.  
51  
52 Gerdol M, et al. 2014. RNA sequencing and de novo assembly of the digestive gland transcriptome in  
53 *Mytilus galloprovincialis* fed with toxinogenic and non-toxic strains of *Alexandrium minutum*. BMC  
54 Research Notes 7: 722.  
55  
56 Gerdol M, De Moro G, Manfrin C, Venier P, Pallavicini A 2012. Big defensins and mytimacins, new AMP  
57 families of the Mediterranean mussel *Mytilus galloprovincialis*. Dev Comp Immunol. 36: 390-399.  
58  
59  
60

1  
2  
3 Gerdol M, et al. 2011. The C1q domain containing proteins of the Mediterranean mussel *Mytilus*  
4 *galloprovincialis*: A widespread and diverse family of immune-related molecules. Dev Comp Immunol. 35:  
5 635-643.

6  
7 Gerdol M, Venier P An updated molecular basis for mussel immunity. Fish Shellfish Immun (in press). doi:  
8 <http://dx.doi.org/10.1016/j.fsi.2015.02.013>

9  
10 González VL, et al. 2015. A phylogenetic backbone for Bivalvia: an RNA-seq approach. Procl Bio Sci (in  
11 press). doi: doi: 10.1098/rspb.2014.2332

12  
13 Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference  
14 genome. Nat Biotechnol 29: 644-652.

15  
16 Gruber CW, Čemažar M, Anderson MA, Craik DJ 2007. Insecticidal plant cyclotides and related cystine  
17 knot toxins. Toxicon. 49: 561-575.

18  
19 Guder A, Wiedemann I, Sahl HG 2000. Posttranslationally modified bacteriocins - The lantibiotics.  
20 Biopolymers - Peptide Science Section. 55: 62-73.

21  
22 Gupta A, Van Vlijmen HWT, Singh J 2004. A classification of disulfide patterns and its relationship to  
23 protein structure and function. Protein Sci. 13: 2045-2058.

24  
25 Hartig GRS, Tran TT, Smythe ML 2005. Intramolecular disulphide bond arrangements in nonhomologous  
26 proteins. Protein Sci. 14: 474-482. doi: 10.1110/ps.04923305

27  
28 Herbst R, Marciano-Cabral F, Leippe M 2004. Antimicrobial and pore-forming peptides of free-living and  
29 potentially highly pathogenic *Naegleria fowleri* are released from the same precursor molecule. J Biol Chem.  
30 279: 25955-25958. doi: 10.1074/jbc.M401965200

31  
32 Hilbish TJ, et al. 2000. Origin of the antitropical distribution pattern in marine mussels (*Mytilus* spp.): routes  
33 and timing of transequatorial migration. Mar Biol. 136: 69-77. doi: 10.1007/s002270050010

34  
35 Imperial JS, et al. 2003. The augertoxins: Biochemical characterization of venom components from the  
36 toxoglossate gastropod *Terebra subulata*. Toxicon. 42: 391-398.

37  
38 Kaas Q, Westermann JC, Craik DJ 2010. Conopeptide characterization and classifications: An analysis using  
39 ConoServer. Toxicon. 55: 1491-1509.

40  
41 Käll L, Krogh A, Sonnhammer ELL 2004. A combined transmembrane topology and signal peptide  
42 prediction method. J Mol Biol. 338: 1027-1036. doi: <http://dx.doi.org/10.1016/j.jmb.2004.03.016>

43  
44 Kelley LA, Sternberg MJ 2009. Protein structure prediction on the Web: a case study using the Phyre server.  
45 Nat Protoc. 4: 363-371.

46  
47 Kubo S, Chino N, Kimura T, Sakakibara S 1996. Oxidative folding of  $\omega$ -conotoxin MVIIC: Effects of  
48 temperature and salt. Biopolymers. 38: 733-744.

49  
50 Li H, Parisi MG, Parrinello N, Cammarata M, Roch P 2011. Molluscan antimicrobial peptides, a review  
51 from activity-based evidences to computer-assisted sequences. Invertebrate Survival Journal. 8: 85-97.

52  
53 Liao Z, et al. 2013. Molecular characterization of a novel antimicrobial peptide from *Mytilus coruscus*. Fish  
54 Shellfish Immun. 34: 610-616.

55  
56 Ma Y, et al. 2009. Transcriptome analysis of the venom gland of the scorpion *Scorpiops jendeki*: Implication  
57 for the evolution of the scorpion venom arsenal. BMC Genomics. 10: 290.

- 1  
2  
3 Marshall E, Costa LM, Gutierrez-Marcos J 2011. Cysteine-Rich Peptides (CRPs) mediate diverse aspects of  
4 cell-cell communication in plant reproduction and development. *J ExpBot.* 62: 1677-1686.  
5  
6 Mayer AMS, Rodríguez AD, Berlinck RGS, Fusetani N 2011. Marine pharmacology in 2007-8: Marine  
7 compounds with antibacterial, anticoagulant, antifungal, anti-inflammatory, antimalarial, antiprotozoal,  
8 antituberculosis, and antiviral activities; Affecting the immune and nervous system, and other miscellaneous  
9 mechanisms of action. *Comp Biochem Phys C.* 153: 191-222.  
10  
11 Mitta G, Hubert F, Noël T, Roch P 1999. Myticin, a novel cysteine-rich antimicrobial peptide isolated from  
12 haemocytes and plasma of the mussel *Mytilus galloprovincialis*. *Eur J Biochem.* 265: 71-78.  
13  
14 Mitta G, Vandenbulcke F, Hubert F, Salzet M, Roch P 2000a. Involvement of mytilins in mussel  
15 antimicrobial defense. *J Biol Chem.* 275: 12954-12962.  
16  
17 Mitta G, et al. 2000b. Differential distribution and defence involvement of antimicrobial peptides in mussel.  
18 *J Cell Sci.* 113: 2759-2769.  
19  
20 Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B 2011. How Many Species Are There on Earth and in  
21 the Ocean? *PLoS Biol.* 9: e1001127. doi: 10.1371/journal.pbio.1001127  
22  
23 Nguyen TTT, Hayes BJ, Ingram BA 2014. Genetic parameters and response to selection in blue mussel  
24 (*Mytilus galloprovincialis*) using a SNP-based pedigree. *Aquaculture.* 420-421: 295-301. doi:  
25 <http://dx.doi.org/10.1016/j.aquaculture.2013.11.021>  
26  
27 Nijhout HF, Grunert LW 2002. Bombyxin is a growth factor for wing imaginal disks in lepidoptera. *P Natl*  
28 *Acad Sci USA.* 99: 15446-15450. doi: 10.1073/pnas.242548399  
29  
30 Olivera BM, et al. 2012. Adaptive radiation of venomous marine snail lineages and the accelerated evolution  
31 of venom peptide genes. *Ann NY Acad Sci.* 1267: 61-70. doi: 10.1111/j.1749-6632.2012.06603.x  
32  
33 Otero-González AJ, et al. 2010. Antimicrobial peptides from marine invertebrates as a new frontier for  
34 microbial infection control. *FASEB J.* 24: 1320-1334. doi: 10.1096/fj.09-143388  
35  
36 Petersen TN, Brunak S, von Heijne G, Nielsen H 2011. SignalP 4.0: discriminating signal peptides from  
37 transmembrane regions. *Nat Methods.* 8: 785-786.  
38  
39 Philipp EER, et al. 2012. Massively parallel rna sequencing identifies a complex immune gene repertoire in  
40 the lophotrochozoan *Mytilus edulis*. *PLoS ONE.* 7: e33091.  
41  
42 Puillandre N, Koua D, Favreau P, Olivera BM, Stöcklin R 2012. Molecular phylogeny, classification and  
43 evolution of conopeptides. *J Mol Evol.* 74: 297-309. doi: 10.1007/s00239-012-9507-2  
44  
45 Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM 2003. Assumption-free analysis of quantitative real-  
46 time polymerase chain reaction (PCR) data. *Neurosci Lett* 339: 62-66. doi: [http://dx.doi.org/10.1016/S0304-](http://dx.doi.org/10.1016/S0304-3940(02)01423-4)  
47 [3940\(02\)01423-4](http://dx.doi.org/10.1016/S0304-3940(02)01423-4)  
48  
49 Ranasinghe S, McManus DP 2013. Structure and function of invertebrate Kunitz serine protease inhibitors.  
50 *Dev Comp Immunol.* 39: 219-227.  
51  
52 Ratzka C, et al. 2012. Molecular Characterization of antimicrobial peptide genes of the carpenter ant  
53 *Camponotus floridanus*. *PLoS ONE* 7: e43036. doi: 10.1371/journal.pone.0043036  
54  
55 Reese MG, Eeckman FH, Kulp D, Haussler D 1997. Improved Splice Site Detection in Genie. *J Comput*  
56 *Biol.* 4: 311-323. doi: 10.1089/cmb.1997.4.311  
57  
58  
59  
60

- 1  
2  
3 Rice P, Longden I, Bleasby A 2000. EMBOSS: The European Molecular Biology Open Software Suite.  
4 Trends Genet. 16: 276-277. doi: [http://dx.doi.org/10.1016/S0168-9525\(00\)02024-2](http://dx.doi.org/10.1016/S0168-9525(00)02024-2)  
5  
6 Rimphanitchayakit V, Tassanakajon A 2010. Structure and function of invertebrate Kazal-type serine  
7 proteinase inhibitors. Dev Comp Immunol. 34: 377-386.  
8  
9 Rodríguez de la Vega RC, Possani LD 2005. On the evolution of invertebrate defensins. Trends Genet. 21:  
10 330-332. doi: <http://dx.doi.org/10.1016/j.tig.2005.03.009>  
11  
12 Romiguier J, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic  
13 diversity. Nature. 515: 261-263. doi: 10.1038/nature13685  
14  
15  
16 Rosani U, et al. 2011. Massively parallel amplicon sequencing reveals isotype-specific variability of  
17 antimicrobial peptide transcripts in *Mytilus galloprovincialis*. PLoS ONE. 6: e26680.  
18  
19 Saez NJ, et al. 2010. Spider-venom peptides as therapeutics. Toxins. 2: 2851-2871.  
20  
21 Sperstad SV, et al. 2011. Antimicrobial peptides from marine invertebrates: Challenges and perspectives in  
22 marine antimicrobial peptide discovery. Biotechnol Adv. 29: 519-530.  
23  
24 Stamatakis A 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of  
25 taxa and mixed models. Bioinformatics. 22: 2688-2690. doi: 10.1093/bioinformatics/btl446  
26  
27 Steiner AM, Bulaj G, Puillandre N 2013. On the importance of oxidative folding in the evolution of  
28 conotoxins: cysteine codon preservation through gene duplication and adaptation. PLoS ONE. 8: e78456.  
29 doi: 10.1371/journal.pone.0078456  
30  
31 Suárez-Ulloa V, et al. 2013a. The CHROMEVALOA database: A resource for the evaluation of Okadaic  
32 Acid contamination in the marine environment based on the chromatin-associated transcriptome of the  
33 mussel *Mytilus galloprovincialis*. Mar Drugs. 11: 830-841.  
34  
35 Suárez-Ulloa V, et al. 2013b. Bivalve omics: state of the art and potential applications for the biomonitoring  
36 of harmful marine compounds. Mar Drugs. 11: 4370-4389.  
37  
38 Takeuchi T, et al. 2012. Draft genome of the pearl oyster *Pinctada fucata*: A platform for understanding  
39 bivalve biology. DNA Res. 19: 117-130. doi: 10.1093/dnares/dss005  
40  
41 Taylor K, Barran PE, Dorin JR 2008. Review: Structure-activity relationships in  $\beta$ -defensin peptides.  
42 Biopolymers - Peptide Science Section. 90: 1-7.  
43  
44 Tedford HW, Fletcher JI, King GF 2001. Functional significance of the  $\beta$ -Hairpin in the insecticidal  
45 neurotoxin  $\omega$ -Atracotoxin-Hv1a. J Biol Chem. 276: 26568-26576. doi: 10.1074/jbc.M102199200  
46  
47 Terlau H, Olivera BM 2004. Conus Venoms: A Rich Source of Novel Ion Channel-Targeted Peptides.  
48 Physiol Rev. 84: 41-68.  
49  
50 Toubiana M, et al. 2013. Toll-like receptors and MyD88 adaptors in *Mytilus*: Complete cds and gene  
51 expression levels. Dev Comp Immunol. 40: 158-166.  
52  
53 Toubiana M, et al. 2014. Toll signal transduction pathway in bivalves: Complete cds of intermediate  
54 elements and related gene transcription levels in hemocytes of immune stimulated *Mytilus galloprovincialis*.  
55 Dev Comp Immunol. 45: 300-312. doi: <http://dx.doi.org/10.1016/j.dci.2014.03.021>  
56  
57 Venier P, et al. 2009. MytiBase: a knowledgebase of mussel (*M. galloprovincialis*) transcribed sequences.  
58 BMC Genomics. 10: 72.  
59  
60

1  
2  
3  
4 Wang C-Z, Jiang H, Ou Z-L, Chen J-S, Chi C-W 2003. cDNA cloning of two A-superfamily conotoxins  
5 from *Conus striatus*. *Toxicon*. 42: 613-619. doi: 10.1016/j.toxicon.2003.08.005  
6

7 Wang J, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*.  
8 490: 49-54.  
9

10 Xiong Y-M, Ling M-H, Wang D-C, Chi C-W 1997. The cDNA and genomic DNA sequences of a  
11 mammalian neurotoxin from the scorpion *Buthus martensii* Karsch. *Toxicon*. 35: 1025-1031. doi:  
12 10.1016/s0041-0101(96)00224-3  
13

14 Xu B, Yang Z 2013. pamlX: A Graphical User Interface for PAML. *Mol Biol Evol*. doi:  
15 10.1093/molbev/mst179  
16

17 Yang S, et al. 2012. Chemical punch packed in venoms makes centipedes excellent predators. *Mol Cell*  
18 *Proteomics*. 11: 640-650. doi: 10.1074/mcp.M112.018853  
19

20 Yang Z 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24: 1586-1591. doi:  
21 10.1093/molbev/msm088  
22

23 Zhang G, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*.  
24 490: 49-54.  
25

26 Zhang Y, et al. 2010. Transcriptome analysis of the venom glands of the Chinese wolf spider *Lycosa*  
27 *singoriensis*. *Zoology*. 113: 10-18.  
28

29 Zhong Y, et al. 2014. A novel neurotoxin from venom of the spider, *Brachypelma albopilosum*. *PLoS ONE*.  
30 9: e110221. doi: 10.1371/journal.pone.0110221  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1:** primers designed for assessing the tissue-specific expression levels of MgCRP-I genes by real-time PCR.

Primer name	Primer sequence
MgCRP-I 1 for	TGTGTGTTGTTGGTCGTCGT
MgCRP-I 1 rev	GTAACCGGAACGACAAAAGC
MgCRP-I 2 for	AGCCTCAAGTAAGAAGTAAAACAGA
MgCRP-I 2 rev	CAGCTTMTTCTACCGCATCC
MgCRP-I 3/25 for	GACAAAGTGAAGTAAAGCATTTC
MgCRP-I 3/25 rev	CTCCGTTTTCTCCAAAGCTG
MgCRP-I 4 for	CATGGCACATGAMGAAATGC
MgCRP-I 4 rev	TTAGCCACCATAGCGTTTGC
MgCRP-I 5 for	TGGATAAAAAGGTGACCCACAG
MgCRP-I 5 rev	TCTTCCAGCATTTTCGTCCTT
MgCRP-I 6 for	AAYATGGCGAAGGAAGACAT
MgCRP-I 6 rev	AAGTTCAGTCGCGCCTACAT
MgCRP-I 7 for	GTTGGAGTCAACATGGCAAA
MgCRP-I 7 rev	GCGCATGCATTTTCTGTAAG
MgCRP-I 8 for	GCATTTGCTTATAGTGTTCAGAG
MgCRP-I 8 rev	TKCAAATGATGGATGGCTAA
MgCRP-I 9 for	GCTTTTTGTTTGTGGTAGCC
MgCRP-I 9 rev	CGAACACATCTTCTGTATGAGCA
MgCRP-I 10/26 for	GGCACATGAAGAAATGTTTCG
MgCRP-I 10/26 rev	CCTGCATACGCCAAAACAT
MgCRP-I 11 for	TAAACCCCTTGTTTCGGTCAC
MgCRP-I 11 rev	AGTGTGACGGATGCAAACAA
MgCRP-I 14 for	AGCCTTCGTTGGAAGTAGCA
MgCRP-I 14 rev	TCGAGCGAGATTGACATCTG
multi-MgCRP-I 1 for	CTGACGAAATGGTGGAGGAT
multi-MgCRP-I 1 rev	TACAGCATTGACGGCTGTTT
multi-MgCRP-I 2 for	GCAAACATGGCCAAAGAAGT
multi-MgCRP-I 2 rev	GTCACGGGTCTTTTTGCATT
multi-MgCRP-I 3 for	AAGAGCTCCTGCATGTGGAT
multi-MgCRP-I 3 rev	TCCTCCTCCGTTCTCTTTT
EF-1 alpha for	CCTCCCACCATCAAGACCTA
EF-1 alpha rev	GGCTGGAGCAAGGTAACAA

**Table 2:** list of the MgCRP-I sequences identified in the present work. \*complete sequence correspond to a full-length coding sequence, from the initial ATG to the STOP codon; \*\*T: transcriptome; G: genome.

Sequence name	Status*	Evidence**	Genomic scaffold	Cysteine-rich domains
MytiCRP-I 1	complete	T	/	1
MytiCRP-I 2	complete	G, T	APJB011836849.1	1
MytiCRP-I 3	complete	T	/	1
MytiCRP-I 4	complete	G, T	APJB011405634.1	1
MytiCRP-I 5	complete	G, T	APJB010137175.1	1
MytiCRP-I 6	complete	T	/	1
MytiCRP-I 7	complete	T	/	1
MytiCRP-I 8	complete	G, T	APJB0118677191.1	1
MytiCRP-I 9	complete	G, T	APJB010390130.1	1
MytiCRP-I 10	complete	T	/	1
MytiCRP-I 11	incomplete	T	/	1
MytiCRP-I 12	complete	T	/	none
MytiCRP-I 13	complete	G, T	APJB010539225.1	1
MytiCRP-I 14	complete	G, T	APJB011981014.1	1
MytiCRP-I 15	complete	G, T	APJB012303283.1	1
MytiCRP-I 16	complete	T	/	1
MytiCRP-I 17	complete	T	/	1
MytiCRP-I 18	complete	G, T	APJB010096024.1	1
MytiCRP-I 19	complete	T	/	1
MytiCRP-I 20	incomplete	T	/	1
MytiCRP-I 21	complete	T	/	1
MytiCRP-I 22	complete	G, T	APJB010149223.1	1
MytiCRP-I 23	complete	T	/	none
MytiCRP-I 24	complete	G, T	APJB011420996.1	1
MytiCRP-I 25	complete	G, T	APJB011525896.1	1
MytiCRP-I 26	complete	G, T	APJB011451595.1	1
MytiCRP-I 27	incomplete	G	APJB010022937.1	1
MytiCRP-I 28	complete	G	APJB010019019.1	1
MytiCRP-I 29	incomplete	G, T	APJB010215939.1	1
MytiCRP-I 30	incomplete	G	APJB010309773.1	1
MytiCRP-I 31	complete	G, T	APJB010337167.1	1
MytiCRP-I 32	incomplete	G	APJB010405325.1	1
MytiCRP-I 33	incomplete	G	APJB010538560.1	1
MytiCRP-I 34	complete	G, T	APJB010602145.1	1
MytiCRP-I 35	incomplete	G	APJB010726714.1	1
MytiCRP-I 36	complete	G, T	APJB010858750.1	1
MytiCRP-I 37	incomplete	G	APJB011013544.1	1
MytiCRP-I 38	incomplete	G, T	APJB011377302.1	1
MytiCRP-I 39	incomplete	G, T	APJB011417411.1	1
MytiCRP-I 40	complete	G	APJB011602152.1	1
MytiCRP-I 41	incomplete	G	APJB01171896.1	1
MytiCRP-I 42	incomplete	G	APJB011833940.1	1

1					
2					
3	MytiCRP-I 43	incomplete	G	APJB011892489.1	1
4	MytiCRP-I 44	incomplete	G	APJB011902451.1	1
5	MytiCRP-I 45	complete	G, T	APJB012001676.1	1
6	MytiCRP-I 46	incomplete	G	APJB012002994.1	1
7	MytiCRP-I 47	incomplete	G	APJB012084462.1	1
8	MytiCRP-I 48	incomplete	G	APJB011591868.1	1
9	MytiCRP-I 49	incomplete	G	APJB011815456.1	1
10	MytiCRP-I 50	complete	T	/	1
11	MytiCRP-I 51	complete	T	/	1
12	MultiMytiCRP-I 1	complete	T	/	3
13	MultiMytiCRP-I 2	complete	G, T	APJB011508508.1	4
14	MultiMytiCRP-I 3	complete	G, T	APJB012209485.1	2
15	MultiMytiCRP-I 4	incomplete	T	/	2
16	MultiMytiCRP-I 5	complete	T	/	2
17	MultiMytiCRP-I 6	incomplete	G	APJB010388843.1	2 or more
18	MultiMytiCRP-I 7	incomplete	G	APJB010167718.1	2 or more
19	MultiMytiCRP-I 8	incomplete	G	APJB010303175.1	3 or more
20	MultiMytiCRP-I 9	incomplete	G	APJB010305277.1	4
21	MultiMytiCRP-I 10	incomplete	G	APJB010449694.1	4
22	MultiMytiCRP-I 11	complete	G	APJB010750334.1	2
23	MultiMytiCRP-I 12	complete	G	APJB011083975.1	2
24	MultiMytiCRP-I 13	complete	G	APJB011153262.1	2
25	MultiMytiCRP-I 14	complete	G, T	APJB011903515.1	4
26	MultiMytiCRP-I 15	incomplete	G	APJB011965594.1	2
27	MultiMytiCRP-I 16	complete	T	/	2
28					
29					
30					
31					
32					
33					

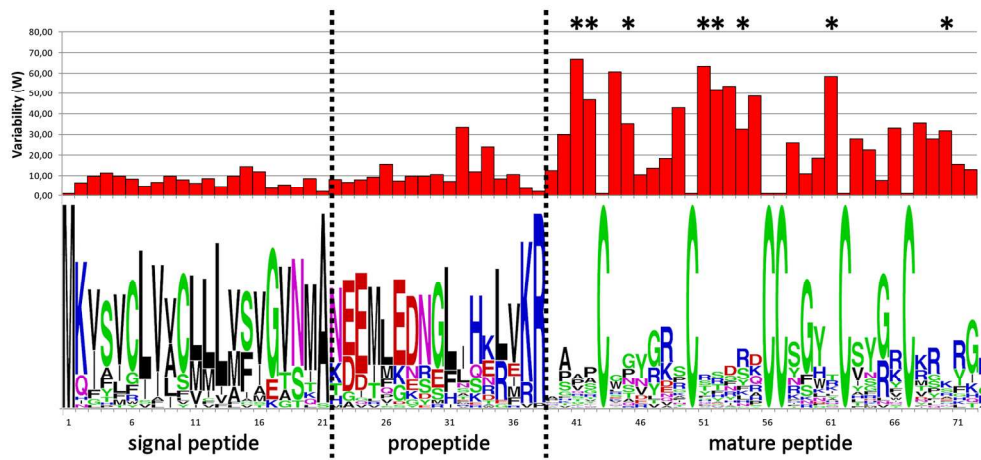
**Table 3:** Number of peptides with a MgCRP-I-like array found in the UniProtKB/Swiss-Prot protein sequence database, listed per taxonomic group (function is indicated if available).

Taxonomic group	Number of identified CRPs in UniprotKB*	Molecular function
Cone snails	163	conotoxins
Spiders	223	venom toxins
Fungi	2	uncharacterized
Viruses		
(Buculoviridae)	4	uncharacterized
Insects	35	hormones/venom toxins
Green Plants	5	antimicrobial peptides
Scorpions	8	venom toxins
Terebrids	2	augertoxins
Horseshoe crabs	1	antimicrobial peptides
Nematodes	1	Insulin-like
Vertebrates	6	unknown/antimicrobial peptides
Turrids	2	turritoxins

\*Non-redundant positive matches based on threshold criteria of a 95% sequence identity.

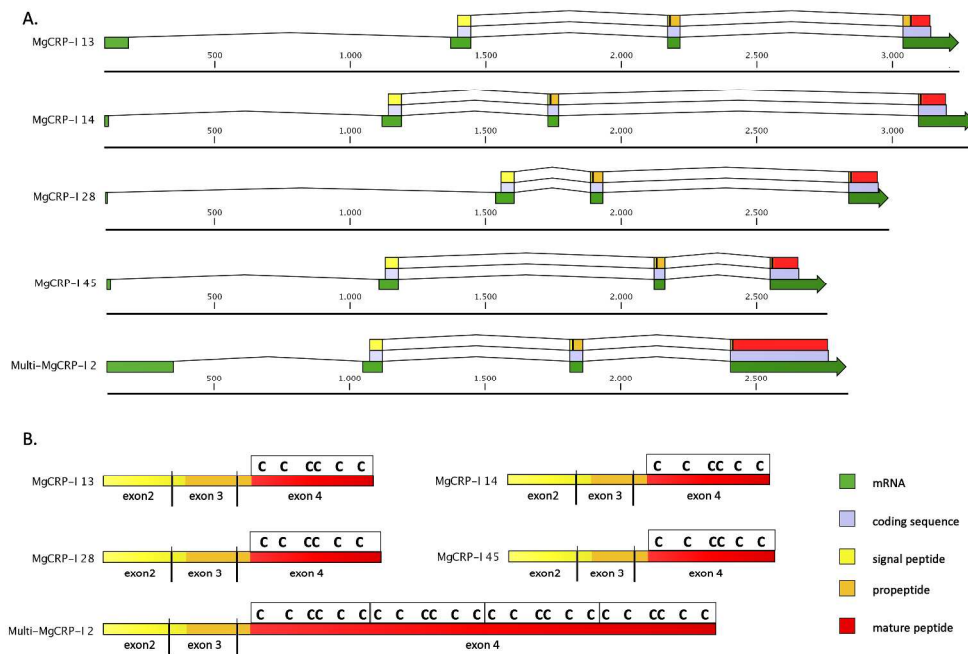
**Table 4:** Number of CRP-I sequencing reads identified in the publicly available transcriptome datasets from *Mytilus* spp. (retrieved from NCBI SRA, February 2015).

Database	REF	tissue	sequencing strategy	total number of sequences	sequences related to MgCRP-I	
<i>M. galloprovincialis</i>	(Venier, et al. 2009)	mixed tissues	Sanger	19,617	0	0
<i>M. californianus</i>	NA	mixed tissues	Sanger	42,354	0	0
<i>M. coruscus</i>	NA	foot	Sanger	719	0	0
<i>M. galloprovincialis</i>	(Craft, et al. 2010)	foot	454	31,227	0	0
<i>M. galloprovincialis</i>	(Craft, et al. 2010)	mantle	454	52,057	0	0
<i>M. galloprovincialis</i>	(Suárez-Ulloa, et al. 2013a)	digestive gland	454	2,206,478	0	0
<i>M. trossulus</i>	(Romiguier, et al. 2014)	mixed tissues	Illumina	~58 million	142	<0
<i>M. galloprovincialis</i>	NA	hemocytes	Illumina	~106 million	490	<0
<i>M. galloprovincialis</i>	NA	gills	Illumina	~52 million	182	<0
<i>M. edulis</i>	(Philipp, et al. 2012)	hemocytes	454	407,061	2	<0
<i>M. edulis</i>	(Bassim, et al. 2014)	larvae	Illumina	~295 million	3,423	0.0
<i>M. californianus</i>	(Romiguier, et al. 2014)	mixed tissues	Illumina	~78 million	644	0.0
<i>M. edulis</i>	(Philipp, et al. 2012)	mixed tissues	454	365,626	3	0.0
<i>M. edulis</i>	(Freer, et al. 2014)	mantle	454	494,391	8	0.0
<i>M. galloprovincialis</i>	(Craft, et al. 2010)	gill	454	58,271	1	0.0
<i>M. galloprovincialis</i>	NA	gills	Illumina	~120 million	3,103	0.0
<i>M. edulis</i>	(Philipp, et al. 2012)	digestive gland	454	1,112,061	30	0.0
<i>M. galloprovincialis</i>	(Gerdol, et al. 2014)	digestive gland	Illumina	~54 million	3,269	0.0
<i>M. galloprovincialis</i>	NA	posterior adductor muscle	Illumina	~103 million	9,429	0.0
<i>M. galloprovincialis</i>	(Romiguier, et al. 2014)	mixed tissues	Illumina	~108 million	11,695	0.0
<i>M. edulis</i>	(Philipp, et al. 2012)	inner mantle	454	323,482	46	0.0
<i>M. edulis</i>	(Romiguier, et al. 2014)	mixed tissues	Illumina	~103 million	15,086	0.0
<i>M. edulis</i>	(González, et al. 2015)	mantle/foot	Illumina	~49 million	8,492	0.0
<i>M. edulis</i>	NA	mixed tissues	Sanger	5,300	1	0.0
<i>M. galloprovincialis</i>	NA	mantle	Illumina	~108 million	30,802	0.0
<i>M. galloprovincialis</i>	(Craft, et al. 2010)	digestive gland	454	33,992	2	0.0
<i>M. edulis</i>	(Philipp, et al. 2012)	mantle rim	454	324,592	299	0.0

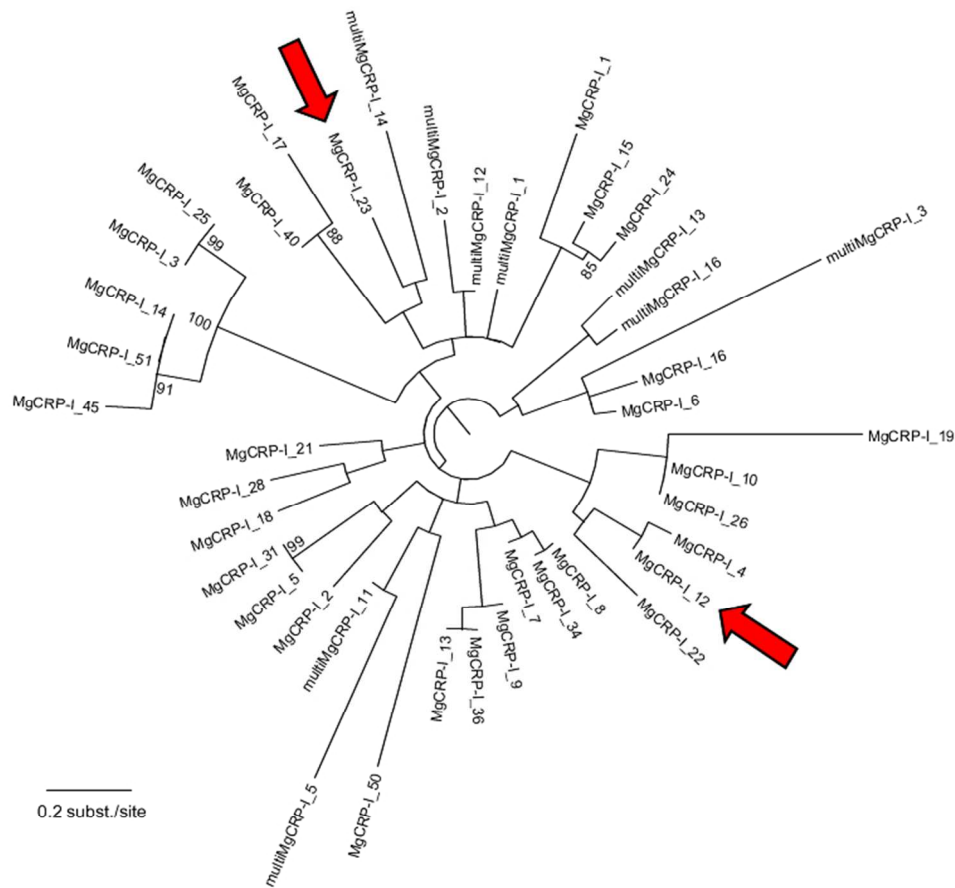


Sequence variability of MgCRP-I sequences; variability index  $W$  is plotted in the upper panel, while the sequence consensus, obtained with Weblogo (<http://weblogo.berkeley.edu>) is shown in the lower panel. Only positions covered by at least 50% sequences in the global alignment of MgCRP-I peptides are shown. Sites under positive selection are indicated by an asterisk.

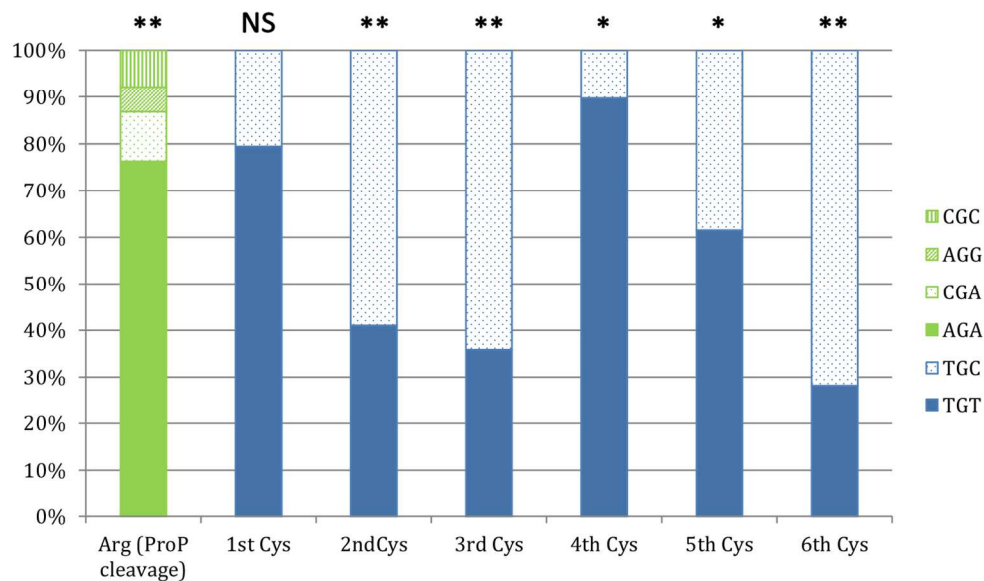
Peer Review



Exon/intron structure of the complete coding regions of the MgCRP-I 13, 14, 28, 45 and multi-MgCRP-I 2 genes (panel A) and corresponding organization of the encoded peptide precursors (panel B). The positions of the signal peptide, pro-region and mature peptide regions are highlighted and each cysteine-rich module is marked by a box.

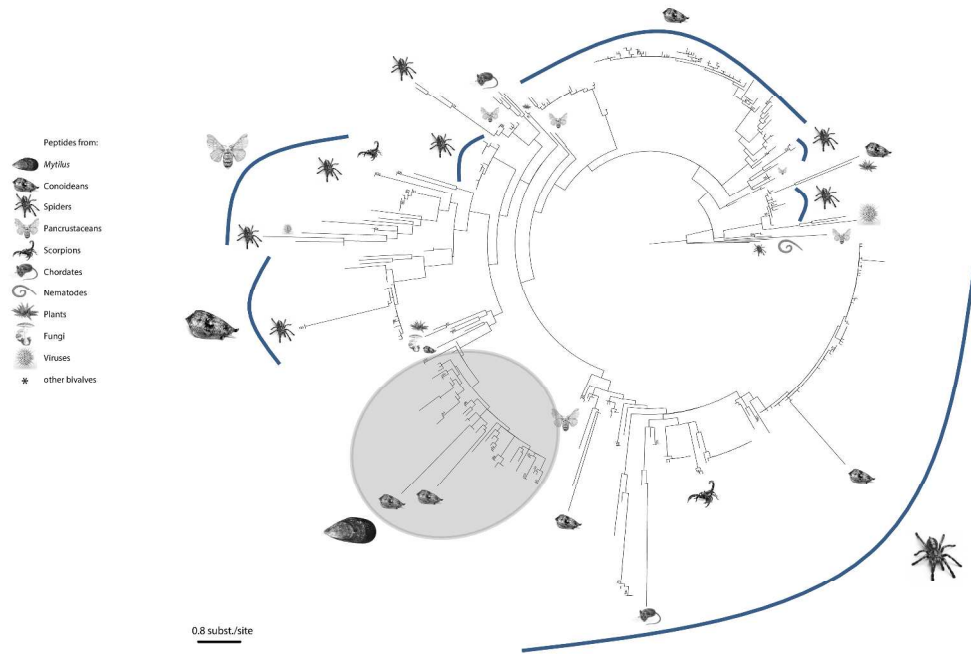


Maximum likelihood tree obtained with the MgCRP-I peptides based on the alignment of the signal peptide only. Only Booststraps values > 75 are shown. Some sequences were not considered in this analysis as their N-terminal region was incomplete (see Table 2). Arrows indicate two MgCRP-I peptides



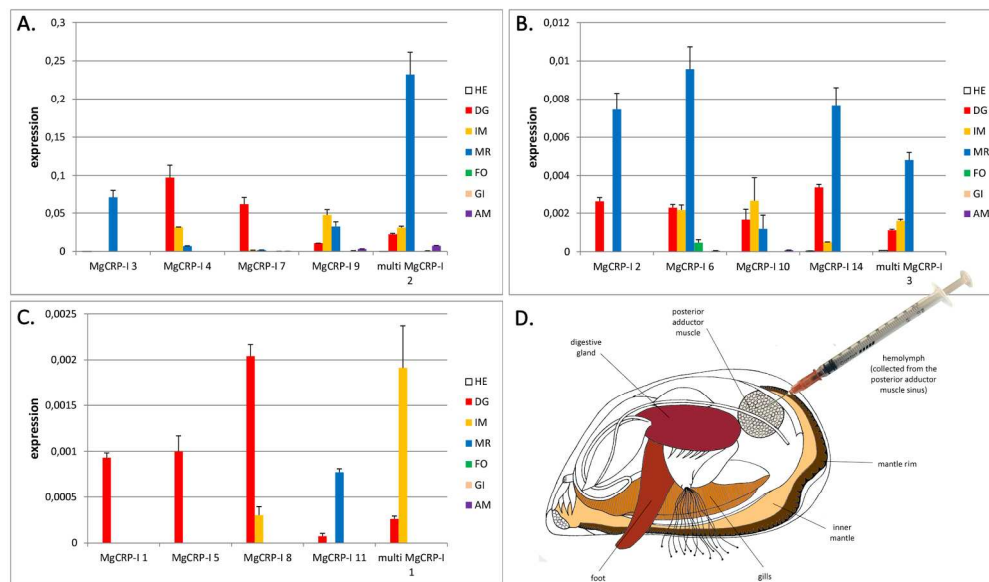
codon usage for the Arg residue responsible of the pro-peptide cleavage site and for the 6 cysteine residues involved in the disulfide bridges, calculated on *M. galloprovincialis* MgCRP-I peptides. The probabilities of finding the observed codon biases were calculated assuming a binomial distribution and the codon usage (a priori probabilities) inferred from the transcriptome published by Gerdol et al. 2014 (75.1%-24.9% for TGT-TGC, encoding Cys, 49.3%-16.3%-13.3%-12.5%-4.7%-3.8% for AGA-AGG-CGA-CGT-CGG-CGC, encoding Arg). Significant ( $p < 0.05$ ) and highly significant ( $p < 0.01$ ) deviations from the expected distributions ( $p < 0,01$ ) are marked by \* and \*\* respectively. NS = not significant.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



Maximum likelihood tree obtained with the MgCRP-I, the orthologous sequences from other mussel species and all the CRPs mined from UniProtKB/Swiss-Prot (see Materials and methods). Peptides from pancrustaceans, conoideans, spiders, scorpions, nematodes, chordates, plants, fungi, viruses and mussels are shown. Mussel CRP-I peptides are highlighted in a grey background.

Review



gene expression of 17 selected MgCRP-I genes in six tissues (H: hemolymph; DG: digestive gland; IM: inner mantle; MR: mantle rim; F: foot; G: gills; AM: posterior adductor muscle); primers for MgCRP-I 3 also target MgCRP-I 25, primers for MgCRP-I 10 also target MgCRP-I 26. Bars represent the expression relative to EF-1 alpha; results are mean ± standard deviation of three replicates. MgCRP-I sequences are divided into three panels based on their expression level: panel A – genes with maximum relative expression value comprised between 0.05 and 0.25; panel B – genes with maximum relative expression value comprised between 0.004 and 0.01; panel C - with maximum relative expression value lower than 0.003. Panel D shows a schematic representation of a *M. galloprovincialis* anatomical features, highlighting the sampled tissues.

review