



**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA E BIOCÊNCIAS**

Cairé Barreto Vieira

**EVOLUÇÃO E DIVERSIDADE MOLECULAR DA FAMÍLIA DE PEPTÍDEOS
ANTIMICROBIANOS CRUSTINAS DO CAMARÃO *Litopenaeus vannamei***

Florianópolis
2021

Cairé Barreto Vieira

**EVOLUÇÃO E DIVERSIDADE MOLECULAR DA FAMÍLIA DE PEPTÍDEOS
ANTIMICROBIANOS CRUSTINAS DO CAMARÃO *Litopenaeus vannamei***

Tese de Doutorado submetida ao Programa de Pós-Graduação em Biotecnologia e Biociências da Universidade Federal de Santa Catarina para obtenção do título de Doutor em Biotecnologia e Biociências.

Orientador: Prof. Dr. Rafael Diego da Rosa

Florianópolis
2021

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Barreto, Cairé

Evolução e diversidade molecular da família de peptídeos antimicrobianos crustinas do camarão *Litopenaeus vannamei* / Cairé Barreto ; orientador, Rafael Diego da Rosa, 2021. 149 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências Biológicas, Programa de Pós Graduação em Biotecnologia e Biociências, Florianópolis, 2021.

Inclui referências.

1. Biotecnologia e Biociências. 2. Crustinas. 3. Peptídeos Antimicrobianos. 4. Sistema Imune. 5. *Litopenaeus vannamei*. I. Rosa, Rafael Diego da. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Biotecnologia e Biociências. III. Título.

Cairé Barreto Vieira

**EVOLUÇÃO E DIVERSIDADE MOLECULAR DA FAMÍLIA DE PEPTÍDEOS
ANTIMICROBIANOS CRUSTINAS DO CAMARÃO *Litopenaeus vannamei***

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Glauber Wagner, Dr.
Universidade Federal de Santa Catarina

Prof. Octávio Luiz Franco, Dr.
Universidade Católica Dom Bosco

Prof. Rafael Dias Mesquita, Dr.
Universidade Federal do Rio de Janeiro

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de doutor em Biotecnologia e Biociências.

Coordenação do Programa de Pós-Graduação

Prof. Rafael Diego da Rosa, Dr.
Orientador

Florianópolis, 2021

Dedico este trabalho à minha amada sobrinha,
Lia da Nóbrega Barreto, que, tão cedo na vida, já
me ensinou o profundo sentimento de amor e
união familiar. Titio te amará para sempre.

AGRADECIMENTOS

Primeiramente, gostaria de agradecer profundamente ao meu orientador Rafael Diego da Rosa pelas inestimáveis contribuições para minha formação acadêmica. Agradeço imensamente a paciência, o carinho e atenção durante todos esses anos. Você foi o grande mediador das mais incríveis experiências que pude obter na vida, desde minha ida à Floripa até minha viagem à França e eu quero deixar minha mais genuína gratidão pela confiança e pela oportunidade de me auxiliar nessa longa e prazerosa caminhada. Guardarei para sempre seus conselhos, pessoais e profissionais, e o rigor científico do qual tanto me espelho. Espero tê-lo como amigo e colaborador no futuro que me espera. Meu muito obrigado, Rafa!

À professora Luciane Maria Perazzolo por todas as contribuições, carinho e companheirismo durante esses anos. Você também fez parte dessa importante história na minha vida e eu deixo aqui meu mais profundo agradecimento.

À Dr. Delphine Destoumieux-Garzón pelas incríveis contribuições como minha orientadora durante meu período de doutorado sanduíche na Université de Montpellier e, também, pelo carinho singular com o qual fui recebido. Foram tempos de muito aprendizado que estarão eternamente guardados na minha memória afetiva.

Aos membros da banca, Dr. Glauber Wagner, Dr. Rafael Mesquita, Dr. Octavio Franco, Dra. Luciane Maria Perazzolo e Dr. Edroaldo Lummertz da Rocha, pela disponibilidade de participar da avaliação deste trabalho e de contribuir para minha formação profissional.

À CAPES por todo o suporte financeiro para a realização deste trabalho.

À minha família por todo o apoio, carinho e amor durante esses duros momentos e por compartilharem sempre minhas vitórias e momentos de felicidade. Vocês representam a base mais sólida que suporta minha existência.

Aos meus irmãos e irmãs de bancada Gabriel, Natan, Mariana, Nicolas, Leo, Gustavo, Flavinho, Talita, Breno pela maravilhosa presença diária que tornou esse período tão especial.

Aos meus amigos e amigas Maiara, Fernando, Anderson, Renan, Laura, Thiago, Chicão e tantos outros que participaram da minha vida diariamente durante todo esse tempo. Guardarei com muito carinho os maravilhosos dias que compartilhamos juntos!

Aos professores e professoras do PPGBCT pela partilha do conhecimento científico e pela contribuição na minha formação intelectual.

A todos os funcionários e funcionárias da universidade que promoveram a manutenção de um ambiente propício ao trabalho, em especial, aos servidores do Restaurante Universitário pela inestimável importância no meu cotidiano.

Meus profundos agradecimentos a todos e todas que participaram direta ou indiretamente da execução deste trabalho!

RESUMO

As crustinas representam a maior e mais diversa família de peptídeos antimicrobianos (AMPs) descrita em crustáceos. Elas são classicamente definidas como peptídeos/polipeptídeos ricos em resíduos de cisteína contendo um típico domínio WAP (*Whey Acidic Protein*) na região C-terminal e são classificadas em cinco principais grupos, referidos como Tipos I a V. As crustinas são diversas em termos de estrutura primária, tamanho e características bioquímicas, exibindo, portanto, múltiplas funções biológicas que vão além de suas propriedades antimicrobianas. O presente estudo buscou dimensionar a diversidade molecular, transcricional e genética de crustinas de *Litopenaeus vannamei*, espécie de camarão mais cultivada no mundo. Inicialmente, foi explorada a diversidade das crustinas mais abundantes e diversas em camarões peneídeos (Tipo II) (**Capítulo 1**). Foram caracterizadas, pela primeira vez, crustinas do Tipo IIb em uma espécie de camarão ocidental e comparadas em nível molecular, transcricional e genético com crustinas do Tipo IIa. Embora codificadas por genes distintos, ambas apresentaram organizações genômicas, características bioquímicas e perfis transcricionais similares. As análises *in silico* levaram, ainda, à identificação de assinaturas moleculares específicas que definem os Tipos IIa e IIb em camarões e proveram evidências de que essas moléculas evoluíram a partir de um gene ancestral comum. Para estudar a diversidade das crustinas de maneira mais ampla e explorar a existência de novos membros, buscas sistemáticas foram realizadas no recém-publicado genoma de *L. vannamei* (**Capítulo 2**). As sequências foram prospectadas através de múltiplas abordagens *in silico* que culminaram com a identificação de 61 genes, incluindo crustinas que não foram ainda descritas. As análises de bioinformática revelaram que as crustinas de *L. vannamei* são bioquimicamente diversas e apresentam uma extensa variedade de organizações genômicas, cuja diversidade deriva, possivelmente, de um extensivo processo de expansão gênica que remonta, pelo menos, a espécies ancestrais da família Penaeidae. Ademais, a identificação de novas crustinas denunciou uma clara fragilidade dos atuais sistemas de classificação que parecem não abarcar a real diversidade molecular dessa família de AMPs e ressaltou a necessidade de um novo modelo, baseado em assinaturas moleculares específicas. Esses achados fornecem uma importante contribuição para a compreensão do sistema imune de camarões e oferece subsídios para estudos futuros que visem o controle de enfermidades nos cultivos.

Palavras-chave: Crustinas. Peptídeos antimicrobianos. Camarão. Domínio WAP. Sistema imune de invertebrados.

ABSTRACT

Crustins represent the largest and most diverse family of antimicrobial peptides (AMPs) described in crustaceans. They are classically defined as cysteine-rich peptides/polypeptides containing a typical WAP (Whey Acidic Protein) domain at C-terminal region and are classified into five main groups, referred to as Types I to V. Crustins are diverse in terms of primary structure, size and biochemical features, thus exhibiting multiple biological functions that go beyond their antimicrobial properties. The present study aimed to estimate the molecular, transcriptional and genetic diversity of crustins from *Litopenaeus vannamei*, the most cultured shrimp species worldwide. Firstly, the diversity of the most abundant and diverse crustins in penaeid shrimp (Type II) was explored (**Chapter 1**). For the first time, Type IIb crustins were identified and characterized in a western shrimp species and compared at the molecular, transcriptional and genetic level with Type IIa crustins. Although encoded by distinct genes, both crustins showed similar genomic organizations, biochemical features and transcriptional profiles. *In silico* analyzes have also shown that each sub-type displays specific amino acid signatures and provided evidence that these molecules evolved from a common ancestral gene. To further study the crustin diversity and explore the existence of novel crustins members, searches were carried out on recently published genome of *L. vannamei* (**Chapter 2**). The sequences were retrieved using multiple *in silico* approaches that led to the identification of 61 genes, including crustins that have not yet been described. Bioinformatics analyzes revealed that *L. vannamei*'s crustins are biochemically diverse and exhibit a broad variety of genomic organizations. The diversity observed among coding genes was possibly associated to an extensive process of gene expansion that traces back, at least, to ancestral species of the Penaeidae family. Furthermore, the identification of novel crustin members revealed a clear weakness of the current classification systems, which seemly do not embrace the real molecular diversity of this family of AMPs and highlighted the need for a new model, based on specific molecular signatures. These findings provide an important contribution to the understanding of the shrimp immune system and offer subsidies for future studies aiming the control of diseases in shrimp farming.

Keywords: Crustins. Antimicrobial Peptides. Shrimp. WAP domain. Invertebrate immune system.

LISTA DE FIGURAS

- Figura 1.** Esquema ilustrativo dos principais mecanismos de ação de peptídeos antimicrobianos na superfície ou no interior de células bacterianas.24
- Figura 2.** Representação esquemática do domínio *Whey Acidic Protein* (WAP) obtido no banco de proteínas PFAM ilustrando a posição dos oito resíduos conservados de cisteína e as ligações dissulfeto envolvidas na formação do “*four-disulfide core*” (4-DSC).29
- Figura 3.** Representação esquemática da estrutura primária dos diferentes Tipos de crustinas descritas em Crustáceos (Tipos I a IV).34
- Figura 4.** Árvore filogenética das crustinas de crustáceos e insetos.36
- Figura 5.** *Litopenaeus vannamei* Type IIb crustins. Amino acid sequences alignment of Type IIb crustins from *L. vannamei* (Crustin-like *Lv*) and the crustin $Pm7$ from *Penaeus monodon*. Identical residues are highlighted in black. Triangles (▼) indicate the 12 conserved cysteine residues found in crustins. The whey acidic protein (WAP) domain is underlined by a solid red line.61
- Figura 6.** *Litopenaeus vannamei* Type II crustins are encoded by distinct genomic sequences. A not-to-scale representation of *L. vannamei* Type II crustins, Type IIa (Crustin *Lv*) and Type IIb (Crustin-like *Lv*). White boxes indicate the exons and the black line indicates the intron. The numbers show the size (in base pairs) of the exons and introns.62
- Figura 7.** Gene expression distribution of Type II crustins in shrimp tissues. Semiquantitative reverse transcriptase polymerase chain reaction (RT-PCR) analysis of Crustin *Lv* (Type IIa) and Crustin-like *Lv* (Type IIb) transcript levels in different tissues from naïve (N) and *Vibrio*-stimulated (S) shrimp. The expression of the β -actin gene was used as an endogenous control. HE: hemocytes, GL: gills, ML: muscle, NC: nerve cord, FG: foregut, HP: hepatopancreas, MG: midgut, HG: hindgut.63
- Figura 8.** Relative expression profile of Crustin *Lv* (Type IIa) and Crustin-like *Lv* (Type IIb) genes in (A) circulating hemocytes and (B) midgut of shrimp at 48 h after experimental infections with the Gram-negative *Vibrio harveyi* ATCC 14126 (white bars) or the White spot syndrome virus (black bars). Results are presented as mean \pm standard deviation of relative expressions (three biological replicates). WSSV: White spot syndrome virus. W-free: tissue homogenate inoculum prepared from WSSV-free shrimp.65
- Figura 9.** Relative abundance of Crustin-like *Lv* (Type IIb) transcripts during shrimp development. Results are present as mean \pm standard deviation (three biological replicates). The red dotted line indicates the basal expression level of the Crustin-like *Lv* gene in hemocytes from juvenile shrimp. EI: fertilized eggs at 0–4 h post-spawning; EII: fertilized eggs at 7–11 h post-spawning; NI: nauplius I; NV: nauplius V; ZI: protozoa I; ZIII: protozoa III; MI: mysis I; MIII: mysis III; PL2: postlarva 2; PL9: postlarva 9; PL17: postlarva 17. Different letters indicate significant differences among the developmental stages

from EII to PL17 (one-way ANOVA/Tukey, $p < 0.05$). Asterisks (*) shows significant differences between each developmental stage and hemocytes from juveniles (one-way ANOVA/Tukey, $p < 0.05$)..... 67

Figura 10. The molecular signature of shrimp Type II crustins. (A) Amino acid sequence alignments of the C-terminal region holding the 12 conserved cysteine residues (the crustin signature [39]) of Type IIa and Type IIb crustins from penaeid shrimp: *Litopenaeus vannamei* (Crustin Lv1, Crustin Lv3, LvCrustin I, LvCrustin P, Crustin-like Lv), *Litopenaeus setiferus* (Crustin Ls1, Crustin Ls2, Crustin Ls3), *Litopenaeus schmitti* (CrusLsch), *Farfantepenaeus paulensis* (CrusFpau), *Farfantepenaeus brasiliensis* (CrusFbra), *Farfantepenaeus subtilis* (CrusFsub), *Fenneropenaeus chinensis* (Fc-crus 2, CruFc), *Fenneropenaeus indicus* (Fi-crustin), *Penaeus monodon* (crustinPm1, crustinPm4, crustinPm5, crustinPm6, crustinPm7, Crustin-like Pm), and *Marsupenaeus japonicus* (Crustin Mj, Crustin-like Mj). (B) Consensus amino acid sequence of shrimp Type IIa and Type IIb crustins. X indicates any amino acid. Identical residues are highlighted in black. Specific amino acid residues conserved in shrimp Type IIa and Type IIb crustins are highlighted in blue and yellow, respectively. Triangles (▼) indicate the 12 conserved cysteine residues found in crustins. The whey acidic protein (WAP) domain is underlined by a solid red line. (C) A not-to-scale representation of shrimp Type IIa and Type IIb crustins indicating the N-terminal glycine-rich region and the C-terminal crustin signature (cysteine-rich region + WAP domain). MW: molecular weight. pI: theoretical isoelectric point. * Sequence obtained from [16] (not deposited in any database). 68

Figura 11. Penaeid shrimp Type II crustins form two distinct phylogenetic clades. The phylogenetic tree was constructed using the Neighbor-Joining method in MEGA 6. Bootstrap sampling was reiterated 1000 times. Sequences included in analyses were the following: (i) Type I crustins (“Carcinins”): the Pp-crustin from *Portunus pelagicus* (GenBank: JQ965930), the Sc-crustin from *Scylla serrata* (GenBank: HQ638025), the CrusSp from *Scylla paramamosain* (GenBank: EU161287), the CruHal from *Hyas araneus* (GenBank: EU921641), PtCrustin from *Portunus trituberculatus* (GenBank: FJ612106), the carcinin (11.5-kDa peptide) from *Carcinus maenas* (GenBank: AJ427538), the PjCrustin2 from *Pacifastacus leniusculus* (GenBank: EF523613), the Pc-crustin 2 from *Procambarus clarkii* (GenBank: GQ301202), the MjCruI-1 and MjCruI-2 from *Marsupenaeus japonicus* [11], and the carcininPm2 from *Penaeus monodon* [10]; (ii) Type IIa crustins (“Crustins”): the LvCrustin I (GenBank: AY488492), LvCrustin P (GenBank: AY488494), Crustin Lv1 (GenBank: AF430071) and Crustin Lv3 (GenBank: AF430073) from *Litopenaeus vannamei* (indicated by black circles, ●), the Crustin Ls1 (GenBank: AF430078), Crustin Ls2 (GenBank: AF430077) and Crustin Ls3 (GenBank: AF430079) from *Litopenaeus setiferus*, the CrusLsch (GenBank: EF182748) from *Litopenaeus schmitti*, the CrusFpau (GenBank: EF182747) from *Farfantepenaeus paulensis*, the CrusFbra (GenBank: EF601055) from *Farfantepenaeus brasiliensis*, the CrusFsub (GenBank: EF450744) from *Farfantepenaeus subtilis*, the Crustin Mj (GenBank: AB121741) from *M. japonicus*, the crustinPm1 (GenBank: FJ686014), crustinPm4 (GenBank: FJ686015), crustinPm5 (GenBank: FJ380049) and crustinPm6 (GenBank: GH717910) from *P. monodon*, the Fc-crus 2 (GenBank: FJ853147) from *Fenneropenaeus chinensis* and the PJC1 (GenBank: FJ797417), PJC2 (GenBank: FJ797418), PJC3 (GenBank: FJ797419), and PJC4 (GenBank: FJ797420) from *Panulirus japonicus*; (iii) Type IIb crustins (“Crustin-like”): the Crustin-like Lv (GenBank: JQ824114, FE049920 and FE049921) from *L. vannamei* (indicated by black squares, ■), the Crustin-like Mj from *M. japonicus* [16], the CruFc (GenBank: DQ097703) from *F. chinensis*, the Fi-crustin (GenBank: GQ469987) from *Fenneropenaeus indicus* and the Crustin-like Pm

(GenBank: GQ334395, JX912161 and GU299808) and crustin*Pm7* (GenBank: EF654658) from *P. monodon*; (iv) Type III crustins (“Single WAP domain-containing proteins or SWD”): the *LvSWD* (GenBank: AY464465) and *LvSWDi* [40] from *L. vannamei*, the *SWDPm1* (GenBank: EU623979) and *SWDPm2* (GenBank: EU623980) from *P. monodon*, the *FcSWD* (GenBank: EF216349) from *F. chinensis* and the *MjSWD* (GenBank: AU176270) from *M. japonicus*; (v) Type IV crustins (“Double WAP domain-containing proteins or DWD”): the *LvSLPI* (GenBank: EF467169) from *L. vannamei*, the *PmDWD* (GenBank: BI784457) from *P. monodon*, the *FcDWD* (GenBank: GQ303571) from *F. chinensis*, and the *MjDWD* (GenBank: EU095018) from *M. japonicus*. 71

Figura 12. Alinhamento aminoacídico da região referente ao “domínio crustina” das crustinas identificadas em *L. vannamei*. 84

Figura 13. Alinhamento das crustinas do Tipo II identificadas em *L. vannamei*. 88

Figura 14. Grupos de crustinas do Tipo II reconhecidos em *L. vannamei*. 90

Figura 15. Árvore filogenética das crustinas do Tipo II de camarões peneídeos. 93

Figura 16. Distribuição topológica dos genes codificantes de crustinas dos grupos 1 e 2 no genoma de *L. vannamei*. 95

Figura 17. Organização genômica dos genes codificantes de crustinas do Tipo II identificados em *L. vannamei*. 97

Figura 18. Representação tridimensional do espectro de diversidade bioquímica entre as crustinas do Tipo II de *L. vannamei*. 99

Figura 19. Grupos de crustinas do Tipo I identificadas em *L. vannamei*. 101

Figura 20. Reconstrução filogenética e estrutura primária de crustinas do Tipo I identificadas em camarões peneídeos. 106

Figura 21. Organização genômica dos genes codificantes de crustinas do Tipo I identificados em *L. vannamei*. 108

Figura 22. Comparação das sequências nucleotídicas dos genes codificantes das crustinas do grupo 7 (Tipo I) e sua distribuição topológica no genoma de *L. vannamei*. 109

Figura 23. Comparação da composição aminoacídica e arquitetura molecular dos genes do grupo 1 (Tipo I). 110

Figura 24. Representação tridimensional do espectro de diversidade bioquímica entre as carcininas e crustinas similares ao Tipo I de *L. vannamei*. 112

| | |
|---|-----|
| Figura 25. Alinhamento aminoacídico comparativo entre o domínio crustina presente no grupo 4 e nas autênticas crustinas do Tipo I e II. | 115 |
| Figura 26. Assinatura aminoacídica e distribuição de crustinas WAPd 1 e WAPd 2 em Dendrobranchiata..... | 117 |
| Figura 27. Distribuição taxonômica dos genes codificantes de WAPd1 e WAPd2 em crustáceos dendrobranquiados..... | 119 |
| Figura 28. Esquema ilustrativo mostrando as novas crustinas recuperadas nas buscas pelo genoma de <i>L. vannamei</i> | 121 |
| Figura 29. Comparação aminoacídica e reconstrução filogenética das novas crustinas identificadas em <i>L. vannamei</i> | 123 |

LISTA DE QUADROS

| | |
|--|-----|
| Quadro 1. Repertório completo de crustinas recuperadas no proteoma predito de <i>L. vannamei</i> | 85 |
| Quadro 2. Validação <i>in silico</i> das crustinas do Tipo II reconhecidas em <i>L. vannamei</i> | 92 |
| Quadro 3. Validação <i>in silico</i> dos subtipos de crustinas do Tipo I reconhecidos em <i>L. vannamei</i> nos bancos transcritômicos TSA e EST desta espécie..... | 102 |

LISTA DE TABELAS

| | |
|--|-----|
| Tabela 1. Sequências dos iniciadores utilizados neste trabalho. | 52 |
| Tabela 2. Lista de sequências homólogas às crustinas do Tipo II identificadas em <i>L. vannamei</i> e em diferentes espécies de camarões peneídeos. | 91 |
| Tabela 3. Lista de prováveis ortólogos das crustinas do Tipo I identificadas em <i>L. vannamei</i> em diferentes espécies de camarões peneídeos. | 104 |
| Tabela 4. Validação <i>in silico</i> em bancos transcritômicos de <i>L. vannamei</i> das novas crustinas identificadas e identificação de prováveis ortólogos em camarões peneídeos. | 122 |

LISTA DE ABREVIATURAS E SIGLAS

- ALF** Fatores anti-lipopolissacarídeos
ALP Antileucoprotease
AMPs Peptídeos antimicrobianos
CFU Unidades formadoras de colônia
Cq Ciclo de quantificação
dNTPs Deoxinucleotídeos trifosfatados
DWD Proteínas contendo dois domínios WAP (*Double WAP Domain*)
EDTA Ácido etilenodiamino tetra-acético
EST Etiqueta de sequência expressa (*Expressed Sequence Tag*)
FAO Organização das Nações Unidas para Agricultura e Alimentação
kDa kilodaltons
MAS Solução de Alsever Modificada
pb Pares de base
RNAi RNA de interferência
RT-qPCR Transcrição reversa seguida de PCR quantitativa em tempo real
SLPI Inibidores de Proteases Secretados por Leucócitos (*Secretory Leukocyte Protease Inhibitors*)
SWD Proteínas contendo um domínio WAP (*Single WAP Domain*)
TSA *Transcriptome Shotgun Assembly*
WAP *Whey Acidic Protein*
WSSV Vírus da Síndrome da Mancha Branca
4-DSC *Four-Disulfide Core*

L - litro
mL – mililitro
μL – microlitro
g – grama
μg – micrograma
μM – micromolar
mM - milimolar
DO – densidade óptica
U – unidade de atividade enzimática
v/v – diluição volume/volume

h – hora(s)
min – minuto(s)
s – segundo(s)

LISTA DE SÍMBOLOS

°C - graus Celsius
< - maior
> - menor
% - porcentagem
 α - letra grega alfa
 β - letra grega beta
 Δ - letra grega delta
 μ - micro
 \pm - mais ou menos

SUMÁRIO

| | |
|--|----|
| I. INTRODUÇÃO GERAL | 21 |
| II. ESTADO DA ARTE | 27 |
| II.1. Aspectos gerais das crustinas: definição, histórico e distribuição taxonômica..... | 28 |
| II.2. Domínio <i>Whey Acidic Protein</i> (WAP): uma característica marcante das crustinas..... | 29 |
| II.3. Sistemas de classificação das crustinas..... | 33 |
| II.4. Aspectos funcionais das crustinas..... | 37 |
| II.5. As crustinas do camarão <i>Litopenaeus vannamei</i> | 40 |
| III. OBJETIVOS | 42 |
| III.1. Objetivo geral..... | 43 |
| III.2. Objetivos específicos..... | 43 |
| IV. METODOLOGIA | 44 |
| IV.1. Análises de sequências..... | 45 |
| IV.1.1. Buscas <i>in silico</i> | 45 |
| IV.1.2. Caracterização molecular e reconstruções filogenéticas..... | 46 |
| IV.2. Animais..... | 47 |
| IV.3. Desafio microbiano e coleta dos tecidos..... | 47 |
| IV.4. Infecções experimentais..... | 48 |
| IV.5. Purificação de RNA total e síntese de cDNA..... | 49 |
| IV.6. Transcrição reversa seguida de PCR semiquantitativa..... | 50 |
| IV.7. Transcrição reversa seguida de PCR quantitativa em tempo real..... | 50 |
| IV.7.1. Perfil transcricional frente a infecções experimentais..... | 50 |
| IV.7.2. Perfil transcricional durante o desenvolvimento..... | 51 |
| V. RESULTADOS | 53 |
| CAPÍTULO 1 | 55 |
| 1. Introduction | 58 |
| 2. Materials and Methods | 60 |
| 3. Results and Discussion | 60 |
| 3.1. Identification and molecular characterization of Type IIb crustins in <i>L. vannamei</i> .. | 60 |
| 3.2. <i>L. vannamei</i> Type II crustins are encoded by distinct genes..... | 61 |
| 3.3. Gene expression distribution of Type II crustins in shrimp tissues..... | 62 |
| 3.4. Expression profile of Type II crustins in response to bacterial and viral infections .. | 63 |
| 3.5. Expression of Type IIb crustins during shrimp ontogenesis..... | 65 |

| | |
|--|-----|
| 3.6. Molecular signatures for Type II crustins in penaeid shrimp..... | 66 |
| 4. Conclusions | 72 |
| Supplementary Data | 73 |
| References | 76 |
| CAPÍTULO 2 | 79 |
| 1. Introdução | 81 |
| 2. Materiais e Métodos | 83 |
| 3. Resultados e Discussão | 83 |
| 3.1. Descrição geral do repertório molecular das crustinas de <i>L. vannamei</i> | 83 |
| 3.2. Identificação e caracterização das crustinas do Tipo II..... | 87 |
| 3.2.1. As crustinas do Tipo II são diversas e constituídas por múltiplos grupos..... | 87 |
| 3.2.2. Crustinas do Tipo II experimentaram uma ampla diversificação em camarões peneídeos..... | 92 |
| 3.2.3. As crustinas do Tipo II de camarões são codificadas por genes que apresentam organizações genômicas diversas..... | 96 |
| 3.2.4. As crustinas do Tipo II são bioquimicamente diversas e exibem características moleculares contrastantes..... | 98 |
| 3.3. Identificação e caracterização das crustinas do Tipo I..... | 100 |
| 3.3.1. As crustinas do Tipo I de camarões são, também, formadas por grupos distintos..... | 100 |
| 3.3.2. A arquitetura molecular das crustinas do Tipo I não é exclusiva às carcininas.... | 104 |
| 3.3.3. As autênticas carcininas e crustinas similares ao Tipo I são codificadas por genes distintos..... | 107 |
| 3.3.4. As autênticas carcininas e crustinas similares ao Tipo I compreendem um grupo bioquimicamente diverso..... | 112 |
| 3.3.5. Alguns membros específicos de crustinas similares ao Tipo I são, na verdade, crustinas do Tipo II desprovidas da região <i>gly-rich</i> N-terminal..... | 113 |
| 3.3.6. Crustinas atípicas contendo um domínio WAP degenerado provavelmente evoluíram especificamente na linhagem Dendrobranchiata..... | 116 |
| 3.4. Identificação e caracterização molecular de novas crustinas em <i>L. vannamei</i> | 120 |
| VI. DISCUSSÃO GERAL | 125 |
| VII. CONCLUSÕES E PERSPECTIVAS | 131 |
| CONCLUSÕES..... | 132 |
| PERSPECTIVAS..... | 132 |
| REFERÊNCIAS | 134 |
| ANEXOS | 141 |
| ANEXO 1. Folha de rosto do artigo 1, referente ao capítulo 1 da tese..... | 142 |

| | |
|---|-----|
| ANEXO 2. Artigo publicado em co-autoria durante período de pós-graduação no LIAA. | 143 |
| ANEXO 3. Artigo publicado em co-autoria durante período de pós-graduação no LIAA. | 144 |
| ANEXO 4. Artigo publicado em co-autoria durante período de pós-graduação no LIAA. | 145 |
| ANEXO 5. Artigo publicado em co-autoria durante período de pós-graduação no LIAA. | 146 |
| ANEXO 6. Artigo publicado em co-autoria durante período de pós-graduação no LIAA. | 147 |
| ANEXO 7. Artigo publicado em co-autoria durante período de pós-graduação no LIAA. | 148 |

I. INTRODUÇÃO GERAL

A carcinicultura é uma atividade em franca expansão no Brasil. Diversos estados da costa brasileira realizam o cultivo de espécies marinhas de crustáceos, sendo o camarão branco do Pacífico (*Litopenaeus vannamei*) a espécie de maior proeminência no cenário nacional em virtude de suas excelentes características zootécnicas, incluindo o rápido ritmo de crescimento e facilidade de manejo (FAO, 2004). Entretanto, as práticas convencionais de cultivo aliadas a outras perturbações ambientais podem levar os animais cultivados a um intenso estresse que se reflete na depleção de suas condições de saúde, deixando-os mais susceptíveis ao acometimento de doenças infecciosas (LIGHTNER *et al.*, 2012). Nesse cenário, destaca-se o vírus causador da Síndrome da Mancha Branca (WSSV) que no Brasil trouxe impactos irreparáveis na produção de camarões no estado de Santa Catarina no início dos anos 2000 (EPAGRI, 2018) e que atualmente se encontra na região nordeste, onde a produção carcinícola é mais expressiva (NUNES; FEIJÓ, 2016). Além disso, doenças de origens bacterianas também têm ameaçado a produção, em especial aquelas causadas por bactérias do gênero *Vibrio*, como as espécies *V. harveyi* e *V. parahaemolyticus* (TRAN *et al.*, 2013).

Os ambientes marinhos são abundantemente habitados por uma ampla variedade de microrganismos que podem apresentar modos de vida livre ou colonizando diferentes tecidos e órgãos dos animais que ali vivem. Por essa razão, os invertebrados marinhos estão constantemente em contato com essa comunidade microbiana, que inclui tanto espécies comensais quanto espécies patogênicas (SCHMITT *et al.*, 2012). Embora esses animais apresentem naturalmente uma densa e diversificada microbiota residente na superfície de seus corpos e também no interior de seus tecidos (RUNGRASSAMEE *et al.*, 2013), em situações de estresse esse delicado equilíbrio pode ser perturbado. As alterações fisiológicas decorrentes desses estresses e concomitante disbiose abrem portas para a infiltração e multiplicação de patógenos oportunistas e o consequente aparecimento de doenças que causam sérios prejuízos à produção (SCHMITT *et al.*, 2012).

Para lidar com essas ameaças, os crustáceos contam com eficientes estratégias de defesa que conferem uma proteção essencial contra a abundante diversidade microbiana dispersa no ambiente (ROSA; BARRACCO, 2010). Essas estratégias de defesa são vinculadas a um sistema imunológico do tipo inato, ausente de linhagens linfocíticas, anticorpos e memória imunológica, e são baseadas na participação dos hemócitos (células imunes circulantes) e de uma miríade de moléculas solúveis imunofetoras. Nos últimos anos, o desenvolvimento de abordagens moleculares como técnicas genômicas e transcritômicas em larga escala permitiu a identificação de importantes efetores do sistema imune de camarões,

que têm sido visados para aplicação em biotecnologia marinha (ROBALINO *et al.*, 2009). Nesse cenário, os peptídeos antimicrobianos têm se mostrado fortemente promissores para o desenvolvimento de estratégias otimizadas no controle e combate a enfermidades nos cultivos (DESTOUMIEUX-GARZÓN *et al.*, 2016).

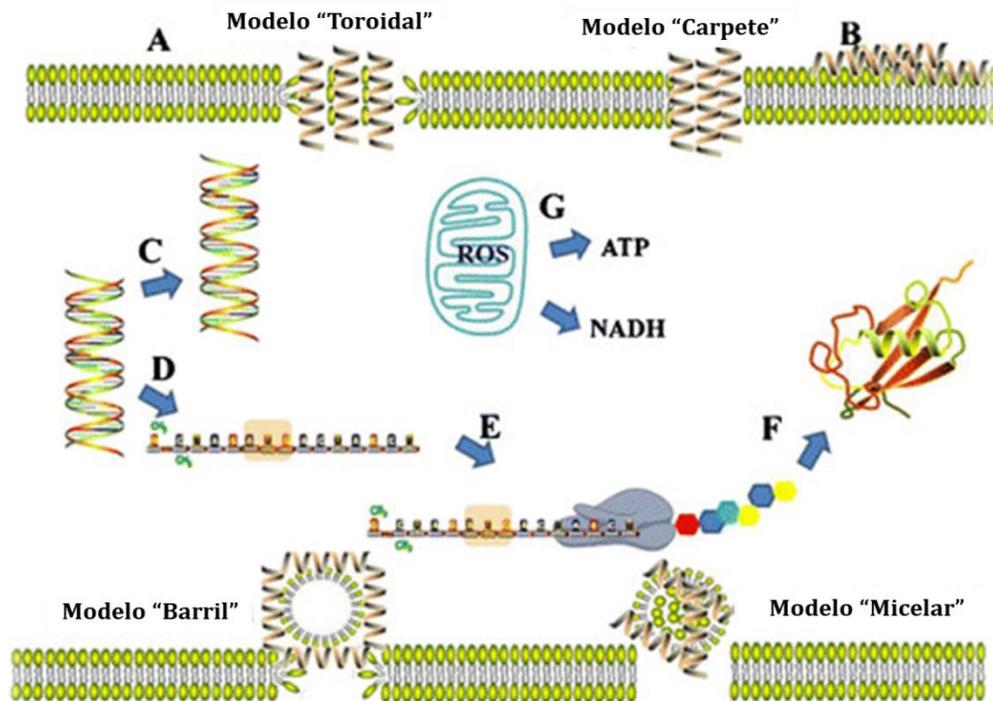
Peptídeos antimicrobianos (AMPs, do inglês *antimicrobial peptides*) são amplamente difundidos nas mais variadas formas de vida e constituem um componente importante da imunidade inata a infecções microbianas (GUANÍ-GUERRA *et al.*, 2010). Essas moléculas, comumente referidas como “antibióticos naturais”, se adaptaram às diversas pressões seletivas impostas pela composição microbiana do ambiente para continuar servindo como protagonistas na imunidade dos metazoários, especialmente naqueles desprovidos de um sistema imunológico adaptativo, como é o caso dos invertebrados (DESTOUMIEUX-GARZÓN *et al.*, 2016). Além de atuarem de maneira rápida e eficiente na primeira linha de defesa contra microrganismos invasores como bactérias Gram-positivas, Gram-negativas, leveduras, fungos filamentosos e vírus envelopados (SCHMITT; ROSA; DESTOUMIEUX-GARZÓN, 2016) e desempenharem, também, papéis multifuncionais nas interações patógeno-hospedeiro, os AMPs abarcam um grande potencial biotecnológico no sentido de representarem fortes candidatos para a exploração em uma nova geração de antibióticos (COSTA *et al.*, 2019). No preocupante cenário atual onde o aparecimento de bactérias resistentes a múltiplas drogas é uma realidade, os AMPs têm se mostrado fortemente promissores, em virtude de seu amplo espectro antimicrobiano, alta seletividade, baixa toxicidade em células eucarióticas e, também, da dificuldade do desenvolvimento de resistência bacteriana quando comparados aos antibióticos convencionais (ZASLOFF, 2019).

Quimicamente, os AMPs são definidos como curtas cadeias peptídicas (<10 kDa) com tamanho entre 15 e 100 resíduos aminoácídicos e compostos majoritariamente por resíduos hidrofóbicos e catiônicos. Essa organização estrutural faz dos AMPs moléculas anfipáticas com propriedades detergentes, capazes de se difundir igualmente em ambientes aquosos e lipídicos (ZASLOFF, 2002). A presença marcante de resíduos catiônicos torna os AMPs moléculas carregadas positivamente em pH fisiológico ao passo que a composição rica em resíduos hidrofóbicos permite também sua difusão em ambientes apolares hidrocarbonados através de interações hidrofóbicas. Entretanto, alguns AMPs fogem a esse perfil estrutural clássico, podendo exibir características aniônicas que surgem a partir da representação massiva de resíduos aminoácídicos negativamente carregados (BROGDEN, 2005). De maneira ainda mais interessante, alguns AMPs não são sequer codificados por genes no genoma. Nesses casos, esses AMPs são fragmentos derivados de proteínas maiores

cujas funções podem ser alternativas à função imune, como é o caso das lactoferricinas B e H derivadas da clivagem amino-terminal da lactoferrina bovina e humana, respectivamente, ou do peptídeo PvHCt, um curto fragmento com potente atividade antifúngica oriundo da clivagem C-terminal da hemocianina de *L. vannamei*, proteína responsável pelo transporte de gases respiratórios (BELLAMY *et al.*, 1992; DESTOUMIEUX-GARZÓN *et al.*, 2001). Por serem constituídos a partir dos 20 resíduos de aminoácidos comumente presentes em proteínas e igualmente encontrados em células microbianas, a notável propriedade antimicrobiana desses peptídeos instigou a busca pela compreensão dos mecanismos bioquímicos associados à sua atividade biológica.

Estudos funcionais *in vitro* mostram que os AMPs atuam, de modo geral, em nível de membrana através de sua inserção e concomitante desestabilização da superfície dos microrganismos. Essa capacidade deriva, em grande parte, da combinação de suas propriedades bioquímicas elementares e da composição química das superfícies microbianas. Por razões ainda desconhecidas, as membranas celulares microbianas apresentam fosfolipídeos carregados negativamente expostos à face externa das células, enquanto que as células de organismos multicelulares expõem fosfolipídeos zwitteriônicos nessa face e segregam aqueles de natureza negativa na face intracelular (BROGDEN, 2005). Esse importante contraste entre as superfícies desses dois tipos celulares fundamenta o mecanismo clássico dos AMPs: a superfície negativamente carregada das bactérias atrai os AMPs através de forças eletrostáticas de modo que, após esse primeiro contato, as porções hidrofóbicas dos AMPs se inserem nas bicamadas lipídicas, formando poros na superfície bacteriana e comprometendo irreversivelmente sua função biológica. No entanto, embora esse represente um mecanismo de ação clássico, diversos estudos mostram a participação de AMPs em nível intracelular. Nesse caso, eles são translocados através da membrana microbiana e atuam sobre alvos específicos, promovendo a inibição de funções metabólicas indispensáveis, como a síntese da parede celular, síntese proteica e de ácidos nucleicos (AISENBREY; MARQUETTE; BECHINGER, 2019) (**Figura 1**).

Figura 1. Esquema ilustrativo dos principais mecanismos de ação de peptídeos antimicrobianos na superfície ou no interior de células bacterianas.



Diversidade de mecanismos de ação de peptídeos antimicrobianos. A figura ilustra dois tipos principais de AMPs: peptídeos ativos contra membrana e peptídeos ativos a nível intracelular. (A) Os mecanismos dos AMPs ativos contra membranas celulares incluem os modelos "barril", "toroidal", "carpete" e "micelar". (B-G) ilustram os mecanismos de ação intracelular de AMPs. (B) Inibição de enzimas necessárias para a ligação de proteínas estruturais da parede celular, (C) inibição da síntese de DNA, (D) inibição da síntese de RNA, (E) inibição da função ribossômica e síntese de proteínas, (F) bloqueio de proteínas chaperonas necessárias para o dobramento adequado de proteínas, (G) inibição da respiração celular e indução da produção de espécies reativas de oxigênio (ROS), dano da integridade da membrana mitocondrial e efluxo de ATP e NADH.

Uma vez que não há restrições específicas em relação à composição aminoacídica além de apresentarem uma natureza química anfipática, os AMPs apresentam uma vasta diversidade de sequência (ZASLOFF, 2019). De fato, os AMPs despontam como moléculas funcionalmente plásticas, uma vez que substituições discretas em resíduos aminoacídicos podem alterar seu espectro antimicrobiano sem alterar suas características biofísicas elementares. Como consequência, organismos que se deparem com alterações no microbioma circundante para o qual seu repertório de AMPs é inadequado, novas moléculas podem evoluir prontamente através de modificações pontuais em sua composição para lidar com os novos desafios. O mecanismo de atuação dos AMPs oferece, ainda, uma grande vantagem evolutiva: bactérias apresentam uma baixa capacidade de evasão ao ataque dessas moléculas. Uma vez que os AMPs atuam majoritariamente em nível de membrana celular, o desenvolvimento de resistência torna-se mais custoso para os microrganismos, visto que um conjunto amplo de mutações associadas à reorganização estrutural da membrana celular se faz

necessário. Além disso, a ausência de assinaturas moleculares específicas dos AMPs impede o surgimento de proteases que atuem sobre eles, mas que sejam inócuas às próprias proteínas bacterianas (ZASLOFF, 2019).

Atualmente, diferentes famílias de AMPs foram identificadas em crustáceos (ROSA; BARRACCO, 2010). Por razões econômicas, as mais bem caracterizadas vêm de espécies cultivadas, como os decápodes marinhos (camarões, lagostas, caranguejos etc), e muitas são específicas a determinados grupos. Em decápodes, as famílias de AMPs codificadas por genes mais bem descritas são as crustinas, as peneidinas e os fatores anti-lipopolissacarídeo (ALFs) (DESTOUMIEUX-GARZÓN *et al.*, 2016). Dessas três, as crustinas são as mais representativas, visto que ocorrem amplamente na classe Malacostraca, maior classe dentre os crustáceos a qual reúne cerca de 30 mil espécies viventes. De fato, além das cerca de 50 crustinas documentadas na Ordem Decapoda, sequências de genes similares às crustinas ocorrem em todo grupo Pancrustacea, sobretudo nas Ordens Amphipoda, Isopoda e até mesmo em alguns insetos (LAI; ABOOBAKER, 2017; TASSANAKAJON; SOMBOONWIWAT; AMPARYUP, 2015).

As demandas por novas ferramentas de combate e controle de doenças infecciosas nos cultivos tornam imprescindíveis estudos voltados para a exploração de moléculas com potencial para utilização na carcinicultura visando à mitigação dos severos impactos causados no setor produtivo. Atualmente, o uso contínuo de antibióticos na água tem sido a estratégia utilizada para combater e prevenir surtos de doenças, uma vez que não há alternativas concretas para lidar com esse problema. No entanto, essa prática tem se tornado fortemente desaconselhada em virtude de (i) seus severos impactos ao ambiente, (ii) criação de um ambiente favorável ao desenvolvimento de cepas bacterianas resistentes a antibióticos e (iii) contaminação dos animais ali imersos, tornando-os impróprios para o consumo humano. Por essa razão, a descoberta de moléculas inócuas ao meio ambiente, capazes de reduzir a viabilidade de microrganismos patogênicos sem induzir resistência e produzidas naturalmente pelos animais cultivados são de grande valia para a exploração na aquicultura mundial e para aplicação na biotecnologia marinha. Nesse contexto, estudos sobre a diversidade de efetores imunológicos em crustáceos, em especial de crustinas, são importantes para ampliar o conhecimento acerca dos mecanismos moleculares de defesa desses animais e, também, para ampliar o repertório de imunofatores candidatos para uso em diferentes terapêuticas, como através da indução seletiva de genes imunes de interesse. Ademais, tendo em vista o potencial biotecnológico dos AMPs, a exploração da diversidade dessas moléculas multifuncionais

pode pavimentar novos rumos de pesquisa direcionados para a sua aplicação na saúde humana e animal.

Ainda que uma miríade de crustinas tenha sido descrita em diferentes espécies de crustáceos, estudos amplos dimensionando a diversidade dessa família de peptídeos antimicrobianos são escassos, sobretudo em nível intraespecífico. Nesse trabalho, a diversidade molecular e evolução de crustinas em camarões peneídeos foram investigadas partindo de análises sistemáticas em bancos genômicos e transcritômicos disponíveis de *L. vannamei*, espécie de camarão mais cultivada no mundo. As buscas por crustinas em *L. vannamei* revelaram, pela primeira vez, o extenso repertório molecular dessa família de AMPs em uma espécie de camarão peneídeo e proveram *insights* sobre possíveis mecanismos evolutivos que fizeram das crustinas a maior e mais diversa família de peptídeos antimicrobianos que se tem conhecimento em crustáceos.

Uma vez que a diversidade molecular das crustinas representa o objeto central de estudo desta tese de doutorado, uma breve apreciação da bibliografia atual incluindo os aspectos moleculares, funcionais e evolutivos está descrita na seção seguinte.

II. ESTADO DA ARTE

II.1. Aspectos gerais das crustinas: definição, histórico e distribuição taxonômica

Crustinas são classicamente definidas como peptídeos/polipeptídeos antimicrobianos (6-22 kDa) catiônicos e anfipáticos ricos em resíduos de cisteína presentes em crustáceos e insetos e que contêm uma sequência sinal na região N-terminal e um domínio WAP conservado (do inglês, *Whey Acidic Protein*) na região C-terminal. Historicamente, as crustinas foram uma das primeiras famílias de AMPs codificados por genes descritas em crustáceos. Sua história começa a partir da identificação de um peptídeo antibacteriano de 11,5 kDa enriquecido em resíduos de cisteína isolado dos hemócitos granulares do caranguejo *Carcinus maenas* (RELF *et al.*, 1999). Embora apenas uma caracterização parcial da sequência desse peptídeo tenha sido realizada no momento, essas moléculas exibiram uma estrutura primária marcadamente distinta de qualquer proteína antimicrobiana descrita em crustáceos e apresentaram atividade antibacteriana especificamente contra bactérias marinhas Gram-positivas ou tolerantes ao sal (RELF *et al.*, 1999). Os então chamados “peptídeos antibacterianos de 11,5 kDa” foram convenientemente denominados “carcininas” por Smith e Chisholm (2001) em homenagem ao gênero *Carcinus* e mostraram ser moléculas dependentes de elevadas concentrações salinas e altamente estáveis permanecendo ativas mesmo após a fervura. Anos mais tarde, Bartlett e colaboradores (2002) identificaram sequências homólogas nos camarões peneídeos *L. vannamei* e *L. setiferus* e cunharam o termo “crustinas” para descrevê-las, termo que logo se tornou universalmente utilizado na literatura para a descrição de moléculas similares em outras espécies de crustáceos. De fato, após essa primeira descrição em camarões, vários estudos descreveram moléculas semelhantes a crustinas (*crustin-like peptides*) em espécies relacionadas de crustáceos decápodes (RATTANACHAI *et al.*, 2004; ROSA, R. D.; BANDEIRA; BARRACCO, 2007; SUPUNGUL *et al.*, 2004; ZHANG, JIQUAN *et al.*, 2007). O número de relatos descrevendo novas crustinas aumentou substancialmente nos anos seguintes, de modo que, em menos de duas décadas, mais de 150 sequências foram descritas na literatura ou depositadas em bancos de dados públicos.

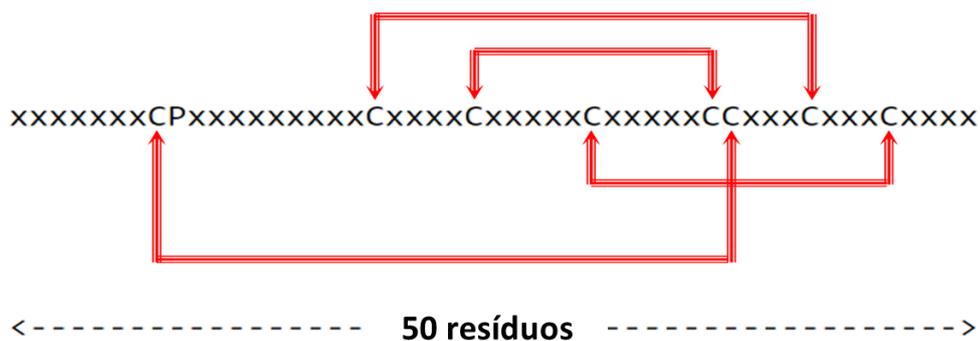
As crustinas representam a maior e mais diversa família de peptídeos antimicrobianos entre os Decapoda, tendo sido encontradas em ambas as subordens Pleocyemata e Dendrobranchiata (SMITH; DYRYNDA, 2015). Atualmente, 183 sequências aminoacídicas e 215 sequências nucleotídicas encontram-se depositadas em bancos de dados públicos sob descrição do termo “*crustin*”, oriundas de 35 espécies de invertebrados (<https://www.ncbi.nlm.nih.gov/search/all/?term=crustin>). A maioria expressiva das crustinas descritas foi identificada e caracterizada em espécies de crustáceos decápodes como camarões, lagostas, lagostins e caranguejos. No entanto, sequências referentes a crustinas

foram identificadas também em crustáceos não decápodes, especialmente naqueles pertencentes às ordens Amphipoda e Isopoda. Recentemente, Becking e colaboradores (2020) identificaram transcritos codificantes de crustinas em bancos transcritômicos de 21 espécies de crustáceos isópodes pertencentes a seis diferentes famílias. Além disso, abordagens similares empregadas por Lai e Aboobaker (2017) levaram à identificação de transcritos de crustinas em 55 espécies de crustáceos, das quais 37 representaram espécies não decápodes. Ademais, sequências de crustinas foram também identificadas em insetos himenópteros a partir de análises de genômica comparativa, mostrando que essa família de AMPs, mais do que restrita a um grupo seletivo de crustáceos, está presente em diferentes espécies dentro de Pancrustacea (ZHANG, Z.; ZHU, 2012).

II.2. Domínio *Whey Acidic Protein* (WAP): uma característica marcante das crustinas

As crustinas são peptídeos antimicrobianos peculiares na imunidade dos crustáceos em nível estrutural. Ao contrário de outros AMPs codificados por genes descritos em crustáceos, as crustinas pertencem a um grupo de proteínas que possuem um domínio *Whey Acidic Protein* (WAP) em sua estrutura. De fato, uma das características marcantes é a presença de um domínio WAP conservado na região C-terminal do peptídeo maduro. O domínio WAP é um motivo proteico de 50 resíduos aminoácídicos (<https://pfam.xfam.org/family/PF00095>) que possui oito resíduos de cisteína conservados envolvidos em quatro ligações dissulfeto intramoleculares. A ligação dissulfeto interna cria uma estrutura compacta que, em última análise, forma um arranjo tridimensional típico, descrito no PROSITE como *four-disulfide core* ou 4-DSC (SMITH *et al.*, 2008) (**Figura 2**).

Figura 2. Representação esquemática do domínio *Whey Acidic Protein* (WAP) obtido no banco de proteínas PFAM ilustrando a posição dos oito resíduos conservados de cisteína e as ligações dissulfeto envolvidas na formação do “*four-disulfide core*” (4-DSC).



X representa qualquer resíduo de aminoácido, C representa resíduos de cisteína, P representa resíduos de prolina enquanto que as setas em vermelho indicam as pontes de dissulfeto formadas entre os pares de cisteínas (adaptado de <https://pfam.xfam.org/family/pf00095#tabview=tab0>).

Apesar de todas as crustinas compartilharem esse motivo proteico comum, o domínio WAP não é exclusivo dessa família de peptídeos antimicrobianos. Na verdade, o termo “WAP” foi cunhado por Piletz e colaboradores (1981) para descrever uma nova proteína presente no soro do leite de roedores a partir do qual essas proteínas foram originalmente purificadas. Os transcritos codificantes das proteínas WAP presentes nos modelos murinos foram clonados e caracterizados e sua sequência aminoacídica deduzida foi identificada contendo domínios 4-DSC canônicos que haviam sido descritos em proteínas não lácteas, como em neurotoxinas associadas ao veneno de serpentes (HENNIGHAUSEN; SIPPEL, 1982). Uma vez que os termos “domínio WAP” e “4-DSC” denotam o mesmo motivo proteico eles passaram a ser usados como sinônimos e, para os fins desta tese, não serão feitas distinções entre eles.

A superfamília de proteínas WAP é amplamente encontrada em uma vasta diversidade de organismos, embora ironicamente uma fração pouco expressiva seja derivada de artrópodes, os animais mais abundantes do planeta. De fato, das mais de 3.400 sequências de proteínas WAP depositadas no banco de sequências aminoacídicas PFAM (ID: PF00095), apenas 15% (516 sequências) são oriundas desse grupo de metazoários. Dentre as proteínas WAP em invertebrados estão as *pearlwapins*, um pequeno grupo de proteínas descritas em moluscos. Essas proteínas são compostas por três domínios WAP e têm sido associadas à formação de conchas por meio da regulação da deposição de cristais de calcita em abalones (MARIE *et al.*, 2010; TRECCANI *et al.*, 2006). Curiosamente, os organismos mais basais que codificam proteínas pertencentes à superfamília WAP são bactérias pertencentes a diferentes filos, sugerindo que os genes codificantes de proteínas WAP surgiram precocemente na natureza e, por desempenharem funções biológicas importantes, alcançaram uma ampla gama de organismos ao longo do curso de evolutivo (SMITH, 2011).

Funcionalmente, a superfamília WAP agrega proteínas biologicamente diversas, incluindo membros com um ou mais domínios WAP. Na verdade, à exceção dos resíduos conservados de cisteína, a composição aminoacídica do domínio WAP pode ser bastante divergente e essa diversidade é refletida em sua amplitude funcional. Exemplos de proteínas WAP incluem a proteína semelhante à caltrina II (*caltrin-like protein II*), um inibidor do transporte de cálcio para os espermatozoides em porquinhos da Índia (CORONEL; SAN AGUSTIN; LARDY, 1990); SPAI-1, um inibidor de bombas de sódio/potássio no intestino de

porcos (ISHIZUKA *et al.*, 1991); ANOS1, uma importante proteína envolvida na formação de vários órgãos durante o desenvolvimento embrionário e potencialmente associada à síndrome de Kallman em humanos (SOUSSI-YANICOSTAS *et al.*, 1998); WDNM1 envolvida no potencial metastático de adenocarcinoma em ratos (DEAR; KEFFORD, 1991) e as próprias proteínas WAP do leite de mamíferos, cuja função biológica ainda não foi atribuída. Curiosamente, algumas proteínas contendo domínios WAP, como as proteínas SWAM1 e SWAM2 presentes em camundongos e as *omwaprinas* isoladas do veneno da serpente *Oxyuranus microlepidotus* apresentam atividades antibacterianas contra bactérias Gram-positivas e/ou Gram-negativas (HAGIWARA *et al.*, 2003; NAIR *et al.*, 2007). No entanto, proteínas WAP estão principalmente associadas à atividade antiproteolítica, conforme evidenciado pelas bem conhecidas antileucoproteínas (ALPs) de mamíferos que desempenham um papel fundamental durante a inflamação, evitando a degradação dos tecidos do hospedeiro por proteases secretadas por neutrófilos (HIEMSTRA *et al.*, 1996); elafinas, um grupo de inibidores de serino-proteases do tipo elastase (SCHALKWIJK; WIEDOW; HIROSE, 1999) e papilinas, um grupo de inibidores de metaloproteinase envolvidos na regulação do desenvolvimento embrionário em invertebrados (KRAMEROVA *et al.*, 2000).

As atividades antiproteolíticas dos inibidores de proteases mais bem estudados pertencentes à superfamília WAP, elafina e ALPs, foram associadas a uma região específica denominada “região de contato primária” (OTA *et al.*, 2002). Essa região fica em torno do segundo resíduo de cisteína do domínio WAP e contém uma metionina conservada adjacente à ligação peptídica cindível. Curiosamente, naquelas proteínas que apresentam propriedades antimicrobianas, a região correspondente à atividade antiproteolítica contém substituições nos resíduos aminoácídicos, sendo particularmente enriquecida em resíduos catiônicos e hidrofóbicos. A substituição dos resíduos nessa região pode responder pelo desvio de atividade biológica uma vez que o posicionamento adequado de aminoácidos catiônicos e hidrofóbicos tende a tornar a proteína anfipática, uma característica importante para aumentar sua capacidade detergente e torná-la capaz de se inserir e desestabilizar as membranas microbianas (ZASLOFF, 2002). Normalmente, as crustinas são desprovidas da região associada à inibição de proteases, mas possuem os resíduos catiônicos e hidrofóbicos que conferem o caráter antimicrobiano. De fato, reconstruções filogenéticas realizadas com um pequeno grupo de crustinas juntamente com algumas proteínas WAP de vertebrados sugerem que as crustinas são mais semelhantes às proteínas que exibem atividade antimicrobiana do que daquelas com propriedades antiproteolíticas (VARGAS-ALBORES; YEPIZ-PLASCENCIA; JIMÉNEZ-VEGA; ÁVILA-VILLA, 2004).

Embora certo grau de liberdade seja admitido para a composição aminoacídica no domínio WAP, o arranjo dos resíduos de cisteína parece ser precisamente conservado. Ranganathan e colaboradores (1999) realizaram uma extensa análise das sequências aminoacídicas de diversas proteínas WAP presentes em vertebrados com o objetivo de identificar motivos proteicos específicos que compõem o 4-DSC. O alinhamento de 84 membros da família 4-DSC, incluindo proteínas WAP do leite e várias outras proteínas com atividade antiproteolítica confirmada ou putativa, levou os autores a proporem uma assinatura específica do domínio WAP, cuja posição dos oito resíduos conservados de cisteína está descrita como:

$$C_1 - (X_n) - C_2 - (X_n) - C_3 - (X_5) - C_4 - (X_5) - C_5 - C_6 - (X_3, X_5) - C_7 - (X_3, X_4) - C_8$$

, onde X representa qualquer resíduo, X_n denota um trecho de n resíduos e (X_m, X_n) representa um comprimento variável de m a n resíduos. Alguns outros resíduos são também fortemente representativos nessa assinatura e tendem a estarem presentes na grande maioria das proteínas WAP, mas para o estabelecimento de uma assinatura universal precisa para 4-DSC eles não foram contemplados no motivo proposto.

Embora a assinatura proposta por Ranganathan e colaboradores (1999) tenha sido identificada a partir de proteínas WAP de vertebrados, a estrutura do domínio WAP presente nas crustinas proposta no presente trabalho está de certa forma em conformidade com a assinatura acima, exceto por pequenas diferenças nos intervalos inter-cisteínas. De fato, como as crustinas representam um grupo único de peptídeos antimicrobianos contendo domínios WAP em invertebrados, é plausível admitir algumas dissimilaridades em relação à assinatura WAP proposta por Ranganathan e colaboradores (1999). O domínio WAP conservado na região C-terminal das crustinas as coloca como um grupo singular de AMPs em invertebrados e, mais importante, as difere de outras famílias de AMPs ricos em cisteína que carecem de um domínio WAP, como as peneidinas (DESTOUMIEUX *et al.*, 1997) e as stylicinas (ROLLAND *et al.*, 2010) presentes em crustáceos.

Nas crustinas, evidências suportam o domínio WAP como um motivo proteico essencial para a sua atividade, uma vez que alterações estruturais nesse domínio resultaram em um impacto significativo na sua atividade antimicrobiana. Zhang e colaboradores (2007) identificaram e caracterizaram duas crustinas do camarão asiático *Fenneropenaeus chinensis*, uma das quais diferiu unicamente em virtude de uma deleção de 30 resíduos aminoacídicos na região inicial do domínio WAP. Interessantemente, a então denominada CshFc mostrou um potencial antimicrobiano drasticamente inferior quando comparada à sua forma completa

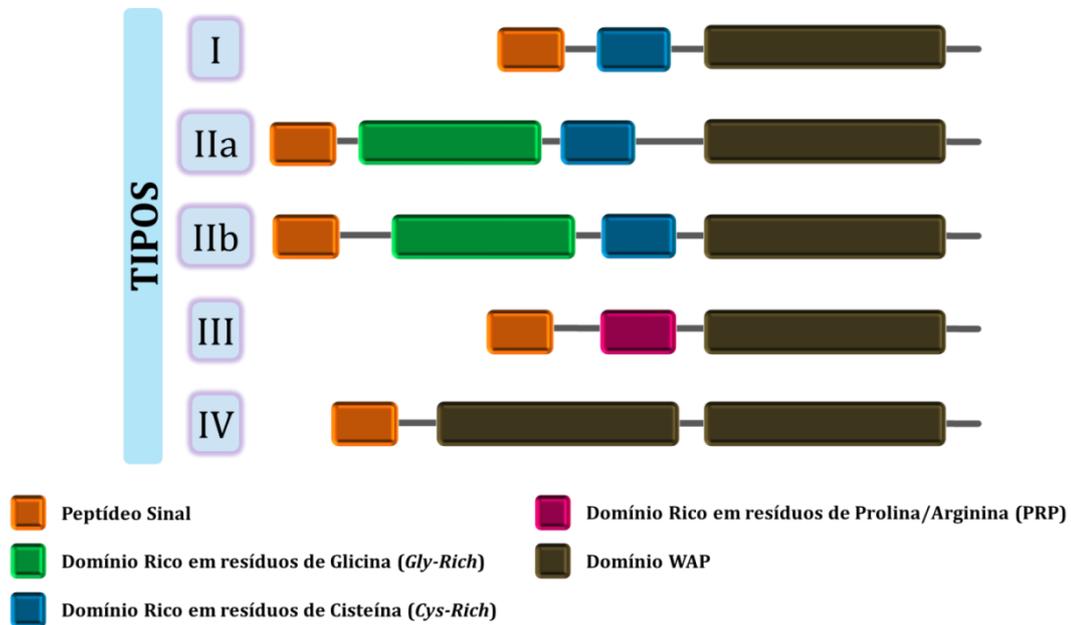
CruFc. Esses resultados destacam, portanto, a importância da composição aminoacídica do domínio WAP na função biológica das crustinas e ressaltam o impacto funcional de alterações estruturais nesse domínio.

II.3. Sistemas de classificação das crustinas

A publicação massiva de sequências similares a crustinas nas mais diversas espécies de crustáceos ao longo dos anos 2000, aliada à ausência de caracterizações funcionais dessas moléculas, levaram a alguma confusão sobre o que deve ser considerado como um membro da família crustina e, particularmente, qual a dimensão de sua diversidade. Essa complexidade crescente foi sistematicamente organizada por Smith e colaboradores em 2008 quando, pela primeira vez, propuseram um sistema objetivo de classificação das crustinas. Nesse sistema, as crustinas foram divididas em três grandes grupos, referidos como “Tipos” (Tipos I, II e III), que diferem entre si em função da presença/ausência de domínios estruturais situados na região amino-terminal do peptídeo maduro (SMITH *et al.*, 2008) (**Figura 3**).

As crustinas do Tipo I são caracterizadas essencialmente pela presença de um domínio de tamanho variado situado na região N-terminal contendo quatro resíduos conservados de cisteína (*cys-rich domain*) arranjados em posições específicas à montante do domínio WAP. Embora esses resíduos de cisteína encontrem-se invariavelmente presentes entre as crustinas do Tipo I, não há estudos mostrando seu engajamento em ligações dissulfeto. Esse subtipo de crustinas reúne sequências similares à carcinina e encontra-se presente predominantemente em crustáceos da Subordem Pleocyemata, como caranguejos, lagostas e lagostins, embora alguns estudos reportem igualmente sua presença em camarões peneídeos (DONPUDSA *et al.*, 2014; JIANG *et al.*, 2015) (**Figura 3**).

Figura 3. Representação esquemática da estrutura primária dos diferentes Tipos de crustinas descritas em Crustáceos (Tipos I a IV).



As crustinas do Tipo II, por outro lado, apresentam um domínio hidrofóbico anterior ao domínio *cys-rich* igualmente presente nas crustinas do Tipo I. Esse domínio hidrofóbico N-terminal possui normalmente entre 20 e 80 resíduos aminoácidos e é abundante em resíduos de glicina (*gly-rich domain*) (**Figura 3**). O número de resíduos de glicina pode variar consideravelmente entre diferentes espécies e tende a aparecer em blocos de quatro a cinco resíduos aminoácidos (tetra- ou pentapeptídeos, como VGGGLG) que se repetem sequencialmente. Apesar de ser uma característica de fácil detecção, nenhuma função em particular foi atribuída a esse domínio hidrofóbico estrutural, embora alguns autores hipotetizem sua participação na ancoragem desses peptídeos nas membranas celulares de suas células produtoras. Crustinas do Tipo II ocorrem mais amplamente entre camarões peneídeos (subordem Dendrobranchiata) e muito raramente em crustáceos pleociemados como no lagostim *P. leniusculus* e no caranguejo *Paralithodes camtschaticus* (JIRAVANICHPAISAL *et al.*, 2007; SPERSTAD *et al.*, 2009). Em virtude do apelo comercial da produção de camarões, as crustinas do Tipo II representam o grupo mais bem caracterizado e é, de longe, o mais diverso dentro da família. Curiosamente, reconstruções filogenéticas produzidas a partir do alinhamento de sequências aminoácidas de crustinas do Tipo II reportadas na literatura apontam para a existência de dois subtipos que diferem basicamente na distância entre o domínio *cys-rich* e o domínio WAP: as crustinas do Tipo IIa, referidas como “*Crustins*” e as crustinas do Tipo IIb referidas como “*Crustin-like*” (TASSANAKAJON;

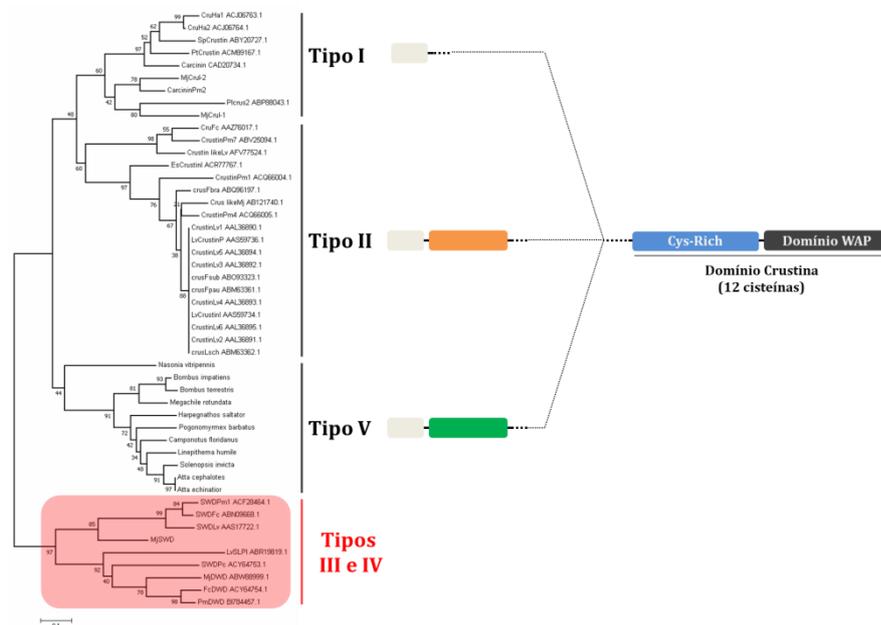
SOMBOONWIWAT; AMPARYUP, 2015) (**Figura 3**). No entanto, embora a distinção entre crustinas do Tipo IIa e IIb tenha sido previamente baseada no número de resíduos entre os dois domínios, a atual classificação desses subgrupos ainda permanece confusa em virtude da ausência de assinaturas moleculares precisas que definam cada subtipo.

As crustinas do Tipo III (também referidas como *Single WAP Domain* ou SWD) apresentam um arranjo estrutural peculiar contendo unicamente o domínio WAP em sua composição. Esse subtipo de crustinas é desprovido dos demais domínios observados nas crustinas dos Tipos I e II, ainda que seja possível encontrar em algumas espécies a presença de domínios curtos ricos em resíduos de prolina e arginina na região N-terminal (**Figura 3**). O arranjo estrutural típico das crustinas do Tipo III é igualmente encontrado em uma vasta diversidade de proteínas presentes em vertebrados, como as proteínas SWAM1 e SWAM2 (HAGIWARA *et al.*, 2003), as *nawaprinas* e *omwaprinas* descritas em serpentes (NAIR *et al.*, 2007) e o eNAP-2, um peptídeo com atividade microbicida descrita em neutrófilos de cavalos (COUTO *et al.*, 1992). Crustinas do Tipo III são igualmente encontradas em crustáceos pertencentes às subordens Pleocyemata e Dendrobranchiata, como nos camarões *L. vannamei*, *P. monodon*, *F. chinensis* e no lagostim *Procambarus clarkii* (TASSANAKAJON; SOMBOONWIWAT; AMPARYUP, 2015).

Após a proposição do que seria o primeiro sistema de classificação das crustinas, Smith e colaboradores (2011) publicaram uma revisão sobre a diversidade de proteínas WAP em metazoários e seu papel funcional em vertebrados inferiores e invertebrados. Nessa revisão, os autores discutem a recente descrição de proteínas contendo dois domínios WAP nos camarões *L. vannamei* (JIMÉNEZ-VEGA; VARGAS-ALBORES, 2007) e *M. japonicus* (CHEN; HE; XU, 2008). Essas proteínas foram inicialmente descritas com base na similaridade de sua sequência aminoacídica com a proteína SLPI presente em vertebrados e foram batizadas como “proteínas similares à SLPI” (*SLPI-like proteins*). Embora contenham um perfil estrutural marcadamente distinto daquele observado para os Tipos I a III, as *SLPI-like proteins* foram inclusas como um quarto grupo dentro das crustinas, designado, então, como Tipo IV (do inglês, *Double WAP Domain* ou DWD) (**Figura 3**). Finalmente, Tassanakajon e colaboradores (2015) adaptaram o sistema de classificação incluindo um quinto e último grupo (Tipo V). Essas crustinas foram identificadas unicamente em insetos himenópteros (formigas, vespas e abelhas) e apresentaram um perfil estrutural similar àquele visto para as crustinas do Tipo I, diferindo em virtude da presença de um domínio rico em resíduos de aminoácidos aromáticos anterior ao domínio *cys-rich* (ZHANG, Z.; ZHU, 2012). Até o presente, esse é o único grupo de crustinas encontrado fora do Subfilo Crustacea.

A fim de estabelecer a relação evolutiva entre os diferentes Tipos de crustinas postulados no sistema de classificação, reconstruções filogenéticas foram persistentemente realizadas em diversos trabalhos a partir da crescente quantidade de sequências reportadas na literatura ou depositadas em bancos de dados públicos (SMITH *et al.*, 2008; TASSANAKAJON; SOMBOONWIWAT; AMPARYUP, 2015; VARGAS-ALBORES; MARÍNEZ-PORCHAS, 2017; ZHAO; WANG, 2008). De maneira interessante, crustinas dos Tipos I, II e V formam invariavelmente um grupo monofilético, sugerindo uma origem evolutiva comum, enquanto que crustinas dos Tipos III e IV formam um grupo externo, distinto das crustinas presentes em crustáceos e insetos (**Figura 4**). Essa unidade filogenética observada entre os Tipos I, II e V mostrou estar possivelmente relacionada a uma característica marcante em sua estrutura primária: a presença de 12 resíduos conservados de cisteína distribuídos de modo específico no peptídeo maduro (quatro resíduos componentes do domínio *cys-rich* N-terminal somados aos oito resíduos do domínio WAP C-terminal) (**Figura 4**).

Figura 4. Árvore filogenética das crustinas de crustáceos e insetos.



Sequências aminoacídicas de crustinas de crustáceos e insetos descritas na literatura ou depositadas em bancos de dados públicos foram recuperadas e analisadas filogeneticamente. A arquitetura molecular típica das crustinas dos Tipos I, II e V e o domínio crustina estão representados esquematicamente. Os retângulos ilustrados representam: peptídeo sinal (cinza), domínio *gly-rich* (laranja) e domínio rico em resíduos aromáticos (verde).

Essa região, situada entre o primeiro e último resíduo de cisteína foi denominada “domínio crustina” por Zhao e Wang (2008) que desconsideraram a região *cys-rich* como um elemento de classificação para, então, incluí-la como uma parte fundamental e estruturante

deste domínio. Nessa perspectiva, a classificação das proteínas *Single WAP Domain* e *Double WAP Domain* como crustinas dos Tipos III e IV, respectivamente, torna-se controversa uma vez que essas proteínas são ausentes da região *cys-rich* componente do característico domínio crustina.

Embora o sistema de classificação de crustinas proposto por Smith e colaboradores (2008, 2011) e adaptado por Tassanakajon e colaboradores (2015) seja central na literatura atual, sistemas alternativos de classificação das crustinas foram igualmente propostos por outros autores, considerando eventuais inconsistências que emergiram com a crescente descrição de novas sequências. Zhao e Wang (2008) propuseram um sistema onde as crustinas são definidas em função da presença de um domínio crustina conservado e classificadas em três classes: classe I, classe II e carcininas. As crustinas das classes I e II seriam correspondentes às crustinas do Tipo IIa e IIb, respectivamente, e as carcininas seriam correspondentes às crustinas do Tipo I. Analogamente, Vargas-Albores e Martínez-Porchas (2017) propuseram um sistema que define a família também com base na presença de um domínio crustina conservado, mas as classifica a partir de outro critério, considerando características específicas da região *gly-rich*.

Apesar da proposição de diferentes sistemas de classificação para as crustinas, todos são essencialmente baseados na presença/ausência de domínios estruturais na região N-terminal. Esse critério tem sido amplamente utilizado devido à relação filogenética direta entre a estrutura primária e o grupo ao qual a molécula pertence. De fato, crustinas descritas como pertencentes ao Tipo I (ausentes de um domínio *gly-rich* N-terminal) de diferentes espécies de crustáceos tendem a se agrupar em um clado comum em reconstruções filogenéticas e o mesmo ocorre com as crustinas do Tipo II (Smith et al., 2008; Zhao & Wang, 2008; Tassanakajon et al., 2015; Vargas-Albores & Martínez-Porchas, 2017). No entanto, a arquitetura molecular pode representar um atributo frágil para a classificação de peptídeos antimicrobianos considerando sua vasta diversidade físico-química e de sequência, de modo que sistemas baseados em assinaturas moleculares específicas podem representar uma estratégia mais acurada para essa finalidade.

II.4. Aspectos funcionais das crustinas

Funcionalmente, as crustinas representam moléculas biologicamente diversas podendo apresentar atividades antimicrobianas e/ou antiproteolíticas (TASSANAKAJON; SOMBOONWIWAT; AMPARYUP, 2015). De fato, estudos *in vitro* utilizando proteínas

recombinantes apontam que esses peptídeos atuam contra bactérias Gram-positivas, Gram-negativas e fungos unicelulares (*Saccharomyces cerevisiae* e *Candida albicans*), embora sua atividade seja marcadamente superior contra bactérias Gram-positivas. Atividades contra vírus e protozoários são, até o presente, desconhecidas. Relatos de atividade antimicrobiana contra bactérias Gram-negativas foram reportadas unicamente em crustinas de camarões peneídeos, como a crustin $Pm7$ do camarão *P. monodon* (Tipo IIb) e a SWDFc do camarão *F. chinensis* (Tipo III) (AMPARYUP *et al.*, 2008; JIA *et al.*, 2008). Índícios da participação de crustinas na imunidade como inibidores de proteases, por outro lado, foram descritos apenas para os Tipos III e IV e parecem estar associados predominantemente ao domínio WAP. Amparyup e colaboradores (AMPARYUP; DONPUDSA; TASSANAKAJON, 2008) descreveram em *P. monodon* três isoformas de crustinas do Tipo III, uma das quais, expressa em sistema recombinante e referida como rSWDPm2, exibiu atividade antiproteolítica contra a subtilisina A, uma serino-protease isolada da bactéria *Bacillus licheniformis*. Analogamente, a crustina recombinante rSWDFc do camarão *F. chinensis* apresentou forte capacidade inibitória contra as proteases subtilisina A, tripsina e proteinase K (JIA *et al.*, 2008). Ambas rSWDPm2 e rSWDFc apresentam também atividades antibacterianas e compartilham um arranjo estrutural típico de crustinas do Tipo III contendo uma curta região N-terminal rica em resíduos de prolina e arginina.

Duas crustinas do Tipo III identificadas no camarão *M. japonicus* e no lagostim *P. clarkii*, apresentam apenas atividade antiproteolítica, sendo desprovidas de atividade antimicrobiana (DU *et al.*, 2010; JIANG *et al.*, 2013). Curiosamente, essas crustinas contêm unicamente o domínio WAP em sua composição e são ausentes de uma região rica em resíduos de prolina e arginina, sugerindo uma importante contribuição dessa região na atividade antimicrobiana dessas moléculas. Em contrapartida, até o presente, todas as crustinas do Tipo IV reportadas em camarões apresentam apenas atividades antiproteolíticas, como MjDWD de *M. japonicus*, FcDWD de *F. chinensis* e PmDWD de *P. monodon* o que indica um possível envolvimento no controle de cascatas proteolíticas das respostas de defesa (CHEN; HE; XU, 2008; DU *et al.*, 2009; SUTHIANTHONG *et al.*, 2011). Embora em camarões crustinas do Tipo IV sejam desprovidas de atividade antimicrobiana, o mesmo não vale para outras espécies de crustáceos decápodes. Li e colaboradores (2013) reportaram uma crustina do Tipo IV no caranguejo *Eriocheir sinensis* (EsDWD) com evidente potencial antimicrobiano contra bactérias Gram-positivas, Gram-negativas e leveduras. Curiosamente, EsDWD apresentou ainda forte capacidade aglutinante contra a levedura *Pichia pastoris* e a

bactéria *Bacillus subtilis* e também potencial inibitório contra proteases bacterianas endógenas e comerciais.

Embora escassos, alguns estudos foram publicados nos últimos anos investigando a participação de crustinas *in vivo* através de silenciamento pós-transcricional via RNA de interferência (RNAi) e mostram uma importante contribuição dessas moléculas na imunidade antibacteriana em decápodes. Shockey e colaboradores (2009) demonstraram o envolvimento de crustinas do Tipo II nas respostas de defesa de *L. vannamei* após infecções experimentais com a bactéria patogênica *V. penaeicida* ou com o fungo *Fusarium oxysporum*. Nesse trabalho, os autores reportaram um aumento significativo na susceptibilidade dos animais após silenciamento gênico seguido da infecção com *V. penaeicida*, mas não com o fungo *F. oxysporum* sugerindo, então, um importante papel de crustinas do Tipo II especificamente no combate a infecções bacterianas em camarões peneídeos. Abordagens similares mostraram um aumento significativo nas taxas de mortalidade de *M. japonicus* submetidos ao silenciamento pós-transcricional de crustinas do Tipo II e experimentalmente desafiados com a bactéria Gram-negativa patogênica *V. penaeicidae*, fenômeno não observado após o desafio com o vírus causador da Síndrome da Mancha Branca (HIPOLITO *et al.*, 2014). A fim de investigar a contribuição de crustinas do Tipo I na imunidade, Du e colaboradores (2019) induziram o silenciamento pós-transcricional da crustina *Pc-crustin 4* no lagostim de água doce *Procambarus clarkii* e investigaram as taxas de sobrevivência após o desafio com bactérias Gram-positivas (*Staphylococcus aureus*) e Gram-negativas (*Edwardsiella ictaluri*). A redução nos níveis transcricionais de *Pc-crustin 4* culminou em um aumento significativo na mortalidade de animais desafiados em relação aos grupos controle, indicando que essas moléculas desempenham uma função chave nas respostas de defesa antibacterianas em crustáceos decápodes.

A participação das crustinas como efetores imunológicos não foi assumida unicamente à luz de sua atividade antimicrobiana e/ou antiproteolítica, mas também porque na vasta maioria das espécies de crustáceos as crustinas são expressas pelos hemócitos circulantes, células especializadas na resposta imune (SMITH, 2011). Suleiman e colaboradores (2017) reportaram experimentalmente através de ensaios de imunolocalização e microscopia confocal a presença de carcininas, autênticas crustinas do Tipo I, majoritariamente nos hemócitos granulares de caranguejos *C. maenas* não estimulados. No entanto, diversos estudos apontam igualmente para a presença de transcritos de crustinas em outros tecidos e órgãos de animais saudáveis como brânquias, intestino, hepatopâncreas, órgãos hematopoiéticos e gônadas, sugerindo também seu possível envolvimento em funções

biológicas alternativas. De fato, considerando que as crustinas despontam como a família de AMPs mais diversa dentre os crustáceos, não é de se surpreender que essa vasta diversidade molecular se reflita em sua multifuncionalidade. Embora os papéis mais bem caracterizados das crustinas sejam suas propriedades antimicrobianas e antiproteolíticas, diversos estudos mostraram sua relevante participação em processos de regeneração e reparo tecidual (DURICA *et al.*, 2006; STOSS *et al.*, 2004; SULEIMAN; SMITH; DYRYNDA, 2017), manutenção da função antimicrobiana (CHEN *et al.*, 2019) e regulação da hematopoiese em camarões peneídeos (CHANG *et al.*, 2013). Interessantemente, a expressão de crustinas foi, ainda, correlacionada espaço-temporalmente com o estabelecimento de comunidades microbianas ectossimbiontes no camarão extremófilo de águas profundas *Rimicaris exoculata*, sugerindo uma importante participação dessas moléculas na regulação da microbiota em crustáceos (LE BLOA *et al.*, 2020).

II.5. As crustinas do camarão *Litopenaeus vannamei*

No camarão peneídeo *L. vannamei*, espécie de maior importância na carcinicultura mundial, foram reportadas crustinas do Tipo I, IIa, Tipo III e Tipo IV (BARTLETT *et al.*, 2002; JIMÉNEZ-VEGA *et al.*, 2004; LI *et al.*, 2020; VARGAS-ALBORES; JIMÉNEZ-VEGA, 2007). Através de abordagens transcritômicas, Bartlett e colaboradores (2002) encontraram pela primeira vez membros dessa família nessa espécie de camarão onde foram identificadas seis isoformas distintas, então nomeadas LvCrustin 1-6 e mais tarde classificadas como crustinas do Tipo IIa. Vargas-Albores e colaboradores (2004) descreveram, ainda, duas crustinas do Tipo IIa que diferem em sua estrutura pela substituição pontual de uma isoleucina (LvCrustin I) por uma prolina (LvCrustin P). Essas duas isoformas apresentam um perfil transcricional distinto após um desafio experimental com a bactéria *Vibrio alginolyticus*, onde a isoforma I apresenta uma redução marcante nos níveis transcricionais enquanto a isoforma P não é modulada, indicando que as duas isoformas podem participar de maneiras distintas na imunidade de *L. vannamei*.

Crustinas do Tipo III em *L. vannamei*, referidas como LvSWD, foram descritas pela primeira vez através de abordagens transcritômicas e sua participação nas respostas de defesa foi sugerida após uma infecção experimental com a bactéria Gram-negativa *V. alginolyticus*, onde um aumento nos níveis transcricionais foi observado após 3 e 6 horas do desafio bacteriano (JIMÉNEZ-VEGA *et al.*, 2004). No entanto, sua possível atuação na imunidade antibacteriana veio a ser investigada anos mais tarde quando Du e colaboradores (2015)

expressaram em sistema heterólogo uma isoforma da proteína LvSWD identificada em hemócitos e denominada LvSWDi. LvSWDi foi a primeira crustina de *L. vannamei* produzida e purificada em laboratório e apresentou atividade antiproteolítica sobre proteases bacterianas e, ainda, uma alta capacidade aderente sobre bactérias Gram-positivas e Gram-negativas. Esses resultados indicam, portanto, que a adesão sobre a superfície bacteriana pode ser um fator importante na atividade biológica das crustinas de *L. vannamei*.

As crustinas do Tipo IV de *L. vannamei*, caracterizadas em nível molecular e transcricional, foram reportadas pela primeira vez como *SLPI-like proteins* em virtude da sua estrutura primária contendo unicamente dois domínios WAP (JIMÉNEZ-VEGA; VARGAS-ALBORES, 2007). Embora crustinas do Tipo IV tenham sido descritas em outras espécies de crustáceos decápodes nos últimos anos, esse foi o primeiro relato de *SLPI-like proteins* em invertebrados, o que abriu um interessante caminho de investigação do papel dessas proteínas na imunidade desses animais. Estudos transcricionais mostraram um aumento nos níveis transcricionais de LvSLPI 3 horas após um desafio experimental com a bactéria *V. alginolyticus*, retornando aos níveis basais 24 horas após o inóculo. A modulação na transcrição desses genes pode representar um forte indício de seu papel imunofetor atuando, possivelmente, no controle proteolítico das reações de defesa.

Apesar da grande diversidade reportada na literatura para essa família de AMPs, o real repertório de crustinas em camarões peneídeos pode ser subestimado devido à escassez de trabalhos em genômica e transcritômica que explorem a diversidade dessas moléculas. Atualmente, mais de 100 sequências aminoacídicas correspondentes a “crustinas” estão disponíveis em bancos de dados públicos (<https://www.ncbi.nlm.nih.gov/protein/?term=crustin>) das quais apenas 15 foram descritas em *L. vannamei*. Recentemente, o genoma dessa espécie foi disponibilizado nas plataformas públicas permitindo o acesso à diversidade em nível gênico e a inferências acerca de sua história evolutiva. Considerando a diversidade de crustinas reportadas nas mais diversas espécies de crustáceos, é plausível admitir que *L. vannamei*, uma espécie modelo de camarão peneídeo, porte um amplo repertório ainda não devidamente explorado dessa família de AMPs e que abarque um forte potencial para aplicação em biotecnologia marinha. A crescente demanda por novas ferramentas para o controle de enfermidades nos cultivos de camarão torna, portanto, imprescindível a identificação de novas moléculas de importância na imunidade, de modo que estudos exploratórios são capazes de prover, em última análise, estratégias otimizadas de produção carcinícola e sustentabilidade.

III. OBJETIVOS

III.1. Objetivo geral

Explorar a diversidade e os processos moleculares possivelmente envolvidos na evolução de crustinas no camarão *Litopenaeus vannamei*.

III.2. Objetivos específicos

- (a) Explorar a diversidade molecular das crustinas do Tipo II em *L. vannamei* e estabelecer assinaturas moleculares específicas para os subtipos IIa e IIb em camarões peneídeos;
- (b) Caracterizar o perfil transcricional das crustinas do Tipo II de *L. vannamei* em termos de distribuição da expressão gênica, em resposta a desafios microbianos e durante o desenvolvimento;
- (c) Dimensionar o repertório completo de crustinas em *L. vannamei*;
- (d) Identificar, caracterizar e classificar as crustinas dos Tipos I e II de camarões peneídeos.

IV. METODOLOGIA

IV.1. Análises de sequências

IV.1.1. Buscas *in silico*

Dados genômicos, transcritômicos e proteômicos foram obtidos a partir dos repositórios públicos disponíveis no NCBI (*National Center for Biotechnology Information*) (<https://www.ncbi.nlm.nih.gov/genome/?term=litopenaeus+vannamei>). Sequências de crustinas foram, então, sistematicamente coletadas por uma combinação de três abordagens distintas. Primeiramente, sequências aminoacídicas (n=178) oriundas de 59 espécies de crustáceos e insetos foram obtidas a partir de bancos de dados públicos (*NCBI nr protein database*) ou recuperadas da literatura e, em seguida, alinhadas utilizando-se o software MAFFT v.7 (<https://mafft.cbrc.jp/alignment/server/>). O alinhamento resultante foi utilizado, então, para a identificação dos resíduos conservados em todas as sequências e, por conseguinte, para o estabelecimento de assinaturas aminoacídicas universais do (1) domínio WAP das crustinas e (2) “domínio crustina”, que compreende a região situada entre o primeiro e último resíduo de cisteína (região *12-cys*) do peptídeo maduro. As assinaturas moleculares foram definidas como seguem:

- $CPX_{(5,8)}CX_{(8,31)}CXXDXXCX_{(4,5)}KCCXDXXCX_{(5,6)}C$, para o domínio WAP;
- $CX_{(2,3)}CX_{(7,14)}CCX_{(4,33)}CPX_{(5,8)}CX_{(8,31)}CXXDXXCX_{(4,5)}KCCXDXXCX_{(5,6)}C$, para o domínio crustina

, onde X representa qualquer resíduo aminoacídico e os números entre parênteses (*n,m*) representam intervalos de *n* a *m* resíduos. Finalmente, o software Patmatdb v. 6.6.0.0 foi utilizado para a busca por ambas as assinaturas no proteoma predito de *L. vannamei*.

Paralelamente, sequências de crustinas foram também obtidas através da busca pelo domínio WAP no proteoma predito de *L. vannamei*, uma vez que esse domínio é ubíquo a todas as crustinas. Para tal, as buscas foram realizadas a partir de modelos matemáticos probabilísticos utilizando-se o perfil HMM (do inglês, *Hidden Markov Model*) específico do domínio WAP (ID: PF00095), publicamente disponível no banco de sequências proteicas PFAM (<https://pfam.xfam.org/>). Nesse método, as probabilidades para a presença de cada resíduo em uma determinada sequência são estimadas a partir de um alinhamento múltiplo de sequências homólogas. A frequência de cada resíduo em cada sítio estabelece, então, um sistema de “pontuação” (*score*) baseado em probabilidades: resíduos aminoacídicos conservados recebem *scores* mais elevados, uma vez que sua alta frequência reduz a probabilidade da presença de um resíduo alternativo naquela mesma posição. Em

contrapartida, resíduos aminoacídicos pouco representados recebem *scores* inferiores. Dessa maneira, é possível criar um modelo probabilístico personalizado e específico para um determinado *dataset*, o que permite a realização de buscas por sequências homólogas de maneira mais precisa comparativamente às tradicionais abordagens baseadas no uso do BLAST. Nessa abordagem, as buscas foram realizadas pela aplicação dos algoritmos *hmmsearch* e *hmmscan* disponíveis no pacote do software HMMER v. 3.1b2. Para realizar essas análises, foram considerados os parâmetros *default* estabelecidos pelo software, utilizando um valor de corte (*e-value*) de $1E^{-5}$ tanto para as sequências quanto para os domínios proteicos.

As sequências aminoacídicas obtidas por ambas as abordagens foram, em seguida, superpostas em um diagrama de Venn utilizando seus respectivos identificadores (*i.e.*, números de acesso atribuídos para cada sequência no proteoma) a fim de mapear e remover duplicatas para a obtenção de um banco único de sequências não redundantes. Sequências relativas apenas às crustinas foram, então, manualmente selecionadas à luz de: (1) presença da assinatura conservada do “domínio crustina” e (2) predição do domínio WAP C-terminal usando o programa *Conserved Domain Database* (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

As crustinas recuperadas foram utilizadas como sequências de entrada (*query*) para uma busca final por sequências similares no genoma de *L. vannamei* utilizando-se a ferramenta tBLASTn. Foram desconsideradas nessa busca sequências com porcentagem de identidade inferior a 30% e coeficientes de confiabilidade (*e-value*) superiores a 10^{-3} . Três abordagens de busca complementares foram empregadas em virtude da presunção de um alto grau de diversidade molecular de crustinas no genoma de *L. vannamei* e, por consequência, da possibilidade de obtenção do repertório incompleto dessas moléculas nessa espécie. Finalmente, os transcritos correspondentes a cada sequência aminoacídica recuperada foram obtidos através de análises de similaridade utilizando a ferramenta tBLASTn contra o transcrito predito de *L. vannamei* e validadas através de buscas nos bancos transcritômicos *Expressed Sequence Tags* (EST) e *Transcriptome Shotgun Assembly* (TSA). Para cada sequência obtida, seu respectivo gene foi mapeado no genoma.

IV.1.2. Caracterização molecular e reconstruções filogenéticas

As sequências aminoacídicas referentes às crustinas foram analisadas quanto à: (1) presença de peptídeo sinal através do programa SignalP 4.1

(<http://www.cbs.dtu.dk/services/SignalP/>) e (2) presença de um domínio WAP conservado na região C-terminal utilizando o programa *Conserved Domain Database* do NCBI. Apenas sequências completas que cumpriram os requisitos acima foram submetidas às análises seguintes. As sequências aminoacídicas obtidas foram, então, analisadas no programa *ExPasy ProtParam Tool* (<http://web.expasy.org/protparam/>) para a determinação de parâmetros bioquímicos como massa molecular e ponto isoelétrico teórico (pI) e comparadas com crustinas de outras espécies de crustáceos através do alinhamento e reconstruções filogenéticas utilizando os programas MAFFT v.7 e MEGA 7.0 (TAMURA *et al.*, 2015), respectivamente. O método estatístico empregado para a reconstrução das árvores foi o método de Máxima Verossimilhança (*Maximum likelihood*) e o teste filogenético foi realizado utilizando o Método *Bootstrap*. As árvores foram reiteradas 1.000 vezes. Para estimar a divergência entre as sequências, o *dataset* obtido foi submetido à predição do melhor modelo de substituição aminoacídica utilizando o software MEGA 7.0, sendo o modelo WAG (assumindo distribuição Gamma e considerando sítios invariáveis) aquele preferencialmente utilizado para as reconstruções filogenéticas subsequentes.

IV.2. Animais

Camarões juvenis da espécie *Litopenaeus vannamei* (10 ± 2 g) foram cedidos pelo Laboratório de Camarões Marinhos do Departamento de Aquicultura (UFSC/CCA/LCM) e mantidos no Setor de Aquários do Laboratório de Imunologia Aplicada à Aquicultura (UFSC/CCB/LIAA). Os animais foram aclimatados por um período mínimo de 7 dias em tanques contendo aproximadamente 100 L de água salgada (com aeração constante e renovação diária da água de 50%) e alimentados uma vez ao dia com ração comercial. A temperatura do ar foi mantida a 26°C, sendo que a temperatura da água permaneceu a 23°C durante os experimentos.

Amostras de camarões em diferentes estágios do desenvolvimento foram cedidas e preparadas conforme descrito por Quispe e colaboradores (2016). Cada estágio do desenvolvimento foi identificado microscopicamente e três amostras independentes de cada estágio foram coletadas. As amostras coletadas representam o começo e o fim de cada estágio do desenvolvimento: ovos fertilizados de 0-4 h (EI) e de 7-11 h pós-desova (EII), Náuplios I e V (NI e NV), Protozoas I e III (ZI e ZIII), Mysis I e III (MI e MIII) e Pós-larvas com idade de 2, 9 e 17 dias (PL2, PL9 e PL17).

IV.3. Desafio microbiano e coleta dos tecidos

Para investigar a distribuição dos transcritos em diferentes tecidos de *L. vannamei*, foi realizado um estímulo microbiano com a bactéria marinha Gram-negativa *Vibrio harveyi* ATCC 14126 inativada por calor. Brevemente, uma cultura de *V. harveyi* foi crescida em meio LB-NaCl (1% peptona, 0,5% extrato de levedura, 3% NaCl, pH 7,2) a $20 \pm 2^\circ\text{C}$ por um período de 12-16 h sob agitação, centrifugada ($1.000 \times g$ por 10 min) e suspensa em água do mar filtrada ($0,45 \mu\text{m}$). A concentração bacteriana foi, então, ajustada para a densidade óptica $\text{DO}_{600\text{nm}} = 0,52$, correspondente a 5×10^8 unidades formadoras de colônias (*Colony Forming Units* ou CFU) por mL. Em seguida, a suspensão bacteriana foi incubada a 70°C por 20 min para inativação das células e posteriormente semeadas em meio LB-NaCl ágar (LB-NaCl acrescido de 1,5% de ágar) para a confirmação do tratamento. Por fim, camarões ($n=5$) foram injetados intramuscularmente entre o primeiro e o segundo segmentos da região dorsal com $100 \mu\text{L}$ dessa suspensão ($\sim 5 \times 10^7$ CFU/animal). Como controle, foram utilizados animais não manipulados (grupo *naïve*; $n=5$).

Após 48 h da injeção, a hemolinfa dos 5 animais de cada grupo foi coletada na presença de solução anticoagulante e agrupada em uma única amostra por condição. As amostras foram centrifugadas ($1.500 \times g$ por 10 min a 4°C) e o precipitado celular (hemócitos) foi suspenso em 1 mL do reagente TRIzol[®] (Thermo Scientific[®]). Após a extração da hemolinfa, os camarões foram anestesiados por hipotermia (-20°C por 20 min) e posteriormente sacrificados para a dissecação e coleta dos seguintes tecidos/órgãos: estômago, hepatopâncreas, intestino médio, intestino posterior, músculo, brânquias e cordão nervoso. Os tecidos coletados foram lavados em solução Tris-NaCl (10 mM Tris, 330 mM NaCl, pH 7,4), macerados na presença de 1 mL de TRIzol[®] (Thermo Scientific[®]), centrifugados ($12.000 \times g$ por 10 min a 4°C) e os sobrenadantes conservados a -20°C para a posterior extração de RNA total.

IV.4. Infecções experimentais

As infecções experimentais foram realizadas com dois patógenos de interesse na carcinicultura: o vírus causador da Síndrome da Mancha Branca (WSSV) e a bactéria Gram-negativa *V. harveyi*. O inóculo viral foi preparado a partir da maceração de 10 g de tecido muscular de camarões previamente diagnosticados positivos para o WSSV e as partículas virais obtidas foram armazenadas em nitrogênio líquido para posterior uso, segundo protocolo estabelecido por Gonçalves e colaboradores (2014). Seguindo a mesma metodologia, foi preparado um inóculo livre de WSSV, utilizando tecido muscular de animais diagnosticados

como negativos para o vírus. Após um período de aclimação de 7 dias, os animais foram injetados intramuscularmente com 100 µL do inóculo viral (3×10^2 cópias genômicas/animal). Para a infecção bacteriana, os camarões foram igualmente aclimatados por 7 dias e, após, injetados intramuscularmente com 100 µL de um inóculo contendo 6×10^7 unidades formadoras de colônia (UFC) de *V. harveyi* ATCC 14126 suspensas em água do mar filtrada.

Após 48 h da injeção, os camarões infectados de cada condição foram submetidos à coleta de hemolinfa e da porção medial do intestino. Os hemócitos e os intestinos médios coletados foram homogeneizados em 1 mL do reagente TRIzol para a posterior extração de RNA total. Como controle, foram utilizados animais não manipulados (grupo *naïve*) e animais injetados com 100 µL do inóculo livre de WSSV para a infecção viral ou 100 µL de água do mar filtrada para a infecção bacteriana. Todas as amostras (grupos controle e infectado) foram preparadas na forma de 3 *pools* de 5 animais.

IV.5. Purificação de RNA total e síntese de cDNA

As amostras homogeneizadas em TRIzol foram submetidas à purificação do RNA total com clorofórmio e isopropanol. Brevemente, as amostras foram incubadas por 3 min à temperatura ambiente na presença de clorofórmio (1:5, v/v) e centrifugadas ($12.000 \times g$ por 15 min a 4°C) para a recuperação da fase aquosa contendo o RNA total. A fase aquosa foi, então, transferida para um novo microtubo e o RNA foi precipitado com isopropanol absoluto (1:1, v/v), lavado com etanol 75% e dissolvido em água livre de DNase/RNase. Em seguida, as amostras foram tratadas com a enzima DNase I (Thermo Scientific®) por 30 min a 37°C para eliminação de DNA genômico contaminante e, após o tratamento, a enzima foi inativada a 65°C por 10 min na presença de 4,5 mM EDTA, conforme instruções do fabricante. As amostras foram precipitadas com isopropanol absoluto (1:1, v/v) e acetato de sódio (300 mM, pH 5,4) e sua concentração e qualidade foi avaliada por espectrofotometria. A integridade do RNA total foi confirmada por eletroforese em gel de agarose 0,8%.

As amostras de RNA total foram, finalmente, submetidas à síntese de cDNA (DNA complementar) com a enzima RevertAid Reverse Transcriptase (Thermo Fisher Scientific). Para tal, reações de 20 µL contendo 1 µg de RNA total e 0,5 µg de oligo (dT)₁₂₋₁₈ foram incubadas a 65°C por 5 min e, em seguida, resfriadas a 4°C por 2 min. Após estas incubações, foram adicionadas às reações 200 U da enzima RevertAid Reverse Transcriptase, 0,5 mM de dNTP Mix e 20 U do inibidor de RNase RiboLock RNase Inhibitor (Thermo Scientific). As amostras foram incubadas a 42°C por 60 min e a enzima inativada a 65°C por 15 min.

IV.6. Transcrição reversa seguida de PCR semiquantitativa

A distribuição dos transcritos em diferentes tecidos de *L. vannamei* foi analisada por transcrição reversa seguida de PCR semiquantitativa (RT-PCR). Brevemente, as reações foram realizadas em um volume final de 15 μL contendo 2 mM de MgCl_2 , 0,4 mM de dNTP Mix, 0,4 μM dos iniciadores senso e antissenso (**Tabela 1**), 1 U da enzima *Taq* DNA Polymerase (Sinapse Inc[®]) e 1 μL de cDNA (diluído 10 vezes). As condições para amplificação dos fragmentos de interesse foram as seguintes: desnaturação a 95°C por 10 min, seguida de 30 ciclos de desnaturação a 95°C por 20 s, hibridização a 60°C por 20 s, extensão a 72°C por 20 s e extensão final a 72°C por 7 min. Os produtos de amplificação foram analisados por eletroforese em gel de agarose (1,5%) corados com brometo de etídeo (Amresco[®]) e seus comprimentos comparados com o padrão molecular de 100 pb (Sinapse Inc[®]). Como controle de expressão e de qualidade das amostras, foi avaliada a expressão do gene β -actina (*LvActin*) nas seguintes condições: desnaturação a 95°C por 10 min, seguida de 40 ciclos de desnaturação a 95°C por 45 s, hibridização a 50°C por 45 s, extensão a 72°C por 1 min e extensão final a 72°C por 5 min. Os géis foram visualizados por fluorescência UV e digitalizados no programa DigiDoc-ItLS.

IV.7. Transcrição reversa seguida de PCR quantitativa em tempo real

IV.7.1. Perfil transcricional frente a infecções experimentais

A determinação do perfil transcricional após as infecções experimentais foi realizada através de RT-qPCR. As reações foram realizadas utilizando o sistema StepOnePlus™ Real-Time PCR Systems (Applied Biosystems[®]) e preparadas em um volume final de 15 μL (em duplicatas) na presença de 1 μL de cDNA (diluído 10 vezes), 0,2 μM de cada iniciador (**Tabela 1**) e 7,5 μL Maxima[®] SYBR Green/ROX qPCR Master Mix 2 \times (Thermo Scientific[®]). Para a amplificação, foram utilizadas as seguintes condições: desnaturação a 95°C por 10 min, seguido por 40 ciclos de desnaturação a 95°C por 15 s e hibridização/extensão a 60°C por 1 min. A especificidade das reações foi confirmada pela análise de curvas de dissociação (60-95°C; 0,3°C/s), realizada após as etapas de amplificação. A eficiência dos iniciadores foi avaliada através da amplificação das sequências de interesse utilizando-se diluições seriadas (curva padrão: 1:8, 1:16, 1:32, 1:64, 1:128) de um *pool* de todos os cDNAs das amostras de hemócitos das infecções virais e bacterianas. Os valores de Cq (ciclo de quantificação) de cada diluição foram utilizados para a construção de uma curva

linear, cujo valor de *slope* (coeficiente angular da reta) foi utilizado para o cálculo da eficiência (E) dos pares de iniciadores segundo a fórmula: $E = 10^{(-1/\text{slope})} - 1$.

Para a normalização dos dados foram escolhidos quatro genes candidatos (*LvEF1 α* , *LvL40*, *LvRpS6* e *LvRpS3A*) e o programa RefFinder (<https://www.heartcure.com.au/reffinder>) foi utilizado para a escolha da melhor condição de referência. Os níveis transcricionais das crustinas foram calculados utilizando-se o grupo *naïve* como condição calibradora, conforme a fórmula $2^{-\Delta\Delta Cq}$ (LIVAK; SCHMITTGEN, 2001). Os dados de RT-qPCR foram submetidos ao teste de normalidade de Shapiro-Wilk e analisados utilizando o teste de análise de variância (ANOVA), seguido pelo teste de comparação de médias de Tukey. As análises estatísticas e a construção dos gráficos de expressão foram realizadas no programa GraphPad Prism 8. As diferenças entre os grupos experimentais foram consideradas significativas quando as variações de expressão gênica mostraram-se superiores a 1,5 vezes e o $p < 0,05$.

IV.7.2. Perfil transcricional durante o desenvolvimento

Os níveis transcricionais de crustinas durante o desenvolvimento foram quantificados por RT-qPCR e as reações foram preparadas e analisadas conforme descrito acima. Os dados foram normalizados com a expressão gênica dos genes de referência *LvRpS6* e *LvActin*. Amostras de hemócitos de camarão juvenil (3 *pools* de 5 animais) foram utilizadas como controle para calibração dos dados de expressão gênica. A significância estatística foi considerada em $p < 0,05$ após teste de análise de variância (ANOVA) seguida pelo teste de Tukey.

Tabela 1. Sequências dos iniciadores utilizados neste trabalho.

| Gene | Referência | Iniciadores (5' - 3') | Amplicon |
|---|---------------------------------|--|----------|
| Iniciadores para análise de distribuição nos tecidos | | | |
| <i>LvActin</i> | WANG; CHANG; CHEN, 2007 | S: TAATCCACATCTGCTGGAAGGTGG AS: TCACCAACTGGGATGACATGG | 846 pb |
| <i>CrustinLv</i> | Este trabalho | S: CGAACCAGAGACACCTGTTG AS: CAGCACACTTGTAGTCGTTG | 113 pb |
| <i>Crustin-like Lv</i> | Este trabalho | S: GCAGGATAAAGACAAGGC AS: GTAATTGCAGTTGAATCCGCC | 199 pb |
| Iniciadores para PCR quantitativa em tempo real | | | |
| <i>LvActin</i> | WANG; CHANG; CHEN, 2007 | S: CCACGAGACCACCTACAAC AS: AGCGAGGGCAGTGATTTC | 142 pb |
| <i>LvEF1α</i> | DE LA VEGA <i>et al.</i> , 2008 | S: TGGCTGTGAACAAGATGGACA AS: TTGTAGCCCACCTTCTTGACG | 103 pb |
| <i>LvL40</i> | COELHO <i>et al.</i> , 2016 | S: GAGAATGTGAAGGCCAAGATC AS: TCAGAGAGAGTGCGACCATC | 104 pb |
| <i>LvRpS6</i> | GONÇALVES <i>et al.</i> , 2014 | S: AGCAGATACCCTTGGTGAAG AS: GATGCAACCACGGACTGAC | 193 pb |
| <i>LvRpS3A</i> | DE LA VEGA <i>et al.</i> , 2008 | S: GGCTTGCTATGGTGTGCTCC AS: TCATGCTCTTGGCTCGCTG | 101 pb |
| <i>CrustinLv</i> | Este trabalho | S: CGAACCAGAGACACCTGTTG AS: CAGCACACTTGTAGTCGTTG | 113 pb |
| <i>Crustin-like Lv</i> | Este trabalho | S: GCAGGATAAAGACAAGGC AS: GTAATTGCAGTTGAATCCGCC | 199 pb |

S: iniciador senso; AS: iniciador antissenso.

V. RESULTADOS

Os resultados obtidos nesta tese de doutorado foram organizados em dois capítulos apresentados da seguinte forma:

CAPÍTULO I: Artigo científico intitulado “*Specific Molecular Signatures for Type II Crustins in Penaeid Shrimp Uncovered by the Identification of Crustin-Like Antimicrobial Peptides in Litopenaeus vannamei*”.

Esse artigo foi publicado no periódico *Marine Drugs*, em janeiro de 2018.

CAPÍTULO II: Identificação e caracterização da diversidade molecular de crustinas em *L. vannamei*: uma abordagem genômica.

RESULTADOS

CAPÍTULO 1

O primeiro capítulo desta tese foi voltado para o estudo da diversidade molecular e transcricional de crustinas do Tipo II em *L. vannamei*. As crustinas do Tipo II representam o grupo mais diverso e bem estudado em virtude de sua presença majoritária em camarões peneídeos. Crustinas do Tipo IIb, em particular, foram, até o presente, descritas em apenas quatro espécies de camarões asiáticos e representam uma pequena fração das amplamente distribuídas crustinas do Tipo II em camarões peneídeos. Nesse sentido, o presente capítulo foi destinado a identificar e caracterizar em nível molecular e transcricional crustinas do Tipo IIb em *L. vannamei*, prover análises moleculares e transcricionais comparativas com crustinas do Tipo IIa (Crustin *Lv*) previamente reportadas na literatura e, com base nas sequências disponíveis em bancos de dados públicos, estabelecer assinaturas moleculares precisas para o Tipo II em camarões peneídeos.

Specific molecular signatures for Type II crustins in penaeid shrimp uncovered by the identification of crustin-like antimicrobial peptides in *Litopenaeus vannamei*

Cairé Barreto^a, Jaqueline R Coelho^a, Jianbo Yuan^b, Jianhai Xiang^b, Luciane M Perazzolo^a, Rafael D Rosa^{a,*}

^aLaboratory of Immunology Applied to Aquaculture, Department of Cell Biology, Embryology and Genetics, Federal University of Santa Catarina, 88040-900, Florianópolis, SC, Brazil

^bKey Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

* Corresponding author. E-mail address: rafael.d.rosa@ufsc.br (R.D. Rosa).

Artigo científico publicado na revista *Marine Drugs* (ISSN 1660-3397): Barreto C, Coelho JR, Yuan J, Xiang JH, Perazzolo LM, Rosa RD (2018) Specific molecular signatures for Type II crustins in penaeid shrimp uncovered by the identification of crustin-like antimicrobial peptides in *Litopenaeus vannamei*. *Marine Drugs*, 16 (1): 31 (<https://doi.org/10.3390/md16010031>).

Abstract

Crustins form a large family of antimicrobial peptides (AMPs) in crustaceans composed of four sub-groups (Types I-IV). Type II crustins (Type IIa or “Crustins” and Type IIb or “Crustin-like”) possess a typical hydrophobic N-terminal region and are by far the most representative sub-group found in penaeid shrimp. To gain insight into the molecular diversity of Type II crustins in penaeids, we identified and characterized a Type IIb crustin in *Litopenaeus vannamei* (Crustin-like *Lv*) and compared Type II crustins at both molecular and transcriptional levels. Although *L. vannamei* Type II crustins (Crustin *Lv* and Crustin-like *Lv*) are encoded by separate genes, they showed a similar tissue distribution (hemocytes and gills) and transcriptional response to the shrimp pathogens *Vibrio harveyi* and White spot syndrome virus (WSSV). As Crustin *Lv*, Crustin-like *Lv* transcripts were found to be present early in development, suggesting a maternal contribution to shrimp progeny. Altogether, our *in silico* and transcriptional data allowed to conclude that (1) each sub-type displays a specific amino acid signature at the C-terminal end holding both the cysteine-rich region and the whey acidic protein (WAP) domain, and that (2) shrimp Type II crustins evolved from a common ancestral gene that conserved a similar pattern of transcriptional regulation.

Keywords: invertebrate immunity; host defense peptide; crustacean; WAP domain; molecular diversity; host-pathogen interaction.

1. Introduction

Crustins are cysteine-rich antimicrobial peptides (AMPs) holding a typical whey acidic protein (WAP) domain [1]. The WAP domain is a conserved motif containing eight cysteine residues, forming a characteristic four-disulfide core (4DSC) arrangement, that is also found in many proteins exhibiting protease inhibitory properties or regulatory functions in growth and tissue differentiation [2]. These bioactive molecules were originally isolated from the hemolymph of the shore crab *Carcinus maenas* as a cationic 11.5-kDa peptide with specific activity against marine Gram-positive bacteria [3]. Homologues of the *C. maenas* 11.5-kDa peptide were identified some years later in two shrimp species (*Litopenaeus vannamei* and *Litopenaeus setiferus*) and tentatively named as “Crustins” [4], a nomenclature that was subsequently accepted and adopted for this AMP family [1,5–7]. Indeed, crustins comprise a large and diverse family of gene-encoded AMPs in decapod crustaceans [8], but they are also present in non-decapod species, such as amphipods, copepods and isopods, and even in some hymenopteran insects [1].

Based on the presence or absence of structural domains lying at the N-terminal region, Smith et al. [5] classified crustins into three sub-groups, designated as Types I to III. Type I crustins are most related to the *C. maenas* 11.5-kDa peptide (later termed as “carcinin” [9]) and are characterized by the presence of four conserved cysteine residues, the cysteine-rich domain, upstream to the C-terminal WAP domain. They occur mainly in Pleocyemata (crabs, lobsters and crayfish), but some members were also reported in penaeid shrimp [10,11]. On the other hand, Type II crustins are mainly present in penaeids (Dendrobranchiata) and harbor a hydrophobic region rich in glycine residues, the glycine-rich domain, positioned at the N-terminal end of the cysteine-rich domain that is also found in Type I crustins. By contrast, Type III crustins (also known as single WAP domain-containing proteins or SWD) are only composed of a single WAP domain, and display both antimicrobial and antiprotease activities [12]. Proteins containing two WAP domains (double WAP domain-containing proteins or DWD) and crustin homologues from hymenopteran insects have been recently classified as Type IV and Type V, respectively [13].

Due to the economic importance of penaeid shrimp worldwide, Type II is the most well characterized sub-group of crustins. They are usually active against Gram-positive bacteria, but *in vivo* studies using gene silencing revealed that Type II crustins play a key role in shrimp defense against pathogenic Gram-negative bacteria [14–16]. Interestingly, the knockdown of Type II crustins in the Pacific white shrimp *L. vannamei* causes an increase in mortality after infections with the bacterial pathogen *Vibrio penaeicida*, but not in response to

the fungal pathogen *Fusarium oxysporum* [14]. The antibacterial activity of Type II crustins appears to be related to the WAP domain. Indeed, it has been shown that Type II crustins with an incomplete WAP domain have impaired antimicrobial activities [17]. This large crustin sub-group showed to be diverse in terms of both structure and function. For instance, in the black tiger shrimp *Penaeus monodon*, at least ten different Type II crustins (crustin Pm 1-10) have been identified, displaying diverse molecular structures and biological activities [7]. In the literature, this heterogeneous and complex crustin Type has been also subdivided into two sub-types, namely Type IIa or “Crustins”, most related to the first glycine-rich peptides identified in *Litopenaeus* species (Crustin Lv and Crustin Ls) [4], and Type IIb or “Crustin-like”, homologues of the Cru Fc peptide from *Fenneropenaeus chinensis* [17] and the Crus-like Pm (or crustin Pm 7) peptide from *P. monodon* [18]. Interestingly, while Type IIa crustins have been identified in different penaeids [4,19–21], Type IIb crustins were only reported in Oriental (Asian) species [16–18,22].

Although a distinction between Type IIa and Type IIb crustins has been previously established (based on differences in the amino acid length between the cysteine-rich and WAP domains) [7], the current classification of these sub-types in penaeid shrimp remains confused and controversial, leading to misleading categorizations. In order to explore and refine this issue, we have firstly identified and characterized at both molecular and transcriptional levels a Type IIb crustin from the most commonly cultivated penaeid shrimp (*L. vannamei*), and then provided amino acid signatures specific to Type IIa and Type IIb crustins in penaeid shrimp. By taking advantage of publicly accessible databases, we have identified seven nucleotide sequences in *L. vannamei* transcriptomes showing high similarities to Type IIb crustins that were opportunely named Crustin-like Lv . *L. vannamei* Type IIa and Type IIb crustins are encoded by distinct genomic sequences and differ not only in the length of their structural domains, but with each sub-type possessing a specific amino acid signature at the C-terminal region containing the 12 conserved cysteine residues. A comparative gene expression analysis showed that *L. vannamei* Type IIa (Crustin Lv) and Type IIb (Crustin-like Lv) crustins are mainly expressed in circulating hemocytes and gills of juveniles and displayed a similar transcriptional pattern in response to two unrelated shrimp pathogens, the Gram-negative *Vibrio harveyi* and the White spot syndrome virus (WSSV). Finally, as observed for Crustin Lv , Crustin-like Lv transcripts were detected in all stages of *L. vannamei* development, from fertilized eggs to larval and post larval stages. We provided here the first molecular characterization of a Type IIb crustin in an Occidental (non-Asian) species and the

identification of molecular amino acid signatures specific to Type II crustins in penaeid shrimp.

2. Materials and Methods

A metodologia deste artigo encontra-se detalhada na seção IV da tese.

3. Results and Discussion

3.1. Identification and molecular characterization of Type IIb crustins in *L. vannamei*

By taking advantage of publicly accessible databases, we have characterized for the first time Type IIb crustins in the most commonly cultivated penaeid shrimp, *L. vannamei*. *In silico* analysis led to the identification of three complete (one from the GenBank Nucleotide database and two from Expressed Sequence Tag libraries) and four incomplete/partial (Transcriptome Shotgun Assembly database) nucleotide sequences. All sequences corresponded to Type IIb crustins that were tentatively named Crustin-like *Lv. L. vannamei* Type IIa crustins were previously termed as Crustin Lv [4]. The complete full-length cDNA sequences (GenBank: JQ824114, FE049920 and FE049921) encode 147 amino acid precursors, starting with a predicted 17-residue signal peptide followed by a putative cationic (pI ~8) mature peptide of 130 amino acid residues (Figure 5). The presence of this signal peptide is necessary to direct the precursors to the endoplasmic reticulum and the Golgi apparatus. Then, the mature peptides are targeted to intracellular granules where they are stored [23]. All mature peptides possess a glycine-rich region containing three repeats of the pentapeptide Val-Phe-Pro-Gly-Ala (VFPGA) at the N-terminal end, followed by a C-terminal region containing twelve conserved cysteine residues, eight of them comprising a single WAP domain (Figure 5). For the incomplete Crustin-like *Lv* sequences (GenBank: GETD01016911, HAAW01014776, GETZ01011995 and GDUV01030786), only the C-terminal region, holding both the cysteine-rich region and the WAP domain, was identified. These incomplete sequences lack the signal peptide and part of the glycine-rich region (Figure S1).

Crustin-like *Lv* deduced amino acid sequences were compared with Type II crustins from penaeid shrimp and with crustins from other crustacean species (Figure S1). Mature Crustin-like *Lv* sequences shared a high amino acid sequence identity with other Type IIb crustins, such as Crustin-like from *Marsupenaeus japonicus* (75%; [16]), the *Fi*-crustin from

Fenneropenaeus indicus (73%;[22]), crustin*Pm7* (Crus-like*Pm*) from *P. monodon* (70%; [18]) and Cru*Fc* from *F. chinensis* (69%; [17]). Within Type IIa crustins, Crustin-like *Lv* was 44% identical to Crustin *Lv* sequences [4] and 40% to 54% identical to other Type IIa crustins from penaeids [10,19–21]. Regarding Type II crustins from non-penaeid species (Pleocyemata), Crustin-like *Lv* showed 43% to 51% amino acid identity with the crustins (PJC1 to PJC4) from the Japanese spiny lobster *Panulirus japonicus* [24]. Finally, *L. vannamei* Type IIb sequences showed less identity with the other crustins sub-groups found in crustaceans, namely Type I (30-43%), Type III (20-25%) and Type IV (25-27%).

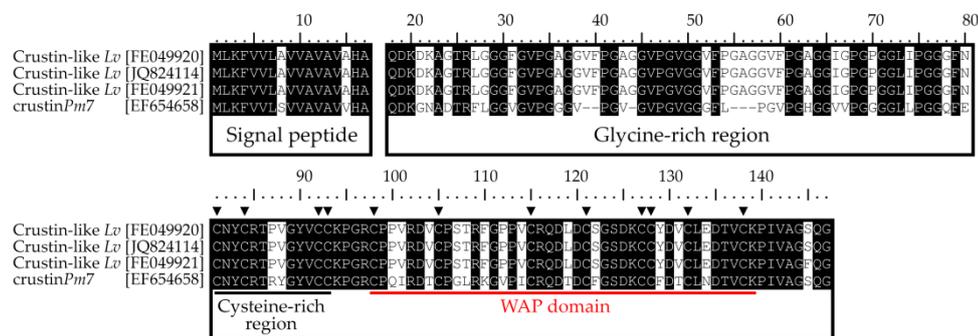


Figure 5. Litopenaeus vannamei Type IIb crustins. Amino acid sequences alignment of Type IIb crustins from *L. vannamei* (Crustin-like *Lv*) and the crustin*Pm7* from *Penaeus monodon*. Identical residues are highlighted in black. Triangles (▼) indicate the 12 conserved cysteine residues found in crustins. The whey acidic protein (WAP) domain is underlined by a solid red line.

3.2. *L. vannamei* Type II crustins are encoded by distinct genes

The *L. vannamei* Type IIa gene that encodes Crustin *Lv* is 1651 bp in size and consists of two exons (36 bp and 680 bp, respectively) interrupted by a single intron (935 bp) (Figure S2). The first exon covers the 5' untranslated region (5'-UTR) and only two codons of the signal peptide (ATGAAG), while the second exon encodes the remainder of the signal peptide sequence, and the complete mature peptide and the 3' untranslated region (3'-UTR) (Figures 2 and S2). Likewise, the Type IIb gene (742 bp) (Figure S2), that encodes Crustin-like *Lv*, is also composed of two exons (39bp and 508 bp, respectively) separated by one intron (195 bp). However, in the Type IIb gene, whereas the first exon covers only the 5'-UTR, the second exon covers the signal peptide, the complete mature peptide and the 3'-UTR (Figures 2 and S2). All splice sites in both Type II crustin genes followed the canonical GT/AG splicing recognition rule. Interestingly, *L. vannamei* Type IIb genomic organization was quite similar to that found in the crustin*Pm7*, a Type IIb crustin gene from *P. monodon* [18]. By contrast, the *P. monodon* Type IIa crustin*Pm5* gene [25] showed an exon-intron

organization different from that found in the Type IIa gene from *L. vannamei* (Figure 6). Altogether, these results suggest that penaeid Type II crustins are encoded by distinct genomic sequences, and that the molecular diversity found in this crustin sub-group was driven by gene duplication and mutations [26].

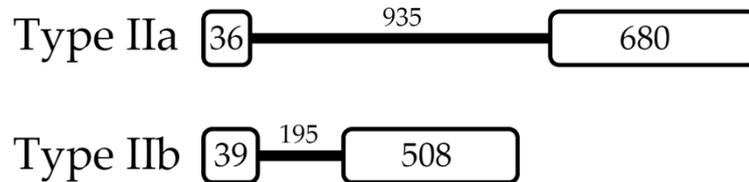


Figura 6. *Litopenaeus vannamei* Type II crustins are encoded by distinct genomic sequences. A not-to-scale representation of *L. vannamei* Type II crustins, Type IIa (Crustin *Lv*) and Type IIb (Crustin-like *Lv*). White boxes indicate the exons and the black line indicates the intron. The numbers show the size (in base pairs) of the exons and introns.

3.3. Gene expression distribution of Type II crustins in shrimp tissues

In our transcriptional analyses, we compared the *L. vannamei* Type II crustins in terms of (1) tissue distribution, (2) transcriptional response to two unrelated pathogens, and (3) expression profile during shrimp development. Firstly, the gene expression distribution of Crustin *Lv* (Type IIa) and Crustin-like *Lv* (Type IIb) were evaluated in eight different tissues (hemocytes, gills, muscle, nerve cord, foregut, hepatopancreas, midgut and hindgut) from naïve (unchallenged) animals and shrimp at 48 h post-stimulation with heat-killed bacteria (*V. harveyi*). Interestingly, both *L. vannamei* Type II crustin sub-types showed a similar gene expression distribution in shrimp tissues. Transcripts of Crustin *Lv* and Crustin-like *Lv* were detected in circulating hemocytes and gills of both unchallenged and *Vibrio*-stimulated animals (Figure 7). Remarkably, whereas the highest expression levels of the Crustin *Lv* gene (Type IIa) were found in circulating hemocytes, the Crustin-like *Lv* gene (Type IIb) showed to be predominantly expressed in gills (Figure 7). In *Vibrio*-stimulated animals, the expression of the Crustin *Lv* gene was also detected in the nerve cord and in the two gut portions (midgut and hindgut) (Figure 7). For both Type II crustins, no signals were observed in muscle, foregut, and hepatopancreas (Figure 7). Comparatively, Type II crustins from *P. monodon* were heterogeneously distributed among shrimp tissues. For instance, while crustin $Pm5$ appeared to be constitutively transcribed in the epipodite and eyestalks [25], the other *P. monodon* Type II crustins (crustin $Pm1$, crustin $Pm4$ and crustin $Pm7$) were mainly expressed in hemocytes [10,18,27].

In *L. vannamei*, the expression of Type IIa crustins was restricted to both circulating and tissue-infiltrating hemocytes [28]. More specifically, the expression of this gene was

apparently higher in semi-granular hemocytes than in the granular cells [29]. Thus, the detection of Crustin *Lv* transcripts in shrimp tissues was undoubtedly the result of infiltrating hemocytes. By contrast, in this study, Crustin-like *Lv* transcripts were mainly detected in gills. This could be the result of a distinct migratory behavior displayed by circulating hemocytes in shrimp tissues, or the presence of specific tissue-resident hemocyte populations. We cannot rule out the hypothesis of the existence of crustin-expressing hemocytes in shrimp gills distinct from those found circulating in hemolymph. Actually, the presence of tissue-specific subsets of a particular immune cell type has been reported in mammals [30]. Thus, the next import steps will be (1) to colocalize Type II crustins in shrimp hemocytes to verify whether they are expressed in the same hemocyte populations, and (2) to define the precise site of expression of the Crustin-like *Lv* gene by using physical mapping techniques (in situ hybridization and immunohistochemistry) as previously performed for other crustin genes [23,27,28].

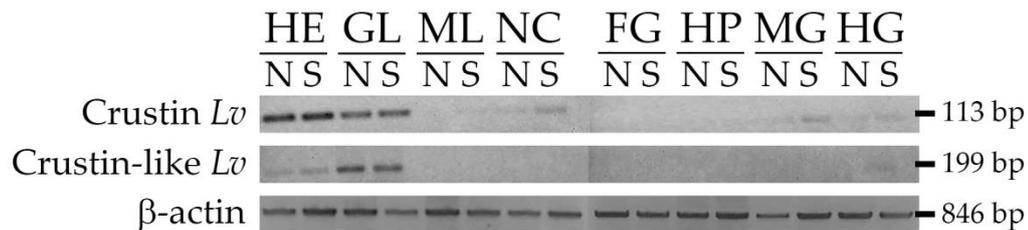


Figura 7. Gene expression distribution of Type II crustins in shrimp tissues. Semiquantitative reverse transcriptase polymerase chain reaction (RT-PCR) analysis of Crustin *Lv* (Type IIa) and Crustin-like *Lv* (Type IIb) transcript levels in different tissues from naïve (N) and *Vibrio*-stimulated (S) shrimp. The expression of the β -actin gene was used as an endogenous control. HE: hemocytes, GL: gills, ML: muscle, NC: nerve cord, FG: foregut, HP: hepatopancreas, MG: midgut, HG: hindgut.

3.4. Expression profile of Type II crustins in response to bacterial and viral infections

We have further investigated the transcriptional response of Type II crustins in two important immune tissues after experimental infections with the Gram-negative *V. harveyi* and the WSSV. Their expression levels were quantified at 48 h after infections by fluorescence-based reverse transcriptase quantitative PCR (RT-qPCR) in circulating hemocytes and midgut. In a previous study from our group, the expression of gene-encoded AMPs (penaeidins, crustins, anti-lipopolysaccharide factors and stylicins) showed to be modulated in *L. vannamei* shrimp at 48 h post-infections [31]. Importantly, whereas hemocytes are the main site for the expression of AMPs in shrimp [6], the midgut represents

an important route of pathogen entry in invertebrates [32]. The transcriptional levels of both Type II crustins were not affected by the bacterial or by the viral infection in circulating hemocytes or in the midgut (Figure 8). Furthermore, no obvious differences in gene expression were observed between unchallenged shrimp and the controls (shrimp injected with SSW or the WSSV-free inoculum) (Figure 8). At this same time point, the expression of the Crustin *Lv* gene (Type IIa) showed to be drastically reduced in circulating hemocytes of shrimp succumbing to a lethal infection by the opportunistic filamentous fungus *Fusarium solani* but not by the WSSV [33]. In our study, the lack of induction of *L. vannamei* Type II crustins could be partly attributed to the time course response of Type II crustins in the analyzed tissues but also to the route of infection. Indeed, we have evaluated the transcriptional response of *L. vannamei* Type II crustins at 48 h post-infections and it is probable that the modulation of these genes has occurred earlier. For instance, in many shrimp species, the expression of crustins showed to be modulated in the first hours post-infections [10–12,18]. Moreover, our experimental procedure consisted of injection of the pathogens directly into the shrimp hemocel in order to standardize a same inoculum load per animal [33]. By using a natural route of infection (immersion method), Soonthornchai et al. [34] could observe an increase in crustin expression in the midgut of *P. monodon* shrimp experimentally infected with *V. harveyi*.

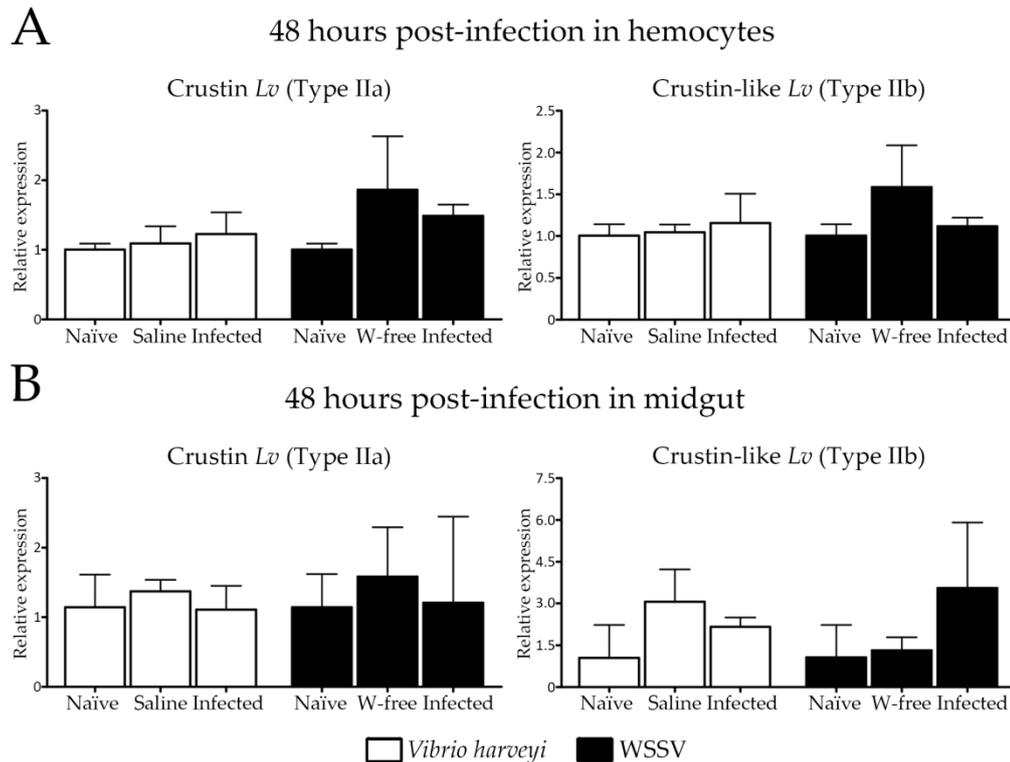


Figure 8. Relative expression profile of Crustin *Lv* (Type IIa) and Crustin-like *Lv* (Type IIb) genes in (A) circulating hemocytes and (B) midgut of shrimp at 48 h after experimental infections with the Gram-negative *Vibrio harveyi* ATCC 14126 (white bars) or the White spot syndrome virus (black bars). Results are presented as mean \pm standard deviation of relative expressions (three biological replicates). WSSV: White spot syndrome virus. W-free: tissue homogenate inoculum prepared from WSSV-free shrimp.

Despite the lack of regulation at the transcriptional level, both Type IIa and Type IIb crustins showed to be directly involved in shrimp antimicrobial defense against pathogenic *Vibrio* infections [14–16]. Besides, the regulation of these molecules may occur at the post-translational level (maturation and/or trafficking), rather than at the transcriptional level, as observed for penaeidins. Indeed, penaeidins are shrimp gene-encoded AMPs that are constitutively expressed and stored in circulating hemocytes, then released to hemolymph in response to microbial challenges [35]. As with penaeidins, crustins are also produced and stored in hemocytes [23,27], so it is plausible to suppose that they are also released towards the plasma by a regulated secretion pathway triggered by infectious agents.

3.5. Expression of Type IIb crustins during shrimp ontogenesis

Another important result from this study was the first characterization of the transcriptional profile of a Type IIb crustin during shrimp development. The transcript abundance of the Crustin-like *Lv* gene (Type IIb) was quantified by RT-qPCR in the 12 developmental stages of *L. vannamei*, from fertilized eggs to larval and post larval stages, and

also in circulating hemocytes from juveniles. Crustin-like *Lv* expression was detected in all shrimp developmental stages, but only quantified from fertilized eggs at 7–11 h post-spawning (Figure 9). Interestingly, Crustin-like *Lv* gene expression profile during shrimp ontogenesis was quite similar to that which was previously observed for the Crustin *Lv* gene [36]. Both Type II crustins were transcribed at variable levels during shrimp development. However, the highest mRNA levels were observed in hemocytes from juveniles (Figure 9), which is in accordance with our results of tissue expression distribution (Figure 7). Moreover, transcripts for both genes were found in fertilized eggs, revealing a maternal contribution of these AMPs to the shrimp offspring [36]. Interestingly, Suleiman et al. [23] detected the presence of the carcinin antimicrobial peptide (Type I crustin) at protein levels in the ovaries and oocytes of the shore crab *C. maenas*. The expression of crustins in early larval stages of *L. vannamei* suggests the participation of gene-encoded AMPs in a critical phase where the immune system has not been fully developed. Therefore, the specific role of Type II crustins during shrimp development claims for further investigations.

3.6. Molecular signatures for Type II crustins in penaeid shrimp

With the comparative transcriptional profiling of Crustin *Lv* and Crustin-like *Lv* genes in hand, we focused our attention on the molecular diversity of Type II crustins in penaeid shrimp (Decapoda: Penaeidae). Full coding sequences of Type II crustins were systematically collected from both annotated (GenBank Nucleotide database) and non-annotated (EST and TSA libraries) nucleotide databases and used for multiple-sequence alignments (Figures 10 and S1) and phylogenetic reconstructions (Figure 11). An important contribution to this analysis was undoubtedly the identification of Crustin-like sequences in an Occidental penaeid species, *L. vannamei*, since Type IIb crustins have been only reported in the Asian shrimps *P. monodon*, *F. chinensis*, *F. indicus* and *M. japonicus* [16–18,22]. However, our *in silico* mapping method failed to recover Type IIb crustin sequences in other penaeid shrimp from the Western Hemisphere (as well as shrimp from other genera), due to the lack of genomic and transcriptomic data available for those species. Moreover, it is important to point out the confused nomenclature adopted in the literature concerning the use of the terms “crustin” and “crustin-like peptides” [15,37,38], evidencing that the terminology of this AMP family need to be fully revised.

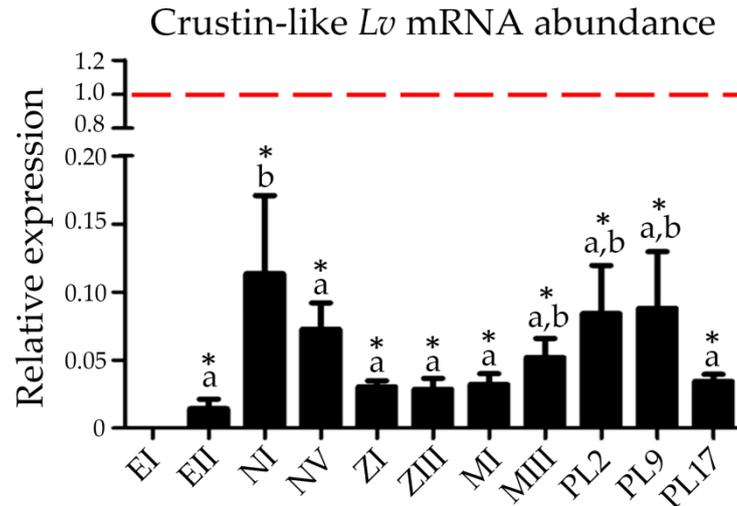


Figura 9. Relative abundance of Crustin-like *Lv* (Type IIb) transcripts during shrimp development. Results are present as mean \pm standard deviation (three biological replicates). The red dotted line indicates the basal expression level of the Crustin-like *Lv* gene in hemocytes from juvenile shrimp. EI: fertilized eggs at 0–4 h post-spawning; EII: fertilized eggs at 7–11 h post-spawning; NI: nauplius I; NV: nauplius V; ZI: protozoa I; ZIII: protozoa III; MI: mysis I; MIII: mysis III; PL2: postlarva 2; PL9: postlarva 9; PL17: postlarva 17. Different letters indicate significant differences among the developmental stages from EII to PL17 (one-way ANOVA/Tukey, $p < 0.05$). Asterisks (*) shows significant differences between each developmental stage and hemocytes from juveniles (one-way ANOVA/Tukey, $p < 0.05$)

The striking information given by the sequence analysis was the identification of amino acid signatures specific to Type IIa and Type IIb crustins in penaeid shrimp. We found a molecular pattern based on conserved amino acid residues that can successfully discriminate Type IIa and Type IIb crustins at the primary structure level (Figure 10). The classification of Type II crustins into two sub-groups (Type IIa and IIb) was initially proposed by Tassanakajon et al. [13], that is essentially based on differences in the amino acid length of the glycine-rich region and on the distance between the cysteine-rich region and the WAP domain. By contrast, multiple alignments of the amino acid sequences of all available shrimp Type II crustins revealed that the differences between Type IIa and Type IIb crustins lie in (1) the presence of specific amino acid residues positioned at precise locations of the C-terminal region holding the cysteine-rich region and the WAP domain, and (2) the arrangement (amino acid intervals) of the 12 conserved cysteine residues (Figure 10A,B). These 12 conserved cysteine residues (four from the cysteine-rich region and eight from the WAP domain) comprise a region previously termed as “crustin domain” (the crustin signature) [39] that is exclusively found in Type I (carcinins) and Type II (crustins and crustin-like peptides) crustins. On the other hand, Type III and Type IV crustins possess one and two WAP domains, respectively, but not the cysteine-rich region found in Type I and

Type II crustins [5]. Thus, both Type III (SWD) and Type IV (DWD) crustins lack the crustin signature proposed by Zhao and Wang [39].

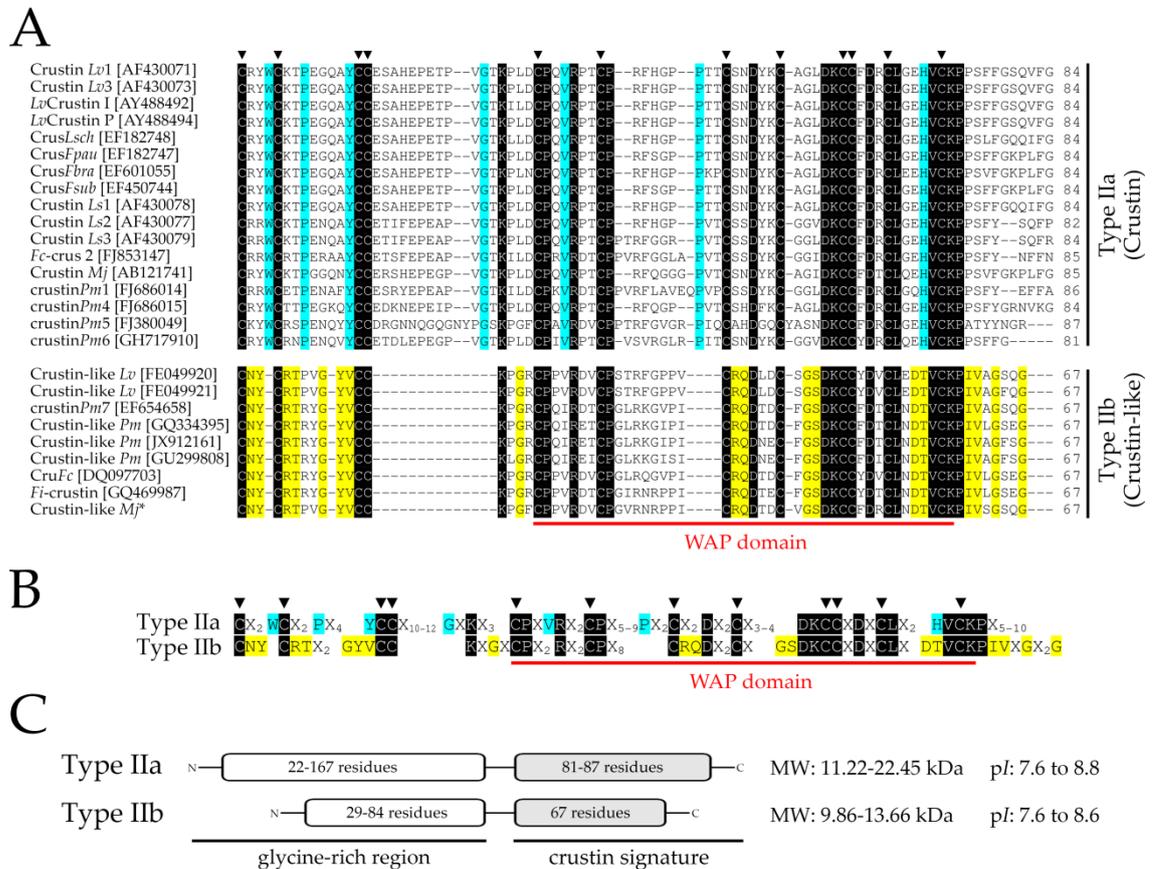


Figure 10. The molecular signature of shrimp Type II crustins. (A) Amino acid sequence alignments of the C-terminal region holding the 12 conserved cysteine residues (the crustin signature [39]) of Type IIa and Type IIb crustins from penaeid shrimp: *Litopenaeus vannamei* (Crustin Lv1, Crustin Lv3, LvCrustin I, LvCrustin P, Crustin-like Lv), *Litopenaeus setiferus* (Crustin Ls1, Crustin Ls2, Crustin Ls3), *Litopenaeus schmitti* (CrusLsch), *Farfantepenaeus paulensis* (CrusFpau), *Farfantepenaeus brasiliensis* (CrusFbra), *Farfantepenaeus subtilis* (CrusFsub), *Fenneropenaeus chinensis* (Fc-crus 2, CruFc), *Fenneropenaeus indicus* (Fi-crustin), *Penaeus monodon* (crustinPm1, crustinPm4, crustinPm5, crustinPm6, crustinPm7, Crustin-like Pm), and *Marsupenaeus japonicus* (Crustin Mj, Crustin-like Mj). (B) Consensus amino acid sequence of shrimp Type IIa and Type IIb crustins. X indicates any amino acid. Identical residues are highlighted in black. Specific amino acid residues conserved in shrimp Type IIa and Type IIb crustins are highlighted in blue and yellow, respectively. Triangles (\blacktriangledown) indicate the 12 conserved cysteine residues found in crustins. The whey acidic protein (WAP) domain is underlined by a solid red line. (C) A not-to-scale representation of shrimp Type IIa and Type IIb crustins indicating the N-terminal glycine-rich region and the C-terminal crustin signature (cysteine-rich region + WAP domain). MW: molecular weight. pI: theoretical isoelectric point. * Sequence obtained from [16] (not deposited in any database).

Type IIa crustins from penaeid shrimp contain three amino acid residues between the first two cysteine residues of the crustin signature (Cys1-Cys2) within a conserved tryptophan (Trp or W) preceding Cys2. Additionally, a sequence of 16–18 amino acids in length is found between Cys4-Cys5 and a sequence of 8–12 residues between Cys6-Cys7 (Figure 10A). Comparatively, Type IIb crustins contain an asparagine (Asn or N) and a tyrosin (Tyr or Y)

between Cys1-Cys2 and only four residues between Cys4-Cys5 (Lys-Pro/Leu-Gly-Arg/Phe) (Figure 10A). Moreover, all Type IIb crustins contain a sequence of eight residues between Cys6-Cys7 and characteristic amino acid residues in the WAPdomain that are not found in Type IIa crustins (Figure 10A). The specific amino acid signature of each sub-group of shrimp Type II crustins is presented in Figure 10B. Finally, whereas the N-terminal glycine-rich region of Type IIb crustins are composed of 29–84 amino acid residues, the glycine-rich region of Type IIa crustins can reach up to 167 residues in length (e.g., crustin*Pm4* from *P. monodon* [10]). On the other hand, whereas the amino acid length of the C-terminal region of Type IIa crustins showed to be variable (81–87 residues), the C-terminal region (holding the crustin signature) of all Type IIb crustins is composed of 67 amino acid residues (Figure 10C). Despite their differences in terms of size and molecular weight, shrimp Type II crustins are cationic peptides/polypeptides with theoretical isoelectric point (pI) ranging from 7.6 to 8.8.

Phylogenetic reconstructions clearly supported the distinction of shrimp Type II crustins into two separate groups. The obtained phylogenetic trees first showed that the four crustin Types found in crustaceans (Types I to IV) clustered into two distinct main clades: a first clade covering both Type I and Type II crustins and a second clade containing the sequences that lack the crustin signature (Type III and Type IV crustins) (Figure 11). Regarding the first clade, Type I and Type II crustins were split into distinct groups. Within the Type II group, shrimp Type IIa crustins (Dendrobranchiata) and crustins from the spiny lobster *P. japonicus* (Pleocyemata) formed a separate clade from Type IIb crustins (Figure 11). Indeed, as shrimp Type II crustins, *P. japonicus* crustins (PJC1 to PJC4) also harbor a portion rich in glycine residues at the N-terminal region [24]. From our analysis, *P. japonicus* crustins correspond to authentic Type II crustins since they hold the conserved amino acid residues found in the Sub-type IIa (Figure 10B). Moreover, while Type IIa crustins showed to be present in different decapod groups (Pleocyemata and Dendrobranchiata), Type IIb crustins appear to be exclusive of penaeid shrimp (Dendrobranchiata). Interestingly, the Type IIa crustin*Pm5* from *P. monodon* formed a distinct clade from the other Type IIa crustins from penaeid shrimp (Dendrobranchiata) and *P. japonicus* (Pleocyemata). Indeed, crustin*Pm5* is a unique crustin member within the Type IIa clade, apparently exclusive of the black tiger shrimp *P. monodon*, that is diverse not only in terms of sequence, but also in terms of genomic organization and tissue expression distribution [25]. Both sequence and phylogenetic data showed that Type IIa crustins display a high degree of diversity at both inter- and intraspecific levels when compared to Type IIb crustins. For instance, Type IIb crustins from

distant shrimp species have been shown to be more similar to each other than Type IIa crustins from the same shrimp. This is the case of *P. monodon* Type IIa crustins that are diverse not only in terms of sequence, but also in genomic organization and gene expression regulation [10,27].

Why Type IIb crustins are quite similar among distant shrimp species is still an open question. On one hand, Type IIb crustins could be recently evolved from an ancestor shrimp Type II gene and, on the other hand, the interspecific conservation of Type IIb crustin sequences could be the result of selective pressures (environmental stressors, pathogens, etc.). In any cases, phylogenetic and maximum likelihood-based codon substitution analyses have shown that diversity in crustin family has been driven by successive gene duplications and positive Darwinian selection [26]. Our transcriptional data revealed that *L. vannamei* Type II crustins display a very similar pattern of gene expression in terms of tissue distribution, transcriptional response to pathogens, and mRNA abundance during shrimp development. Taken together, it is plausible to hypothesize that Type IIa and Type IIb crustins evolved by gene duplication from a common ancestral gene that conserved a similar pattern of transcriptional regulation. The apparent lack of Type IIb crustin sequences in available databases of other crustacean groups suggests that the gene duplication event relative to Type IIa–Type IIb divergence occurred exclusively in the Dendrobranchiata lineage.

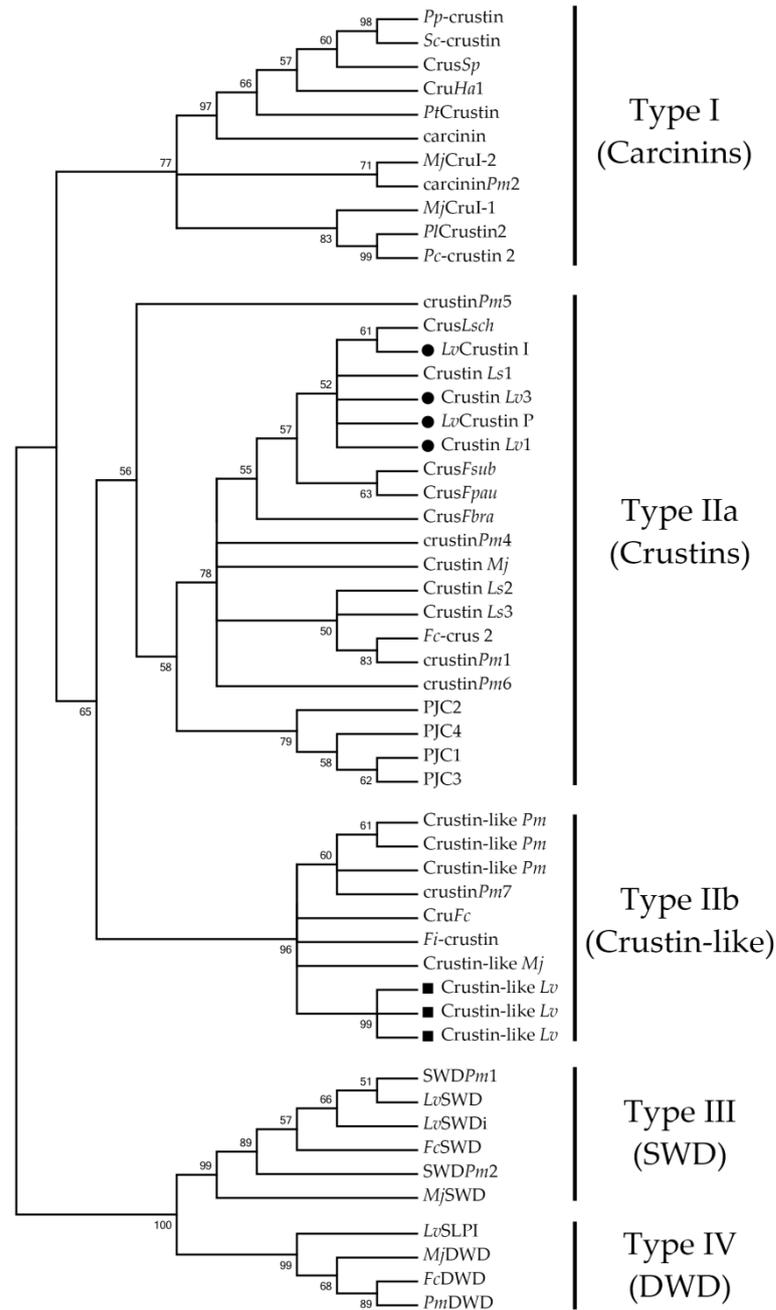


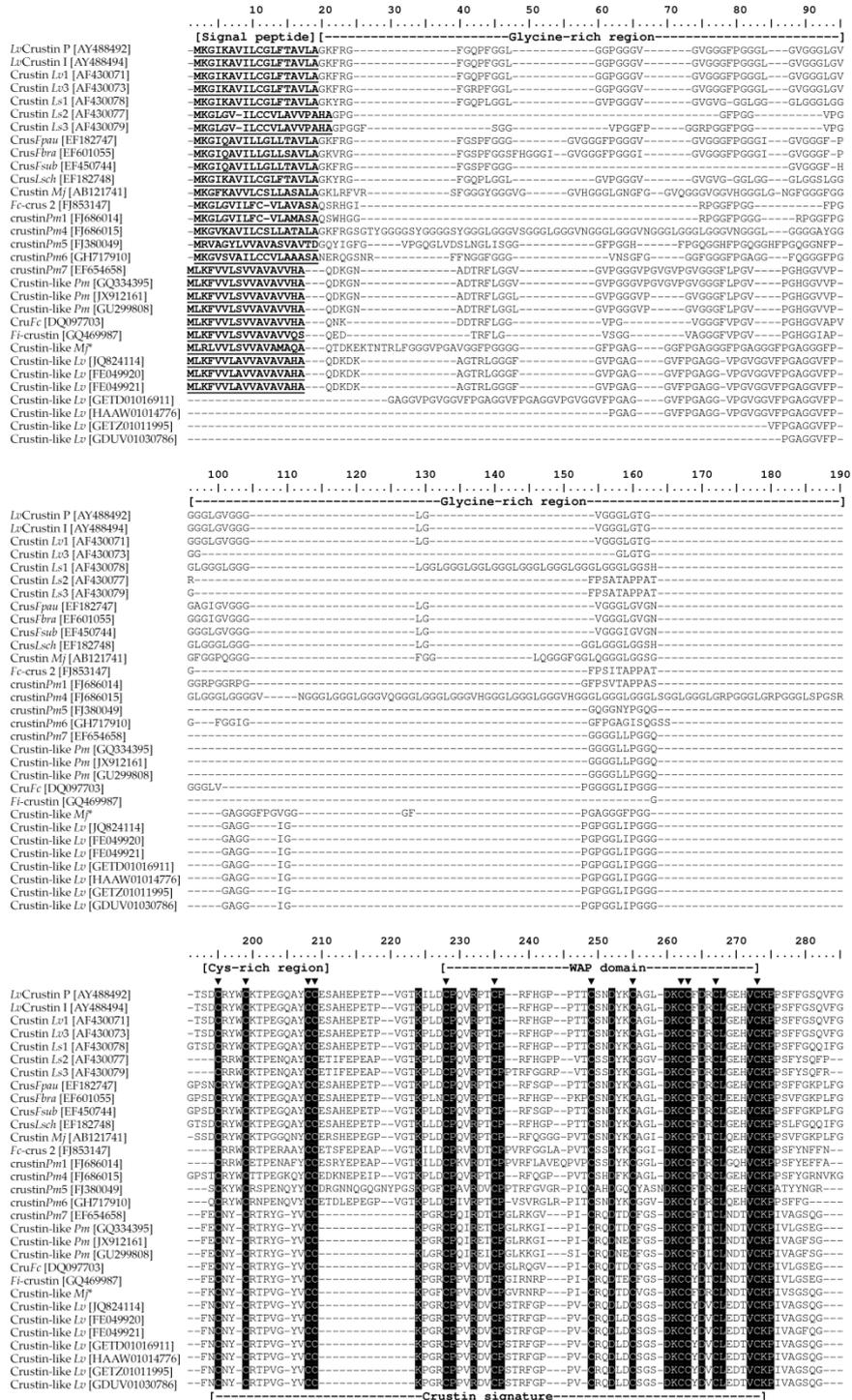
Figure 11. Penaeid shrimp Type II crustins form two distinct phylogenetic clades. The phylogenetic tree was constructed using the Neighbor-Joining method in MEGA 6. Bootstrap sampling was reiterated 1000 times. Sequences included in analyses were the following: (i) Type I crustins (“Carcinins”): the *Pp*-crustin from *Portunus pelagicus* (GenBank: JQ965930), the *Sc*-crustin from *Scylla serrata* (GenBank: HQ638025), the *CrusSp* from *Scylla paramamosain* (GenBank: EU161287), the *CruHa1* from *Hyas araneus* (GenBank: EU921641), *PtCrustin* from *Portunus trituberculatus* (GenBank: FJ612106), the carcinin (11.5-kDa peptide) from *Carcinus maenas* (GenBank: AJ427538), the *PtCrustin2* from *Pacifastacus leniusculus* (GenBank: EF523613), the *Pc*-crustin 2 from *Procambarus clarkii* (GenBank: GQ301202), the *MjCruI-1* and *MjCruI-2* from *Marsupenaeus japonicus* [11], and the carcinin*Pm2* from *Penaeus monodon* [10]; (ii) Type IIa crustins (“Crustins”): the *LvCrustin I* (GenBank: AY488492), *LvCrustin P* (GenBank: AY488494), *Crustin Lv1* (GenBank: AF430071) and *Crustin Lv3* (GenBank: AF430073) from *Litopenaeus vannamei* (indicated by black circles, ●), the *Crustin Ls1* (GenBank: AF430078), *Crustin Ls2* (GenBank: AF430077) and *Crustin Ls3* (GenBank: AF430079) from *Litopenaeus setiferus*, the *CrusLsch* (GenBank: EF182748) from *Litopenaeus schmitti*, the *CrusFpau* (GenBank: EF182747) from *Farfantepenaeus paulensis*, the *CrusFbra* (GenBank: EF601055) from *Farfantepenaeus brasiliensis*, the *CrusFsub* (GenBank: EF450744) from *Farfantepenaeus subtilis*, the *Crustin Mj* (GenBank: AB121741) from *M. japonicus*, the crustin*Pm1* (GenBank: FJ686014), crustin*Pm4* (GenBank: FJ686015), crustin*Pm5* (GenBank: FJ380049) and crustin*Pm6* (GenBank: GH717910) from *P. monodon*, the *Fc*-crus 2

(GenBank:FJ853147) from *Fenneropenaeus chinensis* and the PJC1 (GenBank: FJ797417), PJC2 (GenBank: FJ797418), PJC3 (GenBank: FJ797419), and PJC4 (GenBank: FJ797420) from *Panulirus japonicus*; (iii) Type IIb crustins (“Crustin-like”): the Crustin-like *Lv* (GenBank: JQ824114, FE049920 and FE049921) from *L. vannamei* (indicated by black squares, ■), the Crustin-like *Mj* from *M. japonicus* [16], the *CruFc* (GenBank: DQ097703) from *F. chinensis*, the *Fi*-crustin (GenBank: GQ469987) from *Fenneropenaeus indicus* and the Crustin-like *Pm* (GenBank: GQ334395, JX912161 and GU299808) and crustin*Pm7* (GenBank: EF654658) from *P. monodon*; (iv) Type III crustins (“Single WAP domain-containing proteins or SWD”): the *LvSWD* (GenBank: AY464465) and *LvSWDi* [40] from *L. vannamei*, the *SWDPm1* (GenBank: EU623979) and *SWDPm2* (GenBank: EU623980) from *P. monodon*, the *FcSWD* (GenBank:EF216349) from *F. chinensis* and the *MjSWD* (GenBank: AU176270) from *M. japonicus*; (v) Type IV crustins (“Double WAP domain-containing proteins or DWD”): the *LvSLPI* (GenBank: EF467169) from *L. vannamei*, the *PmDWD* (GenBank: BI784457) from *P. monodon*, the *FcDWD* (GenBank: GQ303571) from *F. chinensis*, and the *MjDWD* (GenBank: EU095018) from *M. japonicus*.

4. Conclusions

In conclusion, we have identified for the first time a Type IIb crustin (Crustin-like *Lv*) in an Occidental (non-Asian) penaeid, the Pacific white shrimp *L. vannamei*. The molecular characterization and comparison of Crustin-like *Lv* with other crustin sequences allowed to the identification of molecular amino acid signatures specific to Type II crustins in penaeid shrimp. Despite their similar patterns of gene expression, Type IIa and Type IIb form two distinct groups of crustins in penaeid shrimp with potential specific biological activities. Results from the *in silico* and phylogenetic analyses combined with our gene expression data suggested that shrimp Type II crustins evolved from a common ancestral gene that conserved a similar pattern of transcriptional regulation. The functional implication of the molecular diversity of Type II crustins in the shrimp immune response is an important task to be assessed by reverse genetic approaches. These bioactive molecules from marine invertebrates could find applications not only in aquaculture and shrimp farming, but also in both human and veterinary medicine.

Supplementary Data



*Sun et al. Fish Shellfish Immunol. 2017 Jul;66:487-496.

Figure S1. Amino acid sequence alignments of shrimp Type II crustins: *Litopenaeus vannamei* (LvCrustin P, LvCrustin I, Crustin Lv1, Crustin Lv3, Crustin-like Lv), *Litopenaeus setiferus* (Crustin Ls1, Crustin Ls2, Crustin Ls3), *Farfantepenaeus paulensis* (CrusFpau), *Farfantepenaeus brasiliensis* (CrusFbra), *Farfantepenaeus subtilis* (CrusFsub), *Litopenaeus schmitti* (CrusLsch), *Marsupenaeus japonicus* (Crustin Mj, Crustin-like Mj), *Fenneropenaeus chinensis* (Fc-crus 2, CruFc), *Penaeus monodon* (crustinPm1, crustinPm4, crustinPm5, crustinPm6, crustinPm7), Crustin-like Pm, and *Fenneropenaeus indicus* (Ficrustin). The predicted signal peptides are in bold and underlined. Identical amino acid residues are shadowed with black backgrounds. Triangles (▼) indicate the 12 conserved cysteine residues found in crustins. * Sequence obtained from [16] (not deposited in any database).

A

Crustin *Lv* gene (Type IIa)

GCACACCACTAGCTTGTACTGGAGGCAACCATGAAGGTATGAGCTTGTGTTTTGTTTTCGT 60

M K

TGACTTCCGGTCTTTTTTTTTTTTTTTTTTCTCTCTTTTTTTTTTAAGTGAGAGAGTTTAA 120
 TGTAAGGAGTTTTTTTTTAATGTTTTTCTTTGTGTACACCGTTTTAAATATCATGTTTTG 180
 AGTCATGATATATACAAATGCGAGATAGTTATTGTTGTGAATGTACTTCCCCTTTCGCGG 240
 GATATTTATAGATTTGACTATCAAATGTGCATTTTCGGATTAAAAAAGATTTTTTAGTCCAA 300
 AATGCAAACCTTCGGTGTAAATCTATAAATAGTATTATTGCGTGATGGTAACGTAAATAC 360
 ACTAAAGATTAAAGCTTATAATCATTCAATTTTGCATGTCTTCGCCACGAAAATTTGCA 420
 ACGAAAATGTCATATACAAACATTTTAGGCTATGCACTGATGTTAGGCATACATATTATT 480
 GTGCATAGCCAGGTTGGATAAAAAACGTTTACAAGATGAAAGGCTTAATCTAGATCCAAGC 540
 TATTACGGTACAATATATTCCGCATTTTCGAAAAGGTGATTA AAAAAGAGATTTTCATTATTGA 600
 ATGCAAAAAGACCATCACACAGGTATTATGGCTTTCTCTCTGTATACAGCATTTTTTTTCCC 660
 TAAAGCTTCCCTAAAAATACCGACGTTACGGTAACAGTAACGTTAAGCAATATAGATTGAT 720
 ATAGAACAAATCGAAATGACGGAAAGTTGATATATAACAATAACACGATAAAAGATTGTTTG 780
 TCTTAAGAATAATCTTTCTCCCGTTTGTCCCACTGCTCAGACTTATAAAAAAGCCATTGG 840
 ACTTTCTCAACAATACAGCAGGGACTCCTTTAACACTGACAACGTTTCCGAACAATGCAA 900
 TGACTTTAATTTCTATTTTACACCAGAGTAATTTCTATTTTTTAATGCTATTTTTTCAATCG 960
 TTCTCTTTTCAGGGCATCAAGGCGGTGATTCTGTGCGGCCTCTTTACGGCGGTTTTGGCTG 1020

G I K A V I L C G L F T A V L A

GCAAGTTTCGCGGCTTCGGACAGCCATTTGGAGGTCTGGGTGGTCCAGGAGGCAGTGTAG 1080
 G K F R G F G Q P F G G L G G P G G S V
 GTGTTGGTGGTGGTTTCCCGGAGGCGGTTTAGGTGTAGGTGGTGGTCTTGGTGTAGGTG 1140
 G V G G G F P G G G L G V G G G L G V G
 GCGGTCTTGGTGTAGGTGGCGGTCTTGGTGTAGGTGGCGGTCTTGGAAGTGGCACAAGCG 1200
 G G L G V G G G L G V G G G L G T G T S
 ACTGCAGGTATTGGTGCAAGACTCCGGAGGGTCAAGCCTACTGCTGCGAGTCGGCCCACG 1260
 D C R Y W C K T P E G Q A Y C C E S A H
 AACCAGAGACACCTGTTGGCACCAAGCCACTCGACTGCCCAAGTCCGTCCCACATGCC 1320
 E P E T P V G T K P L D C P Q V R P T C
 CACGTTTCCATGGGCCCCCACAACCTGTTCCAACGACTACAAGTGTGCTGGCCTCGATA 1380
 P R F H G P P T T C S N D Y K C A G L D
 AGTGTGCTTCGACAGGTGTTTGGGAGAACACGTGTGCAAGCCTCCCTCATTCTTCGGAT 1440
 K C C F D R C L G E H V C K P P S F F G
 CGCAGGTTTTTCGGATGAAGAATAAGCACGAAAGAATTTGAAAGGATGAAGAGAAAGA 1500
 S Q V F G -
 AAAGACCATCTGAAGAACGACCGATGTTTTGGAATTTGACTGAAAAAAGAAAGAAAAACA 1560
 GGGAAATCTTTCTTTCTGTAGGATTTATCTGATTAACATGATTTTTGTTATATGTGAATT 1620
 AGACTATTCTTCTGTCAAAAAGAACTTATAG 1651

B

Crustin-like *Lv* gene (Type IIb)

```

CAGTCCGTTTCGTCGCAGAGCAAGGAGAAGTATTACAAAGGTGAGGAGGAGAAGCCAACAT 60
TACGAGGCTATGCTTTGAGTTACCAGCTTTGGTTGTCCAACGTGTTTCTTTAAAGTCATAG 120
AAACTCTTACCAATGATCGTGTCAATCACTAGGAGACAAAATGCTGTCCTGTGACAATTC 180
AAAGCTCAAATGTCACTCCATGATCTTAATTCCGCGTTCGTTTCTTCCTTGTAGATGCTG 240
                                                                 M L
AAGTTTGTAGTATTAGCCGTTGTCGCCGTGGCCGTGGCGCACGCGCAGGATAAAGACAAG 300
K F V V L A V V A V A V A H A Q D K D K
GCCGGCACTCGCTTAGGAGGAGGATTCGGGGTTCCTGGAGCCGGTGGCGTCTTCCCAGGA 360
A G T R L G G G F G V P G A G G V F P G
GCCGGTGGCGTCCCTGGAGTAGGTGGCGTCTTTCCTGGAGCCGGTGGCGTCTTCCCTGGA 420
A G G V P G V G G V F P G A G G V F P G
GCCGGTGGTATCGGTTCCTGGACCCGGCGGCCTCATCCCCGGAGGCGGATTCAACTGCAAT 480
A G G I G P G P G G L I P G G G F N C N
TACTGCAGGACGCCCGTTCGGGTACGTCTGCTGCAAGCCCGGTAGGTGCCCTCCGGTTCGA 540
Y C R T P V G Y V C C K P G R C P P V R
GACGTCTGCCCGTTCGACCCGCTTCGGACCCCGGTCTGCCGCCAGGACCTGGACTGCTCC 600
D V C P S T R F G P P V C R Q D L D C S
GGCTCCGACAAGTGCTGCTATGACGTCTGCCTGGAAGACACAGTCTGCAAACCCATCGTG 660
G S D K C C Y D V C L E D T V C K P I V
GCAGGTTCACAGGGATAAGCCTGCATGTGAACTTATCAAGCCTTCGTTATCAAATAAAT 720
A G S Q G -
GCTATAACTGTTAATTGTAAAT 742

```

Figure S2. Genomic nucleotide and deduced amino acid sequences of Type IIa Crustin *Lv* (A) and Type IIb Crustin-like *Lv* (B). The predicted signal peptides are in bold and underlined. The exon and intron sequences are shown in black and grey, respectively. A dash (-) marks the stop codon.

References

1. Smith, V.J.; Dyrzynda, E.A. Antimicrobial proteins: From old proteins, new tricks. *Mol. Immunol.* 2015, 68, 383–398.
2. Ranganathan, S.; Simpson, K.J.; Shaw, D.C.; Nicholas, K.R. The whey acidic protein family: A new signature motif and three-dimensional structure by comparative modeling. *J. Mol. Graph. Model.* 1999, 17, 106–113.
3. Relf, J.M.; Chisholm, J.R.; Kemp, G.D.; Smith, V.J. Purification and characterization of a cysteine-rich 11.5-kDa antibacterial protein from the granular haemocytes of the shore crab, *Carcinus maenas*. *Eur. J. Biochem.* 1999, 264, 350–357.
4. Bartlett, T.C.; Cuthbertson, B.J.; Shepard, E.F.; Chapman, R.W.; Gross, P.S.; Warr, G.W. Crustins, homologues of an 11.5-kDa antibacterial peptide, from two species of penaeid shrimp, *Litopenaeus vannamei* and *Litopenaeus setiferus*. *Mar. Biotechnol.* 2002, 4, 278–293.
5. Smith, V.J.; Fernandes, J.M.O.; Kemp, G.D.; Hauton, C. Crustins: Enigmatic WAP domain-containing antibacterial proteins from crustaceans. *Dev. Comp. Immunol.* 2008, 32, 758–772.
6. Rosa, R.; Barracco, M. Antimicrobial peptides in crustaceans. *Invertebr. Surviv. J.* 2010, 7, 262–284.
7. Tassanakajon, A.; Rimphanitchayakit, V.; Visetnan, S.; Amparyup, P.; Somboonwiwat, K.; Charoensapsri, W.; Tang, S. Shrimp humoral responses against pathogens: Antimicrobial peptides and melanization. *Dev. Comp. Immunol.* 2017.
8. Destoumieux-Garzón, D.; Rosa, R.D.; Schmitt, P.; Barreto, C.; Vidal-Dupiol, J.; Mitta, G.; Gueguen, Y.; Bachère, E. Antimicrobial peptides in marine invertebrate health and disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2016.
9. Brockton, V.; Hammond, J.; Smith, V. Gene characterisation, isoforms and recombinant expression of carcinin, an antibacterial protein from the shore crab, *Carcinus maenas*. *Mol. Immunol.* 2007, 44, 943–949.
10. Donpuksa, S.; Visetnan, S.; Supungul, P.; Tang, S.; Tassanakajon, A.; Rimphanitchayakit, V. Type I and type II crustins from *Penaeus monodon*, genetic variation and antimicrobial activity of the most abundant crustinPm4. *Dev. Comp. Immunol.* 2014, 47, 95–103.
11. Jiang, H.-S.; Jia, W.-M.; Zhao, X.-F.; Wang, J.-X. Four crustins involved in antibacterial responses in *Marsupenaeus japonicus*. *Fish Shellfish Immunol.* 2015, 43, 387–395.
12. Amparyup, P.; Donpuksa, S.; Tassanakajon, A. Shrimp single WAP domain (SWD)-containing protein exhibits proteinase inhibitory and antimicrobial activities. *Dev. Comp. Immunol.* 2008, 32, 1497–1509.
13. Tassanakajon, A.; Somboonwiwat, K.; Amparyup, P. Sequence diversity and evolution of antimicrobial peptides in invertebrates. *Dev. Comp. Immunol.* 2015, 48, 324–341.
14. Shockey, J.E.; O’Leary, N.A.; de la Vega, E.; Browdy, C.L.; Baatz, J.E.; Gross, P.S. The role of crustins in *Litopenaeus vannamei* in response to infection with shrimp pathogens: An in vivo approach. *Dev. Comp. Immunol.* 2009, 33, 668–673.
15. Hipolito, S.G.; Shitara, A.; Kondo, H.; Hirono, I. Role of *Marsupenaeus japonicus* crustin-like peptide against *Vibrio penaeicida* and white spot syndrome virus infection. *Dev. Comp. Immunol.* 2014, 46, 461–469.
16. Sun, B.; Wang, Z.; Zhu, F. The crustin-like peptide plays opposite role in shrimp immune response to *Vibrio alginolyticus* and white spot syndrome virus (WSSV) infection. *Fish Shellfish Immunol.* 2017, 66, 487–496.
17. Zhang, J.; Li, F.; Wang, Z.; Xiang, J. Cloning and recombinant expression of a crustin-like gene from Chinese shrimp, *Fenneropenaeus chinensis*. *J. Biotechnol.* 2007, 127, 605–614.
18. Amparyup, P.; Kondo, H.; Hirono, I.; Aoki, T.; Tassanakajon, A. Molecular cloning, genomic organization and recombinant expression of a crustin-like antimicrobial peptide from black tiger shrimp *Penaeus monodon*. *Mol. Immunol.* 2008, 45, 1085–1093.

19. Supungul, P.; Klinbunga, S.; Pichyangkura, R.; Jitrapakdee, S.; Hirono, I.; Aoki, T.; Tassanakajon, A. Identification of immune-related genes in hemocytes of black tiger shrimp (*Penaeus monodon*). *Mar. Biotechnol.* 2002, 4, 487–494.
20. Rosa, R.D.; Bandeira, P.T.; Barracco, M.A. Molecular cloning of crustins from the hemocytes of Brazilian penaeid shrimps. *FEMS Microbiol. Lett.* 2007.
21. Sun, C.; Du, X.-J.; Xu, W.-T.; Zhang, H.-W.; Zhao, X.-F.; Wang, J.-X. Molecular cloning and characterization of three crustins from the Chinese white shrimp, *Fenneropenaeus chinensis*. *Fish Shellfish Immunol.* 2010, 28, 517–524.
22. Antony, S.P.; Bright Singh, I.S.; Philip, R. Molecular characterization of a crustin-like, putative antimicrobial peptide, *Fi*-crustin, from the Indian white shrimp, *Fenneropenaeus indicus*. *Fish Shellfish Immunol.* 2010, 28, 216–220.
23. Pisuttharachai, D.; Fagutao, F.F.; Yasuike, M.; Aono, H.; Yano, Y.; Murakami, K.; Kondo, H.; Aoki, T.; Hirono, I. Characterization of crustin antimicrobial proteins from Japanese spiny lobster *Panulirus japonicus*. *Dev. Comp. Immunol.* 2009, 33, 1049–1054.
24. Vatanavicharn, T.; Supungul, P.; Puanglarp, N.; Yingvilasprasert, W.; Tassanakajon, A. Genomic structure, expression pattern and functional characterization of crustin *Pm5*, a unique isoform of crustin from *Penaeus monodon*. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* 2009, 153, 244–252.
25. Padhi, A. Adaptive evolution of crustin antimicrobial peptides in decapods. *Genetica* 2012, 140, 197–203.
26. Supungul, P.; Tang, S.; Maneeruttanarungroj, C.; Rimphanitchayakit, V.; Hirono, I.; Aoki, T.; Tassanakajon, A. Cloning, expression and antimicrobial activity of crustin *Pm1*, a major isoform of crustin, from the black tiger shrimp *Penaeus monodon*. *Dev. Comp. Immunol.* 2008, 32, 61–70.
27. Wang, Y.-C.; Chang, P.-S.; Chen, H.-Y. Tissue expressions of nine genes important to immune defence of the Pacific white shrimp *Litopenaeus vannamei*. *Fish Shellfish Immunol.* 2007, 23, 1161–1177.
28. Yang, C.-C.; Lu, C.-L.; Chen, S.; Liao, W.-L.; Chen, S. Immune gene expression for diverse haemocytes derived from pacific white shrimp, *Litopenaeus vannamei*. *Fish Shellfish Immunol.* 2015, 44, 265–271.
29. Davies, L.C.; Taylor, P.R. Tissue-resident macrophages: Then and now. *Immunology* 2015, 144, 541–548.
30. Suleiman, S.; Smith, V.J.; Dyrinda, E.A. Unusual tissue distribution of carcinin, an antibacterial crustin, in the crab, *Carcinus maenas*, reveals its multi-functionality. *Dev. Comp. Immunol.* 2017, 76, 274–284.
31. Saraiva, R.G.; Kang, S.; Simões, M.L.; Angleró-Rodríguez, Y.I.; Dimopoulos, G. Mosquito gut antiparasitic and antiviral immunity. *Dev. Comp. Immunol.* 2016, 64, 53–64.
32. Goncalves, P.; Guertler, C.; Bachère, E.; de Souza, C.R.B.; Rosa, R.D.; Perazzolo, L.M. Molecular signatures at imminent death: Hemocyte gene expression profiling of shrimp succumbing to viral and fungal infections. *Dev. Comp. Immunol.* 2014.
33. Soonthornchai, W.; Rungrassamee, W.; Karoonuthaisiri, N.; Jarayabhand, P.; Klinbunga, S.; Söderhäll, K.; Jiravanichpaisal, P. Expression of immune-related genes in the digestive organ of shrimp, *Penaeus monodon*, after an oral infection by *Vibrio harveyi*. *Dev. Comp. Immunol.* 2010, 34, 19–28.
34. Destoumieux, D.; Muñoz, M.; Cosseau, C.; Rodriguez, J.; Bulet, P.; Comps, M.; Bachère, E. Penaeidins, antimicrobial peptides with chitin-binding activity, are produced and stored in shrimp granulocytes and released after microbial challenge. *J. Cell Sci.* 2000, 113, 461–469.
35. Quispe, R.L.; Justino, E.B.; Vieira, F.N.; Jaramillo, M.L.; Rosa, R.D.; Perazzolo, L.M. Transcriptional profiling of immune-related genes in Pacific white shrimp (*Litopenaeus vannamei*) during ontogenesis. *Fish Shellfish Immunol.* 2016, 58, 103–107.
36. Hauton, C.; Brockton, V.; Smith, V.J. Cloning of a crustin-like, single whey-acidic-domain, antibacterial peptide from the haemocytes of the European lobster, *Homarus gammarus*, and its response to infection with bacteria. *Mol. Immunol.* 2006, 43, 1490–1496.

37. Christie, A.E.; Rus, S.; Goiney, C.C.; Smith, C.M.; Towle, D.W.; Dickinson, P.S. Identification and characterization of a cDNA encoding a crustin-like, putative antibacterial protein from the American lobster *Homarus americanus*. *Mol. Immunol.* 2007, 44, 3333–3337.
38. Zhao, X.; Wang, J. The antimicrobial peptides of the immune response of shrimp. *Invertebr. Surviv. J.* 2008, 5, 162–179.
39. Du, Z.-Q.; Yuan, J.-J.; Ren, D.-M. A novel single WAP domain-containing protein isoform with antibacterial relevance in *Litopenaeus vannamei*. *Fish Shellfish Immunol.* 2015, 44, 478–484.
40. Abdelrahman, H.; ElHady, M.; Alcivar-Warren, A.; Allen, S.; Al-Tobasei, R.; Bao, L.; Beck, B.; Blackburn, H.; Bosworth, B.; Buchanan, J.; et al. Aquaculture genomics, genetics and breeding in the United States: Current status, challenges, and priorities for future research. *BMC Genom.* 2017, 18, 191.
41. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 2013, 30, 2725–2729.
42. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* 2001, 25, 402–408.

RESULTADOS

CAPÍTULO 2

As crustinas representam a maior e mais diversa família de AMPs em crustáceos decápodes (SMITH; DYRYNDA, 2015). A plasticidade estrutural das crustinas parece refletir uma incrível adaptação dos organismos a sobreviver em ambientes com diferentes composições microbianas e sugere uma importância fundamental nas respostas de defesa e/ou no controle de processos fisiológicos, como regulação das cascatas proteolíticas. Considerando a surpreendente diversidade das crustinas, foi concebida a hipótese de que essa família de AMPs é igualmente diversa em *L. vannamei*, ainda que os poucos relatos descritos na literatura subestimem seu real repertório. De fato, um vasto número de sequências nucleotídicas e aminoacídicas correspondentes a “crustinas” encontra-se disponível em bancos de dados públicos (<https://www.ncbi.nlm.nih.gov/search/all/?term=crustin>) das quais atualmente pouco mais de 10% correspondem a sequências de *L. vannamei*. Neste capítulo, a diversidade de sequências de crustinas em *L. vannamei* foi dimensionada visitando sistematicamente seu recém-sequenciado genoma (ZHANG *et al.*, 2019).

Identificação e caracterização do repertório molecular das crustinas do camarão *Litopenaeus vannamei*: uma abordagem genômica

Cairé Barreto^a, Guilherme Augusto Maia^b, Glauber Wagner^b, Marco Gerdol^c, Luciane M Perazzolo^a, Rafael D Rosa^{a*}

^aLaboratory of Immunology Applied to Aquaculture, Department of Cell Biology, Embryology and Genetics, Federal University of Santa Catarina, 88040-900, Florianópolis, SC, Brazil

^bLaboratory of Bioinformatics, Department of Microbiology, Immunology and Parasitology, Federal University of Santa Catarina, 88040-900 Florianópolis, Brazil

^cDepartment of Life Sciences, University of Trieste, 34127 Trieste, Italy

*Corresponding author. E-mail address: rafael.d.rosa@ufsc.br (R.D. Rosa).

Artigo científico a ser submetido em sua versão traduzida para o inglês à revista *Developmental and Comparative Immunology*.

1. Introdução

As crustinas são reconhecidas como a maior e mais diversa família de peptídeos/polipeptídeos antimicrobianos (AMPs) de crustáceos (SMITH; DYRYNDA, 2015). Elas foram originalmente purificadas a partir dos hemócitos do caranguejo *Carcinus maenas* como um peptídeo antibacteriano catiônico de 11,5 kDa com atividade específica contra bactérias Gram-positivas marinhas ou tolerantes ao sal (RELF *et al.*, 1999). Após a identificação de sequências homólogas em camarões peneídeos (BARTLETT *et al.*, 2002), o termo crustina foi empregado para nomear essa família de AMPs que se encontra amplamente distribuída em crustáceos (decápodes, anfípodes e isópodes) e até mesmo em alguns insetos (SMITH; DYRYNDA, 2015). As crustinas são sintetizadas como precursores compostos por uma sequência sinal seguida por um peptídeo/polipeptídeo maduro, geralmente catiônico, rico em resíduos de cisteína e contendo um domínio WAP (do inglês, *Whey Acidic Protein*) na região C-terminal. O domínio WAP é usualmente formado por 50 resíduos aminoacídicos que contém oito resíduos conservados de cisteína engajados em quatro ligações dissulfeto. Esse arranjo tridimensional, descrito como um “núcleo com quatro ligações dissulfeto” (do inglês, *four-disulfide core* ou 4DSC), é igualmente encontrado em outras proteínas que desempenham diferentes funções biológicas, desde inibição da protease até diferenciação tecidual (SMITH, 2011). Notavelmente, esse domínio é considerado essencial para a atividade das crustinas, uma vez que sequências contendo domínios WAP incompletos exibem funções antimicrobianas comprometidas (ZHANG *et al.*, 2007).

Apesar da ocorrência sistemática do domínio WAP C-terminal, a região N-terminal das crustinas é variável em termos de tamanho e composição aminoacídica. De fato, com base na presença/ausência de domínios estruturais específicos nessa região as crustinas de crustáceos são agrupadas em quatro grupos, referidos como Tipos I a IV (SMITH, 2011). As crustinas do Tipo I são filogeneticamente mais relacionadas ao peptídeo antibacteriano de 11,5 kDa, posteriormente denominado “carcinina”, e são frequentemente encontradas em decápodes da subordem Pleocyemata (lagostins, camarões de água doce, caranguejos, lagostas etc) (SMITH; DYRYNDA, 2015). Os membros do Tipo I são geralmente referidos como “carcininas” e são compostos por um peptídeo sinal seguido por um peptídeo maduro contendo 12 resíduos conservados de cisteína. As crustinas do Tipo II, por outro lado, possuem uma região hidrofóbica N-terminal adicional enriquecida em resíduos de glicina à montante dos 12 resíduos de cisteína igualmente encontrados entre as crustinas do Tipo I. Essas crustinas ricas em glicina são descritas principalmente em camarões peneídeos (subordem Dendrobranchiata) e são divididas em dois subtipos, Tipo IIa (“*crustins*”) e Tipo

IIb (“*crustin-like*”). Os Tipos III e IV, em contrapartida, são compostos apenas por um e dois domínios WAP, respectivamente. Vale ressaltar que apenas as crustinas dos Tipos I e II possuem um autêntico arranjo estrutural contendo 12 resíduos de cisteína, o qual inclui quatro resíduos de cisteína conservados (região *cys-rich*) seguidos pelas oito cisteínas que compõem o domínio WAP, formando o então chamado “domínio crustina” (BARRETO *et al.*, 2018; ZHAO; WANG, 2008).

Peptídeos antimicrobianos em camarões peneídeos são usualmente codificados por famílias multigênicas, tal como observado para as peneidinas, (O’LEARY; GROSS, 2006), os fatores anti-lipopolissacarídeos (ALFs) (MATOS *et al.*, 2018) e as stylicinas (FARIAS *et al.*, 2019). Essas famílias são comuns em abrigar múltiplas cópias de genes parálogos gerados por eventos sucessivos de duplicação. Eventos de duplicação gênica representam um importante mecanismo para diversificação molecular, uma vez que os genes duplicados estão sujeitos às diversas pressões evolutivas como entidades independentes e podem seguir destinos funcionais distintos, resultando em variações fenotípicas e fornecendo, possivelmente, melhorias no *fitness* do hospedeiro (WANG *et al.*, 2020). No entanto, embora a diversidade de AMPs codificados por genes em camarões seja associada à expansão gênica por paralogia, não há relatos na literatura relativos à possibilidade de a diversidade de crustinas ser um produto de mecanismos genéticos semelhantes. De fato, considerando que a duplicação de genes parece ser um processo comum que impulsiona a diversificação de AMP, é razoável admitir que tal diversidade na família das crustinas resulte de eventos análogos no genoma do camarão. Atualmente, com a grande expansão das ciências “ômicas”, tornou-se possível a realização de buscas ao longo de genomas inteiros, bem como o acesso a uma diversidade genética inteiramente inexplorada de famílias de AMP de crustáceos por meio de abordagens *in silico*.

A publicação recente do genoma de *L. vannamei* (ZHANG *et al.*, 2019) permitiu estimar a diversidade genética de crustinas em camarões peneídeos através da realização de buscas sistemáticas por seus respectivos genes codificantes. Neste capítulo, análises filogenéticas robustas realizadas com o repertório completo de crustinas de *L. vannamei* mostraram uma vasta diversidade em nível genômico, especialmente para as crustinas Tipo II. Além disso, os resultados das análises *in silico* demonstraram uma intensa expansão de crustinas em *L. vannamei*, apontando para a existência de novos membros da família. A caracterização da diversidade genética de AMPs em camarões fornece novos *insights* sobre a evolução molecular de efetores imunes em crustáceos em nível genômico e expande o

repertório de peptídeos antimicrobianos codificados por genes com potencial terapêutico para o futuro desenvolvimento de drogas para a saúde humana e animal.

2. Materiais e Métodos

A metodologia deste artigo encontra-se detalhada na seção IV da tese.

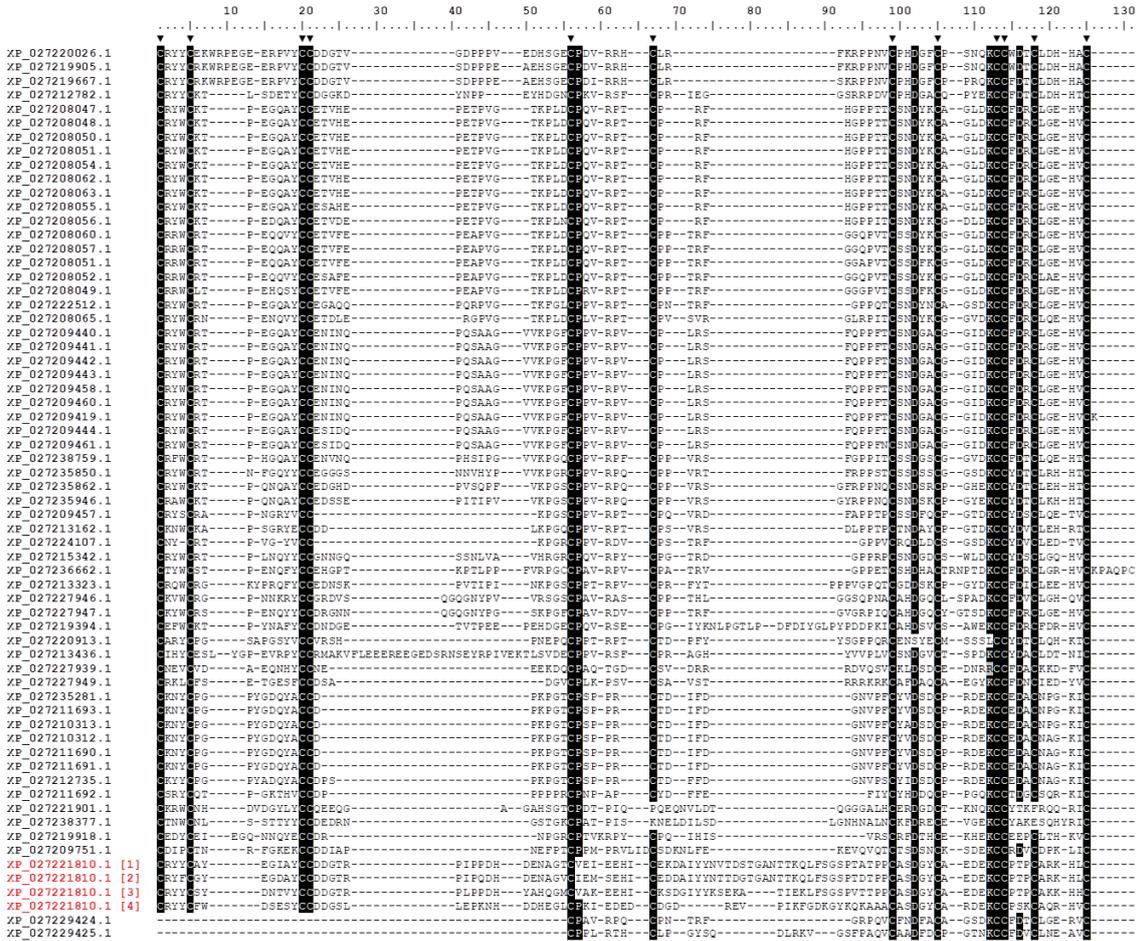
3. Resultados e Discussão

3.1. Descrição geral do repertório molecular das crustinas de *L. vannamei*

Sequências de crustinas foram prospectadas a partir de um universo de 33.260 sequências aminoacídicas preditas do genoma de *L. vannamei* que se encontra atualmente fracionado em 4.683 *scaffolds* genômicos. As buscas levaram à identificação de pelo menos 61 crustinas cujos genes codificantes foram distribuídos em 25 *scaffolds* genômicos (**Figura 12, Quadro 1**). Dessas, 20 sequências foram compostas por um peptídeo sinal seguido de um domínio *cys-rich* e um domínio WAP (10 *scaffolds*); 33 apresentaram um peptídeo sinal seguido por domínios *gly-rich* e *cys-rich* à montante do domínio WAP (11 *scaffolds*) e 2 sequências mostraram a estrutura típica de *LvCrustin A* (1 *scaffold*), cujo peptídeo maduro foi formado unicamente por um domínio *gly-rich* precedendo um domínio WAP C-terminal (LI *et al.*, 2018). Curiosamente, 6 sequências (5 *scaffolds*) apresentaram em sua composição domínios estruturais alternativos àqueles classicamente previstos nos atuais sistemas de classificação das crustinas ou portaram variações no domínio crustina em virtude de polimorfismos no número de resíduos de cisteína (**Quadro 1**). Dentre os 61 genes recuperados, 6 foram associados a mais de um transcrito (*e.g.*, LOC113811778, LOC113814028, LOC113801821, LOC113801828, LOC113801837 e LOC113802945), o que sugere a possível existência de *loci* multialélicos ou mesmo o resultado de eventuais erros de montagem dos transcritos. No entanto, para o estabelecimento de um repertório molecular não redundante, as múltiplas sequências vinculadas a um mesmo gene foram colapsadas em uma única sequência representativa.

As buscas levaram, ainda, à identificação de 5 crustinas do Tipo III (SWD) e de 1 crustina do Tipo IV (DWD), segundo os critérios estabelecidos por Smith e colaboradores (2011). No entanto, o presente trabalho dialoga igualmente com o sistema de classificação proposto por Zhao & Wang (2008), o qual desconsidera as SWD e DWD como crustinas verdadeiras em virtude da ausência de um autêntico domínio crustina. Por essa razão, essas sequências foram manualmente descartadas das análises.

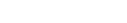
Figura 12. Alinhamento aminoacídico da região referente ao “domínio crustina” das crustinas identificadas em *L. vannamei*.



As seqüências aminoacídicas das regiões conservadas entre as 61 crustinas recuperadas foram alinhadas utilizando o software MAFFT v.7. Os números à esquerda representam os identificadores das seqüências aminoacídicas. Os resíduos conservados na maioria das seqüências estão destacados em preto e os resíduos de cisteína componentes do domínio crustina estão indicados com o símbolo ▼. Os hífen (-) representam *gaps* e a linha vermelha situa o domínio WAP C-terminal. A seqüência XP_027221810.1 que contém múltiplos domínios crustina está evidenciada em vermelho. Os múltiplos domínios crustina estão indicados com os números em colchetes. A régua acima indica a posição relativa dos resíduos aminoacídicos.

Quadro 1. Repertório completo de crustinas recuperadas no proteoma predito de *L. vannamei*.

Continua

| Domínios estruturais | ID proteína | ID locus | scaffold | Referência |
|---|----------------|--------------|-------------------|-----------------------|
|  | XP_027219667.1 | LOC113811998 | LVANscaffold_1143 | Li et al., 2020 |
|  | XP_027219905.1 | LOC113812245 | LVANscaffold_1143 | Li et al., 2020 |
|  | XP_027209751.1 | LOC113803231 | LVANscaffold_1400 | Li et al., 2020 |
|  | XP_027212735.1 | LOC113805871 | LVANscaffold_1678 | Li et al., 2020 |
|  | XP_027212782.1 | LOC113805910 | LVANscaffold_1678 | Li et al., 2020 |
|  | XP_027219394.1 | LOC113811778 | LVANscaffold_2260 | Li et al., 2020 |
|  | XP_027221901.1 | LOC113814028 | LVANscaffold_2487 | Li et al., 2020 |
|  | XP_027238377.1 | LOC113829403 | LVANscaffold_597 | Li et al., 2020 |
|  | XP_027219918.1 | LOC113812253 | LVANscaffold_1143 | Li et al., 2020 |
|  | XP_027235281.1 | LOC113826552 | LVANscaffold_1034 | Este trabalho |
|  | XP_027220026.1 | LOC113812348 | LVANscaffold_1143 | Este trabalho |
|  | XP_027210312.1 | LOC113803699 | LVANscaffold_1454 | Este trabalho |
|  | XP_027210313.1 | LOC113803700 | LVANscaffold_1454 | Este trabalho |
|  | XP_027211690.1 | LOC113804968 | LVANscaffold_1584 | Este trabalho |
|  | XP_027211691.1 | LOC113804971 | LVANscaffold_1584 | Este trabalho |
|  | XP_027211692.1 | LOC113804972 | LVANscaffold_1584 | Este trabalho |
|  | XP_027211693.1 | LOC113804973 | LVANscaffold_1584 | Este trabalho |
|  | XP_027208057.1 | LOC113801827 | LVANscaffold_887 | Este trabalho |
|  | XP_027208058.1 | LOC113801828 | LVANscaffold_887 | Este trabalho |
|  | XP_027208060.1 | LOC113801830 | LVANscaffold_887 | Este trabalho |
|  | XP_027213162.1 | LOC113806236 | LVANscaffold_1709 | Li et al., 2020 |
|  | XP_027222512.1 | LOC113814664 | LVANscaffold_2550 | Li et al., 2020 |
|  | XP_027227947.1 | LOC113819912 | LVANscaffold_3068 | Li et al., 2020 |
|  | XP_027235850.1 | LOC113827151 | LVANscaffold_1288 | Li et al., 2020 |
|  | XP_027235862.1 | LOC113827160 | LVANscaffold_1288 | Li et al., 2020 |
|  | XP_027235946.1 | LOC113827248 | LVANscaffold_1288 | Li et al., 2020 |
|  | XP_027208065.1 | LOC113801837 | LVANscaffold_888 | Li et al., 2020 |
|  | XP_027227946.1 | LOC113819911 | LVANscaffold_3068 | Li et al., 2020 |
|  | XP_027208049.1 | LOC113801818 | LVANscaffold_887 | Li et al., 2020 |
|  | XP_027209444.1 | LOC113802951 | LVANscaffold_1361 | Li et al., 2019 |
|  | XP_027224107.1 | LOC113816267 | LVANscaffold_2687 | Barreto et al., 2018 |
|  | XP_027208055.1 | LOC113801825 | LVANscaffold_887 | Bartlett et al., 2002 |
|  | XP_027209440.1 | LOC113802947 | LVANscaffold_1361 | Este trabalho |
|  | XP_027209441.1 | LOC113802948 | LVANscaffold_1361 | Este trabalho |
|  | XP_027209442.1 | LOC113802949 | LVANscaffold_1361 | Este trabalho |
|  | XP_027209443.1 | LOC113802950 | LVANscaffold_1361 | Este trabalho |
|  | XP_027209457.1 | LOC113802959 | LVANscaffold_1361 | Este trabalho |
|  | XP_027209458.1 | LOC113802960 | LVANscaffold_1361 | Este trabalho |
|  | XP_027209460.1 | LOC113802961 | LVANscaffold_1361 | Este trabalho |
|  | XP_027209461.1 | LOC113802962 | LVANscaffold_1361 | Este trabalho |
|  | XP_027215342.1 | LOC113808202 | LVANscaffold_1889 | Este trabalho |
|  | XP_027238759.1 | LOC113829784 | LVANscaffold_634 | Este trabalho |
|  | XP_027208047.1 | LOC113801814 | LVANscaffold_887 | Este trabalho |
|  | XP_027213323.1 | LOC113806391 | LVANscaffold_1095 | Este trabalho |
|  | XP_027209419.1 | LOC113802945 | LVANscaffold_1361 | Este trabalho |
|  | XP_027208048.1 | LOC113801817 | LVANscaffold_887 | Este trabalho |
|  | XP_027208050.1 | LOC113801819 | LVANscaffold_887 | Este trabalho |
|  | XP_027208051.1 | LOC113801820 | LVANscaffold_887 | Este trabalho |
|  | XP_027208052.1 | LOC113801821 | LVANscaffold_887 | Este trabalho |

Conclusão

| Domínios estruturais | ID proteína | ID locus | scaffold | Referência |
|---|----------------|--------------|-------------------|-----------------|
|  | XP_027208054.1 | LOC113801823 | LVANscaffold_887 | Este trabalho |
|  | XP_027208056.1 | LOC113801826 | LVANscaffold_887 | Este trabalho |
|  | XP_027208062.1 | LOC113801832 | LVANscaffold_888 | Este trabalho |
|  | XP_027208063.1 | LOC113801833 | LVANscaffold_888 | Este trabalho |
|  | XP_027229424.1 | LOC113821158 | LVANscaffold_3186 | Li et al., 2017 |
|  | XP_027229425.1 | LOC113821159 | LVANscaffold_3186 | Este trabalho |
|  | XP_027213436.1 | LOC113806493 | LVANscaffold_1736 | Este trabalho |
|  | XP_027220913.1 | LOC113813163 | LVANscaffold_2401 | Li et al., 2020 |
|  | XP_027227939.1 | LOC113819904 | LVANscaffold_3068 | Este trabalho |
|  | XP_027227949.1 | LOC113819914 | LVANscaffold_3068 | Este trabalho |
|  | XP_027236662.1 | LOC113827916 | LVANscaffold_4539 | Li et al., 2020 |
|  | XP_027221810.1 | LOC113813930 | LVANscaffold_2478 | Este trabalho |



Os domínios estruturais presentes no peptídeo maduro das crustinas identificadas foram representados esquematicamente na primeira coluna. O domínio crustina está representado pelos blocos em azul claro (domínio *cys-rich*) e em cinza (domínio WAP). O domínio *gly-rich* está representado pelo bloco em laranja. Domínios estruturais alternativos que foram identificados entre as sequências, mas que não foram descritos no atual sistema de classificação foram representados em cores distintas (beje, azul escuro, verde e vermelho). A linha em amarelo no domínio WAP da sequência XP_027236662.1 representa a presença de dois resíduos de cisteína adicionais. A coluna REFERÊNCIA indica os trabalhos onde as sequências obtidas foram primeiramente descritas.

A exploração sistemática no proteoma predito de *L. vannamei* revelou uma surpreendente abundância de crustinas, que mostrou ser ainda maior do que aquela previamente descrita para essa espécie. De fato, apenas quatro crustinas haviam sido descritas em *L. vannamei*, a saber: as crustinas dos Tipos IIa e IIb e as *LvCrustin A* e *LvCrustin B* (BARRETO *et al.*, 2018; BARTLETT *et al.*, 2002; LI *et al.*, 2018, 2019). No entanto, em um trabalho recente, Li e colaboradores (2020) descreveram 20 novas crustinas em *L. vannamei*, incluindo sequências que foram descritas como membros dos Tipos I e II. No presente trabalho, todas as crustinas previamente reportadas nessa espécie foram identificadas, além de outros 37 genes que não foram, até o momento, descritos em camarões peneídeos.

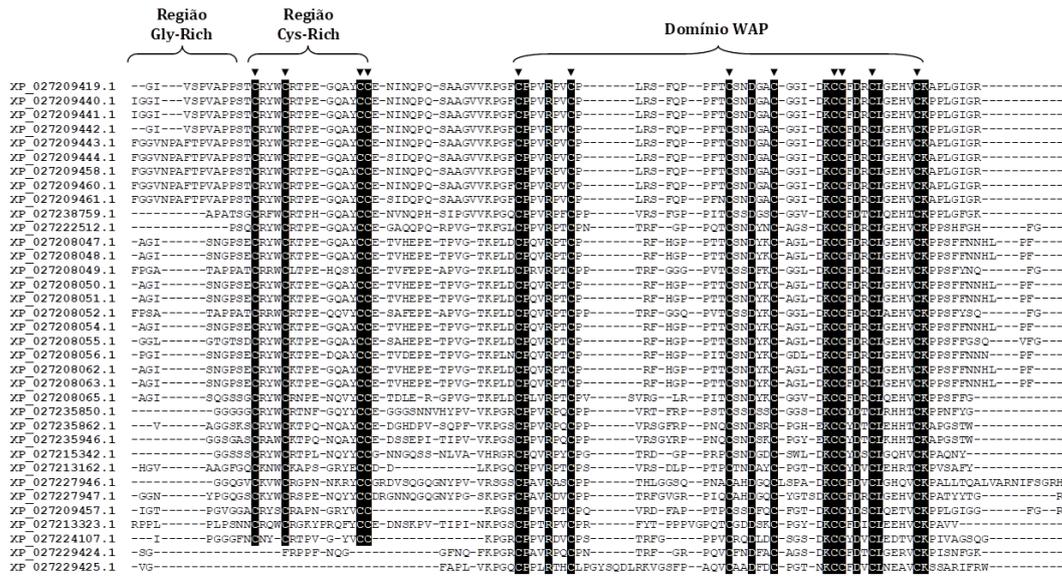
A identificação de pelo menos 61 crustinas a partir das buscas pelo genoma demonstrou objetivamente a suma contribuição das ciências ômicas para a exploração da diversidade molecular de peptídeos antimicrobianos e sua relevância na descoberta de novas moléculas capazes de serem futuramente utilizadas para aplicação biotecnológica. Além disso, o vasto repertório reconhecido em *L. vannamei* sugere uma importante participação das crustinas nas respostas de defesa em camarões peneídeos e pressupõe a ocorrência de possíveis eventos de expansão gênica dirigidos por pressões evolutivas diversas em direção à

formação de um arsenal imunológico abundante e capaz de lidar com a ampla comunidade microbiana dispersa no ambiente.

3.2. Identificação e caracterização das crustinas do Tipo II

3.2.1. As crustinas do Tipo II são diversas e constituídas por múltiplos grupos

Pelo menos 33 sequências aminoacídicas correspondentes a crustinas do Tipo II foram identificadas a partir do proteoma predito de *L. vannamei*, de acordo com os critérios estabelecidos por Smith e colaboradores (2008). De fato, todas as sequências recuperadas apresentaram uma arquitetura molecular consistente com o Tipo II, incluindo a presença de um domínio hidrofóbico amino-terminal rico em resíduos de glicina precedendo um domínio crustina conservado na região C-terminal (**Figura 13**). Além disso, foram identificadas 2 sequências que apresentaram um domínio *gly-rich* típico das crustinas do Tipo II, mas que foram ausentes do característico módulo *cys-rich*. Essa arquitetura molecular foi previamente descrita em uma crustina de *L. vannamei* (*LvCrustin A*) cujas reconstruções filogenéticas mostram uma relação evolutiva próxima com as autênticas crustinas do Tipo II (LI *et al.*, 2018). Nesse sentido, as buscas *in silico* levaram à identificação de 35 genes codificantes de crustinas do Tipo II em *L. vannamei* (**Figura 13**).



As seqüências aminoácidas correspondentes às crustinas do Tipo II foram alinhadas utilizando o software MAFFT v.7 e os domínios estruturais típicos foram assinalados. Os resíduos aminoácidos conservados estão destacados em preto. O símbolo ▼ indica os resíduos conservados de cisteína.

A fim de investigar a diversidade molecular subjacente às crustinas do Tipo II, os 35 genes foram sistematicamente organizados em grupos definidos em função da identidade aminoácida de seus produtos proteicos. Esse modo de agrupamento utilizando a identidade aminoácida como critério de seleção foi aplicado no intuito de trazer validade filogenética aos grupos identificados, uma vez que a conservação aminoácida no peptídeo maduro reflete indiretamente a distância evolutiva entre seus respectivos genes codificantes. Assim, seqüências com identidade igual ou superior a 70% na região referente ao domínio crustina foram reunidas em um mesmo grupo, enquanto que seqüências somando mais de 30% de diferença nessa região foram separadas em grupos distintos. As análises comparativas revelaram a existência de pelo menos 12 crustinas do Tipo II em *L. vannamei* (Grupos 1 – 12) (**Figura 14**). O grupo 1 reuniu as autênticas crustinas do Tipo IIa (Crustin *Lv*) e foi formado por 13 genes, enquanto que o grupo 2 foi constituído por 11 genes que codificaram *Lv*Crustin B, uma autêntica crustina do Tipo II recentemente descrita em *L. vannamei* (LI *et al.*, 2019). O grupo 3 foi representado por um único gene codificante de crustinas do Tipo IIb (Crustin-like *Lv*) e o grupo 4 foi também constituído por um único gene que codificou um provável ortólogo de crustin *Pm5*, descrita apenas no camarão asiático *P. monodon* (VATANAVICHARN *et al.*, 2009). O grupo 11 mostrou codificar a *Lv*Crustin A, portadoras do típico domínio *gly-rich*, mas que são ausentes do módulo *cys-rich* (LI *et al.*, 2018). Todos os genes dos grupos 1 a 4, juntamente com o grupo 11, codificaram crustinas previamente descritas em *L. vannamei*, em detrimento dos demais grupos que codificaram crustinas inéditas nessa espécie (**Tabela 2**).

Tabela 2. Lista de sequências homólogas às crustinas do Tipo II identificadas em *L. vannamei* e em diferentes espécies de camarões peneídeos.

| Grupo | Homólogos | Espécie | Identidade (%) | Tipo |
|--------------|------------------------|---------------------|-----------------------|--|
| 1 | <i>crusFpau</i> | <i>F. paulensis</i> | 82-100% | IIa (Crustin Lv) |
| | <i>crustinPm6</i> | <i>P. monodon</i> | | |
| | <i>crustinPm2</i> | <i>P. monodon</i> | | |
| | GGLH01008401.1 | <i>P. monodon</i> | | |
| 2 | <i>LvCrustin B</i> | <i>L. vannamei</i> | 100% | LvCrustin B |
| | GGLH01076154.1 | <i>P. monodon</i> | | |
| 3 | <i>Crustin-like Lv</i> | <i>L. vannamei</i> | 74-100% | IIb (Crustin-like Lv) |
| | <i>crustinPm7</i> | <i>P. monodon</i> | | |
| | <i>CruFc</i> | <i>F. chinensis</i> | | |
| | <i>Ficrustin</i> | <i>F. indicus</i> | | |
| 4 | <i>MjCrustin-like</i> | <i>M. japonicus</i> | 95% | crustinPm5 |
| | <i>crustinPm5</i> | <i>P. monodon</i> | | |
| 5 | GEUA01090612.1 | <i>F. aztecus</i> | 85-99% | nd |
| | GGLH01067064.1 | <i>P. monodon</i> | | |
| 6 | GGLH01085808.1 | <i>P. monodon</i> | 94% | nd |
| 7 | GGLH01051266.1 | <i>P. monodon</i> | 77% | nd |
| 8 | GGLH01080784.1 | <i>P. monodon</i> | 94% | nd |
| 9 | GGLH01074537.1 | <i>P. monodon</i> | 94% | nd |
| 10 | GGLH01094071.1 | <i>P. monodon</i> | 88% | nd |
| 11 | <i>LvCrustin A</i> | <i>L. vannamei</i> | 100% | LvCrustin A |
| 12 | GGLH01096483.1 | <i>P. monodon</i> | 71% | nd |

nd: não descrito.

Para validar a existência das crustinas identificadas, buscas pelos seus respectivos transcritos codificantes foram realizadas em bancos transcritômicos (repositórios TSA e EST) de *L. vannamei*. As buscas utilizando o algoritmo tBLASTn confirmaram sua existência a partir da identificação de transcritos correspondentes com altos valores de identidade ($\geq 90\%$) e confiabilidade estatística ($e\text{-value} \leq 10^{-24}$) (**Quadro 2**). Curiosamente, nem todas as crustinas do Tipo II apresentaram transcritos correspondentes cujos produtos proteicos expuseram 100% de identidade aminoacídica. Isso ocorreu devido à ocorrência de polimorfismos nas sequências, refletindo, portanto a existência de variabilidade intraespecífica.

Quadro 2. Validação *in silico* das crustinas do Tipo II reconhecidas em *L. vannamei*.

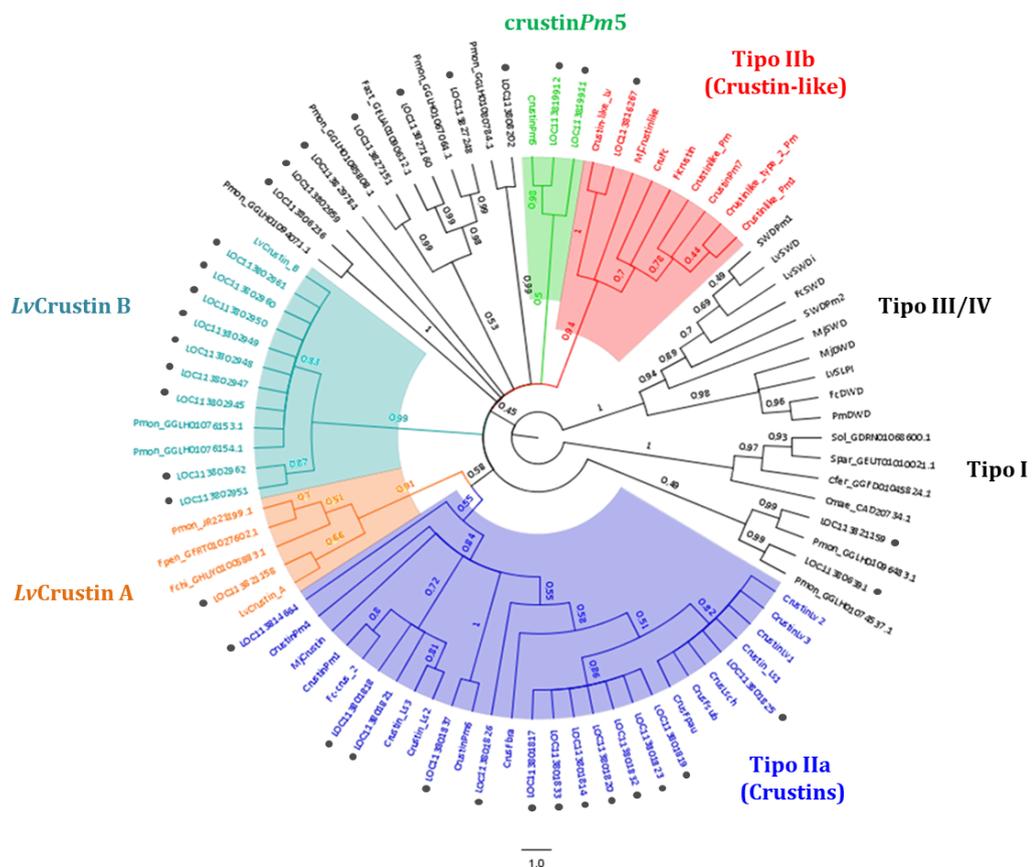
| Crustinas do Tipo II | | | | Validação em bancos transcriptômicos | | |
|----------------------|----------------|-------------------|-------------------|--------------------------------------|----------|------------|
| Grupo | ID proteína | ID locus | id scaffold | GenBank | e-value | Identidade |
| Grupo 1 | XP_027208047.1 | LOC113801814 | LVANscaffold_887 | GGKO01017164.1 | 4,96E-59 | 100% |
| | XP_027208048.1 | LOC113801817 | LVANscaffold_887 | | | |
| | XP_027208049.1 | LOC113801818 | LVANscaffold_887 | | | |
| | XP_027208050.1 | LOC113801819 | LVANscaffold_887 | | | |
| | XP_027208051.1 | LOC113801820 | LVANscaffold_887 | | | |
| | XP_027208052.1 | LOC113801821 | LVANscaffold_887 | | | |
| | XP_027208054.1 | LOC113801823 | LVANscaffold_887 | | | |
| | XP_027208055.1 | LOC113801825 | LVANscaffold_887 | | | |
| | XP_027208056.1 | LOC113801826 | LVANscaffold_887 | | | |
| | XP_027208062.1 | LOC113801832 | LVANscaffold_888 | | | |
| | XP_027208063.1 | LOC113801833 | LVANscaffold_888 | | | |
| | XP_027208065.1 | LOC113801837 | LVANscaffold_888 | | | |
| | XP_027222512.1 | LOC113814664 | LVANscaffold_2550 | | | |
| Grupo 2 | XP_027209419.1 | LOC113802945 | LVANscaffold_1361 | GGKO01020613.1 | 4,03E-56 | 100% |
| | XP_027209440.1 | LOC113802947 | LVANscaffold_1361 | | | |
| | XP_027209441.1 | LOC113802948 | LVANscaffold_1361 | | | |
| | XP_027209442.1 | LOC113802949 | LVANscaffold_1361 | | | |
| | XP_027209443.1 | LOC113802950 | LVANscaffold_1361 | | | |
| | XP_027209444.1 | LOC113802951 | LVANscaffold_1361 | | | |
| | XP_027209458.1 | LOC113802960 | LVANscaffold_1361 | | | |
| | XP_027209460.1 | LOC113802961 | LVANscaffold_1361 | | | |
| | XP_027209461.1 | LOC113802962 | LVANscaffold_1361 | | | |
| | XP_027238759.1 | LOC113829784 | LVANscaffold_634 | | | |
| XP_027209457.1 | LOC113802959 | LVANscaffold_1361 | | | | |
| Grupo 3 | XP_027224107.1 | LOC113816267 | LVANscaffold_2687 | GHXV01014204.1 | 1,10E-44 | 100% |
| Grupo 4 | XP_027227947.1 | LOC113819912 | LVANscaffold_3068 | GGKO01020259.1 | 1,04E-83 | 98,2% |
| Grupo 5 | XP_027235862.1 | LOC113827160 | LVANscaffold_1288 | GFRP01026156.1 | 5,14E-39 | 98,8% |
| | XP_027235946.1 | LOC113827248 | LVANscaffold_1288 | GFRP01034019.1 | 2,75E-37 | 96,4% |
| Grupo 6 | XP_027235850.1 | LOC113827151 | LVANscaffold_1288 | GFRP01042820.1 | 3,97E-24 | 100% |
| Grupo 7 | XP_027227946.1 | LOC113819911 | LVANscaffold_3068 | GDTK01009760.1 | 5,46E-68 | 99% |
| Grupo 8 | XP_027215342.1 | LOC113808202 | LVANscaffold_1889 | GFRP01042651.1 | 1,40E-52 | 97,5% |
| Grupo 9 | XP_027213323.1 | LOC113806391 | LVANscaffold_1095 | GGQV01093230.1 | 7,97E-69 | 91,2% |
| Grupo 10 | XP_027213162.1 | LOC113806236 | LVANscaffold_1709 | GGQV01117632.1 | 2,27E-56 | 98% |
| Grupo 11 | XP_027229424.1 | LOC113821158 | LVANscaffold_3186 | GGQV01050848.1 | 4,37E-64 | 95,3% |
| Grupo 12 | XP_027209425.1 | LOC113821159 | LVANscaffold_3186 | GGKO01020613.1 | 2,44E-56 | 100% |

3.2.2. Crustinas do Tipo II experimentaram uma ampla diversificação em camarões peneídeos

Com o intuito de confirmar a identidade filogenética dos grupos identificados e compreender as relações evolutivas existentes entre eles, as sequências foram analisadas e comparadas com várias crustinas (Tipos I a IV) de crustáceos decápodes através de reconstruções filogenéticas, incluindo os prováveis ortólogos recuperados em camarões peneídeos (**Tabela 2**). Sequências referentes às crustinas dos Tipos III e IV foram utilizadas como grupo externo à árvore. De acordo com o cladograma, as crustinas identificadas em *L. vannamei* mostraram origens evolutivas comuns e foram agrupadas em um único clado correspondente ao Tipo II (**Figura 15**). Ademais, as reconstruções filogenéticas revelaram

uma extensa diversificação genética, tal como evidenciado pela esparsa distribuição dos cladogramas, sugerindo que essas crustinas experimentaram uma explosão de diversidade que remonta, pelo menos, aos ancestrais da família Penaeidae. Dentre os sete genes codificantes de crustinas inéditas em camarões peneídeos, apenas um (*i.e.*, LOC113819911) apresentou relações filogenéticas próximas com uma crustina previamente descrita na literatura (*crustinPm5*) (VATANAVICHARN *et al.*, 2009). Curiosamente, embora tenha apresentado uma arquitetura molecular similar a *LvCrustin A*, sendo igualmente desprovida do módulo *cys-rich*, a crustina pertencente ao grupo 12 (XP_027209425.1) foi clusterizada em um clado separado, indicando que a perda desse módulo amino-terminal pode ter acontecido duas vezes de maneira independente no curso evolutivo e tenha, possivelmente, um valor adaptativo.

Figura 15. Árvore filogenética das crustinas do Tipo II de camarões peneídeos.



As sequências aminoacídicas correspondentes às regiões conservadas foram alinhadas e analisadas filogeneticamente através do método *Maximum Likelihood* com 1.000 replicatas utilizando o modelo WAG (G+I). Os cladogramas correspondentes às crustinas do Tipo II descritas em *L. vannamei* estão evidenciados em cores distintas. As esferas pretas destacam os genes recuperados a partir das buscas. *Sol*: *Scylla olivacea*; *Spar*: *S. paramamosain*; *Cfer*: *Charybdis feriata*; *Cmae*: *C. maenas*; *Pm/Pmon*: *P. monodon*; *Lv*: *L. vannamei*; *Ls*: *L. setyferus*; *Lsch*: *L. schmitti*; *Fsub*: *F. subtilis*; *Fpau*: *F. paulensis*; *Fbra*: *F. brasiliensis*; *Fc/Fchi*: *F. chinensis*; *Mj*: *M. japonicus*; *Fpen*: *F. penicillatus*; *Fazt*: *F. aztecus*; *Fi*: *F. indicus*.

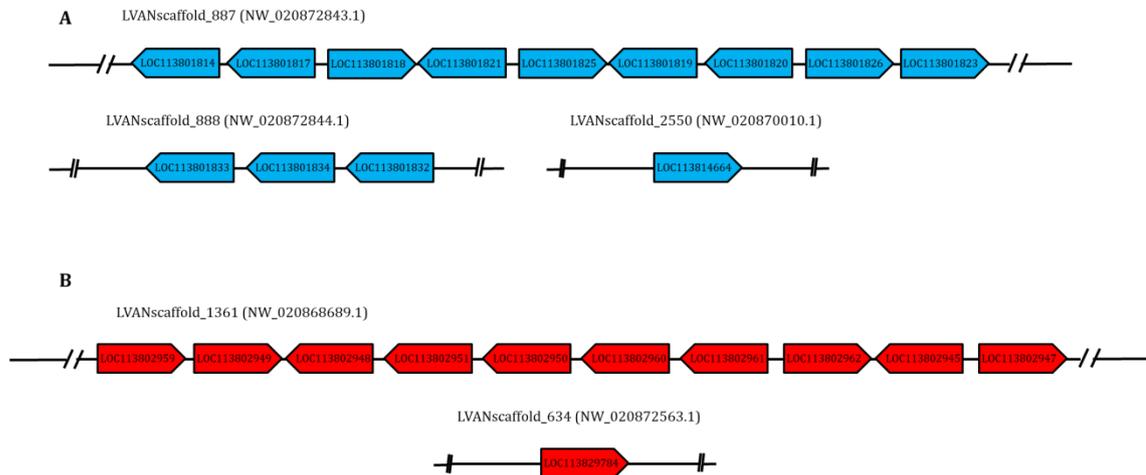
Uma importante informação provida pelas reconstruções filogenéticas foi que o Tipo IIa, definido em função da presença de uma assinatura molecular conservada (vide Capítulo 1), representa um grupo polifilético, uma vez que inclui dois grupos de crustinas do Tipo II (*crustins* e *crustinPm5*) cujos genes codificantes não compartilham um ancestral comum imediato. É fato que a proposição da assinatura molecular conservada descrita no Capítulo 1 para as crustinas do Tipo IIa baseou-se nas sequências disponíveis à época que mostraram consistentemente a formação de um grupo monofilético sólido a partir de sequências de diferentes origens (BARRETO *et al.*, 2018). No entanto, a exploração de crustinas no genoma de *L. vannamei* e a concomitante descoberta de novos representantes demonstraram que a diversidade de crustinas do Tipo II em camarões peneídeos supera aquela previamente descrita e, por isso, ressalta a necessidade de uma reformulação das assinaturas moleculares para abarcar a real diversidade molecular dessas crustinas.

Outra informação revelada a partir das buscas e análises filogenéticas foi que a maioria dos grupos identificados foi formada por um único gene, em detrimento dos grupos 1 (Tipo IIa) e 2 (*LvCrustin B*) que representaram os dois maiores grupos em *L. vannamei*, formados por 13 e 11 genes, respectivamente. A presença de múltiplas cópias gênicas em apenas dois dos doze grupos identificados levanta questões sobre a dinâmica evolutiva das crustinas do Tipo II em camarões peneídeos e sugere uma provável expansão gênica “grupo-específica” derivada de sucessivos eventos de duplicação, possivelmente associados a pressões seletivas para manutenção de múltiplas cópias. Indícios da expansão gênica dos grupos 1 e 2 são suportados pela distribuição topológica dos seus genes codificantes no genoma de *L. vannamei*, os quais localizam-se adjacentes uns aos outros formando densos clusters gênicos (*i.e.*, regiões contendo várias cópias de genes similares) (**Figura 16**).

De fato, nove dentre os treze genes codificantes de crustinas do grupo 1 estão situados no mesmo *scaffold* genômico (LVANscaffold_887) e residem em posições vizinhas, enquanto que três genes estão distribuídos de maneira similar no *scaffold* LVANscaffold_888 (**Figura 16A**). Analogamente, dez dos onze genes codificantes de crustinas do grupo 2 estão situados lado a lado em um *scaffold* genômico comum (LVANscaffold_1361), enquanto que um único gene (LOC113829784) reside separadamente em um *scaffold* genômico individual (LVANscaffold_634) (**Figura 16B**). Eventos de duplicação são comuns em gerar cópias gênicas vizinhas, no qual um gene ou uma região contendo um ou mais genes são duplicados durante a meiose devido a um desalinhamento das regiões homólogas nas cromátides-irmãs. Como consequência, tem-se a formação de clusters gênicos que podem conter dezenas de genes repetidos e que se tornam sujeitos individualmente às pressões seletivas do ambiente,

representando, portanto, uma importante fonte de diversificação molecular (FREELING, 2009).

Figura 16. Distribuição topológica dos genes codificantes de crustinas dos grupos 1 e 2 no genoma de *L. vannamei*.



Os genes codificantes de crustinas dos grupos 1 e 2 foram manualmente coletados a partir do genoma de *L. vannamei* e analisados quanto à sua distribuição topológica. **(A)** Genes codificantes de grupo 1. **(B)** Genes codificantes do grupo 2. As linhas representam os *scaffolds* genômicos. As setas apontando para esquerda ou direita representam o frame de transcrição negativo ou positivo, respectivamente.

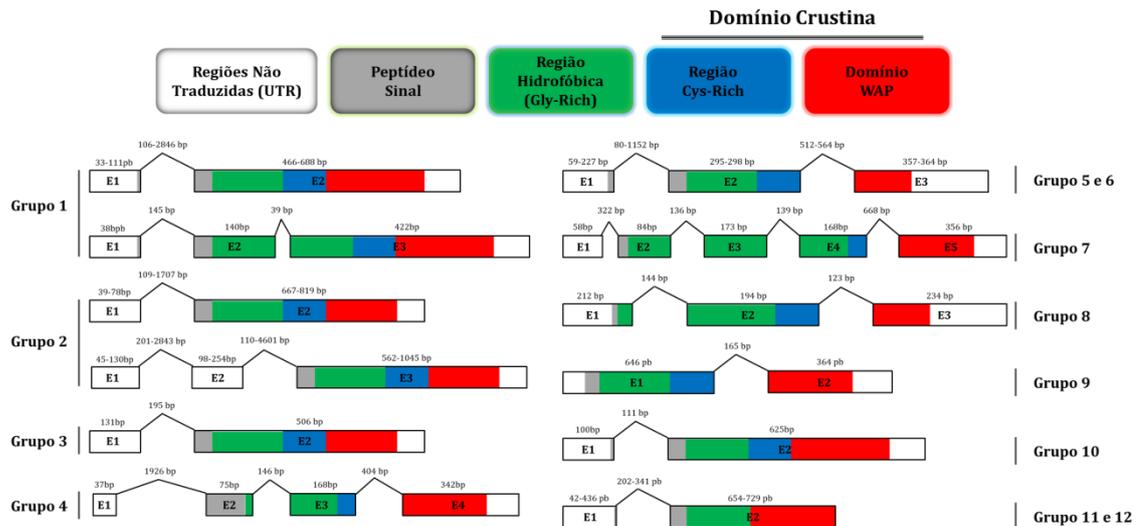
De fato, AMPs são comuns em experimentar expansões gênicas e tornarem-se famílias multigênicas formadas por múltiplos parálogos, embora a redundância funcional dos genes duplicados promova comumente um relaxamento de pressões seletivas que os torna permissivos ao aparecimento de eventuais mutações não sinônimas, pseudogenização ou até mesmo deleção (LAZZARO; ZASLOFF; ROLFF, 2020). O elevado número de genes parálogos observado entre as crustinas dos grupos 1 e 2 pode refletir a necessidade de uma manutenção de altos níveis transcricionais e pressupõe uma provável importância desses genes na homeostase. Nesse cenário, estudos funcionais direcionados à investigação da sua função biológica são importantes para obter pistas sobre sua participação na imunidade, assim como estudos em genômica comparativa visando à investigação do número de cópias gênicas em outras espécies de camarões peneídeos. A identificação de 12 crustinas do Tipo II em *L. vannamei* sugere, então, que a diversidade molecular descrita na literatura é atualmente subestimada e indica que essas crustinas representam um grupo amplo e diverso que experimentou uma profunda expansão gênica em ancestrais da família Penaeidae, sobretudo em *L. vannamei*, cuja expansão gênica parece permanecer acontecendo e dando origem a grupos multigênicos.

3.2.3. As crustinas do Tipo II de camarões são codificadas por genes que apresentam organizações genômicas diversas

Com o intuito de caracterizar a diversidade dos genes codificantes e analisar suas organizações genômicas, os *loci* correspondentes às crustinas do Tipo II foram manualmente coletados a partir do genoma de *L. vannamei*. As análises apontaram que as crustinas do Tipo II formam um grupo multigênico e diverso que reúne pelo menos 35 genes distintos distribuídos em 12 *scaffolds* genômicos. Os genes mostraram um comprimento médio de ca. 3 kb e apresentaram organizações estruturais distintas no que concerne ao número de íntrons e éxons (**Figura 17**).

A arquitetura genômica de crustinas do Tipo II foi previamente analisada nas crustinas dos Tipos IIa e IIb (vide Capítulo 1). Nelas, o primeiro éxon mostrou codificar integralmente a região 5' não traduzida (5'-UTR), enquanto que o peptídeo sinal, todo o peptídeo maduro e a região 3' não traduzida (3'-UTR) mostraram ser codificados no éxon seguinte. Esse arranjo estrutural foi também observado entre as crustinas componentes do grupo 2 (*LvCrustin B*), do grupo 10 e dos grupos 11 e 12. De modo interessante, uma extensa variedade de organizações estruturais foi encontrada entre as crustinas pertencentes aos demais grupos, que incluiu desde genes compostos por dois éxons separados por um íntron até genes formados por cinco éxons, onde o domínio *gly-rich* mostrou estar particionado em três éxons distintos (Grupo 7). Em conformidade com os altos valores de identidade aminoacídica (95%), o gene LOC113819912 (Grupo 4) apresentou a mesma organização genômica da crustin*Pm5* do camarão *P. monodon*, corroborando sua predição enquanto gene ortólogo em *L. vannamei* (**Figura 17**). Todos os íntrons foram flanqueados pelos sítios doadores e aceptores canônicos de *splicing* (*i.e.*, GT/AG).

Figura 17. Organização genômica dos genes codificantes de crustinas do Tipo II identificados em *L. vannamei*.



Os retângulos e as linhas pretas representam, respectivamente, os éxons (E1: éxon 1, E2: éxon 2, E3: éxon 3, E4: éxon 4, E5: éxon 5) e os introns dos genes codificantes. O tamanho das sequências está representado em pares de bases (pb). O comprimento dos introns e éxons não está representado proporcionalmente.

As análises de arquitetura molecular proveram evidências da ocorrência de possíveis eventos de intronização atuando sobre os genes codificantes de crustinas do Tipo II. De fato, na maioria dos genes o módulo *gly-rich* mostrou ser codificado em um único éxon. No entanto, em pelo menos quatro genes (*e.g.*, LOC113801833, LOC113819912, LOC113819911 e LOC113808202) a região codificante desse mesmo módulo mostrou estar particionada em diferentes éxons, sugerindo a possível ocorrência de eventos de intronização que fracionaram essa região em múltiplas unidades independentes. Uma situação similar foi encontrada para a região que codifica o domínio WAP, o qual se encontra normalmente situado à jusante da região *cys-rich* em um éxon comum. Aqui, cinco genes (*e.g.*, LOC113819912, LOC113827160, LOC113827151, LOC113819911 e LOC113808202) mostraram codificar esse domínio separadamente em um éxon individual, permitindo pressupor igualmente um evento de intronização que resultou no isolamento dessas duas regiões. O fracionamento de éxons em múltiplas unidades codificantes representa um mecanismo evolutivo que contribui para o aumento de diversidade molecular de famílias gênicas, uma vez que os novos éxons formados tornam-se sujeitos à diversificação como módulos independentes e permitem o surgimento de variações estruturais pontuais que alteram porções da proteína, mas que conservam as demais regiões, levando ao eventual surgimento de isoformas (SOREK, 2007).

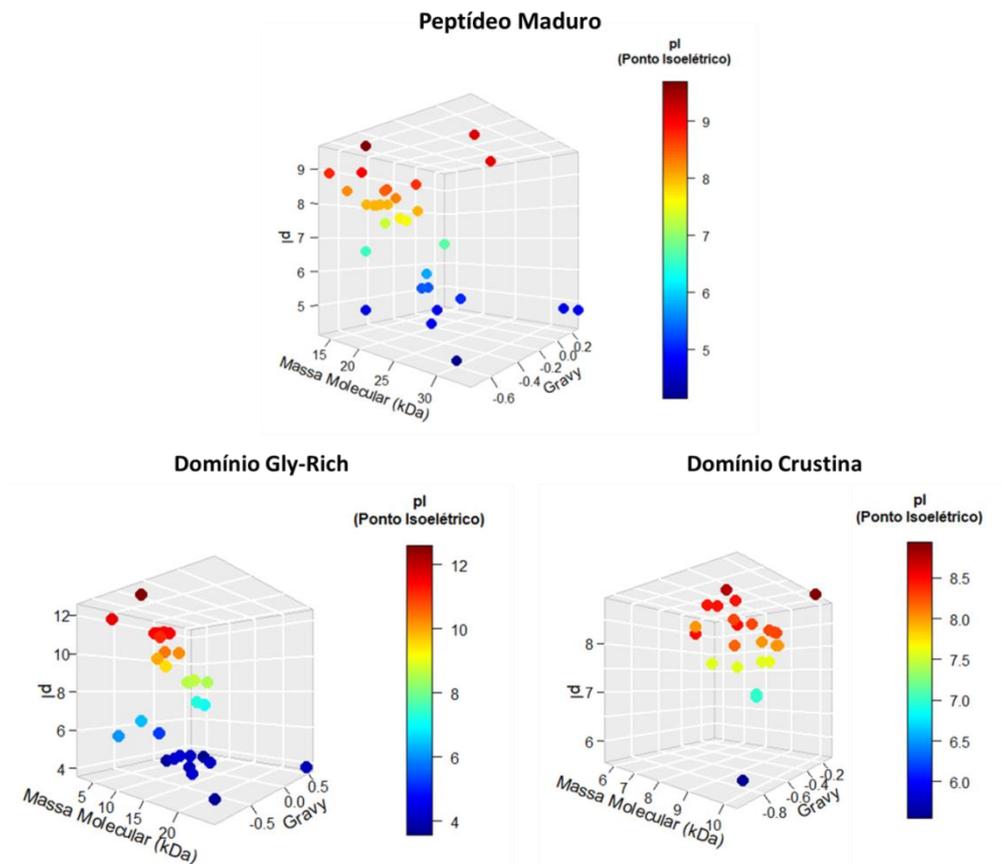
Além disso, outro mecanismo para aumento de diversidade molecular associada à existência de múltiplos éxons é a possibilidade da ocorrência de *splicing* alternativo,

mecanismo que torna possível o surgimento de múltiplas proteínas a partir de um único gene codificante (WANG *et al.*, 2015). Contudo, essas pressuposições representam especulações de eventos evolutivos e exigem análises em genômica comparativa com outras espécies de crustáceos para apropriada inferência da condição ancestral tornando, então, possível reconhecer se os eventos foram associados à perda ou ganho de regiões intrônicas capazes de prover, finalmente, a diversidade estrutural observada entre os genes codificantes.

3.2.4. As crustinas do Tipo II são bioquimicamente diversas e exibem características moleculares contrastantes

Longe de exibirem rigorosamente uma identidade molecular única (*i.e.*, polipeptídeos precursores de natureza catiônica e anfipática), as crustinas do Tipo II de *L. vannamei* apresentaram um amplo espectro de características físico-químicas. As análises *in silico* mostraram que 14 sequências (40%) foram preditas como polipeptídeos aniônicos em pH fisiológico, em virtude de seus pontos isoelétricos teóricos (*pI*) inferiores a 7,0. Três crustinas exibiram uma carga líquida neutra (*pI* 7,3-7,7), enquanto que 18 sequências (51,4%) mostraram uma natureza catiônica (*pI* >7,9). De maneira geral, todos os genes codificaram precursores compostos por um peptídeo sinal de tamanho (ca. 16-24 aac) e composição de aminoácidas variáveis seguidos por polipeptídeos maduros que exibiram distintos níveis de hidropaticidade, conforme indicado pela diversidade dos valores de GRAVY (do inglês, *Grand Average of Hydrophathicity*). Além disso, os peptídeos maduros diferiam substancialmente tanto no tamanho de suas cadeias peptídicas quanto na massa molecular, variando desde polipeptídeos de 12,4 kDa contendo 147 resíduos aminoácidos até moléculas de 33,2 kDa (362 resíduos de aminoácidos) (**Figura 18**). A variedade de massas moleculares observada entre as crustinas do Tipo II se deve predominantemente à diferença no comprimento de suas regiões amino-terminais, que variaram de 49 a 280 resíduos aminoácidos. Em conformidade com o descrito na literatura, essas regiões foram compostas por módulos hidrofóbicos sequenciais ricos em resíduos de glicina, como o pentapeptídeo “VGGGL” identificado entre crustinas do grupo 1 e o hexapeptídeo “SPVGGI” reconhecido entre crustinas do grupo 2 (**Figura 13**).

Figura 18. Representação tridimensional do espectro de diversidade bioquímica entre as crustinas do Tipo II de *L. vannamei*.



As sequências aminoacídicas relativas ao peptídeo maduro, domínio *gly-rich* e domínio crustina das crustinas do Tipo II recuperadas no proteoma de *L. vannamei* foram analisadas quanto às suas características bioquímicas e representadas graficamente em uma distribuição tridimensional utilizando os valores de pI (ponto isoelétrico), massa molecular e GRAVY (*Grand Average of Hydrophaticity*) como eixos do gráfico. Os valores de pI indicam a carga elétrica líquida do peptídeo maduro em pH fisiológico e os valores de GRAVY indicam o grau de hidropaticidade molecular. Valores positivos de GRAVY indicam hidrofobicidade, enquanto que valores negativos indicam hidrofiliidade. Cada ponto do gráfico representa uma sequência aminoacídica.

Até o presente, a expressiva maioria das crustinas do Tipo II reportadas na literatura apresenta uma natureza anfipática e catiônica (TASSANAKAJON; SOMBOONWIWAT; AMPARYUP, 2015). No entanto, os resultados obtidos aqui revelaram que as crustinas do Tipo II dispõem de uma extensa amplitude de características físico-químicas, marcadamente superior àquela descrita na literatura, variando desde polipeptídeos aniônicos hidrofílicos até moléculas catiônicas com caráter mais hidrofóbico. Nessas moléculas, o domínio *gly-rich* amino-terminal mostrou prover uma importante contribuição aos aspectos bioquímicos do peptídeo maduro, uma vez que sua plasticidade em termos de composição aminoacídica se sobrepuja àquela encontrada no domínio crustina, unidade estrutural básica que parece demonstrar menor permissividade a modificações composicionais.

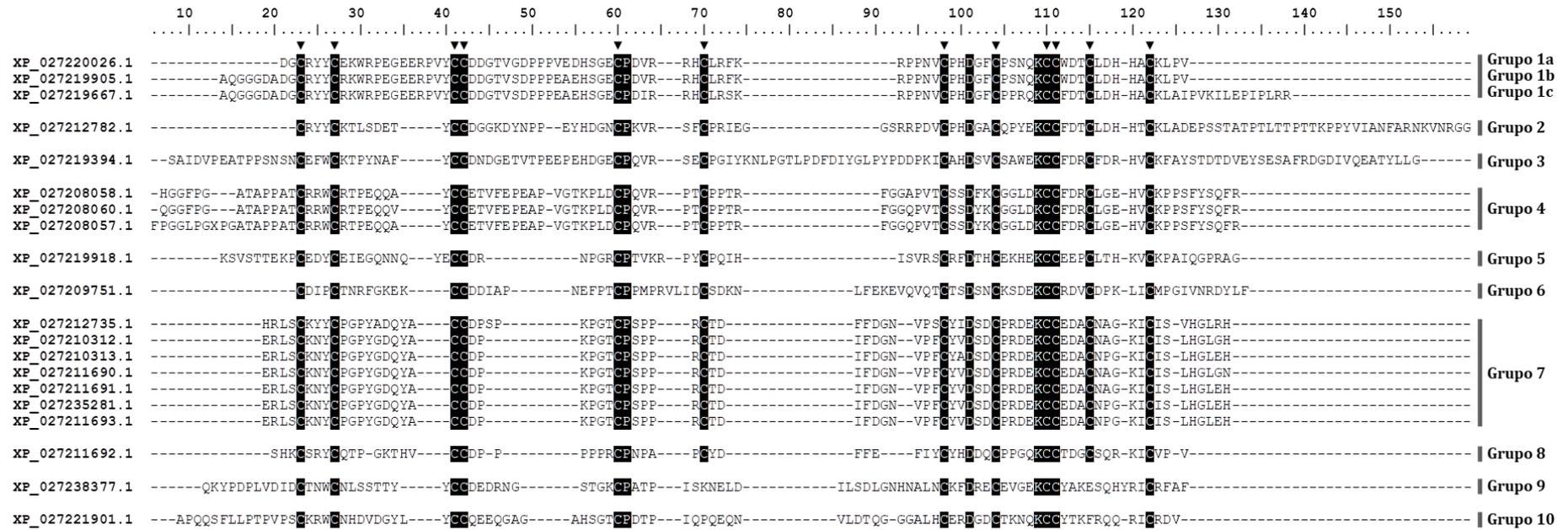
De fato, conforme mostrado na **Figura 18**, o espectro de características bioquímicas, como ponto isoelétrico teórico, massa molecular e índice de hidropaticidade, diferiu substancialmente entre os ambos os domínios e, embora esses dois extremos de diversidade tenham sido reconhecidos, o domínio *gly-rich* amino-terminal mostrou contribuir de maneira mais importante para as características bioquímicas do peptídeo maduro. Nesse sentido, um interessante rumo para estudos futuros seria investigar os aspectos funcionais dessas moléculas visando elucidar a contribuição individual desses dois domínios na atividade biológica das crustinas do Tipo II. Considerando que as características moleculares dos AMPs regem seu espectro de atividades biológicas e que modificações na composição aminoacídica podem influenciar significativamente nos seus mecanismos de ação (ZASLOFF, 2019), a diversidade molecular reconhecida entre as crustinas do Tipo II de *L. vannamei* pode sugerir uma provável participação dessas moléculas em diferentes aspectos da imunidade de camarões peneídeos ou, até mesmo, em atividades alternativas à função imune.

3.3. Identificação e caracterização das crustinas do Tipo I

3.3.1. As crustinas do Tipo I de camarões são, também, formadas por grupos distintos

As buscas *in silico* no proteoma predito de *L. vannamei* levaram à identificação de 20 sequências aminoacídicas que apresentaram uma arquitetura molecular condizente com crustinas do Tipo I (*i.e.*, peptídeos precursores compostos por uma sequência sinal seguida de um peptídeo maduro contendo uma região *cys-rich* contendo quatro resíduos de cisteína à montante de um domínio WAP conservado na região C-terminal). De maneira similar ao que foi realizado para as crustinas do Tipo II, as sequências recuperadas foram organizadas em grupos que foram definidos em função da identidade aminoacídica de suas regiões conservadas, a fim de dimensionar a diversidade molecular e investigar a existência de eventuais subtipos. Aqui, as sequências aminoacídicas que compartilharam identidades superiores a 85% na região referente ao domínio crustina foram reunidas em um grupo comum, enquanto que aquelas que apresentaram composições singulares foram separadas em grupos individuais. Conforme mostrado na **Figura 19**, as crustinas do Tipo I de *L. vannamei* puderam ser organizadas em 10 grupos distintos (Grupos 1 – 10). A maioria dos grupos foi formada por uma única sequência membro, enquanto que três grupos (Grupo 1, Grupo 4 e Grupo 7) foram constituídos por múltiplos membros que compartilharam altas identidades aminoacídicas ($\geq 85\%$).

Figura 19. Grupos de crustinas do Tipo I identificadas em *L. vannamei*.



As seqüências aminoácidas referentes ao peptídeo maduro das crustinas do Tipo I de *L. vannamei* foram alinhadas utilizando o software MAFFT v.7. Os números situados à esquerda correspondem aos identificadores das seqüências aminoácidas correspondentes. Os resíduos conservados em todas as seqüências estão destacados em preto, enquanto que os resíduos de cisteína componentes do domínio crustina estão indicados pelo símbolo ▼. Os hífens (-) representam *gaps*. A régua acima das seqüências indica a posição relativa dos resíduos aminoácidos e a linha vermelha indica o domínio WAP.

Para validar a existência das crustinas identificadas, buscas pelos seus respectivos transcritos codificantes foram realizadas em bancos transcritômicos (repositórios TSA e EST) de *L. vannamei*. Conforme revelado pelas buscas com o algoritmo tBLASTn, todas as sequências apresentaram transcritos correspondentes com altos valores de identidade ($\geq 70\%$) e confiabilidade estatística ($e\text{-value} \leq 10^{-22}$) (**Quadro 3**). Analogamente ao que foi identificado entre as crustinas do Tipo II, algumas sequências não apresentaram transcritos correspondentes com 100% de identidade aminoacídica entre seus produtos proteicos em virtude de polimorfismos na estrutura primária. Essa variabilidade intraespecífica é notória, particularmente, entre as crustinas do grupo 5 cujos valores de identidade com transcritos correspondentes nos bancos transcritômicos não ultrapassam os 75%.

Quadro 3. Validação *in silico* dos subtipos de crustinas do Tipo I reconhecidos em *L. vannamei* nos bancos transcritômicos TSA e EST desta espécie.

| Crustinas do Tipo I | | | | Validação em bancos transcritômicos | | |
|---------------------|-------------------|-------------------|--------------|-------------------------------------|-----------|------------|
| Grupo | ID proteína | ID scaffold | ID locus | GenBank | e-value | Identidade |
| Grupo 1 | XP_027220026.1 | LVANscaffold_1143 | LOC113812348 | GGKO01019693.1 | 3,84E-63 | 100% |
| | XP_027219905.1 | LVANscaffold_1143 | LOC113812245 | GFRP01010583.1 | 2,84E-62 | 100% |
| | XP_027219667.1 | LVANscaffold_1143 | LOC113811998 | GHXV01084268.1 | 1,31E-70 | 100% |
| Grupo 2 | XP_027212782.1 | LVANscaffold_1678 | LOC113805910 | HAAW01015442.1 | 1,89E-65 | 100% |
| Grupo 3 | XP_027219394.1 | LVANscaffold_2260 | LOC113811778 | GDUV01008502.1 | 1,22E-112 | 99,30% |
| Grupo 4 | XP_027208057.1 | LVANscaffold_887 | LOC113801827 | GGTO01010830.1 | 5,99E-71 | 93,70% |
| | XP_027208058.1 | LVANscaffold_887 | LOC113801828 | | | |
| | XP_027208060.1 | LVANscaffold_887 | LOC113801830 | | | |
| Grupo 5 | XP_027219918.1 | LVANscaffold_1143 | LOC113812253 | GGQV01047433.1 | 3,95E-29 | 74,26% |
| Grupo 6 | XP_027209751.1 | LVANscaffold_1400 | LOC113803231 | GFRP01006565.1 | 5,38E-63 | 100% |
| Grupo 7 | XP_027235281.1 | LVANscaffold_1034 | LOC113826552 | GGQV01086552.1 | 1,17E-54 | 98,90% |
| | XP_027210312.1 | LVANscaffold_1454 | LOC113803699 | | | |
| | XP_027210313.1 | LVANscaffold_1454 | LOC113803700 | | | |
| | XP_027211690.1 | LVANscaffold_1584 | LOC113804968 | | | |
| | XP_027211691.1 | LVANscaffold_1584 | LOC113804971 | | | |
| | XP_027211693.1 | LVANscaffold_1584 | LOC113804973 | | | |
| XP_027212735.1 | LVANscaffold_1678 | LOC113805871 | | | | |
| Grupo 8 | XP_027211692.1 | LVANscaffold_1584 | LOC113804972 | GHXV01182119.1 | 4,37E-22 | 84,50% |
| Grupo 9 | XP_027238377.1 | LVANscaffold_597 | LOC113829403 | GFRP01024607.1 | 6,25E-74 | 100% |
| Grupo 10 | XP_027221901.1 | LVANscaffold_2487 | LOC113814028 | GGQV01109124.1 | 5,19E-68 | 100% |

O repertório de crustinas do Tipo I identificado em *L. vannamei* contrasta com o modesto número de relatos dessas crustinas em camarões peneídeos que, até o presente, parece se restringir a apenas duas espécies (*e.g.*, *P. monodon* e *M. japonicus*). De fato, duas crustinas do Tipo I foram identificadas no camarão *P. monodon*, referidas como *carcininPm1* e *carcininPm2*, enquanto que cinco crustinas desse Tipo foram descritas no camarão japonês *M. japonicus* (*i.e.*, *MjCru I-1/5*) (DONPUDSA *et al.*, 2014; JIANG *et al.*, 2015; LIU *et al.*,

2015). Isso vai de encontro com o que se admite na literatura, visto que as crustinas do Tipo I (carcininas) são predominantemente encontradas em crustáceos pleociemados, enquanto que o grupo majoritário de crustinas entre os camarões peneídeos corresponde ao Tipo II, marcadas pela presença adicional de uma região *gly-rich* amino-terminal (TASSANAKAJON; SOMBOONWIWAT; AMPARYUP, 2015).

A identificação de, pelo menos, 20 crustinas do Tipo I em *L. vannamei* representa uma novidade na distribuição taxonômica dessa família de AMPs e oferece *insights* acerca da evolução desse grupo particular de crustinas em crustáceos decápodes. Além disso, a vasta diversidade de crustinas do Tipo I em camarões, outrora consideradas minoritárias e pouco diversas em Dendrobranchiata, propõe interessantes reflexões a respeito da real diversidade dessas crustinas em crustáceos decápodes não peneídeos, como caranguejos, lagostas e lagostins (Subordem Pleocyemata), nos quais a maior parte das crustinas desse Tipo foi descrita. Os resultados obtidos ressaltam, portanto, a necessidade de estudos que visem à exploração de crustinas do Tipo I em bancos genômicos/transcritômicos de outras espécies de crustáceos a fim de alcançar, em última análise, uma ampla compreensão de sua diversidade molecular.

Uma vez estimado o repertório de crustinas do Tipo I em *L. vannamei*, buscas por sequências homólogas em camarões peneídeos foram realizadas, com o intuito de investigar se essa diversidade molecular ocorre unicamente em *L. vannamei* ou estende-se de maneira mais ampla pela família Penaeidae. Os resultados obtidos a partir das buscas *in silico* revelaram a existência de potenciais ortólogos em pelo menos uma outra espécie de camarão peneídeo, indicando que os diferentes subtipos de crustinas do Tipo I reconhecidos em *L. vannamei* estão igualmente presentes em diferentes espécies dentro da família (**Tabela 3**). Oito dentre as 20 crustinas do Tipo I foram recentemente descritas por Li e colaboradores (2020) após uma exploração sistemática em transcritomas de *L. vannamei*, enquanto que dois (*e.g.*, Grupos 4 e 8) representaram crustinas do Tipo I inéditas nessa espécie. Curiosamente, a maioria das crustinas identificadas representaram versões em *L. vannamei* de crustinas reportadas na literatura. Por exemplo, as crustinas que compuseram o grupo 1 foram ortólogas de *MjCru I-1*, descrita no camarão japonês *M. japonicus*, enquanto que a crustina do grupo 2 foi ortóloga de *MjCru I-2* e *carcininPm2*, presentes em *M. japonicus* e *P. monodon*, respectivamente. A crustina pertencente ao grupo 3 mostrou ortologia com *Fc-crus3* reportada no camarão chinês *F. chinensis* e aquela do grupo 5 foi ortóloga de *MjCru I-3* de *M. japonicus*. Finalmente, o grupo 9 mostrou ortologia com duas crustinas descritas em *M.*

japonicus (MjCru I-4) e *P. monodon* (carcininPm1) e o grupo 10 foi ortólogo de MjCru I-5 (Tabela 3).

Tabela 3. Lista de prováveis ortólogos das crustinas do Tipo I identificadas em *L. vannamei* em diferentes espécies de camarões peneídeos.

| Grupo | Gene | Ortólogo | Espécie | Identidade |
|----------|--|--------------------------|------------------------|------------|
| Grupo 1 | LOC113812348 LOC113812245 LOC113811998 | GGLH01123907.1 | <i>P. monodon</i> | 74% |
| | | MjCruI-1 ^c | <i>M. japonicus</i> | 71% |
| | | GFRT01014862.1 | <i>F. penicillatus</i> | 77% |
| | | GGTU01013196.1 | <i>P. longistylus</i> | 69% |
| Grupo 2 | LOC113805910 | carcininPm2 | <i>P. monodon</i> | 70% |
| | | MjCruI-2 ^d | <i>M. japonicus</i> | 70% |
| Grupo 3 | LOC113811778 | GFGA01032989.1 | <i>P. monodon</i> | 85% |
| | | GGTR01011385.1 | <i>P. esculentus</i> | 80% |
| | | GFRT01011061.1 | <i>F. penicillatus</i> | 85% |
| | | Fc-crus 3 ^a | <i>F. chinensis</i> | 75% |
| Grupo 4 | LOC113801830 | GGLH01119800.1 | <i>P. monodon</i> | 85% |
| | | GHUY01006134.1 | <i>F. chinensis</i> | 92% |
| | | GFRT01002739.1 | <i>F. penicillatus</i> | 96% |
| | | GGTN01013329.1 | <i>P. merguensis</i> | 93% |
| Grupo 5 | LOC113812253 | JR220004.1 | <i>P. monodon</i> | 69% |
| | | MjCruI-3 ^d | <i>M. japonicus</i> | 66% |
| | | GFRT01009453.1 | <i>F. penicillatus</i> | 70% |
| | | GIXQ01015951.1 | <i>P. merguensis</i> | 69% |
| Grupo 6 | LOC113803231 | GGLH01182894.1 | <i>P. monodon</i> | 72% |
| | | GHUY01006557.1 | <i>F. chinensis</i> | 69% |
| | | GGTN01003116.1 | <i>P. merguensis</i> | 70% |
| Grupo 7 | LOC113805871 | JR206254.1 | <i>P. monodon</i> | 83% |
| Grupo 8 | LOC113804972 | GFRT01021130.1 | <i>F. penicillatus</i> | 84% |
| | | GGLH01132379.1 | <i>P. monodon</i> | 76% |
| Grupo 9 | LOC113829403 | carcininPm1 ^b | <i>P. monodon</i> | 86% |
| | | MjCruI-5 ^d | <i>M. japonicus</i> | 72% |
| | | GFRT01005273.1 | <i>F. penicillatus</i> | 84% |
| Grupo 10 | LOC113814028 | GGLH01078363.1 | <i>P. monodon</i> | 89% |
| | | MjCruI-4 ^d | <i>M. japonicus</i> | 68% |
| | | GHUY01006495.1 | <i>F. chinensis</i> | 87% |
| | | GFRT01020362.1 | <i>F. penicillatus</i> | 85% |
| | | GGTV01013523.1 | <i>F. semisulcatus</i> | 89% |

Os valores de identidade aminoacídica são referentes à região do domínio crustina. ^aSUN *et al.*, 2010. ^bDONPUDSA *et al.*, 2014. ^cLIU *et al.*, 2015. ^dJIANG *et al.*, 2015.

3.3.2. A arquitetura molecular das crustinas do Tipo I não é exclusiva às carcininas

A diversidade de crustinas do Tipo I identificada em camarões peneídeos levantou questões relativas às suas origens, bem como às relações filogenéticas existentes entre cada um dos grupos identificados. Portanto, visando à compreensão dessas relações evolutivas, as regiões conservadas referentes ao domínio crustina foram alinhadas com a região

correspondente em diferentes crustinas descritas na literatura, incluindo autênticas carcininas (Tipo I), crustinas do Tipo II, crustinas dos Tipos III e IV (grupo externo), e analisadas filogeneticamente. Conforme revelado pelas reconstruções filogenéticas, as crustinas do Tipo I de camarões peneídeos não formaram um clado único e comum às autênticas carcininas presentes em camarões e caranguejos, mas compuseram um grupo diverso com múltiplas origens evolutivas (**Figura 20A**). De fato, as análises apontaram para a existência de, pelo menos, oito clados distintos de crustinas do Tipo I em peneídeos que acomodaram as crustinas de *L. vannamei* juntamente com seus respectivos ortólogos (**Figura 20B**).

Curiosamente, apenas as crustinas dos grupos 1 (XP_027220026.1, XP_027219905.1, XP_027219667.1) e 2 (XP_027212782.1), juntamente com seus ortólogos, mostraram-se filogeneticamente próximas às carcininas presentes em pleociemados, classicamente descritas na literatura como autênticas crustinas do Tipo I (**Figura 20A**). Por essa razão, essas sequências foram particularmente categorizadas como *as verdadeiras carcininas* de camarões peneídeos, em detrimento dos oito demais grupos que formaram clados independentes e filogeneticamente divergentes. De fato, algumas das sequências são verdadeiramente membros de outros Tipos conhecidos de crustinas. Esse é o caso, por exemplo, do grupo 4 (XP_027208060.1), que se agrupou com as autênticas crustinas do Tipo II, e dos grupos 9 e 10 (XP_027238377.1 e XP_027221901.1, respectivamente), estruturalmente similares às carcininas, mas que carecem de dois resíduos de cisteína residentes no domínio WAP (Cys2 e Cys7). Assim, os resultados mostraram claramente que, embora compartilhem uma arquitetura molecular comum, incluindo a presença do domínio crustina contendo 12 resíduos conservados de cisteína (região *cys-rich* seguida por um único domínio WAP) e a ausência de domínios N-terminais adicionais, a maioria das “crustinas Tipo I” de camarões peneídeos não representa autênticas carcininas, mas formam novos grupos até o presente não descritos na literatura.

A identificação desses novos grupos traz importantes implicações para o atual sistema de classificação que reúne as crustinas em função de suas arquiteturas moleculares (SMITH *et al.*, 2008; TASSANAKAJON; SOMBOONWIWAT; AMPARYUP, 2015). De fato, os resultados obtidos indicaram que o perfil estrutural das crustinas do Tipo I não responde necessariamente à relação filogenética, mas deriva, em grande medida, de uma aparente conservação estrutural em genes com diferentes origens evolutivas. Nesse contexto, a descoberta dessa larga diversidade em *L. vannamei*, reforçada pela identificação de potenciais ortólogos em camarões peneídeos, revelou que, as então chamadas “crustinas Tipo I” de Penaeidae, compreendem, na verdade, um grupo altamente heterogêneo de sequências com múltiplas origens evolutivas compostas principalmente por membros filogeneticamente distantes das autênticas carcininas descritas em Pleocyemata.

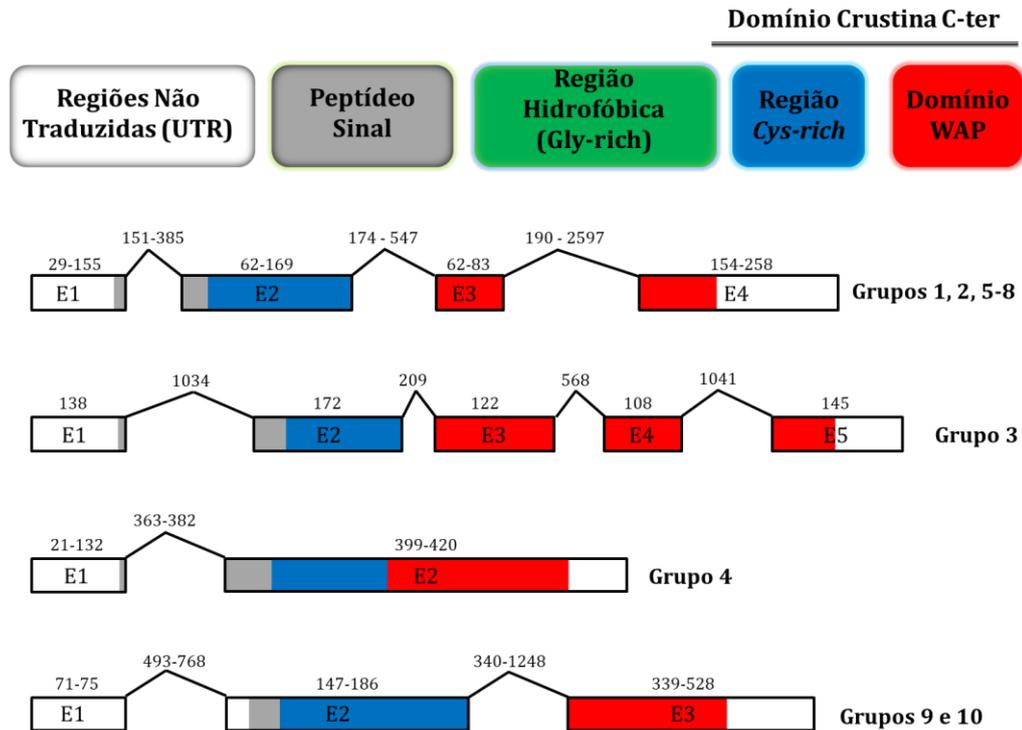
3.3.3. As autênticas carcininas e crustinas similares ao Tipo I são codificadas por genes distintos

De maneira geral, tanto as carcininas autênticas (Grupos 1 e 2) quanto as demais sequências estruturalmente semelhantes às crustinas do Tipo I (Grupos 3 a 10) compreenderam um grupo multigênico e diverso que abrangeu pelo menos 20 genes distintos amplamente distribuídos pelo genoma. De fato, os genes codificantes foram encontrados em, pelo menos, dez diferentes *scaffolds* genômicos, sendo o *scaffold* LVANscaffold_1143 (NW_020868447.1), particularmente, aquele com o maior número de genes (*e.g.*, LOC113812348, LOC113811998, LOC113812245 e LOC113812253). Os genes mostraram um comprimento médio de ca. 2,8 kb e apresentaram organizações estruturais distintas quanto à presença e quantidade de íntrons e éxons (**Figura 21**).

A maioria dos genes foi composta por quatro éxons separados por três íntrons, sendo todos os íntrons flanqueados pelos sítios canônicos de *splicing* GT/AG. No entanto, embora essa distinção estrutural tenha sido observada na organização gênica, os genes mostraram um padrão estrutural comum quanto à organização de regiões exônicas que codificam módulos específicos nos peptídeos. Por exemplo, a região *cys-rich* comum a todas as sequências mostrou ser codificada em um único éxon que incluiu igualmente uma larga porção do peptídeo sinal, enquanto que o domínio WAP C-terminal mostrou ser codificado separadamente em dois éxons adicionais (**Figura 21**). Na maioria dos genes, o primeiro éxon cobriu integralmente a região não traduzida 5' (*Untranslated Region 5'* ou 5'-UTR) e os primeiros códons do peptídeo sinal, enquanto a sequência codificante (*Coding Sequence* ou CDS) encontrou-se distribuída ao longo dos éxons restantes. Via e regra, o último éxon

mostrou codificar a maior parte do domínio WAP e toda a região não traduzida 3' (3'-UTR), exceto para o gene LOC113801830 que codificou o peptídeo maduro integralmente em um único éxon (**Figura 21**).

Figura 21. Organização genômica dos genes codificantes de crustinas do Tipo I identificados em *L. vannamei*.



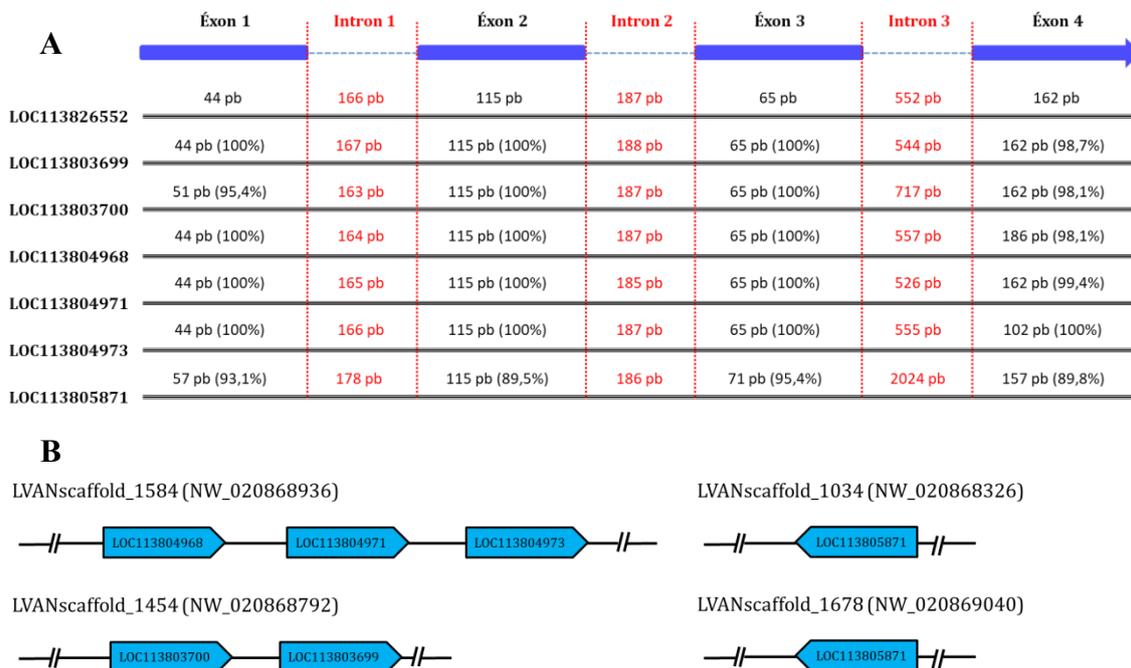
Os retângulos e as linhas pretas representam, respectivamente, os éxons (**E1**: éxon 1, **E2**: éxon 2, **E3**: éxon 3, **E4**: éxon 4, **E5**: éxon 5) e os íntrons dos genes codificantes. O tamanho das sequências está representado em pares de bases. O comprimento dos íntrons e éxons não está representado proporcionalmente.

A recuperação das sequências gênicas proveu, ainda, múltiplas evidências de que parte do amplo repertório identificado em *L. vannamei* é produto de sucessivos processos evolutivos que se encontram ainda em andamento, como eventos de duplicação gênica seguida de mutação. Essas evidências são suportadas (i) pela alta similaridade entre as sequências aminoacídicas e (ii) pela organização estrutural comum dos genes que codificam crustinas pertencentes ao mesmo subtipo. De fato, as análises de diversidade genética revelaram, por exemplo, que as crustinas do grupo 7 são codificadas em pelo menos sete *loci* distintos cujos genes compartilham uma identidade superior a 90% em nível nucleotídico na região codificante e em nível aminoacídico em seus produtos proteicos. Consistentemente, os sete genes retêm uma organização estrutural comum, com o mesmo número (e comprimentos

similares) de íntrons e éxons (à exceção da presença de dois códons adicionais no éxon 3 do gene LOC113805871) (**Figura 22A**).

Contudo, de maneira curiosa, a posição relativa desses genes no genoma impõe dúvidas acerca desse processo: os genes codificantes das crustinas do grupo 7 não se encontram sequencialmente posicionados em um *scaffold* genômico comum, conforme esperado no caso de eventos recentes de paralogia, mas residem em *loci* distribuídos em pelo menos quatro diferentes *scaffolds* genômicos separados por longos segmentos de DNA (**Figura 22B**). Nesse cenário, é plausível inferir desses genes uma origem comum e recente derivada de sucessivos eventos de duplicação seguidos de sua transposição para diferentes regiões do genoma, possivelmente como elementos transponíveis. Contudo, tendo em vista que o fracionamento do genoma do *L. vannamei* em *scaffolds* genômicos não responde por sua distribuição cromossômica, na qual diferentes *scaffolds* podem representar fragmentos de um mesmo cromossomo, a inferência dessas possíveis transposições gênicas é incapaz de determinar com precisão se esses eventos aconteceram a nível intra ou intercromossômico.

Figura 22. Comparação das sequências nucleotídicas dos genes codificantes das crustinas do grupo 7 (Tipo I) e sua distribuição topológica no genoma de *L. vannamei*.

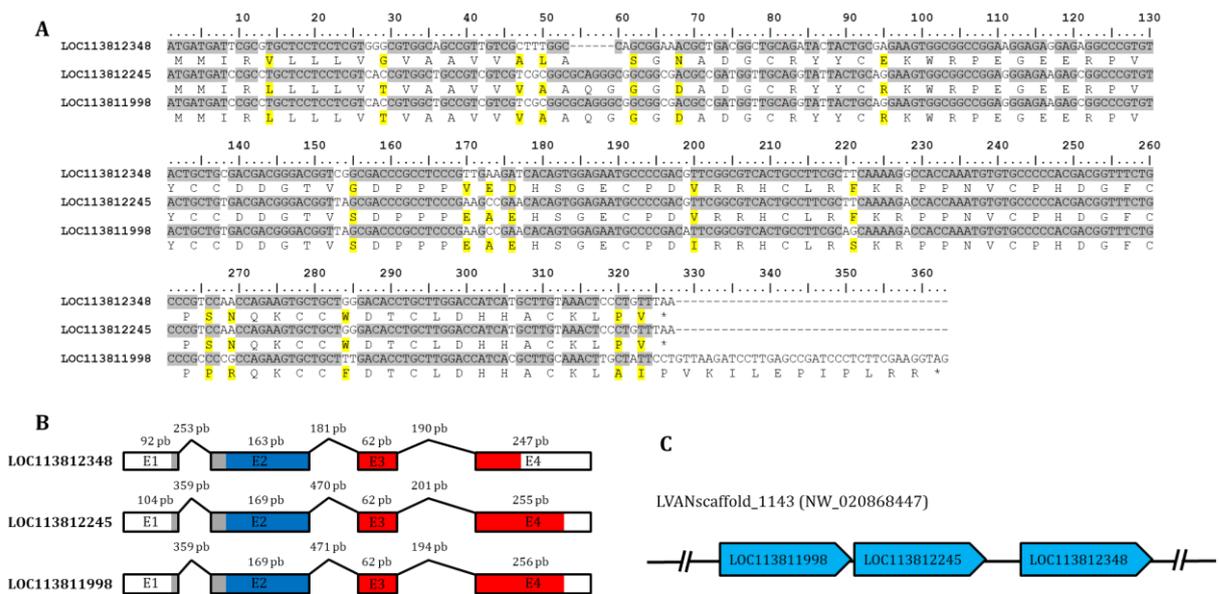


Os composição nucleotídicas referente aos íntrons e éxons dos genes codificantes das crustinas do grupo 7 (Tipo I) foi comparada e sua distribuição topológica no genoma foi investigada. **(A)** Comparação entre as sequências dos sete prováveis parálogos LOC113826552, LOC113803699, LOC113803700, LOC113804968, LOC113804971, LOC113804973 e LOC113805871. As porcentagens mostradas são referentes à identidade nucleotídicas usando a sequência LOC113826552 como referência. Os comprimentos de íntrons e éxons não estão representados proporcionalmente. **(B)** Representação esquemática de parte dos *scaffolds* 1584, 1454, 1034 e 1678 do genoma de *L. vannamei* (GenBank: ASM378908). As setas representam os genes codificantes das

crustinas do grupo 7 (Tipo I). As setas apontando para esquerda ou direita representam o frame de transcrição negativo ou positivo, respectivamente.

Um exemplo adicional que reforça essa hipótese é provido pelas crustinas do grupo 1, onde os três genes identificados (LOC113812348, LOC113812245 e LOC113811998) compartilham pelo menos 84% de identidade tanto na região codificante quanto na estrutura primária de seus produtos proteicos. Os genes apresentam, ainda, rigorosamente a mesma organização estrutural quanto ao número e tamanho de seus íntrons e éxons e residem sequencialmente no mesmo *scaffold* genômico (LVANscaffold_1143), em posições vizinhas (**Figura 23**), sugerindo uma provável relação de homologia derivada de eventos recentes de duplicação.

Figura 23. Comparação da composição aminoacídica e arquitetura molecular dos genes do grupo 1 (Tipo I).



A estrutura primária, arquitetura molecular gênica e distribuição topológica das crustinas do grupo 1 (Tipo I) no genoma foram investigados e comparados entre si. **(A)** Comparação da composição nucleotídica e aminoacídica das crustinas do grupo 1 (Tipo I). Os resíduos nucleotídicos conservados estão destacados em cinza. Os resíduos aminoacídicos divergentes estão destacados em amarelo. Os hífens (-) indicam *gaps* e os asteriscos (*) indicam códons de parada. A régua acima indica a posição relativa dos resíduos nucleotídicos. **(B)** Comparação da arquitetura molecular dos genes codificantes das crustinas do grupo 1 (Tipo I). Os retângulos e as linhas pretas representam, respectivamente, os éxons (E1: éxon 1, E2: éxon 2, E3: éxon 3, E4: éxon 4) e os íntrons dos genes codificantes. Os comprimentos dos éxons e íntrons estão destacados. As regiões codificantes estão representadas em diferentes cores: em cinza está representada a região codificante no peptídeo sinal, em azul a região codificante do domínio *cys-rich* e em vermelha a região codificante do domínio WAP. Em branco estão representadas as regiões não traduzidas (UTR, *untranslated regions*). **(C)** Representação esquemática de parte do *scaffold* 1143 do genoma de *L. vannamei* (GenBank: ASM378908). As setas representam os genes codificantes das crustinas do grupo 1 (Tipo I).

Conforme mencionado, peptídeos antimicrobianos em camarões peneídeos são geralmente codificados por famílias multigênicas, geradas a partir da duplicação sucessiva de

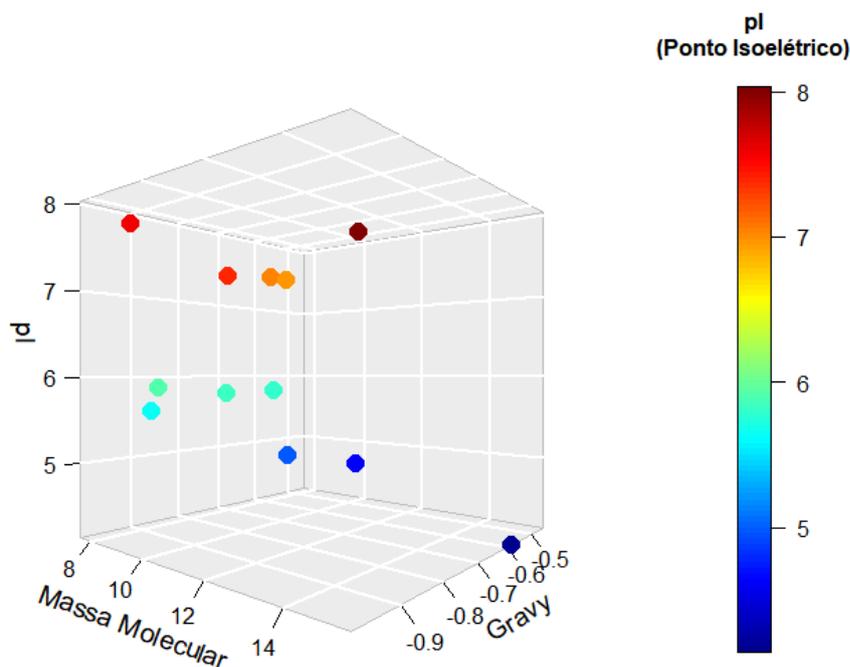
genes individuais cujas cópias evoluem como entidades independentes capazes de desempenhar funções diversas. Esse processo de duplicação gênica seguido de mutação responde, em grande medida, pela surpreendente diversidade molecular encontrada entre peptídeos antimicrobianos (LAZZARO; ZASLOFF; ROLFF, 2020). Em eucariotos, eventos de duplicação gênica podem ocorrer por meio de diferentes mecanismos que emergem principalmente a partir de erros durante a replicação ou reparo do DNA (ZHANG, 2003). Nesses eventos, os genes podem ser duplicados em tandem, resultando no surgimento de dois genes parálogos adjacentes que são usualmente separados por curtos trechos de sequência intergênica ou serem produtos de duplicação da região em que estão embutidos, de modo a serem colocados a distâncias ligeiramente maiores em comparação com as duplicações em tandem, geralmente separados uns dos outros por um ou mais genes (GERDOL; GRECO; PALLAVICINI, 2019). Além disso, um processo alternativo associado ao surgimento de genes parálogos depende da atividade de elementos transponíveis, como os transposons ou retrotransposons. Nesse caso, os genes duplicados são normalmente dispersos em diferentes regiões do genoma e situados distante de seus parálogos, possivelmente em diferentes cromossomos (CUSACK; WOLFE, 2007).

Os resultados obtidos das análises gênicas sugeriram que eventos evolutivos como duplicação gênica em tandem e a atividade de elementos transponíveis podem ter sido prevalentes e inerentemente responsáveis pela expansão e diversificação molecular de carcininas e crustinas similares ao Tipo I em *L. vannamei*, conforme visto entre as crustinas dos grupos 1 e 7, respectivamente. Essa conclusão suporta-se na constatação da existência de uma alta identidade nucleotídica e aminoacídica (>80%) entre os membros desses grupos, bem como na organização estrutural comum de seus respectivos genes codificantes. Aliás, é plausível inferir dois cenários distintos para explicar a elevada identidade observada entre as sequências: (i) os eventos de paralogia têm uma origem recente e a diversificação molecular encontra-se ainda em andamento ou (ii) esses eventos remontam a um passado distante, mas em virtude de pressões evolutivas conservativas para manutenção de suas funções biológicas, os genes retiveram adaptativamente suas informações genéticas. A precisão desses eventos, contudo, exige análises em genômica comparativa, a fim de mapear se tais genes foram igualmente expandidos em outras espécies de camarões peneídeos ou se esse foi um evento exclusivo de *L. vannamei*.

3.3.4. As autênticas carcininas e crustinas similares ao Tipo I compreendem um grupo bioquimicamente diverso

Analogamente às crustinas do Tipo II, as carcininas e crustinas similares ao Tipo I identificadas em *L. vannamei* apresentaram um amplo espectro de propriedades bioquímicas. As análises *in silico* mostraram que a maioria delas (e.g., grupo 1a-c, grupo 3, grupo 6, grupo 7, grupo 9 e grupo 10) foi predita como polipeptídeos aniônicos em pH fisiológico, em virtude de seus pontos isoeletricos teóricos (pI) abaixo de 7,0. Em contraste, quatro crustinas (e.g., grupo 1c, grupo 2, grupo 5 e grupo 8) exibiram uma carga líquida neutra (pI 7-7.6), enquanto uma única sequência (grupo 4) mostrou uma natureza catiônica ($pI >7,9$). De maneira geral, todos os genes codificaram precursores compostos por um peptídeo sinal de tamanho (ca. 15-27 aac) e composição de aminoácidas variáveis seguidos por polipeptídeos maduros hidrofílicos com graus distintos de hidropaticidade, conforme indicado pela diversidade dos valores de GRAVY (do inglês, *Grand Average of Hydrophaticity*). Além disso, os peptídeos maduros diferiam substancialmente tanto no tamanho de suas cadeias peptídicas quanto na massa molecular, variando desde pequenos polipeptídeos de 7,5 kDa contendo 67 resíduos aminoácidos até moléculas de 15,4 kDa (137 resíduos de aminoácidos) (**Figura 24**).

Figura 24. Representação tridimensional do espectro de diversidade bioquímica entre as carcininas e crustinas similares ao Tipo I de *L. vannamei*.



As carcininas e crustinas similares ao Tipo I foram analisadas quanto às suas características bioquímicas e representadas graficamente em uma distribuição tridimensional utilizando os valores de *pI* (ponto isoelétrico), massa molecular e GRAVY (*Grand Average of Hydrophathicity*) como eixos do gráfico. Os valores de *pI* indicam a carga elétrica líquida do peptídeo em pH fisiológico e os valores de GRAVY indicam o grau de hidropaticidade molecular. Valores positivos de GRAVY indicam hidrofobicidade, enquanto que valores negativos indicam hidrofiliçidade. Cada ponto do gráfico representa uma sequência aminoacídica.

Apesar da tradicional natureza catiônica das carcininas em pleociemados, duas das quatro autênticas carcininas de *L. vannamei* (e.g., grupos 1a e 1b) apresentaram pontos isoelétricos teóricos inferiores a 7,0 e representaram, portanto, peptídeos aniônicos. Na verdade, ambas as sequências compartilharam mais de 94% de identidade no peptídeo maduro e foram ausentes de uma cauda C-terminal rica em resíduos catiônicos encontrada unicamente nas crustinas do grupo 1c. Carcininas aniônicas foram igualmente encontradas em outras espécies de camarões peneídeos, como em *F. penicillatus* (TSA: GFRT01014862.1), *P. monodon* (GGLH01123907.1) e no camarão *Metapenaeus bennettiae* (TSA: GHDJ01005232.1). Esse aparente contraste bioquímico entre carcininas de caranguejos e camarões ilustra sua ampla plasticidade molecular e pressupõe uma possível adaptação das carcininas às diferentes pressões seletivas que imperam sobre essas duas linhagens de crustáceos decápodes. Por outro lado, o único grupo de natureza genuinamente catiônica (grupo 4) mostrou uma relação filogenética mais próxima com as crustinas Tipo II de camarões peneídeos, que são majoritariamente peptídeos catiônicos. Os resultados das análises dos parâmetros bioquímicos relativos às sequências recuperadas sugerem, portanto, que as “crustinas do Tipo I” de *L. vannamei* formam um grupo bioquimicamente diverso que inclui majoritariamente polipeptídeos aniônicos.

3.3.5. Alguns membros específicos de crustinas similares ao Tipo I são, na verdade, crustinas do Tipo II desprovidas da região *gly-rich* N-terminal

Embora todas as sequências tenham apresentado uma arquitetura molecular consistente com crustinas do Tipo I, múltiplas evidências sugerem que as sequências componentes do grupo 4 (XP_027208057.1, XP_027208058.1 e XP_027208060.1) representam, na verdade, atípicas crustinas do Tipo II desprovidas do longo domínio hidrofóbico amino-terminal rico em resíduos de glicina. Análises da estrutura primária dessas crustinas revelaram, primeiramente, uma alta similaridade com crustinas do Tipo II, incluindo a presença da assinatura conservada de crustinas do Tipo IIa de camarões peneídeos (BARRETO *et al.*, 2018). De fato, quando comparadas em nível aminoacídico as crustinas do grupo 4 exibiram um domínio crustina similar (74 a 88% de identidade) àquele presente em

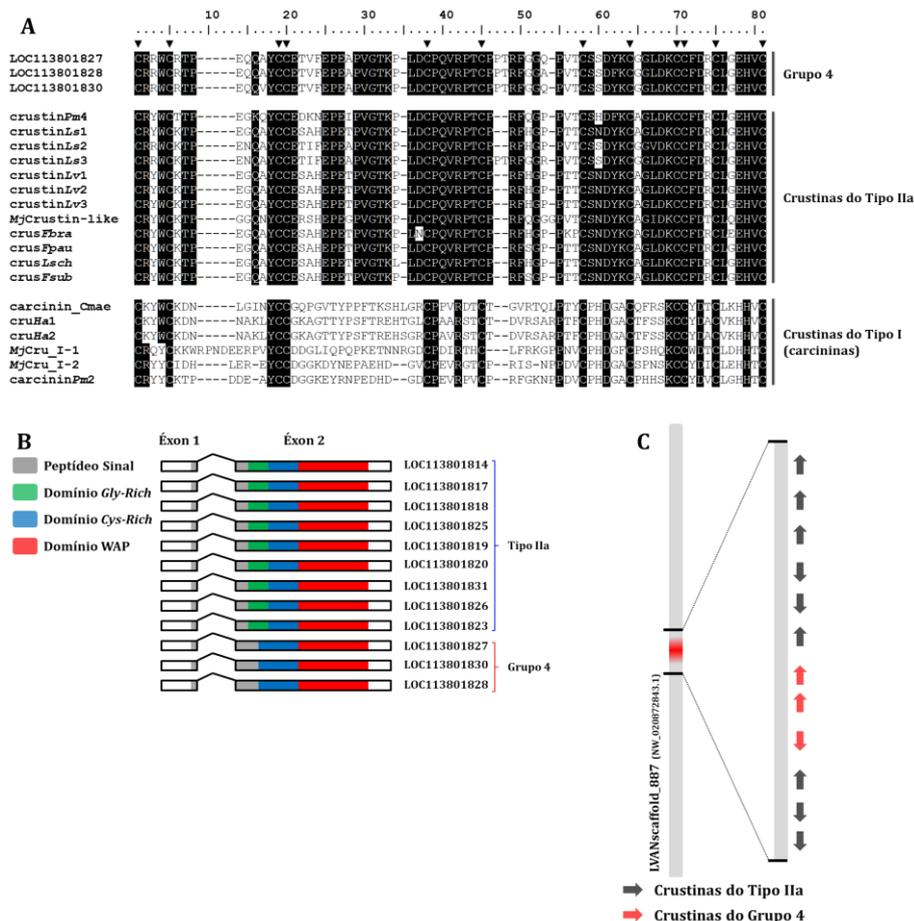
diversas crustinas do Tipo IIa descritas em camarões, incluindo crustin $Pm4$ de *P. monodon*, Crustins Ls de *L. setiferus*, Crustins Lv de *L. vannamei*, “MjCrustin-like peptides” de *M. japonicus* e crus $Fpau$, crus $Lsch$, crus $Fbra$, crus $Fsub$ dos camarões nativos *F. paulensis*, *L. schmitti*, *F. brasiliensis* e *F. subtilis*, respectivamente (BARTLETT *et al.*, 2002; RATTANACHAI *et al.*, 2004; ROSA, R. D.; BANDEIRA; BARRACCO, 2007; TASSANAKAJON *et al.*, 2011). Comparativamente, as crustinas do grupo 4 mostraram uma diferença clara com as autênticas crustinas do Tipo I (carcininas), descritas em caranguejos e camarões, cuja identidade aminoacídica no domínio crustina foi de aproximadamente 36% (**Figura 25A**).

Além da notável semelhança na composição aminoacídica, as crustinas do grupo 4 foram codificadas por três genes (*e.g.*, LOC113801827, LOC113801828 e LOC113801830) que apresentaram a mesma organização estrutural dos genes codificantes das autênticas crustinas do Tipo IIa de *L. vannamei*, então referidas como Crustin Lv (BARRETO *et al.*, 2018). De fato, em ambos os casos os genes foram compostos por dois éxons intervalados por um íntron, sendo que o primeiro éxon cobriu toda a região 5'-UTR e os dois primeiros códons do peptídeo sinal, enquanto que o segundo éxon abrangeu o restante do peptídeo sinal, o peptídeo maduro completo e a região 3'-UTR (**Figura 25B**). Além disso, a distribuição topológica dos genes do grupo 4 no genoma de *L. vannamei* corrobora sua proposição como prováveis crustinas do Tipo IIa: eles residem lado a lado das autênticas Crustin Lv no mesmo *scaffold* genômico, sendo separados por curtas regiões espaçadoras não transcritas (**Figura 25C**). Tal distribuição topológica no genoma sugere, portanto, uma possível origem gênica comum cuja diversidade deriva de sucessivos eventos de duplicação gênica. Esse conjunto de características compartilhadas entre os genes codificantes das Crustin Lv e grupo 4, aliada à similaridade bioquímica de seus produtos proteicos e proximidade filogenética descritas nas seções anteriores implica, finalmente, que as crustinas do grupo 4 representam autênticas crustinas do Tipo II em camarões peneídeos desprovidas do longo domínio hidrofóbico rico em resíduos de glicina.

A identificação dessas crustinas trouxe importantes implicações para o atual sistema de classificação das crustinas que considera o domínio *gly-rich* um atributo determinante para a classificação do Tipo II. Embora virtualmente todas as crustinas do Tipo II descritas na literatura possuam esse domínio, pouco destaque foi dado até o presente acerca do substancial polimorfismo no seu comprimento e suas reais implicações para o sistema de classificação. De maneira geral, o domínio *gly-rich* é formado pela repetição sequencial de módulos hidrofóbicos curtos (tetra- ou pentapeptídeos) ricos em resíduos de glicina, sendo que o

número de módulos define seu comprimento final. Desse modo, as crustinas do Tipo II podem apresentar desde domínios hidrofóbicos curtos contendo 21 resíduos, como na Crustin *Ls2* do camarão *L. setiferus* (BARTLETT et al., 2002), até domínios longos formados pelo acoplamento de mais de 30 módulos hidrofóbicos que somam 167 resíduos aminoacídicos, como observado na crustin *Pm4* do camarão *P. monodon* (TASSANAKAJON et al., 2011). A expressiva plasticidade estrutural observada no comprimento do domínio *gly-rich* entre diferentes crustinas do Tipo II denuncia uma fragilidade no sistema de classificação vigente e torna necessária uma revisão sistemática de seus critérios a fim de evitar eventuais incongruências.

Figura 25. Alinhamento aminoacídico comparativo entre o domínio crustina presente no grupo 4 e nas autênticas crustinas do Tipo I e II.



A estrutura primária, arquitetura genômica e distribuição topológica de crustinas do grupo 4 no genoma foram investigados e comparados com autênticas crustinas do Tipo I e IIa. **(A)** As sequências aminoacídicas referentes ao domínio crustina de autênticas crustinas do Tipo I e II foram alinhadas utilizando o software MAFFT v.7 e comparadas com a região correspondente das crustinas pertencentes ao grupo 4. Os resíduos conservados estão destacados em preto e as cisteínas componentes do domínio crustina estão evidenciadas com o símbolo ▾. Os hífen (-) representam *gaps*. A régua acima indica a posição relativa dos resíduos aminoacídicos. **(B)** Comparação da arquitetura molecular dos genes codificantes de crustinas do grupo 4 e de autênticas crustinas do Tipo IIa. Os retângulos e as linhas pretas representam, respectivamente, os éxons e o íntron dos genes

codificantes. (C) Representação esquemática de parte do *scaffold* 887 do genoma de *L. vannamei* (GenBank: ASM378908). As setas em cinza representam os genes codificantes de crustinas do Tipo IIa e as setas em vermelho representam os genes codificantes de crustinas do grupo 4. As setas apontando para esquerda ou direita representam o frame de transcrição negativo ou positivo, respectivamente.

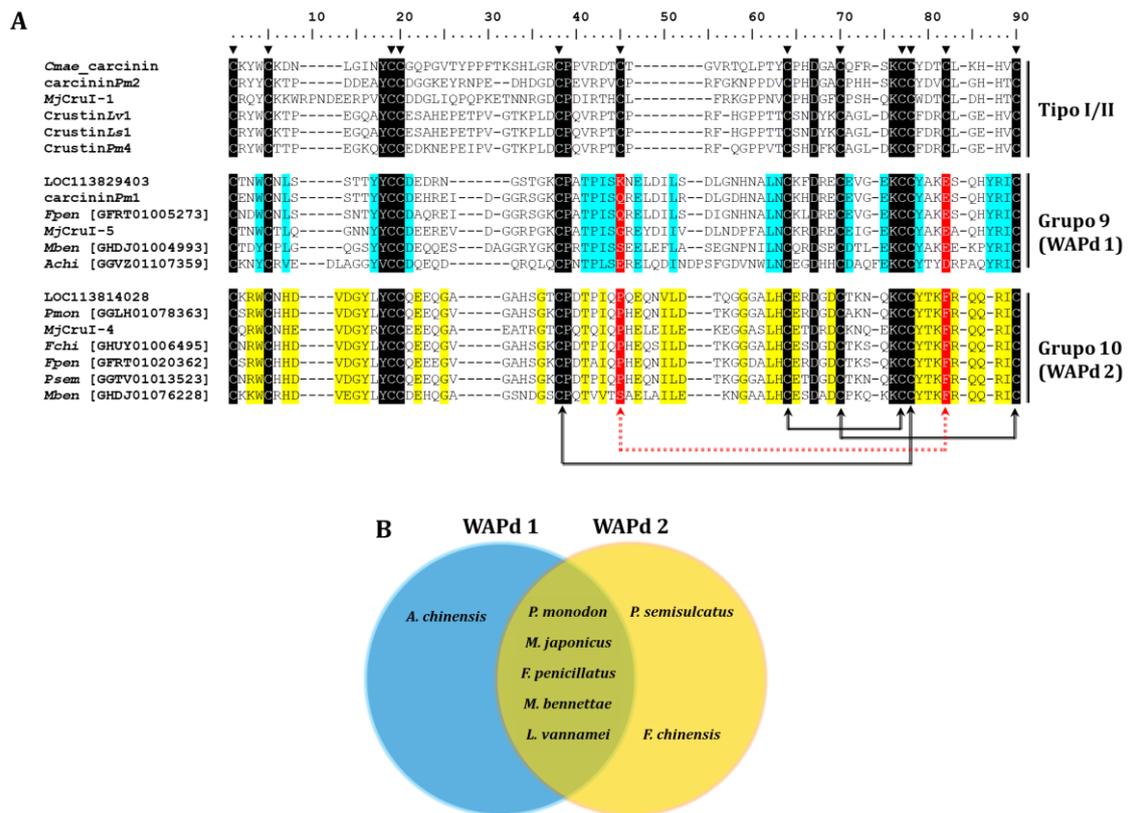
3.3.6. Crustinas atípicas contendo um domínio WAP degenerado provavelmente evoluíram especificamente na linhagem Dendrobranchiata

Entre as sequências identificadas portando a típica arquitetura molecular das crustinas do Tipo I, duas sequências (grupos 9 e 10) apresentaram um arranjo distinto de cisteínas que incluiu 10 resíduos conservados, ao invés dos convencionais 12 resíduos comumente encontrados na maioria dos membros dessa família. As análises *in silico* apontaram que ambas as sequências foram codificadas por genes cujos produtos proteicos apresentaram composições aminoácidas e parâmetros bioquímicos distintos. As sequências aminoácidas precursoras do grupo 9 foram constituídas por um peptídeo sinal de 16 resíduos seguido de um peptídeo maduro aniônico contendo 89 resíduos (10,1 kDa e *pI* de 4,99), enquanto os precursores do grupo 10 foram compostos por um peptídeo sinal de 21 resíduos aminoácidos seguido de um peptídeo maduro igualmente aniônico de 92 resíduos (10,2 kDa e *pI* 5,83). A identificação dessas formas atípicas em *L. vannamei* levou, então, à realização de buscas por sequências homólogas em crustáceos decápodes, incluindo representantes das Subordens Pleocyemata e Dendrobranchiata, a fim de investigar se suas origens remontam a espécies ancestrais da ordem Decapoda ou são, na verdade, produtos de um evento evolutivo restrito à linhagem Dendrobranchiata, que abarca todos os camarões peneídeos e seus parentes próximos.

As buscas por sequências homólogas nos bancos transcritômicos (repositórios TSA e EST) levaram à identificação de prováveis ortólogos exclusivamente Dendrobranchiata. Todas as sequências recuperadas apresentaram igualmente 10 resíduos conservados de cisteína no peptídeo maduro (**Figura 26A**). Nessas análises, quatro espécies (*i.e.*, *P. monodon*, *M. japonicus*, *F. penicillatus* e *M. bennettiae*) foram reconhecidas contendo ortólogos de ambos os grupos 9 e 10, enquanto que duas espécies (*i.e.*, *F. chinensis* e *P. semisulcatus*) foram identificadas portando apenas ortólogos de crustinas do grupo 10 (**Figura 26B**). Dentre os ortólogos identificados foram encontradas sequências previamente descritas na literatura, como a *carcininPm1* e *MjCru I-5* dos camarões asiáticos *P. monodon* e *M. japonicus*, respectivamente (grupo 9) e a *MjCru I-4* do camarão *M. japonicus* (grupo 10). Além da identificação em camarões peneídeos, uma espécie de camarão não peneídeo pertencente à família Sergestidae (*e.g.*, *Acetes chinensis*) foi reconhecida portando uma

crustina do grupo 9. A partir das sequências aminoácídicas precursoras obtidas, as regiões situadas entre o primeiro e último resíduos de cisteína foram alinhadas com autênticas crustinas do Tipo I e II e comparadas a fim de rastrear os resíduos de cisteína degenerados nas crustinas dos grupos 9 e 10. O alinhamento aminoacídico resultante revelou, de fato, a existência de dois grupos distintos que foram marcados pela conservação de resíduos aminoacídicos específicos e mostrou que o arranjo atípico de cisteínas deriva de uma degeneração no domínio WAP em virtude da substituição de dois resíduos conservados residentes nesse domínio (Cys2 e Cys7) que, em tese, estão engajados em uma ligação dissulfeto (**Figura 26**). Por essa razão, os grupos 9 e 10 foram convenientemente designados como crustinas *WAP degenerated* (ou WAPd) 1 e 2, respectivamente.

Figura 26. Assinatura aminoacídica e distribuição de crustinas WAPd 1 e WAPd 2 em Dendrobranchiata.

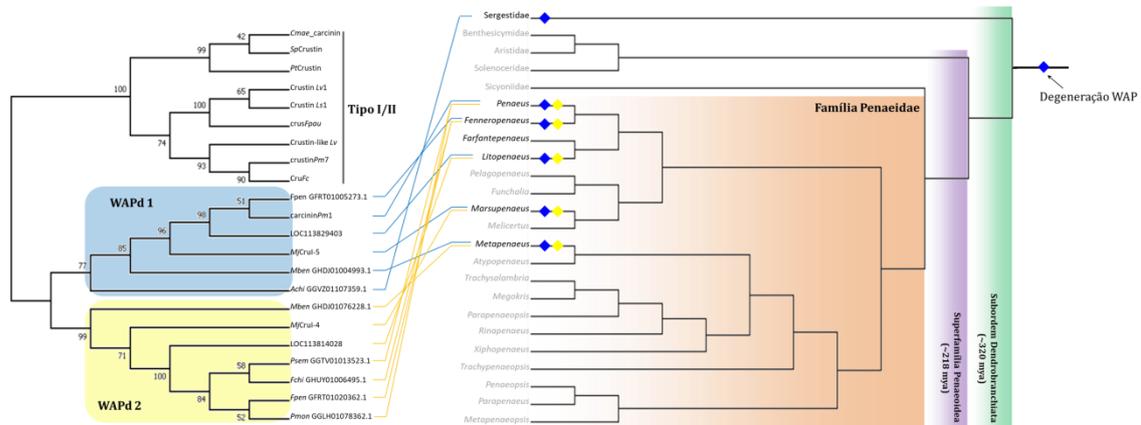


(A) Os resíduos aminoacídicos conservados em todas as sequências estão marcados em preto, sendo que os resíduos conservados de cisteína que compõem o domínio crustina estão indicados com o símbolo ▼. Resíduos específicos de WAPd 1 estão destacados em azul e os de WAPd 2 estão destacados em amarelo. Os sítios da degeneração do domínio WAP estão destacados em vermelho. As setas representam as prováveis ligações dissulfeto que formam o motivo 4-DSC. A régua acima indica a posição relativa dos resíduos aminoacídicos. *Cmae*: *C. maenas*; *Pm*: *P. monodon*; *Lv*: *L. vannamei*; *Ls*: *L. setyferus*; *Fpen*: *F. penicillatus*; *Mj*: *M. japonicus*; *Mben*: *M. bennettiae*; *Achi*: *A. chinensis*; *Fchi*: *F. chinensis*; *Psem*: *P. semisulcatus*. (B) O diagrama de Venn ilustra a distribuição de WAPd 1 e WAPd 2 nas espécies de crustáceos dendrobranquiados.

Os impactos funcionais da substituição de resíduos de cisteína relevantes para a estruturação do domínio WAP na atividade biológica das crustinas ainda são pouco conhecidos e estudos comparativos utilizando proteínas nativas ou recombinantes são fundamentais para a compreensão de suas implicações biológicas. Até o presente, um único estudo funcional comparativo foi realizado com a crustinas WAPd2 no camarão *M. japonicus* (*MjCru* I-4), no qual os autores reportaram a incapacidade de inibição de proteases bacterianas em detrimento das crustinas *MjCru* I-2 e *MjCru* I-3, portadoras de um domínio WAP completo (JIANG *et al.*, 2015), sugerindo a possível participação desses resíduos de cisteína para a manutenção de uma conformação tridimensional adequada à sua função biológica.

Além disso, as buscas por sequências homólogas revelaram que a degeneração nos resíduos Cys2 e Cys7 do domínio WAP não foi um evento evolutivo recente e parece remontar a espécies ancestrais da linhagem Dendrobranchiata. Essa hipótese encontra sustentação a partir da identificação de crustinas WAPd em espécies de camarões peneídeos e também no camarão *Acetes chinensis*, pertencente à família Sergestidae, filogeneticamente mais distante dentre aquelas incluídas na Subordem Dendrobranchiata (**Figura 27**). Além disso, outra importante informação que corrobora tal hipótese é a ausência de crustinas WAPd em crustáceos pertencentes à Subordem Pleocyemata, grupo irmão de Dendrobranchiata, o qual reúne caranguejos, lagostas, lagostins etc. Na verdade, as sequências que apresentaram maior similaridade entre pleociemados incluíram a *Pl-crustin1* do lagostim *Pacifastacus leniusculus* juntamente com seus respectivos ortólogos, sequências providas de um arranjo incomum de cisteínas, mas sem a típica degeneração no domínio WAP encontrada entre as crustinas WAPd de dendrobranquiados (JIRAVANICHPAISAL *et al.*, 2007). Nesse cenário, é plausível supor que o gene codificante de crustinas WAPd emergiu antes da diversificação da Subordem Dendrobranchiata, há aproximadamente 320 milhões de anos e experimentou uma duplicação gênica dando origem aos grupos WAPd1 e WAPd2. Contudo, a precisão temporal desse evento de duplicação não pode ser estimada ante a ausência de genomas e transcritomas disponíveis para os demais táxons da Superfamília Penaeoidea (**Figura 27**).

Figura 27. Distribuição taxonômica dos genes codificantes de WAPd1 e WAPd2 em crustáceos dendrobranquiados.



As seqüências aminoácidas de crustinas WAPd foram alinhadas e analisadas filogeneticamente através do método *Maximum Likelihood* com 1.000 replicatas. A árvore à esquerda representa a filogenia dos genes WAPd 1 e WAPd 2 e a árvore à direita corresponde à filogenia das famílias incluídas na Subordem Dendrobranchiata. Os gêneros incluídos na família Penaeidae estão evidenciados. Os gêneros cujas informações transcritômicas não estão disponíveis estão representados em cinza. Os diamantes (♦) em azul representam o gene que codifica crustinas WAPd1 e os em amarelo WAPd2.

Seqüências codificantes de crustinas WAPd1 não foram identificadas nas espécies *F. chinensis* e *P. semisulcatus*, mas puderam ser reconhecidas nas espécies congênicas *F. penicillatus* e *P. monodon*. Esse aparente contraste pode pressupor uma perda de genes WAPd2 especificamente em *F. chinensis* e *P. semisulcatus* ou, alternativamente, pode se justificar nas limitações impostas pelos bancos de seqüências que foram prospectados. Devido à ausência de genomas disponíveis, as crustinas WAPd foram recuperadas em bancos transcritômicos, os quais carregam inerentemente entraves técnicos para a identificação de novas seqüências. De fato, a depender de sua origem, como tecido biológico ou fase do desenvolvimento é possível que algumas seqüências não estejam contempladas nos transcritomas estudados, levando a interpretações equivocadas quanto à dinâmica evolutiva de alguns genes.

Contudo, diante da ampla identificação de crustinas WAPd1 em camarões penéideos, parece pouco provável que esses genes não estejam presentes em *F. chinensis* e *P. semisulcatus* e sua ausência, possivelmente, justifica-se nas limitações técnicas do transcritoma. Uma situação similar foi encontrada na espécie *A. chinensis*, na qual apenas seqüências codificantes de crustinas WAPd1 foram recuperadas. Considerando-se a distância evolutiva entre as famílias Sergestidae e Penaeidae, presumivelmente separadas há pelo menos 218 milhões de anos, a ausência de crustinas WAPd2 pode indicar que o evento de duplicação gênica ocorreu posteriormente a essa separação ou, tal como mencionado acima,

pode refletir simplesmente a ausência de sequências codificantes nos transcritomas de *A. chinensis*. Nesse contexto, estudos em genômica comparativa se fazem necessários para a apropriada elucidação da dinâmica evolutiva dos genes codificantes dessas crustinas e tornam evidentes as limitações das buscas em bancos transcritômicos para esse propósito.

3.4. Identificação e caracterização molecular de novas crustinas em *L. vannamei*

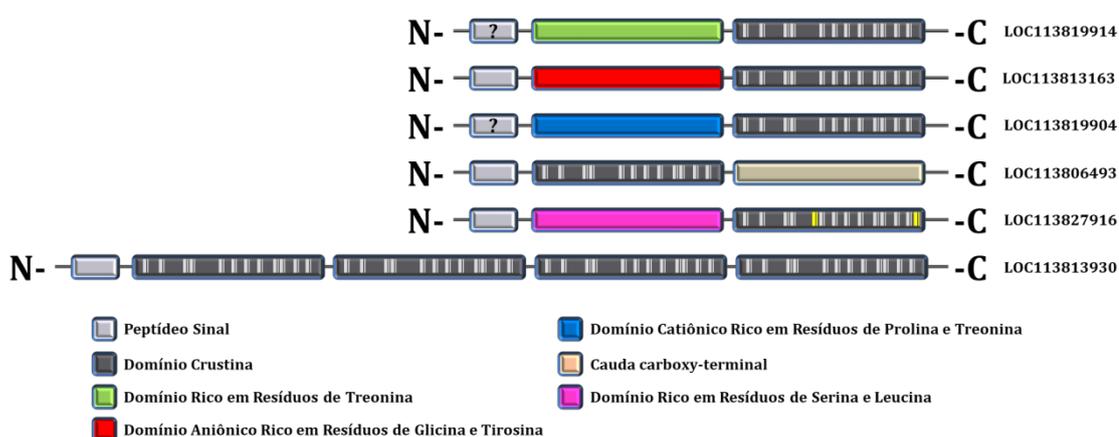
As buscas *in silico* levaram à identificação de 61 crustinas em *L. vannamei*. A maioria delas correspondeu a sequências cuja organização estrutural foi descrita na literatura e obedece à classificação prevista por Smith e colaboradores (2008), enquanto que um pequeno grupo, formado por seis sequências, incluiu crustinas que expuseram arquiteturas moleculares alternativas àquelas previstas no atual sistema de classificação.

Em virtude de suas particularidades estruturais, como a presença de assinaturas específicas, e da clara discrepância com a arquitetura molecular de crustinas reportadas na literatura, essas sequências foram consideradas como novos membros em *L. vannamei*. De fato, três sequências apresentaram em sua região amino-terminal domínios estruturais contendo composições aminoacídicas e propriedades físico-químicas incomuns. Uma sequência (*e.g.*, LOC113819914), por exemplo, foi marcada pela representação massiva de resíduos de treonina que somaram mais 60% dos resíduos nessa região, característica inédita a essa família. Alternativamente, foi identificada uma crustina (*e.g.*, LOC113813163) contendo um domínio amino-terminal fortemente aniônico (*pI*: 3,85) enriquecido em resíduos de glicina (15,5%) e tirosina (10,7%), enquanto que uma terceira sequência (*e.g.*, LOC113819904) apresentou um domínio amino-terminal fortemente catiônico (*pI*: 10) rico em resíduos de prolina (22,4%) e treonina (21,2%). Interessantemente, foi encontrada também uma crustina (*e.g.*, LOC113806493) portando uma longa região C-terminal contendo mais de 480 resíduos aminoacídicos à jusante do domínio crustina conservado (**Figura 28**).

As buscas em *L. vannamei* levaram, ainda, à identificação de sequências que apresentaram variações no domínio crustina. Nessas análises, foi identificada uma sequência (*e.g.*, LOC113827916) marcada pela presença de um domínio amino-terminal enriquecido em resíduos de serina e leucina à montante de um domínio crustina contendo dois resíduos adicionais de cisteína. Apesar do arranjo peculiar contendo 14 resíduos de cisteína ao invés dos convencionais 12 resíduos, essa sequência mostrou uma organização aminoacídica consoante àquela típica das crustinas, incluindo a presença do domínio *cys-rich* comum aos Tipos I e II e de um domínio WAP conservado na região C-terminal (**Figura 28**). Finalmente,

uma sequência (*e.g.*, LOC113813930) foi recuperada apresentando em sua estrutura primária quatro domínios crustina repetidos sequencialmente. Crustinas similares foram igualmente recuperadas em outras duas espécies de camarões peneídeos (*i.e.*, *P. monodon* e *F. chinensis*) e apresentaram também múltiplos domínios crustina em sua estrutura primária. Todos os domínios apresentaram os característicos 12 resíduos conservados de cisteína devidamente espaçados de acordo com os intervalos presentes na assinatura crustina, mas contiveram nesses intervalos composições aminoacídicas únicas, ausentes em quaisquer das demais crustinas identificadas no genoma de *L. vannamei* ou reportadas na literatura (**Figura 28**).

Figura 28. Esquema ilustrativo mostrando as novas crustinas recuperadas nas buscas pelo genoma de *L. vannamei*.



As novas crustinas identificadas em *L. vannamei* foram analisadas quanto à sua estrutura primária e esquematicamente representadas. N- indica a região amino-terminal do peptídeo enquanto que C- indica a região carboxy-terminal. Peptídeos sinal que não foram preditos estão indicados com sinal de interrogação (?). Os resíduos de cisteína que compõem o domínio crustina estão representados pelas barras brancas. Os resíduos adicionais de cisteína presentes em LOC113827916 estão evidenciados em amarelo. Os números à direita representam os identificadores dos respectivos loci.

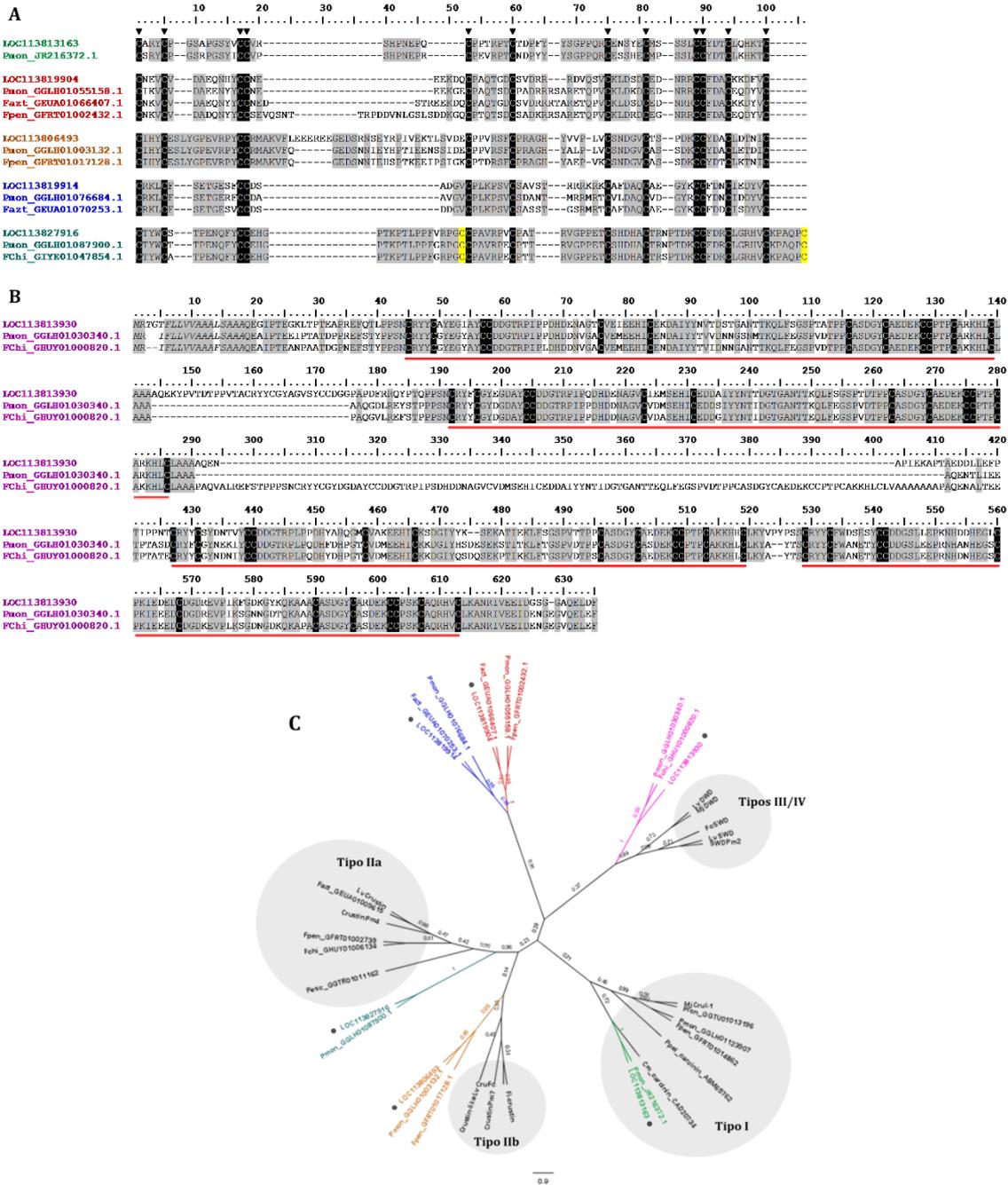
A fim de validar a existência desses novos membros em peneídeos, buscas por sequências homólogas foram realizadas em bancos transcritômicos (banco TSA) de *L. vannamei*, bem como buscas por ortólogos em transcritomas de outras espécies de camarão. Os resultados dessas buscas confirmaram a presença das novas crustinas em *L. vannamei* e revelaram a existência de prováveis ortólogos em pelo menos uma outra espécie de camarão peneídeo, sugerida a partir da identificação de sequências que compartilharam identidades aminoacídicas superiores a 70% nas suas regiões conservadas (**Tabela 4**).

Tabela 4. Validação *in silico* em bancos transcritômicos de *L. vannamei* das novas crustinas identificadas e identificação de prováveis ortólogos em camarões peneídeos.

| ID proteína | Espécie | Número de Acesso (TSA) | Identidade (%) |
|----------------|------------------------|------------------------|----------------|
| XP_027227949.1 | <i>L. vannamei</i> | GGKO01018945.1 | 100,0 |
| | <i>P. monodon</i> | GGLH01076684.1 | 75,0 |
| | <i>F. aztecus</i> | GEUA01070253.1 | 85,9 |
| XP_027220913.1 | <i>L. vannamei</i> | GFRP01036185.1 | 100 |
| | <i>P. monodon</i> | JR216372.1 | 81,2 |
| XP_027227939.1 | <i>L. vannamei</i> | GFRP01026759.1 | 93,8 |
| | <i>P. monodon</i> | GGLH01055158.1 | 75,4 |
| | <i>F. penicillatus</i> | GFRT01002432.1 | 70,8 |
| | <i>F. aztecus</i> | GEUA01066407.1 | 90,8 |
| XP_027213436.1 | <i>L. vannamei</i> | GDTK01001065.1 | 99,0 |
| | <i>P. monodon</i> | GGLH01003132.1 | 82,0 |
| | <i>F. penicillatus</i> | GFRT01017128.1 | 74,2 |
| XP_027236662.1 | <i>L. vannamei</i> | GFRP01043947.1 | 100 |
| | <i>P. monodon</i> | GGLH01087900.1 | 95,1 |
| XP_027221810.1 | <i>L. vannamei</i> | GHXV01004488.1 | 100 |
| | <i>P. monodon</i> | GGLH01030340.1 | 74 |
| | <i>F. chinensis</i> | GHUY01000820.1 | 75,5 |

Análises comparativas da estrutura primária das sequências recuperadas em *L. vannamei* e seus ortólogos mostraram um alto grau de conservação de resíduos aminoácidos e revelaram claramente a existência dos seis grupos discretos contendo assinaturas moleculares distintas (**Figura 29A/B**). Para confirmar as relações de homologia, as sequências foram filogeneticamente analisadas e comparadas com outras crustinas descritas em crustáceos a partir do alinhamento de suas regiões conservadas. As análises filogenéticas demonstraram, de fato, a formação de seis clados bem definidos que reuniram as sequências identificadas em *L. vannamei* e seus respectivos ortólogos. Curiosamente, três crustinas mostraram relações filogenéticas próximas de grupos previamente descritos na literatura, como as sequências LOC113806493 e LOC113827916, que mostraram proximidade filogenética com crustinas do Tipo II e a sequência LOC113813163 que se agrupou com autênticas crustinas do Tipo I. Os resultados revelados pelas reconstruções filogenéticas confirmaram, portanto, as relações de homologia sugeridas entre as sequências recuperadas, bem como a presença dessas novas crustinas em diferentes espécies da família Penaeidae (**Figura 29C**).

Figura 29. Comparação aminoacídica e reconstrução filogenética das novas crustinas identificadas em *L. vannamei*.



As novas crustinas identificadas em camarões peneídeos foram analisadas e comparadas quanto à sua estrutura primária e relação filogenética. **(A)** Composição aminoacídica da região referente ao domínio crustina de cinco dos seis grupos identificados. Os resíduos conservados de cisteína estão destacados em preto, enquanto que os demais resíduos comuns a todas as sequências de cada grupo estão destacados em cinza. Os resíduos de cisteína adicionais na sequência LOC113827916 estão evidenciados em amarelo. A régua acima indica a posição relativa de cada resíduo aminoacídico. **(B)** Composição aminoacídica da sequência LOC113813930 e seus ortólogos em camarões peneídeos. Os resíduos conservados de cisteína estão destacados em preto, enquanto que os demais resíduos comuns a todas as sequências de cada grupo estão destacados em cinza. Os múltiplos domínios crustina reconhecidos nas sequências estão indicados pelas linhas vermelhas. A régua acima indica a posição relativa de cada resíduo aminoacídico. **(C)** Árvore Neighbor-Joining mostrando a relação filogenética entre as novas crustinas identificadas em camarões peneídeos e crustinas autênticas previamente descritas em crustáceos. As esferas pretas representam as sequências recuperadas em *L. vannamei*. Cm: *Carcinus maenas*; Pmon: *Penaeus monodon*; Ppel: *Portunus pelagicus*; Fpen: *Fenneropenaeus penicillatus*; Plong: *Penaeus longystilus*; Mj:

Marsupenaeus japonicus; Fz: *Farfantepenaeus aztecus*; Fc/Fchi: *Fenneropenaeus chinensis*; Pesc: *Penaeus semisulcatus*; Fi: *Fenneropenaeus indicus*.

De acordo com o sistema de classificação vigente, as crustinas de crustáceos são divididas em quatro Tipos (Tipos I - IV), sendo que apenas dois (Tipos I e II) portam os 12 resíduos conservados de cisteína (TASSANAKAJON; SOMBOONWIWAT; AMPARYUP, 2015). Estruturalmente, esses dois grupos diferem unicamente pela presença/ausência de um domínio hidrofóbico amino-terminal rico em resíduos de glicina, decisivo para sua classificação (SMITH *et al.*, 2008; TASSANAKAJON; SOMBOONWIWAT; AMPARYUP, 2015). A descoberta de crustinas em camarões peneídeos que exibiram assinaturas moleculares únicas e arquiteturas alternativas representadas pela existência de domínios N- ou C-terminais ricos em resíduos aminoacídicos incomuns atesta a existência de novos Tipos dentro dessa família de peptídeos antimicrobianos e reforça a diversidade dessas moléculas, antes subestimada pela ausência de relatos na literatura. Além disso, os resultados até antes obtidos sugeriam, em conformidade com o sistema de classificação, que a diversidade das crustinas advinha da variabilidade de domínios N-terminais que eram, por sua vez, acoplados a um arcabouço estrutural constante fundamentado em um domínio crustina conservado (SMITH *et al.*, 2008). No entanto, a identificação de sequências aminoacídicas portadoras de domínios crustina ‘semiestruturados’ revelou que as pressões seletivas imperantes sobre essa família atuam igualmente sobre os resíduos componentes desse domínio tão conservado. De fato, a presença de resíduos de cisteína adicionais e a repetição sequencial de domínios crustina evidencia uma diversidade molecular ainda mais surpreendente e reforça a conjectura de que essas moléculas podem participar efetivamente em processos fisiológicos alternativos à função imune.

VI. DISCUSSÃO GERAL

Desde sua primeira descrição há pouco mais de 20 anos, as crustinas vêm sendo extensivamente estudadas nas mais diversas espécies de crustáceos decápodes. Relatos foram publicados, inclusive, em algumas espécies de insetos himenópteros, revelando a ampla distribuição taxonômica dessa família de AMPs em invertebrados (SMITH; DYRYNDA, 2015; TASSANAKAJON; SOMBOONWIWAT; AMPARYUP, 2015). Em virtude do apelo comercial para a carcinicultura, as crustinas foram mais bem estudadas em espécies que agregam grande valor econômico, como camarões, lagostas, caranguejos e lagostins. Entretanto, a identificação e caracterização de novos membros dessa família foram pontualmente realizadas ao longo dos anos e foram baseadas, normalmente, em métodos convencionais de clonagem e sequenciamento nucleotídico. Esses métodos carregam inerentemente entraves técnicos para a identificação massiva de novas sequências, uma vez que se baseiam no isolamento das sequências de interesse, impedindo o amplo acesso à diversidade molecular. Com o advento das ciências ômicas, tornou-se possível explorar de maneira consistente a diversidade molecular e genética de famílias gênicas, bem como contemplar, pela primeira vez, o repertório integral de seus membros constituintes.

No presente trabalho, a diversidade molecular e genética de crustinas em camarões peneídeos foi investigada a partir da prospecção de sequências aminoacídicas preditas do recém-publicado genoma de *L. vannamei*. O repertório molecular de crustinas nessa espécie mostrou ser surpreendentemente vasto e diverso, revelando um extenso processo de diversificação que imperou durante a evolução dessa família de AMPs e que, baseado na identificação de prováveis ortólogos em outras espécies de camarão, parece remontar a espécies ancestrais da família Penaeidae. De fato, foram identificadas em *L. vannamei* pelo menos 61 crustinas com arquiteturas moleculares, estruturas gênicas e propriedades bioquímicas diversas, dentre as quais 20 apresentaram um perfil estrutural consistente com o Tipo I, 35 com o Tipo II e 6 expuseram estruturas peculiares ainda não descritas na literatura, de acordo com o sistema de classificação definido por Smith e colaboradores (2008). Esse extenso repertório molecular de crustinas, contrastante com os modestos relatos descritos em outras espécies de camarão que normalmente portam não mais do que uma dezena de membros, parece representar um traço adaptativo que parte de sua provável importância para as respostas de defesa. De fato, proteínas do sistema imune, sobretudo aquelas que interagem diretamente com moléculas microbianas, costumam evoluir adaptativamente de modo a portar composições aminoacídicas e organizações estruturais otimizadas para o reconhecimento e combate de patógenos (TENNESSEN, 2005).

Um dos grandes achados desse trabalho foi que as crustinas do Tipo I em camarões peneídeos são abundantes, diversas e apresentam um amplo espectro de características bioquímicas e organizações gênicas. Pelo menos 10 grupos foram reconhecidos em *L. vannamei*, sendo que a identificação dessas crustinas, bem como de seus ortólogos, trouxe à tona um importante aspecto filogenético até o presente despercebido: apenas uma pequena fração das sequências que portam a típica arquitetura molecular do Tipo I (*i.e.*, peptídeo sinal amino-terminal seguido de um peptídeo maduro contendo um módulo *cys-rich* anterior a um domínio WAP C-terminal conservado) em camarões peneídeos é, de fato, próxima das carcininas descritas em crustáceos pleociemados, consideradas como autênticas crustinas do Tipo I. Na verdade, as reconstruções filogenéticas mostraram que a maioria das sequências recuperadas representou novas crustinas em camarões e apresentou múltiplas origens evolutivas, abrangendo, inclusive, membros pertencentes ao Tipo II. Assim como observado em outras famílias multigênicas de AMPs de peneídeos, a definição de grupos leva em consideração a homologia entre seus membros, de modo que são definidos em função de suas relações filogenéticas (GUEGUEN *et al.*, 2006; MATOS *et al.*, 2018). Nesse sentido, a existência de crustinas em *L. vannamei* que são estruturalmente similares, mas filogeneticamente distantes do Tipo I denunciou uma inconsistência no atual sistema de classificação e revelou que crustinas como a *carcininPm 1* e *carcininPm 2* do camarão *P. monodon* e as *MjCru I-3*, *MjCru I-4* e *MjCru I-5* de *M. japonicus* são exemplos de ortólogos em peneídeos que foram erroneamente classificados como membros desse grupo.

Como esperado para camarões peneídeos, as crustinas do Tipo II apresentaram a maior diversidade molecular. Pelo menos 12 grupos foram reconhecidos em *L. vannamei*, os quais apresentaram, também, um amplo espectro de propriedades físico-químicas e estruturas gênicas. De maneira interessante, a região amino-terminal mostrou ser a principal responsável pela diversidade molecular das crustinas do Tipo II, tendo em vista a amplitude de comprimentos, composições aminoacídicas e cargas, em detrimento do domínio crustina que apresentou características bioquímicas fortemente conservativas. Dado esse extenso grau de variabilidade, é plausível considerar que a região amino-terminal experimente de maneira mais imperativa as pressões seletivas que moldam a diversidade molecular nesse grupo, sugerindo sua imprescindível participação na atividade biológica das crustinas. No entanto, essa divergência de permissividade para diversificação encontrada entre ambos os domínios traz um aspecto curioso no que concerne à evolução molecular de seus genes codificantes, uma vez que as regiões gênicas que codificam ambos os domínios se encontram, na maioria dos grupos, precisamente no mesmo éxon, imediatamente adjacentes umas às outras. Assim,

os mecanismos evolutivos envolvidos na manutenção dessa diversidade parecem ser extremamente específicos à região codificante do domínio hidrofóbico amino-terminal.

Análises comparativas dos genes codificantes de crustinas em *L. vannamei* revelaram indícios da ocorrência de uma massiva expansão gênica em alguns dos grupos identificados, como o grupo 7 encontrado entre as crustinas similares ao Tipo I e os grupos 1 e 2 de crustinas do Tipo II, sugerindo que eventos de duplicação gênica têm atuado como a principal força motriz para diversificação. De fato, eventos de duplicação gênica em pequena escala (*i.e.*, independente da duplicação completa de genomas) representam um dos processos evolutivos predominantes para geração de diversidade, uma vez que as cópias gênicas evoluem como unidades independentes e tornam-se passíveis ao acúmulo de mutações que, em última análise, dão origem a novos grupos (WANG *et al.*, 2020). Além disso, os resultados obtidos mostraram que esses eventos parecem ser recentes e encontram-se ainda em andamento, tendo em vista a identificação de cópias gênicas que compartilharam altíssimos níveis de identidade, retêm organizações gênicas comuns quanto às suas regiões codificantes e que estão presentes em densos clusters gênicos. No entanto, essa notável expansão gênica marcada pela extensiva retenção de cópias gênicas requer explicações adicionais, uma vez que parece contradizer com a observação de que genes duplicados que apresentam funções biológicas redundantes evoluem normalmente sob pressões seletivas mais relaxadas, tornando-os mais propensos ao aparecimento de mutações deletérias que culminam na perda de função, pseudogenização e consequente deleção (LYNCH; CONERY, 2000; WAGNER, 1998). Nesse sentido, é possível inferir que a retenção de parálogos similares carrega um valor adaptativo e, baseado na sua participação em um sistema plástico que responde adaptativamente às injúrias do ambiente como o sistema imune, essa vantagem pode emergir da neofuncionalização ou subfuncionalização das novas cópias, que podem estender o leque de microrganismos suscetíveis à sua função antimicrobiana.

A ampla diversidade de crustinas encontradas em camarões peneídeos revelou também uma grande limitação dos atuais sistemas de classificação. Até o presente, três modelos foram propostos para classificar as crustinas, sendo que todos são baseados em caracteres estruturais, como a presença/ausência de domínios ou a abundância relativa de resíduos aminoacídicos (SMITH *et al.*, 2008; VARGAS-ALBORES; MARÍNEZ-PORCHAS, 2017; ZHAO; WANG, 2008). A prospecção de sequências no genoma de *L. vannamei*, juntamente com a recuperação de prováveis ortólogos em camarões peneídeos, revelou que os atuais sistemas de classificação não abarcam a verdadeira diversidade das crustinas e são, por isso, insuficientes para descrever toda a diversidade molecular presente nessa família de

AMPs. De fato, foram identificadas em camarões peneídeos seis crustinas que apresentaram arquiteturas moleculares inéditas, formadas pelo acoplamento ao domínio WAP de domínios amino- ou carboxy-terminais contendo composições aminoacídicas singulares. Em se baseando nos critérios utilizados pelos sistemas estruturais de classificação, a identificação dessas sequências mostra, de imediato, que esses sistemas são restritivos e resumem a diversidade das crustinas apenas à presença ou ausência de um domínio rico em resíduos de glicina, encontrada apenas em uma fração delas. Além disso, os resultados mostraram a existência de uma miríade de membros que apresentam arquiteturas moleculares condizentes com grupos previamente propostos, mas que possuem origens evolutivas filogeneticamente distantes. Essa incongruência foi particularmente revelada a partir da identificação de sequências que compartilharam uma estrutura típica de crustinas do Tipo I, mas que apresentaram múltiplas origens evolutivas. Nesse sentido, a arquitetura molecular que define o “Tipo I” mostrou representar apenas um arcabouço estrutural comum a moléculas de diferentes origens e, como consequência direta, revelou que o sistema de classificação predominante, definido por Smith e colaboradores (2008), para definição desse grupo responde simplesmente pela similaridade estrutural das sequências, mas não pelas suas relações filogenéticas.

Considerando que a quantidade de grupos identificados supera a quantidade de grupos descritos na literatura, o presente trabalho revela a fragilidade dos sistemas estruturais de classificação e reflete sobre a iminente necessidade da proposição de novos modelos que sejam, então, baseados em assinaturas moleculares específicas, tal como proposto no Capítulo 1 para as crustinas do Tipo II e em linha com o que foi realizado para outras famílias de AMPs de camarões peneídeos, como as penaeídeos e os fatores anti-lipopossacarídeos (ALFs) (GUEGUEN *et al.*, 2006; MATOS *et al.*, 2018). Para tal, faz-se necessária a exploração do repertório integral de crustinas em outras espécies de invertebrados para, assim, propor assinaturas consistentes que contemplem todos os grupos presentes nesses animais. Por essa razão, a proposição de um sistema de assinaturas moleculares utilizando-se apenas o genoma de uma única espécie seria incipiente de modo que, por ora, o presente trabalho se reservou a não fazê-la.

Além disso, tornou-se evidente neste trabalho que a unidade estrutural das crustinas é o conspicuo domínio crustina, formado pela presença de um módulo *cys-rich* que contém quatro resíduos de cisteína localizado à montante de um domínio WAP conservado portando oito cisteínas. De fato, ainda que mostrem uma multiplicidade de relações filogenéticas, todas as 61 crustinas recuperadas em *L. vannamei*, assim como aquelas descritas em outras espécies

de invertebrados, apresentaram esse típico arranjo 12-cys (à exceção das crustinas do Tipo II pertencentes aos grupos 11 e 12), sugerindo que o surgimento dessa família de AMPs pode ter sido possivelmente associado o acoplamento de um módulo *cys-rich* a uma linhagem específica de proteínas WAP presente em espécies ancestrais do subfilo Pancrustacea. Nessa perspectiva hipotética, a adição do módulo *cys-rich* a proteínas WAP teria provido uma importante função na fisiologia desses animais e se expandido massivamente ao longo do curso evolutivo para, mais tarde, dar origem a maior e mais diversa família de AMPs existente em crustáceos. Assim, por serem amplamente encontradas em outros grupos e não desempenharem necessariamente nesses grupos funções antimicrobianas, as crustinas do Tipo III (SWD) e IV (DWD), ausentes da unidade estrutural conservada que define essa família, são desconsideradas como crustinas verdadeiras.

Os resultados obtidos nesta tese de doutorado trouxeram um novo olhar para as crustinas e revelaram, pela primeira vez, seu repertório integral em uma espécie de camarão peneídeo através de uma busca criteriosa pelo genoma. A identificação e caracterização da diversidade molecular das crustinas abriram imensos horizontes para o estudo funcional dessas moléculas, permitindo investigar a participação dos diferentes grupos em distintos aspectos da fisiologia de camarões peneídeos e galgar, em etapas futuras, sua possível aplicação para biotecnologia marinha. Na verdade, essa ampla diversidade de moléculas pode encontrar inúmeras outras aplicações biotecnológicas, inclusive seu possível uso para tratamento de moléstias que afligem a saúde humana e animal. De fato, diante da diversidade molecular encontrada nesse trabalho e da descrição de crustinas portando atividades biológicas alternativas à função antimicrobiana (CHANG *et al.*, 2013; CHEN, Y. H. *et al.*, 2019; LIU *et al.*, 2015; LV *et al.*, 2020; STOSS *et al.*, 2004), é possível inferir que a extensiva diversificação molecular que parece ter acompanhado a história evolutiva das crustinas em peneídeos pode ter implicado igualmente na diversificação de suas funções biológicas através de mecanismos de neofuncionalização. Assim, seria plausível classificar as crustinas como “peptídeos de defesa” (HDPs, do inglês *host defense peptides*) e não apenas como peptídeos antimicrobianos. O presente trabalho representa, portanto, o primeiro passo na caracterização da diversidade de crustinas em camarões peneídeos e oferece subsídios para eventual utilização em programas de seleção assistida por marcadores genéticos e para a compreensão da história evolutiva das crustinas nos crustáceos decápodes.

VII. CONCLUSÕES E PERSPECTIVAS

CONCLUSÕES

- As crustinas do Tipo IIa e IIb de camarões peneídeos apresentam assinaturas moleculares específicas de cada grupo e provavelmente evoluíram a partir de um gene ancestral comum que conservou perfis similares de regulação transcricional.
- O camarão *Litopenaeus vannamei* possui um repertório extenso e bioquimicamente diverso de crustinas, codificadas por múltiplos genes que apresentam organizações estruturais diferentes e que são amplamente distribuídos no genoma.
- A diversidade genética de crustinas em *Litopenaeus vannamei* deriva, em grande medida, de mecanismos evolutivos predominantes como duplicação gênica seguida de mutação, eventos comuns na evolução de peptídeos antimicrobianos de camarões peneídeos.
- Os atuais sistemas de classificação das crustinas são restritivos e incapazes de abarcar toda a diversidade molecular existente nessa família, tornando necessária uma revisão sistemática desses modelos de classificação através da utilização de assinaturas moleculares específicas para definição de grupos.

PERSPECTIVAS

- Investigar a diversidade molecular de crustinas em outras espécies de crustáceos decápodes através da prospecção em genomas disponíveis.
- Estudar o espectro de atividades biológicas *in vitro* dos diferentes grupos de crustinas reconhecidos por meio da produção de proteínas recombinantes em sistemas heterólogos.
- Investigar a participação *in vivo* das crustinas na imunidade de *L. vannamei* através de silenciamento pós-transcricional via RNA de interferência.
- Analisar os mecanismos de regulação de expressão gênica de cada um dos grupos para obter pistas de sua participação em diferentes aspectos da fisiologia de *L. vannamei*.

- Estabelecer possíveis rotas evolutivas associadas ao surgimento e diversificação das crustinas em invertebrados.

REFERÊNCIAS

- AISENBREY, C.; MARQUETTE, A.; BECHINGER, B. The Mechanisms of Action of Cationic Antimicrobial Peptides Refined by Novel Concepts from Biophysical Investigations. In: MATSUZAKI, K. (Org.). . *Antimicrobial peptides: basics for clinical application*. [S.l.]: Springer, 2019. p. 33–64.
- AMPARYUP, P. *et al.* Molecular cloning, genomic organization and recombinant expression of a crustin-like antimicrobial peptide from black tiger shrimp *Penaeus monodon*. *Molecular Immunology*, v. 45, n. 4, p. 1085–1093, 2008.
- AMPARYUP, P.; DONPUDSA, S.; TASSANAKAJON, A. Shrimp single WAP domain (SWD)-containing protein exhibits proteinase inhibitory and antimicrobial activities. *Developmental and comparative immunology*, v. 32, n. 12, p. 1497–509, jan. 2008.
- BARRETO, C. *et al.* Specific molecular signatures for type II crustins in penaeid shrimp uncovered by the identification of crustin-like antimicrobial peptides in *Litopenaeus vannamei*. *Marine Drugs*, v. 16, n. 1, p. 1–16, 2018.
- BARTLETT, T. C. *et al.* Crustins, homologues of an 11.5-kDa antibacterial peptide, from two species of penaeid shrimp, *Litopenaeus vannamei* and *Litopenaeus setiferus*. *Marine biotechnology*, v. 4, n. 3, p. 278–93, jun. 2002.
- BECKING, T. *et al.* Shedding light on the antimicrobial peptide arsenal of terrestrial isopods: Focus on armadillidins, a new crustacean AMP family. *Genes*, v. 11, n. 1, 2020.
- BELLAMY, W. *et al.* Identification of the bactericidal domain of lactoferrin. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*, v. 1121, n. 1–2, p. 130–136, 1992.
- BROGDEN, K. A. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature reviews. Microbiology*, v. 3, n. 3, p. 238–50, mar. 2005.
- CHANG, Y. T. *et al.* The New Face of the Old Molecules: crustin $Pm4$ and Transglutaminase Type I Serving as RNPs Down-Regulate Astakine-Mediated Hematopoiesis. *PLoS ONE*, v. 8, n. 8, 2013.
- CHEN, D.; HE, N.; XU, X. Mj-DWD, a double WAP domain-containing protein with antiviral relevance in *Marsupenaeus japonicus*. *Fish & shellfish immunology*, v. 25, n. 6, p. 775–81, dez. 2008.
- CHEN, Y. H. *et al.* Functional characterization of an ER-stress responding Crustin gene in *Litopenaeus vannamei*. *Fish and Shellfish Immunology*, v. 84, n. July 2018, p. 541–550, 2019.
- CORONEL, C. E.; SAN AGUSTIN, J.; LARDY, H. A. Purification and structure of caltrin-like proteins from seminal vesicle of the guinea pig. *Journal of Biological Chemistry*, v. 265, n. 12, p. 6854–6859, 1990.
- COSTA, F. *et al.* Clinical Application of AMPs. In: MATSUZAKI, K. (Org.). . *Antimicrobial peptides: basics for clinical application*. [S.l.]: Springer, 2019. p. 281–298.
- COUTO, M. A. *et al.* eNAP-2, a novel cysteine-rich bactericidal peptide from equine leukocytes. *Infection and Immunity*, v. 60, n. 12, p. 5042–5047, 1992.
- CUSACK, B. P.; WOLFE, K. H. Not born equal: Increased rate asymmetry in relocated and retrotransposed rodent gene duplicates. *Molecular Biology and Evolution*, v. 24, n. 3, p. 679–686, 2007.
- DEAR, T. N.; KEFFORD, R. F. The WDNM1 gene product is a novel member of the “four-disulphide core” family of proteins. *Biochemical and Biophysical Research Communications*, v. 176, n. 1, p. 247–254, 1991.
- DESTOUMIEUX-GARZÓN, D. *et al.* Antimicrobial peptides in marine invertebrate health and disease. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 371, n. 1695, 2016.
- DESTOUMIEUX-GARZÓN, D. *et al.* Crustacean immunity: Antifungal peptides are generated from the C-terminus of shrimp hemocyanin in response to microbial challenge. *Journal of Biological Chemistry*, v. 276, n. 50, p. 47070–47077, 2001.
- DESTOUMIEUX, D. *et al.* Penaeidins, a New Family of Antimicrobial Peptides Isolated from the Shrimp

- Penaeus vannamei* (Decapoda). *Journal of Biological Chemistry*, v. 272, n. 45, p. 28398–28406, 7 nov. 1997.
- DONPUDSA, S. *et al.* Type I and type II crustins from *Penaeus monodon*, genetic variation and antimicrobial activity of the most abundant crustinPm4. *Developmental and comparative immunology*, v. 47, n. 1, p. 95–103, nov. 2014.
- DU, Z.-Q. *et al.* A double WAP domain (DWD)-containing protein with proteinase inhibitory activity in Chinese white shrimp, *Fenneropenaeus chinensis*. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology*, v. 154, n. 2, p. 203–10, out. 2009.
- DU, Z.-Q. *et al.* A single WAP domain (SWD)-containing protein with antipathogenic relevance in red swamp crayfish, *Procambarus clarkii*. *Fish & shellfish immunology*, v. 28, n. 1, p. 134–42, jan. 2010.
- DU, Z.-Q.; YUAN, J.-J.; REN, D.-M. A novel single WAP domain-containing protein isoform with antibacterial relevance in *Litopenaeus vannamei*. *Fish & shellfish immunology*, v. 44, n. 2, p. 478–84, 2015.
- DU, Z. QIANG *et al.* A new antimicrobial peptide isoform, *Pc*-crustin 4 involved in antibacterial innate immune response in fresh water crayfish, *Procambarus clarkii*. *Fish and Shellfish Immunology*, v. 94, n. June, p. 861–870, 2019.
- DURICA, D. S. *et al.* EST library sequencing of genes expressed during early limb regeneration in the fiddler crab and transcriptional responses to ecdysteroid exposure in limb bud explants. *Integrative and Comparative Biology*, v. 46, n. 6, p. 948–964, 2006.
- EPAGRI, E. DE P. A. E E. R. DE S. C. Síntese Informativa da Maricultura 2017. *Centro de Desenvolvimento em Aquicultura e Pesca*, p. 1–8, 2018.
- FAO *et al.* Introductions and movement of *Penaeus vannamei* and *Penaeus stylirostris* in Asia and the Pacific. *RAP publication 2004/10*, v. 75, n. January, p. 1–12, 2004.
- FARIAS, N. D. *et al.* *Litopenaeus vannamei* stylicins are constitutively produced by hemocytes and intestinal cells and are differentially modulated upon infections. *Fish and Shellfish Immunology*, v. 86, n. October 2018, p. 82–92, 2019.
- FREELING, M. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology*, v. 60, p. 433–453, 2009.
- GERDOL, M.; GRECO, S.; PALLAVICINI, A. Extensive tandem duplication events drive the expansion of the C1q-domain-containing gene family in bivalves. *Marine Drugs*, v. 17, n. 10, p. 1–13, 2019.
- GONCALVES, P. *et al.* Molecular signatures at imminent death: hemocyte gene expression profiling of shrimp succumbing to viral and fungal infections. *Developmental and comparative immunology*, v. 42, n. 2, p. 294–301, fev. 2014.
- GUANÍ-GUERRA, E. *et al.* Antimicrobial peptides: general overview and clinical implications in human health and disease. *Clinical immunology (Orlando, Fla.)*, v. 135, n. 1, p. 1–11, abr. 2010.
- GUEGUEN, Y. *et al.* PenBase, the shrimp antimicrobial peptide penaeidin database: sequence-based classification and recommended nomenclature. *Developmental and comparative immunology*, v. 30, n. 3, p. 283–8, jan. 2006.
- HAGIWARA, K. *et al.* Mouse SWAM1 and SWAM2 Are Antibacterial Proteins Composed of a Single Whey Acidic Protein Motif. *The Journal of Immunology*, v. 170, n. 4, p. 1973–1979, 2003.
- HENNIGHAUSEN, L. G.; SIPPEL, A. E. Mouse whey acidic protein is a novel member of the family of “four-disulfide core” proteins. *Nucleic Acids Research*, v. 10, n. 8, p. 2677–2684, 1982.
- HIEMSTRA, P. S. *et al.* Antibacterial activity of antileukoprotease. *Infection and Immunity*, v. 64, n. 11, p. 4520–4524, 1996.

- HIPOLITO, S. G. *et al.* Role of *Marsupenaeus japonicus* crustin-like peptide against *Vibrio penaeicida* and white spot syndrome virus infection. *Developmental and comparative immunology*, v. 46, n. 2, p. 461–9, out. 2014.
- ISHIZUKA, N. *et al.* Na⁺,K⁺-ATPase inhibition by an endogenous peptide, SPAI-1, isolated from porcine duodenum. *BBA - Biomembranes*, v. 1069, n. 2, p. 259–266, 1991.
- JIA, Y.-P. *et al.* A single whey acidic protein domain (SWD)-containing peptide from fleshy prawn with antimicrobial and proteinase inhibitory activities. *Aquaculture*, v. 284, n. 1–4, p. 246–259, nov. 2008.
- JIANG, H.-S. *et al.* Four crustins involved in antibacterial responses in *Marsupenaeus japonicus*. *Fish & shellfish immunology*, v. 43, n. 2, p. 387–95, abr. 2015.
- JIANG, H. S. *et al.* A single whey acidic protein domain containing protein (SWD) inhibits bacteria invasion and dissemination in shrimp *Marsupenaeus japonicus*. *Fish and Shellfish Immunology*, v. 35, n. 2, p. 310–318, 2013.
- JIMÉNEZ-VEGA, F. *et al.* A single WAP domain-containing protein from *Litopenaeus vannamei* hemocytes. *Biochemical and Biophysical Research Communications*, v. 314, n. 3, p. 681–687, fev. 2004.
- JIMÉNEZ-VEGA, F.; VARGAS-ALBORES, F. A secretory leukocyte proteinase inhibitor (SLPI)-like protein from *Litopenaeus vannamei* haemocytes. *Fish & shellfish immunology*, v. 23, n. 5, p. 1119–26, nov. 2007.
- JIRAVANICHPAISAL, P. *et al.* Antibacterial peptides in hemocytes and hematopoietic tissue from freshwater crayfish *Pacifastacus leniusculus*: characterization and expression pattern. *Developmental and comparative immunology*, v. 31, n. 5, p. 441–55, jan. 2007.
- KRAMEROVA, I. A. *et al.* Papilin in development; a pericellular protein with a homology to the ADAMTS metalloproteinases. *Development*, v. 127, n. 24, p. 5475–5485, 2000.
- LAI, A. G.; ABOOBAKER, A. A. Comparative genomic analysis of innate immunity reveals novel and conserved components in crustacean food crop species. *BMC Genomics*, v. 18, n. 1, p. 1–26, 2017.
- LAZZARO, B. P.; ZASLOFF, M.; ROLFF, J. Antimicrobial peptides: Application informed by evolution. *Science*, v. 368, n. 6490, 2020.
- LE BLOA, S. *et al.* Antimicrobial Peptides and Ectosymbiotic Relationships: Involvement of a Novel Type IIa Crustin in the Life Cycle of a Deep-Sea Vent Shrimp. *Frontiers in Immunology*, v. 11, n. July, p. 1–18, 2020.
- LI, M. *et al.* A new crustin is involved in the innate immune response of shrimp *Litopenaeus vannamei*. *Fish and Shellfish Immunology*, v. 94, n. June, p. 398–406, 2019.
- LI, M. *et al.* Molecular cloning, expression, promoter analysis and functional characterization of a new Crustin from *Litopenaeus vannamei*. *Fish and Shellfish Immunology*, v. 73, n. September 2017, p. 42–49, 2018.
- LI, SHIHAO *et al.* Molecular and Functional Diversity of Crustin-Like Genes in the Shrimp *Litopenaeus vannamei*. *Marine Drugs*, v. 18, n. 7, p. 1–17, 2020.
- LI, SHUANG *et al.* A double WAP domain-containing protein *Es-DWD1* from *Eriocheir sinensis* exhibits antimicrobial and proteinase inhibitory activities. *PloS one*, v. 8, n. 8, p. e73563, jan. 2013.
- LIGHTNER, D. V. *et al.* Historic emergence, impact and current status of shrimp pathogens in the Americas. *Journal of Invertebrate Pathology*, v. 110, n. 2, p. 174–183, 2012.
- LIU, N. *et al.* A novel crustin from *Marsupenaeus japonicus* promotes hemocyte phagocytosis. *Developmental and comparative immunology*, v. 49, n. 2, p. 313–22, abr. 2015.
- LIVAK, K. J.; SCHMITTGEN, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, v. 25, n. 4, p. 402–8, 2001.
- LV, X. *et al.* The immune function of a novel crustin with an atypical WAP domain in regulating intestinal

microbiota homeostasis in *Litopenaeus vannamei*. *Developmental and Comparative Immunology*, 2020.

LYNCH, M.; CONERY, J. The Evolutionary Fate and Consequences of Duplicate Genes. *Science*, v. 290, p. 1151–1155, 2000.

MARIE, B. *et al.* Proteomic analysis of the organic matrix of the abalone *Haliotis asinina* calcified shell. *Proteome Science*, v. 8, n. 1, p. 54, 2010.

MATOS, G. M. *et al.* Massive gene expansion and sequence diversification is associated with diverse tissue distribution, regulation and antimicrobial properties of anti-lipopolysaccharide factors in shrimp. *Marine Drugs*, v. 16, n. 10, p. 1–17, 2018.

NAIR, D. G. *et al.* Antimicrobial activity of omwaprin, a new member of the waprin family of snake venom proteins. *Biochemical Journal*, v. 402, n. 1, p. 93–104, 2007.

NUNES, A.; FEIJÓ, R. G. Convivência com o vírus da mancha branca no cultivo de camarão marinho no Brasil. *Revista da Associação Brasileira de Criadores de Camarão*, v. 2, p. 30–36, 2016.

O'LEARY, N. A.; GROSS, P. S. Genomic structure and transcriptional regulation of the penaeidin gene family from *Litopenaeus vannamei*. *Gene*, v. 371, n. 1, p. 75–83, 2006.

OTA, Y. *et al.* The expression of secretory leukocyte protease inhibitor (SLPI) in the Fallopian tube: SLPI protects the acrosome reaction of sperm from inhibitory effects of elastase. *Human Reproduction*, v. 17, n. 10, p. 2517–2522, 2002.

PILETZ, J. E.; HEINLEN, M.; GANSCHOW, R. E. Biochemical characterization of a novel whey protein from murine milk. *Journal of Biological Chemistry*, v. 256, n. 22, p. 11509–11516, 1981.

QUISPE, R. L. *et al.* Transcriptional profiling of immune-related genes in Pacific white shrimp (*Litopenaeus vannamei*) during ontogenesis. *Fish and Shellfish Immunology*, v. 58, n. September, p. 103–107, 2016.

RANGANATHAN, S.; SIMPSON, K. The whey acidic protein family: a new signature motif and three-dimensional structure by comparative modeling. *Journal of Molecular Graphics and Modeling*, v. 3263, n. 99, p. 106–113, 1999.

RATTANACHAI, A. *et al.* Cloning of kuruma prawn *Marsupenaeus japonicus* crustin-like peptide cDNA and analysis of its expression. *Fisheries Science*, v. 70, n. 5, p. 765–771, out. 2004.

RELF, J. *et al.* Purification and characterization of a cysteine-rich 11.5-kDa antibacterial protein from the granular haemocytes of the shore crab, *Carcinus maenas*. *European Journal of Biochemistry*, v. 357, n. June, p. 350–357, 1999.

ROBALINO, J. *et al.* Contributions of functional genomics and proteomics to the study of immune responses in the Pacific white leg shrimp *Litopenaeus vannamei*. *Veterinary Immunology and Immunopathology*, v. 128, n. 1–3, p. 110–118, 2009.

ROLLAND, J. L. *et al.* Stylicins, a new family of antimicrobial peptides from the Pacific blue shrimp *Litopenaeus stylirostris*. *Molecular immunology*, v. 47, n. 6, p. 1269–77, mar. 2010.

ROSA, R.; BARRACCO, M. Antimicrobial peptides in crustaceans. *Inv Surv J*, p. 262–284, 2010.

ROSA, R. D.; BANDEIRA, P. T.; BARRACCO, M. A. Molecular cloning of crustins from the hemocytes of Brazilian penaeid shrimps. *FEMS microbiology letters*, v. 274, n. 2, p. 287–90, set. 2007.

RUNGRASSAMEE, W. *et al.* Bacterial Population in Intestines of the Black Tiger Shrimp (*Penaeus monodon*) under Different Growth Stages. *PLoS ONE*, v. 8, n. 4, 2013.

SCHALKWIJK, J.; WIEDOW, O.; HIROSE, S. The trappin gene family: proteins defined by an N-terminal transglutaminase substrate domain and a C-terminal four-disulphide core. *The Biochemical journal*, v. 340 (Pt 3, p. 569–77, 1999.

SCHMITT, P. *et al.* The antimicrobial defense of the Pacific oyster, *Crassostrea gigas*. How diversity may

compensate for scarcity in the regulation of resident/pathogenic microflora. *Frontiers in Microbiology*, v. 3, n. MAY, p. 1–17, 2012.

SCHMITT, P.; ROSA, R. D.; DESTOUMIEUX-GARZÓN, D. An intimate link between antimicrobial peptide sequence diversity and binding to essential components of bacterial membranes. *Biochimica et Biophysica Acta - Biomembranes*, v. 1858, n. 5, p. 958–970, 2016.

SHOCKEY, J. E. *et al.* The role of crustins in *Litopenaeus vannamei* in response to infection with shrimp pathogens: An in vivo approach. *Developmental and comparative immunology*, v. 33, n. 5, p. 668–73, maio 2009.

SMITH, V. J. *et al.* Crustins: enigmatic WAP domain-containing antibacterial proteins from crustaceans. *Developmental and comparative immunology*, v. 32, n. 7, p. 758–72, jan. 2008.

SMITH, V. J. Phylogeny of whey acidic protein (WAP) four-disulfide core proteins and their role in lower vertebrates and invertebrates. *Biochemical Society transactions*, v. 39, n. 5, p. 1403–8, 2011.

SMITH, V. J.; CHISHOLM, J. Antimicrobial Proteins in Crustaceans. In: BECK, G.; SUGUMARAN, M.; COOPER, E. L. (Org.). *Phylogenetic Perspectives on the Vertebrate Immune System*. [S.l.]: y Kluwer Academic/Plenum Pres, 2001. p. 95–112.

SMITH, V. J.; DYRYNDA, E. A. Antimicrobial proteins: From old proteins, new tricks. *Molecular Immunology*, v. 68, n. 2, p. 383–398, 2015.

SOREK, R. The birth of new exons: Mechanisms and evolutionary consequences. *Rna*, v. 13, n. 10, p. 1603–1608, 2007.

SOUSI-YANICOSTAS, N. *et al.* Anosmin-1 underlying the X chromosome-linked Kallmann syndrome is an adhesion molecule that can modulate neurite growth in a cell-type specific manner. *Journal of Cell Science*, v. 111, n. 19, p. 2953–2965, 1998.

SPERSTAD, S. V. *et al.* Characterization of crustins from the hemocytes of the spider crab, *Hyas araneus*, and the red king crab, *Paralithodes camtschaticus*. *Developmental and comparative immunology*, v. 33, n. 4, p. 583–91, abr. 2009.

STOSS, T. D. *et al.* Inducible Transcript Expressed by Reactive Epithelial Cells at Sites of Olfactory Sensory Neuron Proliferation. *Journal of Neurobiology*, v. 58, n. 3, p. 355–368, 2004.

SULEIMAN, S.; SMITH, V. J.; DYRYNDA, E. A. Unusual tissue distribution of carcinin, an antibacterial crustin, in the crab, *Carcinus maenas*, reveals its multi-functionality. *Developmental and Comparative Immunology*, v. 76, p. 274–284, 2017.

SUN, C. *et al.* Molecular cloning and characterization of three crustins from the Chinese white shrimp, *Fenneropenaeus chinensis*. *Fish and Shellfish Immunology*, v. 28, n. 4, p. 517–524, 2010.

SUPUNGUL, P. *et al.* Antimicrobial peptides discovered in the black tiger shrimp *Penaeus monodon* using the EST approach. *Diseases of aquatic organisms*, v. 61, n. 1–2, p. 123–35, 21 out. 2004.

SUTHIANTHONG, P. *et al.* A double WAP domain-containing protein *PmDWD* from the black tiger shrimp *Penaeus monodon* is involved in the controlling of proteinase activities in lymphoid organ. *Fish and Shellfish Immunology*, v. 30, n. 3, p. 783–790, 2011.

TASSANAKAJON, A. *et al.* Cationic antimicrobial peptides in penaeid shrimp. *Marine biotechnology (New York, N.Y.)*, v. 13, n. 4, p. 639–57, ago. 2011.

TASSANAKAJON, A.; SOMBOONWIWAT, K.; AMPARYUP, P. Sequence diversity and evolution of antimicrobial peptides in invertebrates. *Developmental and comparative immunology*, v. 48, n. 2, p. 324–41, fev. 2015.

TENNESSEN, J. A. Molecular evolution of animal antimicrobial peptides: Widespread moderate positive selection. *Journal of Evolutionary Biology*, v. 18, n. 6, p. 1387–1394, 2005.

TRAN, L. *et al.* Determination of the infectious nature of the agent of acute hepatopancreatic necrosis syndrome

affecting penaeid shrimp. *Diseases of Aquatic Organisms*, v. 105, n. 1, p. 45–55, 2013.

TRECCANI, L. *et al.* Perlwapin, an abalone nacre protein with three four-disulfide core (whey acidic protein) domains, inhibits the growth of calcium carbonate crystals. *Biophysical journal*, v. 91, n. October, p. 2601–2608, 2006.

VARGAS-ALBORES, F.; YEPIZ-PLASCENCIA, G.; JIMÉNEZ-VEGA, F.; AVILA-VILLA, A. Structural and functional differences of *Litopenaeus vannamei* crustins. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology*, v. 138, n. 4, p. 415–22, ago. 2004.

VARGAS-ALBORES, F.; YEPIZ-PLASCENCIA, G.; JIMÉNEZ-VEGA, F.; AVILA-VILLA, A. Structural and functional differences of *Litopenaeus vannamei* crustins. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, v. 138, n. 4, p. 415–422, 2004.

VARGAS-ALBORES, F.; MARÍNEZ-PORCHAS, M. Crustins are distinctive members of the WAP-containing protein superfamily: An improved classification approach. *Developmental and Comparative Immunology*, v. 76, p. 9–17, 2017.

VATANAVICHARN, T. *et al.* Genomic structure, expression pattern and functional characterization of crustinPm5, a unique isoform of crustin from *Penaeus monodon*. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology*, v. 153, n. 3, p. 244–52, jul. 2009.

WAGNER, A. The fate of duplicated genes: Loss or new function? *BioEssays*, v. 20, n. 10, p. 785–788, 1998.

WANG, Q. *et al.* Diversity of antimicrobial peptides in three partially sympatric frog species in northeast asia and implications for evolution. *Genes*, v. 11, n. 2, 2020.

WANG, Y. *et al.* Mechanism of alternative splicing and its regulation. *Biomedical Reports*, v. 3, n. 2, p. 152–158, 2015.

ZASLOFF, M. Antimicrobial Peptides of Multicellular Organisms: My Perspective. In: MATSUZAKI, K. (Org.). *Antimicrobial peptides: basics for clinical application*. [S.l.]: Springer, 2019. p. 3–6.

ZASLOFF, M. Antimicrobial peptides of multicellular organisms. *Nature*, v. 415, p. 389–395, 2002.

ZHANG, JIANZHI. Evolution by gene duplication: An update. *Trends in Ecology and Evolution*, v. 18, n. 6, p. 292–298, 2003.

ZHANG, JIQUAN *et al.* Cloning and recombinant expression of a crustin-like gene from Chinese shrimp, *Fenneropenaeus chinensis*. *Journal of Biotechnology*, v. 127, n. 4, p. 605–614, 2007.

ZHANG, X. *et al.* Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nature Communications*, v. 10, n. 1, 2019.

ZHANG, Z.; ZHU, S. Comparative genomics analysis of five families of antimicrobial peptide-like genes in seven ant species. *Developmental and comparative immunology*, v. 38, n. 2, p. 262–74, out. 2012.

ZHAO, X.; WANG, J. The antimicrobial peptides of the immune response of shrimp. *Invertebrate Survival Journal*, p. 162–179, 2008.

ANEXOS

ANEXO 1. Folha de rosto do artigo 1, referente ao capítulo 1 da tese.

Article

Specific Molecular Signatures for Type II Crustins in Penaeid Shrimp Uncovered by the Identification of Crustin-Like Antimicrobial Peptides in *Litopenaeus vannamei*

Cairé Barreto ¹, Jaqueline da Rosa Coelho ¹, Jianbo Yuan ², Jianhai Xiang ²,
Luciane Maria Perazzolo ¹ and Rafael Diego Rosa ^{1,*} 

¹ Laboratory of Immunology Applied to Aquaculture, Department of Cell Biology, Embryology and Genetics, Federal University of Santa Catarina, Florianópolis 88040-900 SC, Brazil; cairebarreto@gmail.com (C.B.); jaquesombrio@gmail.com (J.d.R.C.); lm.perazzolo@ufsc.br (L.M.P.)

² Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China; yuanjb@qdio.ac.cn (J.Y.); jxiang@qdio.ac.cn (J.X.)

* Correspondence: rafael.d.rosa@ufsc.br; Tel: +55-48-37216163

Received: 30 August 2017; Accepted: 16 October 2017; Published: 16 January 2018

Abstract: Crustins form a large family of antimicrobial peptides (AMPs) in crustaceans composed of four sub-groups (Types I–IV). Type II crustins (Type IIa or “Crustins” and Type IIb or “Crustin-like”) possess a typical hydrophobic N-terminal region and are by far the most representative sub-group found in penaeid shrimp. To gain insight into the molecular diversity of Type II crustins in penaeids, we identified and characterized a Type IIb crustin in *Litopenaeus vannamei* (Crustin-like *Lv*) and compared Type II crustins at both molecular and transcriptional levels. Although *L. vannamei* Type II crustins (Crustin *Lv* and Crustin-like *Lv*) are encoded by separate genes, they showed a similar tissue distribution (hemocytes and gills) and transcriptional response to the shrimp pathogens *Vibrio harveyi* and White spot syndrome virus (WSSV). As Crustin *Lv*, Crustin-like *Lv* transcripts were found to be present early in development, suggesting a maternal contribution to shrimp progeny. Altogether, our *in silico* and transcriptional data allowed to conclude that (1) each sub-type displays a specific amino acid signature at the C-terminal end holding both the cysteine-rich region and the whey acidic protein (WAP) domain, and that (2) shrimp Type II crustins evolved from a common ancestral gene that conserved a similar pattern of transcriptional regulation.

Keywords: invertebrate immunity; host defense peptide; crustacean; WAP domain; molecular diversity; host-pathogen interaction

ANEXO 2. Artigo publicado em co-autoria durante período de pós-graduação no LIAA.

Fish & Shellfish Immunology 56 (2016) 123–126



ELSEVIER

Contents lists available at ScienceDirect

Fish & Shellfish Immunology

journal homepage: www.elsevier.com/locate/fai



Short communication

A hemocyte-expressed fibrinogen-related protein gene (*LvFrep*) from the shrimp *Litopenaeus vannamei*: Expression analysis after microbial infection and during larval development

 CrossMark

Jaqueline da Rosa Coelho, Cairé Barreto, Amanda da Silva Silveira, Crazuela Cleusa Vieira, Rafael Diego Rosa, Luciane Maria Perazzolo*

Laboratory of Immunology Applied to Aquaculture, Department of Cell Biology, Embryology and Genetics, Federal University of Santa Catarina, 88046-900 Florianópolis, SC, Brazil

| | |
|--|---|
| <p>ARTICLE INFO</p> <p><i>Article history:</i> Received 24 May 2016 Received in revised form 23 June 2016 Accepted 27 June 2016 Available online 2 July 2016</p> <p><i>Keywords:</i> Pattern recognition protein/receptor Fibrinogen-related protein Crustacean immunity Gene expression <i>Vibrio harveyi</i></p> | <p>ABSTRACT</p> <p>Fibrinogen-related proteins (FREPs) comprise a large family of microbial recognition proteins involved in many biological functions in both vertebrate and invertebrate animals. By taking advantage of publicly accessible databases, we have identified a FREP-like homolog in the most cultivated penaeid shrimp, <i>Litopenaeus vannamei</i> (<i>LvFrep</i>). The obtained sequence showed a conserved fibrinogen-related domain (FRE-D) and displayed significant similarities to FREP-like proteins from other invertebrates and to ficolins from crustaceans. The expression of <i>LvFrep</i> appeared to be limited to circulating hemocytes. Interestingly, <i>LvFrep</i> gene expression was induced in shrimp hemocytes only in response to a <i>Vibrio</i> infection but not to the White spot syndrome virus (WSSV). Moreover, <i>LvFrep</i> transcript levels were detected early in fertilized eggs, suggesting the participation of this immune-related gene in the antimicrobial defenses during shrimp development.</p> <p style="text-align: right;">© 2016 Elsevier Ltd. All rights reserved.</p> |
|--|---|

ANEXO 3. Artigo publicado em co-autoria durante período de pós-graduação no LIAA.**PHILOSOPHICAL
TRANSACTIONS B**

rstb.royalsocietypublishing.org

**Review**

Cite this article: Destoumieux-Garzón D, Rosa RD, Schmitt P, Barreto C, Vidal-Dupiol J, Mitta G, Gueguen Y, Bachère E. 2016 Antimicrobial peptides in marine invertebrate health and disease. *Phil. Trans. R. Soc. B* **371**: 20150300. <http://dx.doi.org/10.1098/rstb.2015.0300>

Accepted: 10 February 2016

One contribution of 13 to a theme issue 'Evolutionary ecology of arthropod antimicrobial peptides'.

Subject Areas:

biochemistry, environmental science, immunology, microbiology

Keywords:

invertebrate immunity, polymicrobial disease, vibrio, immune suppression, resistance, abiotic stress

Antimicrobial peptides in marine invertebrate health and disease

Delphine Destoumieux-Garzón^{1,2,3,4}, Rafael Diego Rosa⁵, Paulina Schmitt⁶, Cairé Barreto⁵, Jeremie Vidal-Dupiol⁷, Guillaume Mitta^{1,2,3,4}, Yannick Gueguen^{1,2,3,4} and Evelyne Bachère^{1,2,3,4}

¹ORIS, ²themer, ³UPVD, and ⁴Université de Montpellier, Interactions Hôte-Pathogènes-Environnements (IHPE, UMRS244), Place Eugène Bataillon, 34090 Montpellier cedex, France

⁵Laboratory of Immunology Applied to Aquaculture, Department of Cell Biology, Embryology and Genetics, Federal University of Santa Catarina, 88040-900 Florianópolis, Santa Catarina, Brazil

⁶Laboratorio de Genética e Imunologia Molecular, Instituto de Biología, Pontificia Universidad Católica de Valparaíso, Avenida Universidad 330, 2373223 Valparaíso, Chile

⁷themer, UMR 241 BI, LabexCoval, BP 7004, 98719 Taravao, Tahiti, French Polynesia

Aquaculture contributes more than one-third of the animal protein from marine sources worldwide. A significant proportion of aquaculture products are derived from marine protostomes that are commonly referred to as 'marine invertebrates'. Among them, penaeid shrimp (Ecdysozoa, Arthropoda) and bivalve molluscs (Lophotrochozoa, Mollusca) are economically important. Mass rearing of arthropods and molluscs causes problems with pathogens in aquatic ecosystems that are exploited by humans. Remarkably, species of corals (Cnidaria) living in non-exploited ecosystems also suffer from devastating infectious diseases that display intriguing similarities with those affecting farmed animals. Infectious diseases affecting wild and farmed animals that are present in marine environments are predicted to increase in the future. This paper summarizes the role of the main pathogens and their interaction with host immunity, with a specific focus on antimicrobial peptides (AMPs) and pathogen resistance against AMPs. We provide a detailed review of penaeid shrimp AMPs and their role at the interface between the host and its resident/pathogenic microbiota. We also briefly describe the relevance of marine invertebrate AMPs in an applied context.

This article is part of the themed issue 'Evolutionary ecology of arthropod antimicrobial peptides'.

ANEXO 4. Artigo publicado em co-autoria durante período de pós-graduação no LIAA.



Article

Massive Gene Expansion and Sequence Diversification Is Associated with Diverse Tissue Distribution, Regulation and Antimicrobial Properties of Anti-Lipopolysaccharide Factors in Shrimp

Gabriel Machado Matos ¹, Paulina Schmitt ², Caíne Barreto ¹, Natanael Dantas Farias ¹, Guilherme Toledo-Silva ³, Fanny Guzmán ⁴, Delphine Destoumieux-Garzón ⁵, Luciane Maria Perazzolo ¹ and Rafael Diego Rosa ^{1,*}

¹ Laboratory of Immunology Applied to Aquaculture, Department of Cell Biology, Embryology and Genetics, Federal University of Santa Catarina, Florianópolis SC 88040-900, Brazil; gabrielmatos92@gmail.com (G.M.M.); cainebarreto@gmail.com (C.B.); natan.chio@gmail.com (N.D.F.); lm.perazzolo@ufsc.br (L.M.P.)

² Laboratorio de Genética e Inmunología Molecular, Instituto de Biología, Facultad de Ciencias, Pontificia Universidad Católica de Valparaíso, Valparaíso 2373223, Chile; paulina.schmitt@puvc.cl

³ Department of Cell Biology, Embryology and Genetics, Federal University of Santa Catarina, Florianópolis SC 88040-900, Brazil; guilherme.toledo@ufsc.br

⁴ Núcleo Biotecnología Curauma, Pontificia Universidad Católica de Valparaíso, Valparaíso 2373223, Chile; fanny.guzman@puvc.cl

⁵ Interactions Hôtes-Pathogènes-Environnements, Université de Montpellier, CNRS, IREM, Université de Perpignan Via Domitia, CEDEX 5, 34090 Montpellier, France; Delphine.Destoumieux-Garzon@fremer.fr

* Correspondence: rafael.d.rosa@ufsc.br; Tel.: +55-48-37216163

Received: 15 September 2018; Accepted: 9 October 2018; Published: 11 October 2018



Abstract: Anti-lipopolysaccharide factors (ALFs) are antimicrobial peptides with a central β -hairpin structure able to bind to microbial components. Mining sequence databases for ALFs allowed us to show the remarkable diversity of ALF sequences in shrimp. We found at least seven members of the ALF family (Groups A to G), including two novel Groups (F and G), all of which are encoded by different loci with conserved gene organization. Phylogenetic analyses revealed that gene expansion and subsequent diversification of the ALF family occurred in crustaceans before shrimp speciation occurred. The transcriptional profile of ALFs was compared in terms of tissue distribution, response to two pathogens and during shrimp development in *Litopenaeus setiferus*, the most cultivated species. ALFs were found to be constitutively expressed in hemocytes and to respond differently to tissue damage. While synthetic β -hairpins of Groups E and G displayed both antibacterial and antifungal activities, no activity was recorded for Group F β -hairpins. Altogether, our results showed that ALFs form a family of shrimp AMPs that has been the subject of intense diversification. The different genes differ in terms of tissue expression, regulation and function. These data strongly suggest that multiple selection pressures have led to functional diversification of ALFs in shrimp.

Keywords: host defense peptide; antimicrobial peptide; anti-LPS factor; host-microbe relationship; functional diversity; invertebrate immunity; crustacean; antimicrobial activity

ANEXO 5. Artigo publicado em co-autoria durante período de pós-graduação no LIAA.

Fish and Shellfish Immunology 86 (2019) 82–92



Contents lists available at ScienceDirect

Fish and Shellfish Immunology

Journal homepage: www.elsevier.com/locate/fsi



Full length article

***Litopenaeus vannamei* stylicins are constitutively produced by hemocytes and intestinal cells and are differentially modulated upon infections**

Natanael Dantas Farias^{a,1}, Marcelo Falchetti^{a,1}, Gabriel Machado Matos^a, Paulina Schmitt^b,
Cairé Barreto^a, Nicolas Argenta^a, Jean-Luc Rolland^c, Evelyne Bachère^c, Luciane Maria Perazzolo^a,
Rafael Diego Rosa^{a,*}

^a Laboratory of Immunology Applied to Aquaculture, Department of Cell Biology, Biotechnology and Genetics, Federal University of Santa Catarina, 88040-900, Florianópolis, SC, Brazil
^b Laboratorio de Genética e Immunología Molecular, Instituto de Biología, Facultad de Ciencias, Pontificia Universidad Católica de Valparaíso, 2370023, Valparaíso, Chile
^c Institut des Hautes Études Biologiques-Interactions, Université de Montpellier, CNRS, IFREMER, Université de Perpignan Via Domitia, 34090, Montpellier Cedex 5, France

| ARTICLE INFO | ABSTRACT |
|--|--|
| <p>Keywords:</p> <p>Crustacean</p> <p>Penaeid shrimp</p> <p>Invertebrate immunity</p> <p>Antimicrobial peptide</p> <p>Host defense peptide</p> <p>Molecular diversity</p> | <p>Stylicins are anionic antimicrobial host defense peptides (AAMPs) composed of a proline-rich N-terminal region and a C-terminal portion containing 13 conserved cysteine residues. Here, we have increased our knowledge about these unexplored crustacean AAMPs by the characterization of novel stylicin members in the most cultivated penaeid shrimp, <i>Litopenaeus vannamei</i>. We showed that the <i>L. vannamei</i> stylicin family is composed of two members (<i>Lvan-Stylicin1</i> and <i>Lvan-Stylicin2</i>) encoded by different loci which vary in gene copy number. Unlike the other three gene-encoded antimicrobial peptide families from penaeid shrimp, the expression of <i>Lvan-Stylicins</i> is not restricted to hemocytes. Indeed, they are also produced by the columnar epithelial cells lining the midgut and its anterior caecum. Interestingly, <i>Lvan-Stylicins</i> are simultaneously transcribed at different transcriptional levels in a single shrimp and are differentially modulated in hemocytes after infections. While the expression of both genes showed to be responsive to damage-associated molecular patterns, only <i>Lvan-Stylicin2</i> was induced after a <i>Vibrio</i> infection. Besides, <i>Lvan-Stylicins</i> also showed a distinct pattern of gene expression in the three portions of the midgut (anterior, middle and posterior) and during shrimp development. We provide here the first evidence of the diversity of the stylicin antimicrobial peptide family in terms of sequence and gene expression distribution and regulation.</p> |

ANEXO 6. Artigo publicado em co-autoria durante período de pós-graduação no LIAA.

RESEARCH ARTICLE
Applied and Environmental Science

The Ancestral N-Terminal Domain of Big Defensins Drives Bacterially Triggered Assembly into Antimicrobial Nanonets

Karine Loth,^{a,b} Agnès Vergnes,^c Caíre Barreto,^{c,d} Sébastien N. Voisin,^e Hervé Meudal,^a Jennifer Da Silva,^f Albert Bressan,^{c,d} Nawal Belmadi,^f Evelynne Bachère,^c Vincent Aucagne,^a Chantal Cazeville,^g Hélène Marchandin,^h Rafael Diego Rosa,^d Philippe Bulet,^{a,i} Lhoussine Toulet,^j Agnès F. Delmas,^k Delphine Destoumieux-Garzon^l

^aCentre de Biophysique Moléculaire UPR4301 CNRS, Orléans, France

^bURR CoST, Université d'Orléans, Orléans, France

^cHPE, Université de Montpellier, CNRS, Ifremer, Université de Perpignan Via Domitia, Montpellier, France

^dLaboratory of Immunology Applied to Aquaculture, Department of Cell Biology, Embryology and Genetics, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil

^ePlateforme BioPark d'Archamps, Archamps Technopole, Archamps, France

^fEquipe mBio Institut Pasteur/Paris V Mucoviscidose et Bronchopathies Chroniques, Institut Pasteur, Paris, France

^gCOMET, Plateau de microscope électronique, Plateforme Montpellier RIO Imaging, Montpellier, France

^hHydrosciences Montpellier, Département de Microbiologie, CHU Nîmes, CNRS, IRD, Université de Montpellier, Nîmes, France

ⁱInstitute for Advanced Biosciences, CR Université Grenoble Alpes, CNRS UMR5309, La Tronche, France

ABSTRACT Big defensins, ancestors of β -defensins, are composed of a β -defensin-like C-terminal domain and a globular hydrophobic ancestral N-terminal domain. This unique structure is found in a limited number of phylogenetically distant species, including mollusks, ancestral chelicerates, and early-branching cephalochordates, mostly living in marine environments. One puzzling evolutionary issue concerns the advantage for these species of having maintained a hydrophobic domain lost during evolution toward β -defensins. Using native ligation chemistry, we produced the oyster *Crassostrea gigas* BigDef1 (Cg-BigDef1) and its separate domains. Cg-BigDef1 showed salt-stable and broad-range bactericidal activity, including against multidrug-resistant human clinical isolates of *Staphylococcus aureus*. We found that the ancestral N-terminal domain confers salt-stable antimicrobial activity to the β -defensin-like domain, which is otherwise inactive. Moreover, upon contact with bacteria, the N-terminal domain drives Cg-BigDef1 assembly into nanonets that entrap and kill bacteria. We speculate that the hydrophobic N-terminal domain of big defensins has been retained in marine phyla to confer salt-stable interactions with bacterial membranes in environments where electrostatic interactions are impaired. Those remarkable properties open the way to future drug developments when physiological salt concentrations inhibit the antimicrobial activity of vertebrate β -defensins.

IMPORTANCE β -Defensins are host defense peptides controlling infections in species ranging from humans to invertebrates. However, the antimicrobial activity of most human β -defensins is impaired at physiological salt concentrations. We explored the properties of big defensins, the β -defensin ancestors, which have been conserved in a number of marine organisms, mainly mollusks. By focusing on a big defensin from oyster (Cg-BigDef1), we showed that the N-terminal domain lost during evolution toward β -defensins confers bactericidal activity to Cg-BigDef1, even at high salt concentrations. Cg-BigDef1 killed multidrug-resistant human clinical isolates of *Staphylococcus aureus*. Moreover, the ancestral N-terminal domain drove the assembly of the big defensin into nanonets in which bacteria are entrapped and killed. This discovery may explain why the ancestral N-terminal domain has been maintained in diverse marine phyla and creates a new path of discovery to design β -defensin derivatives active at physiological and high salt concentrations.

Citation Loth K, Vergnes A, Barreto C, Voisin SN, Meudal H, Da Silva J, Bressan A, Belmadi N, Bachère E, Aucagne V, Cazeville C, Marchandin H, Rosa RD, Bulet P, Toulet L, Delmas AF, Destoumieux-Garzon D. 2019. The ancestral N-terminal domain of big defensins drives bacterially triggered assembly into antimicrobial nanonets. mBio 10:e01821-19. <https://doi.org/10.1128/mBio.01821-19>.

Invited Editor Bryan W. Davies, University of Texas at Austin

Editor M. Stephen Trent, University of Georgia

Copyright © 2019 Loth et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Delphine Destoumieux-Garzon, ddestoum@ipromet.fr.

Received 10 July 2019

Accepted 17 September 2019

Published 22 October 2019

ANEXO 7. Artigo publicado em co-autoria durante período de pós-graduação no LIAA.

Journal of Invertebrate Pathology 182 (2021) 107586



ELSEVIER

Contents lists available at [ScienceDirect](#)

Journal of Invertebrate Pathology

journal homepage: www.elsevier.com/locate/jip





Farfantepenaeus gene-encoded antimicrobial peptides: Identification, molecular characterization and gene expression in response to fungal infections

Luiz Gustavo Vasconcelos Machado, Priscila Goncalves¹, Cairé Barreto, Luciane Maria Perazzolo, Rafael Diego Rosa^{*}

Laboratory of Immunology Applied to Aquaculture, Department of Cell Biology, Embryology and Genetics, Federal University of Santa Catarina, 88040-900 Florianópolis, SC, Brazil

ARTICLE INFO

Keywords:
Crustacean
Host defense peptide (HDP)
Invertebrate immunity
Molecular diversity
Fusariosis

ABSTRACT

The aim of this study was to identify and characterize, at the molecular and transcriptional levels, sequences encoding the different members of the four families of shrimp antimicrobial peptides (AMPs) in species of the genus *Farfantepenaeus*. The identification of the AMP sequences was performed by *in silico* analysis as well as by molecular cloning and nucleotide sequencing. We identified all seven shrimp ALFs (ALF-A to ALF-G), both Type IIa and Type IIb crustins as well as two stylicins (STY1 and STY2) in *Farfantepenaeus*. Only two genes (PEN1/2 and PEN4) of the four-member penaeidin family (PEN1/2 to PEN5) were found and this is the first report of stylicins as well as of several additional members of ALFs, crustins and penaeidins in species of the genus *Farfantepenaeus*. All AMP genes have shown to be constitutively transcribed in the shrimp immune cells (hemocytes), except for ALF-G. Finally, the transcriptional profile of the different AMPs was assessed in the hemocytes of *F. paulseni* (pink shrimp) following an experimental infection with the opportunistic filamentous fungus *Fusarium solani*. We found that while the expression of ALF-B was induced at 24 h, the STY2 gene was down-regulated at 48 h post-challenge. These results provide evidence of the molecular diversity of AMPs from shrimp of the genus *Farfantepenaeus* in terms of sequences, biochemical properties and expression profiles in response to infectious diseases.
