

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
Programa de Pós-graduação em Parasitologia

**Sobre as Lipocalinas expressas em *Rhodnius prolixus*, um vetor da
doença de Chagas**

Daniela Viana dos Santos

Belo Horizonte

2021

Daniela Viana dos Santos

**Sobre as Lipocalinas expressas em *Rhodnius prolixus*, um vetor da
doença de Chagas**

Versão final

Dissertação apresentada ao Programa de Pós-Graduação em Parasitologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito à obtenção do título de Mestre em Parasitologia.

Área de concentração: Entomologia

Orientador: Prof. Dr. Leonardo Barbosa Koerich

Belo Horizonte

2021

- 043 Santos, Daniela Viana dos.
Sobre as Lipocalinas expressas em *Rhodnius prolixus*, um vetor da doença de Chagas [manuscrito] / Daniela Viana dos Santos. - 2021.
116 f. : il. ; 29,5 cm.
- Orientador: Prof. Dr. Leonardo Barbosa Koerich.
Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Parasitologia.
1. Parasitologia. 2. Lipocalinas. 3. *Rhodnius*. I. Koerich, Leonardo Barbosa. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 576.88/.89

**ATA DE DEFESA DE DISSERTAÇÃO DE MESTRADO****Daniela Viana dos Santos**424/2021/07
entrada
1º/2019
2019654010

Às nove horas do dia 30 de junho do ano de 2021, reuniu-se, por videoconferência, sala <https://meet.google.com/beh-qmhg-yca>, a Comissão Examinadora da Dissertação, indicada pelo Colegiado do Programa de Pós-Graduação em Parasitologia, para julgar, em exame final, o trabalho final intitulado: “**Sobre as Lipocalinas expressas em *Rhodnius prolixus*, um vetor da doença de Chagas**” área de concentração: **Entomologia**. Abrindo a sessão, o(a) Presidente da Comissão, **Dr. Leonardo Barbosa Koerich**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao(a) candidato(a), para apresentação de seu trabalho. Seguiu-se a argüição pelos examinadores, com a respectiva defesa do(a) candidato(a). Logo após, a Comissão se reuniu, sem a presença do(a) candidato(a) e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:

| Prof./Pesq. | Instituição | Indicação |
|---------------------------------------|-------------|-----------|
| Dr. Leonardo Barbosa Koerich | UFMG | Aprovada |
| Dra. Daniella Castanheira Bartholomeu | UFMG | Aprovada |
| Dr. Nelder Figueiredo Gontijo | UFMG | Aprovada |

Expedição do resultado final.

CONSIDERAÇÕES GERAIS – SOBRE TESE APRESENTADA PELO (A) O (A) CANDIDATO (A):

| | |
|---------------|--|
| Exigências | --- |
| Recomendações | --- |
| Sugestões | A banca repassou para a aluna e o orientador sugestões que serão incorporadas a versão final da dissertação. |

Pelas indicações, o(a) candidato(a) foi considerado(a): **Aprovada**

O resultado final foi comunicado publicamente ao(a) candidato(a) pelo Presidente da Comissão. Nada mais havendo a tratar o(a) Presidente encerrou a reunião e lavrou a presente ATA, que será assinada digitalmente por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 30 de junho de 2021.**Dr. Leonardo Barbosa Koerich**
(Orientador)**Dra. Daniella Castanheira Bartholomeu****Dr. Nelder Figueiredo Gontijo**

Aos meus professores.

Agradecimentos

Agradeço ao meu orientador, professor Leonardo Koerich, pela parceria, dedicação, paciência e entusiasmo. Obrigada pela oportunidade de aprender e pela confiança em mim depositada.

Aos demais professores e colegas do LFIH por tornarem essa experiência prazerosa e por compartilharem, de forma tão generosa, os seus conhecimentos. Foi inspirador estar numa equipe com tanto profissionalismo, competência e paixão.

A todos os professores, técnicos e demais funcionários da UFMG, do Departamento de Parasitologia e do Programa de Pós-graduação, por contribuírem para a minha formação científica e humana e por fazerem a diferença dentro e fora da instituição.

Aos animais utilizados nos experimentos laboratoriais, por permitirem o avanço da Ciência.

Às agências de fomento, CNPq, CAPES e FAPEMIG, pelo apoio financeiro.

Aos meus alunos, por serem minha esperança de um futuro melhor.

Aos meus colegas da turma de mestrado, em especial minhas companheiras Kenya Malta e Larissa Procópio, pelas risadas e pela cumplicidade.

Aos meus queridos amigos, Taci, Larissa, Ruan, Pedro, Laíza, Natália, Hortência, Aninha, Ágata, Julinha, Thais e Cris, por serem minha segunda família e me apoiarem em todas as áreas e fases da minha vida.

Ao Cassiano, pelo amor e companheirismo de todos os dias e por acreditar em mim como ninguém.

E por fim, ao meu pai, minha mãe e minha irmã, por serem minha base e por terem me ensinado o que realmente importa.

RESUMO

O processo da hematofagia por artrópodes tem sido um dos principais alvos de estudos na interação parasito-hospedeiro e o triatomíneo *Rhodnius prolixus*, transmissor do protozoário *Trypanosoma cruzi*, tem sido um dos principais modelos para essas pesquisas. Ainda na década de 1980 foi identificado que, entre as proteínas salivares para ruptura da homeostasia do hospedeiro vertebrado, as lipocalinas estavam entre as proteínas mais relevantes para este processo. Desde então, 38 lipocalinas foram identificadas nas glândulas salivares de *R. prolixus*, incluindo 19 Nitroforinas (que atuam na vasodilatação) e outras 19 lipocalinas que atuam em diferentes mecanismos para inibição de agregação plaquetária. Em 2014, um abrangente sequenciamento de RNA de diferentes tecidos de *R. prolixus* foi publicado, sugerindo diversidade e abundância das lipocalinas. O presente trabalho tem como objetivo realizar uma revisão das lipocalinas de *R. prolixus*, combinando dados do transcriptoma e genoma para identificar e descrever, estrutural e evolutivamente, novas lipocalinas expressas em *R. prolixus*. Para isso, tanto as lipocalinas já descritas quanto os transcritos inicialmente identificados como lipocalinas foram mapeados ao genoma utilizando a ferramenta BLAST. Genes identificados como novos tiveram sua estrutura anotada com a ferramenta GeneWise. Com o objetivo de inferir a relação evolutiva de ortologia entre as lipocalinas, e dar suporte as funções previstas para cada lipocalina, todas as sequências de aminoácidos (das já descritas na literatura e das recém-identificadas) foram utilizadas para a construção de árvores filogenéticas. Foram identificadas um total de 29 lipocalinas inéditas, sendo 5 novos vasodilatadores (todas nitroforinas), 5 BAPs (que atuam em inibição de agregação plaquetária e vasodilatação), 11 inibidoras de agregação plaquetária e 8 lipocalinas pertencentes a uma nova família (apolipoproteínas) cuja função é desconhecida. Além disso, foi observado que várias das lipocalinas também apresentam expressão em diferentes tecidos de *R. prolixus*, incluindo intestino e órgãos reprodutores. Para mais que novas lipocalinas identificadas e um mapeamento que localiza as novas e antigas lipocalinas no genoma de *R. prolixus*, o presente estudo também permitiu a revisão funcional e de nomenclatura de algumas das lipocalinas já identificadas. Este trabalho reforça que estamos distantes de compreender o papel de lipocalinas na fisiologia de *R. prolixus*, e que estudos funcionais dessa família ainda são de grande relevância.

Palavras-Chaves: Lipocalinas; Nitroforina; Triabina; Pallidipina; *Rhodnius prolixus*

ABSTRACT

The hematophagy process by arthropods has been one of the main targets of studies in the parasite-host interaction, and the kissing-bug *Rhodnius prolixus*, vector of the protozoan *Trypanosoma cruzi*, has been one of the main models for this research. Still in the 1980s, it was identified that, among the salivary proteins for disrupting vertebrate host homeostasis, lipocalins were among the most relevant proteins for this process. Since then, 38 lipocalins have been identified in the salivary glands of *R. prolixus*, including 19 nitrophorins (which act on vasodilation) and 19 lipocalins that act on different mechanisms to inhibit platelet aggregation. In 2014, an RNA sequencing of different tissues of *R. prolixus* was published, suggesting diversity and abundance of lipocalins. The present work aims to carry out a review of lipocalins from *R. prolixus*, combining transcriptome and genome data to annotate, structurally and evolutionarily, new lipocalins expressed in *R. prolixus*. For this, the lipocalins already known and the new transcripts, identified as lipocalins, were mapped to the genome using a BLAST tool. Identified new genes had their structure annotated with GeneWise. To infer an evolutionary orthological relationship between lipocalins, and to support the predicted functions for each lipocalin, all amino acid sequences (those existing in the literature and those recently identified) were used to construct phylogenetic trees. We identified a total of 29 new lipocalins, 5 of which were new vasodilators (all nitrophorins), 5 BAPs (which act to inhibit platelet aggregation and vasodilation), 11 inhibitors of platelet aggregation and 8 lipocalins belonging to a new family (apolyproteins) with function unknown. In addition, we observed that several of the lipocalins are also expressed in different *R. prolixus* tissues, including gut and reproductive organs. In addition to newly identified lipocalins and a mapping the new and old lipocalins in the genome of *R. prolixus*, our study also carried out a review on functional status and nomenclature of some of the already identified lipocalins. Our study reinforces that we are far from understanding the role of lipocalins in the physiology of *R. prolixus*, and that studies of this family are still of great relevance.

Key words: Lipocalins; Nitrophorin; Triabin; Pallidipin; *Rhodnius prolixus*

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1. Ciclo de desenvolvimento do <i>R. prolixus</i> | 13 |
| Figura 2. Modelo tridimensional de uma lipocalina | 17 |
| Figura 3. Relação evolutiva das Lipocalinas de <i>Rhodnius prolixus</i> | 47 |
| Figura 4. Relações evolutivas das Nitroforinas de <i>R. prolixus</i> | 50 |
| Figura 5. Alinhamento das Nitroforinas de <i>Rhodnius prolixus</i> | 52 |
| Figura 6. Relações evolutivas das Proteínas Biogênicas de Ligação a Amina de <i>R. prolixus</i> | 56 |
| Figura 7. Alinhamento das Proteínas Biogênicas de Ligação a Amina de <i>Rhodnius prolixus</i> | 59 |
| Figura 8. Relações evolutivas das Inibidoras de Agregação Plaquetária (PAIs), Triabinas, Pallidípinas e Procalinas de <i>R. prolixus</i> | 61 |
| Figura 9. Alinhamento e análise de domínio das Triabinas de função hipotética em <i>R. prolixus</i> – grupo Indefinido II | 63 |
| Figura 10. Alinhamento e análise de domínio das Triabinas de função hipotética em <i>R. prolixus</i> – grupo Indefinido III e IV | 64 |
| Figura 11. Relações evolutivas das Apolipoproteínas de <i>R. prolixus</i> | 68 |
| Figura 12. Alinhamento e análise de domínio das Apolipoproteínas de <i>R. prolixus</i> | 69 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1. Lipocalinas salivares de <i>Rhodnius prolixus</i> descritas até o presente estudo .. | 19 |
| Tabela 2. Mapeamento das lipocalinas depositadas no GeneBank ao genoma e proteoma de <i>R. prolixus</i> | 32 |
| Tabela 3. Transcritos de <i>R. prolixus</i> similares a lipocalinas da família de Nitroforinas | 34 |
| Tabela 4. Transcritos de <i>R. prolixus</i> similares a lipocalinas da família de Inibidores de Agregação Plaquetária (PAI)..... | 35 |
| Tabela 5. Transcritos de <i>R. prolixus</i> similares a lipocalinas da família de Pallidipinas. | 35 |
| Tabela 6. Transcritos de <i>R. prolixus</i> similares a lipocalinas da família de Triabinas..... | 36 |
| Tabela 7. Transcritos de <i>R. prolixus</i> similares a lipocalinas da família de Apolipoproteínas | 36 |
| Tabela 8. Transcritos de <i>R. prolixus</i> similares a lipocalinas inespecíficas | 36 |
| Tabela 9. Mapeamento das lipocalinas inicialmente identificadas como Nitroforinas... | 39 |
| Tabela 10. Mapeamento e identificação das lipocalinas inicialmente identificadas como PAI..... | 40 |
| Tabela 11. Mapeamento das lipocalinas inicialmente identificadas como Pallidipinas. | 41 |
| Tabela 12. Mapeamento das lipocalinas inicialmente identificadas como Triabinas..... | 42 |
| Tabela 13. Mapeamento das lipocalinas inicialmente identificadas como Apolipoproteínas | 43 |
| Tabela 14. Mapeamento das lipocalinas inicialmente identificadas como “Inespecíficas” | 44 |
| Tabela 15. Nitroforinas expressas em diferentes tecidos de <i>Rhodnius prolixus</i> | 54 |
| Tabela 16. Proteínas Biogênicas de Ligação Amina expressas em diferentes tecidos de <i>Rhodnius prolixus</i> | 58 |
| Tabela 17. Resultado do BLAST recíproco para lipocalinas indefinidas | 62 |
| Tabela 18. PAI, Pallidipinas, Procalinas e Triabinas expressas em diferentes tecidos de <i>Rhodnius prolixus</i> | 66 |
| Tabela 19. Apolipoproteínas expressas em diferentes tecidos de <i>R. prolixus</i> | 70 |
| Tabela 20. Revisão das lipocalinas de <i>Rhodnius prolixus</i> | 91 |

Lista de abreviações

ADP – adenosina difosfato

APO – apolipoproteína

BAP – proteína biogênica de ligação a amina (do inglês biogenic amine-binding protein)

CG – corpo gorduroso

DNA – ácido desoxirribonucleico

HPLC – cromatografia líquida de alta eficiência (do inglês high performance liquid chromatography)

IA – intestino anterior

IP – intestino posterior

kDa – quilodalton

NP – nitroforina

OV - ovário

PAI – Inibidora de agregação plaquetária

RNA – ácido ribonucleico

RPK – reads por kilobase

Rt – reto

Te – testículo

TM – túbulo de Malpighi

Sumário

| | |
|---|----|
| 1. INTRODUÇÃO..... | 13 |
| 1.1. <i>Rhodnius prolixus</i> e a doença de Chagas | 13 |
| 1.2. Hematofagia..... | 15 |
| 1.3. A família das Lipocalinas | 17 |
| 1.3.1. Nitroforinas | 19 |
| 1.3.2. Proteínas Biogênicas de Ligação a Amina | 21 |
| 1.3.3. Inibidoras de agregação plaquetária (PAI)..... | 22 |
| 1.3.4. Pallidipinas | 23 |
| 1.3.5. Procalinas | 23 |
| 1.3.6. Triabinas..... | 23 |
| 1.3.7. Apolipoproteína D..... | 24 |
| 1.4. As lipocalinas esquecidas do transcriptoma de <i>Rhodnius prolixus</i> | 25 |
| 1.5. Justificativa | 26 |
| 2. OBJETIVOS..... | 27 |
| 2.1. Objetivo geral..... | 27 |
| 2.2. Objetivos específicos | 27 |
| 3. MATERIAIS E MÉTODOS..... | 28 |
| 3.1. O transcriptoma de <i>Rhodnius prolixus</i> | 28 |
| 3.2. Anotação estrutural das lipocalinas expressas em <i>R. prolixus</i> | 28 |
| 3.3. Análise evolutiva das lipocalinas | 29 |
| 3.4. Identificação de isoformas e duplicações gênicas..... | 29 |
| 4. RESULTADOS | 30 |
| 4.1. Mapeamento das lipocalinas conhecidas ao genoma e proteoma de <i>R. prolixus</i> . 30 | |
| 4.2. Identificação de Lipocalinas no transcriptoma de <i>R. prolixus</i> | 33 |
| 4.3. Mapeamento e anotação das lipocalinas transcritas ao genoma e proteoma predito de <i>R. prolixus</i> | 37 |
| 4.4. Análise inicial de ortologia das lipocalinas do transcriptoma de <i>R. prolixus</i> | 44 |
| 4.5. Estrutura, evolução e expressão das Nitroforinas no transcriptoma de <i>R. prolixus</i> | 49 |
| 4.6. Estrutura, evolução e expressão das Proteínas Biogênicas de Ligação a Amina (BAP) no transcriptoma de <i>Rhodnius prolixus</i> | 55 |
| 4.7. Estrutura, evolução e expressão das proteínas Inibidoras de agregação plaquetária (PAI), Triabinas, Pallidipinas e Procalinas no transcriptoma de <i>R. prolixus</i> | 60 |

| | |
|--|-----|
| 4.8. Estrutura, evolução e expressão das Apolipoproteínas no transcriptoma de <i>R. prolixus</i> | 67 |
| 4.9. Genes não resolvidos | 71 |
| 5. DISCUSSÃO | 72 |
| 5.1. Uma revisão do transcriptoma de <i>Rhodnius prolixus</i> | 73 |
| 5.2. Sobre as lipocalinas já conhecidas em <i>R. prolixus</i> | 74 |
| 5.3. Novas lipocalinas de <i>Rhodnius prolixus</i> | 76 |
| 5.3.1. Novas Nitroforinas em <i>R. prolixus</i> | 80 |
| 5.3.2. Novas Proteínas Biogênicas de Ligação a Amina (BAP) em <i>R. prolixus</i> | 82 |
| 5.3.3. As Apolipoproteínas, uma nova família de lipocalinas em <i>R. prolixus</i> | 83 |
| 5.3.4. O caso das lipocalinas inibidoras de agregação plaquetária: diferenças e semelhanças entre PAIs, triabinas, pallidipinas e procalinas | 84 |
| 5.4. Sobre as lipocalinas não salivares de <i>Rhodnius prolixus</i> | 87 |
| 5.5. Uma revisão da nomenclatura e possível função das lipocalinas de <i>R. prolixus</i> . 90 | |
| 6. CONCLUSÃO | 93 |
| REFERÊNCIAS | 94 |
| ANEXO A | 105 |
| ANEXO B..... | 107 |
| ANEXO C..... | 108 |
| ANEXO D | 110 |
| ANEXO E..... | 111 |
| ANEXO F..... | 112 |
| ANEXO G | 114 |

1. INTRODUÇÃO

1.1. *Rhodnius prolixus* e a doença de Chagas

Rhodnius prolixus, um dos insetos conhecidos popularmente como barbeiros, é um inseto pertencente à ordem Hemiptera, família Reduviidae, subfamília Triatominae (GALVÃO, 2003). Tem uma distribuição geográfica restrita à América Latina, com maior abundância na Venezuela, Colômbia e em países da América Central (DUJARDIN *et al.*, 1998; DIOTAIUTI & BARBOSA, 2015). Apresenta hábitos noturnos e tem o ciclo de vida do tipo paurometábolo, no qual o inseto passa pelos estágios de ovo, ninfa (com cinco estádios ninfais) e adulto (Figura 1). Tanto as formas imaturas (a partir do primeiro estágio ninfal) quanto os adultos, machos e fêmeas, apresentam o mesmo hábito alimentar: a hematofagia. O repasto sanguíneo é realizado em animais vertebrados endotérmicos (aves e mamíferos) (LENT & WYGODZINSKY, 1979; NEVES, 2016).

No ambiente silvestre, *R. prolixus* vive na copa das árvores, especialmente palmeiras, próximo a ninhos e tocas de animais que servirão para ele de fonte alimentar. No entanto, devido aos processos antropológicos de invasão e destruição do ambiente natural, esse inseto passou a colonizar também o domicílio e o peridomicílio, tanto no ambiente rural quanto no urbano, se aproximando do homem e, ocasionalmente, fazendo dele uma fonte alimentar (COURA, 2003; LAZZARI *et al.*, 2013).



Figura 1. Ciclo de desenvolvimento do *R. prolixus*. De ovo a adulto, passando por cinco estádios ninfais. Foto retirada de Paim (2014).

Os nutrientes presentes no sangue do hospedeiro são necessários para a mudança de estágio evolutivo do inseto e, no caso das fêmeas adultas, são essenciais para a maturação dos ovários e para a ovoposição, sendo, portanto, importantes para a fertilidade (LENT & WYGODZINSKY, 1979). Já para os machos, a alimentação

sanguínea estimula a espermatogênese, o acasalamento e a produção de fatores masculinos importantes para a fertilidade feminina (SONENSHINE *et al.*, 2011).

Os ovos de *R. prolixus* são colocados separados ou, mais comumente, aderidos ao substrato. São ovos operculados, de coloração esbranquiçada logo após a postura e que se tornam rosados à medida que o embrião se desenvolve. A ovoposição é feita entre 20 e 30 dias após a cópula e a incubação dos ovos dura cerca de 20 dias (DIOTAIUTI & BARBOSA, 2015). As ninfas que eclodem dos ovos se diferem dos adultos, dentre outras características, pelo menor tamanho, ausência de ocelos, ausência de asas e pela imaturidade sexual (ROCHA & SANTOS, 2014).

Rhodnius prolixus é um dos mais importantes vetores biológicos do protozoário *Trypanosoma cruzi*, agente etiológico da doença de Chagas (NATURE, 2010). Essa enfermidade, considerada negligenciada pela Organização Panamericana de Saúde - OPAS, é endêmica do continente americano e possui grande importância epidemiológica, especialmente na América Latina (DIAS, 2015).

O barbeiro adquire o *T. cruzi* durante um repasto sanguíneo infectante, ingerindo as formas tripomastigotas do protozoário presentes no sangue do hospedeiro infectado. Após se desenvolver no trato intestinal do inseto, formas tripomastigotas metacíclicas do protozoário são eliminadas nas fezes e urina do inseto durante um repasto sanguíneo. Essas formas infectantes, em contato com mucosa ou pele lesionada, podem penetrar no hospedeiro vertebrado e, após uma infecção local, alcançar sua corrente sanguínea e desencadear o ciclo da doença de Chagas (GARCIA *et al.*, 2007).

Já foi observado que uma fêmea de *R. prolixus* infectada com *T. cruzi* não é capaz de transmitir o parasito aos seus ovos durante o desenvolvimento destes, ou seja, não ocorre transmissão vertical (DIOTAIUTI & BARBOSA, 2015). No entanto, a partir da fase de ninfa de primeiro estágio, o inseto já é capaz de adquirir o parasito durante um repasto sanguíneo infectante e transmiti-lo a outro hospedeiro vertebrado (DIOTAIUTI & BARBOSA, 2015).

Para além da transmissão vetorial, de caráter contaminativo, *T. cruzi* pode ser transmitido verticalmente (transmissão congênita), por transfusão sanguínea ou transplante de órgãos infectados, ou ainda por via oral, através a ingestão de alimentos contaminados com insetos infectados ou seus dejetos (DIAS *et al.*, 2011). Esta última via de transmissão tem se tornado relevante nos últimos anos, sendo responsável por pequenos surtos em regiões endêmicas (FERREIRA *et al.*, 2014).

Estima-se que haja, em todo o globo, cerca de 8 milhões de pessoas acometidas pela doença de Chagas, que leva a óbito, aproximadamente, 10 mil pessoas por ano. A incidência anual da doença, considerando apenas a transmissão por via vetorial, é de cerca de 24 mil novos casos, o que mostra a importância do controle vetorial ainda na atualidade (WHO, 2019).

Além da sua importância vetorial, *Rhodnius prolixus* é uma espécie modelo para o estudo da fisiologia de insetos desde a década de 1950 (WIGGLESWORTH, 1976; CHIANG *et al.*, 2013). Atualmente, o estudo dessa espécie é facilitado pelo fato de seu genoma e transcriptoma estarem sequenciados e disponíveis nos bancos de dados, bem como pela facilidade de se manter uma colônia em ambiente de laboratório quando comparado a outros triatomíneos, já que *R. prolixus* é menos exigente quanto à fonte alimentar e possui um ciclo biológico que se desenvolve mais rapidamente.

Os conhecimentos adquiridos a partir desse organismo modelo têm possibilitado a compreensão de diversos sistemas fisiológicos, não só dos triatomíneos como um todo, mas também de outros grupos de insetos hematófagos (BEAMENT *et al.*, 1980). Na fisiologia alimentar, a saliva e seus componentes moleculares antihemostáticos têm sido alvo de estudos cada vez mais elucidativos, com grandes contribuições do ponto de vista terapêutico e farmacológico (HERNANDEZ-VARGAS *et al.*, 2016; MONTFORT *et al.*, 2000).

1.2. Hematofagia

Se alimentar de sangue de um animal vertebrado exige uma complexa interação entre o ectoparasito e seu hospedeiro. Para que a hematofagia funcione alguns desafios devem ser enfrentados. Primeiramente, é necessário que o inseto hematófago tenha um aparelho bucal capaz de penetrar a epiderme, encontrar o vaso sanguíneo e sugar o sangue do hospedeiro sem incomodá-lo a ponto de fazê-lo interromper a investida. Além disso, deve-se garantir que o fluxo sanguíneo no local da alimentação permaneça intenso e constante sem que o sangue do hospedeiro coagule, o que inviabilizaria a sua sucção (GORDON & CREWE, 1948; SANT'ANNA *et al.*, 2017).

Os animais vertebrados desenvolveram, ao longo de seu processo evolutivo, respostas fisiológicas complexas para impedir a perda de sangue, como a hemostasia e a resposta imune. Tais processos são ativados não somente em decorrência de lesões acidentais, mas também naquelas provocadas pela picada de insetos hematófagos em busca de um repasto sanguíneo (ANDRADE *et al.*, 2005; ARAÚJO *et al.*, 2012).

A hemostasia, após uma lesão de vaso sanguíneo, controla a perda de sangue por meio da vasoconstrição, agregação plaquetária e cascata da coagulação sanguínea. A inflamação, ativada em resposta a uma lesão tecidual, provoca dor, vermelhidão e calor, aumenta a permeabilidade vascular e a liberação de citocinas, ativa a via alternativa ou clássica do complemento e recruta células de defesa a fim de promover o reparo tecidual (DE ARAÚJO *et al.*, 2012; BASANOVA *et al.*, 2002; RIBEIRO *et al.*, 1998; RIBEIRO *et al.*, 2004).

A resposta imune, especialmente quando se defronta com antígenos dérmicos recorrentes, desencadeia reações que podem levar a uma mudança de comportamento do hospedeiro, que passa, por exemplo, a identificar e remover o inseto hematófago. Além disso, a presença constante de antígenos do inseto pode promover respostas imunológicas mais rápidas e eficientes, levando à neutralização de moléculas que são importantes para o seu sucesso alimentar (GILLESPIE *et al.*, 2000; RIBEIRO *et al.*, 2004).

No entanto, os insetos hematófagos contam com um arsenal de substâncias químicas presentes em sua saliva capazes de driblar os desafios impostos pela hemostasia e resposta imune do hospedeiro vertebrado. Durante o repasto sanguíneo, a saliva, contendo tais substâncias, é inoculada no local da picada e desencadeia diversas reações químicas que possibilitam uma alimentação sanguínea bem-sucedida (CHAMPAGNE, 1994; ANDERSEN *et al.*, 2005).

Substâncias vasodilatadoras, antiplaquetárias e anticoagulantes estão presentes na saliva de todos os animais hematófagos. Na maioria dos casos, essas substâncias são diversas e abundantes e, muitas vezes, apresentam funções redundantes. O maxadilan, por exemplo, é um peptídeo de 6,5 kDa presente na saliva do flebotomíneo *Lutzomyia longipalpis* que atua nos receptores PACAP¹ (LERNER *et al.*, 2007) e é considerado o mais potente vasodilatador natural já descrito. Mosquitos do gênero *Anopheles*, por sua vez, possuem uma proteína peroxidase que atua na destruição de norepinefrina e serotonina e, portanto, tem função vasodilatadora (RIBEIRO & NUSSENZVEIG, 1993).

O óxido nítrico é outro exemplo de substância antiplaquetária e vasodilatadora comum na saliva de alguns insetos hematófagos. Por ser um gás instável, cada inseto

¹ O receptor PACAP (Peptídeo Ativador da Adenilato Ciclase Pituitária) possui diversas funções biológicas, dentre elas a estimulação da produção de óxido nítrico.

desenvolveu mecanismos para estabilizar e transportar essa molécula. *Rhodnius prolixus*, por exemplo, apresenta a proteína nitroforina – pertencente à família das lipocalinas – como protagonista nesse processo de transportar o óxido nítrico da saliva até o sítio de alimentação no hospedeiro vertebrado (RIBEIRO *et al.*, 1993; RIBEIRO *et al.*, 2004).

1.3. A família das Lipocalinas

As lipocalinas são proteínas encontradas em bactérias, plantas e animais e se caracterizam pela presença de uma estrutura terciária de β -barril formado por uma cadeia polipeptídica com 6 a 8 segmentos em conformação de folha β , que interagem entre si por ligações de hidrogênio. Toda a estrutura terciária do β -barril é estabilizada por uma, duas ou três pontes dissulfeto, formando uma espécie de cálice que é internamente revestido por resíduos de aminoácidos hidrofóbicos (Figura 2). Tal conformação confere às lipocalinas sua principal característica bioquímica: a capacidade de interagir e carrear pequenas moléculas de caráter hidrofóbico (GRZYB *et al.*, 2006; LAKSHMI *et al.*, 2015).

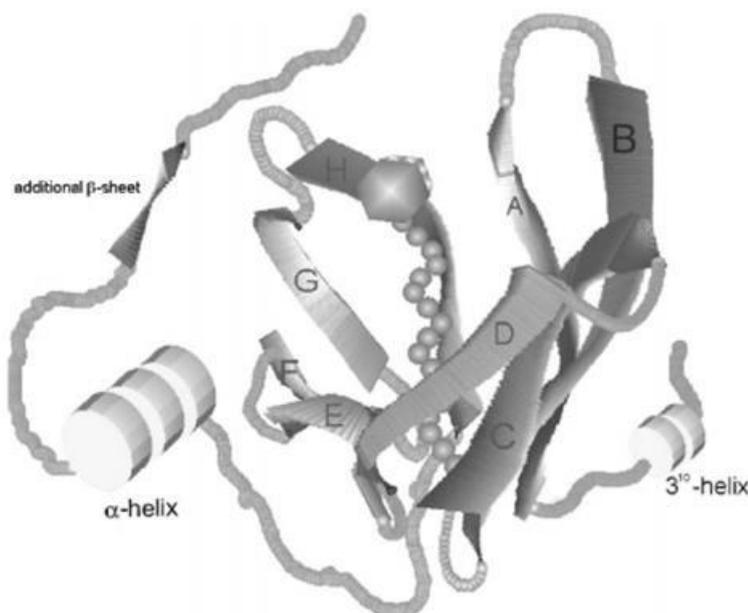


Figura 2. Modelo tridimensional de uma lipocalina. Estrutura de β -barril formada por oito folhas β (indicadas pelas letras de A a H). Baseado na estrutura cristalina da b-lactoglobulina bovina (KONTOPIDIS *et al.*, 2002).

Apesar de ser uma família proteica altamente conservada, com estruturas secundárias e terciárias muito semelhantes, as lipocalinas apresentam uma baixa

identidade na sequência de aminoácidos (<30%) e são bastante heterogêneas (MANS *et al.*, 2008; GRZYB *et al.*, 2006; FLOWER *et al.*, 2000). O domínio, de massa molecular entre 18 e 20 kDa, apresenta uma alta diversidade que, junto aos elementos estruturais adicionais, determinam as funções fisiológicas específicas de cada lipocalina (GRZYB *et al.*, 2006), que vão além da simples capacidade de transportar lipossolúveis.

Já foram descritas, em diversos organismos, lipocalinas associadas a importantes processos biológicos, como resposta imune, transporte de retinol, síntese biológica de prostaglandinas, transporte de feromônios, maturação de células e tecidos e interação espermatozoide-óvulo (GRZYB *et al.*, 2006; LEONE *et al.*, 2002; RIBEIRO *et al.*, 2014; SCHIEFNER *et al.*, 2015; URADE & EGUCHI, 2002). Os triatomíneos, especificamente, desenvolveram ao longo do processo evolutivo, lipocalinas salivares com funções anti-hemostáticas, como vasodilatação, anti-agregação plaquetária e anti-coagulação sanguínea (ANDERSEN *et al.*, 2005). Dentre as lipocalinas salivares mais estudadas em triatomíneos estão as nitroforinas, PAI (inibidoras de agregação plaquetária), pallidipinas e triabinas (BAO *et al.*, 2015; NOESKE-JUNGBLUT *et al.*, 1994; NOESKE-JUNGBLUT *et al.*, 1995; PADDOCK *et al.*, 2001).

Em *Rhodnius prolixus*, 24 lipocalinas salivares já foram descritas (Tabela 1). Durante o repasto sanguíneo, essas lipocalinas salivares são capazes de modular a hemostasia do hospedeiro de modo a manter o sangue em condições adequadas para a ingestão e, assim, viabilizar a hematofagia (RIBEIRO *et al.*, 2004). Tal participação das lipocalinas, observada não somente em triatomíneos como também em outros insetos hematófagos, como culicídeos e flebotomíneos, tem sido amplamente abordada em estudos de fisiologia alimentar de insetos hematófagos (SILVA *et al.*, 2009).

Tabela 1. Lipocalinas salivares de *Rhodnius prolixus* descritas até o presente estudo.

| Nº de acesso | Função hipotética | Referência |
|--------------|---|--------------------------------|
| Q26239.1 | Nitroforina 1 (NP-1) | Champagne <i>et al.</i> , 1995 |
| Q94733.1 | Nitroforina 3 (NP-3) | Champagne <i>et al.</i> , 1995 |
| Q94734.1 | Nitroforina 4 (NP-4) | Andersen <i>et al.</i> , 1998 |
| Q26241.1 | Nitroforina 2 (NP-2) | Andersen <i>et al.</i> , 2000 |
| Q6PQK2.1 | Nitroforina 7 (NP-7) | Andersen <i>et al.</i> , 2004 |
| AAO25746.1 | Proteína biogênica de ligação a amina-1 (BAP-1) | Andersen <i>et al.</i> , 2003 |
| 4GET_A | Proteína biogênica de ligação a amina-2 (BAP-2) | Xu <i>et al.</i> , 2013 |
| 4HFO_A | Proteína biogênica de ligação a amina-3 (BAP-3) | Xu <i>et al.</i> , 2013 |
| AAB09091.1 | Inibidora de agregação plaquetária-1 (PAI-1) | Francischetti, 2000 |
| AAB09090.1 | Inibidora de agregação plaquetária-2 (PAI-2) | Francischetti, 2000 |
| AAQ20838.1 | Inibidora de agregação plaquetária-3 (PAI-3) | Ribeiro <i>et al.</i> , 2004 |
| AAQ20820.1 | Inibidora de agregação plaquetária-4 (PAI-4) | Ribeiro <i>et al.</i> , 2004 |
| AAQ20819.1 | Inibidora de agregação plaquetária-5 (PAI-5) | Ribeiro <i>et al.</i> , 2004 |
| AAQ20818.1 | Inibidora de agregação plaquetária-6 (PAI-6) | Ribeiro <i>et al.</i> , 2004 |
| AAQ20817.1 | Inibidora de agregação plaquetária-7 (PAI-7) | Ribeiro <i>et al.</i> , 2004 |
| 5H9L_A | Inibidora de agregação plaquetária-4B (PAI-4B) | Jablonka <i>et al.</i> , 2016 |
| AAQ20826.1 | Pallidipina-1 | Ribeiro <i>et al.</i> , 2004 |
| AAQ20825.1 | Pallidipina-2 | Ribeiro <i>et al.</i> , 2004 |
| AAQ20828.1 | Procalina-1 | Ribeiro <i>et al.</i> , 2004 |
| AAQ20827.1 | Procalina-2 | Ribeiro <i>et al.</i> , 2004 |
| AAQ20824.1 | Triabina-1 | Ribeiro <i>et al.</i> , 2004 |
| AAQ20823.1 | Triabina-2 | Ribeiro <i>et al.</i> , 2004 |
| AAQ20822.1 | Triabina-3 | Ribeiro <i>et al.</i> , 2004 |
| AAQ20821.1 | Triabina-4 | Ribeiro <i>et al.</i> , 2004 |

1.3.1. Nitroforinas

Um estudo realizado em 1981 por Ribeiro e Garcia buscou compreender a importância da saliva na alimentação de insetos hematófagos. Barbeiros adultos da espécie *R. prolixus* foram salivarectomizados, ou seja, tiveram os ductos salivares cortados e as glândulas salivares removidas. Ao serem alimentados em coelhos, observou-se que os insetos salivarectomizados ingeriam menos sangue, numa menor velocidade, e precisavam picar o hospedeiro mais vezes em relação aos insetos controle (com ducto salivar e glândula salivar íntegros). Essas diferenças, no entanto, não foram observadas quando os mesmos insetos eram alimentados artificialmente (RIBEIRO & GARCIA, 1981). Esse trabalho mostrou que a saliva apresentava uma importante ação anti-hemostática e facilitava uma melhor canulação dos vasos sanguíneos do

hospedeiro. Mais tarde, em 1990, Ribeiro e colaboradores evidenciaram a presença de uma proteína salivar vasodilatadora capaz de se ligar ao óxido nítrico (NO) e promover o relaxamento da musculatura lisa e uma ação anti-plaquetária (RIBEIRO *et al.*, 1990). Nos trabalhos de 1993 e 1994 essa proteína nitrovasodilatadora foi mais bem caracterizada e passou a ser chamada de nitroforina (RIBEIRO *et al.*, 1993; RIBEIRO & WALKER., 1994).

As nitroforinas são lipocalinas de 20 kDa que apresentam um grupamento heme férrico (FerIII) em sua estrutura de β -barril. São as lipocalinas mais estudadas em insetos e uma das proteínas mais abordadas em estudos de fisiologia alimentar em *R. prolixus*, sendo descritas pela primeira vez em 1993 (Ribeiro *et al.*, 1993). Nesta espécie, já foram descritas sete nitroforinas, nomeadas de NP-1 (CHAMPAGNE *et al.*, 1995), NP-2 – ou prolixina-S (RIBEIRO *et al.* 1995; ANDERSEN *et al.*, 2000), NP-3 (CHAMPAGNE *et al.*, 1995), NP-4 (ANDERSEN *et al.*, 1998), NP-5, NP-6 e NP-7 (ANDERSEN *et al.*, 2004). A maioria destas nitroforinas foi sequenciada (Tabela 1), com exceção das NP-5 e NP-6.

Sabe-se que as nitroforinas apresentam três atividades distintas, sendo a mais bem estudada a capacidade de armazenar e transportar o óxido nítrico (NO). Esse composto, ao ser liberado no tecido-alvo do hospedeiro durante o repasto sanguíneo, se liga à guanilato ciclase solúvel (sGC), desencadeando uma série de sinais entre os quais o relaxamento da musculatura lisa e, com isso, a vasodilatação (MONCADA *et al.*, 1988). Essas lipocalinas também são capazes de se ligar e neutralizar a histamina, substância liberada por mastócitos com o objetivo de sinalizar uma resposta inflamatória e imunológica. Dessa maneira, as nitroforinas agem dificultando a inflamação e a resposta imune (MONTFORT *et al.*, 2000). Além disso, a nitroforina NP-2 já mostrou atividade anticoagulante por ser capaz de impedir a conversão do fator X em Xa², o que interfere na continuidade da cascata da coagulação (HELLMAN & HALKINS, 1965; RIBEIRO *et al.*, 1995; ZHANG *et al.*, 1998).

Já a nitroforina NP-7, descrita mais tarde por Andersen e colaboradores (2004) em uma biblioteca de cDNA criada a partir do material de glândulas salivares de ninfas de quinto estágio de *R. prolixus*, apresenta função vasodilatadora, anti-histamínica, anticoagulante e inibidora de agregação plaquetária. Essa nitroforina, além de se ligar ao óxido nítrico (NO) e à histamina, tem afinidade por membranas fosfolipídicas aniônicas,

² Fator X e fator Xa são fatores de coagulação que, na cascata da coagulação sanguínea, convertem a protrombina em trombina.

como as superfícies carregadas negativamente das plaquetas ativadas (ANDERSEN *et al.*, 2004). Devido a esta afinidade, a NP-7 bloqueia os locais de ligação dos fosfolipídios para o complexo da protrombinase nas superfícies das plaquetas ativadas, inibindo, assim, a ativação da protrombina. Carreando o NO, ela inibe a atividade do colágeno e a agregação plaquetária induzida por ADP, além de induzir a desagregação de plaquetas estimulada por ADP por meio de um mecanismo mediado por ADP (ANDERSEN *et al.*, 2004).

Em 2003, Moreira e colaboradores purificaram o conteúdo de glândulas salivares de *R. prolixus* em diferentes estádios de desenvolvimento. As frações de hemoproteínas foram coletadas separadamente e submetidas a fracionamento por meio de cromatografia em coluna de troca catiônica forte. Este trabalho identificou a presença de duas novas lipocalinas sendo expressas na glândula salivar de ninfas de primeiro estágio. Elas foram denominadas de nitroforina 5 (NP-5) e nitroforina 6 (NP-6), mas, até o presente trabalho, elas ainda não foram sequenciadas (MOREIRA *et al.*, 2003). Em 2004, Ribeiro e colaboradores sugeriram que a sequência de aminoácidos da NP-6 recém-descoberta era, na verdade, igual a da já descrita NP-2, e que a sequência da NP-5 difere apenas em um aminoácido da sequência da NP-4, o que levantou questionamentos sobre a existência dessas novas nitroforinas (RIBEIRO *et al.*, 2004).

Moreira e colaboradores também mostraram que as nitroforinas são expressas diferentemente em cada estágio de desenvolvimento do inseto (MOREIRA *et al.*, 2003), corroborando os resultados de trabalhos anteriores (SUN *et al.*, 1998). Em ninfas de primeiro estágio é possível encontrar, em suas glândulas salivares, a nitroforina NP-2, e as duas novas evidenciadas por eles (NP-5 e NP-6). No segundo estágio ninfal aparece a nitroforina NP-4. NP-1 surge apenas no terceiro estágio ninfal e NP-3 e NP-7 no quinto estágio ninfal (MOREIRA *et al.*, 2003, ANDERSEN *et al.*, 2004). À medida que o inseto se desenvolve, as recém-encontradas nitroforinas NP-5 e NP-6 têm a sua expressão reduzida, de modo que na fase adulta o inseto apresenta em suas glândulas salivares ao menos as outras cinco nitroforinas já descritas (NP-1, NP-2, NP-3, NP-4 e NP-7). Isso demonstra que, ao longo do seu desenvolvimento, o inseto muda (e aumenta) seu arsenal de substâncias anti-hemostáticas, que está longe de ser algo estático.

1.3.2. Proteínas Biogênicas de Ligação a Amina

Em 1982, Ribeiro observou a inibição da contração muscular induzida por serotonina em preparados de úteros de camundongos incubados com saliva de *R.*

prolixus, sugerindo uma atividade anti-serotonina da saliva desse inseto (RIBEIRO, 1982). Mais tarde, Andersen e colaboradores clonaram uma nova lipocalina salivar de *R. prolixus* e realizaram bioensaios que demonstraram resultados semelhantes. Essa nova lipocalina foi denominada “proteína biogênica de ligação a amina”, ou BAP (ANDERSEN, 2003). Ao todo, já foram descritas na literatura 3 proteínas biogênicas de ligação a amina em *R. prolixus* (Tabela 1).

As BAPs são proteínas que se ligam a serotonina, bem como a epinefrina e a norepinefrina, impedindo que ela se ligue ao receptor da 5-HT, o que resulta na inibição da contração da musculatura lisa (ANDERSEN *et al.*, 2005). Quando se pensa na fisiologia alimentar de insetos hematófagos, esse grupo de lipocalinas exerce um importante papel anti-hemostático, uma vez que serotonina e epinefrina, quando secretadas pelas plaquetas, contribuem significativamente para a ativação e agregação das mesmas. Ao se ligarem a esses agonistas, portanto, as BAPs neutralizam os seus efeitos na agregação plaquetária, resultando em uma atividade anti-hemostática. Outra interação importante pode ser observada quando ocorre uma lesão no vaso sanguíneo do hospedeiro vertebrado. Nessa ocasião, norepinefrina é liberada pelos nervos simpáticos na parede desse vaso e interage com receptores adrenérgicos, resultando na vasoconstrição. Esse importante mecanismo de defesa contra perda de sangue é, no entanto, bloqueado quando as BAPs capturam a norepinefrina circulante (ANDERSEN *et al.*, 2005).

1.3.3. Inibidoras de agregação plaquetária (PAI)

As lipocalinas inibidoras de agregação plaquetária (PAI) são proteínas de 19 kDa que foram isoladas da glândula salivar de *Rhodnius prolixus*. Foram descritas como sendo as primeiras lipocalinas a atuarem sobre o ADP, uma molécula que ativa vias de sinalização que medeiam a agregação plaquetária. Ao se ligar ao ADP, a PAI impede a ativação de vias de sinalização específicas que levam à expressão funcional do receptor de fibrinogênio nas plaquetas, prejudicando, assim, a agregação plaquetária (FRANCISCHETTI *et al.*, 2000).

Como será visto ao longo do presente trabalho, na saliva dos insetos hemófagos existem outras lipocalinas que inibem a agregação plaquetária, seja ela induzida por colágeno, pela inibição da trombina ou pela liberação de NO. As lipocalinas tratadas como PAI, no entanto, promovem o efeito anti-agregação via ADP. Na literatura já foram descritas, ao todo, oito lipocalinas do tipo PAI em *R. prolixus* (Tabela 1).

1.3.4. Pallidipinas

As pallidipinas são lipocalinas de 19 kDa que foram assim nomeadas por terem sido encontradas, inicialmente, no triatomíneo *Triatoma pallidipennis*. Noeske-Jungblut e colaboradores encontraram essa proteína ao estudarem a saliva dos insetos adultos. Os ensaios realizados nesse trabalho consistiram em incubar a proteína purificada da saliva do inseto com sangue humano (plasma rico em plaquetas e plaquetas lavadas) e monitorar a agregação plaquetária. Foi observado que a proteína, chamada então de pallidipina, foi capaz de inibir a agregação plaquetária induzida por colágeno, mas nenhum efeito foi observado na agregação plaquetária induzida por ADP ou por trombina (NOESKE-JUNGBLUT *et al.*, 1994).

Ao se ligar ao colágeno, a pallidipina impede que ele se ligue às integrinas alfa e beta das plaquetas, impedindo, assim, a agregação plaquetária por essa via. Por isso, as pallidipinas ficaram conhecidas por inibirem a agregação plaquetária induzida, exclusivamente, pelo colágeno (NOESKE-JUNGBLUT *et al.*, 1994). Até o presente trabalho, duas Pallidipinas já haviam sido descritas em *R. prolixus* (Tabela 1).

1.3.5. Procalinas

As procalinas foram identificadas pela primeira vez por Paddock e colaboradores (2001) em *Triatoma protracta*. Inicialmente, ela foi identificada apenas como uma lipocalina salivar, com 18 a 20 kDa, responsável por fortes reações alérgicas em humanos picados por *T. protracta*. Apesar de os autores não terem realizado estudos funcionais da procalina identificada, eles conseguiram obter a sequência de aminoácidos e as análises de homologia sugeriram alta similaridade com as triabinas. Em *R. prolixus*, duas sequências ortólogas as procalinas foram identificadas no transcriptoma de glândula salivar (RIBEIRO *et al.*, 2004), recebendo os nomes de Procalina-1 e Procalina-2. Desde então, nenhum estudo foi realizado sobre a função das procalinas, ficando elas classificadas como possíveis inibidoras de agregação plaquetária.

1.3.6. Triabinas

Inicialmente, as triabinas foram classificadas como pertencentes à família das lipocalinas por apresentarem uma estrutura tridimensional em β -barril muito semelhante (FUENTES-PRIOR *et al.*, 1997). Mais tarde, elas foram reclassificadas como uma família distinta das lipocalinas, devido a uma diferença na topografia da estrutura de β -

barril. Com isso, elas foram classificadas como uma nova família dentro da superfamília das calcinas, onde se encontra também a família das lipocalinas (FLOWER *et al.*, 2000). Dessa forma, muitos autores consideram as triabinas uma família a parte (HERNANDEZ-VARGAS *et al.*, 2016).

No entanto, as triabinas exibem uma conservação de sequência nas regiões estruturalmente conservadas das lipocalinas, além de uma forte homologia com os fatores antiplaquetários das pallidipinas e PAIs (FUENTES-PRIOR *et al.*, 1997; FLOWER, 1996). Por isso, optou-se por incluir as triabinas nesse estudo considerando-as proteínas anti-hemostáticas derivadas das lipocalinas.

Triabinas foram descritas a partir da saliva do *Triatoma pallidipennis* como potentes inibidoras de trombina, a principal promotora da coagulação sanguínea (NOESKE-JUNGBLUT *et al.*, 1995). Neste estudo, a saliva do inseto adulto foi extraída e utilizada em ensaios de tempo de coagulação da trombina, agregação das plaquetas e atividade de coagulação do fibrinogênio. Os resultados dos ensaios mostraram que a saliva de *T. pallidipennis* inibiu a agregação plaquetária induzida pelo colágeno e também a agregação mediada pela trombina. Ensaios posteriores isolaram a proteína inibidora de agregação plaquetária induzida por trombina, agora chamado de triabina, e mostraram que a triabina inibia não apenas a agregação plaquetária, mas também interferia em outras atividades conhecidas da trombina. Ao mesmo tempo, não foi observada nenhuma atividade anti-agregação plaquetária da triabina induzida por colágeno ou por ADP (NOESKE-JUNGBLUT *et al.*, 1995).

Mais tarde foi demonstrado que, ao se ligar à trombina, a triabina compromete a atividade de conversão do fibrinogênio e a capacidade de ativação da proteína C mediada por trombomodulina. Com isso, a agregação plaquetária induzida por trombina é inibida e o tempo para coagulação por trombina é prolongado (FUENTES-PRIOR *et al.*, 1997). A ligação da triabina não se dá no sítio ativo da trombina, o que evidenciou um novo mecanismo de inibição de trombina que pode ser relevante para potenciais aplicações terapêuticas (FUENTES-PRIOR *et al.*, 1997). Em *R. prolixus* já foram descritas 4 triabinas (Tabela 1).

1.3.7. Apolipoproteína D

A apolipoproteína D (ApoD) é uma proteína plasmática pertencente à família das lipocalinas. Nenhuma apolipoproteína foi descrita em *R. prolixus* até o presente estudo, mas como será visto nos resultados desta revisão, é importante fazer uma breve

introdução sobre estas lipocalinas. Elas estão presentes na composição das lipoproteínas de alta densidade (HDL) e estão envolvidas em diversos processos fisiológicos. O HDL, além de ser um importante marcador de proteção para doenças cardiovasculares, é conhecido por inibir a agregação plaquetária (LIMA & COUTO, 2006). Por esse motivo, levantou-se a hipótese de a ApoD também ser capaz de se ligar a prostaglandinas pró-agregadoras e, assim, modular a função plaquetária (CABRAL *et al.* 1995; MA *et al.*, 2012).

No entanto, estudos *in vitro* realizados por Ma e colaboradores (2012) refutaram essa hipótese. A ApoD foi expressa em células HEK293, purificada em coluna Ni-NTA seguida por cromatografia de fase reversa e utilizada em experimentos de calorimetria de titulação isotérmica (ITC). Os resultados não detectaram a interação de ApoD com prostaglandina F2alfa (PGF2 α), tromboxano A2 (TXA2) ou ácido araquidônico (AA), importantes fatores hemostáticos do hospedeiro vertebrado. Além disso, não foi observada inibição da agregação plaquetária desencadeada por agonistas em baixas doses como colágeno, U46619 (tromboxano A2 mimético) ou AA (MA *et al.*, 2012).

Tais resultados divergem dos obtidos em estudos anteriores, onde, utilizando outra metodologia experimental, foi possível observar a ligação de ApoD humana ao ácido araquidônico (AA) (CABRAL *et al.* 1995). Essa não congruência de resultados evidencia a necessidade de mais estudos sobre o papel das apolipoproteínas D no controle da hemostasia.

1.4. As lipocalinas esquecidas do transcriptoma de *Rhodnius prolixus*

Em 2014 foi publicado o sequenciamento do RNA de diversos tecidos de *Rhodnius prolixus*, incluindo intestino anterior, médio e posterior, ovários, testículos, túbulos de malpighi, corpo gorduroso, reto e corpo inteiro (que incluiu os ovos e os diferentes estádios ninfais) (RIBEIRO *et al.*, 2014). As leituras obtidas foram montadas em 25.673 contigs. Mais de 2.900 sequências codificantes (CDS) completas foram obtidas (da metionina inicial até o stop códon) (RIBEIRO *et al.*, 2014).

Apesar de o transcriptoma ter incluído diferentes órgãos e estágios de desenvolvimento dos insetos, os dados foram analisados sob a perspectiva da fisiologia do intestino dos insetos, buscando entender aspectos importantes da digestão de sangue. Assim, as lipocalinas (proteínas primariamente salivares) não foram analisadas no estudo de Ribeiro e colaboradores de 2014. Este trabalho identificou 112 transcritos como possíveis proteínas da família das lipocalinas, sendo muitos deles expressos nos

mais diferentes tecidos do inseto, como intestino, testículos e ovários. A observação de lipocalinas expressas em tecidos diferentes da glândula salivar, por si só, é algo inédito para *R. prolixus*.

1.5. Justificativa

Tendo em vista a abundância e diversidade de lipocalinas encontradas em *R. prolixus*, o pouco conhecimento sobre a atuação delas para além da função alimentar abre um amplo e interessante campo de estudo. Compreender a atuação das lipocalinas em processos fisiológicos até então ignorados pode ser útil para entender a biologia da espécie *R. prolixus* e de outras relacionadas, bem como desenvolver novas ferramentas para o controle da doença de Chagas e de outras doenças transmitidas por insetos hematófagos. Assim, esta dissertação tem como objetivo fazer uma revisão das lipocalinas de *R. prolixus* a partir dos resultados brutos obtidos no transcriptoma de 2014, buscando saber se existem no transcriptoma lipocalinas novas (ainda não descritas e caracterizadas), caracterizando o perfil das diferentes lipocalinas nos diferentes tecidos de *R. prolixus*, e analisando evolutivamente o possível papel de cada lipocalina encontrada.

2. OBJETIVOS

2.1. Objetivo geral

- I. Identificar e caracterizar, estrutural e evolutivamente, novas lipocalinas expressas em diferentes tecidos do inseto *Rhodnius prolixus*.

2.2. Objetivos específicos

- I. Identificar e anotar estruturalmente (*in silico*) as lipocalinas expressas em *R. prolixus*.
- II. Determinar as relações evolutivas e de ortologia entre as lipocalinas expressas em *R. prolixus*.
- III. Identificar a existência de isoformas e duplicações gênicas das lipocalinas de *R. prolixus*.
- IV. Predizer a função putativa das lipocalinas ainda não descritas, de acordo com ortologia e expressão.

3. MATERIAIS E MÉTODOS

3.1. O transcriptoma de *Rhodnius prolixus*

Para o presente estudo, foram utilizadas as sequências do transcriptoma depositadas no VectorBase (<https://vectorbase.org/vectorbase/app/>) baixado em fevereiro de 2019. Os dados de expressão e de função dos genes foram adquiridos de material suplementar do artigo do transcriptoma de *R. prolixus* (RIBEIRO *et al.*, 2014). A identificação inicial de Lipocalinas no material suplementar foi realizada utilizando-se palavras chaves como busca, sendo elas: lipocalin, nitrophorin, platelet aggregation inhibitor, PAI, triabin, pallidipin.

3.2. Anotação estrutural das lipocalinas expressas em *R. prolixus*

Para a anotação estrutural das lipocalinas, foram usadas como modelo as lipocalinas já sequenciadas e caracterizadas em *R. prolixus* (Tabela 1). As sequências genômicas foram primeiro identificadas por homologia usando a ferramenta BLASTn (ALTSCHUL *et al.*, 1990) (parâmetros Word 26, e-value 0.00001) tendo o transcrito como modelo (*query*) e o genoma (versão 3.1) como alvo. Os resultados da sequência *query* foram filtrados de acordo com a cobertura (>90%) e identidade de nucleotídeos (>98%). Para verificar se já existiam proteínas anotadas para as sequências do transcriptoma, realizou-se uma busca com a ferramenta BLASTp, utilizando a sequência de aminoácidos anotada para os transcritos como modelo e a versão 3.1.1 do proteoma de *R. prolixus* como alvo (parâmetros Word 3, e-value 0.00001, identidade de aminoácidos >98%). Todos os BLASTs foram realizados diretamente no site VectorBase (www.vectorbase.org). Nos casos em que uma proteína do proteoma tenha sido identificada como pertencente a um transcrito, a sequência anotada (RPRCXXXX) passou a ser utilizada como a sequência de referência. Nos outros casos, a nomenclatura do transcriptoma (rp_asb-XXXX) foi mantida.

Nos casos em que a região genômica do transcrito foi identificada, mas não havia proteína anotada, realizou-se a anotação manual da proteína com a ferramenta GeneWise2 (www.ebi.ac.uk) (JAREBORG *et al.*, 1999), sendo utilizada como modelo a proteína mais próxima identificada nas análises evolutivas. A região genômica foi identificada por BLAST e a sequência foi obtida contendo 5000 nucleotídeos a mais nas pontas 5' e 3', usando a informação do BLAST para corretamente obter a sequência original ou seu complemento reverso (nos casos em que foi identificado que a proteína

estava codificada na fita oposta). Os parâmetros do GeneWise foram mantidos como o padrão com exceção ao parâmetro *splice-site* em que foi escolhido “modeled”. Os domínios e sítios ativos das proteínas anotadas foram identificados por análise no Conserved Domain Database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Os resultados da anotação manual são mostrados no ANEXO G.

3.3. Análise evolutiva das lipocalinas

A fim de prever a relação evolutiva entre as lipocalinas expressas em *R. prolixus*, as sequências de aminoácidos de todas as lipocalinas já descritas (em *R. prolixus* e outros triatomíneos) foram obtidas no GenBank (www.ncbi.nlm.nih.gov) e utilizadas, junto às sequências daquelas identificadas no transcriptoma, para análise das relações evolutivas. As proteínas foram alinhadas com Muscle (EDGAR, 2004). Utilizou-se o programa MEGAX (KUMAR *et al.*, 2018) que inferiu a história evolutiva entre todas as sequências de aminoácidos usando o método Neighbor-Joining (Bootstrap com 10000 réplicas). As análises foram realizadas com deleção completa ou deleção par a par dos sítios ambíguos dependendo do grupo de proteínas a serem estudadas e a escolha do tratamento das posições sem informação foi especificado na legenda das respectivas figuras. Em todos os casos também foi testada a relação evolutiva das sequências por Máximo Verosimilhança e por Máximo Parsimônia (10000 réplicas) buscando comparar com a análise por Neighbor-Joining.

A identificação de resíduos conservados nas proteínas foi realizada por alinhamento das proteínas pelo Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (SIEVERS *et al.*, 2011) e consequente consulta ao Conserved Domains Database.

3.4. Identificação de isoformas e duplicações gênicas

Com o auxílio da ferramenta tBLASTn (e-value = 0.01) as sequências peptídicas dos transcritos derivados do mesmo gene e/ou da mesma região gênica foram alinhadas contra o genoma de *R. prolixus* (versão 3.1). As coordenadas dos transcritos, ou seja, suas posições no genoma (considerando seus nucleotídeos iniciais e terminais e regiões de *introns* e *exons*) foram comparadas a fim de se identificar a existência de isoformas originadas por *splicing* alternativo ou duplicações gênicas desses transcritos. Todos os BLASTs foram realizados diretamente no site VectorBase (www.vectorbase.org).

4. RESULTADOS

4.1. Mapeamento das lipocalinas conhecidas ao genoma e proteoma de *R. prolixus*

Como primeira etapa desta revisão, realizou-se o mapeamento das lipocalinas já conhecidas de *R. prolixus* (ver Tabela 1) ao genoma e proteoma do inseto, já que este trabalho ainda não havia sido realizado (Tabela 2). Para isso, as sequências depositadas no GeneBank foram utilizadas para buscar por proteínas no proteoma de *R. prolixus* depositado no VectorBase. Utilizou-se a ferramenta BLASTp juntamente com o BLAST recíproco, a fim de confirmar a relação de homologia entre as sequências. Assim, foi possível mapear tanto ao genoma quanto ao proteoma 13 das 24 lipocalinas já descritas. São elas: a NP-4 Q94734.1, que apresentou maior correspondência ao gene RPRC000023; a NP-2 Q26241.1, que correspondeu ao gene RPRC000072; a PAI-1 AAB09090.1, que correspondeu ao gene RPRC002102; a PAI-2 AAB09091.1, com correspondência ao gene RPRC000175; a PAI-5 AAQ20819.1, que correspondeu ao gene RPRC002584; a PAI-6 AAQ20820.1, que correspondeu ao gene RPRC000677; a PAI-7 AAQ20838.1, com maior correspondência ao gene RPRC000463; a Pallidipina-1 AAQ20825.1, com correspondência ao gene RPRC000012; a Procalina-1 AAQ20827.1 com correspondência ao gene RPRC010444; a Triabina-4 AAQ20824.1, que correspondeu ao gene RPRC002583; a Triabina-3 AAQ20823.1, com correspondência ao gene RPRC000147; a Triabina-2 AAQ20822.1, que apresentou correspondência ao gene RPRC012954; e por fim a Triabina-1 AAQ20821.1, com correspondência ao gene RPRC012954.

Das lipocalinas já descritas, 9 foram mapeadas apenas ao genoma, não havendo proteínas correspondentes anotadas no proteoma do inseto. A NP-1 Q26239.1 foi anotada no scaffold ACPB03038429 que apresenta um gap, o que pode justificar a ausência de proteínas anotadas nessa região. A NP-7 Q6PQK2.1 está anotada em uma região gênica que foi quebrada em dois scaffolds (ACPB03033880 e KQ034096), por isso a proteína não foi anotada no proteoma. Já a PAI-3 AAQ20817.1 está anotada no scaffold KQ035719. Neste scaffold encontra-se o gene RPRC006975 que pode ter sido anotado errado, o que justificaria a não correspondência com a PAI-3. As lipocalinas PAI-4 e PAI-4B estão anotadas em uma mesma região gênica formada por dois scaffolds que foram quebrados (KQ036102 e ACPB03037847), por isso não há proteína anotada nessa região. A Pallidipina-2 AAQ20826.1 foi anotada no scaffold KQ034748, onde nenhum gene foi encontrado. Por fim, as BAP-1 AAO25746.1, BAP-2 4GET_A e

BAP-3 4HFO_A foram anotadas na mesma região gênica (scaffold KQ034229), onde há apenas o gene RPRC000114 anotado, que não apresentou correspondência às BAPs.

Duas lipocalinas já descritas não foram mapeadas nem no genoma nem no proteoma do inseto. São elas a NP-3 Q94733.1 e a Procalina-2 AAQ20828.1. Para essas duas lipocalinas não foram encontradas uma região gênica ou uma proteína correspondente. Os dados brutos do mapeamento de todas as lipocalinas já descritas ao genoma e ao proteoma do inseto são mostrados, respectivamente, nos ANEXOS A e B.

Tabela 2. Mapeamento das lipocalinas depositadas no GeneBank ao genoma e proteoma de *R. prolixus*.

| Lipocalina | Nº de acesso | Melhor correspondência ao proteoma | ID% | Cobertura | Melhor correspondência ao genoma | ID% | Cobertura |
|-----------------------------|--------------|------------------------------------|-------|-----------|----------------------------------|-----------|-----------|
| Nitroforina 1 (NP-1) | Q26239.1 | Sem correspondência | | | ACPB03038429 | 97,5% | 38% |
| Nitroforina 3 (NP-3) | Q94733.1 | Sem correspondência | | | Sem correspondência | | |
| Nitroforina 4 (NP-4) | Q94734.1 | RPRC000023 | 99,5% | 100% | ACPB03031753 | 48 a 100% | 100% |
| Nitroforina 2 (NP-2) | Q26241.1 | RPRC000072 | 98,5% | 100% | KQ035484 | 59 a 97% | 97% |
| Nitroforina 7 (NP-7) | Q6PQK2.1 | Sem correspondência | | | ACPB03033880 e KQ034096 | 74 a 100% | 100% |
| BAP-1 | AAO25746.1 | Sem correspondência | | | KQ034229 | 66 a 92% | 98% |
| BAP-2 | 4GET_A | Sem correspondência | | | KQ034229 | 66 a 92% | 100% |
| BAP-3 | 4HFO_A | Sem correspondência | | | KQ034229 | 66 a 92% | 100% |
| PAI-1 | AAB09090.1 | RPRC002102 | 85,2% | 98% | KQ034920 | 91 a 100% | 97% |
| PAI-2 | AAB09091.1 | RPRC000175 | 77,5% | 100% | KQ036605 | 57 a 100% | 100% |
| PAI-3 | AAQ20817.1 | Sem correspondência | | | KQ035719 | 85 a 100% | 100% |
| PAI-4 | AAQ20818.1 | Sem correspondência | | | KQ036102 e ACPB03037847 | 97 a 100% | 59,0% |
| PAI-4B | 5H9L_A | Sem correspondência | | | KQ036102 e ACPB03037847 | 97 a 100% | 59,0% |
| PAI-5 | AAQ20819.1 | RPRC002584 | 94,7% | 95% | ACPB03035352 | 94 a 100% | 100% |
| PAI-6 | AAQ20820.1 | RPRC000677 | 92,8% | 100% | KQ035263 | 75 a 100% | 100% |
| PAI-7 | AAQ20838.1 | RPRC000463 | 91,0% | 95,8% | KQ037461 | 83 a 95% | 100% |
| Pallidipina-1 | AAQ20825.1 | RPRC000012 | 85,5% | 100% | KQ034748 | 67 a 96% | 100% |
| Pallidipina-2 | AAQ20826.1 | Sem correspondência | | | KQ034748 | 58 a 100% | 90% |
| Procalina-1 | AAQ20827.1 | RPRC010444 | 98,3% | 100% | KQ034856 | 100% | 100% |
| Procalina-2 | AAQ20828.1 | Sem correspondência | | | Sem correspondência | | |
| Triabina-1 | AAQ20821.1 | RPRC006167 | 94,5% | 93,0% | ACPB03037937 | 91% | 76% |
| Triabina-2 | AAQ20822.1 | RPRC012954 | 92,9% | 94,0% | KQ035237 | 51 a 96% | 90% |
| Triabina-3 | AAQ20823.1 | RPRC000147 | 77,0% | 99,0% | KQ035329 | 60 a 80% | 98% |
| Triabina-4 | AAQ20824.1 | RPRC002583 | 100% | 100% | KQ034856 | 95 a 100% | 100% |

4.2. Identificação de Lipocalinas no transcriptoma de *R. prolixus*

O transcriptoma de *R. prolixus* gerou um total de 25.673 transcritos. A função hipotética dos transcritos foi determinada por similaridade a proteínas anotadas no GeneBank (Non-redundant), Pfam (Protein families), CDD (conserved domain databases), Uniprot e SwisProt. Foi identificado um total de 112 transcritos com similaridades a lipocalinas, sendo que foi possível classificá-los inicialmente como: 41 nitroforinas (Tabelas 3), 22 PAI (Tabela 4), 12 pallidipinas (Tabela 5), 24 triabinas (Tabela 6) e 11 apolipoproteínas (Tabela 7). Além disso, duas lipocalinas foram classificadas apenas como lipocalinas salivares (Tabela 8). Nenhuma lipocalina foi classificada inicialmente como parte do grupo das procalinas. As tabelas de 1 a 8 também mostram em qual base de dados foi encontrada a melhor homologia (*best-match*) pelo BLAST. Estas tabelas contêm apenas parte da informação das buscas por blast e as tabelas completas podem ser obtidas no material suplementar do artigo de Ribeiro e colaboradores (2014).

Tabela 3. Transcritos de *R. prolixus* similares a lipocalinas da família das Nitroforinas.

| Contig | Tamanho pb | Tamanho aa | Similaridade | Base de Dados | E-value | Cobertura* |
|---------------|------------|------------|----------------|---------------|---------|------------|
| rp_asb-48127 | 955 | 121 | Nitrophorin 1A | NR-LIGHT | 0 | 0.56 |
| rp_asb-48695 | 858 | 122 | Nitrophorin-4 | NR-LIGHT | 0 | 0.47 |
| rp_asb-3238 | 904 | 116 | Nitrophorin-3 | SWISSP | 8E-71 | 0.58 |
| rp_asb-48123 | 633 | 115 | Nitrophorin 1A | NR-LIGHT | 2E-66 | 0.53 |
| rp_asb-3237 | 858 | 106 | Nitrophorin-2 | NR-LIGHT | 2E-47 | 0.90 |
| rp_asb-101622 | 386 | 97 | Nitrophorin 3B | NR-LIGHT | 3E-33 | 0.60 |
| rp_asb-25981 | 1160 | 215 | Nitrophorin | PFAM | 1E-33 | 1.02 |
| rp_asb-74536 | 830 | 217 | Nitrophorin-2 | SWISSP | 1E-12 | 1.04 |
| rp_asb-3236 | 920 | 117 | Nitrophorin-2 | NR-LIGHT | 0 | 0.58 |
| rp_asb-104982 | 885 | 173 | Nitrophorin-7 | NR-LIGHT | 0 | 0.84 |
| rp_asb-48696 | 721 | 124 | Nitrophorin-4 | SWISSP | 3E-97 | 0.55 |
| rp_asb-90144 | 1171 | 117 | Nitrophorin | PFAM | 2E-09 | 0.93 |
| rp_asb-48697 | 774 | 128 | Nitrophorin-1 | NR-LIGHT | 0 | 0.55 |
| rp_asb-25624 | 831 | 211 | Nitrophorin | PFAM | 2E-33 | 1.02 |
| rp_asb-48698 | 858 | 221 | Nitrophorin 4A | NR-LIGHT | 0 | 1.00 |
| rp_asb-1685 | 1006 | 105 | Nitrophorin-2 | SWISSP | 2E-58 | 0.50 |
| rp_asb-48694 | 899 | 202 | Nitrophorin-1 | NR-LIGHT | 0 | 0.91 |
| rp_asb-85899 | 824 | 150 | Nitrophorin 4B | NR-LIGHT | 0 | 0.69 |
| rp_asb-89581 | 815 | 195 | Nitrophorin 3B | NR-LIGHT | 3E-62 | 0.99 |
| rp_asb-48699 | 896 | 98 | Nitrophorin | PFAM | 2E-42 | 0.51 |
| rp_asb-105554 | 538 | 107 | Nitrophorin 4B | NR-LIGHT | 2E-30 | 0.28 |
| rp_asb-104455 | 542 | 96 | Nitrophorin 1A | NR-LIGHT | 4E-15 | 0.49 |
| rp_asb-25025 | 756 | 168 | Nitrophorin | PFAM | 2E-31 | 0.97 |
| rp_asb-3232 | 1212 | 138 | Nitrophorin-3 | NR-LIGHT | 0 | 0.59 |
| rp_asb-48126 | 818 | 154 | Nitrophorin 1A | NR-LIGHT | 1E-65 | 0.69 |
| rp_asb-48125 | 822 | 219 | Nitrophorin 1A | NR-LIGHT | 1E-97 | 1.00 |
| rp_asb-85900 | 859 | 193 | Nitrophorin 4B | NR-LIGHT | 5E-84 | 0.88 |
| rp_asb-48128 | 873 | 169 | Nitrophorin 1A | NR-LIGHT | 6E-53 | 0.61 |
| rp_asb-48124 | 807 | 123 | Nitrophorin 1A | NR-LIGHT | 1E-64 | 0.81 |
| rp_asb-16225 | 489 | 98 | Nitrophorin 1A | NR-LIGHT | 2E-36 | 0.46 |
| rp_asb-48741 | 310 | 77 | Nitrophorin 1A | NR-LIGHT | 7E-24 | 0.33 |
| rp_asb-104644 | 686 | 91 | Nitrophorin-7 | GO | 3E-25 | 0.28 |
| rp_asb-3239 | 850 | 87 | Nitrophorin-2 | SWISSP | 3E-68 | 0.62 |
| rp_asb-89584 | 670 | 191 | Nitrophorin 3B | NR-LIGHT | 6E-77 | 0.95 |
| rp_asb-3235 | 885 | 110 | Nitrophorin-2 | SWISSP | 7E-69 | 0.66 |
| rp_asb-101226 | 625 | 112 | Nitrophorin 4B | NR-LIGHT | 3E-41 | 0.48 |
| rp_asb-101256 | 934 | 150 | Nitrophorin-2 | SWISSP | 2E-71 | 1.00 |
| rp_asb-89583 | 731 | 187 | Nitrophorin 3B | NR-LIGHT | 3E-65 | 0.96 |
| rp_asb-89582 | 698 | 97 | Nitrophorin 3B | NR-LIGHT | 2E-56 | 0.64 |
| rp_asb-104643 | 751 | 123 | Nitrophorin-7 | NR-LIGHT | 1E-60 | 0.98 |
| rp_asb-105049 | 432 | 58 | Nitrophorin | PFAM | 9E-21 | 0.56 |

* Cobertura do alinhamento do transcrito com a proteína anotada na referida base de dados

Tabela 4. Transcritos de *R. prolixus* similares a lipocalinas da família das Inibidoras de Agregação Plaquetária (PAI).

| Contig | Tamanho pb | Tamanho aa | Similaridade | Base de Dados | E-value | Cobertura* |
|---------------|------------|------------|----------------|---------------|---------|------------|
| rp_asb-34424 | 1619 | 177 | lipocalin AI-7 | NR-LIGHT | 1E-63 | 0.58 |
| rp_asb-115859 | 503 | 70 | lipocalin AI-6 | NR-LIGHT | 1E-36 | 0.34 |
| rp_asb-15376 | 774 | 175 | lipocalin AI-5 | NR-LIGHT | 9E-85 | 0.93 |
| rp_asb-15398 | 738 | 168 | lipocalin AI-5 | NR-LIGHT | 2E-65 | 0.71 |
| rp_asb-97660 | 637 | 85 | lipocalin AI-6 | NR-LIGHT | 3E-11 | 0.46 |
| rp_asb-15397 | 624 | 100 | lipocalin AI-5 | NR-LIGHT | 6E-74 | 0.39 |
| rp_asb-34427 | 450 | 149 | lipocalin AI-7 | NR-LIGHT | 5E-73 | 0.92 |
| rp_asb-97659 | 664 | 103 | lipocalin AI-6 | NR-LIGHT | 1E-13 | 0.28 |
| rp_asb-34426 | 661 | 159 | lipocalin AI-7 | NR-LIGHT | 3E-71 | 0.96 |
| rp_asb-15399 | 522 | 103 | lipocalin AI-5 | NR-LIGHT | 1E-30 | 0.67 |
| rp_asb-34429 | 614 | 161 | lipocalin AI-7 | NR-LIGHT | 1E-73 | 1.00 |
| rp_asb-15404 | 320 | 61 | lipocalin AI-5 | NR-LIGHT | 3E-20 | 0.44 |
| rp_asb-34425 | 612 | 171 | lipocalin AI-7 | NR-LIGHT | 2E-72 | 0.93 |
| rp_asb-105584 | 654 | 102 | lipocalin AI-4 | NR-LIGHT | 3E-15 | 0.42 |
| rp_asb-105551 | 671 | 105 | lipocalin AI-6 | NR-LIGHT | 6E-40 | 0.90 |
| rp_asb-34428 | 657 | 195 | lipocalin AI-7 | NR-LIGHT | 5E-70 | 0.95 |
| rp_asb-15400 | 665 | 130 | lipocalin AI-5 | NR-LIGHT | 8E-68 | 0.56 |
| rp_asb-15401 | 649 | 120 | lipocalin AI-5 | NR-LIGHT | 8E-55 | 0.60 |
| rp_asb-44356 | 725 | 193 | lipocalin AI-6 | NR-LIGHT | 7E-48 | 1.03 |
| rp_asb-44609 | 705 | 212 | lipocalin AI-6 | NR-LIGHT | 2E-45 | 0.98 |
| rp_asb-34430 | 506 | 149 | lipocalin AI-7 | NR-LIGHT | 2E-48 | 0.90 |
| rp_asb-104785 | 526 | 130 | lipocalin AI-3 | NR-LIGHT | 7E-75 | 0.75 |

* Cobertura do alinhamento do transcrito com a proteína anotada na referida base de dados

Tabela 5. Transcritos de *R. prolixus* similares a lipocalinas da família das Pallidipinas.

| Contig | Tamanho pb | Tamanho aa | Similaridade | Base de Dados | E-value | Cobertura* |
|---------------|------------|------------|-----------------|---------------|---------|------------|
| rp_asb-73730 | 470 | 73 | pallidipin-like | NR-LIGHT | 8E-10 | 0.17 |
| rp_asb-109013 | 408 | 60 | pallidipin-like | NR-LIGHT | 1E-09 | 0.24 |
| rp_asb-73727 | 619 | 122 | pallidipin-like | NR-LIGHT | 1E-69 | 0.55 |
| rp_asb-109014 | 365 | 117 | pallidipin-like | NR-LIGHT | 1E-27 | 0.45 |
| rp_asb-109015 | 427 | 51 | pallidipin-like | NR-LIGHT | 2E-16 | 0.22 |
| rp_asb-73728 | 811 | 188 | pallidipin-like | NR-LIGHT | 7E-82 | 0.94 |
| rp_asb-73726 | 967 | 206 | pallidipin-like | NR-LIGHT | 0.0 | 1.00 |
| rp_asb-104042 | 1146 | 187 | pallidipin-like | NR-LIGHT | 0.0 | 0.79 |
| rp_asb-73729 | 484 | 107 | pallidipin-like | NR-LIGHT | 9E-42 | 0.54 |
| rp_asb-112121 | 454 | 149 | pallidipin-like | NR-LIGHT | 3E-43 | 0.65 |
| rp_asb-72343 | 453 | 120 | pallidipin-like | NR-LIGHT | 9E-19 | 0.45 |
| rp_asb-73858 | 1215 | 254 | pallidipin-like | NR-LIGHT | 2E-54 | 0.93 |

* Cobertura do alinhamento do transcrito com a proteína anotada na referida base de dados

Tabela 6. Transcritos de *R. prolixus* similares a lipocalinas da família das Triabinas.

| Contig | Tamanho pb | Tamanho aa | Similaridade | Base de Dados | E-value | Cobertura* |
|---------------|------------|------------|--------------|---------------|---------|------------|
| rp_asb-37664 | 695 | 205 | Triabin | PFAM | 1E-15 | 1.01 |
| rp_asb-89585 | 663 | 145 | triabin-like | NR-LIGHT | 5E-74 | 0.63 |
| rp_asb-73238 | 945 | 204 | triabin-like | NR-LIGHT | 4E-23 | 1.02 |
| rp_asb-30197 | 1565 | 125 | triabin-like | NR-LIGHT | 8E-21 | 0.55 |
| rp_asb-102751 | 659 | 177 | triabin-like | NR-LIGHT | 5E-75 | 0.92 |
| rp_asb-105585 | 658 | 187 | triabin-like | PFAM | 8E-21 | 0.99 |
| rp_asb-102621 | 645 | 182 | triabin-like | NR-LIGHT | 7E-76 | 1.00 |
| rp_asb-33412 | 872 | 208 | triabin-like | NR-LIGHT | 2E-21 | 0.95 |
| rp_asb-30198 | 1730 | 201 | triabin-like | NR-LIGHT | 8E-25 | 0.99 |
| rp_asb-102716 | 435 | 143 | triabin-like | NR-LIGHT | 4E-74 | 0.76 |
| rp_asb-30450 | 668 | 134 | triabin-like | NR-LIGHT | 0.00006 | 0.59 |
| rp_asb-65689 | 674 | 203 | triabin-like | PFAM | 9E-18 | 1.00 |
| rp_asb-102620 | 359 | 109 | triabin-like | NR-LIGHT | 7E-53 | 0.59 |
| rp_asb-33413 | 856 | 160 | triabin-like | NR-LIGHT | 1E-19 | 0.70 |
| rp_asb-40554 | 956 | 182 | triabin-like | NR-LIGHT | 1E-61 | 1.00 |
| rp_asb-102753 | 387 | 65 | triabin-like | NR-LIGHT | 5E-21 | 0.36 |
| rp_asb-40553 | 488 | 104 | triabin-like | NR-LIGHT | 1E-54 | 0.58 |
| rp_asb-110283 | 628 | 77 | triabin-like | NR-LIGHT | 4E-32 | 0.50 |
| rp_asb-102749 | 719 | 170 | triabin-like | NR-LIGHT | 2E-75 | 0.90 |
| rp_asb-40550 | 459 | 106 | triabin-like | NR-LIGHT | 6E-33 | 0.59 |
| rp_asb-105577 | 439 | 94 | triabin-like | NR-LIGHT | 4E-22 | 0.49 |
| rp_asb-102752 | 554 | 116 | triabin-like | NR-LIGHT | 7E-38 | 0.85 |
| rp_asb-102750 | 668 | 116 | triabin-like | NR-LIGHT | 1E-31 | 0.50 |
| rp_asb-103687 | 275 | 46 | triabin-like | NR-LIGHT | 0.00002 | 0.19 |

* Cobertura do alinhamento do transcrito com a proteína anotada na referida base de dados

Tabela 7. Transcritos de *R. prolixus* similares a lipocalinas da família de Apolipoproteínas.

| Contig | Tamanho pb | Tamanho aa | Similaridade | Base de Dados | E-value | Cobertura* |
|---------------|------------|------------|------------------|---------------|---------|------------|
| rp_asb-6793 | 741 | 226 | Apolipoprotein D | KOG | 3E-09 | 0.80 |
| rp_asb-14972 | 2410 | 178 | Apolipoprotein D | KOG | 3E-20 | 0.73 |
| rp_asb-4751 | 1333 | 212 | Apolipoprotein D | KOG | 0.00002 | 0.96 |
| rp_asb-7470 | 1235 | 143 | Apolipoprotein O | PFAM | 7E-16 | 0.98 |
| rp_asb-1368 | 1745 | 262 | Apolipoprotein D | KOG | 1E-14 | 0.91 |
| rp_asb-48 | 1325 | 217 | Apolipoprotein D | KOG | 0.00002 | 0.92 |
| rp_asb-40377 | 1080 | 283 | Apolipoprotein D | KOG | 2E-20 | 0.87 |
| rp_asb-47 | 1277 | 217 | Apolipoprotein D | KOG | 0.00002 | 0.92 |
| rp_asb-6771 | 1025 | 222 | Apolipoprotein D | KOG | 2E-10 | 0.70 |
| rp_asb-105382 | 402 | 86 | Apolipoprotein D | KOG | 0.00007 | 0.28 |
| rp_asb-7471 | 824 | 164 | Apolipoprotein O | PFAM | 2E-14 | 0.60 |

* Cobertura do alinhamento do transcrito com a proteína anotada na referida base de dados

Tabela 8. Transcritos de *R. prolixus* similares a lipocalinas inespecíficas.

| Contig | Tamanho pb | Tamanho aa | Similaridade | Base de Dados | E-value | Cobertura* |
|---------------|------------|------------|--------------------|---------------|---------|------------|
| rp_asb-105519 | 737 | 125 | salivary lipocalin | NR-LIGHT | 3E-16 | 0.53 |
| rp_asb-61810 | 1058 | 147 | salivary lipocalin | NR-LIGHT | 2E-14 | 0.57 |

* Cobertura do alinhamento do transcrito com a proteína anotada na referida base de dados.

4.3. Mapeamento e anotação das lipocalinas transcritas ao genoma e proteoma predito de *R. prolixus*

Todos os transcritos identificados inicialmente como lipocalinas foram alinhados ao genoma e ao proteoma de *R. prolixus*, com o intuito de identificar a região genômica de origem do transcrito e se alguma proteína já havia sido identificada como produto de expressão do mesmo. Inicialmente, as buscas realizadas gerariam três tipos de resultados: *i*) transcritos que mapeiam ao genoma e ao proteoma; *ii*) transcritos que mapeiam ao genoma, mas não ao proteoma (por falha de anotação do proteoma); e *iii*) transcritos que não mapeiam nem ao genoma e nem ao proteoma (ou por falha de montagem do genoma, ou do transcriptoma, ou ainda possíveis pseudogenes). Os resultados destas buscas estão sumarizados nas Tabelas 9 – 14.

Ao todo 11 transcritos mapearam no genoma, mas não no proteoma. Estes tiveram suas sequências de aminoácidos anotadas pelo GeneWise de acordo com a metodologia descrita anteriormente. Foi observado que 15 transcritos apresentavam diversos stop-codons ao longo de sua sequência. Estes transcritos foram então classificados, inicialmente, como pseudogenes. Além disso, foram observadas duas situações que inicialmente não foram previstas. A primeira foi de múltiplos transcritos que mapearam a uma mesma região gênica ou a uma mesma proteína. Estes 14 transcritos foram categorizados como “transcritos de mesmo locus” e nas tabelas 9-14 foi apontado a qual locus o transcrito mapeia. O segundo caso é mais curioso: transcritos que alinharam especificamente com uma proteína, mas não possuíam similaridade com nenhuma região do genoma. Dez transcritos apresentaram este perfil e os possíveis motivos para essa observação serão discutidos adiante.

A tabela 9 mostra o resultado do mapeamento para as lipocalinas do grupo nitroforina. Destes transcritos, apenas 6 deles (rp_asb-48127, rp_asb-48695, rp_asb-3238, rp_asb-25981, rp_asb-74536 e rp_asb-3236) mapearam ao genoma e ao proteoma; três transcritos (rp_asb-48123, rp_asb-3237 e rp_asb-101622) alinharam a proteínas já anotadas, mas não alinharam a nenhuma região do genoma. Vinte transcritos (rp_asb-48696, rp_asb-48697, rp_asb-48698, rp_asb-104982, rp_asb-48128, rp_asb-104644, rp_asb-104643, rp_asb-104455, rp_asb-85900, rp_asb-101226, rp_asb-90144, rp_asb-89581, rp_asb-89583, rp_asb-89582, rp_asb-1685, rp_asb-3232, rp_asb-3235, rp_asb-101256, rp_asb-48741, rp_asb-105554) mapearam ao genoma, mas não ao proteoma. Destes, 7 transcritos (rp_asb-48697, rp_asb-104982, rp_asb-104455, rp_asb-90144, rp_asb-89581, rp_asb-89582, rp_asb-3235) apresentaram, na anotação

estrutural, diversos stop-codons em suas sequências e, por isso, eles foram classificados como pseudogenes (recebendo o símbolo genético para pseudogene, Ψ). Os transcritos rp_asb-48696 e rp_asb-48698 mapearam no mesmo locus do gene RPRC000023-PA, podendo ser isoformas deste gene. O mesmo ocorre com o transcrito rp_asb-104644, que mapeou no mesmo locus do gene RPRC000380-PA, podendo ser uma isoforma deste gene; com os transcritos rp_asb-104643, rp_asb-1685, rp_asb-3232 e rp_asb-101256, que mapearam no mesmo do locus do gene RPRC000072-PA, podendo representar isoformas deste gene; e com o transcrito rp_asb-89583, que mapeou no mesmo locus do gene RPRC002583-PA, podendo ser uma isoforma deste. A anotação estrutural evidenciou ainda que os transcritos rp_asb-48128 e rp_asb-48741 possuem estruturas gênicas similares ao do gene RPRC000114, além de estarem localizados próximos no scaffold, o que indica que estes transcritos podem ser originados de cópias do gene RPRC000114. O mesmo foi observado para os transcritos rp_asb-85900, rp_asb-101226 e rp_asb-105554, que aparentam serem cópias do gene RPRC010481.

Tabela 9. Mapeamento das lipocalinas inicialmente identificadas como Nitroforinas.

| Transcrito | Alinhamento ao genoma | | Alinhamento ao proteoma | | Comentário |
|----------------------|-----------------------|--------|-------------------------|--------|------------------|
| | Scaffold | ID (%) | Proteína | ID (%) | |
| rp_asb-48127 | KQ034229 | 100 | RPRC000114-PA | 100 | é RPRC000114 |
| rp_asb-48695 | ACPB03031753 | 100 | RPRC000023-PA | 100 | é RPRC000023 |
| rp_asb-3238 | KQ035220 | 99,45 | RPRC000380-PA | 100 | é RPRC000380 |
| rp_asb-25981 | KQ035220 | 99,72 | RPRC000077-PA | 99,48 | é RPRC000077 |
| rp_asb-74536 | KQ034712 | 98,76 | RPRC000163-PA | 99,43 | é RPRC000163 |
| rp_asb-3236 | KQ035220 | 100 | RPRC000072-PA | 98,29 | é RPRC000072 |
| rp_asb-48123 | KQ034229 | 95,58 | RPRC000114-PA | 100 | é RPRC000114 |
| rp_asb-3237 | KQ034096 | 94,32 | RPRC000367-PA | 100 | é RPRC000367 |
| rp_asb-101622 | KQ034856 | 93,33 | RPRC010566-PA | 100 | é RPRC010566 |
| rp_asb-48696 | ACPB03031753 | 100 | RPRC000023-PA | 96,33 | locus RPRC000023 |
| rp_asb-48697 Ψ | ACPB03031753 | 100 | RPRC004617-PA | 93,33 | Pseudogene |
| rp_asb-48698 | ACPB03031753 | 98,31 | RPRC000023-PA | 91,33 | locus RPRC000023 |
| rp_asb-104982 Ψ | ACPB03033880 | 100 | RPRC000153-PA | 96,69 | Pseudogene |
| rp_asb-48128 | ACPB03047519 | 100 | RPRC000114-PA | 78,79 | Novo |
| rp_asb-104644 | KQ034096 | 100 | RPRC000380-PA | 75 | locus RPRC000380 |
| rp_asb-104643 | KQ034096 | 100 | RPRC000072-PA | 59,7 | locus RPRC000072 |
| rp_asb-104455 Ψ | KQ034229 | 100 | RPRC007298-PA | 84,26 | Pseudogene |
| rp_asb-85900 | KQ034229 | 100 | RPRC010481-PA | 79,31 | Novo |
| rp_asb-101226 | KQ034229 | 100 | RPRC010481-PA | 70 | Novo |
| rp_asb-90144 Ψ | KQ034748 | 100 | RPRC000250-PA | 94,9 | Pseudogene |
| rp_asb-89581 Ψ | KQ034856 | 100 | RPRC000299-PA | 89,07 | Pseudogene |
| rp_asb-89583 | KQ034856 | 100 | RPRC002583-PA | 67,13 | locus RPRC002583 |
| rp_asb-89582 Ψ | KQ034856 | 100 | RPRC000299-PA | 66,13 | Pseudogene |
| rp_asb-1685 | KQ035220 | 99,25 | RPRC000072-PA | 90,91 | locus RPRC000072 |
| rp_asb-3232 | KQ035220 | 100 | RPRC000072-PA | 82,95 | locus RPRC000072 |
| rp_asb-3235 Ψ | KQ035220 | 100 | RPRC000072-PA | 70,45 | Pseudogene |
| rp_asb-101256 | KQ035220 | 98,7 | RPRC000072-PA | 69,31 | locus RPRC000072 |
| rp_asb-48741 | KQ036574 | 100 | RPRC000114-PA | 76,06 | Novo |
| rp_asb-105554 | KQ037843 | 98,91 | RPRC010481-PA | 84,44 | Novo |
| rp_asb-25624 | KQ035220 | 91,12 | RPRC000077-PA | 93,26 | |
| rp_asb-48694 | ACPB03031753 | 97,78 | RPRC000023-PA | 90 | |
| rp_asb-85899 | KQ034229 | 93,28 | RPRC010481-PA | 89,66 | |
| rp_asb-48699 | ACPB03031753 | 97,46 | RPRC000023-PA | 86,25 | |
| rp_asb-25025 | KQ035220 | 96,59 | RPRC000077-PA | 84,12 | |
| rp_asb-48126 | KQ034229 | 94,17 | RPRC000114-PA | 82,43 | |
| rp_asb-48125 | KQ034229 | 97,93 | RPRC000114-PA | 81,86 | |
| rp_asb-48124 | KQ034229 | 94,83 | RPRC000114-PA | 78,26 | |
| rp_asb-16225 | KQ034229 | 86,6 | RPRC000313-PA | 77,55 | |
| rp_asb-3239 | KQ035220 | 93,33 | RPRC000367-PA | 74,36 | |
| rp_asb-89584 | ACPB03035352 | 95,86 | RPRC002583-PA | 73,1 | |
| rp_asb-105049 | KQ035220 | 84,35 | RPRC000072-PA | 47,06 | |

Entre as Inibidoras de Agregação Plaquetária (Tabela 10) apenas três (rp_asb-34424, rp_asb-15376 e rp_asb-15398) mapearam ao genoma e ao proteoma; dois transcritos alinharam a proteínas já anotadas, mas não ao genoma (rp_asb-115859, rp_asb-97660); seis transcritos mapearam ao genoma, mas não ao proteoma (rp_asb-97659, rp_asb-15399, rp_asb-105551, rp_asb-44356, rp_asb-44609, rp_asb-104785), e os demais 11 não alinharam nem ao genoma, nem ao proteoma (rp_asb-15397, rp_asb-34427, rp_asb-34426, rp_asb-34429, rp_asb-15404, rp_asb-34425, rp_asb-105584,

rp_asb-34428). A anotação estrutural dos seis transcritos que mapearam no genoma, mas não ao proteoma, mostrou que os transcritos rp_asb-15399, rp_asb-44356 e rp_asb-44609 possuem diversos stop-codons e, assim, foram classificados como pseudogenes (recebendo o símbolo genético para pseudogene, Ψ). Já o transcrito rp_asb-105551 mapeou no mesmo locus do gene RPRC000446-PA, podendo então ser uma isoforma deste gene. Por fim, os transcritos rp_asb-97659 e rp_asb-104785 mapearam em locus sem genes anotados e não possuem nenhum stop-codon (ANEXO A), sugerindo serem novos genes.

Tabela 10. Mapeamento e identificação das lipocalinas inicialmente identificadas como PAI.

| Transcrito | Alinhamento ao genoma | | Alinhamento ao proteoma | | Comentário |
|---------------------|-----------------------|--------|-------------------------|--------|------------------|
| | Scaffold | ID (%) | Proteína | ID (%) | |
| rp_asb-34424 | KQ034115 | 100 | RPRC010695-PA | 100 | é RPRC010695 |
| rp_asb-15376 | ACPB03035352 | 99,19 | RPRC002584-PA | 98,75 | é RPRC002584 |
| rp_asb-15398 | ACPB03035352 | 100 | RPRC010444-PA | 98,33 | é RPRC010444 |
| rp_asb-115859 | ACPB03039915 | 94,9 | RPRC000677-PA | 100 | é RPRC000677 |
| rp_asb-97660 | KQ035263 | 92,86 | RPRC008117-PA | 98,21 | é RPRC008117 |
| rp_asb-97659 | KQ035263 | 100 | RPRC008117-PA | 91,23 | Novo |
| rp_asb-15399 Ψ | ACPB03035352 | 98,9 | RPRC012540-PA | 90 | Pseudogene |
| rp_asb-105551 | KQ035237 | 100 | RPRC000446-PA | 83,33 | locus RPRC000446 |
| rp_asb-44356 Ψ | ACPB03037469 | 100 | RPRC000446-PA | 68,06 | Pseudogene |
| rp_asb-44609 Ψ | ACPB03046785 | 100 | RPRC000446-PA | 61,81 | Pseudogene |
| rp_asb-104785 | KQ035719 | 100 | RPRC000463-PA | 55,17 | Novo |
| rp_asb-15397 | KQ038166 | 95,74 | RPRC010444-PA | 94,12 | |
| rp_asb-34427 | ACPB03042619 | 97,39 | RPRC000463-PA | 94 | |
| rp_asb-34426 | ACPB03042619 | 96,52 | RPRC000463-PA | 90,72 | |
| rp_asb-34429 | ACPB03042619 | 94,21 | RPRC000463-PA | 86 | |
| rp_asb-15404 | ACPB03035352 | 97,8 | RPRC012540-PA | 85,42 | |
| rp_asb-34425 | KQ035064 | 91,06 | RPRC000463-PA | 85,26 | |
| rp_asb-105584 | ACPB03040145 | 90,14 | RPRC000347-PA | 84,52 | |
| rp_asb-34428 | ACPB03040249 | 96,69 | RPRC000463-PA | 82 | |
| rp_asb-15400 | ACPB03035352 | 97,7 | RPRC002584-PA | 78,75 | |
| rp_asb-15401 | KQ036005 | 86,64 | RPRC002584-PA | 75 | |
| rp_asb-34430 | KQ035064 | 92,31 | RPRC000463-PA | 60,19 | |

Para as Pallidipinas (Tabela 11), dois transcritos (rp_asb-73727 e rp_asb-109013) mapearam ao genoma e ao proteoma do inseto. Outros quatro transcritos (rp_asb-73726, rp_asb-109014, rp_asb-104042 e rp_asb-73730) mapearam ao genoma, mas não ao proteoma; destes transcritos, rp_asb-104042 e rp_asb-73730 foram classificados como pseudogenes por apresentarem, na anotação estrutural, stop-codons em suas sequências (recebendo o símbolo genético para pseudogene, Ψ). O transcrito rp_asb-73726 mapeou no mesmo locus do gene RPRC000012-PA, podendo então ser uma isoforma deste gene. Similarmente, observou-se o mesmo para o transcrito rp_asb-109014, que mapeou no mesmo locus do gene RPRC002137 e pode, portanto, ser uma isoforma deste.

Tabela 11. Mapeamento das lipocalinas inicialmente identificadas como Pallidipinas.

| Transcrito | Alinhamento ao genoma | | Alinhamento ao proteoma | | Comentário |
|----------------------|-----------------------|--------|-------------------------|--------|------------------|
| | Scaffold | ID (%) | Proteína | ID (%) | |
| rp_asb-73727 | ACPB03037858 | 100 | RPRC000012-PA | 100 | é RPRC000012 |
| rp_asb-109013 | KQ034748 | 100 | RPRC002137-PA | 100 | é RPRC002137 |
| rp_asb-73726 | KQ034748 | 98,46 | RPRC000012-PA | 85,52 | locus RPRC000012 |
| rp_asb-109014 | KQ034748 | 99,32 | RPRC002137-PA | 94,25 | locus RPRC002137 |
| rp_asb-104042 Ψ | KQ034748 | 99,52 | RPRC002137-PA | 82,61 | Pseudogene |
| rp_asb-73730 Ψ | KQ034748 | 100 | SC* | | Pseudogene |
| rp_asb-73729 | KQ034748 | 94,61 | RPRC000012-PA | 76,92 | |
| rp_asb-73728 | KQ034748 | 95,72 | RPRC000012-PA | 88,65 | |
| rp_asb-73858 | KQ035329 | 93,54 | RPRC002137-PA | 51,41 | |
| rp_asb-112121 | KQ034748 | 93,98 | RPRC002137-PA | 73,12 | |
| rp_asb-109015 | KQ034748 | 93,38 | RPRC002137-PA | 92,31 | |
| rp_asb-72343 | KQ035064 | 97,45 | RPRC002138-PA | 55,41 | |

*Sem Correspondência

Assim como as Nitroforinas, as Triabinas foram um grupo com muitos transcritos identificados. Cinco transcritos mapearam ao genoma e ao proteoma de *R. prolixus* (rp_asb-102621, rp_asb-73238, rp_asb-89585, rp_asb-33412 e rp_asb-37664); 3 transcritos (rp_asb-102751, rp_asb-105585 e rp_asb-30197) mapearam ao proteoma, mas não ao genoma; e 9 transcritos (rp_asb-33413, rp_asb-102750, rp_asb-110283, rp_asb-30450, rp_asb-65689, rp_asb-102716, rp_asb-102620, rp_asb-105577 e rp_asb-40554) mapearam ao genoma e não ao proteoma (Tabela 12). Dos transcritos mapeados apenas ao genoma, dois deles (rp_asb-102750 e rp_asb-102716) foram identificados,

pela anotação estrutural, como pseudogenes (recebendo o símbolo genético para pseudogene, Ψ), pois apresentaram diversos stop-codons em suas sequências. Os transcritos rp_asb-33413 e rp_asb-65689 não apresentaram stop-codons (ver ANEXO A), sugerindo serem novos genes. O transcrito rp_asb-110283 apresenta uma estrutura gênica semelhante ao do gene RPRC000275 e ambos estão localizados próximos no scaffold, o que sugere que o transcrito seja originado de uma cópia deste gene; o mesmo ocorre com o transcrito rp_asb-105577, que é possivelmente originado de uma cópia do gene RPRC012954. O transcrito rp_asb-30450 mapeou no mesmo locus do gene RPRC000603, podendo representar uma isoforma deste gene. Já o transcrito rp_asb-105577 mapeou no mesmo locus do gene RPRC012954, o que pode indicar se tratar de uma isoforma deste gene.

Tabela 12. Mapeamento das lipocalinas inicialmente identificadas como Triabinas.

| Transcrito | Alinhamento ao genoma | | Alinhamento ao proteoma | | Comentário |
|----------------------|-----------------------|--------|-------------------------|--------|------------------|
| | Scaffold | ID (%) | Proteína | ID (%) | |
| rp_asb-102621 | KQ035237 | 100 | RPRC000275-PA | 99,35 | é RPRC000275 |
| rp_asb-73238 | KQ034856 | 99,68 | RPRC000418-PA | 100 | é RPRC000418 |
| rp_asb-89585 | KQ034856 | 100 | RPRC002583-PA | 100 | é RPRC002583 |
| rp_asb-33412 | KQ034748 | 99,4 | RPRC002646-PA | 98,81 | é RPRC002646 |
| rp_asb-37664 | KQ034748 | 100 | RPRC002647-PA | 100 | é RPRC002647 |
| rp_asb-102751 | ACPB03037937 | 96,92 | RPRC000147-PA | 100 | é RPRC000147 |
| rp_asb-105585 | ACPB03047339 | 91,15 | RPRC000347-PA | 100 | é RPRC000347 |
| rp_asb-30197 | KQ035220 | 97,22 | RPRC000603-PA | 100 | é RPRC000603 |
| rp_asb-33413 | KQ034748 | 99,4 | RPRC000038-PA | 88,89 | Novo |
| rp_asb-102750 Ψ | KQ035329 | 100 | RPRC000147-PA | 62,32 | Pseudogene |
| rp_asb-110283 | KQ035237 | 100 | RPRC000275-PA | 76,47 | Novo |
| rp_asb-30450 | KQ035220 | 100 | RPRC000603-PA | 94,69 | locus RPRC000603 |
| rp_asb-65689 | ACPB03046557 | 100 | RPRC000613-PA | 93,64 | Novo |
| rp_asb-102716 Ψ | ACPB03037937 | 100 | RPRC006167-PA | 95,59 | Pseudogene |
| rp_asb-102620 | KQ035237 | 100 | RPRC012954-PA | 89,16 | é RPRC012954 |
| rp_asb-105577 | ACPB03037685 | 100 | RPRC012954-PB | 65,28 | Novo |
| rp_asb-40554 | KQ035237 | 100 | RPRC012954-PB | 84,09 | locus RPRC012954 |
| rp_asb-102753 | SC* | | RPRC000147-PA | 82,61 | |
| rp_asb-103687 | SC* | | RPRC000147-PA | 58,06 | |
| rp_asb-40550 | SC* | | RPRC000275-PA | 70,09 | |
| rp_asb-30198 | KQ035220 | 95,7 | RPRC000603-PA | 97,37 | |
| rp_asb-102749 | KQ035064 | 98,46 | RPRC006167-PA | 74,03 | |
| rp_asb-102752 | KQ035064 | 97,65 | RPRC006167-PA | 63,51 | |
| rp_asb-40553 | KQ035237 | 93,2 | RPRC012954-PA | 81,25 | |

*Sem Correspondência

Como dito anteriormente, pela primeira vez tem-se a sugestão de que lipocalinas de *R. prolixus* possam pertencer a família das Apolipoproteínas. A Tabela 13 resume o resultado do mapeamento para as lipocalinas pretendentes a este grupo. Nele, 6 transcritos (rp_asb-6793, rp_asb-14972, rp_asb-4751, rp_asb-48, rp_asb-40377 e rp_asb-47) mapearam ao genoma e ao proteoma do inseto. Destes, o transcrito rp_asb-47 mapeou no mesmo locus do gene RPRC015421-PA, podendo ser uma isoforma deste. Três transcritos (rp_asb-6771, rp_asb-105382 e rp_asb-7471) mapearam ao genoma, mas não mapearam ao proteoma. A anotação estrutural destes transcritos mostrou que o rp_asb-105382 possui diversos stop-codons em sua sequência e, por isso, ele foi classificado como um pseudogene (recebendo o símbolo genético para pseudogene, Ψ); já o transcrito rp_asb-6771 foi identificado como uma cópia do gene RPRC015421, pois os genes de ambos têm estruturas muito parecidas e estão anotados lado a lado no genoma; o transcrito rp_asb-7471, por sua vez, mapeia no locus do gene RPRC000339, podendo representar uma possível isoforma deste gene. Os outros dois transcritos restantes (rp_asb-7470 e rp_asb-1368) mapearam ao proteoma, mas não ao genoma do inseto.

Tabela 13. Mapeamento das lipocalinas inicialmente identificadas como Apolipoproteínas.

| Transcrito | Alinhamento ao genoma | | Alinhamento ao proteoma | | Comentário |
|----------------------|-----------------------|--------|-------------------------|--------|------------------|
| | Scaffold | ID (%) | Proteína | ID (%) | |
| rp_asb-6793 | KQ034497 | 100 | RPRC003587-PA | 100 | é RPRC003587 |
| rp_asb-14972 | KQ034137 | 99,7 | RPRC010099-PA | 100 | é RPRC010099 |
| rp_asb-4751 | KQ034633 | 99,36 | RPRC000014-PA | 100 | é RPRC000014 |
| rp_asb-48 | KQ034457 | 98,1 | RPRC015421-PA | 99,53 | é RPRC015421 |
| rp_asb-40377 | KQ034133 | 99,47 | RPRC004179-PA | 99,19 | é RPRC004179 |
| rp_asb-47 | KQ034457 | 99,28 | RPRC015421-PA | 99,05 | locus RPRC015421 |
| rp_asb-7470 | KQ034069 | 97,73 | RPRC000339-PA | 100 | é RPRC000339 |
| rp_asb-1368 | KQ034130 | 97,05 | RPRC004463-PA | 100 | é RPRC004463 |
| rp_asb-6771 | KQ034457 | 100 | RPRC015420-PA | 97,6 | novo |
| rp_asb-105382 Ψ | KQ034133 | 100 | RPRC004250-PA | 73,91 | pseudogene |
| rp_asb-7471 | KQ034069 | 100 | RPRC000339-PA | 71,82 | locus RPRC000339 |

Por fim, dois transcritos, rp_asb-105519 e rp_asb-61810, apresentaram similaridade às lipocalinas, mas não foram classificados em nenhum grupo funcional específico pelas análises iniciais. Essas lipocalinas foram chamadas de “Inespecíficas” e o resultado do mapeamento delas é mostrado na Tabela 14. Ambos os transcritos foram mapeados ao genoma, mas não ao proteoma de *R. prolixus*. Porém, ao analisar o locus específico do gene com o GeneBrowser do Vectorbase, foi observado que as proteínas RPRC000438-PA e RPRC000418-PA estão anotadas no mesmo local. As diferenças de similaridade observadas parecem ser fruto de erro de anotação do transcrito (dados não mostrados). Desta forma, consideraram-se as proteínas anotadas como sendo as lipocalinas salivares inespecíficas e suas sequências foram utilizadas nas análises posteriores.

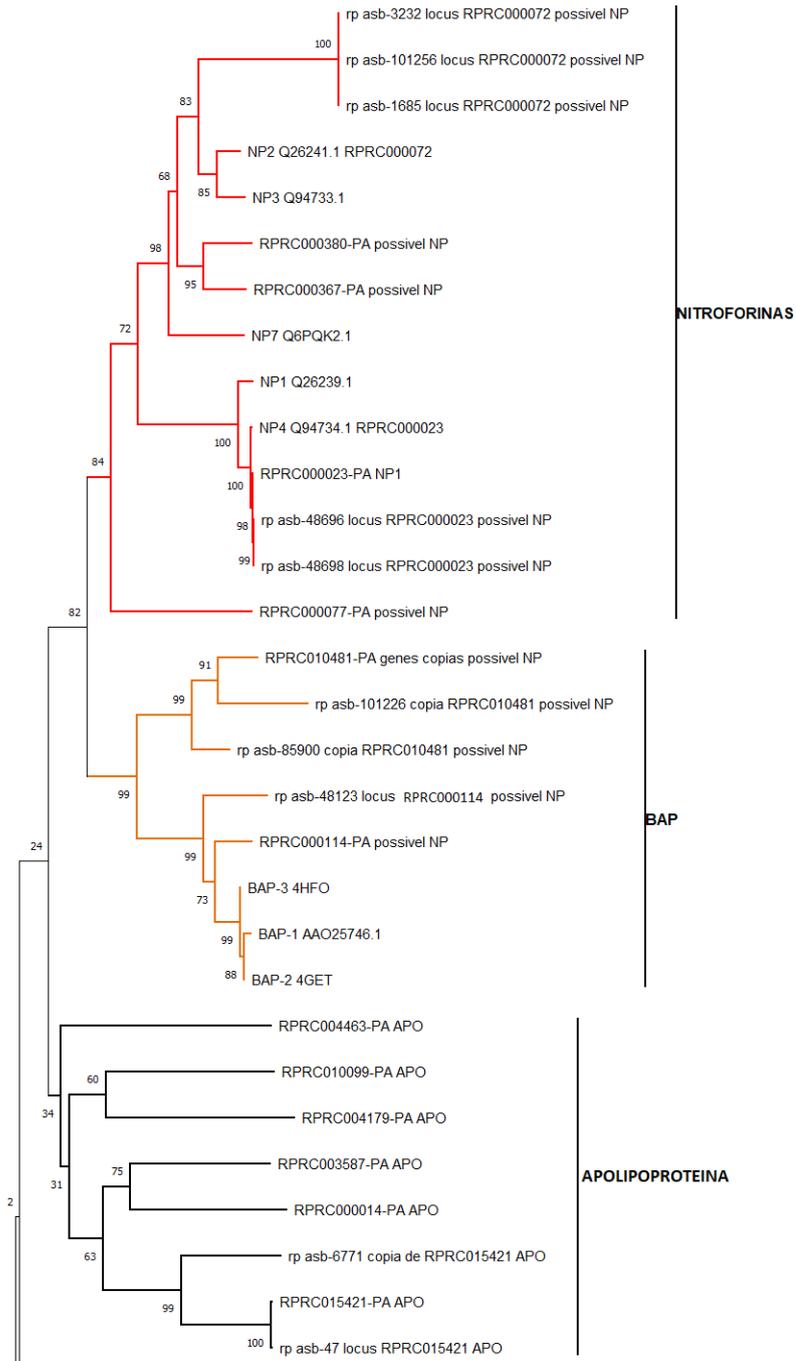
Tabela 14. Mapeamento das lipocalinas inicialmente identificadas como “Inespecíficas”.

| Transcrito | Alinhamento ao genoma | | Alinhamento ao proteoma | | Comentário |
|---------------|-----------------------|--------|-------------------------|--------|--------------|
| | Scaffold | ID (%) | Proteína | ID (%) | |
| rp_asb-105519 | KQ034856 | 100 | RPRC000438-PA | 59,04 | é RPRC000438 |
| rp_asb-61810 | KQ034856 | 100 | RPRC000418-PA | 42,25 | é RPRC000418 |

4.4. Análise inicial de ortologia das lipocalinas do transcriptoma de *R. prolixus*.

A identificação da função hipotética de cada lipocalina por meio dos resultados iniciais do BLAST é importante para dar um direcionamento a respeito do perfil dos transcritos identificados. No entanto, uma análise evolutiva por filogenia com todas as lipocalinas descritas pode oferecer a robustez necessária para uma melhor classificação da função putativa de cada transcrito. Assim, todas as lipocalinas descritas no transcriptoma e anotadas no presente trabalho foram analisadas evolutivamente em conjunto com as lipocalinas já descritas anteriormente (Tabela 1). Devido a diversidade de sequências e diferenças na qualidade de anotação, optou-se por realizar uma análise inicial (ANEXO C) contendo todas as sequencias identificadas (incluindo cópias e transcritos que mapearam no mesmo locus de genes já anotados), com exceção de pseudogenes. Desta análise, a sequência RPRC008117 foi removida por ser muito curta (62 aa) e estar limitando o número de sítios para construção da filogenia. Apesar da diversidade das sequências, muitas delas puderam ser agrupadas em grupos funcionais distintos de acordo com a proximidade evolutiva apresentada entre elas. Também foi

possível confirmar os resultados do BLAST mapeando as lipocalinas já conhecidas ao proteoma. Já as sequências RPRC0110695, RPRC006975, RPRC000418, RPRC000163, rp_asb-89583, rp_asb-65689 e rp_asb-110283 claramente não agruparam com nenhuma das lipocalinas conhecidas e, assim, elas foram excluídas das análises por não ser possível, neste momento, concluir se são de fato lipocalinas. Também foi possível observar que as proteínas preditas como apolipoproteínas não formaram um clado coeso. Com base então na figura do ANEXO C, criou-se um arquivo revisado, contendo apenas as sequências que se mostraram evolutivamente mais próximas a lipocalinas já descritas, assim como as apolipoproteínas (já que para estas não se possui um modelo). Essas sequências foram novamente analisadas evolutivamente e uma nova árvore filogenética foi construída (Figura 3).



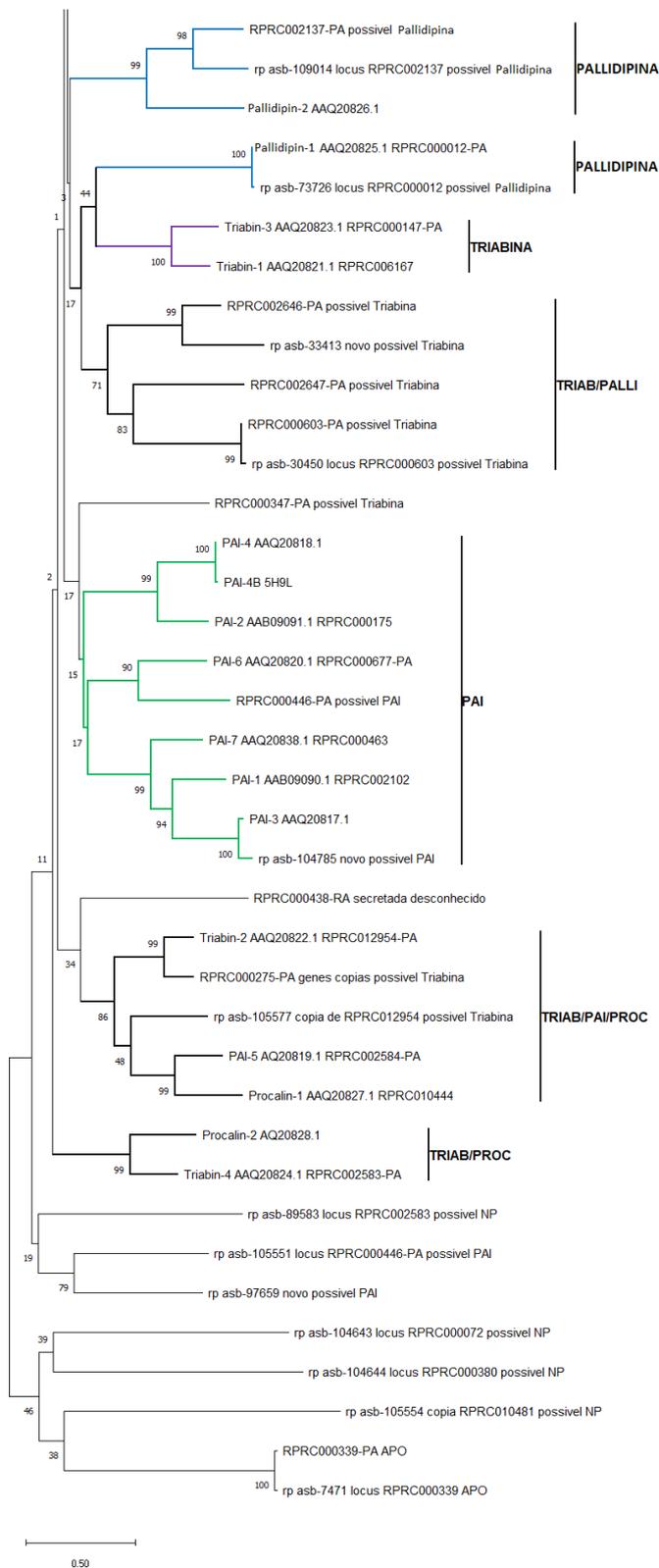


Figura 3. Relação evolutiva das Lipocalinas de *Rhodnius prolixus*. A história evolucionaria foi inferida usando o método de Neighbor-Joining, e o teste de bootstrap a partir de 10000 réplicas. Todas as posições ambíguas foram removidas para cada par de sequências. Os números de acesso das sequências estão demonstrados em cada ramo da árvore filogenética.

Com base na análise evolutiva das lipocalinas (Figura 3), foi possível identificar grupos distintos para as nitroforinas, BAP e PAI. Várias sequências que foram identificadas como possíveis nitroforinas apresentaram maior relação evolutiva com as BAPs, sendo elas: RPRC010481, rp_asb-101226, rp_asb-85900, rp_asb-48123 e RPRC000114. Assim, nas análises futuras, foi considerado que essas sequências seriam possíveis BAPs ao invés de nitroforinas. As análises também mostraram que a proteína RPRC000023, anotada no VectorBase como sendo a Nitroforina-1, é muito mais similar a Nitroforina-4. As apolipoproteínas ficaram, em sua maioria, agrupadas em um mesmo ramo, porém com um valor de bootstrap baixo, de apenas 34. Já as triabinas e pallidipinas se distribuíram em vários ramos menores, distintos apenas em grupos de duas ou três sequências, ficando incerta a relação evolutiva funcional destes genes. Também foi identificado um ramo distinto que contém a Triabina-2 (AAQ20822.1), a PAI-5 (AQ20819.1), a Procalina-1 (AAQ20827) e dois novos genes que haviam sido identificados como possíveis triabinas (RPRC000275 e rp_asb-105577). Da mesma forma, a Procalina-2 (AQ20828.1) e a Triabina-4 (AAQ20824.1) agruparam num mesmo ramo (bootstrap 99). É importante lembrar que triabinas e pallidipinas, na realidade, são proteínas que inibem a agregação plaquetária, sendo que as pallidipinas receberam esta nomenclatura por terem sido isoladas de *Triatoma pallidipenis*. Esta questão será discutida de forma mais aprofundada adiante. Por fim, a análise evolutiva sugere que as sequências rp_asb-89583, rp_asb-105551, rp_asb-97659, rp_asb-104643, rp_asb-104644, rp_asb-105554, rp_asb-7471 e RPRC000339 não são evolutivamente próximas a nenhuma das lipocalinas já conhecidas e, desta forma, foram excluídas de análises futuras. Com um melhor entendimento das relações de similaridade e evolutiva das lipocalinas, seguiu-se para as análises específicas de cada grupo funcional.

4.5. Estrutura, evolução e expressão das Nitroforinas no transcriptoma de *R. prolixus*

Como mostrado nos itens 4.3 e 4.4, as anotações e análises evolutivas iniciais identificaram um total de 10 transcritos com ortologia a nitroforinas já conhecidas (NP-1, NP-2, NP-3, NP-4 e NP-7). Dos 10 transcritos identificados, cinco já estavam anotados no proteoma (RPRC000072, RPRC000380, RPRC000367, RPRC000023 e RPRC000077), sendo que RPRC000023 e a RPRC000072 já haviam sido identificadas, respectivamente, como NP-1 e NP-2. As análises iniciais (Figura 3 e ANEXO C) sugerem que RPRC000072 e NP-2 são de fato o mesmo gene. Por outro lado, RPRC000023 se mostrou mais similar a NP-4 em ambas as análises evolutivas. Aqui foi realizada uma análise evolutiva exclusiva das nitroforinas (como definido pelas análises anteriores) com a inclusão da NP-5 e NP-6 (Figura 4). As sequências de NP-5 e NP-6 foram obtidas diretamente no artigo de Moreira e colaboradores (2003) já que estas sequências nunca foram depositadas no GeneBank, e elas possuem, respectivamente, 33 e 29 aminoácidos. Dessa forma, ressaltamos que a análise foi realizada com a opção de deleção par a par, já que a deleção completa limitaria toda a análise a menos de 30 aminoácidos, o que diminuiria consideravelmente o poder de resolução do estudo.

Mais uma vez, assim como na figura 3, a NP-4 e a RPRC000023 (identificada como NP-1 no VectorBase) ficaram agrupadas, sugerindo que estas sequências se referem a mesma proteína. As nitroforinas RPRC000380, RPRC000367 e RPRC000077 identificadas anteriormente por Mesquita e colaboradores (2015), ficaram em ramos isolados. Por fim, cinco transcritos foram mapeados no mesmo locus que nitroforinas já anotadas, sendo eles: *rp_asb-48696* e *rp_asb-48698* no locus de RPRC000023 (possivelmente NP-4); e *rp_asb-1685*, *rp_asb-3232*, *rp_asb-101256* no locus de RPRC000072. Quanto a NP-5 e NP-6, esta é a primeira análise evolutiva destas proteínas. A filogenia sugere que a NP-6 é altamente relacionada com os transcritos de mesmo locus da RPRC000072. Já a NP-5 fica agrupada no mesmo ramo que a NP-1 e a NP-4, não ficando bem definido se esta proteína estaria mais relacionada a NP-1 ou NP-4.

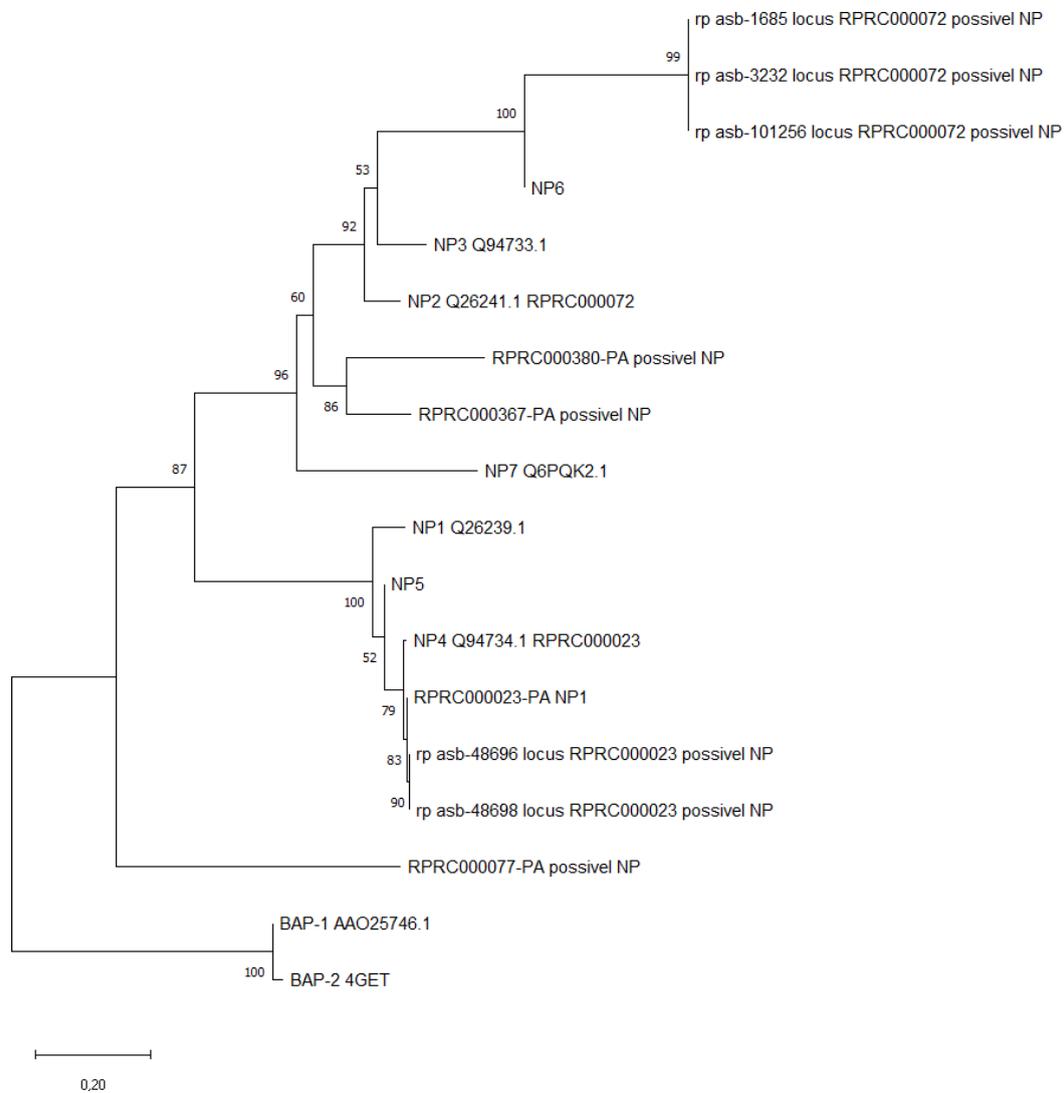


Figura 4. Relações evolutivas das Nitroforinas de *R. prolixus*. A história evolucionária foi inferida usando o método de Neighbor-Joining, e o teste de bootstrap a partir de 10000 réplicas. Todas as posições contendo gaps ou dados perdidos foram eliminadas (opção de deleção completa). Os números de acesso das sequências estão indicados em cada ramo da árvore filogenética. Duas proteínas BAP foram usadas como grupo externo.

As 13 nitroforinas identificadas (com exceção as NP-5 e NP-6 que possuem sequências muito curtas e não acrescentam informações a esta análise) foram então alinhadas pelo programa CLUSTAL Omega para uma análise mais detalhada de domínios e conservação das proteínas (Figura 5). De uma forma geral, o domínio é conservado em todas as proteínas e foi possível identificar cinco regiões conservadas de ligação ao grupo heme na maioria das nitroforinas. As nitroforinas NP-2-PB, NP-2-PC e NP-2-PD foram as nitroforinas mais diferentes dentre as observadas. O domínio destas três proteínas termina prematuramente, no aminoácido 118, interrompendo a conservação do quarto sítio de ligação do grupo heme. Ainda assim, estas três nitroforinas mantêm conservado o quinto sítio de ligação do grupo heme, o que faz então que as nitroforinas NP-2-PB, NP-2-PC e NP-2-PD tenham um sítio de ligação do grupo heme a menos que as demais nitroforinas. Por fim, as duas isoformas de NP-4 (NP-4-PB e NP-4-PC) aparentam se diferenciar da NP-4 devido a uma tradução prematura do gene, causando uma extensão de 35 aminoácidos na região N-terminal destas nitroforinas.

| | | |
|------------|---|-----|
| RPRC000077 | -----SIALVLAAVTFWQLAAERETAECTSS | 26 |
| NP-1 | -----MKSYTALLAVAILCLFAAVGVSGKCTKN | 28 |
| NP-4 | -----MKSYTSLLAVALILCLFG--GVNGACTKN | 26 |
| NP-4-PB | MYLRVEFKLYFFIVPXSLKKIPELQFKTINVCSRYTSLLAVAILCLFA--GVNGACTKN | 58 |
| NP-4-PC | MYLRVEFKLYFFIVPXSLKKIPELQFKTINVCSRYTSLLAVAILCLFA--GVNGACTKN | 58 |
| NP-2-PD | -----CDDYLNCTLSIIIFPSRYSAALLAVTILCLTSTMGVSGDCSTN | 42 |
| NP-2-PB | -----CDDYLNCTLSIIIFPSRYSAALLAVTILCLTSTMGVSGDCSTN | 42 |
| NP-2-PC | -----CDDYLNCTLSIIIFPSRYSAALLAVTILCLTSTMGVSGDCSTN | 42 |
| NP-7 | -----MELYTALLAVTILSPSSIVGLPGECSVN | 28 |
| RPRC000380 | -----SALLAVTILCLTSTMGVSEQTQG | 24 |
| RPRC000367 | -----MEPYSALLAVTILCLTSTMGVSEECSTN | 28 |
| NP-2 | -----MELYTALLAVTILCLTSTMGVSGDCSTN | 28 |
| NP-3 | -----MEPYSALLAVTILCLTSTMGVSGDCSTN | 28 |
| | : ** . . : . * : . | |
| RPRC000077 | VTATTKLNKDKYFKN-GWYATHFKAVQVPTKHSATASKYCAITVGVKVNNEVTEVLSHYN | 85 |
| NP-1 | ALAQTGFNKDKYFNGDVWYVTDYLDLEP --- DDVPKRYCAALAAGTASGKLEALYHYD | 84 |
| NP-4 | AIAQTGFNKDKYFNGDVWYVTDYLDLEP --- DDVPKRYCAALAAGTASGKLEALYHYD | 82 |
| NP-4-PB | AIAQTGFNKDKYFNGDVWYVTDYLDLEP --- DDVPKRYCAALAAGTASGKLEALYHYD | 114 |
| NP-4-PC | AIAQTGFNKDKYFNGDVWYVTDYLDLEP --- DDVPKRYCAALAAGTASGKLEALYHYD | 114 |
| NP-2-PD | ISPKKGLDKAKYFSG-TWYVTHYLDKDP --- Q-VTDPYCSSFTPKESGGTVKEALYHFN | 96 |
| NP-2-PB | ISPKKGLDKAKYFSG-TWYVTHYLDKDP --- Q-VTDPYCSSFTPKESGGTVKEALYHFN | 96 |
| NP-2-PC | ISPKKGLDKAKYFSG-TWYVTHYLDKDP --- Q-VTDPYCSSFTPKESGGTVKEALYHFN | 96 |
| NP-7 | VIPKKNLDDKAKFFSG-TWYETHYLDMDP --- Q-ATEKFCFSFAPRESGGTVKEALYHFN | 82 |
| RPRC000380 | ISPKSGLDKEKYSG-TWYVTHYLDLEP --- Q-GSDKYCSSFAPQTQGNKVNKEIFHNN | 78 |
| RPRC000367 | ISPKSGLDKEKYSG-TWYVTHYLDKDP --- Q-VTDKYCSSFAPRKDGNVREALYHYK | 82 |
| NP-2 | ISPKQGLDKAKYFSG-KWYVTHFLDKDP --- Q-VTDQYCSSFTPRESDGTVKEALYHYN | 82 |
| NP-3 | ISPKKGLDKAKYFSG-TWYVTHYLDKDP --- Q-VTDPYCSSFTPKESGGTVKEALYHFN | 82 |
| | : * * : . . * * : . : * : . . : * : * . . | |
| RPRC000077 | SKTQSSIFDISKV--TVSGGKYTANYRRVDKDGQDKGQ--SGTYTIVLDTDND----- | 135 |
| NP-1 | PKTQDTFYDVSELQEESSPG-KYTANFKKVEKNGNVKVDVTSGNYYTFTVMYADDS----- | 138 |
| NP-4 | PKTQDTFYDVSELQVEESLG-KYTANFKKVDKNGNVKVAVTAGNYTFTVMYADDS----- | 136 |
| NP-4-PB | PKTQDTFYDVSELQVEESLG-KYTANFKKVDKNGNVKVAVTAGNYTFTVMYADDS----- | 168 |
| NP-4-PC | PKTQDTFYDVSELQVEESLG-KYTANFKKVDKNGNVKVAVTAGNYTFTVMYADDS----- | 168 |
| NP-2-PD | SKPNFIYKDDQSKRYKSIICLVNNAKFFEQPRG ----- SFFEKMTRVERSKSVTGRH | 147 |
| NP-2-PB | SKPNFIYKDDQSKRYKSIICLVNNAKFFEQPRG ----- SFFEKMTRVERSKSVTGRH | 147 |
| NP-2-PC | SKPNFIYKDDQSKRYKSIICLVNNAKFFEQPRG ----- SFFEKMTRVERSKSVTGRH | 147 |
| NP-7 | VDSKVSFYNTGTGPLESNGAKYTAKFNTVDPKKGKIKPADEKYSYTVTVIEAAKQ----- | 137 |
| RPRC000380 | TNKNQKMYHIGEGKLGSLGAKYTAKYKTVDSGKEIGQPNEKNSYTVTVMEADES----- | 133 |
| RPRC000367 | TNDKMKMYHVGEGTLESSGAKYTAKYSTVDSSEKELVAANEKNTYTVTVMEADDS----- | 137 |
| NP-2 | ANKKTSFYNIGEGKLESSGLQYTAKYKTVDKKAVLKEADEKNSYTLTVLEADDS----- | 137 |
| NP-3 | SKKKTSTFYNIGEGKLGSSGVQYTAKYNTVDKRKEIEPADPKDSYTLTVLEADDS----- | 137 |
| | . . . : . * : . . | |
| RPRC000077 | --NYAITHLCVTRG-DDVFEFYTIANRQKDGTH-AGVSSGLSSASLTLDFTSSKDLGCN | 191 |
| NP-1 | ---SALIHHTCLHKGKNDLGDLYAVLNRNKDNTAGDKVKGAVTAAALSKFESDFISTKDNKCE | 195 |
| NP-4 | ---SALIHHTCLHKGKNDLGDLYAVLNRNKDAAAGDKVKSASVSAATLEFSKFISTKENNCA | 193 |
| NP-4-PB | ---SALIHHTCLHKGKNDLGDLYAVLNRNKDAAAGDKVKSASVSAATLEFSKFISTKENNCA | 225 |
| NP-4-PC | ---SALIHHTCLHKGKNDLGDLYAVLNRNKDAAAGDKVKSASVSAATLEFSKFISTKENNCA | 225 |
| NP-2-PD | CINKRAVSTEGLSDQSRYSRY-QATTLQSRSETESR-RAAPESGGGQSQLEGHTQLKGV | 205 |
| NP-2-PB | CINKRAVSTEGLSDQSRYSRY-QATTLQSRSETESR-RAAPESGGGQSQLEGHTQLKGV | 205 |
| NP-2-PC | CINKRAVSTEGLSDQSRYSRY-QATTLQSRSETESR-RAAPESGGGQSQLEGHTQLKGV | 205 |
| NP-7 | ---SALIHICLQEDGKDIGDLYSVLNRNKALPNKKIKKALNKVSLVLTKFVVTKDLDC | 194 |
| RPRC000380 | ---SALVHICLRKNGNELGDLFTVLSHQKTGPSATVTNAMSQTSLDLNKFFVTKDLGCE | 190 |
| RPRC000367 | ---SALIHICLKNNGEELGDLYTVLNRDKDTEPNKVKSAVTAQAGLQLSQFVGTGKTLGCG | 194 |
| NP-2 | ---SALVHICLREGSKDLGDLYTVLTHQKDAEPSAKVKSASVTAQAGLQLSQFVGTGKTLGCG | 194 |
| NP-3 | ---SALVHICLREGPKDLGDLYTVLSHQKTGPSATVKNVAQAAGLKLNDVDTKTLSC | 194 |
| | * . . . : . : | |
| RPRC000077 | YEEEKAKKLLTK 203 | |
| NP-1 | YDNVSLKSLTK 207 | |
| NP-4 | YDNDSLKSLTK 205 | |
| NP-4-PB | YDNDSLKSLTK 237 | |
| NP-4-PC | YDNDSLKSLTK 237 | |
| NP-2-PD | LNELLISF -----213 | |
| NP-2-PB | LNELLISF -----213 | |
| NP-2-PC | LNELLISF -----213 | |
| NP-7 | YDDKFLSSWQK- 205 | |
| RPRC000380 | YDDKFTS -----197 | |
| RPRC000367 | YDDEFTS -----201 | |
| NP-2 | YDDQFTSL -----202 | |
| NP-3 | YDDQFTSM -----202 | |
| | : : . | |

Figura 5. Alinhamento das Nitroforinas de *Rhodnius prolixus*. As sequências de aminoácidos foram alinhadas com o programa CLUSTAL-Omega, usando os parâmetros pré-definidos. Em cinza está marcada a região do domínio para a maioria das proteínas (predito pelo CDD). Em vermelho os sítios de ligação do grupo heme.

A tabela 15 mostra os dados de expressão das nitroforinas aqui identificadas. Como não foram identificados transcritos correspondentes a NP-1, NP-3 e NP-7, não foi possível avaliar a expressão destes genes nos diferentes tecidos, mas é sabido que as três nitroforinas são expressas em glândula salivar. Além disso, a partir deste ponto, decidiu-se por chamar os transcritos que se originam do mesmo locus de proteínas anotadas como isoformas (variações transcricionais, originadas por *splicing* alternativo, de um mesmo gene), adotando, por convenção, o nome do gene original com as variações PA, PB, PC, etc., para as diferentes formas. Apesar das nitroforinas serem consideradas proteínas características de glândula salivar, foram observados três transcritos de nitroforinas sendo expressos em tecidos diferentes da glândula salivar (levando em conta que a amostra de corpo inteiro das formas imaturas (NI) contém glândula-salivar e outros tecidos, sendo impossível diferenciar de qual tecido o transcrito seria proveniente). Uma das isoformas da NP-2, a NP-2-PB, foi detectada em intestino anterior, intestino posterior, reto e testículos. A RPRC000077-PA foi detectada em intestino anterior e posterior, assim como nos órgãos sexuais masculino (testículos) e feminino (ovários). Por fim, a RPRC000380 foi detectada também em amostras de testículo.

Tabela 15. Nitroforinas expressas em diferentes tecidos de *Rhodnius prolixus*.

| Transcrito | CG | TM | IA | IP | Rt | OV | Te | NI |
|----------------------------|----|----|-------|------|------|------|-------|-------|
| NP-2 (RPRC000072-PA) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 18,54 |
| NP-2-PB (rp_asb-1685) | 0 | 0 | 24,21 | 2,20 | 7,45 | 0 | 13,90 | 9,23 |
| NP-2-PC (rp_asb-3232) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6,30 |
| NP-2-PD (rp_asb-101256) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4,90 |
| NP-4 (RPRC000023-PA) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15,43 |
| NP-4-PB (rp_asb-48696) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4,41 |
| NP-4-PC (rp_asb-48698) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5,78 |
| RPRC000077-PA | 0 | 0 | 3,00 | 8,93 | 0 | 0,87 | 10,71 | 1,09 |
| RPRC000367-PA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4,30 |
| RPRC000380-PA | 0 | 0 | 0 | 0 | 0 | 0 | 0,85 | 3,09 |

CG – Corpo Gorduroso; TM – Túbulo de Malpighi; IA – Intestino Anterior; IP – Intestino Posterior; Rt – Reto; OV – Ovário; Te – Testículo; NI – Corpo Inteiro de formas imaturas (ovos e ninfas). Em verde significa transcrição não detectada no tecido pelo transcriptoma, em vermelho significa transcrição detectada (em Reads por Kilobase – RPK).

4.6. Estrutura, evolução e expressão das Proteínas Biogênicas de Ligação a Amina (BAP) no transcriptoma de *Rhodnius prolixus*

Como mostrado nas análises evolutivas iniciais (item 4.4), 5 transcritos (RPRC000114, rp_asb-48123, rp_asb-101226, RPRC010481 e rp_asb-85900) que anteriormente haviam sido classificados como possíveis nitroforinas (ver Tabela 9) se relacionaram evolutivamente a lipocalinas do tipo BAP. Dos 5 transcritos identificados, dois já estavam anotados no proteoma (RPRC000114 e RPRC010481). Foi realizada uma análise evolutiva exclusiva das BAPs (incluindo as BAPs já descritas em *R. prolixus* e os 5 transcritos aqui identificados) e foi observado que os 5 transcritos ficaram em um ramo isolado da árvore filogenética, sugerindo serem novas BAPs de *R. prolixus* (Figura 6). Como visto no resultado do mapeamento (ver item 4.3), duas BAPs foram mapeadas em locus de lipocalinas já anotadas, sendo elas: rp_asb-48123, mapeada no mesmo locus da lipocalina RPRC000114; e rp_asb-85900, mapeada no mesmo locus de RPRC010481. Essas duas BAPs foram consideradas, portanto, possíveis isoformas dos respectivos genes.

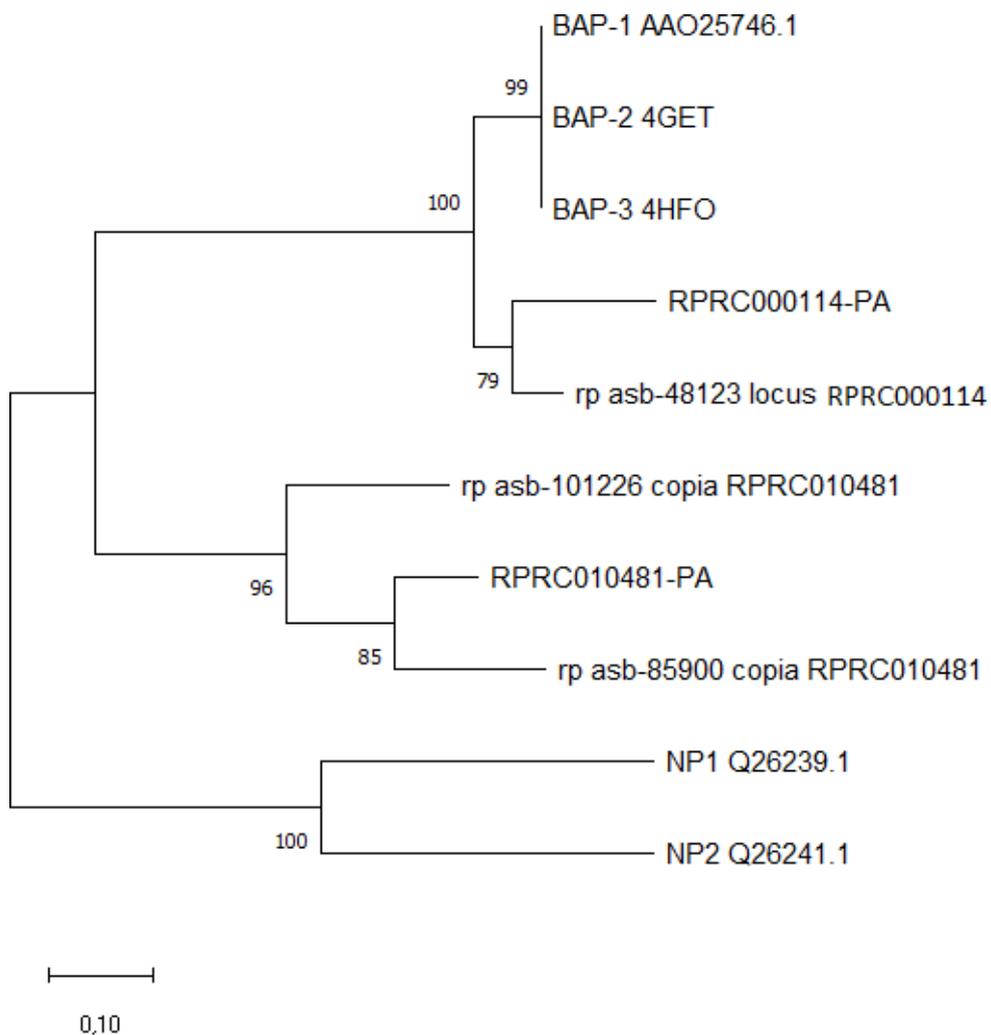


Figura 6. Relações evolutivas das Proteínas Biogênicas de Ligação a Amina de *R. prolixus*. A história evolucionaria foi inferida usando o método de Neighbor-Joining, e o teste de bootstrap a partir de 10000 réplicas. Todas as posições contendo gaps ou dados perdidos foram eliminadas (opção de deleção completa). Os números de acesso das sequências estão demonstrados em cada ramo da árvore filogenética. Duas nitroforinas foram usadas como grupo externo.

A Tabela 16 mostra os dados de expressão das BAPs aqui identificadas. Por não terem sido identificados transcritos correspondentes a BAP-1, BAP-2, BAP-4 e RPRC010481-PA, não foi possível avaliar a expressão destes genes nos diferentes tecidos do inseto, mas é sabido que as três BAPs já descritas (BAP-1, BAP-2, BAP-4)

são expressas em glândula salivar. A partir deste ponto, os transcritos que se originaram do mesmo locus de proteínas já anotadas passaram a ser denominados de isoformas (variações transcricionais de um mesmo gene originadas por *splicing* alternativo). Para estes transcritos foram adotados, por convenção, o nome do gene original com as variações PA, PB, PC etc., para as diferentes formas. Apesar das BAPs serem proteínas descritas em glândula salivar, foi observado a expressão de um dos transcritos (RPRC000114-PA) nos órgãos sexuais masculino (testículos) e feminino (ovários) de insetos adultos. Os demais transcritos tiveram expressão observada apenas na amostra de corpo inteiro das formas imaturas (NI) de *R. prolixus*.

Tabela 16. Proteínas Biogênicas de Ligação Amina expressas em diferentes tecidos de *Rhodnius prolixus*.

| Transcrito | CG | TM | IA | IP | Rt | OV | Te | NI |
|---|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| rp_asb-101226 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4,07 |
| RPRC000114-PA | 0 | 0 | 0 | 0 | 0 | 1,06 | 0,81 | 22,66 |
| RPRC000114-PB (rp_asb-48123) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2,41 |
| RPRC010481-PB (rp_asb-85900) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4,29 |

CG – Corpo Gorduroso; TM – Túbulo de Malpighi; IA – Intestino Anterior; IP – Intestino Posterior; Rt – Reto; OV – Ovário; Te – Testículo; NI – Corpo Inteiro de formas imaturas (ovos e ninfas). Em verde significa transcrição não detectada no tecido pelo transcriptoma, em vermelho significa transcrição detectada (em Reads por Kilobase – RPK).

As 8 BAPs analisadas foram então alinhadas pelo programa CLUSTAL Omega para uma análise mais detalhada de domínios e conservação das proteínas (Figura 7). A região de domínio é conservada em todas as proteínas e foi possível identificar 10 segmentos do sítio de ligação conservados entre elas. As já descritas BAP-1, BAP-2 e BAP-3 são praticamente idênticas à proteína RPRC000114-PA já anotada no proteoma. Além disso, elas se encontram no mesmo locus do gene RPRC000114, indicando se tratar da mesma proteína.

| | | | | | | |
|---------------|--|----------------------|------------------|-----------|--------------|-----|
| RPRC000114-PA | MRAYAALVVFVALWMSGAEGASGCLTVDTVKDFNKDN | FTTGSWY | --- | ITHYKLGGS | 54 | |
| rp_asb-48123 | ----- | ----- | ----- | ----- | 0 | |
| BAP-1 | MRAYAALVLFVVALWMSAEAGASGCVDTVKDFNKDN | FTTGSWY | --- | ITHYKLGDS | 54 | |
| BAP-2 | -----MAGCSTVDTVKDFNKDN | FTTGSWY | --- | ITHYKLGDS | 34 | |
| BAP-3 | -----SGCSTVDTVKDFNKDN | FTTGSWY | --- | ITHYKLGDS | 32 | |
| rp_asb-101226 | -----MSCYTNTI | FTSMFYFE | SEGVAVS | SLTFSGK | 30 | |
| RPRC010481-PA | ----- | ----- | ----- | ----- | 0 | |
| rp_asb-85900 | -----KCANPEPMTNFKIDN | FTKGANV | --- | ITHYKFGTD | 31 | |
| | | | | | | |
| RPRC000114-PA | TLQD-----IDKNCTKFLHKKTNDGKIREVFSNY--NPNGGTYSYDI | | | | 94 | |
| rp_asb-48123 | ----- | | | | 0 | |
| BAP-1 | TLEV-----GDKNCTKFLHQKTADGKIKEVFSNY--NPNAKTYSYDI | | | | 94 | |
| BAP-2 | TLEV-----GDKNCTKFLHQKTADGKIKEVFSNY--NPNAKTYSYDI | | | | 74 | |
| BAP-3 | TLEV-----GDKNCTKFLHQKTADGKIKEVFSNY--NPNAKTYSYDI | | | | 72 | |
| rp_asb-101226 | GFANGTGHNDISTYKNFGHHFERIHSCSVGI-KS-NNSPLSLIFFLILPIEKKD | NSYYI | | | 88 | |
| RPRC010481-PA | ----- | | | EA | 7 | |
| rp_asb-85900 | TLNS-----DDKNCTKLLQKK-QNDNIKEVFSIY--SPTTQTYSYYL | | | | 70 | |
| | | | | | | |
| RPRC000114-PA | SFASVKTFDGNNGKYTAF | NVIVNQDGTIKIDRRK | QVSYIDTDYSK | TSVW | YVCDPSA--PE | 152 |
| rp_asb-48123 | ----AAFEGND | NVIVNQDGTIKIDERT | KVSYIDTDYSK | | HVCDPSA--PD | 53 |
| BAP-1 | SFAKVSDFDGNNG | NVIVEKDGRKIDERT | QVSYIDTDYSK | | HVCDPAA--PD | 152 |
| BAP-2 | SFAKVSDFDGNNG | NVIVEKDGRKIDERT | QVSYIDTDYSK | | HVCDPAA--PD | 132 |
| BAP-3 | SFAKVSDFDGNNG | NVIVEKDGRKIDERT | QVSYIDTDYSK | | HVCDPAA--PD | 130 |
| rp_asb-101226 | SLSNNTSFYETY | HIQVNDKGTKIEERS | SVTYLDDTDYSS | | YVCCERMKDGKL | 148 |
| RPRC010481-PA | SFSKSSFDTTY | HIQVDKVGKELEHS | TVTYLDDTDYDS | | YVCGEIME-NL | 66 |
| rp_asb-85900 | SFSKTTSEVEDY | HIQVDKDGKELKHS | TVTYLDDTDYDS | | YVCGEIME-NL | 129 |
| | * | *****: : * : : : : * | *: * : * : * : * | ** | ** | |
| | | | | | | |
| RPRC000114-PA | YYLYAVQSRNENIN-GVKDKVETALGKVNLLKDLFDATTL-SCKYDEDTLKKLLDRSY | | | | 210 | |
| rp_asb-48123 | YYLYAVQSRKENVN-DVKDKVEAALGKVLKLSGLFDATTL--SCKYDEDTLKNLLAQQF | | | | 110 | |
| BAP-1 | YYLYAVQSRTEENVKEDVKSKEVAALGKVLKLSGLFDATTLGNKCQYDDETLQKLLKQSF | | | | 212 | |
| BAP-2 | YYLYAVQSRTEENVKEDVKSKEVAALGKVLKLSGLFDATTLGNKCQYDDETLQKLLKQSF | | | | 192 | |
| BAP-3 | YYXYAVQSRTEENVKEDVKSKEVAALGKVLKLSGLFDATTLGNKCQYDDETLQKLLKQSF | | | | 190 | |
| rp_asb-101226 | FSLYAVQSRREKTLSENIKTKVNSVLQSLNLDKLSIKD--GVCKFDDSTLNTLLSQTK | | | | 206 | |
| RPRC010481-PA | FSLYAVQSRSQTLNQDVETKVKLSALNGINLKDLSIKD--FGCKYDDSTL----- | | | | 116 | |
| rp_asb-85900 | FSLYAVQSRNKTLDSNVETKVKSVLKSINLNDKLSSTKR--VNCFEFDSTLNELLSKPF | | | | 187 | |
| | : ***** : : : : : ** : : * : : : * : : * : : * : : * | | | | | |
| | | | | | | |
| RPRC000114-PA | PEYEKE----- | | | | 216 | |
| rp_asb-48123 | SEYEK----- | | | | 115 | |
| BAP-1 | PNYEK----- | | | | 217 | |
| BAP-2 | PNYEK----- | | | | 197 | |
| BAP-3 | PNYEK----- | | | | 195 | |
| rp_asb-101226 | KEFKTQEKQCGFMKIGCF | | | | 224 | |
| RPRC010481-PA | ----- | | | | 116 | |
| rp_asb-85900 | TYET-EG----- | | | | 193 | |

Figura 7. Alinhamento das Proteínas Biogênicas de Ligação a Amina de *Rhodnius prolixus*. As sequências de aminoácidos foram alinhadas com o programa CLUSTAL-Omega, usando os parâmetros pré-definidos. Em cinza está marcada a região do domínio para a maioria das proteínas e em vermelho os sítios de ligação (ambos preditos pelo CDD).

4.7. Estrutura, evolução e expressão das proteínas Inibidoras de agregação plaquetária (PAI), Triabinas, Pallidipinas e Procalinas no transcriptoma de *R. prolixus*

Na análise evolutiva com todas as lipocalinas (Figura 3) foi possível identificar um grupo bem distinto de PAIs. No entanto, as triabinas, pallidipinas e procalinas (além da PAI-5) não ficaram bem resolvidas. Apesar de terem recebido diferentes nomes, os estudos funcionais sugerem que tanto triabinas quanto pallidipinas e PAIs são proteínas que atuam como inibidores da coagulação, seja pela inibição de trombina – no caso das triabinas –, ou de agregação plaquetária – no caso das PAIs e pallidipinas. Já as procalinas, foram identificadas apenas como fatores alergênicos, mas as análises iniciais do presente trabalho (Figura 3) sugerem que as procalinas são evolutivamente relacionadas às triabinas e PAIs. Sendo assim, devido à similaridade funcional e evolutiva, decidiu-se por analisar estas proteínas em conjunto. Na construção da árvore filogenética específica para esses grupos foram incluídas as sequências de uma triabina e uma pallidipina de *Triatoma pallidipennis* com o objetivo de analisar as relações de ortologia

Como esperado pelas análises anteriores, a árvore filogenética específica das PAIs, triabinas, pallidipinas e procalinas não resolve bem alguns dos genes estudados (Figura 8). Foi possível identificar: *i*) três grupos de PAI com a maioria das PAIs já descritas (PAI-1 a PAI-7, com exceção a PAI-5) e mais dois genes novos, RPRC000446 e rp_asb104785 (ambos já haviam sido identificados como possíveis PAIs pelo BLAST recíproco); *ii*) dois grupos de pallidipinas contendo as duas pallidipinas já descritas (Pallidipina-1 e Pallidipina-2) e dois novos genes, RPRC002137 e rp_asb-73726 (um transcrito originado do mesmo locus que a Pallidipina-1); e *iii*) apenas um ramo definido contendo Triabina-1 e Triabina-3. As duas procalinas descritas não possuem função definida e a Procalina-1 se mostrou evolutivamente mais próxima a PAI-5 (grupo Indefinido II da Figura 8), enquanto a Procalina-2 é mais próxima da Triabina-4 (grupo Indefinido I). A maioria das triabinas putativas identificadas no transcriptoma (rp_asb-30450, rp_asb-33413, RPRC000603, RPRC002646 e RPRC002647) se agruparam em dois ramos (grupo Indefinido III e IV da Figura 8), mas sem nenhuma triabina já caracterizada nos mesmos, sendo estes ramos evolutivamente equidistantes das Triabinas-1 e 3 e da Pallidipina-1. Por fim, o ramo caracterizado como Indefinido II (Figura 8) foi agrupado com bootstrap de 97 e nele está duas triabinas putativas (rp_asb-105577 e RPRC000275) e três lipocalinas salivares já caracterizadas, a Triabina-2, a

PAI-5 e a Procalina-1 (sendo que esta última foi identificada apenas como proteína alergênica).

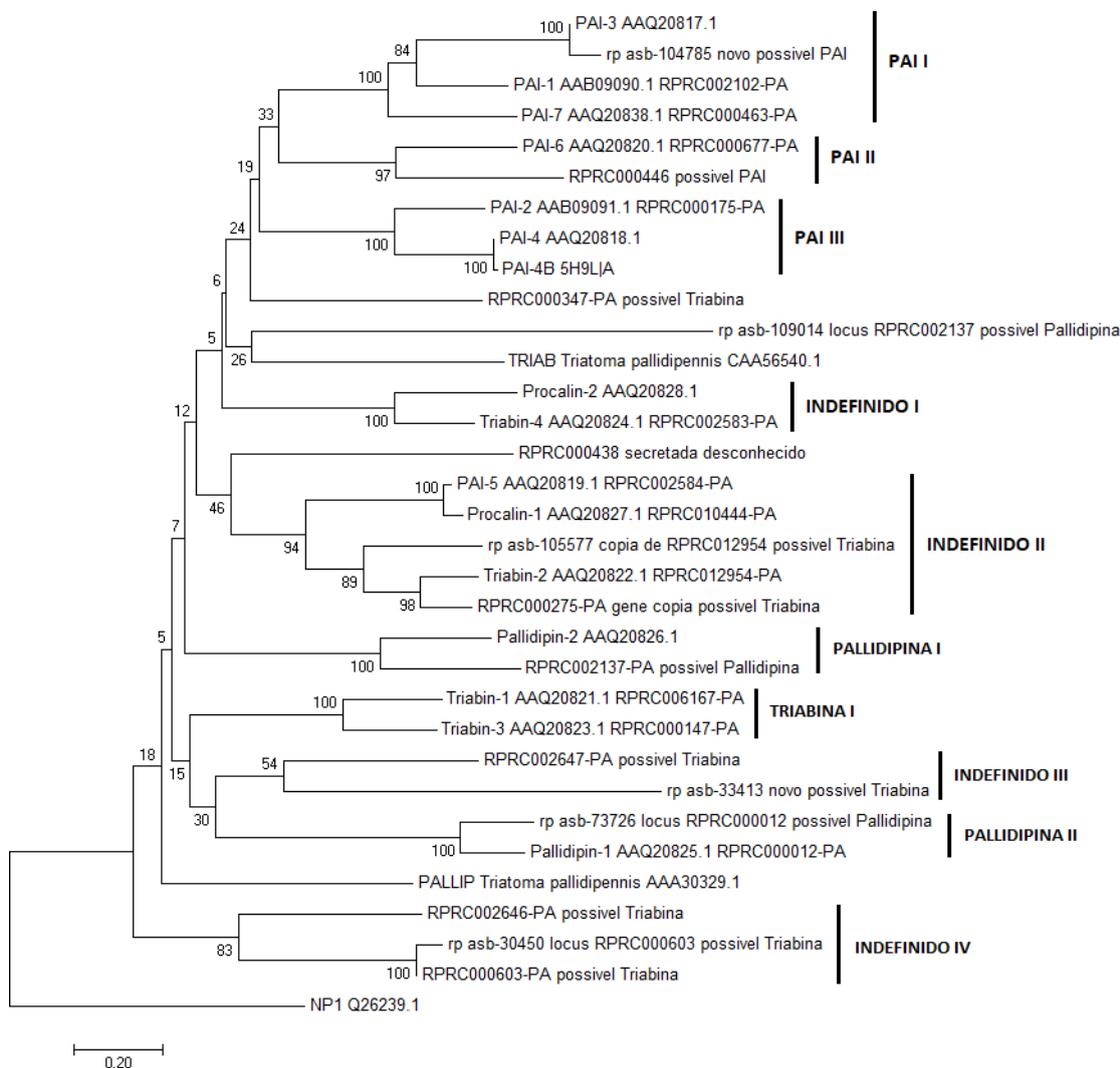


Figura 8. Relações evolutivas das Inibidoras de Agregação Plaquetária (PAIs), Triabinas, Pallidipinas e Procalinas de *R. prolixus*. A história evolucionária foi inferida usando o método de Neighbor-Joining, e o teste de bootstrap a partir de 10000 réplicas. Os números de acesso das sequências estão demonstrados em cada ramo da árvore filogenética. Uma Nitroforina foi usada como grupo externo.

A mesma análise foi realizada utilizando máximo-verossimilhança e máximo-parcimônia (ANEXOS D e E respectivamente), mas em ambos os casos a árvore obtida apresentou uma resolução pior que a análise por Neighbour-Joining. Assim, decidiu-se por analisar mais profundamente, tanto por BLAST recíproco (apenas para os genes sem função caracterizada, incluindo as procalinas), quanto pelo alinhamento e análise de domínios, as contradições encontradas na análise evolutiva.

O BLAST recíproco das proteínas indefinidas (Tabela 17) confirma que RPRC000275 e rp_asb-105577 são de fato mais similares a Triabina-2, enquanto a Procalina-1 é praticamente idêntica a PAI-5. Por outro lado, o grupo de triabinas isoladas (grupo Indefinido III e IV da Figura 8) não se mostrou similar a nenhuma lipocalina de *R. prolixus*, sendo que as proteínas retornadas mais próximas foram triabinas de outros reduviídeos, como o predador *Pristhesancus plagipennis*, e os barbeiros *Triatoma infestans* e *T. dimidiata*.

Tabela 17. Resultado do BLAST recíproco para lipocalinas indefinidas.

| Grupo (Fig. 8) | Gene | Resultado BLAST | | |
|----------------|---------------|---|--------|-----------|
| | | Proteína | ID% | Cobertura |
| Indefinido II | RPRC000275 | Triabina-2 (AAQ20822.1) | 77,97% | 98% |
| | rp_asb-105577 | Triabina-2 (AAQ20822.1) | 59,77% | 89% |
| | Procalina-1 | PAI-5 (AAQ20819) | 100% | 91% |
| Indefinido III | RPRC002647 | Triabina (BAI50840.1) <i>Triatoma dimidiata</i> | 32,49% | 97% |
| | rp_asb-33413 | Triabina (AQM58444.1) <i>Pristhesancus plagipennis</i> | 46,3% | 60% |
| Indefinido IV | RPRC000603 | Triabina (AQM58444.1) <i>Pristhesancus plagipennis</i> | 39,7% | 97% |
| | RPRC002646 | Triabina (AQM58444.1) <i>Pristhesancus plagipennis</i> | 46,3% | 94% |
| | rp_asb-30450 | Triabina (ABR27959.1) de <i>Triatoma infestans</i> | 36,14% | 96% |

Quando analisada a conservação das proteínas PAI-5, Procalina-1, Triabina-2 e as novas proteínas RPRC000275 e rp_asb-105577 do grupo Indefinido II (Figura 9), foi observada uma alta conservação dos sítios ativos, tanto em sua posição quanto no aminoácido, entre as proteínas analisadas (indicado em preto na Figura 9). Todos os 23 segmentos do sítio ativo (100%) de PAI-5 e Triabina-2 são localizados na mesma posição. E 17 dos 23 (73,9%) segmentos do sítio ativo apresentaram resíduos conservados entre as proteínas analisadas. Nos sítios de variação, a proteína rp_asb-105577 possui um resíduo idêntico a Triabina (a.a. 64) e um idêntico a PAI (aa 85). Já a RPRC000275 possui dois resíduos idênticos a Triabina (aa 60 e 140) e um idêntico a PAI (aa 83).

| | | |
|---------------|---|-----|
| Pallidipin-1 | MNTTFAVIFLGLVVTNAAS-----CQKKEPKDNFDS--NKYFSAKLEYVQRVSEG--- | 49 |
| Triabin-1 | MKTIIVALIFLGLTLAHLSS-----SKKCEAMKNIDS--QRFFSGT--FVAHAKNG--- | 48 |
| RPRC002647-PA | ---NTVLIILGTIAFAFAEQPPTVQECNLNTAKANFEP--QNYKGV--YLSYIKYA--- | 51 |
| RPRC000603-PA | --MIIVLAIFAILTFATADSQIKPKKECLKVPVKQNFEP--QKYRGN--FLFNKREEGQ | 55 |
| rp_asb-30450 | MKMIIVLAIFAILTFATADSQIKPKKECLKVPVKQNFEP--QKYRGN--FLFNKREEGQ | 57 |
| RPRC002646-PA | MKIIFSLTFVAIALASAAATVPDESGCQSVVGKETYEP--NKYHKG--YLTHAQKT-PS | 56 |
| rp_asb-33413 | MKRNLQYPYR-----LLSATNLYSSLRYRLVGHT--FKVHHSKG-DH | 40 |
| . | | |
| Pallidipin-1 | PKQTVCIKFDKFRGSDGKVTLS--NYDYVGSQV-NYHSCNGTQRSNNKGFSEFKCQQTS | 107 |
| Triabin-1 | SSTILCREITTKNGTVMKSDTKGKFKQGM-KEYTHCCKE--KKGKPFKCTREE | 104 |
| RPRC002647-PA | DPTGICRVSKDITLSCGSVQKKTGYTESGKGTDFFHNCNGTLKSGGA-IFSHCEQSG | 110 |
| RPRC000603-PA | SNTTVCQESKSRMLEGTISHVIYAYSVDLK-PEFIQINCTGNVKNTQE-KILFQCTLR- | 112 |
| rp_asb-30450 | SNTTVCQESKSRMLEGTISHVIYAYSVDV----- | 86 |
| RPRC002646-PA | SPTDVCRESKNEVLEIGSVIHNIYAYS DKAT-QQFYLLECTTNLKDVKDGTSVLNCKST- | 114 |
| rp_asb-33413 | TVL----KQKQIKETEVIGSDAFGIETEAAT-QQFYLLECTTNLKDVKDGTSVLNCKST- | 93 |
| . | | |
| Pallidipin-1 | DKS--SATPNFQDEMTVMETDYNFSSVSRCVRS--GGFSSGSLVLSKTDAGNGD-SSEL | 163 |
| Triabin-1 | ADRSIRNNEKEDTVMETDYNSEFVVCT--KK--EGKEENVLEINKNKADDFPSA--A | 158 |
| RPRC002647-PA | NEG--GEPKIFEMDGKVLLETDYNNESVYICYKS--GENLKAFLVASRHKGGEVTDSDKI | 167 |
| RPRC000603-PA | -KE--EVTENFQMEGTVAESDYENFVYFICSKE----GDGFVLSRKEDSGPT-DPRV | 163 |
| rp_asb-30450 | ----- | 86 |
| RPRC002646-PA | -KD--GKVNEFOLTATVLETDYEHFTVLYVCGK-LEGKNYGFVFLNRDKNGEPT-DPKI | 169 |
| rp_asb-33413 | -KD--GKVNEFOLTATVLETDYEHFTVLYVCKWESGSGKGFVFLNRDKNGEPT-DPKI | 149 |
| . | | |
| Pallidipin-1 | KKYLEEHS DITFE--HERSKANCTVSTIKNK- | 192 |
| Triabin-1 | KSTLE-KAELKSENLNTRKLYDCK | 181 |
| RPRC002647-PA | AETLKTH-GQYLDKFVSNKNVICKDHPDF | 195 |
| RPRC000603-PA | TETLKKY-GMNL EEFISRKNVKCKSHPD | 190 |
| rp_asb-30450 | ----- | 86 |
| RPRC002646-PA | AETLKKH-GLDLSTFTSRKMFTAKTILAL--- | 197 |
| rp_asb-33413 | AETLKKH-GLDLSTFTSRKNVHCKDHPSTVSN | 180 |

Figura 10. Alinhamento e análise de domínio das Triabinas de função hipotética em *R. prolixus* – grupo Indefinido III e IV. As triabinas hipotéticas (grupo Indefinido III e IV, Fig. 8) foram alinhadas com a Pallidipina-1 e a Triabina-1 de *R. prolixus*. Em cinza estão as regiões de domínios identificadas na Conserved Domains Database. Em amarelo estão os sítios ativos identificados para a Pallidipina-1 e os mesmos sítios ativos conservados nas triabinas hipotéticas. Em vermelho estão os sítios ativos da Triabina-1 e os mesmos sítios ativos conservados nas triabinas hipotéticas. Em preto estão os sítios ativos conservados na Pallidipina-1 e na Triabina-1.

Nossas observações sugerem que triabinas, pallidipinas, PAIs e procalinas são lipocalinas com alto grau de conservação, tanto de domínios quanto de sítios ativos, o que será discutido com mais profundidade mais adiante.

Devido a sua similaridade evolutiva e conservação de aminoácidos, decidiu-se por analisar a expressão de PAI, Triabinas, Pallidipinas e Procalinas de forma conjunta (Tabela 18). Como esperado, a grande maioria das PAI, Triabinas, Pallidipinas e Procalinas foram detectadas apenas na amostra de corpo inteiro das formas imaturas do inseto (inclui-se aqui todas as proteínas que já tiveram suas funções caracterizadas anteriormente). A isoforma da Pallidipina-1, a Pallidipina-1-PB, foi detectada tanto em ovário quanto em testículo. A maior exceção parece ser das Triabinas que formaram o Grupo Indefinido II na filogenia, sendo que quatro das cinco identificadas foram

detectadas em outros tecidos além do corpo inteiro das formas imaturas do inseto (NI). A Triabina RPRC000603-PA foi encontrada em corpo gorduroso e nos tecidos do aparelho digestivo (intestino anterior, intestino posterior e reto). Já sua isoforma, a RPRC000603-PB foi encontrada apenas no corpo gorduroso (além do corpo inteiro de ninfas), assim como a Triabina rp_asb-33413. A Triabina RPRC002646-PA foi detectada em todos os tecidos analisados, sendo que aparenta ter expressão mais elevada no corpo gorduroso e no corpo inteiro das formas imaturas (que inclui as glândulas salivares).

Tabela 18. PAI, Pallipidinas, Procalinas e Triabinas expressas em diferentes tecidos de *Rhodnius prolixus*

| Gene | Grupo | Função Predita | CG | TM | IA | IP | Rt | OV | Te | NI |
|---------------------------------|----------------|----------------|-------|------|------|------|------|------|------|-------|
| PAI-5 (RPRC002584) | Indefinido-II | PAI | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 36.52 |
| rp_asb-104785 | PAI-I | PAI | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.73 |
| RPRC000446 | PAI-I | PAI | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 2.09 |
| Pallidipina 1-PA (RPRC000012) | Pallidipina-I | Pallidipina | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 2.06 |
| Pallidipina 1-PB (rp_asb-73726) | Pallidipina-I | Pallidipina | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 2.41 | 5.40 |
| RPRC002137-PA | Pallidipina-II | Pallidipina | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 4.99 |
| Procalina-1 (RPRC010444) | Indefinido-II | Procalina | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 5.18 |
| Triabina-1 (RPRC000147-PA) | Triabina-I | Triabina | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 14.30 |
| Triabina-4 (RPRC002583) | Indefinido-I | Triabina | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.58 |
| Triabina-2 (RPRC012954) | Indefinido-II | Triabina | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 2.84 |
| RPRC000275 | Indefinido-II | Triabina | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 2.57 |
| rp_asb-105577 | Indefinido-II | Triabina | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.87 |
| RPRC002647-PA | Indefinido-III | Triabina | 3.56 | 0.00 | 0.00 | 0.00 | 0.00 | 1.47 | 0.00 | 3.66 |
| rp_asb-33413 | Indefinido-III | Triabina | 0.96 | 0.00 | 0.81 | 0.00 | 0.00 | 0.00 | 0.00 | 2.98 |
| RPRC000603-PA | Indefinido-IV | Triabina | 0.53 | 0.00 | 0.44 | 0.47 | 0.40 | 0.65 | 0.00 | 12.12 |
| RPRC000603-PB (rp_asb-30450) | Indefinido-IV | Triabina | 1.23 | 0.00 | 1.04 | 1.11 | 0.00 | 0.00 | 0.00 | 0.95 |
| RPRC002646-PA | Indefinido-IV | Triabina | 21.74 | 5.61 | 0.80 | 4.24 | 1.43 | 1.17 | 0.89 | 23.51 |

CG – Corpo Gorduroso; TM – Túbulo de Malpighi; IA – Intestino Anterior; IP – Intestino Posterior; Rt – Reto; OV – Ovário; Te – Testículo; NI – Corpo Inteiro de formas imaturas (ovos e ninfas). Em verde significa transcrição não detectada no tecido pelo transcriptoma, em vermelho significa transcrição detectada (em Reads por Kilobase – RPK).

4.8. Estrutura, evolução e expressão das Apolipoproteínas no transcriptoma de *R. prolixus*

As anotações e análises evolutivas iniciais realizadas anteriormente (itens 4.3 e 4.4 do presente trabalho) identificaram um total de 10 transcritos com similaridade a apolipoproteínas já conhecidas em outros insetos (o transcrito rp_asb-105382 é um pseudogene). Em *Rhodnius prolixus*, é a primeira vez que lipocalinas desse grupo funcional são encontradas. Como visto na Tabela 13, dos 10 transcritos identificados, 6 já estavam anotados no proteoma (RPRC003587-PA, RPRC000014-PA, RPRC015421-PA, RPRC000339 e RPRC004463) sendo que os transcritos rp_asb-47 e rp_asb-48 mapeiam a mesma proteína (RPRC015421-PA) com 100% de identidade. Na primeira análise evolutiva (Figura 3), oito das 11 apolipoproteínas se agruparam em um único ramo, porém com baixo suporte (bootstrap = 34). A exceção foram as RPRC015421-PA e rp_asb-47. Nesta análise inicial não foi utilizada nenhuma outra apolipoproteína caracterizada em insetos, por isso foi construída uma nova filogenia específica das apolipoproteínas (Figura 11) incluindo as apolipoproteínas do percevejo de cama *Cimex lectularis* (XP_014244096.1) e do hemíptero fitófago *Halyomorpha halys* (XP_014288060.2), que estão disponíveis na base de dados REF-SEQ (sequências referência, com evidência de transcrição).

A árvore filogenética das apolipoproteínas formou dois ramos distintos com bootstrap acima de 80, incluindo um ramo bem definido (bootstrap = 100) com as apolipoproteínas de *C. lectularius*, *H. halys* e a RPRC10099 de *R. prolixus*. Similar a estas proteínas está a RPRC004179. Em um segundo ramo (bootstrap = 89) ficaram agrupadas as proteínas RPRC003587, RPRC000014, RPRC015421, rp_asb-47 e rp_asb_6771. Como visto anteriormente, já foi apontado que a rp_asb-47 é transcrita do mesmo locus da RPRC015421 (também vista como rp-asb-48), sugerindo isoformas. Embora não tenha sido observado um ramo coeso das apolipoproteínas, todas se distinguiram bem das nitroforinas e BAPs, que eram os grupos funcionais mais próximos conforme sugerido na Figura 3.

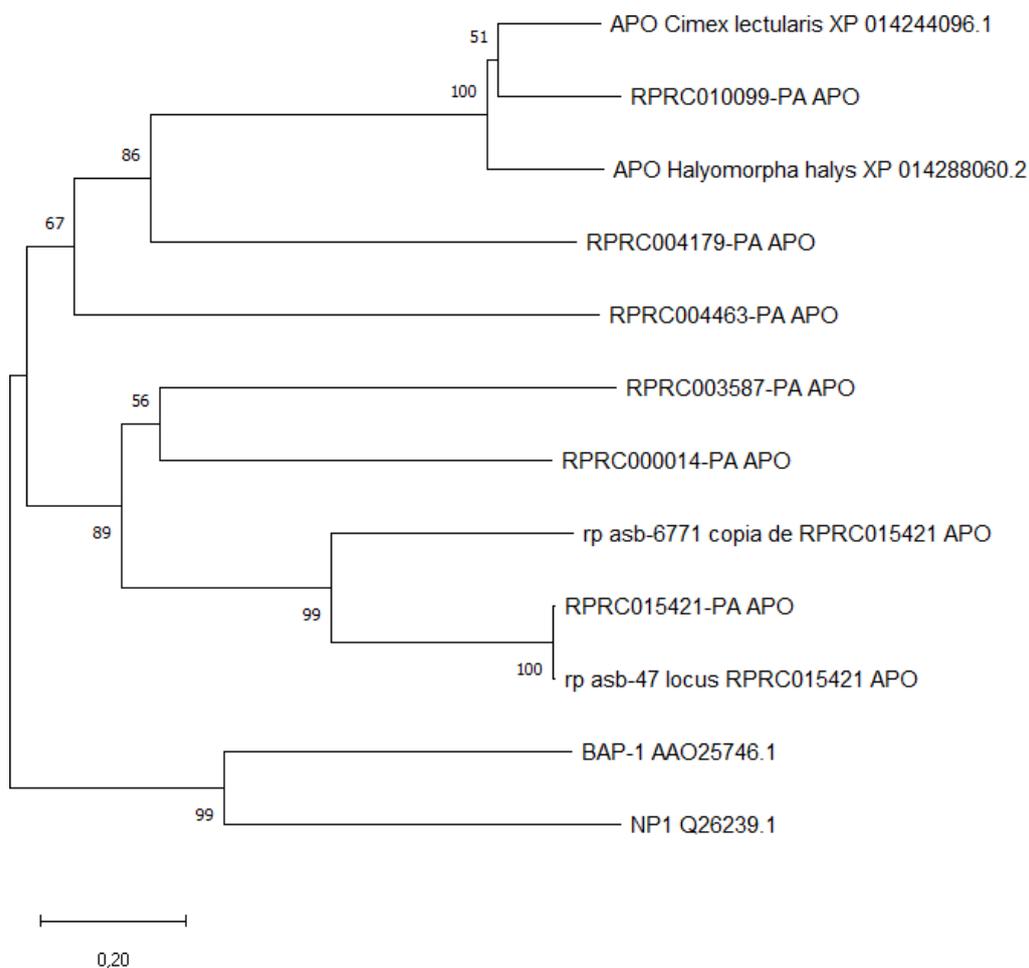


Figura 11. Relações evolutivas das Apolipoproteínas de *R. prolixus*. A história evolucionária foi inferida usando o método de Neighbor-Joining, e o teste de bootstrap a partir de 10000 réplicas. Os números de acesso das sequências estão demonstrados em cada ramo da árvore filogenética. Uma BAP e uma nitroforina foram usadas como grupo externo.

Uma análise da conservação do domínio e sítios ativos das apolipoproteínas mostrou que a RPRC003587 é bastante diferente das outras apolipoproteínas (ANEXO F). De fato, a RPRC003587 possui apenas 114 aminoácidos, enquanto as demais apolipoproteínas possuem de 193 a 324 aminoácidos. Assim, um novo alinhamento foi realizado (Figura 12), excluindo-se esta proteína. Apesar de ter se observado vários resíduos conservados entre as proteínas (incluindo a proteína referência de *C. lectularius*), apenas três dos sítios ativos possuem aminoácidos conservados em todas as proteínas.

| | | |
|---------------|--|-----|
| APO_Cimex | -----MISSVQLV-VFLLVATSAAQRP-----AFGRCPTVE-PKEN | 35 |
| RPRC000014-PA | -----MAMAKLVLTATLAY----TCAAKNVGTHPIIPFKQGPCLPESIAMQK | 44 |
| rp_asb-6771 | -----MNSKLAYIFLI VCGLSFLQAQPNDRKYDRLRSYLPWKIGACLQPS-VVEN | 49 |
| RPRC015421-PA | -----MISKITFIFLSVSALSLSQAFLFGFGPFTPGDYKWPRLGTCLVPT-PVKN | 49 |
| rp_asb-47 | -----RTHRKMISKITFIFLSVSALSLSQAFLFGFGPFTPGDYKWPRLGTCLVPT-PVKN | 54 |
| RPRC004463-PA | -----MVREVLAVLSSAIIAII LGHTY-----HLGECPAVE-PMNG | 36 |
| RPRC004179-PA | -----MKELILI-ICTLTTAISRAQMP-----GFGWCPDKR-AMPG | 34 |
| RPRC010099-PA | MKLNWAATVSWGYQLLV-IAICFFSLSSGQRP-----AFGRCPDVT-PKKD | 45 |
| | : : : | |
| | : : * | |
| APO_Cimex | FDVEKTVGNVEIERSF-YLFESSLSCTKVNFTLYENS-TLKADVNYRVPWRGTSSVSS | 92 |
| RPRC000014-PA | FDPEKFTGTVEYERSSFESDITQEDGNCSGMKARIVRDS-VEILMYHYDSAL-----NK | 97 |
| rp_asb-6771 | FKPKEFKGHVYEMSSFGNFWQVDGICPTIEYNVTDAGEILELDYHFESSL-----GK | 103 |
| RPRC015421-PA | FQPEKFKGHVYVQNHCGNFEWILPGICPTTEYNVTQDGDILELDYHYIPMT-----GR | 103 |
| rp_asb-47 | FQPEKFKGHVYVQNHCGNFEWILPGICPTTEYNVTQDGDILELDYHYIPMT-----GR | 108 |
| RPRC004463-PA | FVMSKFLGVYAYAQKT-----STGSRLTYNFTVGAEPGEYLLQVSEHPVLGVASVDN | 90 |
| RPRC004179-PA | FDIDKFLGTVEAERYF-TVLEAGSCLARTNYTKAVDG-RLLVTEINRLTGIKRVLD | 91 |
| RPRC010099-PA | FDVSKFTGVEIERSF-YLFESSLSCTKLNFTLYSND-TMVADVNYRAPWRGTSSSTSE | 102 |
| | * .: : * * * . : | |
| APO_Cimex | Y-FVVDTKTP-----GELNVLASTLPSLIARMAPGSKYIVDTDYDDEALVYSCDT | 145 |
| RPRC000014-PA | YTSVIGEANLVEAKTARFSIVLTYD-----PILIQEFPIYVVDTDYENYAFYSCR | 149 |
| rp_asb-6771 | YVTIRGQSYTQFIKNDRAFPPTYK-----VLGGLLTFDLPTYIIGTDYETWAVVYTC | 158 |
| RPRC015421-PA | YKEKRAHSNTKLIKNGEGHLPPTFE-----FLRGLHPMDVPTFIIGTDYENWAVVYCKQ | 158 |
| rp_asb-47 | YKEKRAHSNTKLIKNGEGHLPPTFE-----FLRGLHPMDVPTFIIGTDYENWAVVYCKQ | 163 |
| RPRC004463-PA | KYHYTGELKAS-----SDIPAKMSVK-----FPLSLAGKASFTVEMTDYDTYAGVYTCQK | 140 |
| RPRC004179-PA | G-EIRNVPK-----GGVDSKISVK---YSTLFPFMETEYIIGTDYDYSYVWVWCSG | 139 |
| RPRC010099-PA | Y-IIKDVSKTP-----GELNVLASTLPALIARMAPGSKYIVDTDYIDYALVYSCDT | 155 |
| | : : . : . : * * * : * : * | |
| APO_Cimex | LR---FVHADFIWVLGRGSDISVEARTVVYSTLDKKNINRDRLLLTNNKDC----- | 193 |
| RPRC000014-PA | VPGSSSLNDEMAWTFTRKRGDA-SNDGSVDRILKLSGLDPAKFARYEEMTCVDEPTL-- | 206 |
| rp_asb-6771 | ML---FLKSEMSWILTRQRNST-DQLKPVLETLLKSHLSYGEYKPAVNVGCGPNEPHLTS | 214 |
| RPRC015421-PA | QQ---VLKFEESLILTRKRDE-EIWQPVLDTLKLNKLDYRYFKPTVNNGCAPNEPHL-- | 212 |
| rp_asb-47 | QQ---VLKFEESLILTRKRDE-EIWQPVLDTLKLNKLDYRYFKPTVNTGCAPNEPHL-- | 217 |
| RPRC004463-PA | LP---ASNRRSATILSRKSLDKQVVKIRSRSLSLFGVSPYDLSIIDQSKCPGANDDTNV | 197 |
| RPRC004179-PA | FG---PLVNTQTAWVMTRELERLPPGVTLQKAYAVLDKNNKISRTFFVRAEQECCNLGEALSEK | 197 |
| RPRC010099-PA | LR---LVHADFIWVLGRSKDISIDARTIISTLDKKNINRDRLLLSKTKDC----- | 203 |
| | : : * : . : . * | |

Figura 12. Alinhamento e análise de domínio das Apolipoproteínas de *R. prolixus*.

As sequências de aminoácidos foram alinhadas com o programa CLUSTAL-Omega, usando os parâmetros pré-definidos. Em cinza está marcada a região do domínio para a maioria das proteínas e em vermelho os sítios de ligação (ambos preditos pelo CDD).

Os dados de expressão gênica obtidos do transcriptoma mostraram que as oito apolipoproteínas são expressas em mais de um tecido do inseto adulto, além da amostra de corpo inteiro das formas imaturas (NI) (Tabela 19). Esse é um contraste em relação a maioria das lipocalinas estudadas até aqui, sendo que a apolipoproteína RPRC015421-PB apresenta níveis de RPK em intestino anterior, intestino posterior, reto e testículo na ordem de quase 10x maior que as demais apolipoproteínas.

Tabela 19. Apolipoproteínas expressas em diferentes tecidos de *R. prolixus*.

| Transcrito | CG | TM | IA | IP | Rt | OV | Te | NI |
|---|-------|-------|---------|--------|---------|------|---------|-------|
| RPRC010099-PA | 0 | 0 | 3.17 | 0.30 | 0 | 0 | 1.61 | 1.00 |
| RPRC004179-PA | 0 | 3.02 | 0 | 0 | 0 | 0 | 0 | 11.31 |
| RPRC004463-PA | 47.22 | 56.58 | 1.59 | 1.27 | 6.08 | 6.41 | 0 | 12.25 |
| RPRC003587-PA | 0 | 0 | 12.21 | 12.98 | 0.84 | 0 | 0 | 0.68 |
| RPRC000014-PA | 0 | 0 | 120.62 | 50.53 | 34.22 | 0.76 | 25.06 | 3.62 |
| RPRC015421-PA (rp_asb-48) | 0 | 0 | 135.53 | 89.94 | 77.35 | 0 | 89.14 | 3.07 |
| RPRC015421-PB (rp_asb-47) | 0 | 0 | 1318.57 | 992.99 | 1101.66 | 0 | 1041.73 | 26.92 |
| rp_asb-6771 (cópia de RPRC015421-PA) | 0 | 4.77 | 159.58 | 156.71 | 170.11 | 0 | 97.03 | 22.36 |

CG – Corpo Gorduroso; TM – Túbulo de Malpighi; IA – Intestino Anterior; IP – Intestino Posterior; Rt – Reto; OV – Ovário; Te – Testículo; NI – Corpo Inteiro de formas imaturas (ovos e ninfas). Em verde significa transcrição não detectada no tecido pelo transcriptoma, em vermelho significa transcrição detectada (em Reads por Kilobase – RPK).

4.9. Genes não resolvidos

Das lipocalinas estudadas no presente trabalho, duas delas não foram classificadas em nenhum grupo funcional pelas análises iniciais e, por isso, foram chamadas de “inespecíficas” (Tabela 14). São elas os transcritos rp_asb-105519 e rp_asb-61810. Os resultados do mapeamento sugerem que o transcrito rp_asb-61810 possa ser uma isoforma da lipocalina RPRC000418-PA – inicialmente identificada como sendo uma triabina – já que há uma equivalência entre as coordenadas de suas sequências na região gênica a qual elas pertencem. Já o transcrito rp_asb-105519 foi identificado como sendo a proteína RPRC000438. Porém, as análises evolutivas não incluíram estes genes em nenhum grupo funcional, embora RPRC000438 pareça estar relacionada as inibidoras de agregação plaquetária (PAI, triabinas e pallidipinas).

Para alguns dos transcritos que, segundo as análises iniciais, apresentaram similaridade com lipocalinas, as análises evolutivas também não foram conclusivas. Como mostrado na árvore filogenética inicial (ANEXO C), 7 transcritos não se relacionaram a nenhuma lipocalina já descrita. São eles: RPRC010695, RPRC006975, RPRC000418, RPRC000163, rp_asb-89583, rp_asb-65689 e rp_asb-110283. Claramente essas sequências não agruparam de forma coesa a nenhuma lipocalina conhecida e, por isso, não foi possível confirmar se são de fato lipocalinas, tão pouco indicar uma função putativa para elas. Da mesma forma, como já dito anteriormente, não podemos descartar que estes transcritos tenham a função de uma lipocalina.

5. DISCUSSÃO

Desde que *Rhodnius prolixus* se consolidou como uma espécie modelo para os estudos de fisiologia de insetos hematófagos (WIGGLESWORTH, 1976; CHIANG *et al.*, 2013), muito se pergunta a respeito do papel biológico de algumas moléculas encontradas nesse organismo, principalmente quando se trata de moléculas ligadas à fisiologia alimentar. Isso se dá devido à importância da hematofagia no âmbito da transmissão de patógenos, uma vez que é no momento do repasto sanguíneo que vetores hematófagos como o *Rhodnius prolixus* estão mais intimamente relacionados aos seus hospedeiros e os expõem à infecção pela via vetorial.

Nesse cenário, diversos estudos demonstraram a importância da saliva dos insetos hematófagos para o sucesso do repasto sanguíneo e, conseqüentemente, para um bom desempenho vetorial (SILVA, 2009; TIRLONI *et al.*, 2021). É nesse contexto que surgiram os primeiros estudos descrevendo a importância e a função de lipocalinas salivares em *R. prolixus* e outras espécies de insetos hematófagos (RIBEIRO & GARCIA, 1981). Tais estudos deram início a uma sequência de outros que, classicamente, passaram a abordar a família das lipocalinas no âmbito da fisiologia alimentar. As lipocalinas salivares são provavelmente a família de proteínas mais bem estudada em *R. prolixus*, com 38 proteínas funcionalmente caracterizadas (e 24 sequenciadas antes do genoma), incluindo as Nitroforinas (CHAMPAGNE *et al.*, 1995; ANDERSEN *et al.*, 1998; ANDERSEN *et al.*, 2000; ANDERSEN *et al.*, 2004; ANDERSEN *et al.*, 2003; XU *et al.*, 2013; FRANCISCHETTI, 2000; RIBEIRO *et al.*, 2004; JABLONKA *et al.*, 2016). Apesar disso, a presença e função dessa família proteica em outros tecidos e processos fisiológicos distintos daqueles relacionados a alimentação foram negligenciados nos estudos sobre insetos hematófagos. O sequenciamento de RNA mensageiro de diferentes tecidos, estágios fisiológicos e de desenvolvimento de *R. prolixus* publicado em 2014 (Ribeiro *et al.*, 2014) levantou indícios de que muitas outras lipocalinas, expressas em diferentes tecidos, poderiam existir. Porém, desde sua publicação, não houve nenhum aprofundamento no transcriptoma dos outros tecidos e nem na diversidade de lipocalinas presentes. É neste contexto que surge o interesse, no presente trabalho, em revisar o conhecimento que se tem das lipocalinas de *R. prolixus*, agora com a incorporação dos dados do transcriptoma, genoma e proteoma (RIBEIRO *et al.*, 2014; MESQUITA *et al.*, 2015), com o objetivo de identificar novas lipocalinas, avaliar sua origem evolucionária e

funcional em relação as lipocalinas salivares já descritas, e levantar, mesmo que hipoteticamente, possíveis papéis de lipocalinas não salivares.

5.1. Uma revisão do transcriptoma de *Rhodnius prolixus*.

As análises aqui realizadas partiram dos resultados do sequenciamento de RNA de *R. prolixus* publicado por Ribeiro e colaboradores (2014). Como já havia um transcriptoma publicado a partir da biblioteca de ESTs de glândula salivar (RIBEIRO *et al.*, 2004), este transcriptoma se concentrou no sequenciamento de outros tecidos, sendo eles: corpo gorduroso, intestino anterior, intestino posterior, reto, túbulos de malpighi, ovário, testículo e corpo inteiro de ninfas e embriões. Dos 25.673 mil transcritos montados, observamos que, dentre as proteínas secretadas, 112 delas foram inicialmente identificadas como proteínas da família das lipocalinas. Elas foram encontradas em todos os tecidos do inseto que foram sequenciados e, em alguns casos, apresentando uma alta taxa de expressão. Ficou claro, portanto, que essa família proteica é abundante, diversa e amplamente distribuída no corpo do barbeiro.

Apesar de este transcriptoma ter gerado uma quantidade enorme de dados que ajudou em diversos estudos da fisiologia de barbeiros nos anos seguintes, alguns problemas foram observados: o primeiro foi o método de sequenciamento escolhido, o pirosequenciamento por sequenciadores Roche 454. Hoje se sabe que estes sequenciadores produzem uma grande quantidade de erros em regiões de homopolímeros de DNA, dificultando a montagem de genomas e transcriptomas (SCHLEBUSCH & ILLING, 2012). Além disso, os pirosequenciadores não produziam seqüências em volumes suficientes para uma posterior análise de expressão diferencial de genes. O segundo problema foi na escolha do método de montagem do transcriptoma. Neste caso, o transcriptoma foi montado com o mesmo algoritmo criado por Ribeiro para montagem de transcriptomas anteriores (RIBEIRO *et al.*, 2014). Porém, os transcriptomas anteriores foram sequenciados pelo método de Sanger, que produzia leituras longas (maiores de 700 nucleotídeos) e de alta qualidade. Além disso, até hoje não se sabe como o montador criado por Ribeiro lida com *splicing* alternativo. Assim, o transcriptoma de *R. prolixus* é um reflexo das escolhas feitas a sua época. O número de transcritos (25.673) muito superior ao número de proteínas anotadas no genoma (9.311) sugere erros de montagem, seja por erros de sequenciamento ou do montador, ou um grande número de isoformas não anotadas. Além disso, apesar de o transcriptoma nos dar uma ideia se um gene é mais ou menos expresso em um

determinado tecido, não é possível analisar isso de forma objetiva, com análises estatísticas de confiança, limitando sua análise a detecção ou não detecção de um transcrito em um determinado tecido. Por estes motivos, um dos processos cruciais dessa revisão, foi realizar uma análise minuciosa do transcriptoma (focado nas lipocalinas), realizando um cruzamento de dados através do mapeamento do transcriptoma com o genoma e o proteoma.

Todos os transcritos identificados inicialmente como lipocalinas foram usados em buscas por BLAST para mapeá-los ao genoma e/ou ao proteoma. Para isso, definimos como ponto de corte 98% de identidade e 90% de cobertura no alinhamento das sequências de nucleotídeos e proteínas. Este ponto de corte é mais conservador que os normalmente usados para mapeamento de transcritos em genomas. Por exemplo, o alinhador BWA (Borrows-Wheeler Aligner) tem como parâmetro permitir 4% de *mismatch* no alinhamento de *reads* ao genoma. Porém, considerando que as sequências de transcritos já foram montadas e erros foram “removidos” na montagem optamos por uma abordagem mais conservadora, diminuindo assim ambiguidades. Os resultados do mapeamento sugerem que muitos transcritos não alinham a nenhuma região do genoma ou do proteoma de inseto (foram ao todo 11 PAIs putativas, 12 nitroforinas putativas, 6 pallidipinas putativas e 7 triabinas putativas), sugerindo que tais transcritos possam ser artefatos de montagem. E apesar de nossa análise conservadora, não podemos descartar possíveis falhas no genoma e/ou proteoma e, como mostraremos no caso das nitroforinas, de fato existem falhas de montagem no genoma (gaps ou quebras de scaffolds) que impediram a anotação de proteínas muito bem caracterizadas e sequenciadas. Desta forma, os transcritos que não alinharam ao genoma ou ao proteoma foram descartados desta revisão, já que não temos como, neste momento, definir se os transcritos são reais ou se são artefatos de sequenciamento ou montagem. Para os transcritos estudados que foram alinhados ao genoma e ao proteoma ou apenas ao genoma (64 no total, sendo 11 PAI putativas, 8 apolipoproteínas putativas, 29 nitroforinas putativas, 6 pallidipinas putativas, 8 triabinas putativas e 2 inespecíficas), foi possível prosseguir com as análises estruturais e evolutivas que possibilitaram uma caracterização inicial dessas sequências.

5.2. Sobre as lipocalinas já conhecidas em *R. prolixus*.

Trinta e oito lipocalinas já haviam sido descritas em *R. prolixus*, sendo elas: dezoito nitroforinas (NP-1, NP-2, NP-3, NP-4, NP-5, NP-6, NP-7, RPRC000367-PA,

RPRC000380-PA, RPRC000077-PA, RPRC000250-PA, RPRC000537-PA, RPRC000999-PA, RPRC000313-PA, RPRC000034-PA, RPRC000498-PA, RPRC000058-PA, RPRC000284-PA e RPRC000061-PA); oito inibidoras de agregação plaquetárias (PAI-1, PAI-2, PAI-3, PAI-4, PAI-4B, PAI-5, PAI-6 e PAI-7); duas pallidipinas (Pallidipina-1 e Pallidipina-2); duas procalinas (Procalina-1 e Procalina-2); 3 proteínas biogênicas de ligação a amina (BAP-1, BAP-2 e BAP-3) e quatro triabinas (Triabina-1; Triabina-2; Triabina-3 e Triabina-4). Destas, apenas as nitroforinas NP-5 e NP-6 nunca foram sequenciadas.

As nitroforinas são, provavelmente, uma das proteínas salivares mais características de *R. prolixus*, e foram as primeiras lipocalinas salivares a serem identificadas e caracterizadas neste inseto (RIBEIRO & WALKER, 1994). Estas lipocalinas foram identificadas apenas em cimicídeos e em barbeiros da tribo Rhodinini até o momento. As nitroforinas possuem um grupamento heme que se liga a óxido nítrico (NO), um potente vasodilatador que é liberado na corrente sanguínea do hospedeiro vertebrado durante o repasto sanguíneo. Devido a sua importância para a fisiologia do inseto e para a história dos estudos de proteínas salivares de *R. prolixus*, seria de se esperar que todas as sete Nitroforinas conhecidas fossem identificadas e anotadas no genoma de *R. prolixus*. No entanto, apenas duas Nitroforinas foram identificadas, a NP-1 e a NP-2, respectivamente mapeadas aos genes RPRC000023 e RPRC000072. Esse achado levantou a pergunta sobre o que teria acontecido com as outras nitroforinas. Utilizando as sequências das cinco nitroforinas descritas anteriormente, realizamos uma série de buscas por similaridade (BLAST) no genoma e no proteoma. O primeiro fato que nos chamou a atenção é que os resultados sugerem que a proteína RPRC000023 não seria a NP-1, mas sim a NP-4. Já a NP-2 parece ser mesmo a RPRC000072. Assim, tanto a NP-1, NP-3 e NP-7 ficaram sem correspondência no proteoma. Porém, na procura pelo genoma, foi possível identificar sequências com quase 100% de identidade para a NP-1 e NP-7. Se as sequências estão no genoma, por que não foram anotadas no proteoma? Uma análise minuciosa do mapeamento mostrou que a NP-1 alinha com uma região contendo um *gap* no scaffold (ACPB03038429). Tais *gaps*, ou buracos, são formados durante a fase de scaffolding, quando sequências contíguas (contigs) são unidas “artificialmente” baseadas na informação de leituras-pares (mate-pairs). A junção é artificial, mas baseada em evidência biológica, na qual o montador não sabe a sequência da junção, mas sabe exatamente quantos nucleotídeos existem separando os contigs unidos, formando os

gaps. Assim, os *gaps* são preenchidos por um número definido de “N”, simbolizando que não sabemos os nucleotídeos daquela região. Infelizmente, na fase de identificação de genes, os programas de computador (anotadores) não têm como lidar com essa falta de informação e, por isso, genes cuja alguma parte ficam no *gap* geralmente não são anotados (ou são anotados com stop-codons prematuros). Esse seria o motivo mais provável para que a NP-1 não tenha sido corretamente identificada no genoma. Já a NP-7 também foi localizada no genoma, mas nesse caso a sequência do gene está quebrada em dois scaffolds (ACPB03033880 e KQ034096). Este é um caso similar ao descrito para a NP-1, porém aqui não havia evidência biológica para juntar as sequências. A única Nitroforina que não foi encontrada nem no genoma nem no proteoma foi a NP-3, o que sugere uma falha de amostragem ou de montagem que englobou o gene por completo.

Da mesma forma que observado para as nitroforinas, nossos resultados mostram situações similares para as outras lipocalinas. Das oito PAIs, apenas cinco (PAI-1, PAI-2, PAI-5, PAI-6 e PAI-7) foram identificadas no genoma. Somente a Pallidipina-1 e a Procalina-1 foram mapeadas. Por outro lado, todas as quatro Triabinas foram mapeadas a proteínas anotadas no genoma. Curiosamente, nenhuma das três BAPs foram identificadas entre as proteínas anotadas, mas todas elas mapearam no mesmo locus cromossômico, sugerindo que estas três lipocalinas sejam isoformas de um mesmo gene. Além disso, neste locus há uma proteína anotada com estrutura levemente diferente (RPRC000114) e que está identificada no VectorBase como sendo uma nitroforina.

O fato de que, de 36 proteínas muito bem caracterizadas, com sequências disponíveis, apenas 16 foram localizadas no genoma (e algumas anotadas com função equivocada), sugere que o genoma de *R. prolixus* ainda se encontra demasiadamente fragmentado. Tecnologias atuais de sequenciamento de leituras ultralongas, como o MinION (LU *et al.*, 2016) ou o PacBio (RHOADS & AU, 2015) poderiam ajudar a melhorar a montagem atual, criando sequências contíguas maiores e resolvendo os *gaps* existentes.

5.3. Novas lipocalinas de *Rhodnius prolixus*

Nossa primeira preocupação nessa revisão foi de localizar corretamente a origem destes transcritos, assim como sua integralidade. Assim, partimos de seis situações hipotéticas, sendo elas: *i*) Lipocalinas Conhecidas (transcritos que mapeiam com a

proteína anotada de uma das 36 lipocalinas já descritas e sequenciadas); *ii*) Lipocalinas novas já anotadas (transcritos que mapeiam com proteínas anotadas do genoma); *iii*) Lipocalinas novas não anotadas (transcritos que mapeiam no genoma mas não a proteínas anotadas); *iv*) Lipocalinas de mesmo locus (transcritos que mapeiam na mesma região de um transcrito anterior); *v*) Pseudogenes (transcritos que mapeiam no genoma mas não no proteoma, mas produzem stop-codons prematuros); e *vi*) Origem desconhecida (transcritos que não mapearam nem ao genoma nem ao proteoma). Dos 112 transcritos inicialmente identificados como possíveis lipocalinas, um total de 76 deles mapeou ao genoma e/ou ao proteoma, sendo eles: 29 nitroforinas (5 já conhecidas, 4 novas já anotadas, 5 novas não anotadas, 8 de mesmo locus e 7 pseudogenes), 11 PAI (5 já conhecidas, 2 novas não anotadas, 1 de mesmo locus e 3 pseudogenes), 6 pallidipinas (1 já conhecida, 1 nova já anotada, 2 de mesmo locus e 2 pseudogenes), 17 triabinas (4 já conhecidas, 5 novas já anotadas, 4 novas não anotadas, 2 de mesmo locus e 2 pseudogenes), 11 apolipoproteínas (7 novas já anotadas, 1 nova não anotada, 2 de mesmo locus e 1 pseudogene) e 2 inespecíficas (excluídas das análises posteriores por terem sequências muito curtas). Mais adiante, discutiremos em mais detalhes estes mapeamentos e anotações para cada uma das classes de lipocalinas, porém vale a pena aqui ressaltar alguns detalhes. Primeiramente que o mapeamento dos transcritos confirmou os resultados das análises das lipocalinas já conhecidas, em especial no caso da Nitroforina-1 e Nitroforina-4, em que, mais uma vez, o gene RPRC000023 (anotado como NP-1 no VectorBase) foi mais similar a NP-4, enquanto não encontramos transcritos similares a NP-1. O segundo ponto que nos chamou atenção foi o número de lipocalinas novas não anotadas. Muitas destas são bastante similares a genes já anotados e, por algum motivo, não foram anotados pelo projeto genoma de *R. prolixus*. Por fim, identificamos um grande número de transcritos (15 ao todo) que mapearam no mesmo locus que um gene já anotado. Em alguns casos, como será mostrado mais adiante, estes transcritos são quase idênticos a proteínas já anotadas, em outros são bastante diferentes. Ficamos tentados, então, a considerar que tais transcritos de mesmo locus seriam isoformas, produtos de *splicing* alternativo de um gene já descrito. É notório que não foram anotadas nenhuma proteína proveniente de *splicing* alternativo no genoma de *R. prolixus*, de modo que encontrar evidências de *splicing* alternativo aqui seria algo inédito (MESQUITA *et al.*, 2015). Porém, para comprovar que estes transcritos são fruto de *splicing* alternativo e não de erros de montagem ou até mesmo variação alélica (já que o DNA sequenciado de *R. prolixus* veio de uma colônia que não era 100%

homozigota), seriam necessários estudos de biologia molecular que não puderam ser escopo deste trabalho. Apesar disso, baseados em nossas análises evolutivas e no alinhamento genômico, e para facilitar a compreensão por parte da comunidade científica, decidimos adotar a nomenclatura de isoformas, e estes transcritos então receberam o nome do gene originalmente anotado no *locus* adicionando o sufixo PB, PC, etc., para cada uma das isoformas. Após essa decisão de normalização de nomenclatura, continuamos, portanto, com nossa filtragem focando agora mais especificamente na função predita destas lipocalinas.

A princípio, poder-se-ia argumentar que o resultado dos BLASTs (confirmados por BLAST recíproco), seria suficiente para definir que estes transcritos são definitivamente lipocalinas com suas respectivas funções. Porém, nem sempre os resultados de BLAST são confiáveis, em especial quando se tem sequências fragmentadas e bases de dados limitadas. É comum nas bases de dados a função de gene ser definida de forma hipotética, apenas com base em similaridade. No caso das lipocalinas, poucas das 26 já descritas têm função comprovada em ensaios biológicos (mesmo assim, as Procalinas foram identificadas apenas como alergênicas, não tendo sua função real definida), havendo muitas outras com função determinada apenas com base em similaridade de sequência. Já a limitação do tamanho das sequências dos transcritos pode alterar estatísticas do BLAST, como *e-value* e *score*. Alinhamentos inespecíficos (baixa identidade), porém longos, tendem a ser melhor classificados (menor chance de ocorrer ao acaso) que alinhamentos curtos com alta identidade (ALTSCHUL *et al.*, 1990). Assim, é preciso que os resultados do BLAST sejam confirmados por outras metodologias, como análise evolutiva (filogenia) ou análises de sintenia (conservação do mapa genético entre espécies comparadas) (PASSARGE *et al.*, 1999).

Quando analisamos a evolução de genes de um ponto de vista filogenético, buscamos responder a pergunta sobre a origem de um ou mais genes. Se o estudo for realizado entre genes de uma mesma espécie, buscamos identificar genes que possuem uma relação por duplicação, também chamados de parálogos (KOONIN, 2005). A hipótese mais aceita é que genes duplicados recentemente provavelmente ainda possuem a mesma função que sua cópia original. Na filogenia, podemos observar isso

quando genes formam grupos coesos (com valores de *bootstrap*³ acima de 90). Se este grupo, ou ramo filogenético, conter um ou mais genes com função caracterizada comprovada, então os genes do mesmo ramo provavelmente possuem uma função semelhante. É importante ressaltar que estudos de paralogia, ortologia (genes que se originaram de um único gene ancestral nos genomas comparados) e sintenia não substituem os estudos biológicos (experimentais) na determinação funcional de genes, mas são um bom ponto de partida para o estudo funcional de genes.

A primeira análise evolutiva realizada após o mapeamento dos transcritos estudados resultou na construção de uma árvore filogenética que permitiu uma visão mais ampla da história evolutiva das lipocalinas de *R. prolixus*. Apesar das divergências entre as sequências, a árvore foi capaz de sugerir agrupamentos funcionais, confirmando (ou refutando) as funções hipotéticas determinadas pelo BLAST. No caso das nitroforinas, todas as já descritas foram agrupadas em um único ramo da árvore. Neste mesmo ramo foram incluídos também os transcritos RPRC000380-PA, RPRC000367-PA e RPRC000077, o que sugere que estes poderiam ser nitroforinas até então desconhecidas. Outros agrupamentos bem delimitados foram evidenciados: um grupo de BAP e um de PAI. As apolipoproteínas putativas formaram um grupo de baixa coesão (bootstrap abaixo de 60). Já as triabinas, pallidipinas, procalinas e até mesmo uma PAI (PAI-5) ficaram distribuídas de uma forma mais indefinida, pois em mais de um caso, duas proteínas caracterizadas com funções diferentes acabaram se agrupando em um mesmo ramo filogenético. Nesse ponto, vale ressaltar o poder do uso de filogenias para a assertividade de relações evolutivas dos genes. Observamos que vários dos transcritos tidos como possíveis nitroforinas pelo BLAST se alinharam em um ramo muito bem definido das proteínas biogênicas de ligação a amina (BAP), sugerindo, com mais fidedignidade, que estes transcritos são na verdade BAPs. Mas não é de se estranhar que tais transcritos tenham sido confundidos com nitroforinas, já que a filogenia sugere que BAPs e nitroforinas tenham uma origem em comum. Entraremos em mais detalhes nessa discussão mais adiante, pois neste momento nosso principal objetivo foi identificar genes que claramente não se alinhavam e nem se agrupavam com nenhuma lipocalina já caracterizada. Ao todo 26 transcritos não tiveram um grupo funcional putativo sugerido pelo estudo evolutivo, o que demonstra a necessidade de estudos complementares para melhor caracterizar essas moléculas. Na realidade, não

³ Bootstrap é o valor estatístico indicando qual a porcentagem de vezes que um ramo filogenético foi formado em N simulações (ZHARKIKH & LI, 1995).

podemos descartar que estes genes sejam, ou não sejam, lipocalinas (é possível que sejam lipocalinas com funções ainda desconhecidas) e por conta desta ambiguidade é que decidimos não nos aprofundar sobre estes transcritos, pois, pela natureza deste trabalho, teríamos muito pouco a acrescentar sobre os mesmos.

Nas seções subsequentes vamos discutir os resultados das análises realizadas para cada grupo funcional de lipocalinas de *R. prolixus* e, ao final, apresentamos uma tabela (Tabela 20) com as lipocalinas revisadas de *R. prolixus*.

5.3.1. Novas Nitroforinas em *R. prolixus*

A partir das análises aqui apresentadas, cinco novas nitroforinas puderam ser identificadas no presente trabalho. São elas: rp_asb-48696, rp_asb-48698, rp_asb-1685, rp_asb-3232, rp_asb-101256 (anotadas no mesmo locus de nitroforinas já identificadas). As proteínas RPRC000380-PA, RPRC000367-PA e RPRC000077 não se relacionaram a nenhuma nitroforina já descrita, mostrando que são de fato nitroforinas inéditas como proposto por Mesquita e colaboradores (2015). Pela filogenia é possível observar que RPRC000380-PA e RPRC000367-PA estão mais próximas evolutivamente da Nitroforina-2. Na literatura, duas nitroforinas – NP-5 e NP-6 – foram descritas em 2003, mas ainda não haviam sido sequenciadas (MOREIRA *et al.*, 2003). É possível que essas duas sequências (RPRC000380-PA e RPRC000367-PA) correspondam as Nitroforinas 5 e 6.

Os estudos filogenéticos mostraram que NP-7 Q6PQK2.1 está evolutivamente mais próxima da NP-2, corroborando o que já foi descrito anteriormente na literatura. (ANDERSEN *et al.*, 2004). A árvore filogenética mostrou ainda a proximidade evolutiva entre as nitroforinas NP-1 e NP-4, o que também já havia sido observado em trabalhos anteriores (MONTFORT *et al.*, 2000). Nossas análises mostraram que a proteína RPRC000023-PA, anteriormente identificada como NP-1, está muito proximamente relacionada à Nitroforina 4 Q94734.1. É possível que, por NP-1 e NP-4 serem muito similares do ponto de vista evolutivo, a proteína RPRC000023-PA tenha sido erroneamente identificada como uma NP-1. Sugerimos, portanto, que a proteína RPRC000023-PA seja na verdade uma NP-4, e não uma NP-1 como está descrito no banco de dados.

Nós também encontramos três transcritos que se originam do mesmo locus que a NP-2 e outros dois transcritos que se originam do mesmo locus que a NP-4. Estes diferentes transcritos que se originam de um mesmo locus de um gene anotado

poderiam ser: *i*) erros de sequenciamento/montagem; *ii*) variantes alélicas ou populacionais (já que os insetos usados para o genoma e o transcriptoma são de populações diferentes); e *iii*) isoformas gênicas produzidas por *splicing* alternativo. Idealmente, para se descartar ou aceitar uma das alternativas possíveis, seriam necessários experimentos de sequenciamento e estudos de bioinformática para alinhamento dos diferentes transcritos ao genoma. Infelizmente não tivemos tempo hábil para nos aprofundar em nenhuma das análises, mas algumas análises preliminares (dados não mostrados) sugerem que os transcritos adicionais possam realmente ser produto de *splicing* alternativo, pois, aparentemente, adicionam exons na região C-terminal das proteínas. Assim, apesar de não termos evidências significativas de que tais transcritos sejam de fato isoformas de um mesmo gene, decidimos, para facilitar o entendimento (como explicado anteriormente), chamar estes transcritos pelo mesmo nome que o gene anotado originalmente, acrescentando o sufixo PB, PC, PD, etc., para os diferentes transcritos, como preconizado pelo GeneBank.

Esta discussão sobre nomenclatura nos leva então de volta a questão das nitroforinas NP-5 e NP-6. Apesar destas nitroforinas terem sido identificadas por HPLC, Moreira e colaboradores (2004) conseguiram sequenciar menos de 35 aminoácidos de cada proteína, que sequer foram depositadas no GeneBank. Um ano depois, em 2004 (RIBEIRO, *et al.* 2004) levantou a possibilidade de que NP-6 seria na realidade a NP-2 e que NP-5 talvez fosse originada de um erro experimental, ou seja, não existiria. Em nossa revisão, nós observamos uma isoforma de NP-2 que chamamos de NP2-PB e, em nossa segunda análise filogenética incluindo as sequências curtas de NP-5 e NP-6, temos alguma evidência de que NP-6 possa ser mesmo uma isoforma de NP-2. Já a NP-5 estaria relacionada entre a NP-4 e NP-1 e, infelizmente, não temos informação suficiente para confirmar que NP-5 seja um gene independente ou uma isoforma de NP-4. Assim, nos sentimos seguros de, nesta revisão, retirar o status de NP-6 como uma nova nitroforina, já que nossos resultados corroboram a sugestão de Ribeiro e colaboradores (2004) de que NP-6 é uma isoforma de NP-2. Já NP-5 seguirá existindo como uma lipocalina existente. Como NP-5 e NP-6 nunca foram depositados no GeneBank, também não temos como corrigir ou apontar mudanças nestes genes. Assim, decidimos que RPRC000380 será a nova NP-6, que RPRC000367 será a NP-8 e RPRC000077 a NP-9 (Tabela 20).

Quanto aos dados de expressão gênica das nitroforinas, foi uma surpresa observar que algumas nitroforinas são expressas em vários tecidos do inseto e não

somente na amostra contendo a glândula salivar, tendo em vista o papel bastante específico das nitroforinas na vasodilatação do hospedeiro vertebrado. Nitroforinas, inclusive a isoforma NP-2-PB, foram encontradas em intestinos, reto, testículo e/ou ovário dos insetos adultos. Estes dados sugerem que essas proteínas, tão bem estudadas na fisiologia alimentar de insetos hematófagos, possam ter outras funções até então ignoradas.

5.3.2. Novas Proteínas Biogênicas de Ligação a Amina (BAP) em *R. prolixus*

As análises identificaram 4 lipocalinas do tipo BAP, os transcritos rp_asb-101226 (não relacionada a nenhuma BAP já descrita), RPRC000114-PA (já anotada no proteoma), rp_asb-48123 e rp_asb-85900 (anotadas no mesmo locus de BAPs já identificadas). Os resultados iniciais de mapeamento tinham sugerido que esses transcritos eram nitroforinas hipotéticas. Somente na análise filogenética observou-se que os mesmos se agruparam em um ramo coeso, próximo ao das nitroforinas verdadeiras, mas distinto delas. Foi então, com base na filogenia, que essas lipocalinas passaram a ser consideradas proteínas biogênicas de ligação a amina (BAP).

É sabido que BAPs e nitroforinas apresentam uma grande proximidade evolutiva (ANDERSEN *et al.*, 2003), o que pode ter levado aos erros de anotação que acabaram por classificar as BAPs, numa primeira análise, em nitroforinas hipotéticas. Tal proximidade evolutiva pode indicar também uma similaridade funcional, uma vez que ambas têm expressão gênica já bem definida em glândula salivar e já foram caracterizadas funcionalmente como proteínas anti-hemostáticas. Enquanto as nitroforinas promovem a vasodilatação e inibem a agregação plaquetária via óxido nítrico, as BAPs o fazem via inibição de serotonina, noradrenalina e epinefrina (ANDERSEN *et al.*, 2005). Nos dois casos, há uma congruência no resultado fisiológico gerado: inibição da contração da musculatura lisa dos vasos sanguíneos e da agregação plaquetária no hospedeiro vertebrado. Dessa maneira, nitroforinas e BAPs agem em conjunto para evitar uma resposta fisiológica antagônica à ação do óxido nítrico (ANDERSEN *et al.*, 2005). É de se esperar, portanto, que houve no passado uma proteína ancestral comum as nitroforinas e BAPs que apresentava atividade anti-hemostática. Dela se originaram proteínas que, ao longo da evolução, passaram por processos de diversificação da via de atuação, se especializando em alvos distintos, mas que resultariam num mesmo resultado biológico.

Um fato interessante é que as nitroforinas são encontradas exclusivamente na tribo Rhodinini sugerindo que as BAPs sejam as proteínas ancestrais que deram origem as nitroforinas. O estudo de BAPs em outros triatomíneos e uma análise evolutiva mais abrangente (em número de espécies) pode nos dar mais informações sobre como e quando as nitroforinas surgiram. Além disso, análises funcionais nas BAPs de triatomíneos que não possuem nitroforinas seriam interessantes para saber se estas BAPs possuem alguma função de vasodilatação mais próxima da realizada pelas nitroforinas e, assim, quem sabe, identificarmos qual (ou quais) BAPs deram origem às nitroforinas de Rhodinini.

5.3.3. As Apolipoproteínas, uma nova família de lipocalinas em *R. prolixus*

Pela primeira vez foram identificadas, em *R. prolixus*, lipocalinas com similaridade a apolipoproteínas. Uma das dificuldades observadas após a primeira análise filogenética é que as possíveis apolipoproteínas não se agrupavam com um *bootstrap* significativo (>90) e, mais importante, não possuíamos nenhuma apolipoproteína funcionalmente caracterizada em *R. prolixus* para usar como um modelo. Assim, em uma segunda análise, decidimos incluir apolipoproteínas de outros hemípteros, o percevejo de cama *Cimex lectularius* e o fitófago *Halymorpha halys* (conhecido popularmente como Fede-fede). Ressaltamos que as apolipoproteínas de percevejo e Fede-fede também não foram caracterizadas funcionalmente (sua função é predita por similaridade), mas ao menos teríamos mais informações sobre este grupo de proteínas. Ao todo, 11 transcritos com essa função hipotética foram sugeridos, e nossas análises agruparam oito transcritos em basicamente dois grupos de apolipoproteínas. A filogenia sugere que a Apolipoproteína RPRC010099 é ortóloga as apolipoproteínas de *C. lectularius* e de *H. halis*, ou seja, esta proteína já existia no ancestral comum destas espécies, o que nos leva a uma situação interessante. Como explicado anteriormente, esta é a primeira vez que se identifica apolipoproteínas em *R. prolixus* e, pelo fato da maioria dos estudos de lipocalinas em barbeiros focarem no papel salivar das mesmas, uma especulação inicial é de que apolipoproteína de *R. prolixus* teria um papel na hematofagia. Assim, outros estudos com apolipoproteínas (CABRAL *et al.* 1995) nos levariam a hipotetizar que estas proteínas possam ter um papel na inibição de agregação plaquetária. No entanto, analisando a expressão de RPR010099 vemos que transcritos de ApoD foram detectados em intestino anterior, intestino posterior e testículos. Na realidade, a maioria das ApoD hipotéticas identificadas aqui possuem expressão em

diversos tecidos, sendo que todos os tecidos analisados tiveram a expressão de pelo menos uma ApoD. Assim, levando em consideração que RPR010099 não é específica de glândula salivar (na realidade não sabemos sequer se essa proteína é expressa em glândula salivar) e de que ela é um gene ancestral ainda presente em um hemíptero fitófago, é preciso aceitar que esta ApoD (e provavelmente as outras apolipoproteínas aqui identificadas) possui uma função ainda desconhecida.

De certo ponto de vista, as apolipoproteínas de *R. prolixus* são as lipocalinas mais intrigantes identificadas nesta revisão, pois, em relação as outras lipocalinas, elas são os genes com a expressão mais diversificada, e são as mais prováveis de terem funções não relacionadas a hematofagia. Assim, estas seriam excelentes alvos para estudos funcionais. De fato, as proteínas RPRC015421-PA e sua isoforma RPRC015421-PB foram as lipocalinas que nos levaram a realizar esta revisão. Inicialmente, nosso objetivo era estudar funcionalmente estas duas lipocalinas, em especial a RPRC015421-PB que é a apolipoproteína com maior nível de expressão. Esta proteína possui expressão elevada nas diferentes partes do intestino e, mais interessante ainda, nos testículos de *R. prolixus*. Este dado, juntamente com o que já foi discutido anteriormente, nos levou a especular que esta apolipoproteína poderia ter um papel na fertilidade masculina de *R. prolixus*. Infelizmente não pudemos dar continuidade aos estudos funcionais, que ficarão para um estudo futuro. Mais adiante discutiremos em mais detalhes sobre estas lipocalinas que foram detectadas em outros tecidos que não a glândula salivar, onde especularemos sobre possíveis funções não relacionadas à hematofagia.

5.3.4. O caso das lipocalinas inibidoras de agregação plaquetária: diferenças e semelhanças entre PAIs, triabinas, pallidipinas e procalinas.

Os estudos funcionais mostram que PAI, triabinas e pallidipinas são, cada qual à sua maneira, proteínas inibidoras da agregação plaquetária. Foram Noeske-Jungblut e colaboradores (1994) que isolaram a primeira lipocalina inibidora de agregação plaquetária em barbeiros, uma pallidipina (que recebeu este nome por ter sido identificada na saliva de *Triatoma pallidipenis*). Em seu trabalho, foi evidenciado que a pallidipina inibe a agregação plaquetária induzida por colágeno. Um ano depois, em 1995, Noeske-Jungblut e colaboradores isolaram uma nova inibidora de agregação plaquetária, que desta vez atuava na inibição da trombina. Por atuar em uma via diferente das pallidipinas, estas inibidoras de agregação plaquetária receberam o nome

de Triabinas. Em 2000, Francischetti e colaboradores identificaram uma nova lipocalina que inibia a agregação plaquetária em *R. prolixus*. Em seus experimentos, os autores demonstraram que estas proteínas atuavam na inibição de agregação plaquetária pela via de ADP. Diferente dos pesquisadores anteriores, Francischetti e colaboradores decidiram nomear estas novas lipocalinas por sua função, sendo chamadas de PAI (Platelet Aggregation Inhibitor ou Inibidoras de Agregação de Plaquetas). Assim, foram identificadas em barbeiros as lipocalinas que atuavam nas três vias de indução da agregação plaquetária. Por fim, as procalinas, lipocalinas alergênicas de *T. protracta*, foram identificadas apenas como similares a triabinas (PADDOCK *et al.*, 2001). Então, pallidipinas, triabinas, PAI e procalinas possuem em comum o tamanho (todas possuem ~19 kDa) e a função de inibir a agregação plaquetária. A similaridade estrutural e evolutiva neste grupo de proteínas foi ressaltada em nossas análises, que muitas vezes não conseguiram distinguir um tipo funcional de proteína de outro.

Ao analisar, por alinhamento, os domínios e conservação de sítio ativo das pallidipinas, triabinas, PAI e procalinas, foi possível perceber a similaridade estrutural, que pode justificar a similaridade funcional já observada. As sequências das proteínas são muito similares e tanto triabinas quanto PAI, pallidipinas e procalinas compartilham entre si as regiões de domínio e diversos sítios ativos conservados. É possível que essas proteínas tenham origem num mesmo ancestral comum e a diversificação funcional observada tenham se originado de duplicações gênicas posteriores.

Os resultados obtidos a partir da análise filogenética não foram muito conclusivos para os grupos PAI, pallidipina, triabina e procalina. A árvore filogenética sugeriu: um ramo definido para as triabinas, contendo apenas as já descritas Triabina-1 e Triabina-3; um ramo definido para as pallidipinas, contendo Pallidipina-1-PA, Pallidipina-2 e os dois novos genes RPRC002137-PA e Pallidipina-1-PB; e três ramos distintos agrupando a maioria das PAIs (incluindo os genes novos RPRC000446 e rp_asb104785 e as já descritas PAI-1, PAI-2, PAI-3, PAI-4, PAI-4B, PAI-6 e PAI-7). Os demais transcritos inicialmente identificados como PAI (incluindo a PAI-5 já descrita), triabinas e pallidipinas, além das duas procalinas já descritas se agruparam em dois ramos mistos classificados como “Indefinido I” e “Indefinido II”. Além disso, dois ramos contendo apenas triabinas novas, putativas, foram classificados como “Indefinido III” e “Indefinido IV” por não ter inserido nestes ramos nenhuma das triabinas já identificadas, o que impede a determinação funcional baseado em ortologia. No entanto,

análises de blasts recíprocos sugerem que estas novas triabinas são ortólogas a triabinas de outros triatomíneos e, por isso, estamos seguros em sugerir que os genes RPRC000603, RPRC002646, RPRC002647 e os transcritos rp_asb-30450 e rp_asb-33413 são de fato triabinas. Além disso, estas triabinas também formam um grupo coeso no sentido fisiológico, já que a maioria delas (4 de 5) apresentam expressão em corpo gorduroso e outros tecidos além da glândula salivar (amostra de corpo inteiro das formas imaturas), o que não foi observado para nenhuma das outras inibidoras de agregação plaquetária.

Nos resta, então, uma análise mais crítica sobre os grupos Indefinidos I e II. No grupo I, a Procalina-2 se mostrou relacionada à Triabina-4. Isto corrobora a sugestão de Paddock e colaboradores (2001) de que as procalinas identificadas por eles eram muito similares às triabinas. Desta forma, é quase certo que a Procalina-2 tenha uma função de inibição da agregação plaquetária, e possivelmente atuando na inibição de trombina. Já o grupo Indefinido II agrupou a PAI-5, a Procalina-1, a Triabina-2 e dois novos genes identificados inicialmente como triabinas. Com um *bootstrap* de 97, este agrupamento não parece se dever ao acaso e, assim, o status funcional fica questionado. Por isso realizamos uma análise específica da conservação do domínio dos sítios ativos destas cinco proteínas. Dos 23 resíduos apontados como sítio ativos, 14 são conservados nas cinco proteínas, os outros resíduos apresentam conservação entre triabinas ou entre PAI/procalina. Então, em um primeiro olhar, parece que estas proteínas seguiriam indefinidas. Porém, é importante ressaltar que PAI-5, Triabina-2 e Procalina-1 nunca tiveram suas funções confirmadas por estudos experimentais e no estudo em que foram identificadas, Ribeiro e colaboradores (2004) descrevem o seguinte em relação as PAIs:

“O papel destas novas proteínas não é conhecido. Elas podem se ligar a adenosina, como demonstrado previamente para RPAI-1, ou ligar a outras substâncias pro-hemostáticas/inflamatórias” (traduzido de RIBEIRO *et al.*, 2004).

Ou seja, não sabemos se PAI-5 realmente inibe a agregação plaquetária se ligando a ADP. Além disso, o estudo evolutivo realizado por Ribeiro e *cols.* (2004) já sugeria que PAI-5 seria a mais divergente das PAI identificadas. Por outro lado, a Triabina-2 também foi identificada no mesmo estudo e, igualmente a PAI-5, não teve sua função comprovada por estudos biológicos. Temos apenas uma evidência de que

todas estas proteínas são, provavelmente, expressas exclusivamente em glândula salivar, já que o mRNA delas não foi detectado em nenhum outro tecido de insetos adultos. Assim, nos resta aceitar que as lipocalinas agrupadas no grupo Indefinido II são, muito provavelmente, inibidoras de agregação plaquetária. Porém, o mecanismo dessa inibição (se por ligação a ADP ou a trombina) ainda precisa ser investigado.

5.4. Sobre as lipocalinas não salivares de *Rhodnius prolixus*.

Os dados de expressão gênica – gerados pelo sequenciamento realizado por Ribeiro e colaboradores em 2010 e analisados no presente estudo – possibilitaram, pela primeira vez, a identificação de lipocalinas sendo expressas fora da glândula salivar do inseto *R. prolixus*. As lipocalinas rp_asb-1685, RPRC000077-PA (nitroforinas, sendo a primeira a NP2-PB), RPRC010099-PA, RPRC003587-PA, RPRC000014-PA, RPRC015421-PA, RPRC015421-PB e rp_asb-6771 (apolipoproteínas) tiveram expressão gênica identificada nos tecidos de intestino anterior, intestino posterior e/ou reto do inseto adulto. Esses dados podem abrir portas para estudos a respeito do papel dessas lipocalinas nesses sítios fisiológicos, seja no processo de digestão do bolo alimentar ou até mesmo na interação com o *Trypanosoma cruzi*, já que nos insetos infectados esses tecidos são colonizados pelo protozoário.

Os dados de expressão gênica mostraram também que as lipocalinas rp_asb-1685, RPRC000077-PA, RPRC000380-PA (nitroforinas, sendo a primeira a NP2-PB), RPRC000114-PA (BAP), rp_asb-73726 (pallidipina), RPRC002646 (triabina), RPRC010099-PA, RPRC000014-PA, RPRC015421-PA rp_asb-47 e rp_asb-6771 (apolipoproteínas) são expressas nos testículos do inseto adulto, o que abre perspectivas para estudos mais aprofundados sobre o papel dessas proteínas na fisiologia reprodutiva do inseto macho. Dentre as lipocalinas citadas, inclusive, RPRC000380-PA (uma nitroforina) e rp_asb-73726 (uma pallidipina) foram expressas exclusivamente nos testículos (além da amostra de corpo inteiro das formas imaturas). Vale destacar que pouco se sabe a respeito da reprodução e fertilidade masculina dos triatomíneos. Historicamente, a fisiologia reprodutiva feminina recebe maior atenção dos estudos entomológicos, já que na maioria dos grupos de insetos hematófagos apenas elas se alimentam de sangue e são, portanto, vetoras de patógenos. Como isso não acontece no grupo dos triatomíneos, no qual tanto fêmeas quanto machos são hematófagos, faz-se necessário incluir a fisiologia masculina nas perguntas levantadas a respeito da reprodução e fertilidade desses vetores, o que até então tem sido negligenciado. Tal

abordagem pode contribuir, inclusive, para uma melhor compreensão da fisiologia reprodutiva de outros grupos de insetos, sejam eles hematófagos ou não. Outro dado interessante, ainda relacionado a uma possível atuação na fisiologia reprodutiva, é que as lipocalinas RPRC000114-PA (BAP), RPRC000014-PA (apolipoproteína), RPRC002646 (triabina) e RPRC000077-PA (nitroforina), além de serem expressas em testículo são também expressas nos ovários das fêmeas adultas, podendo indicar uma função tanto na fisiologia reprodutiva masculina quanto na feminina. Junto aos transcritos RPRC004463-PA (apolipoproteína) e RPRC000603 (triabina), essas foram as únicas lipocalinas, das analisadas no presente trabalho, a serem expressas em ovário.

Em 2013, Baldini e colaboradores desvendaram uma interessante interação molecular durante o acasalamento de mosquitos da espécie *Anopheles gambiae*. De acordo com o estudo, o hormônio esteroide 20-hidroxi-ecdisona (20E) produzido pelo macho é transferido para o trato reprodutor feminino durante a cópula e interage com uma proteína estimuladora de oogênese induzida por acasalamento (MISO) presente na fêmea. Tal interação desencadeia uma série de eventos moleculares que culmina no aumento da expressão de um importante transportador de lipídios vitelogênicos, resultando em uma maior produção de ovos pela fêmea alimentada (BALDINI *et al.*, 2013). Aqui levantamos a hipótese de que, nessa via fisiológica elucidada, é possível que haja a participação de lipocalinas no transporte do hormônio esteroide 20E dos testículos e/ou glândulas anexas do macho para o útero da fêmea anofelina, e que uma via similar possa estar presente também no grupo dos triatomíneos.

Segundo Baldini (2013), a interação entre o hormônio esteroide 20E e a proteína MISO pode ainda desempenhar um papel na modulação do desenvolvimento do protozoário *Plasmodium* em *Anopheles gambiae*. Tal hipótese é sustentada por um estudo que mostrou que a expressão de transportadores de lipídios vitelogênicos reduz a eficiência do TEP-1 (principal fator antiparasitário de *A. gambiae*) em eliminar o plasmódio (RONO *et al.*, 2010). Dessa forma, acredita-se que uma interferência na interação 20E-MISO possa, além de reduzir a fertilidade do inseto, reduzir também a colonização pelo protozoário *Plasmodium*. Se de fato as lipocalinas estiverem envolvidas no transporte de 20E e, portanto, de toda a interação molecular aqui discutida, abrem-se novas perspectivas de estudos acerca das lipocalinas expressas em testículos de *R. prolixus* e de outros grupos de insetos hematófagos.

Já foram descritas, em mamíferos, lipocalinas associadas à síntese de prostaglandinas (URADE & EGUCHI, 2002), moléculas derivadas de ácidos graxos,

pertencentes ao grupo dos eicosanoides, que exercem uma gama de atividades biológicas. Stanley *et al.* (2019) revisou a participação das prostaglandinas e outros eicosanoides na biologia reprodutiva de insetos. Estudos realizados em fêmeas de grilos da espécie *Teleogryllus comodus* concluíram que a prostaglandina E2 (PGE2) é um estimulante da ovoposição (LOHER, 1979). Mais tarde foi evidenciada que a atividade de biossíntese dessa prostaglandina é transferida dos insetos machos para as fêmeas via espermatóforo durante o acasalamento (LOHER *et al.*, 1981). Em 1989, Brener e Bernasconi identificaram a presença de biossíntese de prostaglandina em espermatóforos e testículos de *Triatoma infestans*, tendo sido relacionado também ao aumento da postura de ovos nessa espécie de barbeiro. Levantamos a hipótese de que, caso esse mecanismo também esteja presente em *Rhodnius prolixus*, é possível que as lipocalinas presentes no sistema reprodutor do inseto participem de algum processo envolvendo a síntese ou o transporte de prostaglandinas que estão relacionadas à sua fertilidade.

A apolipoproteína hipotética rp_asb-6771 também foi encontrada sendo expressa nos túbulos de Malpighi, assim como as RPRC004463-PA e RPRC004179-PA (esta última tendo sido detectada apenas nesse tecido, além da amostra de corpo inteiro das formas imaturas). Por ser um órgão responsável pela eliminação de excretas nitrogenadas, sugere-se que essa lipocalina – assim como outras lipocalinas são capazes de fazer –, participe do transporte de alguma substância lipossolúvel que possua nitrogênio em sua composição. A apolipoproteína RPRC004463-PA foi detectada em todos os tecidos testados com exceção do testículo (Te), o que pode indicar que ela seja uma proteína mais generalista. Em outros organismos as apolipoproteínas já foram bem caracterizadas pela sua participação no transporte de lipídios, como é o caso da molécula HDL (lipoproteína de alta densidade) que possui apolipoproteína D em sua composição. Por terem sido encontradas em diversos tecidos do inseto, é possível sugerir que essas apolipoproteínas possam atuar em diversas vias metabólicas, algumas delas ainda desconhecidas.

Nenhuma das lipocalinas classificadas neste estudo como PAI tiveram expressão gênica identificada em outros tecidos, apenas na amostra de corpo inteiro de formas imaturas de *R. prolixus* que, vale ressaltar, inclui os tecidos de glândula salivar. Isso é um indício de que esse grupo de lipocalinas com função anti-hemostática bem caracterizada deve ter expressão e atuação restrita à glândula salivar do inseto. Muitas das triabinas aqui identificadas, por outro lado, tiveram expressão gênica identificada

em diversos tecidos do inseto. Foi o caso das triabinas rp_asb-33413, RPRC000603, rp_asb-30450 e RPRC002646. Esta última, inclusive, foi expressa em todos os tecidos testados. Tal resultado indica que as triabinas podem apresentar um papel biológico mais amplo do que aquele já conhecido na fisiologia alimentar.

Temos consciência de que as hipóteses levantadas neste tópico a respeito de uma possível atividade das lipocalinas (para além do que já é conhecido na fisiologia alimentar) necessitam de um maior aprofundamento teórico e experimental. Fato é que muito há para se descobrir sobre essas proteínas e o presente trabalho visa, mais do que responder, levantar novas perguntas que poderão, mais tarde, ser objeto de estudo do nosso grupo de pesquisa e de outros cuja curiosidade tenha sido instigada.

5.5. Uma revisão da nomenclatura e possível função das lipocalinas de *R. prolixus*.

Após nossa análise revisando as lipocalinas que já haviam sido identificadas em *R. prolixus*, e com a identificação de novas lipocalinas, gostaríamos então de terminar este estudo com uma proposta para melhor classificação das lipocalinas, baseando-se, claro, nas evidências evolutivas e funcionais aqui observadas. O principal motivo para tal é nossa preocupação com a falta de padronização na nomenclatura das inibidoras de agregação plaquetária. O fato que temos uma família de proteínas chamada de Inibidora de Agregação Plaquetária (PAI) e outras duas famílias que fazem a mesma função mas possuem o nome de Triabinas e Pallidipinas (sem mencionar as procalinas), nos leva a crer que há a possibilidade de pesquisadores serem induzidos ao erro por não identificar, de imediato, que Triabinas e Pallidipinas também inibem agregação plaquetária, mesmo que por vias distintas. Por outro lado, a comunidade científica já está usando a nomenclatura existente a mais de duas décadas e uma mudança de nomenclatura poderia causar mais confusão do que resolução. Assim, na tabela 20, resumimos os resultados, revisados, obtidos nesse estudo, com uma proposta de revisão de nomenclatura para algumas das lipocalinas estudadas.

Tabela 20. Revisão das lipocalinas de *Rhodnius prolixus*

| Função principal | Mecanismo proposto | Nome | Gene/Transcrito | Detectada em: | Comentários |
|--|---|---------------------|-----------------|----------------------------|---|
| Vasodilatador | Liberação de óxido nítrico | Nitroforina-1 | Q26239 | GS | |
| | | Nitroforina-2 | RPRC000072 | GS, NI | |
| | | Nitroforina-2 PB | rp_asb-1685 | GS, NI, CG, IA, IP, Rt, Te | Nova É a antiga NP6 |
| | | Nitroforina-2 PC | rp_asb-3232 | NI | Nova |
| | | Nitroforina-2 PD | rp_asb-101256 | NI | Nova |
| | | Nitroforina-3 | Q94733 | GS | |
| | | Nitroforina-4 | RPRC000023 | GS, NI | |
| | | Nitroforina-4 PB | rp_asb-48696 | NI | Nova |
| | | Nitroforina-4 PC | rp_asb-48698 | NI | Nova |
| | | Nitroforina-5 | Desconhecido | GS | |
| | | Nitroforina-6 | RPRC000367 | NI | A antiga NP-6 agora é Nitroforina-2 PB |
| | | Nitroforina-7 | Q6PQK2 | GS | |
| | | Nitroforina-8 | RPRC000380 | NI, Te | |
| Nitroforina-9 | RPRC000077 | NI, IA, IP, OV, Te | | | |
| Inibidor de agregação plaquetária e vasodilatadora | Ligação a serotonina e epinefrina diminui a velocidade da agregação plaquetária. Ligação a norepinefrina impede vasoconstrição | BAP-1 | AAO25746 | GS | As BAPs 4GET e 4HFO foram unidas nesse único gene |
| | | BAP-2 | RPRC000114 | NI, OV, Te | Nova |
| | | BAP-2 PB | rp_asb-48123 | NI | Nova |
| | | BAP-3 | RPRC010481 | NI | Nova |
| | | BAP-4 | rp_asb-85900 | NI | Nova |
| | | BAP-5 | rp_asb-101226 | NI | Nova |
| Inibidor de agregação plaquetária | Ligação a ADP | PAI-1 | RPRC002102 | GS | |
| | | PAI-2 | RPRC000175 | GS | |
| | | PAI-3 | AAQ20817 | GS | |
| | | PAI-4 | AAQ20818 | GS | PAI 5H9L foi unida a PAI-4 |
| | | PAI-6 | RPRC000677 | GS | |
| | | PAI-7 | RPRC000463 | GS | |
| | | PAI-8 | RPRC000446 | NI | Nova |
| | | PAI-9 | rp_asb104785 | NI | Nova |
| | | Ligação ao colágeno | Pallidipina-1 | RPRC000012 | GS, NI |
| | Pallidipina-1 PB | | rp_asb-73726 | NI, Te | Nova |
| | Pallidipina-2 | | AAQ20826 | GS | |
| | Pallidipina-3 | | RPRC002137 | NI | Nova |
| | Ligação a trombina | Triabina-1 | RPRC006167 | GS | |

| | | | | | |
|----------|--------------|---------------|---------------|--------------------------------|----------------|
| | | Triabina-3 | RPRC000147 | GS | |
| | | Triabina-4 | RPRC002583 | GS | |
| | | Triabina-5 | RPRC002647 | NI | Nova |
| | | Triabina-6 | rp_asb-33413 | NI, CG, IA | Nova |
| | | Triabina-7 | RPRC000603 | NI, CG, IA, IP, Rt, OV | Nova |
| | | Triabina-7 PB | rp_asb-30450 | NI, CG, IA | Nova |
| | | Triabina-8 | RPRC002646 | NI, CG, IA, IP, Rt, TM, OV, Te | Nova |
| | | Procalina-2 | AQ20828 | GS | |
| | Desconhecido | PAI-5 | RPRC002584 | GS | |
| | | Triabina-2 | RPRC012954 | GS | |
| | | Procalina-1 | RPRC010444 | GS | |
| | | RPRC000275 | RPRC000275 | NI | Nova |
| | | rp_asb-105577 | rp_asb-105577 | NI | Nova |
| | Desconhecido | Desconhecido | APO-1 | RPRC010099 | NI, IA, IP, Te |
| APO-2 | | | RPRC004179 | NI, TM | Nova |
| APO-3 | | | RPRC004463 | NI, CG, IA, IP, Rt, TM, OV | Nova |
| APO-4 | | | RPRC003587 | NI, IA, IP, Rt, | Nova |
| APO-5 | | | RPRC000014 | NI, IA, IP, Rt, OV, Te | Nova |
| APO-6 | | | rp_asb-6771 | NI, IA, IP, Rt, Te | Nova |
| APO-7 | | | RPRC015421 | NI, IA, IP, Rt, Te | Nova |
| APO-7 PB | | | rp_asb-47 | NI, IA, IP, Rt, Te | Nova |

GS – Glandula Salivar; IA – Intestino Anterior; CG – Corpo Gorduroso; TM – Tubulo de Malpighi; OV – Ovário; IP – Intestino Posterior; Rt – Reto; Te – Testículo; NI – Corpo Inteiro de formas imaturas (ovos e ninfas).

6. CONCLUSÃO

Algumas das lipocalinas já conhecidas e melhor estudadas são essenciais para o sucesso de insetos hematófagos durante o repasto sanguíneo. O presente trabalho mostra que a importância dessa família pode estar além do conhecimento científico até então adquirido. Foi possível identificar 29 lipocalinas inéditas no inseto *Rhodnius prolixus*, sendo 5 nitroforinas, 5 proteínas biogênicas de ligação a amina (BAP), 11 inibidoras de agregação plaquetária (incluindo 2 PAI, 2 pallidipinas, 5 triabinas e 2 com mecanismo de ação desconhecido), e 8 apolipoproteínas.

A análise das estruturas gênicas e das relações evolutivas das lipocalinas mostrou um alto grau de conservação entre elas, especialmente das que fazem parte dos grupos com função de inibição de agregação plaquetária bem caracterizada (PAI, triabinas e pallidipinas). Tal análise foi útil para, com base na história evolutiva da família, sugerir a revisão da nomenclatura dos grupos de lipocalinas inibidoras de agregação plaquetária, reclassificar lipocalinas erroneamente identificadas e prever as funções biológicas de algumas delas.

Muitas das proteínas aqui evidenciadas foram expressas em diversos tecidos do barbeiro, como intestino, reto, túbulo de mapighi, testículo e ovário. Encontrá-las fora da glândula salivar, além de ser algo inédito na literatura, abre perspectivas para novos estudos funcionais, pois está claro que as funções desempenhadas pelas lipocalinas não se resumem à fisiologia alimentar. É possível que algumas das lipocalinas com expressão em testículo tenham um papel importante na fisiologia reprodutiva masculina, com influência na fertilidade do inseto.

Compreender aspectos particulares da reprodução dos machos de insetos hematófagos pode abrir caminhos para o desenvolvimento de novas técnicas de controle vetorial que sejam complementares as já existentes. Por isso, estudos experimentais de função gênica se fazem necessários para evidenciar o papel das lipocalinas em todo o organismo do *R. prolixus* e de outros insetos hematófagos.

REFERÊNCIAS

ALTSCHUL, Stephen F. *et al.* Basic local alignment search tool. **Journal Of Molecular Biology**, [S.L.], v. 215, n. 3, p. 403-410, out. 1990. Elsevier BV. [http://dx.doi.org/10.1016/s0022-2836\(05\)80360-2](http://dx.doi.org/10.1016/s0022-2836(05)80360-2).

ANDERSEN, John F. *et al.* Inhibition of Hemostasis by a High Affinity Biogenic Amine-binding Protein from the Saliva of a Blood-feeding Insect. **Journal Of Biological Chemistry**, v. 278, n. 7, p. 4611-4617, fev. 2003. Elsevier BV. <http://dx.doi.org/10.1074/jbc.m211438200>.

ANDERSEN, J. F. *et al.* Recognition of Anionic Phospholipid Membranes by an Antihemostatic Protein from a Blood-Feeding Insect. **Biochemistry**, v. 43, n. 22, p. 6987-6994, jun. 2004. American Chemical Society (ACS). <http://dx.doi.org/10.1021/bi049655t>.

ANDERSEN, John F. *et al.* The role of salivary lipocalins in blood feeding by *Rhodnius prolixus*. **Archives Of Insect Biochemistry And Physiology**, [s.l.], v. 58, n. 2, p.97-105, 2005. Wiley. <http://dx.doi.org/10.1002/arch.20032>.

ANDRADE, Bruno B. *et al.* Haematophagous arthropod saliva and host defense system: a tale of tear and blood. **Anais da Academia Brasileira de Ciências**, [S.L.], v. 77, n. 4, p. 665-693, dez. 2005. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0001-37652005000400008>.

ARAÚJO, R.N. *et al.* Interação entre os Insetos Hematófagos e seus Hospedeiros Vertebrados. **In Tópicos Avançados em Entomologia Molecular**. Instituto Nacional de Ciência e Tecnologia em Entomologia Molecular, pp. 1–19. 2012. Disponível em: http://www.inctem.bioqmed.ufrj.br/images/documentos/biblioteca/Capitulo_11_Interacao_entre_os_Insetos_Hematofagos_e_seus_Hospedeiros_Vertebrados.pdf. Acesso em: 03/02/2021.

BALDINI, F. *et al.* The interaction between a sexually transferred steroid hormone and a female protein regulates oogenesis in the malaria mosquito *Anopheles gambiae*. **PLoS Biol.** 2013 Oct;11(10):e1001695. doi: <http://dx.doi.org/10.1371/journal.pbio.1001695>.

BAO, Guan-hu *et al.* The ligands of neutrophil gelatinase-associated lipocalin. **Rsc Advances**, [s.l.], v. 5, n. 126, p.104363-104374, 2015. Royal Society of Chemistry (RSC). <http://dx.doi.org/10.1039/c5ra18736b>.

BASANOVA, A. V. *et al.* Vascular-platelet and plasma hemostasis regulators from bloodsucking animals. **Biochemistry** (Moscow), v. 67, n. 1, p. 143-150, 2002. Springer Science and Business Media LLC. <http://dx.doi.org/10.1023/a:1013968618120>.

BEAMENT, James *et al.* Wigglesworth's contribution to insect biology and the state of the art. **Insect Biology In The Future**, [s.l.], p.1-17, 1980. Elsevier. <http://dx.doi.org/10.1016/b978-0-12-454340-9.50005-7>.

BRENNER, Rodolfo R.; BERNASCONI, Ana. Prostaglandin biosynthesis in the gonads of the hematophagus insect *Triatoma infestans*. **Comparative Biochemistry And Physiology Part B: Comparative Biochemistry**, v. 93, n. 1, p. 1-4, jan. 1989. Elsevier BV. [http://dx.doi.org/10.1016/0305-0491\(89\)90207-1](http://dx.doi.org/10.1016/0305-0491(89)90207-1).

CABRAL, João H. Morais *et al.* Arachidonic acid binds to apolipoprotein D: implications for the protein's function. **Febs Letters**, v. 366, n. 1, p. 53-56, 5 jun. 1995. Wiley. [http://dx.doi.org/10.1016/0014-5793\(95\)00484-q](http://dx.doi.org/10.1016/0014-5793(95)00484-q).

CHAMPAGNE, D. E. The role of salivary vasodilators in bloodfeeding and parasite transmission. **Parasitology Today**, v. 10, n. 11, p. 430-433, jan. 1994. Elsevier BV. [http://dx.doi.org/10.1016/0169-4758\(94\)90173-2](http://dx.doi.org/10.1016/0169-4758(94)90173-2).

CHIANG, Ralem *et al.* Exploring the Role of Rhodtestolin, A Cardio-Inhibitor from the Testes of *Rhodnius prolixus*, in Relation to the Structure and Function of Reproductive Organs in Insect Vectors of Chagas Disease. **Insects**, [s.l.], v. 4, n. 4, p.593-608, 30 out. 2013. MDPI AG. <http://dx.doi.org/10.3390/insects4040593>.

COURA, José Rodrigues. Tripanosomose, doença de Chagas. **Cienc. Cult.**, São Paulo, v. 55, n. 1, p. 30-33, Jan. 2003. Disponível em: <http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S0009-67252003000100022&lng=en&nrm=iso>. Acesso em: 26 de maio de 2021.

DE ARAÚJO, C. N. *et al.* Interactome: smart hematophagous triatomine salivary gland molecules counteract human hemostasis during meal acquisition. **Journal Of**

Proteomics, v. 75, n. 13, p. 3829-3841, jul. 2012. Elsevier BV.
<http://dx.doi.org/10.1016/j.jprot.2012.05.001>.

DIAS, João Carlos Pinto *et al.* Mecanismos alternativos de transmissão do *Trypanosoma cruzi* no Brasil e sugestões para sua prevenção. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 44, n. 3, p. 375-379, 27 maio 2011. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0037-86822011005000032>.

DIAS, João Carlos Pinto *et al.* Evolution of Chagas Disease Screening Programs and Control Programs: Historical Perspective. **Global Heart**, [s.l.], v. 10, n. 3, p.193- 202, 1 set. 2015. Ubiquity Press, Ltd.. <http://dx.doi.org/10.1016/j.gheart.2015.06.003>.

DIOTAIUTI, Liléia; BARBOSA, Silvia H. Triatomíneos. Belo Horizonte: CPqRR, 2015.

DUJARDIN, J. P. *et al.* The origin of *Rhodnius prolixus* in Central America. **Medical And Veterinary Entomology**, v. 12, n. 1, p. 113-115, jan. 1998. Wiley.
<http://dx.doi.org/10.1046/j.1365-2915.1998.00092.x>.

EDGAR, R. C.. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, [S.L.], v. 32, n. 5, p. 1792-1797, 8 mar. 2004. Oxford University Press (OUP). <http://dx.doi.org/10.1093/nar/gkh340>.

FERREIRA, R. T. B. *et al.* Transmissão oral da doença de Chagas pelo consumo de açaí: um desafio para a Vigilância Sanitária. **Vigilância Sanitária em Debate**, Rio de Janeiro, v. 2, n. 4, p. 4-11, 2014.

FLOWER, Darren R. The lipocalin protein family: structure and function. **Biochemical Journal**, v. 318, n. 1, p. 1-14, 15 ago. 1996. Portland Press Ltd..
<http://dx.doi.org/10.1042/bj3180001>.

FLOWER, Darren R. *et al.* The lipocalin protein family: structural and sequence overview. **Biochimica Et Biophysica Acta** (bba) - Protein Structure And Molecular Enzymology, [s.l.], v. 1482, n. 1-2, p.9-24, out. 2000. Elsevier BV.
[http://dx.doi.org/10.1016/s0167-4838\(00\)00148-5](http://dx.doi.org/10.1016/s0167-4838(00)00148-5).

FRANCISCHETTI, Ivo M. B. *et al.* Purification, Cloning, Expression, and Mechanism of Action of a Novel Platelet Aggregation Inhibitor from the Salivary Gland of the

Blood-sucking Bug, *Rhodnius prolixus*. **Journal Of Biological Chemistry**, [S.L.], v. 275, n. 17, p. 12639-12650, abr. 2000. Elsevier BV. <http://dx.doi.org/10.1074/jbc.275.17.12639>.

FUENTES-PRIOR, P. *et al.* Structure of the thrombin complex with triabin, a lipocalin-like exosite-binding inhibitor derived from a triatomine bug. **Proceedings Of The National Academy Of Sciences**, v. 94, n. 22, p. 11845-11850, 28 out. 1997. Proceedings of the National Academy of Sciences. <http://dx.doi.org/10.1073/pnas.94.22.11845>.

GALVÃO, Cleber. A sistemática dos triatomíneos (Hemiptera, Reduviidae), de De Geer ao DNA. **Entomol Vect.** v. 10, n. 4, p. 511-530, 2003.

GARCIA, E. S. *et al.* Exploring the role of insect host factors in the dynamics of *Trypanosoma cruzi*–*Rhodnius prolixus* interactions. **Journal Of Insect Physiology**, v. 53, n. 1, p. 11-21, jan. 2007. Elsevier BV. <http://dx.doi.org/10.1016/j.jinsphys.2006.10.006>.

GILLESPIE, R. D. *et al.* The immunomodulatory factors of bloodfeeding arthropod saliva. **Parasite Immunology**, v. 22, n. 7, p. 319-331, jul. 2000. Wiley. <http://dx.doi.org/10.1046/j.1365-3024.2000.00309.x>.

GORDON, R. M.; CREWE, W. The Mechanisms by Which Mosquitoes and Tsetse-Flies Obtain Their Blood-Meal, the Histology of the Lesions Produced, and the Subsequent Reactions of the Mammalian Host; Together with Some Observations on the Feeding of Chrysops and Cimex. **Annals Of Tropical Medicine & Parasitology**, v. 42, n. 3-4, p. 334-356, dez. 1948. Informa UK Limited. <http://dx.doi.org/10.1080/00034983.1948.11685382>.

GRZYB, Joanna *et al.* Lipocalins – a family portrait. **Journal Of Plant Physiology**, [s.l.], v. 163, n. 9, p.895-915, set. 2006. Elsevier BV. <http://dx.doi.org/10.1016/j.jplph.2005.12.007>.

HELLMANN, K; HALKINS, R. I. Prolixin-S and Prolixin-G; Two Anticoagulants from *Rhodnius prolixus* Stål. **Nature**, v. 207, n. 4994, p. 265-267, jul. 1965. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/207265a0>.

HERNÁNDEZ-VARGAS, M. J. *et al.* An Insight into the Triabin Protein Family of American Hematophagous Reduviids: functional, structural and phylogenetic analysis. **Toxins**, v. 8, n. 2, p. 44, 15 fev. 2016. MDPI AG.

<http://dx.doi.org/10.3390/toxins8020044>.

JAREBORG, N. *et al.* Comparative Analysis of Noncoding Regions of 77 Orthologous Mouse and Human Gene Pairs. **Genome Research**, [S.L.], v. 9, n. 9, p. 815-824, 1 set. 1999. Cold Spring Harbor Laboratory. <http://dx.doi.org/10.1101/gr.9.9.815>.

KONTOPIDIS, George *et al.* The Ligand-binding Site of Bovine β -Lactoglobulin: Evidence for a Function?. **Journal Of Molecular Biology**, [s.l.], v. 318, n. 4, p.1043-1055, maio 2002. Elsevier BV. [http://dx.doi.org/10.1016/s0022-2836\(02\)00017-7](http://dx.doi.org/10.1016/s0022-2836(02)00017-7).

KOONIN, Eugene V.. Orthologs, Paralogs, and Evolutionary Genomics. **Annual Review Of Genetics**, [S.L.], v. 39, n. 1, p. 309-338, dez. 2005. Annual Reviews. <http://dx.doi.org/10.1146/annurev.genet.39.073003.114725>.

KUMAR, Sudhir *et al.* MEGA X: molecular evolutionary genetics analysis across computing platforms. **Molecular Biology And Evolution**, [S.L.], v. 35, n. 6, p. 1547-1549, 2 maio 2018. Oxford University Press (OUP). <http://dx.doi.org/10.1093/molbev/msy096>.

LAKSHMI, Balasubramanian *et al.* Structure-Based Phylogenetic Analysis of the Lipocalin Superfamily. **Plos One**, [s.l.], v. 10, n. 8, p.1-18, 11 ago. 2015. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0135507>.

LAZZARI, Claudio Ricardo *et al.* Behavioural biology of Chagas disease vectors. **Memórias do Instituto Oswaldo Cruz**, v. 108, n. 1, p. 34-47, 2013. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/0074-0276130409>.

LENT, H. & WYGODZINSKY, P. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas disease. **Bull. Am. Mus. Nat. Hist.** v. 163, n. 3, p. 123-520, jul. 1979.

LEONE, Maria Grazia *et al.* Lipocalin type prostaglandin D-synthase: which role in male fertility?. **Contraception**, v. 65, n. 4, p. 293-295, abr. 2002. Elsevier BV. [http://dx.doi.org/10.1016/s0010-7824\(02\)00280-9](http://dx.doi.org/10.1016/s0010-7824(02)00280-9).

LERNER, Ethan A. *et al.* Maxadilan, a PAC1 receptor agonist from sand flies. **Peptides**, v. 28, n. 9, p. 1651-1654, set. 2007. Elsevier BV. <http://dx.doi.org/10.1016/j.peptides.2007.06.021>.

LIMA, Emerson Silva; COUTO, Ricardo David. Estrutura, metabolismo e funções fisiológicas da lipoproteína de alta densidade. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, v. 42, n. 3, p. 169-178, jun. 2006. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1676-24442006000300005>.

LOHER, W. The influence of Prostaglandin E2 on oviposition in *Teleogryllus commodus*. **Entomologia Experimentalis Et Applicata**, v. 25, n. 1, p. 107-109, jan. 1979. Wiley. <http://dx.doi.org/10.1111/j.1570-7458.1979.tb02853.x>.

LOHER, W. *et al.* Prostaglandins: their role in egg-laying of the cricket *Teleogryllus commodus*. **Proceedings Of The National Academy Of Sciences**, v. 78, n. 12, p. 7835-7838, 1 dez. 1981. Proceedings of the National Academy of Sciences. <http://dx.doi.org/10.1073/pnas.78.12.7835>.

LU, Hengyun *et al.* Oxford Nanopore MinION Sequencing and Genome Assembly. **Genomics, Proteomics & Bioinformatics**, [S.L.], v. 14, n. 5, p. 265-279, out. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.gpb.2016.05.004>.

MA, Dongying *et al.* Triplatin, a platelet aggregation inhibitor from the salivary gland of the triatomine vector of Chagas disease, binds to TXA2 but does not interact with glycoprotein PVI. **Thrombosis And Haemostasis**, v. 107, n. 01, p. 111-123, 2012. Georg Thieme Verlag KG. <http://dx.doi.org/10.1160/th11-10-0685>.

MANS, B. J. *et al.* Structure, Function, and Evolution of Biogenic Amine-binding Proteins in Soft Ticks. **Journal Of Biological Chemistry**, [s.l.], v. 283, n. 27, p.18721-18733, 29 abr. 2008. American Society for Biochemistry & Molecular Biology (ASBMB). <http://dx.doi.org/10.1074/jbc.m800188200>.

MANS, Ben J.. Evolution of Vertebrate Hemostatic and Inflammatory Control Mechanisms in Blood-Feeding Arthropods. **Journal Of Innate Immunity**, [S.L.], v. 3, n. 1, p. 41-51, 2011. S. Karger AG. <http://dx.doi.org/10.1159/000321599>.

MESQUITA, Rafael D. *et al.* Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. **Proceedings**

Of The National Academy Of Sciences, [S.L.], v. 112, n. 48, p. 14936-14941, 16 nov. 2015. Proceedings of the National Academy of Sciences. <http://dx.doi.org/10.1073/pnas.1506226112>.

MONCADA, Salvador *et al.* Endothelium-derived relaxing factor: identification as nitric oxide and role in the control of vascular tone and platelet function. **Biochemical Pharmacology**, v. 37, n. 13, p. 2495-2501, jul. 1988. Elsevier BV. [http://dx.doi.org/10.1016/0006-2952\(88\)90236-5](http://dx.doi.org/10.1016/0006-2952(88)90236-5).

MONTFORT, W. R. *et al.* Nitrophorins and related antihemostatic lipocalins from *Rhodnius prolixus* and other blood-sucking arthropods. **Biochimica Et Biophysica Acta (Bba) - Protein Structure And Molecular Enzymology**, v. 1482, n. 1-2, p. 110-118, out. 2000. Elsevier BV. [http://dx.doi.org/10.1016/s0167-4838\(00\)00165-5](http://dx.doi.org/10.1016/s0167-4838(00)00165-5).

MOREIRA, Mônica F. *et al.* Changes in salivary nitrophorin profile during the life cycle of the blood-sucking bug *Rhodnius prolixus*. **Insect Biochemistry And Molecular Biology**, v. 33, n. 1, p. 23-28, jan. 2003. Elsevier BV. [http://dx.doi.org/10.1016/s0965-1748\(02\)00163-7](http://dx.doi.org/10.1016/s0965-1748(02)00163-7).

NATURE. Who, how, what and where?. **Nature**, [s.l.], v. 465, n. 7301, p.8-9, jun. 2010. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/nature09222>.

NEVES, David Pereira. Parasitologia humana. 13. ed. São Paulo: Atheneu, 2016.

NOESKE-JUNGBLUT, Christiane *et al.* An inhibitor of collagen-induced platelet aggregation from the saliva of *Triatoma pallidipennis*. **The Journal Of Biological Chemistry**, [s.i.], v. 269, n. 7, p.5050-5053, 18 fev. 1994.

NOESKE-JUNGBLUT, C. *et al.* Triabin, a Highly Potent Exosite Inhibitor of Thrombin. **Journal Of Biological Chemistry**, [s.l.], v. 270, n. 48, p.28629-28634, 1 dez. 1995. American Society for Biochemistry & Molecular Biology (ASBMB). <http://dx.doi.org/10.1074/jbc.270.48.28629>.

PADDOCK, Christopher D. *et al.* Identification, Cloning, and Recombinant Expression of Procalin, a Major Triatomine Allergen. **The Journal Of Immunology**, [s.l.], v. 167, n. 5, p.2694-2699, 1 set. 2001. The American Association of Immunologists. <http://dx.doi.org/10.4049/jimmunol.167.5.2694>.

PAIM, Rafaela M. M. Genômica funcional de *Rhodnius prolixus* (Hemiptera, Reduviidae): avaliação da persistência do silenciamento (RNAi) e caracterização in vivo de genes importantes na interação do inseto com hospedeiros e parasitos. 2014. 160 f. Tese (Doutorado em Ciências) - Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, 2014.

PASSARGE, Eberhard *et al.* Incorrect use of the term synteny. *Nature Genetics*, [S.L.], v. 23, n. 4, p. 387-387, dez. 1999. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/70486>.

RHOADS, Anthony; AU, Kin Fai. PacBio Sequencing and Its Applications. **Genomics, Proteomics & Bioinformatics**, [S.L.], v. 13, n. 5, p. 278-289, out. 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.gpb.2015.08.002>.

RIBEIRO, J.M. C.; GARCIA, E. S.. The Role of the Salivary Glands in Feeding in *Rhodnius Prolixus*. **Journal Of Experimental Biology**, v. 94, n. 1, p. 219-230, 1 out. 1981. The Company of Biologists. <http://dx.doi.org/10.1242/jeb.94.1.219>.

RIBEIRO, J. M. C. The antiserotonin and antihistamine activities of salivary secretion of *Rhodnius prolixus*. **Journal Of Insect Physiology**, [S.L.], v. 28, n. 1, p. 69-75, jan. 1982. Elsevier BV. [http://dx.doi.org/10.1016/0022-1910\(82\)90024-5](http://dx.doi.org/10.1016/0022-1910(82)90024-5).

RIBEIRO, José M.C. *et al.* A salivary vasodilator in the blood-sucking bug, *Rhodnius prolixus*. **British Journal Of Pharmacology**, v. 101, n. 4, p. 932-936, dez. 1990. Wiley. <http://dx.doi.org/10.1111/j.1476-5381.1990.tb14183.x>.

RIBEIRO, J. M.; NUSSENZVEIG, R. H.. The salivary catechol oxidase/peroxidase activities of the mosquito *Anopheles albimanus*. **Journal Of Experimental Biology**, v. 179, n. 1, p. 273-287, 1 jun. 1993. The Company of Biologists. <http://dx.doi.org/10.1242/jeb.179.1.273>.

RIBEIRO, J. *et al.* Reversible binding of nitric oxide by a salivary heme protein from a bloodsucking insect. **Science**, v. 260, n. 5107, p. 539-541, 23 abr. 1993. American Association for the Advancement of Science (AAAS). <http://dx.doi.org/10.1126/science.8386393>.

RIBEIRO, J M; A WALKER, F. High affinity histamine-binding and antihistaminic activity of the salivary nitric oxide-carrying heme protein (nitrophorin) of *Rhodnius*

prolixus. **Journal Of Experimental Medicine**, v. 180, n. 6, p. 2251-2257, 1 dez. 1994. Rockefeller University Press. <http://dx.doi.org/10.1084/jem.180.6.2251>.

RIBEIRO, J. M. C. *et al.* Purification and characterization of prolixin S (nitrophorin 2), the salivary anticoagulant of the blood-sucking bug *Rhodnius prolixus*. **Biochemical Journal**, v. 308, n. 1, p. 243-249, 15 maio 1995. Portland Press Ltd.. <http://dx.doi.org/10.1042/bj3080243>.

RIBEIRO, José M. C. *et al.* Role of Salivary Antihemostatic Components in Blood Feeding by Triatomine Bugs (Heteroptera). **Journal Of Medical Entomology**, v. 35, n. 4, p. 599-610, 1 jul. 1998. Oxford University Press (OUP). <http://dx.doi.org/10.1093/jmedent/35.4.599>.

RIBEIRO, J. M. C. *et al.* Exploring the sialome of the blood-sucking bug *Rhodnius prolixus*. **Insect Biochemistry And Molecular Biology**, v. 34, n. 1, p. 61-79, jan. 2004. Elsevier BV. <http://dx.doi.org/10.1016/j.ibmb.2003.09.004>.

RIBEIRO, José M. C. *et al.* An Insight into the Transcriptome of the Digestive Tract of the Bloodsucking Bug, *Rhodnius prolixus*. **Plos Neglected Tropical Diseases**, [s.l.], v. 8, n. 1, p.1-31, 9 jan. 2014. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pntd.0002594>.

ROCHA, D.S.; SANTOS, C. M.. Morfologia dos ovos e ninfas. In: GALVÃO, C., org. Vetores da doença de chagas no Brasil [online]. Curitiba: Sociedade Brasileira de Zoologia, 2014, pp. 40-63. Zoologia: guias e manuais de identificação series. ISBN 978-85-98203-09-6. Available from SciELO Books.

RONO, M. K. *et al.* The major yolk protein vitellogenin interferes with the anti-plasmodium response in the malaria mosquito *Anopheles gambiae*. **PLoS Biol.** 2010 Jul 20;8(7):e1000434. doi: <http://dx.doi.org/10.1371/journal.pbio.1000434>.

SANT'ANNA, M. R. V. *et al.* Triatomines (Hemiptera, Reduviidae) blood intake: physical constraints and biological adaptations. **Journal Of Insect Physiology**, v. 97, p. 20-26, fev. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.jinsphys.2016.08.004>.

SCHIEFNER, André *et al.* The dimeric crystal structure of the human fertility lipocalin glycodelin reveals a protein scaffold for the presentation of complex glycans.

Biochemical Journal, [s.l.], v. 466, n. 1, p.95-104, 6 fev. 2015. Portland Press Ltd..
<http://dx.doi.org/10.1042/bj20141003>.

SCHLEBUSCH, Stephen; ILLING, Nicola. Next generation shotgun sequencing and the challenges of de novo genome assembly. **South African Journal Of Science**, [S.L.], v. 108, n. 11/12, p. 1-8, 25 out. 2012. Academy of Science of South Africa.
<http://dx.doi.org/10.4102/sajs.v108i11/12.1256>.

SIEVERS, Fabian *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, [S.L.], v. 7, n. 1, p. 539, jan. 2011. EMBO. <http://dx.doi.org/10.1038/msb.2011.75>.

SILVA, Francinaldo S. A importância hematofágica e parasitológica da saliva dos insetos hematófagos. **Revista Trópica: Ciências Agrárias e Biológicas**, Chapadinha, MA, v. 3, n. 3, p.3-18, fev. 2009.

SONENSHINE, Daniel E. *et al.* First Transcriptome of the Testis-Vas Deferens-Male Accessory Gland and Proteome of the Spermatophore from *Dermacentor variabilis* (Acari: Ixodidae). **Plos One**, [s.l.], v. 6, n. 9, p.1-16, 16 set. 2011. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0024711>.

STANLEY, D.; KIM, Y. Prostaglandins and Other Eicosanoids in Insects: Biosynthesis and Biological Actions. **Front Physiol.** 2019 Feb 7;9:1927. doi:
<http://dx.doi.org/10.3389/fphys.2018.01927>.

SUN, Jianxin *et al.* Characterization and cDNA cloning of a hemoprotein in the salivary glands of the blood-sucking insect, *Rhodnius prolixus*. **Insect Biochemistry And Molecular Biology**, v. 28, n. 3, p. 191-200, mar. 1998. Elsevier BV.
[http://dx.doi.org/10.1016/s0965-1748\(97\)00115-x](http://dx.doi.org/10.1016/s0965-1748(97)00115-x).

TIRLONI, Lucas *et al.* Editorial: The Role of Saliva in Arthropod-Host-Pathogen Relation. **Frontiers in Cellular and Infection Microbiology**. 10:630626. 28 jan. 2021. doi: <http://dx.doi.org/10.3389/fcimb.2020.630626>.

URADE, Yoshihiro; EGUCHI, Naomi. Lipocalin-type and hematopoietic prostaglandin D synthases as a novel example of functional convergence. **Prostaglandins & Other Lipid Mediators**, v. 68-69, p. 375-382, ago. 2002. Elsevier BV.
[http://dx.doi.org/10.1016/s0090-6980\(02\)00042-4](http://dx.doi.org/10.1016/s0090-6980(02)00042-4).

WIGGLESWORTH, V.B. *Insects and the Life of Man*; John Wiley & Sons, Inc.: New York, NY, USA, 1976.

WORLD HEALTH ORGANIZATION. Chagas disease (American trypanosomiasis): Fact Sheet No 340. Geneva, 2019. Disponível em: <https://www.who.int/chagas/en/>. Acesso em: 28 oct. 2019.

ZHANG, Yan *et al.* Nitrophorin-2: a novel mixed-type reversible specific inhibitor of the intrinsic factor-x activating complex. **Biochemistry**, v. 37, n. 30, p. 10681-10690, jul. 1998. American Chemical Society (ACS). <http://dx.doi.org/10.1021/bi973050y>.

ZHARKIKH, Andrey; LI, Wen-Hsiung. Estimation of Confidence in Phylogeny: the complete-and-partial bootstrap technique. **Molecular Phylogenetics And Evolution**, [S.L.], v. 4, n. 1, p. 44-63, mar. 1995. Elsevier BV. <http://dx.doi.org/10.1006/mpev.1995.1005>.

ANEXO A

Mapeamento da Lipocalinas já descritos ao Genoma de *R. prolixus*:

| Proteína | Scaffold | ID | Algn | Qstart | Qend | Hstart | Hend | Evalue | Score |
|----------|--------------|-------|------|--------|------|---------|---------|----------|-------|
| BAP-1 | KQ034229 | 91.67 | 24 | 15 | 38 | 866282 | 866353 | 1.00E-24 | 48.5 |
| BAP-1 | KQ034229 | 82.98 | 47 | 39 | 85 | 866431 | 866571 | 1.00E-24 | 85.9 |
| BAP-1 | KQ034229 | 81.82 | 33 | 89 | 121 | 866945 | 867043 | 4.00E-21 | 61.2 |
| BAP-1 | KQ034229 | 82.35 | 34 | 120 | 153 | 867118 | 867219 | 4.00E-21 | 62 |
| BAP-1 | KQ034229 | 66.67 | 63 | 155 | 217 | 868287 | 868469 | 2.00E-15 | 78.2 |
| BAP-2 | KQ034229 | 94.12 | 17 | 2 | 18 | 866303 | 866353 | 1.00E-20 | 35.4 |
| BAP-2 | KQ034229 | 82.98 | 47 | 19 | 65 | 866431 | 866571 | 1.00E-20 | 85.9 |
| BAP-2 | KQ034229 | 81.82 | 33 | 69 | 101 | 866945 | 867043 | 4.00E-21 | 60.5 |
| BAP-2 | KQ034229 | 82.35 | 34 | 100 | 133 | 867118 | 867219 | 4.00E-21 | 62 |
| BAP-2 | KQ034229 | 66.67 | 63 | 135 | 197 | 868287 | 868469 | 1.00E-15 | 78.2 |
| BAP-3 | KQ034229 | 93.75 | 16 | 1 | 16 | 866306 | 866353 | 2.00E-19 | 33.9 |
| BAP-3 | KQ034229 | 80.85 | 47 | 17 | 63 | 866431 | 866571 | 2.00E-19 | 83.2 |
| BAP-3 | KQ034229 | 81.82 | 33 | 67 | 99 | 866945 | 867043 | 3.00E-21 | 60.8 |
| BAP-3 | KQ034229 | 76.32 | 38 | 98 | 135 | 867118 | 867231 | 3.00E-21 | 62.4 |
| BAP-3 | KQ034229 | 66.13 | 62 | 134 | 195 | 868290 | 868469 | 6.00E-15 | 76.3 |
| NP-1 | ACPB03038429 | 97.92 | 48 | 39 | 86 | 622 | 479 | 6.00E-38 | 103 |
| NP-1 | ACPB03038429 | 97.37 | 38 | 4 | 41 | 798 | 685 | 6.00E-38 | 75.9 |
| NP-2 | KQ035484 | 70.73 | 123 | 115 | 202 | 17435 | 17067 | 1.00E-50 | 157 |
| NP-2 | KQ035484 | 59.68 | 62 | 61 | 114 | 17709 | 17524 | 1.00E-50 | 64.3 |
| NP-2 | KQ035484 | 97.78 | 45 | 39 | 83 | 19645 | 19511 | 9.00E-37 | 100 |
| NP-2 | KQ035484 | 94.74 | 38 | 4 | 41 | 19821 | 19708 | 9.00E-37 | 75.1 |
| NP-3 | KQ035484 | 54.17 | 120 | 118 | 202 | 17426 | 17067 | 1.00E-36 | 116 |
| NP-3 | KQ035484 | 87.5 | 32 | 85 | 116 | 17613 | 17518 | 1.00E-36 | 58.2 |
| NP-3 | KQ035484 | 82.61 | 46 | 39 | 84 | 19645 | 19508 | 3.00E-32 | 88.2 |
| NP-3 | KQ035484 | 89.47 | 38 | 4 | 41 | 19821 | 19708 | 3.00E-32 | 71.6 |
| NP-4 | ACPB03031753 | 100 | 53 | 153 | 205 | 5925 | 5767 | 6.00E-55 | 103 |
| NP-4 | ACPB03031753 | 88.89 | 45 | 114 | 158 | 6160 | 6026 | 6.00E-55 | 83.2 |
| NP-4 | ACPB03031753 | 48.72 | 78 | 45 | 118 | 6462 | 6229 | 6.00E-55 | 69.7 |
| NP-7 | ACPB03033880 | 74.07 | 108 | 4 | 84 | 4357 | 4680 | 2.00E-43 | 158 |
| NP-7 | KQ034096 | 100 | 31 | 86 | 116 | 2070729 | 2070821 | 7.00E-50 | 64.7 |
| NP-7 | KQ034096 | 92.68 | 41 | 115 | 155 | 2070907 | 2071029 | 7.00E-50 | 80.9 |
| NP-7 | KQ034096 | 92.31 | 52 | 154 | 205 | 2071125 | 2071280 | 7.00E-50 | 94.7 |
| PAI-1 | KQ034920 | 91.67 | 36 | 101 | 136 | 58028 | 57921 | 9.00E-14 | 72 |
| PAI-1 | KQ034920 | 100 | 30 | 49 | 78 | 60193 | 60104 | 1.00E-12 | 68.9 |
| PAI-1 | KQ034920 | 100 | 33 | 4 | 36 | 60838 | 60740 | 8.00E-14 | 72 |
| PAI-2 | KQ036605 | 92.68 | 41 | 139 | 179 | 6548 | 6426 | 9.00E-17 | 80.9 |
| PAI-2 | KQ036605 | 97.22 | 36 | 108 | 143 | 7010 | 6903 | 2.00E-14 | 74.3 |
| PAI-2 | KQ036605 | 88.57 | 35 | 75 | 109 | 7470 | 7366 | 7.00E-11 | 63.9 |
| PAI-2 | KQ036605 | 100 | 44 | 3 | 46 | 9838 | 9707 | 7.00E-11 | 63.9 |
| PAI-3 | KQ035719 | 97.44 | 39 | 135 | 173 | 18019 | 17903 | 9.00E-18 | 83.6 |
| PAI-3 | KQ035719 | 94.74 | 38 | 100 | 137 | 18594 | 18481 | 9.00E-15 | 74.7 |
| PAI-3 | KQ035719 | 89.29 | 28 | 73 | 100 | 19698 | 19615 | 5.00E-08 | 55.5 |
| PAI-3 | KQ035719 | 100 | 33 | 44 | 76 | 21517 | 21419 | 3.00E-14 | 73.6 |
| PAI-4 | ACPB03037847 | 100 | 36 | 138 | 173 | 184 | 77 | 1.00E-14 | 74.3 |
| PAI-4 | ACPB03037847 | 97.14 | 35 | 105 | 139 | 644 | 540 | 3.00E-14 | 73.2 |
| PAI-4 | ACPB03037847 | 100 | 28 | 78 | 105 | 1087 | 1004 | 3.00E-09 | 58.5 |
| PAI-4 | KQ036102 | 92.5 | 40 | 3 | 42 | 3275 | 3394 | 4.00E-15 | 75.9 |
| PAI-4B | ACPB03037847 | 100 | 36 | 121 | 156 | 184 | 77 | 7.00E-15 | 74.3 |
| PAI-4B | ACPB03037847 | 97.14 | 35 | 88 | 122 | 644 | 540 | 2.00E-14 | 73.2 |
| PAI-4B | ACPB03037847 | 100 | 28 | 61 | 88 | 1087 | 1004 | 3.00E-09 | 58.2 |
| PAI-4B | KQ036102 | 92 | 25 | 1 | 25 | 3320 | 3394 | 5.00E-06 | 49.3 |
| PAI-5 | ACPB03035352 | 100 | 41 | 37 | 77 | 189 | 311 | 2.00E-19 | 89 |
| PAI-5 | ACPB03035352 | 100 | 33 | 97 | 129 | 1257 | 1355 | 1.00E-13 | 71.6 |
| PAI-5 | ACPB03035352 | 94.87 | 39 | 130 | 168 | 2336 | 2452 | 2.00E-14 | 73.9 |
| PAI-6 | KQ035263 | 92.31 | 26 | 141 | 166 | 25843 | 25766 | 9.00E-07 | 51.6 |
| PAI-6 | KQ035263 | 81.58 | 38 | 105 | 142 | 27104 | 26991 | 4.00E-11 | 64.3 |
| PAI-6 | KQ035263 | 91.67 | 24 | 80 | 103 | 28742 | 28671 | 1.00E-05 | 48.1 |
| PAI-6 | KQ035263 | 75.47 | 53 | 34 | 81 | 29570 | 29412 | 2.00E-17 | 82.4 |

| | | | | | | | | | |
|--------------|--------------|-------|-----|-----|-----|--------|--------|----------|------|
| PAI-6 | KQ035263 | 100 | 39 | 4 | 42 | 30325 | 30209 | 3.00E-18 | 84.7 |
| PAI-7 | KQ035064 | 83.33 | 36 | 102 | 137 | 33484 | 33591 | 5.00E-12 | 66.2 |
| PAI-7 | KQ035064 | 86.67 | 30 | 79 | 108 | 43167 | 43081 | 2.00E-07 | 53.1 |
| PAI-7 | KQ035064 | 77.5 | 40 | 39 | 78 | 45695 | 45576 | 1.00E-12 | 68.2 |
| PAI-7 | KQ035064 | 88.37 | 43 | 1 | 43 | 46518 | 46390 | 1.00E-17 | 82.8 |
| Pallidipin-1 | KQ034748 | 82.05 | 39 | 107 | 145 | 56228 | 56112 | 2.00E-12 | 69.3 |
| Pallidipin-1 | KQ034748 | 71.43 | 35 | 75 | 109 | 57774 | 57670 | 2.00E-07 | 54.3 |
| Pallidipin-1 | KQ034748 | 93.18 | 44 | 33 | 76 | 58989 | 58858 | 1.00E-18 | 86.7 |
| Pallidipin-1 | KQ034748 | 80.65 | 31 | 4 | 34 | 61667 | 61575 | 9.00E-08 | 55.1 |
| Pallidipin-2 | KQ034748 | 100 | 45 | 32 | 76 | 99037 | 99171 | 3.00E-22 | 99 |
| Pallidipin-2 | KQ034748 | 82.05 | 39 | 77 | 113 | 100577 | 100693 | 3.00E-10 | 63.5 |
| Pallidipin-2 | KQ034748 | 58.87 | 124 | 68 | 175 | 103052 | 103423 | 3.00E-18 | 87 |
| Pallidipin-2 | KQ034748 | 96 | 50 | 180 | 229 | 105187 | 105336 | 2.00E-23 | 101 |
| Procalin-1 | ACPB03035352 | 100 | 41 | 37 | 77 | 189 | 311 | 2.00E-20 | 88.2 |
| Procalin-2 | KQ035873 | 73.68 | 38 | 44 | 81 | 7903 | 7793 | 1.00E-09 | 60.5 |
| Procalin-2 | KQ035873 | 72.97 | 37 | 3 | 39 | 10292 | 10182 | 2.00E-08 | 57 |
| Triabin-1 | ACPB03037937 | 90 | 30 | 78 | 107 | 689 | 600 | 4.00E-09 | 58.5 |
| Triabin-1 | ACPB03037937 | 91.11 | 45 | 37 | 81 | 1700 | 1566 | 1.00E-18 | 86.3 |
| Triabin-2 | KQ035237 | 95.45 | 22 | 16 | 37 | 25658 | 25723 | 5.00E-06 | 49.7 |
| Triabin-2 | KQ035237 | 93.02 | 43 | 38 | 80 | 26504 | 26629 | 8.00E-17 | 80.9 |
| Triabin-2 | KQ035237 | 96.15 | 26 | 81 | 106 | 28778 | 28855 | 3.00E-08 | 56.2 |
| Triabin-2 | KQ035237 | 75 | 52 | 95 | 146 | 30057 | 30206 | 1.00E-14 | 74.7 |
| Triabin-3 | KQ035329 | 80.49 | 41 | 37 | 77 | 634 | 756 | 7.00E-13 | 69.3 |
| Triabin-3 | KQ035329 | 78.12 | 32 | 78 | 109 | 1662 | 1757 | 1.00E-07 | 53.9 |
| Triabin-3 | KQ035329 | 60.53 | 38 | 107 | 144 | 2745 | 2858 | 3.00E-06 | 49.7 |
| Triabin-4 | KQ034856 | 100 | 33 | 114 | 146 | 12588 | 12490 | 7.00E-13 | 68.9 |
| Triabin-4 | KQ034856 | 96.43 | 28 | 86 | 113 | 13486 | 13403 | 1.00E-09 | 59.7 |
| Triabin-4 | KQ034856 | 100 | 43 | 45 | 87 | 15044 | 14916 | 1.00E-21 | 94.4 |
| Triabin-4 | KQ034856 | 95.24 | 42 | 3 | 44 | 23991 | 23866 | 2.00E-18 | 85.5 |

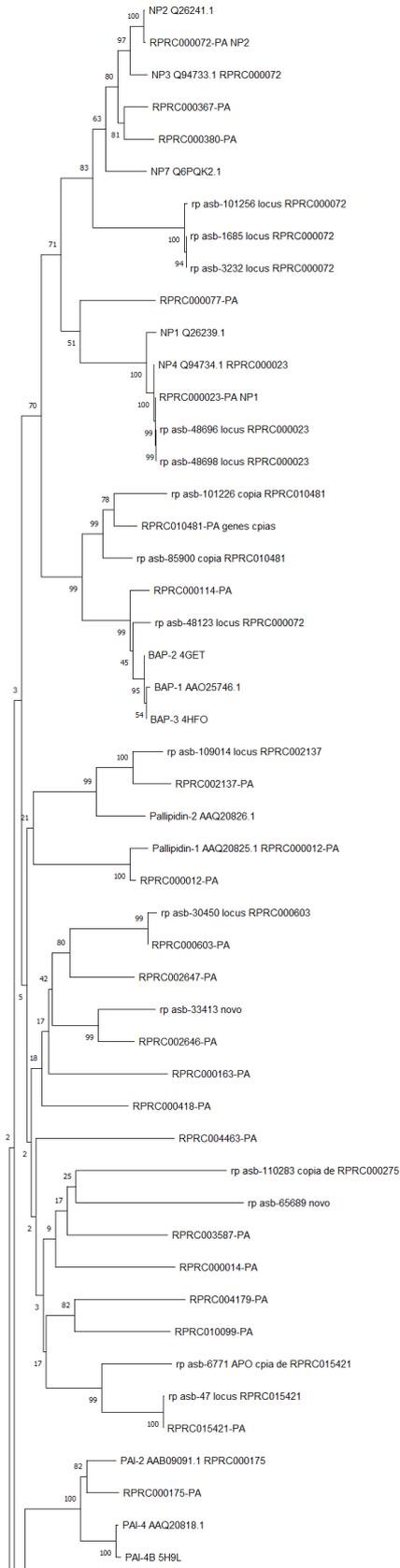
ANEXO B

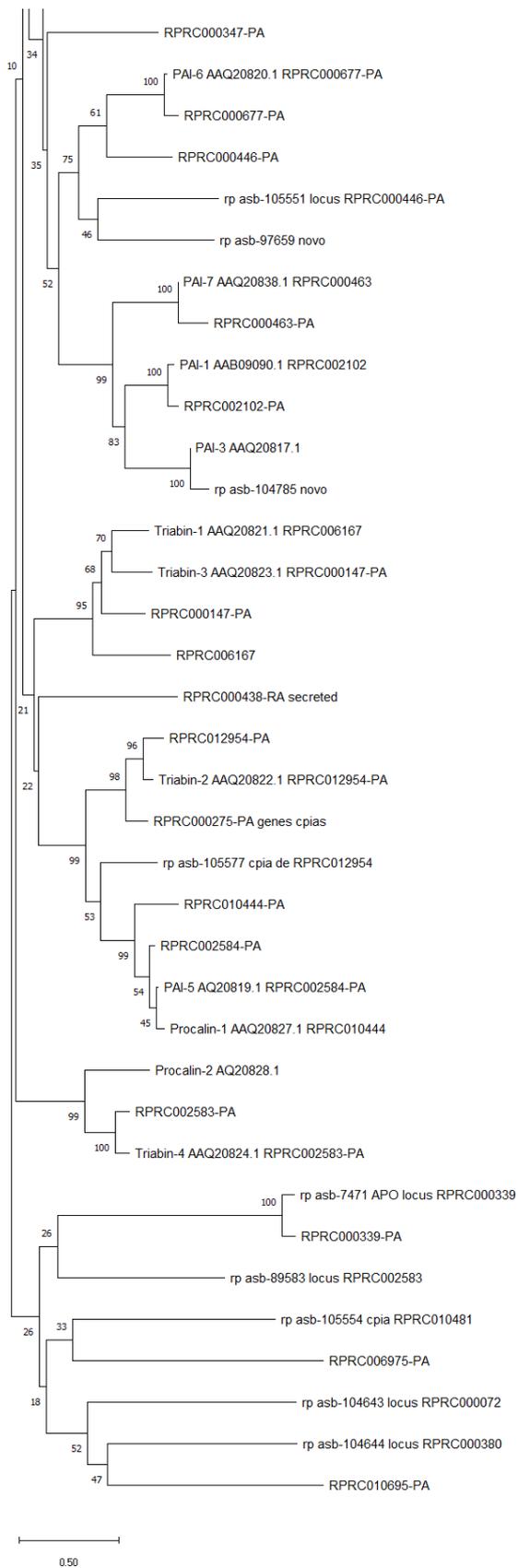
Mapeamento das Lipocalinas já descritas ao proteoma de *R. prolixus*:

| Proteína | Proteoma | ID | Algn | Qstart | Qend | Hstart | Hend | Evalue | Score |
|--------------|---------------|-------|------|--------|------|--------|------|-----------|-------|
| PAI-7 | RPRC000463-PA | 91 | 100 | 4 | 103 | 1 | 100 | 4.00E-66 | 195 |
| PAI-7 | RPRC002102-PA | 51.02 | 147 | 3 | 149 | 6 | 140 | 7.00E-45 | 143 |
| PAI-6 | RPRC000677-PA | 92.77 | 166 | 1 | 166 | 1 | 165 | 5.00E-112 | 315 |
| PAI-6 | RPRC000446-PA | 51.22 | 123 | 43 | 165 | 1 | 122 | 6.00E-42 | 135 |
| PAI-5 | RPRC013562-PA | 89.04 | 73 | 96 | 168 | 2 | 74 | 8.00E-42 | 134 |
| PAI-5 | RPRC010444-PA | 98.33 | 60 | 37 | 96 | 1 | 60 | 1.00E-37 | 122 |
| PAI-4 | RPRC000175-PA | 68.79 | 173 | 8 | 173 | 1 | 172 | 3.00E-78 | 229 |
| PAI-4 | RPRC006738-PA | 100 | 96 | 78 | 173 | 23 | 118 | 2.00E-65 | 195 |
| PAI-3 | RPRC002102-PA | 52.6 | 173 | 3 | 173 | 6 | 165 | 2.00E-51 | 161 |
| PAI-3 | RPRC000463-PA | 45.92 | 98 | 4 | 98 | 1 | 98 | 1.00E-26 | 95.9 |
| PAI-4B | RPRC000175-PA | 68.79 | 157 | 3 | 156 | 17 | 172 | 4.00E-72 | 213 |
| PAI-4B | RPRC006738-PA | 100 | 96 | 61 | 156 | 23 | 118 | 2.00E-65 | 194 |
| PAI-2 | RPRC000175-PA | 77.46 | 173 | 8 | 179 | 1 | 172 | 4.00E-87 | 252 |
| PAI-2 | RPRC000504-PA | 78.77 | 179 | 1 | 179 | 1 | 157 | 3.00E-85 | 248 |
| PAI-2 | RPRC000504-PA | 58.82 | 34 | 47 | 80 | 158 | 191 | 7.00E-09 | 51.6 |
| PAI-1 | RPRC002102-PA | 85.96 | 171 | 3 | 173 | 6 | 165 | 4.00E-92 | 264 |
| PAI-1 | RPRC000463-PA | 53.54 | 99 | 4 | 101 | 1 | 99 | 4.00E-26 | 94.7 |
| Procalin-1 | RPRC010444-PA | 98.33 | 60 | 37 | 96 | 1 | 60 | 4.00E-38 | 121 |
| Procalin-1 | RPRC002584-PA | 93.44 | 61 | 16 | 76 | 20 | 80 | 4.00E-38 | 121 |
| Procalin-2 | RPRC000299-PA | 56.52 | 184 | 1 | 178 | 1 | 178 | 3.00E-64 | 195 |
| Procalin-2 | RPRC002583-PA | 62.14 | 140 | 44 | 183 | 1 | 139 | 2.00E-57 | 176 |
| Pallidipin-2 | RPRC002137-PA | 57.01 | 221 | 15 | 222 | 41 | 260 | 6.00E-72 | 219 |
| Pallidipin-2 | RPRC002138-PA | 50 | 218 | 15 | 227 | 33 | 244 | 1.00E-65 | 202 |
| Pallidipin-1 | RPRC000012-PA | 85.52 | 145 | 1 | 145 | 1 | 145 | 1.00E-90 | 260 |
| Pallidipin-1 | RPRC000418-PA | 32.96 | 179 | 1 | 169 | 1 | 173 | 1.00E-13 | 65.1 |
| Triabin-4 | RPRC002583-PA | 100 | 102 | 45 | 146 | 1 | 102 | 1.00E-73 | 216 |
| Triabin-4 | RPRC000299-PA | 59.18 | 147 | 1 | 145 | 1 | 141 | 7.00E-53 | 165 |
| Triabin-3 | RPRC000147-PA | 74.78 | 115 | 35 | 149 | 1 | 115 | 3.00E-60 | 182 |
| Triabin-3 | RPRC006167-PA | 78.95 | 76 | 35 | 109 | 1 | 76 | 1.00E-36 | 120 |
| Triabin-2 | RPRC000275-PA | 80.13 | 156 | 1 | 156 | 1 | 155 | 5.00E-80 | 234 |
| Triabin-2 | RPRC012954-PA | 92.94 | 85 | 16 | 100 | 14 | 96 | 2.00E-52 | 161 |
| Triabin-1 | RPRC000147-PA | 77.03 | 148 | 35 | 180 | 1 | 148 | 1.00E-80 | 235 |
| Triabin-1 | RPRC006167-PA | 94.52 | 73 | 35 | 107 | 1 | 73 | 4.00E-46 | 145 |
| BAP-1 | RPRC000114-PA | 79.26 | 217 | 1 | 217 | 1 | 215 | 7.00E-115 | 325 |
| BAP-1 | RPRC000313-PA | 71.21 | 198 | 20 | 217 | 1 | 194 | 9.00E-97 | 278 |
| BAP-2 | RPRC000114-PA | 78.06 | 196 | 2 | 197 | 22 | 215 | 2.00E-107 | 306 |
| BAP-2 | RPRC000313-PA | 70.92 | 196 | 2 | 197 | 3 | 194 | 4.00E-95 | 273 |
| BAP-3 | RPRC000114-PA | 76.92 | 195 | 1 | 195 | 23 | 215 | 8.00E-105 | 299 |
| BAP-3 | RPRC000313-PA | 69.74 | 195 | 1 | 195 | 4 | 194 | 1.00E-92 | 267 |
| NP-1 | RPRC000023-PA | 89.22 | 204 | 4 | 207 | 1 | 202 | 4.00E-131 | 365 |
| NP-1 | RPRC000072-PA | 49.25 | 199 | 1 | 198 | 1 | 196 | 3.00E-59 | 183 |
| NP-2 | RPRC000072-PA | 98.51 | 202 | 1 | 202 | 1 | 201 | 4.00E-146 | 403 |
| NP-2 | RPRC000367-PA | 71.64 | 201 | 1 | 201 | 1 | 201 | 4.00E-106 | 302 |
| NP-3 | RPRC000072-PA | 80.2 | 202 | 1 | 202 | 1 | 201 | 4.00E-118 | 332 |
| NP-3 | RPRC000367-PA | 67.16 | 201 | 1 | 201 | 1 | 201 | 3.00E-99 | 285 |
| NP-4 | RPRC000023-PA | 99.5 | 202 | 4 | 205 | 1 | 202 | 2.00E-147 | 406 |
| NP-4 | RPRC000072-PA | 48.74 | 199 | 1 | 196 | 1 | 196 | 1.00E-56 | 176 |
| NP-7 | RPRC000072-PA | 60.7 | 201 | 1 | 201 | 1 | 200 | 2.00E-88 | 257 |
| NP-7 | RPRC000367-PA | 57.21 | 201 | 1 | 201 | 1 | 201 | 2.00E-83 | 244 |

ANEXO C

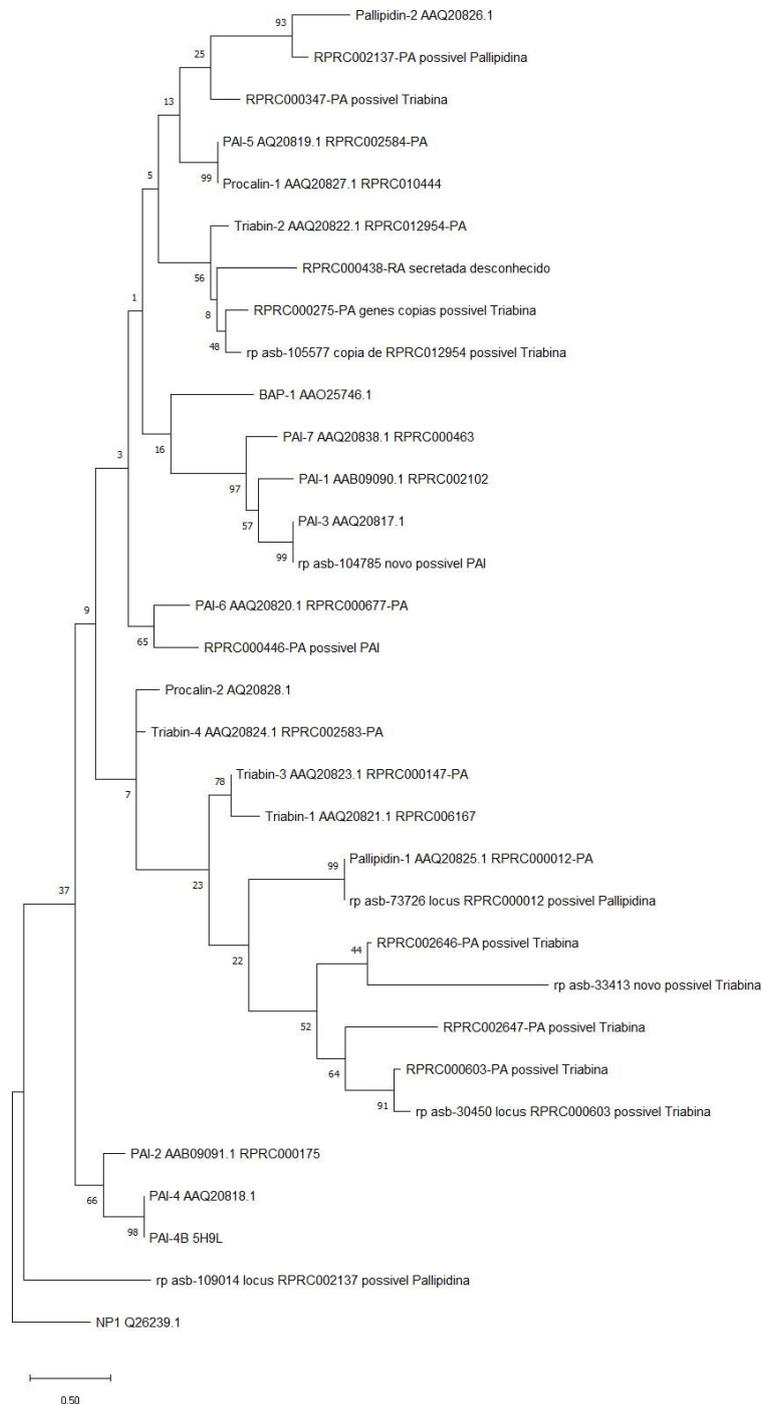
Relação evolutiva das Lipocalinas de *R. prolixus*





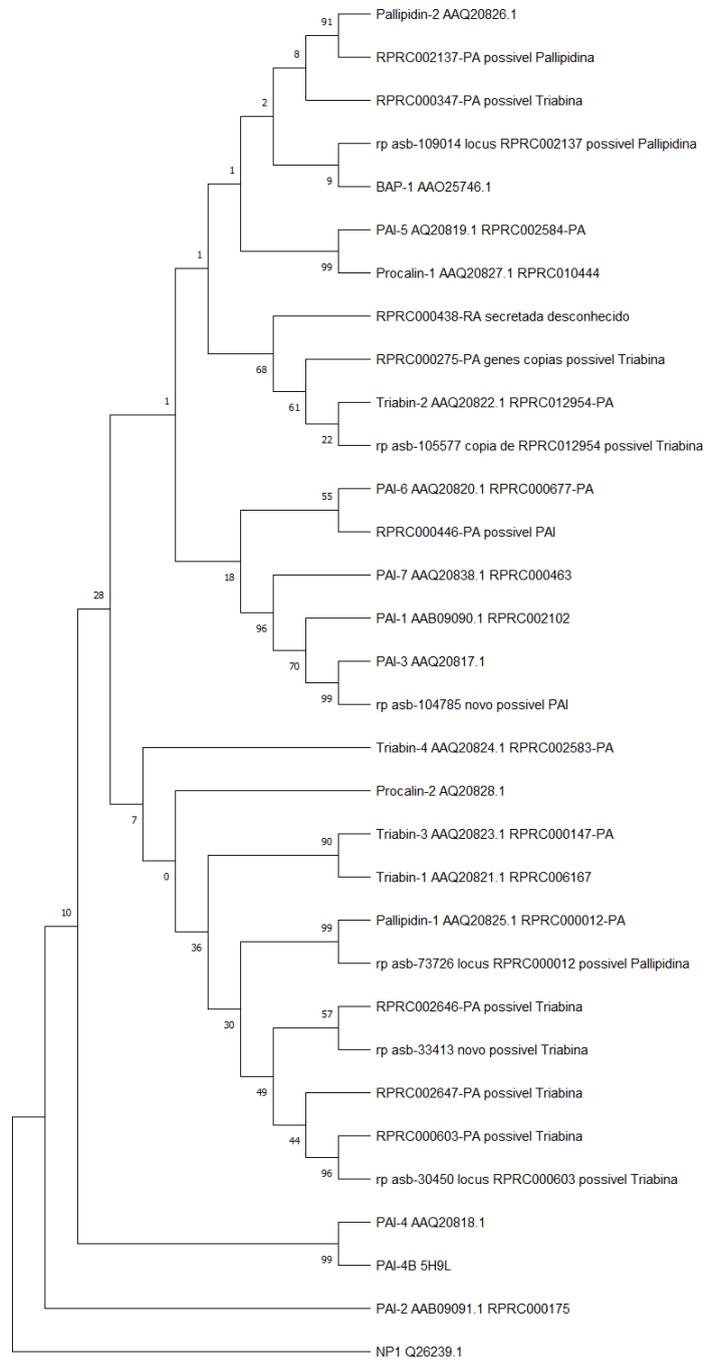
A história evolucionaria das lipocalinas foi inferida usando o método de Neighbor-Joining, e o teste de bootstrap a partir de 1000 réplicas. Os números de acesso das sequencias estão demonstrados em cada ramo da árvore filogenética.

ANEXO D



Filogenia das lipocalinas inibidoras de agregação plaquetária de *Rhodnius prolixus* por Máximo-Verossimilhança

ANEXO E



Filogenia das lipocalinas inibidoras de agregação plaquetária de *Rhodnius prolixus* por Máximo-Parsimônia.

ANEXO F

Análise da conservação do domínio e sítios ativos das Apolipoproteínas

| | | |
|---------------|---|-----|
| rp_asb-6771 | -----MNSKLAYIFLIVCGLSFLQAQPNDRKYDLRSYLPWKIGACLQPS-VVEN | 49 |
| RPRC015421-PA | -----MISKITTFIFLSVSALSLSQAFLFGFGPFTPGDYKPWRLGTCLVPT-PVKN | 49 |
| rp_asb-47 | -----RTHRKMISKITTFIFLSVSALSLSQAFLFGFGPFTPGDYKPWRLGTCLVPT-PVKN | 54 |
| RPRC003587-PA | ----- | 0 |
| RPRC000014-PA | -----MAMAKLVLTATLAY----TCAAKNVGTHPIIPFKQGCLPESIAMQK | 44 |
| RPRC004463-PA | -----MVREVLLLAVLSSAI IAILGHTY-----HLGECPAVE-PMNG | 36 |
| RPRC004179-PA | -----MKELILI-ICTLTI AISRAQMP-----GFGWCPDKR-AMPG | 34 |
| APO_Cimex | -----MISSVQLV-VFLLVATSAARP-----AFGRCPVE-PKEN | 35 |
| RPRC010099-PA | MKLNWAATVSWGYQLLV-IAICFFSLSSGQRP-----AFGRCPDVT-PKKD | 45 |
| | | |
| rp_asb-6771 | FKPKEFFKGHWYEMSSFGNFWLQVDGICPTIEYNVTDAGEILELDYHFESSLGKYVTIRG | 109 |
| RPRC015421-PA | FQPEKFFKGHWYVQNHCGNFEWILPGICPTIEYNVTQDGDILELDYHIIPMTGRYKEKRA | 109 |
| rp_asb-47 | FQPEKFFKGHWYVQNHCGNFEWILPGICPTIEYNVTQDGDILELDYHIIPITGRYKEKRA | 114 |
| RPRC003587-PA | -----MVNMCFKWMAVQL----- | 14 |
| RPRC000014-PA | FDPEKFFTGTYERSFESDITQEDGNCSGMKARIVRD-SVEILMYHYDSALNKYTSVIG | 103 |
| RPRC004463-PA | FVMSKF-LGVWYAIQKT----STGSRCLTYNFTVGAEPGEYLLQVSEHPVLGVASVDN | 90 |
| RPRC004179-PA | FDIDKY-LGTWYEAERYF-TVLEAGSRCAARTNYTKAVDG-RILVTNETNRLTGKRVLD | 91 |
| APO_Cimex | FDVEKF-VGNWVEIERSF-YLFESSLSCTKVNFTLYENS-TLKADVNYRVPWRGTSSVSS | 92 |
| RPRC010099-PA | FDVSQF-TGEWNEIERSF-YLFESSLSCTKLNFTLYSND-TMADVNYRAPWRGTSSSTSE | 102 |
| | | |
| rp_asb-6771 | Q-SYTFIKNDRAFLPFT--YK---VLGGLLTFDLPTYILGTDYETWAVVYTCKPML-F- | 161 |
| RPRC015421-PA | H-SNTKLIKNGEGHLPMT--FE---FLRGLHPMDVPTFILGTDYENWAVVYCKQQQ-V- | 161 |
| rp_asb-47 | H-SNTKLIKNGEGHLPMT--FE---FLRGLHPMDVPTFILGTDYENWAVVYCKQQQ-V- | 166 |
| RPRC003587-PA | --LNQIFKKTISIYYSI----N---MHCLANVKLPLRVLATDYDNYAIFYSCQEA-F- | 63 |
| RPRC000014-PA | E-ANLVEAKTA--RFSI---VL---TYDPILIQEFPYVVDTDYENYAIIFYSCRVPV-GS | 153 |
| RPRC004463-PA | KYHYTGELKAS-SDIPAKMSVK---FPLSLAGKASFTVFMTDYDITYAGVYTCQKLP-A- | 143 |
| RPRC004179-PA | G-EIRNVPK---GGVDSKISVK---YSTLPFPMTEYIILDTDYDSYAVVWVSCSGFGL- | 143 |
| APO_Cimex | Y-FVKDATKTP-GELNMVLASTLPSLIARMAPGSGKYIVLDTDYDFALIYSCTDLR-F- | 148 |
| RPRC010099-PA | Y-IKDVSKTP-GILNMVLASTLPALIARMAPGSGKYIVLDTDYIDYALIYSCTDLR-L- | 158 |
| | * : * : * : * : * | |
| | | |
| rp_asb-6771 | --LKSEMSWILTRQRNS-TDQLKPVLETLKKSHLSGEYKPAVNVGCGPNEPHLTSITIT | 218 |
| RPRC015421-PA | --LKFEESLILTRKRDD-EEIWQPVLDLTKLNKLDYRYFKPTVNNGCAPNEPHL----- | 212 |
| rp_asb-47 | --LKFEESLILTRKRDD-EEIWQPVLDLTKLNKLDYRYFKPTVNTGCAPNEPHL----- | 217 |
| RPRC003587-PA | --FKAEMAWIATREKGD-HSHDNKIKQVTKAGFNVDHFIPOKSTDCGPDEPRL----- | 114 |
| RPRC000014-PA | SSLNDEMAWTFTRKGD-ASNDGVDRIKLSGLDPAKFARYEHMTCVDEPTL----- | 206 |
| RPRC004463-PA | --SNRRSATILSRKSLDKQVVDKIRSRLSLFGVSPYDLSIIDQSKCPGANDTNV---- | 197 |
| RPRC004179-PA | --VNTQTAWVMTRELRPPGTVLQKAYAVLDKNKISRTFFVRAEQEECNLGEALSEKSRAS | 201 |
| APO_Cimex | --VHADFIWVLRGSDISVEARTVVYSTLDKNKINRDRLLLTNNKDC----- | 193 |
| RPRC010099-PA | --VHADFIWVLRGSKDISIDARTIYSTLDKLNKINRDRLLLSKTKDC----- | 203 |
| | : * : * | |
| | | |
| rp_asb-6771 | VF QRSLV----- | 225 |
| RPRC015421-PA | ----- | 212 |
| rp_asb-47 | ----- | 217 |
| RPRC003587-PA | ----- | 114 |
| RPRC000014-PA | ----- | 206 |
| RPRC004463-PA | -----DININPETFSSHNAEAVRK-----AGEKIGDGE-----YVAEGVGKVVN | 238 |
| RPRC004179-PA | DAANQNVESATTSASAVNSSFHHPSPVAPVETSSTGNEVYMPQKKEGVSVSSPEAVGVNSE | 261 |
| APO_Cimex | ----- | 193 |
| RPRC010099-PA | ----- | 203 |
| | | |
| rp_asb-6771 | ----- | 225 |
| RPRC015421-PA | ----- | 212 |
| rp_asb-47 | ----- | 217 |
| RPRC003587-PA | ----- | 114 |
| RPRC000014-PA | ----- | 206 |
| RPRC004463-PA | -----KVHK---DSEVDDL----- | 250 |
| RPRC004179-PA | SSNNAEETVRPDSNDSSRKPAVAVTAGPPALSSQHAVVAGSNSVSEQTKPTVSTSEAME | 321 |
| APO_Cimex | ----- | 193 |
| RPRC010099-PA | ----- | 203 |
| | | |
| rp_asb-6771 | --- | 225 |
| RPRC015421-PA | --- | 212 |
| rp_asb-47 | --- | 217 |
| RPRC003587-PA | --- | 114 |
| RPRC000014-PA | --- | 206 |
| RPRC004463-PA | --- | 250 |
| RPRC004179-PA | LKN | 324 |
| APO_Cimex | --- | 193 |
| RPRC010099-PA | --- | 203 |

```

RPRC000014-PA -----MAMAKLVLTATLAY----TCAAKNVGTHPIIPFKQGCPCLPESIAMQK      44
rp_asb-6771 -----MNSKLAYIFLIVCGLSFLQAQPNDRYKDYDLRSYLPWKIGACLQPS-VVEN      49
RPRC015421-PA -----MISKITFIFLSVSALSLSQAFLFGFGPFTPGDYKWPRLGTCLVPT-PVKN      49
rp_asb-47 -----RTHRKMISKITFIFLSVSALSLSQAFLFGFGPFTPGDYKWPRLGTCLVPT-PVKN      54
RPRC004463-PA -----MVREVLALLAVLSSAIIAALGHTY-----HLGECPAVE-PMNG      36
RPRC004179-PA -----MKELILI-ICTLTIATSRAQMP-----GF'GWCPDKR-AMPG      34
APO_Cimex -----MISSVQLV-VFLLVATSAAQRP-----AFGRCPVE-PKEN      35
RPRC010099-PA MKNLWAAATVSWGQQLLV-IAICFFSLSSGQRP-----AFGRCPDVT-PKKD      45
: : : . * *

RPRC000014-PA FDPEKFFTGTWYERSSFESDITQEDGNCSCGMKARIVRDS-VEILMYHYDSAL-----NK      97
rp_asb-6771 FKPKEFFKGHWYEMSSFGNFWLQVDGICPTIEYNVTDAGEILELDYHFESSL-----GK      103
RPRC015421-PA FQPEKFFKGHWYVQNHCGNFEWILPGICPTTEYNVTQDGDILELDYHYIPMT-----GR      103
rp_asb-47 FQPEKFFKGHWYVQNHCGNFEWILPGICPTTEYNVTQDGDILELDYHYIPIT-----GR      108
RPRC004463-PA FVMSKF-LGVWYAIQKT-----STGSRCLTYNFVGAEPGEYLLQVSEHPVLGVASVDN      90
RPRC004179-PA FDIDKY-LGTWYEAERYF-TVLEAGSRCARTNYTKAVDG-RILVTNEITNRLTGIKRVLD      91
APO_Cimex FDVEKF-VGNWVEIERSF-YLFESSLSCTKVNFTLYENS-TLKADVNYRVFWRGTSSVSS      92
RPRC010099-PA FDVSQF-TGEWNEIERSF-YLFESSLSCTKLNFTLYNSD-TMADVNYRPAWRGTSSSTSE      102
* .: : * * . * :

RPRC000014-PA YTSVIGEANLVEAKTARFISIVLTYD-----PILIQEFPYVVDTDYENYAIIFYSCR      149
rp_asb-6771 YVTIRGQSYTQFIKNDRAFLPFTYK-----VLGGLTFDLPYILGTDYETWAVVYTCKP      158
RPRC015421-PA YKEKRAHSNTKLIKNGEGHLPMTFE-----FLRGLHPMDVPTFILGTDYENWAVVYCKQ      158
rp_asb-47 YKEKRAHSNTKLIKNGEGHLPMTFE-----FLRGLHPMDVPTFILGTDYENWAVVYCKQ      163
RPRC004463-PA KYHYTGELKAS-----SDIPAKMSVK-----FPLSLAGKASFTVMTDYDYAGVYTCQK      140
RPRC004179-PA G-EIRNVPK-----GGVDSKISVK---YSTLFPFMEYIILLDTDYDSYAVVWSCSG      139
APO_Cimex Y-FVKDATKTP-----GELNMVLASTLPSLIARMAPGSGKYIVLDTDYDFALIYSCTD      145
RPRC010099-PA Y-IIKDVSKTP-----GILNMVLASTLPALIARMAPGSGKYIVLDTDYIDYALIYSCTD      155
: : . :. * * : * : *

RPRC000014-PA VPGSSSLNDEMWTFRKRGDASNDGSDVDRILKLSGLDPAKFARYEHMTCPVDEPTL--      206
rp_asb-6771 ML---FLKSEMSWILTRQRNST-DQLKPVLETLKKSLSYGEYKPAVNVGCGPNEPHLTS      214
RPRC015421-PA QQ---VLKFEESLILTRKRDE-EIWQPVLDTLKLNKLDYRYFKPTVNNGCAPNEPHL--      212
rp_asb-47 QQ---VLKFEESLILTRKRDE-EIWQPVLDTLKLNKLDYRYFKPTVNTGCAPNEPHL--      217
RPRC004463-PA LP---ASNRRSATILSRKSLDKQVVDKIRSRLSLFGVSPYDLSIIDQSKCPGANDDTNV      197
RPRC004179-PA FG--PLVNTQTAWVMTRELERLPPGTVLQKAYAVLDKNI SRTFFVRAEQEENLGEALSEK      197
APO_Cimex LR---FVHADFIWVLRGSDISVEARTVVYSTLDKKNINRDRLLLTNNKDC-----      193
RPRC010099-PA LR---LVHADFIWVLRGSKDISIDARTIISTLDKKNINRDRLLLSKTKDC-----      203
: : * * . :. *

RPRC000014-PA -----      206
rp_asb-6771 LTIITVF-----QPSLV-----      225
RPRC015421-PA -----      212
rp_asb-47 -----      217
RPRC004463-PA -----DININPETFSSHNI AEAVRK-----AGEKIGDVE-----YVAEGVG      234
RPRC004179-PA SRASDAANQNVESATTSASAVNSSFHHPSVAPVETSSTGNEVYMPQKKEGVSVPVAVG      257
APO_Cimex -----      193
RPRC010099-PA -----      203

RPRC000014-PA -----      206
rp_asb-6771 -----      225
RPRC015421-PA -----      212
rp_asb-47 -----      217
RPRC004463-PA KVVN-----KVHK---DSEVDDL-----      250
RPRC004179-PA VNSESSNAEETVRPDSNDSSRKPAPAVTAGPPALSSQHAVVAGSNSVSEQTKPTVSTS      317
APO_Cimex -----      193
RPRC010099-PA -----      203

RPRC000014-PA -----      206
rp_asb-6771 -----      225
RPRC015421-PA -----      212
rp_asb-47 -----      217
RPRC004463-PA -----      250
RPRC004179-PA EAMELKN      324
APO_Cimex -----      193
RPRC010099-PA -----      203

```

ANEXO G

Resultados da anotação manual dos transcritos realizada no GeneWise de acordo com metodologia descrita no item 3.4.

a) Transcritos inicialmente identificados como possíveis PAI:

```
>rp_asb-15399
MKSLSLTQLLAEQF*VYVDEVNVRLYVEVYLQFTPQWKISLLPLGKDAGLVHI*TNA*LL
LSVLTTLKSPVILLSETQORSKYQICNVIKVNSISFQFVHSVNFVGGLCWEREPR*IEI
RHLENFSPKKNKNSLTYNYSWXVI*NTEK*VLPNL*CY*SKFYFISVSTFS*FSRRTL
LGERTAIN*DSPSRKIFTLRKTKNNSLTYNYSWQCVNREIQEVLLFIIIFLIILNKEYKY
VKYMMKMEPVTQNTSDIKIISMLTHSKIVPLFLLTFFCFTITFLASCTFDLQSTTMKTE
FMLIFTQTINYKIFF
```

```
>rp_asb-44356
MSQVPTSCIKLEKKYQSKKLGLEHNFA*D*SQFTLC*YSINEIFCIIITVVQCSYIRTNKI
KIFSTFMILKCKNQPLPLNFNLICRGRSDILFSMFNGFYLLKKN SNILLCILASFRFSF
WITKFKVRI SFFVYI STMFWSRNMCFRCSTKCTSQNTKKCQYDYPVTILFILVNFNFEF
CIFQKKEVPYSKLLPTPFFTKSILLNRMKRELSKLNLLFSCSQQTITLNNFEIILKIY
LTIK
```

```
>rp_asb-44609
M*FVRVMSYQILYIVMCDFDIFRKSRTNFLYKTGKIKYHDVRLIKIKYNQHSRVFTNND
FSIRVELLLNCGTVNFS SAHFLSSYGCRRWKIYRAVLEKSIKII CRNFVLELHVQVSDSD
KRHASRSVTLRHRVLSFKGSFKSIGLYNRRNKRKYLQPNNFIFNDLFCY
```

```
>rp_asb-97659
MKTIIAFAFFGLLGRFSAADANTALDSSECQDAKSIAMKDFKPEQFFTGTWYLTHLKYR
NHSSICPKFDIKPDPLTIDYEYGVLGNTQIKKKLAGSFRAGWYRLGRPFFLIQVEITVE
KTGSSLFPNASCKSSGKCRNYGLGSPRKGDLNSLNSFQQIHSNENLCS*NIFFSFNFFIYN
LYIYFFLQRFDISGLQSSDLRLSSDSFNFTFNGVFTLSGISVVDCLRNINIFSEKRK
```

```
>rp_asb-104785
MYSKQFCRIVALFMFGFLAEAQYGGPPKMSRGCKDIYNRGVDNLNYKQFFTGWFLTHGE
RVSSKCDTVSVNGDKITFKLRGAQINCQLENVPDAKYTKFNCKKSVSKTFSTEISVLATD
NNNYALVYRCGVLEDDNYKDNI VVMQRQKQAPFPFALESEVKGKFGHGLNERTFPKFRFLN
SISVRYVEKTKYLPN
```

```
>rp_asb-105551
MKAIIILVSFFGILGCTFCKHVRTGPEGCRDVYDQADPKYKRRDFYTGSWYLTHAKYRNHS
TLCTKFDMTFNPLQIKYDANETS A IETILPEL PKLFQIYIQGVQHKT TVTSGGNSGHHL
QQKIPYKLWLD SAPFCFFKVL LLLFSKQLSTFLKLL
```

b) Transcritos inicialmente identificados como possíveis Apolipoproteínas:

```
>rp_asb-6771
MNSKLAYIFLIVCGLSFLQAQPNDRKYDLRSYLPWKIGACLQPSVVENFKPKKEFFKGHW
YEMSSFGNFWLQVDGICPTIEYNVTDAGEILELDYHFESSLGKYVTIRGQSYTQFIKNDR
AFLPFTYKVLGGLLTFDLPTYILGTDYETWAVVYTCKPMLFLKSEMSWILTRQRNSTDQL
KPVLETLLKSHLSYGEYKPAVNVGCGPNEPHLTSLTITVFQRSLV
```

```
>rp_asb-7471
MSNKDSQRKYVPIRKEENCRFWFLSPLL PVASQPNCNTNCGFPKIHPSDTECRYRPLLS
LGLYLLQILRRIFLPASGCGFAVVHAKENEKEDKCDPERKLVACDLPVYAEKKCLDC
TPPKNEPDTLFTETVKARKEMWYVMDQARIVKQQIEDGLEDLKSSSEEYLYLREESN
VAPRIGAI AVGGLSGYILALRRGIFRKLIIYTSIGAGGMAALCYPREAKEYSHEALQLLKR
```

YIIIGYHFINGVTRDLTGYQLPPLPNGHLVISDSAHPGSEAVPSSAHILPPTATPHTIPL
PHEDPHEVISFSKQHQSVGRELTKTSYLRISTQLSLLRSTLFFLEAD

>rp_asb-105382

MLE*LRYNK*KLLFKIDI*DHKKFNVTVTEMLLLLILSIWVDTSSSQVPYLGSCPDLKTV
DNFDQTRYMGKWYEAERYFSLFEFGGKCINSNYTDGIDSINIISKQTSTLTGIHSTIEGE
VQKTPNGFLSKMNLKYPYLPPIQIKIEFAVISSNH

c) Transcritos inicialmente identificados como possíveis Nitroforinas:

>rp_asb-48696

MYLRVEFKLYYFFIVP*SLKKIPELQFKTINVCSRYTSLLAVAILCLFAGVNGACTKNAI
AQTGFNKDKYFNGDVWYVTDYLDLEPDDVPKRYCAALAAGTASGKLKEALYHYDPKTQDT
FYDVSELQVESLQVSLGKYTANFKKVDKNGNVKVAVTAGNYTFTVMYADDSSALIHTCLHKGN
KDLGDLIYAVLNRNKDAAAGDKVKSASVSAATLEFSKFI STKENNCAYDNDSLKSLLT

>rp_asb-48697

MYLRVEFKLYYFFIVP*SLKKIPELQFKTINVCSRYTSLLAVAILCLFAGVNGACTKNAI
AQTGFNKDKYFNGDVWYVTDYLDLEPDDVPKRYCAALAAGTASGKLKEALYHYDPKTQDT
FYDVSELQVESLQVSLGKYTANFKKVDKNGNVKVAVTAGNYTFTVMYADDSSALIHTCLHKGN
KDLGDLIYAVLNRNKDAAAGDKVKSASVSAATLEFSKFI STKENNCAYDNDSLKSLLT

>rp_asb-48698

MYLRVEFKLYYFFIVP*SLKKIPELQFKTINVCSRYTSLLAVAILCLFAGVNGACTKNAI
AQTGFNKDKYFNGDVWYVTDYLDLEPDDVPKRYCAALAAGTASGKLKEALYHYDPKTQDT
FYDVSELQVESLQVSLGKYTANFKKVDKNGNVKVAVTAGNYTFTVMYADDSSALIHTCLHKGN
KDLGDLIYAVLNRNKDAAAGDKVKSASVSAATLEFSKFI STKENNCAYDNDSLKSLLT

>rp_asb-104982

MIHQNPLAGGGLALPPIGENQDS**CLIKITLSKK*VRISLRINYKHRILAPQEIIFHCF
LTNRFYCVRNLLCVLQRCNNTNIAANPSTFAFLFFKGFCTIRHYLLIRPKGITFTLHSP
GSPTIDDLGRMVTANNAVYLDGNINILKKIRVDVKKSSFLAYRTFFPYGSLHLLLVDMINR
RVW*LILIESFKTFSLTEVCILEYNYTVALISDTRLRYKQLYEREKSKSALNQRWLISLS
S

>rp_asb-48128

GRKIDERTLQVSYIDTDYSKYSVVHVCDPAAPDYCKYLSVLSQYIYSPFLLVKIIL

>rp_asb-104644

MYKNLHQFPPIIFIRVLSSNENHSVSVIFATKK*IYYRTCSQDLW*QRT*SSLN*PYSKLF
*SSY*VAHFYCDLKRNRDL*I*KIKKVNENIALININIYTKINNNIVKELNTSNI FSVFP
QTYVDQSRFLGFSFYNVYFAPFDSKGPVPLKEGIITMFKLFSKSRFYNGGFFLILKNNY
EFRTKYFDLIRNKYIFQMPSTANTANYFEAGLNSMLAHTLILLKSVNKGEIYKYIYFKKNTS
VSNEYLILVNLVIQFYNNFVINL

>rp_asb-104643

MRV*RVLWLSLTEVSWLEDRVLSLLEDVVLESVISRGFVTLRLY*GRIVNATGKKTRP*GK
TRKSLPVFSKILYLLHLTFYFKAGARYYNLSGPFVWSSL*IKFIKECG*E*LVPLIIWGY
LCLLHILKPVWAYGIEMWGSTKRSNLQRIQSFQYLKFFAQ*YFLEYLICGKGCISALTKP
DTDLLLMYKNLHQFPPIIFIRVLSSNENHSVSVIFATKKTFQNKGKSLHTGVLTIIELIYG
AIQEFRTKYFDLIRNKYIFQMPSTANTANYFEAGLNSMLAHTLILLKSVNKGEIYKYIYFKKNTS
VDESKRGSSSPRDTLQKLVSSFSSENYPLAFQSTWHLHSNLYL

>rp_asb-104455

MKTCVALVLLSAILRLSTAGPRPSRDRKCRNPDPKENFEPEKFFKNKWTTHYQLTKYGL
EGNTMNCIAFSSIVFSNGRIKEVFSNYQPKGNAYFTVSSLVSGETYYGINSNGKH*KSHIL
KNLFYDKITTFALQVDFIYKNIIVMLVMDTI PHKMSAICSQNEGLMNVDAVLLEVLSSIH
SWIKGEIALQ*LSKTTFTITHSTLRTLPLRGLPLLLRRVVVSLF

>rp_asb-85900
MNYE*VTIVL*FL*HF*PL*NSSKQHRQPARYK*SFTIFADFFLIVHPLTVNIIKEIIVT
FFFFSHCAYAYDVSYSSLDKDRVGGKYTAQNVVDRKGRLEEAHEYNITIIDTDYTR
YAFVHLCGQDKEGLYFDYFFINGRNOEDKVHPKITEAMKQVIYSLKELLCSNCFNQKDKI
KDI IQQCRPKPYLLQFSLSVGTMRFYVP

>rp_asb-101226
MSCYTNTIFFSMFYFFSEGVAVSSLTFSGKGFANGTGHNDISTYKNFGHHFERIHSVSG
IKSNNSPLSLIFFLILPIEKKKDYSSYISLSNNTSFYETYGKYTAKHIQVKNKDGTKIEER
SLSVTYLDTDYSSYSVVYVCERMKDGKLFSLYAVQSREKTLESNIKTKVNSVLQSINLNL
DKLSSIKDGVCKFDDSTLNTLLSQTKKEFKTQEKQCGFMKIGCF

>rp_asb-90144
LRRYVALFLAVVALSAAAPKSKNNKCPDLKPFKKFDMNQYNGTWYMTHAKIDGIHIPGK
NMFVCGKFSRSLGNGTIREFVTHYLVNSNLFFYVENHIDLKSSTGKGEYIVNSRISSKDG
TPIEDFFPIKYTNLGTDYKTYSVFNMCHEELDNDNVTSFYFIASRTKDARKVPPILKKVLA
KNGLKLMDFSTTEKIKCKP*QKSSLKIVDIFKLYQYNKE

>rp_asb-89581
MNANC*PNLKNFSFKVRYKFEPSCK*KFNLSLIF*FDATNF*I**FFRGDWFVTHANDGT
VSTLCRKYSISENSDDKLIVQYGTRGNK

>rp_asb-89583
MCLDVQVVRHLQRFHGLGNAYTTQAVFFLKKILISLEGKTL SRLTIFILSKDLTNFFSG
DWFVTHAKDETEAILCRKFTTSVNSGGKIEFYGSIREGNGLYSIILTIILILSSHLKSN
KITITRLFVLFGTQKSVMQKNQNFILRLAYTGFQCIYSGFESSYWPFRFSLPIASLLSVN
ERTLFSKGGSVLIKLTVIVEPQIFKRSRTYSPWLKKNRSENIYEKFEYIYKDKIR

>rp_asb-89582
MTWIGT*KQTKLKDGLREPFLLSFIHYFFRGDWFVTHANDGTVSTLCRKYSISENSDDKL
IVQYGTRGNK

>rp_asb-1685
MCIFYIYIYIYIYIYILFIYKYNIFIYIHAYLN*DKKSSYAISSRKVDHEESTSGLNSIS
TF*ADYSV*SAARG*YE*CDDYLNCTLSIIIFPSRYALLAVTILCLTSTMGVSGDCSTN
ISPCKGLDKAKYFSGTWYVTHYLDKDPQVTDYPCSSFTPKESGGTVKEALYHFNSKNPNI
YKDDQSKRYKSIICLNVAKFTEQPRGSFFEKMTRVERSKSVTGRHCINKRAVSTCGREGL
SDQSRQATTLQSRSETESRRRAAPESGGGQSQLEGHTQLKGVNLNELLISF

>rp_asb-3232
MCIFYIYIYIYIYIYILFIYKYNIFIYIHAYLN*DKKSSYAISSRKVDHEESTSGLNSIS
TF*ADYSV*SAARG*YE*CDDYLNCTLSIIIFPSRYALLAVTILCLTSTMGVSGDCSTN
ISPCKGLDKAKYFSGTWYVTHYLDKDPQVTDYPCSSFTPKESGGTVKEALYHFNSKNPNI
YKDDQSKRYKSIICLNVAKFTEQPRGSFFEKMTRVERSKSVTGRHCINKRAVSTCGREGL
SDQSRQATTLQSRSETESRRRAAPESGGGQSQLEGHTQLKGVNLNELLISF

>rp_asb-3235
MILVDLNVNQHITIFLQQRSSHREKYNTLKNV*WHFLSANFFFNLENLMTQLIDDSTT
HK*KKTSFYNIGEGKLGSSGVQYTAKYNTVDKKSFRHVFI RLWFGFLX

>rp_asb-101256
MCIFYIYIYIYIYIYILFIYKYNIFIYIHAYLN*DKKSSYAISSRKVDHEESTSGLNSIS
TF*ADYSV*SAARG*YE*CDDYLNCTLSIIIFPSRYALLAVTILCLTSTMGVSGDCSTN
ISPCKGLDKAKYFSGTWYVTHYLDKDPQVTDYPCSSFTPKESGGTVKEALYHFNSKNPNI
YKDDQSKRYKSIICLNVAKFTEQPRGSFFEKMTRVERSKSVTGRHCINKRAVSTCGREGL
SDQSRQATTLQSRSETESRRRAAPESGGGQSQLEGHTQLKGVNLNELLISF

>rp_asb-48741
MLS*Y*RDN**IGVTL*RKERFESFLKSNFQFMHKEYESLSMSEMPGTMWCRSAIVIRNY
YQPDLSIGSITVSKLNAMFLKRIFSPIQNILQRIVKSTKLFVSTDEHPVAPSTADSHN
ATTNKTSAAYLHKILSEPLIIALILS*LTLRQLQYGFKTKFTKGIFFLLKHI VPKIYKY

SLNFKGNSLDKLYPLRKRVRTKMLDLSISNSLASINIQIATLKGELDKVSEVKQINSEQ

>rp_asb-105554

ML*LSIKS*FYRNNWYFAAFFRTEAYSYYLSFSKKSFFDTTYGKYTAKHIQVDKVGKE
LEEHSITVTYLDTDYDSYSVVYVCGEIMENLFSLYAVQSRSQTLNQDVETKVKSALNGIN
LKLDKLSSIKDFGCKYDDSTLNALLSKSFTHET

d) Transcritos inicialmente identificados como possíveis Pallidipinas:

>rp_asb-73726

MK*NFIQIKIN**TSICLNENCLSSVFEPTYWKLILMFNCTVCCLOISLKNFDQCKYFSA
KLEYVQRVSEGPQKQTVCIKFDKREPDGKVTSNYDYYGLENQKYHISCNGTERSNNKKG
FSFNCKQTSKNSATPIFQEEEMVMEVDYKKEFSVNRVCKSGGFSSGN

>rp_asb-73730

MYKLLKQI*IKSVKY*YKFKININVIYVK*VSNFTNYIK*DNLSPCKKVSFLVTRVMHIQ
RKNKQNLNMFNQSYFKFKKCCIEEKKNLAILTTSKADFLQHIQIKVYYYYFKTDYNSL
MQLASFLKCSNLNTQKNKNSIFGCHNFMITLSPKNCYSNYHTKVIKIGAMVTEWFEARIANL
RCRVVLGSIISGLGLGFYLMFNMRTL

>rp_asb-104042

MITESFIINFFNVSTVETFKFTFKTISFNLVCGKYSQF*NVISMGIPWETSHGMGWDGTA
HIYISHGSRYLILLKFAVANFPSSMLYHCFWRCYQKPEQGFQKYTFWRSSDYKK*NGRK
RISTVYKWFQNFRRSGHMGTSDAERSGRSVEVTTPEIIDKIHDMVMDRRVNVREIASAVG
ISNERVHNIHLQHLDMTKLSARWVPRLLTLAQKRNRKNPQDFKRRFVTVDETWIHHYTPE
TKEQSKQWVTKGESAPKKGEDRFFGRKDYGDCLLGFARNNPHRLSGKG

>rp_asb-109014

MCI*SLGSKLTQRYQLKRF*GYQLRVKDVQ*YC*VLYVTY*IEEYNYLGDGSFSIFTYS
NSILFITITITTYFFNDDFLQFGYEDYLLVCSVDVGSKLIIIVQAIIVSLPTKLMVLSSCV
ISAH

e) Transcritos inicialmente identificados como possíveis Triabinas:

>rp_asb-33413

MKRNLQYPYRLLSATNLYSSLRYLVLGHTWFKVHHSKGDHTVLKQKQIKEIEVIGSDAFG
IETEATQQFYLLECTTNLKDVKDGTSVLNCKSTKDGKVNFEQLTATVLETDYEHFTVLYV
CVKKWESGSKGNFLVLRNDRKNGEPTDPKIAETLKKHGLDLSTFTSRKNVHCKDHPSTVSN

>rp_asb-30450

MKMIIVLAIIFAILTFATADSQIKPKECLKVPVKQNFEPQKYFRGNWFLFNKREEGQSNT
TVCQESKSRMLEDGTISHVIYAYSVD

>rp_asb-40554

MLVAINTWQ*FGLIYVRDVLILNNETHLNK*TSAVFPQVPPNIDFDDWIS*GSSRINN*D
LISLLNLINETIVFTYQIIIGFIFLRSFNLTNPSFKNEICIRNMDNIFYSIISFASTFKN
SLLQLIKWHNDQRFISVHCSDDPGSSLSVSNNTIFHRTMSCKKNLRVVAEFESGISACTHI
IKYYLHVRCSEFFKLFINKIKSISYPSSRLRYLISLVPSYNTVYHRTMSCKKLIRRNKVRV
VAEFELGLGFLYANHLAKAWPNIKY

>rp_asb-65689

KKIVNKLFFVLFCLTVGTHAPEAYGESAMSKTRIYEWYKRFQDGRENVEDDERPGRPSTS
TTDENVEKVKEMVMNDRRITIGEVADDVAISIGSCHEIFSDVLGMNVWQQNLFQNC

>rp_asb-102620

MLVAINTWQ*FGLIYVRDVLILNNETHLNK*TSAVFPQVPPNIDFDDWIS*GSSRINN*D

LISLLNLINEIVFTYQIIIGFIFLRSEFNSLTNPRIFSVHCSDDPGSSLVSNNTIFHRTMS
CKKNLRVVAEFESGISAFIYNNQTIKYRLKRKTKYVILHCSVDPGSSLVSDNTVYHRTMS
CKKLIRRKNTIFNKFGAILEDSENEIILSLKTLLPIFNCDISKIKLKKKVSDIDINHVYNH
YCKLIIDSVCDTIKENKYRRRRKASFGKFRP

>rp_asb-102716

QHKYTSITQLYFSTLKLHKILTGRELYHTFTMVKSRAINGIFFSGTWVTHAKNGSSTIL
CREITLTKNGDVTVKSZYLIKKTSEYTVHCTGKEKKGKVPFKCTREEADVSGPSK*LFT
LEIVSFDNRSWRLFVQSSKNQRRNAQSANSYVSLSTISLDLGHFCAFVSFAIQSTSQEAL
QLDRPEYFHNVDLFLFCSAANQ

>rp_asb-102750

MGTSDAERSGRPVEVTTPEIIDKIHDMVMDRRVKVREIASAVGISNERSLCFGFFFGFT
SNS*GNKHYKNYLLFYRKTKKDYTVNCTGTENNGKGDGSTNNYNIYSYSN

>rp_asb-110283

YLLTITGTLLVGLRSSFYLPSSINSLILENEFYPSKDFLGKLALVELTSLSNKEGCFQ
NNDEVLSRKSVDREDCLVFYDCLSCVISLNSIVFF