

# Massively Parallel Amplicon Sequencing Reveals Isotype-Specific Variability of Antimicrobial Peptide Transcripts in *Mytilus galloprovincialis*

Umberto Rosani<sup>1\*</sup>, Laura Varotto<sup>1</sup>, Alberta Rossi<sup>1</sup>, Philippe Roch<sup>2</sup>, Beatriz Novoa<sup>3</sup>, Antonio Figueras<sup>3</sup>, Alberto Pallavicini<sup>4</sup>, Paola Venier<sup>1\*</sup>

**1** Department of Biology, University of Padua, Padova, Italy, **2** Ecologie des Systèmes Marins et Côtiers, CNRS-IRD-University of Montpellier 2, Montpellier, France, **3** Instituto de Investigaciones Marinas, CSIC, Vigo, Spain, **4** Department of Life Sciences, University of Trieste, Trieste, Italy

## Abstract

**Background:** Effective innate responses against potential pathogens are essential in the living world and possibly contributed to the evolutionary success of invertebrates. Taken together, antimicrobial peptide (AMP) precursors of defensin, mytilin, myticin and mytimycin can represent about 40% of the hemocyte transcriptome in mussels injected with viral-like and bacterial preparations, and unique profiles of myticin C variants are expressed in single mussels. Based on amplicon pyrosequencing, we have ascertained and compared the natural and *Vibrio*-induced diversity of AMP transcripts in mussel hemocytes from three European regions.

**Methodology/Principal Findings:** Hemolymph was collected from mussels farmed in the coastal regions of Palavas (France), Vigo (Spain) and Venice (Italy). To represent the AMP families known in *M. galloprovincialis*, nine transcript sequences have been selected, amplified from hemocyte RNA and subjected to pyrosequencing. Hemolymph from farmed (offshore) and wild (lagoon) Venice mussels, both injected with  $10^7$  *Vibrio* cells, were similarly processed. Amplicon pyrosequencing emphasized the AMP transcript diversity, with Single Nucleotide Changes (SNC) minimal for mytilin B/C and maximal for arthropod-like defensin and myticin C. Ratio of non-synonymous vs. synonymous changes also greatly differed between AMP isotypes. Overall, each amplicon revealed similar levels of nucleotidic variation across geographical regions, with two main sequence patterns confirmed for mytimycin and no substantial changes after immunostimulation.

**Conclusions/Significance:** Barcoding and bidirectional pyrosequencing allowed us to map and compare the transcript diversity of known mussel AMPs. Though most of the genuine cds variation was common to the analyzed samples we could estimate from 9 to 106 peptide variants in hemolymph pools representing 100 mussels, depending on the AMP isoform and sampling site. In this study, no prevailing SNC patterns related to geographical origin or *Vibrio* injection emerged. Whether or not the contact with potential pathogens can increase the amount of AMP transcript variants in mussels requires additional study.

**Citation:** Rosani U, Varotto L, Rossi A, Roch P, Novoa B, et al. (2011) Massively Parallel Amplicon Sequencing Reveals Isotype-Specific Variability of Antimicrobial Peptide Transcripts in *Mytilus galloprovincialis*. PLoS ONE 6(11): e26680. doi:10.1371/journal.pone.0026680

**Editor:** Christos A. Ouzounis, The Centre for Research and Technology, Greece

**Received:** July 2, 2011; **Accepted:** October 2, 2011; **Published:** November 7, 2011

**Copyright:** © 2011 Rosani et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was funded by the European project FOOD-CT-2005-007103 (<http://imaquanim.dfvf.dk/info>) and, in part, by the Italian Project 20084BEJ9F-PRIN08 (<http://prin.miur.it>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: [umberto.rosani@unipd.it](mailto:umberto.rosani@unipd.it) (UR); [paola.venier@unipd.it](mailto:paola.venier@unipd.it) (PV)

## Introduction

*Mytilus* species (Phylum Mollusca, Class Bivalvia) are intertidal filter-feeders distributed worldwide, anchored to hard substrates in dense communities and widely used as bio-sensors of coastal pollution. Mussel populations of the northern and southern hemisphere probably separated 0.54–1.31 million years ago, far after the trans-Arctic expansion towards North America, and before the divergence between the Atlantic and Mediterranean ecotypes [1]. *M. galloprovincialis* hybridizes with *M. edulis* in southwest England and the Mediterranean mussel is now reported in Eastern Asia, California, Chile and Western Australia [1–2].

At different latitudes, mussels face tidal and seasonal fluctuations, changeable pollutant loads and also the surrounding bioma

with behavioral changes [3], metabolic adjustments [4] and a variety of defense reactions [5–7]. With the exception of a few metazoan parasites which also somewhat affect the *Mytilus* species [8–9], mussels seem refractory to diseases and could instead influence the prevalence of pathogens such as *Perkinsus* spp. and *Betanodavirus* (Nodaviridae) in other bivalves and fishes, respectively [10]. Like other invertebrates, bivalve molluscs rely on ancient and rapid defenses to fight potential pathogens, and gene-encoded antimicrobial peptides (AMPs) are major humoral components of their immune system.

Host defense peptides are present in virtually all living organisms, with more than 30 AMPs expressed in humans and about 200 peptides identified in insects (approximately 1500 molecules very diverse in sequence and secondary structures are

reported in specific databases). [11–13]. Among other structural features, a conserved  $\gamma$ -core motif originated from the bidirectional orientation of specific aminoacid residues including an invariant cysteine array has indicated the evolutionary relatedness of cysteine-stabilized  $\alpha$ - $\beta$  (CS- $\alpha\beta$ ) AMPs, kinocidins, invertebrate toxins and snake venoms: such unifying structure provides an interesting hypothesis for context-specific action modes, from the perturbation of negatively charged cell membranes and ion channels to the immunoregulatory functions [14].

In the continuous fight with competitors, predators and pathogens, the evolutionary diversification of AMP types and gene families likely occurred through events of gene duplication, shuffling of functional elements and selection for variation at positions adjacent, or integral to, the conserved structural motifs [13,15]. In *Crassostrea gigas*, combined mechanisms of sequence diversification (e.g. recombination, parallel homoplastic mutations, indel events) and directional selection have been suggested to explain the remarkable gene multiplicity and variable copy number of defensins and proline-rich peptides, whereas the marked transcript diversity of Cg-bpi, a bactericidal permeability protein, has been mainly referred to the allelic polymorphism of one single gene [16]. Microsatellite-mediated mosaics of sequence elements, low-transcription fidelity and transcript editing support the evidence of about 50 polymorphic genes, and an extraordinary diverse set of Sp185/333 proteins expressed in response to pathogens by the purple sea urchin [17–18]. Worthy of note, the copy number polymorphism of  $\alpha$  and  $\beta$  defensin genes with proportional peptide levels in neutrophilic granulocytes, has been related to the individual risk of infection in humans [19–20].

Tens of different AMPs or AMP families have been discovered in marine invertebrates [21]. In the mussels *M. galloprovincialis* and *M. edulis*, four different groups of CS-  $\alpha\beta$  AMPs with multiple isoforms have been discovered and classified according to their primary sequence and secondary structure: defensins reported as MGD1 and MGD2, mytilin A, B, C, D and E, myticin A, B and C, mytimycin, the only strictly antifungal peptide with an EF-hand like domain [22–25]. These AMPs share small size (3.7–4.5 kDa, except mytimycin of 6.2 kDa), positive charge and amphiphilic behavior. Their precursors (pre-pro-peptides) consist of an N-terminal signal peptide, a central mature peptide and a C-terminal extension. Each family is characterized by a cysteine array of 8 (12 in mytimycin) cysteines engaged in intramolecular disulfide bonds.

A broad spectrum of activity, often complementary and not strictly antibacterial, was reported for the mussel defensins, mytilins and myticins [26–27] whereas mytimycin, a 6.2 kDa peptide isolated from normal and immunostimulated mussels, selectively inhibited *Neurospora* and *Fusarium* growth [22]. Whether purified in sufficient amounts from cellular fractions or obtained in stable conformations by chemical synthesis or recombinant system, pure peptides are essential to investigate the antibiotic power of the different mussel AMPs.

*In situ* hybridization and immunolocalization assays performed on mussel hemocytes demonstrated a partially overlapping expression of defensins and mytilins [28]. AMP expression and stored peptides have been observed in several tissues and developmental stages [28–29] to indicate that cells other than hemocytes can produce and release AMPs, a phenomenon well known from frog skin [30] and the male reproductive system of rats [31]. Overall, mussel AMPs display rather complex expression patterns, dependent on developmental stage, seasonality and immunostimulation [32–34]. In *M. galloprovincialis*, massive EST sequencing confirmed the abundance and transcript diversity of AMPs and other key players of the innate immunity [35]. The AMP precursors represented 26–43% of the hemocyte transcripts

in mussels injected with viral-like and bacterial preparations; in particular, 74 precursor and 25 mature peptide variants of myticin C were detected in a sample of only 100 mussels, with unique profiles of transcript variants in single mussels and less common alleles differing at single nucleotide positions from the two most common ones [29,36].

The myticin C variation is also remarkable compared to mytilin B, as one mussel can produce 2–10 different mytilin B transcripts but silent substitutions restrict the peptide variants to only a few [37]. In spite of the abundance of other AMPs, just one singleton plus 4 similar sequences denote mytimycin in Mytibase, interactive catalogue including 18788 expressed sequence tags (ESTs) of *M. galloprovincialis* [25]. Sequencing and Southern blot data indicate one gene copy per genome for defensin MGD2, mytilin B and myticin C [38,36]. Two gene copies or allelic polymorphism could explain the simultaneous presence of two length variants of the mytimycin gene per mussel [39]. The gene copy number of the mussel AMPs need verification since partial gene sequences covering the coding sequence (cds) are only available for MGD1, mytilin B, myticin C and mytimycin [29,35–39]. Two 3D structures have been established by NMR spectrometry, defensin MGD1 [40] and mytilin B [41].

Thus far no *Mytilus* genome has been sequenced and, compared to better known model organisms, a limited number of genes have been investigated: for instance those concerning defensin and mytilins [38], heat shock proteins [42], metallothioneins [43] and apoptotic caspases [44]. Also gene-centered studies take advantage of the massive production of ESTs which currently contributes to the identification of molecules and pathways underlying the mussel response to various natural and experimental conditions [45–48]. Among the 67,726 ESTs and 4680 aminoacid sequences publicly available for the *Mytilus* genus, about 29 and 32%, respectively, refer to *M. galloprovincialis* (www.ncbi.nlm.nih.gov, June 2011). Recently, new ESTs have been produced from the digestive gland, foot, gill and mantle of *M. galloprovincialis* by advanced sequencing [49]. The so called ‘pyrosequencing’ was the first alternative to the use of chain-terminating inhibitors [50] and it has radically increased the sequencing power as well as the resolution of low-abundance variants [51–53]. Adequate read coverage can assure reliable quantification of single nucleotide changes (SNC) when seeking critical mutations or sequence polymorphisms.

Based on 454 pyrosequencing, we have thoroughly studied the sequence diversity of 9 different AMP precursors expressed in hemocytes of mussels (*M. galloprovincialis*) farmed in three European regions. Similarly, we have analyzed and compared mussels farmed offshore or living inside the Venice Lagoon (Italy), before and after injection with live *Vibrio* cells.

## Results

In Table 1 we summarized the main features of nine CS-  $\alpha\beta$  AMPs expressed in *M. galloprovincialis*.

Following appropriate primer design, we amplified the related transcript sequences from hemolymph pools representing groups of 100 mussels farmed in south France (Pa), northwest Spain (Vi) and northeast Italy (Ve) or native from the industrial canals of the Lagoon of Venice (Ve nc). In addition, groups of 40 offshore-farmed (Ve ft) and lagoon-native (Ve nt) mussels were injected with  $10^7$  live *Vibrio splendidus* cells and similarly processed (Table 2). The resulting 78 PCR products (13 amplicons $\times$ 6 samples) were then purified, quantified, diluted to the appropriate concentration ( $\sim 7 \times 10^9$  molecules/ $\mu$ l) to compose two equimolecular pools for the emulsion PCR (cDNA concentration was 2.3 and 2.5 ng/ $\mu$ l, respectively) and bidirectional sequencing.

**Table 1.** Features of the selected mussel AMPs.

AMP	Precursor length (aa)	Mature peptide length (aa)	Cysteine array	Weight (kDa)	Isoelectric point (pH)	Hydrophobicity ratio (%)
<b>MytA</b>	96	40	8	4.5	9.1	35
<b>MytB</b>	96	40	8	4.6	9.2	30
<b>MytC</b>	100	40	8	4.4	8.8	35
<b>MytIB</b>	103	34	8	4.0	9.7	27
<b>MytIC</b>	100	34	8	4.2	9.9	24
<b>MytID</b>	97	34	8	3.9	10.4	32
<b>MytM</b>	152	54	12	6.4	8.4	26
<b>MGD1</b>	82	39	8	4.4	9.4	36
<b>MGDt</b>	61	38	8	4.4	9.3	45

MGDt lacks of C-terminal extension. Table data are calculated from Mytibase ESTs. See also AF162334.1, AF162335.1, AF162336.1, AF162337.1, EU810204.1 and EU927448.1 at NCBI.

doi:10.1371/journal.pone.0026680.t001

Overall, massively parallel sequencing produced 359,867 output reads and more than 73 Mbases with good quality scores, for a total of 304,621 trimmed reads with 226 bp average length. We discarded about 15% of reads per sample (88% of them shorter than 70 bp) which possibly originated in the PCR amplification or sequencing reaction (short sequences are not expected to bias the amplicon coverage nor the accuracy of SNCs detection).

Total or partial overlapping of the forward and reverse reads allowed the complete coverage of the 13 reference sequences. Hence, 97.5% of the good quality reads correctly mapped against the 9 selected AMPs and could be attributed to the 6 original samples. With the exception of MytM in the sample Ve ft, at least 1034 reads mapped on each AMP precursor transcript (range 1034–10814, File S3). Figure 1 shows relevant differences in the average base coverage calculated per AMP precursor transcript in each sample (3378 $\times$ , the average read depth per AMP).

Separately for each AMP and sample, we subsequently grouped the reads having the same length and 100% identity in two cluster types: (1) equal to the original transcript and (2) with at least 1 SNC. The last ones were employed for SNC detection and related analysis. Only the SNCs covered 30 $\times$ , and representing at least 3% of the reads mapping a given AMP, were considered genuine and counted per AMP precursor (Table 3). Based on the SNC counts, the average value of SNC per base calculated for the coding sequence was 0.18. Considering all AMPs together, as much as 134 SNCs were common to the 6 samples and represented 86% of the genuine cds variation (File S4). The SNC frequency values (cds) indicated MytC and MGD1 as the most polymorphic AMP transcripts, opposite to those of mytilins B

and C. The fraction of non-synonymous changes ranged from null (MytIC) to 25 (MytC) whereas the related  $\omega$  values highlighted the opposite cases of MytIC (null) and MytID (9–15).

All SNC frequency values were used to test statistically the sequence diversity of the 9 selected AMPs (one-way ANOVA,  $\alpha = 0.001$ , followed by Tukey's Honestly Significant Difference test,  $\alpha = 0.05$ ). The null hypothesis (equal SNC frequency between AMPs) was rejected and, according to the Tukey's HSD test we classified the 9 cases from the least changeable mytilins, to the most polymorphic myticins and defensins (Figure 2).

Common and exclusive SNCs are reported in Figure 3 per geographical region: the common changes represent the majority, with 68, 74 and 78% in the Vi, Ve and Pa samples, respectively. Similar percentages of common SNCs were found in mussels farmed offshore or living wild in the industrial canals of the Venice lagoon (Ve versus Ve nc), injected or not with  $10^7$  live *V. splendidus* (Ve versus Ve ft, Ve nc versus Ve nt). Details on the immune stimulation and related host response are reported elsewhere [34,35].

Finally, we analyzed the nucleotidic and aminoacidic substitution patterns of each AMP precursor, separately in the samples from Palavas, Vigo and Venice. The percentage of read clusters differing at least 1 SNC ranged from 56% (MytID) to 100% (MytB and MGDt). The read clusters covered at least 3 $\times$  were virtually translated into amino acids, and redundancy due to silent substitutions was removed. As reported in Table 4 we could estimate a number of AMP transcript variants ranging from 9 (Pa MytM) to nearly 100 (Ve MytB, Vi MGD1, Ve MGD1 and Ve MGDt).

The facts emerging from 454 pyrosequencing of the Mytimycin and Myticin C amplicons are reported here below in more detail as instructive examples.

The two gene sequences publicly available for the MytM of *M. galloprovincialis* (FJ804479.1, FJ804478.1) denote 3 exons and 2 introns, and differ only in the length of intron 2; the short and long version of it occurring simultaneously in single mussels [39]. The amplicon designed in the present work covered 429/456 bp, i.e. 94% of the cds.

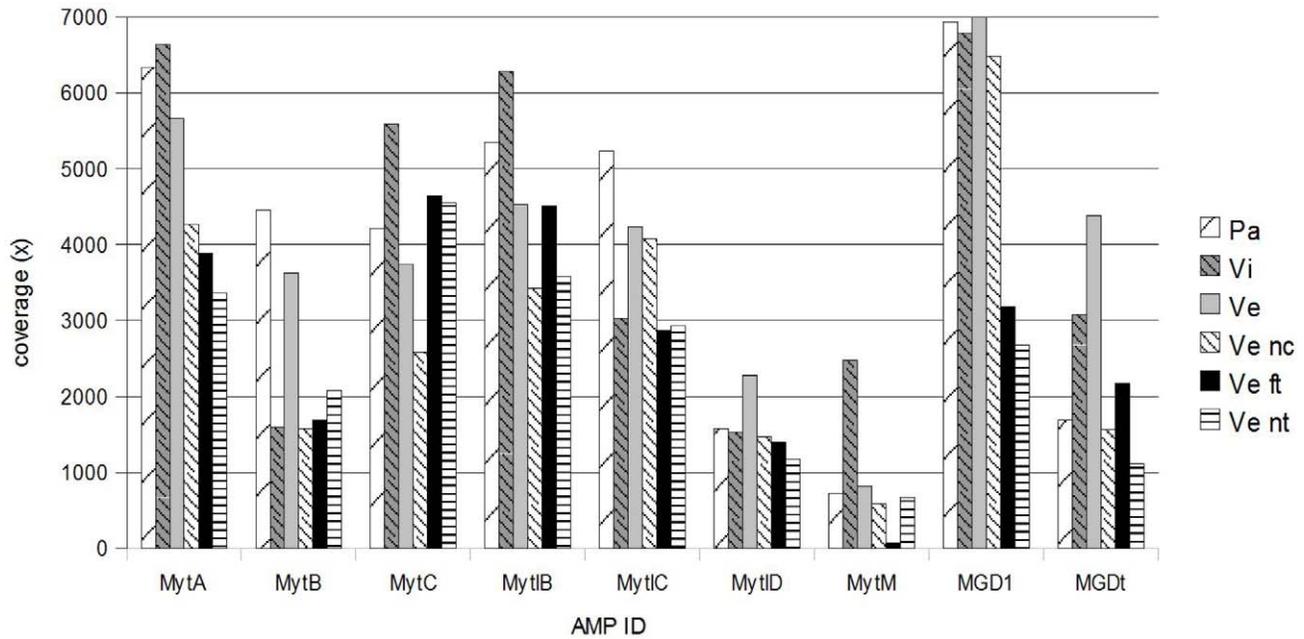
The MytM pyrosequencing yielded 6645 aligned reads (33, 30 and 35% of them differing from the reference sequence in the Pa, Vi and Ve samples, respectively).

A nucleotide change in position 58 (Thymine in the place of Cytosine) was detected in all reads and was not considered as SNC because it could represent an error occurred during Sanger

**Table 2.** Description of the mussel haemolymph samples processed for 454 pyrosequencing.

Sample ID	Origin	No. of mussels	Status	Treatment
Pa	Palavas, F	100	farmed	untreated
Vi	Ria de Vigo, S	100	farmed	untreated
Ve	Offshore Venice, I	100	farmed	untreated
Ve nc	Venice Lagoon, I	100	native	untreated
Ve ft	Offshore Venice, I	40	farmed	<i>Vibrio</i> injection
Ve nt	Venice Lagoon, I	40	native	<i>Vibrio</i> injection

doi:10.1371/journal.pone.0026680.t002



**Figure 1. Average base coverage (x) per AMP and sample.** Coverage is calculated as total sequenced base divided by the length of the amplified transcript.  
doi:10.1371/journal.pone.0026680.g001

sequencing of the original Mytibase singleton (MGC05878, File S4 in Supplementary Materials). The AVA software (Roche Life Sciences) grouped the Pa, Vi and Ve MytM reads in 31, 206 and 32 clusters or singletons (27, 199 and 31 of high quality), respectively. In total, 5690 high quality sequences (86%) were translated in amino acids and produced 9–49 expected peptide variants (Table 4).

Irrespective of the geographical origin, we could consistently identify the two most abundant MytM types: a consensus very similar to the original sequence (MytM\_1, MGC05878) and a second one (MytM\_2) similar to the sequence MytM-P recently described [39]. Jointly, MytM\_1 and MytM\_2 represent the 87, 78 and 76% of the MytM reads in the samples Pa, Vi and Ve, respectively (Vi MytM display higher sequence variability than the other two samples).

MytM\_2 represents 20, 8 and 11% of the MytM reads in samples Pa, Vi and Ve, respectively, and displays 25 SNCs. Eighteen out of 25 changes are non-synonymous (7 in the signal peptide, 8 in the mature peptide and 3 in the C-terminal extension) and confirm the substitutions detected in MytM-P by Sonthi *et al.* using a Sanger approach. Figure 4 locates all genuine SNCs along the original MytM sequence (cds).

Partial gene sequences including the cds have been reported for MytC and denote 3 exons and 2 introns, with the mature peptide entirely located in exon 2 [29,36].

The MytC pyrosequencing yielded 22,119 aligned reads fully covering the cds (71%, 88% and 79% of them differing from the original sequence in the Pa, Vi and Ve samples, respectively). The Pa, Vi and Ve MytC reads could be grouped in 844, 777 and 468 clusters or singletons (823, 752 and 466 of high quality), respectively. In total, 18113 high quality reads (82%) were translated in amino acids and produced 41–93 expected peptide variants (Table 4) without evidence of prevailing variation patterns.

Despite the remarkable number of sequence variants, 98.8% of the MytC peptide clusters retained the typical cysteine array.

Figure 5 locates all genuine SNCs along the original sequence of MytC. Amplicon pyrosequencing confirmed 3/5, 17/27 and 19/33 changes previously detected in the signal mature peptide, and C-terminal regions, respectively [36] and revealed additional changeable positions. As observed in MytM, different SNC combinations increase the total number of possible peptide variants.

The maps locating all synonymous and non-synonymous SNCs along the transcript sequence (cds) of the remaining AMPs are reported in Supplementary Materials (File S5) for comparison.

## Discussion

This study intended to assess, by high-throughput amplicon sequencing, the natural variability of nine AMP precursor sequences found expressed in the Mediterranean mussel. For this purpose, we sampled mussels from farming sites subjected to common European regulations and sanitary controls in three producer countries.

Thirteen sequences almost covering the AMP cds (File S1) have been successfully amplified from hemocyte RNA samples representing mussels farmed in France, Spain and Italy, as well as native mussels of the Venice lagoon area, before and after *Vibrio* injection, for coupled comparison.

Sample preparation is a decisive step of the sequencing workflow, due to the difficulty in preparing a well-balanced unique amplicon pool for the emulsion PCR and subsequent pyrosequencing. For this purpose, we measured and equalized the concentrations of each amplicon with great attention before pooling. Although the two sequencing half plates produced similar read numbers per sample, a general variability of coverage depth between AMP amplicons was finally evident (Table 3, Figure 1). Nevertheless, we obtained at least 500× amplicon coverage with one only exception (MytM in the Ve ft sample).

Stringent criteria were then used to identify genuine SNCs, removing false positives without losing substantial information,

**Table 3.** Number of SNCs detected in the whole AMP transcript precursor and cds, related frequencies, non-synonymous changes and  $\omega$  values.

AMP precursor	Sample	SNCs (total)	SNCs (cds)	Frequency (SNC/base)	Freq/AMP (Mean $\pm$ SD)	ns SNCs*	$\omega$ **
MytA	Pa	38	25	0.26	0.25 $\pm$ 0.04	10	0.7
	Vi	36	20	0.21		8	0.7
	Ve	41	28	0.29		12	0.8
	Ve nc	38	23	0.24		10	0.8
	Ve ft	30	19	0.20		10	1.1
	Ve nt	39	26	0.27		12	0.9
MytB	Pa	21	17	0.18	0.20 $\pm$ 0.05	12	2.4
	Vi	40	29	0.30		15	1.1
	Ve	22	17	0.18		11	1.8
	Ve nc	30	19	0.20		11	1.4
	Ve ft	27	18	0.19		12	2.0
	Ve nt	19	14	0.14		7	1.0
MytC	Pa	33	27	0.27	0.26 $\pm$ 0.04	19	2.4
	Vi	41	33	0.33		25	3.1
	Ve	31	23	0.23		17	2.8
	Ve nc	34	26	0.26		17	1.9
	Ve ft	32	24	0.24		19	3.8
	Ve nt	27	22	0.22		15	2.1
MytB	Pa	11	5	0.05	0.09 $\pm$ 0.03	1	0.3
	Vi	17	10	0.10		5	1.0
	Ve	16	9	0.09		3	0.5
	Ve nc	23	13	0.13		9	2.3
	Ve ft	15	10	0.10		5	1.0
	Ve nt	16	9	0.09		4	0.8
MytC	Pa	8	1	0.01	0.01 $\pm$ 0.01	0	0.0
	Vi	8	1	0.01		0	0.0
	Ve	11	2	0.02		0	0.0
	Ve nc	7	1	0.01		0	0.0
	Ve ft	6	1	0.01		0	0.0
	Ve nt	8	2	0.02		0	0.0
MytD	Pa	23	16	0.17	0.15 $\pm$ 0.04	15	15.0
	Vi	25	20	0.22		18	9.0
	Ve	21	12	0.13		11	11.0
	Ve nc	19	11	0.12		11	[11]
	Ve ft	18	13	0.14		13	[13]
	Ve nt	19	10	0.11		9	9.0
MytM	Pa	28	28	0.20	0.21 $\pm$ 0.01	19	2.1
	Vi	29	29	0.21		21	2.6
	Ve	31	31	0.22		23	2.9
	Ve nc	30	30	0.21		22	2.8
	Ve ft	29	29	0.21		21	2.6
	Ve nt	28	28	0.2		20	2.5
MGD1	Pa	28	21	0.26	0.28 $\pm$ 0.02	14	2.0
	Vi	29	23	0.28		15	1.9
	Ve	28	21	0.26		14	2.0
	Ve nc	31	24	0.30		15	1.7
	Ve ft	28	22	0.27		15	2.1
	Ve nt	30	23	0.28		15	1.9

**Table 3.** Cont.

AMP precursor	Sample	SNCs (total)	SNCs (cds)	Frequency (SNC/base)	Freq/AMP (Mean $\pm$ SD)	ns SNCs*	$\omega$ **
MGDt	Pa	43	12	0.19	0.18 $\pm$ 0.02	11	11.0
	Vi	39	13	0.21		10	3.3
	Ve	40	12	0.19		10	5.0
	Ve nc	41	10	0.16		9	9.0
	Ve ft	35	12	0.19		10	5.0
	Ve nt	36	10	0.16		9	9.0

Frequency values, ns SNCs and  $\omega$  values refer to the cds; \* ns SNCs, non-synonymous changes;

\*\* $\omega$ , ratio between non-synonymous and synonymous SNCs (in brackets, cases without synonymous substitutions).

doi:10.1371/journal.pone.0026680.t003

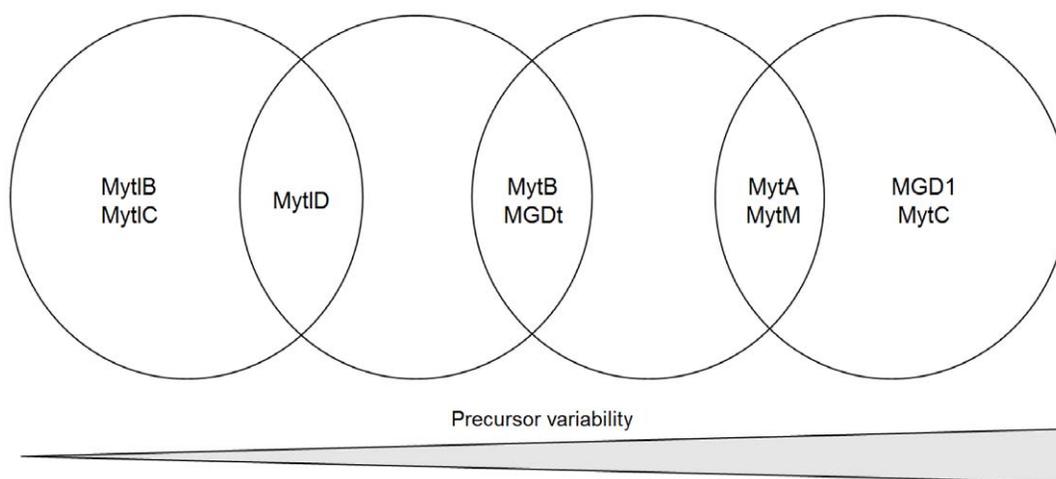
and we could retain most of the output reads (82% and 86% to exemplify Myticin C and Mytimycin, respectively). At first look, the output reads of each AMP appeared very diverse, 88% with at least one SNC on average (100% for MGDt, no matter from which sample). The analysis of SNC frequency per base enabled us to rank the selected AMPs on the basis of the transcript variability (Figure 4). Despite the invariance of the cysteine array, each AMP showed typical levels of diversity irrespective of the geographical origin, with a majority of common SNCs present in all samples (86%, i.e. 134 SNCs) and the Vi sample showing the greatest number of SNCs (292 in total). Moreover, we did not see evidence of increased AMP sequence diversity in farmed and native mussels injected with a high dose of live *Vibrio* cells. Compared to the samples prepared from 100 mussels (Pa, Vi, Ve, Ve nc), those prepared from 40 mussels (Ve ft, Ve nt) showed fewer SNCs (-11%); a fact indicating that the sample size can limit the amount of detectable sequence variants.

The ratio of non-synonymous vs. synonymous changes ( $\omega$ ) substantiates the evolutionary diversification of the mussel AMP isotypes and suggests the functional advantage of transcript variability for most of the analyzed AMPs ( $\omega$  values higher than 1, indicative of positively selected residues, were frequently detected). However, the  $\omega$  values did not reflect precisely the classification based on SNC frequencies and some AMP (e.g. MytD) may have been subjected to higher evolutionary pressure than others.

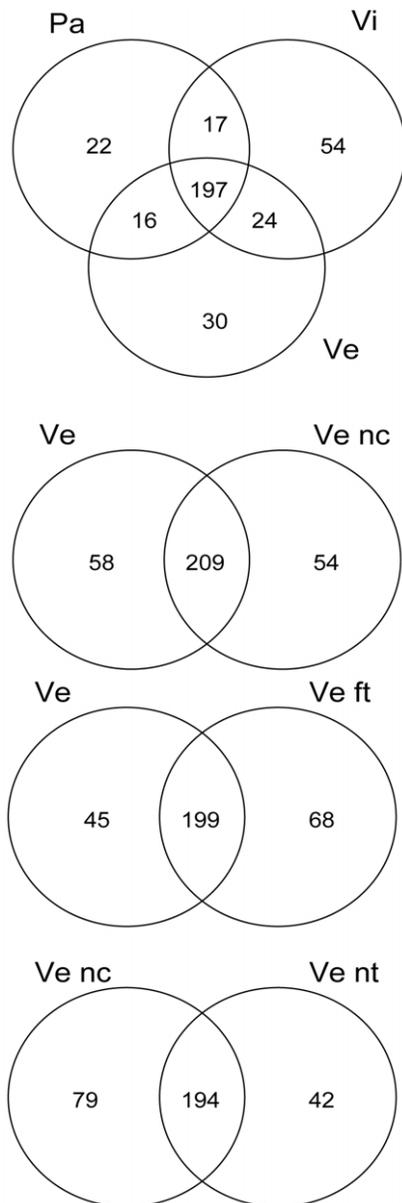
The virtual translation of the transcript consensus resulting from read clustering allowed the identification of AMP isotypes with a relatively low number of SNCs and high sequence diversity according to different SNC combinations (e.g. MytB and MGDt). In the case of mytimycin, we confirmed two major sequence types previously described [39] with no evidence of additional variants. For the remaining AMPs, it was not possible to identify specific patterns of variation (amino acid changes combined together without scheme).

In conclusion, the sequence data reported in this study further emphasize the sequence diversity of mussel AMP precursors. Redundant expression of diverse AMPs with a broad range of action could be regarded as a strategy to reinforce the host response against invaders (foes trying in their turn to escape detection and the host reactions) while the immune system also has to maintain the organism homeostasis with appropriate responses to commensal microbes (friends) and to danger signals released by damaged host cells [54]. On the other hand, environmental factors act as selective force only if they change the distribution of host genotypes (affecting only some genotypes, not all), thus influencing the immune system evolution of the host in the context of its life-history and population traits [55].

The isotype diversity levels found in this study might result from events occurring at DNA level as well as post-transcriptional changes such as deaminase-mediated cytidine to uridine transitions [14,56–57]. Hence, targeted sequence enrichment and



**Figure 2. AMP precursors grouped on the basis of their variability.** Grouping based on HSD test,  $\alpha=0.05$ ; sample-specific and averaged diversity levels are reported in Table 3.  
doi:10.1371/journal.pone.0026680.g002



**Figure 3. Common and exclusive SNCs in the Pa, Vi and Ve samples and paired samples from native and immunostimulated mussels.**

doi:10.1371/journal.pone.0026680.g003

extension strategies applied to genomic DNA could identify active and remnant gene copies of each AMP isotype, and reveal the mechanisms underlying the observed sequence variation.

## Materials and Methods

### Sampling sites, treatment and RNA extraction

Adult mussels (*Mytilus galloprovincialis*) with a shell length of 6–8 cm and mixed sex were obtained from commercial shellfish stocks near Palavas (Pa, Mediterranean Sea, France; 43°31'49 N, 03°54'53 E), Ria de Vigo (Vi, Atlantic Ocean, Spain; 42°14'32 N, 08°48'26 E) and off-shore Venice (Ve, North Adriatic Sea, Italy; 45°18'29.8 N, 12°21'32.0 E). In addition, we collected wild mussels from the industrial canals of Porto Marghera (Lagoon of

Venice, Italy; 45°27'33.5 N, 12°15'41 E). More than 100 animals per group were sampled.

According to the EU Directive 91/492, mussels cultivated in waters classified A (e.g. ¼ mile off-shore in the Adriatic Sea) can be marketed without depuration and are assumed not to contain potential pathogens nor biotoxins. Due to heavy mixed pollution, shellfishing was prohibited since 1996 in the area from the industrial district (P. Marghera) to the Venice town, though the overall shellfish quality can be improved by 2 month-depuration in type A waters. Mussels farmed offshore or living in the industrial canals (Venice lagoon area) were acclimatized for one week in sea water collected at flood tide (32‰, 22°C) and fed with *Isochrysis galbana*. Following shell notching, 0.1 ml of exponentially growing bacteria ( $10^7$  *V. splendidus* LGP32 cells) were injected into the posterior adductor muscle (samples Ve ft and Ve nt).

Hemolymph (1 ml per animal) was withdrawn from the posterior adductor muscle with a syringe containing 0.2 ml of Alsever solution (27 mM sodium citrate, 2.6 mM citric acid, 114 mM glucose and 72 mM NaCl in distilled water) adjusted at pH 7.4, and used to compose pools, each representative of 10 animals.

Haemocytes were pelleted by 15 min centrifugation at 800 xg (4°C), carefully resuspended in 200 µl of TRIZOL reagent (Invitrogen, Carlsbad, USA) and stored at –80°C until use. Total RNA was isolated according to the manufacturer's instructions and resuspended in RNase-free water. A further purification step with LiCl 2 M was applied to remove possible contaminants. RNA concentration was measured by UV-spectrometry (ND1000, NanoDrop Technologies, Wilmington, USA) and the RNA integrity was verified by microcapillary electrophoresis (RNA 6000 Nano LabChip, Agilent Technologies, Palo Alto, USA).

Finally, equal quantities of each RNA pool (N = 10) were mixed together to compose a unique pool per sample (N = 100 mussels for samples Pa, Vi, Ve and Ve nc; N = 40 for samples Ve ft and Ve nt).

### cDNA

cDNA was synthesized from 1 µg of total RNA using SuperScript II enzyme and oligo(dT)<sub>18</sub> primers (Invitrogen), following the manufacturer's instruction. To increase cDNA yield, the reaction was extended for a second hour, adding 0.5 µl of enzyme. cDNA was then purified with MinElute PCR Purification Kit (Qiagen, Hilden, Germany).

### Primer design

Primer design was performed on the raw EST sequences denoting 9 mussel AMP isoforms in Mytibase (<http://mussel.cribi.unipd.it>).

All available ESTs for each selected AMP precursor were aligned using ClustalW [58] and primers were designed on conserved regions flanking the cds, whenever possible, with Primer3 [59]. Due to the pyrosequencing limit of about 250 bp, read length in forward and reverse direction, the maximum length of the PCR products was set at 440 bp (File S1). Degenerated primers have been designed to consider the whole ESTs variability (MytC, MGD1 and MGDt). Two partially overlapping amplicons were designed for longer cds (MytC, MytB, and MytC) or in cases where high sequence variability made the definition of a single primer pair difficult (MytA). Thereby, 13 amplicons were designed in total. Amplicon's specificity was tested using the BlastX algorithm [60].

A tagged sequencing strategy with 5' nucleotide barcodes was implemented to facilitate the parallel processing of multiple samples [61–62]. Briefly, the forward and reverse PCR primers

**Table 4.** AMP transcript variants and expected number of non-redundant peptides per geographical region.

AMP ID	Sample	Total read clusters*	Fraction with at least 1 SNC (%)	Non-redundant peptides**
<b>MytA</b>	Pa	300	83	50
	Vi	684	90	68
	Ve	486	86	51
<b>MytB</b>	Pa	763	100	95
	Vi	280	100	63
	Ve	654	100	106
<b>MytC</b>	Pa	844	97	41
	Vi	777	100	93
	Ve	484	99	59
<b>MytIB</b>	Pa	206	70	18
	Vi	338	90	40
	Ve	190	88	28
<b>MytIC</b>	Pa	240	68	37
	Vi	138	83	23
	Ve	177	79	32
<b>MytID</b>	Pa	61	64	17
	Vi	59	69	17
	Ve	61	56	14
<b>MytM</b>	Pa	31	94	9
	Vi	206	97	49
	Ve	32	94	12
<b>MGD1</b>	Pa	410	99	85
	Vi	557	99	100
	Ve	686	89	101
<b>MGDt</b>	Pa	204	100	45
	Vi	729	100	84
	Ve	793	100	103

\*Total cluster number refers to reads differing in length and/or sequence.

\*\*Virtual number of peptides differing at least for one residue.

doi:10.1371/journal.pone.0026680.t004

were modified by 5'-addition of 39 unique 5-mer barcodes (File S2). Barcodes enable the identification of the 454 reads corresponding to specific AMP, amplicon and sample so that PCR amplicons derived from multiple reactions could be combined for the sequencing run. To reduce the likelihood of misidentification, barcodes were designed not to contain homopolymers and to differ each one by at least two bases according to Roche Life Science protocols. Finally, 19-mer sequences corresponding to either the 454 Roche A Adaptor (for forward primers) or B Adaptor (for reverse primers) were fused to each PCR primer (Fusion primer).

Thermodynamic properties of Fusion primers were controlled to avoid the formation of hetero- or homo-dymer (OligoAnalyzer 3.1, <http://eu.idtdna.com>). Melting temperatures were fixed according to Primer3 software [63], adding 5 ( $\pm 1$ ) °C following HF Phusion polymerase instructions (Finnzymes, Espoo, Finland).

Fusion primers were designed in two sets of 39 (13 amplicons  $\times$  3 samples), with each primer pair having a unique barcode.

### PCR amplification

The PCR amplifications of 78 amplicons were carried out individually in a PCR volume of 50  $\mu$ l with 20 ng of cDNA template, 1  $\times$  Phusion HF buffer, 0.2 mM dNTPs, 1 U HF

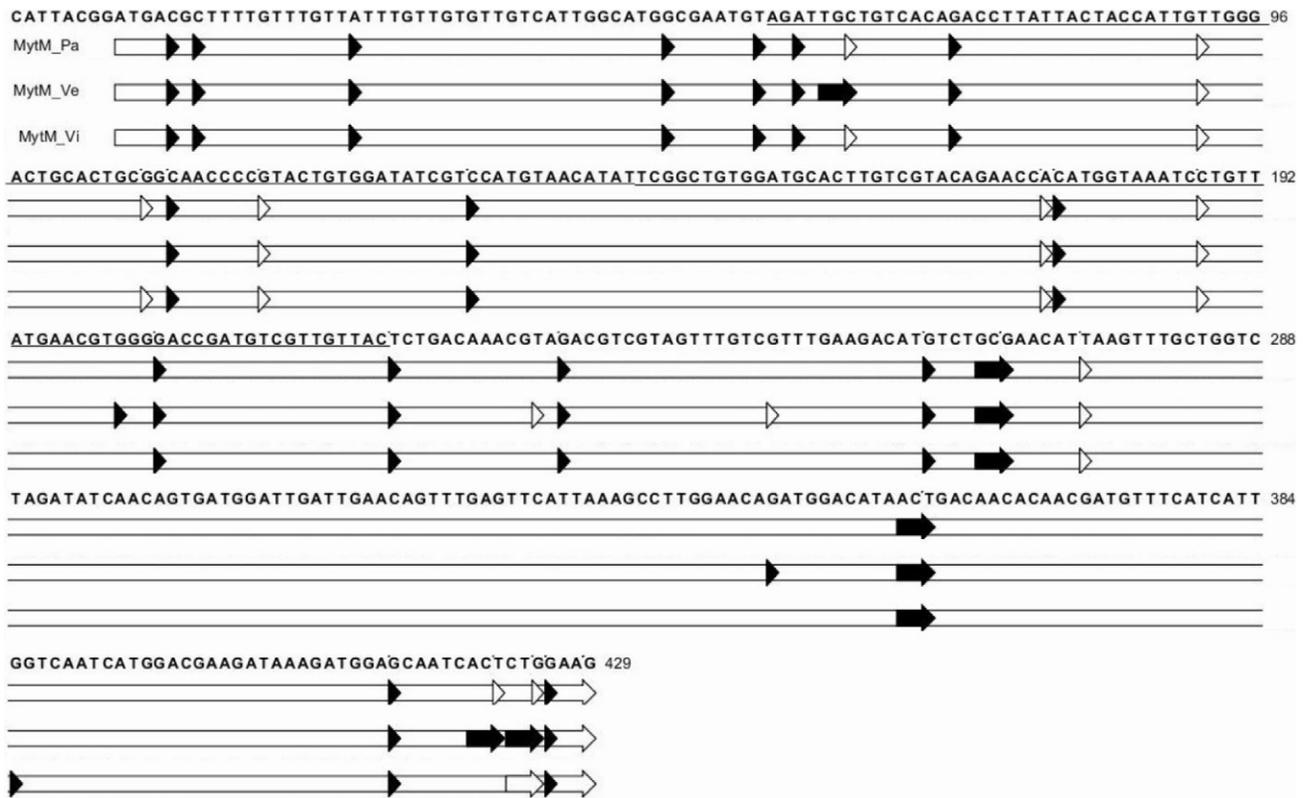
Phusion DNA polymerase, 1.5  $\mu$ l DMSO and 0.2  $\mu$ M of both forward and reverse primers. Amplification was performed in a Mastercycler Gradient Thermal Cycler (Eppendorf, Hamburg, Germany) programmed as follows: 98°C for 30 s followed by 35 cycles of 98°C for 10 s, 60–65°C for 20 s, 72°C for 30 s and a final extension step at 72°C for 5 min.

The resulting amplification products were run on a 2% agarose gel and visualized by SYBR Gold staining (Invitrogen) using UV light transillumination (Gel Doc XR System, Bio-Rad, Hercules, USA). Unspecific small products and primer-dimers were removed using the Agencourt AMPure system (AmPure PCR Purification kit, Brea, USA) and amplicons integrity was confirmed with Agilent 2100 Bioanalyzer (DNA-1000 chip). Good quality amplicons were finally used to compose an equimolecular pool; the number of molecules of each amplicon was calculated with the following formula:

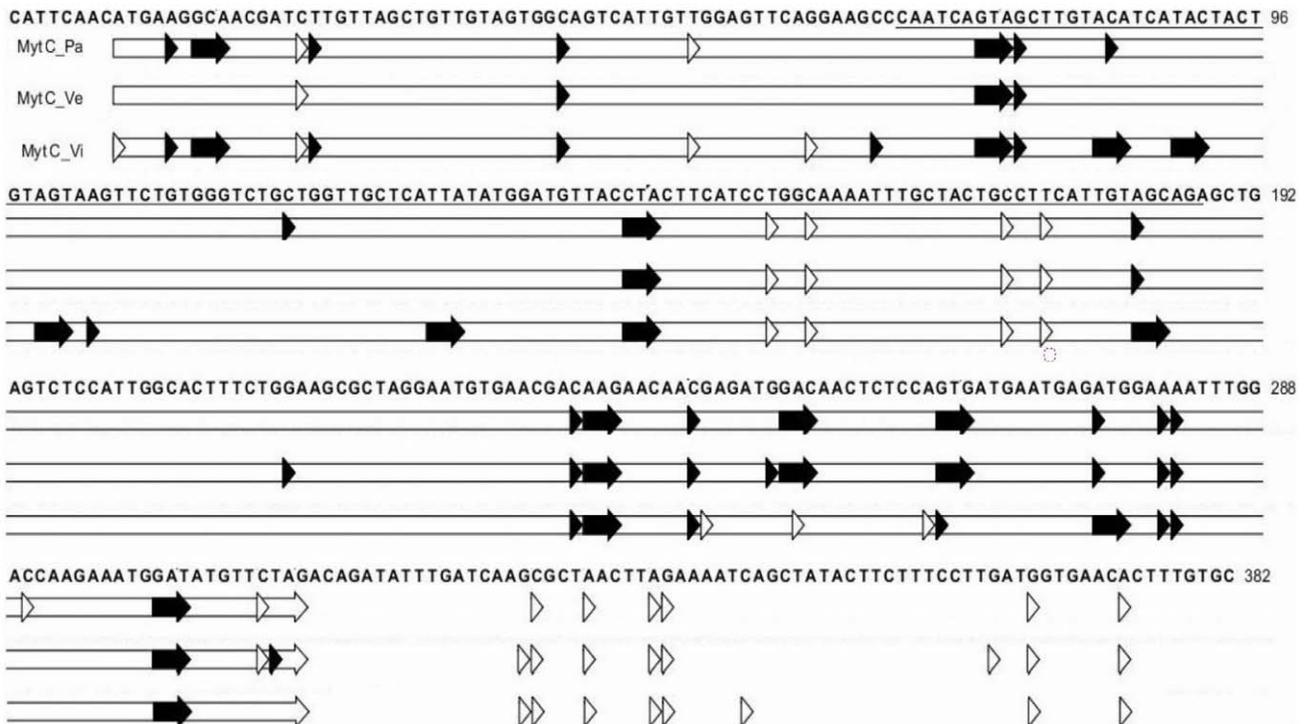
$$\text{Molecules}/\mu\text{l} = C \times N_A / (\text{bp}_w \times 10^9 \times \text{bp})$$

C: sample concentration (ng/ $\mu$ l)

$N_A$ : Avogadro constant



**Figure 4. SNCs mapped along the MytM sequence in the Pa, Ve and Vi samples.** Synonymous (empty) and non-synonymous (filled) changes are located on the AMP precursor sequence. Empty horizontal bars indicates cds, mature peptide sequence is underlined.  
doi:10.1371/journal.pone.0026680.g004



**Figure 5. SNCs mapped along the MytC sequence in the Pa, Ve and Vi samples.** Synonymous (empty) and non-synonymous (filled) changes are located on the AMP precursor sequence. Empty horizontal bars indicates cds, mature peptide sequence is underlined.  
doi:10.1371/journal.pone.0026680.g005

bp<sub>w</sub>: average pair basis weight (g)  
bp: pair basis number

Massively parallel 454 pyrosequencing (FLX-System, Roche Life Sciences, Branford, USA) was performed by BMR-Genomics ([www.bmr-genomics.it](http://www.bmr-genomics.it)) using two PicoTiter half plates. Reads have been recorded at the Sequence Read Archive accessible at <http://www.ncbi.nlm.nih.gov/Traces/sra> (submission ID: SRA038518.3). Six 454-output files corresponding to the 6 sequenced samples are available (SRR286638.1, SRR286639.1, SRR286640.1, SRR286641.1, SRR287657.1 and SRR287658.1).

## Data analysis

Tag sequence were used as keys to part the unprocessed 454 reads into the 6 different samples by means of the GS Amplicon Variant Analyzer Software (AVAST, Roche Life Sciences). Reads in raw format were trimmed using quality score (limit 0.05) and minimum length equal to 100 bp. Subsequently, the output reads were aligned to a backbone consisting of the 9 AMP original transcripts, as obtained from the 13 reference sequences, with CLC Genomic Workbench version 4.6 (CLC Bio, Katrinbjerg, Denmark). The total number of sequenced bases divided by the length of the amplified transcript provided the average base coverage per AMP.

The reads of each mapping (AMP isotype) showing the same length and 100% similarity were clustered together. Single nucleotide changes (SNCs) were detected considering all the aligned reads of each mapping. Non-specific and low quality matches are ignored during the process and SNCs were considered genuine only when covered at least 30×, with a minimum frequency of 3%, and setting the quality level of the changed base and surrounding bases to at least 20 and 15, respectively. SNCs located in the same codon were merged. The expected amino acid changes in the precursor and cds peptide sequences were deduced by virtual translation, and the ratio (ω) between non-synonymous and synonymous changes was also computed. SNC frequency per base in the cds region was calculated for each AMP and sample with the following formula:

$$\text{SNC frequency} = (n \text{ SNC}_{(\text{AMP})}) / (\text{nt}_{\text{seq}})$$

nt<sub>seq</sub>: cds sequenced nucleotides

n SNC<sub>(AMP)</sub>: number of genuine SNC per amplicon in each sample

To assess possible differences in levels of sequence diversity between the AMP precursors amplified from each of the 6 samples, data were analysed with 1-way ANOVA ( $\alpha = 0.001$ ). The

null hypothesis predicted that all AMPs had the same variation rate in all samples. If not, Tukey's Honestly Significant Difference test (HSD,  $\alpha = 0.05$ ) could then discriminate different AMP groups. Genuine SNCs were mapped on the 9 AMP sequences using CLC Genomic Workbench. The cds, signal peptide, mature peptide with the cysteine array and C-terminal regions were systematically localized.

Once the correct sequence reading frame was defined the sequences covered at least 3× were virtually translated to investigate the overall patterns of variation for each AMP transcript precursor in the Pa, Vi and Ve samples. Leftover low quality read ends were manually trimmed and redundancy was removed by using Jalview [64].

## Supporting Information

**File S1 Number of reads per AMP and sample.**  
(XLS)

**File S2 Detailed list of all SNCs detected per AMP precursor in the 6 samples, with type of change and related frequency.**  
(XLS)

**File S3 The maps locating all synonymous and non-synonymous SNCs along the transcript sequence (cds) of the AMPs.**  
(TIF)

**File S4 The 13 reference sequences designed to cover 9 AMP precursor transcripts of *M. galloprovincialis*.** Sequence number and expected amplicon length with related transcript ID, precursor length, and number of clustered ESTs. All reference sequences are also reported in detail.  
(DOC)

**File S5 Fusion primer pairs used for the AMP amplification, primer length, melting temperature and GC percentage.**  
(XLS)

## Acknowledgments

A special thanks to Francesca Dal Pero for helpful advice.

## Author Contributions

Conceived and designed the experiments: PV AP. Performed the experiments: UR. Analyzed the data: UR. Contributed reagents/materials/analysis tools: PV. Wrote the paper: UR PV AR. Evaluated the AMP ESTs, cured primer design and collected Ve hemolymph samples: LV. Provided Pa and Vi samples: PR BN. Critically contributed to the revision of the work: AF BN AP PR PV.

## References

- Gérard K, Bierre N, Borsa P, Chenuil A, Féral JP (2008) Pleistocene separation of mitochondrial lineages of *Mytilus* spp. mussels from Northern and Southern Hemispheres and strong genetic differentiation among southern populations. *Mol Phylogenet Evol* 49(1): 84–91.
- Gilg MR, Hilbish TJ (2003) Patterns of larval dispersal and their effect on the maintenance of a blue mussel hybrid zone in southwestern England. *Evolution* 57(5): 1061–77.
- Galler S (2008) Molecular basis of the catch state in molluscan smooth muscles: a catchy challenge. *J Muscle Res Cell Motil* 29(2–5): 73–99.
- Babarro JM, De Zwaan A (2008) Anaerobic survival potential of four bivalves from different habitats. A comparative survey. *Comp Biochem Physiol A Mol Integr Physiol* 151(1): 108–13.
- Bers AV, D'Souza F, Klijnstra JW, Willemsen PR, Wahl M (2006) Chemical defence in mussels: antifouling effect of crude extracts of the periostracum of the blue mussel *Mytilus edulis*. *Biofouling* 22(3–4): 251–9.
- Akaishi FM, St-Jean SD, Bishay F, Clarke J, da S Rabitto I, et al. (2007) Immunological responses, histopathological finding and disease resistance of blue mussel (*Mytilus edulis*) exposed to treated and untreated municipal wastewater. *Aquat Toxicol* 82(1): 1–14.
- Moore MN (2008) Autophagy as a second level protective process in conferring resistance to environmentally-induced oxidative stress. *Autophagy* 4(2): 254–6.
- Rayyan A, Damianidis P, Antoniadou C, Chintiroglou CC (2006) Protozoan parasites in cultured mussels *Mytilus galloprovincialis* in the Thermaikos Gulf (north Aegean Sea, Greece). *Dis Aquat Organ* 70(3): 251–4.
- Francisco CJ, Hermida MA, Santos MJ (2010) Parasites and symbionts from *Mytilus galloprovincialis* (Lamarck, 1819) (Bivalves: Mytilidae) of the Aveiro Estuary Portugal. *J Parasitol* 96(1): 200–5.
- Gomez DK, Baeck GW, Kim JH, Choresca CH, Jr., Park SC (2008) Molecular detection of betanodaviruses from apparently healthy wild marine invertebrates. *J Invertebr Pathol* 97(3): 197–202.

11. Takahashi D, Shukla SK, Prakash O, Zhang G (2010) Structural determinants of host defense peptides for antimicrobial activity and target cell selectivity. *Biochimie* 92(9): 1236–41.
12. Bucki R, Leszczynska K, Namiot A, Sokolowski W (2010) Cathelicidin LL-37: a multitask antimicrobial peptide. *Arch Immunol Ther Exp (Warsz)* 58(1): 15–25.
13. Tian C, Gao B, Fang Q, Ye G, Zhu S (2010) Antimicrobial peptide-like genes in *Nasomia vitripennis*: a genomic perspective. *BMC Genomics* 11: 187.
14. Yeaman MR, Yount NY (2007) Unifying themes in host defence effector polypeptides. *Nat Rev Microbiol* 5(9): 727–40.
15. Pujol N, Zugasti O, Wong D, Couillaud C, Kurz CL, et al. (2008) Anti-fungal innate immunity in *C. elegans* is enhanced by evolutionary diversification of antimicrobial peptides. *PLoS Pathog* 4(7): e1000105.
16. Smith LC (2010) Diversification of innate immune genes: lessons from the purple sea urchin. *Disease Models & Mechanisms* 3: 1–6.
17. Buckley KM, Terwilliger DP, Smith LC (2008) Sequence variations in 185/333 messages from the purple sea urchin suggest posttranscriptional modifications to increase immune diversity. *J Immunol* 181(12): 8585–94.
18. Hamilton CE, Papavasiliou FN, Rosenberg BR (2010) Diverse functions for DNA and RNA editing in the immune system. *RNA Biol* 7(2): 220–8.
19. Linzmeier RM, Ganz T (2005) Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23. *Genomics* 86(4): 423–30.
20. Chen Q, Hakimi M, Wu S, Jin Y, Cheng B, et al. (2010) Increased genomic copy number of DEFA1/DEFA3 is associated with susceptibility to severe sepsis in Chinese Han population. *Anesthesiology* 112(6): 1428–34.
21. Sperstad SV, Haug T, Blencke HM, Styrvold OB, Li C, Stensvåg K (2011) Antimicrobial peptides from marine invertebrates: Challenges and perspectives in marine antimicrobial peptide discovery. *Biotechnol Adv*; in press.
22. Charlet M, Chernysh S, Philippe H, Hétru C, Hoffmann J, 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.
23. Hubert F, Noël T, Roch P (1996) A member of the arthropod defensin family from edible Mediterranean mussels (*Mytilus galloprovincialis*). *Eur J Biochem* 15;240(1): 302–6.
24. Mitter G, Vandenbulcke F, Roch P (2000) Original involvement of antimicrobial peptides in mussel innate immunity. *FEBS Lett* 486(3): 185–90.
25. Venier P, De Pittà C, Bernante F, Varotto L, De Nardi B, et al. (2009) MytiBase: a knowledge base of mussel (*M. galloprovincialis*) transcribed sequences. *BMC Genomics* 10: 72.
26. Mitter G, Hubert F, Noël T, Roch P (1999) Myticin, a novel cysteine-rich antimicrobial peptide isolated from haemocytes and plasma of the mussel *Mytilus galloprovincialis*. *Eur J Biochem* 265(1): 71–8.
27. Mitter G, Vandenbulcke F, Hubert F, Salzet M, Roch P (2000) Involvement of mytilins in mussel antimicrobial defense. *J Biol Chem* 275(17): 12954–62.
28. Mitter G, Vandenbulcke F, Noël T, Romestand B, Beauvillain JC, et al. (2000) Differential distribution and defence involvement of antimicrobial peptides in mussel. *J Cell Sci* 113(Pt 15): 2759–69.
29. Costa MM, Dios S, Alonso-Gutierrez J, Romero A, Novoa B, et al. (2009) Evidence of high individual diversity on myticin C in mussel (*Mytilus galloprovincialis*). *Dev Comp Immunol* 33(2): 162–70.
30. Berkowitz BA, Bevins CL, Zasloff MA (1990) Magainins: a new family of membrane-active host defense peptides. *Biochem Pharmacol* 39(4): 625–9.
31. Li P, Chan HC, He B, So SC, Chung YW, et al. (2001) An antimicrobial peptide gene found in the male reproductive system of rats. *Science* 291(5509): 1783–5.
32. Cellura C, Toubiana M, Parrinello N, Roch P (2007) Specific expression of antimicrobial peptide and HSP70 genes in response to heat-shock and several bacterial challenges in mussels. *Fish Shellfish Immunol* 22(4): 340–50.
33. Li H, Toubiana M, Monfort P, Roch P (2009) Influence of temperature, salinity and *E. coli* tissue content on immune gene expression in mussel: results from a 2005–2008 survey. *Dev Comp Immunol* 33(9): 974–9.
34. Li H, Venier P, Prado-Alvarez M, Gestal C, Toubiana M, et al. (2010) Expression of *Mytilus* immune genes in response to experimental challenges varied according to the site of collection. *Fish Shellfish Immunol* 28(4): 640–8.
35. Venier P, Varotto L, Rosani U, Millino C, Celegato B, et al. (2011) Insights into the innate immunity of the Mediterranean mussel *Mytilus galloprovincialis*. *BMC Genomics* 12: 69.
36. Pallavicini A, Costa MM, Gestal C, Dreos R, Figueras A, et al. (2008) High sequence variability of myticin transcripts in hemocytes of immune-stimulated mussels suggests ancient host-pathogen interactions. *Dev Comp Immunol* 32(3): 213–26.
37. Parisi MG, Li H, Toubiana M, Parrinello N, Cammarata M, et al. (2009) Polymorphism of mytilin B mRNA is not translated into mature peptide. *Mol Immunol* 46(3): 384–92.
38. Mitter G, Hubert F, Dyrzynda EA, Boudry P, Roch P (2000) Mytilin B and MGD2, two antimicrobial peptides of marine mussels: gene structure and expression analysis. *Dev Comp Immunol* 24(4): 381–93.
39. Sonthi M, Toubiana M, Pallavicini A, Venier P, Roch P (2011) Diversity of Coding Sequences and Gene Structures of the Antifungal Peptide Mytimycin (MytM) from the Mediterranean Mussel, *Mytilus galloprovincialis*. *Mar Biotechnol* (NY) 13: 857–867.
40. Yang YS, Mitter G, Chavanieu A, Calas B, Sanchez JF, et al. (2000) Solution structure and activity of the synthetic four-disulfide bond Mediterranean mussel defensin (MGD-1). *Biochemistry* 39(47): 14436–47.
41. Roch P, Yang Y, Toubiana M, Aumelas A (2008) NMR structure of mussel mytilin, and antiviral-antibacterial activities of derived synthetic peptides. *Dev Comp Immunol* 32(3): 227–38.
42. Pantzartz C, Drosopoulou E, Yiangou M, Drozdov I, Tsoka S, et al. (2010) Promoter complexity and tissue-specific expression of stress response components in *Mytilus galloprovincialis*, a sessile marine invertebrate species. *PLoS Comput Biol* 6(7): e1000847.
43. Aceto S, Formisano G, Carella F, De Vico G, Gaudio L (2011) The metallothionein genes of *Mytilus galloprovincialis*: genomic organization, tissue expression and evolution. *Mar Genomics* 4(1): 61–8.
44. Romero A, Estévez-Calvar N, Dios S, Figueras A, et al. (2011) New insights into the apoptotic process in mollusks: characterization of caspase genes in *Mytilus galloprovincialis*. *PLoS One* 6(2): e17003.
45. Gracey AY, Chaney ML, Boomhower JP, Tyburczy WR, Connor K, et al. (2008) Rhythms of gene expression in a fluctuating intertidal environment. *Curr Biol* 18(19): 1501–7.
46. Banni M, Negri A, Mignone F, Boussetta H, Viarengo A, et al. (2011) Gene expression rhythms in the mussel *Mytilus galloprovincialis* (Lam.) across an annual cycle. *PLoS One* 6(5): e18904.
47. Lockwood BL, Somero GN (2011) Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*). *Mol Ecol* 20(3): 517–29.
48. Manfrin C, Dreos R, Battistella S, Beran A, Gerdol M, et al. (2011) Mediterranean mussel gene expression profile induced by okadaic acid exposure. *Environ Sci Technol* 44(21): 8276–83.
49. Craft JA, Gilbert JA, Temperton B, Dempsey KE, Ashelford K, et al. (2010) Pyrosequencing of *Mytilus galloprovincialis* cDNAs: Tissue-Specific Expression Patterns. *PLoS One* 5(1): e8875.
50. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci* 74: 5463–67.
51. Thomas RK, Nickerson E, Simons JF, Janne PA, Tengs T, et al. (2006) Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat Med* 12(7): 852–5.
52. Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW (2007) Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome Res* 17(8): 1195–201.
53. Benaglio P, Rivolta C (2010) Ultra high throughput sequencing in human DNA variation detection: a comparative study on the NDUFA3-PRPF31 region. *PLoS One* 5(9): e13071.
54. Lazzaro BP, Roff J (2011) Immunology. Danger, microbes, and homeostasis. *Science* 332(6025): 43–4.
55. Schulenburg H, Kurtz J, Moret Y, Siva-Jothy M (2009) Introduction. Ecological immunology. *Philos Trans R Soc Lond B Biol Sci* 364: 3–14.
56. Gao B, Zhu S (2010) Identification and characterization of the parasitic wasp *Nasomia* defensins: positive selection targeting the functional region? *Dev Comp Immunol* 34(6): 659–68.
57. Zhang Y, He X, Li X, Fu D, Chen J, et al. (2011) The second bactericidal permeability increasing protein (BPI) and its revelation of the gene duplication in the Pacific oyster, *Crassostrea gigas*. *Fish Shellfish Immunol* 30(3): 954–63.
58. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 11; 22(22): 4673–4680.
59. Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. Krawetz S, Misener S, eds. *Bioinformatics Methods and Protocols: Methods in Molecular Biology Humana Press, Totowa, NJ*. pp 365–386.
60. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17): 3389–402.
61. Binladen J, Gilbert MT, Bollback JP, Panitz F, Bendixen C, et al. (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2(2): e197.
62. Parameswaran P, Jalili R, Tao L, Shokrala S, Gharizadeh B, et al. (2007) A pyrosequencing tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res* 35: 130.
63. Breslauer KJ, Frank R, Blocker H, Marky LA (1986) Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci* 83(11): 3746–50.
64. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9): 1189–1191.